Expression of eukaryotic membrane proteins in eukaryotic and prokaryotic hosts

Athanasios Kesidis*, Peer Depping*, Alexis Lodé*, Afroditi Vaitsopoulou*, Roslyn M. Bill, Alan D. Goddard & Alice J. Rothnie**

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

*** these authors contributed equally**

**** corresponding author**

Highlights

- **Membrane proteins can be produced in a range of expression systems**
- **We compare bacterial, yeast, insect and mammalian cell expression systems**
- **The influence of construct design on protein folding and yield are discussed**

Abstract

The production of membrane proteins of high purity and in satisfactory yields is crucial for biomedical research. Due to their involvement in various cellular processes, membrane proteins have increasingly become some of the most important drug targets in modern times. Therefore, their structural and functional characterization is a high priority. However, protein expression has always been more challenging for membrane proteins than for soluble proteins. In this review, we present four of the most commonly-used expression systems for eukaryotic membrane proteins. We describe the benefits and drawbacks of bacterial, yeast, insect and mammalian cells. In addition, we describe the different features (growth rate, yield, post-translational modifications) of each expression system, and how they are influenced by the construct design and modifications of the target gene. Cost-effective and fast-growing *E. coli* is mostly selected for the production of small, simple membrane proteins that, if possible, do not require post-translational modifications but has the potential for the production of bigger proteins as well. Yeast hosts are advantageous for larger and more complex proteins but for the most complex ones, insect or mammalian cells are used as they are the only hosts able to perform all the post-translational modifications found in human cells. A combination of rational construct design and host cell choice can dramatically improve membrane protein production processes.

1. Introduction

Over the last few years, there has been a high demand for obtaining in-depth information for membrane proteins due to their association with human pathophysiology, which makes them targets for drug discovery [1-3]. The accessibility of membrane proteins, many of which are located on the cell surface in comparison with internal cellular targets, and the great variety of protein families and functions makes them highly important as potential drug targets. Examples of these targets are receptors (e.g. G protein-coupled receptors (GPCRs)), transporters (e.g. ATP-binding cassette (ABC) transporters or secondary active transporters) and channels (e.g. aquaporins or ion channels). Targeting these proteins can influence biological processes such as cell communication, secondary signalling and the transport or facilitated diffusion of nutrients and waste products in and out of the cell [4].

Despite recent, significant technical progress, membrane protein structural and functional characterization remains challenging. The major difficulty is obtaining large amounts of pure protein. This requires high yields produced by suitable hosts. A single optimised system that produces all membrane proteins is unlikely to be found and expression is still dependent on trial and error experiments. Several options have been explored in order to overcome this bottleneck in research. In some cases, researchers have focused on the use of modified expression systems that would improve protein yield, thanks to a better understanding of signalling pathways, protein folding and post-translational modifications [5-7]. Another option lies in protein engineering by using stabilizing mutant proteins which has allowed successful crystallization of the target protein, although it can have a detrimental impact on protein activity [8-10].

One of the key features in membrane protein production is the design of the construct that will be used to express the target protein. Thought must be given to specific modifications that could enhance the production of the protein such as promoter strength and fusion partners [11]. Another essential point is the selection of the organism in which the target protein will be expressed. Potential expression systems usually range from prokaryotic organisms such as *Escherichia coli* (*E. coli*) to eukaryotic systems such as yeast, insect cells or mammalian cells. The relationship between the lipid composition of the membrane in each of these organisms and the requirements of the recombinant protein in terms of its functionality is of paramount importance in determining its eventual stability and activity. The nature of any post-translational modifications can also be vital [12]. This review aims to cover the different aspects of the production of membrane proteins, from the design of the construct to the selection of a suitable expression system, comparing the four main ones: *E. coli*, yeast, insect cells and mammalian cells. We consider how the transcription and translation of the target gene is influenced by the choice of signal sequence, promoter and host strain, which combinations of host strain and expression vector are most commonly used and how they affect the proper folding and yield of the membrane protein. Furthermore, we review the value of multiple affinity tags and their use during and after expression.

2. Changing trends in membrane protein expression over the last decade

Thanks to technological and methodological advances, 76.5% of all known unique membrane protein structures were solved in the past decade. 48.7% of them are of eukaryotic proteins and

46.6% are of bacterial proteins (Figure 1A). The number of eukaryotic membrane protein structures elucidated per year and as a proportion of the total number of structures is continuously increasing (Figure S1A/B). In comparison, the number of bacterial structures per year has remained fairly constant, but as a proportion of all elucidated structures, bacterial ones are declining. In 2019, the proportion of eukaryotic structures was nearly 3 times higher than bacterial structures. This is an eight-fold increase since 2010 when the proportion of eukaryotic structures was three times lower (Figure S1C). This can be explained by an increased use of eukaryotic expression systems over the years (Figure 1B). 26.0% of eukaryotic membrane proteins of known structure have been expressed in cells from *S. frugiperda*, followed by HEK 293 (25.3%) and yeast cells (16.2%) including *S. cerevisiae* (8.5%) and *P. pastoris* (7.7%). The consistent increase in the number of eukaryotic membrane protein structures in comparison to bacterial structures is representative of the great interest and importance of their functional and structural investigation.

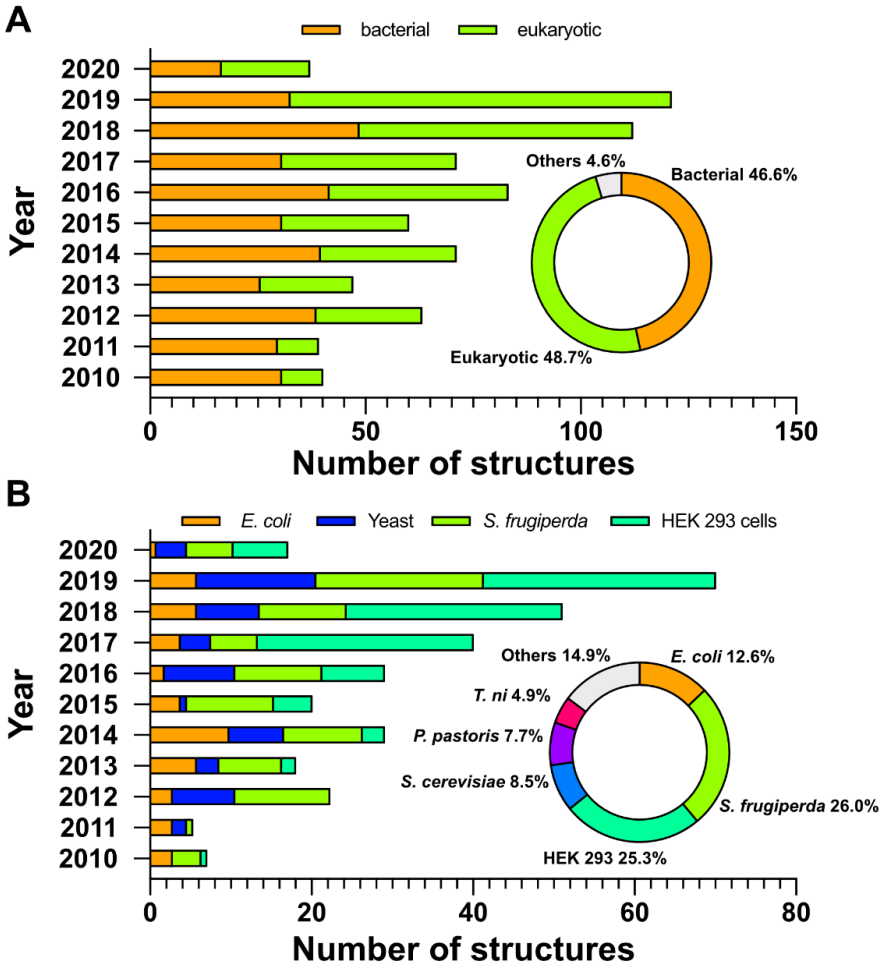

Figure 1: Analysis of eukaryotic membrane proteins of known structure from 2010 to 2020. (A) *796 structures have been solved. 48.7% of them were of eukaryotic and 46.6% of bacterial origin (see pie chart). Since 2015, the total number of elucidated structures per year is dominated by eukaryotic structures. The proportion of elucidated eukaryotic to bacterial structures has increased every year.* ***(B)*** *Predominantly,* S. frugiperda *and HEK 293 cells have been used for eukaryotic membrane protein expression in recent years with* E. coli *and yeast also being common hosts. The number of eukaryotic membrane proteins expressed in yeast cells only takes* S. cerevisiae *and* P. pastoris *expression systems into account. The pie chart considers all unique eukaryotic membrane protein structures elucidated and their expression systems from 2010 to 2020. Data were obtained from https://blanco.biomol.uci.edu/mpstruc/ [13].*

3. Construct design

The selection of an appropriate host strain for the expression of the target membrane protein is critical. However, this alone does not ensure the successful production of the protein. The design of the expression construct is essential and constitutes a key element for all recombinant expression systems (Figure 2). The overall aim is to generate a construct that is optimal for all stages of membrane protein production from transcription to translation and translocation-insertion into the membrane. Furthermore, downstream applications should be considered in the optimal design. Following expression, the protein will need to be detected and isolated from the membrane. A variety of tags are available, however they should be designed to have minimal effect on the function and structure of the protein. The final application, whether functional or structural study, dictates the modifications of the membrane protein. Structural studies demand high concentrations of pure, stable and homogenous protein preparations. Therefore, modifications of the protein coding sequence aim to increase the stability and the homogeneity of the overexpressed protein during purification and final applications.

Traditionally, constructs have been generated through molecular biology techniques such as restriction enzyme cloning, gateway recombination cloning, topoisomerase based cloning, Gibson Assembly, etc., [14-16] but this is increasingly being replaced with *de novo* synthesis of DNA. Even though these services facilitate the process, the manipulations of the gene need to be carefully evaluated. A study in 2012 showed that single synonymous codon substitutions can influence mRNA stability and structure, influence translation initiation and elongation and affect protein folding and translocation [17]. The rate and accuracy of translation are directly connected with the expression of the recombinant membrane protein and therefore the mRNA sequence plays a crucial role [18]. Several adjustments can be made to optimise the gene in order to match the transcription and translation machineries of the expression host. In principle, optimisation of the translation initiation sequence and adaptation to the codon usage of the expression host can have a positive impact on the production of the protein target. The translation initiation sequence has been shown to be influenced by the sequence around the start ATG (methionine) [19]. A study conducted in 2009, where 13 human aquaporins (AQP) were produced in *P. pastoris,* showed that the usage of a Kozak sequence instead of the yeast consensus sequence resulted in higher recombinant protein yields [20]. The Kozak sequence for mammalian proteins is gccRccatgG, where R represents purine (most often adenosine). A G (guanine) in the +4 position was common among the AQP genes with increased yields suggesting that the presence of alanine, and to a smaller extent glycine, at the second codon has a positive influence on protein expression. In a different report from 2013, synonymous codon substitutions in the 5' coding region adjacent to the AUG start significantly improved the expression of membrane proteins in *E. coli* [21]. However, this is not always the case. A study in 2012 showed that the translation initiation could be up-, or down-, regulated for the optimisation of the expression of each target protein. CD20 (containing 4-transmembrane (TM) helixes) and the 7TM GPCRs, EG-VEGFR1 and RA1c, were expressed in *E. coli* using different leader sequences. The proteins expressed with weaker translational initiation sequences had higher yields because they reduced the overload of the translocation machinery and allowed continuous membrane protein accumulation over several hours of induction [22].

There are 64 possible codons present in the genetic code but only 20 amino acids, meaning there is more than one codon for each amino acid (except methionine and tryptophan). The preference of organisms for the usage of codons varies for each amino acid. The mRNA sequences of highly-expressed genes contain codons associated with abundant tRNAs. Therefore, based on the codon bias of each expression system there should be appropriate codon optimisation of the DNA construct to maximize the expression of the protein. However, codon optimisation can have unexpected effects on the membrane protein expression. Codon optimisation was shown to have positive, negative or no effect on overexpressed membrane protein yield [23]. This highlights that any modifications should be target specific and the influence of codon optimisation on final protein yield should be carefully monitored. The frequency of each codon used in different organisms can be found at http://www.kazusa.or.jp/codon [24] and can be used for gene design.

A number of additional sequences are typically incorporated into a recombinant membrane protein expression construct, especially when the proteins are expressed heterologously in a different expression system. In order to control membrane protein targeting in recombinant expression systems, native signal sequences are often modified/replaced. Amino-terminal signal peptides, which are usually short 20 to 30 amino acid helices, mediate the trafficking of the protein to the membrane through a complex cellular machinery. The choice of signal sequence can have a dramatic effect on protein production. A study in 2013 showed that signal peptides can significantly increase expression levels of recombinant membrane proteins [25]. The most commonly-used signal sequence in yeast cells is the *S. cerevisiae* α-mating factor pre-pro-sequence on account of its high secretion efficiency; it is also used for expression in *P. pastoris*. The periplasmic leader sequences derived from ompT, ompA, pelB, phoA, malE, lamB and β-lactamase are used to direct proteins to the inner membrane of the periplasm in *E. coli*. In mammalian and insect cells, proteins are often expressed with their native signal sequence, although there are proteins, such as many family A GPCRs, which express in mammalian cells without a signal sequence. However addition of a signal peptide can help increase expression [26]. Examples of signal peptides that can be used include the baculoviral gp64, or in mammalian cells the modified human serum albumin (mSA), human azurocidin (AZ), modified *Cricetulus griseus* Ig kappa chain V III region MOPC 63 like (mIgkC) and modified human Ig kappa chain V III region VG (mIgkH) [26, 27].

Furthermore, the protein coding sequence can be modified to facilitate structural studies. As stated previously, these studies require high concentrations of pure and stable homogeneous protein preparations. One approach is the introduction of point mutations. Individual or multiple point mutations can increase the stability of membrane proteins. The alanine-scanning method has been used in the past to find point mutations that can increase the stability and facilitate the crystallisation of GPCRs as shown for the β_1-adrenergic receptor (β1 AR) [28], as well as A_{2A} adenosine receptor (A2A AR [29, 30]. This approach has also been used to avoid heterogeneity generated by post-translational modifications (PTMs). Two possibilities are the disruption of the N-glycosylation site sequence (Asn-X-Ser/Thr), usually achieved by mutating the Asn to Gln, as well as the substitution of palmitoylated Cys residues [31]. Caution should be exercised if the Cys residues are replaced by Ser to ensure that a potential kinase site is not inadvertently created. This can be performed for crystallography studies, where the goal is to obtain the structure of the protein. The glycosylation is therefore not needed in this particular case. Another approach is the replacement of unstructured loops with stably-folding domains such as T4-lysozyme (T4L) and the thermostabilized

apocytochrome b562RIL (BRIL) [32]. This approach has been mostly used for the stabilization of GPCRs expressed in insect cells, however it can also be applied in *E. coli* [33].

