



CONTENTS

Change of needs for protein expression in the postgenomic era

Cloning & optimization of expression constructs

Parallelization & automation of bacterial expression systems

Novel baculoviral expression technologies for insect cells

Novel transient expression technologies for insect cells

Viral expression technologies for mammalian cells

Novel transient expression technologies for mammalian cells

Cell-free protein synthesis (bacterial & eukaryotic)

Databases for project tracking in protein expression

Expert commentary

Five-year view

Key issues

References

Affiliations

[†] Author for correspondence
Novartis Institutes of BioMedical Research, Protein Expression & Purification, Discovery Technologies/Lead Discovery Center, CH-4002 Basel, Switzerland
Tel.: +41 6132 43426
Fax: +41 6132 45155
michael.forstner@novartis.com

KEYWORDS:
automation, baculovirus, cell-free transcription, cell-free translation, *Escherichia coli*, mammalian, recombinant protein expression

Optimization of protein expression systems for modern drug discovery

Michael Forstner[†], Lukas Leder and Lorenz M Mayr

The expression of high levels of stable and functional proteins remains a bottleneck in many scientific endeavors, including the determination of structures in a high-throughput fashion or the screening for novel active compounds in modern drug discovery. Recently, numerous developments have been made to improve the production of soluble and active proteins in heterologous expression systems. These include modifications to the expression constructs, the introduction of new and/or improved pro- and eukaryotic expression systems, and the development of improved cell-free protein synthesis systems. The introduction of robotics has enabled a massive parallelization of expression experiments, thereby vastly increasing the throughput and, hopefully, the output of such experiments. In addition, the big challenges of recombinant overexpression of membrane and secreted proteins are tackled, and some new methods are reviewed.

Expert Rev. Proteomics 4(1), 67–78 (2007)

Change of protein expression requirements in the postgenomic era

With the huge number of genes identified in the human genome project (HGP), a certain necessity has grown to investigate the function of the gene products on a biochemical level. During the completion phase of the HGP, it had become clear that the methods used previously for analyzing individual proteins would not be sufficient for the requirements of researchers in industry and academia to study thousands of proteins and their variants efficiently and quickly. High-throughput methods have been developed for all kinds of tasks, from the search for crystallization conditions for structural genomics (SG) projects [1], to technologies for an integration of biological and chemical space in an attempt to find new drug-like substances in high-throughput screening [2]. However, while all these technologies have seen dramatic developments over the last few years, the expression of the proteins required for these studies was hardly optimized. On the contrary, expressing high levels of stable and functional proteins remains the bottleneck of many scientific and biotechnological endeavors.

Early SG efforts relied on the use of bacterial expression; basically accepting a huge attrition rate in the process from identified gene to crystal structure. Soon thereafter, different attempts were made to utilize simple eukaryotic expression systems, such as yeasts, the baculovirus expression system or *Leishmania* for the production of target proteins that were not easily amenable to expression in prokaryotic hosts. However, the main paradigm remained to screen a large number of genes of interest for expression, and drop those that would not express and proceed with the remainder; thereby finding those proteins that were easy to prepare and to investigate. This approach, while having merits in showing up the limits of the methods and demonstrating the need for improvement, could not be sustainable for drug discovery, where relatively large amounts of a specific protein (or at least a functional domain thereof) in active form must be prepared. Nevertheless, the key challenge remains somewhat similar: how to investigate many different expression constructs in several different expression systems in parallel in a comparatively short amount of time with limited resources. It becomes immediately clear that such an approach can only be feasible

through the application of massive parallelization, utilizing robotics wherever possible, and by the improvement of existing expression technologies or the development of novel systems for the efficient production of recombinant protein.

A recombinant protein-production process consists of gene cloning, expression and purification. While many solutions exist for the application of parallelization of the latter, developments have only recently been made to improve the throughput in cloning, and especially in overexpression. Therefore, we will focus on some recent developments in the improvement of proteins for heterologous expression and the introduction of some novel expression technologies that have either been shown to be amenable to parallelization or that can at least be foreseen to be suitable for parallel, high-throughput and high-output approaches.

Cloning & optimization of expression constructs

The most important change in the cloning of expression constructs, which enabled a massive and highly productive parallelization, was probably the introduction of ligation-free cloning. Several approaches have been introduced over the last years, the most prominent probably being the Gateway® system (Invitrogen Corp., CA, USA), which enables fast and efficient shuttling between different expression vectors for different expression systems. As the procedure is also amenable to automation, it has become the method of choice for most industrial high-throughput molecular biology laboratories. Many modifications to the original method have been described since its introduction, and we believe that efficient expression and solubility screening would not be possible without the availability of rapid, automatable cloning procedures. The necessity for the generation of a large number of expression constructs becomes obvious when one considers the difficulties in finding a well-expressed and soluble protein construct for a certain target gene. A recent study of more than 200 genes coding for nonmembrane proteins from a bacterial species indicated that at least 50% of the genes would require further modification to obtain soluble or stable protein [3]. Most likely, the percentage of mammalian genes that would require some modification to optimize their recombinant expression is even higher, given the greater complexity of eukaryotic proteins and the requirements of several post-translational modifications.

Many approaches have been developed to improve recombinant protein expression by varying promoter strength, enhancing solubility with different tags, or modifying growth conditions, such as temperature or media composition. If some information regarding the structure and function of the target protein is available, rational design of expression constructs can be performed. This process of protein engineering introduces several mutations, thereby replacing individual amino acids. These replacements lead to a cumulative stabilizing effect, even if the individual contribution of each mutation is rather small. Alternatively, directed evolution has been suggested as a means to improve structural and functional properties of target proteins. In this approach, an iterative evolutionary process is applied, introducing random mutations and selecting the fittest variants from an ensemble of mutants. Contrary to classical protein engineering, this approach

works even when no knowledge about the structure and function of the protein investigated is available. Several recent examples that prove the power and potential of this methodology have been described in a recent review [4]. We believe that, besides providing insight into the general mechanisms governing protein stability and solubility, random evolution will be an important and broadly applicable method for the optimization of expression constructs.

Very often, the problems with the expression of a recombinant protein do not stem from the protein itself, but from the DNA or mRNA involved. We have seen a number of cases in which the design of a synthetic gene has aided in overcoming expression problems without any changes to the expression protocol. In the process of gene optimization, ideally, several parameters are modified in parallel. These comprise an optimization of codon usage for a particular host, the removal of sequence repeats, the adaptation of the GC content, a knockout of cryptic splice sites and RNA destabilizing sequence elements, a general optimization of mRNA stability and of the mRNA secondary structure. Given the cost of recombinant gene expression, the rather expensive step of having a synthetic gene prepared is very often heavily outweighed by the gain in throughput achievable, and thus should be considered whenever there is a reasonable suspicion that expression problems might stem from nonoptimal gene structure.

