DR ALI ASGHAR KERMANI (Orcid ID : 0000-0002-4994-9730)

Received Date : 30-Aug-2020
Revised Date : 17-Nov-2020
Accepted Date : 16-Dec-2020
Article type : A Guide To…

Title: A guide to membrane protein X-ray crystallography

Running Title: A guide to membrane protein X-ray crystallography

Author
Ali A. Kermani
Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

Corresponding author: kermania@umich.edu
Phone number: +1 734-764-3631

Abbreviation list: NMR, nuclear magnetic resonance spectroscopy; cryo-EM, cryogenic electron microscopy; *E. coli*, *Escherichia coli*; GFP, green fluorescent protein; FSEC, fluorescence-detection size exclusion chromatography; CMC, critical micelle concentration; SDS, sodium dodecyl sulfate; DDM, *n*-dodecyl-ß-D-maltoside; DM, *n*-decyl- ß-D-maltoside; OG, *n*-Octyl- ß-D-glucopyranoside; NG, *n*-Nonyl- ß-D-glucopyranoside; LDAO, lauryldimethylamine-N-oxide; SMALPs, styrene malic acid lipid particles; MSP, membrane scaffold protein; Nb, nanobody; PDC, protein-detergent complexes; SEC, size exclusion chromatography; DLS, dynamic light

scattering; GPCRs, G protein-coupled receptors; IMAC, immobilized metal affinity chromatography; CPM, N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl] maleimide; PEG, polyethylene glycol; LCP, lipidic cubic phase; MAGs, monoacylglycerol; VH, variable domain of heavy-chain; VL, variable domain of light-chain; ICL3, third intracellular loop; FN3, human fibronectin type III domain; $ß_2$AR, $ß_2$-adrenergic receptor; A_{2A}AR, A_{2A} adenosine receptor; MR, molecular replacement; HA, heavy atom; SIR, single isomorphous replacement; MIR, multiple isomorphous replacement; SAD, single-wavelength anomalous dispersion; MAD, multiple-wavelength anomalous dispersion; SeMet, selenomethionine; I-SAD, iodide single-wavelength anomalous diffraction; NaI, sodium iodide.

Keywords: Membrane proteins, X-ray crystallography, detergents, crystallization chaperones, *in meso* crystallization

Conflicts of interest: The author declares no conflict of interest.

Abstract

Membrane proteins play critical physiological roles in all organisms, from ion transport and signal transduction to multidrug resistance. Elucidating their 3D structures is essential for understanding their functions and this information can also be exploited for structure-aided drug discovery efforts. In this regard, X-ray crystallography has been the most widely used technique for determining the high-resolution 3D structures of membrane proteins. However, the success of this technique is dependent on efficient protein extraction, solubilization, stabilization, and generating diffracting crystals. Each of these steps can impose great challenges for membrane protein crystallographers. In this review, the process of generating membrane protein crystals from protein extraction and solubilization to structure determination are discussed. In addition, the current methods for pre-crystallization screening and a few strategies to increase the chance of crystallizing challenging membrane proteins are introduced.

Introduction

It is well established that membrane proteins play an essential role in a wide range of biological processes and their improper folding or mutation is associated with diseases such as cancer, cystic fibrosis, Alzheimer and obesity [1]. Therefore, it is not surprising to see that a large percentage of genome (20-30 %) in most organisms is dedicated to producing transmembrane proteins [2, 3]. Elucidating the 3D structure of membrane proteins is a key to understanding their function and assisting structure-based drug design. X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and cryogenic electron microscopy (cryo-EM) are the main techniques that have been used to determine the 3D structure of transmembrane proteins. Among those, X-ray crystallography is the leading technique by contributing to solving ~80% of membrane protein structures (https://blanco.biomol.uci.edu/mpstruc/#Latest). However, obtaining high-resolution diffracting crystals of membrane proteins is notoriously difficult. The first bottleneck is to generate milligram amounts of target membrane protein. Native expression levels of most membrane proteins, particularly eukaryotic origin, are low and bacterial expression hosts often fail to produce functional forms of eukaryotic proteins [4, 5]. Insect and mammalian cells are more suitable alternatives for producing functional eukaryotic transmembrane proteins, but using them requires cell culture facilities and specialized media and

therefore they provide more expensive expression systems with lower yield per liter of cell culture [6].

The next obstacle is to extract transmembrane proteins from host cell membrane, while maintaining their integrity and function. Historically, detergents have been the most common tool for this purpose. Despite their efficacy and ease of use in extracting membrane proteins, they can negatively impact the stability and function of these macromolecules. Another challenge is to identify the correct condition that leads to crystallization, often requiring screening of several hundred to thousands of different conditions. This could take from months to years. Finally, membrane protein crystals are typically very fragile, sensitive to X-ray damage, diffract only to low resolution, and display crystallization defects such as anisotropy [7, 8], requiring specialized handling, collection, and processing techniques. Due to these obstacles, other structure determination methods such as cryo-EM are becoming increasingly more popular [9]. However, cryo-EM also comes with its own challenges such as detection threshold for membrane proteins smaller than 100 kDa [10, 11], atomic resolution limit [11], and high cost of purchasing, running and maintaining cryo-EM instruments [12]. Significant technical advances in sample preparation, data processing and hardware development during the last few years have improved the size limit to ~50 kDa [13] and in rare cases has generated near-atomic resolution [14]. Until these limitations are completely overcome, X-ray crystallography remains a powerful technique for determining the 3D structure of membrane proteins.

This review will provide a thorough overview of membrane protein crystallography, from identifying the most tractable target through expression host selection, to solubilization, crystallization, and structure determination of the target protein will be provided (Figure 1). In addition, as part of this review, current methods of pre-crystallization screening to identify the most suitable buffer conditions for stabilizing purified proteins and ultimately increasing crystallization likelihood will be discussed. This review is targeted for graduate students, researchers, and drug designers interested in solving the 3D structure of transmembrane proteins using X-ray crystallography.

1. Membrane protein overexpression

1.1 Homolog screening

Membrane proteins typically express with low yield. Therefore, it is essential to screen several homologs of the target membrane protein to identify highly expressed, biochemically tractable homologs [15]. Generating a phylogenic tree is an invaluable tool for this purpose. Initial screening of representative members from separate clades of the phylogenetic tree allows the investigator to determine what clade produces the highest amount of the macromolecule. It is very likely that other members of the same clade behave in a similar fashion and generate high amounts of the target macromolecule [15].

1.2 Expression hosts

Prokaryotic and eukaryotic membrane proteins often require distinct expression hosts, such as *Escherichia coli* (*E. coli*), yeast, insect, and mammalian cells. Each expression host offers distinct strengths and weaknesses over other expression hosts for each group. These expression hosts will be reviewed here to enable researchers to choose the most suitable expression host based on their target membrane protein (Table 1).

1.2.1 Prokaryotic expression hosts

Escherichia coli (*E. coli*) is the most widely used expression host for producing recombinant proteins. This is mainly because *E. coli* is capable of producing high levels of recombinant protein over a short period of time (a few hours to one day). In addition, the low cost of *E. coli* growth and its well-characterized genetics and physiology [16] have made this bacterium the most attractive expression system for generating prokaryotic membrane proteins for structural studies (Figure 2A). However, recombinant protein production in *E. coli* is associated with some drawbacks. Aggregation of large amounts of overexpressed membrane protein in the form of inclusion bodies is one of them [4]. The aggregation of recombinant membrane protein, which is the direct result of misfolding, has been attributed to the saturation of the translocon pathway during membrane protein insertion into the phospholipid bilayer [17]. To overcome this obstacle new strains of *E. coli* with the ability to tune the transcription rate, and hence the level of protein production have been developed. *E. coli* strains BL21(DE3)plysS [18, 19] and Lemo21(DE3)

[20] prevent inclusion body formation by producing T7 lysozyme, a natural inhibitor of T7 polymerase, and fine-tuning the expression levels of this enzyme, respectively (Table 2). Other *E. coli* strains C41(DE3) and C43(DE3), known as Walker strains, harbor a weakening mutation in *lac*UV5 promoter, leading to a reduction in T7 RNA polymerase expression and therefore able to overexpress high levels of toxic and transmembrane proteins [20, 21] (Table 2).

1.2.2 Eukaryotic expression hosts

Although *E. coli* is convenient for producing bacterial membrane proteins, it is usually a poor host for expressing eukaryotic membrane proteins. Many eukaryotic membrane proteins undergo post-translational modifications, such as phosphorylation, glycosylation and ubiquitylation, *E. coli* may not harbor the necessary machinery for applying these post-translational modifications to overexpressed macromolecules [4]. Given that some of these post-translational modifications are necessary for eukaryotic proteins to function properly, eukaryotic hosts are more suitable for this purpose (Figure 2B).

Of eukaryotic membrane proteins with known 3D structures, 17% have been produced in yeast (Figure 2B), representing a very diverse range of transmembrane proteins [22]. Yeast combines the unique features of *E. coli cells*, low cost and feasibility of growth, together with eukaryotic folding capabilities [22]. Among yeast species, *Saccharomyces cerevisiae* and *Pichia pastoris* are the most widely used species for generating eukaryotic membrane proteins for structural determination purposes [22, 23].

Of eukaryotic membrane proteins with known 3D structures, 35% and 27% have been produced in insect and mammalian cells, respectively, (Figure 2B) Insect cells are easy to scale up, they share similar codon-usage with mammalian cells, and they can implement post-transitional modifications to the overexpressed eukaryotic proteins more efficiently than *E. coli* or yeast [16]. On the other hand, no expression system is able to compete with mammalian cell lines in producing functional eukaryotic membrane proteins. Although yeast and insect cells possess the necessary machinery for decorating the eukaryotic membrane proteins with post-translational modifications, these functional groups are not exactly identical to their eukaryotic counterparts

[24]. Yeast and insect cells often fail to correctly fold complex eukaryotic membrane proteins, which can lead to aggregation and non-functional forms of these proteins [24].

1.3 Membrane protein expression screening

Following the selection of an appropriate expression host and overexpression of the membrane protein of interest, it is essential to examine the level of protein expression. Fusion of a Green Fluorescent Protein (GFP) to the membrane protein terminus allows monitoring the amount of recombinant protein in whole cells at levels as low as 10 μg of GFP per liter of culture [25]. SDS-PAGE gel provides higher sensitivity and can detect GFP as low as 5 ng [25]. Excitation of GFP at a wavelength of 395 nm or 498 nm triggers the emission of a green fluorescent light at 509 nm [26]. Rapid screening of multiple clones using this method facilitates the identification of the most promising targets and improves the yield for scaling up and purification. This technique can also assist solubilization and purification steps. For instance, monodispersity and stability of detergent-solubilized targets can be monitored using Fluorescence-detection Size Exclusion Chromatography (FSEC) system [27]. Monodisperse proteins appear as single symmetrical peaks, whereas polydisperse or denatured polypeptides produce several asymmetric peaks corresponding to different states of the protein.

2. Membrane protein extraction, solubilization and purification

2.1 Extraction

Membrane proteins are composed of two regions: hydrophobic and hydrophilic. The hydrophobic core is embedded in lipid bilayers, whereas the hydrophilic region is exposed to aqueous solvent on either side of the membrane. Structural and functional studies of membrane proteins require extraction of the entire macromolecule from the cell membrane following overexpression, without disturbing its integrity. Extracted proteins need to maintain their stability and remain functional throughout the purification process. Detergents have been historically the most commonly used tool for this purpose (Figure 3). However, due to the potential adverse effects of detergents on the stability and function of membrane proteins, other membrane mimetic environments have been developed (Figure 3). In this guide some of the most

successful and efficient detergents used in solubilizing membrane proteins will be discussed and some of the most promising membrane mimetics will be reviewed.