After expression and extraction, the next step is detection and purification of the target protein. For that reason, a broad range of epitope tags and fusion partners has been developed to simplify the process using quick and generic techniques and tools [34]. This is especially important for membrane proteins such as ion channels that lack significant antigenic epitopes. Chosen tags should have minimal effects on the protein structure and function. They are usually added onto the amino- or carboxy-terminus of a target protein through the addition of their cDNA sequence into the construct through gene synthesis or polymerase chain reaction (Figure 2). The carboxy-terminus is preferred when a signal peptide has been added to the amino-terminus for membrane translocation. However, a study in 2018 showed that the addition of more than one tag has been successful for the production and purification of complex membrane proteins including GPCRs and RAMPs [35]. The most commonly-used protein tag is the poly-histidine tag (His). The number of the histidine residues varies from 3 to 12. The His tag offers several advantages because it does not typically interfere with the protein due to its small size, it has low toxicity and offers a one-step purification by metal chelate chromatography using IMAC resins such as nickel-charged affinity resin Ni-NTA (nitrilotriacetic acid). Other frequently-used tags for recombinant membrane proteins are the poly-Arg, FLAG, c-myc, streptavidin-based, GST and green fluorescent protein (GFP) tags. The GFP tag is becoming an increasingly popular choice; it is used to monitor production yields and as a folding reporter [36]. Tags like the His, FLAG and GST can be used for protein binding on a surface e.g. for ligand binding experiments or the analysis of protein interactions with surface plasmon resonance spectroscopy or for antibody drug development by phage or mammalian display [37-41]. However, if tighter binding is required, the AviTag may be a better option for these kinds of experiments [42]. Biotinylating the AviTag makes it possible to bind the protein to streptavidin which is bound to a surface. The recently-developed SpyTag shows promising features for purification and surface binding by using the Spy&Go or SpyCatcher system [43, 44]. In the future, this may be a useful alternative to current, commonly-used tags.

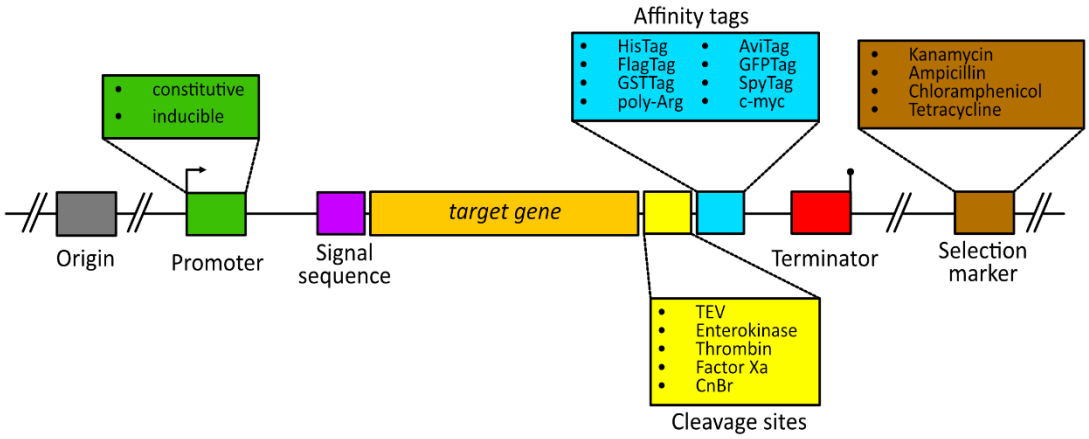

Figure 2: Construct design. *The cloning of the target gene into the expression plasmid can be achieved by various methods such as restriction enzyme, gateway recombination or topoisomerase based cloning. A signal sequence is inserted upstream of the gene so that the resultant membrane protein is translocated to the membrane. Tags can be cloned at the 5´- or 3´-end of the gene encoding the membrane protein. It is advisable to insert a cleavage site between the sequences in order to retain the option of removing the tag after protein purification. The sequence is flanked by the chosen promoter and terminator sequences. The expression vector must include a selection marker and an origin for replication.*

The presence of tags may influence the structure and/or function of the target membrane protein and therefore it is important to have the option to cleave the tag following purification. Tags can be removed by either enzymatic cleavage or chemical cleavage. Cleavage sites are used to enzymatically remove tags by exploiting the ability of enzymes to facilitate site-specific proteolysis. A commonly-used protease is tobacco etch virus (TEV) protease that can easily be produced in-house, is active in the presence of detergents and has high specificity. Alternatives such as enterokinase, thrombin and factor Xa have also been used successfully [45]. Chemical reagents can also be used for the removal of tags. The most commonly-used reagent is cyanogen bromide (CNBr). However, even though chemicals are cheaper than proteolytic enzymes, they are not as popular because their reaction conditions are harsh and can cause protein modifications. Furthermore, CNBr cleaves peptide bonds at the C-terminus of methionine (Met) residues and that can result in fragmented proteins that contain Met residues in their sequence. Nevertheless, they have been used successfully in the cleavage of fusion proteins from inclusion bodies [46].

4. Microbial expression systems

The two yeast species *Saccharomyces cerevisiae* and *Pichia pastoris* together with the bacterial *E. coli* system have been used for the production of the majority of membrane proteins expressed in microbes. A successful strategy starts with the selection of an appropriate host strain and continues with the vector, promoter, selection markers, signal sequence and the optimal gene sequence. The different factors in this design strategy for the overexpression of membrane proteins in bacteria and yeast are described in the following sections.

4.1 *E. coli* expression systems

Due to its potential for high expression efficiency, the possibilities for genetic manipulation and its fast growth rate, *E. coli* is an attractive expression system not only for soluble proteins but also for membrane proteins [47-49]. Overall, from 2010 to 2020, 49.9% of all recombinantly-expressed membrane proteins used *E. coli* as a host (Figure S2) (https://blanco.biomol.uci.edu/mpstruc/)[13]. 12.6% of eukaryotic membrane proteins of known structure have been expressed in *E. coli* (Figure 1B). These include small transmembrane units of membrane proteins used in NMR studies [50-52] but also large oligomers such as the retromer Vps5, the ion channel $Na_v1.4$ from electric eel and the 27 homomer Gasdermin A3 which were analysed using X-ray crystallography and electron microscopy [53-55]. However, most subunits of larger complexes were membrane proteins of smaller size. Overall, monomeric protein sizes lie between 40 and 750 residues (Figure S3A). Normally, protein expression takes 2-3 days in *E. coli,* the cost of medium is below £10/L and the protein yield is considerably higher in comparison to other conventionally used expression systems [56].

Even though *E. coli* provides many advantages for expression of eukaryotic membrane proteins, it may be undesirable as a host because of differences in membrane composition between eukaryotes and prokaryotes and the fact that the lack of eukaryotic post-translational modifications and processing can lead to non-functional and incorrectly-folded proteins [49, 57]. The overexpression of certain membrane proteins can be toxic for bacterial cells which can result in low or no yields at all [58]. Nevertheless, therapeutically-relevant eukaryotic membrane proteins have been successfully expressed in *E. coli* [59-64]. An alternative approach can be to create chimeric proteins combining

elements of the target eukaryotic membrane protein with a bacterial membrane protein which can more easily be expressed in bacteria [65, 66]. In the following sub-sections, the most common *E. coli* expressions systems are introduced.

4.1.1. Choosing the right E. coli *expression system*

The appropriate combination of expression strain and promoter is essential. A study in 2018 listed a total of 477 possible combinations of promoters and *E. coli* strains [48]. The most exploited bacterial strain for the production of membrane proteins is BL21(DE3), followed by C43(DE3), C41(DE3), pLysS-containing BL21(DE3), BL21(DE3) CodonPlus and Rosetta™ 2(DE3) (Figure 3A). BL21(DE3) cells have been used successfully in the homologous and heterologous production of membrane proteins [48, 67]. The majority of these proteins were expressed using the T7 RNA polymerase (T7RNAP)-based promoter system followed by the ara-, T5- and tet-promoters (Figure 3B). Additional commonly used plasmids and promoters for *E. coli* membrane protein expression can be found in tables 1 and 2.

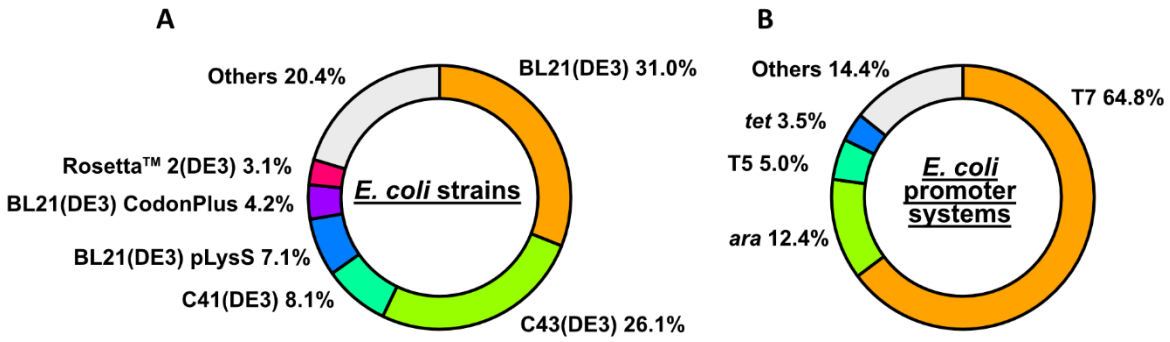

Figure 3: Most used* E. coli *strains (A) and promoter systems (B) used in the successful membrane protein expression of known structure. *Using data from Dilworth et al. 2018, bacterial expression experiments were analysed for strain and promoter usage in the time period 1985-2017* [48]. *BL21(DE3) and the T7 promoter are the most used strain and promoter, respectively.*

Table 1: Commonly used bacterial plasmids.

Plasmid Series	Promoter	Inducer	Repressor System	Tag	Selection Marker	Source
pET	T7/lac, T7	IPTG	LacI	None	Ampicillin (AmpR)	Invitrogen
pBAD	araBAD	Arabinose	araC	None	Ampicillin (AmpR)	Invitrogen
pRSET	T7/lac	IPTG	LacI	6xHis	Ampicillin (AmpR)	Invitrogen
pASK-IBA	tet	Anhydrotetracycline (AHT)	Tet-repressor	Various	Chloramphenicol (CamR)	IBA Life Sciences
pMAL	tac	IPTG	LacI	MBP	Ampicillin (AmpR)	NEB
pQE	T5	IPTG	LacI	8xHis	Ampicillin (AmpR)	Qiagen

pGEX	tac	IPTG	LacI	GST fusion	Ampicillin (AmpR)	GE Healthcare

Table 2: Commonly used bacterial promoters.

Promoter	Description	Regulation	Reference
lac	Promoter from lac operon	Inducible by lactose or IPTG. Repressed by lacI protein.	[68]
T7	Promoter from T7 bacteriophage	Constitutive	[69]
T7/lac	Promoter from T7 bacteriophage plus lac operators	Inducible by IPTG. Repressed by lacI protein.	[70]
araBAD	Promoter of the arabinose metabolic operon	Inducible by arabinose. Repressed by araC dimer.	[71]
tet	Promoter of tetracycline promoter	Inducible by AHT. Repressed by tetracycline transactivator (tTA) protein.	[72]
T5/lac	Promoter from T5 bacteriophage	Inducible by IPTG. Repressed by lacI protein.	[73]
trp	Promoter from E. coli tryptophan operon	Inducible by 3-β-indoleacrylic acid (IAA). Repressible by high levels of cellular tryptophan.	[74]
tac	Combination of promoters from the trp and lac operons	Inducible by IPTG. Repressed by lacI protein.	[75]

T7-based expression systems take advantage of the DE3 gene, encoding the T7RNAP and originating from T7 bacteriophages. In contrast, the bacteriophage T5 promoter can be used in any *E. coli* strain because it facilitates the native *E. coli* RNA polymerase and is not dependent on the DE3 gene. *E. coli* RNA polymerase is eight times slower than the T7RNAP thus producing less protein but decreasing the pressure on the expression machinery of the cell [76]. As already stated above, producing large amounts of membrane proteins can be toxic for host cells. To overcome this challenge, the more tolerant C41(DE3), C43(DE3), C44(DE3) and C45(DE3) strains (which are mutants of BL21(DE3) cells) can be used [6, 77]. In C41(DE3) and C43(DE3), mutations in the lac operon can be found [78]. These mutations reduce the basal expression of the target membrane protein, thereby diminishing any toxicity. Additionally, *lon* promoter activity in C43(DE3) cells is restored resulting in the expression of the *lon* protease. This helps with the removal of misfolded or partially translated proteins so that the Sec pathway is not blocked. In C44(DE3) and C45(DE3), the T7RNAP gene is mutated [77]. An additional stop codon in the sequence results in the production of truncated, non-functional polymerase. Therefore, the transcription of the target gene is dependent on the suppression of the stop codon. In comparison, BL21(DE3), containing the pLysS plasmid, represses toxic basal background expression of the target membrane protein by co-expressing T7 lysozyme which inhibits the transcription activity of T7RNAP [79]. Basal background expression is an undesired effect of the *lacI* repression system which is used by T7-based expression systems (Figure 4A). Basal expression is a function of the "leakiness" of an expression system or the production of the target membrane protein in the absence of an inducer (Figure 5B). The *lacI* gene, encoding the

lacI tetramer, is positioned on the host DNA and the expression vector [58]. The transcription of the target membrane protein is inhibited by the tetramer which blocks the T7RNAP binding site on the DNA strand. Adding isopropyl β-*D*-1-thiogalactopyranoside (IPTG) removes *lacI*. By using varying concentrations of IPTG, the amount of overexpressed membrane protein can be titrated. The *ara* and *tet* expression systems provide much stronger repressors. The *ara* expression system comprises the *ara*BAD promoter, *araC* promoter and *araC* gene (Figure 4B). Transcription from the *araC* promoter creates the *araC* dimer which binds to the I_1 binding site upstream of the *ara*BAD promoter and the O_2 binding site in front of the *araC* promoter, which results in a DNA loop that blocks both of the promoters [80]. The binding of *L*-arabinose to the *araC* dimer is followed by the displacement of the dimer from the O_2 binding site to the I_2 binding site which sits between the $O_{1L/1R}$ binding sites and the *ara*BAD promoter. The loop dissipates so that the transcription from the *ara*BAD promoter can be started. This repressor system is sometimes used with the MC1061 host strain because it does not metabolize *L*-arabinose [48, 81, 82]. Nevertheless, this repressor system often gives an "all-or-nothing" induction instead of a gradually increasing protein expression with different inducer concentrations [83]. The *tet* expression system is based on tetracycline binding and takes advantage of the *tet* operon which contains the *tetA* and *tetR* genes. *TetA* introduces an antibiotic resistance against tetracycline whereas *tetR* encodes the repressor which is blocking the RNA polymerase binding site but can be displaced by adding tetracycline [84, 85]. Similar to the T5 expression system, the *tet* expression system can be used in any *E. coli* strain.

