ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Advances in recombinant protein expression for use in pharmaceutical research

Rene Assenberg[1], Paul T Wan[2], Sabine Geisse[3] and Lorenz M Mayr[4]

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, Fabrikstrasse-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 corner stone 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 disulfide bond formation, such as disulphide isomerases (Dsb proteins) and peptidyl-prolyl isomerases (PPIase) [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 (PDIs 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].

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

Figure 1

Detection Readout	Comment	Information	Protein requirement (per data point)
NMR (Nuclear Magnetic Resonance)	In solution	Binding, K_D, cpd integrity, dynamics	**10-500 μg**
Affinity MS (Mass Spectrometry)	In solution	Binding	**1-5 μg**
ITC (Isothermal Titration Calorimetry)	In solution	Binding, K_D, stoichiometry, ΔH, ΔS	**200-5000 μg**
DSC (Differential Scanning Calorimetry)	In solution	ΔT_m, ΔH, ΔS	**100-500 μg**
DSF (Differential Scanning Fluorimetry; aka *Thermofluor*™)	In solution with tracer dye	ΔT_m	**0.5-5 μg**
DSLS (Differential Static Light Scattering; aka *Stargazer*™)	In solution	ΔT_{agg}	**1-10 μg**
SPR (Surface Plamon Resonance aka *Biacore*™)	Immobilized	Binding, K_D, k_{on}, k_{off}, "stoichiometry"	**0.001-0.1 μg**
RWG (Resonant Waveguide aka Corning *Epic*™ or SRU *BIND*™)	Immobilized	Binding, K_D, "stoichiometry"	**0.5-2 μg**
BLI (Biolayer Interferometry aka *Fortebio*™)	Immobilized	Binding, K_D, k_{on}, k_{off}	**0.2-1 μg**

Current Opinion in Structural Biology

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 (25 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: Cpd = 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 (*trxB*) and glutathione reductase (*gor*) genes, which leads to an accumulation of oxidised thioredoxin (Trx) and glutaredoxin (Grx). In addition, co-expression of Dsbs and PDIs 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 SHuffle™ [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 sulfhydryl 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

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

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 disulphide 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_{600} 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_2 and the lowering of pH caused by CO_2 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_{600} 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^{-1}.

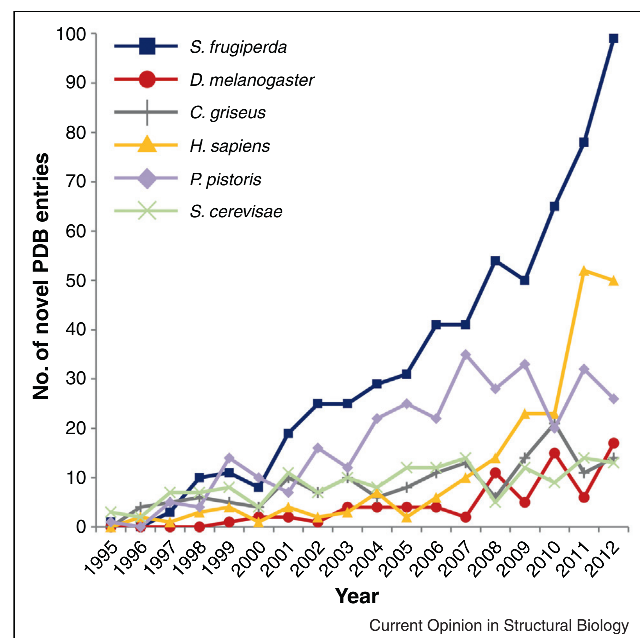
Sivashanmugam and colleagues have conducted extensive studies in finding optimal IPTG inducible media compositions and conditions for high density cultures [28], achieving approximately 15–35 mg of NMR triple-labeled and unlabeled proteins from a 50 mL cell culture for seven test proteins. Their screening 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® system), or, more versatile and less process intensive, *in vivo* recombination in insect cells [30,31] (commercialised as the Flashbac® 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

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

Figure 2



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.

development of efficient recombineering (recombination-mediated genetic engineering) tools such as ACEMBL have paved the way to improved production of protein complexes, but are also increasingly applied to engineer the BV genome to aid protein production (see below) [44–46] (Berger et al., this issue).

