Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Advances and challenges of membrane–protein complex production

Sarah Zorman[1,2], Mathieu Botte[1,2], Qiyang Jiang[1,2], Ian Collinson[3] and Christiane Schaffitzel[1,2,3]

Most membrane–proteins exist in complexes rather than isolated entities. To fully understand their biological function it is essential to study the intact membrane–protein assemblies. The overexpression and purification of many essential membrane–protein complexes is still a considerable and often unsurmountable challenge. In these cases, extraction from source is the only option for many large multi-subunit cellular machines. Here, we describe recent advances in overexpression of multi-subunit membrane–protein complexes, the strategies to stabilize these complexes and highlight major achievements in membrane–protein structural research that were facilitated by the prospect of achieving subnanometer to near-atomic resolution by electron cryo-microscopy.

Addresses
[1] European Molecular Biology Laboratory, Grenoble Outstation, Grenoble 38042, France
[2] Unit for Virus Host-Cell Interactions, University Grenoble Alpes-EMBL-CNRS, Grenoble 38042 France
[3] School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom

Corresponding authors: Collinson, Ian (Ian.Collinson@bristol.ac.uk) and Schaffitzel, Christiane (schaffitzel@embl.fr)

Current Opinion in Structural Biology 2015, **32**:123–130

This review comes from a themed issue on **New constructs and expressions of proteins**

Edited by **Imre Berger** and **Roslyn M Bill**

http://dx.doi.org/10.1016/j.sbi.2015.03.010

Introduction

About one third of the proteome of every cell is translocated into a membrane. Recent developments in high-throughput methods to detect and quantify protein–protein interactions described the interactome of soluble proteins in bacteria, yeast, and human [1–3]. However, complexes involving membrane–proteins are much more difficult to identify, produce and characterize due to their hydrophobic nature (Figure 1a). A genome-wide approach revealed the membrane–protein interaction landscape of *Saccharomyces cerevisiae* using affinity-purification in the presence of different non-denaturing detergents followed by mass spectrometry [4]. Remarkably, in the corresponding interactome almost two-thirds of the interactions relate hitherto unassigned functions. Evidently, yeast membrane–proteins have on average 2.1 interaction partners, around half of those prescribed to globular proteins [4,5], demonstrating that, as in the cytosol, most proteins occur and function in complexes rather than as isolated entities in the membrane.

While systems biology approaches help to catalogue the membrane content and membrane–protein interactions, significant progress is needed for the determination of their stoichiometry, structure and cellular function. Here, we review recent approaches to produce, purify and stabilize membrane–protein complexes for these analyses.

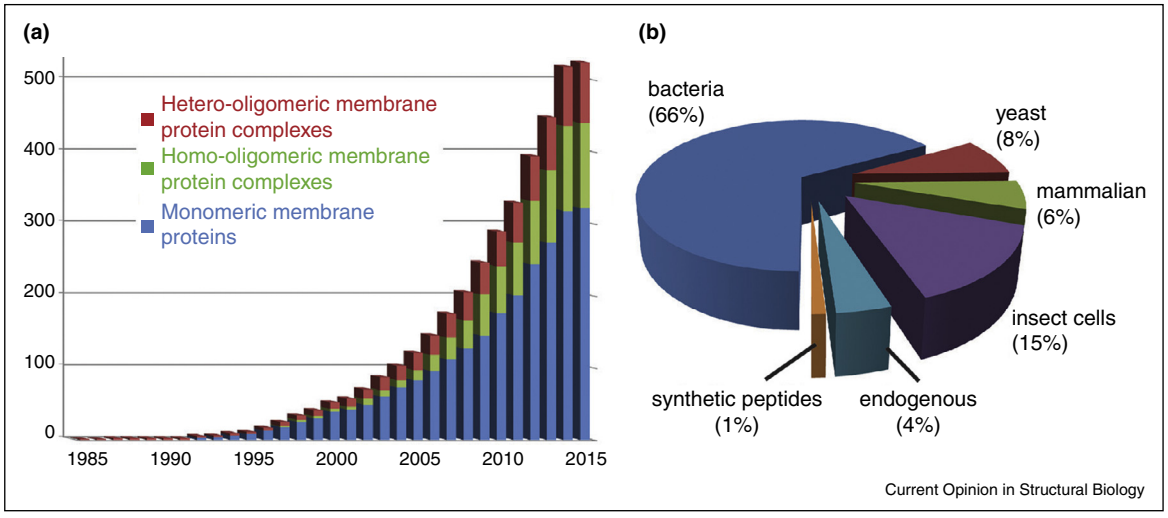
Recombinant membrane–protein complex production in prokaryotes

Escherichia coli has traditionally been used as expression host of choice for membrane–proteins, alongside Gram-positive alternatives [6] (Figure 1b). Recombinant production allows introducing truncations, mutations and tags for affinity-purification. In fact, the first crystal structure of a membrane–protein complex was the *E. coli* fumarate reductase respiratory complex [7,8].

