Improving recombinant eukaryotic membrane protein yields in *Pichia pastoris*: the importance of codon optimisation and clone selection

Running head: Optimisation of AQP production

FREDRIK ÖBERG[1], JENNIE SJÖHAMN[1], MATTHEW T. CONNER[2], ROSLYN M. BILL[2], KRISTINA HEDFALK[1]

[1] *Department of Chemistry/Biochemistry, University of Gothenburg, P.O. Box 462, SE-405 30 Göteborg, Sweden*

[2] *School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK*

fredrik.oberg@chem.gu.se +46 31 786 3941

jennie.sjohamn@chem.gu.se +46 31 786 3938

m.t.conner@aston.ac.uk +44 121204 3869

r.m.bill@aston.ac.uk +44 121204 4274

kristina.hedfalk@chem.gu.se +46 31 786 3923

Correspondence: Kristina Hedfalk; Department of Chemistry/Biochemistry, University of Gothenburg, PO Box 462, SE-405 30 Göteborg, Sweden. Tel: +46(0)31 786 3923. Fax +46(0)31 786 3910. E-mail kristina.hedfalk@chem.gu.se

Key words: recombinant protein production, integral membrane proteins, *P. pastoris*

Abstract

In the last 15 years, 80 % of all recombinant proteins reported in the literature were produced in the bacterium, *Escherichia coli*, or the yeast, *Pichia pastoris*. Nonetheless, developing effective general strategies for producing recombinant eukaryotic membrane proteins in these organisms remains a particular challenge. Using a validated screening procedure together with accurate yield quantitation, we therefore wished to establish the critical steps contributing to high yields of recombinant eukaryotic membrane protein in *P. pastoris*. Whilst the use of fusion partners to generate chimeric constructs and directed mutagenesis have previously been shown to be effective in bacterial hosts, we conclude that this approach is not transferable to yeast. Rather, codon optimisation and the preparation and selection of high-yielding *P. pastoris* clones are effective strategies for maximising yields of human aquaporins.

Introduction

Membrane proteins generally constitute about 30 % of the proteins in a proteome. These proteins are key molecules in the communication and transport processes of a cell and consequently the majority (60-70 %) of current drug targets are membrane proteins. However, membrane proteins are difficult to produce, purify and crystallize [1, 2], which hampers structure-aided drug design. Essentially, there are no more than 277 unique membrane protein structures (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html) in the Protein Data Bank (http://www.rcsb.org/) of which 35 % are eukaryotic and one quarter of these (9 %) are of human origin. Obtaining sufficient functional membrane protein for subsequent structural analysis remains an enduring bottleneck in this drug discovery pipeline.

The bacterium *E. coli* is the most common host for the heterologous production of soluble proteins, especially those of bacterial origin [3], but there is still limited success in the use of bacteria for the production of eukaryotic integral membrane proteins. In comparison, yeast is very suitable for structure-activity analyses of membrane proteins, having the advantages of a eukaryote, whilst being a microbe with well-established genetic tools [4]. For eukaryotic membrane proteins, there are good reasons to choose a eukaryotic rather than a prokaryotic host: the rate of polypeptide elongation is 4-10 times slower in eukaryotes compared to prokaryotes and the translocon is different [5]. Studies in yeast have shown that translation rather than transcription or targeting is the main limiting factor for membrane protein production [6, 7]. In addition, the lipid composition in prokaryotes and

eukaryotes differs, which has consequences for the localisation, stability and functionality of the recombinant membrane protein [8].

In the last 15 years, 80 % of all recombinant proteins reported in the literature were produced in the bacterium, *E. coli*, or the yeast, *Pichia pastoris* [9]. *P. pastoris* has gained increasing attention for its ability to drive expression from its powerful and tightly-regulated *AOX*1 promoter as well as its ability to reach very high cell densities in bioreactors [10]. Very high yields have been reported for soluble proteins expressed and purified from *P. pastoris*, e.g. 3-6 g/L of synthetic gelatine [11] and human growth factor [12]. While yields as high as 25 mg pure protein/L have been reported from *Saccharomyces cerevisiae* [13], 115 mg highly-pure membrane protein per litre culture are achievable in *P. pastoris* [14, 15]. Moreover, *P. pastoris* is a common source for eukaryotic membrane protein structures, a potassium channel [16] and a gated plant aquaporin [15] being just two of eight examples to date.

The *P. pastoris* system requires that the gene of interest is integrated into the host genome. Transformation of the corresponding plasmid DNA can be achieved by two different methods; chemical transformation using lithium chloride or electroporation. Electroporation is commonly recommended since the higher transformation frequency achieved by this method more likely results in a higher frequency of multiple insertions of the gene to be integrated [17]. This higher gene dosage, resulting in so-called 'jack-pot' clones, is assumed to further result in a higher yield of the final protein product [18]. However, for integral

membrane proteins this correlation has not been systematically examined, and indeed may not hold since an overloading of the transcription and translation machinery might occur. In this study, members of the aquaporin family that give different overproduction yields in *P. pastoris*, were therefore used to assess the impact of the transformation method on the final protein yield as quantified using our previously-validated method for accurate yield determination [14, 19].

A common strategy for the recombinant production of novel targets with an intrinsically low production yield is fusion with a stable and highly-produced peptide sequence. This strategy has been especially successful in *E. coli*. For soluble proteins, fusions to the maltose binding protein (MBP) as well as other fusion partners (e.g. the carboxyl terminal fragment of the *Mycobacterium tuberculosis* antigen) have significantly increased the overall yields of soluble recombinant proteins in both *E. coli* and yeast [20-22]. Recently, the protein Mistic was suggested to perform the same role for eukaryotic integral membrane proteins in bacteria [23, 24]. Further investigations of Mistic homologues revealed that the hydrophobic core and the length of the linker between the proteins are crucial for successful production of fusion proteins in *E. coli* [25]. The development of a similar approach for the production of eukaryotic membrane proteins in yeast was therefore evaluated in this study.

We and others have previously shown that there is substantial variation in production yields between highly-related proteins [19], which might suggest that the limiting factor lies within

the protein sequence itself, although the precise 'code' is not yet known. Knowledge of such codes could be used to drive high yielding protein production. In comparison, certain mutations could give rise to substantially higher production levels, like the redirection of the hAQP4 folding pathway resulting in yields enough for purification and characterisation [19]. However, such specific alterations of the sequence are rare based on limited information of the target of interest. A more general approach is gene optimisation which is mostly done in order to adapt the target sequence to the host based on the codon usage, mainly to avoid shortage of tRNAs. The codon usage in a certain organism is commonly referred to as Codon Adaptation Index (CAI) [26]. For situations when the gene and host are markedly different in CAI, codon adaptation is an appropriate strategy as exemplified by the successful production of the AQP from the malaria parasite *Plasmodium falciparum* in *P. pastoris* [27]. However, for genes and hosts that are very similar in CAI, gene optimisation might not be the obvious way forward, as exemplified by the human aquaporins produced in *P. pastoris*. Notably, even though the human aquaporins represent a fairly homogenous group of targets, their yield in *P. pastoris* showed a surprisingly high variation [19]. Therefore, gene optimisation in the context of differences in CAI, gene sequence repeats, codon usage, GC content, polyA sites, splice sites and RNA secondary structure (Geneart; based on intellectual property in patent WO2004059556 'Method and device for optimising a nucleotide sequence for the purpose of expression of a protein') was tested for one of the poorest producing human aquaporins.

Here, we combined a rapid small-scale screening procedure with accurate yield quantitation to identify the critical steps contributing to increased final eukaryotic membrane protein

yield in *P. pastoris*. We show that the use of a fusion strategy is not transferable from *E. coli* to yeast. Instead the choice of transformation protocol, the clone selection step and codon optimisation are critical in establishing a high-yielding production experiment.

