Overcoming bottlenecks in the membrane protein structural biology pipeline

David Hardy¹,², Roslyn M Bill¹, Anass Jawhari² & Alice J Rothnie¹

¹Life & Health Sciences, Aston University, Birmingham, B4 7ET, UK
²Calixar, Lyon, France

Correspondence to: Alice Rothnie (+44 121 204 4013  a.rothnie@aston.ac.uk) or Anass Jawhari (+33 649 555 606  ajawhari@calixar.com)

Key words:
Membrane proteins
Structural biology
SMALP
Calixarenes
MNG
Solubilisation

Abbreviations:
EM-Electron Microscopy
GPCR- G protein-coupled receptors
GNG- Glucose Neopentyl Glycol
MSP – Membrane Scaffold Protein
MNG- Maltose Neopentyl Glycol
NMR- Nuclear magnetic resonance
SMA- Styrene Maleic Acid
SMALPs –SMA Lipid Particles

Abstract

Membrane proteins account for a third of the eukaryotic proteome, but are greatly under-represented in the Protein Data Bank. Unfortunately, recent technological advances in X-ray crystallography and electron microscopy cannot account for the poor solubility and stability of membrane protein samples. A limitation of conventional detergent-based methods is that detergent molecules destabilize membrane proteins, leading to their aggregation. The use of orthologues, mutants and fusion tags has helped improve protein stability, but at the expense of not working with the sequence of interest. Novel detergents such as GNG, MNG and calixarene-based detergents can improve protein stability without compromising their solubilising properties. SMALPs focus on retaining the native lipid bilayer of a membrane protein during purification and biophysical analysis.
Overcoming bottlenecks in the membrane protein structural biology pipeline, primarily by maintaining protein stability, will facilitate the elucidation of many more membrane protein structures in the near future.

**Introduction**

**Membrane proteins – why are they important?**

Membranes are a vital component of any cell creating a barrier between the intracellular components of the cell and its extracellular environment. This barrier allows the cell to regulate what enters and exits through membrane proteins. Not only are membrane proteins localised in the cell surface membranes but also internal membranes of organelles such as mitochondria, endoplasmic reticulum, golgi and nuclei [1]. Membrane proteins make up between 20 and 30% of all proteins in eukaryotic cells [2] and have a wide range of functions as exhibited by transporters (e.g. ABCs), enzymes, receptors (e.g. GPCRs) and anchoring proteins. Because of the importance of these functions, when mutations occur they can cause severe disease states such as cystic fibrosis, neurodegenerative diseases, cancers, cardiovascular diseases and immunological disorders [3-5].

**Membrane proteins – what are the difficulties?**

**Expression**

One bottleneck in membrane protein structural biology pipelines is the low abundance of the membrane protein of interest. Membrane proteins are naturally expressed at low levels (with some exceptions such as rhodopsin, which is very abundant in the retina) [6], so the separation of endogenous membrane proteins from biological sources in sufficient quantity represents a real limitation. Overexpression of membrane proteins increases the yield per cell and has the advantage of creating recombinant membrane proteins, which can incorporate tags to aid purification and stability. *Escherichia coli* is a commonly-used host for prokaryotic membrane protein production whereas eukaryotic membrane proteins often need a eukaryotic host such as mammalian cells, yeast or insect cells [7, 8]. Membrane proteins can sometimes be toxic to the cells used for their expression. The use of inducible expression systems and some specific inhibitors and/or cell-free expression can represent a solution to this toxicity. Correct folding and post translational modifications are key for obtaining functional proteins [9]. Notably, compared to proteins expressed in the cytosol or secreted into the growth medium, there is limited space in the plasma membrane when over-expressing membrane proteins, which therefore limits the maximum theoretical yield.

