Unlocking the eukaryotic membrane protein structural proteome
John Kyongwon Lee and Robert Michael Stroud

Most of the 231 unique membrane protein structures (as of 3/2010) are of bacterial membrane proteins (MPs) expressed in bacteria, or eukaryotic MPs from natural sources. However, eukaryotic membrane proteins, especially those with more than three membrane crossings rarely succumb to any suitable expression in bacterial cells. They typically require expression in eukaryotic cells that can provide appropriate endoplasmic reticulum, chaperones, targeting and post-translational processing. In evidence, only ~20 eukaryotic MP structures have resulted from heterologous expression. This is required for a general approach to target particular human or pathogen membrane proteins of importance to human health. The first of these appeared in 2005. Our review addresses the special issues that pertain to the expression of eukaryotic and human membrane proteins, and recent advances in the tool kit for crystallization and structure determination.

Address
Department of Biochemistry & Biophysics, University of California, San Francisco, S-412C Genentech Hall, 600 16th Street, San Francisco, CA 94158-2517, United States

Corresponding author: Stroud, Robert Michael (stroud@msg.ucsf.edu)

Eukaryotic expression systems
Yeast: Saccharomyces cerevisiae and Pichia pastoris
The methanotrophic yeast Pichia pastoris [3] and the budding yeast Saccharomyces cerevisiae [4,5] are suitable for overexpression and functional analysis of eukaryotic MPs. At least seven even of the first thirteen eukaryotic MP structures expressed heterologously were produced in some form of yeast, though so far only two unique eukaryotic MP structures have been from expression in S. cerevisiae.

We designed a high-throughput S. cerevisiae pipeline that minimizes effort in uncovering high-quality proteins for crystallization [6,7,8]. A screen of 384 rationally selected eukaryotic MPs that entered this pipeline demonstrate that ~25% of yeast MPs, 10 solubilized and purified in dodecyl-β-t-o-maltoside, displayed sufficient purity and stability to enter crystallization trials. Genes are inserted into a S. cerevisiae LIC expression plasmid based on the yeast two-micrometer (2μ) plasmid. This naturally occurring extrachromosomal DNA plasmid within S. cerevisiae replicates under strict cell cycle control and serves as the backbone for most episomal methods within yeast. Cell toxicity is a common problem with the overexpression of MPs and the tight control of induction within the system is important [5,6,10].

Expression of MPs in P. pastoris benefits from the highly inducible methanol oxidase promoter. It has been used successfully for a number of eukaryotic MP crystal structures including the rat Voltage-dependent Shaker K+ channel Kv1.2 at 2.9 Å resolution [9], human aquaporin 4 at 1.8 Å resolution [10], and the yeast aquaporin at 1.15 Å resolution [11]. This system is robust, and inducible, which alleviates some problems of toxicity that might ensue from overexpression during the expansion phase.

The HEK system
Expression in HEK293S cells grown in suspension is a promising system for the expression of higher eukaryotic integral MPs. This expression method is time consuming and requires much care and attention on each individual target, however it can provide high-quality MP in the plasma membrane. The plasmid and HEK293 cell line (HEK293S GnTT) developed by Khorana is made deficient for the enzyme N-acetylglucosaminyl transferase I, thereby limiting the extent to which proteins are glycosylated [12]. This modification results in greater uniformity of MPs, which is an important feature that can play a critical role in the successful crystallization of proteins produced in these cells. In addition, these cells have been adapted to...
growth in suspension and can reach cell density of up to 10 million cells per ml of culture.

The number of atomic MP structures today derived from protein generated from HEK293 is only one (hRhCG) [13**]. However, over the past two years, we have cloned over 30 human MPs (ion channels, transporters, and GPCRs) into the pACMV-tetO inducible expression plasmid, and have proceeded to the stage of stable HEK cell lines with confirmed expression. A high volume oscillating bioreactor-based growth system (8–20 l) enables the production of biochemical amounts of a given MP under a variety of growth conditions. Milligram quantities of several of these MPs have been produced in this system. Gel filtration and ion exchange experiments indicate that the proteins are well behaved and of a size consistent with their expected monomeric or multimeric stoichiometries. Optimization of suspension growth conditions and refinement of post-affinity purification steps are required to ensure highest expression and stability of the purified material. When possible additional testing includes functional assays. For example, TRPV1 expressed in HEK cells was functionally active as a calcium channel by measuring agonist-induced calcium influx. We used this system to produce a crystal structure of human RhCG at 2.1 Å resolution [13**].