Parallelization & automation of bacterial expression systems

The expression of recombinant protein is still widely performed in bacterial hosts, with *Escherichia coli* being by far the most prominent. The advantages include ease of handling in standard laboratories, inexpensive culture media, very well-known genetics, fast growth of the cells and high rates of protein production attaining up to 20% of total cellular protein. Nevertheless, there are also some severe drawbacks upon heterologous expression of target proteins in bacterial hosts, especially with regards to targets from eukaryotic origin. The lack of most of the important post-translational modification steps (i.e., glycosylation), eukaryotic chaperones, high intracellular protein concentration, targeting to subcellular compartments and formation of complexes with certain stabilizing binding partners often lead to accumulation of misfolded and aggregated recombinant protein. This is manifested by the formation of insoluble protein aggregates in the cytoplasm of the bacterial cells, also known as 'inclusion bodies'. Another problem encountered in *E. coli* is sometimes the very low expression yield of heterologous target protein, due to biased codon usage in *E. coli* and/or secondary structures of the mRNA in the 5'-region. In order to raise the chances of obtaining sufficient amount of biological active proteins expressed in *E. coli*, several strategies have been developed. As one of the most important approaches is the attachment of fusion tags, preferably to the N-terminus of the protein of interest. These tags should improve the yield of soluble protein by two different mechanisms:

- Improving the translational efficacy of the target protein by creating more favorable mRNA structures at the 5'-end of the mRNA;

- Enhancing the solubility of the target protein by stabilizing nonaggregated forms of the protein.

Today, a large variety of such fusion tags are used, with the most popular being thioredoxin (Trx), glutathione-S-transferase (GST), maltose binding proteins (MBP) and NusA [5]. Besides these widely used fusion tags, other proteins have been described to have expression and/or solubility enhancing properties, such as: SUMO [6], the N-terminal domain of translation initiation factor 2 [7], rubredoxin [8], protein disulfide isomerase [9], a fragment of bacteriophage T7 kinase [10] and *E. coli* chaperone Skp [10]. Several studies were performed to compare the effect of solubility enhancement of the different fusion tag [11–14], which by-and-large identified MBP and NusA as the most potent tags for solubilization of target proteins. It is assumed that most of these highly soluble fusion tags act in a passive mode by keeping their passenger protein in solution and sequestering it from aggregates [15]. However, there are also some indications that at least MBP has also some chaperone-like functions and plays a more active role in the solubility enhancement of a target protein [16]. If a strategy with a fusion tag is chosen, the tag must usually be removed from the target protein after purification in order to exclude interference of the tag with subsequent studies. This removal is usually performed with highly sequence-specific proteases that cleave at particular amino acid motifs present within polypeptide chains. In the earlier days, proteases such as thrombin, Factor Xa or enterokinase were used; however, their specificity to substrate recognition sites is limited, which can result in unwanted cleavage fragments. To overcome this problem, new proteases (often from viral origin) with a much higher cleavage stringency such as PreScission™ (GE Healthcare Bio-Sciences Corp., NJ, USA), tobacco etch virus (TEV) or 3C, have been developed. A more detailed overview of tag removal strategies are described by Arnau and colleagues [17].

In parallel to the various fusion tags, other strategies are also pursued in order to increase the chances of obtaining soluble and functionally active protein in *E. coli*. For instance, lowering the temperature from 37°C to 20–25°C during induction, and decreasing the amount of the inducing agent, such as isopropyl β-D-1-thiogalactopyranoside (IPTG), often leads to higher percentage of soluble protein, which is most probably caused by slower translation rates and diminished formation of aggregation-prone folding intermediates. Special plasmids have been engineered that contain the promoter of the cold-shock protein *cspA* rather than the classical *lac* or T7 promoters, which allow the expression of recombinant proteins, even at a lower temperature around 15°C [18]. The *cspA* promoter turns on transcription of the mRNA containing the gene of interest in high levels by simply lowering the incubation temperature to lower than 15°C without adding an inducer such as IPTG. Another well-known strategy to augment the yield of soluble proteins is to co-express bacterial chaperonins such as GroEL/GroES, DnaK, DnaJ and GrpE. The genes of these chaperonins are encoded on special expression plasmids with different replication of origin

and antibiotic resistance markers. Elevated levels of these chaperones should enhance proper folding and/or prevent aggregation of newly synthesized proteins in the bacterial cell [19–21]. A combination of incubation at low temperatures and the help of chaperonins for correct folding a new system has been described in which the chaperonins Cpn60/Cpn10 from the psychrophilic organism *Oleispira antarctica* are inserted on a single plasmid from which expression of the target gene is performed together with the chaperonins. The advantage of this approach is that cold-adapted chaperonins are much more active at temperatures between 10–15°C than GroEL/GroES, for instance, which has its optimum temperature around 37°C [22,23]. The choice with a set of N-terminal fusion tags, host strains, incubation temperatures and soon opens a variety of different combinations that can all be tested in parallel in order to determine the most favorable conditions for obtaining high amounts of soluble target protein. Since *E. coli* can be transformed very easily with plasmids, cultivated in 48 or even 96 deep-well plates, and lysed with simple chemicals, almost every necessary step can be implemented on an automated robotics system and within short timelines up to a few hundred different combinations can be evaluated. Some examples of (semi)automated small-scale experiments have been described [12,24–27].

Besides improving the levels of soluble recombinant protein during expression with the various approaches described earlier, refolding of aggregated, insoluble protein from inclusion bodies is still a very interesting and promising alternative. Usually, the inclusion bodies are resolubilized in a solution containing a denaturing agent such as urea or guanidine hydrochloride. Refolding is then performed by fast dilution or dialysis of the denatured protein into a buffer that often contains reagents such as glycerol (poly)-ethyleneglycol, arginine, glycine, glutathione, detergents and so on. All these additives should help to stabilize the protein in a nonaggregated state and increase the chance that the protein can assume its correct conformation. The additives and other parameters, such as pH value, can be combined, resulting in a variety of conditions that can be tested in parallel in order to determine the most promising combination. This approach has been described with the successful refolding of various proteins [28–30]. Besides adding small-molecular-weight compounds, refolding can also be performed in the presence of immobilized protein catalysts such as chaperonin fragments, protein disulfide isomerase and/or protein prolyl isomerase, which should all accelerate correct folding of the target protein [31,32]. A quite different approach has been described recently, where refolding of proteins can be directly achieved from pure inclusion bodies by applying very high (up to 3000 bar) hydrostatic pressures. Under such high pressures, hydrophobic and electrostatic interactions within protein aggregates are disrupted and the monomeric forms of the protein are stabilized, which enables the correct folding to its native structure. The procedure is simply performed by resuspending the inclusion bodies into the appropriate buffer (often containing additives like those for classical refolding) and incubating it for several hours at high pressure. This technology

has been successfully applied for refolding of, for instance, nuclear receptors [33], human phosphatases and *Drosophila* Gram-negative binding proteins [34].

Novel baculoviral expression technologies for insect cells

The baculovirus expression vector system (BEVS) has long been a workhorse for recombinant protein expression as the system is relatively easy to handle. The cultivation of the insect host cells: *Spodoptera frugiperda* 9 and 21 (Sf9, Sf21), *Trichoplusia ni* (Tn5, HighFive™, Invitrogen) requires only very basic cell culture equipment compared with what is needed for mammalian cells, uses comparatively cheap cell culture media and has been shown to be easily scalable to the needs of different applications from the milliliter scale for the preparation of small amounts of protein for functional studies, up to the bioreactor scale for the production of therapeutic protein. In addition, the BEVS can be used under minimal safety precautions. Furthermore, insect cells enable post-translational modifications similar to those of higher eukaryotes, including *Homo sapiens*, such as phosphorylation, myristylation or disulfide bridge formation, with the notable exception of glycosylation. While N-linked glycosylation occurs, insect cells lack the ability to synthesize branched, multiantennar glycans (see later).

