Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Expression strategies for structural studies of eukaryotic membrane proteins

Joseph A Lyons[1,2,3,6], Azadeh Shahsavar[1,2,3,4,6], Peter Aasted Paulsen[3], Bjørn Panyella Pedersen[3,5] and Poul Nissen[1,2,3]

Integral membrane proteins in eukaryotes are central to various cellular processes and key targets in structural biology, biotechnology and drug development. However, the number of available structures for eukaryotic membrane protein belies their physiological importance. Recently, the number of available eukaryotic membrane protein structures has been steadily increasing due to the development of novel strategies in construct design, expression and structure determination. Here, we examine the major expression systems exploited for eukaryotic membrane proteins. Additionally we strive to tabulate and describe the recent expression strategies in eukaryotic membrane protein structural biology. We find that a majority of targets have been expressed in advanced host systems and modified from their wild-type form with distinct focus on conformation and thermostabilisation. However, strategies for native protein purification should also be considered where possible, particularly in light of the recent advances in single particle cryo electron microscopy.

Addresses
[1] Danish Research Institute of Translational Neuroscience (DANDRITE), Nordic EMBL Partnership for Molecular Medicine, Aarhus University, Aarhus, Denmark
[2] Centre for Membrane Pumps in Cells and Disease — PUMPkin, Danish National Research Foundation, Aarhus, Denmark
[3] Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark
[4] EMBL, Notkestrasse 85, 22607 Hamburg, Germany
[5] Aarhus Institute of Advanced Studies, Aarhus University, Denmark

Corresponding author: Nissen, Poul (pn@mbg.au.dk)
[6] These authors contributed equally.

Current Opinion in Structural Biology 2016, **38**:137–144

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

Edited by **Rob Meijers** and **Anastassiss Perrakis**

http://dx.doi.org/10.1016/j.sbi.2016.06.011

Introduction

Integral membrane proteins (IMPs) undertake a myriad of cellular functions, which range from sensory stimuli transduction and catalysis to transport and energy transduction. IMPs account for approximately 30% of all protein-coding genes in humans and are important targets for pharmaceuticals and drug development. The structures obtained to date reflect only a tiny fraction and generally represent the most tractable targets, that is, those expressed to high abundance either natively or through recombinant systems. However, the rate by which new structures of IMPs and in particular eukaryotic integral membrane proteins (eIMP) emerge is steadily increasing due in no small part to innovative protein expression, engineering and structure determination strategies developed over the last few years [1].

In this review, we compare and contrast the various expression systems available for eIMPs with a focus on the last 2 years (Supplementary Table 1). We extract recent advances and summarise the general principles emerging in the field of eIMP structural biology.

Expressions systems

Early structures of eukaryotic membrane proteins exploited the natural abundance of selected targets. Proteins isolated from native sources are a significant source of structural information (Supplementary Table 1, Figure 1a), and will likely remain so also in the future with the recent rise of single particle cryo-electron microscopy (cryo-EM) as typified by the structures of the inositol-1,4,5-trisphosphate (IP3) receptor [2], the ryanodine receptor [3••,4••] and the voltage-gated calcium channel Cav1.1 complex [5••]. For the latter two, the respective native proteins were purified by a pull-down purification approach using a recombinant GST-fused accessory protein [4••,5••]. CRISPR technology may also be helpful to introduce affinity tags or proteolytic sites into target proteins in a native background.

The majority of eIMPs are expressed at very low levels in native sources, so recombinant over-expression is required for structural and functional characterisation. The choice of an expression system for an eIMP target is generally dictated by first, the locally accessible expression systems, where already present knowledge and experience can significantly improve productivity, second, the nature and complexity of the chosen target, and third, cost. Typically, the expression of a range of

homologs representing a variety of species are screened for the best and most stable expressing proteins, although for drug development a close homology to the pharmacological target will be required. Indeed, for bacterial to mammalian expression systems, high-throughput pipelines exist for the evaluation of GFP fusion constructs based on expression and sample homogeneity using fluorescence size exclusion chromatography as reviewed elsewhere [6,7,8••].

Bacteria

The most common bacterial expression system, *Escherichia coli* (*E. coli*) is a rapid, versatile and cost effective expression system. The system has limitations due to different protein folding chaperones and a lack of essential lipids and post-translational modifications (PTMs) required for proper eIMP expression. Despite the many attempts to improve the expression of active eIMPs in *E. coli* including co-expression of molecular chaperones, tagging the target protein with a fusion protein and co-expression of post-translational machineries [9], *E. coli* remains a challenged system for the over-expression of eIMPs (Supplementary Table 1, Figure 1). Indeed, of the 11 eIMP structures derived from *E. coli* expression (Supplementary Table 1), over half were either fragments or required refolding.