Insect cells
Baculovirus expression is often used to obtain increased protein yields over that of HEK cell systems [14,15**,16*]. This method often produces more protein per liter of culture than HEK-based systems because of their ability to be grown in higher density suspension cultures. Initial ramp up time is approximately 45 days for SF9 expression of a particular protein but then shorter time is required for subsequent growths. The cells can be grown up continuously, and then infected with virus when the cells are grown to high density. Thus a strategy used very successfully by Gouaux screens transient expression and correct insertion into the membrane in HEK cells, and then reverts to insect cells for high-level preparative samples. As a reflection of the maturity of the insect cell system, the recent structures of higher eukaryotic GPCRs [17,18], P2X [19*], ASIC [20*], AMPA glutamate receptor [21**], connexin, and aquaporin 4 have all been produced in the insect cell cultures.

Cell-free expression
*Escherichia coli* based cell-free (CF) expression systems have successfully produced up to 6 mg of MP per ml of reaction mixture in an individual continuous exchange system [22]. This system has been used for functional expression of small multidrug transporters [23], β-barrel type nucleoside transporters [22], and G-protein-coupled receptors [24]. An especially relevant advantage of CF expression is the complete control of the amino acid pool afforded by this system. This provides unique isotopic labeling possibilities for NMR [23].

Four different modes of expression have been reported for CF production of MPs. First, no additional detergent or lipid is included, and MPs are produced as a ‘soft’ precipitate, which can be readily solubilized in mild detergents [23,25]. Second, addition of certain detergents that do not interfere with the protein expression machinery allows direct insertion of MPs into detergent micelles [22,25–27]. Third, lipids are added so that MPs are directly reconstituted into lipid bilayers, lipid-detergent micelles [28,29*] or nanolipoprotein particles [30]. Fourth, NVoy™ (Expedeon), a linear carbohydrate-based polymer that facilitates soluble expression is added. These modes have been set up in a parallelized preparative scheme allowing overnight expression screening for 24–48 MPs in the four CF modes.

This technique is clearly making a large impact since the Doetsch, Choe and Riek groups have produced some five structures of human membrane proteins in 2010 using NMR. The ingenious apposition of cell-free synthesis with specific amino acids that are labeled with 15N, and others with 13C provide technology capable of rapidly determining the structure of smaller membrane proteins, typically <30 kDa per monomer.

Crystallization
A powerful strategy to determination of the structure of a particular MP is to select a single target protein and pursue its orthologs in various species. This often includes bacterial homologs that have sometimes led the way to structural understanding of the function of eukaryotic targets. For example, there are two structures for close homologs of human health-related proteins, namely P2X(4) from zebrafish [20*], and P-glycoprotein from mouse [31**] that has 87% identity to human PGP. Both were determined in 2009. The latter followed from earlier related structures determined in the same laboratory that eventually succeeded in the higher eukaryote; these two groups in general approached the eukaryotic MP, benefiting from specialized family focus. In the other example, the proton-activated Na+ conducting ASIC channels from chicken [32], is related to the ATP-gated cation conducting P2X family, and was produced by the Gouaux laboratory before their P2X(4) structure. Lessons learned from these two outstanding landmark membrane structures [20*,32] will help to enable structure determination of other family members. Our own structures of ammonia transporters went from *E. coli*, through Nitrosomonas, to human (Figure 1).

A different approach is that of protein engineering, taken by Kobilka, Stevens, Shertler and colleagues with the β-adrenergic receptor and subsequent GPCR structures, where the focus remained on the human targets alone.
This approach varied the use of ligands, insertion of bacteriophage lysozyme into flexible loops [17], mutations, Fab fragment conjugates as crystallization chaperones [18], and use of lipids as platforms for crystallization.

