Review TRENDS in Biotechnology Vol.24 No.8

Rationalizing membrane protein overexpression

Samuel Wagner[1], Mirjam Lerch Bader[1], David Drew[2] and Jan-Willem de Gier[1]

[1] Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden
[2] Division of Molecular Biosciences, Membrane Protein Crystallography Group, Imperial College London, SW7 2AZ, UK

Functional and structural studies of membrane proteins usually require overexpression of the proteins in question. Often, however, the 'trial and error' approaches that are mainly used to produce membrane proteins are not successful. Our rapidly increasing understanding of membrane protein insertion, folding and degradation means that membrane protein overexpression can be more rationalized, both at the level of the overexpression host and the overexpressed membrane protein. This change of mindset is likely to have a significant impact on membrane protein research.

Introduction

In both prokaryotes and eukaryotes, 20–30% of all genes encode membrane proteins, which often form supramolecular complexes that act in many different and often essential capacities [1]. Membrane proteins have key roles in many diseases, and more than half of all drug targets are membrane proteins. Although biomembrane research has always been an important area in cell biology and biochemistry, recent advances in proteomics and successes in structural biology have put membrane proteins in the limelight.

There are two classes of membrane proteins: β-barrel membrane proteins and helical bundle membrane proteins [2]. β-barrel membrane proteins can be expressed in inclusion bodies, from where they can be readily isolated, subsequently refolded and used for further studies [3]. In contrast to β-barrel membrane proteins – and despite tremendous efforts – there are very few examples of helical bundle membrane proteins that have been successfully refolded after denaturing isolation from inclusion bodies (e.g. see Ref. [4]). Therefore, the preference is to overexpress helical membrane proteins in a membrane system, from which they can be purified after detergent extraction. In this review, we discuss only the overexpression of helical bundle membrane proteins, which hereafter are referred to simply as 'membrane proteins'.

The natural abundance of most membrane proteins is usually too low to isolate sufficient material for functional and structural studies. The use of natural sources also excludes the possibility of genetically modifying proteins to improve their stability and to facilitate their detection and purification. Both the necessity and feasibility of overexpression are illustrated by the now steadily growing number of high-resolution structures of helical bundle membrane proteins obtained through overexpression (Figure 1).

Corresponding author: de Gier, J.-W. (degier@dbb.su.se).
Available online 3 July 2006

A recent overexpression screen of 70 *Mycobacterium tuberculosis* membrane proteins in *Escherichia coli* concluded that well-expressed integrated membrane proteins tend to be small and to have relatively few transmembrane helices [5]. Juxtaposed to this, the analysis of a large-scale overexpression screen of *E. coli* membrane proteins found that levels of membrane-integrated proteins do not correlate with – and thus cannot be predicted by – obvious sequence characteristics such as codon usage, protein size, hydrophobicity and number of transmembrane helices [6]. These conflicting conclusions might be explained by the different approaches that were used to monitor membrane protein overexpression (Box 1). In practice, the expression of many homologs of a target is tested in a given host to select the best expressing (and functional) candidate. It should be noted that the expression of almost identical proteins can vary significantly (e.g. see Ref. [7]).

Although Figure 1 might suggest the opposite, many different systems have been tried in the attempt to overexpress both prokaryotic and eukaryotic membrane proteins. Here, we do not aim to give a complete overview of all systems and their possible advantages and disadvantages, as this topic has been covered recently in a special issue of *Biochimica et Biophysica Acta* [8]. Instead, we discuss what seems to be a novel and very encouraging trend – namely, that membrane protein overexpression is considered more and more a scientific problem to be systematically and rationally addressed, rather than an obligatory first step in functional and structural studies of membrane proteins. Already this change in mindset has led to some remarkable results, and it is probable that membrane protein research will continue to benefit from this in the future.

Bottlenecks affecting membrane protein overexpression

In bacteria, membrane proteins are overexpressed in the cytoplasmic membrane, whereas in eukaryotes they are typically overexpressed in the endoplasmic reticulum (ER) membrane, from where they can be transported further. For clarity, the overexpression of membrane proteins foreign to the expression host is hereafter referred to as 'heterologous overexpression'.

How are membrane proteins targeted to, and inserted and folded into, the bacterial cytoplasmic and eukaryotic ER membranes? Figure 2 provides a simplified overview of our current knowledge of the biogenesis of membrane proteins in these membrane systems (for more detailed information, see the reviews [9,10] and references herein).

 0167-7799/$ – see front matter doi:10.1016/j.tibtech.2006.06.008

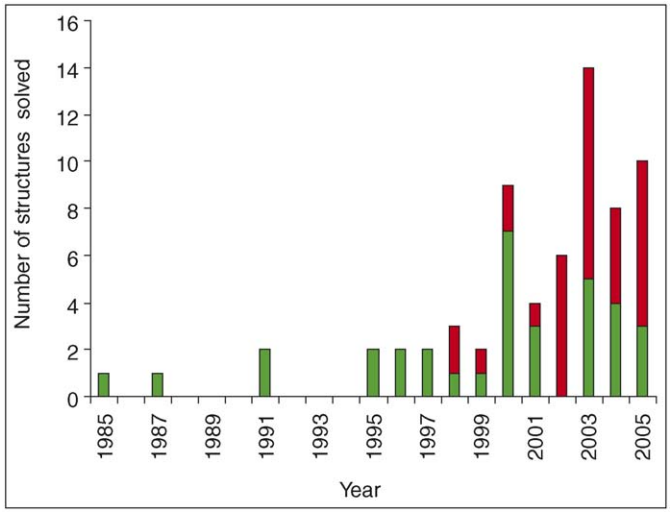

Figure 1. Unique membrane protein structures solved over the years. Structures obtained from membrane proteins isolated from natural sources are in green and those obtained from overexpressed material are in red. The data are taken from the 'Membrane Proteins of Known Structure' website (http://blanco.biomol.uci.edu/Membrane_proteins_xtal.html). Notably, proteins have so far almost exclusively been overexpressed in *E. coli* except in two cases (rat potassium channel [70] and plant aquaporin [71], both overexpressed in yeast) in which the structures were solved from proteins overexpressed in *E. coli*, and >70% of the structures are of non-eukaryotic origin. Proteins of the same type but from different species are considered as unique, whereas mutated proteins and different conformations of a protein are not; only membrane proteins that had been deposited in the Protein Data Bank by the end of 2005 are included.

