Mechanisms driving membrane protein targeting and insertion in a synthetic minimal cell

Ross Eaglesfield, BSc, MSc

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Molecular, Cell and Systems Biology

College of Medical, Veterinary and Life Sciences

University of Glasgow

September 2019

Abstract

The generation of an amphiphilic, semi-permeable membrane was one of the key events in the evolution of cellular life. This event must have been linked to the emergence of integral membrane proteins providing cells with the capability to control the exchange of materials with the environment. Modern cells have evolved complex mechanisms to recruit, insert and assemble these highly hydrophobic proteins into the membrane. Given that many of the proteins involved with these mechanisms are themselves membrane-integral it seems likely that early in evolutionary history cells used more fundamental processes for the insertion of membrane proteins. Identifying such processes is essential for building up an understanding of the more complex mechanisms operating in modern cells and for building artificial cells, both of which will contribute to the advance of medicine and biotechnology.

This thesis has examined the fundamental characteristics responsible for the targeting and insertion of two polytopic α-helical membrane proteins, proteorhodopsin and galactose permease, in a minimal cell-mimicking system. A recombinant cell-free protein synthesis system was used in conjunction with giant unilamellar lipid vesicles (GUVs) to analyse the innate membrane targeting and insertion capability of both proteins in the absence of other targeting and insertion pathways. The results obtained show that both proteins are able to localise and insert co-translationally into the vesicle membrane despite the large aqueous volume of the GUV lumen. Removal of the N-terminal hydrophobic regions of both proteins led to mislocalisation and increased aqueous aggregation, highlighting the importance of effective recruitment of translating ribosomes to the membrane and of the N-terminal regions in this process, even in the absence of other chaperoning mechanisms. Follow-up experiments using truncated proteins, as well as stalled and bound ribosomes, unravelled the functions of the N-terminal domains in ribosome recruitment and insertion.

Taken together, the data obtained supports a model in which the first hydrophobic portion of the protein to emerge from the ribosome effectively recruits the translating ribosome nascent chain complex to the membrane, followed by spontaneous membrane insertion of the entire protein. This protein-inherent mechanism represents a simple, fundamental mode of membrane

protein biogenesis, which is likely to be masked by the existence of more sophisticated chaperoning and translocation pathways in the complex sub-cellular context of modern cells. The technology developed in this thesis offers new opportunities to de-construct *in vivo* pathways and to build artificial cells from the bottom up.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to start by saying a huge thank you to my supervisors Prof. Anna Amtmann and Dr. Julien Reboud. Special thanks to Prof. Anna Amtmann, your support and guidance have been invaluable throughout this project. I will be forever grateful for the independence you allowed me throughout my time, and for the encouragement to pursue my ideas (even some of the crazier ones). You created a wonderful research environment which helped me to become a better scientist.

I would also like to thank all the members of the Stevenson lab, past and present, including - Maria Papanatsiou, Zaigham Shahzad, Alex Mihnev, Jordan Twigg, Mary Ann Madsen, Emily May Armstrong, Naomi Donald, George Boswell and Craig Carr for all the help in the lab and for making it a wonderful place to work.

Special thanks also to David Paterson and Hang Zhao from Engineering for showing me the ropes and getting me excited about working with ‘droplets’.

To all of my family and friends who have always supported me. Especially my parents who have always accepted my decisions and encouraged me to pursue my ambitions, even when some of them were probably questionable.

A big thank you to the University of Glasgow and the Lord Kelvin Adam Smith PhD scholarship for funding this challenging and exciting project and for giving me the opportunity to carry out this work. It has been a trully rewarding experience.

And finally, but by no means least, my wife Zofia. Without your constant support I would never have managed to get this far, you’ve always encouraged me to do what I love and have always been there to listen to my troubles as I try to be for you.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Ross Eaglesfield

Abbreviations

µ	Micro
Ala/A	Alanine
ATP	Adenosine Triphosphate
bp	Base Pair
CL	Cardiolipin
CTP	Cytidine Triphosphate
CX	Connexin
DGS-NTA(Ni)	1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt)
DHFBI	(5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene] -3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPG	1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EGFP	Enhanced Green Fluorescent Protein
EmrE	Multidrug Transporter Protein
ER	Endoplasmic Reticulum
Ffh	Fifty Four Homologue
g	Gram
GalP	Galactose Permease
gDNA	Genomic DNA

Gly/G	Glycine
GPCR	G-Protein Coupled Receptor
GTP	Guanosine Triphosphate
GUV	Giant Unilamellar Vesicle
HA	Hemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic Acid
His/H	Histidine
IMV	Inverted Membrane Vesicle
kbp	Kilo Base Pairs
kDa	Kilo Dalton
kHz	Kilohertz
L	Litre
LB	Luria Broth
Leu/L	Leucine
m	Milli or Metre
M	Molar
Ma	Million Years Ago
ml	Millilitre
MLV	Multilamellar Vesicle
mm	Millimetre
mRNA	Messenger RNA
MW	Molecular Weight
MWCO	Molecular Weight Cut-off
n	Nano
NTA	Nitrilotriacetic Acid

PAGE	Polyacrylamide Gel Electrophoresis
PC	Phosphocholine
PCR	Polymerase Chain Reaction
pDNA	Plasmid DNA
PI	Propidium Iodide
POPC	1-palmitoyl-2-oleoyl-glycero-3-phosphocholine
PR	Proteorhodopsin
PURE	Protein Synthesis Using Recombinant Elements
RNA	Ribonucleic Acid
RNase	Ribonuclease
RNC	Ribosome Nascent Chain Complex
rpm	Revolutions per Minute
rRNA	Ribosomal RNA
s	Seconds
SDS	Sodium Dodecyl Sulphate
SEIRAS	Surface-enhanced Infrared Absorption Spectroscopy
SR	SRP Receptor
SRP	Signal Recognition Particle
SUV	Small Unilamellar Vesicle
TCEP	Tris(2-carboxyethyl)phosphine Hydrochloride
Thr/T	Threonine
TM	Transmembrane Domain/Helix
tRNA	Transfer RNA
UTP	Uridine Triphosphate
V	Volts

1. Introduction

1.1 Origins of cellular life

How cellular life first emerged and expanded on the Earth remains one of the biggest mysteries in biology. Unicellular microorganisms have been identified within fossilised rock samples dating from around 3,500 million years ago (Ma) (Schopf et al., 2018). There is also non-fossil evidence of even older microorganisms identified from sea bed sedimentary deposits dating from, potentially, at least 3,800 Ma (Dodd et al., 2017). This is remarkable given that the earth had only cooled to allow the formation of bodies of liquid water around 3,800-4,280 Ma (O'Neil et al., 2012), indicating that the emergence of life in these primordial oceans was rapid, at least in geological terms. The oldest fossilised microorganisms bear a striking resemblance to modern day filamentous prokaryotes (Schopf, 1993; Schopf et al., 2018), suggesting that by this point in evolutionary time they were already relatively complex and contained many of the structural and metabolic features of modern day cells. Being able to look beyond this 3,500 Ma time-point using the fossil record is unlikely given the metamorphic nature of more ancient rocks (Bernard and Papineau, 2014).

The lack of evidence relating to the emergence of life has led to a multitude of theories on its origins, such as the ‘RNA world’ and ‘lipid world’ hypotheses (Copley et al., 2007; Segre et al., 2001). One thing however remains clear, that at some point in evolutionary history membrane-bound protocells emerged that were able to self-replicate, produce energy and exchange material with the environment (Morowitz et al., 1988). Controlled solute exchange with the environment is intimately linked with energy production in modern cells and is catalysed primarily by polytopic α-helical membrane proteins. The plethora of essential functions mediated by these proteins, as well as their conservation throughout all cellular life, suggests that their origins lie deep in evolutionary history, and their emergence may even have been a fundamental requirement for the expansion of life (Lane and Martin, 2012).

1.2 Bottom-up artificial cells

Even the simplest modern cells are incredibly complex and contain a myriad of proteins of unknown function (Price et al., 2018). Hence, in order to understand the fundamental requirements of cellular life, synthetic biology researchers have begun attempts to create a simplified artificial cell by reducing the inherent biological complexity of modern cells (Xu et al., 2016). Two main approaches have been employed in these attempts, the top-down and bottom-up approaches (Figure 1.1).

1.2.1 Top-down and bottom-up approaches for reducing biological complexity

The top-down approach takes an existing organism and strips away the genome to the minimal number of genes required for maintenance of the characteristics of life (Jewett and Forster, 2010). This idea has been used to design and synthesise a number of minimal genomes that, following transplantation into a chassis organism that had been stripped of its own genome, resulted in viable cells (Hutchison et al., 2016; Richardson et al., 2017). It should be noted that these genomes still contain many genes of unknown function and that the recipient cells still contained their complex cytoplasmic and membrane environments, thus facilitating the synthesis of these synthetic genomes. It is expected that future efforts will be made to link 'cell-free' protein synthesis systems with the synthesis of a minimal genome as has been achieved for the T7 and ΦX174 phage genomes (Shin et al., 2012; Jia and Schwille, 2019). 'Cell-free' approaches refer to *in vitro* techniques that utilise specific cellular components to examine biological pathways and reactions in a controlled environment lacking the complexity of the parent cell. For example, the use of cell extracts to synthesise high yields of proteins that would be impossible *in vivo*, or the use of purified enzymatic pathways for high yield product synthesis (Hodgman and Jewett, 2012). Cell-free technology has been extensively used for investigations of membrane protein folding and assembly and will be discussed in detail in section 1.5.

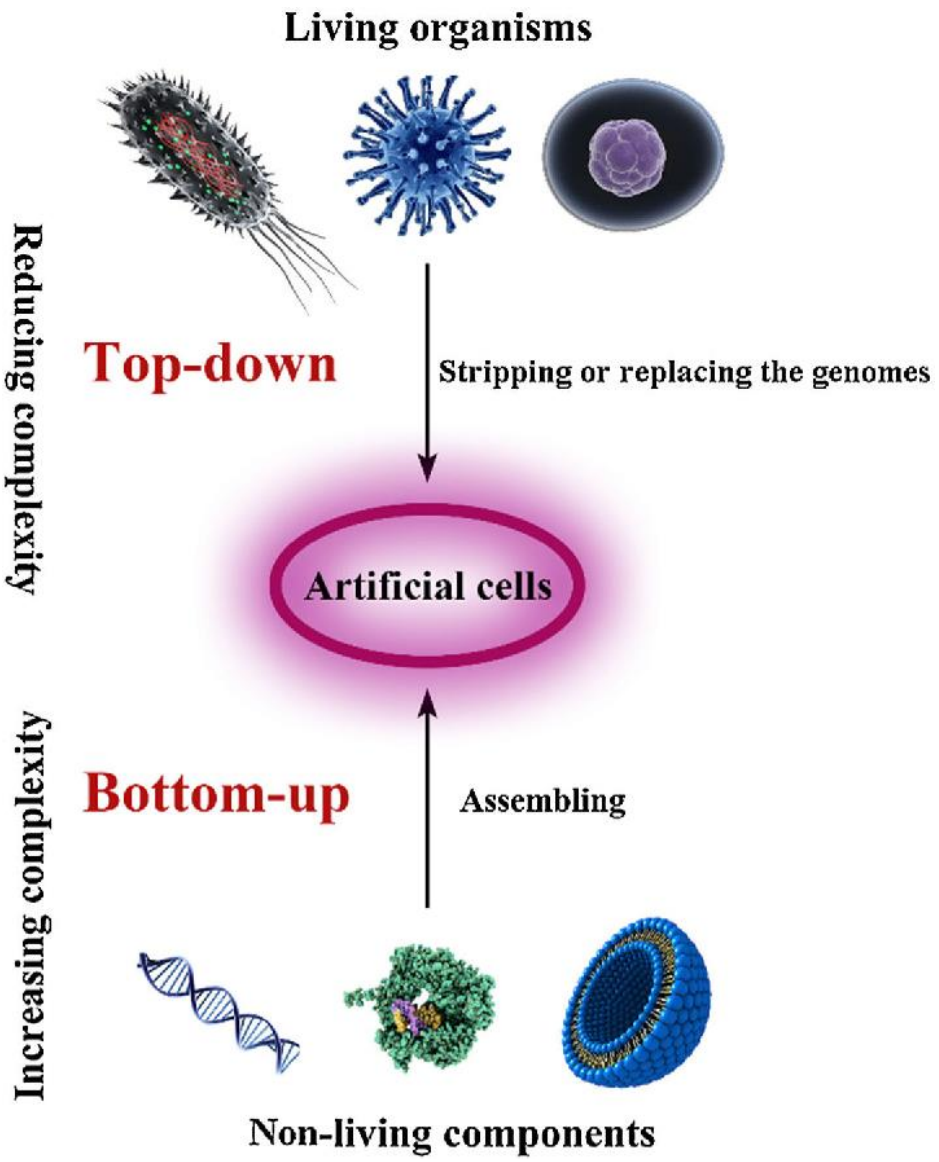


Figure 1.1 The two approaches used in attempts to generate artificial cells.

The top-down approach utilises an existing organism and reduces its existing complexity through genetic manipulations. The aim of the top-down approach is to generate artificial cells with the minimal genetic requirements. The bottom-up approach aims to generate artificial cells through systematic assembly of individual biological components and thus increasing the complexity of the system. The aims of the bottom-up strategy are currently to assemble protocells mimicking individual properties of life such as energy production. Figure adapted from Xu et al. (2016).

The bottom-up approach on the other hand aims to reconstruct a living cell from individual biological and chemical components. This is a daunting task given the large number of required components and achieving a functional cell has so far remained elusive. The main effort currently goes into reconstructing individual 'modules' of life, for example membrane synthesis and division or energy production. Membrane proteins are essential to these processes and their reconstitution into vesicular systems such as liposomes and polymersomes has become an important technology for enabling this effort.

1.2.2 Energy generation in artificial cells

The reconstruction of energy generation pathways has been the focus of a number of studies due to the essential requirement for energy gradients and energy storage molecules in cells. The majority of work has focussed on the reconstitution of active transport proteins into synthetic membrane environments for the generation of energy in the form of ionic gradients as well as the synthesis of the major biological energy storage molecule adenosine triphosphate (ATP).

Key studies have achieved the reconstitution of purified photosynthetic reaction centres for the generation of proton gradients across liposome and polymersome membranes as a stepping-stone to the reconstruction of more complex electron transport pathways (Milano et al., 2012; Tangorra et al., 2015; Altamura et al., 2017). Similarly, the bacteriorhodopsin (BR) protein has been reconstituted alongside ATP-synthase for the light-driven synthesis of ATP in both polymer and lipid vesicles (Steinberg-Yfrach et al., 1998; Choi and Montemagno, 2005; Berhanu et al., 2019). These studies have all relied on detergent-mediated reconstitution techniques, which use *in vivo* produced and assembled membrane proteins, purified and solubilised in detergents prior to their integration into vesicles. This limits the complexity of these synthetic systems due to the varying detergent compatibilities of different membrane proteins (Seddon et al., 2004). Berhanu et al. (2019) partially circumvented this problem through the cell-free synthesis and assembly of BR directly into ATP-synthase containing liposomes, although the ATP-synthase itself had been reconstituted using a classical detergent-based strategy. Cell-free systems for protein synthesis will likely be

key to future developments in this area but a better understanding of the requirements for direct membrane protein insertion will be required to fulfil the potential of these technologies.

1.2.3 Microbial rhodopsins

The microbial rhodopsins are an interesting family of proteins found in a large proportion of prokaryotic and eukaryotic microorganisms. They are closely related to the rhodopsins of higher eukaryotes and G-protein coupled receptors (GPCRs) (Palczewski, 2006). The proteins in this family share a remarkably similar tertiary structure containing a bundle of seven transmembrane α-helices with an extracellular N-terminus and an intracellular C-terminus (Figure 1.2 A). Microbial rhodopsins are activated by light and function as ion transporters and channels, as well as in signal transduction. They contain a retinal chromophore, responsible for their light responsiveness, that is covalently bound via a Schiff base to the ε-amino group of a conserved lysine residue within the seventh transmembrane helix (TM7). This linkage is in the vast majority of cases protonated, and alterations to this protonation state are vital to transport and signalling functions (Ernst et al., 2014) (Figure 1.2 B).

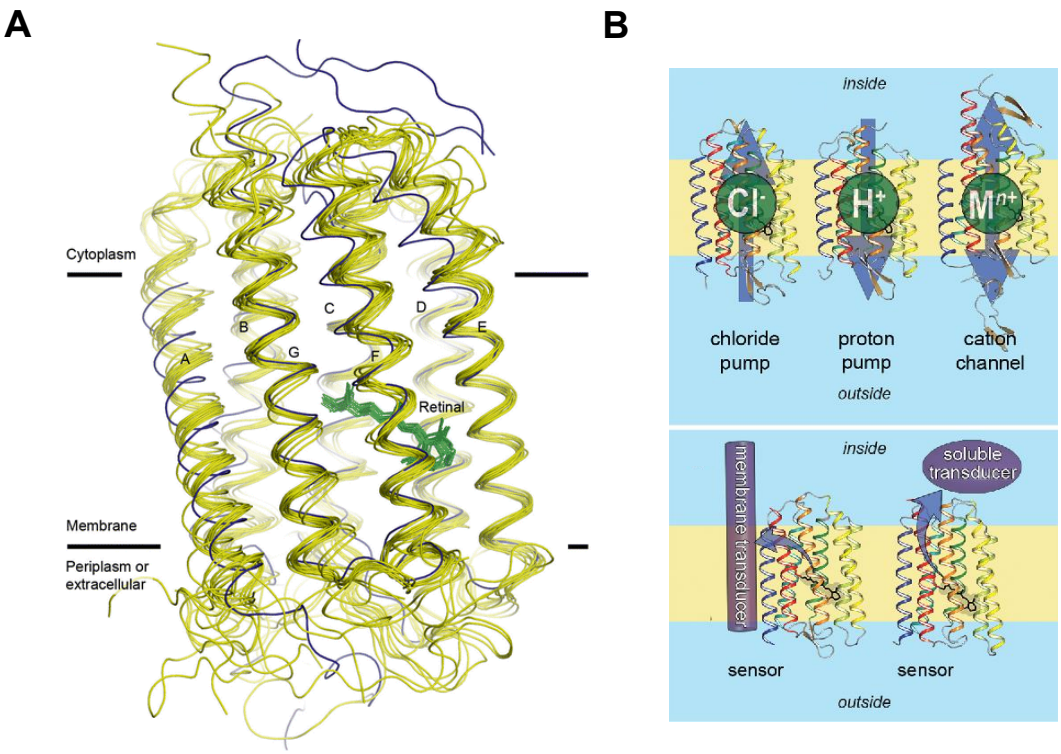


Figure 1.2 Structural homology and functional diversity of microbial rhodopsins.

(A) Known structures of microbial rhodopsins are superimposed to show structural homology. Channelrhodopsin structure is shown in blue as it represents the most divergent structure. The majority of structural differences are located in the exposed loop regions with the transmembrane helices being relatively stable. **(B)** Microbial rhodopsins are known to function as chloride and proton pumps, cation channels and photosensors. Figure adapted from Gushchin and Gordeliy (2018) and Ernst et al. (2014).

The first identified and most widely studied microbial rhodopsin is bacteriorhodopsin (BR) from the archaeon *Halobacterium salinarum* (Oesterhelt and Stoeckenius, 1973). BR is an outward directed, light-activated proton pump that has been used as a model for membrane protein folding studies and as a building block for the generation of light-induced ionic gradients in artificial cells (Yu et al., 2017; Berhanu et al., 2019).

Proteorhodopsin (PR) was the first example of a microbial rhodopsin discovered in the eubacterial domain (Beja et al., 2000). It is similar to BR in that it functions as a light-activated proton pump, but it is more amenable to recombinant expression in *Escherichia coli* making it of interest for biotechnological applications and artificial cell studies (Walter et al., 2007). The bacterial origin of PR means it is also of interest as a model protein for

understanding the targeting and insertion of bacterial inner membrane α-helical proteins. PR contains a short, cleavable N-terminal hydrophobic domain, which is an unusual feature for an inner membrane α-helical protein. Recent evidence has shown that the presence of this hydrophobic domain is not required for protein function when overexpressed in *E. coli,* however yield was greatly enhanced when it was present. The same study showed that post-translational cleavage of the signal peptide sequence was not required for protein function, indicating that it did not interfere with the adoption of tertiary structure (Soto-Rodriguez and Baneyx, 2019).

1.2.4 The major facilitator superfamily of transporters

The major facilitator superfamily (MFS) comprises a large group of membrane transport proteins found ubiquitously throughout bacteria, archaea and eukaryotes (Pao et al., 1998). They are known to function as uniporters, symporters and antiporters for a wide array of molecules such as sugars, nucleotides, phosphates and various drugs (Henderson and Maiden, 1990; Marger and Saier, 1993) (Figure 1.3 A). The vast majority of MFS members are 12-transmembrane α-helical proteins whose structures are organised in two repeating units of six helices known as the N-domain, encompassing the six N-terminal helices, and the C-domain, containing the six C-terminal helices (Figure 1.3 B). Structural studies have proposed an alternating access model to explain transport function, where the cytoplasmic and extracellular regions can be reversibly switched between an open and closed state dependent on substrate binding and, in some cases, an ionic energy gradient (Quistgaard et al., 2013; Kumar et al., 2014) (Figure 1.3 C).

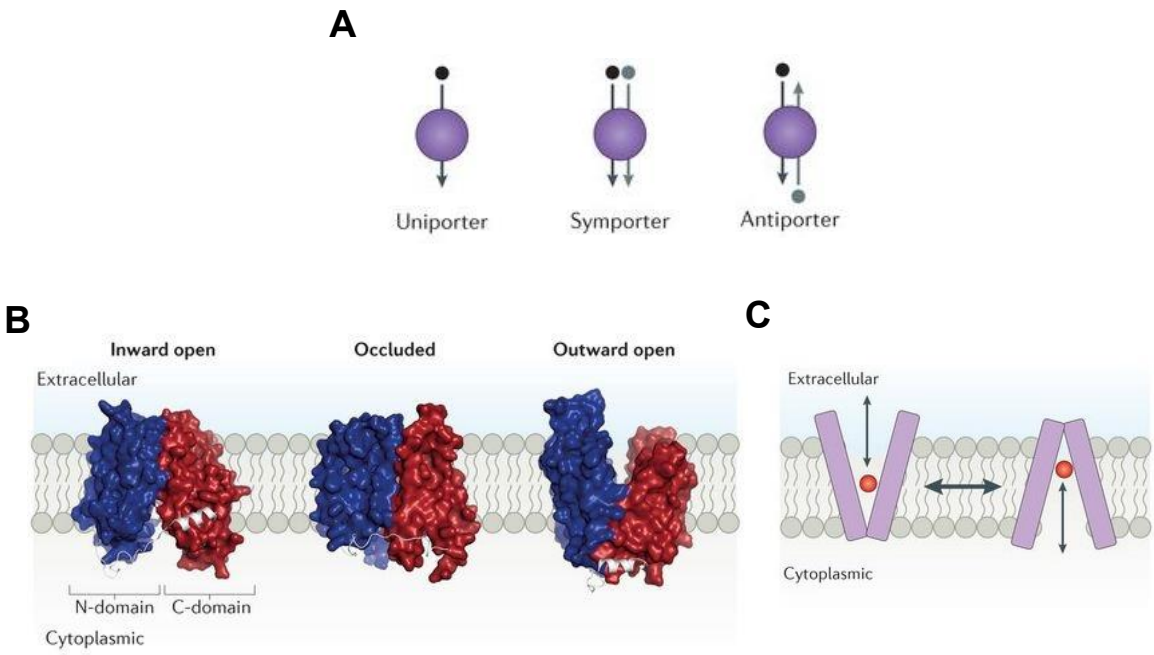


Figure 1.3 Transport principles of the major facilitator superfamily (MFS) of membrane proteins.

(A) Schematic representation of a uniporter which transports its substrate alone, a symporter which transports its substrate along with a coupled ion, and an antiporter which transports its substrate in one direction and a coupling ion in the opposite direction **(B)** Structural representation of an MFS transporter (hsGLUT3, PDB entry 4ZW9) in open and closed states with the two domains containing six helices each labelled (N-domain and C-domain) (Deng et al., 2015) **(C)** A cartoon representation of the alternative access hypothesis for substrate transport by MFS members. Model is based on the structures observed in (B). Rearrangement of the two domains is able to shield the substrate following binding and allow it to be released on the opposite side of the membrane. Figure adapted from Quistgaard et al. (2016).

The bacterial MFS members lactose permease (LacY) and galactose permease (GalP) have been used as models for membrane protein folding studies and LacY has been reconstituted *de novo* in artificial cells using cell-free methods (He and Kaback, 1998; Findlay et al., 2010; Findlay et al., 2016). The high affinity of GalP for glucose makes it an interesting candidate for generating artificial cells containing a simple glucose-based metabolism without the extra enzymatic step required to split lactose (Sahin-Toth et al., 2001).

1.3 Membrane protein biogenesis *in vivo*

The insertion and folding of polytopic membrane proteins is a highly controlled and efficient process *in vivo* that utilises a number of highly conserved proteins. A membrane embedded protein complex known as the Sec-translocon and a translocase of the YidC/Oxa1/Alb3 family mediate the majority of membrane protein insertion and folding, whether it be in the endoplasmic reticulum (ER) of eukaryotes or the cytoplasmic membrane of bacteria (Komar et al., 2016). The Signal recognition particle receptor (SR) is also associated with the translocon complex providing assistance for co-translational protein recruitment prior to insertion (Petriman et al., 2018).

1.3.1 The Sec Translocon

Early in the study of membrane protein insertion and folding, two models were proposed. The direct transfer model hypothesised that membrane proteins were able to spontaneously insert into the membrane due to the thermodynamic driving forces exerted by their hydrophobic regions (Von Heijne and Blomberg, 1979). The second model, known originally as the signal hypothesis, postulated the presence of a water filled pore in the membrane which could facilitate the translocation of highly polar protein regions across this hydrophobic barrier (Blobel and Dobberstein, 1975). In subsequent years, genetic and biochemical evidence accumulated for the presence of a highly conserved protein conducting channel which became known as Sec61 in eukaryotes and SecYEG in prokaryotes (Ito et al., 1983; Deshaies et al., 1991; Stirling et al., 1992).

The *E. coli* translocon is made up of three integral α-helical membrane proteins; SecY, SecE and SecG that associate to form a stable complex in the membrane. High-resolution structures have been solved for a number of prokaryotic SecYEG translocons (Breyton et al., 2002; Tsukazaki et al., 2008; Tanaka et al., 2015). These structures indicate that the 10 TM SecY protein forms the core of the channel, and that SecE surrounds and stabilises this central pore. SecY also contains a small periplasmic plug region, presumably to seal the inactive translocon in order to prevent ion diffusion through the water-filled cavity. SecG is peripherally associated with SecYE and is not essential for protein

translocation activity (Nishiyama et al., 1994). Figure 1.4 shows a model of the first SecYEG translocon structure solved from *Methanocaldococcus jannaschii* (van den Berg et al., 2004).

It is easy to visualise how such a channel could facilitate the translocation of polar, periplasmic proteins across the cytoplasmic membrane. However, membrane protein insertion also requires access to the hydrophobic core of the lipid bilayer. The most accepted hypothesis is the presence of a 'lateral gate' involving helices 2b and 7 of SecY, which is partially open and exposes the water-filled channel to the surrounding amphiphile during protein translocation (du Plessis et al., 2009; Park et al., 2014). However, the passage of transmembrane helices through the channel itself and their exit through the lateral gate is still under debate since none of the available structures of SecYEG contain a substrate protein transmembrane region within the channel itself. An alternative model has been proposed suggesting that only the hydrophilic loops of the protein pass through the SecY channel whereas the helices slide directly into the membrane adjacent to the lateral gate and avoid contact with the water-filled core of the channel (Cymer et al., 2015) (Figure 1.5).

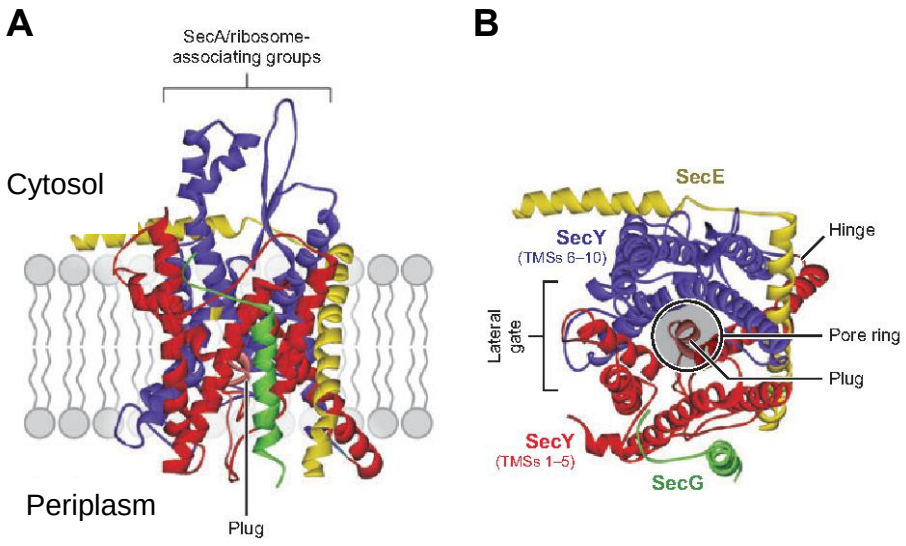


Figure 1.4 Model of the structure of SecYEG from *Methanocaldococcus jannaschii*.

A membrane cross-section **(A)** and a cytosolic **(B)** view of SecYEG from *Methanocaldococcus jannaschii* (van den Berg et al., 2004). The two 5-helix domains of SecY are shown in red and blue, SecE is shown in yellow and SecG is shown in green. The lateral gate located between helices 2b and 7 of SecY is shown in (B), along with the plug domain. Figure adapted from Driessen and Nouwen (2008).

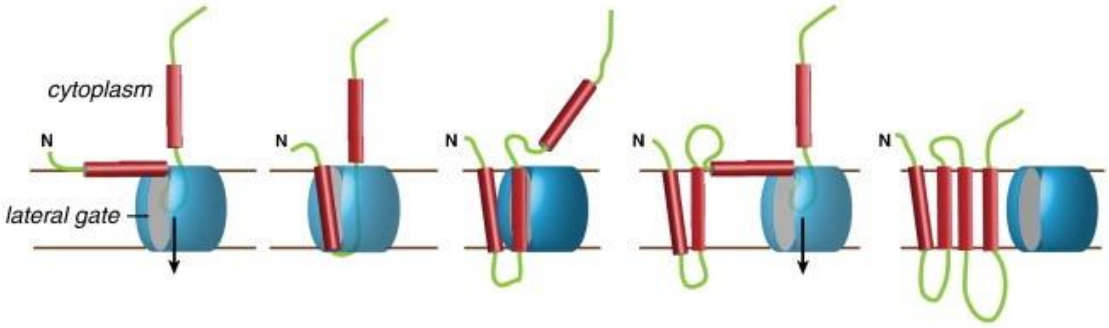


Figure 1.5 Alternative model for polytopic α-helical membrane protein insertion through the Sec translocon.

The alternative model proposes that hydrophobic helices form at the bilayer interface and never pass through the central pore the Sec translocon. Instead, they slide down the exterior of the lateral gate, which is open to allow the passage of soluble protein loops within the water-filled cavity of the translocon. Figure adapted from Cymer et al. (2015).

This hypothesis may explain why in SecE-depleted *E. coli* cells the inner membrane proteome is largely unaffected with regard to membrane proteins with small periplasmic regions (Baars et al., 2008). The findings from this study indicate that the Sec translocon is not ubiquitously required for membrane protein insertion *in vivo*.