Methods

Cloning

All construct are based on the *P. pastoris* vector pPICZB (Invitrogen) for intracellular protein production. Primers used for specific constructs are listed in Table S1. For all PCR reactions and ligations, Phusion DNA polymerase (Finnzymes) and Quick T4 DNA Ligase (New England Biolabs), respectively, were used. For all constructs, C-terminal hexa-histidine tag directly followed by a stop codon, was included. In brief, a codon optimised gene of Mistic for production in *P. pastoris* was ordered from GeneArt. The gene was amplified by PCR and cloned into the *P. pastoris* vector pPICZB creating a 16 amino acid linker, containing a spacer region and a TEV protease cleavage site (see figure S1), between the protein and the *Xba*I site. This construct, Opt-Mistic/pPICZB, was used as vector for the M1 and M2 construct (see Fig. 3A). The full length hAQP1 and hAQP8 was amplified and inserted downstream of the linker in Opt-Mistic/pPICZB. A codon optimised gene of hAQP4 for production in *P. pastoris* was ordered from GenScript. The gene was flanked by restriction sites for *Sfu*I and *Xba*I which were used to sub-clone the gene into the *P. pastoris* vector pPICZB, creating the construct Opt-hAQP4/pPICZB. For imaging, a pPICZB vector containing Strongly enhanced Green Fluorescence Protein (SGFP2) [28] was made by amplification of SGFP2. The resulting vector, SGFP2/pPICZB, was used in creation of hAQP1-SGFP2 and hAQP8-SGFP2 where the hAQP1 and the hAQP8 gene, respectively, was cloned into SGFP2/pPICZB upstream of the SGFP2 gene creating a 12 amino acid long linker with a spacer region (see figure S1). For F1 (see Fig. 4A) two primers were annealed to each other creating a double stranded DNA fragment with two cleavage sites for restriction enzymes. This fragment was inserted into a new vector, hAQP8M/pPICZB (described below), upstream of the hAQP8 gene with

restriction enzymes creating a linker of 16 amino acids, containing a spacer region and a TEV protease cleavage site (see figure S1), between the two genes. The hAQP8M/pPICZB vector contains hAQP8 in pPICZB with the first methionine removed to prevent transcription start. F2 was made in one single PCR reaction where the forward primer contained a 17 amino acid long linker with a spacer region and a TEV protease cleavage site (see figure S1). The PCR product was inserted into pPICZB-hAQP1-Myc-His_6 [14] downstream the hAQP1 gene. For C1 (see Fig. 5A) the N-terminus of hAQP1 was included in the forward primer, whereas for C2-C4 (see Fig. 5A) two different gene fragments were amplified in separate reactions and ligated in a blunt end ligation. The resulting DNA fragments were used as a template for a new PCR reaction and the resulting chimeric fragment were cloned into the *P. pastoris* vector pPICZB. S1 (see Fig. 6A) was made in the same way as C1 whereas for S2 (see Fig. 6A) two different reactions where needed to create two PCR fragments with complementary parts. A third PCR reaction with the $S2_{SoPIP2;1\ -260}$ forward primer and the $S2_{hAQP1\ 228-}$ reverse primer created the final fragment which was subsequently cloned into the *P. pastoris* vector pPICZB.

All constructs were confirmed correct by sequencing (Eurofins MWG Operon).

High zeocin concentration screen

In the general screen, 20 transformants were screened for growth on YPD plates containing high concentration of zeocin (2000 µg/mL) and analysed after two to seven days. The protein production yields of selected transformants were quantified using the small scale screening method for protein production as described in previous work [12, 19]. To optimise

the screen for growth on high zeocin YPD plates, 40 colonies were compared in three different variants of the method. In the first screen, a pipette tip was used to streak cells on the high zeocin plate, having the intrinsic disadvantage that slightly different amounts of cells are streaked out which in turn affect the accurate comparison. For a more reliable comparison, the screen was improved by resuspending the cells in YPD media and adjusting OD_{600} to 20 mOD units in each drop, assuring that the same amount of cells were plated for each transformant. To improve the screen even further, the volume in each drop was adjusted in addition to the amount of cells. In this case 10 mOD of cells where plated in each drop. The plates where incubated at 30°C, and visually observed for seven days.

Protein production in P. pastoris

All constructs were linearized using the *Pme*I restriction endonuclease (Fermentas) before transformation to the *P. pastoris* wild type strain X-33. The substitution constructs S1 and S2 in figure 6 was transformed using the lithium chloride method, since this was the original transformation method for the reference protein SoPIP2;1. The lithium chloride transformation method in brief; *P. pastoris* cells are grown to an OD600 of 0.9, washed with water and LiCl and resuspended in LiCl. Cells, PEG, carrier DNA and the plasmid DNA is mixed in an eppendorf tube, incubated at 30°C for 1h and then subjected to a heat shock at 42°C for 25min. Finally, the cells are incubated for up to 4h at 30°C and spread on zeocin containing YPD plates. Also, for comparison of transformation methods, the lithium chloride method was also used for selected constructs (hAQP1, hAQP4 and hAQP8) in Fig 1. Based on the results from Fig. 1, all other constructs were transformed by electroporation (Invitrogen). In brief, *P. pastoris* cells are grown to an OD_{600} of 1.3, washed by ice cold

distilled water and resuspended in cold sorbitol. Cells and DNA is mixed in an electroporation cuvette and subjected to an electric pulse of 2.2kV directly followed by addition of sorbitol. Finally, the cells are incubated for one hour at 30°C and spread on zeocin containing YPDS plates. Protein production, quantitation of total production yields and densitometry measurements was made as described earlier [19].

Determination of expression profile

To measure the membrane expression profile, line intensity profiles were generated of confocal images of individual live cells. A minimum of three line profiles distributed at regular intervals covering the entire cell were measured in at least three independent experiments. The fluorescence intensity over this distance was also measured and the difference between the mean of the two outermost fluorescence peaks and the mean of the fluorescence between these two peaks was divided by the maximum fluorescence along the line scan to calculate the percentage of fluorescence at the outermost membrane. This was termed the relative expression (RE) [29, 30].An RE of 100% indicated that all fluorescence was detected in the outer circumference of the cell. An RE of 0% indicated that all fluorescence was detected in the cytoplasm and/or within invaginated membranes with no outer circumference fluorescence.

Results

Electroporation is superior to lithium chloride transformation for generating high-yielding P. pastoris *clones*

Chemical transformation and electroporation were evaluated for the generation of high-yielding *P. pastoris* clones. Members of the aquaporin family, for which we had previously established accurate production yields in *P. pastoris* were selected as targets in order to determine whether their yields could be further improved in the clone generation step. hAQP1, hAQP8 and hAQP4 had previously been shown to be produced respectively in high-, moderate- and no-yield following chemical transformation [19]. Total production yields were therefore quantified after either chemical transformation or electroporation. Using our established method for reliable quantitation of membrane protein production yields, we estimated the relative yields in total cell extracts using immunoblots of triplicate cultures compared with a reference target and an internal standard [19]. The resulting yield of each target from whole cell lysates was quantified as shown in Fig. 1A. hAQP1 was seen in its monomeric form, hAQP4 was distributed as both monomers and dimers and hAQP8 was monomeric with some degradation. For hAQP1, the production yield was 40 % that of our highest yielding clone (SoPiP2:1 [15]), irrespective of the transformation method used (data not shown). For hAQP8, which is produced to moderate yields in *P. pastoris* after chemical transformation, the production was increased by a factor of five when using electroporation (Fig. 1B). The yield of hAQP4 was below the detection limit after chemical transformation, while electroporation resulted in yields promising for scale up and purification trials (Fig. 1B).

Screening at high zeocin concentrations is an efficient way to identify P. pastoris *clones with high membrane protein yield*

In order to increase the efficiency of screening for positive transformants, we examined the relationship between the final recombinant protein yield in whole cell lysates and the growth of the recombinant host at high zeocin concentrations (2000 µg/mL; Fig. 2A). For the aquaporins under study, a clear correlation between the ability to grow on high concentrations of antibiotic and membrane protein production yields emerged, where the largest colonies also showed the highest protein production yield, consistent with the observation that multiple insertion events can lead to high protein yields. To make this qualitative assessment more quantitative, an equal density of cells was plated (Fig. 2C) and compared to plates in which the volume was additionally adjusted to be equivalent for all samples (Fig. 2D). In the simplest variant of the screening method (Fig. 2B), differences were difficult to identify on account of differences in cell density. A more accurate variant therefore compared an equal density of cells (Fig. 2C). However, the most reliable screen was achieved when both the cell density and volume were adjusted (Fig 2D). Based on this screen, the crude extract from the two best growing colonies, 11 and 21 (Fig. 2D; Table 1), were compared on an immunoblot confirming that production was high, with colony 21 giving a better yield than colony 11 (data not shown).

Recombinant hAQP1 and hAQP8 are localized to the P. pastoris *membrane*

Confocal imaging of recombinant *P. pastoris* cells showed membrane localization of both hAQP1-GFP and hAQP8-GFP (Fig. 3). This supports the production of correctly-folded,

recombinant human aquaporins in this host. Bleaching experiments showed that bleaching of hAQP8-GFP was more rapid than hAQP1-GFP suggesting that hAQP1-GFP is produced at higher yields than hAQP8-GFP, as previously shown by immunoblot for the same proteins not fused to GFP (Fig. 1A). These data also indicate that the expression profiles for the two aquaporins are different at 48h post-induction, with an RE of 79±2% and 27±2% hours for hAQP8-GFP and hAQP1-GFP, respectively.

