Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Pichia pastoris as an expression host for membrane protein structural biology

Bernadette Byrne

The methylotrophic yeast *Pichia pastoris* is a widely used recombinant expression host. *P. pastoris* combines the advantages of ease of use, relatively rapid expression times and low cost with eukaryotic co-translational and post-translational processing systems and lipid composition. The suitability of *P. pastoris* for high density controlled culture in bioreactors means large amounts of protein can be obtained from small culture volumes. This review details the key features of *P. pastoris*, which have made it a particularly useful system for the production of membrane proteins, including receptors, channels and transporters, for structural studies. In addition, this review provides an overview of all the constructs and cell strains used to produce membrane proteins, which have yielded high resolution structures.

Addresses
Department of Life Sciences, Imperial College London, SW7 2AZ, UK

Corresponding author: Byrne, Bernadette (b.byrne@imperial.ac.uk)

Current Opinion in Structural Biology 2015, **32**:9–17

This review comes from a themed issue on **New constructs and expressions of proteins**

Edited by **Imre Berger** and **Roslyn M Bill**

For a complete overview see the Issue and the Editorial

Available online 3rd February 2015

http://dx.doi.org/10.1016/j.sbi.2015.01.005

Introduction

The number of membrane protein structures has significantly increased in recent years, due in part at least, to the development of a range of novel methodologies to stabilise membrane proteins in solution [1]. However our understanding of eukaryotic membrane proteins in particular remains somewhat limited. One of the key issues associated with eukaryotic membrane proteins is the difficulty in expressing the significant amounts of material required for structural biology. Eukaryotic membrane proteins have specific co-translational and post-translational processing and membrane lipid requirements meaning that bacterial expression systems are often unsuitable. The higher eukaryotic insect cell based systems have an excellent track record for structural studies of GPCRs [2] although are much more expensive and time-consuming than the bacterial systems.

A good middle-ground is achieved with the yeast expression systems which combine the advantages of ease of manipulation and low cost of production with eukaryotic protein processing. The two major yeast expression systems are *Saccharomyces cerevisiae* and *Pichia pastoris*. The bakers yeast, *S. cerevisiae*, has a good track record with respect to heterologous production of membrane protein for structural studies (http//blanco.biomol.uci.edu/mpstruc/). Some recent examples include the PiPT high affinity phosphate transporter from the fungus, *Piriformospora indica* [3] expressed with N and C-terminal FLAG and His tags, and the NRT1.1 nitrate transporter from the model plant *Arabidopsis thaliana* [4] expressed as a fusion protein with a C-terminal GFP-His8 tag [5]. However markedly better success has been achieved with membrane proteins heterologously produced in *P. pastoris* (http//blanco.biomol.uci.edu/mpstruc, Table 1 with example structures in Figure 2). The aim of this review is to provide a brief overview to the key features of the *P. pastoris* expression system, with particular emphasis on the successful examples of eukaryotic membrane proteins produced for structural studies, using this system. It is hoped that this will provide a useful starting point for researchers embarking on membrane protein expression studies.

General features of the *P. pastoris* expression system

P. pastoris is a methylotrophic yeast, able to utilise methanol as the sole carbon source. There are a comparatively limited number of vectors for *P. pastoris* expression but the most widely used incorporate the very strong, inducible *AOX1* promoter. This promoter controls expression of genes central to the metabolism of methanol, *AOX1* and *AOX2*. The *AOX1* promoter is repressed in the presence of glucose or glycerol but strongly upregulated in the presence of methanol, making it a relatively tightly controlled expression system. However, recent research has identified protein expression occurring during the pre-induction phase in cultures grown in bioreactors but not in shaker flasks indicating that the promoter is leaky under certain conditions [6].

The *P. pastoris* vectors usually integrate into the host cell genome to produce a stably expressing clone. It is not possible to control the number of copies which integrate and so the optimal clone must be experimentally determined. Copy number can be easily estimated through colony screening in the presence of increasing concentrations of antibiotic [7•] although other methods including real time PCR are also available [8]. However, higher

Table 1

Membrane protein structures obtained from protein expressed in P. pastoris.

Protein	Source organism	Expression vector	Expression strain	Genetic modifications[a]	Extra sequences	PDB Accession Code and reference
Membrane associated and membrane bound enzymes						
Monoamine oxidase B	*Homo sapiens*	pPIC3.5K	GS115/KM71		No tag	1GOS, [32] 1OJA, [33]
Monoamine oxidase A	*Homo sapiens*	pPIC9K	KM71		No tag	2BXR, [34]
Leukotriene LTC_4 synthase	*Homo sapiens*	pPICZ	KM71H		N-terminal 6His tag	2UUH, [51] 4JCZ, 4JRZ, 4J7T, 4J7Y, [52]
G-protein coupled receptors						
Adenosine A_{2A} receptor	*Homo sapiens*	pPIC9K	SMD1163	Codon optimised, C-terminal truncation, mutated to remove N-linked glycosylation site	Alpha factor signal sequence, N-terminal FLAG tag, C-terminal 10His	3VG9, [21]
Histamine H_1 receptor	*Homo sapiens*	pPIC9K	SMD1163	Codon optimised, mutated to remove N-linked glycosylation sites, N-terminal truncation, insertion of sequence encoding T4 lysozyme into the third intracellular loop	C-terminal GFP-8His	3RZE, [17••]
Ion channels						
Two-pore domain potassium channel $K_{2P}1.1$	*Homo sapiens*	pPICZ	SMD1163	N and C terminal truncation, mutated to removed N-linked glycosylation site and prevent non-specific disulphide bridge formation	PreScission cleavage site, C-terminal GFP-His8	3UKM, [28]
Two-pore domain potassium channel $K_{2P}4.1$	*Homo sapiens*	pPICZ	SMD1163	Codon optimised, mutated to remove N-linked glycosylation sites, C-terminal truncation	PreScission cleavage site, C-terminal GFP-His8	3UM7, [16] 4I9W, [38]
Kv1.2 voltage gated potassium channel	*Rattus norvegicus*	pPICZ	SMD1163	Mutated to remove glycosylation sites	N-terminal 8His tag followed by TEV cleavage site	2A79, [53]
Kv1.2 voltage gated potassium channel	*Rattus norvegicus*					3LUT, [54]
Kir2.2 inward rectifier potassium channel	*Gallus gallus*	pPICZ	SMD1163	N and C-terminally truncated	C-terminal PreScission protease cleavage site followed by GFP and a 1D4 antibody recognition site	3JYC, [55] 3SPI, [56]
Kv1.2/Kv2.1 voltage gated potassium channel chimera	*Rattus norvegicus*	pPICZ	SMD1163[b]	Mutated to remove N-linked glycosylation sites and reduce formation of non-specific disulphide bonds	N-terminal 10His tag followed by a Thrombin protease cleavage site	2R9R, [25] 3LNM, [57]
GIRK2 (Kir 3.2) G-protein gated potassium channel	*Mus musculus*	pPICZ	SMD1163	N and C-terminally truncated	C-terminal PreScission protease cleavage site followed by GFP and a 10His tag	3SYO, [22] 4KFM, [58]