The drawback of the BEVS is the need for the generation of a recombinant baculovirus, which, depending on the method used, takes several weeks and includes a number of handling steps that are relatively labor-intensive and not easily amenable to automation, such as the picking of viral plaques for the clonal selection of a virus population, or the determination of virus titers using a plaque assay. A relatively recent review describes the application of classic baculovirus protocols [35].

Several years ago, significant improvement was made by the introduction of the Bac-to-Bac™ (Invitrogen) method, which shifts the recombination process from the eukaryotic host (i.e., the insect host cells) to *E. coli* using a plasmid–bacmid recombination process to introduce the gene of interest into the baculoviral genome and β-galactosidase complementation for the selection of positive clones. The bacmid DNA can then be prepared in larger quantities and sufficient purity to be transfected into insect host cells (Sf9), where the generation of baculovirus occurs. After two to three rounds of virus amplification, sufficient amounts of recombinant baculovirus can be obtained to infect billions of insect cells for the production of recombinant protein sufficient for structure determination or drug discovery projects. For details of the method, see, for example, the manufacturer's website [101].

While the Bac-to-Bac method has eliminated one bottleneck in the throughput obtainable with the BEVS, it remains a very time-consuming method that still requires up to 3 weeks before sufficient amounts of virus have been obtained for protein production. One of the main reasons for this time lag is the fact that only a limited percentage of insect cells is transfected in the first step of the procedure. We have thus introduced two new vectors, termed pMACSiBac and pUltraBac, to overcome this problem and quickly select for positively transfected

cell populations [36,37]. Both vectors are based on a bicistronic plasmid, thereby enabling the cloning of a gene of interest under the control of the very later p10 promoter and a marker gene (CD4–enhanced green fluorescent protein [eGFP] fusion peptide in pMACSiBac, eGFP in pUltraBac, respectively) under the control of the late pBasic (Promega Corp., WI, USA) promoter. Within 24 h after transfection, the genes under the control of pBasic will be expressed in sufficient amounts to utilize them for the selection of successfully transfected cells. The selection process is either by the application of magnetic cell separation (pMACSiBac) or by dilution and reseedling of host cells (pUltraBac), thereby eliminating nontransfected cells efficiently from the cell culture setup (for details, see [36,37]). The introduction of eGFP as a marker gene enables fast titer determination using simple dilution and counting of fluorescent cells, and for a very efficient means of monitoring efficient infection of host cells for production, as well as for control of maintenance of infectivity upon virus storage. We also observed no negative effects of eGFP expression on the expression of the target gene of interest, probably because the two promoters are active at different times in the viral reproduction cycle, and thus there is no competition for the available transcription/translation machinery. The introduction of these new vectors in our laboratories has efficiently decreased the time required for the generation of production high-titer virus stock from previously at least 15 days to 5 days, while the expression of eGFP enabled a very careful optimization and monitoring of expression conditions. It should be noted that, in our laboratories, the use of the UltraBac vector and its derivatives has overseeded the use of the conventional Bac-to-Bac technology, largely, but not only, due to the ease of use and the straightforward approach to overcome the tedious approach of the cumbersome traditional titer determination procedure.

The expression of several genes of interest at the same time and from a single baculovirus vector using bicistronic vectors containing two promoters (pFastBac DUAL™, Invitrogen) has been an important feature of the BEVS for some years now. However, this approach has been limited by the possible number of expression cassettes that can be cloned at the same time. To overcome that issue, a bicistronic baculovirus transfer vector has been described based on the 5′-untranslated region (UTR) internal ribosome entry site (IRES) of the *Rhopalosiphum padi* virus (RhPV) [38]. Recombinant baculoviruses containing the red fluorescent protein gene and green fluorescent protein gene flanking the RhPV 5′UTR IRES can simultaneously produce dual fluorescence in recombinant virus-infected Sf21 cells under the control of a polyhedrin promoter with the translational efficiency of the RhPV 5′UTR IRES being about threefold weaker than cap-dependent translation. These results suggest that the RhPV IRES can be used in the development of bicistronic baculovirus expression vectors for production of heterologous multiprotein complexes. It remains to be shown whether this vector can be expanded for the expression of three or more genes of interest at the same time, combining the approaches in the pFastBac DUAL vector and the IRES approach described here.

Another interesting approach in this direction, MultiBac, a simple and versatile system for generating recombinant baculovirus DNA to express protein complexes comprising many subunits, was described recently [39]. This method uses transfer vectors containing a multiplication module that can be nested to facilitate assembly of polycistronic expression cassettes, thereby minimizing requirements for unique restriction sites. The transfer vectors access a modified baculovirus DNA through Cre-loxP site-specific recombination or Tn7 transposition. Furthermore, specific viral genes were eliminated to give this baculovirus improved protein expression characteristics.

A novel technology utilizing homologous recombination, *FlashBAC*TM (Oxford Expression Technologies, UK) has recently been introduced, which, due to its single-step procedure, offers the potential for automation and parallelization. In this system, the baculoviral polyhedrin coding region has been replaced by a bacterial artificial chromosome and the essential open reading frame (ORF) 1629 of the baculovirus genome, required for viral replication, has been deleted. The gene of interest is then inserted by homologous recombination with an expression plasmid containing ORF 1629, thus restoring the capability for DNA replication and putting the gene of interest under the control of the strong polyhedrin promoter. After 5 days, this virus can then be directly added to insect cells for production of a recombinant protein without the need for plaque purification. Adaptation of the vector to a ligation independent cloning system, such as Gateway, might make this new technology the most promising innovation in recent years in the BEVS field.

The strength of the BEVS has long been the expression of soluble, intracellular proteins. Although several examples of successfully produced secreted proteins exist, one of the limitations in their preparation is the lack of more than a basic glycosylation machinery. Cells from *S. frugiperda* and *T. ni* provide *N*-glycosylation with high mannose and *N*-acetyl-glucosamine containing monoantennar glycans, whereas multiantennar branching and the incorporation of *N*-acetyl-neuraminic acid do not exist. This can cause problems in the stability and biological function of secreted recombinant proteins prepared using the BEVS. Recently, novel, modified insect cell strains have been engineered that are designed to overcome this limitation and provide a humanized glycosylation machinery [40]. Given the simplicity of the BEVS, this may provide a simple means for the preparation of fully glycosylated recombinant proteins that hitherto had only been available from nonrecombinant sources or from more difficult to handle mammalian expression systems.

In order to improve throughput in eukaryotic expression experiments and to decide quickly on which expression system to use, we (pMamBac [41]) and others (pBacMam [42]) have also developed vectors that can be used for the preparation of baculovirus suitable for both the infection of insect cells and the transduction of mammalian cells. In these vectors, the gene of interest is cloned under the control of both, the very late baculoviral p10 promoter and the cytomegalovirus (CMV)

promoter, enabling expression in infected insect cells or transduced mammalian cells. The expression levels in insect cells are comparable to those obtained with a vector containing just the baculoviral promoter, whereas those in mammalian cells are in the range obtained with a similar vector using pCMV. However, the virus amount required for efficient transduction of mammalian cells is very high, making a multiplicity of infection in the range of 100–400 necessary for efficient production of the target protein. Therefore, these vectors can certainly be used for protein expression in insect cells, whereas in mammalian cells, one will only get an assessment of feasibility of expression of a certain target gene in a certain cell line. Nevertheless, as this result can be obtained without the need for cloning the gene of interest in another expression vector, significant time savings can be obtained.