The bacterial SecYEG translocon complex also contains an essential peripherally associated protein SecA and an associated membrane protein complex containing SecD, SecF and YajC (Du Plessis et al., 2011). SecA is a translocon associated ATPase which is essential for cell viability (Sardis and Economou, 2010). It is known to assist in the energy driven, post-translational translocation of secretory proteins via the SecB pathway as well as the translocation of large hydrophilic loops of polytopic membrane proteins being inserted co-translationally (Froderberg et al., 2003). The SecDFYajC complex has been shown to conduct protons when engaged with the translocon and thus appears to assist translocation in an ATP-independent, but proton motive force-dependent capacity (Tsukazaki et al., 2011).

1.3.2 The YidC/Oxa1/Alb3 family of membrane protein insertases

Alongside the Sec translocon, members of the YidC/Oxa1/Alb3 family play an essential role in membrane protein biogenesis. Oxa1 family members are found in the mitochondrial inner membrane and facilitate both co-translational and post-translational folding events (Szyrach et al., 2003; Preuss et al., 2005). Alb3 members are found in the chloroplast membrane of plants and facilitate at least post-translational membrane protein folding (Moore et al., 2000). YidC family members are found in both gram-positive and gram-negative bacteria and facilitate co- and post-translational membrane protein folding (Nagamori et al., 2004; Robinson and Woolhead, 2013). These proteins contain a conserved core of five transmembrane α-helices responsible for insertase activity (Shanmugaw and Dalbey, 2019). Interestingly, a YidC-like protein known as Duf106 was recently identified in the archaeon *Methanocaldococcus jannaschii*. This protein contained only three of the five conserved helices with low sequence homology to YidC, but showed high structural homology to the core region of *E. coli* YidC (Borowska et al., 2015). These data, combined with the discovery of Duf106 homologues in other archaea, suggest that YidC homologues are found throughout all prokaryotes (Kuhn and Kiefer, 2017).

Proteins of the YidC/Oxa1/Alb3 family are known to be essential for the insertion and assembly of various membrane complexes involved in energy transduction. For example, YidC and Oxa1 are required for the insertion and assembly of respiratory complexes such as cytochrome bo3 oxidase in *E. coli* and cytochrome c oxidase in mitochondria (Bonnefoy et al., 1994; Bauer et al., 1994; van der Laan et al., 2005). Alb3 is important for the post-translational insertion and assembly of the photosynthetic apparatus, specifically the light-harvesting chlorophyll binding protein (LHCBP) (Moore et al., 2000).

YidC is known to interact and cooperate with SecYEG for membrane protein biogenesis in bacteria (Scotti et al., 2000), and Alb3 has been shown to interact directly with the chloroplast translocon cpSecYE (Klostermann et al., 2002). Interestingly, the mitochondrial inner membrane does not contain a Sec-type translocon (Glick and VonHeijne, 1996) and Oxa1 is expected to be solely responsible for the insertion of membrane proteins encoded by the mitochondrial genome (Hell et al., 2001). The ER on the other hand does not

contain a strict YidC/Oxa1/Alb3 homologue. Instead, recent evidence has implicated the ER membrane protein complex (EMC) in biogenesis of polytopic α-helical membrane proteins (Chitwood et al., 2018), and EMC subunit 3 does weakly resemble the core region of bacterial YidC. The ER resident proteins Get1 and TMCO1 also show weak homology to bacterial YidC (Anghel et al., 2017) (Figure 1.6).

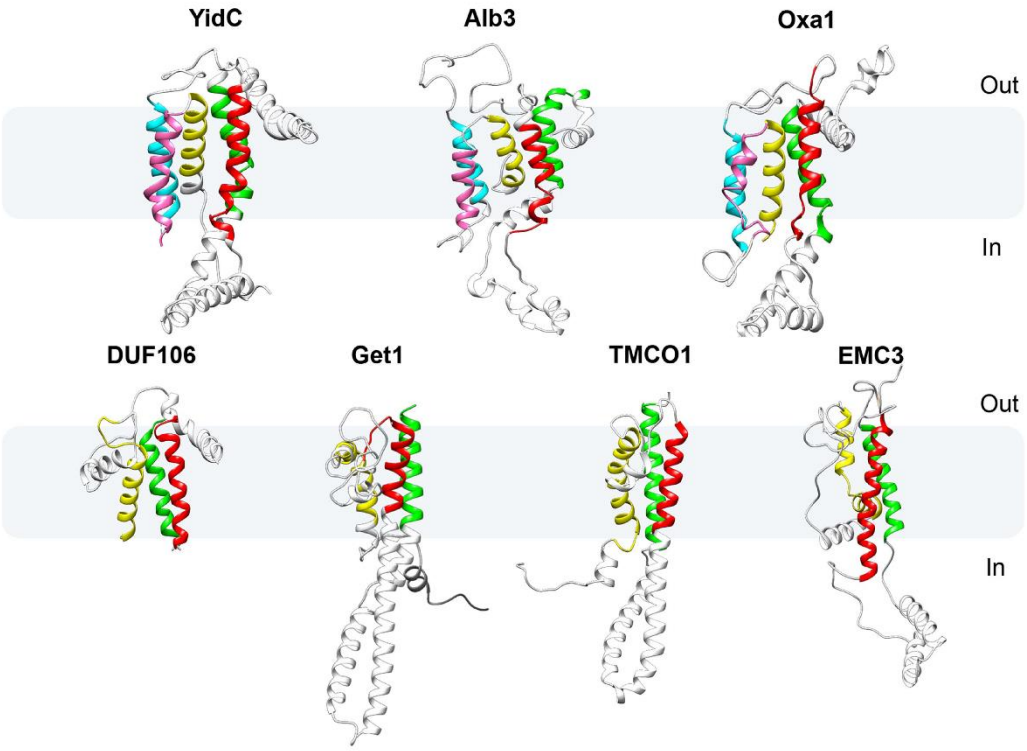


Figure 1.6 Structural homology of archaeal DUF106 and ER residents Get1, TMCO1 and EMC3 to YidC/Oxa1/Alb3 insertases.

The top row shows structural homology of bacterial YidC, chloroplast Alb3 and mitochondrial Oxa1. Related TMs are colour coded green for TM1, red for TM2, cyan for TM3, purple for TM4 and yellow for TM5. The bottom row shows structural homology of newly discovered YidC/Oxa1/Alb3 homologues from archaea (DUF106) and the eukaryotic ER colour coded as before. Figure was adapted from Shanmugaw and Dalbey (2019) where information can be found on the generation of protein structural models.

1.3.3 Membrane protein targeting

The combined efforts of the Sec translocon and a membrane insertase of the YidC/Oxa1/Alb3 family are essential for efficient membrane protein insertion and folding *in vivo*, however another important aspect of the biogenesis process is the recruitment of membrane proteins to the membrane for insertion.

In bacteria, proteins destined for secretion or the outer membrane are often targeted post-translationally to the Sec translocon by the soluble chaperone SecB (Valent et al., 1998). The vast majority of polytopic inner membrane proteins however, are targeted to the bacterial plasma membrane co-translationally by the signal recognition particle (SRP) pathway (Akopian et al., 2013b). In eukaryotic cells, membrane proteins encoded in the nucleus are inserted co-translationally into the ER via the Sec61 translocon following recruitment by the SRP and are then trafficked to their target membrane post-translationally (Rapoport, 2007).

The first step in the co-translational membrane recruitment and insertion process is the recognition of a signal sequence by the SRP as it emerges from the translating ribosome. This sequence can be a cleavable signal peptide or a non-cleaved transmembrane helix (Lee and Bernstein, 2001). Once SRP has bound to the targeting sequence, the entire SRP-ribosome nascent chain (RNC) complex is recruited to the membrane via the affinity of SRP for its cognate, membrane-bound receptor; the SRP receptor (SR). SR is known to associate with the Sec translocon in both the eukaryotic ER (Song et al., 2000) and the bacterial cytoplasmic membrane (Petriman et al., 2018). This association is important for the handoff of the RNC to the translocon for subsequent insertion and folding events. Translocon-dependent GTPase activation and subsequent disassembly of the SRP-SR-RNC complex is known to occur although the exact handover mechanism remains to be elucidated (Akopian et al., 2013a).

Bacterial SRP is a relatively simple ribonucleoprotein complex containing 4.5S RNA and a single protein called fifty-four homologue (Ffh) due to its homology to eukaryotic SRP54 (Luirink et al., 1992) (Figure 1.7). A methionine rich region of Ffh known as the M-domain is responsible for signal sequence recognition and binding (Batey et al., 2000), while a specialised GTPase domain known as the NG

domain is required for interaction with the homologous NG region of the SR, known as FtsY in bacteria (Egea et al., 2004).

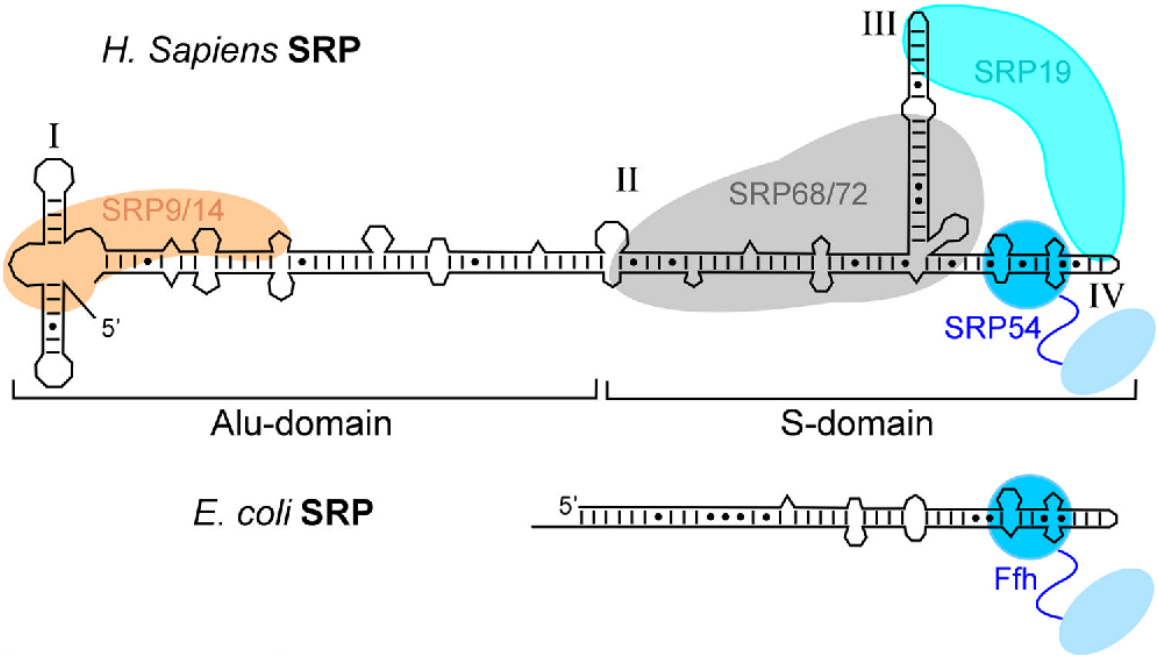


Figure 1.7 Schematic representation of the signal recognition particles (SRPs) from *Homo sapiens* and *E. coli*.

Both SRPs from eukaryotes and prokaryotes are ribonucleoprotein complexes. Eukaryotic SRP54 and bacterial fifty-four homologue (Ffh) are essential protein components that recognise substrates emerging from the ribosome and bind to the membrane associated SRP receptor (SR). Eukaryotic SRP is more complex containing extended RNA features as well as five additional proteins; SRP9 and 14 (orange), SRP68 and 72 (grey), and SRP19 (cyan). Figure adapted from Akopian et al. (2013b).

Eukaryotic SRP is more complex than the bacterial homologue and contains a larger 7S RNA and six proteins: SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72. SRP54 is the homologue of bacterial Ffh and contains the same M- and NG-domains critical for signal sequence recognition and SR binding (Krolkiewicz et al., 1994). The additional SRP proteins are essential for assembly and stability of eukaryotic SRP, but the reasons for this increased complexity are unknown (Akopian et al., 2013b). Eukaryotic SRP is further subdivided into two domains known as the S domain and the Alu domain (Figure 1.7). The S domain contains SRP54 and is responsible for substrate recognition and receptor binding, while the Alu domain slows translation by blocking the entry site for elongation factor

on the ribosome (Halic et al., 2004), although this function has been recently challenged (Chartron et al., 2016). This translational stalling activity likely provides the SRP-RNC complex with more time to localise to the Sec translocon to avoid unwanted membrane protein aggregation (Akopian et al., 2013b). Interestingly, the replacement of eukaryotic SRP and SR with bacterial Ffh and FtsY has no effect on *in vitro* protein translocation, illustrating the functional homology of the SRP system across domains (Powers and Walter, 1997).

Recent studies have found that SRP binding is not limited to the N-terminal signalling sequence, as previously assumed, but instead binds to transmembrane regions along the entire length of the nascent chain as they appear from the ribosome (Schibich et al., 2016; Costa et al., 2018). SRP also appears to bind to upstream untranslated RNA regions of the transcript through unknown mechanisms (Chartron et al., 2016).

Other soluble chaperones may also be able to assist in the targeting and insertion of polytopic inner membrane proteins. The depletion of Ffh from *E. coli* cells has been shown to result in a concomitant increase in the levels of chaperones such as GroEL/ES and DnaK/J, indicating that they may be able to partially assist membrane protein recruitment and compensate for Ffh function *in vivo* (Wickstrom et al., 2011). Studies have also shown that eukaryotic Hsp70, homologous to bacterial DnaK, assists in co-translational targeting of a subset of membrane proteins *in vivo* (Willmund et al., 2013; del Alamo et al., 2011). Additionally, *in vitro* studies have shown that GroEL/ES is able to facilitate the post-translational insertion of both lactose permease (LacY) and bacteriorhodopsin (BR) into pure liposomes (Bochkareva et al., 1996; Deaton et al., 2004).

Another mechanism utilised by cells for the recruitment of membrane proteins to their target membrane is the direct membrane attachment of their mRNA transcripts. This is a well characterised phenomena in eukaryotes (Jagannathan et al., 2014) and a putative mRNA membrane receptor known as p180 has been identified, although it is yet to be sufficiently characterised (Ueno et al., 2012). Targeting of membrane protein mRNAs has also been described in *E. coli,* suggesting that this may be a fundamental mechanism utilised throughout biology to aid translational targeting and membrane protein insertion (Nevo-

Dinur et al., 2011). An overview of the various pathways involved in membrane protein biogenesis in bacteria is shown in Figure 1.8.

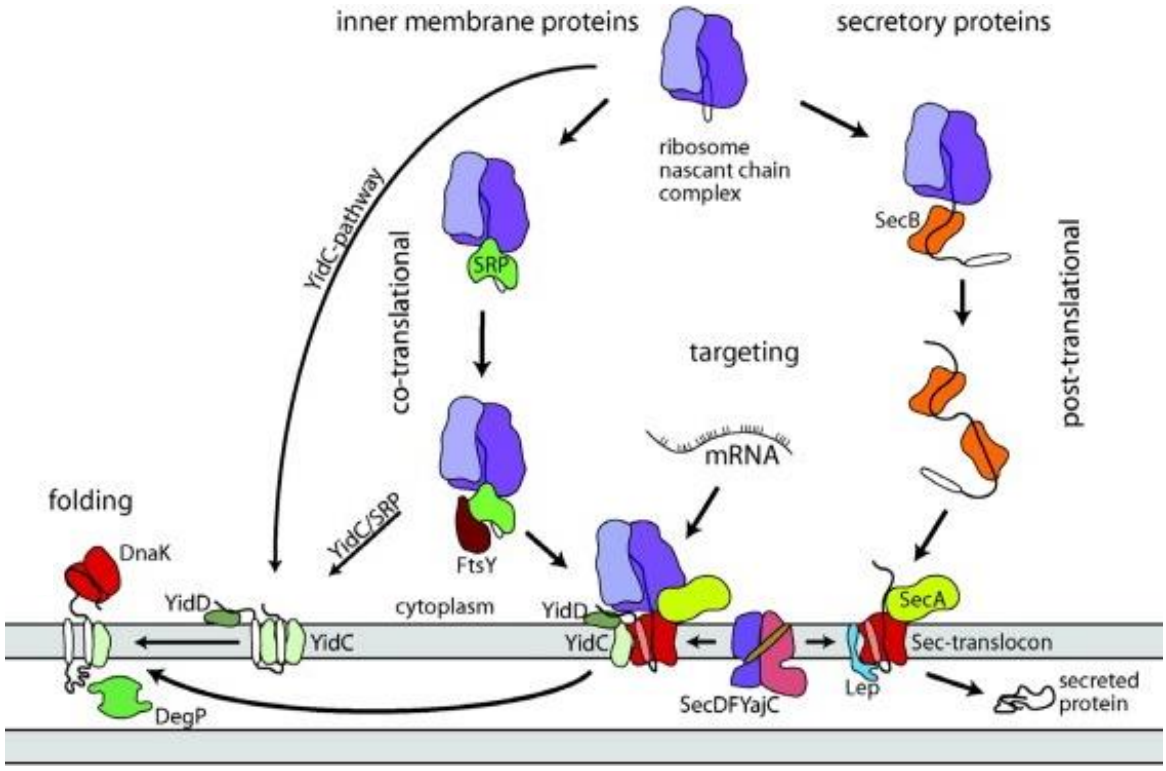


Figure 1.8 Schematic representation of bacterial membrane protein biogenesis.

The targeting of membrane proteins to their cellular destination is mediated by one of two pathways, the co-translational SRP pathway and the post-translational SecB pathway. In the co-translational pathway SRP recognises and binds to a hydrophobic nascent chain as it appears from the ribosome. SRP then binds to its cognate, membrane tethered receptor FtsY which brings the SRP-RNC complex into contact with the translocon complex. This complex contains the SecYEG translocon and the insertase YidC. SRP then hands-off the emerging nascent chain to this complex for the subsequent insertion and assembly steps. A YidC only pathway also exists which likely utilises SRP in an unknown fashion. Soluble chaperones such as DnaK may also play a role in assisting co-translational folding. In the post-translational pathway, an emerging nascent chain is recognised by the soluble chaperone SecB. SecB maintains the protein, either an outer membrane protein or a protein destined for the periplasm, in an unfolded confirmation before handing the protein to the translocon complex for ATP-dependent translocation via SecA. The SecDFYajC complex also plays a role in proton motive force dependent translocation of polar regions through SecYEG. Finally, membrane targeting of membrane protein coding mRNA transcripts has been implicated in translational recruitment. Figure adapted from Luirink et al. (2012).

1.4 *In vitro* membrane protein folding

Proteins such as the Sec translocon and YidC are essential for increasing the efficiency of membrane protein insertion and folding in the complex intra-cellular environment. However, *in vitro* studies have shown that they are not an absolute requirement for membrane proteins to fold into a lipid bilayer and that the native tertiary structures of membrane proteins represent a state of thermodynamic equilibrium. The following sections will review the current state of knowledge regarding the minimal requirements for the recruitment and insertion of proteins into a membrane.

1.4.1 The two-stage model

Popot and Engelman (1990) first proposed the two-stage model of membrane protein folding as an explanation for early *in vitro* studies on the refolding of BR. Initial work had shown that, following protein denaturation by organic solvents or SDS, BR was able to recover native structure and function upon replacement of the denaturing solvents with non-denaturing detergents or lipids (Huang et al., 1981; Lind et al., 1981; London and Khorana, 1982). These initial refolding studies were followed by fragmentation analyses where various helical fragments of BR were reconstituted separately into lipid vesicles. These individual fragment-containing vesicles were then fused to generate mixed fragment populations within the same vesicle. This mixing led to the recovery of native protein tertiary structure and absorption characteristics (Popot et al., 1986; Popot et al., 1987). Based on these studies the two-stage model of membrane protein folding was developed. The model postulates that in stage 1, an α-helical structure is adopted across the membrane due to the stability of such structural elements within the bilayer environment. This is then followed by stage 2, where these independent helices interact within the membrane to form the proteins native tertiary structure which represents a state of thermodynamic equilibrium (Popot and Engelman, 1990) (Figure 1.9).

Since the initial postulation of the two-stage model, further fragmentation studies on GPCRs (Kobilka et al., 1988; Maggio et al., 1993), lactose permease (Wrubel et al., 1994), a eukaryotic anion exchanger (Groves et al., 1998) and

Sec61 (Wilkinson et al., 1997) all seem to confirm the validity of this hypothesis, at least in simple bilayer environments.

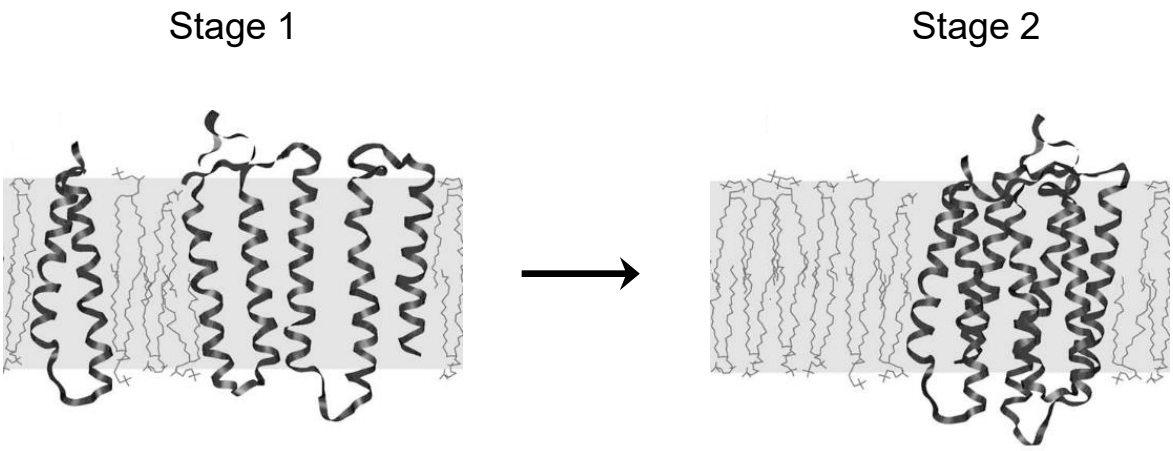


Figure 1.9 The two-stage model of membrane protein folding.

Stage 1 shows a representation of a lipid membrane containing two fragments of BR containing secondary α-helical structure, but no tertiary structure. Stage 2 shows the thermodynamically driven adoption of native tertiary structure through helical interactions. This native state is thought to be an equilibrium structure. Figure adapted from Engelman et al. (2003).

The model has since been altered to account for subsequent higher order folding events as well as the binding of prosthetic groups, however it remains an overly simplified model explaining a complex process (Engelman et al., 2003). For instance, the model deliberately fails to take into account the initial insertion of helices into the bilayer, and assumes that all individual helices of a protein are independently stable, which is known to be untrue for a number of membrane proteins (De Marothy and Elofsson, 2015). It is likely that a multitude of smaller events occur, potentially co-translationally, before the final tertiary structure is formed (Harris and Booth, 2012; Yu et al., 2017).

1.4.2 The four-step model

Fragmentation and denaturation or unfolding studies have been invaluable for probing intermediate protein folding events and determining stability free energies for a number of polytopic membrane proteins, as well as validating the underlying assumptions made by the two-stage model. However, they are less useful for investigating the very early stages of membrane protein folding due to the maintenance of helical structure in chemical denaturants and the post-translational nature of experiments. Meanwhile, the two-stage model fails to account for these early stages of folding and thus a further thermodynamic explanation is needed. Around the same time as the two-stage model was being proposed and tested, another thermodynamic model was postulated to explain the early events of membrane protein folding (Jacobs and White, 1989). This model was the first to implicate the complex environment of the membrane interface region in membrane protein helix folding and insertion. This interface region is known to account for ~50% of total bilayer thickness and provides a complex chemical environment containing water as well as phosphatidylcholine (PC), glyceryl and carbonyl groups from the lipid head and interfacial regions. This provides an environment with abundant potential for peptide interactions and ample space to accommodate α-helical structures (White and Wimley, 1998) (Figure 1.10).

This model has since been integrated with the two-stage model to form the basis of the ‘four-step’ model describing the process of hydrophobic helix insertion, folding and interaction from the stage of an unfolded polypeptide chain to the adoption of tertiary structure within the membrane (White and Wimley, 1999).

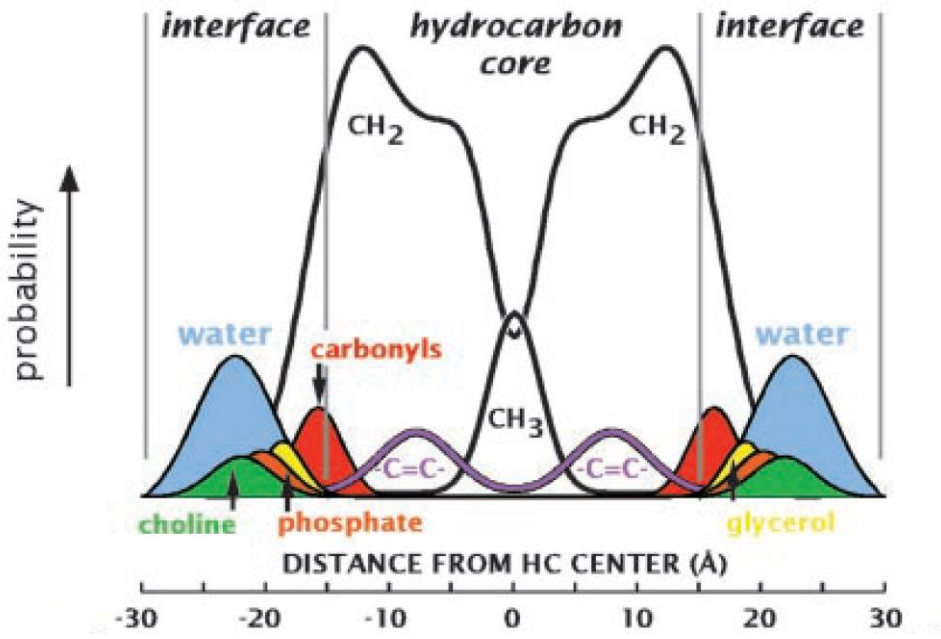


Figure 1.10 Schematic representation of the varying environments existing across a DOPC lipid bilayer.

The structure of a pure DOPC bilayer was determined by the refinement of x-ray and neutron diffraction data (Wiener and White, 1992). These data were then assembled into the present figure to illustrate the predominant chemical environment of the bilayer at a given distance from the core of the structure where the acyl chains of the lipids meet. The hydrocarbon core is dominated by the hydrophobic species present in the acyl chains while the interface regions contain a less hydrophobic environment and contain many chemical groups that may assist the formation of protein α-helical structure. Figure adapted from (White et al., 2001).

Early work focussed on the determination of partitioning free energies for each amino acid into octanol and directly into lipid bilayers using short, leucine rich peptides. Results highlighted the prohibitive free energy cost of partitioning a non-hydrogen bonded peptide into the bilayer (Wimley and White, 1996). Simulations indicated that this free energy cost was around 6.4kcal mol^{-1} per peptide bond, while the cost for a hydrogen bonded peptide was reduced to around 2.1kcal mol^{-1} per peptide bond (BenTal et al., 1997). This suggested that α-helical structure needed to be adopted prior to partitioning into the membrane and that helices would not unfold once within the hydrophobic membrane core. These early models also indicated the importance of side chain hydrophobicity within hydrogen bonded helices to overcome the still prohibitive energy barrier of partitioning the helical backbone into the membrane. To

illustrate this point, Figure 1.11 shows the free energy costs for partitioning the transmembrane helix of glycophorin A into the bilayer (Wimley and White, 2000). Partitioning of the hydrogen bonded backbone, without taking into account sidechain hydrophobicity, results in a prohibitive free energy barrier of ~24 kcal mol^{-1}. This can be overcome by the favourable free energy of partitioning hydrophobic side chains (-36 kcal mol^{-1}) resulting in a free energy balance of ~-12 kcal mol^{-1}, thus favouring membrane partitioning of the helix (Wimley and White, 2000).

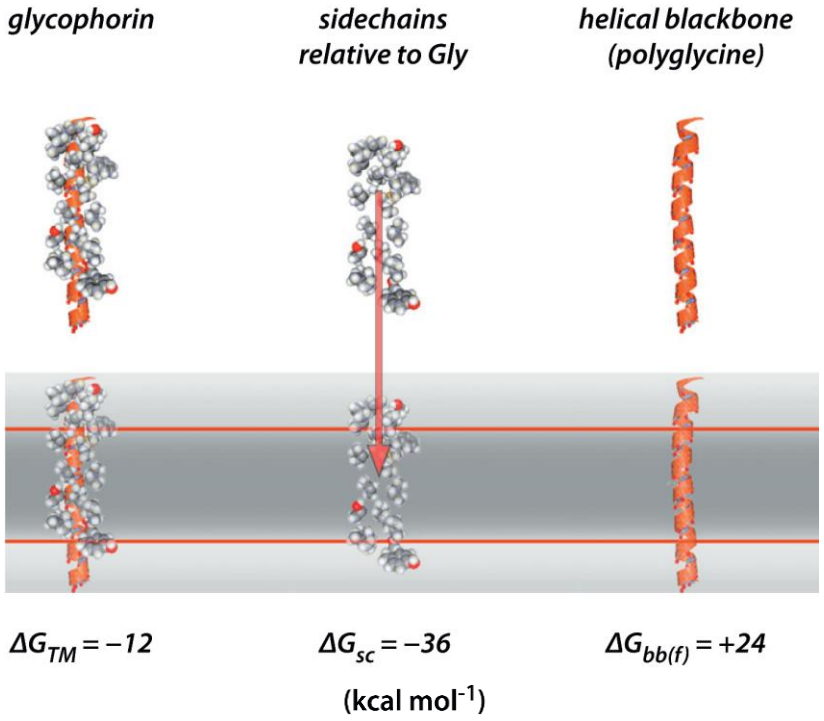


Figure 1.11 Free energy costs for partitioning the transmembrane helix of glycophorin A from water into a lipid bilayer environment.

In order to partition into the bilayer environment, the unfavourable free energy cost of transferring the hydrogen-bonded peptide backbone ($\Delta G_{bb(f)}$) must be overcome by the hydrophobicity of the amino acid side chains (ΔG_{sc}). As shown here for the transmembrane helix of glycophorin A, this side-chain effect is enough to compensate for the unfavourable $\Delta G_{bb(f)}$ leading to a combined free energy cost of -12kcal mol^{-1} (ΔG_{TM}) and thus favouring membrane partitioning of the helix. Figure adapted from Cymer et al. (2015).

It has been shown above that the free energy change from partitioning a hydrophobic peptide from the aqueous phase to the bilayer interface and finally the hydrophobic core of the bilayer is favourable. This free energy change can be described using the following equation:

$$\Delta G = \Delta H - T\Delta S$$

Where ΔG refers to the change in Gibb's free energy due to changes in enthalpy (ΔH) and entropy (ΔS) at a given temperature (T). It is therefore important to consider the relative contributions of both enthalpy and entropy that result in the spontaneous association and partitioning of a hydrophobic peptide into a lipid bilayer. Early work suggested that the major driving force for partitioning amphiphilic solutes into lipid bilayers was a large negative enthalpy and was not driven, as expected, by a large positive entropy from the hydrophobic effect (Seelig and Ganz, 1991). These results were challenged by further studies indicating that the relatively weak entropic force was due to unknown thermodynamic effects generated by the bilayer itself (Wimley and White, 1992; Wimley and White, 1993). In a more recent study by Fernández-Vidal et al. (2011), the authors were able to disentangle bilayer and non-bilayer entropic contributions to the partitioning free energy of melittin. They found that the hydrophobic effect was the dominant driving force responsible for membrane partitioning despite variations in enthalpic values due to lipid composition and headgroup charge (Fernandez-Vidal et al., 2011).