Table 1 (*Continued*)

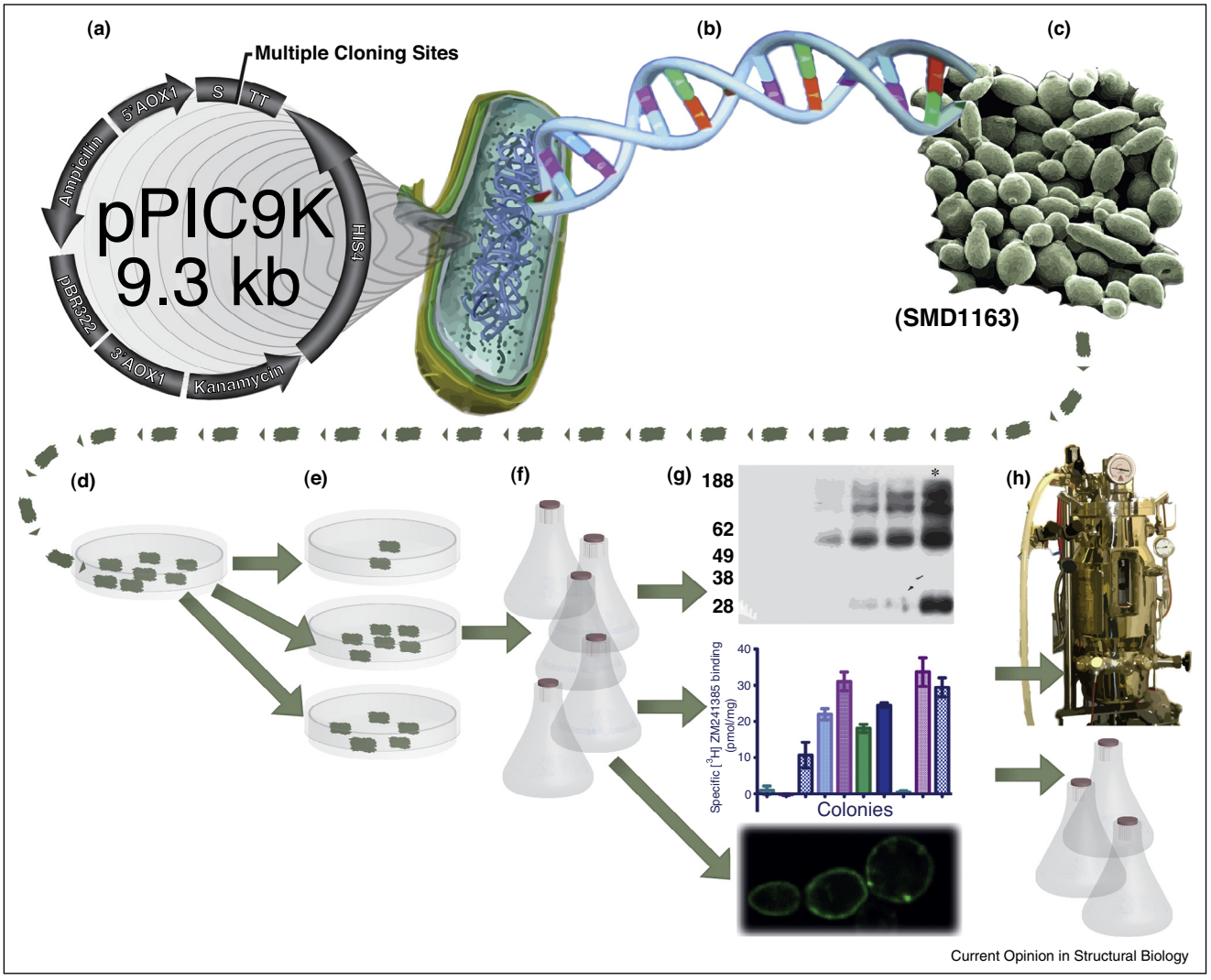
Protein	Source organism	Expression vector	Expression strain	Genetic modifications[a]	Extra sequences	PDB Accession Code and reference
Orai Calcium release activated channel (CRAC)	*Drosophila melanogaster*	pPICZ	SMD1163	N and C terminally truncated. Mutated to facilitate crystallisation and reduce formation of non-specific disulphide bonds	C-terminal antitubulin antibody affinity tag	4HKR, [23]
Aquaporins						
AQP2 Aquaporin	*Homo sapiens*	pPICZ	GS115 *aqy-1Δ*	Codon optimised C-terminally truncated	N-terminal 8 His tag followed by a TEV protease cleavage site	4NEF, [15]
AQP4 Aquaporin	*Homo sapiens*	pPICZ	X-33		N-terminal 8His tag followed by a FLAG tag and a 3C protease cleavage site	3GD8, [59]
AQP5 Aquaporin	*Homo sapiens*	pPICZ	X-33		No tags	3D9S, [36]
SoPIP2;1 aquaporin	*Spinacia oleracea*	pPICZ	X-33		C-terminal c-myc tag followed by a 6His tag	1Z98, 2B5F, [35]
SoPIP2;1 aquaporin	*Spinacia oleracea*	pPICZ	X-33		N-terminal 6His tag followed by a Thrombin protease cleavage site except for double mutant 3CN5	3CLL, 3CN5, 3CN6, [60]
Transporters						
P-glycoprotein	*Mus musculus*	pHIL-D2	GS115	Mutated to remove N-linked glycosylation sites	C-terminal 6 His tag	3G5U, [42]; 4MIM, [61] (structure refined from original data)
P-glycoprotein	*Mus musculus*	pLIC	KM71H	Mutated to remove N-linked glycosylation sites and codon optimised	C-terminal 6 His tag	4K5B, 4K5C, 4K5D, [19]
P-glycoprotein	*Caenorhabditis elegans*	pPICZ	SMD1163		C-terminal PreScission protease cleavage site followed by GFP and a 10 His tag	4F4C, [27]
CmABCB1 (P-glycoprotein homologue)	*Cyanidioschyzon merolae*	pPICZ	SMD1163		C-terminal TEV protease cleavage + 10 His tag	(GSENLYFQGRSH$_{10}$) 3WME, [62]

