Hijacked then lost in translation: the plight of the recombinant host cell in membrane protein structural biology projects
Roslyn M Bill 1 and Tobias von der Haar 2

Membrane protein structural biology is critically dependent upon the supply of high-quality protein. Over the last few years, the value of crystallising biochemically characterised, recombinant targets that incorporate stabilising mutations has been established. Nonetheless, obtaining sufficient yields of many recombinant membrane proteins is still a major challenge. Solutions are now emerging based on an improved understanding of recombinant host cells; as a ‘cell factory’ each cell is tasked with managing limited resources to simultaneously balance its own growth demands with those imposed by an expression plasmid. This review examines emerging insights into the role of translation and protein folding in defining high-yielding recombinant membrane protein production in a range of host cells.

Addresses
1 School of Life & Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK
2 Kent Fungal Group, School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK

Corresponding author: Bill, Roslyn M (r.m.bill@aston.ac.uk)

Recombinant proteins are essential to progress in modern structural biology
Progress in soluble protein structural biology continues at a ‘breathless pace’ [1], while success in the membrane protein field lags behind [2]. However, recent advances in experimental approaches [3] mean that even the most recalcitrant membrane proteins are now tenable structural biology targets; this requires sufficient, stable recombinant protein for biochemical characterisation and crystallisation trials. A widely used strategy is illustrated in Figure 1, where the chosen host cell is regarded as a ‘cell factory’ that can be used to produce the target protein. This typically necessitates mutagenesis of the expression construct to increase protein yield [4] plus the incorporation of additional mutations to stabilise the resultant protein so it is more likely to crystallise [3]. These manipulations can be done pre-translationally by mutating the gene sequence (e.g. [5]) or fusing it with a stabilising partner (e.g. [6*]). Alternatively, the protein can be engineered post-translationally by deglycosylation, proteolysis or by making other chemical modifications [7]. The 3 Å crystal structure of the GABA_A receptor (Figure 2) was solved after approximately one-hundred construct variants of the full-length human β3 subunit were evaluated (in a stable mammalian cell-line [8]); the consequences were examined of N-linked glycosylation site removal, mutation of cysteine residues, amino-terminal and carboxy-terminal truncations, truncations in the intracellular loop connecting transmembrane helices 3 and 4 and introduction of T4 lysozyme to minimise loop flexibility [9*]. The structure of the TRPV1 channel at 3.4 Å was possible because recent developments in electron cryo-microscopy meant that absolute conformational homogeneity was not required and only small amounts of protein were needed; consequently a less extensive mutational strategy was necessary [10]. Even in this case, however, the recombinant TRPV1 variant (produced using a modified baculovirus system) was composed of amino acids 110–603 and 627–764, demonstrating the importance of prior biochemical characterisation [10]. Similar methodological approaches have been used for proteins produced in Escherichia coli [e.g. 11], Saccharomyces cerevisiae [e.g. 12] and Pichia pastoris [e.g. 13] (see also Figure 2, Table 1).

Understanding the host cell provides new opportunities to improve functional yields
A complementary, host–cell-centric approach to producing recombinant proteins for structural analysis focuses on maximising the functional yield of every cell; in this case, the philosophy is to understand the workings of the ‘cell factory’ (Figure 3). Any recombinant host cell must simultaneously balance its requirements for cellular growth with the metabolic burden imposed by the expression plasmid. Consequently, if it were possible to uncouple growth from recombinant synthesis in a host cell that remained metabolically capable of transcription and translation, this might enable metabolic fluxes to be entirely diverted to the production of a recombinant protein [14]. This concept has been demonstrated in E. coli to produce soluble chloramphenicol acetyltransferase to more than
Recombinant protein production for structural biology. Heterologous DNA is introduced into a host cell with the aim of producing a recombinant protein. This typically necessitates mutagenesis of the expression construct to increase protein yield plus the incorporation of additional mutations to stabilise the resultant protein so it is more likely to crystallise. These manipulations can be done pre-translationally by mutating the gene sequence or fusing it with a stabilising partner. Alternatively, the protein can be engineered post-translationally. In many cases, many tens or even hundreds of constructs are examined before proceeding to structural studies.