The requirement for α-helical structure formation prior to membrane partitioning further highlights the importance of the interface region to this spontaneous process. This region is known to enhance adoption of helical secondary structure in mildly hydrophobic proteins such as melittin by encouraging the formation of hydrogen bonds (Ladokhin and White, 1999). It therefore seems likely that contact with the interface acts in a chaperoning capacity to drive secondary structure formation of more hydrophobic helices that would aggregate outside of this environment prior to the adoption of helical structure. Figure 1.12 illustrates the four-step model of helix folding, partitioning and structure formation.

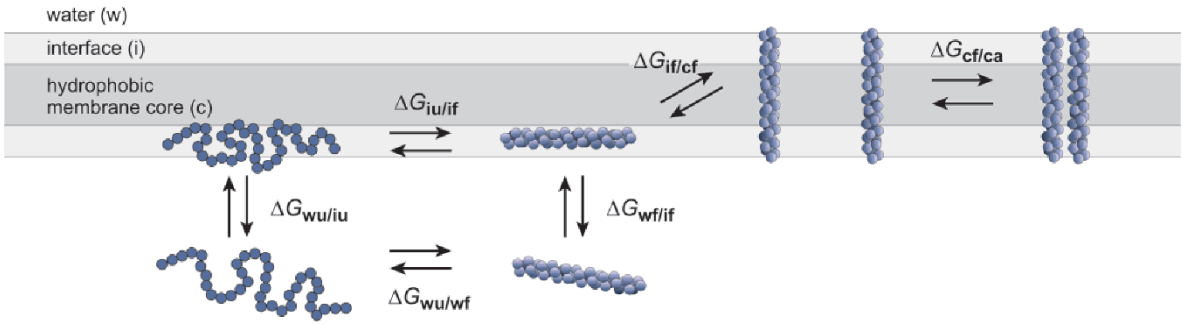


Figure 1.12 The four-step thermodynamic model of hydrophobic helix folding, insertion and interaction.

A hydrophobic peptide initially in the water phase can either partition to the membrane interface region in an unfolded state ($\Delta G_{wu/iu}$) or adopt some helical structure in the water phase ($\Delta G_{wu/wf}$) prior to transitioning to the interface ($\Delta G_{wf/if}$). If the peptide is in an unfolded state, the properties of the interface allow the adoption of α-helical structure ($G_{iu/wf}$) which is required to allow partitioning of the peptide into the core of the bilayer ($G_{if/cf}$). These stably inserted helices can then interact within the membrane to form tertiary structures ($G_{cf/ca}$). Figure adapted from MacKenzie (2006).

Molecular dynamics simulations have been used to interrogate this four-step folding pathway and examine the importance of the interfacial region. An illuminating study from Ulmschneider et al. (2011) placed an unfolded polyleucine (Leu10) peptide 10 Å from a POPC bilayer and ran μs-scale simulations. Results showed that within 2 ns the protein was adsorbed to the interface of the bilayer in the unfolded state. The peptide then adopted helical structure at the interface and subsequently partitioned into the core of the bilayer. While there was some transition between interfacial helix and membrane helix throughout the experiment, the peptide never lost its helical secondary structure or partitioned back into the aqueous phase in accordance with the free energies discussed previously (Ulmschneider et al., 2011) (Figure 1.13). Further simulations using different hydrophobic peptides have confirmed this behaviour and results were independent of the force fields applied to experiments (Ulmschneider et al., 2014; Ulmschneider et al., 2017; Ulmschneider et al., 2018; Gumbart et al., 2018).

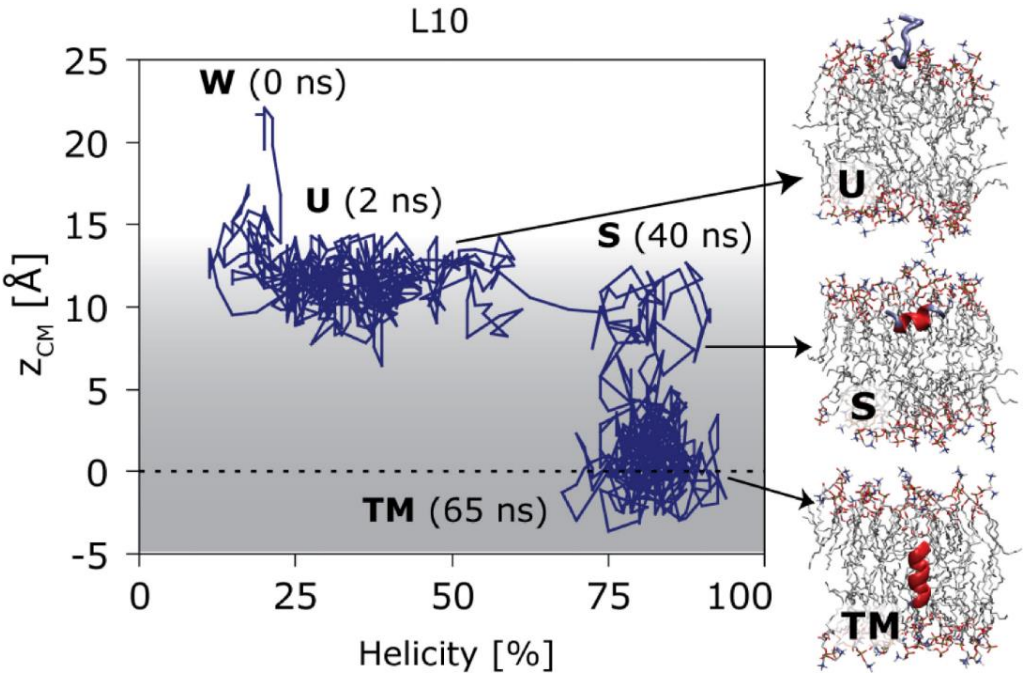


Figure 1.13 Simulation data showing the folding and insertion of a short, hydrophobic polypeptide.

The simulation shows the folding path undertaken by a polyleucine (Leu10) peptide introduced in an unfolded state to a POPC bilayer. Initially, the unfolded protein adsorbed irreversibly to the interface region. This was followed by the adoption of helical structure and the subsequent transition between interfacial and bilayer integrated states. Following the initial attachment to the interface, the peptide was never again able to unfold into the aqueous phase. Figure adapted from Ulmschneider et al. (2011).

One of these studies examined the insertion of hydrophobic peptides of varying length, both *in silico* and *in vitro* using a previously established glycosylation assay designed to measure translocon-guided translocation (van Geest and Lolkema, 2000). The authors compared the free energies for spontaneous helix insertion and translocon-guided insertion. Their results indicated a remarkable correlation between values calculated from *in vitro* experiments and those from *in silico* simulations, suggesting that the translocon may be more important for the transfer of polar protein regions across the membrane than for partitioning into the membrane, and that the membrane interface may well play a key role in membrane protein folding *in vivo* (Ulmschneider et al., 2014).

It is also important to understand how helices interact with each other within the membrane, since these interactions are vital to the adoption of membrane protein tertiary structure. Helix-helix interaction corresponds to the final free

energy transition in the four-step model (Figure 1.9). Original studies interrogating the two-stage model clearly indicated that polytopic proteins contain intrinsic properties that drive helical interaction and the adoption of native tertiary structure (see section 1.4.1). The protein glycophorin A has been used as a simple model for helical interaction studies since it forms a stable dimer in both SDS micelles and in lipid membranes (Bormann et al., 1989). Evidence suggested that the dimer was held together by van der Waals interactions and a tight steric interaction known as 'knobs-into-holes packing' (Cohen and Parry, 1990; Eriksson et al., 1992; Lemmon and Engelman, 1994). These data led to the identification of the GXXXG helical interaction motif which is common to many polytopic α-helical membrane proteins (Russ and Engelman, 2000). The formation of intra-helical hydrogen bonds may also be important for the stabilisation of helical interactions. For example, the inclusion of hydrogen bonding polar residues within synthetic transmembrane peptides resulted in dimerization following their expression *in vivo* in *E. coli* (Zhou et al., 2001). The design of synthetic membrane proteins has also noted the stabilising effects of intrahelical hydrogen bonds formed by polar residue containing helices (Mravic et al., 2019).

What is becoming clear from the majority of studies on membrane protein thermodynamics is that if membrane protein helices are able to partition into the bilayer, they will likely adopt their native tertiary structure, which appears to be in a state of thermodynamic equilibrium. Initial helical structure may form within the interfacial region of the bilayer, especially when dealing with moderately hydrophobic helices. Finally, it should be noted that the propensity for hydrophobic helices to aggregate in the aqueous phase reiterates the importance of the initial recruitment of these peptide chains to the correct location, be it the membrane interface or the translocon.

1.5 Cell-free membrane protein synthesis

What had become clear from fragmentation and unfolding studies was that individually inserted helices, or bundles of helices, could spontaneously form the necessary contacts for the adoption of native tertiary structure. The vast majority of α-helical membrane proteins *in vivo* insert into the membrane co-translationally. The synthesis of polytopic membrane proteins using cell-free systems has thus emerged as an exciting tool allowing researchers to examine membrane protein insertion in a more biologically relevant, co-translational, context.

1.5.1 Using cell lysate-based systems

So-called cell-free protein synthesis kits contain all the necessary components for the synthesis of proteins directly from either DNA or RNA *in vitro*. This is most commonly achieved through the use of cell lysates produced from *E. coli* cells, wheat germ cells or rabbit reticulocyte cells. Cell-free kits are available commercially, but detailed protocols for their in-house production are also available (Kigawa et al., 1999). The resulting mixtures contain all the soluble components of the cells while removing the membrane fractions and allow the synthesis of proteins from DNA or RNA following the addition of various translation factors such as nucleotides, amino acids and energy substrates (Figure 1.14). The major disadvantage of cell-lysate kits is that the exact molecular composition of the lysate is not known. In order to circumvent this issue, an *E. coli*-based recombinant system has been developed known as the Protein synthesis Using Recombinant Elements (PURE) system, which contains a minimal set of highly purified cellular components, thus removing the vast majority of cellular proteins (Shimizu et al., 2001). This recombinant approach will be discussed later.

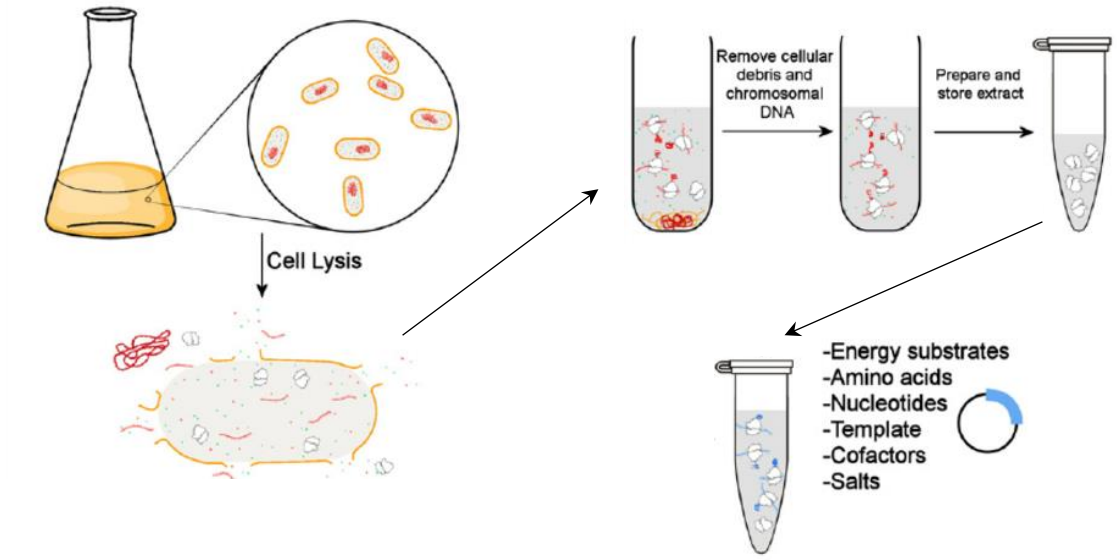


Figure 1.14 Schematic representation of the preparation of a cell lysate for cell-free protein synthesis.

Chosen cells are cultured according to standard methods followed by lysis to release cytoplasmic components. Membrane fragments are then removed by ultracentrifugation and genomic DNA (gDNA) is removed by precipitation. The membrane and gDNA-free extract can then be stored at -80°C prior to use. Energy substrates, amino acids, nucleotides, salts, translation cofactors and template DNA or RNA are added to the extract to induce protein synthesis. Figure adapted from Carlson et al. (2012).

Pioneering studies were performed in the 1990s using mainly rabbit reticulocyte lysate (RRL) to produce functional membrane proteins in microsomes. Microsomes are vesicle structures derived from fragments of the ER of eukaryotic cells such as *Xenopus laevis* oocytes. These microsomes contain all of the proteinaceous membrane components of the parent cell's ER including the Sec translocon. An initial study in 1990 showed for the first time the synthesis and assembly of a polytopic membrane protein using such a cell-free approach (Kobilka, 1990). Their results indicated native ligand binding and post-translational modification of the human GPCR, ß-2 adrenergic receptor, suggesting the formation of native structure (Kobilka, 1990). This initial study was followed in 1992 by the cell-free synthesis of the first functional ion channel (the shaker potassium channel from *Drosophila melanogaster*) in microsomes and transfer to planar lipid bilayers (PLBs) (Rosenberg and East, 1992). These seminal studies led to a number of groups showing the native folding of cell-free synthesised proteins such as the sodium channel α-bENaC (Awayda et al., 1995),

connexins CX26, CX32 and CX43 (Falk et al., 1997) and multiple receptor proteins (Huppa and Ploegh, 1997; Joseph et al., 1997; Lyford and Rosenberg, 1999).

Up until this point, the cell-free synthesis of polytopic membrane proteins had relied on the use of model membrane systems containing all of the translocation machinery, such as microsomes or inverted membrane vesicles (IMVs). The first example of spontaneous insertion and oligomerisation of a polytopic membrane protein, synthesised using an *E. coli*-derived lysate, was published in 2002 (van Dalen et al., 2002). The authors observed the direct, co-translational insertion and oligomerisation of the prokaryotic potassium channel KcsA into liposomes generated from a purified lipid extract. The insertion was reliant on the presence of Ffh in the lysate, but not on the presence of the post-translational chaperone SecB (van Dalen et al., 2002). These experiments showed for the first time that the insertion of *de novo* synthesized helices into the membrane did not absolutely require a proteinaceous insertion apparatus but seemed to be driven by fundamental thermodynamic constraints. The need for Ffh also indicated the importance of chaperoning these aggregation-prone transmembrane segments through the aqueous phase.

A number of studies followed showing the functional synthesis of a wide variety of polytopic membrane proteins directly into detergent micelles, adding further weight to the assumption that the thermodynamic equilibrium of membrane proteins favours spontaneous membrane insertion as hypothesised by the two-stage and four-step models (Berrier et al., 2004; Klammt et al., 2004; Elbaz et al., 2004; Ishihara et al., 2005) (see section 1.4). While valuable, these studies did not address the process of helical insertion into a sealed lipid bilayer such as a liposome.

The direct insertion and folding of a polytopic membrane protein into an empty lipid bilayer was achieved again in 2007 (Kalmbach et al., 2007). The authors were able to demonstrate the spontaneous insertion and functional activity of cell-free synthesised BR in small unilamellar vesicles (SUVs) using an *E. coli* lysate. They also examined the effect of altered lipid acyl chain length, and thus altered hydrophobic mismatch, on insertion efficiency. The highest rate of insertion was observed using 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) which has an acyl chain length of 18 carbons and one unsaturated chain

containing a cis-double bond at carbon 9 (Kalmbach et al., 2007). DOPC bilayers exist in the L_α or liquid disordered phase at room temperature and are known to be ~60 Å thick including the interface region, with a hydrocarbon core thickness of ~30 Å (Wiener and White, 1992).

More recent lysate-based studies have shown the direct insertion and assembly of a number of polytopic membrane proteins and protein complexes. The ATP-synthase complex from the thermophilic bacterium *Caldalkalibacillus thermarum* was successfully inserted into LUVs generated from a mixture of DOPC and 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DOPG) (Matthies et al., 2011). The *E. coli* mechanosensitive channel MscL, previously identified as a YidC substrate, was shown to insert functionally into asolectin liposomes in the absence of YidC (Berrier et al., 2011). The bacterial peptidoglycan synthesis enzyme MraY from both *E. coli* and *Bacillus subtilis* was functionally inserted into soybean phosphatidylcholine (PC) liposomes (Ma et al., 2011). The mitochondrial ADP/ATP carrier (AAC) from *S. cerevisiae* was directly inserted into LUVs generated from a mixture of 1-palmitoyl-2-oleoyl (PO) lipids and cardiolipin (CL) and exhibited transport function (Long et al., 2012). The spontaneous integration and pore formation of connexin 43 (CX43) was also achieved inside giant unilamellar vesicles (GUVs) made from an egg PC mixture (Liu et al., 2013). This study represented the first reported case of the direct insertion of a membrane protein synthesised from inside a liposome.

This non-exhaustive list indicates the wide range of polytopic α-helical membrane proteins that are able to spontaneously insert into pure liposome membranes without the aid of membrane insertases such as YidC or SecYEG, although the use of cell-lysates makes it impossible to completely rule out the presence of small quantities of these proteins. It should be noted that lipid nanodiscs have also become a popular folding medium, with various studies using them in conjunction with lysate-based cell-free for the analysis of polytopic membrane protein folding (Gao et al., 2012; Komar et al., 2016; Winterstein et al., 2018).

1.5.2 The PURE system for membrane protein synthesis

The major drawback of using cell-lysates is the presence of the entire soluble proteome of the cell. This contains various chaperones and many proteins of unknown function. It is also difficult to completely rule out the presence of small membrane fragments containing translocon components in these lysates. In order to address these issues, the PURE system was developed in 2001 (Shimizu et al., 2001). This recombinant system contains 36 individually purified enzymes as well as extensively purified 70S ribosomes, and when supplemented with small molecular weight components such as NTPs and amino acids, allows the *in vitro* synthesis of proteins from DNA under the control of a T7 promoter (Shimizu et al., 2001). The first study to use this system for the synthesis of polytopic membrane proteins was performed in 2005 and examined the insertion of the *E. coli* beta barrel outer membrane protein A (OmpA), the inner membrane polytopic α-helical PTS system mannitol-specific EIICBA component (MtlA) and the single-pass transmembrane cell division protein (FtsQ) into IMVs (Kuruma et al., 2005). The first example of direct integration into empty liposomes using the PURE system showed that connexin 43 (CX43) could integrate and form functional channels in DOPC LUVs and GUVs (Moritani et al., 2010) and was followed by a number of papers showing the direct integration of polytopic membrane proteins into various liposome-based systems. For example, the multidrug transporter EmrE was functionally inserted directly into POPC GUVs when synthesised in the lumen (Soga et al., 2014). SecY, SecE, SecG and YidC were also successfully inserted into soybean PC LUVs using the PURE system, although YidC required SecYEG for the translocation of its large periplasmic domain (Matsubayashi et al., 2014). *E. coli* LacY was shown to be functional following PURE synthesis inside a droplet interface bilayer system (Findlay et al., 2016). The rhomboid protease GlpG and the thiol:disulfide interchange protein DsbA were recently integrated directly into LUVs and nanodiscs with structure formation being temporally analysed by surface-enhanced infrared absorption spectroscopy (SEIRAS) (Harris et al., 2017). In a recent study, BR was synthesised using the PURE system in LUVs that also contained detergent-reconstituted ATP synthase. The resulting protein/lipid assembly was used as a simple energy-producing 'artificial organelle' encapsulated within a GUV chassis (Berhanu et al., 2019). These studies clearly show that a diverse range of polytopic

membrane proteins can insert directly into various liposome-based systems without the aid of the translocon or any other soluble or membrane-bound chaperoning components. This implies that thermodynamic driving forces alone account for both the insertion of helices and the adoption of native tertiary structure in these simplified membrane mimetic environments.

1.5.3 Amphiphilic environments

Regardless of the experimental methodology being followed, the study of membrane protein insertion and folding requires the selection of a suitable amphiphilic environment that will allow the hydrophobic and hydrophilic regions of the protein to fold correctly. Many of these have been introduced in previous sections but are explained in more detail here.

Amphiphiles used in the study of membrane proteins can be split into three groups: detergents, lipids and polymers. All have the essential hydrophobic and hydrophilic regions in their molecular structure that define an amphiphile and provide an accommodating environment for membrane proteins to fold. The vast majority of studies have utilised detergents and lipids, however there is a growing trend towards the use of polymers for the generation of highly stable protocells functionalised with embedded membrane proteins as artificial cell models (Martino et al., 2012; Nallani et al., 2011).

Many detergents have been designed and optimised for solubilising membrane proteins while allowing them to maintain their tertiary structures, and their use has been key to the determination of high resolution structures (Garavito and Ferguson-Miller, 2001). While these structures provide an invaluable source of information, it should not be forgotten that the *in vivo* membrane environment is quite different. All cells are bound by a lipid-based membrane, and the lipid composition influences membrane protein properties through a number of important features.

Lipids can influence protein insertion, folding and topology through headgroup charge and occupied area as well as acyl chain features such as hydrophobic length and saturation. For example, anionic lipids can interact with positively charged amino acid residues to influence the topology of membrane proteins

during co-translational insertion and during post-translational detergent reconstitution experiments (vanKlompenburg et al., 1997; Tunuguntla et al., 2013; Rues et al., 2016). Acyl chain saturation leads to changes in the lateral pressure profile of the membrane's hydrophobic interior and has been shown to affect insertion and function of membrane proteins such as MscL, GlpG and DsbB (Roos et al., 2013; Harris et al., 2017). Hydrophobic mismatch due to acyl chain length has also been shown to alter the insertion efficiency of the single transmembrane Pf3 coat protein (Ridder et al., 2002). One of the most striking examples of lipid-dependent effects on a membrane protein emerged from studies of the MFS transporter LacY (Bogdanov et al., 2008). This protein has been shown to alter its own topology post-insertion based on the surrounding lipid environment. The topological switching occurs *in vivo* but also *in vitro* in empty liposomes containing no other membrane protein components (Bogdanov et al., 2008; Vitrac et al., 2013). These results were remarkable given the large energetic barrier that must be overcome to transfer large polar regions of the protein across the membrane. Since the initial study on LacY, topological flipping has been observed for other unrelated membrane proteins upon alterations to the lipid environment (Bogdanov et al., 2014).

Another key factor that is rarely discussed is the size and curvature of the mimetic system being used. For example, the majority of cell-free studies have utilised either SUVs or LUVs as the membrane-mimetic chassis for protein insertion, with proteins being synthesised and inserted from the outside of these vesicles. The nanometre scale of SUVs and LUVs means they have a very high positive curvature unlike living cells, and curvature differences have been shown to affect membrane protein function (Botelho et al., 2006; Fujii et al., 2015; Li et al., 2017).

GUVs have become a popular tool due to their large size (>1 μm diameter) and volume/surface ratio which more realistically mimics bacterial cells or mitochondria (Figure 1.15). Their size also makes them amendable to optical microscopy techniques and allows the encapsulation of reactions within their lumenal space (Walde et al., 2010). However, this larger size potentially favours aggregation of hydrophobic proteins before they reach the membrane environment. Soga et al. (2014) investigated the relationship between increasing surface area to volume ratio on the insertion of the polytopic membrane protein

EmrE using the PURE system within GUVs. They found that, as lumenal volume increased, the efficiency of EmrE insertion decreased. This observation could be due to the lack of any chaperones or membrane recruitment components in the PURE system. It would be logical to assume that as the aqueous volume increased the amount of protein being in direct contact with the membrane would decrease and the rates of protein aggregation would increase. Therefore, GUVs are a good system to study the minimum requirements for both the recruitment and the insertion of proteins into the membrane in a cell-size mimicking, minimal system.

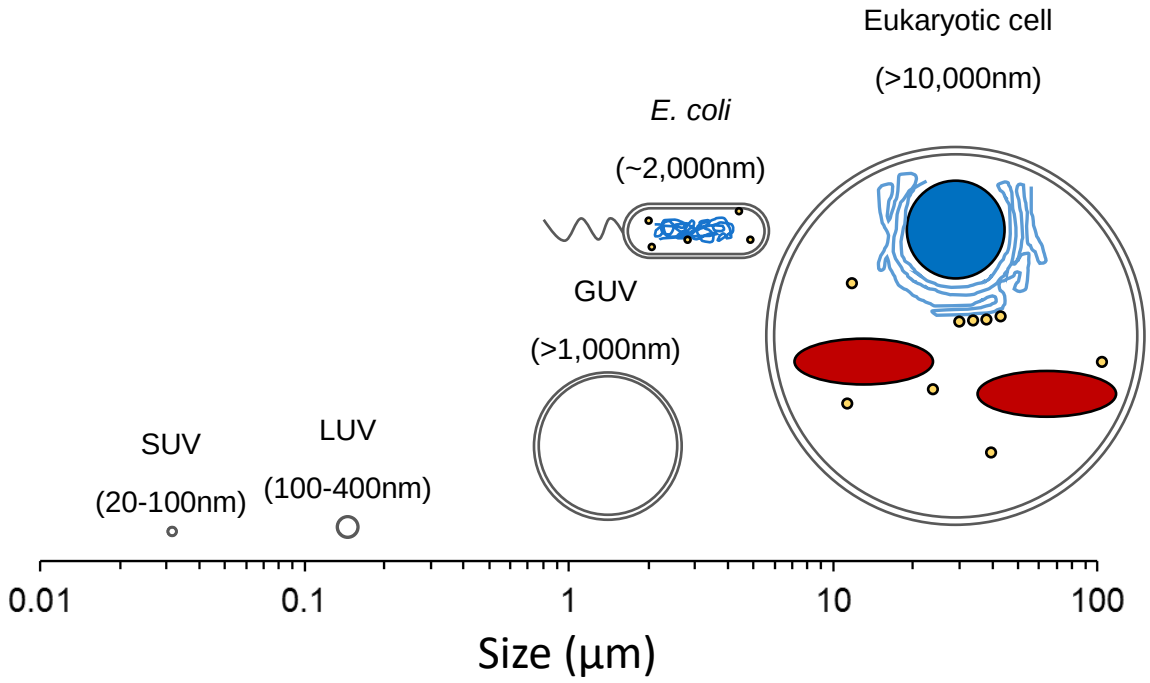


Figure 1.15 Schematic representation of the scale of liposomes in comparison to bacterial and eukaryotic cells.

The most commonly used liposomes for *in vitro* membrane protein studies are SUVs, which range from 20 - 100 nm in diameter, and LUVs, which range from 100 - 400 nm in diameter. GUVs of > 1µm in diameter are better able to mimic the volume to surface area ratio of cellular organisms such as *E. coli*. The schematic representation of an *E. coli* cell shows the genome in blue and ribosomes as yellow circles. The schematic representation of a typical eukaryotic cell shows the nucleus and ER in blue, mitochondria in red and ribosomes as yellow circles. Representations are not drawn to scale.

1.6 Thesis aims

An increasing number of polytopic α-helical membrane proteins have been shown to spontaneously insert and fold into pure lipid membranes without the aid of the translocation apparatus or any other chaperones, driven purely by thermodynamic effects. Given the propensity of hydrophobic domains to aggregate in the aqueous phase, it is logical to assume that the availability of a membrane environment is key to this process. Another possibility is that the emerging nascent chains themselves are able to act as primitive targeting signals and direct the translating ribosome to the membrane in the absence of more complex targeting pathways.

The aim of this thesis was to answer the following questions:

1. What are the minimal requirements for efficient recruitment and insertion of polytopic α-helical proteins into a lipid membrane in a cell-size mimicking system?

2. Are these processes driven by protein-inherent features or do they require other machineries such as chaperones or insertases?

3. How important is it to localise translation to the membrane and how does this affect the process of spontaneous insertion?

To answer these questions the objectives of the thesis were to:

1. Use the PURE system to synthesise the polytopic membrane proteins proteorhodopsin (PR) and galactose permease (GalP) inside GUVs and to assess their membrane localisation and insertion using fluorescence-based microscopy assays.

2. Investigate the importance of the N-terminal hydrophobic domains of PR and GalP to the targeting and insertion of these proteins in GUVs.

3. Synthetically attach ribosomes to the membrane of GUVs in order to determine the effects of localised translation on membrane protein targeting and insertion.

2. Materials and methods

2.1 Materials

2.1.1 Chemicals, reagents and buffers

Unless otherwise stated, all chemicals used were purchased from Sigma-Aldrich (UK) or Thermo-Fisher Scientific (UK). Lipid products were from Avanti-Polar Lipids (USA) and enzymes used for molecular biology were obtained from Promega (UK), Invitrogen (UK) and Qiagen (UK). Plasmid DNA purification kits were from Qiagen (UK) and PCR purification/gel extraction kits were from Invitrogen (UK). Primers were synthesised by Integrated DNA Technologies (Belgium). PURE cell-free protein synthesis kits were purchased from Genefrontier (Japan). A list of the solutions and media used can be found in section 2.1.2 below.

2.1.2 Buffer composition

6x Agarose gel loading buffer

10 mM Tris-HCl (pH 7.6), 30% (v/v) glycerol, 0.15% (w/v) orange G

Competent cell storage buffer

100 mM $CaCl_2$, 15% (v/v) glycerol

Cushion buffer

20 mM HEPES-KOH (pH 7.6), 10 mM magnesium chloride, 150 mM potassium chloride, 30 mM ammonium chloride and 30% (w/v) sucrose.

EGFP elution buffer

50 mM Tris-HCl (pH 8), 150 mM NaCl, 200 mM imidazole

EGFP dialysis buffer

50 mM Tris-HCl (pH 8), 150 mM NaCl

EGFP re-suspension/wash buffer

50 mM Tris-HCl (pH 8), 150 mM NaCl, 10 mM imidazole

Extrusion buffer

50 mM HEPES-KOH (pH 7.6), 100 mM potassium glutamate and 15 mM magnesium acetate.

LB media

0.1% (w/v) tryptone, 0.1% (w/v) NaCl, 0.05% (w/v) yeast extract. (1.5% (w/v) agar added for LB agar).

Polymix buffer

5 mM magnesium acetate, 5 mM ammonium chloride, 95 mM potassium chloride 0.5 mM calcium chloride, 8 mM putrescine, 1 mM spermidine, 5 mM potassium phosphate (pH 7.6) and 1 mM DTT.

PURE feed solution

0.3 mM of each of the 20 amino acids, 1.5 mM spermidine, 3.75 mM ATP, 2.5 mM GTP, 1.25 mM of both CTP and UTP, 25 mM creatine phosphate, 1.5 mM TCEP, 18 mM magnesium acetate, 280 mM potassium glutamate, 50 mM HEPES-KOH pH7.6 and 0.02 mg ml^{-1} folinic acid

Ribosome dialysis buffer

20 mM HEPES-KOH (pH 7.6), 10 mM magnesium chloride, 150 mM potassium chloride and 30 mM ammonium chloride.

Ribosome elution buffer

20 mM HEPES-KOH (pH 7.6), 10 mM magnesium chloride, 150 mM potassium chloride, 30 mM ammonium chloride, 150 mM imidazole.

Ribosome re-suspension buffer

20 mM HEPES-KOH (pH 7.6), 10 mM magnesium chloride, 150 mM potassium chloride, 30 mM ammonium chloride, 0.5 mg ml^{-1} lysozyme and 10 µg ml^{-1} RNase-free DNase I.

Ribosome wash buffer

20 mM HEPES-KOH (pH 7.6), 10 mM magnesium chloride, 150 mM potassium chloride, 30 mM ammonium chloride, 5 mM imidazole.

