Understanding the Yeast Host Cell Response to Recombinant Membrane Protein Production

Zharain Bawa, Charlotte E Bland, Nicklas Bonander, Nagamani Bora, Stephanie P Cartwright, Michelle Clare, Matthew T Conner, Richard AJ Darby, Marvin V Dilworth, William J Holmes, Mohammed Jamshad, Sarah J Routledge, Stephane R Gross and Roslyn M Bill[1]

School of Life and Health Sciences, Aston University, Birmingham, UK

[1]To whom correspondence should be addressed (email r.m.bill@aston.ac.uk)

Short (page heading) title: Recombinant membrane protein production in yeast

Abstract

Membrane proteins are drug targets for a wide range of diseases. Having access to appropriate samples for further research underpins the pharmaceutical industry's strategy for developing new drugs. This is typically achieved by synthesizing a protein of interest in host cells that can be cultured on a large scale, allowing the isolation of the pure protein in quantities much higher than those found in the protein's native source. Yeast is a popular host as it is a eukaryote with similar synthetic machinery to the native human source cells of many proteins of interest,

whilst also being quick, easy and cheap to grow and process. Even in these cells the production of human membrane proteins can be plagued by low functional yields: we wish to understand why. We have identified molecular mechanisms and culture parameters underpinning high yields and have consolidated our findings to engineer improved yeast host strains. By relieving the bottlenecks to recombinant membrane protein production in yeast, we aim to contribute to the drug discovery pipeline, whilst providing insight into translational processes.

Membrane proteins as targets for recombinant protein production

Membrane proteins are central to many cellular processes: they are involved in the uptake and export of diverse charged and uncharged molecules, as well as mediating the interaction of cells with their environment. As a consequence they are of prime importance as drug targets to pharmaceutical companies. In order to structurally or functionally characterize a given member of this important class of protein, a sufficiently stable, active sample is required. In practice this means the requirement for a regular supply of milligram quantities of purified membrane protein [1].

Few membrane proteins have evolved to be naturally abundant in their native membranes with notable exceptions including mammalian and bacterial rhodopsins, aquaporins and complexes involved in respiration and photosynthesis. Inevitably, these proteins were amongst the first to have their structures solved: the first high-resolution crystallographic structure of a membrane protein (that of the photosynthetic reaction centre from *Blastochloris viridis* isolated from natural sources) was published in 1985 [2]. The fact that it remained the sole unique structure of a membrane protein for the next decade is indicative of the inherent challenges of working with this class of protein. The first such challenge is that

recombinant production is required in order to secure the hundreds-of-milligram quantities necessary to complete a successful structural biology project [3].

In 1998, the first structures derived from recombinant membrane proteins were deposited. These were of the prokaryotic proteins MscL [4] and KcsA [5], both produced in *Escherichia coli*. Thereafter, the elucidation of unique structures derived from recombinant sources saw an exponential-like growth rate, with 30 deposited in 2010 [1]. In contrast, the number of unique structures from native sources reached a plateau of around 10 per year in 2002 [1]. Of particular note are eukaryotic membrane protein structures, which represent the most challenging targets: they account for fewer than 20 % of all the unique membrane protein structures deposited to date. The yeasts *Pichia pastoris* and *Saccharomyces cerevisiae* have made an important contribution to this count as they have been used to generate more than half of the recombinant eukaryotic membrane protein samples leading to structures, with the remainder coming from insect or mammalian cells [1]. Specific examples derived from recombinant yeasts include the aquaporins, SoPiP2;1 [6] and AQP5 [7], the potassium channel, Kv1.2 [8], and the calcium ATPase, SERCA1a [9].

Strategies for producing recombinant membrane proteins

The process of identifying a membrane protein target of interest to solving its structure at high resolution is complex, involving several potential bottlenecks [1]. The development of validated protocols and technologies for recombinant membrane protein production that are less reliant on traditional trial and error approaches is therefore key to progress in this area.