**Extraction/solubilisation & purification**

Membrane proteins must be extracted/solubilised from the membrane before being purified. Solubilising a membrane protein whilst maintaining its stability is an enduring challenge in membrane protein structural biology. The conventional and most commonly-used method for solubilising membrane proteins is the addition of surfactants, known as detergents. These
amphipathic molecules are able to replace the lipid bilayer by forming micelles (Figure 1A) that allow the membrane protein to become soluble in aqueous solutions [10, 11]. Unfortunately many detergents do not maintain the structure and function of membrane proteins, which can become misfolded once extracted from their native lipid environment, possibly due to the lack of lateral pressure. Since detergents have varying characteristics, it is hard to predict which detergent will be optimal for a particular membrane protein; the conditions for efficient yet functional extraction are also protein dependent. The mechanism behind this is poorly understood, nevertheless some general rules exist such as short chain detergents are not as effective for solubilisation, but are better for crystallization than long chain detergents, which may interfere with protein-protein interactions within a packed crystal [11]. Some detergents needed for solubilisation and stabilisation of membrane proteins can interfere with affinity purification. Therefore detergent exchange prior to purification may be necessary. Although detergents solubilise and allow membrane protein purification, it is a real challenge to avoid aggregation and maintain structural and functional integrity. Therefore, detergent screening requires functional and structural cross-validation.

Proteoliposomes (Figure 1B) and membrane scaffold proteins (MSP) (Figure 1C) are two model systems in which a membrane protein can be reconstituted back into a lipid bilayer allowing protein function to be examined. For example, the reconstitution of membrane protein transporters into proteoliposomes allows their transport function to be investigated [12]. However, the orientation of the protein cannot be controlled, which can be limiting for some assays and to antibody discovery screening. MSP are capable of self-assembling into discoidal nanoparticles that encapsulate the membrane protein giving it a native-like lipid environment that promotes stability and solubility [13]. Although both approaches allow the membrane protein to be in a lipidic environment, and have allowed many functional and mechanistic details to be examined, they still require initial detergent solubilisation and purification prior to reconstitution.

**Structural biology of membrane proteins**

**X-ray crystallography**

X-ray crystallography has been the single most successful technique in determining high resolution structures of membrane proteins. However, the rate of success is not increasing as quickly as initially expected in comparison to that of soluble proteins. Producing 3D protein crystals for X-ray crystallography requires a membrane protein that is highly pure, stable and homogenous. The optimal detergents used to solubilise and stabilise a membrane protein are not necessarily the same as the best detergents for 3D crystal formation, so detergent exchange and optimisation are necessary. Detergents may also interfere with protein-protein contacts in a crystal with short chain detergents therefore being more suitable for crystallisation [11]. One method that has been used to circumvent some of the problems of using detergents in crystallography is the use of “in cubo” systems that use lipid structures [9] such as the lipidic cubic phase (LCP). The idea behind these systems is that the membrane protein is surrounded by an environment that is close to the constituents of its natural lipid bilayer. The membrane protein should therefore have a more natural and stable conformation and is more likely to produce protein crystals. Other approaches that promote stability during crystallisation trials are to use a truncated and/or mutated version of the
However, this may require a significant deviation from the native protein sequence that may be limiting to understanding the mechanism of action and to drug development.

Electron microscopy

Until recently the use of electron microscopy (EM) as a structural technique for membrane proteins has been limited by low resolution. Recent technical advances in microscopes, classification methods to sort particles, direct-detection cameras and the stabilisation of membrane proteins has led to improvements in resolution [15-17]. Single particle EM can determine the structure of a membrane protein by imaging individual particles. The membrane protein particles can be prepared either by negative stain or freezing in vitreous ice (cryo-EM) [16]. There is less artefactual information produced from cryo-EM, but negative staining has the advantage of maintaining highly-stable sample specimens that are less sensitive to radiation and have much better contrast [16]. Images of membrane proteins by single particle EM can be enhanced by averaging particles within the same class [18]. These electron micrographs can be used to pick particles, obtain class averages and back project them into a 3D density map of the membrane protein using imaging process software such as SPIDER, EMAN or IMAGIC [15]. Another method called electron crystallography consists of producing thin layers of planar, two-dimensional (2D) crystals of membrane proteins which are able to scatter electrons [19]. Electron crystallography can therefore generate structures in a near-to-native environmental state [20]. Notably, these approaches still require high quality crystals, which is not the case for single particle EM. EM can also provide information about different structural states, conformations or oligomers of the protein within a single sample, thanks to classification routines [15]. This is only possible in X-ray crystallography if different structures can be solved.