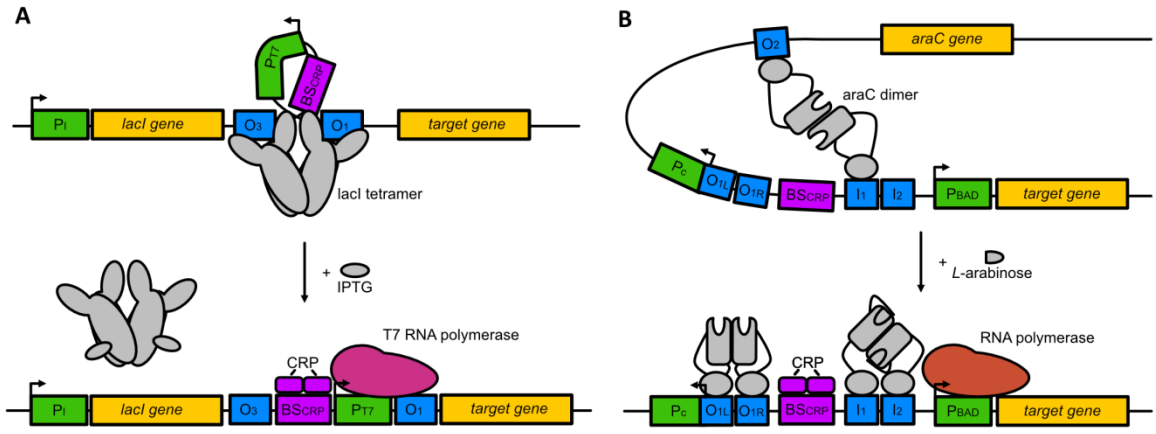

Figure 4: Repressor systems in* E. coli *expression systems. (A) *LacI repressor system. Due to the binding of the lacI tetramer, T7 RNA polymerase cannot bind to the T7 promoter (P_{T7}). Binding of IPTG displaces the tetramer and the transcription of the target gene is started. Additionally, the cAMP receptor binding protein (CRP), which regulates promoter activity using cAMP, can bind to its binding site (BS_{CRP}). **(B)** AraBAD repressor system. The DNA strand forms a loop by binding the araC dimer to the I_1 and O_2 binding site. The loop blocks the araBAD promoter (P_{BAD}). Binding of L-arabinose changes the conformation of the dimer so that it is displaced from O_2 and binds to I_2, thereby giving the RNA polymerase access to P_{BAD}. Excess araC binds to $O_{1L/1R}$ thereby autoregulating its own production by blocking P_C. P_I = lacI promoter; $O_{3/1}$ = lacI tetramer binding sites; BS_{CRP} = CRP binding site; CRP = cAMP receptor binding protein; P_{T7} = T7 promoter; P_C = araC promoter; $O_{2/1L/1R/I1/I2}$ = araC dimer binding sites; P_{BAD} = araBAD promoter.*

The heterologous expression of eukaryotic membrane proteins in *E. coli* is desirable but the lack of rare tRNAs is a limitation. The deficiency in required tRNAs can result in truncations, mutations, misfolding or low yields of the target membrane protein [86]. Therefore, strains such as BL21(DE3) Codon Plus and Rosetta™ 2(DE3) have been generated. They contain codons, encoding rare tRNAs needed for the satisfactory production of eukaryotic membrane proteins.

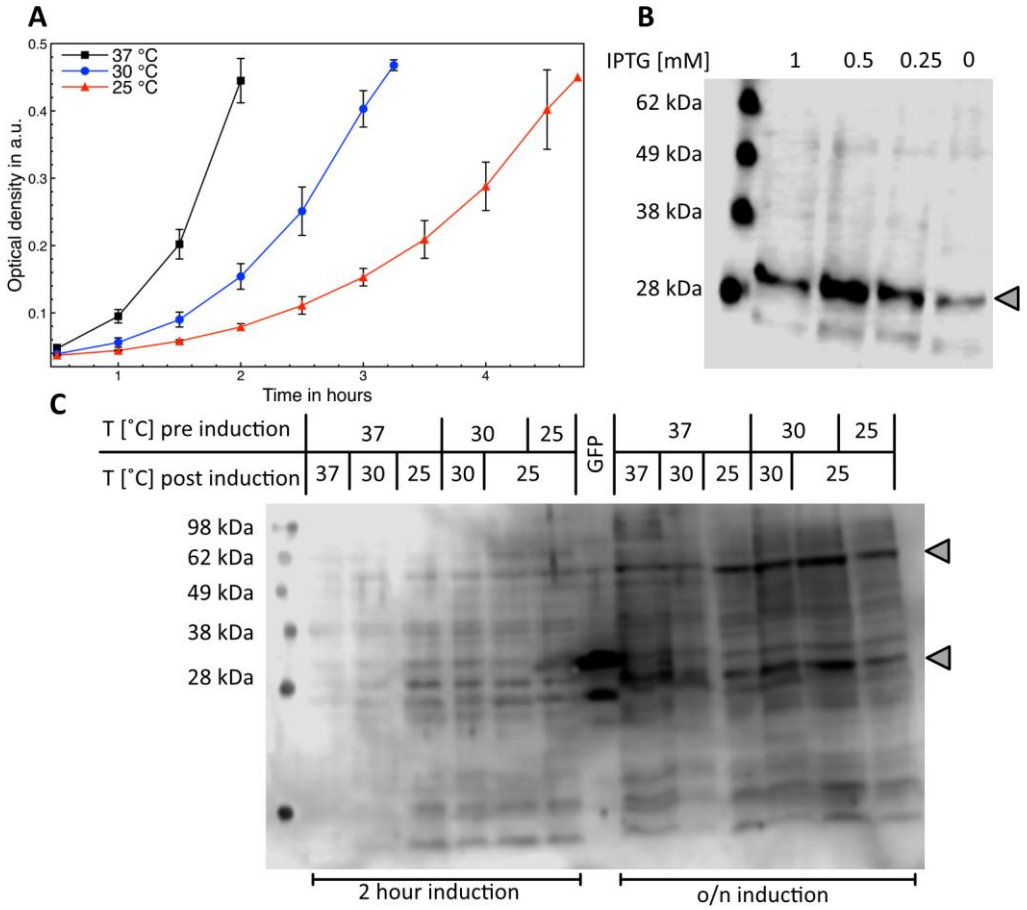

Figure 5: Finding the optimal overexpression conditions for a tetrameric ion channel in* E. coli.** ***(A) *Growth curves for BL21(DE3) cells transformed with the pBAT4 expression vector including the target gene at 37 °C, 30 °C and 25 °C. At 37 °C the desired OD_{600} of around 0.45 for induction is reached after 2 hours, whereas at 30 °C and 25 °C it takes at least 3 hours or 4.75 hours, respectively.* ***(B)*** *Western blot of the cell suspension after induction with either 1 mM, 0.5 mM, 0.25 mM or 0 mM IPTG. The optimal IPTG concentration for this experiment is 0.5 mM, as seen by the densest band on the blot indicated by the grey arrow above the 28 kDa molecular weight marker. Basal expression of the protein can be seen when no inducer was added to the culture. This shows the "leakiness" of the plasmid.* ***(C)*** *Western blot of the cell suspension grown and/or induced at 37 °C, 30 °C or 25 °C for 2 hours and overnight (o/n). In this case, a growth temperature of 30 °C and an o/n induction at 25 °C is optimal for the overexpression. This is indicated by the protein bands below the molecular weight marker of 28 kDa. The densest band can be seen at the described temperatures. In comparison, the protein bands around the 62 kDa marker display no differences. The protein bands are indicated by the grey arrows. His-tagged GFP in combination with anti-His antibodies were used as a positive control on the blot.*

4.1.2. Influence of expression conditions on protein yields

A very important variable in membrane protein production in *E. coli* is the choice of expression conditions. Insufficiently-regulated overproduction of the target membrane protein can result in slow cell growth, cell death and/or the formation of inclusion bodies [87, 88]. In the section above, various promoter and repressor systems have been described that reduce toxicity and regulate expression. The growth rate can be determined by measuring the optical density at 600 nm (OD_{600}) in a time-course experiment (Figure 5A). Even though *E. coli* grow optimally at 37 °C, protein production is not necessarily favoured at that particular temperature (Figure 5C). The reason for that

is the “leakiness” of some expression vectors and the metabolic stress on the bacterial cells [58]. If too much protein is produced at once, the native folding machinery can be overwhelmed [88]. If the protein is not folded correctly, it aggregates and thereby forms inclusion bodies. In addition to lowering the temperature, the concentration of the inducer can be decreased (Figure 5B). Optimizing the induction time can result in an improved functional protein yield as well. Due to its toxicity, the membrane protein may be destroyed by auto-proteolysis [58]. Therefore, shorter induction times e.g. 2 hours should be investigated.

4.2 Yeast expression systems

4.2.1 Yeast host system features

Recombinant membrane protein expression in yeast cells for crystallization studies was first used in 2005 in order to produce the rabbit Ca^{2+}-ATPase SERCA1a in *S. cerevisiae* and the rat voltage-dependant potassium ion channel Kv1.2 in *P. pastoris* [89, 90]. Ever since, yeast has remained a reliable option for eukaryotic membrane protein production. Indeed, membrane protein expression in yeast shows multiple advantages: as microbes, yeasts are inexpensive (<£10/L culture) and easy to grow compared to more complex eukaryotic cells [91]. They are also able to express complex proteins that require specific post-translational modifications such as disulphide bonds, glycosylation or folding that prokaryotic organisms are not able to process [92-94]. Yeast systems are capable of performing phosphorylation, prenylation -as it is a common characteristic of eukaryotic cells- and palmitoylation, of target proteins [95]. Phosphorylation consists in the reversible addition of phosphate groups that can be crucial for protein activity (it is estimated half of the enzymes in *S. cerevisiae* are phosphoproteins [96], while prenylation is the irreversible addition of prenyl groups in order to help protein anchoring on the cell membrane. Finally, palmitoylation is the reversible addition of a fatty acid such as palmitic acid on a cysteine residue that has an impact on subcellular trafficking and increases hydrophobicity. These modifications are essential for the correct function of proteins and their presence must be investigated on the protein of interest, especially for functional studies, before making a choice of expression system. Yeast system were mostly used to produce eukaryotic membrane proteins from 250 to 900 residues (Figure S3B) while *E. coli* systems are typically used for smaller proteins (Figure S3A). It was possible to efficiently produce proteins as large as up to 1,500 residues in length [97]. Yeast can also be very easily genetically modified [11, 94, 98]. Thus, yeast protein expression has become a viable alternative to bacteria for large-scale eukaryotic protein production. However, yeast need a longer time than bacteria in culture in order to reach the required cell density, and they need more time to express the desired recombinant protein (4 to 6 days in total), compared to more simple bacterial hosts. They remain faster than more complex eukaryotic expression systems as it takes about a week to cultivate the cells and express the target protein.

Several species of yeast have been tested as platforms for protein production. The two most successfully-used ones are *S. cerevisiae* and *P. pastoris*. Together, they have been used to produce 16.2% of recombinantly-expressed eukaryotic membrane proteins of known structure during the last decade (Figure 1B) [13]. Other species show interesting features such as *Hansenula polymorpha*, known for its thermostability (between 30 and 50 °C) with an optimal growth temperature around 37-43 °C which allows production of mammalian proteins that require temperatures of 37 °C to preserve their activity [99, 100]. *Candida boidinii* is also an interesting host as it can use different

methanol-inducible promoters that allow the control of the level of protein expression according to the carbon source used [101, 102]. These different yeast species have been reported to be able to produce gram quantities of mammalian soluble proteins per litre of culture supernatant [103].

S. cerevisiae has been very well studied in the literature so that tools and mutations for improved protein production are easily accessible. Indeed, a deletion library containing more than 20,000 strains that carry precise start-to-stop deletions of ~6,000 ORFs exists. This collection represents the first and only complete, systematically-constructed deletion collection available for any organism [104]. This can be extremely beneficial to allow tuning of the expression environment. For example, it is possible to modify sterol production in yeast in order to reach conditions closer to mammalian cells. As the natural sterol found in yeast membranes is ergosterol, and although the structural differences between ergosterol and cholesterol are minor, this may still result in a difference in interactions between the membrane protein and the lipids found in the native versus the host membrane. Therefore, specific strains have been produced that are able to make cholesterol instead of ergosterol in their membranes, thus allowing more native-like interactions for the protein [105]. A study from 2011 found that some proteins expressed in cholesterol-producing yeast strains showed good functionality (such as solute transporters for tryptophan or arginine), whereas others (such as a weak organic acid exporting protein called Pdr12p from the ABC transporters multi-drug resistance family) did not. More generally, the lipid composition in the yeast plasma membrane is very similar to the one found in mammalian cells: phosphatidylinositol has higher abundance in the plasma membrane whereas phosphatidylcholine is found in smaller amounts. In addition, the ratio of sterol/phospholipids is also slightly different; the choice of expression system must therefore take this into account [106]. However, production of mitochondrial membrane proteins is not affected, as the lipid composition of this organelle's membrane is almost identical between yeast and mammalian cells.