The targeting and insertion of heterologously overexpressed membrane proteins have been studied in *E. coli*. *In vitro* translation crosslinking approaches have shown that heterologously expressed membrane proteins make contact with the signal recognition particle (SRP), the Sec translocon and YidC [11]. Nonetheless, this does not necessarily mean that the *E. coli* components are optimal for assisting the targeting, insertion and folding of heterologously expressed membrane proteins into the membrane. The successful overexpression of mitochondrial carriers in the Gram-positive bacterium *Lactococcus lactis* indicates that membrane proteins can be efficiently targeted and inserted in a heterologous system via pathways that they do not use under normal conditions [12].

The focus of this section is on bottlenecks specific to membrane protein overexpression rather than more general problems related to the production of proteins, such as mRNA stability and codon usage, which have been reviewed extensively elsewhere (e.g. see Refs [13,14]). We address these potential bottlenecks using the order of the biogenesis events as a reference.

Heterologous overexpression

Prokaryotic membrane proteins are usually overexpressed in the Gram-negative bacterium *E. coli*, whereas eukaryotic membrane proteins have been also overexpressed in the Gram-positive bacterium *L. lactis* and various eukaryotic systems. Heterologous overexpression of membrane proteins can be hampered by different synthesis, targeting, insertion and folding characteristics in the overexpression host [15].

For example, the polypeptide elongation and protein folding rates are considerably higher in prokaryotes than

Box 1. Monitoring membrane protein overexpression

Monitoring the localization, quantity and quality of overexpressed membrane proteins is important for assessing and optimizing overexpression yields [73]. Because it is unpredictable whether overexpressed membrane proteins will end up in a membrane system or in inclusion bodies, the first step in monitoring membrane protein overexpression is usually fractionation of the overexpression vehicle into soluble, insoluble (inclusion bodies) and membrane fractions [5]. It should be noted that aggregated membrane proteins might associate with membranes rather than producing distinct inclusion bodies [4].

Coomassie- or silver-stained standard SDS–PAGE gels are most often used to detect membrane proteins in (subfractionated) overexpression vehicles and to analyze (partially) purified material. They facilitate assessment of the purity, integrity and the quantity of overexpressed membrane proteins.

Western blotting using antibodies against, for example, an expression or purification tag, is also widely used to detect membrane proteins in (subfractionated) overexpression vehicles and to assess (partially) purified material. The sensitivity of western blotting greatly depends on the antibody used. Posttranslational modifications, such as phosphorylation, can also be detected by western blotting [66,74]. Owing to the hydrophobic nature of membrane proteins, their transfer from a gel to a blotting membrane can be cumbersome, making western blotting less suitable for quantitative purposes. A recent study concluded that well-expressed integrated membrane proteins tend to be small and have relatively few transmembrane helices [5]. The expression of many membrane proteins in this study was screened by means of western blotting. It is possible that the transfer of small membrane proteins with relatively few transmembrane helices from a gel to blotting membrane is simply more efficient than the transfer of other membrane proteins.

Recently, the detection of membrane protein overexpression levels by means of dot-blotting (i.e. spotting samples directly on a nitrocellulose filter) has been explored [43]. In this approach, membrane protein material obtained by detergent extraction of whole cells combined with affinity-tag purification is spotted on a nitrocellulose filter and subsequently detected with antibodies against the affinity tag. The method is fast and avoids the difficult and unreliable transfer from the SDS–PAGE gel to a nitrocellulose filter. Dot-blotting, however, does not provide any information on the integrity of the overexpressed membrane protein (except that the affinity or detection tag is still intact), and the solubilization of aggregated overexpressed protein in harsh detergents might lead to an overestimation of membrane-integrated protein.

Fusing GFP to the C terminus of membrane proteins enables overexpressed proteins to be monitored in intact cells. The GFP moiety folds properly and becomes fluorescent only if a membrane protein is stably inserted into the membrane [62,75]. By using whole cells as the starting material, the membrane-protein–GFP fusion can be visualized directly by in-gel fluorescence in standard SDS–PAGE gels [42]. In addition, the GFP moiety greatly facilitates purification and quality assessment of the membrane-protein–GFP fusion; in-solution and in-gel fluorescence, in addition to size-exclusion chromatography directly coupled to fluorescence detection, can be used to monitor membrane protein purification and to facilitate precrystallization screening [42,76].

Finally, mass spectrometry can be used to assess the integrity and to characterize posttranslational modifications of purified overexpressed membrane proteins, (e.g. see Refs [77–79]).

These methods to monitor the quantity and quality of overexpressed membrane proteins do not provide any information on the proper folding and functionality of the overexpressed material: integration into a membrane system is no guarantee of integrity. To monitor the functionality of a protein, not only should its function be known but also an activity assay should be available: depending on the protein, binding assays with fluorescent or radioactive ligands, or transport assays could be used (e.g. see Ref. [73]).