4x SDS-PAGE resolving buffer

1.5 M Tris-HCl (pH 8.8), 0.4% (w/v) SDS

4x SDS-PAGE sample buffer

200 mM Tris-HCl (pH 6.8), 8% (w/v) SDS, 40% (w/v) glycerol, 0.2% (w/v) bromophenol blue, 400 mM ß-mercaptoethanol

4x SDS-PAGE stacking buffer

0.5 M Tris-HCl (pH 6.8), 0.4% (w/v) SDS

10x SDS-PAGE running buffer

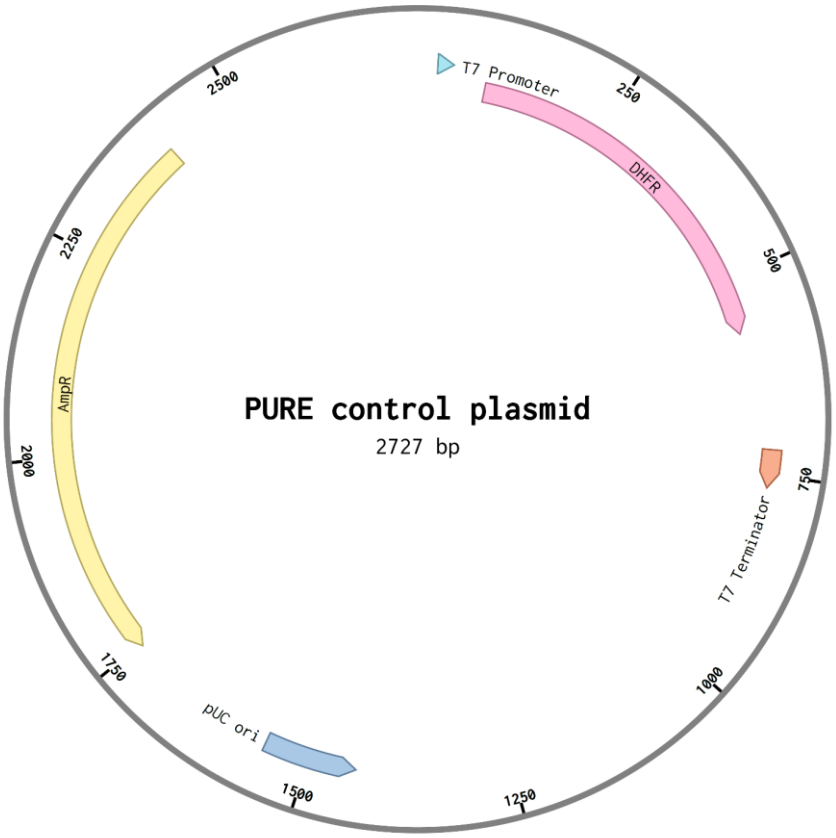
250 mM Tris, 1.89 M glycine, 1% (w/v) SDS

50x TAE buffer

2 M Tris, 5.71% (v/v) glacial acetic acid, 50 mM EDTA.

2.1.3 Plasmid vectors

The plasmid vector used for all work was the control vector supplied as part of the PURE cell-free expression kit. This vector contains a T7 promoter and terminator sequences flanking the gene of interest. It contains an ampicillin resistance gene that was used for selection. The control vector contained the coding sequence for the DHFR protein, which was removed during cloning reactions through linearisation of the vector by PCR.



Professor Edward DeLong (Massachusetts Institute of Technology, USA) kindly provided the vector containing the gene encoding green light-absorbing proteorhodopsin from a member of the SAR86 group of γ-proteobacteria (Beja et al., 2000). This proteorhodopsin gene was amplified by PCR and ligated into the control plasmid in place of DHFR.

2.2 Methods

2.2.1 Microbiology techniques

2.2.1.1 Preparation of LB plates

LB-Agar (1.5% w/v) was autoclaved at 121°C for 20 minutes and subsequently allowed to cool to ~60°C before the addition of ampicillin to a concentration of 100 µg ml^{-1}. Plates were then poured under sterile conditions.

2.2.1.2 Preparation of chemically competent *E. coli*

A 100 µl aliquot of chemically competent Top-10 *E. coli* cells was grown in 5 ml of LB media supplemented with 10 µg ml^{-1} streptomycin shaking at 180 rpm overnight (~16 hours) at 37°C. 1 ml of this overnight culture was used to inoculate 500 ml of streptomycin supplemented LB which was grown shaking at 37°C to an optical density at 600 nm (OD_{600}) of 0.3 - 0.4. Cultures were then transferred to sterile 50 ml falcon tubes and were pelleted by centrifugation at 4,000 rpm for 10 minutes at 4°C using a Sorvall Legend Rt centrifuge. The supernatant was discarded, and pellets were re-suspended and washed thoroughly with ice cold 100 mM $MgCl_2$ before being re-pelleted by centrifugation as before. Cells were re-suspended and washed in ice cold 100 mM $CaCl_2$ before a final centrifugation as before. Pellets containing chemically competent cells were then carefully re-suspended in competent cell storage buffer and aliquoted into sterile 1.5ml eppendorf tubes (50 µl aliquots), snap frozen in liquid N_2 and stored at -80°C.

2.2.1.3 Transformation of competent *E. coli*

A 50 µl aliquot of chemically competent Top-10 cells was thawed on ice before supplementation with ~100 ng of the relevant pDNA, 5 µl of ligation reaction, or 2 µl of Gibson assembly reaction. Suspensions were gently mixed and incubated on ice for 20 minutes prior to a 45-second heat shock at 42°C. Cells were returned to ice for 3 minutes before the addition of 500 µl of antibiotic-free LB media and incubation shaking at 37°C for 45 minutes. A relevant volume of transformed cells was then spread on antibiotic-supplemented LB-Agar plates and incubated at 37°C overnight.

2.2.2 Molecular biology techniques

2.2.2.1 Plasmid miniprep DNA isolation

A single colony was selected from an *E. coli* transformation plate and was used to inoculate 5 ml of LB media containing the relevant antibiotic. This culture was grown shaking overnight at 37°C before cells were pelleted by centrifugation at 4,000 rpm for 15 minutes at room temperature. The Qiagen QIAprep Spin Miniprep Kit was used to isolate pDNA according to the manufacturers' instructions. Briefly, cell pellets were re-suspended in 250 µl of buffer P1 before cell lysis was initiated by the addition of 250 µl buffer P2. 350 µl of buffer N3 was subsequently added to stop the lysis reaction and cell debris was removed by centrifugation at 14,000 xg for 10 minutes at room temperature in a tabletop centrifuge. The supernatant was then transferred to a QIAprep spin column and centrifuged at 14,000 xg for 1 minute. Columns were then washed with 750µl of buffer PE and centrifuged at 14,000 xg for 1 minute. Excess ethanol was removed by a further centrifugation at 14,000 xg for 1 minute before pDNA was eluted into sterile 1.5ml eppendorf tubes by adding 50 µl of elution buffer to the column and centrifuging at 14,000 xg for 1 minute. Purified pDNA concentration was then quantified using a BioPhotometer plus 6132 (Eppendorf) and was stored at -20°C.

2.2.2.2 Polymerase chain reaction (PCR)

Specifically designed forward and reverse oligonucleotide primers were used for the amplification of DNA sequences of interest and for the linearisation of pDNA (see Appendix I for a list of primers). Primers were designed to contain appropriate restriction sites and, where necessary, start and stop codons. All PCR reactions were carried out in a GeneAmp 9700 thermocycler (Applied Biosystems). A standard reaction setup for high fidelity PCR reactions is shown in Table 2.1 and standard temperature-cycling conditions are shown in Table 2.2. An annealing temperature of 5°C below the salt-adjusted melting temperature of the primers was generally chosen, and an extension time of 30 seconds per kbp was used.

Table 2.1 High fidelity (KOD) PCR reaction setup

Reaction component	Stock concentration	Volume (µl)
Template DNA	100 nm μl^{-1}	1
Forward primer	10 µM	1
Reverse primer	10 µM	1
dNTPs	2 mM (each)	5
KOD reaction buffer	10x	5
$MgSO_4$	25 mM	5
KOD Hot Start DNA polymerase	1 unit μl^{-1}	0.5
dH_2O	-	31.5
Total		50

Table 2.2 High fidelity (KOD) PCR thermocycler conditions

Step	Temperature (°C)	Time (s)
1. Initial denaturation	95	30
2. Denaturation	95	30
3. Annealing	Variable	30
4. Extension	72	30 seconds/kb
Repeat steps 2-4 35 times		
5. Final extension	72	600

2.2.2.3 Agarose gel electrophoresis

A 1.5% (w/v) agarose solution was prepared by dissolving agarose powder in 1x TAE buffer by heating in a microwave. Once the solution had cooled to ~60°C ethidium bromide was added to a final concentration of 0.5 µg ml^{-1} and gels were poured into a mould containing a loading comb. Once gels had set, the comb was removed, and gels were placed in electrophoresis tanks submerged in 1x TAE buffer. 6x agarose gel loading buffer was added to DNA samples before

they were loaded into the wells at an appropriate volume. A 1 kb plus DNA ladder (Invitrogen) was included alongside samples for size comparison. Electrophoresis was carried out at 120 volts (V) for ~40 minutes before the gel was visualised using a UV transilluminator (Bio Rad).

2.2.2.4 Gel extraction

DNA bands were purified from agarose gels using the PureLink Quick Gel Extraction/PCR Purification Kit (Invitrogen). Bands visible on a UV light box were excised from the gel using a razor blade and were placed into sterile 2 ml eppendorf tubes. These samples were then weighed and 3 gel volumes of buffer L3 were added before tubes were placed in a heat block set to 65°C. Tubes were mixed every 3 minutes until the gel was completely dissolved. Isopropanol (1 gel volume) was then added before samples were transferred to PureLink spin columns and centrifuged at 14,000 xg for 1 minute at room temperature. Column flow-through was discarded and columns were washed with 750 µl of buffer W1 and centrifuged at 14,000 xg for 1 minute. Flow-through was discarded and columns were spun for 3 minutes at 14,000 xg to remove residual ethanol. DNA was then eluted into sterile eppendorf tubes following the addition of 35 µl of buffer E1 and centrifugation at 14,000 xg for 1 minute. DNA products were then quantified using a BioPhotometer plus 6132 (Eppendorf) and were stored at -20°C.

2.2.2.5 Restriction endonuclease digestion

Restriction digests of pDNA or PCR products were setup as described in Table 2.3 and incubated for 1 hour at 37°C to ensure complete digestion. Reactions were then purified using the PureLink Quick Gel Extraction/PCR Purification Kit as described previously with the following alterations. Buffer B1 was added to reactions at 3x per reaction volume prior to loading on spin columns.

Table 2.3 Restriction endonuclease digestion reaction setup

Component	Stock concentration	Volume (µl)
DNA	Variable	1-25
Restriction enzyme(s)	20 units μl^{-1}	1 each
CutSmart buffer	10x	3
dH_2O	-	0-25
Total		30

2.2.2.6 T4 ligase reactions

Restriction digested DNA (both pDNA and purified PCR fragments) were placed into ligation reactions to allow the insertion of the gene of interest downstream of the promoter sequence. Reactions were incubated for either 1 hour at 25°C or at 4°C overnight. A standard reaction setup can be found in Table 2.4. 5 µl of the completed ligation reaction was then used to transform chemically competent Top-10 cells as described previously. Single colonies were selected and grown overnight for pDNA purification which was sent for DNA sequencing to confirm correct assembly.

Table 2.4 T4 ligation reaction setup

Component	Stock concentration	Volume (µl)
linearise pDNA	50 ng μl^{-1}	1
Insert DNA	Variable	Variable
T4 DNA ligase buffer	10X	2
T4 DNA ligase	400 units μl^{-1}	1
dH_2O	-	Up to 20 µl
Total		20

2.2.2.7 Gibson assembly

Overlapping primers for Gibson assembly were designed using Benchling's Gibson assembly tools. Standard PCR reactions were then carried out to amplify the template DNA and insert DNA with homologous 20 bp overhangs. Following PCR purification or gel extraction and quantification of DNA concentration, these fragments were assembled together with the Gibson assembly master mix (New England Biolabs) and incubated at 50°C for 30 minutes. Gibson assembly reaction setup is shown in Table 2.5. 2 µl of completed Gibson assembly reaction were used to transform Top-10 cells. Colonies were then selected and grown for pDNA isolation following which they were sent for DNA sequencing analysis.

Table 2.5 Gibson assembly reaction setup

Component	Stock concentration	Volume (µl)
Vector DNA	Variable	Variable (50 ng)
Insert DNA	Variable	Variable (150 ng)
Gibson assembly master mix	2x	5
dH_2O	-	Up to 10 µl
Total		10

2.2.2.8 Site directed mutagenesis

Overlapping primers for site-directed mutagenesis were designed using Agilent's QuikChange primer design tool. Primers were added to a standard high-fidelity PCR reaction as outlined before. Following completion of the PCR reaction cycles, reactions were treated with DpnI for 2 hours at 37°C to digest methylated template DNA. 5 µl of DpnI treated reaction was then used to transform Top-10 cells and successful transformants were grown for pDNA isolation and DNA sequencing analysis as before.

2.2.3 General biochemical techniques

2.2.3.1 EGFP-His purification by NTA(Ni) affinity chromatography

E. coli strain BL21 (DE3) were transformed with pDNA (pET21c(+)) containing a C-terminal histidine-tagged enhanced green fluorescent protein (EGFP). Following an initial preculture, cells were grown shaking at 37°C in LB media supplemented with 200 µg ml^{-1} ampicillin to an OD_{600} = 1. Protein expression was then induced by the addition of isopropyl-ß-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cultures were incubated shaking at 37°C for 3 hours to induce protein expression. Cells were pelleted by centrifugation at 4,000 rpm for 30 minutes at 4°C and were resuspended in EGFP re-suspension

buffer supplemented with 0.5 mg ml^{-1} lysozyme and 10 µg ml^{-1} DNase I. Cells were then lysed on ice by three rounds of sonication at 23 kHz using a soniprep 150 tip sonicator (MSE) and insoluble material was removed by centrifugation at 4,000 rpm for 30 minutes at 4°C. Cleared lysates were then applied to a pre-packed NTA(Ni) column equilibrated with EGFP re-suspension buffer without lysozyme or DNase. The column was then washed with EGFP wash buffer before samples were eluted with EGFP elution buffer. Pooled elution fractions were dialyzed twice for 2 hours against 5 L of EGFP dialysis buffer at 4°C to remove residual imidazole. Protein concentration was then measured by the Bradford assay and purified EGFP was concentrated to 20 µM using 10,000 MWCO concentration columns (Generon).

2.2.3.2 70S ribosome purification from *E. coli* strain JE28 by NTA(Ni) affinity chromatography

E. coli strain JE28 was used for isolation of tetra-histidine-tagged 70S ribosomes using a previously established protocol (Ederth et al., 2009). Briefly, following an initial preculture, JE28 cultures were grown shaking at 37°C in LB media supplemented with 50 µg ml^{-1} kanamycin. At OD_{600} = 1 cultures were slowly cooled to 4°C. Cells were then pelleted by centrifugation at 4,000 rpm for 30 minutes at 4°C and re-suspended in ribosome re-suspension buffer. Cell suspensions were further lysed by three rounds of sonication at 23 kHz using a tip sonicator as before. Lysates were subjected to two rounds of centrifugation at 20,000 xg for 30 minutes at 4°C to remove insoluble material. Cleared lysates were then applied to a pre-packed NTA(Ni) column equilibrated with ribosome re-suspension buffer without lysozyme or DNase. The column was washed with ribosome wash buffer before samples were eluted with ribosome elution buffer. Pooled elution fractions were dialyzed twice for 2 hours against 5 L of ribosome dialysis buffer at 4°C to remove residual imidazole. At this stage ribosomes were either pelleted by ultracentrifugation through a 30% sucrose cushion (cushion buffer) at 130,000 xg for 16 hours at 4°C and re-suspended in polymix buffer, or were fluorescently labelled (see below). Ribosome concentration was determined using the Beer-Lambert law by measuring absorbance at 260 nm and using an extinction coefficient of 3.91 × 107 M^{-1} cm^{-1} (Becker et al., 2012). Ribosomes were further concentrated to 20 µM using 10,000 MWCO concentration columns (BioVision).

2.2.3.3 Fluorescent labelling of ribosomal proteins

Alexa Fluor 488 5-SDP ester (Thermo-Fisher Scientific) was added to 5 ml imidazole-free ribosome solutions at a final concentration of 50 μg ml^{-1}. The resulting solution was gently mixed at room temperature for one hour and then dialyzed against 5 L of dye-free solution for 2 hours at 4°C. This solution was then ultracentrifuged through a 30% sucrose cushion (cushion buffer) at 130,000 xg for 16 hours at 4°C to remove residual dye. The resulting pellet containing fluorescently labelled ribosomes was re-suspended in polymix buffer and concentrated to 20 μM as before.

2.2.3.4 SDS-PAGE

Polyacrylamide separating gels were prepared at the relevant percentage by dilution of 30% (w/w) acrylamide/bis-acrylamide solution and 4x SDS separating buffer. Ammonium persulfate (APS) was added to a final concentration of 0.067% (w/v) from a 10% (w/v) stock solution and N,N,N',N'-tetramethylethane-1,2-diamine (TEMED) was added neat at a ratio of 1:1500. Separating gels were then cast using standard 1 mm separation glass plates in Bio Rad gel casting cassettes. Gels were overlaid with 0.1% (w/v) SDS and allowed to set at room temperature. Stacking gels were prepared at 3.9% acrylamide using 4x SDS stacking buffer and, following the addition of APS and TEMED as above, were cast on top of separating gels. A sample comb was inserted immediately after casting and the gels were left to set at room temperature. Protein containing samples were prepared by adding the relevant volume of 4x SDS sample buffer and heating at 65°C for 15 minutes before applying to gels. Gels were run at 75 V for ~20 minutes until the sample had migrated out of the stacking gel, at which point the voltage was increased to 120 V to complete sample separation. Upon completion, the stacking gel was removed, and gels were visualised for BODIPY-FL fluorescence using a Typhoon FLA 9000 (GE Healthcare) with an excitation wavelength of 473 nm and a long-pass emission filter for detection of wavelengths > 510 nm. Following this, gels were recovered and stained with coomassie quick stain (Generon).

2.2.4 Liposome preparation

2.2.4.1 Preparation of lipid stocks

Powdered lipids (Avanti Polar Lipids) were dissolved in chloroform to the relevant stock concentrations, transferred to 7 ml glass vials (Supelco), sealed with parafilm and stored at -20°C until use. Lipid films were generated by transferring stock lipid solutions to a fresh glass vial and evaporating the solvent under a stream of N_2 gas. Following this, vials were further dried under vacuum for 1 hour to remove any residual chloroform. Dried lipid films could then be dissolved in mineral oil for GUV experiments or hydrated with extrusion buffer for the generation of LUVs.

2.2.4.2 LUV generation by extrusion

Extrusion buffer was added to dried lipid films for a final lipid concentration of 10 mg ml^{-1}. The mixture was then vortexed vigorously for 1 minute to generate multilamellar vesicles (MLVs). Using a mini-extruder (Avanti Polar Lipids), MLV solutions were passed 11 times through a 200 nm polycarbonate membrane followed by 13 passes of the same solution through a 100 nm polycarbonate membrane. The resulting solution of LUVs was then used immediately for experiments or was stored at 4°C for no longer than a week.

2.2.4.3 GUV generation by droplet transfer

GUVs were generated using the droplet transfer method as previously described (Noireaux and Libchaber, 2004; Altamura et al., 2017). Briefly, an aqueous/lipid interface was generated by gently layering 300 µl of 0.5 mM POPC (Avanti Polar Lipids) solubilized in light mineral oil on top of an outer aqueous solution supplemented with 200 mM glucose. The resulting two-phase solution was left at room temperature for 2 hours to allow saturation of the interface with POPC molecules. Meanwhile, 20 µl of an inner solution supplemented with 200 mM sucrose was prepared and transferred into 600 µl of 0.5 mM POPC in light mineral oil. This solution was then emulsified by pipetting for 30 seconds and gently layered on top of the previously prepared, now lipid saturated, aqueous/lipid interface. The entire mixture was then centrifuged at 2,500 xg for 10 minutes at room temperature and pelleted GUVs were collected using a pipette following the careful removal of the oil phase.

For cell-free reactions, the inner solution was composed of the PURE cell-free reaction mixture supplemented with 200 mM sucrose, 1 µl murine RNase inhibitor (40,000 units ml^{-1}) (New England Biolabs) and 20 ng μl^{-1} of the relevant pDNA. Additionally, for cell-free reactions with His-tagged ribosomes, the supplied ribosome solution was omitted and 1 µM of purified His-tagged ribosomes was added. PURE feed solution was osmotically matched to the inner solution with NaCl (typically 75 mM) using a Vapro 5520 vapour pressure osmometer (Wescor). For all experiments with PR, all-trans-retinal was added to a final concentration of 100µM and reactions were incubated in the dark.

For ribosome binding experiments the inner solution was composed of 100 mM potassium glutamate, 18 mM magnesium acetate, 50 mM HEPES-KOH pH7.6, 1 µM 70S ribosomes and 200 mM sucrose. The outer solution was composed of 100 mM potassium glutamate, 18 mM magnesium acetate, 50 mM HEPES-KOH pH7.6 and 200 mM glucose and was osmotically matched to the inner solution using NaCl as before.

2.2.5 *In vitro* transcription and translation

De novo protein synthesis was performed using circular pDNA with the PUREfrex version 1.0 recombinant transcription-translation kit (Genefrontier). Reactions were prepared as in Table 2.6 in sterile 200 µl PCR tubes. For GUV experiments PURE reactions were supplemented with 200 mM sucrose. For bulk reactions for SDS-PAGE analysis, fluorotect in-gel fluorescence detection reagent (Promega) was added at a ratio of 0.6:20 and LUVs were added at a final concentration of 2 mg ml^{-1}.

Table 2.6 Standard PURExpress reaction setup

Component	Stock concentration	Volume (µl)
Solution 1	2x	10
Solution 2	20x	1
Solution 3	20 µM	1
RNase inhibitor	40,000 units ml^{-1}	1
pDNA	100 ng μl^{-1}	4
dH_2O	-	Up to 20 µl
Total		20

For mRNA localization experiments, ribosomes were omitted from the reaction mixture to avoid translation of mRNAs and 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DHFBI) was added to a final concentration of 20 µM. PURE reactions were incubated at 37°C for one hour to allow high levels of mRNA transcription.

2.2.6 Confocal microscopy

All experiments were performed using a Leica TCS SP8 laser scanning confocal microscope equipped with a 20x objective and a 63x oil immersion objective. Samples were placed in 8-well polymer µ-slide microscopy chambers (Ibidi) and were allowed to settle to the bottom of these chambers for 10 minutes prior to image acquisition.

2.2.7 Membrane protein insertion assays

2.2.7.1 Ratiometric fluorescence assay using nile red

GUVs collected following synthesis of the relevant protein using the PURE system were re-suspended in fresh PURE feed solution supplemented with 200 mM glucose and 0.1 µM nile red. GUVs were then placed into 8-well live cell imaging

polymer microscopy chambers (Ibidi) and were visualised using confocal microscopy. A 488 nm argon laser was used to excite nile red and emission was collected using band-pass filters for 510 - 590 nm and for 650 – 750 nm. Quantification of ratiometric data was performed using Fiji software.

2.2.7.2 Hemagglutinin epitope surface display assay

GUVs collected following synthesis of the relevant protein using the PURE system were re-suspended in fresh PURE feed solution supplemented with 200 mM glucose and 5 μg ml^{-1} of Alexa Fluor 647 or 488-conjugated HA antibody. Solutions were mixed gently by pipetting and were incubated at 37°C for 30 minutes. GUVs were then pelleted by centrifugation at 2,500 xg for 10 minutes at room temperature, antibody-containing solution was carefully removed, and GUVs were re-suspended in fresh PURE feed solution supplemented with 200 mM glucose. Following this, GUVs were placed into 8-well live cell imaging polymer microscopy chambers (Ibidi) and were analysed using confocal microscopy.

2.2.8 Data analysis

2.2.8.1 Image analysis

For all microscopy imaging data, image analysis was performed using Fiji, version 1.52i. The Radial Profile Angle plugin was used for all radial fluorescence data and standard histograms and plot profiles were used for all other data. Three-dimensional surface plots were generated using a fire LUT to enhance the visualisation of samples with low fluorescence emission. An example of radial fluorescence profiling is shown in Figure 2.1.

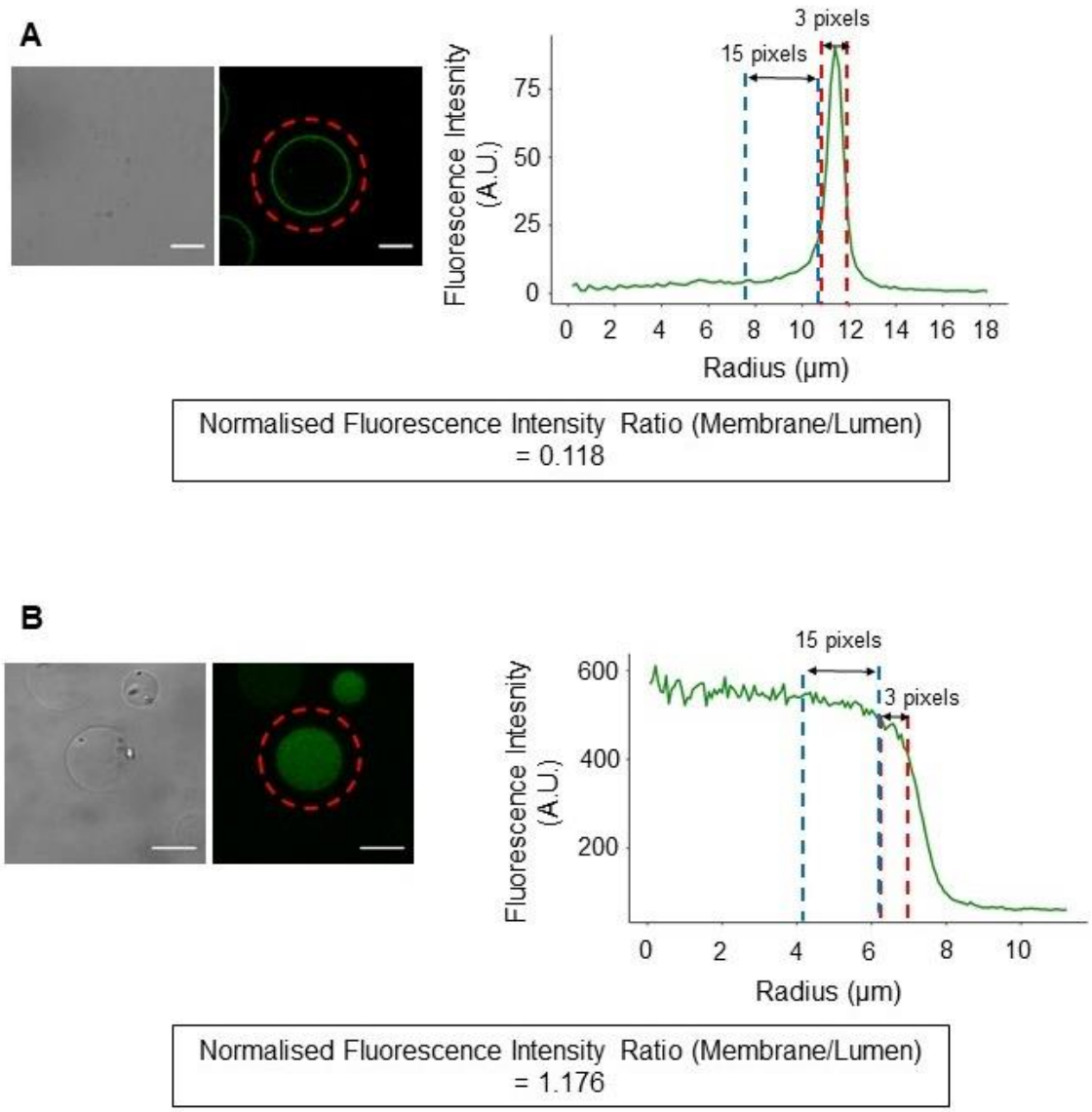


Figure 2.1 Examples of radial profile fluorescence analysis.

(A) Representative confocal microscopy image of a GUV with membrane localised fluorescence. The red circle indicates the region of interest used for radial profiling. Scale bar is 10 µm. The corresponding line plot shows the radial fluorescence profile. Membrane fluorescence was determined by calculating an average of three pixels with the highest signal being the middle pixel. Lumen fluorescence was determined by taking an average of the subsequent 15 pixels moving into the vesicle lumen. Values were normalised to background fluorescence measured as an average of three arbitrary points on the image. **(B)** As in (A) but for a GUV exhibiting lumenal fluorescence.

2.2.8.2 Statistical analysis

Statistical analyses were carried out using R version 3.4.3. Data are presented as the mean ± sample standard deviation and statistical significance was tested

using student's t-test or one-way ANOVA with Tukey (HSD) post-hoc analysis with a significance threshold of $p < 0.001$ unless otherwise stated.

3. *De novo* synthesis of membrane proteins in a minimal environment

3.1 Introduction

Polytopic, alpha-helical membrane proteins are essential for cellular life and play key roles in many fundamental processes such as energy generation, molecular transport and responses to environmental stimuli. In order to perform these essential functions, proteins must be correctly inserted into the amphiphilic environment of the lipid bilayer and adopt the correct tertiary structure. *In vivo,* this biogenesis process involves a large number of chaperones and various membrane spanning insertase proteins (Alder and Johnson, 2004; Luirink et al., 2012).

It has long been argued that the final adopted structure of α-helical membrane proteins represents the most energetically favourable state (White and Wimley, 1999), and that once embedded within the membrane, these proteins will spontaneously adopt their native fold through helical interactions leading to a state of thermodynamic equilibrium (Popot and Engelman, 1990; Popot and Engelman, 2000).

The major thermodynamic hurdle that a membrane protein needs to overcome is therefore the initial insertion of hydrophobic helices into the membrane and not the subsequent folding. It has been shown that, in the vast majority of cases, this insertion occurs in a co-translational manner, most likely to avoid unwanted aggregation of hydrophobic protein domains in the cytoplasm (Cymer and von Heijne, 2013). The Sec-type translocon and other insertases such as the YidC/Oxa1/Alb3 family have evolved specifically to aid this process (Do et al., 1996; Reid and Nicchitta, 2012).

The direct insertion of hydrophobic transmembrane helices into the lipid bilayer without the aid of the translocon is thermodynamically feasible when enough hydrogen bonded helical structure is present to overcome the prohibitive energy barrier of partitioning non-hydrogen bonded peptides into the membrane (Almeida et al., 2012). Simulations have shown that helical structure is first adopted when hydrophobic peptides are adsorbed to the bilayer interface region, and that this structure subsequently partitions into the bilayer

(Ulmschneider et al., 2011). Interestingly, the free energy of partitioning these segments without the translocon closely mimics experimentally determined, Sec-dependent free energy costs (Ulmschneider et al., 2014).

The development of 'cell-free' protein expression systems has provided a means to probe the fundamental requirements for co-translational membrane protein insertion and folding *in vitro*. Initial work focussed on the synthesis of membrane proteins using lysate-based cell-free synthesis systems coupled with either ER-derived microsomes, which contain all of the translocation machinery for protein insertion, or empty liposomes containing no translocation machinery (Rosenberg and East, 1992; Kalmbach et al., 2007). Lysate-based systems contain all of the cytoplasmic components of the parent cell from which they are derived, and as such contain many chaperones and proteins of unknown function which may affect membrane protein integration and folding. The development of the protein synthesis using recombinant elements (PURE) system, an *E. coli*-based system containing purified components, provided an excellent tool for testing the fundamental requirements for membrane protein insertion and folding in the absence of any chaperoning elements (Kuruma et al., 2005). This system has been used in conjunction with empty liposomes in order to analyse the direct integration and folding of a number of polytopic α-helical proteins (Moritani et al., 2010; Soga et al., 2014; Harris et al., 2017; Berhanu et al., 2019).