NMR

Nuclear magnetic resonance (NMR) has produced over 100 membrane protein structures. This has been done mostly through the use of solution-state NMR, but oriented solid-state and magic-angle spinning solid-state NMR are starting to make an increasing contribution [21-23]. Structures of beta-barrel, alpha-helical and membrane-associated proteins have all been solved by NMR. However, most membrane proteins are too large for NMR structural studies: the majority of known structures solved by NMR are of proteins with molecular masses below 20 kDa. Notably, the 34 kDa CXCR1 chemokine G protein-coupled receptor (GPCR) structure was elucidated by rotationally aligned (RA) solid-state NMR [21]. NMR also has an important role to play in investigating protein-protein, and ligand-protein interactions as well as for fragment-based drug discovery [24-26].
**Overcoming problems with stability**

Despite technical improvements in the expression (and purification) of membrane proteins and with structural methods, a major challenge remaining in membrane protein structural biology is the instability of protein samples.

Conventional methods of stabilising membrane proteins are the use of orthologues, ligand binding, mutagenesis and protein engineering, which each come with their own set of problems. Although the use of orthologues means the target sequence of interest will not be studied, an orthologue may be more readily expressed and determination of its structure can be used to generate a molecular model of the target protein, as illustrated by the case of using Sav 1866 to model human MRP4 [27].

The ability of ligands to stabilize membrane proteins can help reduce flexibility and therefore facilitate structure determination, as well as provide details on specific ligand/protein interactions, however it does often provide only a single snapshot of a highly dynamic process. The binding of antibodies or nanobodies has also been used as a way to stabilise membrane proteins [6].

Mutagenesis and protein engineering can produce more stable membrane proteins by alanine scanning [14, 28]. Fusion proteins improve crystallisation by swapping unstable areas with T4 lysozyme, an approach which has been successful with GPCRs such as the adenosine A2a receptor [29].

**Novel extraction/solubilisation approaches**

Although the use of mutants and fusion proteins has been very successful, there is always the risk that they differ from the native structure. One approach to improve stability of wild-type proteins is to develop new methods for extraction and solubilisation which provide greater stability to the protein.

**Glucose neopentyl glycol and maltose neopentyl glycol**

Glucose neopentyl glycol (GNG) (Figure 2A) and maltose neopentyl glycol (MNG) (Figure 2B) are two classes of amphiphiles that are structurally similar, with the hydrophilic groups of GNG derived from glucose and MNG derived from maltose. The difference between these detergents and more conventional detergents with similar structures, such as n-dodecyl β-D-maltoside (DDM) (Figure 2C) is the central quaternary carbon atom that is derived from neopentyl glycol allowing for the addition of two hydrophobic and hydrophilic groups [30]. GNG has four different variants and MNG has three with the difference involving the attachment of the hydrophobic chain to the quaternary carbon [31]. Use of GNG-3 has help to produce the structure of Na^+^-pumping pyrophosphatase [32] and human aquaporin 2 (AQZ 2) [33]. MNG-3 has demonstrated great solubility and stability characteristics and its use facilitated the determination of the high resolution structure of more than 10 GPCRs [34]. Although it should be noted that for many of these structures, mutational/fusion protein approaches were also used in conjunction with the novel detergents.