Mistic-hAQP fusions are stable in P. pastoris *but do not increase yield*

We examined whether production of our low-yielding AQP targets could be improved by fusing them to Mistic, a protein which has previously been shown to be stably produced at high yields in bacterial membranes. Chimeras were made in various combinations (Fig. 4A). For use in *P. pastoris*, Mistic was specifically codon optimised (Opt-Mistic) and its production was found to be remarkably stable (Fig. 4B) with yields in the same range as the highest hAQP1 yield (Fig. 4C). The fusion of Opt-Mistic to hAQP1 (M1) resulted in a stable product of the expected molecular weight where the monomeric band was dominant on the immunoblot but a dimeric form was also observed (Fig. 4B). For Opt-Mistic-hAQP8 (M2) the majority of the product corresponded to the expected molecular weight, but some degradation, as seen for hAQP8 alone, was also observed (Fig. 4B). However, for both M1 and M2, the fusion to Mistic resulted in lower yields compared to non-fused targets (Fig. 4C). These data suggest that the concept of increasing eukaryotic membrane protein yields using Mistic partners cannot be extrapolated from *E. coli* to *P. pastoris*.

The carboxy-terminal hydrophilic extensions of AQP1 may direct high yields of human aquaporins

In order to explore further the use of a fusion strategy, hAQP1 was identified as a highly-homologous partner for low-yielding hAQP8. Full length hAQP8 was therefore fused either to the amino terminus of hAQP1 (F1: Fig. 5A), or to the whole protein (F2; Fig. 5A). As seen in Fig 5B, full-length F1 was produced as well as a smaller protein product of the same size as hAQP8 itself (27 kDa). For F2, the band on the immunoblot was smaller than any of the individual fusion partners consistent with degradation and instability of the product (Fig. 5B). Overall the total yields were not improved (Fig. 5C). For F1 and F2, a linker of 17 amino acids had been included between the two proteins, but when this was removed the same outcome was obtained (data not shown). When hAQP1 TMD1, TMD1-2 or TMD1-3, respectively, were also fused to the amino terminus of hAQP8 there was no improvement of the yield (data not shown).

To more specifically identify regions of AQP1 that might direct high yields, four additional chimeric constructs were made (C1-C4; Fig. 6A). These constructs were comprised of hAQP8 in which the amino-terminus, TMD1, TMD1-2 or TMD1-3 were substituted in turn for the corresponding sequences of hAQP1. When the amino-terminal sequences of hAQP8 were substituted with the corresponding hAQP1 sequences, yields were very low (Fig. 6C). In contrast, exchanging the first two or three helices did not substantially affect the yield (Fig. 6C). This pointed to the importance of the carboxy-terminal half of hAQP1 being a potential determinant of successful membrane protein overproduction.

In order to investigate the role of the termini with respect to production yields, the hydrophilic domains of two highly-produced targets, hAQP1 and SoPIP2;1, were exchanged. The amino- and carboxy-termini of SoPIP2;1 were replaced with the corresponding sequences of hAQP1, to give S1 and S2, respectively (Fig. 7A). For this approach, the hAQP1 construct giving the highest yield in *P. pastoris*; hAQP1-Myc-His_6 with a yield of 1.5 relative to SoPIP2;1 production [19], was used. Notably, S1 and S2 were stably produced, no degradation was observed (Fig. 7B) and the yield of SoPIP2;1 was not affected by the exchange of the amino-terminus. In contrast, exchange of the carboxy-terminus led to an even higher yield than for SoPIP2;1 itself (Fig. 7C) supporting the view that sequences within the carboxy-terminus of hAQP1 can direct high yields of human aquaporins.

Codon optimisation results in substantially higher production yields for hAQP4

As part of this study, the Mistic protein was codon optimised for production in *P. pastoris* and high yields were observed (Fig. 4D). hAQP4 was therefore selected for codon optimisation, giving Opt-hAQP4, on account of its low yields ([19] and Fig. 1). Opt-hAQP4 showed a substantial increase in yield comparable with that of hAQP1 (Fig. 8B), clearly supporting the use of codon optimisation in order to increase initial production yields. As seen in Fig. 8A, Opt-hAQP4 was stably produced in *P. pastoris* and bands corresponding to both the monomeric and oligomeric forms were observed on the immunoblot. In a previous study, two mutations directing the folding pathway were introduced into hAQP4 (M48N, L50K; [19]) resulting in hAQP4* being produced at higher yields than hAQP4. We noted that

codon optimisation exceeded the benefits of this mutational strategy by a factor of two. When hAQP4* in addition was codon optimised, giving Opt-hAQP4*, stable products at slightly higher yields than Opt-hAQP4 was achieved in *P. pastoris* (Fig. 8B). In both cases, applying mutation or optimisation, there is sufficient material for further functional characterisation and crystallisation attempts of the full length hAQP4 protein.

Discussion

In order to optimise any protein production experiment, several parameters must be considered. Here we evaluated the role of clone selection, fusion partners and codon optimisation in obtaining high yields of members of the aquaporin family.

Clone selection

Chemical transformation using lithium chloride has previously been used routinely in generating *P. pastoris* strains [19] despite the fact that electroporation is likely to result in a higher frequency of multi-copy integrants [17], which in turn should result in higher production yields. While several studies have suggested that slowing the rate of translation may increase the functional yield of recombinant membrane proteins [5], including members of the aquaporin family [7], a link has not been established between gene dosage and yield in these reports. Consequently, we investigated the impact of the transformation method on the final yield. For the low-yielding targets, hAQP4 and hAQP8, a significant yield improvement was observed following electroporation (Fig. 1) suggesting that low yielding targets benefit from increasing gene dosage. This was supported by the observation that the highest yielding clones also survived growth on high zeocin concentrations, as multi-copy insertions contain repeated insertions of the gene of interest together with a zeocin resistance gene. The frequency of multi-copy insertion is predicted to be 1-10% and hence it would be necessary to screen up to 1000 transformants in order to identify them. We noted that screening on 2000 μg/mL zeocin is a rapid and effective way to identify these high-yielding clones.

Notably, confocal microscopy imaging of hAQP1-GFP and hAQP8-GFP showed that both high- and moderate-yielding targets were localized to the yeast membrane, supporting proper processing of recombinant integral membrane proteins in the host cell. Moreover, the total amount estimated by imaging is in good agreement with the total production yields quantified by immunoblot (Fig. 4C). Interestingly, the expression profile of the two proteins in *P. pastoris* cells differed (Fig. 3). This observation could be consistent with saturation and overload resulting from a membrane protein production level as high as the one observed for hAQP1 in the *P. pastoris* plasma membrane [14, 19].

Fusion partners

In this study a set of constructs was designed with different combinations of fusion partners, chimeras and domain swaps analogous to previous approaches predominantly in *E. coli* [20-24]. In a first approach, Mistic [23] was codon optimised for production in *P. pastoris* and used as a fusion partner for hAQP1 and hAQP8 representing a high- and a low yielding target, respectively. Even though codon-optimised Mistic itself was produced at high yields in *P. pastoris*, it did not direct high yields of its fusion partners (Fig. 4B). A possible explanation for this is that fusion of the two proteins results in a chimera that cannot be easily processed by the yeast translocon machinery, compared with its prokaryotic counterpart. More likely, however, is that Mistic, which is a prokaryotic protein, does not behave in the same way when produced in a yeast cell and, hence, does not bring the same positive influence in a eukaryotic context. In contrast, a eukaryotic membrane protein that has already been shown to be stably produced in high yields in *P. pastoris* should be a

suitable fusion partner. Consequently, hAQP1 [14, 19] was selected to drive the production of a homologous, but less well produced protein, hAQP8. Fusion with either the full-length protein, or parts of its amino-terminus was not a successful strategy resulting in unstable products and low yields (Fig. 5). Interestingly, different folding routes are observed for aquaporins when their transmembrane domains are inserted in various orders which may explain this observation [31]. In relation to this, shorter amino-terminal exchanges in general resulted in more unstable proteins compared to the situation where half the protein was exchanged (Fig. 5). This coincides well with the fact that aquaporins are comprised of an inverted repeat of three helices [32], and hence chimeras combining two homologous repeats are more stable.

To follow up the observation that shorter swaps were less tolerated when fused to hAQP8, shorter domains were swapped between two proteins both known to be stably produced to high yields in yeast, namely hAQP1 and SoPIP2;1 [14, 33]. Notably, both the N- and the C-terminus of SoPIP2;1 could be replaced by the corresponding termini from hAQP1 without the following decrease in production yield previously observed for hAQP1-hAQP8 chimeras indication stable formation of stable product formation from those domain swaps (Fig. 7). Noteworthy, the substitution of the C-terminus resulted in an even higher yield compared to native SoPIP2;1. Mainly, this observation emphasize that the importance of the hydrophilic extensions for high production yield. In addition, the stable production of domain swaps between hAQP1 and SoPIP2;1 implies that these two proteins share the same folding pathway, while hAQP8 has another folding route giving rise to unstable hAQP1-hAQP8 chimeras (Fig. 6). Our previous results showing that the successful strategy to

improve the production yield of hAQP4 by a supposed redirection of its folding pathway [19] further supports such an explanation.

