Current strategies for protein production and purification enabling membrane protein structural biology

Aditya Pandey^a, Kyungsoo Shin^a, Robin E. Patterson^a, Xiang-Qin Liu^a, and Jan K. Rainey^{a,b,*}

^aDepartment of Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

^bDepartment of Chemistry, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada

Abstract

Membrane proteins are still heavily underrepresented in the protein data bank (PDB) due to multiple bottlenecks. The typical low abundance of membrane proteins in their natural hosts makes it necessary to overexpress these proteins either in heterologous systems or through *in vitro* translation/cell-free expression. Heterologous expression of proteins, in turn, leads to multiple obstacles due to the unpredictability of compatibility of the target protein for expression in a given host. The highly hydrophobic and/or amphipathic nature of membrane proteins also leads to challenges in producing a homogeneous, stable, and pure sample for structural studies. Circumventing these hurdles has become possible through introduction of novel protein production protocols; efficient protein isolation and sample preparation methods; and, improvement in hardware and software for structural characterization. Combined, these advances have made the past 10–15 years very exciting and eventful for the field of membrane protein structural biology, with an exponential growth in the number of solved membrane protein structures. In this review, we focus on both the advances and diversity of protein production and purification methods that have allowed this growth in structural knowledge of membrane proteins through X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM).

Keywords

Membrane protein; structural biology; stable isotope enrichment; protein expression and purification; fusion proteins

Introduction

Membrane proteins play crucial roles in a wide variety of cellular functions, including regulation of ion transport across the membrane; sensing and transmitting chemical and electrical signals; mediating cellular attachment; and, controlling membrane lipid composition (von Heijne 2007). The importance of membrane proteins is also highlighted by

*Corresponding author (jan.rainey@dal.ca).

the facts that they represent ~26% of the human proteome (Fagerberg et al. 2010) and that >50% of currently marketed drugs target this class of proteins (Arinaminpathy et al. 2009; Overington et al. 2006; Yildirim et al. 2007). Membrane proteins can arguably be referred to as the 'Holy Grail' in the field of structural biology. Correspondingly, despite their importance and relative prevalence, <2% of the experimentally determined structures in the protein data bank (PDB) are of membrane proteins (mpstruc database (Table 1) relative to the PDB). Characterization of the structure and dynamics of these proteins not only provides insight into their mechanisms of function, but also aids in rational design of novel drugs (Lounnas et al. 2013; Vinothkumar and Henderson 2010).

Over the past few decades, X-ray diffraction (XRD) and nuclear magnetic resonance spectroscopy (NMR) have been the predominant techniques for atomic resolution structural determination, with electron crystallography and microscopy increasing in prevalence primarily through advances in cryo-electron microscopy (cryo-EM) instrumentation and methodologies. By 2000, there were 26 (9 by NMR) unique integral membrane structures in the PDB (Bowie 2000). Studies to this point often relied upon proteins found naturally at high abundance, meaning that a relatively small number of proteins were accessible to structural studies. Notable early examples include bacteriorhodopsin (Henderson and Unwin 1975) and light harvesting complex proteins (Deisenhofer et al. 1984); channels such as the β -barrel porins (Weiss et al. 1990) or the α -helical aquaporin (Cheng et al. 1997; Li et al. 1997; Walz et al. 1997) and KcsA potassium channel (Doyle et al. 1998); respiratory enzymes (Tsukihara et al. 1995); the G-protein-coupled receptor rhodopsin (Palczewski et al. 2000); and, the sarcolemmal Ca^{2+} -ATPase (Toyoshima et al. 2000).

In the past 15 years, improvements in methods for protein expression, solubilization, and stabilization have led to a tremendous increase in the number of solved membrane protein structures (Pedersen and Nissen 2015). As of now, there are ~587 unique membrane protein structures solved, including >130 by NMR spectroscopy (see Table 1 for useful databases). Notably, in contrast to early studies relying upon natural sources, at least 70% of currently available structures in the PDBTM (Kozma et al. 2013) are of recombinantly expressed proteins (Figure 1).

The clear disparity in the degree of structural information available for membrane proteins relative to soluble proteins stems mainly from their hydrophobicity and low levels of natural abundance, which produce obstacles in the path of studying membrane proteins at all steps from protein expression through to sample preparation. This review focuses on the significant progress made in expression and purification of membrane proteins, relating this to considerations in sample preparation for structural characterization of membrane proteins using XRD, NMR or cryo-EM, with an underlying goal of comparing and contrasting a wide variety of successful strategies.

Expression hosts

Structural studies typically require milligram quantities of protein. Given that membrane proteins are generally scarce in their natural environment, heterologous expression of these proteins has become the predominant method (Figure 1) to obtain enough protein for

characterization (Bill et al. 2011). However, heterologous expression in the host of choice is not always feasible due to the often toxic nature of a membrane protein upon overexpression in that host (Bernaudat et al. 2011). In addition, discrepancies in the levels of expression from one membrane protein to the next are not yet completely understood (Grisshammer 2006). Thus, it is not currently possible to predict the degree of toxicity of a target protein, making its relative expression in a given host still a matter of trial and error.

Correspondingly, a variety of expression hosts have been used successfully for expression of target membrane protein. These differ in their ability to carry out post-translational modifications (PTM), in level of expression, and in sensitivity to modification of culture conditions, which can dictate the selection of one host over the other (summarized in Table 2). In this section, we briefly discuss the advantages and limitations of the most commonly used expressions systems, with Figure 2 demonstrating the relative usage of each for membrane proteins.

Escherichia coli

E. coli is the most widely used host system for expression of recombinant proteins, with membrane proteins being no exception (Figure 2), for five main reasons. (i) *E. coli* has fast growth kinetics with a doubling time of 20 min in optimal conditions (Sezonov et al. 2007). This reduces the time required for expression of the protein of interest (POI). (ii) Cells can be grown to high density to achieve very high protein production (Shiloach and Fass 2005). (iii) Growth medium is inexpensive and can be manipulated without a significant loss in yield, enabling enrichment with stable isotopes for NMR (Marley et al. 2001; Tyler et al. 2005) and selenomethionine for XRD (Sreenath et al. 2005; Stols et al. 2004). (iv) Extensive knowledge of genetics, physiology, and metabolism has enabled intelligent genetic manipulations to produce strains with specific advantages (e.g., (Andersen et al. 2013; Sorensen and Mortensen 2005; Studier 1991; Wagner et al. 2008)). (v) An abundance of expression vectors (https://www.embl.de/pepcore/pepcore_services/cloning/choice_vector/ecoli/index.html) enables easy incorporation of desired properties in the plasmid vector.

T7 RNA polymerase (T7RNAP) from *Enterobacteria phage λ* is often used to drive recombinant protein production in *E. coli* (Studier 1991). T7RNAP exclusively recognizes the T7 promoter and synthesizes RNA several times faster than *E. coli* RNA polymerase with fewer observed incomplete transcripts due to premature termination, allowing higher protein yield (Iost et al. 1992). Typically, T7RNAP is employed under control of the *lac* operon, where addition of isopropyl β-D-thiogalactoside (IPTG) induces expression of T7RNAP and subsequently enables overexpression of the POI (Studier 1991). Since *E. coli* can grow on minimal medium, rich medium can be replaced with a medium containing isotope-enriched precursors or amino acids before inducing expression, enabling cost effective production of proteins amenable to heteronuclear NMR characterization (Marley et al. 2001).

Recently, auto-induction of membrane protein expression has been suggested to be more effective than traditional T7RNAP IPTG-induced expression (Gordon et al. 2008). In this method, protein production is induced by lactose in the medium, which only occurs upon the depletion of metabolites such as glucose. Although the auto-induction protocol seems to

provide an edge over the traditional IPTG induction protocol, its cost-effectiveness is somewhat compromised in producing isotope-enriched proteins for NMR (Tyler et al. 2005). For additional background, the reader is referred to detailed reviews of various aspects of recombinant protein expression using *E. coli* expression systems (Baneyx 1999; Makrides 1996; Rosano and Ceccarelli 2014; Sorensen and Mortensen 2005; Stevens 2000).

The prokaryotic *E. coli* system also has many limitations. Specifically, this organism often lacks the essential lipids, molecular chaperons, and machinery for PTMs required for correct membrane insertion and eukaryotic protein folding. It is important to note that efforts have been made to overcome these barriers by methods such as codon-optimization (Burgess-Brown et al. 2008); addition of fusion tags for a variety of purposes (detailed below); and, co-translation/induction of post-translational machineries to facilitate protein folding (de Marco et al. 2005). However, use of *E. coli* still leads to many challenges in order to be an ideal expression system for a membrane POI. For further detail, Miroux and co-workers (Hattab et al. 2015) have recently provided an extensive analysis of the impact of *E. coli* expression systems in the field of membrane protein structural characterization.

Yeast

Yeast are next in popularity to *E. coli* (Figure 2) and exhibit many similar advantages for heterologous expression. For example, yeast also have rapid growth rate, the ability to be grown to high density, well-studied genetics, and readily available advanced tools for genetic manipulation. However, yeast cells have the added advantage of being capable of performing various eukaryotic PTMs. These modifications include proteolytic processing of signal peptide sequences; disulfide-bond formation; prenylation; phosphorylation; acylation; and, certain types of *O*- and *N*-linked glycosylation that may be essential for activity, correct folding, and membrane insertion of proteins (Boer et al. 2007). It is important to note that glycosylation patterns vary between different yeast strains, with some strains engineered to produce more uniform and human-like glycosylation patterns (Hamilton et al. 2003; Hamilton et al. 2006; Vervecken et al. 2007). These factors make yeast an inexpensive and efficient alternative to prokaryotic expression systems for production of membrane proteins. *Saccharomyces cerevisiae* and *Pichia Pastoris* (*Pichia*) are the most commonly used species for expression of membrane proteins (Bornert et al. 2012; Joubert et al. 2010). *Schizosaccharomyce pombe* is less frequently used, but sometimes outperforms other yeast species in expression of mammalian membrane proteins (Sander et al. 1994b; Takegawa et al. 2009).

Using *S. cerevisiae*, a model eukaryote with a long history of study, comes with a few perks. These include a completely sequenced genome, a variety of genetic manipulation tools, and a large number of available strains available for expression of eukaryotic membrane proteins (Drew et al. 2008; Goffeau et al. 1996; Li et al. 2009; Winzeler et al. 1999). In comparison to other strains of yeast, protocols are available for high-throughput expression of membrane proteins (Newstead et al. 2007). The ability to carry out *in vivo* homologous recombination also makes *S. cerevisiae* an attractive host to produce membrane protein mutants in a high-throughput fashion to optimize expression for downstream applications (Ito et al. 2008).

Despite these advantages, the methylotrophic yeast *Pichia* has been the most frequently and successfully applied membrane protein expression host among all the yeast strains (Figure 2). A *Pichia* expression system was first released for academic use in 1993, following which there has been an exponential increase in the knowledgebase and the number of membrane proteins expressed using this system (Cregg et al. 2000; Ramon and Marin 2011). A key physiological trait of *Pichia* is its strong preference for respiratory growth, facilitating culturing at higher densities than fermentative yeast. Its methylotrophic nature also means that it can derive energy by metabolism of methanol using enzymes such as alcohol oxidase (AOX) and dihydroxyacetone synthase (DHAS) (Stewart et al. 2001). Since the promoters of methanol utilization pathways are very efficient and tightly regulated, they are most commonly used for expression of heterologous proteins (Hollenberg and Gellissen 1997). Using this system, very high expression yields have been achieved for membrane proteins (e.g., 90 mg·L⁻¹ for human aquaporin (Nyblom et al. 2007)). Similar to *E. coli*, *Pichia* can grow on minimal medium that can be modified to produce isotope-enriched protein required for structural characterization using NMR (Denton et al. 1998; van den Burg et al. 2001; Wood and Komives 1999). Although the cost of production of an isotope-enriched protein may be higher compared to *E. coli*, *Pichia* is still a very attractive alternate host for production of isotope-enriched proteins that cannot be efficiently produced in *E. coli* (Clark et al. 2015).

Baculovirus/insect cells

Third in popularity (Figure 2), and providing many advantages, are baculovirus/insect cell expression systems. In comparison to other virus-based methodologies, baculovirus is safe due to its inability to infect mammals. Similarly to other eukaryotic expression systems, baculovirus/insect cell expression of heterologous genes typically permits proper protein folding as a consequence of PTMs that are often identical to those that occur in higher eukaryotes (Shi and Jarvis 2007). Cell cultures from the insects *Spodoptera frugiperda* (Sf9, Sf21) and *Trichoplusia ni* (Hi5) infected by the baculovirus *Autographa californica* multi-nucleopolyhedrovirus (AcMNPV) are the most commonly used systems for membrane protein expression (Hitchman et al. 2009). As in all expression systems, the heterologous expression of proteins varies in yield when expressed in different types of insect cells (Unger and Peleg 2012).

A typically employed commercially available system for production of recombinant baculovirus for expression of proteins in insect cells is the BAC-to-BAC® (Invitrogen by life technologies, Waltham, MA, USA) system. In this protocol, the target gene is sub-cloned into the pFastBac vector, which is then transformed into DH10Bac *E. coli* competent cells harbouring a baculovirus shuttle vector (bacmid) with a transposon site and a helper plasmid. Transposition of the target gene onto the bacmid gives rise to a recombinant bacmid that can be purified and used for production of recombinant virus that can be later used for infection of insect cells for testing protein expression. This entire process takes approximately 3–4 week to reach the stage of testing protein expression. Recently, a faster and more convenient method employing transient transfection was designed for a rapid screening of membrane protein expression (Chen et al. 2013).

To date, baculovirus-insect cell expression has given the highest number of eukaryotic membrane proteins for structural characterization (He et al. 2014). It is, however, relatively expensive to produce proteins by this method in comparison to the systems discussed earlier. Moreover, due to complex metabolism and inability to grow in minimal medium, it is currently prohibitively expensive to produce isotope-enriched proteins for structural characterization by NMR (Gossert and Jahnke 2012; Sitarska et al. 2015). Notably, an economical approach has recently been put forward for expression of isotope-enriched proteins using a homemade isotope-enriched yeast extract (Opitz et al. 2015). Given the reported success of this system, it is highly appealing to devise strategies to reduce cost over currently available protocols.

Mammalian cells

Membrane proteins that require specific PTMs and subcellular environments for proper folding and activity are often impossible to produce in functional form by prokaryotic or lower eukaryotic systems. Interestingly, the difficulty in competent protein production has been stated to not be proportional to the number of transmembrane segments or to the size of the protein; rather, it appears to arise from the complexity of folding process for the protein in question (Tate 2001). It is likely that a combination of various factors decides the fate of protein expression in a specific expression host. Therefore, mammalian cell lines seem to be an obvious choice to obtain functional mammalian proteins. However, these systems also have drawbacks.

One of the main drawbacks with mammalian cell systems is that levels of expression relative to the systems already introduced are typically very low, even under conditions of induced overexpression, making it prohibitively expensive to produce sufficient protein for structural studies (Andrell and Tate 2013). Chinese hamster ovary cells (CHO), human embryonic kidney cells (HEK293), baby hamster kidney cells (BHK-21) and monkey kidney fibroblast cells (COS-7) are the most commonly used cell lines for expression (Andrell and Tate 2013). These different cell lines can vary significantly in terms of expression level depending on the POI, requiring individual optimization for each POI.

Mammalian cell line-based expression can also be performed either by transient transfection or by making stable cell lines. Transient transfection can be either carried out using recombinant viruses (Dukkipati et al. 2008; Hassaine et al. 2006; Matrai et al. 2010) or using chemical agents (Geisse and Fux 2009). Although more time consuming, stable transfection provides a means of producing heterologous proteins with high reproducibility and expression level (Camponova et al. 2007; Chelikani et al. 2006). It is also important to note that there are engineered cell lines that can have beneficial outcome depending on the POI. For example, the HEK293S-TetR cell line, with a tetracycline-inducible expression system (Reeves et al. 2002b), allowed crystallization of a mutant form of rhodopsin that could not be expressed in a constitutive expression system, likely due to the toxic nature of the POI. Cell lines such as CHO Lec.3.2.8.1 (Stanley 1989), HEK293S GnTI(–) (Reeves et al. 2002a), and HEK 293S Lec36 (Crispin et al. 2009) have mutations that lead to modified levels of N-glycosylation and/or decreased complexity, which can lead to improved crystallization (Deupi et al. 2012). To date, due to complex metabolism and prohibitively

expensive expression media, a very limited number of proteins have been expressed using mammalian expression system with isotope-enrichment, illustrating the current challenge of applying this class of expression system for NMR (Sastry et al. 2012).

Cell-free expression systems

In all of the conventional cell-based expression systems, the focus is on overexpression, which often results in improper protein folding and membrane insertion. In the case of membrane proteins, overexpression also frequently leads to formation of protein aggregates, destabilization of cellular membranes, and interaction with host metabolic machinery leading to cytotoxicity (Klammt et al. 2006). Therefore, cell-free expression systems (also referred to as *in vitro* translation systems) were developed to bypass the complexities and sensitivities inherent in dealing with living cells (Nirenberg and Matthaei 1961). In recent years, cell-free expression has gained in popularity for the expression of membrane proteins.

