



The synthesis of recombinant membrane proteins in yeast for structural studies



Sarah J. Routledge^{a,b}, Lina Mikaliunaite^b, Anjana Patel^b, Michelle Clare^b, Stephanie P. Cartwright^b, Zharain Bawa^b, Martin D.B. Wilks^c, Floren Low^b, David Hardy^b, Alice J. Rothnie^b, Roslyn M. Bill^{b,*}

^a Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD, UK

^b School of Life & Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK

^c Smallpeice Enterprises Ltd, 27 Newbold Terrace East, Leamington Spa, Warwickshire CV32 4ES, UK

ARTICLE INFO

Article history:

Received 25 July 2015

Received in revised form 28 September 2015

Accepted 29 September 2015

Available online 30 September 2015

Keywords:

Membrane proteins

Yeast

Recombinant

Expression plasmid vector

Promoter

SMALP

Cholesterol

ABSTRACT

Historically, recombinant membrane protein production has been a major challenge meaning that many fewer membrane protein structures have been published than those of soluble proteins. However, there has been a recent, almost exponential increase in the number of membrane protein structures being deposited in the Protein Data Bank. This suggests that empirical methods are now available that can ensure the required protein supply for these difficult targets. This review focuses on methods that are available for protein production in yeast, which is an important source of recombinant eukaryotic membrane proteins. We provide an overview of approaches to optimize the expression plasmid, host cell and culture conditions, as well as the extraction and purification of functional protein for crystallization trials in preparation for structural studies.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Recombinant membrane protein production in yeast

The first crystal structures of mammalian membrane proteins derived from recombinant sources were solved in 2005 using protein that had been produced in yeast cells: the rabbit Ca²⁺-ATPase, SERCA1a, was overexpressed in *Saccharomyces cerevisiae* [1] and the rat voltage-dependent potassium ion channel, Kv1.2 was produced in *Pichia pastoris* [2] (Fig. 1a and e). Since then, several other host cells have been used for eukaryotic membrane protein production including *Escherichia coli*, baculovirus-infected insect cells and mammalian cell-lines [3]. Whilst all host systems have advantages and disadvantages, yeasts have remained a consistently-popular choice in the eukaryotic membrane protein field [4,5]. As microbes, they are quick, easy and cheap to culture; as eukaryotes they are able to post-translationally process eukaryotic membrane proteins. Very recent crystal structures of recombinant transmembrane proteins produced in yeast include those of human aquaporin 2, chicken bestrophin-1, the human TRAAK channel, human leukotriene C₄ synthase, an algal P-glycoprotein homologue and mouse P-glycoprotein using *P. pastoris*-derived samples; the

structures of the *Arabidopsis thaliana* NRT1.1 nitrate transporter, a fungal plant pathogen TMEM16 lipid scramblase and the yeast mitochondrial ADP/ATP carrier were solved using recombinant protein produced in *S. cerevisiae* (Fig. 1b–d, f–k).

Despite these successes (as well as others using recombinant protein produced in bacteria, insect cells and mammalian cell-lines; see <http://blanco.biomol.uci.edu/mpstruc/>), the overall rate of progress in membrane protein structural biology has, until very recently, been markedly slower than that in the soluble protein field [6]. However, recent experimental breakthroughs mean that the gap is set to narrow. For example, the use of stabilizing mutants has had a revolutionary impact on increasing the crystallization propensity of some membrane protein targets [7], while incorporating fusion partner proteins such as T4 lysozyme (T4L)¹ [8] has been particularly important in structural studies of G protein-coupled receptors (GPCR). From the perspective of the host

¹ Abbreviations: β-OG, *n*-octyl-β-D-glucopyranoside; CPM, N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide; DDM, *n*-dodecyl-β-D-maltopyranoside; DM, *n*-decyl-β-D-maltopyranoside; DMSO, dimethyl sulfoxide; GFP, green fluorescent protein; GPCR, G protein-coupled receptor; HSR, heat shock response; MNG, maltose neopentyl glycol; ORF, open reading frame; SEM, standard error of the mean; SMA, styrene maleic acid; SMALP, SMA lipid particles; T4L, T4 lysozyme; UPR, unfolded protein response; v/v, percent by volume of solute in the total volume of solution.

* Corresponding author.

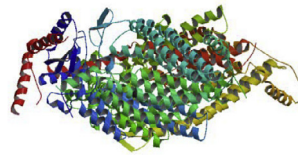
E-mail address: r.m.bill@aston.ac.uk (R.M. Bill).

Crystal structures solved using recombinant proteins produced in *S. cerevisiae*(a) Ca²⁺-ATPase, SERCA1a
(structure modelled on 1T5S)

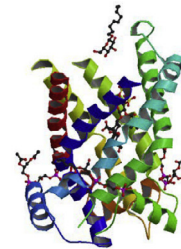
(b) NRT1.1 (4CL4; 2014)



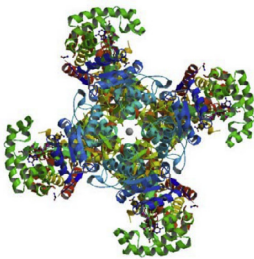
(c) TMEM16 (4WIS; 2014)



(d) ADP/ATP carrier (4C9G; 2014)

**Crystal structures solved using recombinant proteins produced in *P. pastoris***

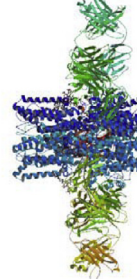
(e) Kv1.2 (2A79; 2005)



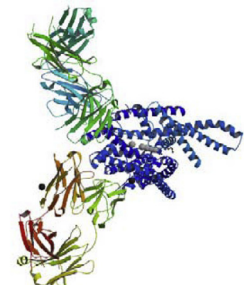
(f) Aquaporin 2 (4NEF; 2014)



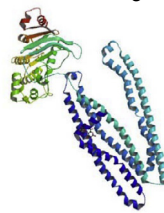
(g) Bestrophin-1 (4RDQ; 2014)



(h) TRAAK (4WFF; 2014)

(i) Leukotriene C₄ synthase (4JCZ; 2014)

(j) P-glycoprotein homologue (3WME; 2014)



(k) P-glycoprotein (4Q9H; 2015)



Fig. 1. Recombinant transmembrane proteins produced in yeast, for which crystal structures were published in 2014 and 2015; the first such structures, published in 2005, are also shown. The name of the protein and the host cell used in its recombinant production are given, together with the PDB code and publication year. Structural images, with the orientations shown, were downloaded directly from the PDB website (<http://www.pdb.org/pdb/home/home.do>) on 24th September 2015; protein chains are coloured from the amino-terminus to the carboxy-terminus using a spectral colour gradient. (a) The structure of recombinant Ca²⁺-ATPase, SERCA1a, produced in *S. cerevisiae* and published in 2005 [1], was not deposited in the PDB; the authors concluded that the structure of the yeast-expressed protein represents in full the native rabbit (*Oryctolagus cuniculus*) protein (PDB code of the native protein: 1T5S). (b) The structure of the *A. thaliana* NRT1.1 nitrate transporter at 3.7 Å (PDB code: 4CL4; [18]) was solved using a fusion protein with carboxy-terminal GFP and hexahistidine tags that had been produced in *S. cerevisiae*. (c) The structure of TMEM16 (from the plant pathogen, *Nectria haematococca*) at 3.30 Å (PDB code: 4WIS; [94]) was solved using protein produced in *S. cerevisiae* following expression screening of more than 80 constructs; the protein was produced as a fusion with a cleavable tag comprising enhanced-GFP and a decahistidine sequence. (d) The structure of the *S. cerevisiae* mitochondrial ADP/ATP carrier at 2.49 Å (PDB code: 4C9G; [111]) was solved using a fusion protein with an amino-terminal histidine tag and a factor Xa cleavage site produced in *S. cerevisiae*. (e) The structure of Kv1.2 at 2.9 Å (PDB code: 2A79; [2]) was solved using a recombinant rat (*Rattus norvegicus*) N207Q mutant with an amino-terminal octahistidine tag and TEV protease cleavage site; this construct was co-expressed with the β2 channel β subunit from rat brain in *P. pastoris*. (f) The structure of the *Homo sapiens* aquaporin 2 water channel was solved to 2.75 Å (PDB code: 4NEF; [112]) using recombinant protein produced in *P. pastoris* following codon optimization of the corresponding gene sequence. (g) The structure of bestrophin-1 at 2.85 Å (PDB code: 4RDQ; [95]) was solved after screening 30 orthologues for expression in *P. pastoris*; the recombinant protein comprised amino acids 1–405 of chicken (*Gallus gallus*) bestrophin-1 followed by an affinity tag (QGQF) that is recognized by an anti-tubulin antibody. (h) The structure of the *H. sapiens* TRAAK K⁺ channel at 2.5 Å (PDB code: 4WFF; [113]) was solved using protein produced in *P. pastoris* as an EGFP-decahistidine fusion protein with a carboxy-terminal truncation and all N-linked glycosylation sites removed. (i) The structure of the *H. sapiens* leukotriene C₄ synthase at 2.75 Å (PDB code: 4JCZ; [114]) was solved using protein produced in *P. pastoris* with a hexahistidine tag. (j) The structure of a P-glycoprotein homologue from *Cyanidioschyzon merolae* at 2.75 Å (PDB code: 3WME; [115]) was solved using protein produced in *P. pastoris* with a carboxy-terminal decahistidine tag. (k) The structure of P-glycoprotein at 3.4 Å (PDB code: 4Q9H; [116]) was solved using a glycosylation-deficient (mutations at N83, N87 and N90) *Mus musculus* amino acid sequence encompassing a carboxy-terminal hexahistidine tag.

cell, our improved understanding of cellular pathways controlling translation and protein folding, and how they influence functional recombinant protein yields, means it is now possible to select (or even design) better expression strains; this knowledge has also allowed a more strategic approach to cell culture in order to maximise the productivity of each cell [3]. Finally, new methods for extracting and solubilizing membrane proteins from the cell membrane using styrene maleic anhydride (SMA) co-polymers have enabled traditional detergents to be circumvented [9,10]. The benefits of this approach include improved thermostability of the solubilized protein and retention of protein–lipid interactions that

are normally disrupted during detergent-extraction [11]. This review focuses on current approaches to optimizing expression plasmids, yeast strains and culture conditions, as well as the extraction and purification of functional membrane proteins for crystallization trials (and subsequent structural studies) using detergents and SMA co-polymers.