[a] Only modifications made to facilitate expression, purification and/or crystallisation are included here, not mutations made to explore protein function.
[b] SB Long (unpublished data).

copy number does not necessarily correlate with higher target protein expression so it is not usually sufficient to identify clones based on copy number alone, ideally functional protein expression levels must also be assessed [9].

One important feature of the *P. pastoris* system is that it is highly suitable for large scale growth and culture using bioreactors [10••]. Bioreactors give the user precise control over culture parameters including pH, aeration, feed rate and temperature as well as allowing real-time monitoring of changes in OD_{600} and dissolved oxygen [11]. This allows tightly regulated growth of the culture and means that ultra-high cell densities (>100 g/L dry cell weight; >500 OD_{600} units/mL) can be achieved [10••]. Medium density cultures (~100 OD_{600} units/mL) are more usual for membrane proteins since this both reduces proteolysis [12] and cellular stress [13] associated with high density cultures. This feature combined with comparatively inexpensive growth medium makes *P. pastoris* a highly efficient and cost-effective expression system.

Figure 1



Schematic illustrating the key steps in expression of an integral membrane protein in the methylotrophic yeast, *P. pastoris*. *E. coli* is typically used for construct generation and amplification **(a)**. Following confirmation of the correct sequence **(b)**, the expression vector containing the gene of interest together with appropriate tags and protease cleavage sites is transformed into electrocompetent *P. pastoris* cells **(c)**. The DNA is linearised before transformation and so integrates into the host vector. In our laboratory all our research involved the SMD1163 protease deficient strain. Depending on the expression vector and cell line combination used initial selection of the colonies can be done using antibiotics (e.g. zeocin) or **(d)** auxotrophy of the strains. Colonies initially selected using cell auxotrophy can then be further selected based on vector copy number using increasing concentrations of the antibiotic G418 or zeocin **(e)**. Although this indicates copy number it is not necessarily indicative of the expression level of the target protein. Small scale expression trials can be carried out in 10–50 mL culture volume **(f)**. Expression levels of the target protein can be assessed using **(g)** functional analysis, Western blot analysis or direct fluorescence measurements where GFP is used as a tag. Once the best expressing clone is identified this can be used for large scale expression either in shaker flasks or bioreactors **(h)**.

A key consideration is that high levels of the methanol used as an inducer are cytotoxic [14] which can reduce culture viability and protein production. We use a sensor to monitor levels of unmetabolised methanol in the bioreactor [11] to reduce the potential risk of cytotoxic effects and help to optimise protein expression levels.

Figure 1 summarises the steps involved in generating and screening *P. pastoris* clones (see figure legend for details).

