Advances in recombinant protein expression for use in pharmaceutical research
Rene Assenberg¹, Paul T Wan², Sabine Geisse³ and Lorenz M Mayr⁴

Protein production for structural and biophysical studies, functional assays, biomarkers, mechanistic studies in vitro and in vivo, but also for therapeutic applications in pharma, biotech and academia has evolved into a mature discipline in recent years. Due to the increased emphasis on biopharmaceuticals, the growing demand for proteins used for structural and biophysical studies, the impact of genomics technologies on the analysis of large sets of structurally diverse proteins, and the increasing complexity of disease targets, the interest in innovative approaches for the expression, purification and characterisation of recombinant proteins has steadily increased over the years. In this review, we summarise recent developments in the field of recombinant protein expression for research use in pharma, biotech and academia. We focus mostly on the latest developments for protein expression in the most widely used expression systems: Escherichia coli (E. coli), insect cell expression using the Baculovirus Expression Vector System (BEVS) and, finally, transient and stable expression of recombinant proteins in mammalian cells.

Addresses
¹ Novartis Pharma AG, Expertise Platform Proteases, Fabrikrasse-16, Basel CH-4002, Switzerland
² AstraZeneca Ltd., Discovery Sciences – Reagents & Assay Development, Pepparedsleden 1, Mölndal SE-43183, Sweden
³ Novartis Pharma AG, Novartis Biologics Center, WSJ-506, Basel CH-4002, Switzerland
⁴ AstraZeneca Ltd., Discovery Sciences – Reagents & Assay Development, Mereside 52G16, Alderley Park, Macclesfield, SK10 4TG, United Kingdom

Corresponding author: Mayr, Lorenz M (Lorenz.Mayr@astraZeneca.com)

Current Opinion in Structural Biology 2013, 23:xx-yy
This review comes from a themed issue on New constructs and expressions of proteins
Edited by Imre Berger and Lorenz M Mayr

0959-440X/ – see front matter, Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.sbi.2013.03.008

Introduction
Protein production remains a fundamental cornerstone of drug development in the pharmaceutical industry. Biophysical techniques for the characterisation of protein and ligand interactions are crucial to this endeavor but have the drawback that they often require, sometimes repeatedly, the supply of substantial quantities of protein (Figure 1). Advances in the sophisticated exploitation of different expression systems ensuring that these protein production requirements can be met is the focus of this article.

Protein expression in Escherichia coli
Expression of heterologous proteins in the E. coli host system often remains the preferred choice for a number of reasons: it is inexpensive, offers rapid culture times and the ability to achieve high biomass and high protein yields. Extensively studied, there is a wealth of biochemical and physiological knowledge available, notwithstanding that E. coli is also highly genetically amenable to manipulation with a variety of genetic tools. However, as the protein targets become more mammalian-like, E. coli faces constraints in expressing such proteins due to their increased complexity (e.g. post-translational modifications). Below we discuss recent innovations that attempt to overcome some of these shortfalls.

Enabling human protein production in E. coli: disulphide formation and glycoengineering
Correct disulphide bond formation is a major post-translational modification that dictates the correct fold of a protein, inferring protein stability and biological function. This is particularly true for many secreted mammalian proteins (hormones, growth factors and immunoglobulins). Correct disulphide formation requires a redox environment, found in the periplasm of E. coli, and enzymes facilitating disulphide bond formation, such as disulphide isomerases (Dsbs) and peptidyl-prolyl isomerases (PPIases) [1].

One approach in obtaining disulphide proteins in E. coli involves targeting heterologously expressed proteins to the periplasm, by the addition of a signal peptide to the N-terminus of the recombinant protein [2], via the SEC-dependent pathway or the signal recognition particle (SRP) dependent translocation machinery. The former requires linear polypeptides to allow translocation and the latter strategy is useful for proteins that rapidly fold in the cytoplasm [3]. This has advantages also in downstream processing, as the periplasm contains fewer endogenous proteins than the cytoplasm, including proteases. Correct folding in the periplasm can be enhanced through co-expression of disulphide bond formation catalysts (Dsbs) and protein disulphide isomerases (PDIis or PPIases) [2]. In addition, these catalysts can ‘proof read’ for correct disulphide formation and some have innate chaperone-like properties. PPIase rescue has been demonstrated for cyclodextrin glycosyltransferase [4].
2 New constructs and expressions of proteins