2. Choice of yeast species: *S. cerevisiae* or *P. pastoris*?

Over 1500 species of yeast are known, but only very few of them have been employed as host organisms for the production

of recombinant proteins [12]. The two most widely used for recombinant membrane protein production are *S. cerevisiae* and *P. pastoris*. These single-celled, eukaryotic microbes grow quickly in complex or defined media (doubling times are typically 2.5 h in glucose-containing media) in formats ranging from multi-well plates to shake flasks and bioreactors of various sizes [12]. *P. pastoris* has the advantage of being able to grow to very high cell densities (>100 g/L dry cell weight; >500 OD₆₀₀ units/mL [5]) and therefore has the potential to produce large amounts of recombinant membrane protein for structural analysis (Fig. 1e–k). This yeast has also been important in generating high-resolution GPCR crystal structures such as the adenosine A_{2A} [13] and the histamine H₁ [14] receptors. However, because it is a strictly aerobic organism, the full benefits of *P. pastoris* are achievable only if it is cultured under highly-aerated conditions; this is usually only possible in continuously-stirred tank bioreactors.

S. cerevisiae has the advantage that its genetics are better understood (<http://www.yeastgenome.org/>) and that it is supported by a more extensive literature than *P. pastoris*. This has led to the development of a much wider range of tools and strains for improved membrane protein production (see Section 4). Consequently, projects requiring specialized strains may benefit from using *S. cerevisiae* as the host. Notably, the structure of the histamine H₁ receptor was obtained from protein produced in *P. pastoris*, although initial screening to define the best expression construct was performed in *S. cerevisiae* [15]. This is presumably because of the greater range of molecular biological tools available for *S. cerevisiae* at the screening stage, coupled with the superior yield characteristics of *P. pastoris* when cultured at larger scale in bioreactors. In principle, many of the tools established for *S. cerevisiae* could be transferred to *P. pastoris* (for which a genome sequence was published in 2009 [16]) combining the strengths of both yeast species, although such work would be time-consuming. In our laboratory, we often start with *P. pastoris* and, if the production is not straightforward, turn to *S. cerevisiae* to troubleshoot thereby benefitting from the best attributes of the two hosts [12].

3. Optimization of the yeast expression plasmid

Having decided which yeast species will be used as the recombinant host, a suitable expression plasmid needs to be selected or designed. Table 1 lists examples of common plasmids that are used for recombinant protein production in *S. cerevisiae* and *P. pastoris*, while Sections 3.1–3.3 briefly review three key elements of such plasmids: the promoter; the nature of any tags and the codon sequence.

Typically, episomal plasmids are used for expression in *S. cerevisiae*, but the expression cassette is integrated into the genome of *P. pastoris*. These continuing preferences may have resulted from the replication of early successes using particular plasmid/species combinations. Since the *P. pastoris* system depends upon very strong promoters, only a few copies of the gene (as typically

present in stably-integrated strains) are required to obtain sufficient levels of mRNA. In contrast, in *S. cerevisiae*, the promoter can be 10- to 100-fold weaker so the use of episomal plasmids with high copy numbers is advantageous; episomal plasmids are available for *P. pastoris*, but are not yet widely used in structural biology projects.

Auxotrophic markers are routinely used in *S. cerevisiae* plasmids to select for successfully-transformed yeast cells. Notably, the yield of the recombinant insulin analogue precursor protein was increased sevenfold simply by using the selection marker *URA3* instead of *LEU2* [17]. Truncations in the promoters of auxotrophic marker genes can further increase recombinant protein yields: by decreasing the promoter length, transcription of the marker gene on the plasmid is reduced and the cell compensates by increasing the plasmid copy number [17]. A truncated *LEU2* promoter was recently used to increase the yields of nine different transporters, including NRT1.1 [18] (Fig. 1b).

3.1. The promoter – constitutive or inducible?

Most recombinant expression systems employed in structural biology pipelines depend upon strong, inducible promoters to drive high rates of mRNA synthesis. For example, the strong *S. cerevisiae* promoter, P_{GAL1}, is induced with galactose while P_{AOX1} (a very strong *P. pastoris* promoter) is induced with methanol [5]. In choosing a strong promoter, the idea is that transcription should not be rate limiting. However, high mRNA synthesis rates may be countered by high rates of mRNA degradation [19]. Moreover, evidence from prokaryotic expression systems suggests that acquired mutations that lower promoter efficiency lead to improved functional yields of membrane proteins for some, but not all, targets [20]. In a separate study, a series of *E. coli* strains that had been evolved to improve their yield characteristics were found to have a mutation in the *hns* gene, which has a role in transcriptional silencing [21]. Together, these results support an emerging view that a suitable balance between mRNA and protein synthesis rates is desirable, although how this might be achieved in practice is not yet understood; one possibility might be a system based on slow, constitutive expression [22].

It has been proposed that the ideal inducible system would completely uncouple cell growth from recombinant synthesis, which requires the host cell to remain metabolically capable of transcription and translation in a growth-arrested state. In this scenario, all metabolic fluxes would be diverted to the production of recombinant protein [23]. While this approach is yet to be demonstrated for membrane protein production in yeast cells, soluble chloramphenicol acetyltransferase was produced to more than 40% of total cell protein in *E. coli* [24] suggesting that this may be a strategy worth exploring in yeast. Indeed, growth rates often (but not always) decline dramatically upon induction of yeast cultures, in part achieving this state.

When wild-type *P. pastoris* cells were cultured in methanol, it was found that a higher proportion of the total mRNA pool was

Table 1
Examples of commonly-used *S. cerevisiae* and *P. pastoris* expression plasmids.

Yeast	Plasmid	Selection marker	Tag	Promoter	Episomal or integrative	Supplier/reference
<i>S. cerevisiae</i>	pYES2	<i>URA3</i>	None	P _{GAL1} ; Inducible	Episomal	Invitrogen ^a
	pVTU260	<i>URA3</i>	His ₆	P _{ADHI} ; Constitutive	Episomal	Euroscarf ^b
<i>P. pastoris</i>	pPICZ	Zeocin	His ₆ ^c	P _{AOX1} ; Inducible	Integrative	Invitrogen ^c
	pGAPZ	Zeocin	His ₆ ^c	P _{GAP} ; Constitutive	Integrative	Invitrogen ^d
	pBGP1	Zeocin	None	P _{GAP} ; Constitutive	Episomal	[106]

^a https://tools.lifetechnologies.com/content/sfs/manuals/pyes2_man.pdf.

^b <http://web.uni-frankfurt.de/fb15/mikro/euroscarf/data/pVTU260.html>; <https://www.addgene.org/63682/>.

^c https://tools.lifetechnologies.com/content/sfs/manuals/ppicz_man.pdf; plasmids containing a 5' α mating factor pre-pro secretion signal sequence are also available.

^d https://tools.lifetechnologies.com/content/sfs/manuals/pgapz_man.pdf; plasmids containing a 5' α mating factor pre-pro secretion signal sequence are also available.

^e These plasmids also contain a c-myc epitope tag.

associated with two or more ribosomes (and therefore judged to be highly translated) compared to the same cells cultured in any other non-inducing growth condition [25]. This observation suggests that high recombinant protein yields in methanol-grown cells are due not just to promoter strength, but also to the global response of *P. pastoris* to growth on methanol [25]. However, P_{AOX1} -driven expression is leaky; the recent characterization of pre-induction expression under the control of P_{AOX1} [26] indicates that the uncoupling of growth and protein synthesis in *P. pastoris* cells has not yet been achieved. The response of a series of inducible *S. cerevisiae* promoters to different carbon sources has also been studied [22]; this type of careful analysis of promoter expression patterns now opens up opportunities for dynamic regulation of recombinant protein production in *S. cerevisiae*.

3.2. Tags and other fusion partners

In addition to the open reading frame (ORF) of the gene of interest, a typical expression plasmid will usually incorporate a number of other sequences in its expression cassette. The *S. cerevisiae* α -mating factor signal sequence is a common addition to commercial expression plasmids (Table 1) because it is believed to correctly-target recombinant membrane proteins to the yeast membrane. For example, its presence had a positive impact on the yield of the mouse 5-HT_{5A} serotonin receptor [27] but dramatically reduced expression of the histamine H₁ receptor [28]. Alternative signal sequences have been used (albeit much less frequently) such as the *STE2* leader sequence of the fungal GPCR, Ste2p [29].

Many expression plasmids contain tags as part of their DNA sequence (Table 1), and it is straightforward to add a range of others [30] by gene synthesis or polymerase chain reaction. Frequently-used tags for recombinant membrane proteins are polyhistidine (hexa-, octa- and decahistidine tags are all common), green fluorescent protein (GFP) and T4L. These and others have been reviewed extensively elsewhere [31,32]. Briefly, polyhistidine tags are routinely fused to recombinantly-produced membrane proteins to facilitate rapid purification by metal chelate chromatography using Ni-NTA resins. In many cases, the tag is not removed prior to crystallization trials, although protease cleavage sites can be engineered into the expression plasmid if this is desired [5]. GFP tags are used differently, typically to assess functional yield or homogeneity of the purified recombinant protein prior to crystallization trials. In the former case, caution must be exercised because GFP tags remain fluorescent in eukaryotic cells irrespective of whether the partner membrane protein is correctly folded in the plasma membrane [33]. GFP is therefore an inappropriate marker to assess the folding status of recombinant membrane proteins produced in yeast prior to extraction, although it is still useful in analyzing the stability of a purified membrane protein by fluorescence size-exclusion chromatography [34]. Finally, most GPCR crystal structures have been obtained using a fusion protein strategy where the flexible third intracellular loop is replaced by T4L, with modified T4L variants having been developed to optimize crystal quality or promote alternative packing interactions [8]. Overall, the precise combination and location (at either terminus or within the protein sequence itself) of any tags needs to be decided based upon their proposed use (for targeting, as an epitope, for purification or as a tool to assess protein quality) and the biochemistry of the recombinant membrane protein (since the exact fusion site may have a major impact on protein yield and quality).