While the developments described earlier tackle some of the historic limitations of the BEVS, massive parallelization has not been described to date. First approaches in this direction were reported recently [43–45], with the conversion of nearly all the steps in the process, including the expression testing and purification to a multi-well plate format. This enables a significant increase in the number of constructs that can be processed in a shorter period of time and an order of magnitude increase in the number of expression conditions that can be analyzed. A key step in the process appears to be that the transfection is performed in suspension rather than adherent cells, which gives a much higher virus titer than in the standard methods. It remains to be shown how the enhanced throughput offered by this approach can be adapted to robotic handling systems and the implications this has on the capacity to conduct multiparallel protein expression studies.

Novel transient expression technologies for insect cells

Recently, plasmid-based transient expression systems have been introduced for both insect and mammalian cells. Most notably, the introduction of the InsectDirectTM system (EMD Biosciences Inc., CA, USA) might change the possibilities of parallelization and automation in eukaryotic protein expression dramatically [102]. While the baculovirus system is not (or only slightly) amenable to parallelization (see earlier) due to the many manual intervention steps in virus propagation and the lengthy incubation steps, a plasmid-based expression system circumvents these problems. After transfection of insect cells in monolayers or suspension culture using chemical transfection agents, expression of the gene of interest is driven by an early baculoviral promoter that does not require the presence of the baculoviral transcription machinery, but can rely on the host cell for efficient and fast expression. In our experience, expression levels obtained are generally lower (with exceptions) than with the baculovirus system, probably also due to the use of an early promoter. However, expression commences after less than 24 h, allowing for a fast evaluation of protein activity and expression level. After some modification to the original supplier's protocol, the expression system can be scaled up for the use in liter-size cultures, and is thus useful for everything from

expression testing in the milliliter scale up to the expression of sufficient protein amounts for high-throughput screening or crystallization trials. However, due to the necessity of using significant amounts of transfection reagent, there is a limitation to the scalability of the expression system and we cannot at the moment envisage the use of this or similar plasmid-based expression methods at the bioreactor scale.

On the contrary, as several expression constructs can be tested in parallel under different conditions very easily, without the need for preparing virus stocks, it is relatively simple to choose the most appropriate construct for the task at hand, selecting for high solubility, expression yield and/or specific activity. In order to determine whether the expression results from a parallel study of different constructs in a plasmid-based system can be transferred to the baculovirus system, and thus enable a better scalability of expression once the optimal expression construct has been determined, we systematically compared the expression of targets of pharmacological interest (kinases, lyases, phosphatases and hydrolases) in the different expression systems. In this limited study, we found that the optimal construct determined using InsectDirect was also the best behaved expression construct in the BEVS in terms of overall expression yields, yield of soluble, nonaggregated material and specific activity in all cases [FORSTNER AND COLLEAGUES, UNPUBLISHED DATA]. Therefore, we assume that it is safe to perform expression testing using the plasmid-based system and after optimizing the construct to either scale up the expression in the same system or prepare a baculovirus using the same construct. This notwithstanding, one must remember that expression using the BEVS is not only dependent on the gene of interest, but also on other physical parameters (e.g., multiplicity of infection, time of infection and temperature), medium composition and so on, and further experiments must be performed to find optimal expression conditions.

It can now easily be envisaged that this expression system is amenable to automation using standard laboratory robotics and cell culture hardware. This might open up the possibility to achieve parallelization rates and expression testing times similar to what has been possible with *E. coli* for some years now, enabling a more efficient, faster tackling of the problem of hard-to-express proteins.

Viral expression technologies for mammalian cells

Viral vectors are an easy and effective way to deliver a gene of interest to a host cell in an efficient and nondestructive way. While some viruses (e.g., retroviruses), promote integration of the viral genome in the host cell chromosome, most others can be used for transient expression with recombinant protein production usually only occurring during certain stages of the viral life cycle. The simplicity of virus-mediated protein expression makes it a useful tool for production in higher eukaryotes, especially as the generation of stable recombinant animal cells is a tedious and lengthy procedure. Furthermore, stable recombinant cell lines are not usually selected for very high protein expression levels, but for stable and therefore sustainable

expression of a recombinant protein. Thus, they may not be the optimal choice for researchers trying to obtain high levels of recombinant protein from low cell-culture volumes. Besides *Retroviridae*, especially *Lentiviruses*, three major classes of viruses have been reported in the literature to be useful for the generation of sufficient amounts of recombinant protein for structural and functional studies: adenovirus, vaccinia virus, and *Alphaviridae* (e.g., Semliki Forest virus [SFV] and Sindbis virus). The major advantage of *Lentiviruses* is their capability to transduce both dividing and nondividing cells, making them a powerful tool for functional studies of proteins in a variety of host cells. They provide stable expression at high levels, and recent developments have enabled a combination of ligation-free cloning (Gateway) with the *Lentivirus* system (ViraPower™). Adenovirus, which has double-stranded linear DNA as genetic material, offers easy manipulation *in vitro* and the possibility to reach very high viral titers in a variety of host cells. Manipulation of the virus with enhancers added to the ectopic major late promoter region of the viral genome yielded the expression vector pAdBM5 [46], which have been used for the generation of a variety of recombinant proteins. Expression levels reported in the literature can reach up to 20% of the total host cell protein [47], albeit most commonly with a few exceptions, and viral proteins were prepared with this vector. Vaccinia virus contains a 200,000-bp, double-stranded DNA genome and replicated in the cytoplasm of a variety of host cells. Late viral promoters can be used to drive the expression of a gene of interest and high levels of recombinant protein can be achieved.

A drawback is the need for increased biosafety levels (BL2) for the handling of the virus. However, the high protein yields achievable, may well warrant investing in these safety measures. Protein production with vaccinia virus has been shown to be scalable up to the 1000 l range; for example, HIV-1 rgp160 [48] or human prothrombin [49] have been produced with this system. Finally, alphaviruses are a class of viruses that contain single-stranded, positive-sense RNA as their genetic material, which, upon host cell infection, immediately becomes the mRNA for viral protein transcription and virus replication. In order to make SFV safe to handle, the viral genome had to be split, and the system consists of two plasmids: the cloning vector pSFV1 contains the replicase gene and the gene of interest, whereas the pSFV-Helper plasmid encodes the structural proteins of the virus [50]. Therefore, the generated SFV-like particles are replication deficient. Further modification of the spike protein makes the generated particles noninfectious, unless activated by proteolytic cleavage for added safety. The extremely efficient SFV 26S promoter and the high number of copies of RNA (>200,000) per infected cell leads to very high expression levels. The system has been shown to be particularly useful for the preparation of membrane proteins, especially G-protein-coupled receptors [51,52], and its scalability to the bioreactor scale has been reported [53]. In our hands, the SFV expression vector system has worked particularly well for the expression of transmembrane and membrane-associated target proteins [FORSTNER AND COLLEAGUES, UNPUBLISHED DATA].

