EXPERT REVIEWS

Relieving the first bottleneck in the drug discovery pipeline: using array technologies to rationalize membrane protein production

Expert Rev. Proteomics 6(5), 501–505 (2009)

Nicklas Bonander and Roslyn M Bill†

†*Author for correspondence*
School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK
Tel.: +44 121 204 4274
Fax: +44 121 204 4175
r.m.bill@aston.ac.uk

The slow down in the drug discovery pipeline is, in part, owing to a lack of structural and functional information available for new drug targets. Membrane proteins, the targets of well over 50% of marketed pharmaceuticals, present a particular challenge. As they are not naturally abundant, they must be produced recombinantly for the structural biology that is a prerequisite to structure-based drug design. Unfortunately, however, obtaining high yields of functional, recombinant membrane proteins remains a major bottleneck in contemporary bioscience. While repeated rounds of trial-and-error optimization have not (and cannot) reveal mechanistic details of the biology of recombinant protein production, examination of the host response has provided new insights. To this end, we published an early transcriptome analysis that identified genes implicated in high-yielding yeast cell factories, which has enabled the engineering of improved production strains. These advances offer hope that the bottleneck of membrane protein production can be relieved rationally.

Keywords: drug discovery • membrane protein • recombinant protein production • transcriptome • yeast

Bottleneck of recombinant membrane protein production

Understanding membrane proteins is vital for the development of new drugs in the fight against human disease [1]. As they are not highly abundant naturally, membrane proteins must be produced recombinantly for the detailed studies that will reveal their biochemical, functional and structural characteristics. Despite this need, obtaining high yields of functional, recombinant membrane protein remains a major bottleneck in contemporary bioscience [2]. We have shown that the root of the problem is understanding the host organism [3] and, in particular, its response to the production of recombinant proteins in its cell membranes. Our rational approach to this problem is in stark contrast to most protein production strategies that rely on repeated rounds of trial-and-error optimization and cannot provide a mechanistic insight, which is also true of approaches that rely on the mutation of the protein target to improve its production yields [4].

The use of postgenomic array methods have been key to this rational approach to optimization [3] in yeast [5]; in one such example, we reported 39 host cell genes whose expression was significantly altered when high-yielding production conditions were compared with low-yielding conditions in *Saccharomyces cerevisiae*. Although similar studies have subsequently been performed in other hosts [6,7], mechanistic insight into successful recombinant protein production has remained elusive [8]. Building on our comparative transcriptome analysis [3], we recently showed that deletion of three *S. cerevisiae* genes could each increase protein yields. In particular, overexpression of a fourth gene from our list, *BMS1*, could be specifically tuned to maximize yields of a range of membrane proteins [9]. By altering the amount of *BMS1* transcript, the metabolism of high-yielding cultures was changed substantially and coincided with the ratio of ribosomal subunits being perturbed, offering, for the first time, an insight into the actual mechanisms involved.