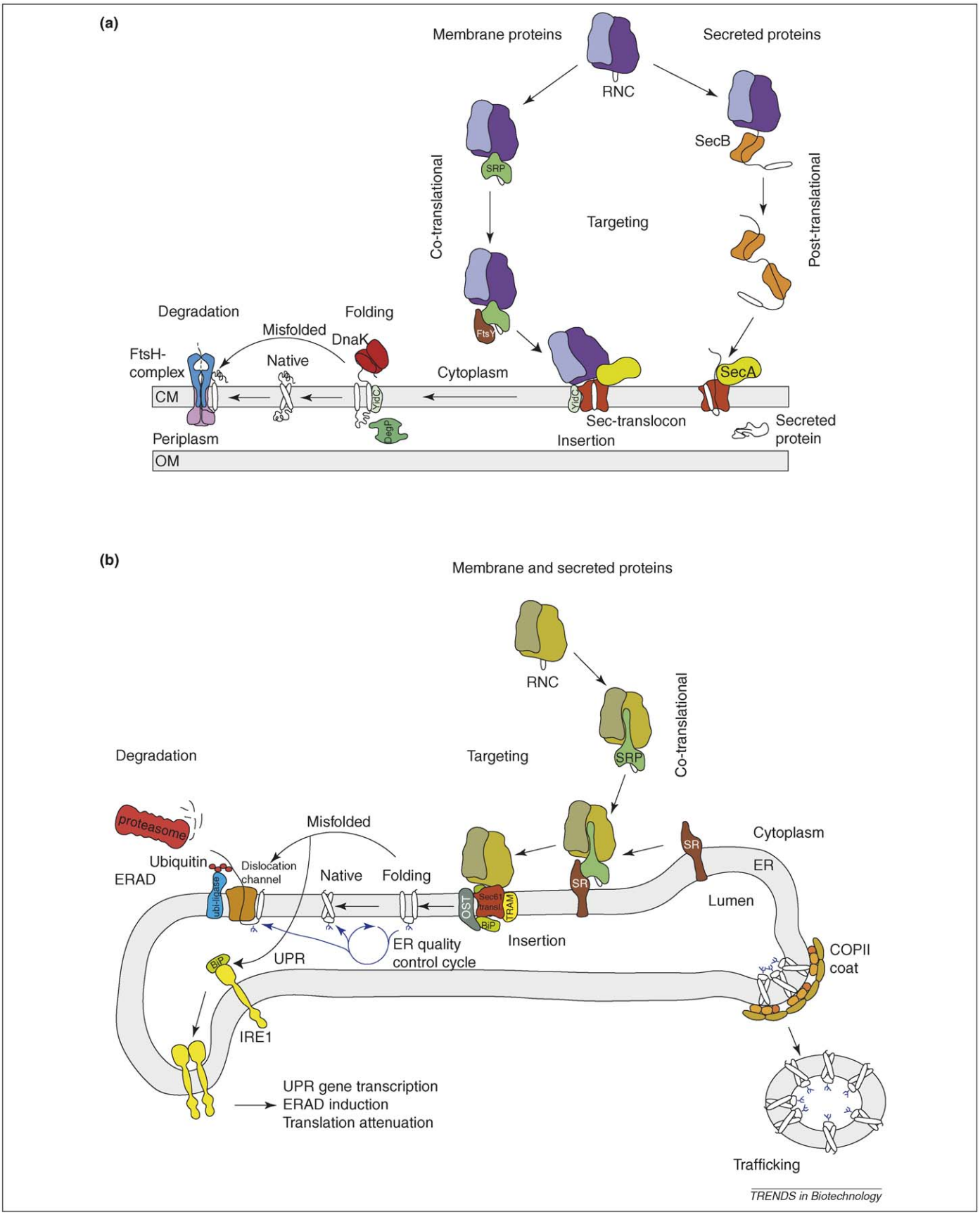

Figure 2. Membrane protein targeting, insertion and degradation in the bacterial cytoplasmic and eukaryotic ER membranes. **(a)** Ribosome–membrane protein nascent chain complexes (RNCs) are targeted in a cotranslational fashion to the bacterial (i.e. *E. coli*) cytoplasmic membrane via the SRP pathway (comprising the signal recognition particle and its receptor FtsY). At the cytoplasmic membrane, the RNC docks at the Sec translocon – a protein-conducting channel that facilitates both the translocation of hydrophilic polypeptide chains across the membrane and the insertion of transmembrane segments into the lipid bilayer. The translocation of sizeable periplasmic loops requires the ATPase SecA. YidC has been proposed to mediate the transfer of transmembrane segments from the Sec translocon into the lipid bilayer and can assist the folding of membrane proteins. The lipid composition of the membrane might also influence membrane protein folding. Folding of soluble cytoplasmic domains might be

in eukaryotes, which might cause mistargeting and misfolding of heterologously expressed membrane proteins. Incompatibility of factors involved in the processing of heterologously expressed membrane proteins indicates a general problem. Some of the charged residues in the yeast Sec61 translocon that are proposed to be important for proper functioning [16] are not conserved in the prokaryotic Sec translocon. Thus, it is very well possible that subtle differences in the insertion process could contribute to the difficulties experienced in the heterologous overexpression of membrane proteins [17]. In addition, the availability of endogenously expressed factors assisting the biogenesis of membrane proteins can be insufficient to support proper processing of the overexpressed protein. Indeed, in *E. coli* it has been shown that, on overexpression of homologous membrane proteins, SRP is titrated out [18].

Protein folding capacity

Increasing evidence indicates that, in both the bacterial cytoplasmic and the eukaryotic ER membranes, folding of membrane proteins is catalyzed by generic integral membrane chaperones. The capacity of these folding mediators might not be sufficient to support the folding of overexpressed membrane proteins.

Some membrane proteins, such as some transporters in yeast and a few rhodopsins, require membrane-protein-specific chaperones for proper folding [19–21]. Thus, it is possible that, for the heterologous overexpression of membrane proteins, the chaperones required to assist folding might be absent. Folding of soluble domains of membrane proteins might require cytoplasmic and periplasmic or luminal chaperones. Again, these chaperones might not always be present in sufficient numbers to support the folding of overexpressed material in addition to their endogenous substrates.