Novel transient expression technologies for mammalian cells

Mammalian expression systems have an undisputed long-standing and very successful history for the generation of recombinant proteins, mainly as biopharmaceuticals. However, for use as tool proteins in (e.g., assay development and screening, for structure elucidation and as antigens), these expression systems were generally regarded as being cumbersome, tedious and expensive. This bias has largely been overcome with the very recent development of large-scale transient transfection (LST) approaches. In particular, the recently described HEK.EBNA expression system has contributed significantly to this success [54]. The simplicity and speed of this approach compares well to expression trials using the widely applied BEVS system. In addition, proteins generated in mammalian cells are usually correctly folded, fully processed and functionally active.

The HEK 293FreeStyle™ expression system (Invitrogen [101]) is a recent development, which can be foreseen to be amenable to massive parallelization and automation. Similar to the InsectDirect expression system described earlier, the FreeStyle system makes use of plasmids as expression vectors, harboring the gene of interest under the control of the highly efficient CMV promoter. Over the last 18 months, we have used this system for the expression of ten different human secreted and membrane proteins in the liter scale in shaker flasks with a 100% success rate. The only limitations of the FreeStyle expression system in terms of scalability are the availability of sufficient amounts of plasmid DNA and the need for relatively high quantities of transfection reagent. However, after some optimization of the transfection method we have been able to use this expression system successfully in the 20-l bioreactor scale for the expression of a secreted human protein (data not shown). Despite the fact that scalability is limited, we believe that this expression system has the potential of making recombinant overexpression of genes in a mammalian host system possible at high throughput, enabling the assessment of optimal choice of construct and so on in timescales similar to those possible with prokaryotic expression systems.

Novel expression systems can contribute greatly to our arsenal of methods to tackle the task of recombinant protein expression, there are also other developments that make use of cell physiology to improve yields, such as the cold-shock response in mammalian cells [55]. While the mechanism involved in the cellular response to elevated temperatures has been extensively studied, relatively little is known about the biochemical and cell biological events in response to subphysiological temperatures. The major stress factors for eukaryotic cells at those temperatures appear to be the changes in temperature themselves and the higher levels of dissolved oxygen at lower temperatures [55]. A study on the cold-shock response in 3T3 cells reports that 10% of the cellular genes were regulated upon a temperature shift from 37 to 32°C, particularly those involving genes that play roles in cell proliferation, cholesterol biosynthesis and cytoskeletal family genes [56]. Several studies reporting an increase in heterologous protein production and improved product quality have been published (for details see [54];

especially Table 1 in [54]), although there does not appear to be a general trend and effects appear to be highly dependent on the cell line and the expression system used. Nevertheless, we believe that harnessing the cold-shock response in eukaryotic cells will be an important tool for improving recombinant protein product yield and quality in an industrial setting in the future. One could even think of a cold-inducible expression system for the recombinant overexpression of genes toxic to the host cell, with the cells growing at optimal temperatures and expression of the gene of interest is induced by simply lowering the ambient temperature. A first step in this direction has already been reported. The combination of two mutations in the viral replicase of Sindbis virus, an alphavirus such as SFV described earlier, which renders the replicase noncytopathic and temperature sensitive, allowed the generation of a DNA-based vector (CytTs) that demonstrates temperature-inducible expression [57]. However, similar to other stable expression systems, tedious screening of individual cell clones is required in order to get a high producer cell clone. To circumvent this, an episomally replicating vector was created by introducing an Epstein-Barr virus minireplicon unit into CytTs. This novel vector allowed rapid generation of cell populations that demonstrated tight regulation of expression and comparable expression levels to the ones achieved with high producer cell clones with CytTs. Protein production could then easily be scaled-up to a fermentation process.

Cell-free protein synthesis (bacterial & eukaryotic)

Cell-free protein synthesis has originally been introduced for small-scale expression of proteins, basically providing sufficient material for an initial characterization of the protein of interest. In the meanwhile, the productivity of coupled transcription-translation systems in the *E. coli* cell-free protein synthesis system has been improved to yield protein amounts in the range of 5–6 mg/ml reaction mixture, which makes the method potentially useful for SG and drug discovery efforts [58]. In general, as PCR-generated DNAs are used for mRNA preparation, it is not necessary to clone ORF in expression vectors for use in cell-free protein synthesis. This cloning-free nature makes cell-free systems particularly useful for the high-throughput production of many proteins (or many expression constructs) in parallel.

Cell-free protein synthesis has been successfully employed for stable isotope labeling of proteins for nuclear magnetic resonance (NMR) spectroscopy in several cases (e.g., amino acid-selective ¹⁵N-labeling, dual amino acid-selective ¹³C-¹⁵N-labeling and uniform ¹³C/¹⁵N-labeling [59]). Most importantly, heteronuclear single quantum coherence (HSQC) spectra can probably be recorded directly from the reaction without the need for purification of the products to quickly identify possible problems due to multiple conformations or sample heterogeneity [60]. Finally, integral membrane proteins that are notoriously difficult to express *in vivo* can probably be produced *in vitro* in ranges of more than a milligram per milliliter, making them amenable to biophysical analysis by NMR spectroscopy [61,62].

There is further use for cell-free systems for the incorporation of selenomethionine (Se-Met), replacing methionine in recombinant proteins for multiwavelength anomalous diffraction (MAD) phasing of protein crystal structures [63]. The high cytotoxicity of Se-Met can cause low production yields of a target protein when methionine in the culture medium is replaced by Se-Met. Since cytotoxicity is not a problem for cell-free systems, the method could thus be applied for the incorporation of Se-Met into proteins [64].

Furthermore, a cell-free protein synthesis system was reconstituted from recombinant His-tagged protein factors purified to homogeneity, termed 'Protein synthesis Using Recombinant Elements' (PURE) [65]. The PURE system was capable of producing recombinant protein at a rate of 0.1–0.2 µg/ml/h in batch mode, and the protein products could easily be purified by affinity chromatography, removing the tagged protein factors in a reverse mode of purification.

In addition, cell-free protein synthesis was achieved with a microfabricated reaction chamber on a plastic chip, thereby enabling further miniaturization of protein synthesis on a cell-free microchip [66].

While the scalability of cell-free protein synthesis has thus been shown to be possible in both directions, miniaturization and scale up, the application range of cell-free methods has been somewhat limited because the reducing condition of the system derived from cytosolic factors is not suitable for the formation of disulfide bridges. Nevertheless, the heavy and light chain genes of the Fab fragment of a catalytic antibody were simultaneously expressed in the *E. coli* cell-free system without the addition of a reducing agent [67]. When reduced and oxidized glutathione, protein disulfide isomerase and molecular chaperones, GroEL and GroES, were added, the solubility and antigen binding activity of the purified Fab fragment were greatly improved and the purified Fab fragment was shown to be catalytically active.