Codon optimisation

Codon bias is an intrinsic problem in heterologous protein production and need to be taken into account in the experimental design [34]. Codon optimisation in order to optimise the sequence for a specific host has been used with great success previously. Especially when the codon usage for the gene and the specific host is severely different, as for *Plasmodium falciparum* AQP produced in *P. pastoris* [27], codon optimisation is the logic approach. On the contrary, a totally different situation emerged for human proteins to be produced in yeast. More specifically, Codon Adaptation Index (CAI) for one of the low yielding human aquaporin targets, hAQP4, is very similar in human and in yeast; 0.760 and 0.707, respectively (http://genomes.urv.es/CAIcal). Hence, the impression is that underlying reason for the low production yield of hAQP4 in *P. pastoris* could not be explained by CAI. Nevertheless, the optimised hAQP4 gene (Opt-hAQP4) showed a dramatic improvement in production yield (Fig. 8), even a bit higher than that of mutated hAQP4 (hAQP4*) [19]. The likely explanation to this is the benefit from improved mRNA stability in the specific host which is the additional improvement besides CAI adaption included in the gene optimisation concept. While the codon optimisation as such is likely to improve the translation process making sure there is no depletion of tRNAs for a specific codon, other strategies can be applied to improve the translocation process specific for integral membrane proteins. Noteworthy, translocation is highlighted as the main bottleneck for successful membrane protein production, especially for heterologous production [5]. Indeed, specific mutations

affecting the folding pathway and tetramer stability of hAQPs has also been showed to have a very positive influence on the production yield [19]. However, for such specific protein dependent strategies, a deep knowledge of the selected target is critical, knowledge that to date could be regarded as luxury taking collection of substantial amounts of functional and structural information on each individual protein. However, for the few cases when such information is available, such a stream lines design, combining strategies improving the translation and translocation processes (Opt-hAQP4*), would have the highest success rate, se Fig. 8B.

Conclusion

When getting as far in the optimisation as possible with the construct design, the other aspect to consider is to optimise growth conditions to maximize the membrane protein production yield. From our previous work we know that optimal growth conditions for the host is not at all the same as optimal production conditions for a specific protein [7]. Hence, especially for an integral membrane protein where the production involves more partners in the cellular machinery compared to their soluble counterparts, growth is a critical consideration. Taken together, a substantial improvement on membrane protein production can be achieved by method optimisation achieving and selecting high yielding transformants. For the gene design, more information has to be gathered in this precious field in order to understand and be able to conduct a high yielding production experiment of a desired membrane protein target. At its current status, it clearly takes directed approaches for a specific protein to be successful. Hence, a deep knowledge about the target of interest is needed. Thus, we are left with the major challenge to find general concepts valid for any membrane protein. However, the collected message from this study, and other studies of this kind, is that large changes on the gene level tend to have a negative influence on the production while subtle changes and directed protein specific strategies appears to be the solution, a conclusion that appears exceptionally valid for an integral membrane protein.

Acknowledgements

We would also like to thank Anna Polyakova for assistance in making the Opt-hAQP4 clone.

Declaration of interest

This work was supported by contracts LSHG-CT-2004-504601 (E-MeP), LSHG-CT-2006-037793 (OptiCryst), HEALTH-F4-2007-201924 (EDICT) and QLG2-CT-2002-00988 (SPINE).

References

[1] White S H. 2009. Biophysical dissection of membrane proteins. Nature 459:344-346.

[2] Lundstrom K. 2006. Structural genomics: the ultimate approach for rational drug design. Mol Biotechnol 34:205-212.

[3] Eshaghi S, Hedren M, Nasser M I, Hammarberg T, Thornell A, Nordlund P. 2005. An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. Protein Sci 14:676-683.

[4] Bill R M. 2001. Yeast--a panacea for the structure-function analysis of membrane proteins? Curr Genet 40:157-171.

[5] Tate C G. 2001. Overexpression of mammalian integral membrane proteins for structural studies. FEBS Lett 504:94-98.

[6] Joseph-Liauzun E, Farges R, Le Fur G, Ferrara P, Loison G. 1995. High-level production of a human membrane protein in yeast: the peripheral-type benzodiazepine receptor. Gene 155:195-199.

[7] Bonander N, Hedfalk K, Larsson C, Mostad P, Chang C, Gustafsson L, Bill R M. 2005. Design of improved membrane protein production experiments: quantitation of the host response. Protein Sci 14:1729-1740.

[8] Tate C G, Haase J, Baker C, Boorsma M, Magnani F, Vallis Y, Williams D C. 2003. Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. Biochim Biophys Acta 1610:141-153.

[9] Sorensen H P. Towards universal systems for recombinant gene expression. Microb Cell Fact 9:27.

[10] Cereghino G P, Cereghino J L, Ilgen C, Cregg J M. 2002. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotechnol 13:329-332.

[11] Werten M W, Wisselink W H, Jansen-van den Bosch T J, de Bruin E C, de Wolf F A. 2001. Secreted production of a custom-designed, highly hydrophilic gelatin in *Pichia pastoris*. Protein Eng 14:447-454.

[12] Fantoni A, Bill R M, Gustafsson L, Hedfalk K. 2007. Improved yields of full-length functional human FGF1 can be achieved using the methylotrophic yeast *Pichia pastoris*. Protein Expr Purif 52:31-39.

[13] Lenoir G, Menguy T, Corre F, Montigny C, Pedersen P A, Thines D, le Maire M, Falson P. 2002. Overproduction in yeast and rapid and efficient purification of the rabbit SERCA1a Ca(2+)-ATPase. Biochim Biophys Acta 1560:67-83.

[14] Nyblom M, Oberg F, Lindkvist-Petersson K, Hallgren K, Findlay H, Wikstrom J, Karlsson A, Hansson O, Booth P J, Bill R M, Neutze R, Hedfalk K. 2007. Exceptional overproduction of a functional human membrane protein. Protein Expr Purif 56:110-120.

[15] Tornroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P. 2006. Structural mechanism of plant aquaporin gating. Nature 439:688-694.

[16] Chen X, Wang Q, Ni F, Ma J. Structure of the full-length Shaker potassium channel Kv1.2 by normal-mode-based X-ray crystallographic refinement. Proc Natl Acad Sci U S A 107:11352-11357.

[17] Invitrogen C. 2005. EasySelect Pichia Expression Kit, version G. http://www.invitrogen.com/content/sfs/manuals/easyselect_man.pdf.

[18] Cregg J M, Tolstorukov I, Kusari A, Sunga J, Madden K, Chappell T. 2009. Expression in the yeast *Pichia pastoris*. Methods Enzymol 463:169-189.

[19] Oberg F, Ekvall M, Nyblom M, Backmark A, Neutze R, Hedfalk K. 2009. Insight into factors directing high production of eukaryotic membrane proteins; production of 13 human AQPs in *Pichia pastoris*. Mol Membr Biol 26:215-227.

[20] Wang A, Clapper J, Guderian J A, Foy T M, Fanger G R, Retter M W, Skeiky Y A. 2003. A novel method for increasing the expression level of recombinant proteins. Protein Expr Purif 30:124-133.

[21] Hennig L, Schafer E. 1998. Protein purification with C-terminal fusion of maltose binding protein. Protein Expr Purif 14:367-370.

[22] Perez-Martin J, Cases I, de Lorenzo V. 1997. Design of a solubilization pathway for recombinant polypeptides in vivo through processing of a bi-protein with a viral protease. Protein Eng 10:725-730.

[23] Roosild T P, Greenwald J, Vega M, Castronovo S, Riek R, Choe S. 2005. NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. Science 307:1317-1321.

[24] Petrovskaya L E, Shulga A A, Bocharova O V, Ermolyuk Y S, Kryukova E A, Chupin V V, Blommers M J, Arseniev A S, Kirpichnikov M P. Expression of G-protein coupled receptors in Escherichia coli for structural studies. Biochemistry (Mosc) 75:881-891.

[25] Roosild T P, Vega M, Castronovo S, Choe S. 2006. Characterization of the family of Mistic homologues. BMC Struct Biol 6:10.

[26] Sharp P M, Li W H. 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281-1295.