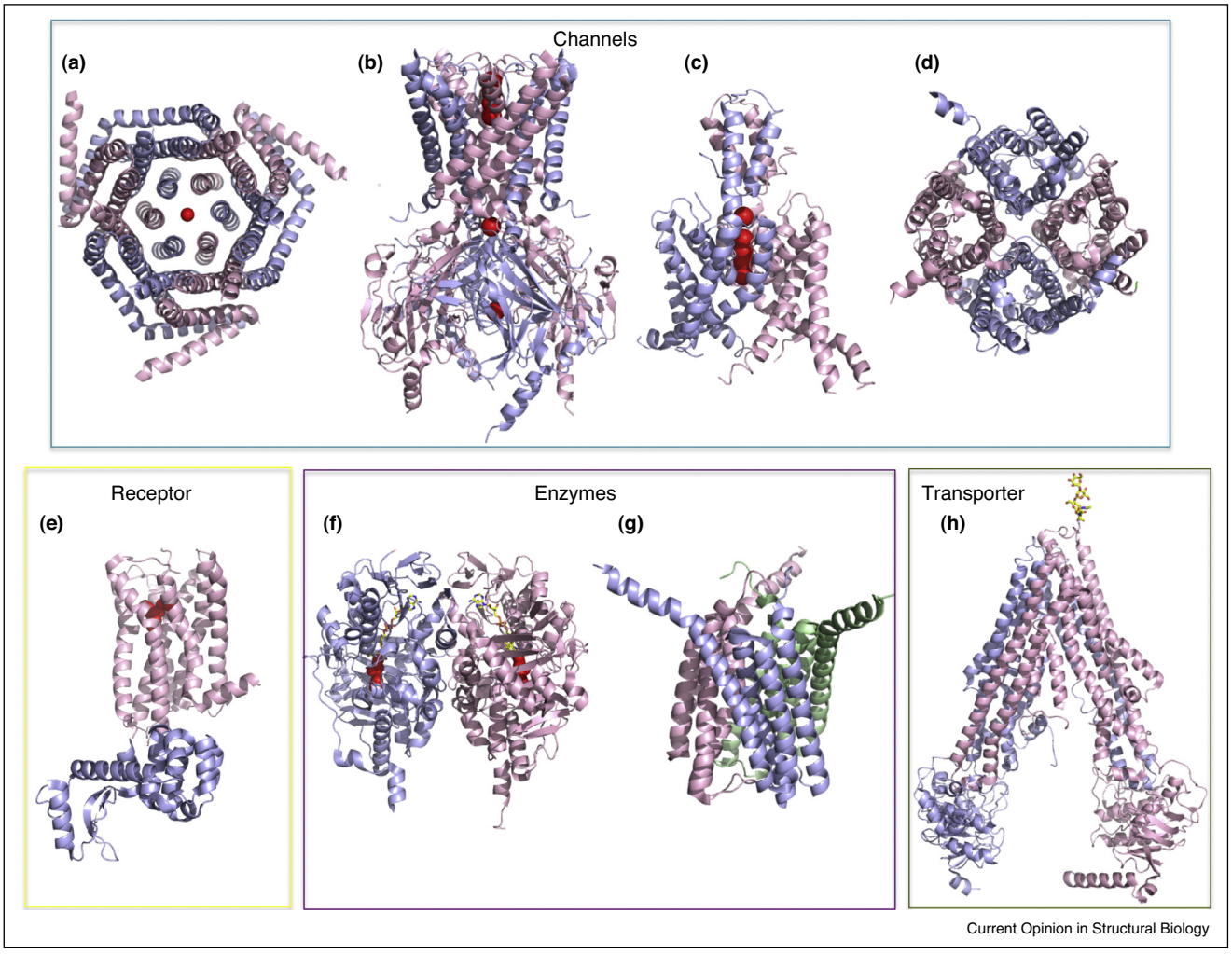
Genetic modifications

There are several examples of membrane proteins successfully produced in *P. pastoris* where the gene has been modified to improve expression levels or facilitate crystallisation (Table 1). Genes of interest are often codon optimised for expression in *P. pastoris* [15,16,17••]. It is not clear in all cases what difference this makes to protein production as a comparison with the non-codon optimised gene is not always reported. However, a recent detailed

study on P-glycoprotein (PgP) from mouse revealed that expression from a gene codon optimised for *P. pastoris* yielded substantially more protein than expression from the wildtype gene [18••]. Intriguingly, size exclusion analysis also revealed that the PgP expressed from the wildtype gene codon was more aggregated than that expressed from the codon optimised gene. These findings indicate that a more stable protein is expressed from the codon optimised gene. This construct has been subsequently used to obtain high resolution structures of the PgP [19].

The GPCRs are typically truncated at the C-terminus. This has been shown to increase both expression and stability of the $A_{2A}R$ produced in *P. pastoris* [20] although removal of this long flexible domain is also important for crystallisation [21]. It is likely that for some GPCRs this

Figure 2



Examples of integral membrane protein structures obtained using protein expressed in *P. pastoris*. **(a)** Orai Calcium release activated channel, (Accession code: 4HKR). The protein is a homohexamer (individual monomers in blue and pink) and is shown from the extracellular side of the membrane. A bound Ba^{2+} ion is indicated in the red sphere. **(b)** GIRK2, G-protein gated K^+ channel (Accession code: 3SYO). The protein is a homotetramer (individual monomers in blue and pink) and is shown looking through the membrane. K^+ ions are shown as red spheres. **(c)** Two pore domain K^+ channel $K_{2P}4.1$ (Accession code: 3UM7). The protein is a dimer (individual monomers in blue and pink) and is shown parallel to the membrane. Associated K^+ ions are indicated in red spheres. **(d)** Aquaporin AQP2 (Accession code: 4NEF). The protein is a homotetramer (individual monomers in blue and pink) and is shown from the intracellular side of the membrane. **(e)** Histamine H_1–T4 lysozyme fusion protein (Accession code: 3RZE). The protein is a monomer with the receptor shown in pink and the T4 lysozyme in blue. The bound ligand, doxepin, is shown in the red spheres. **(f)** Monomamine oxidase B (Accession code: 1GOS). The protein is a dimer (monomers shown in blue and pink). The protein is normally located on the cytosolic side of the outer mitochondrial membrane and is shown with the cytosolic domains at the top of the image. The bound flavin adenine dinucleotide is shown in yellow stick model and the inhibitor, pargyline, is shown in the red spheres. **(g)** Leukotriene synthase (Accession code: 2UUH). The protein is a trimer (individual monomers in blue, pink and green). This protein is normally located in the nuclear outer membrane and the protein is shown with the cytosolic domains at the top of the image. **(h)** P-glycoprotein (Accession code: 4F4C). The protein molecule is a single polypeptide chain comprised of two homologous domains (domains shown in blue and pink) with the glycosylation group shown in yellow stick model. The protein is shown looking through the membrane.

also removes potential palmitoylation sites. The ion channels, for example, the G-protein activated K^+ channel [22] and the Orai Calcium release activated channel [23], are also often truncated at both the N and C-termini removing extended flexible regions to produce a compact domain more suitable for crystallisation (see also Table 1 and Figure 2a,b). Identification of potential flexible regions, which may hinder crystallisation can be carried out using disorder prediction software such as RONN [24].

Minor modifications include removal of N-linked glycosylation sites through mutagenesis [16,17••,25]. Removal of these bulky, flexible groups is thought to be advantageous for crystallisation. Yeast are able to glycosylate but only perform some steps in common with higher eukaryotes, adding the core $(Man)_8$-$(GlcNAc)_2$ groups but not the higher order decorations found in mammals. *P. pastoris* can hyper-glycosylate although to a lesser extent than *S. cerevisiae*. A study by Yurugi-Kobayashi *et al.* [26] expressed 25 non-glycosylated GPCRs in *P. pastoris*. The data demonstrated that in many cases loss of glycosylation resulted in reduced functional expression compared to glycosylated receptor indicating that the effects of such modifications need to be assessed for each individual protein [26]. It was possible to obtain a high resolution structure of a glycosylated form of P-glycoprotein (PgP) [27] produced in *P. pastoris* demonstrating that at least in some cases the presence of the sugar groups does not inhibit the formation of well diffracting crystals (Figure 2h).