 10.1586/EPR.09.65 ISSN 1478-9450

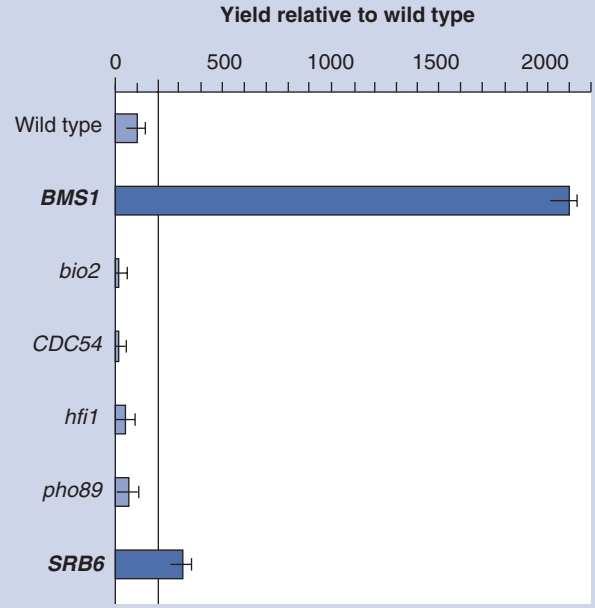

Figure 1. Analysis of yields of Fps1 to validate the array data in Table 1. Fps1 yields, as assessed by immunoblot, are reported relative to wild type. Lower-case letters denote deletion strains, while upper-case letters denote strains with an upregulated expression of that particular gene. *BMS1* and *SRB6* are in bold as the yields obtained for these strains in a production screen exceeded the threshold (more than twice that of wild type) required for further study.

Designing appropriate experiments to address the bottleneck

The predominant technique for obtaining a description of a global cellular response to a genetic or environmental change is to use DNA microarrays. Modern arrays comprise amplified cDNA or oligonucleotides (typically ranging from 25 to 60 nucleotides in length), which are either spotted or directly synthesized onto a solid support in various layouts. Widely used microarray chips are marketed by Agilent Technologies Incorporated [101] and Affymetrix Incorporated [102], although other platforms, as well as home-made options, exist. An article from 2006 in *Expert Review of Molecular Diagnostics*, provides useful pointers in the selection of appropriate microarray platforms and the principles of experimental design [10].

We previously used nylon filter arrays generated in Louise Showe's laboratory [103] in order to understand the yeast response to stresses, such as slow glucose uptake rates [11] and membrane protein production [3]. In each case, our aim has been to achieve a list of all genes that were potentially correlated with these phenotypes, but which was not too large to be subsequently validated by independent analysis, such as quantative (Q)-PCR or strain engineering.

Typically, a pair-wise comparison of RNA preparations in two to six biological repeats is performed for a reference and several specific test conditions. Obtaining a manageable number of genes for validation and testing that are correlated with significantly changed mRNA levels requires a robust experimental design. This can be achieved by sampling cells growing at the same growth rate but over a range of optical densities (ODs; e.g., 0.5, 1.0 and 1.5) rather than multiple sampling at a single OD. In addition, introducing further 'redundancy' into the test conditions can filter out nonspecific effects. For example, we examined two different growth conditions that both led to relatively low protein yields compared with normal conditions, and looked for changes in mRNAs that occurred in the same direction in both sets. This highlighted 39 genes. It is notable that other array studies of *S. cerevisiae*, albeit not focused on protein production, typically present many more genes as being significantly regulated. For example, a shift from 27 to 37°C for 30–80 min yielded 447 genes [12], while there were 150 upregulated genes following a shift to pH 7.4 [13]. A shift from 30 to 10°C for 8 h induced a change in 1024 genes [14].

The 39 genes that we found were validated against the list obtained when comparing the single growth condition that led to higher protein yields compared with normal conditions. Figure 1 shows an example of the data obtained. In all cases, genes that were downregulated under low-yielding conditions were upregulated under high-yielding conditions and *vice versa* (Table 1). This lent confidence that these genes had a role in high-yielding protein production, which was subsequently further validated by strain engineering (Figure 1). Specifically, we used this approach to characterize the role of *BMS1* [9] and to engineer the first yeast production strain (GB0813253.2) by rational design [8].

In another study, we examined the ability of a respiratory yeast strain, V5.TM6*P, to grow on glucose compared to its respiro-fermentative parent, V5. We were able to improve the analysis even further by comparing all previous analyses of the yeast diauxic shift, which had been first studied by microarray in 1997 [15]. Again, we compared the two strains (n = 12) over a range of sampling points (36–5 g/l glucose), rather than taking multiple samples at a single point [11]. Our dataset gives a remarkably complete view of the involvement of genes in the tricarboxylic acid cycle, glyoxylate cycle and respiratory chain in the expression of the phenotype of V5.TM6*P. Furthermore, 88% of the transcriptional response of the induced genes in our dataset could be related to the potential activities of just three proteins: Hap4, Cat8 and Mig1. Overall, the data supported genetic remodeling in V5.TM6*P consistent with a respiratory metabolism that is insensitive to external glucose concentrations.