Typically, a cell-free expression system is an open system enabling *in vitro* protein translation with the help of translation machineries provided by the cell lysates from various organisms (Schwarz et al. 2008). A major attractive feature of cell-free systems for membrane protein expression is the ability to implement a carefully chosen folding environment. This allows for better control over the driving forces of membrane protein folding such as lipid composition (Phillips et al. 2009); hydrophobic mismatch (Cybulski and de Mendoza 2011); membrane lateral pressure and curvature (Botelho et al. 2006; Marsh 2007); membrane elasticity (Lundbaek 2006; Lundbaek et al. 2010); and, cholesterol content (Chini and Parenti 2009; Pucadyil and Chattopadhyay 2006). Cell-free expression systems exhibit suppressed interconversion of amino acids, enabling a wide range of isotope-enrichment patterns that can be exploited in various NMR approaches to solve the signal overlap problem that is common in membrane protein structural characterization (Junge et al. 2011; Ozawa et al. 2006; Parker et al. 2004; Reckel et al. 2011). Cell-free systems also allow production of proteins with Stereo-Array Isotope Labeling (SAIL) (Kainosho et al. 2006). SAIL is a method of producing proteins with amino acids having certain stereospecific patterns of isotopes optimal for protein NMR analysis, particularly for structural characterization of large proteins by NMR (Takeda et al. 2008; Tonelli et al. 2011).

Currently, several types of cell-free system are available, based upon extracts from different sources, such as: *E. coli* (Schwarz et al. 2007), wheat germ (Harbers 2014), rabbit reticulocyte (Anastasina et al. 2014), *Spodoptera frugiperda* (Ezure et al. 2014), CHO cells (Brodel et al. 2014), mouse embryonic fibroblasts (Zeenko et al. 2008), and the HeLa cell line (Mikami et al. 2006). Among these, *E. coli* lysate is currently the most widely, and successfully, applied cell-free expression medium. Variations have been developed, such as the Protein synthesis Using Recombinant Elements (PURE) system (Shimizu et al. 2001) containing a minimal set of purified elements required for the translation reaction, and the Cytomim system (Jewett et al. 2008) where, besides the *E. coli* cell extract, inverted inner membrane vesicles from *E. coli* are also added. Recently, the PURE system has been optimized for rapid production of high quantities of functional membrane proteins by adding liposomes to the reaction mixture (Kuruma and Ueda 2015). Although relatively expensive

in terms of the required isotope-enriched precursors, cell-free expression is a very attractive technique for production of isotope-enriched proteins for structural characterization by NMR. It should also be noted that although production of properly folded membrane proteins has not yet been streamlined in cell-free expression systems, this class of expression has a great deal of potential, especially when it comes to production of membrane proteins for NMR.

Use of fusion proteins - advantages and limitations

As stated earlier, recombinant proteins can be difficult to express and purify. One of the most common strategies to solve this problem is to fuse a proteinaceous tag to the protein of interest (POI), as detailed in Table 3. Such additions have been shown to improve the yield of the target proteins in *E. coli* through increased expression (Kefala et al. 2007; Leviatan et al. 2010; Pandey et al. 2014), allowing for greater potential yield. Another strategy to increase the yield of membrane proteins and other difficult-to-express proteins is to specifically target them to inclusion bodies using insoluble fusion partners, shielding them from proteolytic degradation and preventing other perturbations to cell function (Hwang et al. 2014). Some tags increase solubility (Butt et al. 2005; Marblestone et al. 2006) to improve handling during subsequent purification and/or to enhance purification through affinity chromatography methods (Hochuli et al. 1988). Tags may also be extremely valuable to facilitate crystallization for XRD (Engel et al. 2002; Cherezov et al. 2007).

In the case of membrane proteins, targeted localization of an aggregate-prone and hydrophobic POI to the expression medium or periplasm has also been shown to improve yield (Better et al. 1993). Periplasmic targeting may also be employed to produce disulfide-bonded recombinant proteins (de Marco 2009). In some instances, inner membrane targeting of the POI through fusion tags has proven successful for membrane POI overexpression (Neophytou et al. 2007; Tiralongo and Maggioni 2011). Localization to the outer membrane may also be targeted [Jung, 1998 #280; Francisco, 1992 #281], although this is typically used for surface display of soluble proteins (van Bloois et al. 2011). Fusion tags composed of native yeast signal sequences are also used to target the membrane in yeast expression systems (King et al. 1990; Weiss et al. 1995).

The benefits of fusion tags may potentially be both additive and complementary, meaning that they are often fused together in various arrangements to improve yield (Costa et al. 2013a; Liu et al. 2015; Pryor and Leiting 1997), including for isotope-enriched membrane proteins for NMR (Kohn et al. 1998; Leviatan et al. 2010). Given the difficulties in expressing mammalian (and non-mammalian) membrane proteins in heterologous systems such as *E. coli*, use of fusion tags is usually essential in producing sufficient quantities for structural studies.

Although fusion tags have many advantages, the design of fusion tag-POI constructs demand careful planning with consideration of the exact construct in question. One property of the tag itself that calls for careful selection is its size. Larger tags tend to impose a heavy metabolic burden on the host system, which increases with the complexity of the amino acid composition (Table 3). Use of large tags can also lead to interactions between the fusion tag

and the POI. This phenomenon can be favourable for aggregate-prone proteins, since larger tags such as NusA inhibit the aggregation and promote solubility (Davis et al. 1999). However, these interactions will likely interfere with structure and/or activity (Majtan and Kraus 2012; Sabaty et al. 2013; Singh et al. 2013). This leads to another factor that must be carefully optimized: the length of the linker between the tag and the POI. This linker region must be long enough to prevent unfavourable tag-target interaction and to allow for proper folding of the target protein (Raran-Kurussi and Waugh 2012). Unfortunately, the linker region can also impede structural characterization, since its flexibility can interfere with crystallization. A solution for many of these issues is removal of the tag through various means, as detailed below. An additional factor with direct relation to selection of tag(s) is the optimal purification protocol, where success can vary in the presence or absence of various salts, chelating agents, and/or detergents. This is discussed in greater detail below.

In short, use of fusion tags can cause a great deal of grief due to the complexities, required optimization of conditions, and, nearly limitless number of possible arrangements for fusion tags. However, the advantages afforded are often significant. Given their utility in protein expression upon optimized design, fusion tags are likely to remain a key factor in membrane protein production.

Tag removal strategies

Although fusion tags are highly useful, they often must be removed from the POI prior to structural studies, since they can affect both structure and function of the target protein. This leads to an additional layer of complexity in recombinant protein production and is often the bottleneck in such a procedure. As in the choice of the tags themselves (Table 3), there is a wide variety of choice for tag removal strategies. Possibilities include protease-mediated cleavage, self-cleavage, chemical cleavage, and *in vivo* cleavage. Each of these methods may take advantage of a number of different reagents/proteases to cleave the tag (Table 4 lists a variety of commercially available proteases that have been employed with membrane proteins). With the ongoing introduction of new reagents and proteases, the already wide array of potential tag-target fusion designs continues to increase (Marino et al. 2012).

Further complicating this issue, the efficiency of tag removal can vary dramatically between different target proteins. For example, steric hindrance or unfavourable residues at cleavage sites can hamper protease-mediated tag removal. This may be circumvented by use of a different protease with different active site requirements; lengthening of the linker region; use of a smaller tag; or, through use of chemical cleavage methods involving reagents such as cyanogen bromide (CNBr) or 2-nitro-5-thiocyanatobenzoic acid (NTCB), which cleave C-terminal to Met and Cys, respectively (Crimmins et al. 2005). In addition, different methods present different difficulties. For example, off-target cleavage can occur from proteases that recognize specific short amino acid motifs. Some proteases also leave residual amino acids on the target protein, thus increasing the protein size (Table 4) and, potentially, perturbing function. The removal of a tag can also yield improperly folded or disrupted target protein structures due to protein aggregation and precipitation under cleavage conditions. Thus, cleavage conditions require optimization through trial and error and can be highly time-consuming.

Following optimized cleavage, purification of the POI away from the cleaved tag is usually accomplished in three-steps: i) initial purification of the fusion protein; ii) subsequent *in vitro* removal of the tag; and, iii) separation of the tag and protease from the target protein. As discussed in detail below, each purification step has components that require careful selection and optimization, and again can be very time-consuming. To avoid some of these limitations, a variety of methods have been developed. One example is the IMPACT™ system (Chong et al. 1997), which uses inducible self-cleaving tags to increase yield by decreasing the loss of protein during multiple purification steps and also reduces the overall time required for purification. These tags rely on inteins, an internal peptide segment that can self-excise upon activation and join the remaining portions (exteins), to produce two independent products. Another method is *in vivo* cleavage, also referred to as controlled intracellular processing (CIP) (Kapust and Waugh 2000). In CIP, the tagged protein and protease are co-expressed, but their expression is independent. Upon its induction, the protease can proteolytically cleave the tag from the POI intracellularly, resulting in considerable savings in time and cost. Innovations such as these in tag removal may allow for more rapid and optimal production of difficult membrane proteins for structural study and ongoing developments have great promise for future studies.

Considerations for membrane protein purification

One of the first factors to consider after overexpression of the POI is its localization. This is essential, since localization must be factored into selection of the initial purification method. For example, use of pelB tag can lead to localization of the protein to the medium or periplasm, which is mostly void of intracellular protein, allowing for simpler down-stream purification. Some proteins tend to aggregate into inclusion bodies, either with or without tags, and may require more stringent methods of solubilization prior to initial purification (Carrio and Villaverde 2002). Localization to inclusion bodies can also be highly beneficial, assuming the POI can be refolded, since this may greatly simplify downstream purification (Fan et al. 1996). Depending upon the host, membrane proteins can even be localized to the host cell membrane, which requires purification of cell membrane and subsequent solubilization prior to purification. Thus, identifying or manipulating (through tags) the localization of the target protein is an important step as it can dictate the method of purification.

The next logical consideration in purification is the solubilization process. Similar to tag removal methods, there are various methods of solubilization. The relative effectiveness of these methods is difficult to predict; therefore, trial and error is often involved to identify the most optimal and compatible method. Solubilization of membrane proteins typically relies either upon detergents, which coat exposed hydrophobic regions of the protein to allow solubilization, or chaotropic agents, which disrupt hydrogen bonds and decrease the hydrophobic effect, in turn favouring protein disaggregation. If possible, maintaining protein structure throughout the solubilization and purification process can be advantageous, as the process of refolding a denatured protein can prove to be extremely challenging (Booth 2003). However, membrane proteins have a tendency to aggregate without detergents or chaotropic agents, making subsequent purification more difficult (Smith 2011). Thus, it must be decided at this step whether it is essential to maintain protein structure (and potentially

function) throughout the process, as this can severely limit the scope of available solubilization and purification methods.

Purification techniques can be grouped into two categories: techniques that exploit the intrinsic properties of the protein, and techniques that require the use of a fusion tag. In practice, the purification process for many membrane proteins involves several techniques performed sequentially to achieve the desired purity (Park et al. 2008; Whiteman et al. 2014; Yun et al. 2015). Perhaps the most important consideration in choosing a purification technique is the compatibility of the solubilizing detergent or chaotropic agent with the purification technique of choice. For example, the charge state of a detergent may interfere with any technique that exploits the isoelectric point of the protein. Alternatively, detergent micelles bound to proteins may inhibit interaction in any affinity-based purification technique.

Following these considerations, fusion tag-based purification is the most popular method for initial purification of membrane proteins. Of all the fusion tags, polyhistidine tags are the most commonly used, enabling immobilized metal affinity chromatography (IMAC). IMAC is a versatile technique as it can be used for proteins solubilized using ionic or nonionic detergents alongside chaotropic agents (Bornhorst and Falke 2000). Detergent-solubilized membrane proteins are also susceptible to degradation by proteases, making the use of protease inhibitors necessary. It should be noted that protease inhibitors such as ethylenediaminetetraacetic acid (EDTA) that work by chelation of metal ions are incompatible with IMAC, even at moderately high concentrations. Many alternative fusion tag purification (Table 3) techniques can be applied to overcome this issue. Unfortunately, those techniques may also be hindered by the use of detergents required to solubilize the POI. For example, the interaction between the Strep-II tag and its affinity purification resin is often weakened by the presence of detergent micelles. Repeating the tag sequence is one strategy to counter these negative effects and increase affinity (Yeliseev et al. 2007).

Proteins can also be purified on the basis of differences in isoelectric point using ion exchange chromatography (IEC). Generally, ionic detergents cannot be used for solubilization of membrane proteins when employing this technique, which limits the number of possible detergent choices for solubilization. While using non-ionic or zwitterionic detergents, this method can also be applied to separate the protein-detergent complexes from the homogeneous detergent clusters (Seddon et al. 2004).

Reverse-phase high performance liquid chromatography (RP-HPLC) separates proteins based upon predominantly hydrophobicity and, to some degree, size, providing a purification technique particularly useful for small membrane proteins (Langelaan et al. 2013; Lee et al. 1996). The use of RP-HPLC requires careful consideration, as detergents cannot be used, and membrane proteins will be in a denatured state. This could cause large membrane proteins to exhibit a broad elution profile or to elute as a series of separate populations due to adoption of different conformations.

Size exclusion chromatography (SEC), or gel filtration chromatography, separates proteins based on hydrodynamic volume, in some instances corresponding directly to molecular

weight (Kunji et al. 2008). Sephadex, Superose, Superdex and Sephacryl are the most commonly used matrices for purification of membrane proteins by SEC. In comparison to other purification techniques, it has a lower resolution due to its inability to differentiate between species of similar hydrodynamic volume (Kunji et al. 2008). Therefore, it is often coupled with other methods to obtain the desired purity of the membrane protein. It is also critical to recognize that the apparent molecular weight of the protein will increase if it is bound to detergent micelle(s), complicating purification.

Detailed resources, e.g. (Arnold and Linke 2007; GE Healthcare 2007; Von Jagow et al. 1994), are available to aid in designing purification protocols for membrane proteins. In short, when using any protein purification technique, one must consider the downstream effects that the chosen strategy may have on both the final structure and functional state of the protein.

Rapid screening methods for membrane protein expression

In the case of membrane proteins, it is often challenging to identify a suitable expression host, optimize the level of protein expression and obtain a monodisperse sample for structural characterization. Successful high-level expression of functional membrane proteins for structure determination often requires evaluation of multiple versions of the POI obtained using strategies including additions and deletions in the N- and C- terminus, mutants with variations in natural sequence, and/or different fusion proteins. Therefore, methods for screening of expression of multiple membrane proteins in parallel are very useful. Fusing green fluorescent protein (GFP) to the POI is one of the most commonly used methods to gauge expression levels without purifying the protein (Chalfie et al. 1994). Coupling SEC with fluorescence-detection enables screening for both expression and homogeneity of the produced membrane protein sample. Fluorescence-detection size-exclusion chromatography (FSEC) experiments using a GFP-tagged POI to screen for expression level and monodispersity has been used for various proteins produced using different expression systems (Bird et al. 2015; Chaudhary et al. 2011; Chen et al. 2013; Goehring et al. 2014; Kawate and Gouaux 2006). Proper implementations of such high-throughput methods have great potential to significantly enhance structural biology efforts for membrane proteins.

Considerations in sample preparation for structural studies

If the objective is X-ray crystallography, then crystallization conditions must be optimized to obtain well-diffracting, sufficiently large crystals upon successful purification of a folded membrane protein. For initial crystallization trials, various crystallization conditions are usually screened using the vapour diffusion method (Delmar et al. 2015). Conditions leading to crystal formation are then subsequently optimized. Screening systems such as MemStart, MemGold and MemSys are specifically designed for membrane protein crystallization trials (Carpenter et al. 2008; Newstead et al. 2008). In order to solubilize a membrane POI for crystallization, protein detergent complexes, bicelles, or lipid cubic phases (LPCs) may be employed.

In the case of protein-detergent complexes (the most commonly employed phase for membrane protein crystallization (Loll 2014)), an important point to keep in mind is that the membrane protein must be solubilized with a detergent that forms micelles that are relatively small in size as compared to the size of the protein itself. This is due to the fact that large micelles may inhibit protein crystallization (Prive 2007). The most successfully and frequently applied detergents for membrane protein crystallization (Loll 2014; Prive 2007) are alkyl glycosides such as *n*-octyl- β -D-glucopyranoside (OG), *n*-nonyl- β -D-glucopyranoside (NG), *n*-decyl- β -D-maltopyranoside (DM), *n*-undecyl- β -D-maltopyranoside (UM), *n*-dodecyl- β -D-maltopyranoside (DDM), and the cyclohexyl maltoside (CYMAL) family of detergents alongside the maltose-neopentyl glycol (MNG) amphiphiles; *N,N*-dimethyldodecylamine-*N*-oxide (LDAO); and, octyltetraoxyethylene (C₈E₄) and other polyoxyethylenes.

Alternatively, bicelles (disk or Swiss cheese-like structures comprising a mix of long chain lipids and short chain lipids or detergents (Poulos et al. 2015; Prosser et al. 2006)) may be employed for crystallization. Most frequently, the lipid 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) is mixed with 3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO) or an analogue of CHAPSO (Poulos et al. 2015). Although relatively unpopular and, arguably, underutilized at present (Loll 2014), bicelles provide the potential to maintain a bilayer environment throughout the crystallization process and offer some advantages in sample handling and manipulation over detergent-protein complexes and the LCP. Methodologies and considerations for bicelle-based crystallization and analysis are comprehensively covered in (Poulos et al. 2015).