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 (BIICs) 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 proconvertase 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 knockdown 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,

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

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 (GnTI deficient; [86]) and Lec36 (GnTII 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 isotope 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 isotope enriched membrane proteins for NMR studies, with substantial cost savings [106,107].

In contrast to uniform labeling, amino acid type selective (AATS) isotope 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

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

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 UAAs 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 overexpression 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 increased 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 3×10^{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 Biologics Center' (NBC) in Basel, Switzerland.

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

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Merdanovic M, Clausen T, Kaiser M, Huber R, Ehrmann M: **Protein**
• **quality control in the bacterial periplasm**. *Annu Rev Microbiol* 2011, **65**:149-168.

Review article discussing recent advances in the identification and characterisation of the major cellular factors that are involved in Gram-negative bacteria host multiple signal transduction pathways that sense and respond to protein-folding stresses. It includes mechanistic models explaining how various stress pathways sense and coordinate the responses to the presence of misfolded proteins.

2. de Marco A: **Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli***. *Microb Cell Fact* 2009, **8**:26.

3. Grudnik P, Bange G, Sinning I: **Protein targeting by the signal recognition particle**. *Biol Chem* 2009, **390**:775-782.

4. Kim SG, Kweon DH, Lee DH, Park YC, Seo JH: **Coexpression of folding accessory proteins for production of active cyclodextrin glycosyltransferase of *Bacillus macerans* in recombinant *Escherichia coli***. *Protein Expr Purif* 2005, **41**:426-432.

5. Ow DS, Lim DY, Nissom PM, Camattari A, Wong VV: **Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production**. *Microb Cell Fact* 2010, **9**:22.

6. Choi JH, Lee SY: **Secretory and extracellular production of recombinant proteins using *Escherichia coli***. *Appl Microbiol Biotechnol* 2004, **64**:625-635.

7. Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M: **SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm**. *Microb Cell Fact* 2012, **11**:56.

8. Yamamoto Y, Ritz D, Planson AG, Jonsson TJ, Faulkner MJ, Boyd D, Beckwith J, Poole LB: **Mutant AhpC peroxiredoxins suppress thiol-disulfide redox deficiencies and acquire deglutathionylating activity**. *Mol Cell* 2008, **29**:36-45.

9. Hatahet F, Nguyen VD, Salo KE, Ruddock LW: **Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli***. *Microb Cell Fact* 2010, **9**:67.

10. Nguyen VD, Hatahet F, Salo KE, Enlund E, Zhang C, Ruddock LW:
• **Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of *E. coli***. *Microb Cell Fact* 2011, **10**:1.

Pre-expression by arabinose induction of Erv1p and a catalyst of disulphide 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*.

11. Bandyopadhyay A, Saxena K, Kasturia N, Dalal V, Bhatt N, Rajkumar A, Maity S, Sengupta S, Chakraborty K: **Chemical chaperones assist intracellular folding to buffer mutational variations**. *Nat Chem Biol* 2012, **8**:238-245.

12. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW *et al.*: **N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli***. *Science* 2002, **298**:1790-1793.

13. Schwarz F, Huang W, Li C, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang LX: **A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation**. *Nat Chem Biol* 2010, **6**:264-266.

14. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, Heiss C, Azadi P, Aebi M, DeLisa MP: **An engineered eukaryotic protein glycosylation pathway in *Escherichia coli***. *Nat Chem Biol* 2012, **8**:434-436.

15. Singh SM, Panda AK: **Solubilization and refolding of bacterial inclusion body proteins**. *J Biosci Bioeng* 2005, **99**:303-310.

16. Balduino KN, Spencer PJ, Malavasi NV, Chura-Chambi RM, Lemke LS, Morganti L: **Refolding by high pressure of a toxin containing seven disulfide bonds: bothropstoxin-1 from *Bothrops jararacussu***. *Mol Biotechnol* 2011, **48**:228-234.

17. Okai M, Ohtsuka J, Asano A, Guo L, Miyakawa T, Miyazono K, Nakamura A, Okada A, Zheng H, Kimura K *et al.*: **High pressure refolding, purification, and crystallization of flavin reductase from *Sulfolobus tokodaii* strain 7**. *Protein Expr Purif* 2012, **84**:214-218.

18. Kornblatt J, Hui Bon Hoa G: **The pressure-induced, reversible inactivation of mouse brain enolases**. *Eur J Biochem* 1982, **128**:577-581.