Another reported modification mutates Cys residues to prevent non-specific disulphide bridge formation which could lead to loss of protein through aggregation [23,28].

Importantly, *P. pastoris* seems to be able to cope with more radical genetic modifications including the expression of a Histamine H_1 receptor–T4 lysozyme (H_1R–T4L) fusion protein [17••]. Here the gene encoding the T4L replaced the region of the H_1R encoding the third intracellular loop [29] (Figure 2e). This has the overall effect of stabilising the receptor and also adding an extended soluble domain, key in forming crystal contacts essential for crystal growth. Interestingly, the initial screening of the expression construct for the H_1R–T4L fusion protein was performed in *S. cerevisiae* [30]. The final construct, which also incorporated a C-terminal TEV cleavage site followed by GFP and a 10His tag, was cloned directly from the *S. cerevisiae* vector into the *P. pastoris* vector. The preferential use of *S. cerevisiae* for screening is likely due to the ease of generating constructs by homologous recombination [31] but clearly *P. pastoris* was superior for large scale production purposes. The ease of cloning from the *S. cerevisiae* vector to the *P. pastoris* vector makes this two-step approach attractive for a range of target proteins.

Tagging the construct

Untagged membrane proteins have been successfully isolated from *P. pastoris* culture [32–36] based on the properties of the individual proteins (Figure 2f shows the structure of one of these proteins, monoamine oxidase B). However it is more usual to produce tagged versions of the protein to facilitate both detection and isolation of the target molecule. Expression in *P. pastoris* is compatible with a range of different tags. The vast majority of proteins have been expressed as His tagged constructs (Table 1) to allow affinity chromatography. However FLAG tags have also been used both for analysis of the protein [37] and as a means of isolation [20]. A number of protein constructs additionally incorporate GFP for rapid assessment of optimal expression and solubilisation conditions [28,38]. A range of different cleavage sites have also been employed to successfully remove tags before crystallisation trials (see Table 1 for details).

Some of the *P. pastoris* expression vectors used for heterologous membrane protein expression incorporate the *S. cerevisiae* alpha-mating factor sequence signal upstream of the gene of interest facilitating correct protein targeting. This has been shown to increase expression of the mouse 5-HT_{5A} serotonin receptor [39] and the human μ-opioid receptor [40]. However presence of the alpha-mating factor sequence signal dramatically reduced expression of the Histamine H_1 receptor [29]. Thus, screening is essential to assess the effects of the addition of the signal sequence. The alpha mating factor should be cleaved following expression by the endogenous Kex2 protease. However we have found that if recombinant expression levels are very high then the Kex2 is unable to effectively cleave the signal sequence from all the produced protein leading to a heterogeneous population (Singh *et al.*, unpublished data). Careful construct design incorporating a proteolytic cleavage site upstream of the gene of interest allowed us to deal with this issue [20]. Recent research has produced a novel *P. pastoris* strain with increased *Kex2* copy number which increases production of a secreted protein [41]. Such a strain may also be suitable for membrane protein production.

Expression strains

Compared to *S. cerevisiae* and *E. coli*, there are relatively few expression strains available for *P. pastoris*. All the strains that have been used for production of membrane proteins share the methanol utilisation (Mut) phenotype. Most commonly *P. pastoris* cell strains produce both the AOX1 and AOX2 proteins resulting in a wild-type Mut^+ methanol utilisation phenotype. However the KM71 Mut^S strains contain only the less efficient AOX2 protein and exhibit slow growth on methanol. The wildtype X-33 strain, together with vectors allowing antibiotic selection using Zeocin, was successfully used to express the SoPIP2;1 aquaporin from spinach [35]. A number of membrane proteins have been expressed in

the GS115 (*his4*) and KM71 strains which lack the gene encoding the histidinol dehydrogenase [32,42]. The presence of this gene in the expression vector provides a convenient means of screening for positive clones. The SMD1163 strain (*his4*, *pep4*, *prb1*) lacks both Proteinase A (*pep4*) and Proteinase B (*prb1*) activity. Proteinase A activates endogenous *P. pastoris* proteases including Proteinase B. However even in the absence of Proteinase A, Proteinase B retains some activity. The SMD1163 strain yielded increased expression of the 5-HT_{5A} serotonin receptor, a G-protein coupled receptor (GPCR) compared to the GS115 strain [39]. This strain is the most successful for the production of membrane proteins for structural studies (Table 1) having been utilised for the expression of GPCRs [17••,21], ion channels [16,22] and a transporter [27]. However, despite the successes achieved using SMD1163 cells for membrane protein production, it is important to note that this strain is no longer commercially available. The related SMD1168 strain (*his4*, *ura3*, *pep4*) lacks only Proteinase A activity and is commercially available but there are no examples of this strain being used to successfully produce membrane proteins for structural studies so it is difficult to assess its suitability.

A bespoke *P. pastoris* strain deficient in the native Aquaporin 1 was developed for structural studies on the yeast aquaporin [43••] and subsequently used for the production of the human AQP2 aquaporin [15] (Figure 2d). Further, more recent developments have resulted in a strain producing cholesterol [44••], and the rational identification of the transcription factor, Hac1, co-expression of which may facilitate membrane protein expression [45]. Although neither approach has been utilised for protein production for structural studies as yet, the addition of these to the toolbox may widen the applicability of *P. pastoris* as an expression host for challenging membrane proteins.

Optimisation of culture conditions

Optimisation of membrane protein expression is possible through alteration of *P. pastoris* media components and culture conditions. A range of different temperatures (27–30°C) have been used for initial culture growth with a drop in temperature to 20–27 °C during induction. However many manuscripts report no temperature reduction for induction. The induction time is also a key variable. An analysis of three different induction times 20, 40 or 60 hours indicated that 60 hours was optimal for the production of human CHRM2, a GPCR, in shaker flasks [26]. However optimal functional expression of the $A_{2A}R$ in a bioreactor was achieved after only 18 hours [11]. Induction times of 16–60 hours have been reported for a range of membrane proteins [15,16,42] and although details are often not provided it is inferred that these times were identified by optimisation trials as being the most suitable for the individual target protein.