Unfortunately, producing recombinant membrane proteins can still seem like something of an art-form. There is no guarantee of success when using a previously-successful experimental set-up, even for targets that are highly homologous to those

that worked well in those earlier trials [10]. Suitable production hosts for this purpose include prokaryotic microbes such as bacteria, eukaryotic microbes such as yeast, higher eukaryotes such as insect and mammalian cells and *in vitro* systems. All have their unique advantages and disadvantages [11]. We and others take the view that yeast provides a good compromise, benefitting from the speed, ease of use and low cost of goods of a microbe whilst having the benefits of a eukaryote. As we have already seen, this position is strengthened substantially by its increasingly important contribution to the structural biology of membrane proteins [1].

Our own work in this area was prompted by an inability to produce Fps1, a member of the aquaporin family, despite the fact that we could routinely produce orthologs such as GlpF [10]. We and others have therefore sought to understand the scientific principles that underpin how recombinant host cells can be used optimally for making membrane proteins. Some groups have chosen to tackle this problem by producing between 40-100 orthologs and then focusing on the protein that gives the best yields [12]. This approach has led to structures, but it is labor intensive and by its nature does not necessarily yield the specific protein of interest. Our focus has therefore been to understand yeast cells in order to make them more amenable to producing the target of choice [13]. The strategy has been to increase the total volumetric yield of a production experiment, in particular by increasing the yield per cell or "specific productivity".

Improving volumetric yields in *S. cerevisiae* by metabolic engineering

One of the key features of the respiratory yeast species, *P. pastoris*, is that it can be grown to very high cell densities. Optical densities in the hundreds are achievable, which is a straightforward way of increasing volumetric yield. Conferring this property on respiro-fermentative *S. cerevisiae*, which sacrifices its biomass yield when producing ethanol, would be desirable as its genetics and molecular biology are

much better established than those of *P. pastoris*. This strategy was therefore pursued by constructing a fully respiratory strain of *S. cerevisiae*: Otterstedt and colleagues replaced the endogenous hexose transporters in yeast with a single chimeric transporter comprised of the first 6 transmembrane domains of the low affinity Hxt1 transporter and the last 6 transmembrane domains of the high affinity Hxt7 transporter [14]. The resultant TM6* strain was found not to exhibit the respiro-fermentative behavior of a wild-type yeast, but rather a mono-phasic metabolism consistent with being respiratory. This is most likely due to a restricted glucose consumption and hence a reduced glycolytic rate [14]. However, the important feature in this context is that TM6* was able to produce more protein than a wild-type strain by achieving a higher volumetric yield, albeit in the knowledge that its specific growth rate is about 67% that of the wild-type parent [15].

In the TM6* strain, we found that the yield per cell of Fps1 was the same as that for two wild-type strains [15]. The increase in volumetric yield occurred, as anticipated, because the TM6* strain produces about 2-3 times more biomass (determined by OD_{595} measurements) in culture than a wild-type strain. When it was used to produce two G protein-coupled receptors (GPCRs), the A_{2a} adenosine receptor ($A_{2a}R$) and the cannabinoid receptor 2, the B_{max} and K_d values obtained were comparable with those from a wild-type strain. The increased volumetric yield was due to the fact that there was a doubling of OD_{595} and, more relevantly, the total membrane protein yield in the TM6* strain (determined by protein assay) was a factor of 4 higher than in the wild-type strain [15].

Improving yields per cell of *S. cerevisiae* by strain engineering

Improving the biomass yield of a *S. cerevisiae* strain that also has an increased yield per cell could provide a route to maximal yields. We therefore looked at the global host cell response to producing a membrane protein [13]. Using arrays, we examined

two different growth conditions that both led to relatively low protein yields (35ºC, pH 5 and 35ºC, pH7) compared with normal conditions (30ºC, pH 5) and looked for changes in mRNAs that occurred in the same direction in both sets. This highlighted 39 genes, which were validated against the list obtained when comparing the single growth condition that led to higher protein yields (20ºC, pH 5) with normal conditions (30ºC, pH 5). In all but one case, genes that were down-regulated under low yielding conditions were up-regulated under high-yielding conditions and vice versa [16].