Giant unilamellar vesicles (GUVs) have become a popular membrane mimetic system due to their cell-mimicking size (>1μm) and thus the possibility of using them as simplified cell models. Their size also makes them suitable for single vesicle analysis using optical techniques such as microscopy (Soga et al., 2014; Furusato et al., 2018). Encapsulation of various chemical and biological reactions within GUVs has also grown in popularity and synthesis of membrane proteins within these cell-mimics can provide a more realistic environment to study targeting and folding. The method most favoured for high encapsulation efficiency is the droplet transfer approach (Noireaux and Libchaber, 2004).

The aim of the present chapter was to use the PURE system for the cell-free synthesis of two polytopic α-helical membrane proteins, proteorhodopsin (PR) and galactose permease (GalP), and to determine their membrane insertion propensity when synthesised within GUVs generated using the droplet transfer approach. Stable GUVs were therefore generated and tested using the PURE

system for the encapsulated synthesis of a soluble reporter (EGFP). Membrane proteins PR and GalP were then synthesised in bulk and within GUVs, and protein synthesis, localisation and insertion were probed using a number of fluorescence-based assays.

3.2 Results

3.2.1 Giant vesicle production using the droplet transfer approach

In order to investigate the localisation and insertion of membrane proteins in a cell-size mimicking, minimal environment, a method was required to generate stable giant unilamellar lipid vesicles (GUVs) allowing the encapsulation of solutions within the vesicle lumen. The droplet transfer method was developed with this purpose in mind (Noireaux and Libchaber, 2004). The process is outlined in Figure 3.1 A and can be broken down into a number of methodological steps. First a phospholipid saturated aqueous/oil interface was prepared where the lipids were solubilised in mineral oil and placed on top of an aqueous phase containing glucose. The amphiphilic nature of phospholipids led to the generation of a monolayer at the oil/water interface with the phosphate containing head-groups of the lipids localised to the aqueous phase, and the aliphatic lipid tails facing the oil. The generation of a fully saturated interface required incubation at room temperature for two hours. Near the end of this incubation process a separate solution was prepared which would comprise the inner solution encapsulated within the GUVs. A small volume of an inner solution containing sucrose was emulsified through repeated pipetting in a lipid-containing oil solution to generate lipid-stabilised water in oil droplets. These droplets were then placed on top of the previously prepared interface and were passed through it by centrifugation thanks to the molecular weight difference between the isosmotic glucose and sucrose solutions. GUVs could then be collected following removal of the oil phase and visualised using microscopy techniques. Figure 3.1 B shows a representative confocal microscopy image of GUVs containing a fluorescent lipid, 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(lissamine rhodamine B sulfonyl) (18:1 Liss Rhod PE) at 1 mol%, and encapsulating a solution containing 0.5 mM of the water soluble fluorescent indicator 8-Hydroxypyrene-1,3,6-trisulfonic acid (HPTS).

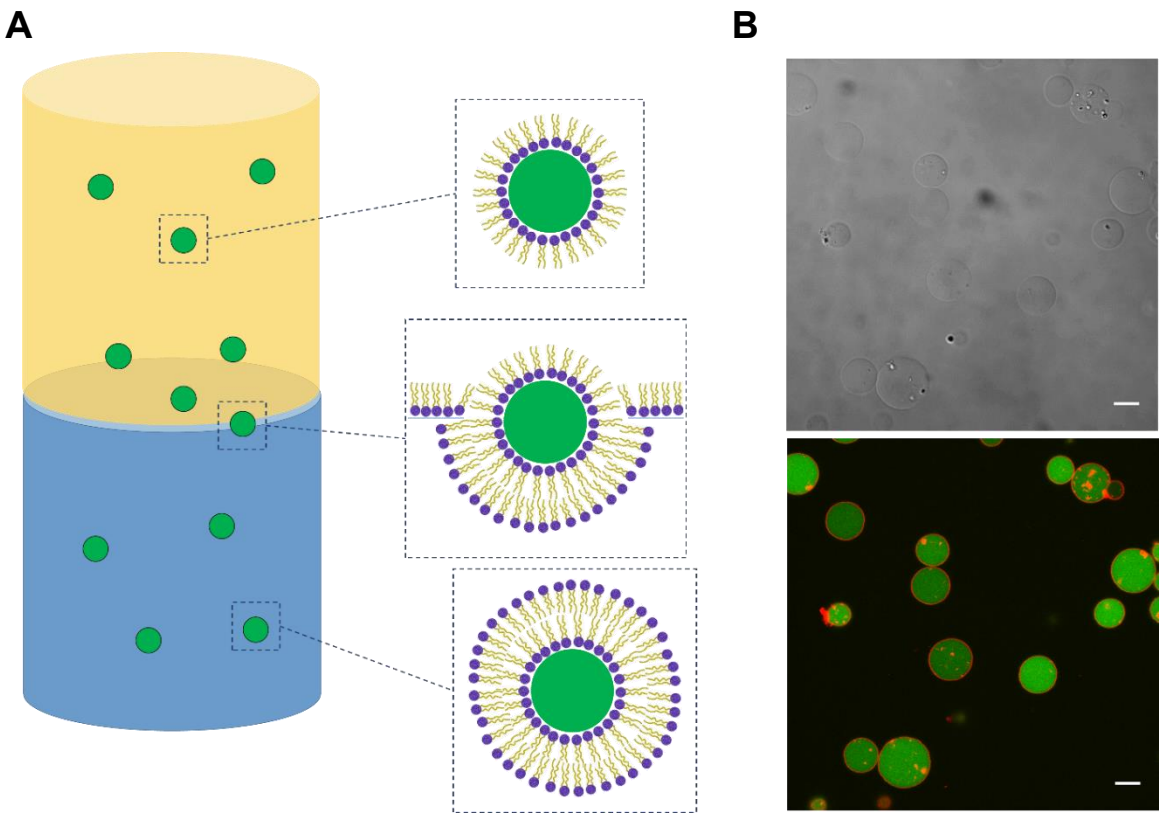


Figure 3.1 Production of giant unilamellar vesicles (GUVs) and encapsulation of solutions.

(A) Schematic representation of the droplet transfer process for the formation of GUVs (see methods for detailed process). Blue colour indicates the outer aqueous phase, yellow the oil/lipid phase and green the inner aqueous phase. Insets illustrate the process of passing a lipid supported oil/water emulsion through a pre-assembled lipid monolayer at a water/oil interface. **(B)** Representative confocal image of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) GUVs generated using this method containing 1 mol% 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(lissamine rhodamine B sulfonyl) (18:1 Liss Rhod PE) in the membrane (red channel) and 0.5 mM 8-Hydroxypyrene-1,3,6-trisulfonic acid (HPTS) in the inner aqueous phase (green channel). Images are bright field (top) and dual-channel fluorescence (bottom). Scale bars are 10 μm.

GUVs prepared using the above approach were tested for stability after a 24 hour incubation at room temperature. Figure 3.2 A shows representative microscopy images of GUVs encapsulating 0.5 mM HPTS at the time of production and after 24 hours of room temperature incubation. The lumenal fluorescence intensity of HPTS and the diameter of GUVs were measured and no statistically significant differences ($p > 0.05$) were observed for either parameter after the 24 hour incubation indicating that the vesicles were stable (Figure 3.2 B).

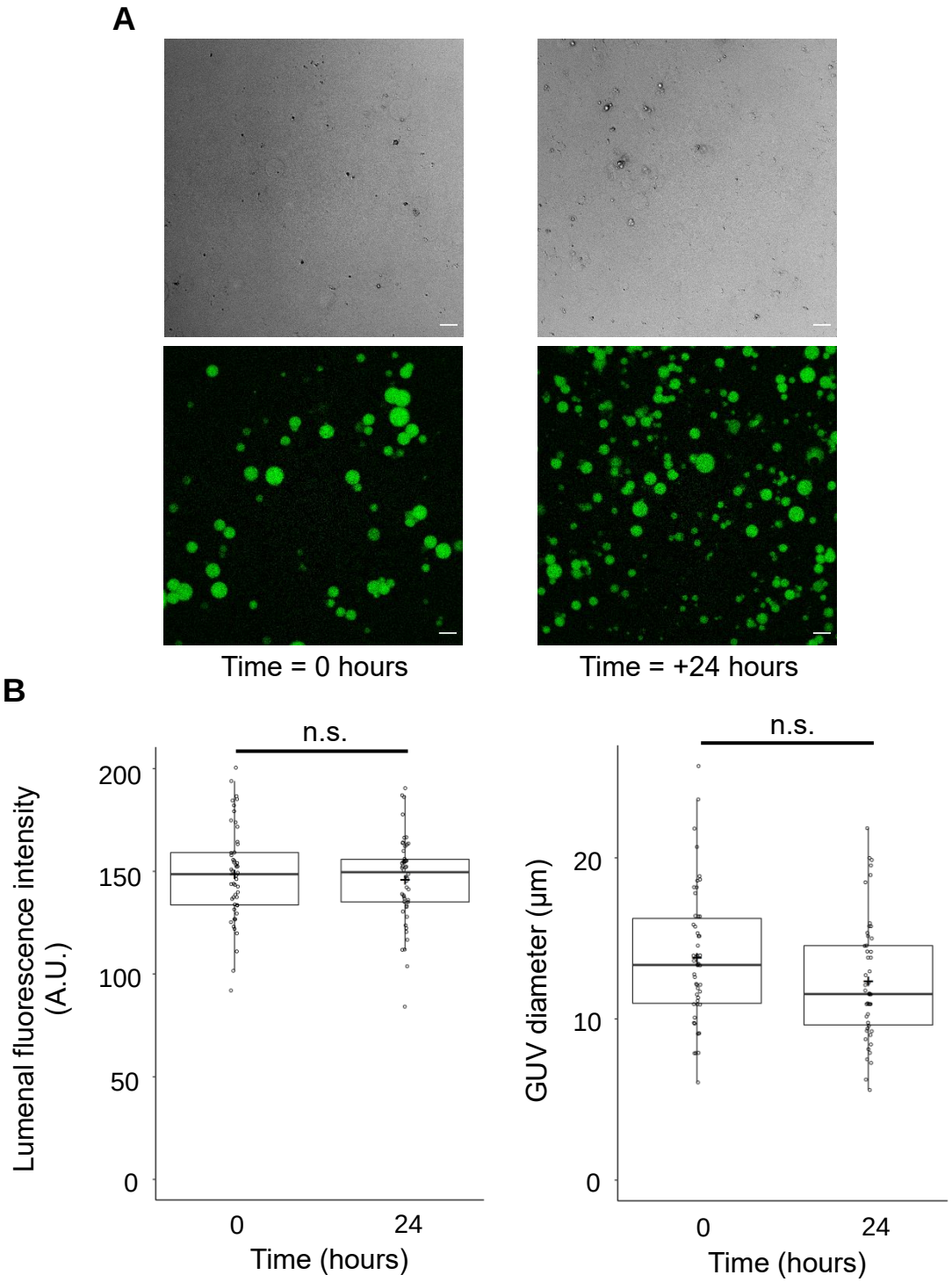


Figure 3.2 Stability of droplet transfer generated GUVs at room temperature.

(A) Representative confocal images of POPC GUVs encapsulating 0.5 mM HPTS taken immediately after generation and following incubation at room temperature for 24 hours. Images are brightfield (top) and fluorescence emission of HPTS following excitation with a 488 nm laser (bottom). Scale bars are 20 µm **(B)** Lumenal fluorescence intensity was measured from 50 individual GUVs and the corresponding diameter was recorded. All individual data points are shown along with the mean (+), median and interquartile range. Whiskers extend to 1.5x the interquartile range. Statistical significance was measured using a two-tailed, paired student's t-test with a significance threshold of $p < 0.05$.

3.2.2 Synthesis of a soluble protein within GUVs using the PURE system

The PURE system, a minimal cell-free protein synthesis system made from purified recombinant proteins from *Escherichia coli,* was selected for all experiments due to its simplicity and lack of chaperoning systems when compared to lysate-based systems. In order to test the suitability of this system for the synthesis of proteins directly from plasmid DNA (pDNA), the gene encoding enhanced green fluorescent protein (EGFP) was inserted into the supplied control expression vector in place of the DHFR control sequence (see section 2.1.3). This vector has been optimised for the expression of genes in the PURE system under the control of a T7 promoter, hence its selection as the expression vector for all subsequent PURE experiments. Initially, a bulk reaction was prepared and supplemented with 20ng μl^{-1} of EGFP pDNA and 0.6 µl of fluorotect reagent containing fluorescently labelled lysine residues for in-gel detection of translation products. Following a 3-hour incubation at 37°C the reaction was used directly for SDS-PAGE analysis on a 10% polyacrylamide gel. Figure 3.3 B shows the synthesis of a single protein band running at a molecular weight of ~35 kDa. A coomassie stain of the same gel is shown as a loading control.

Fluorescence analysis was also performed on bulk PURE reactions in order to determine whether the synthesised protein was indeed EGFP and whether it folded correctly in this minimal environment. Reactions were either supplemented with 20ng μl^{-1} of EGFP pDNA or with an equal quantity of nuclease-free H_2O to act as a negative control. Plates were incubated in a fluorescence plate reader at 37°C for 180 minutes with fluorescence emission measurements being taken every 5 minutes using a standard FITC filter set. EGFP fluorescence increased over time of incubation until a plateau was reached around 100 minutes into the experiment (Figure 3.3 C) indicating successful synthesis and folding of EGFP using the PURE system.

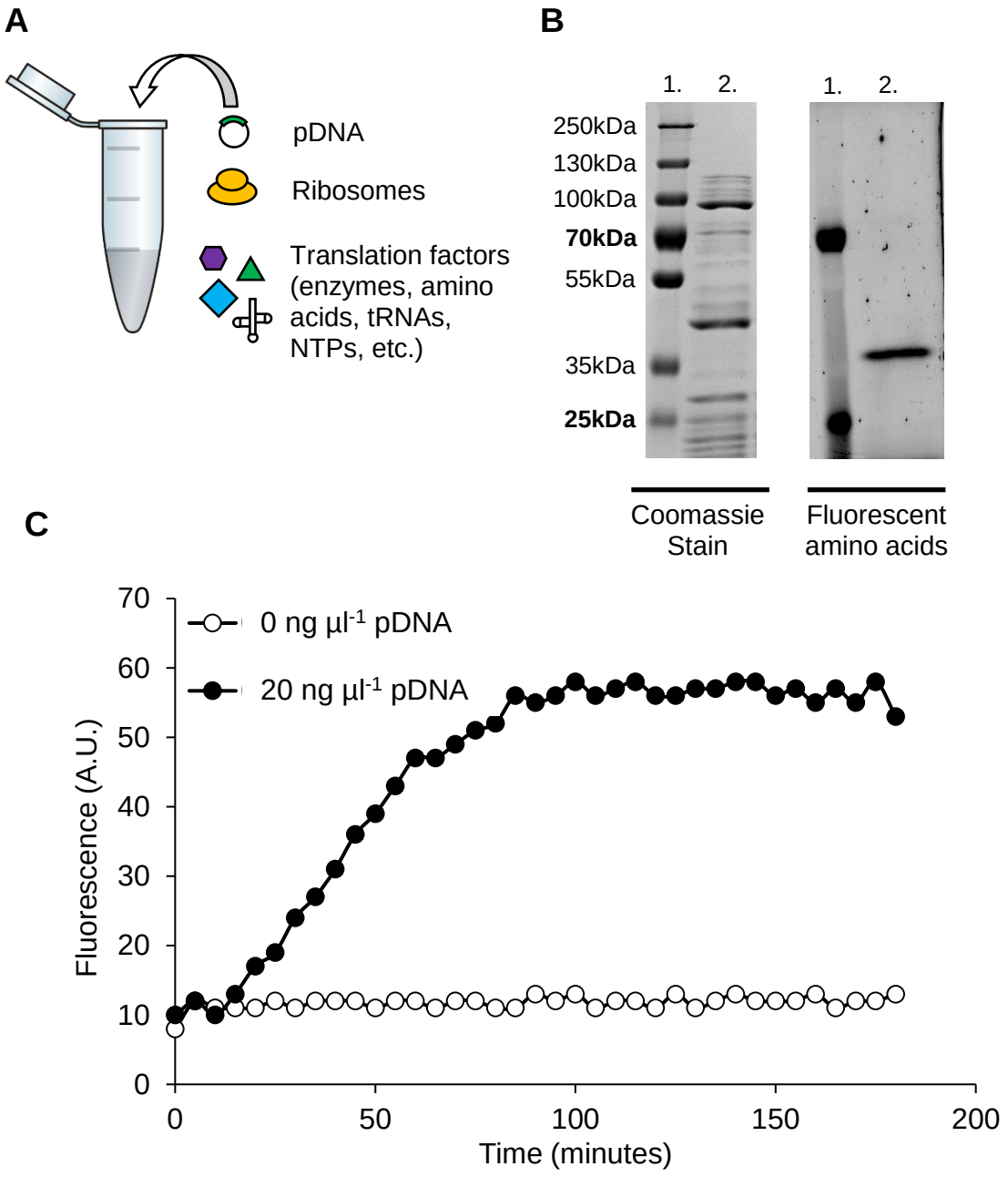


Figure 3.3 Cell-free synthesis and analysis of EGFP in bulk using the PURE system.

(A) Schematic showing the assembly of bulk PURE reactions for soluble protein synthesis. **(B)** SDS-PAGE analysis of a bulk PURE reaction supplemented with 20 ng μl^{-1} EGFP pDNA and fluorotect detection reagent. Images are of the same gel imaged for fluorescence and stained with coomassie quick stain. Lane 1 contains PageRuler Plus prestained molecular weight marker, lane 2 contains the PURE reaction. Fluorescent molecular weight bands are labelled in bold. **(C)** Generation of fluorescence from a bulk PURE reaction synthesising EGFP at 37°C measured using a plate reader with standard filter settings and compared to a control lacking pDNA.

As shown previously, the droplet transfer approach allows the encapsulation of user-defined solutions within the lumen of GUVs. PURE reactions supplemented with 200 mM sucrose were therefore used as the inner solution to generate GUVs for encapsulated protein synthesis (Figure 3.4 A). Following collection, PURE containing GUVs were incubated at 37°C for 3 hours before being analysed by confocal microscopy. EGFP fluorescence was detected in the majority of GUVs following incubation and no fluorescence was observed from control vesicles incubated without pDNA (Figure 3.4 B).

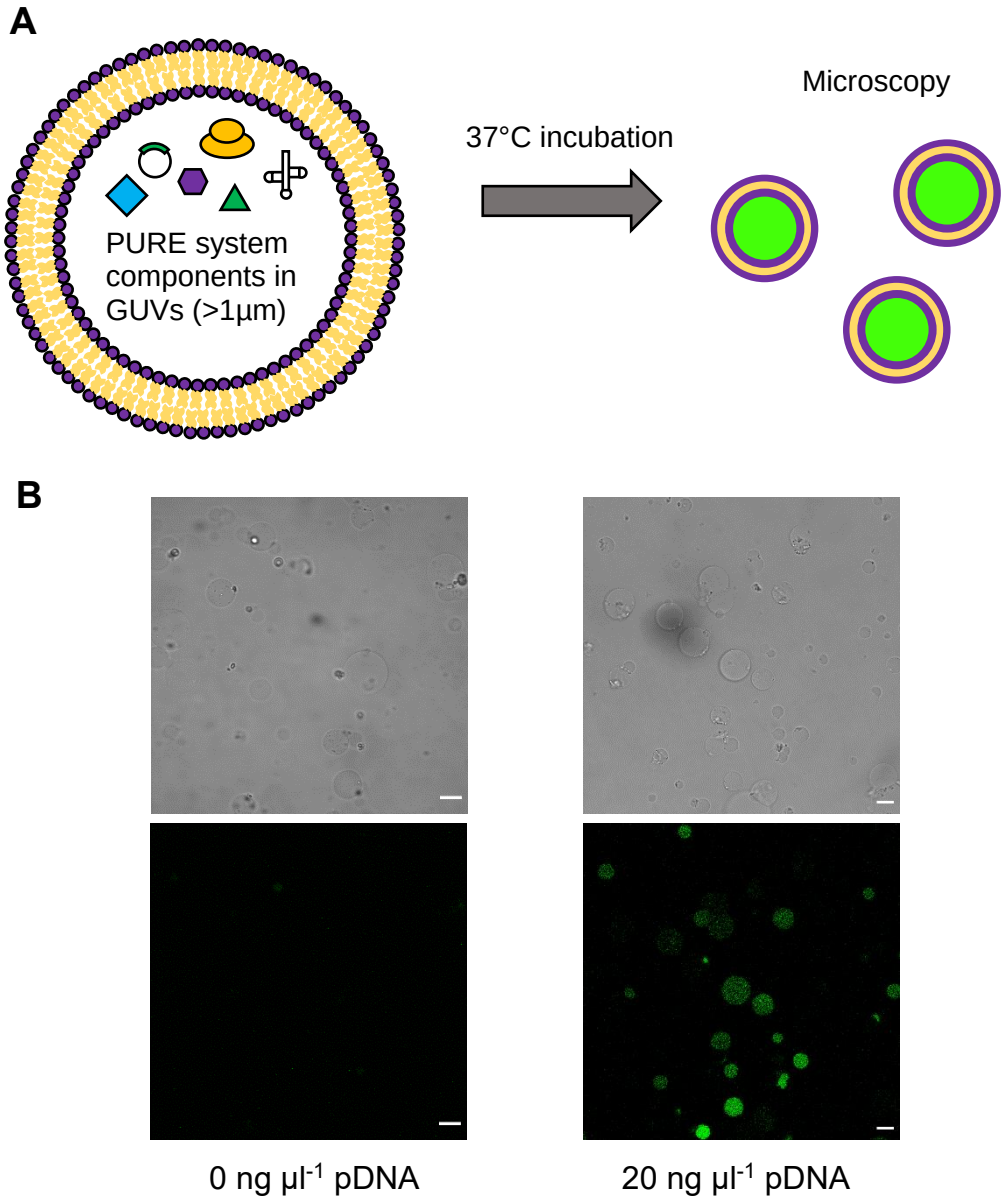


Figure 3.4 Cell-free synthesis of EGFP using GUV encapsulated PURE reactions.

(A) Schematic showing the encapsulation of PURE reactions inside GUVs. **(B)** Representative confocal images of GUVs encapsulating PURE reactions following a 3-hour incubation at 37°C without (left) and with (right) supplementation with 20ng μl^{-1} pDNA. Scale bars are 10 µm.

3.2.3 *De novo* synthesised membrane proteins localise to vesicle membranes

The successful synthesis of a soluble reporter using the PURE system, both in bulk and inside GUVs, indicated that these methods could be used to investigate the localisation and insertion of α-helical membrane proteins. The synthesis of such proteins in a minimal environment devoid of any chaperoning or insertion mediating systems would allow the identification of fundamental processes driving membrane integration which may be masked in the complex cellular environment. The α-helical proteins selected for analysis were the microbial proton pump proteorhodopsin (PR) (Beja et al., 2000) and the *E. coli* monosaccharide transporter galactose permease (GalP) (Zheng et al., 2010).

PR is a member of the highly conserved rhodopsin family of membrane proteins. It contains 7 transmembrane α-helices and a cleavable N-terminal hydrophobic extension signal forming a short eighth α-helix (Figure 3.5 A, B). In order to synthesise membrane proteins in the minimal environment provided by the PURE system, a suitable amphiphile must be supplied so that the hydrophobic transmembrane helices of the protein do not aggregate and lead to translational issues. For the analysis of translation products by SDS-PAGE (as seen in Figure 3.3) one solution is to supplement reactions with large unilamellar vesicles (LUVs), thus providing a bilayer environment for the proteins. LUVs were generated using the extrusion technique to produce vesicles with a diameter of ~100 nm (see section 2.2.4). These vesicles were added to PURE reactions at a concentration of 1 mg ml^{-1} prior to the addition of pDNA. Following a 3-hour incubation at 37°C as before, reactions were used directly for SDS-PAGE analysis on 10% polyacrylamide gels. Figure 3.5 C shows the synthesis of a protein band with a molecular weight of around 50 kDa corresponding to a PR-EGFP fusion product where EGFP was attached to the C-terminus of full-length PR.

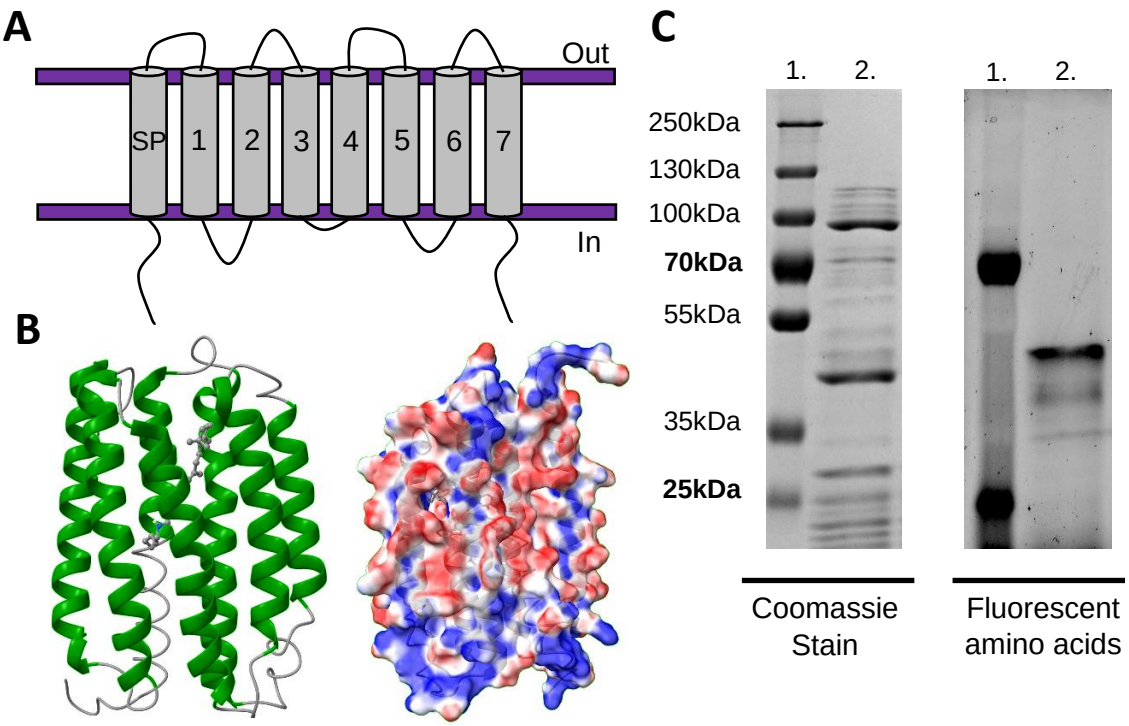


Figure 3.5 Bulk synthesis of a proteorhodopsin-EGFP fusion protein using the PURE system supplemented with large unilamellar vesicles.

(A) Cartoon representation of proteorhodopsin (PR) in a bilayer environment. SP is the signal peptide while numbers 1-7 represent the transmembrane helices. **(B)** Structural representation of PR (PDB entry 2L6X) with transmembrane helices in green. The right-hand image shows the derived protein surface coloured for hydrophobicity, with red representing highly hydrophobic portions and blue representing hydrophilic portions of the surface. **(C)** SDS-PAGE analysis of a bulk PURE reaction supplemented with 20ng μl^{-1} PR-EGFP pDNA, fluorotect detection reagent and 1 mg ml^{-1} of POPC LUVs. Images are of the same gel probed for fluorescence and stained with coomassie quick stain. Lane 1 contains PageRuler Plus prestained molecular weight marker, lane 2 contains the PURE reaction. Fluorescent molecular weight bands are labelled in bold.

The sugar transporter GalP is a member of the highly conserved major facilitator superfamily (MFS) of proteins and contains 12 transmembrane helices and, unlike PR, no N-terminal extension (Figure 3.6 A, B). Bulk PURE reactions were set up, as for PR, with the addition of LUVs. Following incubation as before, reactions were used directly for SDS-PAGE. Figure 3.6 C shows the synthesis single protein bands at ~40 kDa and ~70 kDa corresponding to native GalP and a GalP-mCherry fusion protein where mCherry was attached to the C-terminus of GalP.

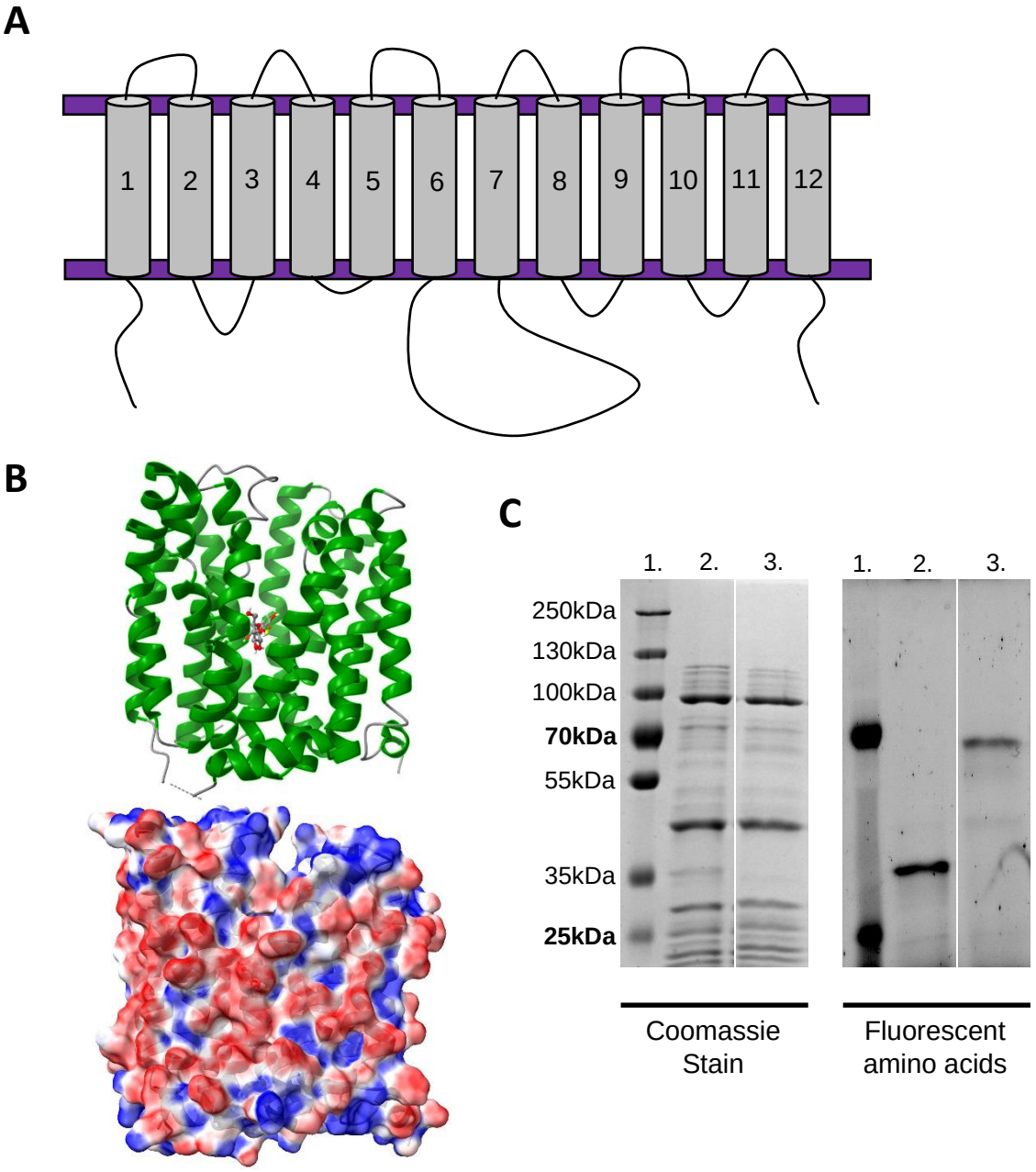


Figure 3.6 Bulk synthesis of GalP and a GalP-mCherry fusion protein using the PURE system supplemented with LUVs.