Glycosylation

Glycosylation of eukaryotic membrane proteins can be essential for proper folding, stability and also function [22]. The composition and number of *N*-glycans can be crucial and depends on the overexpression host used. The most commonly used prokaryotic overexpression vehicle, *E. coli*, cannot glycosylate proteins. This does not disqualify *E. coli per se* as a host for the functional overexpression of eukaryotic membrane proteins that are normally glycosylated: the human CB2 receptor, for example, can be functionally overexpressed in *E. coli* [23].

Lipid composition

Membranes from prokaryotes, yeast or higher eukaryotes differ in their lipid composition, which might lead to problems with the heterologous overexpression of functional membrane proteins. This bottleneck is illustrated by the observation that the mammalian presynaptic serotonin transporter (SERT) is functionally overexpressed only in the presence of cholesterol, which is not present in prokaryotic membranes [15], and that the *E. coli* membrane protein lactose permease (LacY) absolutely requires the lipid phosphatidylethanolamine for proper folding [24,25]. Thus, differences in membrane bilayer properties can have a significant effect on the insertion, folding and functioning of a membrane protein [26,27]. These differences might not necessarily indicate a dead-end pathway, however, because the addition of certain lipids during and/or after purification can restore functionality of overexpressed membrane proteins. For LacY, for example, post-assembly synthesis of phosphatidylethanolamine in intact *E. coli* cells can restore its functionality [25].

Membrane space and accommodation of foreign structures

The membrane space required to accommodate overexpressed membrane proteins represents another potential bottleneck. It has been shown that membrane proliferation during overexpression of the *E. coli* b subunit of the F_1F_0 ATP synthase can improve overexpression yields [28]. The mechanism that induces this effect is not understood.

The accumulation of non-native or foreign structures in a membrane can induce stress responses and activate proteolytic systems of the overexpression host. In *E. coli*, the FtsH complex is involved in the degradation of misfolded overexpressed membrane proteins, as shown for the Sec translocon component SecY, which is rapidly degraded when it is not expressed in the presence of its complex partner SecE [29]. In eukaryotes, saturating the ER folding capacity by membrane protein overexpression can induce the unfolded protein response (UPR) [30]. This response orchestrates the expression of a large set of proteins that are involved in protein folding and degradation of misfolded proteins [31]. In particular, ER stress transiently attenuates protein synthesis and upregulates the ER-associated protein degradation (ERAD) system, which might affect the overexpression of secreted proteins (Figure 2). For this reason, it has been suggested that overexpression of functional membrane proteins is more likely to be successful if the protein load in the ER does not exceed the threshold for UPR induction [30].

Alleviating the bottlenecks

From the previous section it is clear that there are many potential bottlenecks that might affect the overexpression of membrane proteins. What has been done so far to alleviate them? Here, we give examples of how bottlenecks

supported by cytoplasmic chaperones such as DnaK, whereas that of periplasmic domains might be supported by periplasmic chaperones such as DegP (which can also act as a protease). The FtsH complex is involved in quality control and degradation of membrane proteins. Secretory proteins are targeted by the chaperone SecB in a mostly posttranslational manner to the Sec translocon. The translocation of secretory proteins is SecA dependent. **(b)** RNCs are targeted in a cotranslational manner to the ER membrane via the SRP pathway (comprising the signal recognition particle and its receptor, SR). Eukaryotic SRP is, in contrast to the SRP of *E. coli*, equipped with components that on binding of the SRP to the RNC, can halt translation. At the ER membrane, the RNC docks at the Sec61 translocon. On dissociation of the SRP from the RNC, translation is resumed. The Hsp70 chaperone BiP supports both the translocation of polypeptides into the ER by a Brownian ratchet mechanism and the folding of luminal protein domains. TRAM might assist the transfer of transmembrane segments from the Sec61 translocon into the lipid bilayer and the folding of membrane proteins. After translocation, proteins can be glycosylated by the oligosaccharyltransferase (OST). Folding of ER-glyco(membrane)proteins is chaperoned in a complex quality-control cycle [72]. The accumulation of misfolded proteins can elicit the unfolded protein response (UPR) by competing for binding to BiP. In this response, receptor kinases (e.g. IRE1 in yeast) dimerize and transduce a folding stress signal. Misfolded proteins are dislocated from the ER, ubiquitinated and finally degraded by the cytoplasmic proteasome in a process called ERAD. Native conformers leave the ER from exit sites by COPII-mediated trafficking. In principle, every single step of membrane protein biogenesis can malfunction on the overexpression of membrane proteins. Abbreviations: CM, cytoplasmic membrane; OM, outer membrane.

have been addressed at the level of both the overexpression host and the overexpressed membrane protein.

Optimizing overexpression conditions

The conditions used to overexpress membrane proteins in a particular host can greatly influence overexpression yields. It is important to keep in mind that the best growth conditions (i.e. those that generate as much biomass as possible) are not necessarily the optimal conditions for protein production. In many cases, it has been beneficial to decrease the translation rate by lowering the culture temperature or inducer concentration, or by choosing a weaker promoter. As mentioned in the previous section, several potential bottlenecks – such as the titration of components involved in the biogenesis of overexpressed material and the accumulation of non-native and/or misfolded protein in the membrane – might be alleviated by decreased translation rates.

Indeed, a recent study in yeast has reconfirmed that culture conditions can influence not only the amount of produced protein but also its membrane integration [32]. In this example, although the overexpression *per se* was better at higher temperatures, lower temperatures resulted in a larger amount of membrane-integrated overexpressed material. This study also stresses the need for tools to monitor the amount of membrane-integrated overexpressed protein directly (Box 1).