**Antibodies as crystallization partners**

A highly successful strategy to determine the structures of membrane proteins is use of crystallization chaperones [33,34,35,36,37,38] (Figure 2). These partners are generally Fabs prepared from monoclonal antibodies or other binding domains that have been engineered to bind specifically to a given protein target. Fab-based partners have been the enabling factor for determining a number of landmark structures by reducing conformational heterogeneity (i.e. reducing flexibility), by masking hydrophobic surfaces and increasing solubility, and by providing primary contact points between molecules in the crystal lattice. Several exciting prospects are emerging which promise advantages of in vitro selection and recombinesant reagents using bacteriophage display. One approach that yielded a structure for full-length KcsA, synthetic affinity reagents (sABs) were selected from highly functional phage display libraries in vitro [31,37] (Figure 1). Unlike animal immunization, selection can be precisely adjusted for specific requirements for each target for the intended use of the sAB. Such precise biochemical control is particularly important for detergent solubilized MPs, because their conformation is highly sensitive to solution conditions. Such sABs are readily produced in E. coli and stored as an expression vector. Amino acid sequences are determined and the use of an invariant scaffold makes it straightforward to reformat a sAB from one format to another.

This ingenious approach uses libraries based on a reduced genetic code diversity in invariant scaffolds [37,38]. Use of a reduced genetic code allows for introducing diversity into more (up to 20) sites without compromising function. The scaffold can be optimized for stability, expression and efficient crystal lattice formation and the antigen-binding interface can be optimized for maximum binding potency [37]. sABs can be used to lock a protein into a specific conformational state, which allows for selection of different functional states. Phage display selections can also be tuned to direct sAB binding to various regions on a
proteins’ surface. This capability was key to crystallizing the full-length KcsA potassium channel (Figure 2) [34**]. The MacKinnon group determined the first, landmark structure of a K+ channel using a truncated version that lacked the cytoplasmic C-terminal domain. In a strategy to visualize the 40 amino acid C-terminal domain, sABs to the C-terminal domains were generated by eliminating those that bind to truncated KcsA, with the hope of reducing the domain’s inherent flexibility. Three sABs were used for co-crystallization and structures of both the closed and open forms were determined [34**].

In a different type of approach, the Craik lab recently constructed an Fab phage display library from native human peripheral and spleen B cells. Of the scFvs identified, the best were found to have both high affinity ($K_i = 12$ pM) and specificity for the targeted antigen. This led to a structure of a soluble protein to 2.2 Å resolution [39]. Establishing this as a sound approach to chaperone-selection against membrane proteins, will be a high priority in the field.

GFP screening
GFP fusions and fluorescence detected size exclusion chromatography (FSEC) is a robust method for identifying MP constructs that are amenable to crystallization and is a means for screening appropriate detergents [40,41]. This method exploits the unique spectral signature of GFP to detect the size exclusion properties of the test protein from small culture sizes and without requiring extensive purification. When working with higher eukaryotic expression systems such as baculovirus or HEK293, this allows for a quick initial analysis of the expression level of multiple orthologs, screening of detergents for solubilization and stability, and determining the effect of different ligands on the solubilized protein. Such methodology has led to the recent solution of a number of membrane protein structures [19*,20*,32,42*]. For example, the recent structure of the rat GluA2 [21**] shows the power of this screening method as the authors were able to screen multiple orthologs of the receptor, screen for optimal detergent, and to study effects of different ligands on the protein. This enabled the authors to select the best ortholog and detergent early in the project and to follow the effects of several mutations in later stages of the project. FSEC is proving to be a powerful tool speeding up the process of identifying the best ortholog of the target protein and optimizing the conditions to get to the structure of the MP.

Mesophase crystallization
The in meso crystallization methods led to the first high-resolution structures of bacteriorhodopsin, followed by three other haloarchaeal rhodopsins. However it initially seemed to be limited to this class. Now with the advent of cubic lipidic and sponge phase methods [43*] and the structures of several GPCRs starting with the β-adrenergic receptor, the use of lipid environment for MP crystallization platforms has now been established as of broad general applicability [44**]. Resulting crystals are generally ‘type I’ membrane protein crystals in which the MPs are associated laterally in a plane, as if in a bilayer.
throughout the extent of the crystal. Current models for how crystals form by the in meso method invoke a transition between mesophases [45]. A more hydrated and open mesophase, of reduced interfacial curvature, was explored by performing crystallization in the presence of additives that swell the cubic phase. Such swollen mesophase yielded 2.45 Å resolution structure of the light-harvesting II complex (LHII). The structural details of the complex resembled those of crystals grown by the conventional vapor diffusion method, with some important differences. In particular, packing density in the in meso-grown crystals was dramatically higher, more akin to that seen with water-soluble proteins. These results present a rational case for including mesophase-swelling, so-called ‘spongifying’ additives in screens for in meso crystallogenesis.