2.1.1 Detergents

Similar to membrane phospholipids, detergent molecules are composed of a hydrophobic hydrocarbon tail and a hydrophilic head group. This amphiphilic structure enables detergents to obtain a discoidal conformation in the solution, known as micelles. Micelles solubilize membrane proteins by encompassing the transmembrane domains of integral membrane proteins, with the loops and hydrophilic regions exposed to solvent. The minimum concentration of a detergent necessary to form micelles and extract membrane proteins is called critical micelle concentration or CMC.

Depending on the charge of hydrophilic head group, detergents are classified into three groups: ionic, nonionic and zwitterionic detergents [26, 28] (Figure 4). *Ionic detergents* carry a charged head group, either negative (anionic) or positive (cationic) and historically have been the most efficient group of detergents in extracting membrane proteins from lipid bilayers (Figure 4A). However, ionic detergents can have deleterious effects on protein-protein interactions and often lead to membrane protein denaturation [28]. Therefore, their usage has become limited to membrane proteins that are otherwise difficult to extract. Sodium dodecyl sulfate (SDS) and sodium cholate are two common examples of ionic detergents (Figure 4A). Sodium cholate belongs to the bile acids. Unlike ionic detergents, which have a distinct head group and tail, bile acids have a kidney-shaped structure with both hydrophobic and hydrophilic faces [28]. The hydrophilic face harbors several hydroxyl groups and the hydrophobic face is composed of a steroid nucleus (Figure 4A).

Nonionic detergents are currently the most popular and successful group of detergents in extracting and solubilizing membrane proteins for both functional and structure determination purposes. This is due to their non-disruptive nature [29], which enables them to preserve the native structure of the target protein by breaking protein-lipid interactions instead of protein-protein interactions. Alkyl glycoside detergents such as *n*-dodecyl-ß-D-maltoside (DDM), *n*-

decyl- ß-D-maltoside (DM), *n*-Octyl- ß-D-Glucopyranoside (OG), and *n*-Nonyl- ß-D-Glucopyranoside (NG) by contributing to the purification and crystallization of about 70% of membrane proteins [29, 30] are the most successful nonionic detergents (Table 3, Figure 4B). Another advantage of nonionic detergents is that they do not interfere with UV measurements, which enables fluorescence-based experiments on membrane proteins [26].

Zwitterionic detergents can be used as alternative detergents to nonionic detergents because they have an intermediate level of harshness between ionic and nonionic detergents [28]. They carry both positive and negative charged groups in their polar heads with an overall net charge of zero. One of the most successful detergents from this class for purification and crystallization purposes is Lauryldimethylamine-N-oxide or LDAO [29, 30] (Figure 4C).

2.1.2 Membrane mimetics

Membrane mimetic systems, such as nanodiscs and styrene malic acid lipid particles (SMALPs), provide an alternative platform for purification and stabilization of membrane proteins and hence eliminate the deleterious effects of detergents on these macromolecules. Nanodiscs are composed of phospholipid patches surrounded by two copies of membrane scaffold protein (MSP). MSP is a genetically engineered version of human serum apolipoprotein A-I [31]. The target membrane protein is initially extracted in the presence of detergents. The detergent-solubilized protein is mixed with phospholipids and MSP scaffold protein, and then the detergent is gradually removed either by dialysis or biobeads, triggering the self-assembly of nanodiscs [32]. However, the correct ratio of membrane protein, phospholipids and MSP must be carefully optimized to achieve homogeneous nanodiscs assembly and to prevent aggregation. Detailed protocols on the assembly of nanodiscs can be found in these publications [33, 34]. A native-like phospholipid bilayer environment enables packing of some of the lipids necessary for the stability and function of membrane proteins. For instance, it has been shown that incorporating cholesterol in nanodiscs improves the activity of membrane proteins, particularly G protein-couple receptors (GPCRs) [35]. So far very few crystal structures of nanodisc-embedded membrane proteins have been solved using *in meso* X-ray crystallography (e.g., bacteriorhodopsin [36]); the main application of this platform has been functional studies and cryo-EM structure determination [37,

38], where nanodiscs can assist in overcoming the size limit for small transmembrane proteins by maintaining their native oligomeric structures [39].

Although nanodiscs improve the stability of membrane proteins by removing the detergents after protein purification, the initial extraction of these proteins from their native cell membrane is still detergent-dependent. A growing number of studies have shown that the majority of the membrane lipids necessary for the activity and stability of membrane proteins are likely to be removed during initial stages of detergent extraction [40] and the obtained X-ray crystal structures might fail to present the native structure of the protein [41]. Therefore, detergent-free platforms such as styrene malic acid (SMA) copolymers have been developed to address this issue. The amphipathic nature of SMA copolymers enable them to solubilize membrane proteins similar to nanodiscs by encapsulating them along with a portion of surrounding membrane. However, in contrast to nanodiscs, which use a defined composition of lipids, SMA co-polymer lipid particles (SMALPs) will contain native lipid membranes. Teo and his colleagues recently showed that the membrane lipid compositions in different biological systems are very distinct [42], therefore using a defined lipid composition may not compensate for the native lipid environment. Although there are very few crystal structures of SMA-solubilized membrane proteins so far, such as bacteriorhodopsin [43], there are several examples of this platform used in structural studies of membrane proteins with Cryo-EM, such as multidrug exporter AcrB [44], alternative complex III in supercomplex with cytochrome oxidase [45], and glycine receptor [46].

2.2 Purification

Histidine (His)-tagged solubilized membrane proteins are typically purified using immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC). In case detergents are utilized for solubilization purposes, it is essential to ensure the presence of detergents above the CMC throughout the purification process. See Table 3 for the concentrations of most commonly used detergents. For a detailed protocol on membrane protein extraction, solubilization and purification see [15, 47].

3. Membrane protein activity and stabilization measurement

Following protein extraction, solubilization and purification, it is essential to perform pre-crystallization screening in order to assess the stability of proteins. Identifying the conditions that improve the stability of purified proteins increases the likelihood of crystallization and diffraction to a higher resolution. Thermal denaturation assay is a fast, high throughput technique for measuring the thermostability of solubilized proteins [48]. This assay is based on a thiol-specific fluorescent dye called N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl] maleimide (CPM). CPM is nonfluorescent in its unbound state and becomes fluorescent upon binding to cysteine residues. Cysteines are usually embedded in the membrane, located at helix-helix interaction sites and become solvent exposed as a result of unfolding. One can monitor the accessibility of cysteine residues to CPM during unfolding process induced by raising the temperature [48]. In case the target protein does not contain free thiol groups, cysteine residues are required to be engineered into the protein.

Size exclusion chromatography (SEC) is currently the most popular method for assessing the quality of solubilized membrane proteins in terms of homogeneity and aggregation. However, SEC requires a relatively large amount of protein sample and preparing and running the samples might be laborious and time consuming. Therefore, SEC does not possess the necessary requirements for a high-throughput technique. A fast and sensitive alternative method with an established efficacy in characterizing membrane proteins in solution is dynamic light scattering (DLS). DLS monitors the scattered light by macromolecules present in the solution and based on rate of fluctuations in the light scattering can determine the dimensions, homogeneity and stability of samples [49]. A detergent solubilized-membrane protein sample contains protein-detergent complexes (PDC) as well as free detergent micelles and detergent monomers. Protein aggregates formed from unfolded or insolubilized membrane proteins can also exist. Each of these states have a unique behavior in solution and can be distinguished from one another using DLS. PDCs display a dominant single-peak with an average radii size of 5 to 10 nm, whereas empty micelles generate smaller and narrower peaks [49]. Large aggregates have a complex peak distribution with much larger radii. Detergents suitable for membrane protein extraction are not necessarily suitable for stability and activity of proteins and can be exchanged based on SEC or DLS results. Using DLS, Meyer and colleagues showed that different detergents alter the

distribution of hydrodynamic radii and stability of membrane proteins in the solution over time [50]. DLS requires a small amount of sample (0.5 to 2 ul, 0.3 to 50 mg/mL) and can detect small differences in the hydrodynamic radii [50].

4. Membrane protein crystallization

Compared to soluble proteins, crystallization of membrane proteins is notoriously difficult, mainly because membrane proteins are extracted from their native phospholipid environment and transferred to detergents or membrane mimetics (*see section* 2). This can cause several hurdles for membrane protein crystallographers. For instance, protein-free micelles can hamper protein-protein interactions and reduce the success rate of crystallization. Moreover, detergents and membrane mimetics cover most parts of the membrane protein (hydrophobic region) and leave a small surface area (loops and hydrophilic region) for forming crystal contacts. In addition, as will be discussed in the next section, crystals formed from detergent-solubilized proteins are often associated with low-resolution diffraction or crystallography defects such as anisotropy or twinning. In order to overcome these challenges, new technological innovations have been introduced over the past two decades. Vapor diffusion crystallography, however, is often the first choice for membrane protein crystallographers [47].

4.1 Vapor diffusion crystallization

Following pre-crystallization screening of purified membrane proteins (*see section* 3) and identifying the most stable samples, the chosen samples will be concentrated using centrifugal concentrators [15]. High concentration of purified protein is required to achieve supersaturated conditions. However, over-concentrating can lead to protein instability and aggregation. To attain the highest stable concentration of macromolecule, small aliquots can be continuously concentrated and monitored by DLS for any sign of aggregation. The concentration at which the macromolecule starts aggregating determines the concentration threshold. If DLS is not available or the amount of protein is not enough for concentration trials, 10 mg/mL concentration can be used as a rule of thumb for the majority of transmembrane proteins [51]. Large proteins (>30 kDa) require less concentration, 2-5 mg/mL, whereas small proteins (<10 kDa) require higher

concentration, 20-50 mg/mL, to achieve supersaturation solution, as established by Michael Sawaya's laboratory (https://people.mbi.ucla.edu/sawaya/).

Similar to soluble proteins, sitting-drop and hanging-drop vapor diffusion are commonly used for crystallizing membrane proteins (Figure 5A). The concentrated protein is screened manually or using liquid handler robots against a wide range of commercially available or rationally designed screens. The goal of this initial screening is to identify conditions that generate "hits" suitable for optimization into well diffracting crystals. MemGold, MemGold2, MemPlus, and MemTrans (www.moleculardimensions.com) are only a few of these commercially available screens. While these commercially available screens are designed based on the most successful conditions that have led to high-resolution crystal structures of α-helical transmembrane proteins [30, 52] or β-barrel proteins [53], rational screens that systematically investigate a broad range of salts, pH and polyethylene glycol (PEG) type to identify the crystallization conditions can also be used (Figure 5B). Following the identification of the initial hits, a fine grid screen is designed around the hit condition to optimize the pH, salt and PEG size/concentration (Figure 5B). PEG is the most common precipitant used for crystallizing transmembrane proteins.

4.2 *In meso* crystallization

To overcome the obstacles associated with crystallizing detergent-solubilized transmembrane proteins, *in meso* crystallization was developed [54-56]. Lipidic Cubic Phase (LCP) provides a more native-like membrane mimetic environment for crystallizing membrane proteins (Figure 5A), and has promoted crystallization of difficult membrane proteins, particularly GPCRs [57, 58]. Mesophase is formed when detergent-solubilized protein is mixed with neutral lipids such as monoacylglycerol (MAGs). Mesophase is composed of three-dimensional lipid bilayers with separated water channels [59]. Membrane protein is reconstituted from detergent micelles into the bilayer part of the mesophase. Adding a precipitant can trigger a phase separation and formation of a lamellar phase [60]. This lamellar phase is where the protein enrichment, nucleation and subsequently crystal growth takes place [59]. With current advances in LCP forming lipids, this technique can be successfully applied for crystallizing membrane proteins stabilized in a wide range of temperatures, 4-55 ^{o}C [55, 61], pHs, 3.5-9.0 [62] or other harsh

conditions [63]. A detailed protocol for crystallizing membrane proteins in LCP is provided here [64].