40% of total cell protein [15], but has not been widely adopted by structural biologists. A concept that has gained more traction focuses on the idea that the modulation of translation and protein folding may help to further improve host cells [16]. This review examines emerging insights into the role of these dependent pathways in defining high-yielding recombinant membrane protein production experiments.

**Post-transcriptional bottlenecks in recombinant protein production**

Many commonly used recombinant expression systems (Table 1) use strong promoters to drive high rates of mRNA synthesis (although this may be countered by high rates of mRNA degradation [17]). In the *E. coli* Walker strains C41(DE3) and C43(DE3) [18], acquired mutations were found to lower the efficiency of the T7 promoter leading to improved yields of membrane proteins for some but not all targets [19]. Similarly, in the *E. coli* MemStar system, the activity of T7 RNA polymerase was found to be modulated [20]. In a separate study, a series of evolved *E. coli* strains that had been selected based on their resistance to erythromycin were found to have a mutation in the hns gene, which has a role in transcriptional silencing [21]. These examples suggest that a common theme of prokaryotic strains selected for their high-yielding properties is a tendency to rebalance mRNA and protein synthesis rates, although the detail of how this is achieved is not yet understood.

A recent analysis of functional yields (by radioligand binding) and total yields (by immunoblotting) of recombinant angiotensin II type 1 receptor produced in insect cells showed that the majority of the protein was non-functional. When the same protein was produced in a stable, inducible HEK293 cell-line that had been selected by fluorescence-activated cell sorting to identify a high-yielding clone, a similar total yield was obtained, but the majority of the protein was now functional [22**]. Several factors may account for this observation: first, the
Examples of recent structures of recombinant membrane proteins. The name of the protein and the host cell used in its recombinant production are given. Structural images were downloaded from the PDB website (http://www.pdb.org/pdb/home/home.do) on 30th March 2015; protein chains are colour from the amino-terminus to the carboxy-terminus using a spectral colour gradient. (a) The structure of the KcsA potassium channel (PDB code: 2JK5) was solved to 2.4 Å using a truncated protein produced in E. coli in which thirty-five residues of KcsA had been removed with chymotrypsin. (b) The structure of the aquaporin 2 water channel was solved to 2.75 Å (PDB code: 4NEF) using recombinant protein produced in P. pastoris following codon optimisation of the corresponding gene sequence. (c) The structure of the NRT1.1 nitrate transporter at 3.7 Å (PDB code: 4CL4) was solved using a fusion protein with carboxy-terminal GFP and hexahistidine tags that had been produced in S. cerevisiae. (d) The structure of a the voltage-gated calcium channel at 2.75 Å (PDB code: 4MS2) was solved after six mutant forms of the protein were produced using the baculovirus system. (e) The 3 Å crystal structure of the GABA$_A$ receptor (PDB code: 4COF) was solved after approximately one-hundred construct variants of the full-length human β3 subunit were evaluated in a mammalian cell expression system.

strong polyhedrin promoter (polyhedrin mRNA accounts for 20% of polyadenylated mRNA in the cell [23]) is likely to produce very high levels of mRNA that overwhelm post-transcriptional pathways; second, viral-based expression systems impair the secretory pathway of a host cell [24], which will negatively impact on the ability of the nascent protein to fold.

Recovering functional protein from recombinant host cells is critically dependent upon the ability of the host to synthesise an authentically folded protein, which in turn depends on the proper functioning of its secretory pathway [25]. Recombinant eukaryotic proteins are often produced in low yields or are misfolded in prokaryotic host cells [26]. One explanation for this is that polypeptide synthesis rates are faster in prokaryotes (10–20 amino acids per s) than in eukaryotes (3–8 amino acids per s) [27**]. A popular strategy to mitigate this has been to decrease culture temperature, but many cellular functions that impact yield (e.g. transcription, translation, folding rates, membrane composition) are also affected by low temperature stress [28], which means that yield increases do not always transpire. Experiments in E. coli cells with mutant ribosomes (whose translation speed can be modulated) showed that reducing polypeptide elongation rates enhanced the folding efficiency of soluble firefly luciferase, the cycle 3 mutant of *Aequorea victoria* green fluorescent protein (GFP) and *S. cerevisiae* Cdc13p [27]. This suggests that protein folding requires slow translation rates in eukaryotes; in contrast, folding in bacteria is uncoupled from protein synthesis. Notably, in these experiments, decreasing translation rates in E. coli did not result in endogenous protein misfolding or the activation of a bacterial stress response [27].