Yeast

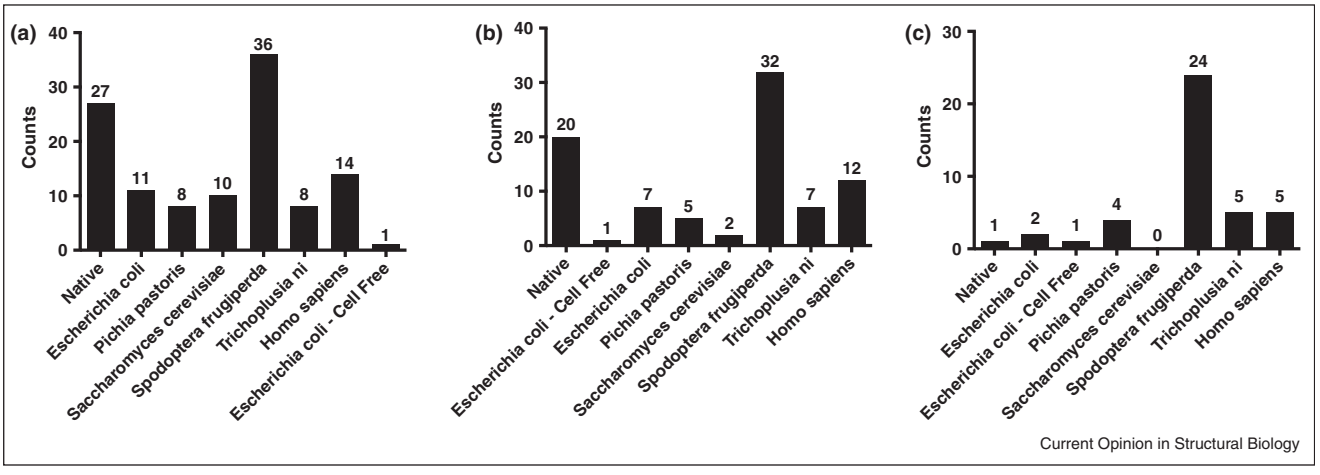
Yeast was the first successful system for recombinant expression of eIMPs for crystallographic studies [10,11]. The most common yeast strains for the over-expression of eIMPs for structural studies are *Saccharomyces cerevisiae* (*S. cerevisiae*) [7,11] and the methylotrophic *Pichia Pastoris* (*P. pastoris*) [12] (Supplementary Table 1). Both strains are cost effective eukaryotic expression systems capable of performing various PTMs including high mannose *N*-glycosylation. However, the absence of specific sterols might be one of the reasons that only a limited number of eIMP structures have been obtained using yeast expression systems (Figure 1b, c, Supplementary Table 1).

Expression in *P. pastoris* uses an integrated vector compared to the multicopy plasmid system in *S. cerevisiae*. The most common vector used in *P. pastoris* expression system, pPICZ, uses a promoter derived from the alcohol oxidase I (AOXI) gene, which is inducible by methanol and carries a simple drug-based selection system (Zeocin). Most vectors used in *S. cerevisiae* expression systems are propagated using the high copy 2-micron plasmid replication origin [13]. Recently, taking advantage of incorporating a defective *leu2-d* gene, in addition to the primary selection marker, has promoted an even higher plasmid copy number under leucine deficient growth conditions leading to improved expression levels [14–16]. Expression in *S. cerevisiae* mostly utilizes GAL1, a strong inducible promoter, which drives expression of the target gene following depletion of glucose and addition of galactose as the carbon source during culturing.

Insect cell

Baculovirus transduction of insect cells is the dominant heterologous expression system for the production of eIMPs, especially mammalian/human targets that have yielded structures over the past 2 years including a surge of G-protein coupled receptors (Figure 1). The gene of interest is cloned into the pFastBac vector, which controls the expression of the target gene by either the strong *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) polyhedrin (PH) or p10 promoters. The vector is then used to produce a recombinant baculovirus shuttle vector (bacmid) [17] to transfect the insect cells

Figure 1



Distribution of expression systems for **(a)** eukaryotic, **(b)** mammalian, and **(c)** human integral membrane proteins over the past 2 years. Values for each entry are indicated over the respective bar. See Supplementary Table 1 for a detailed description of selection criteria.