3.3. Codon optimization

The sequence of an mRNA transcript is critically important in determining the rate and accuracy of translation [35] meaning that

optimal design of the corresponding DNA expression plasmid is essential to the success of a recombinant protein production experiment. Each organism is known to have a preference for some of the 64 available codons over others, but the biological reason for this is not yet clear. One idea is that each codon is decoded at a different rate: codons that are decoded quickly will be more resource efficient [36], while slower decoding will allow time for proper post-translational folding and translocation [37]. Another idea is that different codons are read with different accuracy, which might affect proteolysis and degradation [38]. Codon optimization involves manipulating the sequence of an ORF in order to maximize its expression. Several companies (e.g. GeneArt and GenScript) offer codon optimization services that account for codon bias in the host cell, mRNA GC content and secondary structure while minimizing sites such as internal ribosome entry sites or premature polyA sites that may negatively affect gene expression. However, there is no guarantee that recombinant protein yields will be increased, as demonstrated for the production of two membrane proteins in *E. coli* [39]. In contrast, careful codon optimization of the mouse P-glycoprotein gene for expression in *P. pastoris* led to substantially more recombinant protein compared to expression from the wild-type gene [40] (Fig. 1k). It has been proposed that the mRNA sequence around the translation start site has a bigger influence on membrane protein yields than codon choice in the rest of the ORF both in *E. coli* [39] and *P. pastoris* [41] since strong mRNA structure in this region could affect translation initiation and therefore protein production [39]. The use of degenerate PCR primers to optimize the codon sequence around the start codon therefore offers one approach to improving the expression plasmid [42].

4. Optimization of the host cell

4.1. *Saccharomyces cerevisiae*

As mentioned in Section 2, a wide range of *S. cerevisiae* resources are available, including comprehensive strain collections from which potential expression hosts can be selected. These resources are supported by a wealth of information in the *Saccharomyces* Genome Database (<http://www.yeastgenome.org/>). The yeast deletion collections comprise over 21,000 mutant strains with precise start-to-stop deletions of the approximately 6000 *S. cerevisiae* ORFs [43]. The collections include heterozygous and homozygous diploids as well as haploids of both *MATa* and *MAT α* mating types. Individual strains or the complete collection can be obtained from Euroscarf (<http://web.uni-frankfurt.de/fb15/mikro/euroscarf/>) or the American Type Culture Collection (<http://www.atcc.org/>). Complementing this, Dharmacon sells the Yeast Tet-Promoters Hughes Collection (yTHC) with 800 essential yeast genes under control of a tetracycline-regulated promoter that permits experimental regulation of essential genes. A number of specifically-engineered *S. cerevisiae* strains also exist including those with “humanized” sterol (see Section 4.3) and glycosylation pathways [44]. Notably, protease-deficient strains are a consistently-popular choice in membrane protein structural biology projects (Table 2).

Use of specific strains from these collections offers the potential to gain mechanistic insight into the molecular bottlenecks that preclude high recombinant protein yields; we and others have used transcriptome analysis to guide strain selection. In an early study we were able to identify genes that were up-regulated under high yielding conditions for our target membrane protein (the yeast glycerol facilitator, Fps1p) but down-regulated under low yielding conditions or *vice versa* [45]. This enabled us to select four high yielding strains: *srb5 Δ* , *spt3 Δ* , *gcn5 Δ* and yTHCBMS1. The use

Table 2
Examples of protease-deficient *S. cerevisiae* strains used in structural biology projects.

Membrane protein	Host species	<i>S. cerevisiae</i> production strain	Resolution (Å)	Reference
H ⁺ -translocating pyrophosphatase (H ⁺ -PPase)	<i>Vigna radiate</i>	BJ2168 (<i>MATa leu2 trp1 ura3-52 pep4-3 prb1-1122 prc1-407 gal2</i>)	2.35	[107]
Phosphate transporter (PiPT)	<i>Piriformospora indica</i>	DSY-5 (<i>MATa leu2 trp1 ura3-52 his3 pep4 prb1</i>)	2.9	[108]
CAAX protease, Ste24p	<i>Saccharomyces mikatae</i>	BJ54060 (<i>MATa ura3-52 trp1 lys2-801 leu2-Δ1 his3-Δ200 pep4:: HIS3 prb1-Δ1 6R can1 GAL*</i>)	3.1	[109]
Lipid scramblase, TMEM16	<i>Nectria haematococca</i>	FGY217 (<i>MATa, ura3-52, lys2Δ201, pep4Δ</i>)	3.3	[94]
Monoamine oxidase A	<i>Homo sapiens</i>	BJ2168 (<i>MATa leu2 trp1 ura3-52 prc1-407 ptb1-1122 pep3-3</i>)	2.2	[110]

of the *spt3Δ* strain resulted in the largest yields of Fps1p in shake-flasks (over 40-fold compared to wild-type cells). When the *yTHCBMS1* strain was cultured in the presence of 0.5 μg/ml doxycycline (to regulate the essential gene, *BMS1*, which is involved in the biogenesis of the 40S ribosomal subunit [46]), yields were increased by 30-fold in shake-flasks and over 70-fold in bioreactors compared with wild-type cells [47]. Using the strains *srb5Δ* and *gcn5Δ*, Fps1p yields were increased 5- and 10-fold over wild-type, respectively (Fig. 2). While these strains were originally selected to optimize Fps1p yields, we also noted generic advantages in that functional yields of the adenosine A_{2A} receptor and soluble GFP could be doubled using them [47]. This suggests that both general and target-specific effects are likely to occur during recombinant protein production in yeast (and indeed in any recombinant host cell). It would be desirable to be able to distinguish between the two, but this remains challenging because of the limited number of studies that have been done, including those in yeast.

Specific metabolic pathways have been targeted in order to increase functional recombinant protein yields in yeast cells. For example, exploiting the global cellular stress response to misfolded proteins has been investigated as a route to improving functional yields of recombinant proteins for structural studies [48]; it has even been argued that exposure to mild stress may enhance tolerance to a future stressful stimulus [49] such as that imposed during recombinant protein production. A recent study of recombinant GPCR production in *S. cerevisiae* demonstrated that mislocalized proteins were associated with the endoplasmic reticulum chaperone, BiP [50], providing opportunities to regulate the chaperone

network. The unfolded protein response (UPR) and the heat shock response (HSR) have also been examined; tuning expression levels to avoid or minimize UPR induction has previously been shown to increase functional membrane protein yields [51], while the HSR activates chaperones and the proteasome in order to relieve stress. HSR up-regulation has specifically 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 [52]. Overall, studies such as these demonstrate that the manipulation of stress responses may influence recombinant protein yields in yeast (and other organisms), but that the magnitude of any effect is protein specific.

4.2. *Pichia pastoris*

P. pastoris expression plasmids are usually integrated into the yeast genome to produce a stable production strain. Since it is not possible to precisely control the number of copies that integrate, the optimal clone must be selected experimentally. One approach is to screen on increasing concentrations of antibiotic (usually zeocin; Fig. 3) to obtain so-called “jackpot” clones. Although the results in Fig. 3 suggest a correlation between the copy number of the integrated expression cassette (as determined by resistance to increasing zeocin concentrations) and the final yield of recombinant protein, this is not always the case [53]. Sometimes clones with lower copy numbers are more productive, suggesting that the cellular machinery is overwhelmed in jackpot clones (resulting in misfolded or degraded protein). Consistent with this idea, adenosine A_{2A} receptor yields were increased

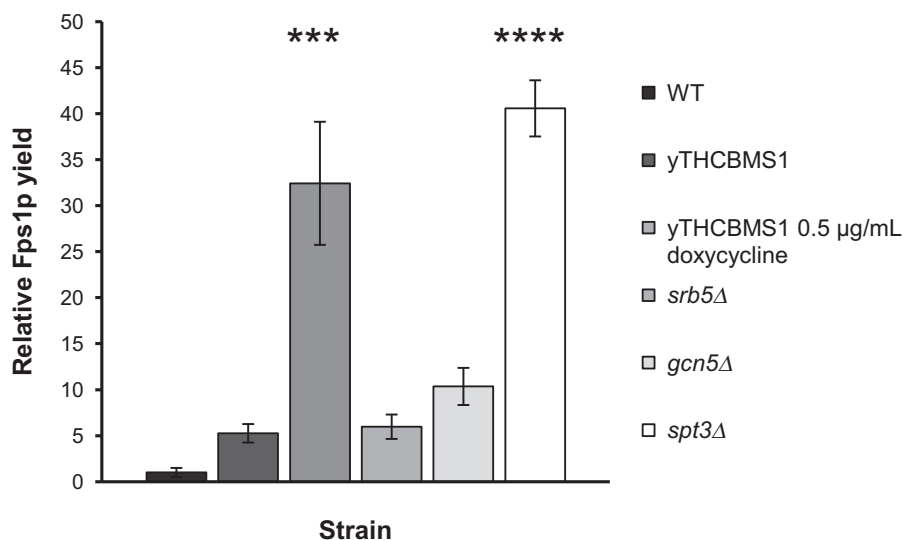


Fig. 2. Relative yields of the membrane protein, Fps1p, in high yielding strains. Yeast were grown in shake flasks in 2 × CBS until an OD₆₀₀ ~4. The data are relative yields from immunoblots where the error bars represent standard error of the mean (SEM; n = 3). Data were analysed using a one-way ANOVA (P < 0.0001). Asterisks show the significance of the relative yields of Fps1p in the different production strains compared to wild-type, as determined by a Dunnett's multiple comparison test, where ***P ≤ 0.001 and ****P ≤ 0.0001.