19. Paladini AA Jr, Weber G: **Pressure-induced reversible dissociation of enolase**. *Biochemistry* 1981, **20**:2587-2593.

20. Cothran A, St John RJ, Schmelzer CH, Pizarro SA: **High-pressure refolding of human vascular endothelial growth factor (VEGF) recombinantly expressed in bacterial inclusion bodies: refolding optimization, and feasibility assessment**. *Biotechnol Prog* 2011, **27**:1273-1281.

21. Seefeldt MB, Ouyang J, Froland WA, Carpenter JF, Randolph TW: **High-pressure refolding of bikunin: efficacy and thermodynamics**. *Protein Sci* 2004, **13**:2639-2650.

22. Qoronfleh MW, Hesterberg LK, Seefeldt MB: **Confronting high-throughput protein refolding using high pressure and solution screens**. *Protein Expr Purif* 2007, **55**:209-224.

23. Huang CJ, Lin H, Yang X: **Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements**. *J Ind Microbiol Biotechnol* 2012, **39**:383-399.

24. Studier FW: **Protein production by auto-induction in high density shaking cultures**. *Protein Expr Purif* 2005, **41**:207-234.

25. Li Z, Kessler W, van den Heuvel J, Rinas U: **Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems**. *Appl Microbiol Biotechnol* 2011, **91**:1203-1213.

26. Ukkonen K, Vasala A, Ojamo H, Neubauer P: **High-yield production of biologically active recombinant protein in shake flask culture by combination of enzyme-based glucose delivery and increased oxygen transfer**. *Microb Cell Fact* 2011, **10**:107.

27. Pilarek M, Brand E, Hillig F, Krause M, Neubauer P: **Enhanced plasmid production in miniaturized high-cell-density cultures of *Escherichia coli* supported with perfluorinated oxygen carrier**. *Bioprocess Biosyst Eng* 2012 http://dx.doi.org/10.1007/s00449-012-0861-7.

28. Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q: **Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli***. *Protein Sci* 2009, **18**:936-948.

29. Luckow VA, Lee SC, Barry GF, Olins PO: **Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli***. *J Virol* 1993, **67**:4566-4579.

30. Kitts PA, Possee RD: **A method for producing recombinant baculovirus expression vectors at high frequency**. *Biotechniques* 1993, **14**:810-817.

31. Zhao Y, Chapman DA, Jones IM: **Improving baculovirus recombination**. *Nucleic Acids Res* 2003, **31** E6–E6.

32. Summers MD: **Milestones leading to the genetic engineering of baculoviruses as expression vector systems and viral pesticides**. *Adv Virus Res* 2006, **68**:3-73.

33. van Oers MM: **Opportunities and challenges for the baculovirus**
•• **expression system**. *J Invertebr Pathol* 2011, **107 (Suppl)**:S3-S15.

An excellent overview of the current state-of-the-art in the baculovirus expression field.

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

34. Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK: **Structure of the delta-opioid receptor bound to naltrindole**. *Nature* 2012, **485**:400-404.

35. Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T *et al.*: **Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist**. *Nature* 2012, **482**:547-551.

36. Hattori M, Gouaux E: **Molecular mechanism of ATP binding and ion channel activation in P2X receptors**. *Nature* 2012, **485**:207-212.

37. Kato K: **Structural biology of post-translational modifications of proteins**. *Yakugaku Zasshi* 2012, **132**:563-573.

38. Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO *et al.*: **Structure and dynamics of the M3 muscarinic acetylcholine receptor**. *Nature* 2012, **482**:552-556.

39. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S: **Crystal structure of the micro-opioid receptor bound to a morphinan antagonist**. *Nature* 2012, **485**:321-326.

40. Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D *et al.*: **Crystal structure of the beta2 adrenergic receptor–Gs protein complex**. *Nature* 2011, **477**:549-555.

41. White JF, Noinaj N, Shibata Y, Love J, Kloss B, Xu F, Gvozdenovic-Jeremic J, Shah P, Shiloach J, Tate CG *et al.*: **Structure of the agonist-bound neurotensin receptor**. *Nature* 2012, **490**:508-513.

42. Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang XP, Carroll FI *et al.*: **Structure of the human kappa-opioid receptor in complex with JDTic**. *Nature* 2012, **485**:327-332.