**Calixarenes**

Another novel class of detergents is based on the calixarene platform. These detergents have the same general structure as conventional detergents (polar head group and hydrocarbon chain) but
also contain a calixarene platform of aromatic rings as shown in Figure 2D. They are capable of making specific contacts with a membrane protein to enable them to solubilise and stabilise at the same time. Specific interactions between the hydrophobic tail and the transmembrane domains of membrane proteins allow competition with lipids to facilitate solubilisation. In addition, interactions between the calixarene platform, carboxylate or sulfonate groups with aromatic and charged amino acid residues (corresponding to \( \pi \) stacking and salt bridge interactions) help to stabilise the protein. The use of these detergents was shown to maintain structure and function of transporter proteins [35] and helped maintain interactions between BAG3 and its corresponding partner at the membrane of macrophages. This allowed the identification of IFITM2 as a specific partner and provided molecular clues to signalling in the context of pancreatic cancer [36]. Some calixarenes have been shown to facilitate crystallisation of proteins [35, 37, 38]. Significant progress has been made in generating different calixarenes with different chemical characteristics to favour solubilisation and stabilisation at the same time, to improve stability after solubilisation or to facilitate crystallisation.

**SMALPs**

Styrene maleic acid (SMA) is a co-polymer made up of alternating units of styrene and maleic acid. This copolymer has the ability to solubilise lipid membranes into nanodisc-like structures termed SMA lipid particles (SMALPs; also known as native nanodiscs or lipodisqs), where a small disc of lipid bilayer is encircled by the SMA co-polymer [39-41]. During the formation of SMALPs from a biological membrane, membrane proteins can become trapped inside these discs (as shown in Figure 1D), solubilising them into small particles (much like nanodiscs when using membrane scaffolding proteins), but without the need for detergent at any stage. The proteins within a SMALP can be effectively purified using affinity chromatography [39, 42-44]. The particles are stable once formed so there is no need to supplement buffers during purification as there is when using detergents; this is a major cost advantage. The membrane protein that is encompassed inside the SMALP is now stable in its native lipid environment, allowing access to both sides of the membrane protein and aiding in functional and structural studies. Proteins within SMALPs have been shown to bind ligands comparably to those in the native membrane [44]. The use of SMALPs has facilitated improved thermostability and homogeneity of membrane proteins such as P-glycoprotein (P-gp) and the adenosine A2a receptor [39, 42, 44]. To date no crystal structure has been reported using SMALPs but both negative stain and cryo-EM low resolution structural maps have been determined [39, 43]. Unlike MSP-nanodiscs, the polymer does not show significant extra-electron density; EM shows a very small annulus of lipids and polymer.

**Conclusion & perspectives**

Membrane protein structural biology is challenging. Not only must the membrane protein be expressed at high enough levels for structural studies, it must also have the correct conformation, homogeneity and stability. Regardless of the structural biology method used, the supply of stable, purified protein is still a major bottleneck, which is now being addressed with a range of novel approaches (Table 1). One of the biggest challenges is caused by the need to solubilise membrane proteins. The most common method of solubilising membrane proteins is with detergents. Although conventional detergents such as DDM are capable of making membrane proteins soluble in aqueous
solution, and much has been achieved using them, they often do not stabilise the membrane protein efficiently and can lead to protein aggregation [11]. To overcome these problems, novel detergents such as GNG, MNG and calixarenes have been produced that enhance stability [30, 34, 35]. Whilst the number of structures obtained using these novel detergents is still much lower than with conventional detergents, this may just be a matter of time. There are reports in the literature of previously challenging proteins that are now able to be extracted and purified to the degree needed for structural studies using these new molecules [45-47]. It is interesting to note that both the GNG/MNGs and Calixarenes contain larger, more structured head groups than conventional detergents, and this may be a key factor in their stabilising properties. In fact a recently published study using an even larger branched pentameric sugar based headgroup reports even greater stabilising properties, but this has not yet been used for structural studies [48].