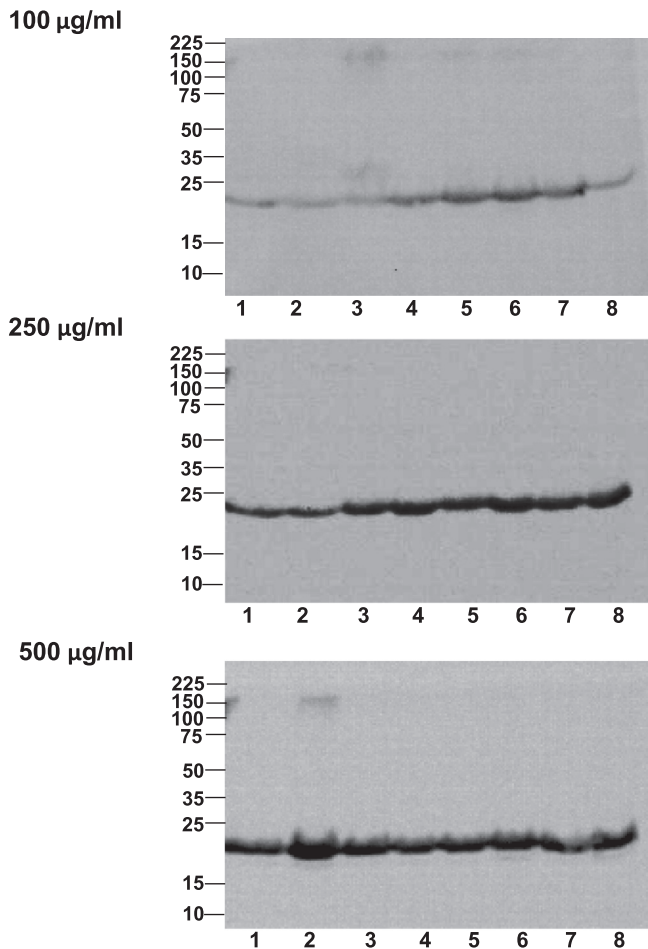


Fig. 3. Screening and scale-up of a *P. pastoris* production strain. A pPICZ α -based expression plasmid containing the human AQP1 gene was integrated into *P. pastoris* strain X33. Colonies were selected on YPD plates containing 100, 250 and 500 µg/ml zeocin. Eight colonies were isolated from each plate and cultured in BMGY to OD₆₀₀ = 1. Expression was induced by culturing the cells in BMMY for 24 h. Protein production was assessed by immunoblot of 100 µg total membrane protein per lane; the images indicate the variation in yield for the different colonies under different zeocin selection pressures. Molecular weight markers are shown in kDa.

1.8-fold when the corresponding gene was co-expressed in *P. pastoris* with the stress-response gene *HAC1* [54]; Hac1 drives transcription of UPR genes (see also Section 4.1).

In contrast to the situation in *S. cerevisiae*, many fewer *P. pastoris* strains are available in which to integrate the expression plasmid for the generation of a recombinant production strain. The wild-type strain, X33, the histidine auxotroph GS115, and the slow-methanol-utilization strain KM71H, have all been used to produce membrane proteins for structural studies [5]. Protease-deficient strains such as SMD1163, which lacks proteinase A and proteinase B, are also available. The structures of recombinant membrane proteins produced using *P. pastoris* that were published in 2014 and 2015 (Fig. 1) were all produced in one of the three mutant strains, SMD1163, KM71H and GS115 [3]. Notably production of human aquaporin 2 was actually done using an engineered GS115 strain in which the native aquaporin gene, *AQY1*, was deleted.

In all these strains, *P. pastoris* (like *S. cerevisiae*) post-translationally glycosylates membrane proteins by adding core (Man)₈-(GlcNAc)₂ groups, but not the higher-order structures found in humans and other mammals; compared to *S. cerevisiae*, the mannose chains also tend to be shorter. However, the effects of these non-native modifications are not necessarily detrimental

and need to be assessed for each individual protein [55]. Indeed, the high-resolution structure of a glycosylated form of the *Caenorhabditis elegans* P-glycoprotein (using recombinant protein produced in *P. pastoris*) demonstrates that yeast glycosylation does not necessarily hinder crystal formation [56]. Nonetheless, in order to overcome potential bottlenecks in producing, purifying, characterizing and crystallizing human proteins in yeast, engineered strains have been developed including strains with “humanized” glycosylation [57,58] and sterol pathways (see Section 4.3).

4.3. Engineering the yeast membrane for membrane protein production

The yeast membrane differs in composition from that of mammalian membranes. This is likely to be highly relevant to subsequent structural and functional studies of recombinant membrane proteins produced in yeast because lipids have a particularly important role in the normal function of membrane proteins; they contribute to membrane fluidity and may directly interact with membrane proteins.

In an attempt to “humanize” the yeast membrane, yeast strains have been developed that synthesize cholesterol rather than the native yeast sterol, ergosterol. This was achieved by replacing the *ERG5* and *ERG6* genes of the ergosterol biosynthetic pathway with the mammalian genes *DHRC24* and *DHRC7* [59–61], respectively. The gene products of *DHRC7* and *DHRC24* were identified as key enzymes that saturate sterol intermediates at positions C7 and C24 in cholesterol (but not ergosterol) synthesis (Fig. 4). Erg5p introduces a double bond at position C22 and Erg6p adds a methyl group at position C24 in the ergosterol biosynthetic pathway and therefore competes with the gene product of *DHRC24* for its substrate.

The yeast tryptophan permease, Tat2p, was unable to function in a yeast strain producing only ergosterol intermediates (instead of ergosterol), but in a cholesterol-producing strain activity was recovered to almost wild-type levels. Localization to the plasma membrane also appeared to correlate with the function of Tat2p [59]. The yeast ABC transporter, Pdr12p, although correctly localized to the plasma membrane, was inactive in a cholesterol-producing strain because of the lack of the key methyl group at position C24 [59] (Fig. 4). A similar scenario was observed with the function of yeast Can1p: the protein was localized to the plasma membrane regardless of the sterol produced, but function was lost when ergosterol production was disrupted [59]. The native yeast GPCR, Ste2p, which is involved in signal transduction, partially retained its function when cholesterol was produced instead of ergosterol. The agonist of Ste2p, MF α , retained potency on this receptor in both wild-type and cholesterol-producing strains. However, the efficacy appeared to be only half of that observed in the wild-type strain [60]. A positive outcome was observed when the human Na,K-ATPase $\alpha 3\beta 1$ isoform was expressed in a cholesterol-producing *P. pastoris* strain [61]: there was an improvement in recombinant yield and radio-ligand binding on intact cells, with the number of ligand binding sites in the cholesterol-producing strain increasing 2.5- to 4-fold compared to wild-type and protease deficient strains, respectively, both of which are ergosterol-containing [61].

Overall, studies on native yeast membrane proteins suggest that cell viability is not impaired in “humanized” yeast cells, although growth rates and densities are somewhat affected. However, this is likely to be an acceptable trade-off in return for higher yields of functional protein. Since a relatively small number of heterologous membrane proteins have been produced in cholesterol-producing yeast strains to date, potential exists to further optimize functional yields by using them.

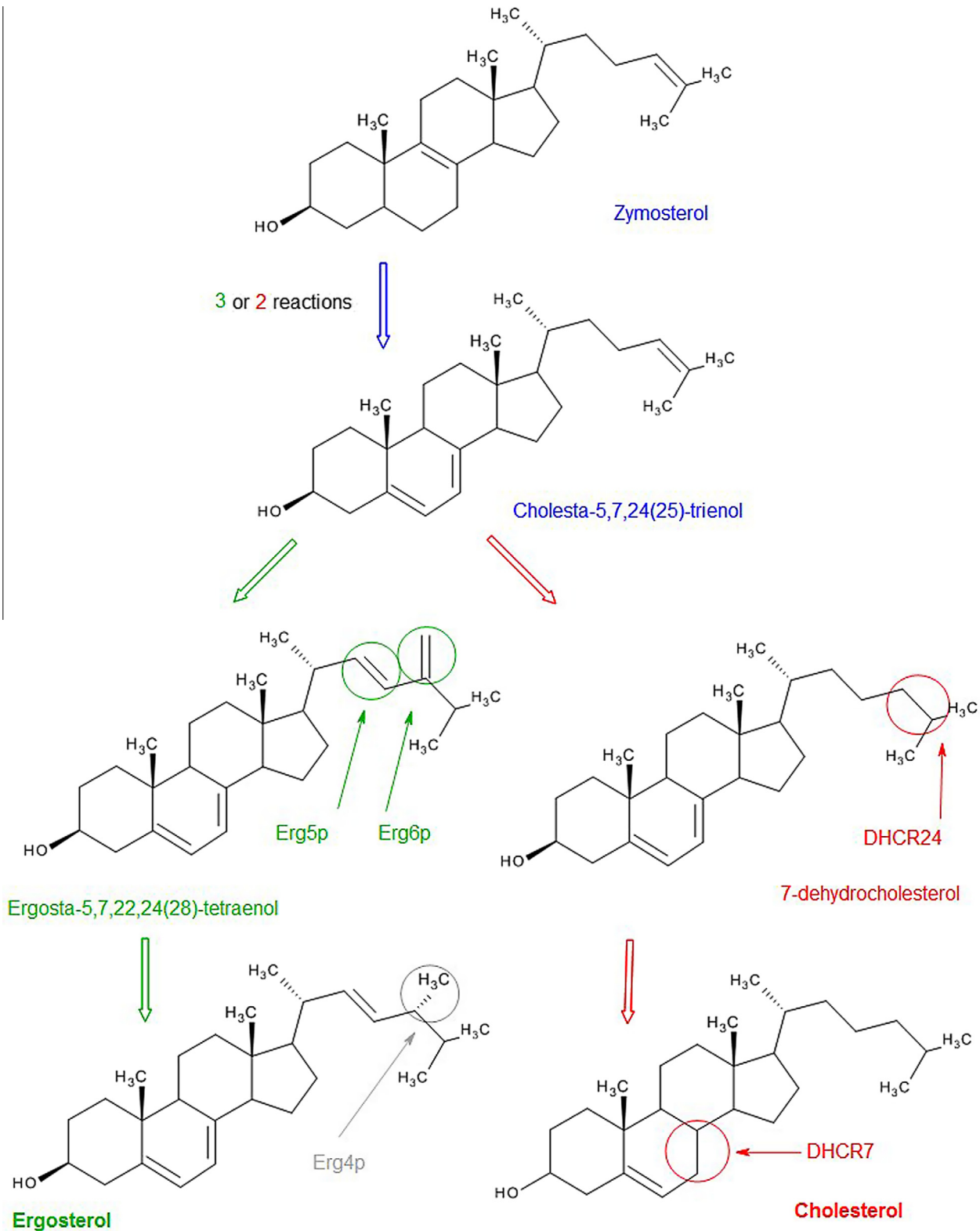


Fig. 4. Humanization of the yeast sterol pathway. The key products and intermediates in the mammalian cholesterol and fungal ergosterol biosynthesis pathways are shown. The pathways share common precursors, labeled in blue font, such as zymosterol which undergoes several chemical reactions (3 in yeast and 2 in mammalian cells) before being converted to cholesta-5,7,24(25)-trienol, which is effectively the branching point of the two pathways. The red circles and arrows indicate where the mammalian enzymes, DHCR7 and DHCR24, saturate bonds at positions 7 and 24 in cholesterol synthesis, respectively; they were identified as the enzymes required for cholesterol production in yeast. Yeast Erg6p and Erg5p of the ergosterol synthesis pathway, whose actions are shown in green, were deleted simultaneously with the introduction of the mammalian genes mentioned above because these enzymes would interfere with cholesterol synthesis in yeast. Erg4p, whose action is circled in grey, is the last enzyme in the pathway; since its substrate is the product of Erg6p catalysis, it does not interfere with cholesterol synthesis.