The lipid cubic phase (LCP), also referred to as *in meso* conditions, is an alternative method of protein crystallization that has gained popularity and shown great success in membrane protein crystallization in the past decade (Loll 2014). For *in meso* crystallization, the standard protocol involves homogenizing the purified protein-detergent complex with a monoacylglycerol (MAG) like monoolein, and then adding the precipitant to initialize crystallization. This is comprehensively reviewed in (Caffrey 2015).

Reduction of protein dynamics is another very important prerequisite in obtaining well-diffracting crystals. There are various factors that can be modified to reduce protein dynamics, including optimization of buffers, detergents, bicelle components, or LCP constituents employed during crystallization. In addition, fusion proteins or antibodies have been beneficial in crystallization. Although the previous section discussed fusion tag removal, retaining the tag has been fundamental for decreased dynamics and improving crystallization for many proteins studied by XRD. For example, T4 lysozyme fusion (Engel et al. 2002) was instrumental in the crystallization of many GPCRs including β 2-adrenergic receptor (Cherezov et al. 2007), δ -opioid receptor (Granier et al. 2012), and M3 muscarinic acetylcholine receptor (Kruse et al. 2012). Antibodies, antibody fragments, or nanobodies have also been used to decrease dynamics to facilitate crystallization. In the case of GPCRs, use of nanobodies (as G-protein mimics) allowed for crystallization of GPCRs in the active conformation to provide valuable information on receptor activation and subsequent G-protein binding (Kruse et al. 2013; Rasmussen et al. 2011). Also, mutating residues with a high degree of dynamics also yielded well-diffracting crystals, as seen in β 1-adrenergic

receptor (Serrano-Vega et al. 2008), neurotensin receptor (Shibata et al. 2009), and adenosine A(2A) receptor (Lebon et al. 2011). Lastly, post-expression modification of proteins has been helpful in promoting crystallization of membrane proteins. Such modifications including limited proteolysis to remove flexible regions (Doyle et al. 1998) and deglycosylation to remove heterogeneous glycosylation (Sui et al. 2000) are frequently employed, as reviewed in (Columbus 2015). Thus, screening different crystallization conditions, antibodies and fusion proteins, and mutations/modifications for decreased dynamics are common and fundamental steps in crystallization (Loll 2014; Prive 2007; Serrano-Vega et al. 2008).

Notably, recent advancements in XRD have been afforded by the introduction of free-electron lasers. This has allowed a significant step forward in structural characterization of membrane proteins, making possible the collection of diffraction data from micron- or submicron-scale crystals alongside time-resolved studies of protein dynamics on an ultrafast time scale (Neutze et al. 2015). This has clear potential for extremely high impact in the field, but is also not a routinely available technique.

NMR is a very versatile technique for atomic-level studies of biomolecular structure and dynamics. Membrane proteins may be characterized in environments amenable to either solution- or solid-state NMR methods. These methods differ in their requirement for isotropic tumbling of the nuclei being probed, with large molecules or complexes (>100 kDa) being extremely challenging to characterize by, if at all amenable to, solution-state NMR (Nietlispach and Gautier 2011). NMR allows study of membrane proteins in their native environment (Brown and Ladizhansky 2015; Murray et al. 2013) or, following reconstitution, in native-like environments, including membrane-mimetics such as organic solvent mixtures (Schwaiger et al. 1998), micelles (Arora et al. 2001), bicelles (Prosser et al. 2006), nanodiscs (Gluck et al. 2009; Hagn et al. 2013), and bilayers (Das et al. 2013). Based on the size of the reconstituted assembly of the membrane protein, either solution-state or solid-state approaches are used for their characterization. For solution-state NMR, approaches such as transverse relaxation optimized spectroscopy (TROSY) (Pervushin et al. 1997) and specific-isotope labeling schemes (Tugarinov et al. 2002) have been used to extend the use of solution-state NMR for larger protein assemblies. These approaches rely upon exploitation of differential/optimal relaxation properties, and have also been exploited to study both structure and dynamics of integral membrane proteins (Gautier et al. 2010; Hwang and Kay 2005). Solution-state NMR is a very powerful technique to study appropriately sized membrane protein samples in mimetics such as an organic solvent mixture, detergent micelles, isotropic bicelles and monodisperse nanodiscs,

For both solution-state NMR and XRD, it is of great importance that the monodispersity and stability of the reconstituted membrane protein (whether in micelles, bicelles, or another environment) be determined prior to attempting structure determination. For XRD, maintaining soluble monodispersed protein is crucial for successful crystallization while for solution-state NMR a polydisperse sample will unduly confound what is likely an already challenging resonance assignment process. SEC may be used to confirm homogeneity of a purified/solubilized POI, monodispersity and oligomeric state of the POI, and as such is frequently used as an analytical technique to validate sample integrity (Kunji et al. 2008),

rather than “just” for purification as discussed above. Hydrodynamics techniques, such as dynamic light scattering (Neale et al. 2013), sedimentation velocity analytical ultracentrifugation (Ebel 2011), or pulsed-field gradient diffusion NMR (Liebau et al. 2015), are highly valuable alternatives or corroborative techniques for SEC-based confirmation of sample monodispersity.

Study of membrane proteins in a native environment typically requires different types of lipids, leading to larger protein-lipid complexes such as non-isotropic bicelles, macrodiscs, or bilayers that are not compatible with solution-state NMR. In the case of these samples, where transverse relaxation is typically extremely rapid while longitudinal relaxation may be extremely slow, one must resort to solid-state NMR approaches. Recent advance in solid-state NMR have enabled structural characterization of membrane proteins in environments very close to the native membranes, as very nicely reviewed in (Opella 2015) and (Judge et al. 2015). Solid-state NMR has strong potential for characterization of both structure and dynamics of membrane proteins, and is still being actively developed.

Cryo-EM is a method which is rapidly growing in popularity in the field of membrane proteins. In cryo-EM, atomic-level structures can be determined either by electron crystallography, in which electron diffraction from 2D crystals afford high-resolution, or single particle analysis, where a 3D structure is determined through computational reconstruction of 2D images from transmission electron microscopy (TEM) of many individual macromolecules in identical or similar conformations but stochastic orientation. Cryo-EM presents a number of practical difficulties to overcome for near-atomic resolution structure calculations. These include sample dehydration under vacuum, radiation damage by the electron beam, and poor electron scattering by biological macromolecules. To get around first two issues, preparations of biological cryo-EM samples involve cryogen-mediated trapping of samples in a thin layer of noncubic amorphous or vitrified ice; thus, giving the name cryo-EM. This step allows for preservation of specimen at near-native, hydrated environment.

Although cryo-EM was once limited by low-resolution arising from a poor signal to noise ratio, current state of the art cryo-EM can characterize protein structures at a single digit Å level and is pushing towards smaller and smaller proteins (Smith and Rubinstein 2014). To date, the best resolution of 1.9 Å (on par with the best XRD structures) was observed for the membrane protein aquaporin using the electron crystallography technique (Gonen et al. 2005). Recently, the single particle reconstruction technique was used to solve the structure of the 170 kDa human γ -secretase to 3.4 Å, an elusive target of XRD, and the smallest protein solved by cryo-EM thus far (Bai et al. 2015). Improved microscope technology (Zhou and Chiu 1993); detector technology (Campbell et al. 2012; Ruskin et al. 2013), and computational resources and image processing strategies (Shigematsu and Sigworth 2013) have all been instrumental in the increased resolution now possible.

In particular, cryo-EM is phenomenally suited to characterization of large molecular weight complexes. It is important to note that such characterization often relies upon structures of some or all of the components making up the complex previously determined by XRD or NMR in order to produce an atomic model. The γ -secretase structure mentioned above, for

example, used the XRD-derived domain structures to model the 170 kDa complex (Bai et al. 2015). The role of cryo-EM in understanding of membrane protein structure will only increase as technology continues to advance. Comprehensive overviews of cryo-EM are provided by (Binshtein and Ohi 2015; Kuhlbrandt 2014; Liao et al. 2014).

Deciding between methods of membrane protein structure determination

If a protein is relatively small (<50 kDa), and amenable to isotopic enrichment through expression in an appropriate host and subsequent solubilization, then solution-state NMR may be preferable as this provides the potential to gain insight into conformational dynamics and ligand binding mechanisms. The structure and dynamics of the heptahelical protein sensory rhodopsin II, for example, proved amenable to structural and dynamic characterization in both detergent micelles and in bicelles (Gautier et al. 2010). Probing of specific regions of β 2-adrenergic receptor was made possible by labeling with NMR-active ^{19}F Trp (Manglik et al. 2015) or ^{13}C Met (Nygaard et al. 2013), allowing elucidation of the effects of agonists, antagonists, and G-proteins upon the conformational equilibrium of the receptor.

If a protein is larger than ~50 kDa, solution-state NMR becomes very difficult (although by no means impossible) for the reasons outlined earlier. Larger proteins amenable to isotope enrichment may then be good targets for solid-state NMR study, allowing solubilization in a bilayer setting and study of both dynamic and more rigid portions of the protein. Nice recent examples of state-of-the-art capabilities of solid-state NMR include the backbone structure of the GPCR CXCR1 (Park et al. 2012b) alongside characterization of heptahelical *Anabaena* sensory rhodopsin structure (Wang et al. 2013) and dynamics (Good et al. 2014) in lipid bilayers.

If the protein is ordered in its nature or can be forced into an ordered structure through use of tags, antibodies, or mutagenesis, then both cryo-EM (2D crystal) and XRD (3D crystal) may be desirable. In particular, XRD allows for relatively rapid determination of structures of unknown novel POIs, as is manifest in its near monopoly in the field of protein structure characterization for many years. Given that XRD is dependent on crystallization, any dynamic regions of the POI will unfortunately not be visible. Ideally, specific regions of this nature could be probed either in isolation or in the full protein by NMR spectroscopy, assuming that POI is amenable to the conditions stated above.

In the case of a relatively large POI that contains domains of solved structure, cryo-EM can be greatly advantageous compared to NMR and XRD in elucidating various conformations simultaneously in a single sample set. A recent example was the differentiation between multiple conformations of V-ATPases, providing valuable insights in the mechanism of ATPase activity (Zhao et al. 2015; Zhou et al. 2015). However, given that cryo-EM does not always present high-resolution images, one could argue that cryo-EM may be most optimal as a complementary technique. For reference, Table 5 provides an overview of pros and cons of each the techniques discussed. Exploiting the advantages of each technique under proper conditions will further our understanding of membrane protein structure and dynamics.

Concluding Remarks

The past 15 years have been truly remarkable for the field of membrane protein structural biology, with tremendous expansion in protein production, sample preparation, and experimental approaches. As an example, this growth has provided structural insight into 31 unique members of one of the most sought after class of membrane proteins, the G-protein-coupled receptors (GPCRs) (Chrencik et al. 2015; Dore et al. 2014; Park et al. 2012b; Srivastava et al. 2014; Tautermann 2014; Yin et al. 2015; Zhang et al. 2015a; Zhang et al. 2015b). However, this set of 31 GPCRs represents only a fraction of all GPCRs (~800 in the human genome; (Fredriksson et al. 2003)), with many receptors remaining uncharacterized due to a lack of high throughput and generalized methods for membrane protein production. As has been highlighted throughout this review, even with all of the recent advances, determining optimal, often protein-specific, conditions for membrane protein production and characterization still tends to heavily rely upon trial and error and remains an extremely time-consuming task.

In this review, we have tried to encapsulate the advancements that have allowed the recent growth in structural knowledge of membrane proteins, with an emphasis on heterologous production, subsequent purification, and screening of conditions for structure characterization (Figure 3). To summarize, *E. coli* tends to be the most cost-effective (and popular, Figure 2) host for membrane protein production. However, due to the lack of eukaryotic PTM machineries, membrane proteins tend to aggregate, misfold, and impair cell viability. In these cases, use of yeast, baculovirus, or mammalian system may be beneficial; however, yields may be substantially lower and cost may be prohibitive. Cell-free expression systems are, therefore, perhaps the most attractive alternative, given the number of different configurations possible. Regardless of expression system, new fusion tags, proteolytic, and chemical cleavage strategies continue to provide new routes to enhanced expression and purification. Unfortunately, until more streamlined methods to screen membrane protein production and purification protocols are developed, obtaining the required quantity of well-folded and functional protein will remain the bottleneck in membrane protein structural biology.

Acknowledgments

This work was supported by a Canadian Institutes of Health Research (CIHR) Operating Grant (MOP-111138 to J.K.R.); a Nova Scotia Health Research Foundation (NSHRF) Scotia Support Grant (to J.K.R.); and, a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (to X.Q.L.). AP was supported by the Beatrice Hunter Cancer Research Institute with funds provided by the Canadian Imperial Bank of Commerce and the Harvey Graham Cancer Research Fund as part of The Terry Fox Strategic Health Research Training Program in Cancer Research at CIHR; KS is supported by an Alexander Graham Bell Canadian Graduate Scholarship from NSERC; and, JKR is supported by a CIHR New Investigator Award.

References

- Abrahmsén L, Moks T, Nilsson B, Uhlén M. Secretion of heterologous gene products to the culture medium of *Escherichia coli*. Nucleic Acids Res. 1986; 14(18):7487–7500. [PubMed: 3020516]
- Achmuller C, Kaar W, Ahrer K, Wechner P, Hahn R, Werther F, Schmidinger H, Cserjan-Puschmann M, Clementschitsch F, Striedner G, Bayer K, Jungbauer A, Auer B. N(pro) fusion technology to produce proteins with authentic N termini in *E. coli*. Nat Methods. 2007; 4(12):1037–1043. [PubMed: 18026112]

- Ahn KY, Song JA, Han KY, Park JS, Seo HS, Lee J. Heterologous protein expression using a novel stress-responsive protein of *E. coli* RpoA as fusion expression partner. *Enzyme Microb Technol.* 2007; 41(6–7):859–866.
- Anastasina M, Terenin I, Butcher SJ, Kainov DE. A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *Biotechniques.* 2014; 56(1):36–39. [PubMed: 24447137]
- Andersen KR, Leksa NC, Schwartz TU. Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins.* 2013; 81(11):1857–1861. [PubMed: 23852738]
- Andrell J, Tate CG. Overexpression of membrane proteins in mammalian cells for structural studies. *Mol Membr Biol.* 2013; 30(1):52–63. [PubMed: 22963530]
- Arinaminpathy Y, Khurana E, Engelman DM, Gerstein MB. Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov Today.* 2009; 14(23–24):1130–1135. [PubMed: 19733256]
- Arnold T, Linke D. Phase separation in the isolation and purification of membrane proteins. *Biotechniques.* 2007; 43(4):427–430. 432. 434 passim. [PubMed: 18019333]
- Arora A, Abildgaard F, Bushweller JH, Tamm LK. Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat Struct Biol.* 2001; 8(4):334–338. [PubMed: 11276254]
- Bai XC, Yan C, Yang G, Lu P, Ma D, Sun L, Zhou R, Scheres SH, Shi Y. An atomic structure of human gamma-secretase. *Nature.* 2015; 525(7568):212–217. [PubMed: 26280335]
- Baneyx F. Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol.* 1999; 10(5): 411–421. [PubMed: 10508629]
- Becker GW, Hsiung HM. Expression, secretion and folding of human growth hormone in *Escherichia coli*. Purification and characterization. *FEBS Lett.* 1986; 204(1):145–150. [PubMed: 3527743]
- Bernaudeau F, Frelet-Barrand A, Pochon N, Dementin S, Hivin P, Boutigny S, Rioux JB, Salvi D, Seigneurin-Berny D, Richaud P, Joyard J, Pignol D, Sabaty M, Desnos T, Pebay-Peyroula E, Darrouzet E, Vernet T, Rolland N. Heterologous expression of membrane proteins: choosing the appropriate host. *PLoS One.* 2011; 6(12):e29191. [PubMed: 22216205]
- Better M, Bernhard SL, Lei SP, Fishwild DM, Lane JA, Carroll SF, Horwitz AH. Potent anti-CD5 ricin A chain immunoconjugates from bacterially produced Fab' and F(ab')₂. *Proc Natl Acad Sci U S A.* 1993; 90(2):457–461. [PubMed: 7678457]
- Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, Newstead S, Poolman B, Tate CG, Vogel H. Overcoming barriers to membrane protein structure determination. *Nat Biotechnol.* 2011; 29(4):335–340. [PubMed: 21478852]
- Binshtein E, Ohi MD. Cryo-electron microscopy and the amazing race to atomic resolution. *Biochemistry.* 2015; 54(20):3133–3141. [PubMed: 25955078]
- Bird LE, Rada H, Verma A, Gasper R, Birch J, Jennions M, Lwe J, Moraes I, Owens RJ. Green fluorescent protein-based expression screening of membrane proteins in *Escherichia coli*. *J Vis Exp.* 2015; (95):e52357. [PubMed: 25590335]
- Boer E, Steinborn G, Kunze G, Gellissen G. Yeast expression platforms. *Appl Microbiol Biotechnol.* 2007; 77(3):513–523. [PubMed: 17924105]
- Booth PJ. The trials and tribulations of membrane protein folding *in vitro*. *Biochim Biophys Acta.* 2003; 1610(1):51–56. [PubMed: 12586379]
- Bornert O, Alkhalfioui F, Logez C, Wagner R. Overexpression of membrane proteins using *Pichia pastoris*. *Curr Protoc Protein Sci.* 2012; Chapter 29(Unit 29.2)
- Bornhorst JA, Falke JJ. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* 2000; 326:245–254. [PubMed: 11036646]
- Botelho AV, Huber T, Sakmar TP, Brown MF. Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophys J.* 2006; 91(12):4464–4477. [PubMed: 17012328]
- Bowie JU. Are we destined to repeat history? *Curr Opin Struct Biol.* 2000; 10(4):435–437. [PubMed: 10981631]
- Brodel AK, Sonnabend A, Kubick S. Cell-free protein expression based on extracts from CHO cells. *Biotechnol Bioeng.* 2014; 111(1):25–36. [PubMed: 24018795]