4.3 Crystallization chaperones

Another technology that has been helpful in crystallizing challenging membrane proteins is crystallization chaperones. Crystallization chaperones are soluble proteins that specifically bind the target membrane protein, expanding the surface area necessary for forming crystal contacts and hence promoting crystallization. One example of a crystallization chaperone that has been successful for membrane protein crystallization is T4 lysozyme [65, 66]. T4 lysozyme has a high tendency to crystallize, which triggers crystallization of its binding partner. For example, crystallization of ß$_2$-adrenergic receptor (ß$_2$AR) was achieved when the entire third intracellular loop (ICL3) was replaced with T4 lysozyme [65, 67] (Figure 6A). ICL3 is highly flexible and plays a key role in interaction with G-protein. T4 fusion provided extra polar surface area and restricted the movement of protein at this region leading to crystals, which diffracted to 2.4 Å resolution. Although this strategy has been crucial for structure determination of several membrane proteins, including A_{2A} adenosine receptor (A_{2A}AR) [68], chemokine CXCR4 receptor [66], dopamine D3 receptor [69], histamine H1 receptor [70], δ–opioid receptor [71] and others, in some cases it has impeded protein functionality [72]. Another challenge associated with T4 lysozyme fusion is the number of constructs that need to be generated, since the placement of T4 lysozyme on the target protein is crucial for the protein's solubility and functionality. Each construct carries the T4 lysozyme fused to a different loop and the overexpression and thermal stability of each recombinant protein need to be determined empirically, making it a time-consuming, laborious and expensive process.

More commonly, proteins that specifically recognize and bind the target membrane protein have been used as crystallization chaperones. One common example is antigen-binding fragment (Fab) (Figure 6B). Fab is a fragment of an antibody, raised in small laboratory animals and selected to recognize a specific epitope on the target protein with extremely high affinity. Fab binding stabilizes the protein in a fixed conformation. Reducing flexibility lowers conformational heterogeneity and together with a larger polar surface area facilitates crystal

contact formation. This has been essential for solving the crystal structures of numerous membrane proteins including K^+ channel KcsA [73] (Figure 6B), ClC chloride channel [74], SecYE protein-conducting channel [75], nitric oxide reductase [76], and bestrophin calcium-activated chloride channel [77] among many others. However, raising antibodies in small laboratory animals is a very time-consuming and costly process and may not be successful for every membrane protein.

An alternative platform to Fab antibody fragments are nanobodies (Nb) (Figure 6C). Conventional antibodies are typically composed of two heavy (H)-chains and two light (L)-chains, and both chains made of constant (C) domains and variable (V) domains. Variable domains from both H chain (VH) and L chain (VL) form the antigen-binding domain in conventional antibodies. In contrast to conventional antibodies, camelid antibodies are composed of only H chains, and the paired VH-VH domains constitute the antigen-binding domain [78]. These unique variable domains from the H chains of camelid antibodies are termed VHH, or nanobodies [78]. Nanobodies comprise nine antiparallel β-strands organized in a 4 + 5 β-sheet conformation and connected by short loops and stabilized by a conserved disulfide bond [78]. The simplicity of nanobody structures, the lower molecular weight, and the biochemical tractability has made them an attractive tool for structural and functional studies of membrane proteins [79, 80]. Libraries of nanobodies, generated by protein engineering and capable of recognizing distinct epitope have been generated for yeast surface display [81, 82] or ribosome display [83], allowing rapid in vitro screening for binders to a target macromolecule. Compared to Fab production, which requires immunizing small laboratory animals, fusing spleen cells with myeloma cell lines and screening for monoclonal antibody production [84] (a several months to years-long process), nanobodies can be generated in only three to four weeks [81], making their production process relatively short and far less costly. The yeast surface display library of nanobodies developed by the Kruse laboratory [81] can be obtained for non-profit research from Kerafast (https://www.kerafast.com/item/1770/yeast-display-nanobody-library-nblib).
More importantly, unlike Fabs, which cannot be expressed in bacteria, high-affinity nanobody binders can be isolated from the library and expressed as soluble, recombinant proteins in *E. coli* [81].

Monobodies present a distinct class of synthetic proteins, which has been successfully employed as crystallization chaperones for membrane proteins (Figure 6D). Similar to Fab and nanobodies, monobodies are highly specific to their targets and provide an excellent platform for stabilizing and expanding the surface area of membrane proteins necessary for forming crystal contacts [85, 86]. Monobodies are based on human fibronectin type III domain (FN3), composed of seven antiparallel β-sheets connected by three loops on each side of the protein [87]. Diversification of two loops on opposite ends of the scaffold protein and the connecting β-sheet enable them to interact with both convex and concave surfaces on the target membrane proteins [88]. This feature gives monobodies a major advantage over the rest of crystallization chaperones in identifying an increased number of epitopes [88]. There are currently ~50 PDB entries for monobody-bound proteins. Unlike antibodies, monobodies do not contain disulfide bonds within their structures, which allows their overexpression in reducing environments, such as the *E. coli* cytoplasm. Their smaller size (~10 kDa) also provides an additional advantage for studying small transmembrane proteins. Using monobodies, the high resolution crystal structures of the SMR family of multidrug transporters [89, 90] (Figure 6D), fluoride channel Fluc [91], adhesion GPCR GPR56 [92], and fluoride H^+ antiporter [93] have been determined.

5. Membrane protein structure determination

Crystals suitable for membrane protein X-ray crystallography are formed when membrane proteins and detergents or lipids pack together in an orderly manner. Depending on the organization of this packing, two types of crystals can form, type I and type II 3D crystals [94]. In type I crystals, which are often obtained from LCP crystallization, protein molecules and lipids are organized in planar sheets through hydrophobic interactions, while protein-protein interactions are stabilized by polar interactions (Figure 7A). In type II 3D crystals, which more frequently form from micellar solutions and vapor diffusion crystallography, the hydrophobic regions of membrane proteins are covered by detergent micelles and only surface hydrophilic regions are accessible to form necessary protein-protein contacts (Figure 7B). Therefore, polar interactions are the main stabilizing force in this type of crystals. These crystals often have high solvent content [26]. As a result, membrane protein crystals grown in vapor diffusion

crystallization are typically very fragile, difficult to handle, diffract to low resolution, and are sensitive to radiation damage during data collection [7].

5.1 Anisotropy

A typical diffraction pattern is shown in Figure 7C, in which the crystal diffracts uniformly in every direction. However, in some cases the reflection intensities are higher in one direction compared to the other directions [95] (Figure 7D). This trait, which is caused by disorder between crystal packing and is very common between membrane protein crystals [8, 96], is termed anisotropy. Previously, anisotropic data were considered as unusable for structural determination purposes, but current developments in the software applications have come to the aid of membrane protein crystallographers to overcome this issue. To this end, the diffraction data are collected to the resolution limit of the best direction and then truncated using anisotropy servers such as STARANISO [97] or UCLA Diffraction Anisotropy Server [98]. The truncation is performed in an ellipsoidal shape instead of spherical averaging (Figure 7D). This enables inclusion of the high-resolution data during data processing, which otherwise would have been discarded. High resolution structural determination (up to 2.2 Å) of the Small Multidrug Resistance family of transporters using anisotropic data is a very good example of the efficacy of this platform [90]. Other examples are mitochondrial complex I [99], cytochrome bd oxidase [100], and $ß_2$-adrenergic receptor-Gs protein complex [101].

5.2 Phasing

Diffraction data (Figure 7C-D) provide the amplitude of the diffraction spots but lacks the phase information. Both the amplitude and phase are required to reconstruct the electron density map. The most convenient and fastest approach to obtain the phase is to calculate it from previously solved homologous structures. This approach, molecular replacement (MR), has led to solving the crystal structure of more than 60% of membrane proteins [102]. In the case of membrane proteins with novel structures, experimental phasing is required. Phasing the structure of membrane transporters, in particular, is highly dependent on experimental phasing [103]. The members of this family of transmembrane proteins display a high level of irregularity in their

helical structures, hence, the MR search programs fail to find an accurate solution for these proteins [103]. Heavy atom (HA) derivatization [104] and selenomethionine incorporation [105] are commonly used as experimental phasing techniques. In the former technique, a heavy atom is incorporated into the protein structure and changes in diffraction amplitude are measured (HAs with higher atomic number and more electrons present a stronger X-ray scattering). Single isomorphous replacement (SIR) and multiple isomorphous replacement (MIR) refer to one and several data sets collected in the presence and absence of HA, respectively. If the anomalous scattering at the absorption edges of the HA is used to calculate the initial phases, this method is termed single-wavelength anomalous dispersion (SAD) or multiple-wavelength anomalous dispersion (MAD) based on the number of wavelengths used.

Incorporating HA can be made during membrane protein purification or by soaking crystals after crystallization. Pretreating the membrane protein with HA allows testing the HA binding and its effects on protein stability and aggregation states, whereas crystal soaking enables parallel derivatization of protein crystals with different HAs. Other benefits of HA derivatization are (i) to map the position of residues in the electron density map, when the diffraction resolution is low, and (ii) identifying small molecule binding pockets using small molecules derivatized by HAs such as bromide or iodine [104]. For a list of HAs successfully employed in membrane protein experimental phasing see [103].

Substituting methionine with selenomethionine (SeMet) is a routine alternative to HA derivatization. Selenium with 34 electrons, compared to sulfur which only has 16 electrons, provides a stronger X-ray scattering and superior phasing power [104]. This technique utilizes a methionine-auxotroph *E. coli* strain B834 (DE3), which is not able to produce its own methionine and is dependent on the growth media for supplying the required methionine. Therefore, providing a growth media supplemented with SeMet, can result in producing recombinant proteins with SeMet residues instead of methionine. Depending on the growth and induction strategies, the level of SeMet incorporation can be different [106]. Although this technique is fully optimized for *E. coli* expression, reliable and consistent SeMet protein production in other expression systems is not always definite. Lack of methionine-auxotrophic

strains, low protein yield and low rate of SeMet incorporation as well as SeMet toxicity are some of these challenges [106, 107].

While HA derivatization is expensive and hazardous, and SeMet derivatization generates very low yield of labelled protein, faster and more efficient experimental phasing techniques for membrane proteins are required. Iodide single-wavelength anomalous diffraction (I-SAD) present a promising alternative to these techniques [108]. This technique is based on the fact that positive residues, arginine and lysine and to a lesser extent histidine, are enriched on the cytoplasmic side of transmembrane proteins (positive-inside rule) [109] and they are eager to bind to negatively charged ions, such as iodide, to compensate their charge. This makes iodide incorporation to the membrane proteins easy, fast and efficient [108]. Moderate concentrations of sodium iodide (NaI), 0.2 to 0.5 mM, are used to co-crystallize it with the membrane protein [110] or in soaking solution for existing membrane protein crystals [111].

Summary

It is well-known among structural biologists that crystallizing membrane proteins can be very difficult. Some of its difficulties stem from (i) poor overexpression, (ii) extraction and solubilization difficulties, (iii) instability or loss of function of solubilized proteins, and (iv) generating high-resolution diffracting crystals. Despite these challenges, X-ray crystallography has been instrumental in our understanding of the structures of transmembrane proteins. Technological developments during the past two decades have come to the aid of membrane protein crystallographers to overcome some of these challenges. Unfortunately, there is not a universal platform for crystallizing all membrane proteins, and each membrane protein has unique characteristics. However, familiarity with available techniques and methods enables the crystallographer to make the most informed decision for each step of the process. In this review, an overview of membrane protein crystallography is provided and some of the inherent obstacles to the practitioner, along with available solutions, are discussed.

Acknowledgments

I would like to thank Dr. Randy Stockbridge and her research group for critical review and editing this manuscript.

References

1. Sanders, C. R. & Myers, J. K. (2004) Disease-Related Misassembly of Membrane Proteins, *Annual Review of Biophysics and Biomolecular Structure.* **33**, 25-51.

2. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J Mol Biol.* **305**, 567-80.

3. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. & Pontén, F. (2015) Tissue-based map of the human proteome, *Science.* **347**, 1260419.

4. Jia, B. & Jeon, C. O. (2016) High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives, *Open Biol.* **6**, 160196.
5. Freigassner, M., Pichler, H. & Glieder, A. (2009) Tuning microbial hosts for membrane protein production, *Microbial Cell Factories.* **8**, 69.
6. Werten, P. J. L., Rémigy, H. W., de Groot, B. L., Fotiadis, D., Philippsen, A., Stahlberg, H., Grubmüller, H. & Engel, A. (2002) Progress in the analysis of membrane protein structure and function, *FEBS Letters.* **529**, 65-72.
7. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. (2008) Overcoming the challenges of membrane protein crystallography, *Current Opinion in Structural Biology.* **18**, 581-586.
8. Robert, X., Kassis-Sahyoun, J., Ceres, N., Martin, J., Sawaya, M. R., Read, R. J., Gouet, P., Falson, P. & Chaptal, V. (2017) X-ray diffraction reveals the intrinsic difference in the physical properties of membrane and soluble proteins, *Scientific Reports.* **7**, 17013.
9. Murata, K. & Wolf, M. (2018) Cryo-electron microscopy for structural analysis of dynamic biological macromolecules, *Biochimica et Biophysica Acta (BBA) - General Subjects.* **1862**, 324-334.
10. Glaeser, R. M. & Hall, R. J. (2011) Reaching the information limit in cryo-EM of biological macromolecules: experimental aspects, *Biophys J.* **100**, 2331-2337.
11. Kühlbrandt, W. (2014) Biochemistry. The resolution revolution, *Science.* **343**, 1443-4.
12. Callaway, E. (2020) Revolutionary cryo-EM is taking over structural biology, *Nature.* **578**, 201.
13. Liu, Y., Huynh, D. T. & Yeates, T. O. (2019) A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold, *Nature Communications.* **10**, 1864.
14. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., Chirgadze, D. Y., Murshudov, G., Aricescu, A. R. & Scheres, S. H. W. (2020) Single-particle cryo-EM at atomic resolution, *Nature.* **587**, 152-156.

15. McIlwain, B. C. & Kermani, A. A. (2020) Membrane Protein Production in Escherichia coli in *Expression, Purification, and Structural Biology of Membrane Proteins* (Perez, C. & Maier, T., eds) pp. 13-27, Springer US, New York, NY.

16. Bernaudat, F., Frelet-Barrand, A., Pochon, N., Dementin, S., Hivin, P., Boutigny, S., Rioux, J.-B., Salvi, D., Seigneurin-Berny, D., Richaud, P., Joyard, J., Pignol, D., Sabaty, M., Desnos, T., Pebay-Peyroula, E., Darrouzet, E., Vernet, T. & Rolland, N. (2011) Heterologous expression of membrane proteins: choosing the appropriate host, *PLoS One.* **6**, e29191-e29191.

17. Wagner, S., Baars, L., Ytterberg, A. J., Klussmeier, A., Wagner, C. S., Nord, O., Nygren, P. A., van Wijk, K. J. & de Gier, J. W. (2007) Consequences of membrane protein overexpression in Escherichia coli, *Mol Cell Proteomics.* **6**, 1527-50.

18. Studier, F. W. & Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *Journal of Molecular Biology.* **189**, 113-130.

19. Zhang, X. & Studier, F. W. (1997) Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme, *J Mol Biol.* **269**, 10-27.

20. Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O. & de Gier, J. W. (2008) Tuning Escherichia coli for membrane protein overexpression, *Proc Natl Acad Sci U S A.* **105**, 14371-6.

21. Miroux, B. & Walker, J. E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, *J Mol Biol.* **260**, 289-98.

22. Parker, J. L. & Newstead, S. (2014) Method to increase the yield of eukaryotic membrane protein expression in Saccharomyces cerevisiae for structural and functional studies, *Protein Sci.* **23**, 1309-1314.

23. Byrne, B. (2015) Pichia pastoris as an expression host for membrane protein structural biology, *Current Opinion in Structural Biology.* **32**, 9-17.

24. Dukkipati, A., Park, H. H., Waghray, D., Fischer, S. & Garcia, K. C. (2008) BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies, *Protein Expr Purif.* **62**, 160-170.

25. Drew, D., Lerch, M., Kunji, E., Slotboom, D.-J. & de Gier, J.-W. (2006) Optimization of membrane protein overexpression and purification using GFP fusions, *Nature Methods.* **3**, 303-313.

26. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. (2014) Membrane protein structure determination - the next generation, *Biochim Biophys Acta.* **1838**, 78-87.

27. Kawate, T. & Gouaux, E. (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins, *Structure.* **14**, 673-81.

28. Seddon, A. M., Curnow, P. & Booth, P. J. (2004) Membrane proteins, lipids and detergents: not just a soap opera, *Biochimica et Biophysica Acta (BBA) - Biomembranes.* **1666**, 105-117.

29. Stetsenko, A. & Guskov, A. (2017) An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization, *Crystals.* **7**.

30. Newstead, S., Ferrandon, S. & Iwata, S. (2008) Rationalizing α-helical membrane protein crystallization, *Protein Science.* **17**, 466-472.

31. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. (2002) Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins, *Nano Letters.* **2**, 853-856.

32. Bayburt, T. H. & Sligar, S. G. (2010) Membrane protein assembly into Nanodiscs, *FEBS letters.* **584**, 1721-1727.

33. Boldog, T., Li, M. & Hazelbauer, G. L. (2007) Using Nanodiscs to create water-soluble transmembrane chemoreceptors inserted in lipid bilayers, *Methods Enzymol.* **423**, 317-35.

34. Goddard, A. D., Dijkman, P. M., Adamson, R. J., dos Reis, R. I. & Watts, A. (2015) Reconstitution of membrane proteins: a GPCR as an example, *Methods Enzymol.* **556**, 405-24.

35. Rouck, J. E., Krapf, J. E., Roy, J., Huff, H. C. & Das, A. (2017) Recent advances in nanodisc technology for membrane protein studies (2012-2017), *FEBS letters.* **591**, 2057-2088.

36. Nikolaev, M., Round, E., Gushchin, I., Polovinkin, V., Balandin, T., Kuzmichev, P., Shevchenko, V., Borshchevskiy, V., Kuklin, A., Round, A., Bernhard, F., Willbold, D., Büldt, G. & Gordeliy, V. (2017) Integral Membrane Proteins Can Be Crystallized Directly from Nanodiscs, *Crystal Growth & Design.* **17**, 945-948.

37. Su, C.-C., Morgan, C. E., Kambakam, S., Rajavel, M., Scott, H., Huang, W., Emerson, C. C., Taylor, D. J., Stewart, P. L., Bonomo, R. A. & Yu, E. W. (2019) Cryo-Electron Microscopy Structure of an Acinetobacter baumannii Multidrug Efflux Pump, *mBio.* **10**, e01295-19.
38. Shen, P. S., Yang, X., DeCaen, P. G., Liu, X., Bulkley, D., Clapham, D. E. & Cao, E. (2016) The Structure of the Polycystic Kidney Disease Channel PKD2 in Lipid Nanodiscs, *Cell.* **167**, 763-773.e11.
39. Frauenfeld, J., Löving, R., Armache, J. P., Sonnen, A. F., Guettou, F., Moberg, P., Zhu, L., Jegerschöld, C., Flayhan, A., Briggs, J. A., Garoff, H., Löw, C., Cheng, Y. & Nordlund, P. (2016) A saposin-lipoprotein nanoparticle system for membrane proteins, *Nat Methods.* **13**, 345-51.
40. Phillips, R., Ursell, T., Wiggins, P. & Sens, P. (2009) Emerging roles for lipids in shaping membrane-protein function, *Nature.* **459**, 379-385.
41. Guo, Y. (2020) Be Cautious with Crystal Structures of Membrane Proteins or Complexes Prepared in Detergents, *Crystals.* **10**, 86.
42. Teo, A. C. K., Lee, S. C., Pollock, N. L., Stroud, Z., Hall, S., Thakker, A., Pitt, A. R., Dafforn, T. R., Spickett, C. M. & Roper, D. I. (2019) Analysis of SMALP co-extracted phospholipids shows distinct membrane environments for three classes of bacterial membrane protein, *Scientific Reports.* **9**, 1813.
43. Broecker, J., Eger, B. T. & Ernst, O. P. (2017) Crystallogenesis of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs, *Structure.* **25**, 384-392.
44. Qiu, W., Fu, Z., Xu, G. G., Grassucci, R. A., Zhang, Y., Frank, J., Hendrickson, W. A. & Guo, Y. (2018) Structure and activity of lipid bilayer within a membrane-protein transporter, *Proc Natl Acad Sci U S A.* **115**, 12985-12990.
45. Sun, C., Benlekbir, S., Venkatakrishnan, P., Wang, Y., Hong, S., Hosler, J., Tajkhorshid, E., Rubinstein, J. L. & Gennis, R. B. (2018) Structure of the alternative complex III in a supercomplex with cytochrome oxidase, *Nature.* **557**, 123-126.
46. Yu, J., Zhu, H., Lape, R., Greiner, T., Shahoei, R., Wang, Y., Du, J., Lü, W., Tajkhorshid, E., Sivilotti, L. & Gouaux, E. (2019) Mechanism of gating and partial agonist action in the glycine receptor, *bioRxiv*, 786632.

47. Newby, Z. E., O'Connell, J. D., 3rd, Gruswitz, F., Hays, F. A., Harries, W. E., Harwood, I. M., Ho, J. D., Lee, J. K., Savage, D. F., Miercke, L. J. & Stroud, R. M. (2009) A general protocol for the crystallization of membrane proteins for X-ray structural investigation, *Nat Protoc.* **4**, 619-37.
48. Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. A. & Stevens, R. C. (2008) Microscale Fluorescent Thermal Stability Assay for Membrane Proteins, *Structure.* **16**, 351-359.
49. Kwan, T. O. C., Reis, R., Siligardi, G., Hussain, R., Cheruvara, H. & Moraes, I. (2019) Selection of Biophysical Methods for Characterisation of Membrane Proteins, *Int J Mol Sci.* **20**, 2605.
50. Meyer, A., Dierks, K., Hussein, R., Brillet, K., Brognaro, H. & Betzel, C. (2015) Systematic analysis of protein-detergent complexes applying dynamic light scattering to optimize solutions for crystallization trials, *Acta Crystallogr F Struct Biol Commun.* **71**, 75-81.
51. Ma, P., Weichert, D., Aleksandrov, L. A., Jensen, T. J., Riordan, J. R., Liu, X., Kobilka, B. K. & Caffrey, M. (2017) The cubicon method for concentrating membrane proteins in the cubic mesophase, *Nature Protocols.* **12**, 1745-1762.
52. Parker, J. L. & Newstead, S. (2012) Current trends in α-helical membrane protein crystallization: an update, *Protein Sci.* **21**, 1358-1365.
53. Newstead, S., Hobbs, J., Jordan, D., Carpenter, E. P. & Iwata, S. (2008) Insights into outer membrane protein crystallization, *Mol Membr Biol.* **25**, 631-638.
54. Landau, E. M. & Rosenbusch, J. P. (1996) Lipidic cubic phases: A novel concept for the crystallization of membrane proteins, *Proceedings of the National Academy of Sciences.* **93**, 14532.
55. Qiu, H. & Caffrey, M. (2000) The phase diagram of the monoolein/water system: metastability and equilibrium aspects, *Biomaterials.* **21**, 223-234.
56. Caffrey, M. (2003) Membrane protein crystallization, *J Struct Biol.* **142**, 108-32.
57. Bertheleme, N., Chae, P. S., Singh, S., Mossakowska, D., Hann, M. M., Smith, K. J., Hubbard, J. A., Dowell, S. J. & Byrne, B. (2013) Unlocking the secrets of the gatekeeper: Methods for stabilizing and crystallizing GPCRs, *Biochimica et Biophysica Acta (BBA) - Biomembranes.* **1828**, 2583-2591.