<table>
<thead>
<tr>
<th>Detection Readout</th>
<th>Comment</th>
<th>Information</th>
<th>Protein requirement (per data point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR (Nuclear Magnetic Resonance)</td>
<td>In solution</td>
<td>Binding, $K_D$, cpd integrity, dynamics</td>
<td>10-500 µg</td>
</tr>
<tr>
<td>Affinity MS (Mass Spectrometry)</td>
<td>In solution</td>
<td>Binding</td>
<td>1-5 µg</td>
</tr>
<tr>
<td>ITC (Isothermal Titraton Calorimetry)</td>
<td>In solution</td>
<td>Binding, $K_D$, stoichiometry, $\Delta H$, $\Delta S$</td>
<td>200-5000 µg</td>
</tr>
<tr>
<td>DSC (Differential Scanning Calorimetry)</td>
<td>In solution</td>
<td>$\Delta T_m$, $\Delta H$, $\Delta S$</td>
<td>100-500 µg</td>
</tr>
<tr>
<td>DSF (Differential Scanning Fluorimetry; aka Thermofluor™)</td>
<td>In solution with tracer dye</td>
<td></td>
<td>0.5-5 µg</td>
</tr>
<tr>
<td>DSLS (Differential Static Light Scattering; aka Stargazer™)</td>
<td>In solution</td>
<td>$\Delta T_{agg}$</td>
<td>1-10 µg</td>
</tr>
<tr>
<td>SPR (Surface Plasmon Resonance aka Biacore™)</td>
<td>Immobilized</td>
<td>Binding, $K_D$, $k_{on}$, $k_{off}$, &quot;stoichiometry&quot;</td>
<td>0.001-0.1 µg</td>
</tr>
<tr>
<td>RWG (Resonant Waveguide aka Corning Epic™ or SRU BIND™)</td>
<td>Immobilized</td>
<td>Binding, $K_D$, &quot;stoichiometry&quot;</td>
<td>0.5-2 µg</td>
</tr>
<tr>
<td>BLI (Biolayer Interferometry aka ForteBio™)</td>
<td>Immobilized</td>
<td>Binding, $K_D$, $k_{on}$, $k_{off}$</td>
<td>0.2-1 µg</td>
</tr>
</tbody>
</table>

A comparison of various biophysical methods applied in hit/lead discovery and the estimated use of protein per sample. We estimated a typical drug discovery project will need large quantities of protein, typically in the range 50-500 mg or more per project. Running a typical fragment-based lead-generation (FBLG) project comprises in general a set of 2500 fragments and a hit rate of 1% will generate 550 data points in total (250 cpds × 22 pt DRC). This will translate into total protein consumption in the range of 30-1 500 000 mg of purified protein. Running a typical surface-plasmon resonance (SPR) project comprises in general a set of 25 000 fragments and a hit rate of 1% will generate 5500 data points in total (250 cpds × 22 pt DRC). This will translate into total protein consumption in the range of 30–3000 mg of purified protein. Running a typical isothermal titration calorimetry (ITC) project in lead optimisation comprises in general around 75 compounds (50 cpds in primary screening, 25 cpds in secondary screening; 50% follow-up rate). This will translate into total protein consumption in the range of 30–3000 mg of purified protein. Abbreviations: Cpds = compound, DRC = dose–response curve, Pt = point.

Alternatively, periplasmic co-expression of the Skp and FkpA chaperones successfully refolded a recombinant antibody [5].

Exploitation of the secretory pathways in *E. coli* may result in over burdening the translocation complexes, leading to the accumulation of recombinant protein, inhibition of the translocase machinery and poorly folded protein. Increasing the levels of the translocation complexes could overcome this hurdle but robust strategies are yet to be developed [6].