- Brown LS, Ladizhansky V. Membrane proteins in their native habitat as seen by solid-state NMR spectroscopy. *Protein Sci.* 2015; 24(9):1333–1346. [PubMed: 25973959]
- Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein Expr Purif.* 2008; 59(1):94–102. [PubMed: 18289875]
- Butt TR, Edavettal SC, Hall JP, Mattern MR. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif.* 2005; 43(1):1–9. [PubMed: 16084395]
- Caffrey M. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr F Struct Biol Commun.* 2015; 71(Pt 1):3–18. [PubMed: 25615961]
- Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkis D, Veessler D, Pan J, Harrison SC, Potter CS, Carragher B, Grigorieff N. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure.* 2012; 20(11):1823–1828. [PubMed: 23022349]
- Camponova P, Baud S, Matras H, Duroux-Richard I, Bonnafous JC, Marie J. High-level expression and purification of the human bradykinin B(2) receptor in a tetracycline-inducible stable HEK293S cell line. *Protein Expr Purif.* 2007; 55(2):300–311. [PubMed: 17561413]
- Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol.* 2008; 18(5):581–586. [PubMed: 18674618]
- Carrio MM, Villaverde A. Construction and deconstruction of bacterial inclusion bodies. *J Biotechnol.* 2002; 96(1):3–12. [PubMed: 12142138]
- Caswell J, Snoddy P, McMeel D, Buick RJ, Scott CJ. Production of recombinant proteins in *Escherichia coli* using an N-terminal tag derived from sortase. *Protein Expr Purif.* 2010; 70(2):143–150. [PubMed: 19883766]
- Chalfie M, Tu Y, Euskirchen G, Ward W, Prasher D. Green fluorescent protein as a marker for gene expression. *Science.* 1994; 263(5148):802–805. [PubMed: 8303295]
- Chatterjee DK, Esposito D. Enhanced soluble protein expression using two new fusion tags. *Protein Expr Purif.* 2006; 46(1):122–129. [PubMed: 16146696]
- Chaudhary S, Pak JE, Pedersen BP, Bang LJ, Zhang LB, Ngaw SM, Green RG, Sharma V, Stroud RM. Efficient expression screening of human membrane proteins in transiently transfected Human Embryonic Kidney 293S cells. *Methods.* 2011; 55(4):273–280. [PubMed: 21925269]
- Chelikani P, Reeves PJ, Rajbhandary UL, Khorana HG. The synthesis and high-level expression of a beta2-adrenergic receptor gene in a tetracycline-inducible stable mammalian cell line. *Protein Sci.* 2006; 15(6):1433–1440. [PubMed: 16731977]
- Chen H, Shaffer PL, Huang X, Rose PE. Rapid screening of membrane protein expression in transiently transfected insect cells. *Protein Expr Purif.* 2013; 88(1):134–142. [PubMed: 23268112]
- Cheng A, van Hoek AN, Yeager M, Verkman AS, Mitra AK. Three-dimensional organization of a human water channel. *Nature.* 1997; 387(6633):627–630. [PubMed: 9177354]
- Cheng Y, Gu J, Wang HG, Yu S, Liu YQ, Ning YL, Zou QM, Yu XJ, Mao XH. EspA is a novel fusion partner for expression of foreign proteins in *Escherichia coli*. *J Biotechnol.* 2010; 150(3):380–388. [PubMed: 20854850]
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weiss WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human β 2-Adrenergic G protein-coupled Receptor. *Science.* 2007; 318(5854):1258–1265. [PubMed: 17962520]
- Chini B, Parenti M. G-protein-coupled receptors, cholesterol and palmitoylation: facts about fats. *J Mol Endocrinol.* 2009; 42(5):371–379. [PubMed: 19131499]
- Choi SI, Song HW, Moon JW, Seong BL. Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. *Biotechnol Bioeng.* 2001; 75(6):718–724. [PubMed: 11745150]
- Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene.* 1997; 192(2):271–281. [PubMed: 9224900]

- Chrencik JE, Roth CB, Terakado M, Kurata H, Omi R, Kihara Y, Warshaviak D, Nakade S, Asmar-Rovira G, Mileni M, Mizuno H, Griffith MT, Rodgers C, Han GW, Velasquez J, Chun J, Stevens RC, Hanson MA. Crystal structure of antagonist bound human lysophosphatidic acid receptor 1. *Cell*. 2015; 161(7):1633–1643. [PubMed: 26091040]
- Chun E, Thompson AA, Liu W, Roth CB, Griffith MT, Katritch V, Kunken J, Xu F, Cherezov V, Hanson MA, Stevens RC. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure*. 2012; 20(6):967–976. [PubMed: 22681902]
- Clark KM, Fedoriw N, Robinson K, Connelly SM, Randles J, Malkowski MG, DeTitta GT, Dumont ME. Purification of transmembrane proteins from *Saccharomyces cerevisiae* for X-ray crystallography. *Protein Expr Purif*. 2010; 71(2):207–223. [PubMed: 20045057]
- Clark L, Zahm JA, Ali R, Kukula M, Bian L, Patrie SM, Gardner KH, Rosen MK, Rosenbaum DM. Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*. *J Biomol NMR*. 2015; 62(3):239–245. [PubMed: 26025061]
- Columbus L. Post-expression strategies for structural investigations of membrane proteins. *Curr Opin Struct Biol*. 2015; 32:131–138. [PubMed: 25951412]
- Cook GA, Stefer S, Opella SJ. Expression and purification of the membrane protein p7 from hepatitis C virus. *Biopolymers*. 2011; 96(1):32–40. [PubMed: 20560141]
- Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ. Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J Biol Chem*. 1990; 265(16):9062–9065. [PubMed: 2160953]
- Costa SJ, Almeida A, Castro A, Domingues L, Besir H. The novel Fh8 and H fusion partners for soluble protein expression in *Escherichia coli*: a comparison with the traditional gene fusion technology. *Appl Microbiol Biotechnol*. 2013a; 97(15):6779–6791. [PubMed: 23160981]
- Costa SJ, Coelho E, Franco L, Almeida A, Castro A, Domingues L. The Fh8 tag: a fusion partner for simple and cost-effective protein purification in *Escherichia coli*. *Protein Expr Purif*. 2013b; 92(2):163–170. [PubMed: 24084009]
- Cregg JM, Cereghino JL, Shi J, Higgins DR. Recombinant protein expression in *Pichia pastoris*. *Mol Biotechnol*. 2000; 16(1):23–52. [PubMed: 11098467]
- Crimmins DL, Mische SM, Denslow ND. Chemical cleavage of proteins in solution. *Curr Protoc Protein Sci*. 2005; Chapter 11(Unit 11.4)
- Crispin M, Chang VT, Harvey DJ, Dwek RA, Evans EJ, Stuart DI, Jones EY, Lord JM, Spooner RA, Davis SJ. A human embryonic kidney 293T cell line mutated at the Golgi α -mannosidase II locus. *J Biol Chem*. 2009; 284(32):21684–21695. [PubMed: 19465480]
- Cybulski LE, de Mendoza D. Bilayer hydrophobic thickness and integral membrane protein function. *Curr Protein Pept Sci*. 2011; 12(8):760–766. [PubMed: 22044142]
- Das N, Murray DT, Cross TA. Lipid bilayer preparations of membrane proteins for oriented and magic-angle spinning solid-state NMR samples. *Nat Protoc*. 2013; 8(11):2256–2270. [PubMed: 24157546]
- Davis GD, Elisee C, Newham DM, Harrison RG. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol Bioeng*. 1999; 65(4):382–388. [PubMed: 10506413]
- de Marco A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb Cell Fact*. 2009; 8:26. [PubMed: 19442264]
- de Marco A, Vigh L, Diamant S, Goloubinoff P. Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperones*. 2005; 10(4):329–339. [PubMed: 16333986]
- Deisenhofer J, Epp O, Miki K, Huber R, Michel H. X-ray structure analysis of a membrane protein complex. Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodospseudomonas viridis*. *J Mol Biol*. 1984; 180(2):385–398. [PubMed: 6392571]
- Delmar JA, Bolla JR, Su CC, Yu EW. Crystallization of membrane proteins by vapor diffusion. *Methods Enzymol*. 2015; 557:363–392. [PubMed: 25950974]
- DelProposto J, Majmudar CY, Smith JL, Brown WC. Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr Purif*. 2009; 63(1):40–49. [PubMed: 18824232]

- Denton H, Smith M, Husi H, Uhrin D, Barlow PN, Batt CA, Sawyer L. Isotopically labeled bovine beta-lactoglobulin for NMR studies expressed in *Pichia pastoris*. *Protein Expr Purif*. 1998; 14(1): 97–103. [PubMed: 9758756]
- Deupi X, Edwards P, Singhal A, Nickle B, Oprian D, Schertler G, Standfuss J. Stabilized G protein binding site in the structure of constitutively active metarhodopsin-II. *Proc Natl Acad Sci U S A*. 2012; 109(1):119–124. [PubMed: 22198838]
- di Guana C, Lib P, Riggs PD, Inouyeb H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene*. 1988; 67(1):21–30. [PubMed: 2843437]
- Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, Errey JC, Jazayeri A, Khan S, Tehan B, Weir M, Wiggin GR, Marshall FH. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature*. 2014; 511(7511):557–562. [PubMed: 25042998]
- Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science*. 1998; 280(5360):69–77. [PubMed: 9525859]
- Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, Iwata S. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat Protoc*. 2008; 3(5):784–798. [PubMed: 18451787]
- Dukkipati A, Park HH, Waghay D, Fischer S, Garcia KC. BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. *Protein Expr Purif*. 2008; 62(2):160–170. [PubMed: 18782620]
- Ebel C. Sedimentation velocity to characterize surfactants and solubilized membrane proteins. *Methods*. 2011; 54(1):56–66. [PubMed: 21112401]
- Ellison MJ, Hochstrasser M. Epitope-tagged ubiquitin. A new probe for analyzing ubiquitin function. *J Biol Chem*. 1991; 266(31):21150–21157. [PubMed: 1718971]
- Engel CK, Chen L, Prive GG. Insertion of carrier proteins into hydrophilic loops of the *Escherichia coli* lactose permease. *Biochim Biophys Acta*. 2002; 1564(1):38–46. [PubMed: 12100994]
- Ezure T, Suzuki T, Ando E. A cell-free protein synthesis system from insect cells. *Methods Mol Biol*. 2014; 1118:285–296. [PubMed: 24395425]
- Fagerberg L, Jonasson K, von Heijne G, Uhlen M, Berglund L. Prediction of the human membrane proteome. *Proteomics*. 2010; 10(6):1141–1149. [PubMed: 20175080]
- Fan QR, Garboczi DN, Winter CC, Wagtman N, Long EO, Wiley DC. Direct binding of a soluble natural killer cell inhibitory receptor to a soluble human leukocyte antigen-Cw4 class I major histocompatibility complex molecule. *Proc Natl Acad Sci U S A*. 1996; 93(14):7178–7183. [PubMed: 8692965]
- Ficca AG, Testa L, Tocchini Valentini GP. The human beta 2-adrenergic receptor expressed in *Schizosaccharomyces pombe* retains its pharmacological properties. *FEBS Lett*. 1995; 377(2): 140–144. [PubMed: 8543037]
- Field J, Nikawa J, Broek D, MacDonald B, Rodgers L, Wilson IA, Lerner RA, Wigler M. Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol Cell Biol*. 1988; 8(5):2159–2165. [PubMed: 2455217]
- Francisco JA, Earhart CF, Georgiou G. Transport and anchoring of beta-lactamase to the external surface of *Escherichia coli*. *Proc Natl Acad Sci U S A*. 1992; 89(7):2713–2717. [PubMed: 1557377]
- Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol*. 2003; 63(6):1256–1272. [PubMed: 12761335]
- Gao D, Wang S, Li H, Yu H, Qi Q. Identification of a heterologous cellulase and its N-terminus that can guide recombinant proteins out of *Escherichia coli*. *Microb Cell Fact*. 2015; 14:49. [PubMed: 25889647]
- Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol*. 2010; 17(6):768–774. [PubMed: 20512150]

- GE Healthcare. Purifying Challenging Proteins: Principles and Methods. GE Healthcare Bio-Sciences AB; Uppsala, Sweden: 2007.
- Geisse S, Fux C. Recombinant protein production by transient gene transfer into mammalian cells. *Methods Enzymol.* 2009; 463:223–238. [PubMed: 19892175]
- Gluck JM, Wittlich M, Feuerstein S, Hoffmann S, Willbold D, Koenig BW. Integral membrane proteins in nanodiscs can be studied by solution NMR spectroscopy. *J Am Chem Soc.* 2009; 131(34):12060–12061. [PubMed: 19663495]
- Goehring A, Lee CH, Wang KH, Michel JC, Claxton DP, Bacongus I, Althoff T, Fischer S, Garcia KC, Gouaux E. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat Protoc.* 2014; 9(11):2574–2585. [PubMed: 25299155]
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. Life with 6000 genes. *Science.* 1996; 274(5287):563–567.
- Goncharuk SA, Goncharuk MV, Mayzel ML, Lesovoy DM, Chupin VV, Bocharov EV, Arseniev AS, Kirpichnikov MP. Bacterial Synthesis and Purification of Normal and Mutant Forms of Human FGFR3 Transmembrane Segment. *Acta Naturae.* 2011; 3(3):77–84.
- Gonen T, Cheng Y, Sliz P, Hiroaki Y, Fujiyoshi Y, Harrison SC, Walz T. Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature.* 2005; 438(7068):633–638. [PubMed: 16319884]
- Good DB, Wang S, Ward ME, Struppe J, Brown LS, Lewandowski JR, Ladizhansky V. Conformational dynamics of a seven transmembrane helical protein *Anabaena* Sensory Rhodopsin probed by solid-state NMR. *J Am Chem Soc.* 2014; 136(7):2833–2842. [PubMed: 24467417]
- Gordon E, Horsefield R, Swarts HG, de Pont JJ, Neutze R, Snijder A. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expr Purif.* 2008; 62(1):1–8. [PubMed: 18692139]
- Gossert AD, Jahnke W. Isotope labeling in insect cells. *Adv Exp Med Biol.* 2012; 992:179–196. [PubMed: 23076585]
- Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK. Structure of the δ -opioid receptor bound to naltrindole. *Nature.* 2012; 485(7398):400–404. [PubMed: 22596164]
- Grisshammer R. Understanding recombinant expression of membrane proteins. *Curr Opin Biotechnol.* 2006; 17(4):337–340. [PubMed: 16777403]
- Haberstock S, Roos C, Hoevels Y, Dötsch V, Schnapp G, Pautsch A, Bernhard F. A systematic approach to increase the efficiency of membrane protein production in cell-free expression systems. *Protein Expr Purif.* 2012; 82(2):308–316. [PubMed: 22342679]
- Hagn F, Etzkorn M, Raschle T, Wagner G. Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *J Am Chem Soc.* 2013; 135(5):1919–1925. [PubMed: 23294159]
- Hamilton SR, Bobrowicz P, Bobrowicz B, Davidson RC, Li H, Mitchell T, Nett JH, Rausch S, Stadheim TA, Wischniewski H, Wildt S, Gerngross TU. Production of complex human glycoproteins in yeast. *Science.* 2003; 301(5637):1244–1246. [PubMed: 12947202]
- Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi BK, Hopkins D, Wischniewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science.* 2006; 313(5792):1441–1443. [PubMed: 16960007]
- Han KY, Seo HS, Song JA, Ahn KY, Park JS, Lee J. Transport proteins PotD and Crr of *Escherichia coli*, novel fusion partners for heterologous protein expression. *Biochim Biophys Acta.* 2007a; 1774(12):1536–1543. [PubMed: 17974510]
- Han KY, Song JA, Ahn KY, Park JS, Seo HS, Lee J. Enhanced solubility of heterologous proteins by fusion expression using stress-induced *Escherichia coli* protein, Tsf. *FEMS Microbiol Lett.* 2007b; 274(1):132–138. [PubMed: 17608803]
- Hansson L, Noppa L, Nilsson AK, Stromqvist M, Bergstrom S. Expression of truncated and full-length forms of the Lyme disease *Borrelia* outer surface protein A in *Escherichia coli*. *Protein Expr Purif.* 1995; 6(1):15–24. [PubMed: 7756835]