58. Ghosh, E., Kumari, P., Jaiman, D. & Shukla, A. K. (2015) Methodological advances: the unsung heroes of the GPCR structural revolution, *Nature Reviews Molecular Cell Biology.* **16**, 69-81.

59. Caffrey, M. (2015) A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes, *Acta crystallographica Section F, Structural biology communications.* **71**, 3-18.

60. Caffrey, M. (2008) On the Mechanism of Membrane Protein Crystallization in Lipidic Mesophases, *Crystal Growth & Design.* **8**, 4244-4254.

61. Salvati Manni, L., Zabara, A., Osornio, Y. M., Schöppe, J., Batyuk, A., Plückthun, A., Siegel, J. S., Mezzenga, R. & Landau, E. M. (2015) Phase behavior of a designed cyclopropyl analogue of monoolein: implications for low-temperature membrane protein crystallization, *Angew Chem Int Ed Engl.* **54**, 1027-31.

62. Zha, J. & Li, D. (2018) Lipid Cubic Phase for Membrane Protein X-ray Crystallography in pp. 175-220.

63. Ishchenko, A., Peng, L., Zinovev, E., Vlasov, A., Lee, S. C., Kuklin, A., Mishin, A., Borshchevskiy, V., Zhang, Q. & Cherezov, V. (2017) Chemically Stable Lipids for Membrane Protein Crystallization, *Crystal growth & design.* **17**, 3502-3511.

64. Caffrey, M. & Cherezov, V. (2009) Crystallizing membrane proteins using lipidic mesophases, *Nat Protoc.* **4**, 706-31.

65. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor, *Science.* **318**, 1258-1265.

66. Wu, B., Chien, E. Y. T., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V. & Stevens, R. C. (2010) Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists, *Science.* **330**, 1066.

67. Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Yao, X. J., Weis, W. I., Stevens, R. C. & Kobilka, B. K. (2007) GPCR engineering yields

high-resolution structural insights into beta2-adrenergic receptor function, *Science.* **318**, 1266-73.

68. Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., Ijzerman, A. P. & Stevens, R. C. (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist, *Science.* **322**, 1211-7.

69. Chien, E. Y., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V. & Stevens, R. C. (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist, *Science.* **330**, 1091-5.

70. Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C. & Iwata, S. (2011) Structure of the human histamine H1 receptor complex with doxepin, *Nature.* **475**, 65-70.

71. Granier, S., Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Weis, W. I. & Kobilka, B. K. (2012) Structure of the δ-opioid receptor bound to naltrindole, *Nature.* **485**, 400-404.

72. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. (2009) The structure and function of G-protein-coupled receptors, *Nature.* **459**, 356-363.

73. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. (2001) Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution, *Nature.* **414**, 43-8.

74. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. (2002) X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity, *Nature.* **415**, 287-94.

75. Tsukazaki, T., Mori, H., Fukai, S., Ishitani, R., Mori, T., Dohmae, N., Perederina, A., Sugita, Y., Vassylyev, D. G., Ito, K. & Nureki, O. (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures, *Nature.* **455**, 988-991.

76. Hino, T., Matsumoto, Y., Nagano, S., Sugimoto, H., Fukumori, Y., Murata, T., Iwata, S. & Shiro, Y. (2010) Structural Basis of Biological N₂O Generation by Bacterial Nitric Oxide Reductase, *Science.* **330**, 1666.

77. Kane Dickson, V., Pedi, L. & Long, S. B. (2014) Structure and insights into the function of a Ca(2+)-activated Cl(-) channel, *Nature.* **516**, 213-8.

78. Muyldermans, S. (2013) Nanobodies: natural single-domain antibodies, *Annu Rev Biochem.* **82**, 775-97.

79. Manglik, A., Kobilka, B. K. & Steyaert, J. (2017) Nanobodies to Study G Protein-Coupled Receptor Structure and Function, *Annu Rev Pharmacol Toxicol.* **57**, 19-37.

80. Uchański, T., Pardon, E. & Steyaert, J. (2020) Nanobodies to study protein conformational states, *Current Opinion in Structural Biology.* **60**, 117-123.

81. McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., Erlandson, S. C., Hilger, D., Rasmussen, S. G. F., Ring, A. M., Manglik, A. & Kruse, A. C. (2018) Yeast surface display platform for rapid discovery of conformationally selective nanobodies, *Nat Struct Mol Biol.* **25**, 289-296.

82. Uchański, T., Zögg, T., Yin, J., Yuan, D., Wohlkönig, A., Fischer, B., Rosenbaum, D. M., Kobilka, B. K., Pardon, E. & Steyaert, J. (2019) An improved yeast surface display platform for the screening of nanobody immune libraries, *Scientific Reports.* **9**, 382.

83. Zimmermann, I., Egloff, P., Hutter, C. A. J., Arnold, F. M., Stohler, P., Bocquet, N., Hug, M. N., Huber, S., Siegrist, M., Hetemann, L., Gera, J., Gmür, S., Spies, P., Gygax, D., Geertsma, E. R., Dawson, R. J. P. & Seeger, M. A. (2018) Synthetic single domain antibodies for the conformational trapping of membrane proteins, *eLife.* **7**, e34317.

84. Partridge, L. J. (1994) The Production of Monoclonal Antibodies to Membrane Proteins in *Biomembrane Protocols: II Architecture and Function* (Graham, J. M. & Higgins, J. A., eds) pp. 65-86, Springer New York, Totowa, NJ.

85. Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. (2007) High-affinity single-domain binding proteins with a binary-code interface, *Proceedings of the National Academy of Sciences.* **104**, 6632.

86. Sha, F., Salzman, G., Gupta, A. & Koide, S. (2017) Monobodies and other synthetic binding proteins for expanding protein science, *Protein Sci.* **26**, 910-924.

87. Karatan, E., Merguerian, M., Han, Z., Scholle, M. D., Koide, S. & Kay, B. K. (2004) Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain, *Chem Biol.* **11**, 835-44.

88. Hantschel, O., Biancalana, M. & Koide, S. (2020) Monobodies as enabling tools for structural and mechanistic biology, *Current Opinion in Structural Biology.* **60**, 167-174.

89. Kermani, A. A., Macdonald, C. B., Gundepudi, R. & Stockbridge, R. B. (2018) Guanidinium export is the primal function of SMR family transporters, *Proceedings of the National Academy of Sciences.* **115**, 3060.

90. Kermani, A. A., Macdonald, C. B., Burata, O. E., Ben Koff, B., Koide, A., Denbaum, E., Koide, S. & Stockbridge, R. B. (2020) The structural basis of promiscuity in small multidrug resistance transporters, *Nature Communications.* **11**, 6064.

91. Stockbridge, R. B., Kolmakova-Partensky, L., Shane, T., Koide, A., Koide, S., Miller, C. & Newstead, S. (2015) Crystal structures of a double-barrelled fluoride ion channel, *Nature.* **525**, 548-551.

92. Salzman, G. S., Ackerman, S. D., Ding, C., Koide, A., Leon, K., Luo, R., Stoveken, H. M., Fernandez, C. G., Tall, G. G., Piao, X., Monk, K. R., Koide, S. & Araç, D. (2016) Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains, *Neuron.* **91**, 1292-1304.

93. Last, N. B., Stockbridge, R. B., Wilson, A. E., Shane, T., Kolmakova-Partensky, L., Koide, A., Koide, S. & Miller, C. (2018) A CLC-type F(-)/H(+) antiporter in ion-swapped conformations, *Nat Struct Mol Biol.* **25**, 601-606.

94. Birch, J., Axford, D., Foadi, J., Meyer, A., Eckhardt, A., Thielmann, Y. & Moraes, I. (2018) The fine art of integral membrane protein crystallisation, *Methods.* **147**, 150-162.

95. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. (2013) Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination, *The FEBS Journal.* **280**, 5705-5736.

96. Kane Dickson, V. (2016) Phasing and structure of bestrophin-1: a case study in the use of heavy-atom cluster compounds with multi-subunit transmembrane proteins, *Acta Crystallogr D Struct Biol.* **72**, 319-325.

97. Tickle, I., J ., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G. (2018) STARANISO in Cambridge, United Kingdom: Global Phasing Ltd. .

98. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from

Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences of the United States of America.* **103**, 8060-8065.

99. Zickermann, V., Wirth, C., Nasiri, H., Siegmund, K., Schwalbe, H., Hunte, C. & Brandt, U. (2015) Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I, *Science.* **347**, 44-9.

100. Safarian, S., Rajendran, C., Müller, H., Preu, J., Langer, J. D., Ovchinnikov, S., Hirose, T., Kusumoto, T., Sakamoto, J. & Michel, H. (2016) Structure of a bd oxidase indicates similar mechanisms for membrane-integrated oxygen reductases, *Science.* **352**, 583-6.

101. Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T. A., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K. & Kobilka, B. K. (2011) Crystal structure of the β2 adrenergic receptor–Gs protein complex, *Nature.* **477**, 549-555.

102. Huang, C.-Y., Olieric, V., Howe, N., Warshamanage, R., Weinert, T., Panepucci, E., Vogeley, L., Basu, S., Diederichs, K., Caffrey, M. & Wang, M. (2018) In situ serial crystallography for rapid de novo membrane protein structure determination, *Communications Biology.* **1**, 124.

103. Parker, J. L. & Newstead, S. (2013) Phasing statistics for alpha helical membrane protein structures, *Protein Science.* **22**, 1664-1668.

104. Pike, A. C. W., Garman, E. F., Krojer, T., von Delft, F. & Carpenter, E. P. (2016) An overview of heavy-atom derivatization of protein crystals, *Acta Crystallogr D Struct Biol.* **72**, 303-318.

105. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure, *EMBO J.* **9**, 1665-1672.

106. Walden, H. (2010) Selenium incorporation using recombinant techniques, *Acta Crystallogr D Biol Crystallogr.* **66**, 352-357.

107. Barton, W. A., Tzvetkova-Robev, D., Erdjument-Bromage, H., Tempst, P. & Nikolov, D. B. (2006) Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells, *Protein Sci.* **15**, 2008-2013.

108. Melnikov, I., Polovinkin, V., Kovalev, K., Gushchin, I., Shevtsov, M., Shevchenko, V., Mishin, A., Alekseev, A., Rodriguez-Valera, F., Borshchevskiy, V., Cherezov, V., Leonard, G. A., Gordeliy,

V. & Popov, A. (2017) Fast iodide-SAD phasing for high-throughput membrane protein structure determination, *Science Advances.* **3**, e1602952.

109. von Heijne, G. (1989) Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues, *Nature.* **341**, 456-8.

110. Gu, Y., Li, H., Dong, H., Zeng, Y., Zhang, Z., Paterson, N. G., Stansfeld, P. J., Wang, Z., Zhang, Y., Wang, W. & Dong, C. (2016) Structural basis of outer membrane protein insertion by the BAM complex, *Nature.* **531**, 64-9.

111. Gushchin, I., Melnikov, I., Polovinkin, V., Ishchenko, A., Yuzhakova, A., Buslaev, P., Bourenkov, G., Grudinin, S., Round, E., Balandin, T., Borshchevskiy, V., Willbold, D., Leonard, G., Büldt, G., Popov, A. & Gordeliy, V. (2017) Mechanism of transmembrane signaling by sensor histidine kinases, *Science.* **356**, eaah6345.

112. Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., Drew, D., Slotboom, D. J. & de Gier, J.-W. (2012) Optimizing Membrane Protein Overexpression in the Escherichia coli strain Lemo21(DE3), *Journal of Molecular Biology.* **423**, 648-659.

113. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes, *Methods in enzymology.* **185**, 60-89.

Table 1. Comparison of expression hosts used for overexpressing membrane proteins.

Characteristics	*Escherichia coli*	Yeast	Insect cells	Mammalian cells
Doubling time	15-20 min	90-120 min	24-72 hours	~13-24 hours
Growth cost	low	low	high	high
Growth difficulty	easy	easy	Requires complex media and cell culture facilities	Requires complex media and cell culture facilities
Expression level	high	moderate to high	low to moderate	low to moderate
Applying post-translational modifications	no	yes	yes	complex post-translational modifications

Table 2. Commonly used E. coli strains for membrane protein overexpression.

Strain	Description	Application	Reference
BL21(DE3)	an *E. coli* strain with DE3, a λ prophage carrying the T7 RNA polymerase gene and *lacI*	Producing large amounts of recombinant protein due to exploiting the T7 RNA polymerase	[18]
BL21(DE3)pLysS	Carries a second plasmid (pLysS), which encodes T7 lysozyme	Lowers the background expression level of target gene by reducing the activity of T7 RNA polymerase	[18], [19]
C41(DE3) and C43(DE3)	Derived from BL21 (DE3) strain with a weakening mutation in *lac*UV5 promoter	Increases the overexpression by preventing the cell death associated with expression of recombinant toxic protein	[20, 21]
Lemo21(DE3)	Expresses the T7 RNA polymerase inhibitor protein (LysY)	Tunes recombinant protein expression by varying the level of lysozyme (LysY) production	[20], [112]
Rosetta(DE3)	Carries pRARE plasmid encoding rare tRNA codons	Suitable for overexpressing eukaryotic proteins in *E. coli*	[113]

Table 1. The most successful purification and crystallization detergents based on [29, 30].

Detergent class/name	Chain length	CMC (mM)	CMC (% w/v)	Extraction (mM)	Purification (mM)	Micelle size (Da)
Nonionic						
n-dodecyl-ß-D-maltoside (DDM)	12C	0.17	0.0087	20	0.6	72,000
n-decyl- ß-D-maltoside (DM)	10C	1.8	0.087	21	5	33,000
n-Octyl- ß-D-Glucopyranoside (OG)	8C	18-20	0.53	68	40	25,000
n-Nonyl- ß-D-Glucopyranoside (NG)	9C	6.5	0.2			90,000
Zwitterionic						
Lauryldimethylamine-N-oxide (LDAO)	12	1-2	0.023	51	1.4-4	21.5

Figure 1. An overview of membrane protein crystallography.

Figure 2. Membrane protein production in different expression hosts. The proportion of solved 3D structure of (A) prokaryotic membrane proteins and (B) eukaryotic membrane proteins overexpressed in different expression hosts based on published unique protein structures as of August 2020 (https://blanco.biomol.uci.edu/mpstruc/query).

Figure 3. Extraction of transmembrane proteins from cell membranes. Proteins in the phospholipid bilayer from left to right were prepared from protein data bank with the accession codes 1OMF, 6WK8 and 4RY2, respectively. Detergents and bicelles, a mixture of detergents and lipids, have been used as conventional tools for extracting and solubilizing membrane proteins. Nanodiscs and SMALPs provide a more native-like environment for this purpose. They are composed of phospholipid patches surrounded by MSP protein (nanodiscs) or styrene malic acid (SMA) copolymer lipid particles (SMALP).

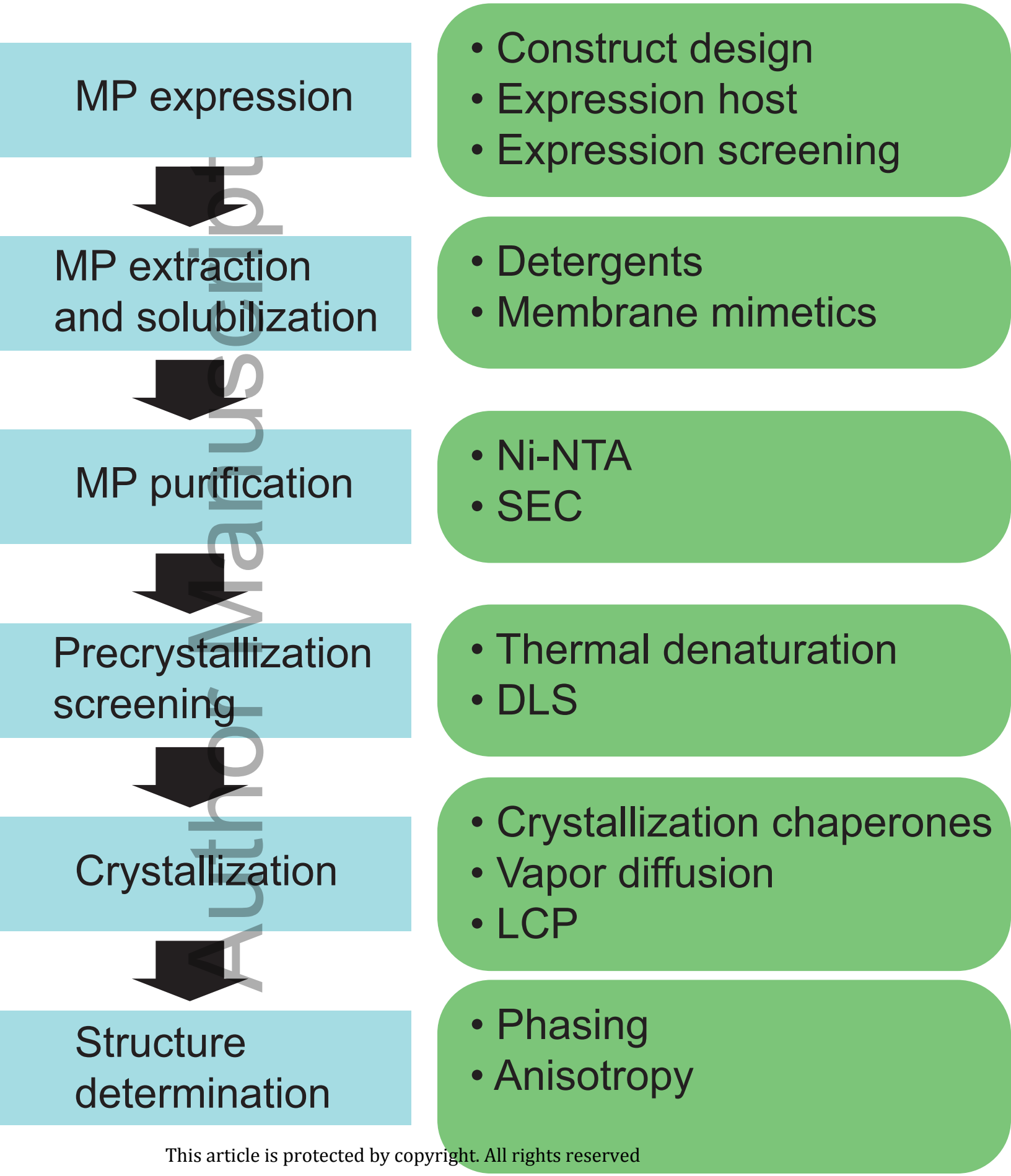
Figure 4. Chemical structures of different classes of detergents, used to extract and solubilize transmembrane proteins.

Figure 5. Schematic representation of vapor diffusion crystallization and lipidic cubic phase (LCP). (A) In vapor diffusion crystallography, the detergent-solubilized membrane protein is mixed with crystallization buffer and is used to set up crystallization drops. In the hanging drop method (left), the crystallization drop is hanging from the cover slide, whereas in the sitting drop method (right) the crystallization drop is placed on a pedestal next to the reservoir solution. In the LCP method, detergent-solubilized protein is mixed with neutral lipids, mainly

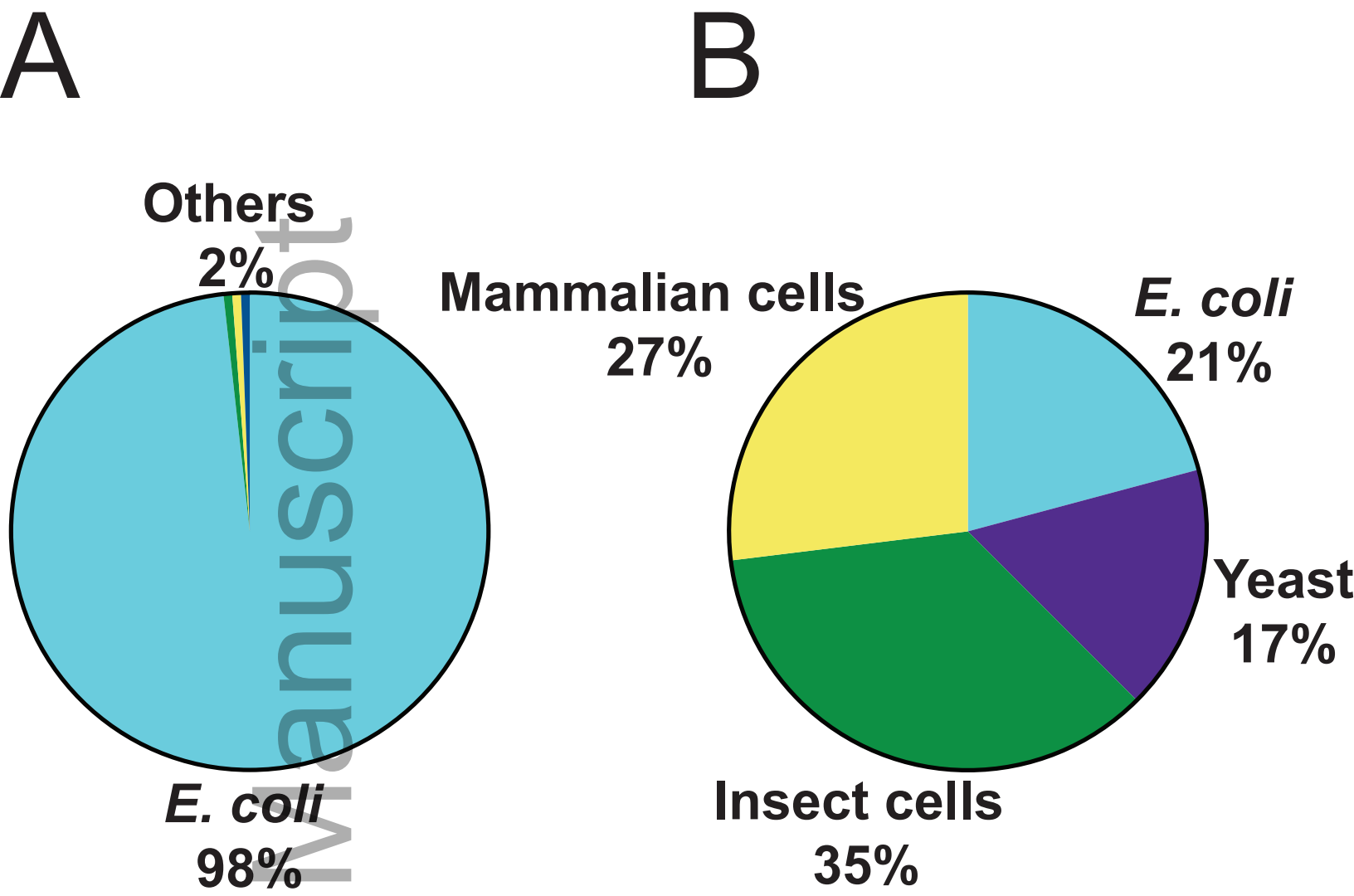
monoacylglycerol (MAGs), by using two syringes and a coupler. The mixed protein-lipid mixture is overlaid with buffer to set up crystallization drops on a glass sandwich plate. (B) A typical crystallization screen where the pH, salt and precipitant type/concentration is rationally modified. Upon identifying initial hits, further screens are designed around the hit condition, in which the pH, salt and PEG type/concentration are modified in small decrements or increments.