Other landmark successes include the Vitamin B12 Transporter/Receptor, BtuB. A short-chained MAG lipid was designed that would enhance in meso crystallization. It was subsequently shown to produce diffraction quality crystals of BtuB [46], notable in that it was the first β-barrel protein to be crystallized by the in meso method. The structure at 1.95 Å differed in several important details from that of its counterpart grown by the more traditional method. Packing in in meso-grown crystals is dense and layered, consistent with the current model for crystallogenesis in lipidic mesophases. It is notable that the BtuB crystals grown by the in meso method did so in the sponge mesophase. Also a small β-barrel protein, the Adhesin, OpcA resides from the outer membrane of Neisseria meningitidis was obtained at 1.95 Å resolution using crystals grown in a lipidic mesophase [47].

Conclusion

The wisdom for each class of membrane proteins seems to permeate from individual laboratories that have spent perhaps decades pursuing structures within a particular family or class. However once the wisdom of experience in expression systems, purification schemes, and crystallizations has been explored, other orthogons become more accessible. The field has matured to the point that now we can again frame the most important questions of biology and have every opportunity of finding the solution at atomic level within a one or two year period. The new technologies include the recent outstanding results with the lipidic mesophases methods of crystallization, methods for screening expression levels of folded proteins in a membrane, use antibody chaperones, and new methods of screening crystals grown in microfluidic environments.

Acknowledgment

This research was supported by National Institute of Health Grant R01 GM24485, GM73210 and GM74929.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


5. Li M, Hays FA, Roe-Zurz Z, Vuong L, Kelly L, Ho CM, Robbins RM.


We designed a high-throughput structural genomics-oriented pipeline that seeks to minimize effort in uncovering high-quality, responsive nonredundant targets for crystallization. This ‘discovery-oriented’ pipeline sidesteps two significant bottlenecks in the IMP structure determination pipeline: expression and membrane extraction with detergent. A screen of 384 rationally selected eukaryotic IMPs in baker’s yeast Saccharomyces cerevisiae is outlined to demonstrate the results expected when applying this discovery-oriented pipeline to whole-organism membrane proteomes.


This reports a membrane protein annotation pipeline to define the integral membrane genome and associations between 21,379 proteins from 34 genomes. This pipeline was used to provide target input for a structural genomics project that successfully cloned, expressed, and purified 61 of our first 96 selected targets in yeast.


The structure of the voltage-dependent Shaker family K+ channel Kv1.2. In addition to the structure, a highlight of the report is the use of Pichia pastoris as the expression system and the use of lipid cocktail in the purification and crystallization steps that was essential for structure determination.

10. Ho JD, Yeh R, Sandstrom A, Chorny I, Harries WE, Robbins RA:


Structure of human aquaporin expressed in Pichia pastoris.

Very high resolution structure for a eukaryotic aquaporin expressed in Pichia pastoris.


A human integral membrane protein structure determined from protein produced in HEK293 cells.


A method paper describing small-scale screening protocols in the baculovirus system to evaluate the functional expression of a large number of mammalian membrane protein variants.


The structure of a human integral membrane protein receptor that brings together all the methods developed to solve the jAR.


A paper detailing the cholesterol binding site of multiple GPCRs and validating the essential role of cholesterol in stabilizing purified GPCRs for structural studies.


The crystal structure of a human/j2-adrenergic receptor-T4 lysozyme fusion protein bound to the partial inverse agonist carazolol at 2.4 A resolution. In addition to the structure, this report brings together the use of intelligently engineered T4 lysozyme fusion for GPCR work and the coming of age of the lipid-mesophase crystallization methods for integral membrane protein crystallography.


The 2.7 A resolution crystal structure of a beta(1)-adrenergic receptor in complex with the high-affinity antagonist cyanopindolol, determined using a specific antibody fragment to bind flexible regions and produce more ordered lattice contacts.