In addition to the already well-developed and established *E. coli*-based cell-free protein synthesis system, eukaryotic cell-free system from wheat-germ extract [68] and rat reticulocyte lysate [69] have also been developed. Due to the costs of goods and ease of use, the wheat-germ lysate system is more commonly used today. This system was prepared from extensively washed wheat embryos to remove translation inhibitors that act as a suicide system directed on ribosomes. Translation could be observed in a batch system for up to 4 h and for up to 3 days in a dialysis system, yielding up to 3 mg of protein. An improved methodology using bilayer diffusion rather than the dialysis method has been successfully applied to high-throughput functional proteomics [68,70]. In this study, more than 50 genes were shown to be translated in parallel, yielding between 0.1 and 2.3 mg of recombinant protein in 48 h.

Post-translational modifications pose another general problem for recombinant protein production. The generation of a Sf-21 cell-based lysate has been described recently, providing core protein glycosylation enzymes [71]. Prior to this development, cell-free protein synthesis of glycosylated proteins had to rely on the supplementation of membrane vesicles derived from

endoplasmic reticulum preparations (i.e., microsomal preparations). Other post-translational modifications have been described in lysates from higher eukaryotes, providing phosphorylation, myristylation, farnesylation, isoprenylation and adenylation capabilities to cell-free systems [72]. On the contrary, the dynamic heterogeneity and complexity of post-translational modifications sometimes makes the production of homogeneous protein samples *in vivo* difficult, and methods have been proposed to create artificial modifications that mimic the structure of the natural ones in order to overcome this obstacle [73]. Obviously, the most favorable platform for this new strategy appears to be cell-free protein synthesis.

There was a certain lag time after the introduction of cell-free systems before their usefulness for protein biochemistry could be shown. However, the reports cited earlier and developments in protein evolution, protein arrays, diagnostics and therapeutics applying cell-free protein expression reactions (described elsewhere in more detail [74]) indicate that cell-free protein synthesis appears to be a technology that is evolving rapidly, and novel applications of this powerful and flexible technology can certainly be expected.

Databases for project tracking in protein expression

With the advent of high-throughput technologies for gene cloning, protein expression and purification, the need for efficient laboratory information systems to track the hundreds, if not thousands, of individual experiments has evolved. Similar needs have become evident in the early 1990s with the advent of high-throughput screening. In that aspect, several of the technologies evolved in the field of high-throughput screening can be easily transferred towards the field of high-throughput protein expression [75].

In high-throughput protein expression, one must track not only the various DNA constructs produced in the process with their genealogy and their sequence information, but also the cloning vectors, oligonucleotides and their storage/sample information. The tracking of the protein expression information, such as expression host, expression conditions, induction/harvesting time and multiplicity of infection is more demanding. Collecting that information is essential for data integrity in a largely automated process, but also important for correct automation of the process via liquid handling and robotics. The technology requires automation of information retrieval and depends mostly upon the use of barcoded information on laboratory consumables, such as Eppendorf tubes, microtiter plates, shake flasks and so forth.

In our laboratories, we routinely track all molecular biology and protein expression information in a rather simple database, allowing for an efficient comparison of results and avoiding the duplication of tedious tasks and avoidable errors. We believe that the collection of this information will provide not only easily accessible experimental data, but more importantly, a wealth of knowledge to be mined for a better understanding of the factors governing the expression of well-folded, soluble recombinant proteins in high yields.

Expert commentary

Recombinant protein expression is most often a difficult venture and requires careful and thorough experimental planning and a variety of technologies. With a variety of host organisms and expression systems available, the choice of the one system most suitable for the task at hand heavily influences the outcome, and it is worthwhile to spend some time thinking about the optimal procedure and the desired specifications of the product. Bacterial and eukaryotic expression systems are constantly being improved in order to meet all the demands of drug discovery and structural biology, and one can choose from a plethora of possibilities. Compared against only a few years ago, when recombinant protein expression was limited to relatively few expression systems, one can now select the optimal system for a given task without having to invest huge amounts of time and money in the set up of new facilities. New developments in the field of cell-free expression enable fast and comparatively cheap isotope labeling for structural studies as well as for the fast generation of small amounts of proteins to begin drug discovery efforts. The availability of sufficient amounts of pure protein of consistent quality will remain most important for these tasks, and only recombinant protein expression can satisfy this need.

Five-year view

Obviously, the field of recombinant protein expression has received increased attention due to the needs of SG and functional proteomics for large quantities of soluble, active proteins. The technologies and methods described earlier are but a few of the recent developments in the production of recombinant proteins, and we had to make deliberate choices about what to include in this review. The need for pure protein in good quality and yields beyond the microgram scale has triggered

developments in the field of recombinant protein expression that are far from being finished. We believe that transient, plasmid-based expression systems and massive parallelization will dominate the developments to emerge, mainly because the handling steps and incubation times involved are much more suitable for robotics than those required when working with viral vectors. In addition, the rise of cell-free expression technologies cannot be neglected as more and more successes with these techniques are being reported. What are crucially needed are new and improved technologies for the expression and display of membrane proteins as they will certainly be the prime drug targets of the future. Developments such as nanodisks are certainly the first steps in this direction, and we assume that the technological hurdles will be overcome once sufficient interest is generated. Furthermore, as regulatory authorities increasingly react to the demands of a risk adverse society, a more thorough biochemical characterization of molecular targets will become a prerequisite for drug approval, and thus more efficient, faster expression technologies will become necessary.

Acknowledgements

We are very grateful to our colleagues at Novartis Institute of Biomedical Research (NIBR) Department of Discovery Technologies, who have contributed with ideas and advice to the successful implementation and application of the various technologies for pro- and eukaryotic protein expression; in particular René Amstutz, Jutta Blank, Felix Freuler, Peter Fuerst, Sabine Geisse, Fraser Glickman, Martin Klumpp, Hans Kocher, Christian Ostermeier, Johannes Ottl, Bjoern Philipps, Paul Ramage, Sabine Wellnitz, Hans Widmer and Hartmut Zehender, and we fully appreciate their continuous support, encouragement and enthusiasm for novel technologies in drug discovery.

Key issues

- Recombinant protein expression is at the foundation of successful drug discovery and development in pharmaceutical companies.
- Classic expression systems such as *Escherichia coli* and baculovirus are constantly being improved by the introduction of new strains, vectors and tags.
- Novel plasmid-based transient expression systems enable expression in mammalian and insect cell host systems with relatively little technical effort.
- The development of cell-free expression systems is of increasing importance, especially for structural genomics.
- Parallelization and automation are both key drivers of and are driven by the new developments in expression technologies.

References

Papers of special note have been highlighted as:

- of interest
- of considerable interest

- McPherson A. Protein crystallization in the structural genomics era. *J. Struct. Funct. Genomics* 5(1–2), 3–12 (2004).
- Mayr LM. Tackling the chemo-genomic space by novel screening technologies: monograph of ESRF Workshop 58.
- In: *Chemical Genomics: Small Molecule Probes to Study Cellular Function*. Jaroch S, Weinmann H (Eds.), Springer Verlag, Weinheim, Germany, 111–173 (2006).
- Christendat D, Yee A, Dharamsi A *et al.* Structural proteomics of an archaeon. *Nature Struct. Biol.* 7, 903–909 (2000).
- Roodveldt C, Aharoni A, Tawfik DS. Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* 15, 50–56 (2005).
- Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60, 523–533 (2003).
- Butt TR, Edavettal SC, Hall JP, Mattern MR. SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* 43, 1–9 (2005).