[27] Hedfalk K, Pettersson N, Oberg F, Hohmann S, Gordon E. 2008. Production, characterization and crystallization of the *Plasmodium falciparum* aquaporin. Protein Expr Purif 59:69-78.

[28] Kremers G J, Goedhart J, van den Heuvel D J, Gerritsen H C, Gadella T W, Jr. 2007. Improved green and blue fluorescent proteins for expression in bacteria and mammalian cells. Biochemistry 46:3775-3783.

[29] Viard P, Butcher A J, Halet G, Davies A, Nurnberg B, Heblich F, Dolphin A C. 2004. PI3K promotes voltage-dependent calcium channel trafficking to the plasma membrane. Nat Neurosci 7:939-946.

[30] Conner M T, Conner A C, Brown J E, Bill R M. Membrane trafficking of aquaporin 1 is mediated by protein kinase C via microtubules and regulated by tonicity. Biochemistry 49:821-823.

[31] Foster W, Helm A, Turnbull I, Gulati H, Yang B, Verkman A S, Skach W R. 2000. Identification of sequence determinants that direct different intracellular folding pathways for aquaporin-1 and aquaporin-4. J Biol Chem 275:34157-34165.

[32] Reizer J, Reizer A, Milton H. Saier J. 1993. The MIP Family of Integral Membrane Channel Proteins: Sequence Comparisons, Evolutionary Relationships, Reconstructed Pathway of Evolution, and Proposed Functional Differentiation of the Two Repeated Halves of the Proteins. Critical Reviews in Biochemistry and Molecular Biology 28:235-257.

[33] Karlsson M, Fotiadis D, Sjovall S, Johansson I, Hedfalk K, Engel A, Kjellbom P. 2003. Reconstitution of water channel function of an aquaporin overexpressed and purified from *Pichia pastoris*. FEBS Lett 537:68-72.

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

Figure legends

Figure 1. Total production yield as a result of different transformation methods.

(A) Immunoblot showing whole cell lysate for hAQP1, hAQP4 and hAQP8, respectively, from triplicate *P. pastoris* cultures. The proteins are detected using the His-tag. Chemical transformation is indicated by c and electroporation with e, respectively, after each protein name. (B) Bar graph showing the quantified production yield relative the reference protein SoPIP2;1. The left bar indicates the total production yield of clones transformed with the lithium chloride method whereas the right bar shows clones transformed with Electroporation

Figure 2. High zeocin growth screen

Showing YPD plates containing high concentration of zeocin to find colonies able to grow during these conditions, suggesting they have multiple inserts of the expression cassette containing the zeocin resistance gene. (A) Immunoblot showing the total production yield after a small scale screen for protein production in *P. pastoris* made from a selection of colonies shown on the plate below. Colony number is indicated above the immunoblot and under the plate, respectively. The correlation between colony size and signal strength has been indicated with two arrows. (B) Screen where different amount of cells are spread on a plate to quickly find colonies which can survive on high concentration of zeocin. Left plate has been incubated for two days and the right plate for seven days. (C) Screen where equal amounts of cells resuspended in liquid media are plated but with a different total volume. Left plate has been incubated for two days and the right plate for three days. (D) Screen

where both the amounts of cells and the volume of the drop has been adjusted. Left plate has been incubated for two days and the right plate for three days.

This Figure is reproduced in colour in Molecular Membrane Biology online.

***Figure 3. Membrane localisation of hAQP1 and hAQP8 in* P. pastoris**

Production profiles of mutant AQP-GFP fusion proteins in *P. pastoris* where a) shows hAQP1 and b) shows hAQP8. Cells shown are the same culture at 24 h and 48 h post-induction sequentially, from left to right. Line scans below each image are indicated with a yellow line.

***Figure 4. Production of Mistic fusions in* P. pastoris**

(A) Fusion constructs of Opt-Mistic and human aquaporin homologues; M1 and M2. hAQP1 (yellow), hAQP8 (green) and the linker (light blue) are indicated. (B) Immunoblots showing the total production in whole *P. pastoris* cell extracts of Mistic fusions; M1 and M2; as well as individual fusion partners. Samples representing the total production yield in *P. pastoris* are loaded in triplicates for each construct. The amount of total protein loaded for each construct is stated above each immunoblot. (C) Bar chart showing the total production yields of Mistic fusions and individual fusion partners.

All yields are relative the SoPIP2;1 production, for which the production is set to one. The y-axis represent average from triplicate cultures, error bars show the standard deviation (n=3).

This Figure is reproduced in colour in Molecular Membrane Biology online.

***Figure 5. Production of hAQP1 fusions in* P. pastoris**

(A) Fusion constructs of human aquaporin homoglous; F1 and F2. hAQP1 (yellow), hAQP8 (green) and the linker (light blue) are indicated. (B) Immunoblots showing the total production in whole *P. pastoris* cell extracts of fusion constructs; F1 and F2. Samples representing the total production yield in *P. pastoris* are loaded in triplicates for each construct. The amount of total protein loaded for each construct is stated above each immunoblot. (C) Bar chart showing the total production yield of fusion constructs; F1 and F2, relative hAQP8. All yields are relative the SoPIP2;1 production, for which the production is set to one. The y-axis represent average from triplicate cultures, error bars show the standard deviation (n=3).

This Figure is reproduced in colour in Molecular Membrane Biology online.

***Figure 6. Production of hAQP chimeras in* P. pastoris**

(A) Chimeric constructs of aquaporin homologues; C1-C4. hAQP1 (yellow) and hAQP8 (green) are indicated. (B) Immunoblots showing the total production in whole *P. pastoris* cell extracts of chimeric constructs; C1-C4. Samples representing the total production yield in *P. pastoris* are loaded in triplicates for each construct. The amount of total protein loaded for each construct is stated above each immunoblot. (C) Bar chart showing the total production yield of chimeric constructs; C1-C4, relative hAQP8. All yields are relative the SoPIP2;1 production, for which the production is set to one. The y-axis represent average from triplicate cultures, error bars show the standard deviation (n=3).

This Figure is reproduced in colour in Molecular Membrane Biology online.

***Figure 7. Production of hAQP domain swaps in* P. pastoris**

(A) Domain swaps of aquaporin homogogues; S1 and S2. SoPIP2;1 (blue) and hAQP1 (yellow) are indicated. (B) Immunoblots showing the total production in whole *P. pastoris* cell extracts of domain swaps; S1 and S2. Samples representing the total production yield in *P. pastoris* are loaded in triplicates for each construct. The amount of total protein loaded for each construct is stated above each immunoblot. (C) Bar chart showing the total production yield of domain swaps; S1 and S2. All yields are relative the SoPIP2;1 production, for which the production is set to one. The y-axis represent average from triplicate cultures, error bars show the standard deviation

This Figure is reproduced in colour in Molecular Membrane Biology online.

***Figure 8. Production of gene optimised hAQP4 in* P. pastoris**

(A) Immunoblot showing the production of hAQP4, hAQP4*, Opt-hAQP4 and Opt-hAQP4*, respectively, in total cell extracts. Samples representing the total production yield are loaded in triplicates for each construct produced in *P. pastoris.* The amount of protein loaded for each target is indicated on top of the immunoblot. (B) Bar chart showing the quantified production yield of hAQP4*, Opt-hAQP4 and Opt-hAQP4*, respectively, compared to hAQP4 and related to the SoPIP2;1 production, for which the production is set

to one. The y-axis represent average from triplicate cultures, error bars show the standard deviation (n=3).

This Figure is reproduced in colour in Molecular Membrane Biology online.

Table I. *High zeocin screen on* Pichia pastoris *transformants*

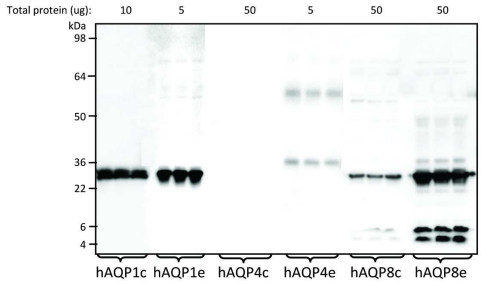
The largest, densest colonies from a total of 40 (Fig. 2) are listed after different incubation times at 30 °C.