43. Zhang C, Srinivasan Y, Arlow DH, Fung JJ, Palmer D, Zheng Y, Green HF, Pandey A, Dror RO, Shaw DE *et al.*: **High-resolution crystal structure of human protease-activated receptor 1**. *Nature* 2012, **492**:387-392.

44. Berger I, Fitzgerald DJ, Richmond TJ: **Baculovirus expression system for heterologous multiprotein complexes**. *Nat Biotechnol* 2004, **22**:1583-1587.

45. Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I: **Protein complex expression by using multigene baculoviral vectors**. *Nat Methods* 2006, **3**:1021-1032.

46. Noad RJ, Stewart M, Boyce M, Celma CC, Willison KR, Roy P: **Multigene expression of protein complexes by iterative modification of genomic Bacmid DNA**. *BMC Mol Biol* 2009, **10**:87.

47. Geisse S: **Insect Cell Cultivation and Generation of Recombinant Baculovirus Particles for Recombinant Protein Production**. In *Animal Cell Biotechnology: Methods and Protocols*, vol. 24. Edited by Pörtner R. Human Press Inc.; 2007:489-507.

48. Wasilko DJ, Lee SE, Stutzman-Engwall KJ, Reitz BA, Emmons TL, Mathis KJ, Bienkowski MJ, Tomasselli AG, Fischer HD: **The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus**. *Protein Expr Purif* 2009, **65**:122-132.

49. Chen H, Shaffer PL, Huang X, Rose PE: **Rapid screening of membrane protein expression in transiently transfected insect cells**. *Protein Expr Purif* 2013, **88**:134-142.

50. Hitchman RB, Possee RD, Siaterli E, Richards KS, Clayton AJ, Bird LE, Owens RJ, Carpentier DC, King FL, Danquah JO *et al.*: **Improved expression of secreted and membrane-targeted proteins in insect cells**. *Biotechnol Appl Biochem* 2010, **56**:85-93.

51. Kaba SA, Salcedo AM, Wafula PO, Vlak JM, van Oers MM: **Development of a chitinase and v-cathepsin negative bacmid for improved integrity of secreted recombinant proteins**. *J Virol Methods* 2004, **122**:113-118.

52. Kool M, Voncken JW, van Lier FL, Tramper J, Vlak JM: **Detection and analysis of *Autographa californica* nuclear polyhedrosis virus mutants with defective interfering properties**. *Virology* 1991, **183**:739-746.

53. Pijlman GP, van den Born E, Martens DE, Vlak JM: **Autographa californica baculoviruses with large genomic deletions are rapidly generated in infected insect cells**. *Virology* 2001, **283**:132-138.

54. Pijlman GP, van Schijndel JE, Vlak JM: **Spontaneous excision of BAC vector sequences from bacmid-derived baculovirus expression vectors upon passage in insect cells**. *J Gen Virol* 2003, **84**:2669-2678.

55. Cieplik M, Klenk HD, Garten W: **Identification and characterization of spodoptera frugiperda furin: a thermostable subtilisin-like endopeptidase**. *Biol Chem* 1998, **379**:1433-1440.

56. Koopmann R, Cupelli K, Redecke L, Nass K, Deponte DP, White TA, Stellato F, Rehders D, Liang M, Andreasson J *et al.*: **In vivo protein crystallization opens new routes in structural biology**. *Nat Methods* 2012, **9**:259-262.

57. • Redecke L, Nass K, DePonte DP, White TA, Rehders D, Barty A, Stellato F, Liang M, Barends TR, Boutet S *et al.*: **Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser**. *Science* 2013, **339**:227-230.
A remarkable observation and successful exploitation of *in vivo* protein crystallization. Intracellular crystallization of proteins has been observed in the past, but the success achieved here in deriving good quality diffraction data from such crystals argues for a more comprehensive analysis to ascertain if this approach can be exploited more generally for protein crystallization.

58. •• Marek M, van Oers MM, Devaraj FF, Vlak JM, Merten OW: **Engineering of baculovirus vectors for the manufacture of virion-free biopharmaceuticals**. *Biotechnol Bioeng* 2011, **108**:1056-1067.
An interesting approach to eliminate Baculovirus virions by deleting the essential structural gene VP80 from the Baculovirus genome, which prevents assembly of the virions. Virus stocks can be produced in special engineered Sf9 cells, encoding the missing VP80 ORF, whilst virion-free protein production can be achieved in normal Sf9 cells.