Figure 6. Crystallization chaperones developed for structural and functional studies of transmembrane proteins. (A) crystal structure of human ß2-adrenergic G protein-coupled receptor (purple) bound to T4 lysozyme (greencyan) (PDB code 2RH1). (B) crystal structure of potassium channel KcsA (yellow) solved in the presence of Fab (pink) (PDB code 1K4C). (C) crystal structure of angiotensin II type 1 receptor (blue) stabilized using nanobody S1I8 (green) (PDB code 6DO1). (D) crystal structure of the Small Multidrug Resistance (orange) in the presence of monobody (cyan) (PDB code 6WK8). Membrane proteins are shown in cylindrical shapes to make an easier distinction between membrane proteins and crystallization chaperones.

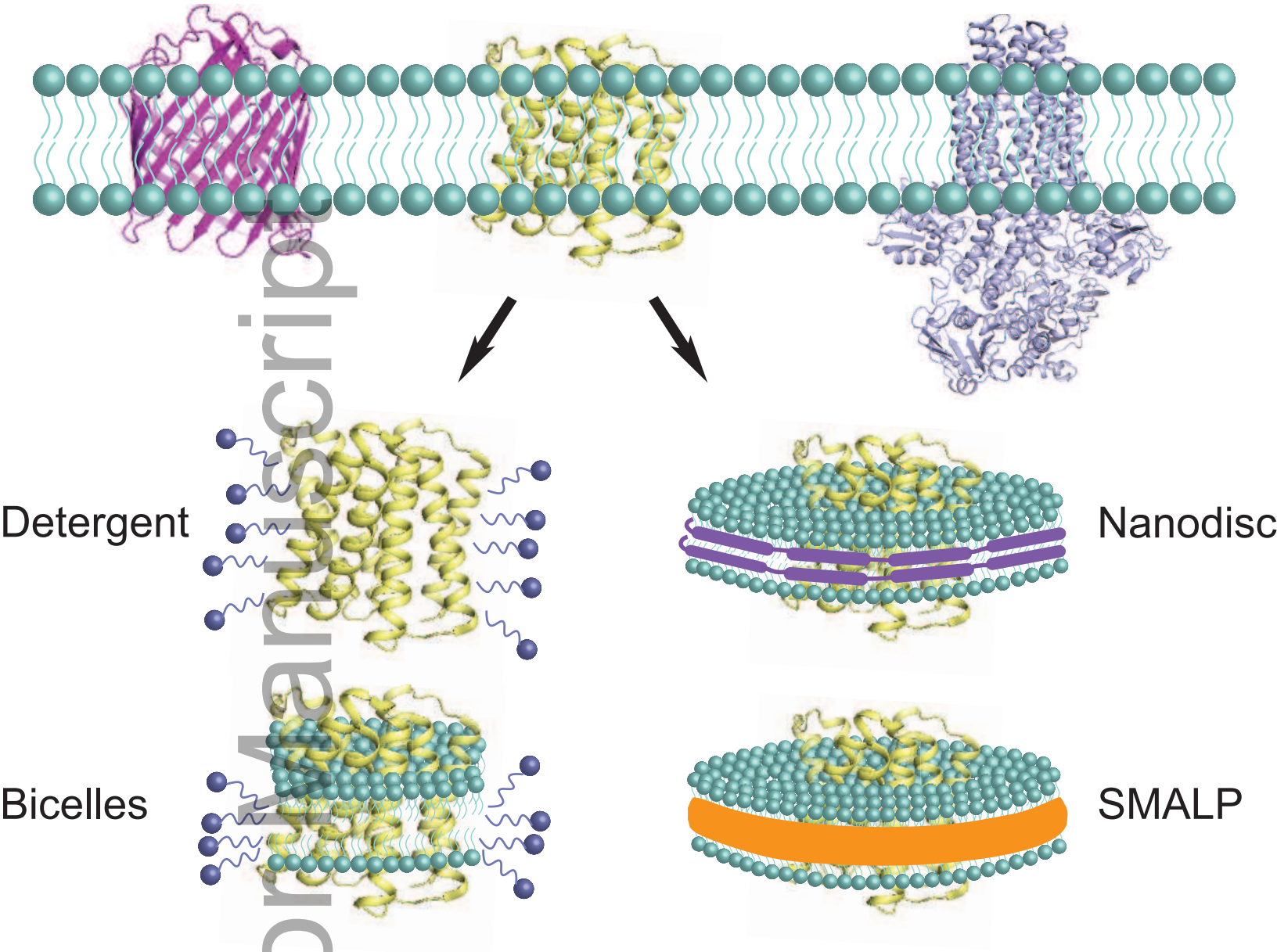
Figure 7. Crystal packing and diffraction in membrane protein crystallization. (A) Crystal packing type I from Bordetella pertussis Fluc channel (yellow) bound to a crystallization chaperone monobody (purple) (PDB accession code: 5A41) [91]. In this type, membrane proteins and lipids are arranged in planar sheets similar to the cell membrane and stacked on top of one another. This assembly is stabilized through hydrophobic and polar interactions. (B) Crystal packing type II from E. coli Fluc channel (yellow) bound to a crystallization chaperone monobody (purple) (PDB accession code: 5A43) [91]. In this type, crystals are formed largely due to the polar interactions between hydrophilic surface of membrane proteins, while the hydrophobic region is concealed by detergent micelles. Formation of large solvent channels is the main characteristic of this type of crystal packing. (C) A typical diffraction pattern of protein crystals, in which the diffraction spots are distributed uniformly in every direction. (D) Anisotropic diffraction pattern, in which the reflection intensities are higher in one direction compared to the other directions. Spherical averaging and ellipsoidal truncation are shown in green circle and black oval, respectively.