Screen number and short description	Day	Largest/densest colonies[1]
Screen #1. Different amount of cells, different area	3	21, 11, 18, 37, 12
	7	11, 21, 18, 37, 04
Screen #2. Same amount of cells, different volume	2	11, 21, 04, 15, 18
	3	11, 21, 04, 15, 18
Screen #3. Same amount of cells, same volume	2	11, 21, 18, 37
	3	11, 21, 18, 37

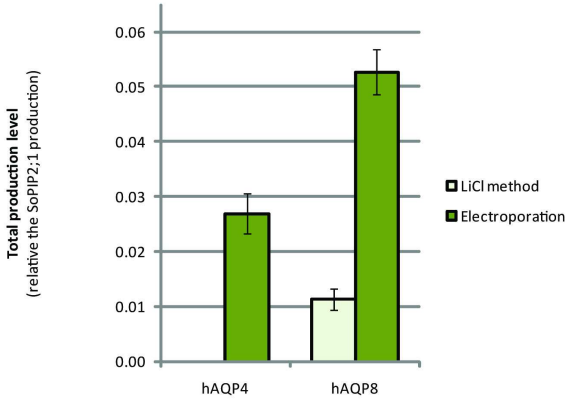
[1] Colonies are referred to by the number of the square where they are growing (Fig. 2).

Figure 1

A)

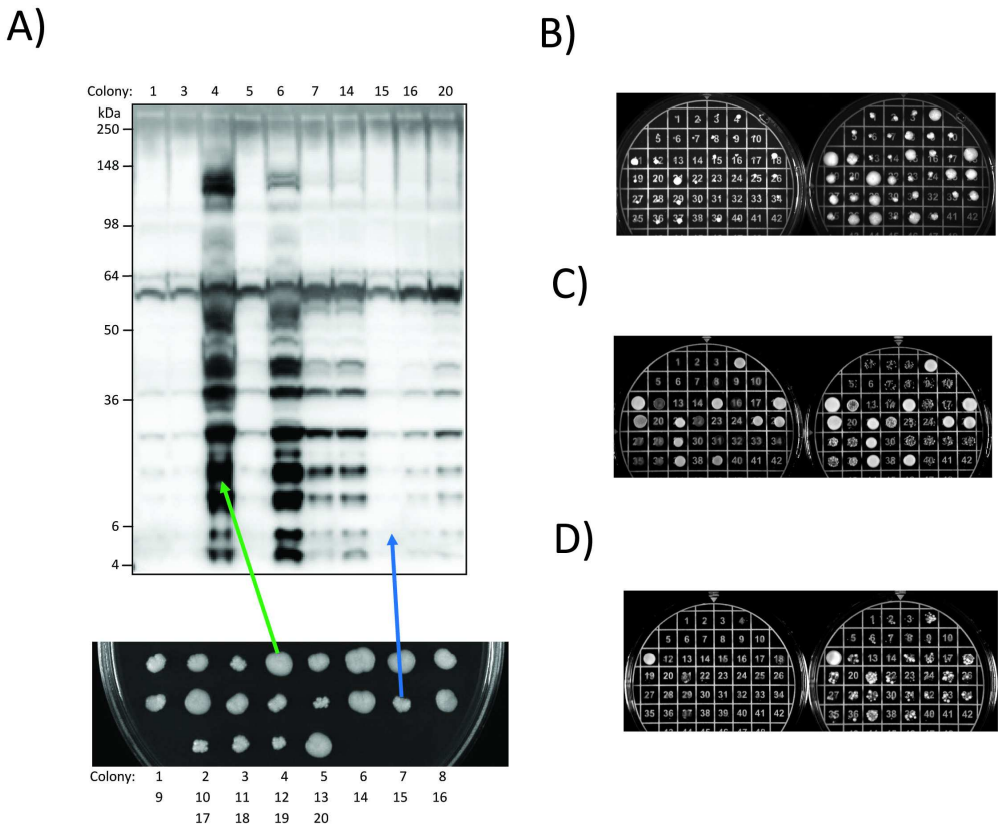


B)



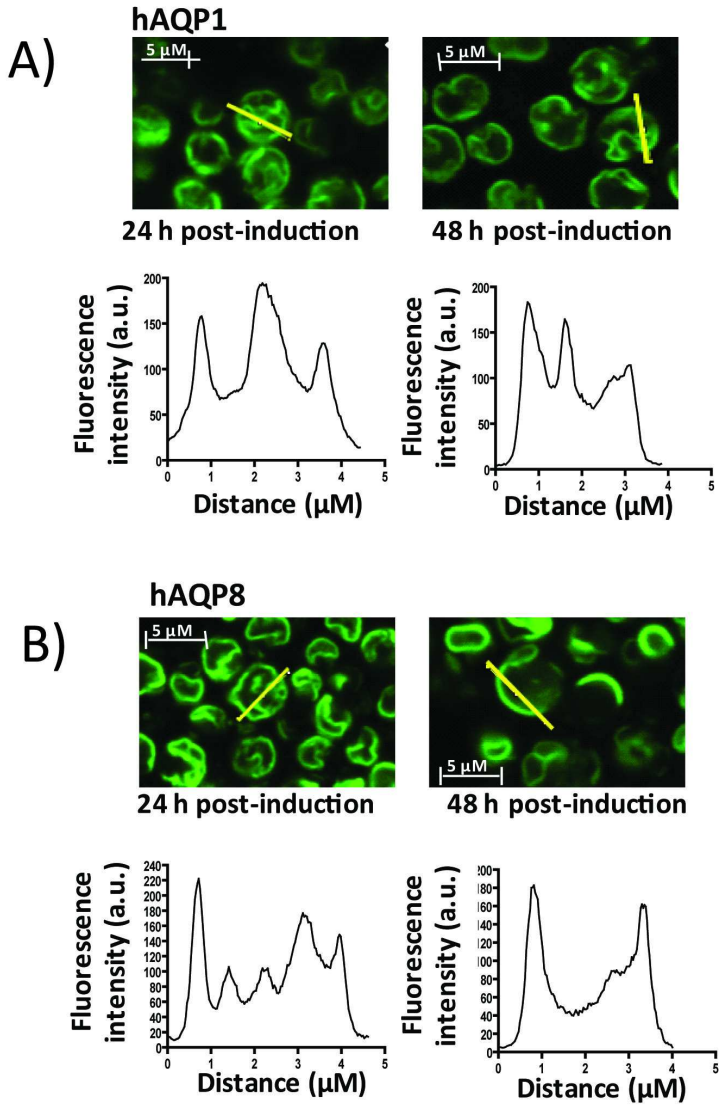
577x986mm (72 x 72 DPI)

Figure 2



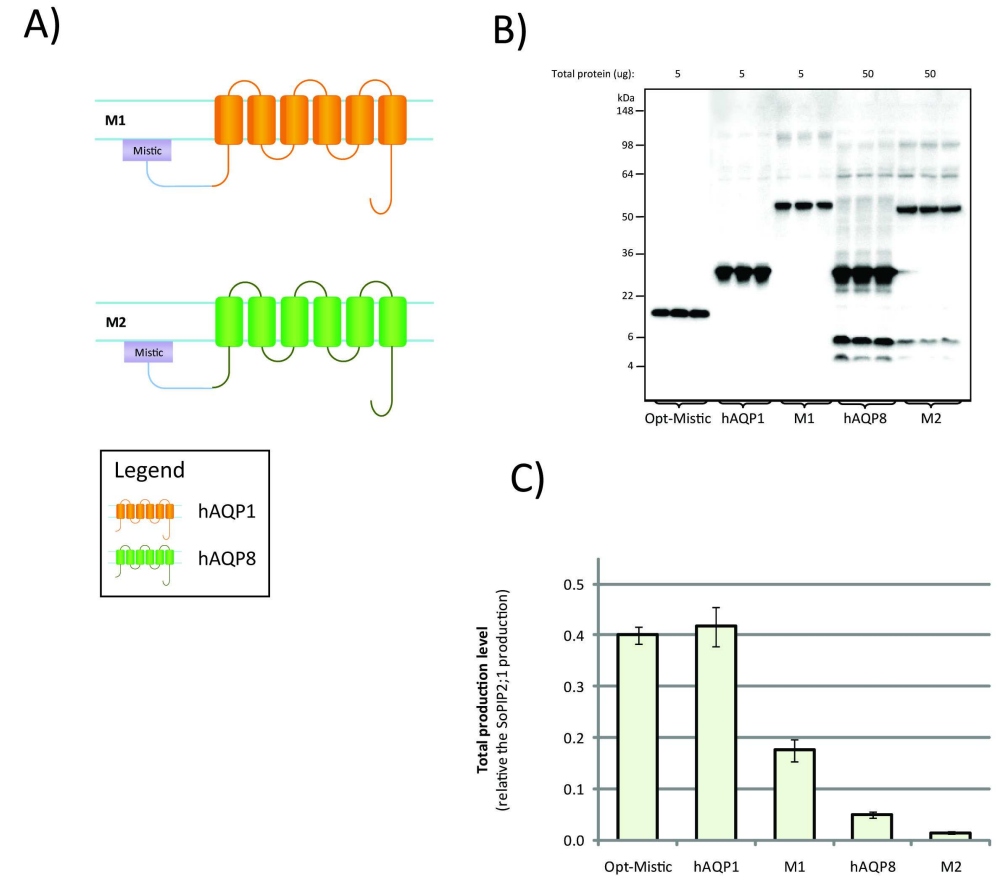
794x722mm (72 x 72 DPI)