An alternative approach is to use SMA co-polymer to solubilise and purify membrane proteins whilst retaining the lipid bilayer environment [39]. This approach has proved successful to date for all cell types tested, and a range of different membrane proteins. One limitation of this approach is that the size of discs formed with SMA are limited to approximately 10nm diameter, thus it is not suitable for large membrane complexes. The number of lipids within a SMALP will depend upon the size of the membrane protein encapsulated, and this may also influence the fluidity of the bilayer. For example it has been shown that in a lipid only SMALP, the central lipids behave much like a native bilayer, whilst the lipids at the edges are much more ordered [41]. Another limitation of SMALPs is their sensitivity to divalent cations such as magnesium which is an important co-factor for some biological reactions. Whilst a crystal structure from SMA purified protein has not yet been achieved, the increased stability, homogeneity and ease of concentration are all promising factors, and the sensitivity to magnesium may provide a way to remove the SMA to aid crystallization and/or reconstitution. SMALPs have however been shown to be suitable for single particle EM studies, and with the excellent advances made recently with enhanced microscopes and classification methods to sort particles, increasing the resolution that can be obtained by EM to almost atomic resolution [15-17], this has enormous potential for the future.
Table 1: A table outlining limitations and current solutions throughout the membrane protein structural biology pipeline.

<table>
<thead>
<tr>
<th>Process</th>
<th>Limitation</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expression</td>
<td>1. Naturally low in abundance</td>
<td>1. Over-expression of membrane protein</td>
</tr>
<tr>
<td></td>
<td>2. Overexpression can lead to misfolded and non-functional membrane proteins</td>
<td>2. Find the best expression host and optimise expression conditions that allow high-yielding, functional expression</td>
</tr>
<tr>
<td>Solubilisation and</td>
<td>1. Loss of native lipid membrane reduces lateral pressure and causes protein to collapse</td>
<td>1. SMA and MSP can encapsulate membrane proteins in their native lipid environment</td>
</tr>
<tr>
<td>stabilisation</td>
<td>2. Detergent solubilisation denatures proteins</td>
<td>2. Screen the best detergent that allows solubilisation and stabilisation, by considering new detergents such as GNG, MNG and calixarenes</td>
</tr>
<tr>
<td>Purification</td>
<td>1. Some detergents interfere with binding</td>
<td>1. Exchange detergents at different steps (solubilisation and purification) or use SMA which allows a detergent-free purification whilst maintaining stability</td>
</tr>
<tr>
<td></td>
<td>2. Samples must be homogeneous and in high concentration for functional and structural studies</td>
<td>2. Fine-tune solubilisation, stabilisation and affinity purification conditions to avoid aggregates (observed in size exclusion chromatography or native PAGE) and maintain functionality</td>
</tr>
<tr>
<td>Functional and Structural</td>
<td>1. Detergents can interfere with functional study</td>
<td>1. Cross validate the solubilisation and purification trials with functional studies and if necessary reconstitute the protein in proteoliposomes or nanodiscs</td>
</tr>
<tr>
<td>Studies</td>
<td>2. No one detergent can solubilise, stabilise and be used for crystallography</td>
<td>2. GNG, MNG and calixarenes can be used to solubilise and stabilise at the same time, improve stability and facilitate crystallisation</td>
</tr>
<tr>
<td></td>
<td>3. Producing 3D crystals requires highly stable and homogenous samples</td>
<td>3. EM can handle some degree of heterogeneity but requires high purity</td>
</tr>
</tbody>
</table>
Figure legends

Figure 1: Cartoons of detergent micelles, proteoliposomes, membrane scaffold proteins (MSP) and SMA lipid particles (SMALPS). These cartoons show how the phospholipids (red circles, black zig-zags) or detergents (orange circles, black zig-zags) interact with the blue membrane proteins. The orange barrels of the membrane scaffold protein represent apolipoprotein A1 (used in MSP) and the green disc in the SMALP cartoon represents SMA.

Figure 2: Chemical structures of DDM, MNG, GNG and calixarene-based detergents. There are similarities in the structures of DDM, MNG and GNG. However, MNG and GNG have maltose and glucose head groups attached to a central quaternary carbon atom that is derived from neopentyl glycol allowing for the addition of two hydrophobic hydrocarbon chains. One calixar-based detergent, so-called calix[4]arene is made of 4 aromatic rings with three hydrophilic sodium carboxylate groups at the para position of the calixarene platform. The hydrophobic part consists of one aliphatic tail (which can be any length, here decyl) grafted onto one of the phenolic groups.

Funding

David Hardy is supported by a BBSRC Industrial CASE studentship with Calixar [BB/L015846/1].

References