- Harbers M. Wheat germ systems for cell-free protein expression. *FEBS Lett.* 2014; 588(17):2762–2773. [PubMed: 24931374]
- Hasenwinkle D, Jervis E, Kops O, Liu C, Lesnicki G, Haynes CA, Kilburn DG. Very high-level production and export in *Escherichia coli* of a cellulose binding domain for use in a generic secretion-affinity fusion system. *Biotechnol Bioeng.* 1997; 55(6):854–863. [PubMed: 18636596]
- Hassaine G, Wagner R, Kempf J, Cherouati N, Hassaine N, Prual C, Andre N, Reinhart C, Pattus F, Lundstrom K. Semliki Forest virus vectors for overexpression of 101 G protein-coupled receptors in mammalian host cells. *Protein Expr Purif.* 2006; 45(2):343–351. [PubMed: 16055346]
- Hattab G, Warschawski DE, Moncoq K, Miroux B. *Escherichia coli* as host for membrane protein structure determination: a global analysis. *Sci Rep.* 2015; 5:12097. [PubMed: 26160693]
- He Y, Wang K, Yan N. The recombinant expression systems for structure determination of eukaryotic membrane proteins. *Protein Cell.* 2014; 5(9):658–672. [PubMed: 25119489]
- Henderson R, Unwin PN. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature.* 1975; 257(5521):28–32. [PubMed: 1161000]
- Hitchman RB, Possee RD, King LA. Baculovirus expression systems for recombinant protein production in insect cells. *Recent Pat Biotechnol.* 2009; 3(1):46–54. [PubMed: 19149722]
- Hochuli E, Bannwarth W, Dobeli H, Gentz R, Stuber D. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Nat Biotechnol.* 1988; 6(11):1321–1325.
- Hollenberg CP, Gellissen G. Production of recombinant proteins by methylotrophic yeasts. *Curr Opin Biotechnol.* 1997; 8(5):554–560. [PubMed: 9353229]
- Hsu MF, Yu TF, Chou CC, Fu HY, Yang CS, Wang AH. Using *Haloarcula marismortui* bacteriorhodopsin as a fusion tag for enhancing and visible expression of integral membrane proteins in *Escherichia coli*. *PLoS One.* 2013; 8(2):e56363. [PubMed: 23457558]
- Huang L, Mao X, Abdulaev NG, Ngo T, Liu W, Ridge KD. One-step purification of a functional, constitutively activated form of visual arrestin. *Protein Expr Purif.* 2012; 82(1):55–60. [PubMed: 22133714]
- Huth JR, Bewley CA, Jackson BM, Hinnebusch AG, Clore GM, Gronenborn AM. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci.* 1997; 6(11):2359–2364. [PubMed: 9385638]
- Hwang PM, Kay LE. Solution structure and dynamics of integral membrane proteins by NMR: a case study involving the enzyme PagP. *Methods Enzymol.* 2005; 394:335–350. [PubMed: 15808227]
- Hwang PM, Pan JS, Sykes BD. A PagP fusion protein system for the expression of intrinsically disordered proteins in *Escherichia coli*. *Protein Expr Purif.* 2012; 85(1):148–151. [PubMed: 22841980]
- Hwang PM, Pan JS, Sykes BD. Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. *FEBS Lett.* 2014; 588(2):247–252. [PubMed: 24076468]
- Iost I, Guillerez J, Dreyfus M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J Bacteriol.* 1992; 174(2):619–622. [PubMed: 1729251]
- Ito K, Sugawara T, Shiroishi M, Tokuda N, Kurokawa A, Misaka T, Makyio H, Yurugi-Kobayashi T, Shimamura T, Nomura N, Murata T, Abe K, Iwata S, Kobayashi T. Advanced method for high-throughput expression of mutated eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun.* 2008; 371(4):841–845. [PubMed: 18474222]
- Jappelli R, Perrin MH, Lewis KA, Vaughan JM, Tzitzilonis C, Rivier JE, Vale WW, Riek R. Expression and functional characterization of membrane-integrated mammalian corticotropin releasing factor receptors 1 and 2 in *Escherichia coli*. *PLoS One.* 2014; 9(1):e84013. [PubMed: 24465390]
- Jenny RJ, Mann KG, Lundblad RL. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr Purif.* 2003; 31(1):1–11. [PubMed: 12963335]
- Jeong KJ, Lee SY. Secretory production of human leptin in *Escherichia coli*. *Biotechnol Bioeng.* 2000; 67(4):398–407. [PubMed: 10620755]
- Jeong KJ, Lee SY. Secretory production of human granulocyte colony-stimulating factor in *Escherichia coli*. *Protein Expr Purif.* 2001; 23(2):311–318. [PubMed: 11676607]

- Jewett MC, Calhoun KA, Voloshin A, Wu JJ, Swartz JR. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol.* 2008; 4:220. [PubMed: 18854819]
- Jidenko M, Lenoir G, Fuentes JM, le Maire M, Jaxel C. Expression in yeast and purification of a membrane protein, SERCA1a, using a biotinylated acceptor domain. *Protein Expr Purif.* 2006; 48(1):32–42. [PubMed: 16603381]
- Joubert O, Nehme R, Bidet M, Mus-Veteau I. Heterologous expression of human membrane receptors in the yeast *Saccharomyces cerevisiae*. *Methods Mol Biol.* 2010; 601:87–103. [PubMed: 20099141]
- Judge PJ, Taylor GF, Dannatt HR, Watts A. Solid-state nuclear magnetic resonance spectroscopy for membrane protein structure determination. *Methods Mol Biol.* 2015; 1261:331–347. [PubMed: 25502207]
- Jung HC, Lebeault JM, Pan JG. Surface display of *Zymomonas mobilis* levansucrase by using the ice-nucleation protein of *Pseudomonas syringae*. *Nat Biotechnol.* 1998; 16(6):576–580. [PubMed: 9624691]
- Junge F, Habersack S, Roos C, Stefer S, Proverbio D, Dotsch V, Bernhard F. Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins. *N Biotechnol.* 2011; 28(3):262–271. [PubMed: 20637904]
- Kaderbhai N, Kaderbhai MA. Expression, isolation, and characterization of a signal sequence-appended chimeric precursor protein. *Protein Expr Purif.* 1996; 7(3):237–246. [PubMed: 8860648]
- Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guntert P. Optimal isotope labelling for NMR protein structure determinations. *Nature.* 2006; 440(7080):52–57. [PubMed: 16511487]
- Kamezaki Y, Enomoto C, Ishikawa Y, Koyama T, Naya S-i, Suzuki T, Sakka K. The Dock tag, an affinity tool for the purification of recombinant proteins, based on the interaction between dockerin and cohesin domains from *Clostridium josui* cellulosome. *Protein Expr Purif.* 2010; 70(1):23–31. [PubMed: 19836451]
- Kang YS, Song JA, Han KY, Lee J. *Escherichia coli* EDA is a novel fusion expression partner to improve solubility of aggregation-prone heterologous proteins. *J Biotechnol.* 2015; 194:39–47. [PubMed: 25486632]
- Kapust RB, Waugh DS. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr Purif.* 2000; 19(2):312–318. [PubMed: 10873547]
- Kawate T, Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure.* 2006; 14(4):673–681. [PubMed: 16615909]
- Kefala G, Kwiatkowski W, Esquivies L, Maslennikov I, Choe S. Application of Mystic to improving the expression and membrane integration of histidine kinase receptors from *Escherichia coli*. *J Struct Funct Genomics.* 2007; 8(4):167–172. [PubMed: 17985211]
- Kim JS, Raines RT. Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Sci.* 1993; 2(3):348–356. [PubMed: 8453373]
- King K, Dohlman HG, Thorner J, Caron MG, Lefkowitz RJ. Control of yeast mating signal transduction by a mammalian beta 2-adrenergic receptor and Gs alpha subunit. *Science.* 1990; 250(4977):121–123. [PubMed: 2171146]
- Klammt C, Schwarz D, Lohr F, Schneider B, Dotsch V, Bernhard F. Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *FEBS J.* 2006; 273(18):4141–4153. [PubMed: 16930130]
- Kohn T, Kusunoki H, Sato K, Wakamatsu K. A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and $^{15}\text{N}/^{13}\text{C}$. *J Biomol NMR.* 1998; 12(1):109–121. [PubMed: 9729791]
- Kozma D, Simon I, Tusnady GE. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* 2013; 41(Database issue):D524–529. [PubMed: 23203988]
- Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weiss WI, Weiss J, Kobilka BK. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature.* 2012; 482(7386):552–556. [PubMed: 22358844]

- Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hubner H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia KC, Wess J, Kobilka BK. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature*. 2013; 504(7478):101–106. [PubMed: 24256733]
- Kuhlbrandt W. Cryo-EM enters a new era. *Elife*. 2014; 3:e03678. [PubMed: 25122623]
- Kuliopulos A, Walsh CT. Production, purification, and cleavage of tandem repeats of recombinant peptides. *J Am Chem Soc*. 1994; 116(11):4599–4607.
- Kumar A, Ward P, Katre UV, Mohanty S. A novel and simple method of production and biophysical characterization of a mini-membrane protein, Ost4p: a subunit of yeast oligosaccharyl transferase. *Biopolymers*. 2012; 97(7):499–507. [PubMed: 22302405]
- Kunji ER, Harding M, Butler PJ, Akamine P. Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography. *Methods*. 2008; 46(2):62–72. [PubMed: 18952172]
- Kurokawa Y, Yanagi H, Yura T. Overproduction of bacterial protein disulfide isomerase (DsbC) and its modulator (DsbD) markedly enhances periplasmic production of human nerve growth factor in *Escherichia coli*. *J Biol Chem*. 2001; 276(17):14393–14399. [PubMed: 11279016]
- Kuruma Y, Ueda T. The PURE system for the cell-free synthesis of membrane proteins. *Nat Protoc*. 2015; 10(9):1328–1344. [PubMed: 26270393]
- Langelaan DN, Reddy T, Banks AW, Dellaire G, Dupre DJ, Rainey JK. Structural features of the apelin receptor N-terminal tail and first transmembrane segment implicated in ligand binding and receptor trafficking. *Biochim Biophys Acta*. 2013; 1828(6):1471–1483. [PubMed: 23438363]
- LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Nat Biotechnol*. 1993; 11(2):187–193.
- Lebon G, Bennett K, Jazayeri A, Tate CG. Thermostabilisation of an agonist-bound conformation of the human adenosine A(2A) receptor. *J Mol Biol*. 2011; 409(3):298–310. [PubMed: 21501622]
- Lee JH, Kim JH, Hwang SW, Lee WJ, Yoon HK, Lee HS, Hong SS. High-level expression of antimicrobial peptide mediated by a fusion partner reinforcing formation of inclusion bodies. *Biochem Biophys Res Commun*. 2000; 277(3):575–580. [PubMed: 11061996]
- Lee JH, Lee JY, Song JA, Han KY, Lee DS, Lee J. A stress-responsive *Escherichia coli* protein, CysQ is a highly effective solubility enhancer for aggregation-prone heterologous proteins. *Protein Expr Purif*. 2014; 101:91–98. [PubMed: 24945073]
- Lee RP, Doughty SW, Ashman K, Walker J. Purification of hydrophobic integral membrane proteins from *Mycoplasma hyopneumoniae* by reversed-phase high-performance liquid chromatography. *J Chromatogr A*. 1996; 737(2):273–279. [PubMed: 8673256]
- Leviatan S, Sawada K, Moriyama Y, Nelson N. Combinatorial method for overexpression of membrane proteins in *Escherichia coli*. *J Biol Chem*. 2010; 285(31):23548–23556. [PubMed: 20525689]
- Li H, Lee S, Jap BK. Molecular design of aquaporin-1 water channel as revealed by electron crystallography. *Nat Struct Biol*. 1997; 4(4):263–265. [PubMed: 9095192]
- Li M, Hays FA, Roe-Zurz Z, Vuong L, Kelly L, Ho CM, Robbins RM, Pieper U, O'Connell JD 3rd, Miercke LJ, Giacomini KM, Sali A, Stroud RM. Selecting optimum eukaryotic integral membrane proteins for structure determination by rapid expression and solubilization screening. *J Mol Biol*. 2009; 385(3):820–830. [PubMed: 19061901]
- Li SJ, Hochstrasser M. A new protease required for cell-cycle progression in yeast. *Nature*. 1999; 398(6724):246–251. [PubMed: 10094048]
- Liao M, Cao E, Julius D, Cheng Y. Single particle electron cryo-microscopy of a mammalian ion channel. *Curr Opin Struct Biol*. 2014; 27:1–7. [PubMed: 24681231]
- Liebau J, Ye W, Maler L. Characterization of fast-tumbling isotropic bicelles by PFG diffusion NMR. *Magn Reson Chem*. 2015; In press. doi: 10.1002/mrc.4399
- Liu X, Huang A, Luo D, Liu H, Han H, Xu Y, Liang P. Use of adenylate kinase as a solubility tag for high level expression of T4 DNA ligase in *Escherichia coli*. *Protein Expr Purif*. 2015; 109:79–84. [PubMed: 25700573]

- Loll PJ. Membrane proteins, detergents and crystals: what is the state of the art? *Acta Crystallogr F Struct Biol Commun.* 2014; 70(Pt 12):1576–1583. [PubMed: 25484203]
- Londesborough J, Richard P, Valkonen M, Viljanen K. Effect of C-terminal protein tags on pentitol and L-arabinose transport by *Ambrosiozyma monospora* Lat1 and Lat2 transporters in *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 2014; 80(9):2737–2745. [PubMed: 24561586]
- Lounnas V, Ritschel T, Kelder J, McGuire R, Bywater RP, Foloppe N. Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Comput Struct Biotechnol J.* 2013; 5:e201302011. [PubMed: 24688704]
- Löwenadler B, Jansson B, Paleus S, Hougren E, Nilsson B, Moks T, Palm G, Josephson S, Philipson L, Uhlén M. A gene fusion system for generating antibodies against short peptides. *Gene.* 1987; 58(1):87–97. [PubMed: 3692175]
- Lundbaek JA. Regulation of membrane protein function by lipid bilayer elasticity-a single molecule technology to measure the bilayer properties experienced by an embedded protein. *J Phys Condens Matter.* 2006; 18(28):S1305–1344. [PubMed: 21690843]
- Lundbaek JA, Collingwood SA, Ingolfsson HI, Kapoor R, Andersen OS. Lipid bilayer regulation of membrane protein function: gramicidin channels as molecular force probes. *J R Soc Interface.* 2010; 7(44):373–395. [PubMed: 19940001]
- Majtan T, Kraus JP. Folding and activity of mutant cystathionine beta-synthase depends on the position and nature of the purification tag: characterization of the R266K CBS mutant. *Protein Expr Purif.* 2012; 82(2):317–324. [PubMed: 2233527]
- Makrides SC. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol Rev.* 1996; 60(3):512–538. [PubMed: 8840785]
- Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics.* 2004; 5(1–2):75–86. [PubMed: 15263846]
- Malik A, Rudolph R, Sohling B. A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm. *Protein Expr Purif.* 2006; 47(2):662–671. [PubMed: 16600627]
- Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL, Prosser RS, Kobilka BK. Structural insights into the dynamic process of beta2-adrenergic receptor signaling. *Cell.* 2015; 161(5):1101–1111. [PubMed: 25981665]
- Marblestone JG, Edavettal SC, Lim Y, Lim P, Zuo X, Butt TR. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* 2006; 15(1):182–189. [PubMed: 16322573]
- Marino J, Bordag N, Keller S, Zerbe O. Mystic's membrane association and its assistance in overexpression of a human GPCR are independent processes. *Protein Sci.* 2015; 24(1):38–48. [PubMed: 25297828]
- Marino J, Geertsma ER, Zerbe O. Topogenesis of heterologously expressed fragments of the human Y4 GPCR. *Biochim Biophys Acta.* 2012; 1818(12):3055–3063. [PubMed: 22867850]
- Marley J, Lu M, Bracken C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR.* 2001; 20(1):71–75. [PubMed: 11430757]
- Marsh D. Lateral pressure profile, spontaneous curvature frustration, and the incorporation and conformation of proteins in membranes. *Biophys J.* 2007; 93(11):3884–3899. [PubMed: 17704167]
- Matrai J, Chuah MK, VandenDriessche T. Recent advances in lentiviral vector development and applications. *Mol Ther.* 2010; 18(3):477–490. [PubMed: 20087315]
- Maurice S, Dekel M, Shoseyov O, Gertler A. Cellulose beads bound to cellulose binding domain-fused recombinant proteins; an adjuvant system for parenteral vaccination of fish. *Vaccine.* 2003; 21(23):3200–3207. [PubMed: 12804848]
- Mikami S, Masutani M, Sonenberg N, Yokoyama S, Imataka H. An efficient mammalian cell-free translation system supplemented with translation factors. *Protein Expr Purif.* 2006; 46(2):348–357. [PubMed: 16289705]