59. Salem TZ, Zhang F, Thiem SM: **Reduced expression of *Autographa californica* nucleopolyhedrovirus ORF34, an essential gene, enhances heterologous gene expression**. *Virology* 2013, **435**:225-238.

60. Higgins MK, Demir M, Tate CG: **Calnexin co-expression and the use of weaker promoters increase the expression of correctly assembled Shaker potassium channel in insect cells**. *Biochim Biophys Acta* 2003, **1610**:124-132.

61. Hong SM, Yamashita J, Mitsunobu H, Uchino K, Kobayashi I, Sezutsu H, Tamura T, Nakajima H, Miyagawa Y, Lee JM *et al.*: **Efficient soluble protein production on transgenic silkworms expressing cytoplasmic chaperones**. *Appl Microbiol Biotechnol* 2010, **87**:2147-2156.

62. Hsu TA, Eiden JJ, Betenbaugh MJ: **Engineering the assembly pathway of the baculovirus-insect cell expression system**. *Ann N Y Acad Sci* 1994, **721**:208-217.

63. Trometer C, Falson P: **Mammalian membrane protein expression in baculovirus-infected insect cells**. *Methods Mol Biol* 2010, **601**:105-117.

64. • Lupardus PJ, Birnbaum ME, Garcia KC: **Molecular basis for shared cytokine recognition revealed in the structure of an unusually high affinity complex between IL-13 and IL-13Ralpha2**. *Structure* 2010, **18**:332-342.
A nice approach to tackling the glycosylation heterogeneity, in order to improve crystallizability. It illustrates the ease in which co-expression can be achieved in insect cells, and the concomittant process simplifications this potentially allows. Here, co-infection of cells with 3 baculoviruses, two expressing interacting proteins and one expressing Endoglycosidase H, facilitated structure resolution of the protein complex.

65. Duffy S, Tsao KL, Waugh DS: **Site-specific, enzymatic biotinylation of recombinant proteins in *Spodoptera frugiperda* cells using biotin acceptor peptides**. *Anal Biochem* 1998, **262**:122-128.

66. Erdmann D, Zimmermann C, Fontana P, Hau JC, De Pover A, Chene P: **Simultaneous protein expression and modification:**

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

an efficient approach for production of unphosphorylated and biotinylated receptor tyrosine kinases by triple infection in the baculovirus expression system. *J Biomol Tech* 2010, **21**:9-17.

67. Aumiller JJ, Mabashi-Asazuma H, Hillar A, Shi X, Jarvis DL: **A new glycoengineered insect cell line with an inducibly mammalianized protein N-glycosylation pathway**. *Glycobiology* 2012, **22**:417-428.

68. Harrison RL, Jarvis DL: **Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins**. *Adv Virus Res* 2006, **68**:159-191.

69. • Palmberger D, Wilson IB, Berger I, Grabherr R, Rendic D: **SweetBac: a new approach for the production of mammalianised glycoproteins in insect cells**. *PLoS ONE* 2012, **7**:e34226.
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.

70. Dukkipati A, Park HH, Waghray D, Fischer S, Garcia KC: **BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies**. *Protein Expr Purif* 2008, **62**:160-170.

71. Hofmann C, Sandig V, Jennings G, Rudolph M, Schlag P, Strauss M: **Efficient gene transfer into human hepatocytes by baculovirus vectors**. *Proc Natl Acad Sci U S A* 1995, **92**:10099-10103.

72. Kost TA, Condreay JP, Ames RS: **Baculovirus gene delivery: a flexible assay development tool**. *Curr Gene Ther* 2010, **10**:168-173.

73. Scott MJ, Modha SS, Rhodes AD, Broadway NM, Hardwicke PI, Zhao HJ, Kennedy-Wilson KM, Sweitzer SM, Martin SL: **Efficient expression of secreted proteases via recombinant BacMam virus**. *Protein Expr Purif* 2007, **52**:104-116.

74. Vijayachandran LS, Thimiri Govinda Raj DB, Edelweiss E, Gupta K, Maier J, Gordeliy V, Fitzgerald DJ, Berger I: **Gene gymnastics: synthetic biology for baculovirus expression vector system engineering**. *Bioengineered* 2013, **4**:1-9.

75. Bollin F, Dechavanne V, Chevalet L: **Design of Experiment in CHO and HEK transient transfection condition optimization**. *Protein Expr Purif* 2011, **78**:61-68.