5. Optimization of yeast culture conditions

Recovering functional protein from recombinant host cells is dependent upon their capacity to synthesize an authentically-folded polypeptide. This requires the proper functioning of the transcription, translation and folding pathways [62]. During a recombinant protein production experiment, the maintenance and processing of an expression plasmid places a substantial metabolic burden on a cell, which means that these pathways must operate under abnormally stressful conditions [3]. A popular strategy to mitigate this burden is to decrease culture temperature; however, transcription, translation, polypeptide folding rates and membrane composition are also affected by low temperature stress [63]. This probably explains why increased yields are not always observed experimentally using that approach. Furthermore, many other variables are likely to affect yields including the composition of the growth medium, the pH and oxygenation of the culture, the inducer concentration and the point of induction.

5.1. Medium composition

Yeast cells grow quickly in complex or defined media; the selection and composition of suitable broths have been discussed elsewhere [12]. While higher yields are typically achieved in complex media, more control is possible in selective media, such as the ability to incorporate selenomethionine for anomalous dispersion phasing [64] in both *S. cerevisiae* [65] and *P. pastoris* [66].

The transcriptional and translational machinery of a cell respond to its growth rate, which is strongly affected by nutrient availability [3]. For example, several inducible and constitutive *S. cerevisiae* promoters have recently been characterised following growth on different carbon sources (glucose, sucrose, galactose and ethanol) and across the diauxic shift in glucose batch cultivation [22]. The study demonstrates that constitutive promoters differ in their response to different carbon sources and that expression under their control decreases as glucose is depleted and cells enter the diauxic shift [22]. Changes in nutrient source have also been found to alter the transcriptome and the global translational capacity of *P. pastoris* [25]. As discussed in Section 3.1, when *P. pastoris* cells were cultured in methanol, the majority of the total mRNA pool was associated with two or more ribosomes per mRNA (and therefore designated as highly-translated). Methanol is used to induce protein production under the control of P_{AOX1} in this yeast, suggesting that high recombinant protein yields may be associated with the global response of *P. pastoris* to methanol as well as promoter activity [26].

5.2. Additives

Several small molecules, sometimes referred to as chemical chaperones, have been investigated for their ability to enhance functional membrane protein yields. Specific improvements in yield have been reported following addition to recombinant yeast cultures of dimethyl sulphoxide (DMSO), glycerol, histidine and protein-specific ligands [67–69]. The effects of antifoams on protein yield, which are added to prevent foaming in bioreactor cultures, are discussed separately in Section 5.3.

The solvent DMSO has numerous biological applications, and is routinely used as a cryoprotectant and a drug vehicle [70]. Addition of DMSO to yeast cultures producing membrane proteins has been reported to have a positive effect on yield [71]. DMSO added at 2.5% v/v more than doubled the yield of 9 GPCRs (out of a panel of 20) produced in *P. pastoris*, with improvements of up to 6-fold [72]. In another study, the production in *S. cerevisiae* of a range of transporters fused to GFP was enhanced on average by 30%

following DMSO addition [73]. However, DMSO has also been reported to have no effect or, in some cases, negative effects upon membrane protein yields [72,74]. The underlying mechanisms are incompletely understood; DMSO is known to increase membrane permeability and cross membranes itself [70]. It has also been shown to upregulate the transcription of genes involved in lipid biosynthesis and increase phospholipid levels in *S. cerevisiae* [75]. When DMSO is added with stabilizing ligands, it may therefore improve the ability of these compounds to pass through the membrane and reach receptors in compartments within the cell [72].

Glycerol has been added to *S. cerevisiae* cultures producing human P-glycoprotein; at 10% v/v, yields were improved by up to 3.3-fold [67]. Glycerol is not as membrane-permeable as DMSO [70], so is thought to exert its effects by stabilizing protein conformation [67,69]. However, in another study, glycerol addition had a negative impact upon the yields of several membrane proteins produced in *S. cerevisiae* [73].

When producing recombinant membrane proteins such as GPCRs, ligands may be added at saturating concentrations to boost yields. Functional yields of GPCRs such as the β_2 -adrenergic receptor were tripled, the 5HT_{5A} receptor doubled [76] and the adenosine A_{2A} receptor doubled [77] by adding receptor-specific ligands. An optimization study demonstrated that addition of ligand could improve functional yields of 18 out of 20 GPCRs, with increases of up to 7-fold. However, a decrease in B_{max} was observed for two of the receptors investigated [72]. It is thought that ligands able to pass through the plasma membrane may bind to receptors as they fold during biosynthesis, thereby stabilizing them in the correctly-folded state. As a result, the level of functional receptors expressed at the plasma membrane is increased [78].

The amino acid, histidine, has been shown to double yields of some GPCRs when added to cultures at 0.04 mg/mL. Notably, its addition positively influenced fewer receptors than other additives such as DMSO [72]. Histidine addition did not have any effect upon the growth of the cells; instead it has been suggested that improved protein yields may result from its ability to protect yeast from oxidative stress [72,79].

Overall, it is clear that the use of a range of additives has improved recombinant membrane protein yields for diverse targets. In some cases additive effects have been synergistic, while in others their addition has been detrimental [72]. It is therefore important to systematically investigate the effects of additives on a case-by-case basis.

5.3. Other factors: temperature; pH; oxygenation

Membrane protein production is often done in bioreactors in order to obtain the large quantities of protein required for crystallization trials. Use of bioreactors enables the tight control of critical parameters, such as culture temperature, pH and the level of dissolved oxygen, thereby enabling the design of highly-reproducible bioprocesses. The most efficient way to select the optimal combination of these parameters is to use a design of experiments (DoE) approach [80]. DoE applies a structured test design to determine how combining input parameters set at different levels (e.g. pH set at 5, 6, 7; temperature set at 20 °C, 25 °C, 30 °C; dissolved oxygen concentration set at 30%, 40%, 50%) affects the output (recombinant protein yield). This efficient test design means that all experimental combinations do not need to be tested in order to derive the empirical relationship between the input parameters and protein yield in the form of a deterministic equation. The DoE approach is therefore a highly efficient way to obtain a quantitative understanding of how each factor and its interaction with all other factors affect final protein yield [80]. While this strategy is ideally executed in a bioreactor format, even in shake flasks yields can be improved by careful control of culture conditions [81].

One of the most important parameters in bioreactor cultures, especially of *P. pastoris* cells, is appropriate oxygenation. This is achieved by vigorous stirring and (if necessary) sparging of gases, which usually leads to foaming. The addition of chemical antifoaming agents is therefore required to manage and prevent the formation of foam. As additives to the process, these chemicals can affect both host cells and the recombinant proteins being produced; yields can be affected by the type of antifoam used, the concentration added, and whether production is undertaken in small shake flasks or in larger-scale bioreactors [82] and unpublished data). Although the biological effects of antifoams are not well understood, they have been shown to affect the volumetric mass oxygen transfer coefficient ($k_L a$) [83], influence growth rates of yeast [82,84] and are thought to alter membrane permeability [85]. While it was possible to more than double the yield of soluble GFP secreted by *P. pastoris* cells following the optimization of antifoam addition, the same conditions had detrimental effects on the functional yield of a recombinant GPCR produced in yeast (unpublished data). These findings highlight the importance of investigating the effects of antifoam addition; this often disregarded experimental parameter can significantly affect recombinant protein yields.

5.4. Induction conditions

In Section 3.1, we highlighted the fact that most recombinant expression systems employed in structural biology pipelines depend upon strong, inducible promoters. All promoters are known to vary in activity over time as well as in response to different carbon sources, which means that the timing of induction (and the concentration of the inducer) can be critical in obtaining the highest yields of functional protein; these parameters must be empirically determined. The response of a series of inducible *S. cerevisiae* promoters to different carbon sources has been studied [22] providing a framework for these types of experiments. We previously demonstrated the major impact of the induction regime on the yield of secreted GFP from *P. pastoris* cultures, showing the importance of matching the composition of the methanol feedstock to the metabolic activity of the cells [86]. P_{AOX1} is induced on methanol; however, when glucose (which has been shown to repress $AOX1$ expression) was the pre-induction carbon source, the adenosine A_{2A} receptor and GFP were still produced in the pre-induction phases of bioreactor cultures [26]. This study also reveals that a range of recombinant membrane proteins can be detected in the pre-induction phases of *P. pastoris* cultures when grown in bioreactors, but not shake-flasks. The results of all these investigations suggest that a DoE approach to selecting and optimizing induction phase conditions might be a particularly effective method of maximizing recombinant protein yields.

6. Extraction and solubilization of functional protein from yeast cells

The first steps in isolating a recombinant membrane protein are to break open the host cells and harvest the membranes. In yeast this requires breaking the cell wall, which needs harsher conditions than those typically used for insect, mammalian or *E. coli* cells. Typical methods for achieving this include high pressure (French Press or Emulsiflex-C3) or homogenization using glass beads shaken at high frequency followed by differential centrifugation [87].

6.1. Detergent-based solubilization of recombinant membrane proteins

In isolating a recombinant membrane protein from yeast membranes [87], the goal is to maintain structural integrity and

functionality. Depending on the protein target, this can be an extremely difficult task. However approaches are available to optimize the extraction process and the environment into which the target protein is being transferred.

Traditionally, detergents have been used for membrane protein extraction, purification and crystallization; the general principles have been reviewed extensively elsewhere [7,88]. Popular detergents include the non-ionic *n*-octyl- β -D-glucopyranoside (β -OG), *n*-decyl- β -D-maltopyranoside (DM) and *n*-dodecyl- β -D-maltopyranoside (DDM) [89]. Interestingly, the most commonly-used detergents to date are the same for yeast as for other expression systems, despite the differences in membrane composition. Optimization of detergent and buffer conditions must be done for each individual target membrane protein by assessment of protein stability and monodispersity. Unfortunately, membrane protein aggregation is a relatively common occurrence in these studies since detergents do not provide an exact mimic of the lipid environment in which the protein natively resides. Alternative amphiphiles have been designed to overcome these limitations and include novel compounds such as maltose neopentyl glycol (MNG) [90]. It has been suggested that for some target membrane proteins, MNG provides increased protein stability in comparison to detergents such as DM [91].