Validation by Q-PCR & subsequent strain engineering

Validation of array results by Q-PCR or strain engineering is an essential next step. Our protein production strain, in which the ribosomal gene *BMS1* was upregulated by a factor of 6–7 compared with wild-type, gave corresponding yield improvements of a factor of 2–70 for a range of membrane proteins [9]. This rationally engineered strain complements strains resulting from previous speculative deletion strategies. In *S. cerevisiae*, for example, deletion of two vacuolar proteases increased yields of the G-protein-coupled receptors, Ste2, by a factor of 10 [16], while in *Pichia pastoris* strain SMD1163, the deletion of two

Table 1. Array data from a transcriptome analysis of yeast strains giving high or low yields of recombinant membrane protein.

Systematic name	Gene	30°C pH 5 to 35°C pH 7 (low yield I)	30°C pH 5 to 35°C pH 5 (low yield II)	30°C pH 5 to 20°C pH 5 (high yield)
Results for verified genes				
YBR253W	*SRB6*	↓ (0.3)	↓ (0.3)	↑ (2.0)
YBR288C	*APM3*	↓ (0.2)	↓ (0.3)	↑ (2.9)
YBR296C	*PHO89*	↓ (0.1)	↓ (0.04)	↑ (14.0)
YFL036W	*RPO41*	↓ (0.3)	↓ (0.3)	↑ (5.4)
YGL029W	*CGR1*	↓ (0.2)	↓ (0.3)	↑ (3.0)
YGR286C	*BIO2*	↑ (5.3)	↑ (5.5)	↓ (0.2)
YHR190W	*ERG9*	↑ (3.3)	↑ (3.0)	↓ (0.3)
YMR251W	*HOR7*	↓ (0.1)	↓ (0.2)	↑ (3.0)
YOR303W	*CPA1*	↓ (0.3)	↓ (0.3)	↑ (2.7)
YPL094C	*SEC62*	↓ (0.3)	↓ (0.2)	↑ (2.2)
YPL104W	*MSD1*	↓ (0.3)	↓ (0.3)	↑ (2.4)
YPL187W	*MFα1*	↑ (3.9)	↑ (3.7)	↓ (0.3)
YPL206C	*PGC1*	↓ (0.4)	↓ (0.4)	↑ (2.0)
YPL217C	*BMS1*	↓ (0.2)	↓ (0.2)	↑ (4.1)
YPL219W	*PCL8*	↓ (0.2)	↓ (0.2)	↑ (4.7)
YPL254W	*HFI1*	↓ (0.2)	↓ (0.1)	↑ (5.0)
YPR019W	*CDC54*	↓ (0.2)	↓ (0.2)	↑ (6.6)
Results for putative or dubious genes				
YBL112C	Putative	↓ (0.1)	↓ (0.2)	↑ (5.8)
YCR018C-A	Putative	↓ (0.3)	↓ (0.3)	↑ (2.8)
YDR444W	Putative	↓ (0.3)	↓ (0.2)	↑ (4.5)
YFL066C	Putative	↓ (0.3)	↓ (0.3)	↑ (4.5)
YLR149C	Putative	↓ (0.1)	↓ (0.2)	↑ (4.5)
YLR162W	Putative	↓ (0.1)	↓ (0.1)	↑ (6.0)
YOL098C	Putative	↓ (0.3)	↓ (0.2)	↑ (4.9)
YOR389W	Putative	↓ (0.2)	↓ (0.2)	↑ (4.9)
YPL216W	Putative	↓ (0.3)	↓ (0.2)	↑ (3.9)
YKR040C	Dubious	↓ (0.2)	↓ (0.2)	↑ (2.7)
YLR202C	Dubious	↓ (0.2)	↓ (0.2)	↑ (3.6)
YMR290W-A	Dubious	↓ (0.1)	↓ (0.2)	↑ (2.2)
YOR333C	Dubious	↓ (0.1)	↓ (0.1)	↑ (3.6)