76. Durocher Y, Loignon M: **Process, vectors and engineered cell lines for enhanced large-scale transfection**. *US Patent 2011, US 2011/0039339 A1*.

77. Geisse S, Voedisch B: **Transient expression technologies: past, present, and future**. *Methods Mol Biol* 2012, **899**:203-219.

78. Pereira J, Rajendra Y, Baldi L, Hacker DL, Wurm FM: **Transient gene expression with CHO cells in conditioned medium: a study using TubeSpin(R) bioreactors**. *BMC Proc* 2011, **5 (Suppl 8)**:P38.

79. Fischer S, Charara N, Gerber A, Wolfel J, Schiedner G, Voedisch B, Geisse S: **Transient recombinant protein expression in a human amniocyte cell line: the CAP-T(R) cell system**. *Biotechnol Bioeng* 2012, **109**:2250-2261.

80. Schiedner G, Hertel S, Bialek C, Kewes H, Waschutza G, Volpers C: **Efficient and reproducible generation of high-expressing, stable human cell lines without need for antibiotic selection**. *BMC Biotechnol* 2008, **8**:13.

81. Brown SW, Mehtali M: **The avian EB66(R) cell line, application to vaccines, and therapeutic protein production**. *PDA J Pharm Sci Technol* 2010, **64**:419-425.

82. Hristodorov D, Fischer R, Joerissen H, Muller-Tiemann B, Apeler H, Linden L: **Generation and comparative characterization of glycosylated and aglycosylated human IgG1 antibodies**. *Mol Biotechnol* 2012, **53**:326-335.

83. • Zhang P, Chan KF, Haryadi R, Bardor M, Song Z: **CHO glycosylation mutants as potential host cells to produce therapeutic proteins with enhanced efficacy**. *Adv Biochem Eng Biotechnol* 2013, **131**:63-87.
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 nuclease mediated knockout of genes and their applications.

84. Patnaik SK, Stanley P: **Lectin-resistant CHO glycosylation mutants**. *Methods Enzymol* 2006, **416**:159-182.

85. Stanley P: **Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity**. *Mol Cell Biol* 1989, **9**:377-383.

86. Reeves PJ, Callewaert N, Contreras R, Khorana HG: **Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line**. *Proc Natl Acad Sci U S A* 2002, **99**:13419-13424.

87. Crispin M, Chang VT, Harvey DJ, Dwek RA, Evans EJ, Stuart DI, Jones EY, Lord JM, Spooner RA, Davis SJ: **A human embryonic kidney 293T cell line mutated at the Golgi alpha-mannosidase II locus**. *J Biol Chem* 2009, **284**:21684-21695.

88. Wilke S, Krausze J, Gossen M, Groebe L, Jager V, Gherardi E, van den Heuvel J, Bussow K: **Glycoprotein production for structure analysis with stable, glycosylation mutant CHO cell lines established by fluorescence-activated cell sorting**. *Protein Sci* 2010, **19**:1264-1271.

89. Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, Yu C, Boles KS, Evans EJ, Stuart DI *et al.*: **Glycoprotein structural genomics: solving the glycosylation problem**. *Structure* 2007, **15**:267-273.

90. • Zeltins A: **Construction and characterization of virus-like particles: a review**. *Mol Biotechnol* 2013, **53**:92-107.
All you ever wanted to know about VLiPs: a comprehensive review on the different ways of VLiP generation, their charactertitics and applications, not only in the contexrt of vaccine development.

91. Lattanzi L, Federico M: **A strategy of antigen incorporation into exosomes: comparing cross-presentation levels of antigens delivered by engineered exosomes and by lentiviral virus-like particles**. *Vaccine* 2012, **30**:7229-7237.

92. Sharma S, Gillespie BM, Palanisamy V, Gimzewski JK: **Quantitative nanostructural and single-molecule force spectroscopy biomolecular analysis of human-saliva-derived exosomes**. *Langmuir* 2011, **27**:14394-14400.

93. Serebryany E, Zhu GA, Yan EC: **Artificial membrane-like environments for in vitro studies of purified G-protein coupled receptors**. *Biochim Biophys Acta* 2012, **1818**:225-233.

94. Nagae M, Yamaguchi Y: **Function and 3D structure of the N-glycans on glycoproteins**. *Int J Mol Sci* 2012, **13**:8398-8429.