An alternative approach is to alter the cytoplasmic reducing environment to a more oxidising one, favourable for disulphide formation, for example, through deletion of the thioredoxin reductase (*trx*B) and glutathione reductase (*gor*) genes, which leads to an accumulation of oxidised thioredoxin (Trx) and glutaredoxin (Grx). In addition, co-expression of DsbS and PDI has also been described to increase the yield of disulphide containing proteins. Such genetically engineered strains have been commercialised (Origami™ (and the Rosetta™ family) and SHufle™ [7]), although Origami™ strains have a compromised cellular metabolism that requires further mutational rescue [8]. A complimentary strategy has been described recently, where the Ev1p sulphhydryl oxidase and disulphide bond isomerases are pre-expressed, and has proven to be an innovative and robust technique for the production of disulphide proteins in the cytoplasm of *E. coli* [9,10*].

Non-biological routes have also been used to enhance disulphide bond formation, which include the use of osmolytes and chemical chaperones [2]. It is hoped that a correlation between the stabilising effects of different chemical properties of these non-biological chaperones can be matched with molecular properties of proteins and polypeptides [11].

Finally, another area that has gained attraction over the past years is glycoengineering in *E. coli*, fuelled in large part by the discovery of the *Campylobacter jejuni* N-glycosylation...
system, and its transfer to *E. coli* [12], allowing simple glycosylation of proteins. Although the technological advances in this area are impressive, the low glycosylation efficiency and potential need for *in vitro* processing of the glycans [13,14] remain significant hurdles that need to be overcome for a wider application in protein production.

**High hydrostatic pressure refolding (HHP)**

It is beyond the scope of this paper to provide a full overview of refolding strategies, covered in depth by others [15]. Instead, we focus here on an approach that has been gaining momentum over the past several years, namely HHP refolding. An increasing number of publications demonstrate the utility of HHP refolding in cases where traditional refolding strategies have either failed or produced lower yields [16,17].

HHP refolding techniques were first outlined and later confirmed with enolase in the 1980’s [18,19]. Since then several groups have reported successful HHP refolding of a variety of protein targets, including VEGF with multiple disulfide bonds (including cysteine knots) [20] and heavily glycosylated protein inclusion bodies (IBs) from mammalian cell pellets [21].

At room temperature native proteins undergo denaturation at pressures beyond 500 MPa, whilst pressures between 100–300 MPa lead only to protein disaggregation [22]. HHP disrupts ionic and hydrophobic interactions without affecting hydrogen bond interactions; consequently, the need for high concentrations of chaotropic reagents to dissociate aggregates is not required. HHP refolding of IBs at pressures between 100 and 300 MPa is thermodynamically disadvantageous to the re-formation of insoluble aggregates while favouring and stabilising the formation of native folds. Yet, the requirement to optimise both physical parameters (protein concentration, time, temperature and pressure) and buffer composition (pH, redox reagents, disaggregation additives) remains.

Taken together, this technique shows high promise with the potential to become routinely applied in protein expression laboratories.

**High cell density fermentation**

The ability to grow laboratory scale high density cultures in shake flasks, particularly for NMR labeling studies, has great utility. For a review of high cell density *E. coli* fermentation using bio-reactors readers are directed to reference [23].

Auto-induction strategies and media compositions [24] have allowed lab scale *E. coli* growth to achieve much higher densities than possible using standard rich culture media. OD<sub>600</sub> levels of 10–20 are routinely obtained [24,25], and while lower than densities achieved in bio-reactor cultures the two processes share the same problems associated with high density growth: Plasmid loss, oxygen transfer limitations, increased CO<sub>2</sub> and the lowering of pH caused by CO<sub>2</sub> and acetate production. Some of these limitations have been resolved, for instance the use of Ultra Yield Flasks™ compared to standard Erlenmeyer flasks, which has shown to confer a fourfold higher oxygen transfer rate, whilst maintaining pH values [26]. The use of Ultra Yield Flasks™ in concert with an enzyme-based glucose delivery system (Enbase®) by the same group demonstrated densities comparable to bio-reactor levels (OD<sub>600</sub> levels of 70–80), while achieving higher soluble yields and activity compared to TB cultures where much of the protein was expressed as inclusion bodies [26]. The high densities achieved also allow the screening of Enbase® cultures in 96 well blocks using 500 µl volumes. The oxygen limitation is overcome by the addition of oxygen-saturated liquid perfluorodecalin [27].