Yeast strains have also been engineered in order to induce protease deficiency. Host-specific protease activity can be a limiting factor in the production of high-yields of membrane proteins. A study has shown that the creation of protease-deficient strains, by disruption of two genes essential for maturation and activation of several proteases, can enhance recombinant protein yields by 10-fold [107].

One of the biggest disadvantages for the production of eukaryotic membrane proteins in yeasts is the difference in post-translational modifications, and especially glycosylation patterns, between higher eukaryotes and yeasts. Indeed, while N-linked glycans are quite similar between lower and higher eukaryotes, differences appear during the maturation in the Golgi apparatus. Moreover, there are major differences between O-glycosylation patterns between yeasts and higher eukaryotic systems, from the linkage between the different residues which is more linear in yeasts, to the composition in monosaccharides, which is much more complex in higher eukaryotes while it tends to contain only mannose in yeasts [108]. Protein glycosylation is of major importance as is involved in various pathways such as target receptor engagement, biological activity modulation, immunogenicity potential, tissue distribution or pharmacokinetics and pharmacodynamics [109]. Therefore, efforts have been made in order to engineer yeast strains capable of producing membrane proteins with O- and N-glycosylation patterns identical to human ones [110].

Yeast strains have been engineered to confer a selective advantage to cells producing the target protein, compared to those that do not, thus increasing the production yield [48, 111, 112].

Although the total yield of produced protein reported is higher, a similar yield of functional membrane proteins was produced in the modified and unmodified strains; this could be limited by the lipid composition of the membrane.

P. pastoris and *S. cerevisiae* to a lesser extent are able to grow rapidly in a range of complex media to a very high cell density when presented with high oxygenation rates (Figure 6), which means they are able to produce very large amounts of the target protein. In this particular experiment on average, for 1 L of culture, 10 g of cells are harvested, from which more than 1 g of membranes can be retrieved. This result is in accordance with previously-published studies using *P. pastoris* as an expression system. This is possible thanks to the ability of *P. pastoris* in not generating significant amounts of the toxic fermentative product, ethanol [113]. In addition, protein production in *P. pastoris* relies on very strong, methanol-inducible promoters (as *P. pastoris* strains are methylotrophic), so that there is no need for a high number of copies of the gene to result in a high yield of the protein. *P. pastoris* can also easily be grown in bioreactors, which tend to give higher yields of protein compared with shake flasks [114]. Increasing concentrations of antibiotic (e.g. Zeocin) can be used in order to select cells that have integrated the highest number of copies. However, the relationship between the number of copies of the gene of interest and the yield of recombinant protein is not always positive as a larger amount of protein can result in higher stress levels for the cells, leading to unfolded protein and protein degradation [115, 116]. The rate of protein production must therefore be adapted to the host's production capacity in order not to overwhelm its abilities, especially translocation to the membrane.

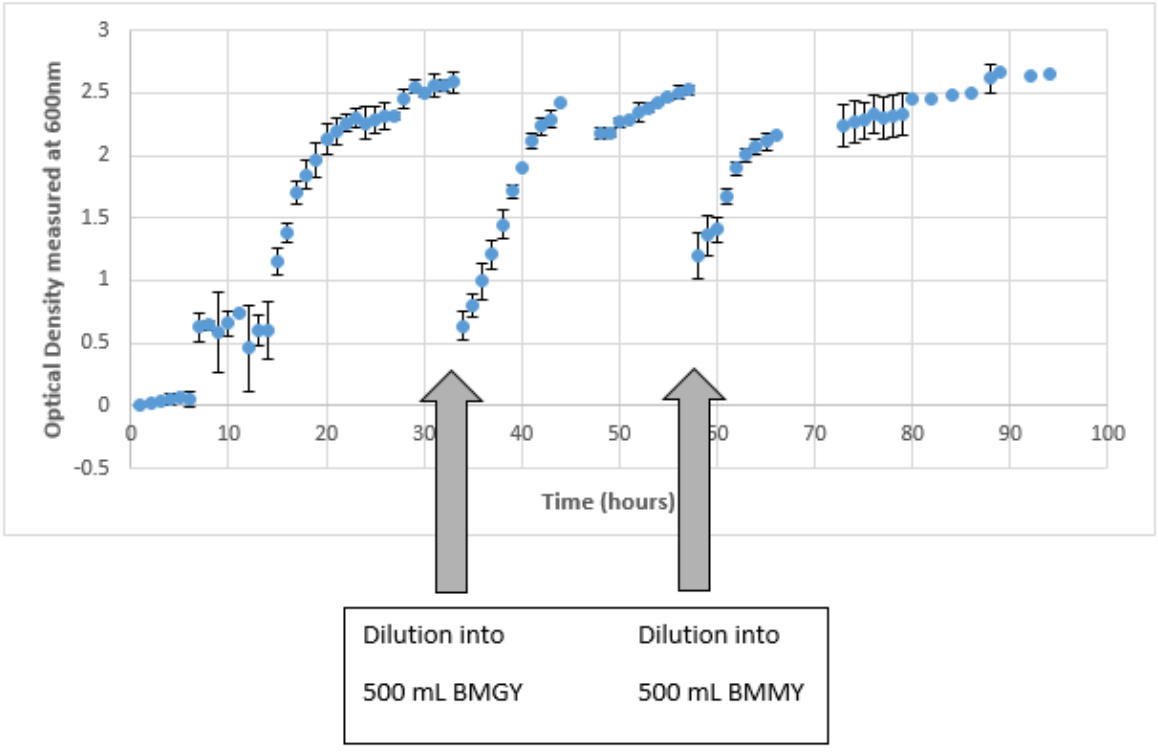

Figure 6: Using* Pichia pastoris *X33 cells to overexpress a membrane protein, monitoring growth by OD measurement at 600 nm. *A single colony was cultured in 50 mL BMGY medium supplemented with Zeocin at 30 °C and 200 rpm. After reaching an OD of 2.5 (after around 30 hours), 5 mL of the culture was diluted into 200 mL BMGY at 30 °C and 200 rpm. After 24 hours, the cells were harvested by centrifugation and diluted in 500 mL of BMMY (which contains methanol) to induce protein expression. After 24 hours at 30 °C and 200 rpm, the medium was supplemented with 5 mL 100% methanol for 24 more hours of induction.*

4.2.2 Overcoming the production of non-functional protein

Functional studies usually require a moderate amount of membrane protein. Structural studies have increased requirements in terms of yield, so agents that generate the UPR (Unfolded Protein Response), which is a suite of cellular responses following the build-up of misfolded proteins could prove beneficial [117]. Inducing a heat shock response can be done either by using a transcription factor inducing the heat shock pathway or by temperature shift. Using chaotropic agents such as urea or arginine can be another way to improve the correct protein folding rate [118].

The two hosts described here have been used together for the production of GPCRs: *S. cerevisiae* was first used to determine the best expression construct while *P. pastoris* was used to obtain the crystals in large amounts for structural studies [119]. The different advantages and disadvantages of using yeast expression systems are summarized in Table 3.

Table 3: Advantages and disadvantages of yeast host systems.

Yeast species	***Saccharomyces cerevisiae***	***Pichia pastoris***
Advantages	Simple growth and handling Low cost Possibility to express large proteins Easy genetic manipulations	
	High cell density Mutant library available	Very strong promoter for protein expression Very high cell density
Disadvantages	Need more time to culture than other microorganisms Some differences with human cells can still lead to improper folding or altered interaction in the membrane	

***Table 4: Commonly used plasmids in* P. pastoris.**

Plasmid Series	**Promoter**	**Antibiotic selection**	**Auxotrophic Marker**	**Tags and Fusion partners**	**Source**
pPICZa	PAOX1 (Inducible)	Zeocin (ZeoR)	None	a-factor, His Tag (6x), c-Myc Epitope Tag	Invitrogen
pPICZ	PAOX1 (Inducible)	Zeocin (ZeoR)	None	His Tag (6x), c-Myc Epitope Tag	Invitrogen
pPIC9K	PAOX1 (Inducible)	Ampicillin (AmpR), Gentamicin (GmR)	His4	None	Invitrogen
pPIC3.5K	PAOX1 (Inducible)	Ampicillin (AmpR), Gentamicin (GmR)	His4	None	Invitrogen
pGAPZ	PGAP (Constitutive)	Zeocin	None	C-His	Invitrogen

pGAPZa	PGAP (Constitutive)	Zeocin	None	a-factor, C-His	Invitrogen

***Table 5: Commonly used plasmids in* S. cerevisiae.**

Plasmid Series	Promoter	Antibiotic selection	Auxotrophic Marker	Tags and Fusion partners	Source
pYES2	PGAL1 (Inducible)	Ampicillin (AmpR)	URA3	None	Invitrogen
pYES3	PGAL1 (Inducible)	Ampicillin (AmpR)	TRP1	His Tag (6x), V5 Epitope Tag	Invitrogen
pRS426	PGAL1 (Inducible), PTEF1(Constitutive)	Not found	URA3, TRP1, HIS3, LEU2	Various	[120], [121]
pDDGFP-2	PGAL1 (Inducible), pLEU2 (Truncated)	Ampicillin (AmpR)	URA3, LEU2	GFP, TEV, His Tag (8x),	[122]
p423	PGAL1 (Inducible)	Ampicillin (AmpR)	HIS3	None	ATCC
p426	PGAL1 (Inducible)	Ampicillin (AmpR)	URA3	None	ATCC

4.2.3 Expression plasmid design

In general, there are two types of yeast vectors commonly used for recombinant expression. Usually, these vectors are shuttle vectors which means that they can be propagated in yeast and bacteria. This attribute is exploited for gene cloning and amplification. Episomal plasmids are more popular for expression in *S. cerevisiae* while integrative vectors are predominantly used in *P. pastoris.* Examples of commonly-used plasmids in *S. cerevisiae* and *P. pastoris* are listed in Table 4 and 5.

4.2.4 Promoter choice

A promoter is one of the key elements of an expression plasmid and therefore choosing a suitable promoter is fundamental to success of any expression experiment. The nature of the protein affects the choice of the promoter, e.g. for membrane proteins that can be toxic to the cell when overexpressed, inducible promoters are preferred because they allow induction of expression after the growth of the cells to a high density. Typically, P. pastoris plasmids harbour strong, inducible promoters such as the PAOX1 from the alcohol oxidase encoding gene (AOX1) which is tightly-repressed by glucose and strongly-induced by methanol as the sole carbon source [123]. However, there are numerous studies where the strong, constitutive promoter glyceraldehyde-3-phosphate dehydrogenase PGAP has produced equally good results [124]. For S. cerevisiae, the use of episomal plasmids with high copy numbers can be more beneficial because the promoters may be 10- to 100- fold weaker than those used in P. pastoris [125, 126]. A method which increases the number of plasmid copies per cell and in turn the recombinant eukaryotic membrane protein production was studied for S. cerevisiae from Parker and Newstead in 2014 [127]. The use of a truncated promoter controlling the expression of the metabolic marker LEU2 gene was shown to put

selective pressure to yeast cells growing in minus leucine medium to maintain a high copy number of the vector pDDGFP2 (80-100 copies per cell). This resulted to up to fourfold increase of the expression. One strong promoter, which is the basis of the most commonly used plasmids used in S. cerevisiae, is the PGAL1, which is induced with galactose and repressed by glucose. Commonly used promoters used in S. cerevisiae and P. pastoris are shown in Table 6.

Table 6: Commonly used yeast promoters.

	Promoter	Description	Regulation	Reference
S. cerevisiae	**Constitutive**			
	ADH1	Alcohol dehydrogenase 1		[128], [129], [130]
	GAP/TDH3	Glyceraldehyde-3-phosphate dehydrogenase		[131]
	TEF1	Translation elongation factor 1		[132]
	Inducible			
	GAL1	Galactokinase	Galactose	[133]
	GAL10	α-D-galactose-1-phosphate uridyltransferase	Galactose	[133]
	MET25	O-acetylhomoserine sulfhydrylase	Methionine	[134]
P. pastoris	**Constitutive**			
	GAP	Glyceraldehyde-3-phosphate dehydrogenase		[135]
	TEF1	Translation elongation factor 1 α		[136]
	SDH	Sorbitol dehydrogenase		[137]
	Inducible			
	AOX1	Alcohol oxidase 1	Methanol	[138]
	AOX2	Alcohol oxidase 2	Methanol	[139]
	FLD1	Formaldehyde dehydrogenase	Methanol, methylamine, choline	[140]

5 Higher eukaryotic expression systems

Higher eukaryotic hosts are frequently used for the production of eukaryotic proteins especially when post-translational modifications and native folding are crucial for the study. In addition, the

lipid environment plays an important role and these systems provide a more native lipid environment for proper protein function. The two most common host types are baculovirus-infected insect cells and various different mammalian cell hosts [141].

5.1 Insect cell expression systems

Over the years, insect cells have been used for a variety of applications including functional assays, structural analyses and antibody generation [142]. Some examples of membrane proteins successfully expressed in insect cells are GPCRs such as β2APT, A2A adenosine receptor, dopamine D3 receptor, muscarinic acetylcholine receptors (M2, M3), opioid receptors (κ-, μ-, δ-), channels including P2X4, GluA2, AQP4 (Aquaporin 4) and ABC transporters including P-glycoprotein, ABCG2 and ABCC4 [26, 143-147].

The cell-lines that are most widely used are Sf9, Sf21 and Hi5. During the last decade, 26.0% of eukaryotic membrane proteins of known structure have been produced in Sf cell lines and 4.9% in Hi5 [13]. Using electron microscopy, it was possible to elucidate the structure of the human transporter ABCA1 with a size of 2,305 residues [148]. However, most elucidated structures are from medium sized receptors (40 to 80 residues; Figure S3C). Sf9 is a cell-line derived from the ovaries of the fall armyworm, *Spodoptera frugiperda*, a lepidopteran insect. Sf21 is another cell line from *Spodoptera frugiperda* and the High five (Hi5) cell-line originates from the ovarian cells of the cabbage looper, *Trichoplusia ni* [149]. Expression of proteins within these insect cells takes advantage of the natural ability of baculoviruses to infect and replicate within insect cells. By incorporating a target gene within the viral genome, the insect cells can be used for the production of eukaryotic proteins [150]. *Drosophila* Schneider 2 (S2) cells are immortalized non-tumorigenic cells isolated from primary cultures of *Drosophila melanogaster* embryos. They can be cultured to high densities and can be transiently- or stably-transfected using a non-lytic plasmid-based system that provides additional benefits over virus-infected insect systems. S2 cells possess all the mammalian-like cell machineries for gene expression, protein processing and trafficking, including post-translational modifications that may be critical for proper maturation, localization and function of target protein [151]. In addition, the ExpiSf expression system was developed in order to culture insect cells in yeastolate-free medium and provides high protein yields as well as consistent cell growth and protein expression [152].