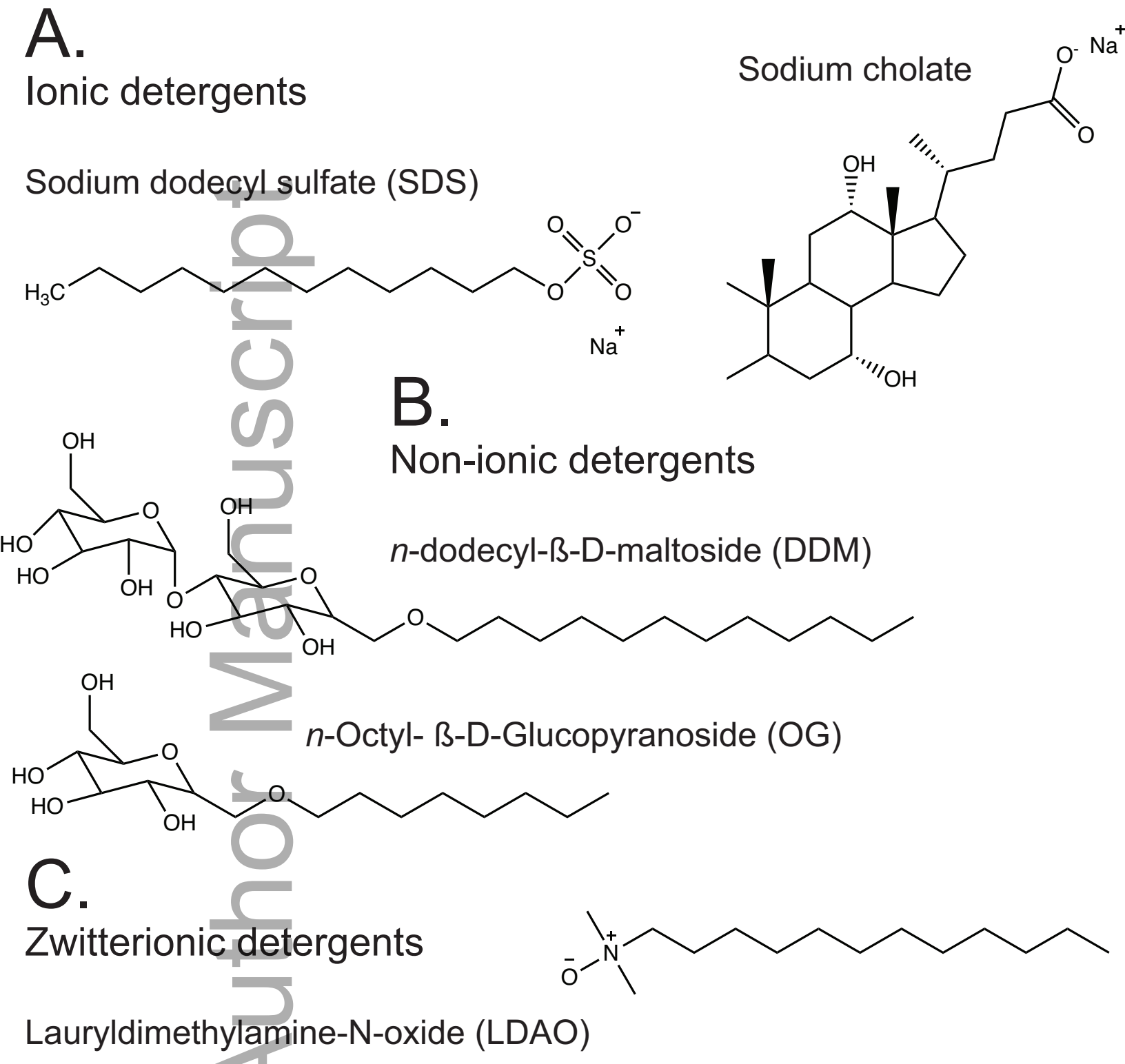
febs_15676_f1.eps



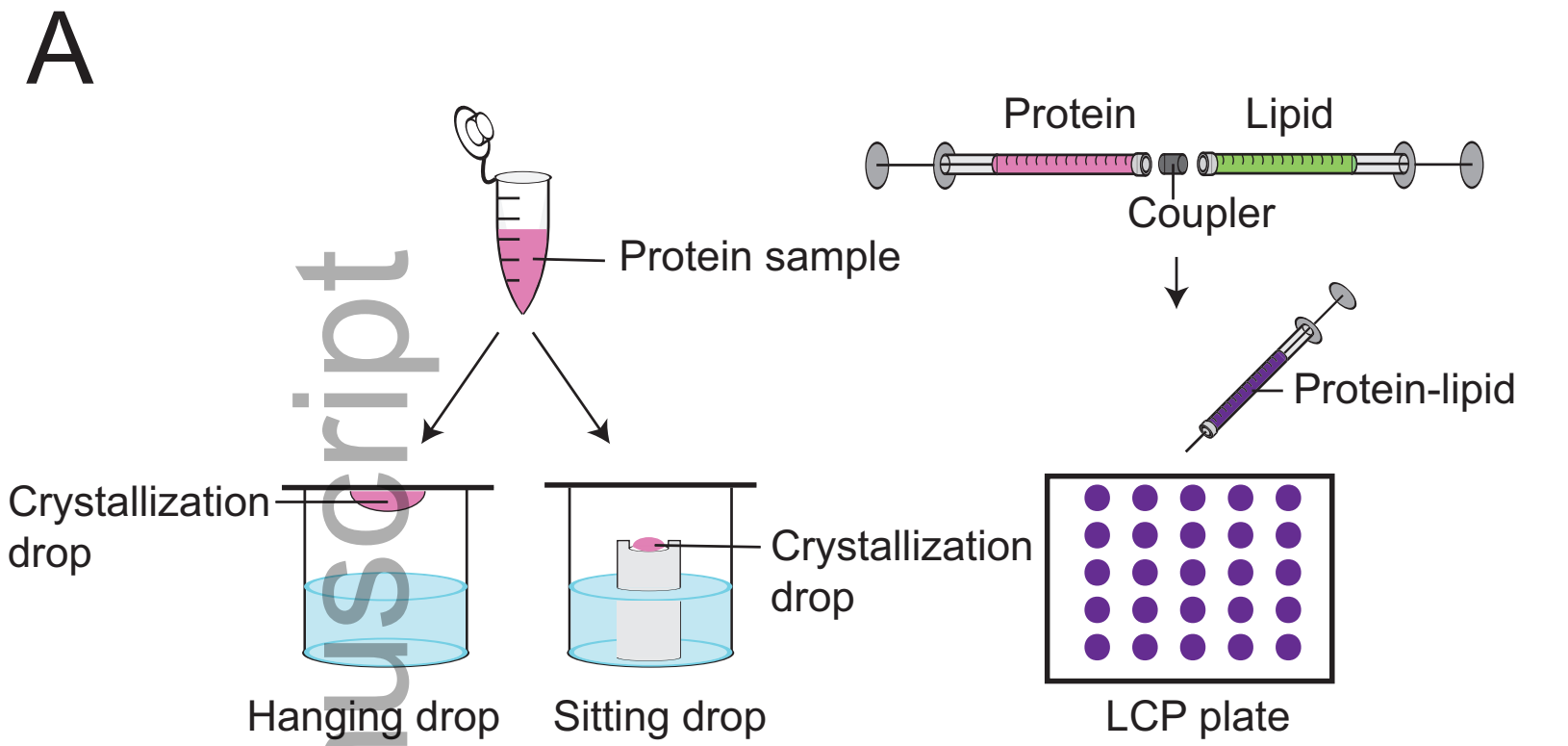
febs_15676_f2.eps

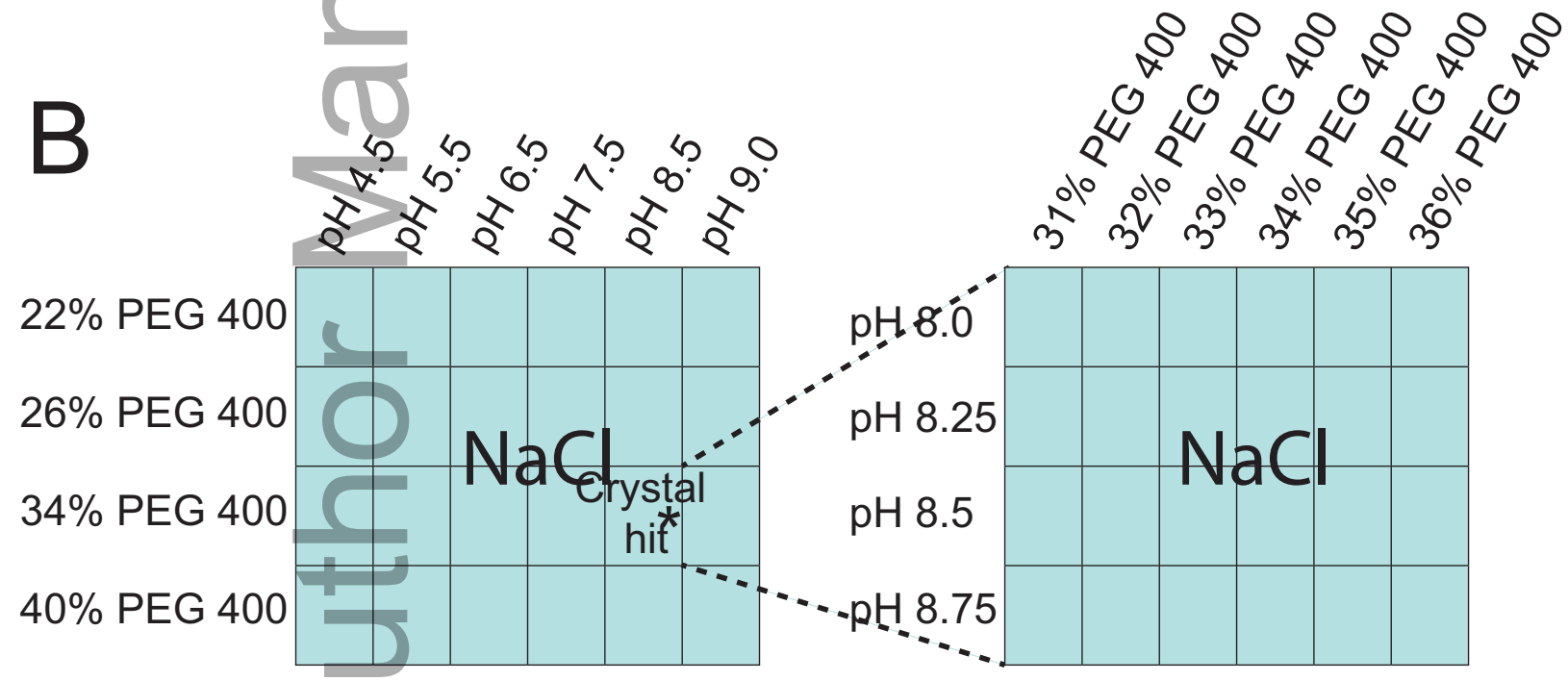


febs_15676_f3.eps

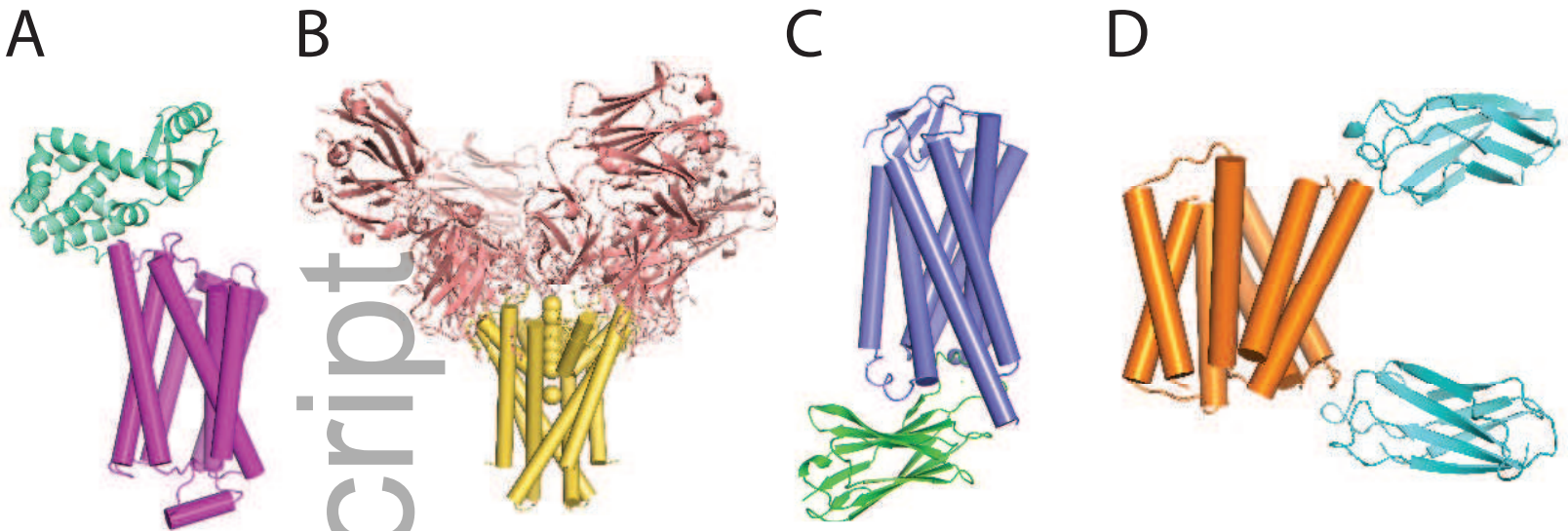


febs_15676_f4.eps

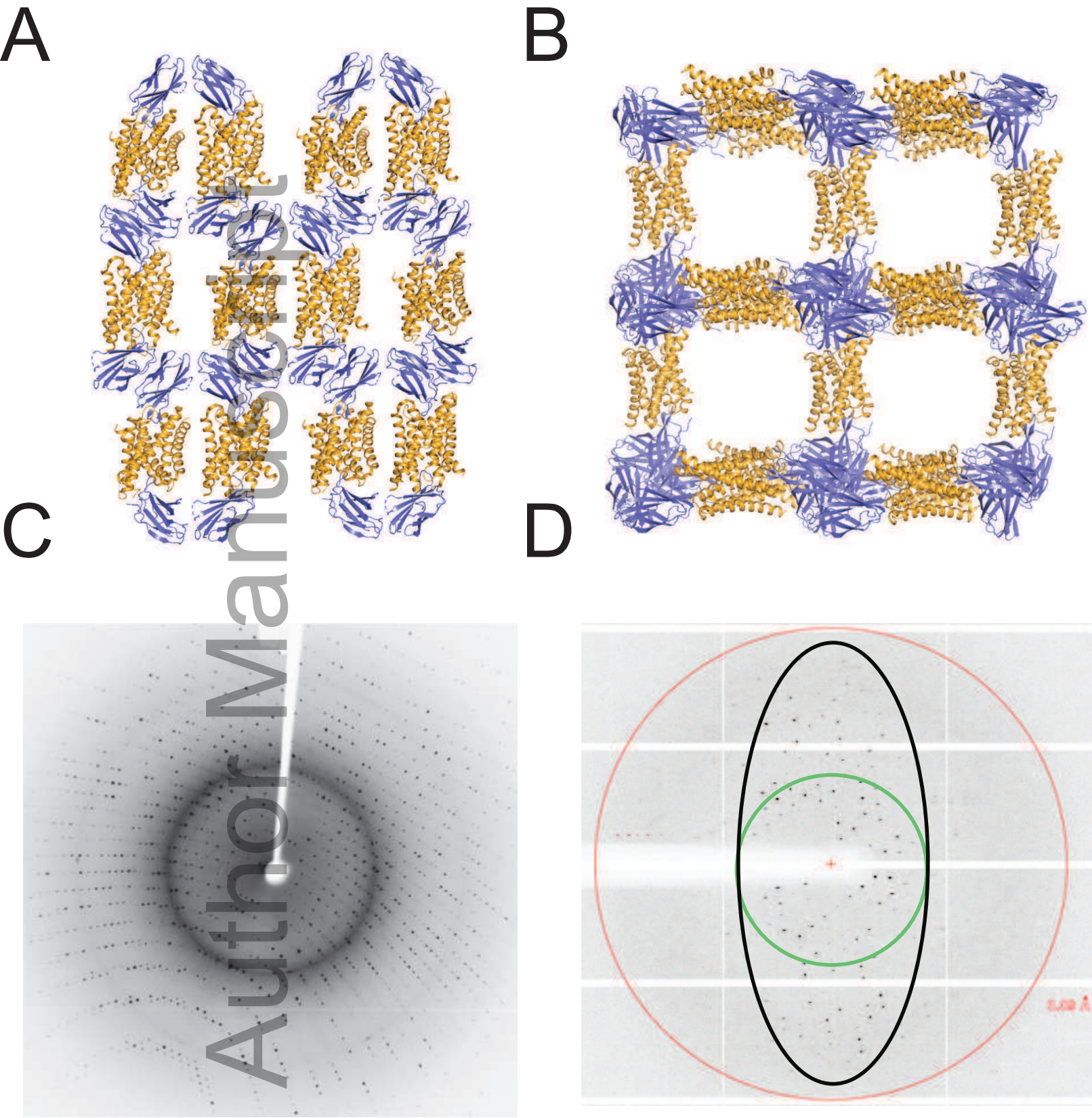




febs_15676_f5.eps



febs_15676_f6.eps



febs_15676_f7.eps