- Munro S, Pelham HR. Use of peptide tagging to detect proteins expressed from cloned genes: deletion mapping functional domains of *Drosophila* hsp 70. *EMBO J.* 1984; 3(13):3087–3093. [PubMed: 6526011]
- Murray DT, Das N, Cross TA. Solid state NMR strategy for characterizing native membrane protein structures. *Acc Chem Res.* 2013; 46(9):2172–2181. [PubMed: 23470103]
- Neale C, Ghanei H, Holyoake J, Bishop RE, Prive GG, Pomes R. Detergent-mediated protein aggregation. *Chem Phys Lipids.* 2013; 169:72–84. [PubMed: 23466535]
- Neophytou I, Harvey R, Lawrence J, Marsh P, Panaretou B, Barlow D. Eukaryotic integral membrane protein expression utilizing the *Escherichia coli* glycerol-conducting channel protein (GlpF). *Appl Microbiol Biotechnol.* 2007; 77(2):375–381. [PubMed: 17828601]
- Neutze R, Branden G, Schertler GF. Membrane protein structural biology using X-ray free electron lasers. *Curr Opin Struct Biol.* 2015; 33:115–125. [PubMed: 26342349]
- Newstead S, Ferrandon S, Iwata S. Rationalizing alpha-helical membrane protein crystallization. *Protein Sci.* 2008; 17(3):466–472. [PubMed: 18218713]
- Newstead S, Kim H, von Heijne 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(35):13936–13941. [PubMed: 17709746]
- Nietlispach D, Gautier A. Solution NMR studies of polytopic alpha-helical membrane proteins. *Curr Opin Struct Biol.* 2011; 21(4):497–508. [PubMed: 21775128]
- Nirenberg MW, Matthaei JH. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A.* 1961; 47:1588–1602. [PubMed: 14479932]
- Nizard P, Chenal A, Beaumelle B, Fourcade A, Gillet D. Prolonged display or rapid internalization of the IgG-binding protein ZZ anchored to the surface of cells using the diphtheria toxin T domain. *Protein Eng.* 2001; 14(6):439–446. [PubMed: 11477224]
- Nyblom M, Oberg F, Lindkvist-Petersson K, Hallgren K, Findlay H, Wikstrom J, Karlsson A, Hansson O, Booth PJ, Bill RM, Neutze R, Hedfalk K. Exceptional overproduction of a functional human membrane protein. *Protein Expr Purif.* 2007; 56(1):110–120. [PubMed: 17869538]
- Nygaard R, Zou Y, Dror RO, Mildorf TJ, Arlow DH, Manglik A, Pan AC, Liu CW, Fung JJ, Bokoch MP, Thian FS, Kobilka TS, Shaw DE, Mueller L, Prosser RS, Kobilka BK. The dynamic process of beta(2)-adrenergic receptor activation. *Cell.* 2013; 152(3):532–542. [PubMed: 23374348]
- Ohana RF, Encell LP, Zhao K, Simpson D, Slater MR, Urh M, Wood KV. HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expr Purif.* 2009; 68(1):110–120. [PubMed: 19464373]
- Olins PO, Devine CS, Rangwala SH, Kavka KS. The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene.* 1988; 73(1):227–235. [PubMed: 3072257]
- Opella SJ. Solid-state NMR and membrane proteins. *J Magn Reson.* 2015; 253:129–137. [PubMed: 25681966]
- Opitz C, Isogai S, Grzesiek S. An economic approach to efficient isotope labeling in insect cells using homemade ¹⁵N-, ¹³C- and ²H-labeled yeast extracts. *J Biomol NMR.* 2015; 62(3):373–385. [PubMed: 26070442]
- Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov.* 2006; 5(12):993–996. [PubMed: 17139284]
- Ozawa K, Wu PS, Dixon NE, Otting G. N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. *FEBS J.* 2006; 273(18):4154–4159. [PubMed: 16930129]
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science.* 2000; 289(5480):739–745. [PubMed: 10926528]
- Pandey A, Sarker M, Liu XQ, Rainey JK. Small expression tags enhance bacterial expression of the first three transmembrane segments of the apelin receptor. *Biochem Cell Biol.* 2014; 92(4):269–278. [PubMed: 24943103]

- Park SH, Casagrande F, Chu M, Maier K, Kiefer H, Opella SJ. Optimization of purification and refolding of the human chemokine receptor CXCR1 improves the stability of proteoliposomes for structure determination. *Biochim Biophys Acta*. 2012a; 1818(3):584–591. [PubMed: 22024025]
- Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM, Opella SJ. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature*. 2012b; 491(7426):779–783. [PubMed: 23086146]
- Park TJ, Choi SS, Gang GA, Kim Y. High-level expression and purification of the second transmembrane domain of wild-type and mutant human melanocortin-4 receptor for solid-state NMR structural studies. *Protein Expr Purif*. 2008; 62(2):139–145. [PubMed: 18809499]
- Parker MJ, Aulton-Jones M, Hounslow AM, Craven CJ. A combinatorial selective labeling method for the assignment of backbone amide NMR resonances. *J Am Chem Soc*. 2004; 126(16):5020–5021. [PubMed: 15099056]
- Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal Biochem*. 1994; 216(2):413–417. [PubMed: 8179197]
- Pedersen BP, Nissen P. Membrane proteins - do we catch up with the breathless pace of soluble protein structural biology? *Biochim Biophys Acta*. 2015; 1850(3):447–448. [PubMed: 25598513]
- Periasamy A, Shadiac N, Amalraj A, Garajova S, Nagarajan Y, Waters S, Mertens HD, Hrmova M. Cell-free protein synthesis of membrane (1,3)-beta-d-glucan (curdian) synthase: co-translational insertion in liposomes and reconstitution in nanodiscs. *Biochim Biophys Acta*. 2013; 1828(2):743–757. [PubMed: 23063656]
- Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A*. 1997; 94(23):12366–12371. [PubMed: 9356455]
- Pestov NB, Rydstrom J. Purification of recombinant membrane proteins tagged with calmodulin-binding domains by affinity chromatography on calmodulin-agarose: example of nicotinamide nucleotide transhydrogenase. *Nat Protoc*. 2007; 2(1):198–202. [PubMed: 17401354]
- Phillips R, Ursell T, Wiggins P, Sens P. Emerging roles for lipids in shaping membrane-protein function. *Nature*. 2009; 459(7245):379–385. [PubMed: 19458714]
- Poulos S, Morgan JL, Zimmer J, Faham S. Bicelles coming of age: an empirical approach to bicelle crystallization. *Methods Enzymol*. 2015; 557:393–416. [PubMed: 25950975]
- Pratap J, Dikshit KL. Effect of signal peptide changes on the extracellular processing of streptokinase from *Escherichia coli*: requirement for secondary structure at the cleavage junction. *Mol Gen Genet*. 1998; 258(4):326–333. [PubMed: 9648736]
- Prive GG. Detergents for the stabilization and crystallization of membrane proteins. *Methods*. 2007; 41(4):388–397. [PubMed: 17367711]
- Prosser RS, Evanics F, Kitevski JL, Al-Abdul-Wahid MS. Current applications of bicelles in NMR studies of membrane-associated amphiphiles and proteins. *Biochemistry*. 2006; 45(28):8453–8465. [PubMed: 16834319]
- Pryor KD, Leiting B. High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr Purif*. 1997; 10(3):309–319. [PubMed: 9268677]
- Pucadyil TJ, Chattopadhyay A. Role of cholesterol in the function and organization of G-protein coupled receptors. *Prog Lipid Res*. 2006; 45(4):295–333. [PubMed: 16616960]
- Qiu J, Swartz JR, Georgiou G. Expression of Active Human Tissue-Type Plasminogen Activator in *Escherichia coli*. *Appl Environ Microbiol*. 1998; 64(12):4891–4896. [PubMed: 9835579]
- Quick M, Wright EM. Employing *Escherichia coli* to functionally express, purify, and characterize a human transporter. *Proc Natl Acad Sci U S A*. 2002; 99(13):8597–8601. [PubMed: 12077304]
- Rahman M, Ismat F, McPherson MJ, Baldwin SA. Topology-informed strategies for the overexpression and purification of membrane proteins. *Mol Membr Biol*. 2007; 24(5–6):407–418. [PubMed: 17710645]
- Ramon A, Marin M. Advances in the production of membrane proteins in *Pichia pastoris*. *Biotechnol J*. 2011; 6(6):700–706. [PubMed: 21567964]

- Rao XC, Li S, Hu JC, Jin XL, Hu XM, Huang JJ, Chen ZJ, Zhu JM, Hu FQ. A novel carrier molecule for high-level expression of peptide antibiotics in *Escherichia coli*. *Protein Expr Purif*. 2004; 36(1):11–18. [PubMed: 15177279]
- Raran-Kurussi S, Waugh DS. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS One*. 2012; 7(11):e49589. [PubMed: 23166722]
- Rasmussen SGF, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK. Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature*. 2011; 469(7329):175–180. [PubMed: 21228869]
- Reckel S, Gottstein D, Stehle J, Lohr F, Verhoeven MK, Takeda M, Silvers R, Kainosho M, Glaubitz C, Wachtveitl J, Bernhard F, Schwalbe H, Guntert P, Dotsch V. Solution NMR structure of proteorhodopsin. *Angew Chem Int Ed Engl*. 2011; 50(50):11942–11946. [PubMed: 22034093]
- Reddi H, Bhattacharya A, Kumar V. The calcium-binding protein of *Entamoeba histolytica* as a fusion partner for expression of peptides in *Escherichia coli*. *Biotechnol Appl Biochem*. 2002; 36(3): 213–218. [PubMed: 12452805]
- Reeves PJ, Callewaert N, Contreras R, Khorana HG. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc Natl Acad Sci U S A*. 2002a; 99(21):13419–13424. [PubMed: 12370423]
- Reeves PJ, Kim JM, Khorana HG. Structure and function in rhodopsin: a tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proc Natl Acad Sci U S A*. 2002b; 99(21):13413–13418. [PubMed: 12370422]
- Rondahl H, Nilsson B, Holmgren E. Fusions to the 5' end of a gene encoding a two-domain analogue of staphylococcal protein A. *J Biotechnol*. 1992; 25(3):269–287. [PubMed: 1368804]
- Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*. 2014; 5:172. [PubMed: 24860555]
- Ruskin RS, Yu Z, Grigorieff N. Quantitative characterization of electron detectors for transmission electron microscopy. *J Struct Biol*. 2013; 184(3):385–393. [PubMed: 24189638]
- Sabaty M, Grosse S, Adryanczyk G, Boiry S, Biaso F, Arnoux P, Pignol D. Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis. *BMC Biochem*. 2013; 14:28. [PubMed: 24180491]
- Sanchez J, Hirst TR, Uhlin BE. Hybrid enterotoxin LTA::STa proteins and their protection from degradation by in vivo association with B-subunits of *Escherichia coli* heat-labile enterotoxin. *Gene*. 1988; 64(2):265–275. [PubMed: 2841198]
- Sander P, Grunewald S, Bach M, Haase W, Reilander H, Michel H. Heterologous expression of the human D2S dopamine receptor in protease-deficient *Saccharomyces cerevisiae* strains. *Eur J Biochem*. 1994a; 226(2):697–705. [PubMed: 8001587]
- Sander P, Grunewald S, Reilander H, Michel H. Expression of the human D2S dopamine receptor in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*: a comparative study. *FEBS Lett*. 1994b; 344(1):41–46. [PubMed: 8181561]
- Sarramegna V, Talmont F, Seree de Roch M, Milon A, Demange P. Green fluorescent protein as a reporter of human mu-opioid receptor overexpression and localization in the methylotrophic yeast *Pichia pastoris*. *J Biotechnol*. 2002; 99(1):23–39. [PubMed: 12204555]
- Sastry M, Bewley CA, Kwong PD. Mammalian expression of isotopically labeled proteins for NMR spectroscopy. *Adv Exp Med Biol*. 2012; 992:197–211. [PubMed: 23076586]
- Schmidt TGM, Koepke J, Frank R, Skerra A. Molecular interaction between the strep-tag affinity peptide and its cognate target, streptavidin. *J Mol Biol*. 1996; 255(5):753–766. [PubMed: 8636976]
- Schwaiger M, Lebendiker M, Yerushalmi H, Coles M, Groger A, Schwarz C, Schuldiner S, Kessler H. NMR investigation of the multidrug transporter EmrE, an integral membrane protein. *Eur J Biochem*. 1998; 254(3):610–619. [PubMed: 9688273]
- Schwarz D, Dotsch V, Bernhard F. Production of membrane proteins using cell-free expression systems. *Proteomics*. 2008; 8(19):3933–3946. [PubMed: 18763710]

- Schwarz D, Junge F, Durst F, Frolich N, Schneider B, Reckel S, Sobhanifar S, Dotsch V, Bernhard F. Preparative scale expression of membrane proteins in *Escherichia coli*-based continuous exchange cell-free systems. *Nat Protoc.* 2007; 2(11):2945–2957. [PubMed: 18007631]
- Seddon AM, Curnow P, Booth PJ. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta.* 2004; 1666(1–2):105–117. [PubMed: 15519311]
- Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG. Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc Natl Acad Sci U S A.* 2008; 105(3):877–882. [PubMed: 18192400]
- Sezonov G, Joseleau-Petit D, D'Ari R. *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol.* 2007; 189(23):8746–8749. [PubMed: 17905994]
- Shi X, Jarvis DL. Protein N-glycosylation in the baculovirus-insect cell system. *Curr Drug Targets.* 2007; 8(10):1116–1125. [PubMed: 17979671]
- Shibata Y, White JF, Serrano-Vega MJ, Magnani F, Aloia AL, Grisshammer R, Tate CG. Thermostabilization of the neurotensin receptor NTS1. *J Mol Biol.* 2009; 390(2):262–277. [PubMed: 19422831]
- Shigematsu H, Sigworth FJ. Noise models and cryo-EM drift correction with a direct-electron camera. *Ultramicroscopy.* 2013; 131:61–69. [PubMed: 23748163]
- Shiloach J, Fass R. Growing *E. coli* to high cell density--a historical perspective on method development. *Biotechnol Adv.* 2005; 23(5):345–357. [PubMed: 15899573]
- Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T. Cell-free translation reconstituted with purified components. *Nat Biotechnol.* 2001; 19(8):751–755. [PubMed: 11479568]
- Singh P, Sharma L, Kulothungan SR, Adkar BV, Prajapati RS, Ali PS, Krishnan B, Varadarajan R. Effect of signal peptide on stability and folding of *Escherichia coli* thioredoxin. *PLoS One.* 2013; 8(5):e63442. [PubMed: 23667620]
- Sitarska A, Skora L, Klopp J, Roest S, Fernandez C, Shrestha B, Gossert AD. Affordable uniform isotope labeling with (2)H, (13)C and (15)N in insect cells. *J Biomol NMR.* 2015; 62(2):191–197. [PubMed: 25929326]
- Smith DB, Johnson KS. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene.* 1988; 67(1):31–40. [PubMed: 3047011]
- Smith MT, Rubinstein JL. Structural biology. Beyond blob-ology. *Science.* 2014; 345(6197):617–619. [PubMed: 25104368]
- Smith SM. Strategies for the purification of membrane proteins. *Methods Mol Biol.* 2011; 681:485–496. [PubMed: 20978985]
- Soares CR, Gomide FI, Ueda EK, Bartolini P. Periplasmic expression of human growth hormone via plasmid vectors containing the lambdaPL promoter: use of HPLC for product quantification. *Protein Eng.* 2003; 16(12):1131–1138. [PubMed: 14983096]
- Song JA, Lee DS, Park JS, Han KY, Lee J. A novel *Escherichia coli* solubility enhancer protein for fusion expression of aggregation-prone heterologous proteins. *Enzyme Microb Technol.* 2011; 49(2):124–130. [PubMed: 22112398]
- Sorensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol.* 2005; 115(2):113–128. [PubMed: 15607230]
- Sørensen HP, Sperling-Petersen HU, Mortensen KK. A favorable solubility partner for the recombinant expression of streptavidin. *Protein Expr Purif.* 2003; 32(2):252–259. [PubMed: 14965771]
- Sreenath HK, Bingman CA, Buchan BW, Seder KD, Burns BT, Geetha HV, Jeon WB, Vojtik FC, Aceti DJ, Frederick RO, Phillips GN Jr, Fox BG. Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr Purif.* 2005; 40(2):256–267. [PubMed: 15766867]
- Srivastava A, Yano J, Hirozane Y, Kefala G, Gruswitz F, Snell G, Lane W, Ivetac A, Aertgeerts K, Nguyen J, Jennings A, Okada K. High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature.* 2014; 513(7516):124–127. [PubMed: 25043059]
- Stanley P. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol Cell Biol.* 1989; 9(2):377–383. [PubMed: 2710109]