5.1.1 Recombinant protein production systems

The most commonly-used baculovirus is AcMNPV (*Autographa californica* multicapsid nucleopolyhedrovirus). Recombinant baculoviruses can be produced using systems such as Bac-to-Bac or FlashBAC. The Bac-to-Bac system works as follows: the gene of interest is cloned into a transfer vector, pFastBac, under the control of the AcMNPV polyhedron promoter [143]. Polyhedrin is a protein involved in forming the coat of the baculovirus, and thus has a strong promoter, however polyhedrin is not needed for infection and replication within cultured cells [150]. The pFastBac vector containing the gene of interest is then transformed into specialised DH10Bac *E. coli* cells. They contain a baculovirus shuttle vector (bacmid) encoding the baculovirus genome with a transposon, and a helper plasmid, which aids with the site-specific transposition of the gene of

interest in the bacmid. This recombinant bacmid is then isolated, purified and transfected into insect cells to produce the recombinant baculovirus which get released into the medium (Figure 7).

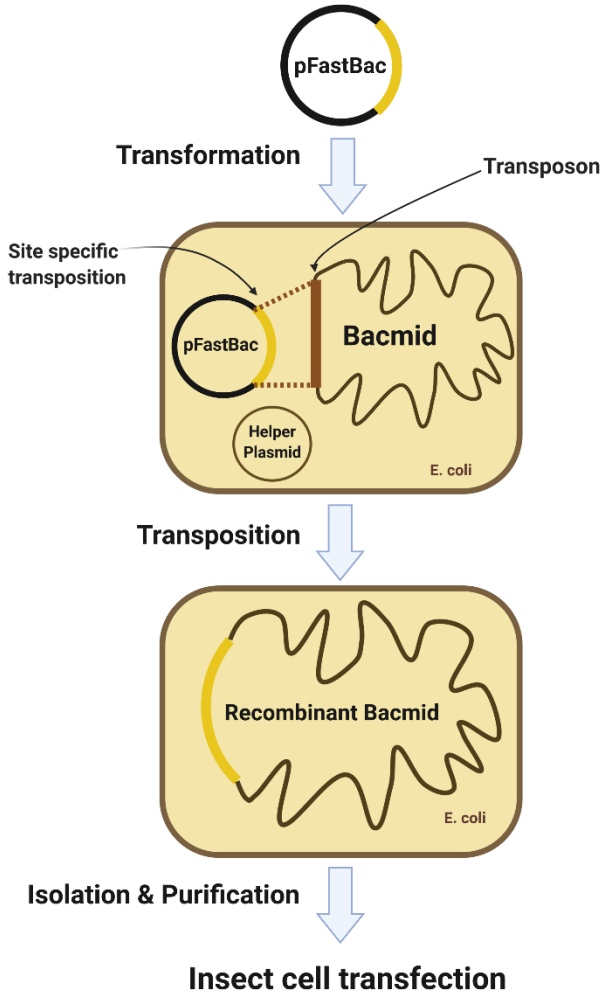

Figure 7: Bac-to-Bac expression system. *The gene of interest is cloned into the pFastBac vector. After transformation of pFastBac into DH10Bac* E.coli *cells that contain the bacmid with a transposon and a helper plasmid, site-specific transposition occurs between the bacmid and the gene of interest in the vector. The recombinant Bacmid is then isolated and purified and used for insect cell transfection. Within the insect cells, the bacmid genome leads to the production of recombinant baculovirus, which lyses the cells and is released in the media.*

The FlashBAC system, on the other hand, is based on a baculovirus genome that lacks part of an essential gene, ORF1629, and contains a bacterial artificial chromosome (BAC) at the polyhedrin locus, replacing the polyhedrin coding region. The virus cannot replicate in insect cells because of the deletion of the essential gene. It would not usually be possible to replicate the baculoviral DNA in bacteria but the insertion of BAC into the AcMNPV gene locus allows viral DNA to be maintained and propagated as a circular genome in bacteria, thus producing the FlashBAC DNA which can be isolated and purified. The gene of interest is cloned into a transfer vector such as pOET, which also contains the essential ORF1629 gene. The transfer vector and the purified FlashBAC bacmid DNA are co-transfected into insect cells. After recombination, within the insect cells, the function of the essential gene is restored, leading to the replication of the viral DNA and the insertion of the gene of interest, removing the BAC sequence. The recombinant virus can be harvested from the culture medium (Figure 8) [153].

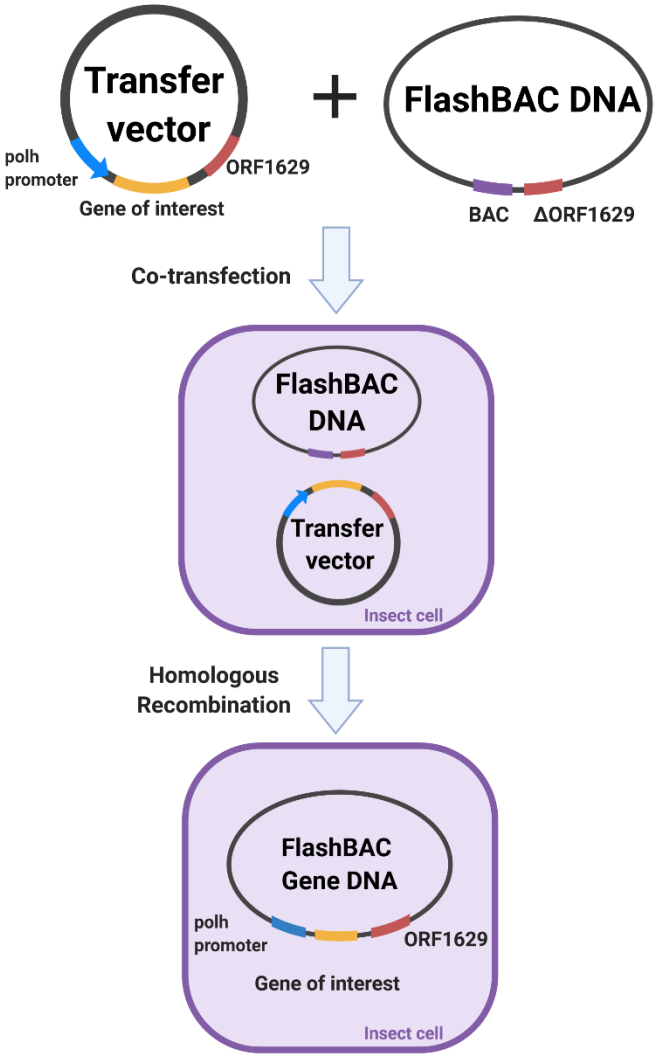

Figure 8: FlashBAC expression system. *To produce the recombinant baculovirus DNA using the FlashBAC system, a transfer vector containing the gene of interest, the polh promoter and ORF1629 gene is co-transfected into insect cells together with the viral FlashBAC DNA. The replication of the recombinant baculovirus DNA is only initiated after homologous recombination of the transfer vector and FlashBAC DNA. After recombination, new viruses are produced and can be harvested so that they can be used for further insect cell infection and protein production. The original FlashBAC DNA contains a truncated version of the ORF1629 gene (ΔORF1629) which is not functional. As ORF1629 is needed for the replication in insect cells, it can only start after the recombination has occurred.*

A different technology for multiprotein complex expression, called MultiBac system, has also been developed [154], which can be used for the expression of protein complexes or proteins with a number of different sub-units. This system uses a number of small donor and acceptor vectors into which the target genes are inserted, either by conventional cloning methods or sequence-independent and ligation-independent cloning methods. Cre-mediated fusion of the donor and acceptor vectors generates a single multigene construct in a single-step reaction. The single multigene construct is inserted into the baculoviral genome by recombination with bacmid DNA in specialised DH10MultiBac *E. coli* cells, which also contain a helper plasmid, similarly to the Bac-to-Bac system. The MultiBac baculoviral genome which contains all desired heterologous genes is then purified from small bacterial cultures and used to transfect insect cells. The resulting baculovirions

are harvested and applied to larger insect cell cultures for heterologous protein production [155]. Notably the MultiBac baculoviral genome was also engineered to remove genes encoding viral protease and apoptotic activities, with the aim of reducing proteolytic breakdown of the target proteins and delaying lysis of the infected cells

Finally, one of the latest technologies that was developed for gene transfection is the SmartBac system [156]. This builds upon the MultiBac approach but simplifies some of the initial cloning, making long DNA sequences encoding polyproteins which are produced by overlapping PCR and Gibson assembly. The final transfer plasmid produced by Cre-LoxP recombination is designed to ensure a size below 25 kb, in order to facilitate efficient transformation. To increase the stability of the large final transfer plasmid, which is propagated in *E. coli*, a low-copy p15A replication origin into the acceptor vectors is introduced and the bacteria are cultured at 30 °C instead of the usual optimal 37 °C.

Table 7: Commonly used insect cell promoters.

Promoter	Description	Regulation	Reference
Polyhedrin	Strong promoter from the polyhedrin coat protein of baculovirus	Constitutive	[157]
p10	Strong promoter from the p10 gene of baculovirus	Constitutive	[158]
p6.9	Strong promoter from the basic protein of Baculovirus. Drives expression at an earlier point of infection than polyhedrin or p10	Constitutive	[159]

Table 8: Commonly used insect cell plasmids.

Plasmid Series	Promoter	Tag	Selection Marker	Source
pFastBac	Polyhedrin	None	Ampicillin (AmpR), Gentamicin	Gibco™
pFastBac-Dual	Polyhedrin, p10	None	Ampicillin (AmpR), Gentamicin	Gibco™
pOET	Polyhedrin, p10	None	Ampicillin (AmpR)	OxfordExpressionTechnologies
pVL1392-3	Polyhedrin	None	Ampicillin (AmpR)	Invitrogen
pBAC-1	Polyhedrin	6xHis	Ampicillin (AmpR)	Novagen

5.1.2 Insect cell culture optimisation

Insect cell cultures require a temperature of 27-28 ^{o}C to grow. There is no need for addition of CO_2 to the incubator as their medium is buffered with phosphate rather than carbonate. Both serum-free and serum-containing media can be used, but insect cells tend to adapt to a specific medium and there may be loss of culture if they are passed abruptly from one medium to another.

Insect cells can be cultured in suspension or be adherent. However, they are loosely adherent so there is no need for EDTA or trypsin to detach them [150].

By using different ratios of virus to cells and different times of infection, insect cells can be used for either viral amplification or protein expression. Typically viral amplification uses a low MOI (multiplicity of infection – number of virus particles per cell) such as 0.1-0.25, for 5-7 days. Protein expression uses a higher MOI (1-5) for shorter time periods such as 24-72 hours. To optimise protein expression, several different factors can be altered. These include the cell density, the ratio of cells to virus and the time length of infection; the precise conditions vary for different proteins. Figure 9 shows an example of the effect of time on expression of the ABC transporter MRP4/ABCC4 (multidrug resistance protein 4) in Sf9 cells. Cells harvested after 72 hours of infection show extensive degradation of the target protein, whereas after 48 hours the protein remains intact. Other changes can be made to improve the quantity and the quality (full-length protein) of the protein such as adding protease inhibitors or altering the growth temperature.

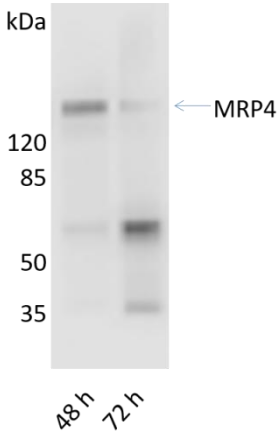

Figure 9: Effect of infection time on the expression of human MRP4/ABCC4 in Sf9 cells. *Sf9 cells at 1x10^6 cells/ml were infected with recombinant MRP4 baculovirus at an MOI of 4 and harvested after either 48h or 72h. Membrane preparations from these cells were tested for MRP4 expression by Western blot using an anti-MRP4 primary antibody, an anti-rat HRP secondary antibody and detection by chemiluminescence.*

5.1.3 Insect cell culture features

Their ease of culture, the fact that they can express large proteins and the lack of human pathogen contaminants make insect cells simple to manipulate and decreases the cost and biosafety required. Also, the lack of the polyhedrin coat protein means the virus can only infect cultured cells, therefore posing no risk to insects in the environment. Another feature making them an important system for protein production is that they produce properly folded proteins and, since they are a eukaryotic system, provide post-translational modifications [149]. However, these modifications may differ from those of mammalian cells as insect cells may lack some mechanisms of post-translational modifications. One example can be the different glycosylation patterns in insect cells. Glycosylated proteins in yeast, insect and mammalian cells have a common sugar core containing two N-acetylglucosamine and three mannose residues. Membrane proteins in mammalian cells are further modified with terminal fucose, sialic acid or galactose residues by the enzymes sialyltransferase and galactosyltransferase. Glycosylation patterns in insect cells depend on the expressed membrane proteins and on the host cell-line. Typical patterns are the high mannose-type or the insect-specific

paucimannose type [160]. Nevertheless, there have been advances towards the humanization of insect cells [161, 162]. Other advantages of this system are that it is available for multiple gene expression since the viral genome has a great degree of capacity and flexibility and that baculovirus does not infect humans which makes its manipulation safe [149].

However, insect cells have some disadvantages such as the high cost for consumables and equipment and the lower yield or protein in comparison to microbial expression systems [56]. Additionally, the generation of the recombinant virus is time consuming and cells can only be infected transiently because the virus causes cell lysis. Furthermore, the lipid composition and the levels of cholesterol in the insect cell membrane differ from the mammalian one since the insect plasma membrane contains higher amounts of unsaturated lipids and lower amounts of cholesterol to ensure membrane fluidity at the organism growth temperature [26]. This may adversely affect the expression of functional mammalian membrane proteins.