One useful technique to assess membrane protein stability prior to crystallization trials exploits a thiol-specific fluorochrome, N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM), which enables the investigator to assess the thermal stability of a recombinant membrane protein in a high-throughput format, therefore requiring small amounts of purified material [92]. In order to use this assay, the target membrane protein must have cysteine residues buried within the hydrophobic interior. Such residues bind thiol-specific CPM upon temperature-induced protein unfolding. CPM is essentially non-fluorescent until it reacts with a cysteine residue; therefore fluorescence can be recorded over time to determine the rate of protein unfolding. The influence of detergent type and concentration, salt concentration, pH, glycerol content and lipid addition on stability can all be investigated. Several studies have found a correlation between protein stability (as determined by the CPM assay) and the likelihood of obtaining well-ordered crystals for high resolution structure determination [93].

6.2. Functional reconstitution of detergent-solubilized recombinant membrane proteins

When determining the structure of a protein it is important to demonstrate that it is functional. For many membrane proteins, measuring function in the detergent-solubilized state can be difficult, either due to detergent effects (e.g. stripping away interacting lipids, lack of lateral pressure, protein denaturation) or because both 'sides' of the membrane are accessible. Therefore reconstitution of detergent-solubilized proteins into proteoliposomes is needed. Typically this involves the following steps: preparation of liposomes comprised of the desired lipids; destabilization of the liposomes with a detergent; mixing of detergent-purified protein with the liposomes; and removal of detergent using methods such as adsorption onto Bio-Beads SM-2 resin [18,94] or dialysis [95]. Several proteins expressed in *S. cerevisiae* [18,94] or *P. pastoris* [95–97] have been reconstituted into proteoliposomes and studied, showing that proteins produced in yeast are fully functional and comparable to those expressed in other cell systems.

6.3. Detergent-free methods – the emergence of SMALPs

Although all crystal structures of membrane proteins to date, including those synthesized in yeast, have used detergents for

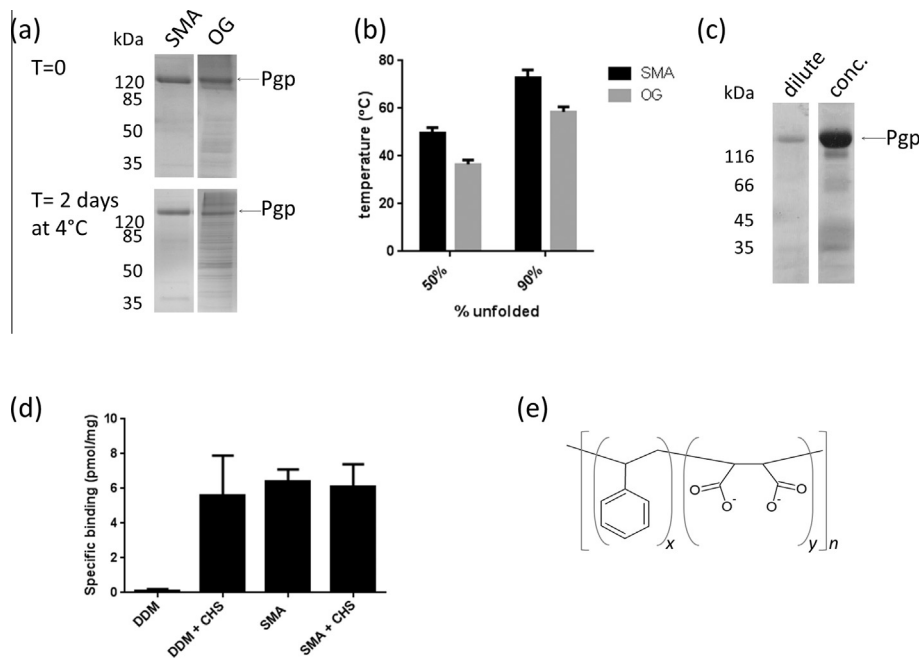


Fig. 5. SMALP-encapsulated membrane protein is more stable than its detergent-solubilized variant, and can easily be concentrated. (a) Stability over time. P-glycoprotein-His₁₂ (Pgp) was extracted from membranes using either 2.5% (w/v) SMA or 2% (w/v) β -OG (OG), and purified using Ni²⁺-NTA affinity chromatography. Samples of purified protein were analysed by SDS-PAGE and stained using Instant Blue. Following storage of the purified protein at 4 °C for 2 days, further samples were analyzed. The SMA-purified sample appears very similar, but the β -OG-purified sample shows significant breakdown over this time. (b) Thermostability. Unfolding of purified Pgp was monitored by fluorescence labeling with CPM after a 10 min incubation at various temperatures. OG-solubilized Pgp was 50% unfolded at a temperature of 37 °C, whereas SMALP-encapsulated Pgp was 50% unfolded after 10 min at 50 °C. (c) Purified SMALP-encapsulated Pgp at a concentration of 35 μ g/ml (dilute) was concentrated using an Amicon Ultra centrifugal concentrator (with a molecular weight cut-off of 30 kDa) to 1 mg/mL (conc), without significant loss or denaturation of protein. (d) Specific binding of [³H] ZM241385 (10 nM) to the adenosine A_{2A} receptor, extracted either with DDM or SMA in the absence or presence of CHS. Data are mean \pm SEM, $n = 3$. (e) Schematic of the structure of SMA.

extraction of the protein from the lipid bilayer, the use of detergents is not without problems. As mentioned in Section 6.1, screening for conditions and detergents that effectively extract the protein yet retain structure and stability can be difficult, time consuming and expensive. The environment produced by a detergent micelle does not fully mimic the lipid bilayer environment, as not only does the bilayer provide lateral pressure to stabilize the protein structure but interactions between the protein and its annular lipids can affect protein function. Notably, the most effective detergents for extraction are often not the best detergents for crystal formation. Recently a new detergent-free method for extraction of membrane proteins has emerged using SMA co-polymers (Fig. 5). The SMA inserts into biological membranes and forms small discs of lipid bilayer (10–12 nm) surrounded by the polymer, termed SMALPs (SMA lipid particles) [9,10], also known as lipodisks [98] or native nanodisks [11]. Membrane proteins within the SMALPs retain their annular lipid bilayer environment [9,11,99], yet the particles are small, stable and water soluble, allowing standard affinity chromatography methods to be used to purify a protein of interest [11,99–102].

To date this approach has been successfully applied to a wide range of transmembrane proteins from many different expression systems including both *S. cerevisiae* [98,100] and *P. pastoris* [102], for protein targets including GPCRs, ABC transporters and ion channels. Proteins within SMALPs have been shown to retain functional activity [11,98,100–102]. The small size of the particle and lack of interference from the polymer scaffold mean the SMALPs are ideal for many spectroscopic and biophysical techniques [9,10,100,102,103]. Importantly for structural studies, SMALP-encapsulated proteins have been found to be significantly more thermostable, less prone to aggregation, and easier to concentrate than detergent-solubilized proteins [11,100–102] (Fig. 5).

The importance of maintaining the lipid bilayer environment and lateral pressure is highlighted in Fig. 5d. When the adenosine A_{2A} receptor is extracted from *P. pastoris* membranes with detergent (DDM), it is necessary to supplement with the cholesterol analogue, cholesterol hemisuccinate (CHS), in order to retain any binding activity. However when the SMA co-polymer is used to extract the receptor, there is no requirement for CHS suggesting that it is not the cholesterol *per se* that is required for function of this protein, but some stabilizing interaction with lipids.

Although as yet, there are no reports of SMALP-encapsulated proteins being used to generate protein crystals, they have been used in both negative stain and cryo-single particle electron microscopy [100,101]. With recent technological and analytical advances within the field of electron microscopy the possibility of high resolution membrane protein structures using electron microscopy has become a reality [104,105]; SMALPs offer the ability for these structures to be obtained without stripping away the membrane environment from a transmembrane protein.

7. Conclusions

Yeast has an important role to play in membrane protein structural biology projects; since *S. cerevisiae* and *P. pastoris* are particularly amenable to genetic study, new insight may emerge that can lead to the design of improved experiments. One challenge is to identify which experimental parameters discussed in Sections 3–5, above, should be the focus in devising a production trial for a novel target. This is particularly demanding since these parameters may affect both host-cell- and target-protein-specific responses. Our understanding of the interlinked processes of transcription, translation and protein folding offers new opportunities to improve functional yields of recombinant membrane proteins

through strain selection and the choice of suitable culture conditions using DoE (Section 5). Coupled with new approaches to extraction and solubilization (Section 6), it is likely that the pace of solving new membrane protein structures is set to increase in the foreseeable future.

Acknowledgements

We acknowledge funding from the Biotechnology and Biological Sciences Research Council (BBSRC; via grants BB/I019960/1, BB/K013319/1 and BB/L502194/1 to RMB and BB/L015846/1 to AR) and the Innovative Medicines Joint Undertaking under Grant Agreement number 115583 to the ND4BB ENABLE Consortium to RMB and AR.