Improving targeting and folding of overexpressed membrane protein

The coexpression of chaperones is now routinely used to improve overexpression yields of challenging soluble proteins (e.g. see Refs [13,14]); however, finding the right coexpression conditions has to be done empirically and can be a tedious process (e.g. see Ref. [33]). Wang and co-workers [34] tried to improve overexpression of the *E. coli* Mg^{2+} transporter CorA in *E. coli* by coexpressing components of the SRP-targeting pathway, but without success. Surprisingly, they found that overexpression levels of CorA could be improved by coexpression of the cytoplasmic DnaK–DnaJ chaperone system [34].

Recently, the structure of CorA was solved [35]. The CorA transporter is a homopentamer, and monomeric CorA consists of a large amino (N)-terminal cytoplasmic domain and two transmembrane segments at the very carboxyl (C) terminus. The architecture of CorA suggests that, unlike most other membrane proteins, it is not targeted via the SRP pathway but is posttranslationally targeted. Thus, it is probable that the DnaK–DnaJ chaperone system is involved in the targeting and folding of CorA, which could explain why coexpression of the DnaK–DnaJ chaperone system improves CorA overexpression yields.

As mentioned above, the unfolded protein response (UPR) in the ER can be induced on the overexpression of membrane proteins. In yeast, it has been shown that monitoring the UPR with a reporter gene can be used to optimize the overexpression of functional membrane proteins. To maximize the production of functional material, the expression, and thereby the protein-folding load, must be fine-tuned to avoid or to minimize induction of the UPR [30].

In eukaryotic overexpression systems, the coexpression of factors involved in protein folding in the ER has been used to improve the overexpression yields of several membrane proteins. For example, coexpression of the ER folding catalyst calnexin in the baculovirus expression system resulted in a threefold increase in functionally expressed SERT; coexpression of the chaperones calreticulin and BiP also improved SERT overexpression, albeit to a lesser extent than coexpression with calnexin [36].

The yields of functionally overexpressed membrane protein can be also increased by adding ligands to the culture medium that assist folding and stabilization of the overexpressed protein [37,38]. This possibility must be tested empirically, however, because it has been shown that plasma membrane transporters in yeast can be downregulated by their substrates through proteolysis at higher substrate levels [39].

Specialized hosts and systems for membrane protein overexpression

Almost a decade ago, a screen was designed to isolate derivatives of the *E. coli* strain BL21(DE3), which is the most widely used overexpression vehicle, with improved characteristics for membrane protein overexpression [40]. The screen was based on the ability of the vehicle to cope with toxic effects of the overexpression of a particular membrane protein. *E. coli* strains C41(DE3) and C43(DE3) are the best known examples of strains isolated from this screen. The toxicity of the target protein is reduced in them, and plasmid stability is improved, particularly in C43(DE3) [41]. For reasons not yet understood, overexpression of many membrane proteins in the cytoplasmic membrane is less toxic in these strains than in BL21(DE3): growth is hardly affected on induction of membrane protein overexpression, thereby improving overexpression yields [42]. The C41(DE3) and C43(DE3) strains are now widely and very successfully used to overexpress membrane proteins (e.g. see Refs [42,43]).

A system that is 'optimal' for overexpressing membrane proteins can also be used; in other words, a system that by nature has a high biogenesis capacity for membrane proteins and sufficient space in the membrane to accommodate the overexpressed material. The most notable example is probably the use of the fruit fly *Drosophila melanogaster* to overexpress a G-protein-coupled receptor (GPCR) [44]. The photoreceptor cells of *D. melanogaster* contain extensive stacks of membranes, where high levels of rhodopsins are naturally present. In transgenic flies, these photoreceptor cells have been successfully used to overexpress the *D. melanogaster* metabotropic glutamate receptor. It remains to be seen whether other membrane proteins can also be successfully overexpressed in photoreceptor cells of flies.

Cell-free translation systems for protein synthesis have been developed to circumvent the damaging consequences of protein production on the overexpression host. Kuruma *et al.* [45] have developed the PURESYSTEM, which is based on *E. coli* and comprises inverted membrane vesicles supplemented with purified factors involved in protein synthesis and targeting. This system can be used to produce small amounts of membrane-integrated protein. Unfortunately, it is costly and difficult to scale up.

By contrast, the cell-free expression systems developed by Klammt *et al.* [46] and Elbaz *et al.* [47] facilitate high-yield expression in the range of milligrams of protein per milliliter of reaction volume. Unfortunately, the amount of functional material that can be obtained from the total protein produced has not been determined for these studies. Recently, the structure of the *E. coli* membrane protein EmrE has been solved by combining material obtained through overexpression of the protein in *E. coli* and material obtained through synthesis in a cell-free system, nicely showing that cell-free translation systems are not just an oddity [48].

Engineering membrane proteins for overexpression

In addition to the optimization of expression systems and/or conditions, the membrane protein itself can be modified to improve its expression yields. In this section, we discuss how such modifications can be achieved: most examples are of eukaryotic proteins expressed in prokaryotes.

N-terminal truncation and signal sequence fusions

In both prokaryotes and eukaryotes, the N-terminal tails of membrane proteins show a strong preference for the cytoplasm [6,49]. Translocation of an N-terminal tail depends on the ability of the N terminus to remain unfolded (which is seemingly easier if the tail is shorter), the number of positively charged residues in the tail region, and the 'strength' of the first transmembrane segment (i.e. the charge difference, the length, and the overall hydrophobicity of the reverse signal anchor) [50]. In short, inability to translocate the N-terminal tail of a membrane protein efficiently might hamper the overexpression of membrane proteins with N-terminal tails that have to be translocated across a membrane.

This notion is supported by many observations. For example, functional overexpression of the yeast mitochondrial carrier AAC2 (ADP–ATP exchanger) can be increased in *L. lactis* if the N terminus is shortened or the N-terminal tail is swapped with a shorter one taken from the isoform AAC3 [51]. Similarly, surface expression of the human cannabinoid receptor CB1 can be increased in BHK cells if its 116-residue N-terminal tail is also truncated and/or an ER signal peptide is engineered to the N terminus of the protein [52]. Fusion of a signal peptide to CB1 facilitates targeting and translocation of the long N terminus of this receptor.