**Bottlenecks in translation**

Translation is a highly regulated process that requires interactions between mRNA, ribosomes and a large number of other molecules in a complex, but optimised
<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Details of host cells used to produce recombinant α-helical transmembrane proteins for structures published in 2014 and 2015; representative structures are illustrated in Figure 2</td>
</tr>
<tr>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><em>Pichia pastoris</em></td>
</tr>
<tr>
<td>Methylo trophic yeast noted for its ability to grow to very high cell densities [52]; expression is typically via stably-integrated expression cassettes [19]</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
</tr>
<tr>
<td>Yeast with a wide range of genetic resources that have enabled host engineering studies [26]; expression is typically via episomal plasmids that are transiently transformed into the host cell, but stable integration is used</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
network that has been well-described for prokaryotes [29] and eukaryotes [30]. In eukaryotic cells, a key player in translational regulation is the mammalian target of rapamycin (mTOR) complex, which controls ribosome and tRNA biogenesis, translation initiation and entry into quiescence (G0 phase) [31]; mTOR is a central regulator of cell metabolism, growth, proliferation and survival, demonstrating the essential link between translation and host cell physiology.

The translational machinery of a cell responds to its growth rate, which is strongly affected by nutrient availability. Changes in nutrient source were recently shown to significantly reprogramme the transcriptome of *P. pastoris* under growth conditions relevant to recombinant protein production [32]. The main response was found to be transcriptional, while translational regulation was global rather than transcript-specific. When *P. pastoris* cells were cultured in methanol, a high proportion of the mRNA pool was associated with two or more ribosomes (and therefore deemed to be highly translated); methanol is typically used to induce protein production in this yeast species, which suggests that high recombinant protein yields may be associated with the slow growth rate observed under these conditions as well as promoter activity [33]. It has been known for decades that ribosomes from slow-growing cells have lower amino acid incorporation rates per second per ribosome than cells that are growing normally [34**]. Slowing translation speed may therefore enhance recombinant protein folding efficiency. In support of this, partial inhibition of translation in mammalian host cells (using the drug, emetine) was shown to result in a substantial reduction in the yield of misfolded *Pontellina plumata* GFP; emetine treatment also increased the functional recombinant yields of both GFP and a ΔF508 mutant of the cystic fibrosis transmembrane conductance regulator [35]. Lower translation speed has been postulated to directly lead to higher translational accuracy [36]. This is difficult to reconcile with current biochemical knowledge of translation but if the relationship does indeed exist, slower translation could be beneficial for the production of difficult-to-fold proteins that are particularly sensitive to translational errors.

The sequence of the mRNA transcript is also important in determining the rate and accuracy of protein translation [37]. It is established that individual species have a preference for certain of the 64 available codons over others, but the biological reason for this is unclear. One idea is that each codon has a different decoding time; ‘faster’ codons lead to a higher translation rate,
which is more resource efficient [38]. Recently, 'speed control signals' have been proposed in the signal peptides of secreted proteins that delay translation to allow co-translational and post-translational translocation, protein processing and folding [39]. Codon usage bias was found be normal in open reading frames, but decreased in signal peptide coding regions [39]. This may have consequences for the optimal design of the heterologous DNA construct in Figures 1 and 3. Another idea is that different codons are read with different degrees of accuracy: when translational accuracy was included as a parameter in the design of expression constructs for a Plasmodium lysyl-tRNA synthetase in E. coli, proteolysis was reduced and solubility increased [40].