Various additions to the media have been reported to increase expression of GPCRs such as DMSO and Histidine [7•,46••], although these have not been widely used.

Comparison with other systems for the expression of membrane proteins

There have been a few studies specifically assessing a range of different systems for the production of membrane proteins for structural studies. Lundstrom *et al.* carried out a study on over 100 GPCRs and showed that *P. pastoris* successfully expressed as many individual receptors as a Semliki Forest Virus mammalian cell system and produced almost twice as many receptors as *Escherichia coli* [46••]. Radioligand binding analysis also revealed that for several of the receptors higher functional expression was achieved in *P. pastoris* compared to the mammalian cell system [46••]. Comparison of muscarinic acetylcholine receptor M2 expression in *P. pastoris* and *Spodoptera frugipera* insect cells revealed similar levels of specific activity and binding affinity but the higher cell densities achieved in *P. pastoris* meant that the overall yield in this system was twice that in the insect cells [47]. The most successful host for the recombinant production of eukaryotic membrane proteins for structural studies is currently the insect cell system [48], however *P. pastoris* is not far behind.

What expression levels can be achieved?

Expression of both the human Aquaporin 1 (hAQP1) [49•] and the human adenosine A_{2A} receptor ($A_{2A}R$) in bioreactor culture [11] produced more than twice the expression level achieved in the equivalent shaker flask culture. In the case of the hAQP1 not only was there more protein produced in the bioreactor but much more of it was correctly localised to the membrane [49•]. Western blot analysis of the $A_{2A}R$ indicated that there was substantially more receptor produced in the shaker flask than suggested by functional analysis by radioligand binding assay whilst the Western blot and binding assay results correlated better for the bioreactor culture [11]. These results show that the bioreactors produce not just higher amounts of membrane protein but higher quality membrane protein. Isolation of the hAQP1 gave the remarkable yield of 90 mg/L. Yields of 13 mg/100 g cells are reported for the codon optimised version of PgP expressed in a bioreactor [18••]. *P. pastoris* is a highly attractive system for production of eukaryotic membrane proteins but successful expression remains target protein dependent [50].

Conclusion

The low cost and high yields together with the eukaryotic processing machinery mean *P. pastoris* is a viable alternative to insect and mammalian cell systems for the large scale production of many integral membrane proteins for structural studies. Whilst there remain no clear rules for the successful expression of membrane proteins in *P. pastoris*, the examples detailed here do provide a useful set of guidelines for researchers. It is also anticipated that

a more expanded set of expression vectors and cell strains will greatly facilitate the production of more challenging membrane proteins in the future.

Conflict of interest

None declared.

Acknowledgements

The author would like to thank Dr Shweta Singh for critically assessing the manuscript and Drs Yilmaz Alguel and Maxime Boulet-Audet for assistance with the figures. This work was supported by the Biotechnology and Biological Sciences Research Council grant BB/K017292/1.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Bertheleme N, Chae PS, Singh S, Mossakowska D, Hann MM, Smith KJ, Hubbard JA, Dowell SJ, Byrne B: **Unlocking the secrets of the gatekeeper: methods for stabilizing and crystallizing GPCRs**. *Biochim Biophys Acta* 2013, **1828**: 2583-2591.

2. Maeda S, Schertler GFX: **Production of GPCR and GPCR complexes for structure determination**. *Curr Opin Struct Biol* 2013, **23**:381-392.

3. Pedersen BP, Kumar H, Waight AB, Risenmay AJ, Roe-Zurz Z, Chau BH, Schlessinger A, Bonomi M, Harries W, Sali A *et al.*: **Crystal structure of a eukaryotic phosphate transporter**. *Nature* 2013, **496**:533-536.

4. Parker JL, Newstead S: **Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1**. *Nature* 2014, **507**:68-72.

5. Drew D, Newstead S, Sonoda Y, Kim H, Heijne von G, Iwata S: **GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae***. *Nat Protoc* 2008, **3**:784-798.

6. Bawa Z, Routledge SJ, Jamshad M, Clare M, Sarkar D, Dickerson I, Ganzlin M, Poyner DR, Bill RM: **Functional recombinant protein is present in the pre-induction phases of *Pichia pastoris* cultures when grown in bioreactors, but not shake-flasks**. *Microb Cell Fact* 2014, **13**:561.

7. • André N, Cherouati N, Prual C, Steffan T, Zeder-Lutz G, Magnin T, Pattus F, Michel H, Wagner R, Reinhart C: **Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen**. *Protein Sci* 2006, **15**:1115-1126.

Detailed study describing optimisation of expression of GPCRs in *P. pastoris* to a level suitable for structural studies.

8. Abad S, Kitz K, Hörmann A, Schreiner U, Hartner FS, Glieder A: **Real-time PCR-based determination of gene copy numbers in *Pichia pastoris***. *Biotechnol J* 2010, **5**:413-420.

9. Aw R, Polizzi KM: **Can too many copies spoil the broth?** *Microb Cell Fact* 2013, **12**:128.

10. •• Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM: **Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris***. *Curr Opin Biotechnol* 2002, **13**:329-332.

A clear summary of the advantages of *P. pastoris* by researchers who pioneered the use of this system for heterologous expression.

11. Singh S, Gras A, Fiez-Vandal C, Ruprecht J, Rana R, Martinez M, Strange PG, Wagner R, Byrne B: **Large-scale functional expression of WT and truncated human adenosine A2A receptor in *Pichia pastoris* bioreactor cultures**. *Microb Cell Fact* 2008, **7**:28.

12. Jahic M, Wallberg F, Bollok M, Garcia P, Enfors S-O: **Oxygen-limited fed-batch process: an alternative control for *Pichia pastoris* recombinant protein processes**. *Microb Cell Fact* 2003, **2**:6.

13. Mattanovich D, Gasser B, Hohenblum H, Sauer M: **Stress in recombinant protein producing yeasts**. *J Biotechnol* 2004, **113**:121-135.