References

- [1] M. Jidenko, R.C. Nielsen, T.L. Sorensen, J.V. Moller, M. le Maire, P. Nissen, C. Jaxel, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 11687–11691.
- [2] S.B. Long, E.B. Campbell, R. Mackinnon, *Science* 309 (2005) 897–903.
- [3] R.M. Bill, T. von der Haar, *Curr. Opin. Struct. Biol.* 32 (2015) 147–155.
- [4] R.M. Bill, *Front Microbiol.* 5 (2014) 85.
- [5] B. Byrne, *Curr. Opin. Struct. Biol.* 32c (2015) 9–17.
- [6] B.P. Pedersen, P. Nissen, *Biochim. Biophys. Acta* 1850 (2015) 447–448.
- [7] I. Moraes, G. Evans, J. Sanchez-Weatherby, S. Newstead, P.D. Stewart, *Biochim. Biophys. Acta* 1838 (2014) 78–87.
- [8] T.S. Thorsen, R. Matt, W.I. Weis, B.K. Kobilka, *Structure* 22 (2014) 1657–1664.
- [9] T.J. Knowles, R. Finka, C. Smith, Y.P. Lin, T. Dafforn, M. Overduin, *J. Am. Chem. Soc.* 131 (2009) 7484–7485.
- [10] M. Jamshad, V. Grimard, I. Idini, T.J. Knowles, M.R. Dowle, N. Schofield, P. Sridhar, Y.P. Lin, R. Finka, M. Wheatley, O.R.T. Thomas, R.E. Palmer, M. Overduin, C. Govaerts, J.M. Ruyschaert, K.J. Edler, T.R. Dafforn, *Nano Res.* 8 (2015) 774–789.
- [11] J.M. Dorr, M.C. Koorengel, M. Schafer, A.V. Prokofyev, S. Scheidelaar, E.A. van der Cruysen, T.R. Dafforn, M. Baldus, J.A. Killian, *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) 18607–18612.
- [12] R.A. Darby, S.P. Cartwright, M.V. Dilworth, R.M. Bill, *Methods Mol. Biol.* 866 (2012) 11–23.
- [13] T. Hino, T. Arakawa, H. Iwanari, T. Yurugi-Kobayashi, C. Ikeda-Suno, Y. Nakada-Nakura, O. Kusano-Arai, S. Weyand, T. Shimamura, N. Nomura, A.D. Cameron, T. Kobayashi, T. Hamakubo, S. Iwata, T. Murata, *Nature* 482 (2012) 237–240.
- [14] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G.W. Han, T. Kobayashi, R.C. Stevens, S. Iwata, *Nature* 475 (2011) 65–70.
- [15] M. Shiroishi, H. Tsujimoto, H. Makyio, H. Asada, T. Yurugi-Kobayashi, T. Shimamura, T. Murata, N. Nomura, T. Haga, S. Iwata, T. Kobayashi, *Microb. Cell Fact.* 11 (2012) 78.
- [16] K. De Schutter, Y.C. Lin, P. Tiels, A. Van Hecke, S. Glinka, J. Weber-Lehmann, P. Rouze, Y. Van de Peer, N. Callewaert, *Nat. Biotechnol.* 27 (2009) 561–566.
- [17] A. Kazemi Seresht, P. Norgaard, E.A. Palmqvist, A.S. Andersen, L. Olsson, *Appl. Microbiol. Biotechnol.* 97 (2013) 3939–3948.
- [18] J.L. Parker, S. Newstead, *Nature* 507 (2014) 68–72.
- [19] A. Siwaszek, M. Ukleja, A. Dziembowski, *RNA Biol.* 11 (2014) 1122–1136.
- [20] S. Schlegel, J. Lofblom, C. Lee, A. Hjelm, M. Klepsch, M. Strous, D. Drew, D.J. Slotboom, J.W. de Gier, *J. Mol. Biol.* 423 (2011) 648–659.
- [21] N. Gul, D.M. Linares, F.Y. Ho, B. Poolman, *J. Mol. Biol.* 426 (2014) 136–149.
- [22] B. Peng, T.C. Williams, M. Henry, L.K. Nielsen, C.E. Vickers, *Microb. Cell Fact.* 14 (2015) 91.
- [23] S. Mahalik, A.K. Sharma, K.J. Mukherjee, *Microb. Cell Fact.* 13 (2014) 177.
- [24] D.C. Rowe, D.K. Summers, *Appl. Environ. Microbiol.* 65 (1999) 2710–2715.
- [25] R. Prielhofer, S.P. Cartwright, A.B. Graf, M. Valli, R.M. Bill, D. Mattanovich, B. Gasser, *BMC Genomics* 16 (2015) 167.
- [26] Z. Bawa, S.J. Routledge, M. Jamshad, M. Clare, D. Sarkar, I. Dickerson, M. Ganzlin, D.R. Poyner, R.M. Bill, *Microb. Cell Fact.* 13 (2014) 127.
- [27] H.M. Weiss, W. Haase, H. Michel, H. Reilander, *FEBS Lett.* 377 (1995) 451–456.
- [28] M. Shiroishi, T. Kobayashi, S. Ogasawara, H. Tsujimoto, C. Ikeda-Suno, S. Iwata, T. Shimamura, *Methods* 55 (2011) 281–286.
- [29] P. Sander, S. Grünwald, M. Bach, W. Haase, H. Reiländer, H. Michel, *Eur. J. Biochem.* 226 (1994) 697–705.
- [30] B. Miladi, C. Dridi, A. El Marjoui, G. Boeuf, H. Bouallagui, F. Dufour, P. Di Martino, A. Elm'selmi, *Mol. Biotechnol.* 55 (2013) 227–235.
- [31] K. Terpe, *Appl. Microbiol. Biotechnol.* 60 (2003) 523–533.
- [32] B.A. Manjasetty, A.P. Turnbull, S. Panjikar, K. Bussow, M.R. Chance, *Proteomics* 8 (2008) 612–625.
- [33] J. Thomas, C.G. Tate, *J. Mol. Biol.* 426 (2014) 4139–4154.
- [34] D. Drew, S. Newstead, Y. Sonoda, H. Kim, G. von Heijne, S. Iwata, *Nat. Protoc.* 3 (2008) 784–798.
- [35] S. Reuveni, I. Meilijson, M. Kupiec, E. Ruppim, T. Tuller, *PLoS Comput. Biol.* 7 (2011) e1002127.
- [36] D. Chu, T. von der Haar, *Nucleic Acids Res.* 40 (2012) 10098–10106.
- [37] S. Mahlab, M. Linial, *PLoS Comput. Biol.* 10 (2014) e1003294.
- [38] L. Ragionieri, R. Vitorino, J. Frommlet, J.L. Oliveira, P. Gaspar, L. Ribas de Poupplana, M.A. Santos, G.R. Moura, *FEBS J.* 282 (2015) 769–787.
- [39] M.H. Norholm, S. Toddo, M.T. Virkki, S. Light, G. von Heijne, D.O. Daley, *FEBS Lett.* 587 (2013) 2352–2358.
- [40] J. Bai, D.J. Swartz, I.I. Protasevich, C.G. Brouillette, P.M. Harrell, E. Hildebrandt, B. Gasser, D. Mattanovich, A. Ward, G. Chang, I.L. Urbatsch, *PLoS One* 6 (2011) e22577.
- [41] F. Oberg, M. Ekvall, M. Nyblom, A. Backmark, R. Neutze, K. Hedfalk, *Mol. Membr. Biol.* 26 (2009) 215–227.
- [42] K. Mirzadeh, V. Martinez, S. Toddo, S. Guntur, M.J. Herrgard, A. Elofsson, M.H. Norholm, D.O. Daley, *ACS Synth. Biol.* (2015).
- [43] G. Giaever, C. Nislow, *Genetics* 197 (2014) 451–465.
- [44] C. Arico, C. Bonnet, C. Javaud, *Methods Mol. Biol.* 988 (2013) 45–57.
- [45] N. Bonander, K. Hedfalk, C. Larsson, P. Mostad, C. Chang, L. Gustafsson, R.M. Bill, *Protein Sci.* 14 (2005) 1729–1740.
- [46] D. Gelperin, L. Horton, J. Beckman, J. Hensold, S.K. Lemmon, *RNA* 7 (2001) 1268–1283.
- [47] N. Bonander, R.A. Darby, L. Grgic, N. Bora, J. Wen, S. Brogna, D.R. Poyner, M.A. O'Neill, R.M. Bill, *Microb. Cell Fact.* 8 (2009) 10.
- [48] C.L. Young, A.S. Robinson, *Curr. Opin. Biotechnol.* 30 (2014) 168–177.
- [49] D. Chasman, Y.H. Ho, D.B. Berry, C.M. Nemecek, M.E. MacGilvray, J. Hose, A.E. Merrill, M.V. Lee, J.L. Will, J.J. Coon, A.Z. Ansari, M. Craven, A.P. Gasch, *Mol. Syst. Biol.* 10 (2014) 759.
- [50] M.A. O'Malley, J.D. Mancini, C.L. Young, E.C. McCusker, D. Raden, A.S. Robinson, *Protein Sci.* 18 (2009) 2356–2370.
- [51] D.A. Griffith, C. Delipala, J. Leadsham, S.M. Jarvis, D. Oesterhelt, *FEBS Lett.* 553 (2003) 45–50.
- [52] J. Hou, T. Osterlund, Z. Liu, D. Petranovic, J. Nielsen, *Appl. Microbiol. Biotechnol.* 97 (2013) 3559–3568.
- [53] R. Aw, K.M. Polizzi, *Microb. Cell Fact.* 12 (2013) 128.
- [54] M. Guerfal, S. Ryckaert, P.P. Jacobs, P. Ameloot, K. Van Craenenbroeck, R. Derycke, N. Callewaert, *Microb. Cell Fact.* 9 (2010) 49.
- [55] T. Yurugi-Kobayashi, H. Asada, M. Shiroishi, T. Shimamura, S. Funamoto, N. Katsuta, K. Ito, T. Sugawara, N. Tokuda, H. Tsujimoto, T. Murata, N. Nomura, K. Haga, T. Haga, S. Iwata, T. Kobayashi, *Biochem. Biophys. Res. Commun.* 380 (2009) 271–276.
- [56] M.S. Jin, M.L. Oldham, Q. Zhang, J. Chen, *Nature* 490 (2012) 566–569.
- [57] B. Laukens, C. De Wachter, N. Callewaert, *Methods Mol. Biol.* 1321 (2015) 103–122.
- [58] S.R. Hamilton, P. Bobrowicz, B. Bobrowicz, R.C. Davidson, H. Li, T. Mitchell, J.H. Nett, S. Rausch, T.A. Stadheim, H. Wischniewski, S. Wildt, T.U. Gerngross, *Science* 301 (2003) 1244–1246.
- [59] C.M. Souza, T.M. Schwabe, H. Pichler, B. Ploier, E. Leitner, X.L. Guan, M.R. Wenk, I. Riezman, H. Riezman, *Metab. Eng.* 13 (2011) 555–569.
- [60] S. Morioka, T. Shigemori, K. Hara, H. Morisaka, K. Kuroda, M. Ueda, *Appl. Microbiol. Biotechnol.* 97 (2013) 4013–4020.
- [61] M. Hirz, G. Richter, E. Leitner, T. Wriessnegger, H. Pichler, *Appl. Microbiol. Biotechnol.* 97 (2013) 9465–9478.
- [62] H. Hussain, R. Maldonado-Agurto, A.J. Dickson, *Biotechnol. Lett.* 36 (2014) 1581–1593.
- [63] C. Barria, M. Malecki, C.M. Arraiano, *Microbiology* 159 (2013) 2437–2443.
- [64] M.G. Malkowski, E. Quartley, A.E. Friedman, J. Babulski, Y. Kon, J. Wolfley, M. Said, J.R. Luft, E.M. Phizicky, G.T. DeTitta, E.J. Grayhack, *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007) 6678–6683.
- [65] D.A. Bushnell, P. Cramer, R.D. Kornberg, *Structure* 9 (2001) R11–14.
- [66] A.M. Larsson, T.A. Jones, *Methods Mol. Biol.* 389 (2007) 165–174.
- [67] R.A. Figler, H. Omote, R.K. Nakamoto, M.K. Al-Shawi, *Arch. Biochem. Biophys.* 376 (2000) 34–46.
- [68] M. Freigassner, H. Pichler, A. Glieder, *Microb. Cell Fact.* 8 (2009).
- [69] J.W. Welch, C.R. Brown, *Cell Stress Chaperon.* 1 (1996) 109–115.
- [70] Z.W. Yu, P.J. Quinn, *Biosci. Rep.* 14 (1994) 259–281.
- [71] A.K. Shukla, W. Haase, C. Reinhart, H. Michel, *Protein Expr. Purif.* 55 (2007) 1–8.
- [72] N. André, N. Cherouati, C. Prual, T. Steffan, G. Zeder-Lutz, T. Magnin, F. Pattus, H. Michel, R. Wagner, C. Reinhart, *Protein Sci.* 15 (2006) 1115–1126.
- [73] S. Newstead, H. Kim, G. von Heijne, S. Iwata, D. Drew, *Proc. Natl. Acad. Sci.* 104 (2007) 13936–13941.
- [74] A. Antaloae, V.C. Montigny, M. le Maire, K.A. Watson, T.L.-M. Sørensen, *PLoS ONE* 8 (2013).
- [75] Y. Murata, T. Watanabe, M. Sato, Y. Momose, T. Nakahara, S. Oka, H. Iwahashi, *J. Biol. Chem.* 278 (2003) 33185–33319.
- [76] H.M. Weiss, W. Haase, H. Michel, H. Reilander, *Biochem. J.* 330 (1998) 1137–1147.
- [77] N.J. Fraser, *Protein Expr. Purif.* 49 (2006) 129–137.
- [78] V. Bernier, D.G. Bichet, M. Bouvier, *Curr. Opin. Pharmacol.* 4 (2004) 528–533.
- [79] K. Murakami, Y. Onoda, J. Kimura, M. Yoshino, *IUBMB Life* 42 (1997) 1063–1069.
- [80] N. Bora, Z. Bawa, R.M. Bill, M.D. Wilks, *Methods Mol. Biol.* 866 (2012) 115–127.
- [81] K. Ukkonen, A. Vasala, H. Ojamo, P. Neubauer, *Microb. Cell Fact.* 10 (2011) 107.
- [82] S.J. Routledge, C.J. Hewitt, N. Bora, R.M. Bill, *Microb. Cell Fact.* 10 (2011).
- [83] H. Yagi, F. Yoshida, *J. Ferment. Technol.* 52 (1974) 905–916.
- [84] H. Mukaiyama, Y. Giga-Hama, H. Tohda, K. Takegawa, *Appl. Microb. Cell Physiol.* 85 (2009) 155–164.