(A) Cartoon representation of GalP in a bilayer environment. Transmembrane helices are numbered 1 through 12. **(B)** Structural representation of the GalP related MFS transporter lacY from *E. coli* (PDB entry 1PV7) with transmembrane helices in green. The bottom image shows the derived protein surface coloured for hydrophobicity, with red representing highly hydrophobic portions and blue representing hydrophilic portions of the surface. **(C)** SDS-PAGE analysis of a bulk PURE reaction supplemented with 20ng μl^{-1} of relevant GalP containing pDNA, fluorotect detection reagent and 1 mg ml^{-1} of POPC LUVs. Images are of the same gel probed for fluorescence and stained with coomassie quick stain. Lane 1 contains PageRuler Plus prestained molecular weight marker, lane 2 contains a PURE reaction supplemented with GalP pDNA, and lane 3 contains a PURE reaction supplemented with GalP-mCherry pDNA. Fluorescent molecular weight bands are labelled in bold.

To investigate the localisation of the membrane proteins PR and GalP in a more relevant cell-mimetic environment, PURE reactions supplemented with either PR-EGFP or GalP-mCherry pDNA were encapsulated within GUVs. The large aqueous compartment within GUVs more accurately represents the hydrophilic challenge that membrane proteins faced due to increasing cell size early in evolution. GUVs encapsulating PURE reactions were again incubated for 3 hours at 37°C and were subsequently collected and visualised by confocal microscopy (Figure 3.7 A). Radial fluorescence profiles (see section 2.2.8) were generated for 20 individual GUVs following the synthesis of PR-EGFP and were compared to control ratios determined for vesicles synthesising soluble EGFP. PR-EGFP exhibited strong membrane localisation resulting in a mean lumen/membrane fluorescence ratio of 0.172 ± 0.034, significantly lower than the ratio determined for soluble EGFP of 1.154 ± 0.084 (Figure 3.7). This lower ratio was indicative of the observed membrane localisation of *de novo* synthesised PR.

A similar trend was observed for GUVs expressing GalP-mCherry, with membrane protein synthesising vesicles showing clear membrane localisation and resulting in a lumen/membrane ratio of 0.365 ± 0.082. This was again significantly lower than the ratio determined for EGFP expressing vesicles of 1.154 ± 0.084 (Figure 3.8). Again, the lower lumen/membrane ratio indicating that GalP was able to localise to the membranes of GUVs when synthesised *de novo* in a minimal system.

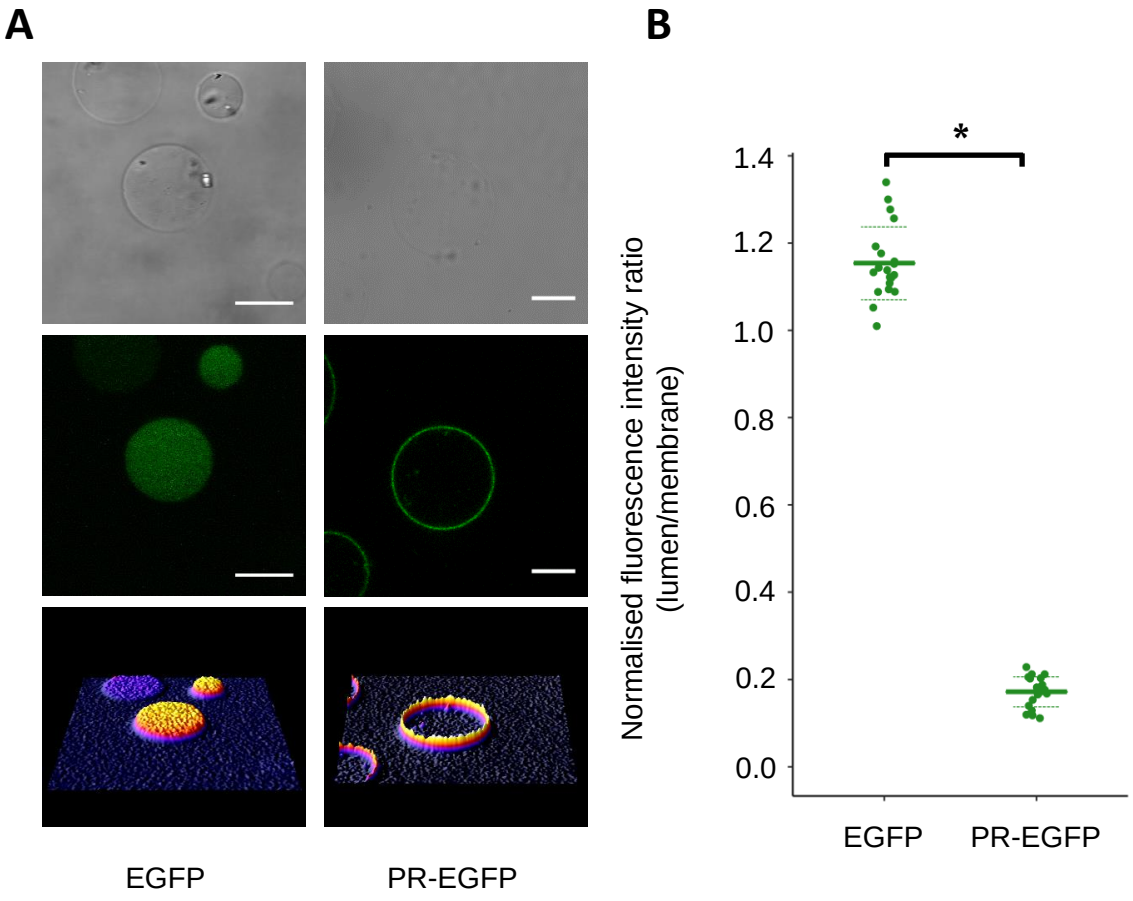


Figure 3.7 *De novo* synthesised proteorhodopsin localises to the membrane of GUVs.

(A) Representative confocal microscopy images of PURE encapsulating GUVs following synthesis of EGFP or PR-EGFP. Rows of images are; bright field, fluorescence emission and 3-dimensional representations of fluorescence from top to bottom respectively. Scale bars are 10 µm. **(B)** Radial profiles of EGFP fluorescence emission were analysed and are represented as a lumen/membrane fluorescence intensity ratio normalized to background. Data shown are individual values (circles), mean values (solid lines) and standard deviations (dashed lines) from 20 individual GUVs. * represents statistical significance ($p < 0.001$) determined using a student's two-tailed, unpaired t-test.

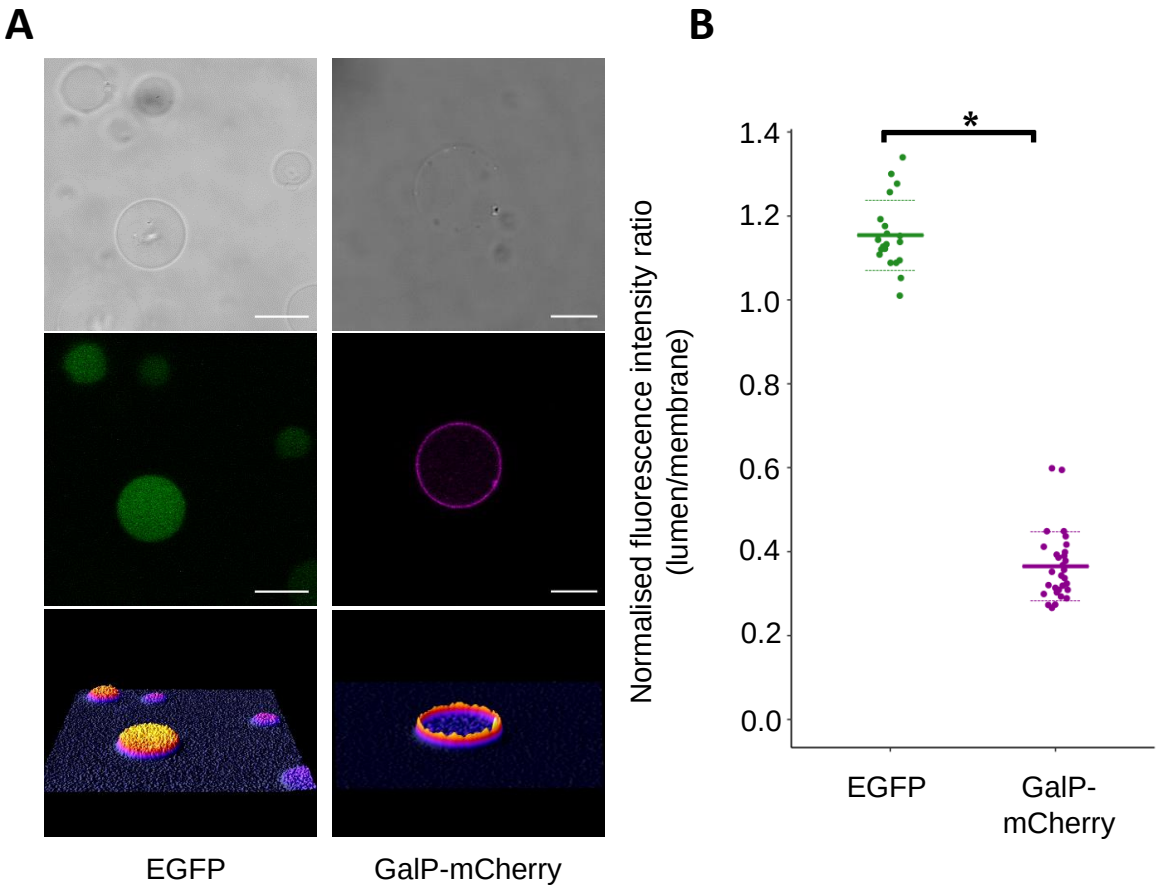


Figure 3.8 *De novo* synthesised galactose permease localises to the membrane of GUVs.

(A) Representative confocal microscopy images of PURE encapsulating GUVs following synthesis of EGFP or GalP-mCherry. Rows of images are; bright field, fluorescence emission and 3-dimensional representations of fluorescence from top to bottom respectively. Scale bars are 10 μm. **(B)** Radial profiles of EGFP and m-Cherry fluorescence emission were analysed and are represented as a lumen/membrane fluorescence intensity ratio normalized to background. Data shown are individual values (circles), mean values (solid lines) and standard deviations (dashed lines) from ≥ 20 individual GUVs. * represents statistical significance ($p < 0.001$) determined using a student's two-tailed, unpaired t-test.

3.2.4 Cell-free synthesised proteorhodopsin and galactose permease insert into empty lipid membranes

As both PR and GalP showed pronounced membrane localisation following *de novo* protein synthesis inside GUVs, it was decided to examine whether this localisation was accompanied by insertion of the proteins transmembrane α-helices into the bilayer. To answer this question the ratiometric fluorescent probe nile red was used. When supplied exogenously, this lipophilic dye inserts into the outer leaflet of bilayer membranes and its fluorescence emission properties are altered depending on the order of lipid tails surrounding the probe. It was hypothesised that the insertion of protein α-helices into an empty lipid membrane would increase the packing order of the lipids sufficiently to alter the emission properties of nile red.

Initially, control experiments were performed by generating GUVs with a defined concentration of cholesterol included in the membrane. Cholesterol is known to increase the order of lipid tail packing when included in a simple two component mixture such as the POPC:cholesterol mixture used here (Marsh, 2010), and thus lead to altered spectral characteristics of nile red (Figure 3.9 A). GUVs were generated using the droplet transfer approach with lipid solutions containing POPC and 0, 10, 20, 30 and 40 mol% of cholesterol. Following collection of vesicles, nile red was added at a concentration of 0.1 μM. GUVs were then gently mixed and allowed to settle to the bottom of microscope chambers before being visualised using confocal microscopy collecting dual emission data from 510 - 590 nm and from 650 - 750 nm. Dual channel images were then used to calculate a fluorescence ratio of emission from the red end of the spectrum to the yellow/green end of the spectrum for individual GUVs. Figure 3.9 shows that, as the concentration of cholesterol was increased above 10 mol% the fluorescence ratio was reduced indicating greater emission from the green region of the spectrum and thus higher packing order of the lipid tails.

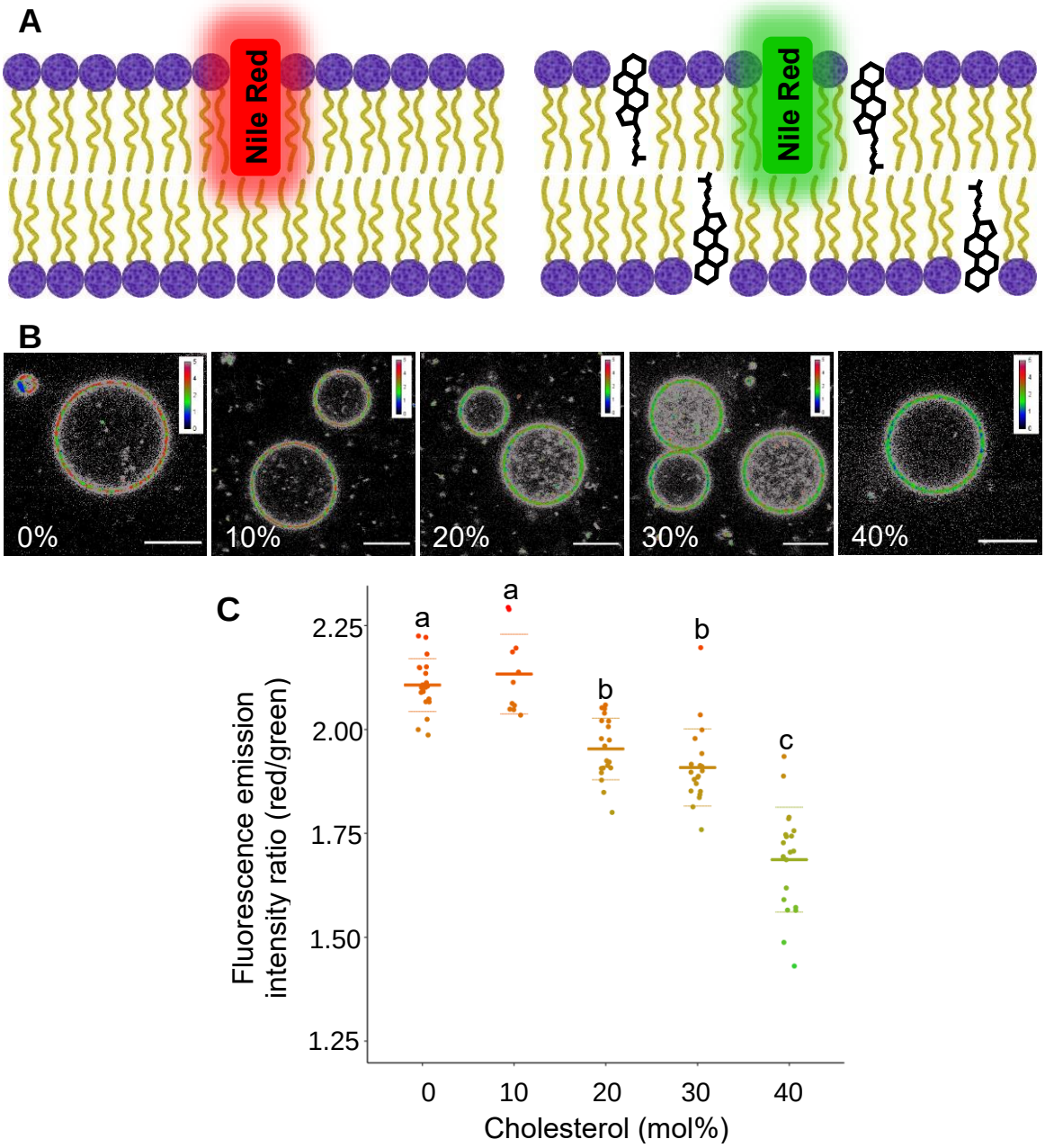


Figure 3.9 Nile red calibration in POPC:cholesterol GUVs.

(A) Schematic showing the nile red fluorescence emission response to increased cholesterol and thus lipid packing order. Loosely packed lipids without cholesterol led to probe emission being red-shifted. Increasing concentrations of cholesterol led to a blue-shift in the emission spectrum due to increased lipid tail order. **(B)** Representative confocal microscopy images of GUVs containing increasing quantities of cholesterol. Images were generated by dividing the red channel by the green channel and applying a rainbow LUT to aid ratio visualisation. Scale bars are 10 µm. **(C)** Quantification of the red/green ratio for ≥ 11 individual GUVs. Data presented are individual, mean and standard deviations. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

The above data indicated that nile red could be used as a reporter of outer leaflet lipid tail order in GUVs, at least when a large perturbation was created by cholesterol. Previous data had shown that PR localised to the membrane when synthesised *de novo*, therefore exogenously supplied nile red was used to report on alterations in lipid order at the outer leaflet of PR containing GUVs in order to determine whether the proteins transmembrane domains were able to penetrate the lipid bilayer. Control vesicles containing PURE reactions with no pDNA were analysed and lead to a nile red ratio of 2.036 ± 0.105. When pDNA encoding PR was included in the inner PURE solution, the ratio was significantly blue-shifted to 1.974 ± 0.130 (Figure 3.10). This drop in ratio suggested that the protein was inserted into the bilayer to such an extent that lipid packing at the outer leaflet became more ordered or constrained.

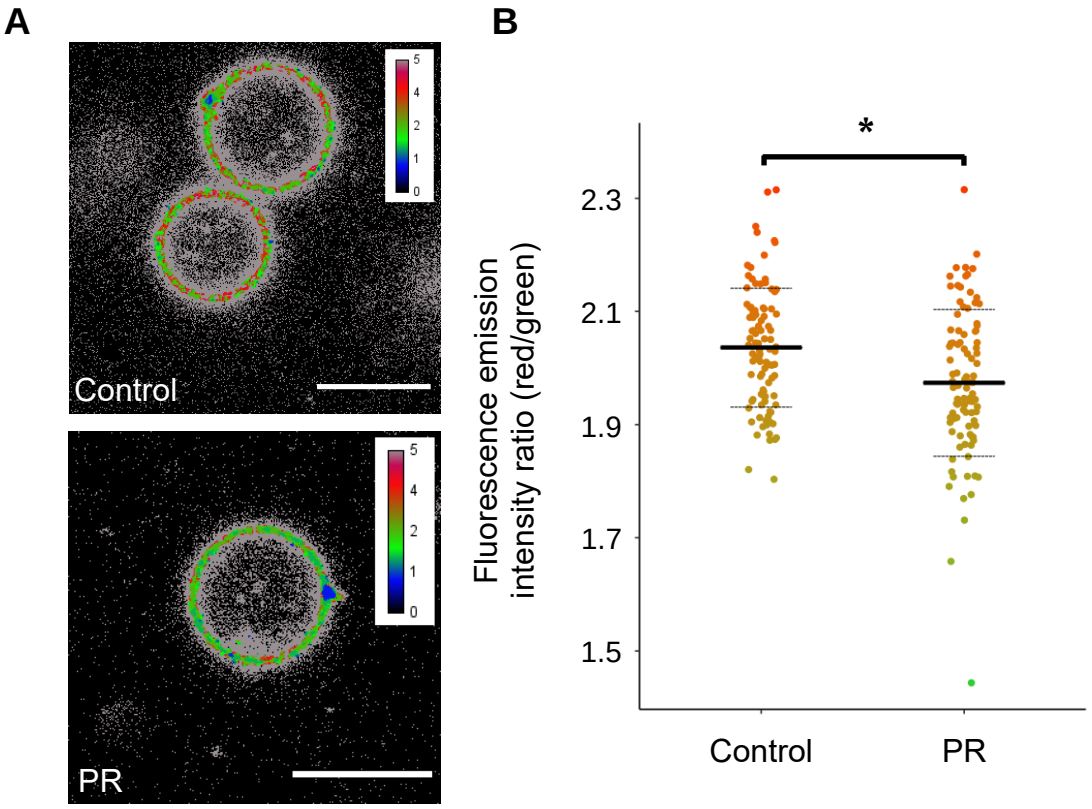


Figure 3.10 Proteorhodopsin synthesis in GUVs results in increased lipid order at the outer leaflet.

(A) Representative confocal microscopy images of GUVs containing PURE reactions with and without PR encoding pDNA. Images were generated by dividing the red channel by the green channel and applying a rainbow LUT to aid ratio visualisation. Scale bars are 10 μm. **(B)** Quantification of nile red fluorescence ratio for ≥ 20 GUVs from three independently generated batches. * represents statistical significance ($p < 0.001$) determined using a student's two-tailed, unpaired t-test.

A similar experiment was performed for GUVs following the synthesis of GalP, which had also been shown to localise to the membrane. This localisation was accompanied by a significant blue-shift in the nile red emission ratio which was determined at 1.969 ± 0.104 (Figure 3.11). This, again when compared to the control ratio of 2.036 ± 0.105, indicated that the transmembrane helices of GalP were, like PR, able to insert into the bilayer to such an extent that they altered the order of lipid tails at the outer leaflet.

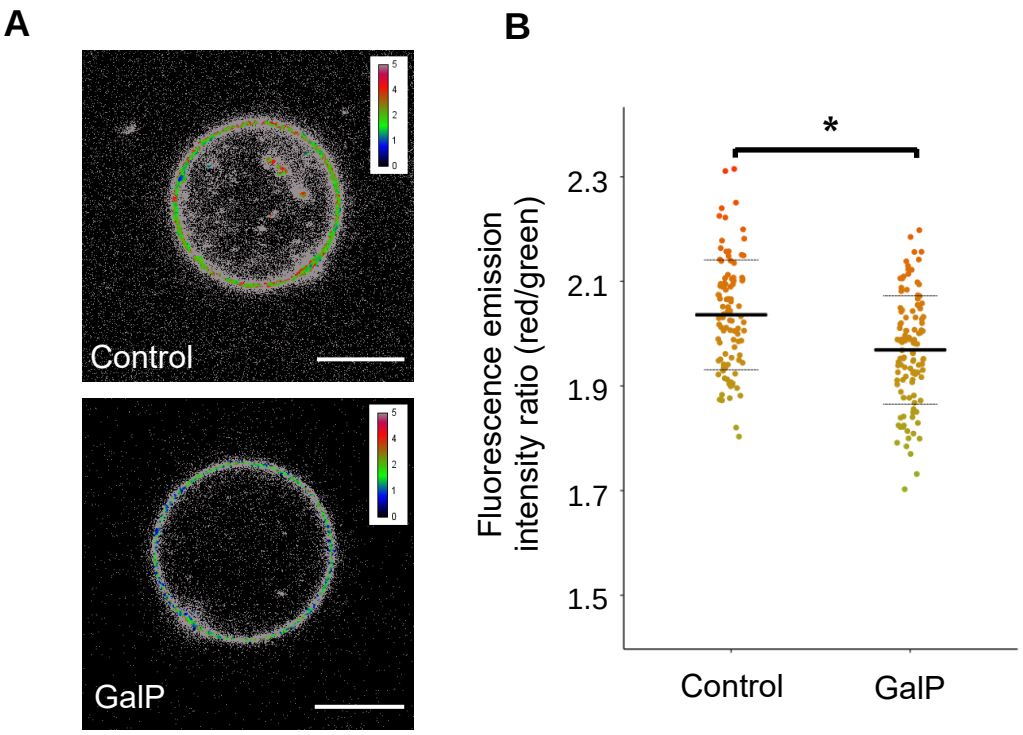


Figure 3.11 Galactose permease synthesis in GUVs results in increased lipid order at the outer leaflet.

(A) Representative confocal microscopy images of GUVs containing PURE reactions with and without GalP encoding pDNA. Images were generated by dividing the red channel by the green channel and applying a rainbow LUT to aid ratio visualisation. Scale bars are 10 μm. **(B)** Quantification of nile red fluorescence ratio for ≥ 20 GUVs from three independently generated batches. * represents statistical significance ($p < 0.001$) determined using a student's two-tailed, unpaired t-test.

The data presented above clearly indicated that both PR and GalP were able to localise themselves to the membrane when synthesised using minimal components in a minimal environment. It was also clear that the transmembrane helices of both of these proteins were penetrating the bilayer to such a degree that lipid order at the outer leaflet was being affected. However, the extent of

this insertion remained unclear. In order to test whether the helices were inserted fully and spanned the membrane correctly an assay was developed utilising hemagglutinin (HA) antibody affinity for the amino acid sequence: YPYDVPDYA. DNA oligonucleotides encoding this sequence were designed with overhanging complementarity to regions within the coding sequences of PR and GalP for assembly of HA epitope containing proteins using Gibson assembly. These constructs could then be used to synthesise HA containing proteins using the PURE system in GUVs to further analyse their insertion and topology through the binding of a fluorescently labelled membrane impermeable HA antibody.

Figure 3.12 shows the locations selected for HA insertion in the amino acid sequence of PR. The epitope was inserted in the extracellular loop located between transmembrane helices 2 and 3, and in the intracellular loop located between helices 5 and 6 to generate constructs PR-EL1HA and PR-IL3HA respectively. The presence of the desired coding sequence was confirmed by DNA sequencing. Figure 3.13 shows the selected sites for HA epitope integration in GalP. Here the epitope was inserted into the extracellular loop located between helices 5 and 6 and the large intracellular loop located between helices 6 and 7 to generate constructs GalP-EL3HA and GalP-IL3HA respectively. Constructs were confirmed by DNA sequencing.

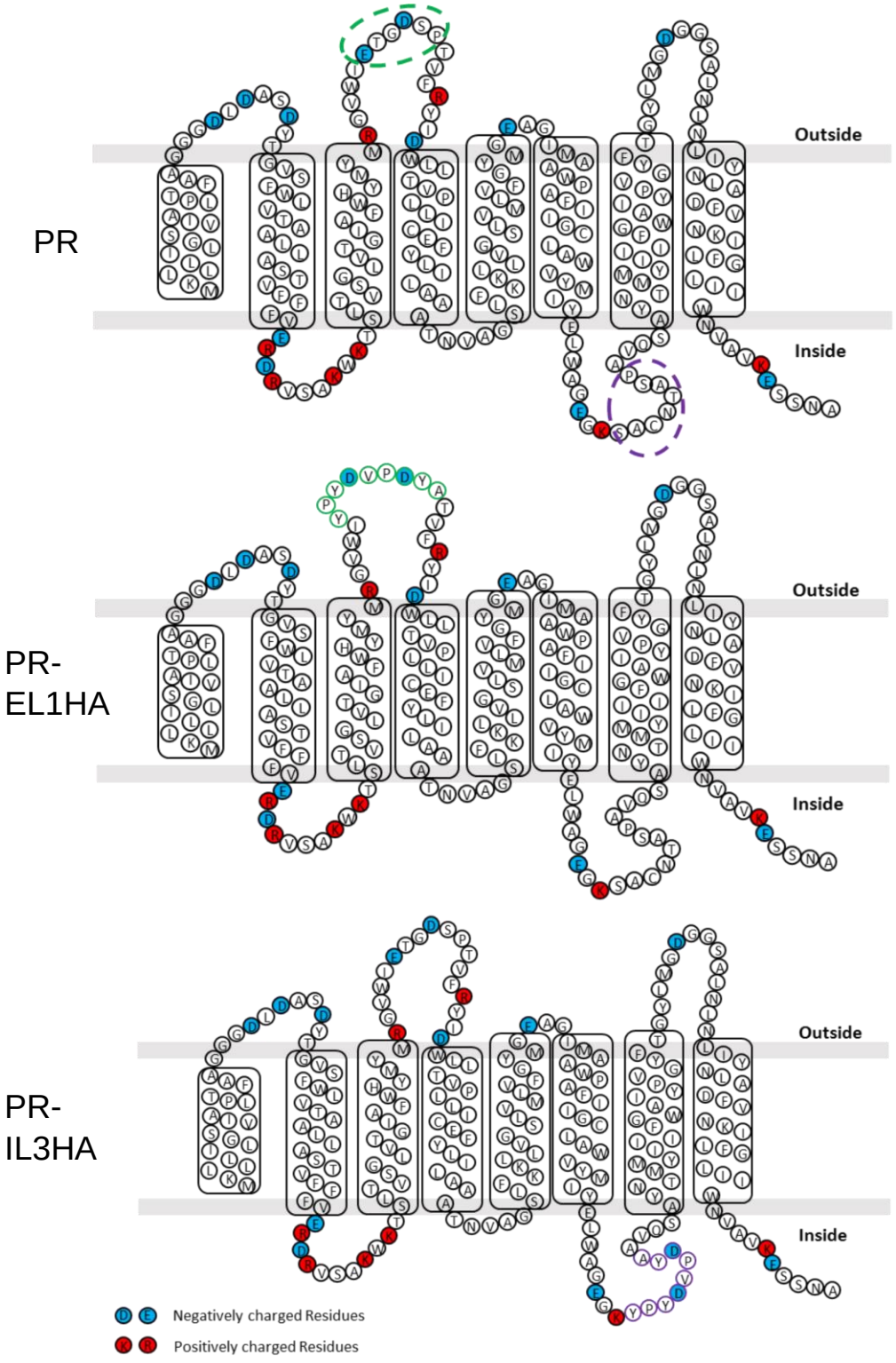
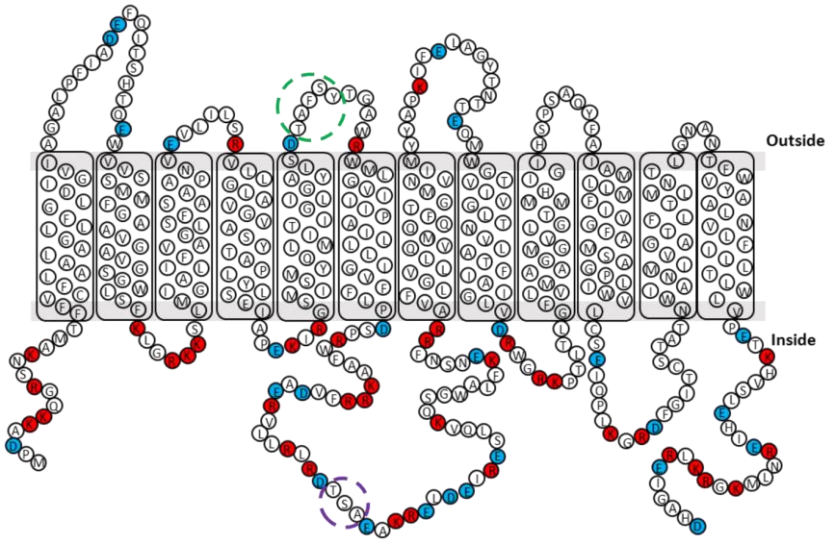


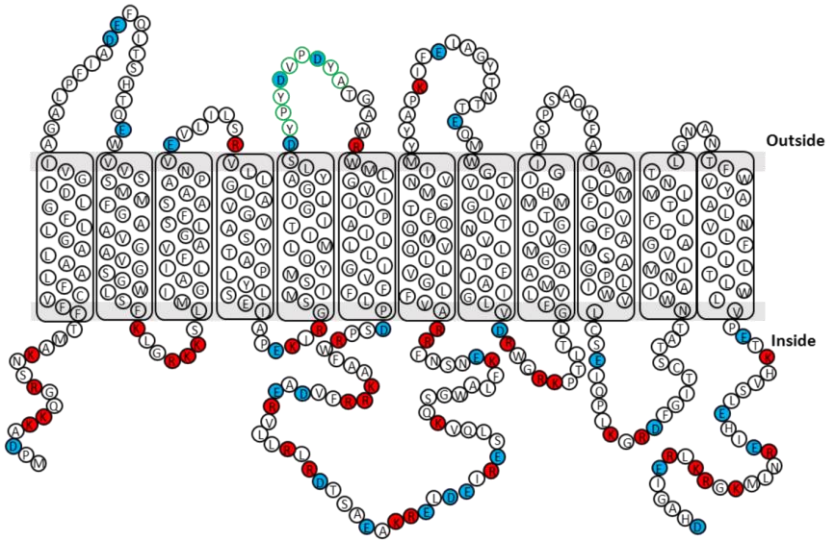
Figure 3.12 Extra- and intracellular loop modification of PR with hemagglutinin epitope.