- 7 Sorensen HP, Sperling-Petersen HU, Mortensen KK. A favorable solubility partner for the recombinant expression of streptavidin. *Protein Expr. Purif.* 32, 252–259 (2003).
- 8 Kohli BM, Ostermeier C. A rubredoxin based system for screening of protein expression conditions and on-line monitoring of the purification process. *Protein Expr. Purif.* 28, 362–367 (2003).
- 9 Liu Y, Zhao TJ, Yan YB, Zhou HM. Increase of soluble expression in *Escherichia coli* cytoplasm by a protein disulfide isomerase gene fusion system. *Protein Expr. Purif.* 44, 155–161 (2005).
- 10 Chatterjee DK, Esposito D. Enhanced soluble protein expression using two new fusion tags. *Protein Expr. Purif.* 46, 122–129 (2006).
- **New fusion tags designed for improvement of soluble expression of recombinant proteins in *Escherichia coli*.**
- 11 Hammarström M, Hellgren N, Van der Berg S *et al.* Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*. *Prot. Sci.* 11, 313–321 (2002).
- 12 Shih YP, Kung WM, Chen JC *et al.* High-throughput screening of soluble recombinant proteins *Prot. Sci.* 11, 1714–1719 (2002).
- 13 Davis GD, Elisee C, Newham DM, Harrison RG. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotech. Bioeng.* 65, 382–388 (1999).
- 14 Dyson MR, Shadbolt PS, Vincent KJ *et al.* Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnology* 4, 32 (2004).
- 15 Nallamsetty S, Waugh DS. Solubility-enhancing protein MBP and NusA play a passive role in the folding of their fusion partner. *Protein Expr. Purif.* 45, 175–182 (2006).
- 16 Bach H, Mazor Y, Shaky S *et al.* *Escherichia coli* maltose binding protein as a molecular chaperone for recombinant intracellular single-chain antibodies. *J. Mol. Biol.* 312, 79–93 (2001).
- 17 Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* 48, 1–13 (2006).
- 18 Qing G, Ma L-C, Khorchid A *et al.* Cold-shock induced high-yield expression protein production in *Escherichia coli*. *Nat. Biotechnol.* 22, 877–882 (2004).
- 19 Nishihiara K, Kanemori M, Kitagawa M *et al.* Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisted refolding of an allergen from Japanese cedar pollen, Cry2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 64, 1694–1699 (1999).
- 20 Xu Y, Weng CL, Narayanan N *et al.* Chaperone-mediated folding and maturation of the penicillin acylase precursor in the cytoplasm of *Escherichia coli*. *Appl. Environ. Microbiol.* 71, 6247–6253 (2005).
- 21 Sung-Gun K, Dae-Hyuk K, Dae-Hee L *et al.* Coexpression of folding accessory proteins for production of active cyclodextrin glycosyltransferase of *Bacillus macerans* in recombinant *Escherichia coli*. *Protein Expr. Purif.* 41, 426–432 (2005).
- 22 Ferrer M, Chernikova TN, Yakimov MM *et al.* Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nat. Biotechnol.* 21, 1266–1267 (2003).
- 23 Strocchi M, Ferrer M, Timmis KN, Golyshin PN. Low temperature-induced systems failure in *Escherichia coli*: insights from rescue by cold-adapted chaperones. *Proteomics* 6, 193–206 (2006).
- **The crucial role of the cold-shock Cpn60/10 chaperone for successful growth of *E. coli* at low temperatures is elucidated.**
- 24 Braun P, Ju Y, Halleck A *et al.* Proteome-scale purification of human proteins from bacteria. *Proc. Natl Acad. Sci. USA* 99, 2654–2659 (2002).
- 25 Knaust RKC, Nordlund P. Screening for soluble expression of recombinant proteins in a 96-well format. *Anal. Biochem.* 297, 79–85 (2001).
- 26 Scheich C, Leitner D, Sievert V *et al.* Fast identification of folded human protein domains expressed in *E. coli* suitable for structural analysis. *BMC Struct. Biol.* 4, 4 (2004).
- 27 Lin CT, Moore PA, Auberry DL *et al.* Automated purification of recombinant proteins: combining high-throughput with high yield. *Protein Expr. Purif.* 47, 16–24 (2006).
- 28 Tobbell DA, Middleton BJ, Raines S *et al.* Identification of *in vitro* folding conditions for procathepsin S and cathepsin S using fractional factorial screens. *Protein Expr. Purif.* 24, 242–252 (2002).
- 29 Vincentelli R, Canaan S, Campanacci V *et al.* High-throughput automated refolding screening of inclusion bodies. *Prot. Sci.* 22, 2782–2792 (2004).
- 30 Willis MS, Hogan JK, Prabhakar P *et al.* Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions. *Prot. Sci.* 14, 1818–1826 (2005).
- 31 Altamirano MM, Consuelo G, Possani LD, Fersht AR. Oxidative refolding chromatography of the scorpion toxin Cn5. *Nat. Biotechnol.* 17, 187–191 (1999).
- 32 Gao YG, Guan YX, Yao SJ, Cho MG. On-column refolding of recombinant human interferon- γ with an immobilized chaperone fragment. *Biotechn. Prog.* 19, 915–920 (2003).
- 33 Schoner BE, Bramlett KS, Guo H, Burris TP. Reconstitution of functional nuclear receptor proteins using high pressure refolding. *Mol. Gen. Metabol.* 84, 318–322 (2005).
- 34 Lee SH, Carpenter JF, Chang BS *et al.* Effects of solutes on solubilization and refolding of proteins from inclusion bodies with high hydrostatic pressure. *Prot. Sci.* 15, 304–313 (2006).
- 35 Ikonomou L, Schneider YJ, Agathos SN. Insect cell culture for industrial production of recombinant proteins. *Appl. Microbiol. Biotechnol.* 62, 1–20 (2003).
- 36 Philipps B, Forstner M, Mayr LM. Baculovirus expression system for magnetic sorting of infected cells and enhanced titer determination. *Biotechniques* 36, 80–83 (2004).
- 37 Philipps B, Rotmann D, Wicki M *et al.* Time reduction and process optimization of the baculovirus expression system for more efficient recombinant protein production in insect cells. *Protein Expr. Purif.* 42, 211–218 (2005).
- **Description of an optimized baculovirus vector and expression methodology that has changed the way recombinant proteins are produced at Novartis International AG (Basel, Switzerland).**
- 38 Chen YJ, Chen WS, Wu TY. Development of a bicistronic baculovirus expression vector by the *Rhopalosiphum padi* virus 5' internal ribosome entry site. *Biochem. Biophys. Res. Commun.* 335, 616–623 (2005).
- 39 Berger I, Fitzgerald DJ, Richmond TJ. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* 22, 1583–1587 (2004).
- 40 Jarvis DL. Developing baculovirus–insect cell expression systems for humanized recombinant glycoprotein production. *Virology* 310, 1–7 (2003).