In order to determine if these genes influenced the yield per cell, we first looked at all non-essential genes that were down-regulated under high-yielding conditions in a screen of deletion strains. This revealed, perhaps surprisingly, that in some cases the deletion of a single gene could increase the yield of our test protein compared to the wild-type strain, as seen for SPT3, SRB5 and GCN5 (Figure 1). The three deleted genes are known to be components of the transcriptional SAGA (for GCN5 and SPT3) and mediator (for SRB5) complexes. SAGA may have a role in the transcription of stress-induced genes in order to balance inducible stress responses with the steady output of housekeeping genes [17]. The mediator complex appears to be required for all transcriptional events and transmits regulatory signals from transcription factors to RNA polymerase II [18]. We found that improved yields of Fps1 of up to a factor of 9 over the corresponding wild-type control were not explained by changes in promoter activity or FPS1 transcript number. This suggested that the improvements we observed were not due to changes at the transcriptional level, but that post-transcriptional events might regulate the production of recombinant Fps1 [16].

Next, we looked at the over-expression of genes that were up-regulated under high-yielding conditions by using the doxycycline-repressible tetO system. This revealed an additional 3 candidate genes that might have a role in increasing the yield per cell

of our membrane protein (Figure 2). Most notable amongst them was *BMS1*, which is involved in ribosome biogenesis. Using qRT-PCR, we noted a clear relationship between Fps1 yield and an increase in *BMS1* transcript number compared to wild-type in all our high-yielding host strains. In particular, we were able to tune the expression of *BMS1* to maximize yields. When the promoter was fully on (with no doxycycline added) and the expression of *BMS1* was fully induced we saw relatively low expression of our target protein. This was also the case when the promoter was fully repressed (at >10µg/mL doxycycline). High-yielding strains and conditions were found to be correlated with 0.5-0.7 copies of *BMS1* mRNA per cell, with yield improvements of up to 70 times. The use of this strain also doubled the yield of the GPCR, $A_{2a}R$, after tuning with a different doxycycline concentration of 10µg/mL [16].

Under high-yielding conditions, the cells grew more slowly and efficiently (indicated by a lower heat output rate as measured by on-line flow microcalorimetry) than under low-yielding or control conditions [16]. This is in agreement with the fact that these cells consume glucose and make ethanol more slowly, which may enable them to accommodate the increased metabolic load of producing higher yields of protein. This phenotype is also seen for our 3 high-yielding deletion strains. These strains also show differences in their translation characteristics and polysome profiles: they have significant accumulation of the polysomes into 80S peaks, strongly hinting at a block in initiation. This indicates changes in their translational properties compared to wild-type cells and links to an over-arching role for *BMS1*. Selective translation of specific proteins is well documented following stress when ribosome functions are usually stalled at the initiation stage and where the GCN2/GCN4 complex plays a pivotal role [19-21]. It would therefore be interesting to establish whether such a regulatory complex is involved in the direct regulation of membrane protein production.

Improving yields per cell of *P. pastoris* and *S. cerevisiae* by bioprocess control

In addition to genetic engineering, controlling culture parameters in both *P. pastoris* and *S. cerevisiae* has the potential to improve the yield per cell. Our initial work in *P. pastoris* was done on secreted green fluorescent protein, which is commonly used to tag membrane proteins [22]. Since we wanted to find a simple method of identifying and scaling up promising production conditions (thereby minimizing trial and error) we adopted a design-of-experiments (DoE) approach, which is often used in industry, to examine the effect of three input parameters on the product yield. The DoE approach has several advantages since it reduces the number of experiments that need to be performed to examine the combined effect of the input parameters, allows any interactions between them to be interrogated and results in a predictive equation.