Figure 3



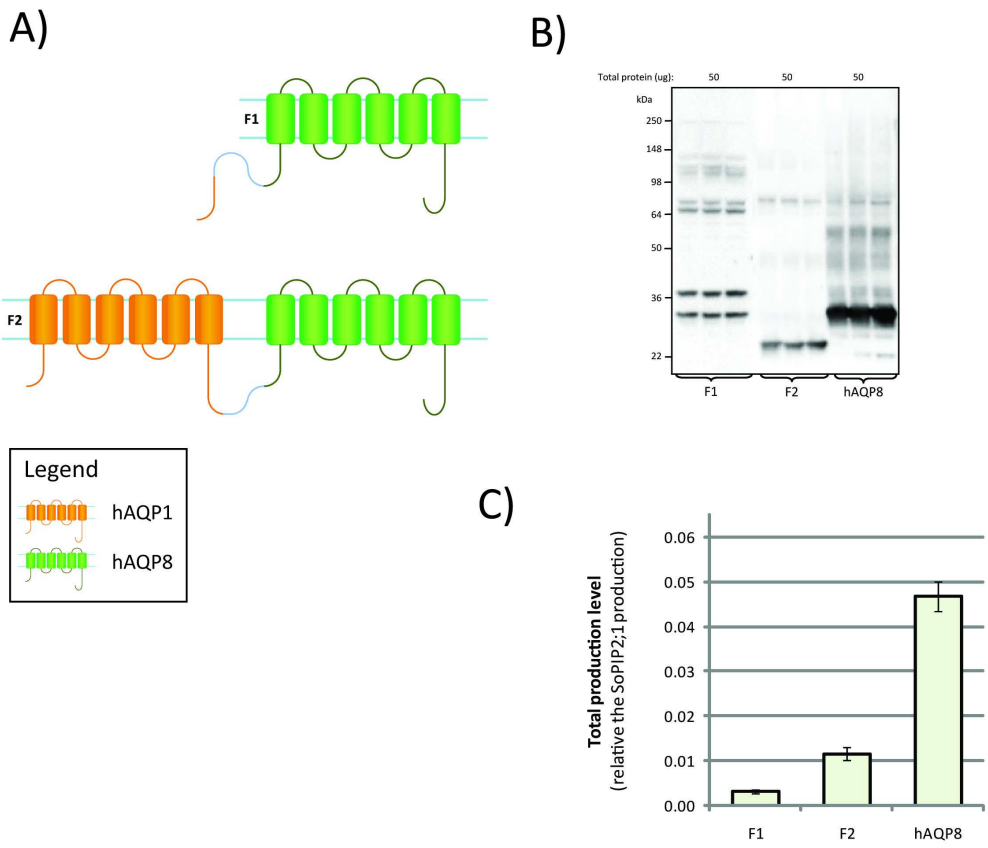
468x777mm (72 x 72 DPI)

Figure 4



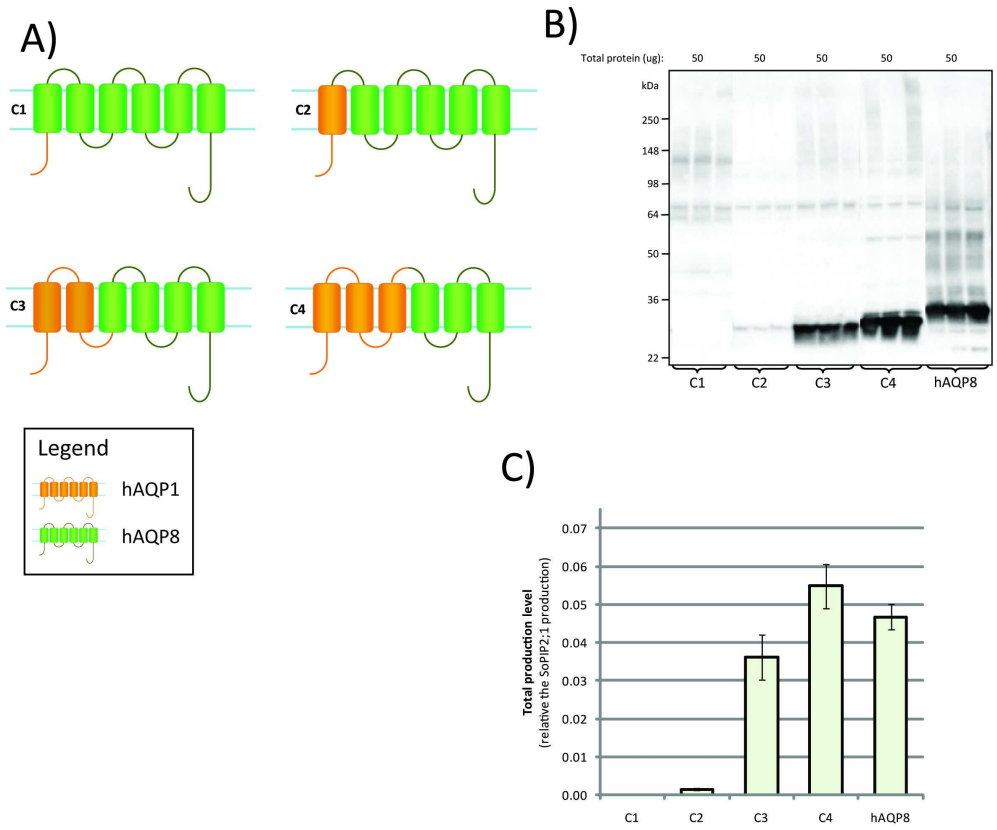
840x807mm (72 x 72 DPI)

Figure 5



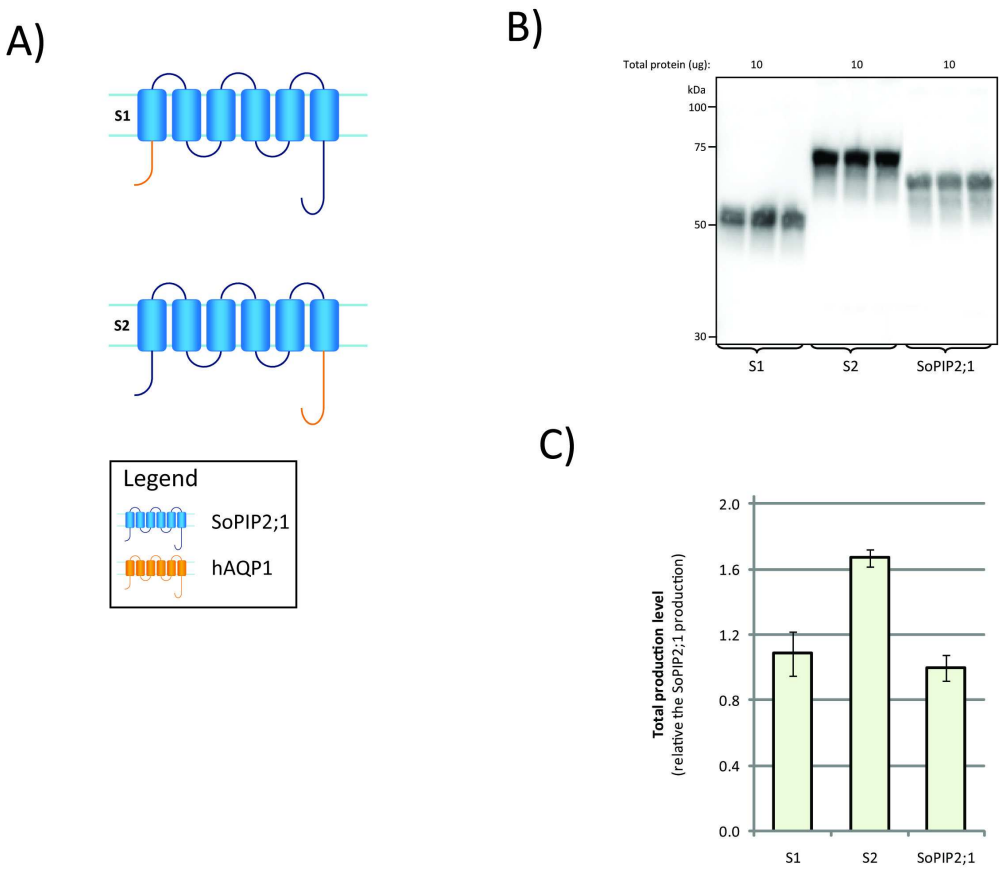
802x745mm (72 x 72 DPI)