A simple defined auto-induction medium that yields high cell densities and recombinant protein expression has been developed by the group of Rinas [25]. As the medium requires no complex medium compounds, amino acids or vitamin supplements and is fully defined, it is suitable for protein labeling, with reported yields of 500 mg L<sup>-1</sup>.

Sivashanmugam and colleagues have conducted extensive studies in finding optimal IPTG inducible media compositions and conditions for high density cultures [28], achieving approximately 15–35 mg of NMR triple-labeled and unlabeled proteins from a 50 mL cell culture for seven test proteins. Their screening protocols ensured high density cultures at optimal growth phases without the risk of plasmid loss.

**Protein production in insect cells: the Baculovirus Expression Vector System (BEVS)**

Baculovirus (BV) technology has matured to a degree of routine implementation in many labs for recombinant protein production, and among the eukaryotic expression systems it leads the way in producing diffraction quality proteins (Figure 2). Key to this development has been the simplification of the generation of recombinant BVs by *in vivo* transposition in *E. coli* [29] (commercialised as the Bac-to-Bac<sup>®</sup> system), or, more versatile and less process intensive, *in vivo* recombination in insect cells [30,31] (commercialised as the Flashbac<sup>®</sup> system). Other important factors are the ease of culturing and infecting insect cells, and the high yields obtainable [32,33]. In addition, the insect cell environment recapitulates many features of mammalian cells, such as post-translational modifications and similar folding machinery. BV expression has proven particularly powerful in the field of membrane structural biology, contributing to many of the recently solved X-ray structures of membrane proteins [34–43]. In addition, the
New constructs and expressions of proteins

Impact of eukaryotic expression systems on structural biology. The graph shows the number of new, unique entries added to the Protein Databank (PDB) per year (unique is defined here as entries with ~90% sequence identity). Note that S. frugiperda equates to baculovirus expression, whilst H. sapiens corresponds primarily to HEK293 cells (PDB was searched using Homo sapiens as a search term for expression system, but no attempt was made to quantify the exact number of entries using HEK293 cells specifically). For a more comprehensive comparison of the impact of different expression hosts on structural biology the reader is referred to: Fernandez & Vega: Technologies to keep an eye on: alternative hosts for protein production in structural biology. Curr Opin Struct Biol 2013, http://dx.doi.org/10.1016/j.sbi.2013.02.002.

The overall process of over-expressing heterologous proteins using BV technology involves (co-)transfection of insect cells, followed by one or two rounds of virus amplification. Hence, the initial timeline compares less favourably with bacterial expression, but this is often compensated by higher yields and an improvement in protein quality. In addition, the recombinant virus is stable for approximately six months, allowing for rapid repeated rounds of production [47]. Alternatively, frozen baculovirus-infected insect cells (BICs) can be used for rapid scale-up in a highly reproducible manner [48], circumventing prolonged storage of virus stocks. To select optimal expression constructs before embarking on large scale protein production using BVs, screening by transient expression in insect or mammalian cells can likewise be employed. Chen et al. and Kato et al. used this approach coupled to fluorescence-detection size exclusion chromatography (FSEC) to optimise expression conditions of membrane proteins [37,49].

