International Tables for Crystallography (2012). Vol. F, Section 3.2.2, pp. 92–93.

Chapter 3.2. Expression and purification of membrane proteins for structural studies

J. A. Ernst, D. G. Yansura and C. M. Koth

3.2.1. Introduction

Integral membrane proteins constitute about a third of the proteome of most organisms but less than 1% of all entries in protein structural databases (Berman *et al.*, 2002). This disparity is largely due to inherent difficulties in their expression, solubilization and purification. The production of sufficient protein for structural studies can be challenging for any target, but several obstacles are unique to membrane proteins. For example, proper insertion into the membrane relies on host cellular machinery that may be limiting or incompatible. Exceeding this capacity can lead to cell death or the accumulation of aggregated and inactive protein within the cell (Geertsma *et al.*, 2008). Also, post-translational modifications such as glycosylation, acylation and sulfation may not be faithfully reproduced (Grisshammer & Tate, 1995). As a further complication, even if sufficient expression can be achieved, most membrane-protein structural studies require that the target be extracted from the cellular membrane using detergents. However, detergents can adversely affect protein structure and function, as well as influence the outcome of crystal trials (Engel *et al.*, 2002; Lemieux *et al.*, 2003; Prive, 2007). Fortunately, if a strategy for purifying sufficient quantities of a given membrane-protein target *can* be established, crystallization strategies largely mimic the standard techniques for soluble proteins (Newby *et al.*, 2009).

Despite the aforementioned challenges, high-resolution structures of almost 200 unique membrane proteins have now been solved (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html), the vast majority using protein produced by recombinant methods (Willis & Koth, 2008). The rate of new structure determinations has also increased dramatically over the last few years, mimicking the exponential growth of soluble protein structures in the early 1980s. Space limitations here preclude a thorough review of all possible membrane-protein production methods. However, an examination of the successful expression, solubilization and purification strategies that have led to membrane-protein structures reveals that, in many cases, remarkably similar methods have been used (Carpenter *et al.*, 2008; Willis & Koth, 2008; Newby *et al.*, 2009). Guided by these methods, the following sections detail a rational and consensus 'first-attempt' strategy that has worked for a broad range of membrane targets with only minor variations in technique (Dobrovetsky *et al.*, 2005, 2007; Lunin *et al.*, 2006). It must be noted that while this represents an evidence-guided approach, the methods provided herein will not work for every membrane protein and, in fact, will fail for many. Thus, prioritized lists of alternative strategies are provided for those targets in which the initial expression or isolation attempts do not succeed or problems are encountered.

3.2.2. A consensus strategy for membrane-protein expression

Historically, the likelihood of successful structure determination has been high for those membrane proteins that can be isolated from readily available abundant natural sources (Sakai & Tsukihara, 1998). Unfortunately, most membrane proteins do not meet this criterion. In reality, the vast majority of high-resolution membrane-protein structures are of prokaryotic targets, expressed in *Escherichia coli* by recombinant methods (Willis & Koth, 2008; Newby *et al.*, 2009). The reasons for this are simple. Attempting recombinant expression in *E. coli* is inexpensive, flexible, simple and easily scaled-up, and many constructs and strains can be screened quickly. As with crystallographic efforts for many soluble proteins (Gräslund *et al.*, 2008), most successful membrane-protein endeavours also commonly screen multiple constructs and/or orthologues for any given target (see, for example, Chang *et al.*, 1998; Doyle *et al.*, 1998), since the greater the number of unique constructs screened, the greater the chance that one will be successfully isolated and crystallized. Given the overwhelming use of *E. coli* for successful membrane-protein structures, a 'first-pass' expression strategy suitable for almost any membrane protein is clear: attempt expression in *E. coli* and, whenever possible, screen multiple constructs, orthologues and strains.

3.2.2.1. Choosing the expression system and affinity tags

In addition to the choice of expression host, one must also consider the expression system and the type and placement of affinity tags or fusion proteins. For *E. coli*, IPTG-inducible T7 polymerase-driven expression systems, such as those based on pET vectors (Studier & Moffatt, 1986) and λDE3 lysogen strains, are the most widely used for membrane proteins, as is the case for soluble targets. Also, it has generally been observed that for most targets, protein expression is optimal at lower temperatures (*i.e.* <20 °C; Christendat *et al.*, 2000; Wang *et al.*, 2003; Dobrovetsky *et al.*, 2005). When structural studies are the desired outcome, the most common tagging strategy for membrane proteins is to engineer a stretch of at least six histidine residues at the amino or carboxyl terminus of the target constructs; this is used for >80% of successful targets (Willis & Koth, 2008). If expression levels are sufficient (*i.e.* $\geq$0.05 mg g^{-1} cell paste), this often permits purification using a general two-step procedure: capture of the tagged protein by immobilized metal affinity chromatography (IMAC), followed by size-exclusion chromatography (SEC) (see below). Remarkably, this basic approach has proved successful for the crystallization of many membrane proteins, as discussed in the following section. If premature termination is observed during expression, the engineering of a carboxyl-terminal tag will ensure that these proteins are not isolated during purification. Also, extending the stretch of histidines to greater than six residues can improve the retention of membrane proteins on immobilized metal affinity resins. This can prove particularly useful, given the modest expression levels and reduced chromatographic resolution and recovery of many membrane targets (Dobrovetsky *et al.*, 2005; Eshaghi *et al.*, 2005; Surade *et al.*, 2006; Lewinson *et al.*, 2008). Other affinity tags are viable options, but are more rarely used. For example, although fusion proteins such as glutathione S-transferase (GST) and thioredoxin are widely used to promote expression and/or simplify the purification of soluble proteins [for an extensive comparison of various affinity tags, see Lichty *et al.* (2005)], they are rarely cited in expression strategies for polytopic membrane proteins.