95. Coleman BM, Hanssen E, Lawson VA, Hill AF: **Prion-infected cells regulate the release of exosomes with distinct ultrastructural features**. *FASEB J* 2012, **26**:4160-4173.

96. Mulder AM, Carragher B, Towne V, Meng Y, Wang Y, Dieter L, Potter CS, Washabaugh MW, Sitrin RD, Zhao Q: **Toolbox for non-intrusive structural and functional analysis of recombinant VLP based vaccines: a case study with hepatitis B vaccine**. *PLoS ONE* 2012, **7**:e33235.

97. • Tochio H: **Watching protein structure at work in living cells using NMR spectroscopy**. *Curr Opin Chem Biol* 2012, **16**:609-613.
A very interesting overview on the potential of in-cell NMR spectroscopy, how it evolved and its future.

98. Pieper M, Betz M, Budisa N, Gomis-Ruth FX, Bode W, Tschesche H: **Expression, purification, characterization, and X-ray analysis of selenomethionine 215 variant of leukocyte collagenase**. *J Protein Chem* 1997, **16**:637-650.

99. Cronin CN, Lim KB, Rogers J: **Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells**. *Protein Sci* 2007, **16**:2023-2029.

100. Nettleship JE, Assenberg R, Diprose JM, Rahman-Huq N, Owens RJ: **Recent advances in the production of proteins in**

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008

insect and mammalian cells for structural biology. *J Struct Biol* 2010, **172**:55-65.

101. Dutta A, Saxena K, Schwalbe H, Klein-Seetharaman J: **Isotope labeling in mammalian cells**. *Methods Mol Biol* 2012, **831**:55-69.

102 •• Gossert AD, Jahnke W: **Isotope labeling in insect cells**. *Adv Exp Med Biol* 2012, **992**:179-196.
A comprehensive overview of current approaches in isotope labelling in insect cells.

103. Saxena K, Dutta A, Klein-Seetharaman J, Schwalbe H: **Isotope labeling in insect cells**. *Methods Mol Biol* 2012, **831**:37-54.

104. Strauss A, Bitsch F, Fendrich G, Graff P, Knecht R, Meyhack B, Jahnke W: **Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells**. *J Biomol NMR* 2005, **31**:343-349.

105. Verardi R, Traaseth NJ, Masterson LR, Vostrikov VV, Veglia G: **Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins**. *Adv Exp Med Biol* 2012, **992**:35-62.

106. Suzuki M, Zhang J, Liu M, Woychik NA, Inouye M: **Single protein production in living cells facilitated by an mRNA interferase**. *Mol Cell* 2005, **18**:253-261.

107. Mao L, Tang Y, Vaiphei ST, Shimazu T, Kim SG, Mani R, Fakhoury E, White E, Montelione GT, Inouye M: **Production of membrane proteins for NMR studies using the condensed single protein (cSPP) production system**. *J Struct Funct Genomics* 2009, **10**:281-289.

108. Gossert AD, Hinniger A, Gutmann S, Jahnke W, Strauss A, Fernandez C: **A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility**. *J Biomol NMR* 2011, **51**:449-456.

109. Strauss A, Bitsch F, Cutting B, Fendrich G, Graff P, Liebetanz J, Zurini M, Jahnke W: **Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies**. *J Biomol NMR* 2003, **26**:367-372.

110. Michel E, Skrisovska L, Wuthrich K, Allain FH: **Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology**. *Chembiochem* 2013, **14**:457-466.

111 • Davis L, Chin JW: **Designer proteins: applications of genetic code expansion in cell biology**. *Nat Rev Mol Cell Biol* 2012, **13**:168-182.
An authorative review of the current approaches and applications in the field of unnatural amino acids.

112. Young TS, Schultz PG: **Beyond the canonical 20 amino acids: expanding the genetic lexicon**. *J Biol Chem* 2010, **285**:11039-11044.

113. Li F, Shi P, Li J, Yang F, Wang T, Zhang W, Gao F, Ding W, Li D, Xiong Y *et al.*: **A genetically encoded F NMR probe for tyrosine phosphorylation**. *Angew Chem Int Ed Engl* 2013, **52**:3958-3962.

114. Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, Hu H, Brittain SM, Brinker A, Caldwell J, Bursulaya B *et al.*: **In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy**. *J Am Chem Soc* 2008, **130**:9268-9281.