In addition, the expression of several other GPCRs, such as human opioid receptor, human dopamine receptor and serotonin receptor in the yeast *Pichia pastoris*, has been improved by fusing the *Saccharomyces cerevisiae* yeast α-factor signal sequence to their N-terminal tails (e.g. see Ref. [53]). A recent study of the optimization of the functional expression of 20 GPCRs in *P. pastoris* used this signal sequence in all constructs [38]. Likewise, functional expression of the GPCRs neurotensin receptor (NTR) and adenosine receptor is improved in *E. coli* by fusing the maltose-binding protein (MBP) with its secretory signal sequence to the N termini of these proteins [54,55].

N-terminal protein fusions

N-terminal soluble protein fusion partners, such as glutathione *S*-transferase, NusA and green fluorescent protein (GFP), are routinely tested for their ability to improve the expression of poorly expressed membrane proteins. For the above-mentioned reasons, without a signal sequence this approach is likely to be more successful if the N-terminal tail is cytoplasmic.

Recently, an N-terminal fusion with Mistic, a small protein that is unique to *Bacillus subtilis*, has been reported to improve the expression of several eukaryotic membrane proteins in *E. coli* [56]. By autonomously associating with the bacterial membrane, Mistic is speculated to chaperone a downstream membrane protein into the lipid bilayer without assistance from the Sec translocon. At this stage, further experiments are needed to verify whether this targeting hypothesis is correct, and more examples of functionally expressed membrane proteins are needed. Thus, it remains to be seen how far- reaching the magic of Mistic is.

C-terminal protein fusions

Because folding through the translocon is by and large unidirectional (i.e. it occurs from the N to the C terminus) [57], the C-terminal tails of multispanning membrane proteins can be threaded through on either side of the membrane in an unfolded state. Thus, the translocation of large extracytoplasmic C-terminal tails is unlikely to be as problematic as that of N-terminal tails, and the attachment of any fusion protein seems reasonable (GFP, which is unable to fold in the periplasm in *E. coli*, is probably an exception [58,59]).

The advantage of attaching a fusion protein is one of stability, because proteases such as the *E. coli* FtsH complex can unravel and degrade membrane proteins from free N- or C-terminal ends [60]. Indeed, Grisshammer and co-workers [55] tested the ability of various fusions and combinations of C-terminal tags (e.g. biotin, polyhistidine, Flag and streptavidin tags) and single fusions of a polyhistidine or c-Myc epitope or *E. coli* thioredoxin (residues 2–109) to improve the functional overexpression of an MBP–NTR receptor. Thioredoxin provided the most significant improvement, which was attributed to the remarkable stability of the globular domain.

The combined use of an N-terminal MBP and a histidine-tagged C-terminal thioredoxin fusion has been successfully used to overexpress the GPCRs CB2 and adenosine receptor in *E. coli* [23,54,61]. Large-scale expression of the NTR fusions in *E. coli* yields up to 10 mg of purified protein per 50 l of culture. TEV protease sites located at NTR boundaries facilitate removal of the fusion partners [61]. Interestingly, fusion of GFP to the C terminus of the human KDEL receptor also improves functional expression levels 15-fold in *L. lactis* [62]. Similar to thioredoxin, GFP is an exceptionally stable protein that might protect overexpressed proteins from proteases.

Improving the quality of overexpressed membrane proteins

Even if yields are satisfactory, the quality of the isolated overexpressed material can be insufficient. Heterogeneity due to proteolysis or posttranslational modifications is likely to affect crystallization adversely. To avoid heterogeneity caused by proteolytic degradation and also to

remove flexible parts of the protein that might prevent crystal formation, the proteolytically resistant core of the protein can be identified by limited proteolysis combined with mass spectrometry. Either a trimmed construct is engineered, as has been done for the *E. coli* glycerol-3-phosphate transporter [63], or the full-length protein can be treated with a suitable protease before crystallization, as exemplified by the KcsA K^+ channel [64].

To reduce heterogeneity due to posttranslational modifications, (potential) glycosylation and phosphorylation sites can be removed, as described for the rat K^+ channel Kv1.2, which has been overexpressed in the yeast *P. pastoris* [65].

Future perspectives

Our rapidly growing knowledge of the biogenesis of membrane proteins, coupled with efforts to identify bottlenecks hampering membrane protein overexpression, has created new possibilities to design strategies for improving yields at the level of both the overexpression host and the target protein.

So far, proteomics and DNA microarrays have not been used to study the effects of membrane protein overexpression on the host; however, these technologies will undoubtedly facilitate the identification of bottlenecks impairing membrane protein overexpression. Recently, the proteomic characterization of *E. coli* cells overexpressing soluble proteins enabled bottlenecks affecting protein synthesis to be identified and, on the basis of these, strains with improved protein production characteristics to be engineered (e.g. see Refs [66,67]), illustrating the potential of such strategies.

Both coexpression of components involved in the biogenesis of overexpressed membrane proteins (or alleviation of secondary effects caused by membrane protein overexpression) and engineering of membrane proteins have been explored to improve overexpression yields – sometimes with very encouraging results. More daring and so far unexplored strategies – for example, tailoring components involved in the biogenesis of overexpressed membrane proteins and inactivating components such as proteases that interfere with membrane protein overexpression – have also great potential to improve overexpression yields.

The exploration of systems that are, by nature, very well suited for membrane production, such as the photoreceptor cells of fruit flies, also deserves more attention. Interestingly, in 1992 attempts were made to overexpress membrane proteins such as the cystic fibrosis transmembrane conductance regulator in the milk of transgenic mammals [68]. The membranes of the secreted milk fat globules are an excellent platform for the production of membrane proteins.