- Stevens RC. Design of high-throughput methods of protein production for structural biology. *Structure*. 2000; 8(9):R177–185. [PubMed: 10986469]
- Stewart MQ, Esposito RD, Gowani J, Goodman JM. Alcohol oxidase and dihydroxyacetone synthase, the abundant peroxisomal proteins of methylotrophic yeasts, assemble in different cellular compartments. *J Cell Sci*. 2001; 114(Pt 15):2863–2868. [PubMed: 11683419]
- Stols L, Millard CS, Dementieva I, Donnelly MI. Production of selenomethionine-labeled proteins in two-liter plastic bottles for structure determination. *J Struct Funct Genomics*. 2004; 5(1–2):95–102. [PubMed: 15263848]
- Studier FW. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol*. 1991; 219(1):37–44. [PubMed: 2023259]
- Su PC, Si W, Baker DL, Berger BW. High-yield membrane protein expression from *E. coli* using an engineered outer membrane protein F fusion. *Protein Sci*. 2013; 22(4):434–443. [PubMed: 23345122]
- Sui H, Walian PJ, Tang G, Oh A, Jap BK. Crystallization and preliminary X-ray crystallographic analysis of water channel AQP1. *Acta Crystallogr D Biol Crystallogr*. 2000; 56(Pt 9):1198–1200. [PubMed: 10957645]
- Suzuki KG, Kasai RS, Fujiwara TK, Kusumi A. Single-molecule imaging of receptor-receptor interactions. *Methods Cell Biol*. 2013; 117:373–390. [PubMed: 24143988]
- Takeda M, Chang CK, Ikeya T, Guntert P, Chang YH, Hsu YL, Huang TH, Kainosho M. Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J Mol Biol*. 2008; 380(4):608–622. [PubMed: 18561946]
- Takegawa K, Tohda H, Sasaki M, Idiris A, Ohashi T, Mukaiyama H, Giga-Hama Y, Kumagai H. Production of heterologous proteins using the fission-yeast (*Schizosaccharomyces pombe*) expression system. *Biotechnol Appl Biochem*. 2009; 53(Pt 4):227–235. [PubMed: 19531030]
- Tate CG. Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett*. 2001; 504(3):94–98. [PubMed: 11532439]
- Tautermann CS. GPCR structures in drug design, emerging opportunities with new structures. *Bioorg Med Chem Lett*. 2014; 24(17):4073–4079. [PubMed: 25086683]
- Tiralongo J, Maggioni A. The targeted expression of nucleotide sugar transporters to the *E. coli* inner membrane. *Methods Mol Biol*. 2011; 705:237–249. [PubMed: 21125390]
- Tomme P, Boraston A, McLean B, Kormos J, Creagh AL, Sturch K, Gilkes NR, Haynes CA, Warren RAJ, Kilburn DG. Characterization and affinity applications of cellulose-binding domains I. *J Chromatogr B*. 1998; 715(1):283–296.
- Tonelli M, Singarapu KK, Makino S, Sahu SC, Matsubara Y, Endo Y, Kainosho M, Markley JL. Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J Biomol NMR*. 2011; 51(4):467–476. [PubMed: 21984356]
- Toyoshima C, Nakasako M, Nomura H, Ogawa H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature*. 2000; 405(6787):647–655. [PubMed: 10864315]
- Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. Structures of metal sites of oxidized bovine heart cytochrome c oxidase at 2.8 Å. *Science*. 1995; 269(5227):1069–1074. [PubMed: 7652554]
- Tugarinov V, Muhandiram R, Ayed A, Kay LE. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc*. 2002; 124(34):10025–10035. [PubMed: 12188667]
- Tyler RC, Sreenath HK, Singh S, Aceti DJ, Bingman CA, Markley JL, Fox BG. Auto-induction medium for the production of [U-¹⁵N]- and [U-¹³C, U-¹⁵N]-labeled proteins for NMR screening and structure determination. *Protein Expr Purif*. 2005; 40(2):268–278. [PubMed: 15766868]
- Unger T, Peleg Y. Recombinant protein expression in the baculovirus-infected insect cell system. *Methods Mol Biol*. 2012; 800:187–199. [PubMed: 21964790]
- Vaillancourt P, Zheng CF, Hoang DQ, Breister L. Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides. *Methods Enzymol*. 2000; 326:340–362. [PubMed: 11036652]

- van Bloois E, Winter RT, Kolmar H, Fraaije MW. Decorating microbes: surface display of proteins on *Escherichia coli*. Trends Biotechnol. 2011; 29(2):79–86. [PubMed: 21146237]
- van den Burg HA, de Wit PJ, Vervoort J. Efficient $^{13}\text{C}/^{15}\text{N}$ double labeling of the avirulence protein AVR4 in a methanol-utilizing strain (Mut+) of *Pichia pastoris*. J Biomol NMR. 2001; 20(3):251–261. [PubMed: 11519748]
- Vervecken W, Callewaert N, Kaigorodov V, Geysens S, Contreras R. Modification of the N-glycosylation pathway to produce homogeneous, human-like glycans using GlycoSwitch plasmids. Methods Mol Biol. 2007; 389:119–138. [PubMed: 17951639]
- Vidovic V, Prongidi-Fix L, Bechinger B, Werten S. Production and isotope labeling of antimicrobial peptides in *Escherichia coli* by means of a novel fusion partner that enables high-yield insoluble expression and fast purification. J Pept Sci. 2009; 15(4):278–284. [PubMed: 19189273]
- Vinothkumar KR, Henderson R. Structures of membrane proteins. Q Rev Biophys. 2010; 43(1):65–158. [PubMed: 20667175]
- von Heijne G. The membrane protein universe: what's out there and why bother? J Intern Med. 2007; 261(6):543–557. [PubMed: 17547710]
- Von Jagow, G., Link, TA., Schägger, H. Purification strategies for membrane proteins. In: Jagow, GV., Schägger, H., editors. A Practical Guide to Membrane Protein Purification. Academic Press; San Diego: 1994. p. 3–21.
- Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Hogbom M, van Wijk KJ, Slotboom DJ, Persson JO, de Gier JW. Tuning *Escherichia coli* for membrane protein overexpression. Proc Natl Acad Sci U S A. 2008; 105(38):14371–14376. [PubMed: 18796603]
- Walz T, Hirai T, Murata K, Heymann JB, Mitsuoka K, Fujiyoshi Y, Smith BL, Agre P, Engel A. The three-dimensional structure of aquaporin-1. Nature. 1997; 387(6633):624–627. [PubMed: 9177353]
- Wang S, Munro RA, Shi L, Kawamura I, Okitsu T, Wada A, Kim SY, Jung KH, Brown LS, Ladizhansky V. Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. Nat Methods. 2013; 10(10):1007–1012. [PubMed: 24013819]
- Weiss HM, Haase W, Michel H, Reilander H. Expression of functional mouse 5-HT_{5A} serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization. FEBS Lett. 1995; 377(3):451–456. [PubMed: 8549774]
- Weiss MS, Wacker T, Weckesser J, Welte W, Schulz GE. The three-dimensional structure of porin from *Rhodobacter capsulatus* at 3 Å resolution. FEBS Lett. 1990; 267(2):268–272. [PubMed: 2165921]
- Whiteman P, Redfield C, Handford PA. Bacterial expression and in vitro refolding of limited fragments of the Notch receptor and its ligands. Methods Mol Biol. 2014; 1187:193–208. [PubMed: 25053491]
- Winzler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentale E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science. 1999; 285(5429):901–906. [PubMed: 10436161]
- Wood MJ, Komives EA. Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation. J Biomol NMR. 1999; 13(2):149–159. [PubMed: 10070756]
- Xu R, Du P, Fan JJ, Zhang Q, Li TP, Gan RB. High-level expression and secretion of recombinant mouse endostatin by *Escherichia coli*. Protein Expr Purif. 2002; 24(3):453–459. [PubMed: 11922762]
- Yeliseev A, Zoubak L, Gawrisch K. Use of dual affinity tags for expression and purification of functional peripheral cannabinoid receptor. Protein Expr Purif. 2007; 53(1):153–163. [PubMed: 17223358]

- Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. Drug-target network. *Nat Biotechnol.* 2007; 25(10):1119–1126. [PubMed: 17921997]
- Yim SK, Kim DH, Jung HC, Pan JG, Kang HS, Ahn T, Yun CH. Surface display of heme- and diflavin-containing cytochrome P450 BM3 in *Escherichia coli*: a whole cell biocatalyst for oxidation. *J Microbiol Biotechnol.* 2010; 20(4):712–717. [PubMed: 20467243]
- Yin J, Mobarec JC, Kolb P, Rosenbaum DM. Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. *Nature.* 2015; 519(7542):247–250. [PubMed: 25533960]
- Yun JH, Kim M, Kim K, Lee D, Jung Y, Oh D, Ko YJ, Cho AE, Cho HS, Lee W. Solution structure of the transmembrane 2 domain of the human melanocortin-4 receptor in sodium dodecyl sulfate (SDS) micelles and the functional implication of the D90N mutant. *Biochim Biophys Acta.* 2015; 1848(6):1294–1302. [PubMed: 25753114]
- Zeenko VV, Wang C, Majumder M, Komar AA, Snider MD, Merrick WC, Kaufman RJ, Hatzoglou M. An efficient in vitro translation system from mammalian cells lacking the translational inhibition caused by eIF2 phosphorylation. *RNA.* 2008; 14(3):593–602. [PubMed: 18230759]
- Zhang D, Gao ZG, Zhang K, Kiselev E, Crane S, Wang J, Paoletta S, Yi C, Ma L, Zhang W, Han GW, Liu H, Cherezov V, Katritch V, Jiang H, Stevens RC, Jacobson KA, Zhao Q, Wu B. Two disparate ligand-binding sites in the human P2Y1 receptor. *Nature.* 2015a; 520(7547):317–321. [PubMed: 25822790]
- Zhang H, Unal H, Gati C, Han GW, Liu W, Zatspein NA, James D, Wang D, Nelson G, Weierstall U, Sawaya MR, Xu Q, Messerschmidt M, Williams GJ, Boutet S, Yefanov OM, White TA, Wang C, Ishchenko A, Tirupula KC, Desnoyer R, Coe J, Conrad CE, Fromme P, Stevens RC, Katritch V, Karnik SS, Cherezov V. Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. *Cell.* 2015b; 161(4):833–844. [PubMed: 25913193]
- Zhao J, Benlekbir S, Rubinstein JL. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature.* 2015; 521(7551):241–245. [PubMed: 25971514]
- Zhou A, Rohou A, Schep DG, Bason JV, Montgomery MG, Walker JE, Grigorieff N, Rubinstein JL. Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. *Elife.* 2015;4.
- Zhou P, Lugovskoy AA, Wagner G. A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J Biomol NMR.* 2001; 20(1):11–14. [PubMed: 11430750]
- Zhou ZH, Chiu W. Prospects for using an IVEM with a FEG for imaging macromolecules towards atomic resolution. *Ultramicroscopy.* 1993; 49(1–4):407–416. [PubMed: 8475604]
- Zou Z, Cao L, Zhou P, Su Y, Sun Y, Li W. Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *J Biotechnol.* 2008; 135(4):333–339. [PubMed: 18599143]
- Zuo X, Li S, Hall J, Mattern MR, Tran H, Shoo J, Tan R, Weiss SR, Butt TR. Enhanced expression and purification of membrane proteins by SUMO fusion in *Escherichia coli*. *J Struct Funct Genomics.* 2005a; 6(2–3):103–111. [PubMed: 16211506]
- Zuo X, Mattern MR, Tan R, Li S, Hall J, Sterner DE, Shoo J, Tran H, Lim P, Sarafianos SG, Kazi L, Navas-Martin S, Weiss SR, Butt TR. Expression and purification of SARS coronavirus proteins using SUMO-fusions. *Protein Expr Purif.* 2005b; 42(1):100–110. [PubMed: 15939295]

Figure 1. Source of protein for all 2631 experimentally solved membrane protein structures currently available in the PDBTM. (Numbers based upon annotations in the PDB file headers, with “unspecified source” based only on the PDB file header.)

Figure 2.

Expression system employed for all solved membrane protein structures produced by recombinant expression currently available in the PDBTM. (Numbers based upon annotations in the PDB file headers; hence, the total of 1719 is slightly lower than the 1855 implied by Figure 1.)

Figure 3. Flowchart outlining various optimization steps allowing for structural characterization of membrane proteins by XRD, NMR spectroscopy, or cryo-EM.

Table 1

List of available databases and resources on membrane protein structure.

Resource description	URL
Membrane Proteins of Known 3D Structure (mpstruc database)	http://blanco.biomol.uci.edu/mpstruc/
PDBTM: Protein Data Bank of Transmembrane Proteins	http://pdbtm.enzim.hu/
Membrane Protein Explorer (MPEx)	http://blanco.biomol.uci.edu/mpex/
Membrane Proteins Of Known Structure Determined By NMR	http://www.drorlist.com/nmr/MPNMR.html
Detection of Transmembrane Regions by Using 3D Structure of Proteins (TMDet)	http://tmdet.enzim.hu/
IUPHAR Database	http://www.guidetopharmacology.org/
Transporter Classification Database	http://www.tcdb.org/
Experimentally Solved GPCR Structures (GPCR-exp)	http://zhanglab.ccmb.med.umich.edu/GPCR-EXP/
GPCR Database (GPCRDB)	http://www.gpcr.org/7tm/
Plant Membrane Protein Database	http://aramemnon.uni-koeln.de/
Human Membrane Protein Analysis System	http://fcode.kaist.ac.kr/hmpas/

Table 2

Pros and cons of various membrane protein expression systems.

Expression system	Pros	Cons
<i>Escherichia coli</i>	High expression levels Low cost Simple culture conditions Rapid growth Scaleable Simple transformation protocols Many parameters can be altered to optimise expression Least expensive for NMR-active isotope labelling Engineered strains can help alleviate the problems with disulfide bond formation (Shuffle and Origami) and codon bias (Rossetta and CodonPlus RIL/RP)	Inefficient disulfide bond formation Poor folding of proteins in the cytoplasm sometimes leading to inclusion body formation Codon usage different to eukaryotes Minimal post-translational modifications (systems available to allow some modifications) Endotoxin present
<i>Saccharomyces cerevisiae</i>	Good expression levels Choice of secreted or cellular expression Low cost Simple culture conditions Scaleable Able to perform most eukaryotic post-translational modifications Efficient protein folding Endotoxin-free	Likely lower expression than with <i>Pichia pastoris</i> Secretion likely lower than with <i>Pichia pastoris</i> Glycosylation differs from mammalian cells Tendency to hyperglycosylate proteins N-glycan structures considered allergenic
<i>Pichia pastoris</i>	High expression levels Low cost Simple culture conditions Relatively rapid growth Scaleable Choice of secreted or intracellular expression Extensive post-translational modification of proteins Relatively inexpensive NMR-active isotope labeling Efficient protein folding N-glycosylation more similar to higher eukaryotes than with <i>Saccharomyces cerevisiae</i> Endotoxin-free	Use of methanol as inducer is a safety (fire) hazard at scale Glycosylation differs from mammalian cells
Baculovirus-infected insect cells	Good expression levels Relatively rapid growth Efficient protein folding Moderately scaleable Extensive post-translational modification of proteins Glycosylation more like mammalian cells Relatively easy enzymatic deglycosylation Endotoxin-free	Expensive culture media Large volumes of virus needed on scale-up Inefficient processing of pro-peptides in secretory pathway Glycosylation still different to mammalian cells Viral infection leads to cell lysis and potential degradation of expressed proteins
Mammalian cells	Good expression levels Moderately scaleable Suspension-adapted cells facilitate scale-up Efficient protein folding All post-translational modifications Endotoxin-free	Expensive culture media Complex growth requirements
Cell-free protein production	<i>E.coli</i> , wheat germ, insect and mammalian systems commercially available Scaleable Protein synthesis conditions can be manipulated Can readily incorporate non-natural amino acids Easy to vary isotope labeling patterns for NMR spectroscopy Can use PCR products as template Endotoxin-free	Limited post-translational modifications Relatively expensive (however, protein yield per unit volume may overcome this)

Table 3

Properties and applications of commonly used fusion tags sorted by class of use in alphabetical order.