Figure 6



839x754mm (72 x 72 DPI)

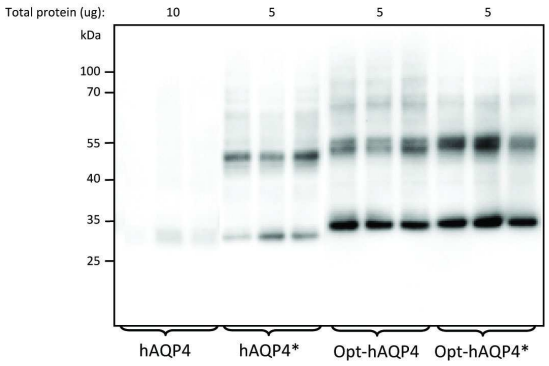
Figure 7



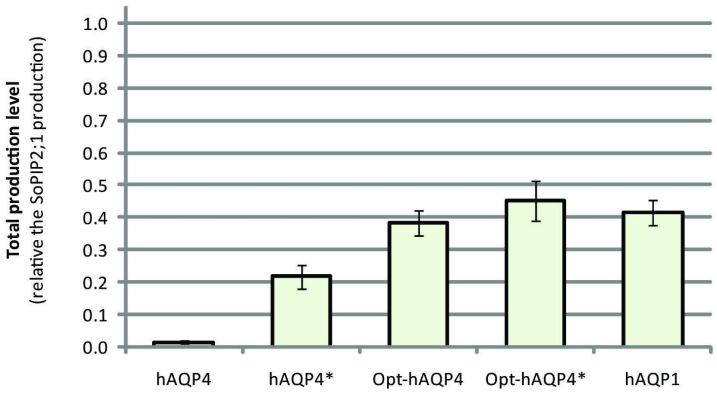
796x759mm (72 x 72 DPI)

Figure 8

A)



B)



696x804mm (72 x 72 DPI)

Table S1.

Primers and restriction enzymes used for the cloning of the constructs.

Construct name[1]	Forward Primer[2]	Reverse primer[3]
S1 $_{hAQP1\ -4,\ SoPIP2;1\ 32-304}$	GCGGAATTCAAAATGGCCAGCGAGCTCAAATTGTGGTC	CGCCATATGGCTCCTAAACATTGGGC[4]
S2$_{SoPIP2;1\ -260}$	AGCCATATGTGGTGTAGGA[4]	GCGTGGGGCCAGGATGAAGTCGTGGTATGCTGCTGC
S2$_{hAQP1\ 228-}$	GCAGTAGCAGCAGCATACCAGGACTTCATGCTGGCC	CGCTCTAGAGCTTTGGGCTTCATCTC[5]
C1$_{hAQP1\ -4,\ hAQP8\ 29-}$	GACGAATTCAAAATGGCCAGCGAGTGGTACGAACGGTTTGTG	CGAATTGCGGCCGCTCAGTGATGGTGATGGTGATGCCGAGCCTTCAGGATGAG
C2$_{hAQP1\ -45}$	GACGAATTCAAAATGGCCAGCGAGTTCAAG	CGCCGTCTGGTTGTTCCCCAC*
C2$_{hAQP8\ 61-}$	GACACTGGGCTGCTGCAGCCG*	CGAATTGCGGCCGCTCAGTGATGGTGATGGTGATGCCGAGCCTTCAGGATGAG
C3$_{hAQP1\ -76}$	GACGAATTCAAAATGGCCAGCGAGTTCAAG	GTTGAGGTGGGCGCCGCTGATG*
C3$_{hAQP8\ 92-}$	CCTGCGGTGTCCCTGGCAGCC*	CGAATTGCGGCCGCTCAGTGATGGTGATGGTGATGCCGAGCCTTCAGGATGAG
C4$_{hAQP1\ -128}$	GACGAATTCAAAATGGCCAGCGAGTTCAAG	GTCATTGCGGCCAAGCGAG*
C4$_{hAQP8\ 147-}$	GTCCAGGAGCAGGGGCAG*	CGAATTGCGGCCGCTCAGTGATGGTGATGGTGATGCCGAGCCTTCAGGATGAG
hAQP8$_{2-}$/pPICZB	GACGAATTCAAATCTGGAGAGATAGCCATG	CGACCTTCTAGATCAGTGATGGTGATGGTGATGCCGAGCCTTCAGGATG[5]
F1$_{hAQP1-4,\ hAQP8}$	GGCTTCGAAAAAATGGCCAGCGAGGATTACGATATCCCAACTACCGAAAACTTGTATTTTCAGGGTGAATTC GAC [6]	GTCGAATTCACCCTGAAAATACAAGTTTTCGGTAGTTGGGATATCGTAATCCTCGCTGGCCATTTTTTCGAA GCC [6]
F2$_{hAQP1,\ hAQP8}$	AGGTCGTCTAGATGATTACGATATCCCAACTACCGAAAACTTGTATTTTCAGGGTTCTGGAGAGATAGCCATG[5]	CGCATAGTCGACTCAGTGATGGTGATGGTGATGCCGAGCCTTCAGGATG [7]
Mistic/pPICZB	GACGAATTCAAAATGTTCTGTACTTTCTTC	CGACCTTCTAGAACCCTGAAAATACAAGTTTTCGGTAGTTGGGATATCGTAATCCTCTTTTTCACCCTCTTC[5]
Mistic-hAQP1	AGGTCGTCTAGAGCCAGCGAGTTCAAGAAG[5]	CGCATAGTCGACTCAGTGATGGTGATGGTGATGTTTGGGCTTCATCTCC[7]
Mistic-hAQP8	AGGTCGTCTAGATCTGGAGAGATAGCCATG[5]	CGCATAGTCGACTCAGTGATGGTGATGGTGATGCCGAGCCTTCAGGATG[7]
SGFP2/pPICZB	GTATCGTCTAGAATGGTGAGCAAGGGCGAGGAG[5]	CGACTTTCTAGATCACTTGTACAGCTCGTCCATGCC
hAQP1-SGFP2	GGCTTCGAAAAAATGGCTAGCGAGTTCAAG[8]	CTTGGTACCGTGATGGTGATGGTGATGTTTGG[9]
hAQP8-SGFP2	GGCTTCGAAAAATGTCTGGAGAGATAGCC[8]	CTTGGTACCGTGATGGTGATGGTGATGCCGAG[9]

[1]The constructs are named as shown in Fig. 3-6 Proteins partially amplified are written in subscript followed by numbers indicating what amino acids this part corresponds to.

[2]Forward primers contain a restriction site for *EcoR*I (underlined) if not stated otherwise. For primers containing two restriction sites, both have been stated in the order they occur. Primers with a phosphorylated 5' end has been marked with a star (*).

[3]Reverse primers contain a restriction site for *Not*I (underlined) if not stated otherwise. Primers are written 5' to 3'. For primers containing two restriction sites, both have been stated in the order they occur. Primers with a phosphorylated 5' end has been marked with a star (*).

[4]This primer has a restriction site for *Nde*I (underlined).

[5]This primer has a restriction site for *Xba*I (underlined).

[6]This primer has a restriction site for *Sfu*I and *EcoR*I (both underlined)

[7]This primer has a restriction site for *Sal*I (underlined).

[8]This primer has a restriction site for *Sfu*I (underlined). [9]This primer has a restriction site for *Kpn*I (underlined).

Figure S1

Linker sequences used in various constructs:

Linker for M1 and M2 (16 aa):

```
==========spacer========== ------------TEV------------ :XbaI::
GAT TAC GAT ATC CCA ACT ACC GAA AAC TTG TAT TTT CAG GGT TCT AGA
  D   Y   D   I   P   T   T   E   N   L   Y   F   Q   G   S   R
```

Linker for GFP tagged constructs (12 aa):

```
====================spacer====================
GGT ACC TCG AGC CGC GGC GGC CGC CAG CTT TCT AGA
  G   T   S   S   R   G   G   R   Q   L   S   R
```

Linker for F1 (16 aa):

```
==========spacer========== ------------TEV------------ :EcoRI:
GAT TAC GAT ATC CCA ACT ACC GAA AAC TTG TAT TTT CAG GGT GAA TTC
  D   Y   D   I   P   T   T   E   N   L   Y   F   Q   G   E   F
```

Linker for F2 (17 aa):

```
  ::XbaI::  ==========spacer========== ------------TEV------------
GCT CTA GAT GAT TAC GAT ATC CCA ACT ACC GAA AAC TTG TAT TTT CAG GGT
  A   L   D   D   Y   D   I   P   T   T   E   N   L   Y   F   Q   G
```