Recently developed methodologies to monitor, rapidly, yields of membrane protein overexpression enable different overexpression conditions and hosts to be screened in a high-throughput manner (Box 1). Such screening will facilitate not only random-based approaches, such as forward evolution for the isolation of membrane protein variants that express better than their wild-type counterparts [69], but also more rational approaches as outlined in this review.

In summary, the rationalization of membrane protein overexpression, combined with novel approaches to monitor membrane protein expression, has shown its strength and holds great promise for membrane protein research in the future.

Acknowledgements

We thank Dirk Jan Slotboom for critically reading the manuscript. Research in the laboratory of J.W.d.G. is supported by the Swedish Research Council and the Marianne and Marcus Wallenberg Foundation. M.L.B. was a recipient of a fellowship from the Swiss National Science Foundation. S.W., M.L.B. and D.D. contributed equally to the production of this review.

References

1 Wallin, E. *et al.* (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* 7, 1029–1038
2 von Heijne, G. (1999) Recent advances in the understanding of membrane protein assembly and structure. *Q. Rev. Biophys.* 32, 285–307
3 Bannwarth, M. *et al.* (2003) The expression of outer membrane proteins for crystallization. *Biochim. Biophys. Acta* 1610, 37–45
4 Kiefer, H. (2003) *In vitro* folding of α-helical membrane proteins. *Biochim. Biophys. Acta* 1610, 57–62
5 Korepanova, A. *et al.* (2005) Cloning and expression of multiple integral membrane proteins from *Mycobacterium tuberculosis* in *Escherichia coli*. *Protein Sci.* 14, 148–158
6 Daley, D.O. *et al.* (2005) Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* 308, 1321–1323
7 Huang, Y. *et al.* (2003) Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* 301, 616–620
8 Tate, C.G. and Grisshammer, R.K. (eds) (2003) Overexpression of integral membrane proteins. *Biochim. Biophys. Acta* 1610, 1–152
9 Wickner, W. *et al.* (2005) Protein translocation across biological membranes. *Science* 310, 1452–1456
10 Luirink, J. *et al.* (2005) Biogenesis of inner membrane proteins in *Escherichia coli*. *Annu. Rev. Microbiol.* 59, 329–355
11 Raine, A. *et al.* (2003) Targeting and insertion of heterologous membrane proteins in *E. coli*. *Biochimie* 85, 659–668
12 Kunji, E.R. *et al.* (2005) Eukaryotic membrane protein overproduction in *Lactococcus lactis*. *Curr. Opin. Biotechnol.* 16, 546–551
13 Baneyx, F. *et al.* (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* 22, 1399–1408
14 Sorensen, H.P. *et al.* (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* 115, 113–128
15 Tate, C.G. (2001) Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett.* 504, 94–98
16 Goder, V. *et al.* (2004) Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol. Biol. Cell* 15, 1470–1478
17 Bowie, J.U. (2005) Solving the membrane protein folding problem. *Nature* 438, 581–589
18 Valent, Q.A. *et al.* (1997) Nascent membrane and presecretory proteins synthesized in *Escherichia coli* associate with signal recognition particle and trigger factor. *Mol. Microbiol.* 25, 53–64
19 Kota, J. *et al.* (2005) Specialized membrane-localized chaperones prevent aggregation of polytopic proteins in the ER. *J. Cell Biol.* 168, 79–88
20 Nagamori, S. *et al.* (2004) Role of YidC in folding of polytopic membrane proteins. *J. Cell Biol.* 165, 53–62
21 Matsumoto, H. *et al.* (1987) Gene encoding cytoskeletal proteins in *Drosophila rhabdomeres*. *Proc. Natl. Acad. Sci. U. S. A.* 84, 985–989
22 Helenius, A. *et al.* (2004) Roles of *N*-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* 73, 1019–1049
23 Yeliseev, A.A. *et al.* (2005) Expression of human peripheral cannabinoid receptor for structural studies. *Protein Sci.* 14, 2638–2653
24 Bogdanov, M. *et al.* (1996) A phospholipid acts as a chaperone in assembly of a membrane transport protein. *J. Biol. Chem.* 271, 11615–11618

25 Bogdanov, M. *et al.* (1999) Phospholipid-assisted refolding of an integral membrane protein. Minimum structural features for phosphatidylethanolamine to act as a molecular chaperone. *J. Biol. Chem.* 274, 12339–12345

26 Lee, A.G. (2004) How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta* 1666, 62–87

27 Hunte, C. (2005) Specific protein–lipid interactions in membrane proteins. *Biochem. Soc. Trans.* 33, 938–942

28 Arechaga, I. *et al.* (2000) Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of $F_{(1)}F_{(o)}$ ATP synthase. *FEBS Lett.* 482, 215–219

29 Kihara, A. *et al.* (1995) FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc. Natl. Acad. Sci. U. S. A.* 92, 4532–4536

30 Griffith, D.A. *et al.* (2003) A novel yeast expression system for the overproduction of quality-controlled membrane proteins. *FEBS Lett.* 553, 45–50

31 Kaufman, R.J. *et al.* (2002) The unfolded protein response in nutrient sensing and differentiation. *Nat. Rev. Mol. Cell Biol.* 3, 411–421

32 Bonander, N. *et al.* (2005) Design of improved membrane protein production experiments: quantitation of the host response. *Protein Sci.* 14, 1729–1740

33 Tolia, N.H. *et al.* (2006) Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods* 3, 55–64

34 Chen, Y. *et al.* (2003) DnaK and DnaJ facilitated the folding process and reduced inclusion body formation of magnesium transporter CorA overexpressed in *Escherichia coli*. *Protein Expr. Purif.* 32, 221–231