115. Pless SA, Ahern CA: **Unnatural amino acids as probes of ligand-receptor interactions and their conformational consequences**. *Annu Rev Pharmacol Toxicol* 2013, **53**:211-229.

116. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW: **Expanding the genetic code of *Drosophila melanogaster***. *Nat Chem Biol* 2012, **8**:748-750.

117. Greiss S, Chin JW: **Expanding the genetic code of an animal**. *J Am Chem Soc* 2011, **133**:14196-14199.

118. Kim CH, Kang M, Kim HJ, Chatterjee A, Schultz PG: **Site-specific incorporation of epsilon-N-crotonyllysine into histones**. *Angew Chem Int Ed Engl* 2012, **51**:7246-7249.

119. Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW: **Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction**. *Nat Chem* 2012, **4**:298-304.

120. Yu Z, Pan Y, Wang Z, Wang J, Lin Q: **Genetically encoded cyclopropene directs rapid, photoclick-chemistry-mediated protein labeling in mammalian cells**. *Angew Chem Int Ed Engl* 2012, **51**:10600-10604.

121 •• Novoa EM, Ribas de Pouplana L: **Speeding with control: codon usage, tRNAs, and ribosomes**. *Trends Genet* 2012, **28**:574-581.
Excellent review on the complexity of the relation between codon bias and protein expression.

122. Wu G, Zheng Y, Qureshi I, Zin HT, Beck T, Bulka B, Freeland SJ: **SGDB: a database of synthetic genes re-designed for optimizing protein over-expression**. *Nucleic Acids Res* 2007, **35**:D76-D79.

123. Xu Y, Ma P, Shah P, Rokas A, Liu Y, Johnson CH: **Non-optimal codon usage is a mechanism to achieve circadian clock conditionality**. *Nature* 2013, **495**:116-120.

124. Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, Sachs MS, Liu Y: **Non-optimal codon usage affects expression, structure and function of clock protein FRQ**. *Nature* 2013, **495**:111-115.

125. Crombie T, Swaffield JC, Brown AJ: **Protein folding within the cell is influenced by controlled rates of polypeptide elongation**. *J Mol Biol* 1992, **228**:7-12.

126. Kramer G, Boehringer D, Ban N, Bukau B: **The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins**. *Nat Struct Mol Biol* 2009, **16**:589-597.

127. Pechmann S, Frydman J: **Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding**. *Nat Struct Mol Biol* 2013, **20**:237-243.

128. Pechmann S, Willmund F, Frydman J: **The ribosome as a hub for protein quality control**. *Mol Cell* 2013, **49**:411-421.

129. Thanaraj TA, Argos P: **Protein secondary structural types are differentially coded on messenger RNA**. *Protein Sci* 1996, **5**:1973-1983.

130. Zhang G, Hubalewska M, Ignatova Z: **Transient ribosomal attenuation coordinates protein synthesis and co-translational folding**. *Nat Struct Mol Biol* 2009, **16**:274-280.

131. Drummond DA, Wilke CO: **Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution**. *Cell* 2008, **134**:341-352.

132. Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM: **A "silent" polymorphism in the MDR1 gene changes substrate specificity**. *Science* 2007, **315**:525-528.

133 •• Allert M, Cox JC, Hellinga HW: **Multifactorial determinants of protein expression in prokaryotic open reading frames**. *J Mol Biol* 2010, **402**:905-918.
Quantatively describes multiple variants that can dictate the three phases of ribosomal function, initiation, elongation and termination. Good correlation found between prescence of high AU content and low secondary structure in the ORF 5′ region.

134. Gustafsson C, Govindarajan S, Minshull J: **Codon bias and heterologous protein expression**. *Trends Biotechnol* 2004, **22**:346-353.

135. Welch M, Villalobos A, Gustafsson C, Minshull J: **You're one in a googol: optimizing genes for protein expression**. *J R Soc Interface* 2009, **6 (Suppl 4)**:S467-S476.

136. Gustafsson C, Minshull J, Govindarajan S, Ness J, Villalobos A, Welch M: **Engineering genes for predictable protein expression**. *Protein Expr Purif* 2012, **83**:37-46.

Please cite this article in press as: Assenberg R, et al.: Advances in recombinant protein expression for use in pharmaceutical research, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.03.008