- 41 Philipps B, Forstner M, Mayr LM. A baculovirus expression vector system for simultaneous protein expression in insect and mammalian cells. *Biotechnol. Prog.* 21, 708–711 (2005).
- 42 Kost TA, Condreay JP, Jarvis DL. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* 23, 567–575 (2005).
- **Excellent paper illustrating the capabilities of baculovirus transduction of mammalian cells.**
- 43 McCall EJ, Danielsson A, Hardern IM *et al.* Improvements to the throughput of recombinant protein expression in the baculovirus/insect cell system. *Protein Expr. Purif.* 42, 29–36 (2005).
- 44 Hunt I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr. Purif.* 40, 1–22 (2005).
- **Very good overview of technological developments.**
- 45 Bahia D, Cheung R, Buchs M *et al.* Optimisation of insect cell growth in deep-well blocks: development of a high-throughput insect cell expression screen. *Protein Expr. Purif.* 39, 61–70 (2005).
- 46 Massie B, Dionne J, Lamarche N *et al.* Improved adenovirus vector provides herpes simplex virus ribonucleotide reductase R1 and R2 subunits very efficiently. *Biotechnology (NY)* 13, 602–608 (1995).
- 47 Garnier A, Cote J, Nadeau I *et al.* Scale up of the adenovirus expression system for the production of recombinant protein in human 293S cells. *Cytotechnology* 15, 145–155 (1994).
- 48 Pialoux G, Excler JL, Riviere Y *et al.* A prime-boost approach to HIV preventive vaccine using a recombinant canarypox virus expressing glycoprotein 160 (MN) followed by a recombinant glycoprotein 160 (MN/LAI). *AIDS Res. Hum. Retroviruses* 11, 373–381 (1995).
- 49 Falkner FG, Turecek PL, MacGillivray RT *et al.* High level expression of active human prothrombin in a vaccinia virus expression system. *Thromb. Haemost.* 68, 119–124 (1992).
- 50 Liljestrom P, Garoff H. A new generation of animal cell expression vectors based on the Semliki Forest virus replicon. *Biotechnology (NY)* 9, 1356–1361 (1991).
- 51 Lundstrom K. Semliki Forest virus vectors for rapid and high-level expression of integral membrane proteins. *Biochim. Biophys. Acta* 1610, 90–96 (2003).
- 52 Hassaine G, Wagner R, Kempf J *et al.* Semliki Forest virus vectors for overexpression of 101 G protein-coupled receptors in mammalian host cells. *Protein Expr. Purif.* 45, 343–351 (2006).
- 53 Lundstrom K. Semliki Forest virus vectors for large-scale production of recombinant proteins. *Methods Mol. Med.* 76, 525–543 (2003).
- 54 Geisse S, Henke M. Large-scale transient transfection of mammalian cells: a newly emerging attractive option for recombinant protein production. *J. Struct. Funct. Genomics* 6, 165–170 (2005).
- 55 Al-Fageeh MB, Marchant RJ, Carden MJ, Smales CM. The cold-shock response in cultured mammalian cells: harnessing the response for the improvement of recombinant protein production. *Biotechnol. Bioeng.* 93, 829–835 (2006).
- 56 Beer C, Buhr P, Hahn H *et al.* Gene expression analysis of murine cells producing amphotropic mouse leukaemia virus at a cultivation temperature of 32 and 37 degrees C. *J. Gen. Virol.* 84, 1677–1686 (2003).
- 57 Ivanova L, Brandli J, Saudan P, Bachmann MF. Hybrid Sindbis/Epstein-Barr virus episomal expression vector for inducible production of proteins. *Biotechniques* 39, 209–212 (2005).
- **Development of a cold-inducible mammalian expression vector – exciting!**
- 58 Kigawa T, Yabuki T, Yoshida Y *et al.* Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* 442, 15–19 (1999).
- 59 Yokoyama S. Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* 7, 39–43 (2003).
- 60 Yee A, Chang X, Pineda-Lucena A *et al.* An NMR approach to structural proteomics. *Proc. Natl Acad. Sci. USA* 99, 1825–1830 (2002).
- 61 Guignard L, Ozawa K, Pursglove SE *et al.* NMR analysis of *in vitro*-synthesized proteins without purification: a high-throughput approach. *FEBS Lett.* 524, 159–162 (2002).
- 62 Elbaz Y, Steiner-Mordoch S, Danieli T, Schuldiner S. *In vitro* synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *Proc. Natl Acad. Sci. USA* 101, 1519–1524 (2004).
- 63 Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9, 1665–1672 (1990).
- 64 Kigawa T, Yamaguchi-Nunokawa E, Kodama K *et al.* Selenomethionine incorporation into a protein by cell-free synthesis. *J. Struct. Funct. Genomics* 2, 29–35 (2002).
- 65 Shimizu Y, Inoue A, Tomari Y *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* 19, 751–755 (2001).
- 66 He M, Taussig MJ. DiscernArray technology: a cell-free method for the generation of protein arrays from PCR DNA. *J. Immunol. Methods* 274, 265–270 (2003).
- 67 Jiang X, Ookubo Y, Fujii I *et al.* Expression of Fab fragment of catalytic antibody 6D9 in an *Escherichia coli in vitro* coupled transcription/translation system. *FEBS Lett.* 514, 290–294 (2002).
- 68 Sawasaki T, Ogasawara T, Morishita R, Endo Y. A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl Acad. Sci. USA* 99, 14652–14657 (2002).
- 69 Tymms MJ (Ed.) *In vitro* transcription and translation protocols. In: *Methods in Molecular Biology. Vol. 37.* Humana Press, NJ, USA (1995).
- 70 Endo Y, Sawasaki T. High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. *J. Struct. Funct. Genomics* 5, 45–57 (2004).
- 71 Tarui H, Murata M, Tani I *et al.* Establishment and characterization of cell-free translation/glycosylation in insect cell (*Spodoptera frugiperda* 21) extract prepared with high pressure treatment. *Appl. Microbiol. Biotechnol.* 55, 446–453 (2001).
- 72 Jackson AM, Boutell J, Cooley N, He M. Cell-free protein synthesis for proteomics. *Brief Funct. Genomic. Proteomic.* 2, 308–319 (2004).
- 73 Davis BG. Biochemistry. Mimicking post-translational modifications of proteins. *Science* 303, 480–482 (2004).
- 74 Katzen F, Chang G, Kudlicki W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* 23, 150–156 (2005).
- **Excellent overview of cell-free transcription/translation.**
- 75 Acton TB, Gunsalus KC, Xiao R *et al.* Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium. *Methods Enzymol.* 394, 210–243 (2005).

Website

- 101 Invitrogen Corp.
www.invitrogen.com
- 102 EMD Biosciences, Inc.
www.emdbiosciences.com

Affiliations

- *Michael Forstner, PhD*
Laboratory Head, Protein Expression & Purification Novartis Institutes of BioMedical Research, Discovery Technologies/Lead Discovery Center CH-4002 Basel, Switzerland
Tel.: +41 613 243 426
Fax: +41 613 245 155
michael.forstner@novartis.com
- *Lukas Leder, PhD*
Laboratory Head, Protein Expression & Purification, Novartis Institutes of BioMedical Research, Discovery Technologies/Lead Discovery Center, CH-4002 Basel, Switzerland
Tel.: +41 613 249 231
Fax: +41 613 245 155
lukas.leder@novartis.com
- *Lorenz M Mayr, PhD*
Executive Director, Head BioChemical Screening Group, Novartis Institutes of BioMedical Research, Discovery Technologies/Lead Discovery Center, CH-4002 Basel, Switzerland
Tel.: +41 613 245 441
Fax: +41 613 246 164
lorenz.mayr@novartis.com