While of great utility, baculovirus technology has a number of limitations. For example, the lytic infection process results in the release of host and viral proteases which may impact on protein quality. Certain viral proteins (e.g. v-cathepsin and v-chitinase) may impede the secretory pathway [50,51], and repeated rounds of virus amplification tend to produce defective interfering particles with concomitant reduction in protein yields [52–54]. The functional and biochemical properties of the recombinant protein can likewise be affected by an insect cell-derived glycosylation pattern. Incomplete processing of larger precursor proteins (e.g. peptide hormones) is another phenomenon observed and likely due to a lack in pro-convertase activity in infected insect cells [55]. Interestingly, impaired secretion has been shown to yield diffraction quality grade in vivo protein crystals, although general exploitation of this approach awaits further studies [56,57]. Given these shortcomings, an increasing number of studies have focused on the elimination of baculovirus genes detrimental to the quality and quantity of the protein produced, or conversely, introduction of foreign proteins to improve the latter. For example, v-cathepsin and v-chitinase deletion has been shown to improve the production of secreted proteins [50,51], more recently deletion of VP80 was used to eliminate the presence of viral particles in the medium, with a beneficial effect on downstream purification of proteins [58**]. Interestingly, ORF34 was recently identified as an essential baculovirus gene, yet siRNA-mediated knock-down nevertheless improved recombinant protein expression, challenging the idea that essential genes should not be manipulated in BV optimisation approaches [59]. Co-expression with chaperones has been shown to aid proper folding of the protein of interest [60–63], although unfolded proteins not amenable to chaperone assisted folding may nevertheless be bound to the chaperones and end up in the soluble fraction with little chance of recovery. Co-expression with glycan trimming enzymes such as endoglycosidase H, in combination with the glycosyltransferase inhibitor kifunensine, has been utilized to produce secreted, crystal grade protein complexes [64*]. In vivo biotinylation of proteins, for example for SPR studies, can be achieved through co-expression of Avi-tagged proteins with BirA ligase [65,66]. Glycoengineering of either the baculovirus genome itself or its host, in order to produce more human-like sugar modifications of secreted proteins, has been described [67,68,69*]. Finally, transduction of mammalian cells with recombinant baculovirus, carrying genes under the control of a mammalian promoter, can apart from assay development,
also be utilized for producing crystal grade proteins [70–73] although its use in large scale protein production has not widely been reported.

The increased knowledge on the function of the baculovirus genes and the role of insect host proteins in the protein production process, coupled to advances in genetic recombination tools, collectively open the possibility to design more streamlined baculovirus and engineered cell lines tailored to specific protein production needs [74].

**Transient and stable expression of recombinant proteins in mammalian cells**
For research purposes, the trend to employ mammalian cells as expression host is intimately linked to the increasing complexity of targets and the need to express and screen a large number of candidate proteins. Yet, the tedious establishment of stable cell lines for production is often not necessary. More than 10 years of transient gene expression (TGE) experiments resulted in thoroughly investigated and well-established transient processes amenable to high throughput, partial automation, and scale-up to >100 L giving rise to gram quantities of protein. In a nutshell, key achievements in this context cover suspension cultivation of cell lines in serum-free, protein-free or chemically defined media allowing for efficient protein capture, high density transfection, cheap transfection reagents such as polyethylenimine (PEI), various additives to boost expression rates, and the availability of disposable bioreactors in different formats [75–78]. The most frequently used cell line(s) in TGE approaches comprise HEK293 derivatives and CHO cells, but also novel cell lines such as CAP-T [79,80] and avian cells [81] are gaining increasingly recognition.

While production of an ‘authentic’ protein is extremely favourable for biologics and screening applications, the inherent inhomogeneity of mammalian-derived proteins, mainly due to glycosylation poses challenges to crystallography and structure elucidation. A small number of glycosylated molecules could be successfully crystallized and gave rise to structures, but a homogenous protein preparation is still considered as advantageous.

This can be achieved in several ways, for instance by genetically modifying the predicted glycosylation sites before expression of the gene — an approach applied to production of α-glycosylated antibodies exhibiting enhanced effector functions [82,83]. Glycosylation pathway-engineered cell lines, such as the lectin-resistant CHO line Lec3.2.8.1 cells [84,85] and the HEK293S cell lines Lec1 (GnTII deficient; [86]) and Lec36 (GnTIV deficient; [87]) were established several years ago. In combination with Flp recombinase mediated cassette exchange (RMCE) into a pre-defined, high-expressing locus in Lec3.2.8.1 cells, this approach represents an elegant way for multiple construct screening [88]. Yet, some proteins require native glycosylation for correct folding, and moreover, as these modifications impact the entire cell metabolism, growth, viability and productivity can be negatively affected [89]. Finally, the glycosylation pattern of the protein can be chemically modified by addition of α-mannosidase inhibitors to the cell culture, such as kifunensine and swainsonine, resulting in high-mannose type glycoforms easily trimmed by endoglycosidase (EndoH) treatment [89].