The topological structure of PR is shown with the signal peptide on the left followed by the 7 transmembrane helices. HA epitope insertion by Gibson assembly yielded the extracellular (green) modification in PR-EL1HA and the intracellular (purple) modification in PR-IL3HA. Negatively charged residues are shown in blue and positively charged residues are shown in red.

GalP



GalP-
EL3HA



GalP-
IL3HA

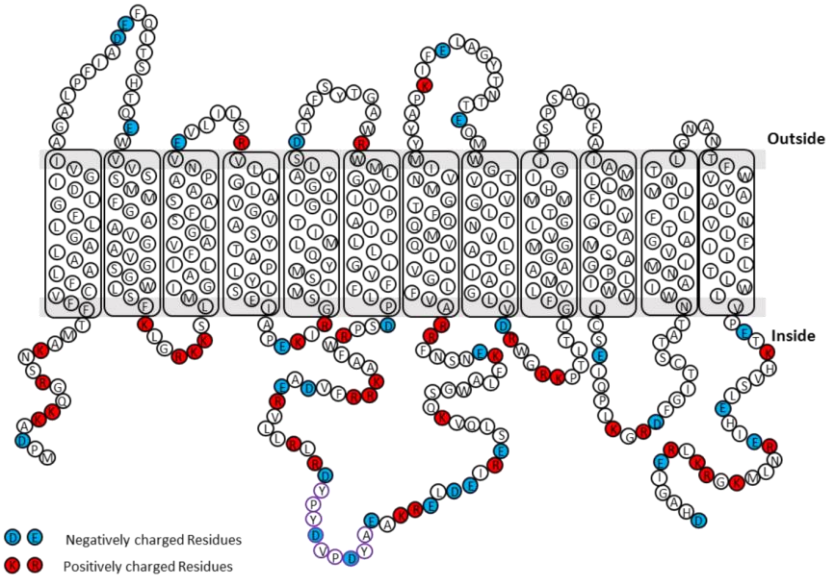


Figure 3.13 (previous page) Extra- and intracellular loop modification of GalP with hemagglutinin epitope.

The topological structure of GalP is shown with the 12 transmembrane helices ordered from N- to C-terminus. HA epitope insertion by Gibson assembly yielded the extracellular (green) modification in GalP-EL3HA and the intracellular (purple) modification in GalP-IL3HA. Negatively charged residues are shown in blue and positively charged residues are shown in red.

The above constructs were used in PURE reactions for the synthesis of HA epitope containing proteins inside GUVs. Following 3 hours of protein synthesis at 37°C, fluorescently labelled HA antibody (Alexa Fluor - 647 for PR and Alexa Fluor - 488 for GalP) was added to the GUV solution at a concentration of 5 μg ml^{-1} and incubated a room temperature for 30 minutes to allow antibody binding. GUVs were then pelleted by centrifugation and the antibody containing solution was removed and replaced by fresh outer solution. GUVs were then placed in microscopy chambers and allowed to settle before being visualised by confocal microscopy. Radial profiles were generated to obtain an extracellular/membrane ratio of antibody fluorescence. All ratios were normalised to the lowest ratio observed in control samples containing the protein of interest with no HA epitope.

Figure 3.14 shows results obtained for PR constructs. Confocal image analysis clearly shows a membrane localised fluorescence signal from Alexa Fluor-647 conjugated HA antibody when applied to PR-EL1HA containing GUVs where the HA epitope should be available on the extra-vesicular surface. A mean ratio of 0.757 ± 0.267 was calculated and this was found to be significantly lower than the ratios determined for both PR-IL3HA (1.017 ± 0.197) and the normalisation control PR (1.192 ± 0.141) when analysed using a one-way ANOVA with Tukey HSD post-hoc analysis. These data indicated that the extracellular loop of PR, located between helices 2 and 3 was correctly positioned on the outside of GUVs while the intracellular loop located between helices 5 and 6 was, in the majority of cases, unavailable for externally supplied HA antibody, again indicating the expected topology.

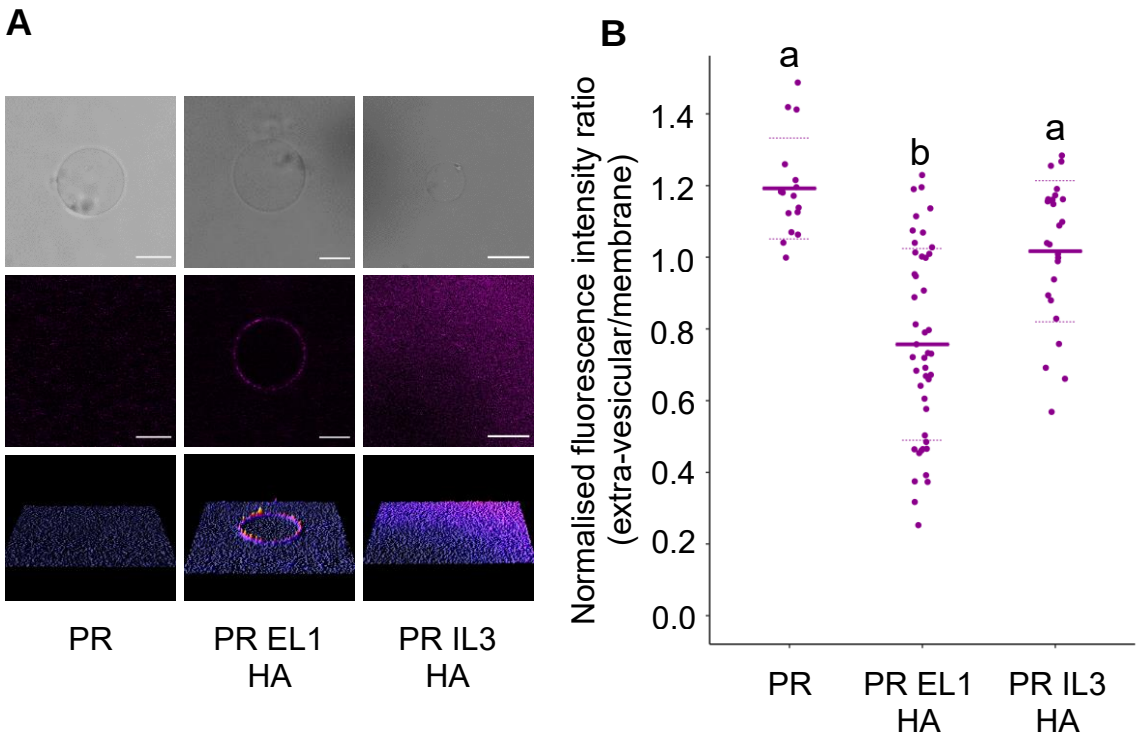


Figure 3.14 *De novo* synthesised PR is inserted into GUV bilayers and exhibits correct topology.

(A) Representative confocal images of GUVs following protein synthesis and incubation with Alexa Fluor 647-conjugated HA antibody. Images are organised as in Figure 3.7. Scale bars are 10 μm. **(B)** Quantification of extra-vesicular/membrane fluorescence ratio of Alexa Fluor 647-conjugated HA antibody fluorescence. Values shown are individual, mean and standard deviation of ≥ 25 GUVs for HA containing samples. All data were normalised to the lowest ratio observed from 16 GUVs containing PR without an HA epitope. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

Figure 3.15 shows results obtained for GalP constructs. Confocal analysis shows the membrane localisation of Alexa Fluor-488 conjugated HA antibody bound to an HA epitope localised on the external side of GUV membranes. A control construct containing GalP with no HA epitope was synthesised and HA binding was quantified. The highest level of binding observed in for these vesicles was used as a control value to normalise the subsequent data. Antibody binding was significantly higher than in control vesicles for both GalP EL3 HA (0.819 ± 0.211) and GalP IL3 HA (0.812 ± 0.273) indicating that the protein was inserted but seemed to demonstrate dual topology.

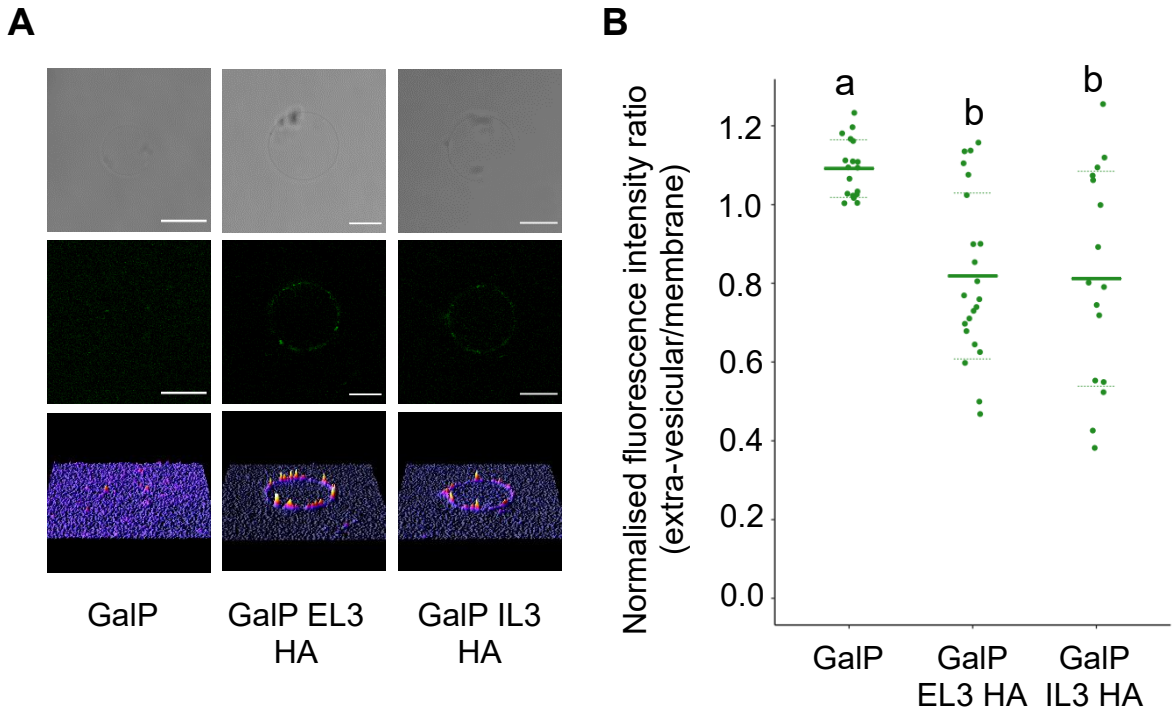


Figure 3.15 *De novo* synthesised GalP is inserted into GUV bilayers and exhibits dual topology.

(A) Representative confocal images of GUVs following protein synthesis and incubation with Alexa Fluor 488-conjugated HA antibody. Images are organised as in Figure 3.7. Scale bars are 10 µm. **(B)** Quantification of extra-vesicular/membrane fluorescence ratio of Alexa Fluor 488-conjugated HA antibody fluorescence. Values shown are individual, mean and standard deviation of ≥ 16 GUVs for HA containing samples. All data were normalised to the lowest ratio observed from 18 GUVs containing GalP without an HA epitope. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

3.3 Discussion

Using a minimal cell-free protein synthesis system to investigate polytopic membrane protein folding *in vitro* has been the focus of a number of studies (Matsubayashi et al., 2014; Kuruma and Ueda, 2015; Harris et al., 2017). However, few have examined this process in a cell-size mimicking environment (Soga et al., 2014; Ohta et al., 2016) and none, until now, has tracked protein localisation within these cell-mimetic systems. Considering the large ratio of aqueous volume to membrane surface area in these systems and the propensity for polytopic α-helical membrane proteins to aggregate in aqueous environments, it would be reasonable to assume that translational recruitment plays an important role even in this minimal environment. This chapter has focussed on using a minimal cell-free system (PURE) for the synthesis of two polytopic membrane proteins; the 7 transmembrane domain protein PR and the 12 transmembrane domain protein GalP; both in bulk solution supplemented with LUVs (< 200 nm diameter) and inside cell-size mimicking GUVs (> 1 μm diameter). Protein localisation and insertion in GUVs was monitored using fluorescence-based confocal microscopy assays with results indicating that both PR and GalP could localise and insert into GUV membranes in the absence of the *in vivo* targeting and insertion pathways.

The results presented in section 3.2.1 show the successful generation of GUVs using the droplet transfer approach (Altamura et al., 2017) and the encapsulation characteristics of the method. The inclusion of the soluble fluorescent marker HPTS in the inner phase and a fluorescently labelled lipid (Rh-PE) in the oil phase led to the clear definition of membrane and aqueous phases using dual fluorescence confocal microscopy (Figure 3.1). The leakage of small molecules such as HPTS has been reported to be higher in micro-scale GUVs as compared to nano-scale LUVs (Braun et al., 2018). Leakage such as this would be detrimental to protein synthesis reactions encapsulated within GUVs and as such size, stability and leakage of HPTS was analysed over 24 hours at room temperature. If vesicles were leaking HPTS, a drop in the mean fluorescence intensity of GUVs would be expected, and if the vesicles were unstable, a change in vesicular diameter would also be detected. Figure 3.2 shows that over a 24 hour period there was no significant reduction in GUV size

or HPTS fluorescence intensity confirming that GUVs generated using this approach were stable and did not leak internal soluble components.

It was then important to test the synthesis capability of the PURE system when synthesising the soluble reporter EGFP, since correct folding of EGFP would be crucial as a marker for membrane protein localisation moving forwards. Figure 3.3 shows the results for the bulk synthesis of EGFP using 20 µl PURE reactions. A single band was clearly visible on an SDS-PAGE gel when synthesised in the presence of fluorescently labelled lysine amino acids. Protein migration through the gel indicated a larger molecular weight than expected at ~35 kDa, however this can be explained by the high salt concentrations of the PURE buffer solution which made boiling of samples impossible. This most likely led to the maintenance of some secondary protein structure leading to the unexpected protein band migration. To test that this product was indeed EGFP and that it was folded correctly, fluorescence analysis of a bulk reaction was performed (Figure 3.3 C) and increasing fluorescence was detected using an FITC filter set suitable for the measurement of EGFP fluorescence. These data indicated that the PURE system could synthesise soluble proteins and that they could adopt their native fold in this minimal environment, confirming previous reports (Shimizu et al., 2001).

EGFP synthesis within GUVs was then carried out to confirm that stable GUVs could be produced encapsulating PURE reactions, and that protein could be synthesised at 37°C in these vesicles. GUVs were stable over this 3-hour time course and encapsulated PURE reactions maintained their functionality for the synthesis of soluble reporters (Figure 3.4) again confirming previous encapsulation studies (Yu et al., 2001). It is likely that some GUVs burst during incubation at this temperature however the dilution factor from the inside to the outside of GUVs is so high that any protein synthesis outside GUVs would be negligible as can be seen from the lack of EGFP signal from the external solution (Figure 3.4).

The synthesis of a number of membrane proteins, both in bulk and encapsulated within GUVs, has been reported for the PURE system (Kuruma and Ueda, 2015; Ohta et al., 2016). Bulk synthesis of PR and GalP was performed in the presence of LUVs to confirm full length protein synthesis was occurring. Protein bands were detected on SDS-PAGE gels and again migration did not correspond to the

expected protein molecular weights (Figures 3.5 and 3.6). This was again due to the lack of sample boiling and the maintenance of membrane protein secondary structure in SDS leading to migration discrepancies. This phenomena for membrane proteins has been well reported (Rath et al., 2009). The existence of smeared bands may indicate the presence of truncated translation products as has been reported previously when using the PURE system (Tuckey et al., 2014), or it may be due to the presence of multiple SDS-stabilised, full-length protein structures being present with differing levels of denaturation and detergent binding.

Analysing membrane fluorescence in GUVs using radial profiles (see section 2.2.8) has been used previously to show membrane localisation of a fluorescent reporter (Peters et al., 2015). Thus, this analysis method was applied to investigate the membrane localisation of both PR and GalP following protein synthesis. Both proteins exhibited strong association to the membrane and little aggregation within the vesicle lumen (Figures 3.7 and 3.8). This was surprising considering the large aqueous volume present within these GUVs. These initial localisation experiments were then followed up with analyses for membrane insertion.

It has been shown that membrane protein helix insertion causes detectable changes to the surrounding lipid packing order (Martyna et al., 2017). The lipophilic, environment sensitive probe nile red has been used to report on lipid packing changes in various membrane mimetic systems (Kucherak et al., 2010). Previous work has also shown that nile red specifically localises to the outer leaflet of artificial membranes (Mukherjee et al., 2007) making it a good choice for the analysis of α-helix insertion deep into the membrane. The fluorescence emission properties of nile red are altered depending on the order of the surrounding lipids with higher order packing leading to a blue-shift in fluorescence emission. These properties allow the ratiometric detection of lipid packing changes (Kucherak et al., 2010). Initial experiments were performed using POPC GUVs with increasing concentrations of cholesterol which is known to increase lipid order (Marsh, 2010). Results confirmed that, with increasing cholesterol there was a significant blue-shift in the fluorescence emission ratio (Figure 3.9). No difference was detected between 0 and 10 mol% cholesterol or between 20 and 30 mol% cholesterol but this may have been due to the low

sample size, and further sampling may have resolved differences. Nevertheless, the technique seemed suitable for the analysis of helix insertion. Synthesis of PR and GalP in POPC GUVs led to a significant blue-shift in the emission ratio of nile red, suggesting that helix insertion was occurring in both of these cases and causing increased lipid order at the outer leaflet (Figures 3.10 and 3.11). The detected change in emission ratio was small and thus required a large number of samples to be analysed in order to resolve differences. Nile red concentration effects were ruled out through sampling from 3 independently prepared experiments for each condition.

Since membrane proteins are expected to be equilibrium structures (White and Wimley, 1999), and that once inserted into the membrane they will adopt a native fold (Pellowe and Booth, 2019), it would be reasonable to assume that the majority of these proteins are inserted and folded correctly. In order to test this, constructs were generated including an HA epitope within an expected intracellular loop and an expected extracellular loop of both PR and GalP (Figures 3.12 and 3.13). Following protein synthesis and incubation with a fluorescent HA antibody, GUVs were analysed for surface exposed HA epitopes by confocal microscopy.

PR showed insertion and translocation of the expected extracellular loop in the majority of cases. In a minority of cases, HA was found to bind to the expected intracellular loop suggesting that this region was translocated across the bilayer in these instances (Figure 3.14). This could be explained by the fact that the HA epitope contains two negatively charged aspartate residues and its insertion into the intracellular loop affected the charge balance, thus leading to increased propensity for the translocation of this loop in accordance with the positive inside rule (Von Heijne, 1992). The hydrophobicity of the surrounding transmembrane helices may have made this translocation event less likely accounting for the limited number of cases detected. It also cannot be ruled out that a small proportion of proteins are misfolded within the membrane, as some misfolded protein has been detected previously using the PURE system and surface enhanced infrared spectroscopy (SEIRAS) (Harris et al., 2017). It would be possible to increase the positive charge of this intracellular loop through the addition of positively charged amino acids flanking the HA epitope. This may re-

balance the overall charge balance of the loop, however would still not represent the native charge state.

Interestingly, GalP showed efficient translocation of both the extracellular loop (as expected) but also the anticipated intracellular loop (Figure 3.15). Unlike PR, the proportions of translocation for both regions were equal, suggesting that both of these regions were surface exposed to an equal extent. This was surprising due to the high polarity of the expected intracellular region and the thermodynamic challenge this poses to translocation (Moon and Fleming, 2011). This relatively simplistic view has however been challenged recently by studies showing the post-translational flipping of entire helical regions and their polar connecting loops due exclusively to lipid environmental effects (Vitrac et al., 2013; Vitrac et al., 2015; LeBarron and London, 2016). These studies have often focussed on the 12 transmembrane *E. coli* protein lacY, which is a member of the same family and thus closely related to GalP. This implies that similar dual topology may well be possible for GalP and would explain the experimental results presented in Figure 3.15. It may also be possible that the introduction of two negatively charged aspartate residues within the HA epitope affected the equilibrium structure of GalP or led to a transient dual topology state as for PR.

In conclusion, the work carried out in this section showed that stable GUVs could be formed using the droplet transfer approach and that both soluble and membrane proteins could be synthesised using the PURE system encapsulated within these cell-size mimicking GUVs. Both PR and GalP were able to localise to the membrane despite the large challenge these proteins must have faced in traversing the aqueous interior of the GUV lumen. PR was shown to insert into the membrane with the expected topology, indicating that the protein was properly folded. The insertion of GalP was also quantified and a dual topology state was identified similar to that observed in previous studies on the closely related protein lacY.

4. Minimal requirements for membrane protein targeting and insertion

4.1 Introduction

The importance of membrane proteins to modern cells is underlined by genomic data estimating that around 30% of all open reading frames in a given organism's genome encode membrane proteins (Wallin and von Heijne, 1998). α-helical proteins perform a diverse range of functions, and despite the heterogeneity of their primary amino acid sequences they all maintain a structural homology in the formation of bundled of hydrophobic helices. In an aqueous environment these helices favour aggregation meaning any subsequent insertion into the lipid bilayer would have to overcome a substantial thermodynamic barrier. Thus, to prevent this detrimental aggregation, mechanisms must have evolved to either maintain α-helical proteins in an unfolded state until they reach the membrane, or to enable rapid co-translational membrane targeting and insertion before aggregates can form. Modern cells contain a number of chaperoning pathways to achieve precisely this, for example the signal recognition particle (SRP) and SecB pathways.

In the classical view of membrane protein biogenesis, the emergence of the first hydrophobic domain from the ribosome exit tunnel acts as a trigger to determine the fate of the protein. For the majority of inner membrane proteins the emergence of this hydrophobic region is followed by its recognition by the SRP, a highly conserved ribonucleoprotein, and the subsequent recruitment of the translating ribosome-SRP-nascent chain complex to the translocon via interaction with the SRP receptor (SR) for the co-translational insertion of the downstream helices (Luirink et al., 2012). A number of recent studies indicating that SRP targets multiple regions along the entire nascent chain, and even untranslated RNA regions upstream of the proteins start codon, has challenged this model (Schibich et al., 2016; Costa et al., 2018). Nevertheless, the emergence of N-terminal transmembrane (TM) regions from the ribosome must still have important implications due to their hydrophobicity and aggregation potential.

As mentioned previously, evidence suggests that the majority of polytopic α-helical membrane proteins are inserted into the membrane co-translationally (Rapoport, 2007). However, in a simplified *in vitro* system such as the PURE system, post-translational insertion cannot be entirely ruled out, and has been observed for small proteins such as subunit c of F_1F_0 ATP synthase and the bacterial potassium channel KcsA (Robinson and Woolhead, 2013; Altrichter et al., 2017).

Interestingly, mRNA localisation has been implicated *in vivo,* in both prokaryotes and eukaryotes, as being important for the localisation of membrane protein synthesis (Nevo-Dinur et al., 2011; Holt and Bullock, 2009). However, the intrinsic properties of membrane protein mRNAs with respect to membrane binding have not yet been investigated *in vitro*.

Proteorhodopsin (PR) is an interesting example of an inner membrane protein that contains a short N-terminal hydrophobic signal peptide, usually reserved for proteins destined for the secretory pathway or the outer membrane. A recent study has suggested that this signal sequence is important for the efficient targeting of PR to the inner membrane through the SRP pathway and that it facilitated protein insertion in a sec-independent manner (Soto-Rodriguez and Baneyx, 2019). Data presented in Chapter 3 have shown that PR, containing the N-terminal signal peptide, can localise and insert into GUV membranes despite the lack of chaperoning systems. The question then arises whether the presence of the signal peptide affects the targeting and insertion of the proteins downstream helices?

The 12-transmembrane helix protein GalP is a more 'traditional' inner membrane protein in so far as it contains no signal peptide sequence, and in this case the helices themselves likely act as SRP substrates. Data presented in Chapter 3 again indicate that the full-length protein can localise and insert itself into GUV membranes spontaneously, however, the question remains whether the first transmembrane helix contains fundamental information aiding this process?

The aim of this chapter was to investigate the fundamental mechanisms responsible for the membrane targeting and insertion of both PR and GalP in cell-sized giant unilamellar vesicles (GUVs). This required the design of experiments to address a number of key questions: 1) Is protein localisation and

insertion a co- or post-translational event? 2) Is membrane binding of mRNAs responsible for localising translation? 3) Are the N-terminal hydrophobic helices of PR and GalP important for localisation and insertion? 4) And if so, what are the physical characteristics of these helices that aid localisation and insertion?

To answer the above questions a number of experiments using the PURE system and GUVs were performed. Co- or post-translational localisation and insertion was tested by tracking EGFP/mCherry-tagged protein, synthesised without a membrane present, and by monitoring the localisation of fluorescently labelled, stalled ribosome nascent-chain complexes. Constructs were generated to enable the tracking of fluorescently labelled mRNAs during transcription using a conditionally fluorescent dye/aptamer combination. Further constructs were generated to investigate the effects of N-terminal helix modifications on protein targeting and insertion using EGFP/mCherry localisation, a previously developed nile red lipid environment assay, and a fluorescent HA antibody binding assay.

4.2 Results

4.2.1 Post-translational addition of a lipid bilayer does not lead to membrane protein integration

The localisation and insertion of α-helical membrane proteins PR and GalP was assessed in Chapter 3 and results showed that the proteins were capable of inserting themselves into an empty bilayer environment without the aid of translocon complexes or other chaperoning systems. However, the mechanisms underpinning this insertion process remained unclear. The first question to be addressed was whether the observed localisation pattern was due to a co-translational or post-translational process, with the hypothesis being that it was happening co-translationally.

To investigate this, PURE reactions supplemented with either PR-EGFP or GalP-mCherry pDNA and 200 mM sucrose were set up in bulk without the addition of a bilayer environment. These bulk reactions were incubated for three hours at 37°C as before. Following incubation, reactions were treated with RNase at room temperature for 30 minutes to stop all protein synthesis and then used directly as the inner solution to generate GUVs by droplet transfer. Following a further 30-minute incubation at 37°C, the GUVs were visualised by confocal microscopy to determine the localisation of PR and GalP when the amphiphilic environment was supplied post-translationally.

Figure 4.1 shows a representative confocal image of GUVs generated using a post-translational reaction containing PR-EGFP. The level of fluorescence, and thus the yield of protein, is clearly low in comparison to co-translational experiments, and no vesicles were observed exhibiting any membrane localisation of EGFP fluorescence. The plot profiles shown are from three arbitrarily selected GUVs from within the selected image and further represent the lack of membrane localisation, with the majority of the fluorescence localised within the lumen of vesicles (Figure 4.1 B). These data indicate that the previously determined insertion of PR into GUV membranes was not a post-translational process and that the lack of a readily available lipid environment seems to have a detrimental effect on protein translation.

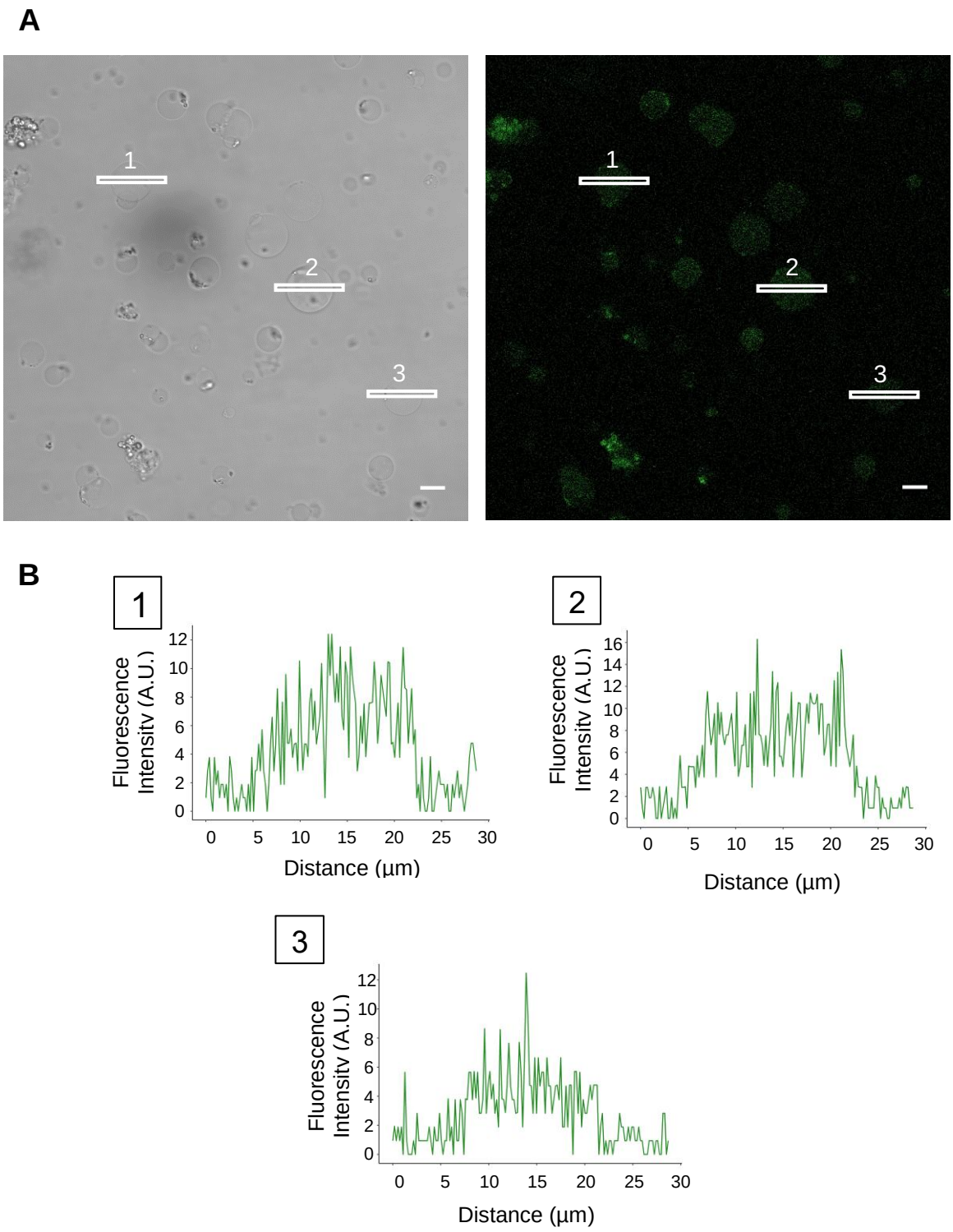


Figure 4.1 PR is unable to insert into GUV membranes supplied post-translationally.

(A) Representative bright field and fluorescence confocal images showing GUVs encapsulating PR-EGFP synthesized prior to vesicle formation without an amphiphile present. Vesicle regions selected for profile analysis are shown and labelled 1-3. Scale bars are 10 µm. **(B)** Fluorescence profile analyses for the three GUVs selected from (A).