We used a Box-Behnken model for our DoE, with three input parameters (temperature, pH and dissolved oxygen) set at three levels. This statistical design requires only a sub-set of all possible combinations to be used to identify the optimal production condition; in this case, 13 experiments out of a possible 27. Our read-outs were the optical density (OD_{595}) of the culture and its fluorescence. We therefore reported the yield per cell as the fluorescence per OD_{595}. Experiments were performed in a parallel mini-bioreactor system that can individually control pH, temperature and dissolved oxygen in each of its 24 vessels. This approach allowed us to develop a model that predicted the optimal yielding conditions for our culture (Figure 3), which we validated by comparing predicted versus experimental yields for additional combinations of factors not used to construct the model. We further demonstrated that the model was scalable from 6mL to 3L, although the relationship here suggested that it was under-predicting the yield. This led us to explore the induction regime, which is a critical component of a *P. pastoris* production experiment [22].

We noted that induction with 100 % methanol did not result in a sustained production, and that after 15 hours the total yield had typically reached a plateau. We therefore examined a mixed induction of 60 % sorbitol and 40 % methanol, and saw that production was sustained, and in the case of our DoE-optimized conditions, the yield was dramatically improved. Gas chromatographic analysis of the residual methanol concentration showed that the best yielding conditions had residual methanol concentrations in the region of 3g/L. At concentrations below this, yields were lower. This optimized feeding regime increased the yield per cell by a factor of about 6 over standard conditions.

We have recently applied the DoE approach to the production of a GPCR in three strains of *S. cerevisiae* using three culture parameters as before. In addition the design includes the absence or presence of additives that have been shown to improve the functional yields of some GPCRs. As predicted, the input parameters did not significantly affect the growth of any of the three strains. However, they had a clear effect on binding activity in the membranes, with each strain giving its highest values under different culture conditions. The complete data set will be used to build a predictive model, which will then be validated at scale.

Conclusions

Yeast is a flexible recombinant host for eukaryotic membrane protein production. The conditions under which yeast cells are grown and harvested are critical in maximizing the total volumetric yield of functional protein. Moreover, differences in yields can be related to the differential expression of genes involved in transcription and translation and these genes have provided the basis for developing new production strains that can improve yields per cell up to 70-fold. We have used these strategies to produce several membrane proteins in yields suitable for further study. Examples include

three 4 transmembrane domain proteins (CD81, CD82 and CLDN1), our initial target, Fps1, and the GPCR, $A_{2a}R$.

Acknowledgements

References

1   Bill, R. M., Henderson, P. J. F., Iwata, S., Kunji, E. R. S., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C. G. and Vogel, H. (2011) Widening bottlenecks in membrane protein structure pipelines. Nature Biotechnol, in press

2   Deisenhofer, J., Epp, O., Sinning, I. and Michel, H. (1995) Crystallographic refinement at 2.3 A resolution and refined model of the photosynthetic reaction centre from Rhodopseudomonas viridis. J Mol Biol. **246**, 429-457

3   Bonander, N. and Bill, R. M. (2009) Relieving the first bottleneck in the drug discovery pipeline: using array technologies to rationalize membrane protein production. Expert Rev Proteomics. **6**, 501-505

4   Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. and Rees, D. C. (1998) Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. Science. **282**, 2220-2226

5   Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science. **280**, 69-77

6   Tornroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R. and Kjellbom, P. (2006) Structural mechanism of plant aquaporin gating. Nature. **439**, 688-694

7   Horsefield, R., Norden, K., Fellert, M., Backmark, A., Tornroth-Horsefield, S., Terwisscha van Scheltinga, A. C., Kvassman, J., Kjellbom, P., Johanson, U. and Neutze, R. (2008) High-resolution x-ray structure of human aquaporin 5. Proc Natl Acad Sci U S A. **105**, 13327-13332

8   Long, S. B., Campbell, E. B. and Mackinnon, R. (2005) Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. Science. **309**, 897-903

9   Jidenko, M., Nielsen, R. C., Sorensen, T. L., Moller, J. V., le Maire, M., Nissen, P. and Jaxel, C. (2005) Crystallization of a mammalian membrane protein overexpressed in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA. **102**, 11687-11691