35 Lunin, V.V. *et al.* (2006) Crystal structure of the CorA Mg^{2+} transporter. *Nature* 440, 833–837

36 Tate, C.G. *et al.* (1999) Molecular chaperones stimulate the functional expression of the cocaine-sensitive serotonin transporter. *J. Biol. Chem.* 274, 17551–17558

37 Weiss, H.M. *et al.* (1998) Expression of an integral membrane protein, the 5HT5A receptor. *Methods Mol. Biol.* 103, 227–239

38 Andre, N. *et al.* (2006) Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Sci.* 15, 1115–1126

39 Stimpson, H.E. *et al.* (2006) Transferrin receptor-like proteins control the degradation of a yeast metal transporter. *EMBO J.* 25, 662–672

40 Miroux, B. *et al.* (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260, 289–298

[41] Dumon-Seignovert, L. *et al.* (2004) The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* 37, 203–206

42 Drew, D. *et al.* (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* 3, 303–313

43 Eshaghi, S. *et al.* (2005) An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. *Protein Sci.* 14, 676–683

44 Eroglu, C. *et al.* (2002) Functional reconstitution of purified metabotropic glutamate receptor expressed in the fly eye. *EMBO Rep.* 3, 491–496

45 Kuruma, Y. *et al.* (2005) Development of a minimal cell-free translation system for the synthesis of presecretory and integral membrane proteins. *Biotechnol. Prog.* 21, 1243–1251

46 Klammt, C. *et al.* (2004) High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* 271, 568–580

47 Elbaz, Y. *et al.* (2004) *In vitro* synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1519–1524

48 Pornillos, O. *et al.* (2005) X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science* 310, 1950–1953

49 Wallin, E. *et al.* (1995) Properties of N-terminal tails in G-protein coupled receptors: a statistical study. *Protein Eng.* 8, 693–698

50 Monne, M. *et al.* (2005) Competition between neighboring topogenic signals during membrane protein insertion into the ER. *FEBS J* 272, 28–36

51 Monne, M. *et al.* (2005) Functional expression of eukaryotic membrane proteins in *Lactococcus lactis*. *Protein Sci.* 14, 3048–3056

52 Andersson, H. *et al.* (2003) Membrane assembly of the cannabinoid receptor 1: impact of a long N-terminal tail. *Mol. Pharmacol.* 64, 570–577

53 Sarramegna, V. *et al.* (2005) Solubilization, purification, and mass spectrometry analysis of the human μ-opioid receptor expressed in *Pichia pastoris*. *Protein Expr. Purif.* 43, 85–93

54 Weiss, H.M. *et al.* (2002) Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*. *Eur. J. Biochem.* 269, 82–92

55 Tucker, J. *et al.* (1996) Purification of a rat neurotensin receptor expressed in *Escherichia coli*. *Biochem. J.* 317, 891–899

56 Roosild, T.P. *et al.* (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* 307, 1317–1321

57 Dalbey, R.E. *et al.* (1995) Directionality in protein translocation across membranes: the N-tail phenomenon. *Trends Cell Biol.* 5, 380–383

58 Feilmeier, B.J. *et al.* (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *J. Bacteriol.* 182, 4068–4076

59 Drew, D. *et al.* (2002) Rapid topology mapping of *Escherchia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2690–2695

60 Ito, K. *et al.* (2005) Cellular functions, mechanism of action, and regulation of Ftsh protease. *Annu. Rev. Microbiol.* 59, 211–231

61 White, J.F. *et al.* (2004) Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS Lett.* 564, 289–293

62 Drew, D. *et al.* (2005) A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* 14, 2011–2017

63 Auer, M. *et al.* (2001) High-yield expression and functional analysis of *Escherichia coli* glycerol-3-phosphate transporter. *Biochemistry* 40, 6628–6635

64 Cohen, S.L. *et al.* (2001) Mass spectrometry as a tool for protein crystallography. *Annu. Rev. Biophys. Biomol. Struct.* 30, 67–85

65 Parcej, D.N. *et al.* (2003) Structural characterisation of neuronal voltage-sensitive K^+ channels heterologously expressed in *Pichia pastoris*. *J. Mol. Biol.* 333, 103–116

66 Han, M.J. *et al.* (2003) Engineering *Escherichia coli* for increased productivity of serine-rich proteins based on proteome profiling. *Appl. Environ. Microbiol.* 69, 5772–5781

67 Lee, S.Y. *et al.* (2005) Systems biotechnology for strain improvement. *Trends Biotechnol.* 23, 349–358

68 DiTullio, P. *et al.* (1992) Production of cystic fibrosis transmembrane conductance regulator in the milk of transgenic mice. *Biotechnology* 10, 74–77

69 Pedelacq, J.D. *et al.* (2002) Engineering soluble proteins for structural genomics. *Nat. Biotechnol.* 20, 927–932

70 Long, S.B. *et al.* (2005) Crystal structure of a mammalian voltage-dependent Shaker family K^+ channel. *Science* 309, 897–903

71 Tornroth-Horsefield, S. *et al.* (2006) Structural mechanism of plant aquaporin gating. *Nature* 439, 688–694

72 Ellgaard, L. *et al.* (2003) Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* 4, 181–191

73 Wang, D.N. *et al.* (2003) Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies. *Biochim. Biophys. Acta* 1610, 23–36

74 Bjorklof, K. *et al.* (2002) Co- and posttranslational modification of the α_{1B}-adrenergic receptor: effects on receptor expression and function. *Biochemistry* 41, 4281–4291

75 Drew, D.E. *et al.* (2001) Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett.* 507, 220–224

76 Kawate, T. *et al.* (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* 14, 673–681

77 Jiang, Y. *et al.* (2002) Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* 417, 515–522

78 Wacker, M. *et al.* (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298, 1790–1793

79 Doyle, D.A. *et al.* (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* 280, 69–77