14. Zhang W, Smith LA, Plantz BA, Schlegel VL, Meagher MM: **Design of methanol feed control in *Pichia pastoris* fermentations based upon a growth model**. *Biotechnol Prog* 2002, **18**:1392-1399.

15. Frick A, Eriksson UK, de Mattia F, Öberg F, Hedfalk K, Neutze R, de Grip WJ, Deen PMT, Törnroth-Horsefield S: **X-ray structure of human aquaporin 2 and its implications for nephrogenic diabetes insipidus and trafficking**. *Proc Natl Acad Sci U S A* 2014, **111**:6305-6310.

16. Brohawn SG, del Mármol J, MacKinnon R: **Crystal structure of the human K2P TRAAK, a lipid- and mechano-sensitive K^+ ion channel**. *Science* 2012, **335**:436-441.

17. •• Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW *et al.*: **Structure of the human histamine H_1 receptor**. *Nature* 2012, **475**:65-70.

Describes the successful expression of a GPCR + T4 lysozyme fusion protein in *P. pastoris*. The protein was used to determine the structure of the receptor.

18. •• Bai J, Swartz DJ, Protasevich II, Brouillette CG, Harrell PM, Hildebrandt E, Gasser B, Mattanovich D, Ward A, Chang G *et al.*: **A gene optimization strategy that enhances production of fully functional P-glycoprotein in *Pichia pastoris***. *PLoS ONE* 2011, **6**:e22577.

Study revealed that codon optimization improved both the levels of the expressed membrane protein in *P. pastoris* and the quality of the produced protein.

19. Ward AB, Szewczyk P, Grimard V, Lee C-W, Martinez L, Doshi R, Caya A, Villaluz M, Pardon E, Cregger C *et al.*: **Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain**. *Proc Natl Acad Sci U S A* 2013, **110**:13386-13391.

20. Singh S, Hedley D, Kara E, Gras A, Iwata S, Ruprecht J, Strange PG, Byrne B: **A purified C-terminally truncated human adenosine A(2A) receptor construct is functionally stable and degradation resistant**. *Protein Expr Purif* 2010, **74**:80-87.

21. Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomura N *et al.*: **G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody**. *Nature* 2012, **482**:237-240.

22. Whorton MR, MacKinnon R: **Crystal structure of the mammalian GIRK2 K^+ channel and gating regulation by G proteins, PIP2, and sodium**. *Cell* 2011, **147**:199-208.

23. Hou X, Pedi L, Diver MM, Long SB: **Crystal structure of the calcium release-activated calcium channel Orai**. *Science* 2012, **338**:1308-1313.

24. Yang ZR, Thomson R, McNeil P, Esnouf RM: **RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins**. *Bioinformatics* 2005, **21**:3369-3376.

25. Long SB, Tao X, Campbell EB, MacKinnon R: **Atomic structure of a voltage-dependent K^+ channel in a lipid membrane-like environment**. *Nature* 2007, **450**:376-382.

26. Yurugi-Kobayashi T, Asada H, Shiroishi M, Shimamura T, Funamoto S, Katsuta N, Ito K, Sugawara T, Tokuda N, Tsujimoto H *et al.*: **Biochem Biophys Res Commun**2009, **380**:271-276.

27. Jin MS, Oldham ML, Zhang Q, Chen J: **Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans***. *Nature* 2012, **490**:566-569.

28. Miller AN, Long SB: **Crystal structure of the human two-pore domain potassium channel K2P1**. *Science* 2012, **335**:432-436.

29. Shiroishi M, Kobayashi T, Ogasawara S, Tsujimoto H, Ikeda-Suno C, Iwata S, Shimamura T: **Production of the stable human histamine H_1 receptor in *Pichia pastoris* for structural determination**. *Methods* 2011, **55**:281-286.

30. Shiroishi M, Tsujimoto H, Makyio H, Asada H, Yurugi-Kobayashi T, Shimamura T, Murata T, Nomura N, Haga T, Iwata S *et al.*: **Platform for the rapid construction and evaluation of GPCRs for crystallography in *Saccharomyces cerevisiae***. *Microb Cell Fact* 2012, **11**:78.

31. Newstead S, Kim H, Heijne von G, Iwata S, Drew D: **High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae***. *Proc Natl Acad Sci U S A* 2007, **104**:13936-13941.

32. Binda C, Newton-Vinson P, Hubálek F, Edmondson DE, Mattevi A: **Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders**. *Nat Struct Biol* 2002, **9**:22-26.

33. Binda C, Li M, Hubálek F, Restelli N, Edmondson DE, Mattevi A: **Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures**. *Proc Natl Acad Sci U S A* 2003, **100**:9750-9755.

34. De Colibus L, Li M, Binda C, Lustig A, Edmondson DE, Mattevi A: **Three-dimensional structure of human monoamine oxidase A (MAO A): relation to the structures of rat MAO A and human MAO B**. *Proc Natl Acad Sci U S A* 2005, **102**:12684-12689.

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

36. Horsefield R, Nordén K, Fellert M, Backmark A, Törnroth-Horsefield S, Terwisscha van Scheltinga AC, Kvassman J, Kjellbom P, Johanson U, Neutze R: **High-resolution X-ray structure of human aquaporin 5**. *Proc Natl Acad Sci U S A* 2008, **105**:13327-13332.

37. Weiss HM, Haase W, Michel H, Reilländer H: **Comparative biochemical and pharmacological characterization of the mouse 5HT5A 5-hydroxytryptamine receptor and the human beta2-adrenergic receptor produced in the methylotrophic yeast *Pichia pastoris***. *Biochem J* 1998, **330**:1137-1147.

38. Brohawn SG, Campbell EB, MacKinnon R: **Domain-swapped chain connectivity and gated membrane access in a Fab-mediated crystal of the human TRAAK K^+ channel**. *Proc Natl Acad Sci U S A* 2013, **110**:2129-2134.

39. Weiss HM, Haase W, Michel H, Reilländer H: **Expression of functional mouse 5-HT5A serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization**. *FEBS Lett* 1995, **377**:451-456.