The structure of the zebrafish P2X(4) ion channel.


The structure of a functional acid-sensing ion channel in a desensitized state at 3 A resolution that identifies the location and composition of the desensitization gate, and the trigonal antiprism coordination of cesium ions bound in the extracellular vestibule.


The structure of the AMPA glutamate receptor. The 40 pages of the supplement details the huge amount of effort and tools used to determine the structure as well as describing key aspects of the function of the channel.


Article that describes the production of integral membrane proteins in a soluble form by adding detergents to the cell-free system.


A report that confirms high-level expression of functional integral membrane proteins in a cell-free system.


A report on the production of six different GPCRs in a cell-free expression system based on Escherichia coli extracts. The long chain polyoxethylene detergent Brij78 resulted in the solubilization of each GPCR at milligram amounts. Single particle analysis ligand interaction studies provide evidence of proper folding and functional activity.


A review describing the in vitro cell-free approach with special emphasis on technical aspects as well as on the functional and structural characterization of cell-free produced membrane proteins.


A report on the production of milligram amounts of functional GPCRs in cell-free expression method with the E. coli S30 extract.


Book chapter that focuses on the protocols for the cell-free production of integral membrane proteins using the Escherichia coli S30-extract. The authors describe the setup and optimization of the cell-free expression technique and provide protocols for the solubilization and reconstitution of membrane proteins directly from the cell-free produced precipitates.


A report describing liposome-assisted cell-free synthesis of functional GPCR, bacteriorhodopsin (bR(C)), in the presence of small unilamellar liposomes.


A cell-free expression method that produces high yields of integral membrane proteins without the use of detergents or refolding steps.


Description of cell-free system for the production of integral membrane proteins using nanolipoprotein particles, which are lipid layers confined within a ring of amphipathic protein of defined diameter.


The structure of mouse P-glycoprotein produced in Pichia pastoris, P-glycoprotein (P-gp) detoxifies cells by exporting hundreds of chemically unrelated toxins, but its role in multidrug resistance (MDR) in the treatment of cancers makes this structure huge in its potential to impact in human health.

The first structure of an acid-sensing ion channels (ASICs). One of the structures solved by the Gouaux lab that utilized the FSEC screening protocol.


The first use of Fv fragments as a crystallization chaperone for membrane proteins.


Authors outline the application of limited alphabet two chain Fv regions to crystallization of membrane proteins, achieving the first structure of full-length KcsA potassium channel tetramer. The crystal contacts are made between the antibody components leaving the transmembrane regions-free of contacts.


Structure of the K+ channel solved with the aid of Fab chaperone.


Early review on the use of Fv or Fab chaperones to crystallize membrane proteins.


A description of the method for generating high affinity, highly selective single domain antibody.


A description of the development of the synthetic antibody system.


Structure of a Fab-protease complex that confirms the viability of this Fab library as a chaperone platform for membrane protein crystallization.


Methods article that discusses the use of membrane protein-GFP fusions which allows direct monitoring and visualization of membrane proteins of interest at any stage during overexpression, solubilization and purification.


Description of a cost-effective fluorescent-based high-throughput approach for rapidly screening membrane proteins that can be over-produced to levels of >1 mg/l in Saccharomyces cerevisiae.


The first structure of an acid-sensing ion channels (ASICs). One of the Gouaux articles, the extensive 31-page supplement is full of important data.


Numerous eukaryotic membrane proteins require specific lipids for their stability and activity, and efforts to crystallize and solve the structures of membrane proteins that do not address the issue of lipids frequently end in failure rather than success. To help address this problem, we have developed a sparse matrix crystallization screen consisting of 48 lipidic-sponge phase conditions.


An alternative in meso approach, which employs a bicontinuous lipidic mesophase, has emerged as a method with considerable promise in part because it involves reconstitution of the solubilized protein back into a stabilizing and organizing lipid bilayer reservoir as a prelude to crystallization. A hypothesis for how the method works at the molecular level and experimental evidence in support of the proposal are reviewed here. The latest advances, successes, and challenges associated with the method are described.


Protocols for in meso crystallization.


The first beta sheet structure of a membrane protein produced by cubic lipidic phase crystallization.


Another successful beta sheet membrane protein structure solved using the in meso crystallization method.