Improved *E. coli* strains for membrane–protein expression

Significant effort has been invested to develop bacterial strains that are tailor-made for membrane–protein production. The T7 RNA polymerase-based expression hosts C41λ(DE3) and C43λ(DE3) have mutations in the *lac*UV5 promoter that governs T7 RNA polymerase expression, resulting in lower amounts of T7 RNA polymerase and consequently in slower transcription and translation rates of the proteins under the control of a T7 promoter [9,10]. Applying a similar strategy, the Lemo21(DE3) strain was designed to precisely control T7 RNA polymerase activity: expression levels of its natural inhibitor T7 lysozyme are titrated by rhamnose induction [10]. Slower membrane–protein production rates can reduce accumulation of misfolded, aggregated proteins in inclusion bodies. The C41λ(DE3) and C43λ(DE3) expression hosts are frequently used; successful examples include the AcrAB-TolC multidrug efflux pump [11•] and the LptD–LptE complex involved

Figure 1



Membrane–protein structures in the Protein Data Base and expression hosts used. **(a)** The number of membrane–proteins deposited in the Protein Data Bank (PDB, www.rcsb.org) is exponentially increasing since 1985, but still presents a small fraction compared to all the proteins in the PDB (currently ~98,770). The majority of the membrane–protein structures are monomeric or from homo-oligomers. Multi-subunit membrane–protein complexes are particularly challenging, and thus only 84 structures are currently available. **(b)** Expression hosts used for membrane–protein production for structures deposited in the PDB between 2010 and 2015. Membrane–proteins are traditionally expressed in bacteria (66%), the vast majority in *Escherichia coli*. However, eukaryotic expression systems are clearly becoming more important; in particular baculovirus-insect cell expression (15%) is successfully used for production of G protein-coupled receptors.

in lipopolysaccharide translocation into the outer membrane of Gram-negative bacteria [12] (depicted in Figure 3b). The twin-arginine translocase inner-membrane subunit (TatC) is one of many examples for successful membrane–protein expression in a Lemo(DE3) strain [13]. Additionally, auto-induction-based media which slow down protein expression are now commonly used to improve membrane–protein production [14].

Co-expression systems for protein complexes

Membrane–protein complexes usually cannot be reconstituted from the purified subunits or subcomplexes. This could be due to the fact that they need interaction partners for stable folding and function or because the protein–protein interactions formed by interfaces within the bilayer are masked by the detergent used for solubilisation. Co-expression of the complex subunits overcomes this problem. Co-transformation of several plasmids with single or multiple expression cassettes and with different origins of replication and antibiotic resistance [15,16] has the drawback that the copy numbers of the respective plasmids can vary significantly. This complicates the balanced expression of the subunits and adjusting the stoichiometry of the complex. Poly-cistronic expression systems using natural or artificial operons where the complex components are expressed from one messenger RNA often result in a more balanced protein production [17]. In the case of artificial operons the order of genes in the construct needs to be experimentally determined because expression levels depend on intrinsic properties of the particular coding sequences [18]. Moreover, larger operons with many genes often lead to low expression levels of the proteins encoded downstream. The trimeric Sec translocon complex was expressed from a synthetic poly-cistronic mRNA, purified and successfully crystallized with and without the partner ATPase SecA [19,20]. The approach is not restricted to *E. coli*; the Na^+-pumping NADH:quinone oxidoreductase complex was produced by homologous expression of the operon encoding its six subunits in *Vibrio cholerae* [21•]. The *E. coli* AcrABZ-TolC multidrug efflux pump recently has been elegantly produced in C43(DE3) using co-expression from a vector encoding an AcrA–AcrB fusion and a pETDUET vector [16] encoding an AcrA–AcrZ fusion-protein and TolC [11•]. By using two different AcrA fusion-proteins and thus providing two AcrA copies, the authors stabilized the AcrA:AcrB:AcrZ:TolC complex for characterization by cryo-EM and favoured the formation of a complex with a 6:3:3:3 stoichiometry [11•].

The ACEMBL system for multi-protein expression in *E. coli* uses small, designed acceptor and donor vectors [22]. Individual genes or poly-cistrons are inserted into the multiple-integration element of the acceptor and into one or several donor vectors with different antibiotic resistance. Donor vectors have a conditional origin of replication (oriR6Kγ in Figure 2a) and need to be propagated in strains encoding the phage R6Kγ *pir* gene. Incubation of