40. Talmont F, Sidobre S, Demange P, Milon A, Emorine LJ: **Expression and pharmacological characterization of the human mu-opioid receptor in the methylotrophic yeast *Pichia pastoris***. *FEBS Lett* 1996, **394**:268-272.

41. Yang S, Kuang Y, Li H, Liu Y, Hui X, Li P, Jiang Z, Zhou Y, Wang Y, Xu A *et al.*: **Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1′ site**. *PLOS ONE* 2013, **8**:e75347.

42. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrell PM, Trinh YT, Zhang Q, Urbatsch IL *et al.*: **Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding**. *Science* 2009, **323**:1718-1722.

43. •• Fischer G, Kosinska-Eriksson U, Aponte-Santamaría C, Palmgren M, Geijer C, Hedfalk K, Hohmann S, de Groot BL, Neutze R, Lindkvist-Petersson K: **Crystal structure of a yeast aquaporin at 1.15 angstrom reveals a novel gating mechanism**. *PLoS Biol* 2009, **7**:e1000130.

The authors generated a *P. pastoris* knockout strain. This strain was subsequently used for the expression of the human AQP2 described in Ref. [15].

44. •• Hirz M, Richter G, Leitner E, Wriessnegger T, Pichler H: **A novel cholesterol-producing *Pichia pastoris* strain is an ideal host for functional expression of human Na,K-ATPase $\alpha 3\beta 1$ isoform**. *Appl Microbiol Biotechnol* 2013, **97**:9465-9478.

Description of a modified *P. pastoris* strain capable of producing cholesterol. The strain has potential for the production of a range of mammalian membrane proteins.

45. Vogl T, Thallinger GG, Zellnig G, Drew D, Cregg JM, Glieder A, Freigassner M: **Towards improved membrane protein production in *Pichia pastoris*: general and specific transcriptional response to membrane protein overexpression**. *New Biotechnol* 2014, **31**:538-552.

46. •• Lundstrom K, Wagner R, Reinhart C, Desmyter A, Cherouati N, Magnin T, Zeder-Lutz G, Courtot M, Prual C, André N *et al.*: **Structural genomics on membrane proteins: comparison of more than 100 GPCRs in 3 expression systems**. *J Struct Funct Genomics* 2006, **7**:77-91.

Detailed comparative analysis of *Escherichia coli*, *P. pastoris* and Semliki Forest virus infected mammalian cells as hosts for GPCR expression.

47. Asada H, Uemura T, Yurugi-Kobayashi T, Shiroishi M, Shimamura T, Tsujimoto H, Ito K, Sugawara T, Nakane T, Nomura N *et al.*: **Evaluation of the *Pichia pastoris* expression system for the production of GPCRs for structural analysis**. *Microb Cell Fact* 2011, **10**:24.

48. He Y, Wang K, Yan N: **The recombinant expression systems for structure determination of eukaryotic membrane proteins**. *Protein Cell* 2014, **5**:658-672.

49. • Nyblom M, Öberg F, Lindkvist-Petersson K, Hallgren K, Findlay H, Wikström J, Karlsson A, Hansson Ö, Booth PJ, Bill RM *et al.*: **Exceptional overproduction of a functional human membrane protein**. *Protein Expr Purif* 2007, **56**:110-120.

Describes very high functional expression level achieved for a human aquaporin in *P. pastoris*.

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

51. Molina DM, Wetterholm A, Kohl A, McCarthy AA, Niegowski D, Ohlson E, Hammarberg T, Eshaghi S, Haeggström JZ, Nordlund P: **Structural basis for synthesis of inflammatory mediators by human leukotriene C4 synthase**. *Nature* 2007, **448**:613-616.

52. Niegowski D, Kleinschmidt T, Olsson U, Ahmad S, Rinaldo-Matthis A, Haeggström JZ: **Crystal structures of leukotriene C4 synthase in complex with product analogs: implications for the enzyme mechanism**. *J Biol Chem* 2014, **289**:5199-5207.

53. Long SB, Campbell EB, Mackinnon R: **Crystal structure of a mammalian voltage-dependent shaker family K^+ channel**. *Science* 2005, **309**:897-903.

54. 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* 2010, **107**:11352-11357.

55. Tao X, Avalos JL, Chen J, MacKinnon R: **Crystal structure of the eukaryotic strong inward-rectifier K^+ channel Kir2.2 at 3.1 Å resolution**. *Science* 2009, **326**:1668-1674.

56. Hansen SB, Tao X, MacKinnon R: **Structural basis of PIP2 activation of the classical inward rectifier K^+ channel Kir2.2**. *Nature* 2011, **477**:495-498.

57. Tao X, Lee A, Limapichat W, Dougherty DA, MacKinnon R: **A gating charge transfer center in voltage sensors**. *Science* 2010, **328**:67-73.

58. Whorton MR, MacKinnon R: **X-ray structure of the mammalian GIRK2-$\beta\gamma$ G-protein complex**. *Nature* 2013, **498**:190-197.

59. Ho JD, Yeh R, Sandstrom A, Chorny I, Harries WEC, Robbins RA, Miercke LJW, Stroud RM: **Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance**. *Proc Natl Acad Sci U S A* 2009, **106**:7437-7442.

60. Nyblom M, Frick A, Wang Y, Ekvall M, Hallgren K, Hedfalk K, Neutze R, Tajkhorshid E, Törnroth-Horsefield S: **Structural and functional analysis of SoPIP2;1 mutants adds insight into plant aquaporin gating**. *J Mol Biol* 2009, **387**:653-668.

61. Li J, Jaimes KF, Aller SG: **Refined structures of mouse P-glycoprotein**. *Protein Sci* 2013, **23**:34-46.

62. Kodan A, Yamaguchi T, Nakatsu T, Sakiyama K, Hipolito CJ, Fujioka A, Hirokane R, Ikeguchi K, Watanabe B, Hiratake J *et al.*: **Structural basis for gating mechanisms of a eukaryotic P-glycoprotein homolog**. *Proc Natl Acad Sci U S A* 2014, **111**: 4049-4054.