**Bottlenecks in folding and secretion**

Translation, folding and secretion are interconnected pathways that draw on the same cellular resources. If the interconnections are not optimised in favour of the recombinant protein, then yields will be low (Figure 3). Most nascent proteins do not fold spontaneously in vivo, but require a network of chaperones that not only facilitate protein folding, but also perform quality control, ensuring that damaged or misfolded proteins are degraded or refolded. Many studies have focused on the
co-expression or down-regulation of individual chaperones to improve recombinant protein yields [41], but a current trend is to exploit a host cell’s global stress response to misfolded proteins to ensure the proper functioning of its secretory pathways; this approach is anticipated to be more effective in improving the quantity and quality of recombinant protein for structural studies [42*]. In eukaryotes, even moderate levels of stress cause a decrease in the rate of translation, the sequestering of mRNA into stress granules and the aggregation of proteins [43]. It still remains unclear how cells regulate this response, but the ability to influence the distribution of diverse chaperones in vivo may offer solutions to improving protein yields [44]. For example, a recent study of recombinant G protein-coupled receptor production in S. cerevisiae demonstrated that mislocalised proteins were associated with the endoplasmic reticulum chaperone, BiP [45].

The impact of two cellular responses is of particular interest: The unfolded protein response (UPR; influencing the early part of the secretory pathway) and the heat shock response (HSR; a response to cytosolic stress). Recombinant protein production is likely to affect ER homeostasis and therefore trigger a UPR, which in turn causes an increase in the folding capacity of a cell [25]; recently biosensors have been used to measure the UPR in mammalian cell-lines producing monoclonal antibodies [46]. During the UPR in yeast, about 5% of the genome is up-regulated (mirroring the situation in higher eukaryotes), ribosomal biogenesis and assembly are translationally repressed and mRNAs encoding the UPR transcription factor Hac1p, the ER-oxidoreductase Erp1p and the ER-associated protein degradation (ERAD) protein Der1p, are enriched in polysomal fractions, indicating translational up-regulation [47]. The HSR activates chaperones and the proteasome in order to relieve stress. HSR up-regulation has been used to increase recombinant yields of soluble α-amylase in S. cerevisiae, but did not increase the yield of a recombinant human insulin precursor [48]. It is now established that the UPR, HSR and other stress responses (such as the environmental stress response; ESR) overlap with each other, providing a hormetic benefit to cells in which mild stress enhances tolerance to future stressful stimuli [49] such as that imposed during recombinant protein production.

Conclusions

The use of stabilising mutations to improve the crystallisation propensity of recombinant membrane proteins has resulted in major breakthroughs in modern structural biology [3]. Our emerging understanding of how the processes of translation and protein folding are affected in recombinant host cells now offers new, complementary opportunities to improve functional yields. However, the interlinked nature of transcription, translation, folding and secretion and their impact on cell physiology means that optimising host cells to be maximally productive with respect to functional recombinant protein yield is both intellectually and technically demanding. Table 1 suggests a dominant role for microbial host cells in membrane protein structural biology projects; since microbes are particularly amenable to genetic engineering, new insight may emerge from their use in the foreseeable future.

Conflict of interest statement

Nothing declared.

Acknowledgements

RMB thanks Kristina Hedfalk, University of Gothenburg, for her insightful comments and acknowledgements funding from the Biotechnology and Biological Sciences Research Council (BBSRC; via grant BB/I019960/1) and the Innovative Medicines Joint Undertaking under Grant Agreement n° 115583 to the ND4BB ENABLE Consortium. TvdH acknowledges funding from BBSRC (via grant BB/I010351/1) and from the Leverhulme Trust (via grant RPG-2014-032).

References and recommended reading

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

- of special interest
- of outstanding interest


Most G protein-coupled receptor crystal structures have been obtained using a fusion protein strategy where the flexible third intracellular loop is replaced by T4 lysozyme. This article reports two modified T4L variants that can be used to optimise crystal quality or promote alternative packing interactions.


An example of the feedback loop illustrated in Figure 1, where approximately one-hundred construct variants of the full-length human β3 subunit were evaluated in a mammalian cell expression system.


22. Thomas J, Tate CG: Quality control in eukaryotic membrane protein overproduction. J Mol Biol 2014, 442:4139-4154. This article describes a simple and rapid test to determine whether recombinant membrane proteins are correctly folded, without requiring radioligand binding assays. It also offers deep insight into the challenges of synthesising correctly folded recombinant proteins.


An insightful overview of the role of protein folding and secretion in recombinant protein production in S. cerevisiae.