In the original study, genes were tabulated if their expression was changed when comparing both of the low-yielding conditions (I and II) with normal-growth conditions: 30°C pH 5 to 35°C pH 7 and 30°C pH 5 to 35°C pH 5. Data for verified genes separately from that for putative or dubious genes are presented. The change is expressed as a factor (in parentheses), where that factor is x when a gene expressed with intensity 1 at 30 °C pH 5 is expressed with intensity x at the new condition. Arrows show whether the gene is up- or down-regulated. The results for the high-yielding conditions are shown for comparison. Since nine genes from the original study did not yield statistically significant data in the high-yielding condition, they are not listed. Remarkably, all genes that are downregulated under low-yielding conditions are upregulated under high-yielding conditions and *vice versa*.
Data from [3].

protease genes (*PEP4* and *PBR*) boosted membrane protein production [17]. Minimizing the unfolded protein response (UPR) of *S. cerevisiae* has also improved the yields of certain functional membrane protein [18] that induce the UPR.

Expert commentary

The production of recombinant membrane proteins is finally being recognized as a science rather than an art. Using systems biotechnology, it is now possible to identify changes in gene expression in the host cell that are associated with high yields of these potential drug targets. We have used the results of one such study to engineer high-yielding yeast strains and are now in a position to gain further mechanistic insights using postgenomic techniques.

Five-year view

We have shown that comparative transcriptome analysis of *S. cerevisiae* host cells during high- versus low-yielding recombinant protein production experiments can guide strain engineering. In particular, by titrating the overexpression of *BMS1*, a ribosome biogenesis gene, the functional yields of a range of membrane proteins can be improved by a factor of 2–70. This discovery is the basis of a recent patent application (GB0813253.2), which also describes three additional engineered strains. With the very recent publication of a curated *P. pastoris* genome [19], we will now be able to transfer this technology to a second industrially relevant yeast species and we further predict that this approach will be used for a range of host cells and protein targets. As array techniques have improved, it is now possible to relate changes in mRNA with changes at the protein level. A comparative proteome analysis of membrane versus soluble protein production in *E. coli* highlighted that the cytoplasmic membrane protein translocation machinery might be limiting in low-yielding bacterial transformants [6]. Interestingly, we found no evidence for this in yeast [9], which highlights clear differences between prokaryotic and eukaryotic cells. With time, the ability to combine the outputs from these types of studies with those from metabolomic approaches will contribute to a true systems biotechnological description of recombinant membrane protein production in yeast.

Financial & competing interests disclosure

The authors are supported by EPSRC CASE awards with Alpha Biologics Ltd and Applikon Biotechnology Ltd, a BBSRC CASE award with AstraZeneca Ltd and a Technology Strategy Board grant with Glycoform Ltd and Polytherics Ltd, as well as European Commission contracts LSHG-CT-2004–504601 (E-MeP), LSHG-CT-2006–037793 (OptiCryst) and Grant Agreement Number 201924 (EDICT) to RMB. The BBSRC supports bioreactors and a flow microcalorimeter in the RMB laboratory through an REI award. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Key issues

- There is a move toward structure-based drug design in modern discovery pipelines.
- Membrane proteins, especially G-protein-coupled receptors and ion channels, are the most important class of new drug targets.
- The recombinant production of membrane proteins is a challenging task.
- Array-based techniques are finally shedding light on the mechanisms underpinning high yields of recombinant proteins in yeast cells.
- Careful experimental design is critical to identifying genes whose expression is correlated with highly recombinant protein yields.
- These experiments are an efficient route to engineering new, high-yielding production strains.

References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest

1 Schlyer S, Horuk R. I want a new drug: G-protein-coupled receptors in drug development. *Drug Discov. Today* 11(11–12), 481–493 (2006).

2 Forstner M, Leder L, Mayr LM. Optimization of protein expression systems for modern drug discovery. *Expert Rev. Proteomics* 4(1), 67–78 (2007).