Novel expression formats for complex targets, for example, the presentation of receptors on virus-like particles (VLPs) derived from mammalian, insect cells [90] or (engineered) exosomes [91,92] and new membrane protein scaffolds [93] are under development. This, as well as the emerging insight into the impact of glycans on protein conformation [94] calls for new innovative strategies in not only obtaining structural information employing traditional X-ray crystallography and NMR but also imaging techniques and mass spectrometry [95,96,97].

**Protein labeling approaches for biophysical studies**
Biophysical characterisation, for example for NMR or X-ray studies of protein-drug interactions, often necessitates the use of isotopic or other amino acid labels. For solving the phase problem in X-ray crystallography, selenomethionine incorporation can be achieved efficiently in *E. coli* (e.g. through use of methionine auxotrophic B834 strains as reported by Pieper et al. [98]), although cell toxicity and lower yields remain a significant problem in eukaryotic hosts [99,100]. Uniform isotopic labeling for NMR studies can be achieved in eukaryotic cells, but for large scale production the cost of goods is prohibitively high, and isotopic incorporation efficiency and yields are often sub-optimal [101,102,103,104]. For achieving uniform labeling *E. coli* or yeast therefore remain the preferred expression hosts, in part due to advances in high density fermentation (see above) and greater versatility in labeling approaches [105]. Another interesting, recent development has been the single protein production (SPP) system in *E. coli*, which in essence is a genetic adaptation allowing the production of only a single protein of interest. The method exploits the use of MazF, an mRNA interferase, which arrests growth through cleavage of mRNAs at ACA nucleotides. Thus, by engineering the gene of interest to be devoid of ACA base triplets, only the protein of interest is produced in the presence of MazF [106]. The method has, for example, been used to produce isocholesterol enriched membrane proteins for NMR studies, with substantial cost savings [106,107].

In contrast to uniform labeling, amino acid type selective (AATS) isocholesterol labeling is efficient and cost-effective in insect and mammalian cells and leads to less metabolic scrambling (transfer of the isotope to a different amino acid)
New constructs and expressions of proteins

Acid) compared to E. coli [101,108,109]. The past several years have also seen advances in the fields of segmental labeling, allowing NMR studies of large, multidomain proteins [110], and in particular the unnatural amino acid (UAA) incorporation into polypeptide chains, to facilitate selective labeling of proteins using bio-orthogonal chemistries [111,112]. Such methods can facilitate introduction of NMR active labels at specific sites in the protein [113], and study receptor-ligand interactions [114,115]. Although initially limited to prokaryotic systems, recent developments have enabled incorporation of UAs in eukaryotic systems and have yielded improved bio-orthogonal chemistries compatible with living cells [49,116–120] promising further advances in this area.

Synthetic biology and engineering of open reading frames (ORFs)

An active field concerns the so-called optimisation of gene sequences in order to improve recombinant protein expression. Such approaches utilize (often proprietary) mathematical algorithms aimed at the optimisation of a diverse range of sequence parameters in order to boost protein expression. These parameters go beyond host codon usage and GC%, although these usually have a prominent role, but in addition include factors such as 5′ end mRNA secondary structures, mRNA levels, transcriptional speed and pausing, cryptic transcriptional terminators, initiation and translation efficiency. However, it is increasingly recognized that in cells transcription, translation and protein folding are dictated by codon bias and tRNA abundance in ways not fully understood [121]. Studies on rare codons have, for example, shown that in some instances the re-design of ORFs using rare codons actually increased expression levels [122].

A particularly striking example concerns the circadian clock proteins of Synechococcus elongatus (KaiA, KaiB and KaiC) and Neurospora crassa (FRQ), which are specifically codon biased to produce less protein. Two recent studies showed that in both cases artificial codon optimisation, although increasing protein yields, negatively affected the clock function of these proteins [123,124]. Strikingly, the FRQ yields led to adverse conformational changes, loss of stability and impeded circadian rhythms [124]. This example illustrates the increasing recognition that codon usage is often intimately linked to the folding and function of a particular protein as both mRNA structure and distribution of rare and frequent codons can have a strong effect on the co-translational folding efficiency of the protein as it emerges from the ribosome tunnel [121,125–130]. Hence, as the FRQ example and other studies have shown, perturbations in the codon distribution can lead to adverse effects [121,124,126,131,132]. This implies that the generally accepted dogma of avoiding rare codons in gene synthesis needs to be re-evaluated. Moreover, it seems likely that ‘gene optimality’, in the sense that it delivers high yields of properly folded protein, may well be protein-dependent, or perhaps protein family dependent, given the link with (co-translational) protein folding. Current in silico gene synthesis algorithms thus cannot cover all parameters that govern the processes underpinning protein expression [133,134]. Conceptually, the systematic use of synonymous substitutions in an ORF would provide valuable information for the generation of synthetic genes, but such undertakings are impractical due to the sheer number of variables in searchable ORF engineering space (for a 100 amino acid protein there are $3 \times 10^{100}$ ways to encode the same protein sequence [135]). Nevertheless, statistical approaches such as Design of Experiment (DoE) are starting to make an impact on the (commercial) design of synthetic genes [136] offering the hope that synthetic gene synthesis may eventually lead to predictable protein expression levels.