10   Hedfalk, K., Bill, R. M., Hohmann, S. and Rydström, J. (2000) Overexpression and purification of the glycerol transport facilitators, Fps1 and GlpF, in *Saccharomyces cerevisiae* and *Escherichia coli*. In Molecular Biology and Physiology of Water and solute Transport (Hohman, S. and Nielsen, S., eds.). pp. 29-34, Kluwer Academic/ Plenum Publishers, Amsterdam

11   Grisshammer, R. (2006) Understanding recombinant expression of membrane proteins. Curr Opin Biotechnol. **17**, 337-340

12   Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T. and MacKinnon, R. (2003) X-ray structure of a voltage-dependent K+ channel. Nature. **423**, 33-41

13   Bonander, N., Hedfalk, K., Larsson, C., Mostad, P., Chang, C., Gustafsson, L. and Bill, R. M. (2005) Design of improved membrane protein production experiments: Quantitation of the host response. Protein Sci. **14**, 1729-1740

14   Otterstedt, K., Larsson, C., Bill, R. M., Stahlberg, A., Boles, E., Hohmann, S. and Gustafsson, L. (2004) Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. EMBO Rep. **5**, 532-537

15   Ferndahl, C., Bonander, N., Logez, C., Wagner, R., Gustafsson, L., Larsson, C., Hedfalk, K., Darby, R. A. and Bill, R. M. (2010) Increasing cell biomass in *Saccharomyces cerevisiae* increases recombinant protein yield: the use of a respiratory strain as a microbial cell factory. Microb Cell Fact. **9**, 47

16   Bonander, N., Darby, R. A., Grgic, L., Bora, N., Wen, J., Brogna, S., Poyner, D. R., O'Neill, M. A. and Bill, R. M. (2009) Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. Microb Cell Fact. **8**, 10

17   Huisinga, K. L. and Pugh, B. F. (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Mol Cell. **13**, 573-585

18   Kornberg, R. D. (2005) Mediator and the mechanism of transcriptional activation. Trends Biochem Sci. **30**, 235-239

19   Yang, R., Wek, S. A. and Wek, R. C. (2000) Glucose limitation induces GCN4 translation by activation of Gcn2 protein kinase. Mol Cell Biol. **20**, 2706-2717

20   Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G. and Marton, M. J. (2001) Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. Mol Cell Biol. **21**, 4347-4368

21   Valenzuela, L., Aranda, C. and Gonzalez, A. (2001) TOR modulates GCN4-dependent expression of genes turned on by nitrogen limitation. J Bacteriol. **183**, 2331-2334

22   Holmes, W. J., Darby, R. A., Wilks, M. D., Smith, R. and Bill, R. M. (2009) Developing a scalable model of recombinant protein yield from *Pichia pastoris*: the influence of culture conditions, biomass and induction regime. Microb Cell Fact. **8**, 35

Figure legends

Figure 1: The deletion of a single gene can increase the yield of a recombinant membrane protein. The yield of a test protein, Fps1, was analyzed from shake flask cultures of *S. cerevisiae* deletion strains using an anti-HA-tag immunoblot as the read-out. Highlighted from the 10 strains tested in this experiment (1 per lane) are the yields from the wild-type parent strain and the three deletion strains, *srb5Δ*, *spt3Δ* and *gcn5Δ*, which clearly show yield improvements.

Figure 2: A selection of deletion and overexpression strains exhibiting improved yield characteristics for recombinant Fps1. Fps1 yields, as assessed by immunoblot, are reported relative to wild-type (black bars). Lower case letters denote deletion strains while upper case letters denote strains with an up-regulated expression of that particular gene.

Figure 3: A response surface contour plot showing how yield per cell changes with each of the input factors. T = temperature (°C), pH = pH and DO = dissolved oxygen tension (%). All hold values are the "0" mid-point values in the DoE matrix.

Figure 1

Fps1 —

Figure 2

Figure 3