3 Bonander N, Hedfalk K, Larsson C *et al.* Design of improved membrane protein production experiments: quantitation of the host response. *Protein Science* 14(7), 1729–1740 (2005).

• **First report of a transcriptome analysis of recombinant membrane protein production in yeast.**

4 Martinez Molina D, Cornvik T, Eshaghi S *et al.* Engineering membrane protein overproduction in *Escherichia coli*. *Protein Sci.* 17(4), 673–680 (2008).

5 Bill RM. Yeast – a panacea for the structure-function analysis of membrane proteins? *Curr. Genet.* 40(3), 157–171 (2001).

6 Wagner S, Baars L, Ytterberg AJ *et al.* Consequences of membrane protein overexpression in *Escherichia coli*. *Mol. Cell. Proteomics* 6(9), 1527–1550 (2007).

• **Proteome analysis of recombinant membrane protein production in the prokaryote, *Escherichia coli*.**

7 Wagner S, Klepsch MM, Schlegel S *et al.* Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl Acad. Sci. USA* 105(38), 14371–14376 (2008).

8 Sevastsyanovich Y, Alfasi S, Cole J. Recombinant protein production: a comparative view on host physiology. *New Biotechnol.* 25(4), 175–180 (2009).

• **Recent and informative view on the status of recombinant protein production.**

9 Bonander N, Darby RA, Grgic L *et al.* Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. *Microb. Cell. Fact.* 8, 10 (2009).

•• **First example of strain engineering based on array analysis.**

10 Ahmed FE. Microarray RNA transcriptional profiling: part I. Platforms, experimental design and standardization. *Expert Rev. Mol. Diagn.* 6(4), 535–550 (2006).

11 Bonander N, Ferndahl C, Mostad P *et al.* Transcriptome analysis of a respiratory *Saccharomyces cerevisiae* strain suggests

the expression of its phenotype is glucose insensitive and predominantly controlled by Hap4, Cat8 and Mig1. *BMC Genomics* 9, 365 (2008).

12 Gasch AP, Spellman PT, Kao CM *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell.* 11(12), 4241–4257 (2000).

13 Serrano R, Ruiz A, Bernal D, Chambers JR, Arino J. The transcriptional response to alkaline pH in *Saccharomyces cerevisiae*: evidence for calcium-mediated signalling. *Mol. Microbiol.* 46(5), 1319–1333 (2002).

14 Sahara T, Goda T, Ohgiya S. Comprehensive expression analysis of time-dependent genetic responses in yeast cells to low temperature. *J. Biol. Chem.* 277(51), 50015–50021 (2002).

15 DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278(5338), 680–686 (1997).

16 David NE, Gee M, Andersen B *et al.* Expression and purification of the *Saccharomyces cerevisiae* α-factor receptor (Ste2p), a 7-transmembrane-segment G protein-coupled receptor. *J. Biol. Chem.* 272(24), 15553–15561 (1997).

17 Andre N, Cherouati N, Prual C *et al.* Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Sci.* 15(5), 1115–1126 (2006).

18 Griffith DA, Delipala C, Leadsham J, Jarvis SM, Oesterhelt D. A novel yeast expression system for the overproduction of quality-controlled membrane proteins. *FEBS Lett.* 553(1–2), 45–50 (2003).

19 De Schutter K, Lin YC, Tiels P *et al.* Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nature Biotechnol.* 27(6), 561–566 (2009).

Websites

101 Agilent Technologies Incorporated www.agilent.com

102 Affymetrix Incorporated www.affymetrix.com

103 Professor Louise Showe's laboratory http://showelab.wistar.upenn.edu/Wistar_Showe_Lab_Gene_Lists.htm

Affiliations

- Nicklas Bonander
School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK
Tel.: +44 121 204 4274
Fax: +44 121 204 4175
n.bonander@aston.ac.uk
- Roslyn M Bill
School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK
Tel.: +44 121 204 4274
Fax: +44 121 204 4175
r.m.bill@aston.ac.uk