Summary

In summary, in the pharmaceutical industry, aiming at the discovery of proprietary targets and development of molecules for therapeutic intervention, a repertoire of various methods for the generation of drug targets and their characterisation by structure and function is indispensable. We see significant advances in the heterologous expression of target proteins in the three main expression systems, bacterial expression in E. coli, insect cell expression in the BEVS system and finally expression in mammalian cells. These novel developments have enabled a steady increase in the yield/liter of recombinantly expressed proteins, shorter timelines, higher quality and overall higher success rates for expression of proteins in a functionally active state suitable for downstream applications.

The strong increase in interest for protein expression as therapeutics, the need to express proteins in large amounts for structural and biophysical studies, the need for protein expression in the form of multi-component protein complexes, and finally the need to work on more demanding protein expression projects with difficult to produce proteins will further accelerate the development of novel expression technologies.

Acknowledgements

The authors would like to acknowledge fruitful discussions and valuable input from various colleagues at AstraZeneca and Novartis. This work was performed at the Department ‘Reagents & Assay Development’ (RAD) at AstraZeneca at Alderley Park (UK) & Mölndal (Sweden) and at the Department ‘Expertise Platform Proteases’ (EPP) and ‘Novartis Biologies Center’ (NBC) in Basel, Switzerland.

We are very grateful for critical reading and helpful comments by Dr. Imre Berger (EMBL, Grenoble), Dr. Nick Dekker (AstraZeneca, Mölndal), Dr. Bruno Martoglio (Novartis, Basel), Dr. Ross Oveman (AstraZeneca, UK), Dr. Thomas Pietzonka (Novartis, Basel), Dr. Paul Ramage (Novartis, Basel) and Dr. Robert Roth (AstraZeneca, Mölndal).
References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- outstanding interest


Pre-expression by arabinose induction of Erv1p and a catalyst of disulfide bond isomerization shows a 800 fold greater yield of the 9 disulfide bond containing eukaryote protein vTPA. These results open up new opportunities for the production of high yield complex disulfide containing proteins in E. coli.


An excellent overview of the current state-of-the-art in the baculovirus expression field.
8 New constructs and expressions of proteins


An innovative approach to generate more mammalian-like glycosylation of recombinant proteins in insect cells, by introducing a combination of C. elegans and bovine glycosyltransferases into the Baculovirus genome.

This very recent article elegantly summarizes the most recent advances in the generation of glycosylation mutants in CHO cells by lectin-selection and zinc finger nucleases mediated knockout of genes and their applications.
All you ever wanted to know about VLPs: a comprehensive review on the different ways of VLP generation, their characteristics and applications, not only in the context of vaccine development.
A very interesting overview on the potential of in-cell NMR spectroscopy, how it evolved and its future.
100. Nettleship JE, Assenberg R, Diprose JM, Rahman-Huq N, Owens RJ: Recent advances in the production of proteins in
New constructs and expressions of proteins


A comprehensive overview of current approaches in isotope labeling in insect cells.


An authoritative review of the current approaches and applications in the field of unnatural amino acids.


Excellent review on the complexity of the relation between codon bias and protein expression.


Quantitatively describes multiple variants that can dictate the three phases of ribosomal function, initiation, elongation and termination. Good correlation found between presence of high AU content and low secondary structure in the ORF 5’ region.