- [85] J.L. Uma Maheswar Rao, T. Satyanarayana, *Lett. Appl. Microbiol.* 36 (2003) 191–196.
- [86] W.J. Holmes, R.A. Darby, M.D. Wilks, R. Smith, R.M. Bill, *Microb. Cell Fact.* 8 (2009) 35.
- [87] M. Jamshad, R.A. Darby, *Methods Mol. Biol.* 866 (2012) 237–246.
- [88] I. Moraes, M. Archer, *Methods Mol. Biol.* 1261 (2015) 211–230.
- [89] R.C. Oliver, J. Lipfert, D.A. Fox, R.H. Lo, S. Doniach, L. Columbus, *PLoS ONE* 8 (2013) e62488.
- [90] P.S. Chae, S.G. Rasmussen, R.R. Rana, K. Gotfryd, R. Chandra, M.A. Goren, A.C. Kruse, S. Nurva, C.J. Loland, Y. Pierre, D. Drew, J.L. Popot, D. Picot, B.G. Fox, L. Guan, U. Gether, B. Byrne, B. Kobilka, S.H. Gellman, *Nat. Methods* 7 (2010) 1003–1008.
- [91] H. Suzuki, T. Nishizawa, K. Tani, Y. Yamazaki, A. Tamura, R. Ishitani, N. Dohmae, S. Tsukita, O. Nureki, Y. Fujiyoshi, *Science* 344 (2014) 304–307.
- [92] A.I. Alexandrov, M. Mileni, E.Y. Chien, M.A. Hanson, R.C. Stevens, *Structure* 16 (2008) 351–359.
- [93] Y. Sonoda, S. Newstead, N.J. Hu, Y. Alguet, E. Nji, K. Beis, S. Yashiro, C. Lee, J. Leung, A.D. Cameron, B. Byrne, S. Iwata, D. Drew, *Structure* 19 (2011) 17–25.
- [94] J.D. Brunner, N.K. Lim, S. Schenck, A. Duerst, R. Dutzler, *Nature* 516 (2014) 207–212.
- [95] V. Kane Dickson, L. Pedi, S.B. Long, *Nature* 516 (2014) 213–218.
- [96] B. Zehnpfennig, I.L. Urbatsch, H.J. Galla, *Biochemistry* 48 (2009) 4423–4430.
- [97] T. Zollmann, G. Moiset, F. Tumulka, R. Tampe, B. Poolman, R. Abele, *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015) 2046–2051.
- [98] A.R. Long, C.C. O'Brien, K. Malhotra, C.T. Schwall, A.D. Albert, A. Watts, N.N. Alder, *BMC Biotechnol.* 13 (2013) 41.
- [99] D.J. Swainsbury, S. Scheidelaar, R. van Grondelle, J.A. Killian, M.R. Jones, *Angew. Chem. Int. Ed. Engl.* 53 (1807) (2014) 11803–11807.
- [100] S. Gulati, M. Jamshad, T.J. Knowles, K.A. Morrison, R. Downing, N. Cant, R. Collins, J.B. Koenderink, R.C. Ford, M. Overduin, I.D. Kerr, T.R. Dafforn, A.J. Rothnie, *Biochem. J.* 461 (2014) 269–278.
- [101] V. Postis, S. Rawson, J.K. Mitchell, S.C. Lee, R.A. Parslow, T.R. Dafforn, S.A. Baldwin, S.P. Muench, *Biochim. Biophys. Acta* 1848 (2015) 496–501.
- [102] M. Jamshad, J. Charlton, Y.P. Lin, S.J. Routledge, Z. Bawa, T.J. Knowles, M. Overduin, N. Dekker, T.R. Dafforn, R.M. Bill, D.R. Poyner, M. Wheatley, *Biosci. Rep.* 35 (2015).
- [103] I.D. Sahu, R.M. McCarrick, K.R. Troxel, R. Zhang, H.J. Smith, M.M. Dunagan, M. S. Swartz, P.V. Rajan, B.M. Kroncke, C.R. Sanders, G.A. Lorigan, *Biochemistry* 52 (2013) 6627–6632.
- [104] Y. Cheng, *Cell* 161 (2015) 450–457.
- [105] M. Liao, E. Cao, D. Julius, Y. Cheng, *Nature* 504 (2013) 107–112.
- [106] C.C. Lee, T.G. Williams, D.W. Wong, G.H. Robertson, *Plasmid* 54 (2005) 80–85.
- [107] S.M. Lin, J.Y. Tsai, C.D. Hsiao, Y.T. Huang, C.L. Chiu, M.H. Liu, J.Y. Tung, T.H. Liu, R.L. Pan, Y.J. Sun, *Nature* 484 (2012) 399–403.
- [108] B.P. Pedersen, H. Kumar, A.B. Waight, A.J. Risenmay, Z. Roe-Zurz, B.H. Chau, A. Schlessinger, M. Bonomi, W. Harries, A. Sali, A.K. Johri, R.M. Stroud, *Nature* 496 (2013) 533–536.
- [109] E.E. Pryor Jr., P.S. Horanyi, K.M. Clark, N. Fedoriw, S.M. Connelly, M. Koszelak-Rosenblum, G. Zhu, M.G. Malkowski, M.C. Wiener, M.E. Dumont, *Science* 339 (2013) 1600–1604.
- [110] S.Y. Son, J. Ma, Y. Kondou, M. Yoshimura, E. Yamashita, T. Tsukihara, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 5739–5744.
- [111] J.J. Ruprecht, A.M. Hellowell, M. Harding, P.G. Crichton, A.J. McCoy, E.R. Kunji, *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) E426–434.
- [112] A. Frick, U.K. Eriksson, F. de Mattia, F. Oberg, K. Hedfalk, R. Neutze, W.J. de Grip, P.M. Deen, S. Tornroth-Horsefield, *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) 6305–6310.
- [113] S.G. Brohawn, E.B. Campbell, R. MacKinnon, *Nature* 516 (2014) 126–130.
- [114] D. Niegowski, T. Kleinschmidt, U. Olsson, S. Ahmad, A. Rinaldo-Matthis, J.Z. Haeggstrom, *J. Biol. Chem.* 289 (2014) 5199–5207.
- [115] A. Kodan, T. Yamaguchi, T. Nakatsu, K. Sakiyama, C.J. Hipolito, A. Fujioka, R. Hirokane, K. Ikeguchi, B. Watanabe, J. Hiratake, Y. Kimura, H. Suga, K. Ueda, H. Kato, *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) 4049–4054.
- [116] P. Szewczyk, H. Tao, A.P. McGrath, M. Villaluz, S.D. Rees, S.C. Lee, R. Doshi, I.L. Urbatsch, Q. Zhang, G. Chang, *Acta Cryst. Sec. D Biol. Crystallogr.* 71 (2015) 732–741.