Table 3.2.2.1
Strategies for improving recombinant membrane-protein expression in *E. coli*

Troubleshooting expression problems: if the 'first-pass' expression strategy is unsuccessful, then some changes in the expression construct, host strain or induction conditions may be helpful. The symptoms of the problem often give a clue as to how to make these corrections.

Symptom	Solution
E. coli colonies small after transformation into expression strain	Need tighter promoter such as phoA, araC *etc.*
Unprocessed signal sequence†	If none is used in native situation, remove from construct† If one is present in native situation, use an expression host co-translational signal sequence† If still unprocessed, lower translation initiation strength
Mostly high-molecular-weight aggregates	Decrease induction time
Rapid cell growth after induction with low expression levels	Increase translation initiation strength
Poor cell growth after induction with low expression levels (*a*) Accumulation maximum after only 1–3 h (*b*) Accumulation increases out to 8–10 h post induction	Decrease translation initiation rates or plasmid copy number Usually successful: (*a*) Re-synthesize the gene, removing rare codons and optimizing codon pairs† (*b*) Convert transmembrane and intracellular domain cysteines (normally reduced) to serines† (*c*) Try different non-DE3 lysogen strains after switching to a non-T7 promoter (i) Cytoplasmic protease deletions (ii) Periplasmic protease deletions (iii) Heat-shock htpR^{TS} Sometimes successful: (*a*) Try different C-terminal tags such as poly-his, Flag *etc.*† (*b*) Try different temperatures of induction such as 20, 25, 30 °C

† These methods are also useful for optimizing eukaryotic gene expression.

3.2.2.2. Strategies for improving membrane-protein expression

Low expression levels are common for many membrane proteins. Before testing alternative hosts, it may be worthwhile first to screen alternative *E. coli* expression systems, such as those using the tightly controlled arabinose promoter (Guzman *et al.*, 1995). Also, other *E. coli* strains can be tested. For example, the mutant C41 and C43 'Walker' expression strains (Miroux & Walker, 1996; Wagner *et al.*, 2008) yield higher expression levels for some membrane proteins. Table 3.2.2.1 summarizes common strategies for improving the expression of membrane proteins in *E. coli*. If reasonable efforts in *E. coli* still do not yield sufficient expression, then an alternative expression host should be considered. The next most successful is yeast. Of the 22 recombinant eukaryotic membrane proteins whose structures have been solved by crystallography, half were produced using *Pichia pastoris* (seven targets), *Saccharomyces cerevisiae* (three targets) or *Saccharomyces pombe* (one target) (Raman *et al.*, 2006). Membrane proteins expressed in yeast are typically fused to an amino-terminal host signal sequence to promote proper membrane targeting. The use of higher eukaryotic hosts, such as insect and mammalian cells, is also increasing as more groups attempt to crystallize recombinant eukaryotic targets. The recent high-resolution structures of several G-protein-coupled receptors are notable examples of targets expressed successfully in insect cells (Cherezov *et al.*, 2007; Rasmussen *et al.*, 2007; Hanson *et al.*, 2008; Jaakola *et al.*, 2008; Warne *et al.*, 2008).

To date, there is only one report of a structure determined from cell-free synthesis of the target membrane protein (EmrE; Chen *et al.*, 2007). However, for several reasons, cell-free approaches should be explored as a complement to traditional expression methods. Firstly, the absence of a host cell negates toxicity problems for some target proteins. Secondly, many targets that do not express in cells can be produced cell-free (Ishihara *et al.*, 2005; Klammt *et al.*, 2007; Savage *et al.*, 2007). Thirdly, these systems allow for considerable flexibility with respect to reaction conditions and components. Cell-free systems for prokaryotic and eukaryotic hosts have been described (Schwarz *et al.*, 2008), some of which function in the presence of solubilizing detergents (Ishihara *et al.*, 2005; Klammt *et al.*, 2007; Savage *et al.*, 2007) or lipids (Kuruma *et al.*, 2005; Schwarz *et al.*, 2007), facilitating proper folding and downstream purification of membrane-protein targets.

Many eukaryotic proteins have undergone extensive construct engineering to improve expression and/or crystallization outcomes. These include fusion to T4 lysozyme, mutation of glycosylation sites, deletion of potentially disordered regions and/or targeted evolution to identify conformationally stable mutants (Long *et al.*, 2005; Cherezov *et al.*, 2007; Rosenbaum *et al.*, 2007; Jaakola *et al.*, 2008; Magnani *et al.*, 2008; Serrano-Vega *et al.*, 2008; Warne *et al.*, 2009). Also worth mentioning is the observation that incorporation of an amino-terminal 'rhodopsin tag' comprising the first 20 amino acids of bovine rhodopsin (amino-acid sequence MNGTEGPNFYVPFSNKTGVV) has been found to boost dramatically the expression levels of eukaryotic membrane targets expressed in insect and mammalian cells (Krautwurst *et al.*, 1998).

3.2.3. A consensus strategy for membrane-protein purification

3.2.3.1. General principles

The purification of an expressed integral membrane protein for structural studies typically involves four discrete steps: cell lysis, membrane isolation, detergent extraction and chromatographic separation. In this section, an evidence-based consensus strategy is outlined which aims to serve as a starting point for membrane-protein purification. Detailed protocols are outlined in Fig.