5.2 Mammalian cell expression systems

Mammalian cells are used for protein production for several applications such as functional assays, protein interaction studies and antibody generation [142]. Some examples of membrane protein expression in mammalian cells are GPCRs (including adrenergic receptors, serotonin 5HT receptors and the μ-opioid receptor) and ABC transporters (such as P-gp) [143, 146, 163, 164].

One of the most commonly used cell-lines is from human embryonic kidney cells; HEK 293 cells have been used for the production of 25.3% of eukaryotic membrane proteins of known structure in the last 10 years, principally thanks to their ease of transfection, the high protein yields [165] and the increased use of electron microscopy for structure elucidation. The size of the elucidated proteins ranged from 300 to 1900 residues (Figure S3D) with a maximum size 2,822 residues[166]. In contrast to insect cells, most of the proteins formed higher oligomeric complexes like tetramers up to hexamers. Other commonly-used cell types are baby hamster kidney cells (BHK-21), monkey kidney fibroblast cells (COS-7), Chinese hamster ovary cells (CHO), the latter being especially used when it comes to therapeutic protein production [167]. Finally, human epithelioid carcinoma cells (HeLa) is a very famous and widely used cell-line extracted from a patient called Henrietta Lacks who died from cancer in the 1950s [141, 160, 163]. Mammalian cells typically have an optimal growth temperature of about 37 ^{o}C and require 5% CO_2. They usually grow in adherent cultures, but HEK 293 and CHO have been adapted to grow in suspension, which greatly facilitates their scale-up [141]. Mammalian cells provide endogenous post-translational modifications needed for expression of mammalian proteins, alongside authentic translocation and trafficking machineries. Expression of GPCRs such as the angiotensin II type 1 receptor, the avian β1 adrenergic receptor and the serotonin transporter in mammalian cell-lines, provides evidence that mammalian cells can be superior for the production of functional and properly-folded proteins compared to insect cells [168]. Proteins expressed in mammalian cells are usually properly folded and functional. Mammalian cells also provide a close-to native lipid environment for mammalian proteins. There are differences between the various cell-lines so one should take this into consideration before starting the protein expression. [144, 160, 169].

Moreover, cell engineering of mammalian cells provided solutions for structural studies and biopharmaceutical manufacturing of proteins. Some examples are the N-

acetylglucosaminyltransferase-I (GnTI) negative HEK293, the 293SGlycoDelete and the ExpiCHO cell lines [165, 167, 170, 171]. The GnTI- HEK293 cell line was designed in order to promote homogenous glycosylation which is essential for the stabilisation of folded proteins and helpful for crystallography attempts [170]. The 293SGlycoDelete cells were generated in order to solve the problem of N-glycosylation heterogeneity and maintain a balance between the N-glycan heterogeneity and their folding-enhancing functions, which are important for the biopharmaceutical industry [171]. Finally, the ExpiCHO cell line was selected for its high transfection efficiency, thus providing a high yield of protein production and better quality for antibody production [165, 167].

Another big advantage of mammalian cells is that they can be transfected either transiently or stably with a plasmid containing the target gene, optionally with inducible promoters. Transiently transfected cells often show very high levels of overexpression in a single cell, but the transfection efficiency (% of total cells that are transfected) can be low. Transiently transfected cells can also only express the protein for a limited period of time and the gene is not integrated into the genome. The genetic material can be lost by environmental factors and cell division. Transient cell-lines are efficient, cost-effective and quick. They can be used to assess the effect of mutations and truncations prior to scale-up of protein production. However, when scale-up is required, there may be some variability in the amount of the protein that is expressed, and the requirement for large amounts of plasmid DNA and transfection reagent can become limiting. Some factors that could affect the level of expression are the plasmid size, the amount of the plasmid used, the strength of the promoter, the cell type, the efficiency of transfection and the toxicity of the transfection reagents [163]. Transfection methods can be biological (transduction, which is discussed below), chemical or physical. Chemical methods commonly use cationic polymers, calcium phosphate, cationic lipids or cationic amino acids. Physical methods use a variety of tools to deliver the gene, the most widely-used being electroporation, during which a short electrical pulse disturbs cell membranes and makes holes in the membrane through which nucleic acids can pass. Other tools are micro-injection, biolistic and laser-mediated transfection which may be more sensitive but usually require skill or expensive equipment. In all cases, the principle is that negatively-charged DNA makes complexes with positively-charged chemicals which pass through the cell membrane [172]. The process is very simple: generation of a plasmid, transfection in log phase and harvest from 48h to 14 days depending on the particular protein, cell line and conditions used [141].

For stable transfection, plasmids contain a marker gene for selection to retain the plasmid. Yields of protein per cell may be lower than with transiently-transfected cells, but the consistent level of target protein production is helpful, especially when scaling up, and negates the need to constantly produce purified plasmid DNA. Stable cell-lines are generated by the integration of the recombinant DNA in the host cells [172]. This can either be random or targeted with specific sequences being recognised by a recombinase. The selection is made using an antibiotic resistant marker so that only the cells that receive the transgene will survive when cultured in medium containing the antibiotic. However, this strategy can be unreliable, especially for high-level transgene expression [163]. In this case, GFP can be a useful tool. Cells that stably integrate a GFP expression vector can be identified by intracellular fluorescence and isolated by fluorescence-activated cell sorting (FACS). Cell sorting can isolate a small number of clones from millions of cells that produce the target protein in high yields. By repeating the process, cells that express GFP stably over time can be isolated. Once the recombinant DNA is established, the expression will be consistent [173].

Different expression vectors have been designed to transfer foreign genes into mammalian cells. The choice of the vector depends on the application, the host cell, the time limitation, the yield of desired product, and safety. Most commonly-used, strong constitutive promoters (Table 8) are from viral genomes including the human cytomegalovirus (CMV) immediate early promoter and the simian virus 40 (SV40) early promoter [174]. CMV is a herpes virus that can infect various human cell types. During productive infection, CMV genes are expressed from immediate early (IE) genes to early genes and then to late genes in a coordinated order. The CMV IE promoter and enhancer are widely used to drive gene expression in a variety of cell types [175]. SV40 infects a wide range of cell types from humans to other mammals, and expresses its genes in them. Plasmids incorporating SV40 genes and/or promoter may express either transiently or stably in cell-lines [176].

Table 9: Commonly used mammalian promoters.

Promoter	Description	Regulation	Reference
Constitutive expression			
CMV	Mammalian expression promoter from the human cytomegalovirus	Constitutive	[174]
EF1a	Mammalian expression promoter from human elongation factor 1 alpha	Constitutive	[177]
SV40	Mammalian expression promoter from the simian vacuolating virus 40	Constitutive	[174]
Inducible expression			
CMV/TO	Hybrid promoter consisting of the human cytomegalovirus immediate-early (CMV) promoter and tetracycline operator 2 (TetO2) sites for high-level tetracycline-regulated expression in a wide range of mammalian cells	Inducible	[163]

Table 10: Commonly used mammalian plasmids.

Plasmid Series	Promoter	Description	Tag	Antibiotic Resistant	Source
pcDNA4/TO	CMV/TO	Hybrid promoter consisting of CMV and tetracycline operator 2 (TetO2).	None	Zeocin (ZeoR)	Invitrogen
pcDNA3.1	CMV	CMV promoter for high-level expression in a wide range of mammalian cells	None	Geneticin (G-418), Ampicillin (AmpR)	Invitrogen
pEF	EF1a	Overproduction of recombinant proteins in mammalian cell lines	6xHis, c-Myc Epitope	Blasticidin (BsdR)	Invitrogen
pACMVtetO	CMV/TO	Hybrid promoter consisting of CMV and tetracycline operator 2 (TetO2).	None	Ampicillin (AmpR)	[178]
pCMV	CMV	Protein expression in mammalian systems driven by CMV	None	Geneticin (G-418), kanamycin (KanR)	Agilent
pEG BacMam	CMV	Baculovirus-mediated gene transfer into mammalian cells (BacMam)	None	Gentamicin (GenR), Ampicillin (AmpR)	[143]

Inducible cell-lines can also be convenient. These cells do not normally produce protein except upon induction, such as via the tetracycline system. The expression vector includes Tet operator sites (TetO) downstream of the promoter. A repressor protein (TetR) binds to TetO, preventing transcription. Addition of tetracycline inactivates TetR allowing transcription from the cytomegalovirus (CMV) promoter. This system is applicable to both transient and stable cell-line production [163]. This can be particularly useful for expression of several membrane proteins, especially channels, for which overexpression can be toxic to the cell. Being able to first grow the cells to a reasonable density before inducing expression of the protein is helpful in that case.

Viral infection methods can also be used. One example is the BacMam system which utilises recombinant baculovirus as with insect cell expression, but with a mammalian promoter incorporated. The baculovirus can transduce mammalian cells to express the target protein, however the baculovirus cannot replicate within mammalian cells which makes its use safe. This system is well-tolerated by cells. Its reproducibility, rapid manipulation and efficiency make it an appealing approach that combines the ease of baculoviral DNA delivery with the mammalian cell machinery [179]. Lentivirus offers an alternative approach, showing highly-efficient infection of mammalian cells to generate stable cell-lines [180], however the inherent safety risks of using a virus that can infect humans must be considered.

The high cost, the demanding culture conditions, the susceptibility to contamination and low protein yields are some factors that may discourage the use of mammalian cells, especially for structural studies [142, 144, 160]. Nonetheless, they remain a very useful alternative to simpler expression systems when it comes to the expression of proteins that require complex post-translational modifications that cannot be performed in other expression systems or for large proteins.

6. Conclusion

The production of eukaryotic membrane proteins remains a challenging task, but a diverse and increasing number of expression systems is available to address that challenge. In this review, we describe the advantages and disadvantages of bacterial, yeast, insect cell and mammalian cell expression systems (Table 11). We discuss the options for construct design and gene modification. There is no expression system that can be used for the production of all eukaryotic membrane proteins. Even though bacterial cells do not provide the perfect cellular environment for eukaryotic membrane proteins, they have been widely-used and have the advantages of cost efficiency, easy handling and modification. They are limited in their expression of large proteins and suffer from a lack of eukaryotic post-translational modifications. In comparison, yeast cells are able to perform simple eukaryotic post-translational modifications but can become quickly overwhelmed by mass production, resulting in non-functional proteins. Eukaryotic cells should be chosen if a correctly-folded and modified eukaryotic membrane protein from higher organisms is desired. However, handling insect and mammalian cells in particular is more challenging than dealing with microbes. Additionally, the financial and temporal burdens are greater than for bacterial and yeast cells. The amount of consumables which are not reusable for insect and mammalian expression systems is much higher than for microbial systems and specialised equipment is needed. Therefore, these systems are not only more expensive but also show a bad ecological sustainability. Overall, it may be

necessary to explore several systems in order to find the ideal host for each membrane protein of interest.

***Table 11: Properties of the expression systems discussed in this article**. Yields and sizes are based on the findings from eukaryotic membrane protein structural elucidations from the last decade. [1]The expression time considers the steps from equilibrating the cells to their harvest after protein expression. [2] The preparation time considers all the steps needed to be done after plasmid generation and before the protein production can be started. [3] Ease of manipulation considers the range of optimisation possibilities.*

Properties	*E. coli*	Yeast	Insect cells	Mammalian cells
Expression time[1]	2 to 3 days	4 to 6 days	2 weeks minimum	A few days to 2 weeks
Preparation time[2]	1 day	3 to 5 days	2 to 4 weeks	Several days to weeks
Ease of manipulation[3]	Very easy	Easy	Difficult	Hard
Relative yield	High	High	Medium	Low
Medium cost [£/L]	<10	<10	8 to 95	5 to 105
Amount of residues	40 - 750 Up to 1850	250 – 900 Up to 1500	350 – 600 Up to 2300	300 – 1900 Up to 2900
Eukaryotic PTMs	None	Simple	Better	Best
Protein folding [56]	Refolding usually required	Refolding may be required	Proper folding	Proper folding
Cost of storage	Low (glycerol; -80 °C)	Low (glycerol; -80 °C)	High (liquid nitrogen)	High (liquid nitrogen)
Ecological sustainability (use of plastic ware, consumables)	Good	Good	Medium	Poor
Advantages	Low cost, fast growth	Low cost, eukaryotic PTMs	Most eukaryotic PTMs, simpler maintenance than mammalian cells	All PTMs, native mammalian cellular environment
Disadvantages	Lack of eukaryotic PTMs	Hyper-glycosylation	High cost, slow growth, limited glycosylation	High cost, slow growth

7. Acknowledgements

We are grateful for funding from the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie grant agreement No 847419 (MemTrain).

8. Bibliography

1. Santos, R., et al., *A comprehensive map of molecular drug targets.* Nat Rev Drug Discov, 2017. **16**(1): p. 19-34.
2. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nature reviews Drug discovery, 2006. **5**(12): p. 993-996.
3. Hutchings, C.J., P. Colussi, and T.G. Clark. *Ion channels as therapeutic antibody targets*. in *MAbs*. 2019. Taylor & Francis.
4. Lodish, H., et al., *Molecular cell biology*. 2008: Macmillan.
5. Ow, D.S.-W., et al., *Co-expression of Skp and FkpA chaperones improves cell viability and alters the global expression of stress response genes during scFvD1.3 production.* Microbial Cell Factories, 2010. **9**(1): p. 22.
6. Miroux, B. and J.E. Walker, *Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels.* Journal of Molecular Biology, 1996. **260**(3): p. 289-298.
7. Arechaga, I., et al., *Characterisation of new intracellular membranes in Escherichia coli accompanying large scale over-production of the b subunit of F1Fo ATP synthase.* FEBS letters, 2000. **482**(3): p. 215-219.
8. Moraes, I., et al., *Membrane protein structure determination — The next generation.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2014. **1838**(1, Part A): p. 78-87.
9. Zou, Y., W.I. Weis, and B.K. Kobilka, *N-terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor.* PloS one, 2012. **7**(10).
10. Boura, E., et al., *Metal ions-binding T4 lysozyme as an intramolecular protein purification tag compatible with X-ray crystallography.* Protein Science, 2017. **26**(6): p. 1116-1123.
11. Yang, Z. and Z. Zhang, *Engineering strategies for enhanced production of protein and bio-products in Pichia pastoris: A review.* Biotechnology Advances, 2018. **36**(1): p. 182-195.
12. Lee, A., *Lipid–protein interactions in biological membranes: a structural perspective.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2003. **1612**(1): p. 1-40.
13. White, S. *Membrane Proteins of Known 3D Structure*. [cited 2020 01/04]; Available from: https://blanco.biomol.uci.edu/mpstruc/.
14. Shuman, S., *Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase.* J Biol Chem, 1994. **269**(51): p. 32678-84.
15. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
16. Ashwini, M., et al., *[Advances in Molecular Cloning].* Mol Biol (Mosk), 2016. **50**(1): p. 3-9.
17. Nørholm, M.H., et al., *Manipulating the genetic code for membrane protein production: what have we learnt so far?* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2012. **1818**(4): p. 1091-1096.
18. Reuveni, S., et al., *Genome-Scale Analysis of Translation Elongation with a Ribosome Flow Model.* PLOS Computational Biology, 2011. **7**(9): p. e1002127.
19. Kozak, M., *Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo.* 1984. **308**(5956): p. 241-246.
20. Öberg, F., et al., *Insight into factors directing high production of eukaryotic membrane proteins; production of 13 human AQPs in Pichia pastoris.* 2009. **26**(4): p. 215-227.
21. Nørholm, M.H., et al., *Improved production of membrane proteins in Escherichia coli by selective codon substitutions.* FEBS letters, 2013. **587**(15): p. 2352-2358.