Tag (Reference(s))	Size*	Organism	Tag placement	Use	Used in membrane protein expression in [expression system]?
Calmodulin binding domain (Vaillancourt et al. 2000)	26 aa	<i>Mammalian</i>	N- or C-terminal	Purification	Yes (Pestov and Rydstrom 2007) [<i>E. coli</i>]
Cellulose binding module (Tomme et al. 1998)	4–20 kDa	<i>Synthetic</i>	N- or C-terminal	Purification	Yes (Maurice et al. 2003) [fish]
c-Myc (Ellison and Hochstrasser 1991)	11 aa	<i>H. sapien</i>	N- or C-terminal	Purification	Yes (Sarramegna et al. 2002) [yeast]
Dock (dockerin domain) (Kamezaki et al. 2010)	22 aa	<i>C. josui</i>	N-or C-terminal	Purification	No
FLAG (Munro and Pelham 1984)	8 aa	<i>Synthetic</i>	N- or C-terminal	Purification	Yes (Quick and Wright 2002) [<i>E. coli</i>]
HA (hemagglutinin) (Field et al. 1988)	9 aa	<i>Influenza Virus</i>	N- or C-terminal	Purification	Yes (Gao et al. 2015) [insect]
IMPACTTM (intein-mediated purification with chitin binding domain) (Chong et al. 1997)	51 aa	<i>B. ciruculans</i> , <i>Pyrococcus sp.</i> , <i>S. cerevisiae</i>	N- or C-terminal	Purification [†]	No
Poly-His (Hochuli et al. 1988; Kohno et al. 1998)	6–12 aa	<i>Synthetic</i>	N or C-terminal	Purification	Yes (Park et al. 2012a) [<i>E. coli</i>]
S (S protein of RNase A) (Kim and Raines 1993)	15 aa	<i>Mammalian</i>	N- or C-terminal	Purification	No
Strep II (Schmidt et al. 1996)	8 aa	<i>S. avidinii</i>	N or C-terminal	Purification	Yes (Rahman et al. 2007) [<i>E. coli</i>]
Fh8 (Fasciolahepatica 8-kDa antigen) (Costa et al. 2013a; Costa et al. 2013b)	69 aa	<i>F. hepatica</i>	N-terminal	Purification & solubility	No
GST (glutathione S-transferase) (Smith and Johnson 1988)	211 aa	<i>S. japonicum</i>	N-terminal	Purification & solubility	Yes (Park et al. 2012a) [<i>E. coli</i>]
HaloTag (Mutated dehalogenase) (Ohana et al. 2009)	~300 aa	<i>Rhodococcus sp.</i>	N-terminal	Purification & solubility	Yes (Suzuki et al. 2013) [mammalian]
MBP (maltose binding protein) (di Guana et al. 1988; Pryor and Leiting 1997)	396 aa	<i>E. coli</i>	N- or C-terminal	Purification & solubility	Yes (Chaudhary et al. 2011) [<i>E. coli</i>]
T7 gene10 (Olins et al. 1988)	260 aa	<i>Bacteriophage T7</i>	N-terminal	Purification & solubility	No
Ubiquitin (Kohno et al. 1998)	76 aa	<i>S. cerevisiae</i>	N-terminal	Purification & solubility	Yes (Bird et al. 2015) [<i>E. coli</i>]

Tag (Reference(s))	Size*	Organism	Tag placement	Use	Used in membrane protein expression in [expression system]?
Acidic proteins (MsbB and YjgD) (Zou et al. 2008)	124, 138 aa	<i>E. coli</i>	N-terminal	Solubility	No
AK (adenylate kinase) (Liu et al. 2015)	362 aa	<i>E. coli</i>	N-terminal	Solubility	Yes (Londesborough et al. 2014) [yeast]
ArsC (Stress-responsive arsenate reductase) (Song et al. 2011)	141 aa	<i>E. coli</i>	N-terminal	Solubility	No
CaBP (Calcium-binding protein) (Reddi et al. 2002)	134 aa	<i>E. histolya</i>	N-terminal	Solubility	No
Crr (glucose-specific phosphotransferase enzyme IIA component) (Han et al. 2007a)	169 aa	<i>E. coli</i>	N-terminal	Solubility	No
CysQ (3'-phosphoadenosine-5'-phosphatase) (Lee et al. 2014)	266 aa	<i>E. coli</i>	N-terminal	Solubility	No
EDA (KDPG aldolase) (Kang et al. 2015)	213 aa	<i>E. coli</i>	N-terminal	Solubility	No
EspA (E. coli secreted protein A) (Cheng et al. 2010)	192 aa	<i>E. coli</i>	N-terminal	Solubility	No
GB1 (IgG domain of B1 of protein G) (Huth et al. 1997; Kumar et al. 2012; Zhou et al. 2001)	56 aa	<i>Streptococcus sp.</i>	N-terminal	Solubility	Yes (Kumar et al. 2012) [<i>E. coli</i>]
IF2-domain I (N-terminal fragment of translation initiation factor IF2) (Sørensen et al. 2003)	158 aa	<i>E. coli</i>	N-terminal	Solubility	No
Mocr (Monomeric bacteriophage T7 0.3 protein) (DelProposto et al. 2009)	117 aa	<i>Bacteriophage T7</i>	N-terminal	Solubility	No
NusA (N-utilization substance) (Davis et al. 1999)	495 aa	<i>E. coli</i>	N-terminal	Solubility	Yes (Goehring et al. 2014) [mammalian]
PotD (spermidine/putrescine-binding periplasmic protein) (Han et al. 2007a)	348 aa	<i>E. coli</i>	N-terminal	Solubility	No
RpoA, S, D (RNA polymerase subunits) (Ahn et al. 2007; Zou et al. 2008)	329, 330, 613 aa	<i>E. coli</i>	N-terminal	Solubility	No

Tag (Reference(s))	Size*	Organism	Tag placement	Use	Used in membrane protein expression in [expression system]?
Skp (Seventeen kDa protein) (Chatterjee and Esposito 2006)	161 aa	<i>E. coli</i>	N-terminal	Solubility	No
SlyD (FKBP-type peptidyl-prolyl cis-trans isomerase) (Han et al. 2007b)	196 aa	<i>E. coli</i>	N-terminal	Solubility	No
SNUT (Solubility enhancing ubiquitous tag) (Caswell et al. 2010)	147 aa	<i>S. aureus</i>	N-terminal	Solubility	No
SUMO (small ubiquitin modified protein) (Zuo et al. 2005a)	100 aa	<i>H. sapiens</i>	N-terminal	Solubility	Yes (Zuo et al. 2005b) [<i>E. coli</i>]
T7PK (Phage T7 protein kinase) (Chatterjee and Esposito 2006)	~240 aa	<i>Bacteriophage T7</i>	N-terminal	Solubility	No
Trx (Thioredoxin) (LaVallie et al. 1993)	109 aa	<i>E. coli</i>	N- or C-terminal	Solubility	Yes (Yeliseev et al. 2007) [<i>E. coli</i>]
Tsf (elongation factor Ts) (Han et al. 2007b)	283 aa	<i>E. coli</i>	N-terminal	Solubility	No
GlpF (the glycerol-conducting channel protein) (Neophytou et al. 2007)	281 aa	<i>E. coli</i>	N-terminal	Inner membrane insertion	Yes (Neophytou et al. 2007) [<i>E. coli</i>]
OmpA signal sequence (Tiralongo and Maggioni 2011)	21 aa	<i>E. coli</i>	N-terminal	Inner membrane insertion	Yes (Tiralongo and Maggioni 2011) [<i>E. coli</i>]
<i>S. cerevisiae</i> α -factor signal sequence (Weiss et al. 1995)	89 aa	<i>S. cerevisiae</i>	N-terminal	Plasma membrane insertion	Yes (Weiss et al. 1995)[yeast]
Yeast STE2 receptor signal sequence (King et al. 1990; LaVallie et al. 1993)	24 aa	<i>S. cerevisiae</i>	N-terminal	Plasma membrane insertion	Yes (Ficca et al. 1995; King et al. 1990; Sander et al. 1994a) [yeast]
Lpp-OmpA (Lpp signal peptide + OmpA) (Francisco et al. 1992)	142 aa	<i>E. coli</i>	N-terminal	Outer membrane display	Yes [#] [<i>E. coli</i>]
Ice nucleation protein (Inp) (Jung et al. 1998)	114 kDa	<i>P. syringae</i>	N-terminal	Outer membrane display/insertion	Yes [#] (Yim et al. 2010)
Ecotin (E. coli trypsin inhibitor) (Malik et al. 2006)	162 aa	<i>E. coli</i>	N-terminal	Solubility & periplasm/media secretion	No
ZZ (IgG repeat domain ZZ of Protein A) (Löwenadler et al. 1987; Rondahl et al. 1992)	116 aa	<i>S. aureus</i>	N- and C-terminal	Solubility & periplasm/media secretion	Yes (Nizard et al. 2001) [mammalian]

Tag (Reference(s))	Size*	Organism	Tag placement	Use	Used in membrane protein expression in [expression system]?
Cell-CD (cellulase catalytic domain) (Gao et al. 2015)	20, 375 aa	<i>Bacillus sp.</i>	N-terminal	Periplasm/media secretion	No
Cex (exoglucanase) (Hasenwinkle et al. 1997)	41 aa	<i>C. fimi</i>	N-terminal (signal sequence)	Periplasm/media secretion	No
DsbA (disulfide oxidoreductase A) (Soares et al. 2003)	19 aa	<i>E. coli</i>	N-terminal (signal sequence)	Periplasm/media secretion	Yes (Jappelli et al. 2014) [<i>E. coli</i>]
Exl/Enx (endoxylanase) (Jeong and Lee 2000, 2001)	12, or 28 aa	<i>Bacillus sp.</i>	N-terminal (signal sequence)	Periplasm/media secretion	No
LamB (λ receptor protein) (Pratap and Dikshit 1998)	25 aa	<i>E. coli</i>	N-terminal (signal sequence)	Periplasm/media secretion	Yes [#]
LTB (heat-labile enterotoxin subunit B) (Sanchez et al. 1988)	21 aa	<i>E. coli</i>	N-terminal (signal sequence)	Periplasm/media secretion	No
Outer membrane protein (OmpC, F) (Becker and Hsiung 1986)	21, 22 aa	<i>E. coli</i>	N-terminal (signal sequence)	Periplasm/media secretion	Yes [#] [<i>E. coli</i>]
OmpT (protease VII) (Kurokawa et al. 2001)	20 aa	<i>E. coli</i>	N-terminal (signal sequence)	Periplasm/media secretion	Yes [#] [<i>E. coli</i>]
PhoA (alkaline phosphatase) (Xu et al. 2002)	21 aa	<i>E. coli</i>	N-terminal (signal sequence)	Periplasm/media secretion	Yes (Kaderbhai and Kaderbhai 1996) [<i>E. coli</i>]
PelB (pectate lyase B) (Better et al. 1993)	22 aa	<i>E. carotov</i>	N-terminal (signal sequence)	Periplasm/media secretion	Yes (Jappelli et al. 2014) [<i>E. coli</i>]
SpA (<i>S. aureus</i> Protein A) (Abrahmsén et al. 1986)	280	<i>S. aureus</i>	N-terminal (signal sequence)	Purification& periplasm/media secretion	Yes (Hansson et al. 1995) [<i>E. coli</i>]
SthII (heat stable enterotoxin II) (Qiu et al. 1998)	23 aa	<i>E. coli</i>	N-terminal (signal sequence)	Periplasm/media secretion	No
AT rich gene tags (PS, T7, AT, SER, H, G, R) (Habersack et al. 2012; Pandey et al. 2014)	8, 11, 6, 6, 6, 6 aa	Derived from various species	5'-end of mRNA	Expression enhancement	Yes (Habersack et al. 2012; Pandey et al. 2014) [<i>E. coli</i>]
HmBRI/D94N (mutated bacteriorhodopsin (BR) (Hsu et al. 2013)	250 aa	<i>H. marismortui</i>	N-terminal	Expression enhancement	Yes (Hsu et al. 2013) [<i>E. coli</i>]
Mistic (Kefala et al. 2007; Marino et al. 2015)	110 aa	<i>B. subtilis</i>	N-terminal	Expression enhancement	Yes (Kefala et al. 2007; Marino et al. 2015) [<i>E. coli</i>]

Tag (Reference(s))	Size*	Organism	Tag placement	Use	Used in membrane protein expression in [expression system]?
YaiN (regulator of frmRAB operon) (Leviatan et al. 2010)	98 aa	<i>E. coli</i>	N- or C-terminal	Expression enhancement	Yes (Leviatan et al. 2010) [<i>E. coli</i>]
YbeL (Leviatan et al. 2010)	120 aa	<i>E. coli</i>	N- or C-terminal	Expression enhancement	Yes (Leviatan et al. 2010) [<i>E. coli</i>]
EDDIE (N-terminal autoprotease NPRO) (Achmuller et al. 2007)	168 aa	<i>Classical swine fever virus</i>	N-terminal	Inclusion body targeting	No
KSI (ketosteroid isomerase) (Kuliopulos and Walsh 1994)	14 kDa	<i>P. testosteroni</i>	N-terminal	Inclusion body targeting	Yes (Chun et al. 2012) [<i>E. coli</i>]
Outer membrane protein F SP (OmpF missing signal peptide)	142 aa	<i>E. coli</i>	N-terminal	Inclusion body targeting	Yes (Su et al. 2013) [<i>E. coli</i>]
PagP SP (palmitoyltransferase missing the signal peptide) (Hwang et al. 2012)	161 aa	<i>E. coli</i>	N-terminal	Inclusion body targeting	No
PaP3.30 (ORF 30) (Rao et al. 2004)	162 aa	<i>P. aeruginosa bacteriophage Pap3</i>	N-terminal	Inclusion body targeting	No
PurF (N-terminal of polypeptide F4) (Lee et al. 2000)	152 aa	<i>E. coli</i>	N-terminal	Inclusion body targeting	No
TAF12 HFD (histone folding domain of human transcription factor) (Vidovic et al. 2009)	72 kDa	<i>H. sapiens</i>	N-terminal	Inclusion body targeting	No
Trp LE (deletion of a part of TrpL from TrpE operon) (Cook et al. 2011)	17 aa	<i>E. coli</i>	N-terminal	Inclusion body targeting	Yes (Cook et al. 2011) [<i>E. coli</i>]
BRIL (thermostabilized apocytochrome b562) (Chun et al. 2012)	10.9 kDa	<i>H. sapiens</i>	Intermolecular	Stabilization & crystallization	Yes (Chun et al. 2012) [insect]
T4 lysozyme (Engel et al. 2002)	15.9 kDa	<i>Bacteriophage T4</i>	Intermolecular	Stabilization & crystallization	Yes (Engel et al. 2002) [<i>E. coli</i>] & (Cherezov et al. 2007; Chun et al. 2012; Granier et al. 2012; Kruse et al. 2012) [insect]

* aa: amino acids

†. intein-mediated removal of chitin binding domain (single step purification, no protease-mediated removal of tag)

Given that the tags are derived from membrane proteins, they have inherently been used for membrane protein expression

Note: tags may not function as specified depending on the effect of POI on the fusion protein.

Table 4

Properties of commercially available proteases reported for cleavage of fusion tags from membrane proteins in alphabetical order. Note that references include both key primary literature around protease and examples of usage for membrane proteins.

Protease ^a (Reference(s))	Cleavage site (at “P”)	Size	Location of tag	pH range	Chaotrope sensitivity	Salt sensitivity	Temperature range (°C)
Enterokinase (Choi et al. 2001; Goncharuk et al. 2011)	DDDDK	31 kDa	N-terminal	7–8			4–37
Factor Xa (Jenny et al. 2003; Periasamy et al. 2013)	IEGR	43 kDa	N- or C-terminal	6–9	0.25 M Urea 0.25 M GnCl 1% Triton X- 100 0.001% SDS	0.25 M NaCl or imidazole	4–37
Rhinovirus 3C protease (PreScission) (Clark et al. 2010; Cordingley et al. 1990)	LEVLFQ GP	46 kDa	N- or C-terminal	7–8	1% Triton, Tween, NP40	0.5 M NaCl	5–15
Subtilisin BPN (Profinity eXact) (Huang et al. 2012)	EEDKLFKAL (C-terminus of propeptide)	27.8 kDa	N-terminal	7.2	2 M Urea (no GnCl)	3 M NaCH ₃ CO O (no Cl [–])	4–30
SUMO protease (Li and Hochstrasser 1999; Malakhov et al. 2004; Zuo et al. 2005a)	GG (C-terminus of SUMO)	72 kDa	N- terminal	7–9	2 M Urea	0.3 M NaCl 0.15 M imidazole	4–30
TEV (tobacco etch virus) (Kapust and Waugh 2000; Parks et al. 1994)	ExxYxQ (G/S)	27 kDa	N- or C-terminal	6–8.5	2 M Urea	0.1 M	4–37
Thrombin (Jenny et al. 2003; Jidenko et al. 2006)	LVPR G	37 kDa	N- or C-terminal	5–10	0.1 M Urea	0.15 M	4–37

^a Cleavage conditions and chemical compatibility based on commercial protocols/handbooks: Sigma/Aldrich - TEV, thrombin and enterokinase; NEB – Factor Xa; Invitrogen – SUMO; BioRad – Subtilisin BPN; GE Healthcare – 3C.

Table 5

Pros and cons of XRD, solution- and solid-state NMR, and cryo-EM.

	Pros	Cons
XRD	Get entire 3D structure at atomic-level resolution if well-diffracting Single model visualization (easy interpretation) Large protein structure determination possible Largely automated Fast data acquisition	Stable and well-diffracting crystals required (long screening time) Cannot resolve dynamic regions Unnatural/non-physiological conditions Potential artefacts during crystallization
Solution-state NMR	Does not require crystals Ability to probe dynamics, conformational equilibrium, protein folding, intra- and intermolecular interactions, reaction kinetics (many different experiments observing various nuclei) Potential for near-native (if not native) conditions/ environment	Optimization of conditions to solubilize samples May require concentrated samples (potential aggregation) Isotope-labeled samples, often requiring multiple variations (costly) Data assignment is time-consuming Produces an ensemble of possible structures (data interpretation less straightforward than single model) Limited to relatively small proteins
Solid-state NMR	Does not require crystals Ability to probe dynamics, conformational equilibrium, protein folding, intra- and intermolecular interactions, reaction kinetics (many different experiments observing various nuclei) Large protein structure determination possible Native or near-native environments quite readily employed	Optimization of sample conditions Isotope-labeled samples (costly) Data assignment is time-consuming Produces an ensemble of possible structures
Cryo-EM	Mostly physiological or native in its condition Do not require crystals Can resolve different conformations in an ensemble (heterogeneous and dynamic samples) Large protein structure and complex determination possible	May not achieve atomic resolution Highly technical sample preparation procedure Require large data set (costly and lengthy acquisition) Electron beam can damage samples and cause particle movement