22. Kim, H.S., et al., *Translation levels control multi-spanning membrane protein expression.* PloS one, 2012. **7**(4).
23. Schlegel, S., et al., *Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21 (DE3).* Journal of molecular biology, 2012. **423**(4): p. 648-659.
24. Nakamura, Y. *Codon Usage Database*. [cited 2020 14/05]; Available from: http://www.kazusa.or.jp/codon.
25. Kober, L., C. Zehe, and J. Bode, *Optimized signal peptides for the development of high expressing CHO cell lines.* Biotechnol Bioeng, 2012. **110**(4): p. 1164-73.
26. Massotte, D., *G protein-coupled receptor overexpression with the baculovirus–insect cell system: a tool for structural and functional studies.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2003. **1610**(1): p. 77-89.
27. Attallah, C., et al., *A highly efficient modified human serum albumin signal peptide to secrete proteins in cells derived from different mammalian species.* Protein Expression and Purification, 2017. **132**: p. 27-33.
28. Warne, T., et al., *Structure of a β 1-adrenergic G-protein-coupled receptor.* Nature, 2008. **454**(7203): p. 486-491.
29. Doré, A.S., et al., *Structure of the adenosine A2A receptor in complex with ZM241385 and the xanthines XAC and caffeine.* Structure, 2011. **19**(9): p. 1283-1293.
30. Lebon, G., et al., *Agonist-bound adenosine A 2A receptor structures reveal common features of GPCR activation.* Nature, 2011. **474**(7352): p. 521-525.
31. Tan, N.Y., et al., *Sequence-based protein stabilization in the absence of glycosylation.* Nature Communications, 2014. **5**(1): p. 3099.
32. Chun, E., et al., *Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors.* Structure, 2012. **20**(6): p. 967-976.
33. Mallipeddi, S., N. Zvonok, and A. Makriyannis, *Expression, purification and characterization of the human cannabinoid 1 receptor.* Scientific reports, 2018. **8**(1): p. 1-11.
34. Stevens, R.C., *Design of high-throughput methods of protein production for structural biology.* Structure, 2000. **8**(9): p. R177-R185.
35. Liang, Y.-L., et al., *Cryo-EM structure of the active, Gs-protein complexed, human CGRP receptor.* Nature, 2018. **561**(7724): p. 492-497.
36. Hattori, M., Ryan, and E. Gouaux, *A Fluorescence-Detection Size-Exclusion Chromatography-Based Thermostability Assay for Membrane Protein Precrystallization Screening.* 2012. **20**(8): p. 1293-1299.
37. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Applied Microbiology and Biotechnology, 2003. **60**(5): p. 523-533.
38. Pande, J., M.M. Szewczyk, and A.K. Grover, *Phage display: Concept, innovations, applications and future.* Biotechnology Advances, 2010. **28**(6): p. 849-858.
39. Bowers, P.M., et al., *Mammalian cell display for the discovery and optimization of antibody therapeutics.* Methods, 2014. **65**(1): p. 44-56.
40. Ceglarek, I., et al., *A novel approach for separating bacteriophages from other bacteriophages using affinity chromatography and phage display.* Scientific Reports, 2013. **3**(1): p. 3220.
41. Pan, Y., et al., *Glutathione (GSH)-decorated magnetic nanoparticles for binding glutathione-S-transferase (GST) fusion protein and manipulating live cells.* Chemical Science, 2011. **2**(5): p. 945-948.
42. Li, Y.-J., et al., *Reversible immobilization of proteins with streptavidin affinity tags on a surface plasmon resonance biosensor chip.* Analytical and Bioanalytical Chemistry, 2006. **386**(5): p. 1321-1326.
43. Khairil Anuar, I.N.A., et al., *Spy&Go purification of SpyTag-proteins using pseudo-SpyCatcher to access an oligomerization toolbox.* Nat Commun, 2019. **10**(1): p. 1734.

44. Fierle, J.K., et al., *Integrating SpyCatcher/SpyTag covalent fusion technology into phage display workflows for rapid antibody discovery.* Sci Rep, 2019. **9**(1): p. 12815.
45. Jenny, R.J., K.G. Mann, and R.L. Lundblad, *A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa.* Protein Expr Purif, 2003. **31**(1): p. 1-11.
46. Hwang, P.M., J.S. Pan, and B.D. Sykes, *Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in Escherichia coli.* FEBS Lett, 2014. **588**(2): p. 247-52.
47. Bill, R.M. and T. von der Haar, *Hijacked then lost in translation: the plight of the recombinant host cell in membrane protein structural biology projects.* Curr Opin Struct Biol, 2015. **32**: p. 147-55.
48. Dilworth, M.V., et al., *Microbial expression systems for membrane proteins.* Methods, 2018. **147**: p. 3-39.
49. Wiseman, D.N., et al., *Expression and purification of recombinant G protein-coupled receptors: A review.* Protein Expr Purif, 2020. **167**: p. 105524.
50. Chadwick, A.C., et al., *NMR Structure of the C-Terminal Transmembrane Domain of the HDL Receptor, SR-BI, and a Functionally Relevant Leucine Zipper Motif.* Structure, 2017. **25**(3): p. 446-457.
51. Nadezhdin, K.D., et al., *Structural basis of p75 transmembrane domain dimerization.* Journal of Biological Chemistry, 2016. **291**(23): p. 12346-12357.
52. Bocharov, E.V., et al., *Left-Handed Dimer of EphA2 Transmembrane Domain: Helix Packing Diversity among Receptor Tyrosine Kinases.* Biophysical Journal, 2010. **98**(5): p. 881-889.
53. Kovtun, O., et al., *Structure of the membrane-assembled retromer coat determined by cryo-electron tomography.* Nature, 2018.
54. Yan, Z., et al., *Structure of the Na_v1.4-β1 Complex from Electric Eel.* Cell, 2017. **170**(3): p. 470-482.e11.
55. Ruan, J., et al., *Cryo-EM structure of the gasdermin A3 membrane pore.* Nature, 2018. **557**(7703): p. 62-67.
56. *Comparison of Expression Systems*. [cited 2020 17/05]; Available from: https://www.genwaybio.com/technologies/protein-expression.
57. Grisshammer, R. and C.G. Tateu, *Overexpression of integral membrane proteins for structural studies.* Quarterly Reviews of Biophysics, 1995. **28**(3): p. 315-422.
58. Sorensen, H.P. and K.K. Mortensen, *Advanced genetic strategies for recombinant protein expression in Escherichia coli.* J Biotechnol, 2005. **115**(2): p. 113-28.
59. Srinivasan, V., A.J. Pierik, and R. Lill, *Crystal structures of nucleotide-free and glutathione-bound mitochondrial ABC transporter Atm1.* Science, 2014. **343**(6175): p. 1137-1140.
60. Weiß, H.M. and R. Grisshammer, *Purification and characterization of the human adenosine A2a receptor functionally expressed in Escherichia coli.* European Journal of Biochemistry, 2002. **269**(1): p. 82-92.
61. Egloff, P., et al., *Structure of signaling-competent neurotensin receptor 1 obtained by directed evolution in Escherichia coli.* Proceedings of the National Academy of Sciences, 2014. **111**(6): p. E655-E662.
62. Bayrhuber, M., et al., *Structure of the human voltage-dependent anion channel.* Proceedings of the National Academy of Sciences, 2008. **105**(40): p. 15370-15375.
63. Hiller, S., et al., *Solution Structure of the Integral Human Membrane Protein VDAC-1 in Detergent Micelles.* Science, 2008. **321**(5893): p. 1206-1210.
64. Park, S.H., et al., *Structure of the chemokine receptor CXCR1 in phospholipid bilayers.* Nature, 2012. **491**(7426): p. 779-783.
65. Legros, C., et al., *Generating a high affinity scorpion toxin receptor in KcsA-Kv1.3 chimeric potassium channels.* The Journal of biological chemistry, 2000. **275**(22): p. 16918-16924.
66. Takacs, Z., et al., *A designer ligand specific for Kv1.3 channels from a scorpion neurotoxin-based library.* Proceedings of the National Academy of Sciences, 2009. **106**(52): p. 22211-22216.

67. Snijder, H.J.A. and J. Hakulinen, *Membrane protein production in E. coli for applications in drug discovery*, in *Advanced Technologies for Protein Complex Production and Characterization*. 2016, Springer. p. 59-77.
68. Müller-Hill, B., *The lac Operon: A Short History of a Genetic Paradigm*. 1996: Berlin: Walter de Gruyter.
69. Studier, F.W., et al., *Use of T7 RNA polymerase to direct expression of cloned genes.* Methods Enzymol, 1990. **185**: p. 60-89.
70. Baneyx, F., *Recombinant protein expression in Escherichia coli.* Curr Opin Biotechnol, 1999. **10**(5): p. 411-21.
71. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* J Bacteriol, 1995. **177**(14): p. 4121-30.
72. Gossen, M. and H. Bujard, *Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.* Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5547-51.
73. Gentz, R. and H. Bujard, *Promoters recognized by Escherichia coli RNA polymerase selected by function: highly efficient promoters from bacteriophage T5.* J Bacteriol, 1985. **164**(1): p. 70-7.
74. Bass, S.H. and D.G. Yansura, *Application of the E. coli trp promoter.* Mol Biotechnol, 2000. **16**(3): p. 253-60.
75. de Boer, H.A., L.J. Comstock, and M. Vasser, *The tac promoter: a functional hybrid derived from the trp and lac promoters.* Proc Natl Acad Sci U S A, 1983. **80**(1): p. 21-5.
76. Baumgarten, T., et al., *Optimizing Recombinant Protein Production in theEscherichia coliPeriplasm Alleviates Stress.* Applied and Environmental Microbiology, 2018. **84**(12): p. e00270-18.
77. Angius, F., et al., *A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins.* Scientific reports, 2018. **8**(1): p. 1-11.
78. Kwon, S.-K., et al., *Comparative genomics and experimental evolution of Escherichia coli BL21 (DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction.* Scientific reports, 2015. **5**: p. 16076.
79. Stano, N.M. and S.S. Patel, *T7 Lysozyme Represses T7 RNA Polymerase Transcription by Destabilizing the Open Complex during Initiation.* 2004. **279**(16): p. 16136-16143.
80. Schleif, R., *AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action.* FEMS Microbiol Rev, 2010. **34**(5): p. 779-96.
81. Mulligan, C., et al., *The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter.* Proceedings of the National Academy of Sciences, 2009. **106**: p. 1778-1783.
82. Clark, M.A., et al., *Regulation and expression of human Fabs under the control of the Escherichia coli arabinose promoter, PBAD.* Immunotechnology, 1997. **3**(3): p. 217-226.
83. Siegele, D.A. and J.C. Hu, *Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations.* Proceedings of the National Academy of Sciences, 1997. **94**: p. 8168-8172.
84. Berens, C. and W. Hillen, *Gene regulation by tetracyclines.* European journal of biochemistry, 2003. **270**(15): p. 3109-3121.
85. Skerra, A., *Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli.* Gene, 1994. **151**(1): p. 131-135.
86. Gustafsson, C., S. Govindarajan, and J. Minshull, *Codon bias and heterologous protein expression.* Trends in biotechnology, 2004. **22**(7): p. 346-353.
87. Rosano, G.N.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Frontiers in Microbiology, 2014. **5**.

88. Kaur, J., A. Kumar, and J. Kaur, *Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements.* International Journal of Biological Macromolecules, 2018. **106**: p. 803-822.
89. Jidenko, M., et al., *Crystallization of a mammalian membrane protein overexpressed in Saccharomyces cerevisiae.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(33): p. 11687-11691.
90. Long, S.B., E.B. Campbell, and R. MacKinnon, *Crystal structure of a mammalian voltage-dependent Shaker family K+ channel.* Science, 2005. **309**(5736): p. 897-903.
91. Green, S.R. and C.M. Moehle, *Media and culture of yeast.* Current protocols in cell biology, 1999. **4**(1): p. 1.6. 1-1.6. 12.
92. Demain, A.L. and P. Vaishnav, *Production of recombinant proteins by microbes and higher organisms.* Biotechnology advances, 2009. **27**(3): p. 297-306.
93. Jigami, Y., *Yeast glycobiology and its application.* Bioscience, biotechnology, and biochemistry, 2008. **72**(3): p. 637-648.
94. Bawa, Z., et al., *Understanding the yeast host cell response to recombinant membrane protein production*. 2011, Portland Press Ltd.
95. Roth, A.F., et al., *Global analysis of protein palmitoylation in yeast.* Cell, 2006. **125**(5): p. 1003-1013.
96. Vlastaridis, P., et al., *The pivotal role of protein phosphorylation in the control of yeast central metabolism.* G3: Genes, Genomes, Genetics, 2017. **7**(4): p. 1239-1249.
97. Timcenko, M., et al., *Structure and autoregulation of a P4-ATPase lipid flippase.* Nature, 2019. **571**(7765): p. 366-370.
98. Cartwright, S.P., et al., *Constitutively-stressed yeast strains are high-yielding for recombinant Fps1: implications for the translational regulation of an aquaporin.* Microbial cell factories, 2017. **16**(1): p. 41.
99. Manfrão-Netto, J.H., A.M. Gomes, and N.S. Parachin, *Advances in using Hansenula polymorpha as chassis for recombinant protein production.* Frontiers in bioengineering and biotechnology, 2019. **7**: p. 94.
100. Van Dijk, R., et al., *The methylotrophic yeast Hansenula polymorpha: a versatile cell factory.* Enzyme and microbial technology, 2000. **26**(9-10): p. 793-800.
101. Yurimoto, H., et al., *Regulation and evaluation of five methanol-inducible promoters in the methylotrophic yeast Candida boidinii.* Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 2000. **1493**(1-2): p. 56-63.
102. Yurimoto, H. and Y. Sakai, *Methanol-inducible gene expression and heterologous protein production in the methylotrophic yeast Candida boidinii.* Biotechnology and applied biochemistry, 2009. **53**(2): p. 85-92.
103. Fleer, R., *Engineering yeast for high level expression.* Current opinion in biotechnology, 1992. **3**(5): p. 486-496.
104. Giaever, G. and C. Nislow, *The yeast deletion collection: a decade of functional genomics.* Genetics, 2014. **197**(2): p. 451-465.
105. Souza, C.M., et al., *A stable yeast strain efficiently producing cholesterol instead of ergosterol is functional for tryptophan uptake, but not weak organic acid resistance.* Metabolic engineering, 2011. **13**(5): p. 555-569.
106. Van Meer, G., D.R. Voelker, and G.W. Feigenson, *Membrane lipids: where they are and how they behave.* Nature reviews Molecular cell biology, 2008. **9**(2): p. 112-124.
107. Tomimoto, K., et al., *Protease-deficient Saccharomyces cerevisiae strains for the synthesis of human-compatible glycoproteins.* Bioscience, biotechnology, and biochemistry, 2013. **77**(12): p. 2461-2466.
108. Lommel, M. and S. Strahl, *Protein O-mannosylation: conserved from bacteria to humans.* Glycobiology, 2009. **19**(8): p. 816-828.

109. Sethuraman, N. and T.A. Stadheim, *Challenges in therapeutic glycoprotein production.* Current opinion in biotechnology, 2006. **17**(4): p. 341-346.
110. Hamilton, S.R. and D. Zha, *Progress in yeast glycosylation engineering*, in *Glyco-Engineering*. 2015, Springer. p. 73-90.
111. Linares, D.M., E.R. Geertsma, and B. Poolman, *Evolved Lactococcus lactis strains for enhanced expression of recombinant membrane proteins.* Journal of molecular biology, 2010. **401**(1): p. 45-55.
112. Gul, N., et al., *Evolved Escherichia coli strains for amplified, functional expression of membrane proteins.* Journal of molecular biology, 2014. **426**(1): p. 136-149.
113. Cregg, J.M., *Introduction: distinctions between Pichia pastoris and other expression systems.* Methods in molecular biology (Clifton, NJ), 2007. **389**: p. 1-10.
114. Byrne, B., *Pichia pastoris as an expression host for membrane protein structural biology.* Current opinion in structural biology, 2015. **32**: p. 9-17.
115. Hohenblum, H., et al., *Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant Pichia pastoris.* Biotechnology and bioengineering, 2004. **85**(4): p. 367-375.
116. Love, K.R., et al., *Systematic single-cell analysis of Pichia pastoris reveals secretory capacity limits productivity.* PloS one, 2012. **7**(6).
117. Rao, R.V. and D.E. Bredesen, *Misfolded proteins, endoplasmic reticulum stress and neurodegeneration.* Current opinion in cell biology, 2004. **16**(6): p. 653-662.
118. Liu, Y. and A. Chang, *Heat shock response relieves ER stress.* The EMBO journal, 2008. **27**(7): p. 1049-1059.
119. Shiroishi, M., et al., *Platform for the rapid construction and evaluation of GPCRs for crystallography in Saccharomyces cerevisiae.* Microbial cell factories, 2012. **11**(1): p. 78.
120. Mumberg, D., R. Müller, and M. Funk, *Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds.* Gene, 1995. **156**(1): p. 119-122.
121. Mumberg, D., R. Müller, and M. Funk, *Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression.* Nucleic Acids Res, 1994. **22**(25): p. 5767-8.
122. Newstead, S., et al., *High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae*, in *Proc Natl Acad Sci U S A*. 2007. p. 13936-41.
123. Hartner, F.S. and A. Glieder, *Regulation of methanol utilisation pathway genes in yeasts.* Microbial Cell Factories, 2006. **5**(1): p. 39.
124. Zhang, A.-L., et al., *Recent advances on the GAP promoter derived expression system of Pichia pastoris.* Molecular Biology Reports, 2009. **36**(6): p. 1611-1619.
125. Darby, R.A., et al., *Which yeast species shall I choose? Saccharomyces cerevisiae versus Pichia pastoris*, in *Recombinant Protein Production in Yeast*. 2012, Springer. p. 11-23.
126. Logez, C., et al., *Preparation of Pichia pastoris Expression Plasmids*. 2012, Humana Press. p. 25-40.
127. Parker, J.L. and S. Newstead, *Method to increase the yield of eukaryotic membrane protein expression in Saccharomyces cerevisiae for structural and functional studies.* 2014. **23**(9): p. 1309-1314.
128. Williamson, V.M., et al., *Isolation of the structural gene for alcohol dehydrogenase by genetic complementation in yeast.* Nature, 1980. **283**(5743): p. 214-6.
129. Bennetzen, J.L. and B.D. Hall, *The primary structure of the Saccharomyces cerevisiae gene for alcohol dehydrogenase.* J Biol Chem, 1982. **257**(6): p. 3018-25.
130. Denis, C.L., J. Ferguson, and E.T. Young, *mRNA levels for the fermentative alcohol dehydrogenase of Saccharomyces cerevisiae decrease upon growth on a nonfermentable carbon source.* 1983.

131. McAlister, L. and M.J. Holland, *Differential expression of the three yeast glyceraldehyde-3-phosphate dehydrogenase genes.* J Biol Chem, 1985. **260**(28): p. 15019-27.
132. Schirmaier, F. and P. Philippsen, *Identification of two genes coding for the translation elongation factor EF-1 alpha of S. cerevisiae.* EMBO J, 1984. **3**(13): p. 3311-5.
133. Johnston, M. and R.W. Davis, *Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae.* 1984. **4**(8): p. 1440-1448.
134. Kerjan, P., H. Cherest, and Y. Surdin-Kerjan, *Nucleotide sequence of the Saccharomyces cerevisiae MET25 gene.* Nucleic Acids Res, 1986. **14**(20): p. 7861-71.
135. Waterham, H.R., et al., *Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter.* Gene, 1997. **186**(1): p. 37-44.
136. Ahn, J., et al., *Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter.* Appl Microbiol Biotechnol, 2007. **74**(3): p. 601-8.
137. Periyasamy, S., et al., *Isolation, characterization and evaluation of the Pichia pastoris sorbitol dehydrogenase promoter for expression of heterologous proteins.* 2013. **92**(1): p. 128-133.
138. Tschopp, J.F., et al., *Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris.* Nucleic Acids Res, 1987. **15**(9): p. 3859-76.
139. Cregg, J.M., et al., *Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris.* Mol Cell Biol, 1989. **9**(3): p. 1316-23.
140. Shen, S., et al., *A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris.* Gene, 1998. **216**(1): p. 93-102.
141. McKenzie, E.A. and W.M. Abbott, *Expression of recombinant proteins in insect and mammalian cells.* Methods, 2018. **147**: p. 40-49.
142. Yin, J., et al., *Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes.* Journal of Biotechnology, 2007. **127**(3): p. 335-347.
143. Goehring, A., et al., *Screening and large-scale expression of membrane proteins in mammalian cells for structural studies.* Nat Protoc, 2014. **9**(11): p. 2574-85.
144. Midgett, C.R. and D.R. Madden, *Breaking the bottleneck: Eukaryotic membrane protein expression for high-resolution structural studies.* Journal of Structural Biology, 2007. **160**(3): p. 265-274.
145. Ozvegy, C., et al., *Functional characterization of the human multidrug transporter, ABCG2, expressed in insect cells.* Biochem Biophys Res Commun, 2001. **285**(1): p. 111-7.
146. Sarramegna, V., et al., *Heterologous expression of G-protein-coupled receptors: comparison of expression systems from the standpointof large-scale production and purification.* Cellular and Molecular Life Sciences CMLS, 2020. **60**(8): p. 1529-1546.
147. Hardy, D., et al., *Functional Expression of Multidrug Resistance Protein 4 MRP4/ABCC4.* SLAS DISCOVERY: Advancing the Science of Drug Discovery, 2019. **24**(10): p. 1000-1008.
148. Qian, H., et al., *Structure of the human lipid exporter ABCA1.* Cell, 2017. **169**(7): p. 1228-1239. e10.
149. Ahn, M.-H., et al., *Production of therapeutic proteins with baculovirus expression system in insect cell.* Entomological Research, 2008. **38**: p. S71-S78.
150. Jarvis, D.L., *Chapter 14 Baculovirus–Insect Cell Expression Systems*. 2009, Elsevier. p. 191-222.
151. Mus-Veteau, I., *Heterologous Expression of Membrane Proteins for Structural Analysis*. 2010, Humana Press. p. 1-16.
152. Yovcheva, M., et al., *High-Titer Recombinant Protein Production: Employing the ExpiSf™ Chemically Defined Sf9 Insect Cell Expression System.* Genetic Engineering & Biotechnology News, 2018. **38**(13): p. 20-21.
153. Hitchman, R.B., et al., *Optimizing the baculovirus expression vector system.* 2011. **55**(1): p. 52-57.

154. Bieniossek, C., et al., *MultiBac: expanding the research toolbox for multiprotein complexes.* Trends in Biochemical Sciences, 2012. **37**(2): p. 49-57.
155. Sari, D., et al., *The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics*. 2016, Springer International Publishing. p. 199-215.
156. Zhai, Y., et al., *SmartBac, a new baculovirus system for large protein complex production.* Journal of Structural Biology: X, 2019. **1**: p. 100003.
157. Smith, G.E., M.J. Fraser, and M.D. Summers, *Molecular Engineering of the Autographa californica Nuclear Polyhedrosis Virus Genome: Deletion Mutations Within the Polyhedrin Gene.* J Virol, 1983. **46**(2): p. 584-93.
158. Knebel, D., H. Lübbert, and W. Doerfler, *The promoter of the late p10 gene in the insect nuclear polyhedrosis virus Autographa californica: activation by viral gene products and sensitivity to DNA methylation.* EMBO J, 1985. **4**(5): p. 1301-6.
159. Hill-Perkins, M.S. and R.D. Possee, *A baculovirus expression vector derived from the basic protein promoter of Autographa californica nuclear polyhedrosis virus.* Journal of General Virology, 1990. **71**(4): p. 971-976.
160. Junge, F., et al., *Large-scale production of functional membrane proteins.* Cellular and Molecular Life Sciences, 2008. **65**(11): p. 1729-1755.
161. Chang, G.-D., et al., *Improvement of glycosylation in insect cells with mammalian glycosyltransferases.* 2003. **102**(1): p. 61-71.
162. Jarvis, D.L., *Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production.* Virology, 2003. **310**(1): p. 1-7.
163. Andrell, J. and C.G. Tate, *Overexpression of membrane proteins in mammalian cells for structural studies.* Mol Membr Biol, 2013. **30**(1): p. 52-63.
164. Shukla, S., et al., *Use of Baculovirus BacMam Vectors for Expression of ABC Drug Transporters in Mammalian Cells.* Drug Metabolism and Disposition, 2012. **40**(2): p. 304-312.
165. Jain, N.K., et al., *A high density CHO-S transient transfection system: Comparison of ExpiCHO and Expi293.* Protein Expression and Purification, 2017. **134**: p. 38-46.
166. Wang, L., et al., *Structure and mechanogating of the mammalian tactile channel PIEZO2.* Nature, 2019. **573**(7773): p. 225-229.
167. Zhong, X., et al., *Transient CHO expression platform for robust antibody production and its enhanced N-glycan sialylation on therapeutic glycoproteins.* Biotechnology Progress, 2019. **35**(1): p. e2724.
168. Thomas, J.A. and C.G. Tate, *Quality Control in Eukaryotic Membrane Protein Overproduction.* Journal of Molecular Biology, 2014. **426**(24): p. 4139-4154.
169. Chaudhary, S., et al., *Overexpressing human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells.* Nature Protocols, 2012. **7**(3): p. 453-466.
170. Reeves, P.J., et al., *Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line.* Proceedings of the National Academy of Sciences, 2002. **99**(21): p. 13419-13424.
171. Meuris, L., et al., *GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins.* Nature Biotechnology, 2014. **32**(5): p. 485-489.
172. Kim, T.K. and J.H. Eberwine, *Mammalian cell transfection: the present and the future.* Analytical and Bioanalytical Chemistry, 2010. **397**(8): p. 3173-3178.
173. Büssow, K., *Stable mammalian producer cell lines for structural biology.* Current Opinion in Structural Biology, 2015. **32**: p. 81-90.
174. Baldi, L., et al., *Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives.* Biotechnology Letters, 2007. **29**(5): p. 677-684.

175. Yu, X., et al., *Using a Commonly Down-Regulated Cytomegalovirus (CMV) Promoter for High-Level Expression of Ectopic Gene in a Human B Lymphoma Cell Line.* Medical Science Monitor, 2017. **23**: p. 5943-5950.
176. Holzman, D., *Gene Therapy Depends on Finding The Right Vector.* 1995. **87**(6): p. 406-410.
177. Kim, D.W., et al., *Use of the human elongation factor 1 alpha promoter as a versatile and efficient expression system.* Gene, 1990. **91**(2): p. 217-23.
178. Reeves, P.J., J.M. Kim, and H.G. Khorana, *Structure and function in rhodopsin: A tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants*, in *Proc Natl Acad Sci U S A*. 2002. p. 13413-8.
179. Ho, F.Y. and B. Poolman, *Engineering Escherichia coli for functional expression of membrane proteins.* Methods Enzymol, 2015. **556**: p. 3-21.
180. Tandon, N., et al., *Generation of Stable Expression Mammalian Cell Lines Using Lentivirus.* Bio Protoc, 2018. **8**(21).