Figure 4.2 shows the results of the same post-translational experiment carried out using GalP-mCherry. The representative confocal image again shows that the majority of fluorescence from mCherry was distributed throughout the vesicle lumen and was not localised to the membrane. This was quantified in plot profiles from three arbitrarily selected GUVs as for the PR-EGFP experiment. A small number of GUVs showed moderate membrane localisation as can be seen in Figure 4.2 A, however, this only occurred in vesicles which had fused together and thus was likely an artefact of the fusion process. In all of these cases, there was still a clear fluorescence signal from within the vesicle lumen. No individual, unfused GUVs exhibited any membrane localisation of fluorescence. These data indicate that GalP, like PR, was unable to localise and insert into the membrane when the lipid environment was supplied post-translationally and was likely misfolded and aggregated within the aqueous lumen.

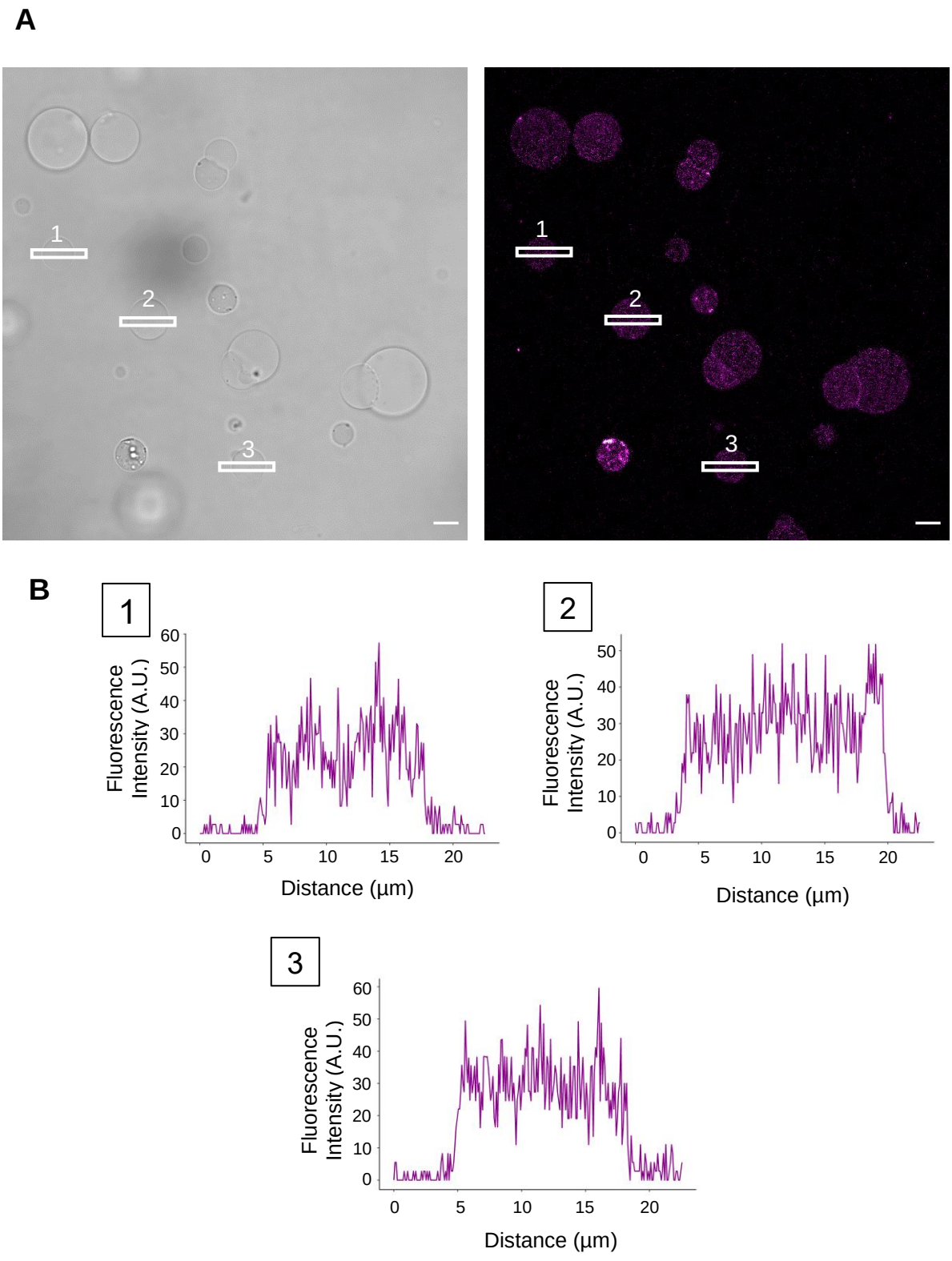


Figure 4.2 GalP is unable to insert into GUV membranes supplied post-translationally.

(A) Representative bright field and fluorescence confocal images showing GUVs encapsulating GalP-mCherry synthesized prior to vesicle formation without an amphiphile present. Vesicle regions selected for profile analysis are shown and labelled 1-3. Scale bars are 10 µm. **(B)** Fluorescence profile analyses for the three GUVs selected from (A).

4.2.2 Membrane protein mRNA is not itself targeted to the membrane of GUVs

The localisation of mRNA transcripts encoding membrane proteins PR and GalP was investigated to determine whether they were able to directly associate with the membrane of GUVs, and thus favour the membrane localisation of translation products. A fluorescence assay was used in which the mRNAs were tagged with an RNA aptamer sequence capable of directly binding and greatly enhancing the fluorescence of a dye molecule. This aptamer, known as Spinach2 (Strack et al., 2013) (Figure 4.3 A), was attached by PCR to the 3' end of both PR and GalP coding sequences, upstream of the T7 terminator, to generate constructs for mRNA visualisation using the conditionally fluorescent dye (5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene] -3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one (DHFBI) (Figure 4.3 B,C). It was essential to uncouple transcription from translation for these mRNA experiments and therefore ribosomes were omitted from the reaction mixture.

The PR-Spinach2 construct was first tested in bulk using a plate reader (Figure 4.3 D). PURE reactions without ribosomes were set up and supplemented with 20 ng μl^{-1} PR-Spinach2 pDNA or nuclease-free H_2O as a negative control. DHFBI was added at a concentration of 20 μM and reactions were incubated at 37°C throughout the experiment with fluorescence emission being measured every 10 minutes. Fluorescence increased up to 80 minutes and then slightly decreased until the end of the experiment at 180 minutes. No increase in fluorescence was observed in the negative control. Figure 4.3 E and F show the results obtained for PR-Spinach2 mRNA localisation in GUVs. Here, PURE reactions were again set up without ribosomes before encapsulation within GUVs. DHFBI was included in both the inner and outer solutions at 20 μM. Vesicles were incubated for 60 minutes at 37°C due to the high level of fluorescence seen in bulk at this time point. A representative confocal image is shown which demonstrates a completely soluble and diffuse pool of PR mRNA within the lumen of GUVs, thus ruling out intrinsic mRNA membrane binding as the cause of the observed protein localisation (Figure 4.3 E). Plot profiles were generated for three arbitrary GUVs from the provided image to illustrate the diffuse nature of the fluorescent mRNA signal (Figure 4.3 F).

Figure 4.4 shows the results for Spinach2 tagged GalP mRNA. Again, bulk experiments were set up and analysed using a plate reader (Figure 4.4 A). As observed for PR, the fluorescence signal from Spinach2 bound DHFBI increased over time. In contrast to PR, there was no loss of fluorescence throughout the experiment but fluorescence plateaued at around 110 minutes. GUV experiments were also carried out to determine the localisation of GalP-Spinach2 mRNA following the same procedure detailed for PR-Spinach2. A representative confocal image is shown with corresponding plot profiles. Again, the fluorescence of GalP mRNA products was present throughout the lumen and showed no membrane localisation (Figure 4.4). These data suggest that, like PR, GalP mRNA itself does not localise to the membrane.

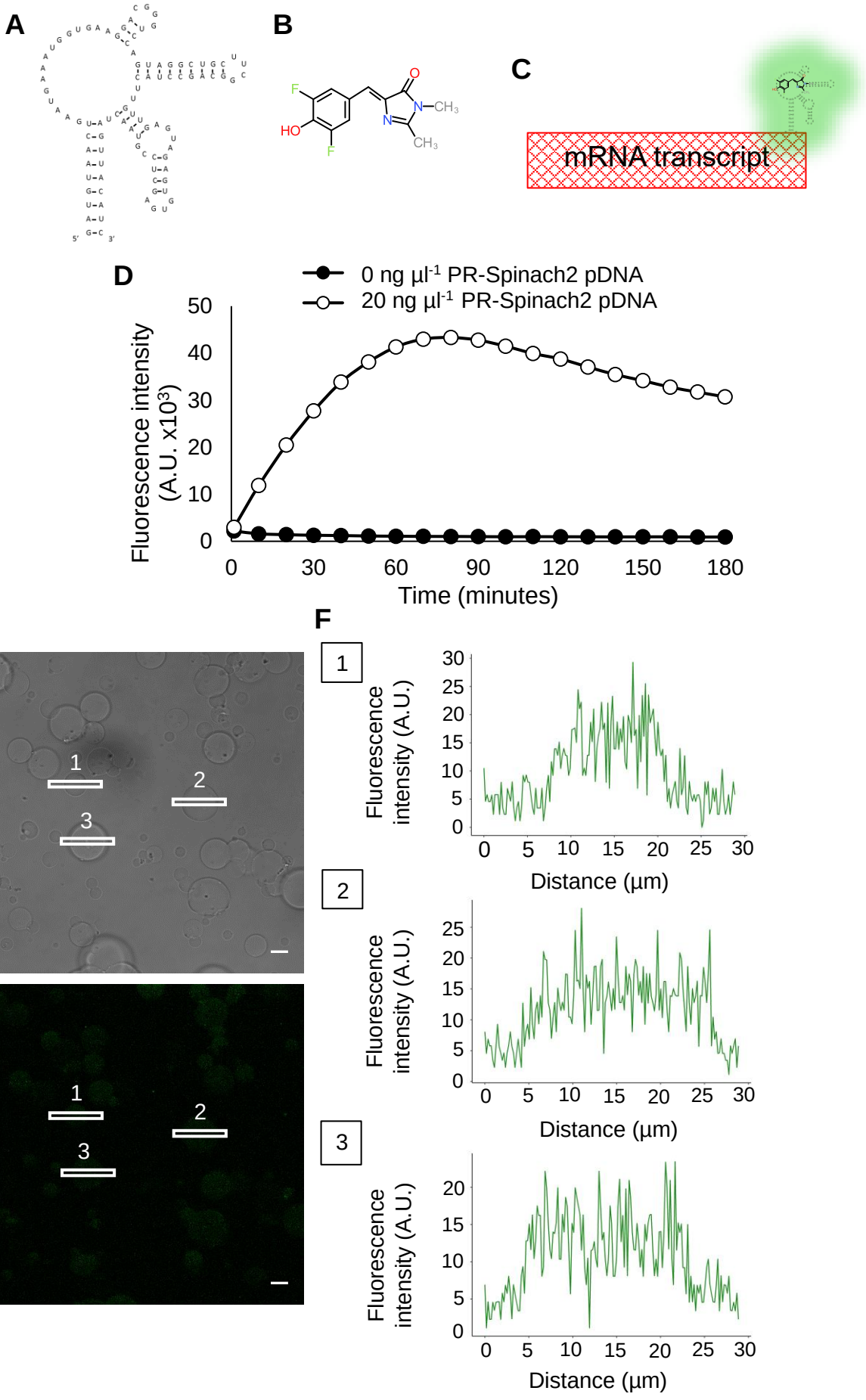

A
B
C
mRNA transcript
D
0 ng µl-1 PR-Spinach2 pDNA
20 ng µl-1 PR-Spinach2 pDNA
Fluorescence intensity (A.U. x10³)
Time (minutes)
E
1
2
3
F
Fluorescence intensity (A.U.)
Distance (µm)

Figure 4.3 (previous page) Fluorescently labelled PR mRNA is not membrane localised.

(A) RNA sequence and structure of the Spinach2 aptamer. **(B)** Chemical structure of the conditionally fluorescent dye DHFBI that fluoresces when bound to the Spinach2 aptamer. **(C)** Schematic representation of an mRNA with DHFBI bound Spinach2 at the 3' end. **(D)** Fluorescence from a bulk PURE reaction synthesising PR-Spinach2 supplemented with 20μM DHFBI at 37°C measured using a plate reader with standard FITC filter settings and compared to a control lacking pDNA. **(E)** Representative bright field and fluorescence confocal images showing GUVs encapsulating PR-Spinach2 mRNA synthesised without ribosomes and in the presence of 20 μM DHFBI both inside and outside. Vesicle regions selected for profile analysis are shown and labelled 1-3. Scale bars are 10 μm. **(F)** Fluorescence profile analyses for the three GUVs selected from (E).

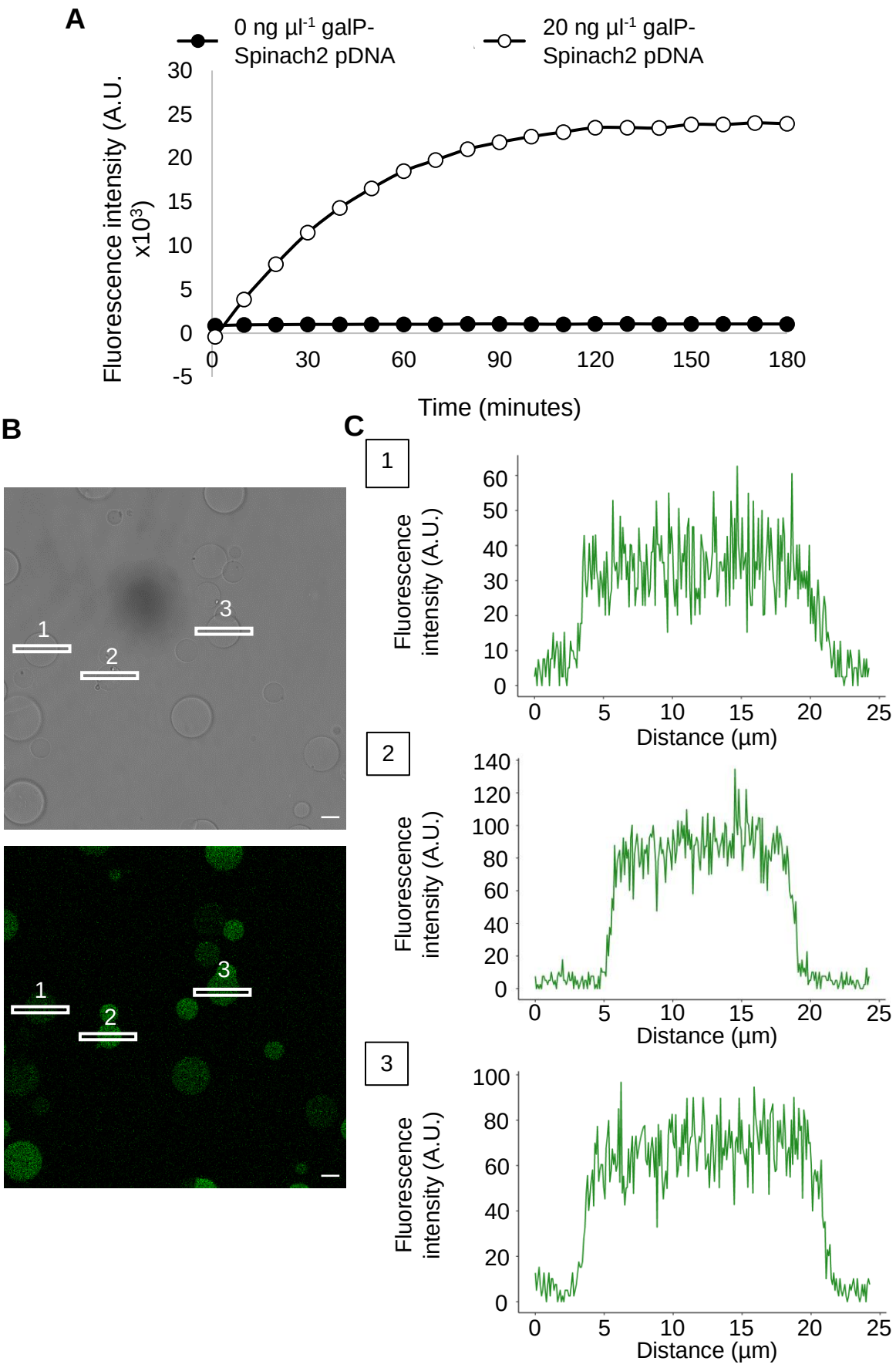

A
0 ng µl-1 galP-Spinach2 pDNA
20 ng µl-1 galP-Spinach2 pDNA
Fluorescence intensity (A.U. x10³)
Time (minutes)
B
C
1
2
3
Fluorescence intensity (A.U.)
Distance (µm)

Figure 4.4 (previous page) Fluorescently labelled GalP mRNA is not membrane localised.

(A) Fluorescence from a bulk PURE reaction synthesising GalP-Spinach2 supplemented with 20 μM DHFBI at 37°C measured using a plate reader with standard filter settings and compared to a control lacking pDNA. **(B)** Representative bright field and fluorescence confocal images showing GUVs encapsulating GalP-Spinach2 mRNA synthesized without ribosomes and in the presence of 20 μM DHFBI both inside and outside. Vesicle regions selected for profile analysis are shown and labelled 1-3. Scale bars are 10 μm. **(C)** Fluorescence profile analyses for the three GUVs selected from (B).

4.2.3 *In situ* observation of ribosome nascent chain complex localisation during protein synthesis

The data presented above clearly indicate that, for both PR and GalP, the observed protein localisation could not occur when the lipid environment was supplied post-translationally. Additionally, intrinsic localisation properties of the protein coding mRNAs were not responsible for the observed protein localisation. These data implied that the localisation and insertion of both PR and GalP occurs in a co-translational manner. To test whether this was indeed the case, an assay was devised to track the localisation of stalled ribosome nascent chain complexes (RNCs) (Figure 4.5 A, B).

To generate stalled RNCs it was first necessary to construct a vector that synthesised the entire protein, followed by a linker region, and finally a stalling sequence. An 84 residue amino acid region from the *E. coli* protein TolA was selected to act as a linker region. TolA had previously been shown to produce an unstructured peptide chain used as a linker for RNC purposes (Amstutz et al., 2002). This linker was then followed by a 16 residue amino acid sequence taken from the *E. coli* SecM protein known to cause translational stalling at a C-terminal proline residue (Zhang et al., 2015).

Two PR containing vectors were constructed, one coding for an un-stalled PR-TolA fusion and another coding for PR-TolA-SecM. These constructs were included in PURE reactions supplemented with LUVs and fluorescently labelled lysine amino acids for in-gel SDS-PAGE analysis of translation products. Samples were collected at 30, 60, 120 and 180 minutes, incubated at 65°C for 10 minutes in 2x SDS sample buffer and then stored at -20°C until preparation of all samples was complete. Following this, samples were run on 10% polyacrylamide gels and

fluorescence was analysed before gels were stained with coomassie to assess protein loading. Figure 4.5 C shows the results of SDS-PAGE analysis. PR-TolA translation products were clearly visible on the gel following fluorescent analysis and generally increased over time as expected. In contrast, PR-TolA-SecM showed greatly reduced production of a fluorescently labelled translation product indicating that translational stalling was occurring and producing arrested RNCs.

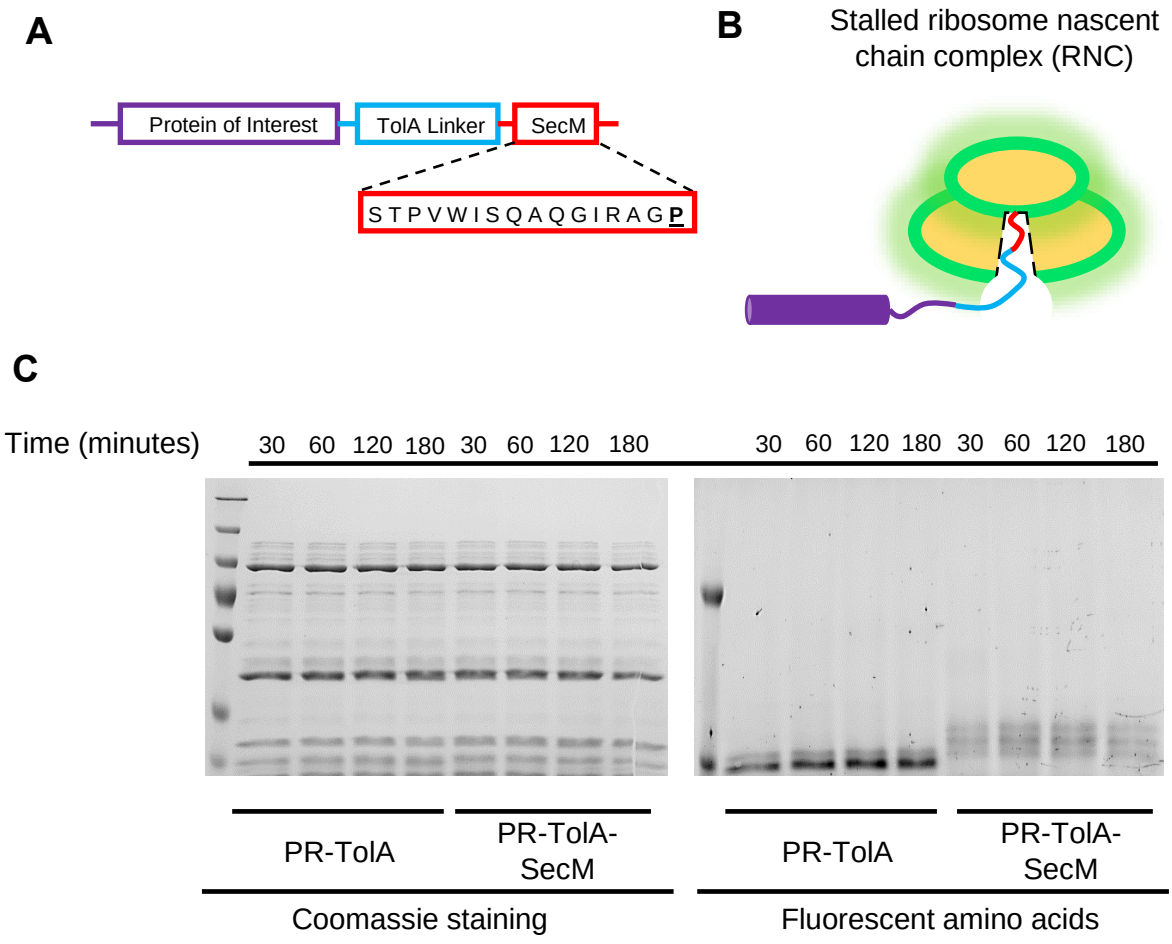


Figure 4.5 Translational stalling assay and production of stalled PR using the SecM peptide.

(A) Schematic showing the construction of a sequence containing an unstructured linker region from the protein TolA and a stalling sequence from the protein SecM attached to the C-terminus of a protein of interest. The sequence of SecM is shown with the proline residue responsible for translational stalling indicated in bold and underlined. **(B)** Cartoon representation of the assay for tracking ribosome nascent chain complex (RNC) localisation by fluorescently labelling ribosomes. **(C)** SDS-PAGE time-series analysis of PURE reactions synthesising PR-TolA with and without the SecM stalling sequence. Images are of the same gel stained with coomassie (left) or probed for BODIPY fluorescence (right).

To visualise the accumulation of stalled RNCs, ribosomes needed to be fluorescently labelled. To achieve this, ribosomal proteins were labelled with Alexa Fluor 488 5-SDP ester, which forms covalent linkages with primary amines found at the N-terminus of all proteins and on lysine residues. Following extensive dialysis and sucrose gradient purification, fluorescently labelled ribosomes were purified and supplemented into PURE reactions, encapsulated within GUVs and used to synthesise PR-TolA-SecM. Vesicles were analysed immediately following formation and again following a 60 minute incubation at 37°C to allow the formation of stalled RNCs. Figure 4.6 clearly shows that, following incubation, a proportion of stalled PR-RNCs accumulated at the membrane. Membrane localised ribosome fluorescence was quantified using radial profiling as before and the lumen/membrane ratio was significantly reduced following incubation. These data, taken together with previous results, confirm that the localisation and insertion of PR is a co-translational event and involves protein-inherent recruitment of translating ribosomes to the membrane.

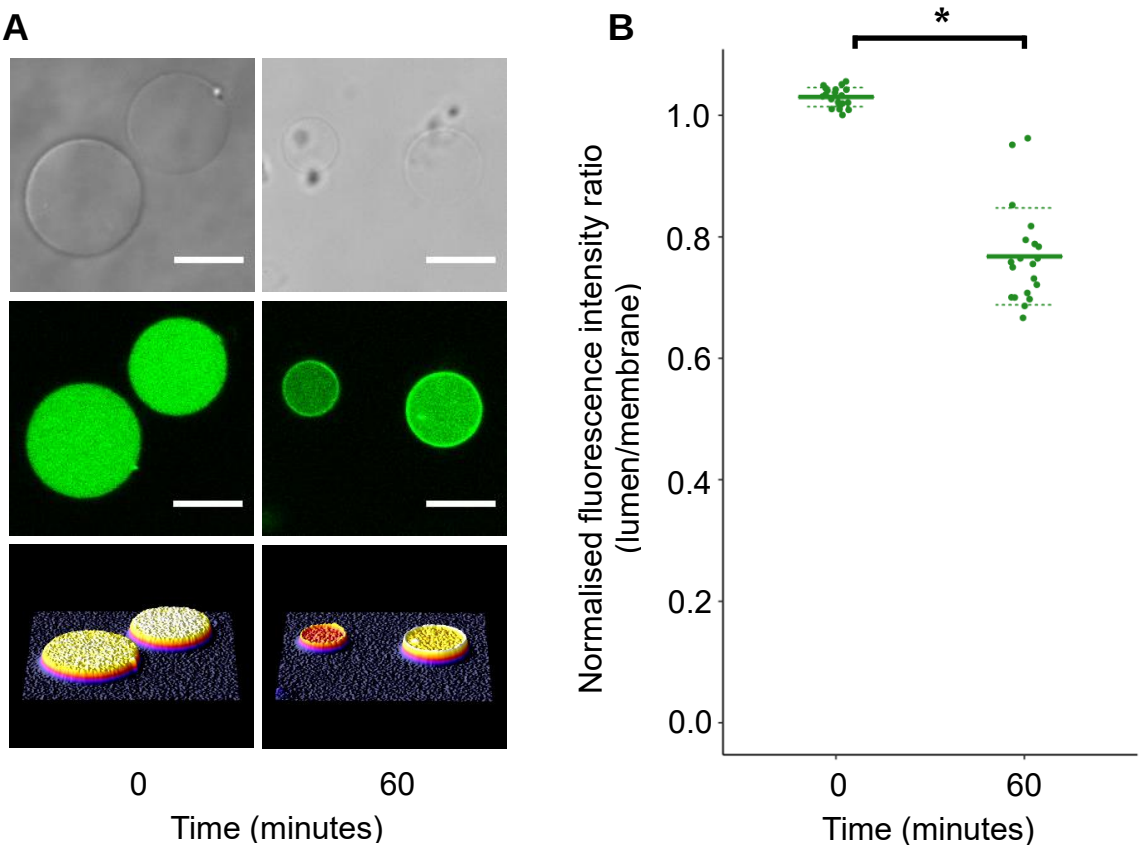


Figure 4.6 Co-translational recruitment of ribosomes to the membrane during translation of PR.

(A) Representative confocal images of GUVs encapsulating PURE reactions supplemented with 1 μM of Alexa Fluor-488 labelled ribosomes synthesising PR-TolA-SecM before and after a 60-minute incubation at 37°C. Scale bars are 10 μm. **(B)** Fluorescence emission radial profiles were generated from 20 individual GUVs from each condition. Data presented are individual, mean and standard deviations. * denotes statistically significant difference ($p < 0.001$) as measured by paired, two-tailed student's t-test.

Previous data had indicated that GalP was also able to localise and insert into GUV membranes and that, like PR, this was not a post-translational event and was not due to localisation of the mRNA. Experiments were therefore performed to analyse the accumulation of stalled GalP RNCs to further confirm that the observed localisation and insertion was co-translational and involved ribosome targeting through a protein-inherent mechanism.

A GalP-TolA-SecM vector was constructed in the same configuration as PR-TolA-SecM (Figure 4.5 A) and was included in bulk PURE reactions alongside GalP as a control for un-stalled translation. Figure 4.7 shows the results of SDS-PAGE

analysis of translational stalling. Translation products for GalP and GalP-TolA-SecM could clearly be observed and band intensity of the SecM containing samples was slightly reduced. Considering that TolA contains 11 lysine residues encoded by AAA codons, which are recognised by fluorescently labelled lysine tRNAs, this suggested that significant translational stalling was taking place.

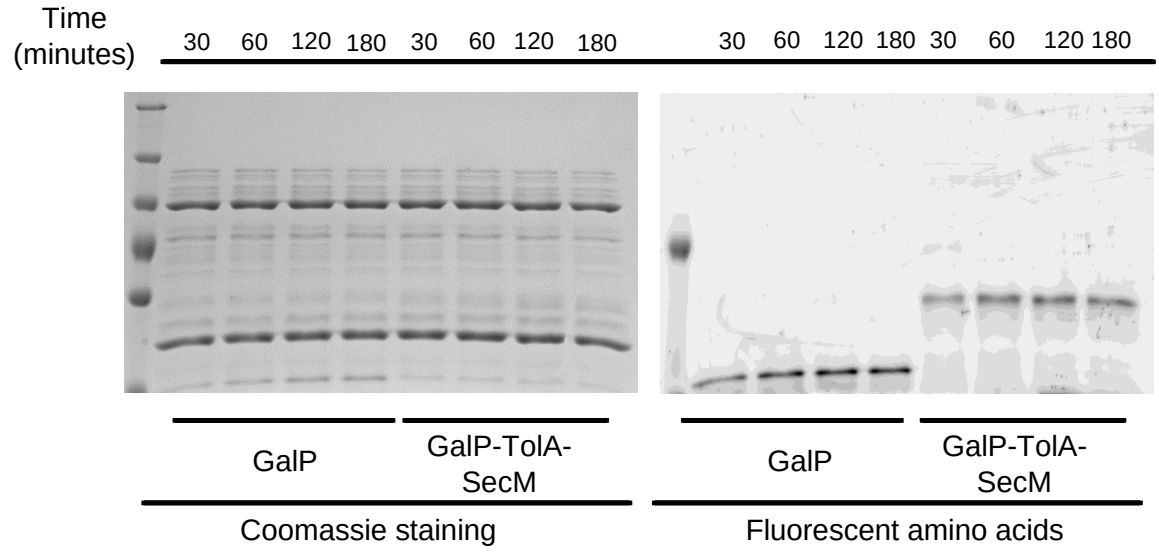


Figure 4.7 Translational stalling of GalP using stalling peptide SecM.

SDS-PAGE time-series analysis of PURE reactions synthesizing GalP and GalP-TolA-SecM. Images are of the same gel stained with coomassie (left) or probed for BODIPY fluorescence (right).

Fluorescently labelled ribosomes were supplemented into PURE reactions, synthesising GalP-TolA-SecM inside GUVs for analysis of membrane localisation of RNCs. Figure 4.8 shows that, following incubation, a proportion of ribosomes were localised to the membrane, as was the case for PR-RNCs observed previously. This confirmed that GalP was also localised and inserted into the membrane in a co-translational manner and that, like PR, this was not due to mRNA localisation but to a protein-inherent ribosome membrane recruitment mechanism.

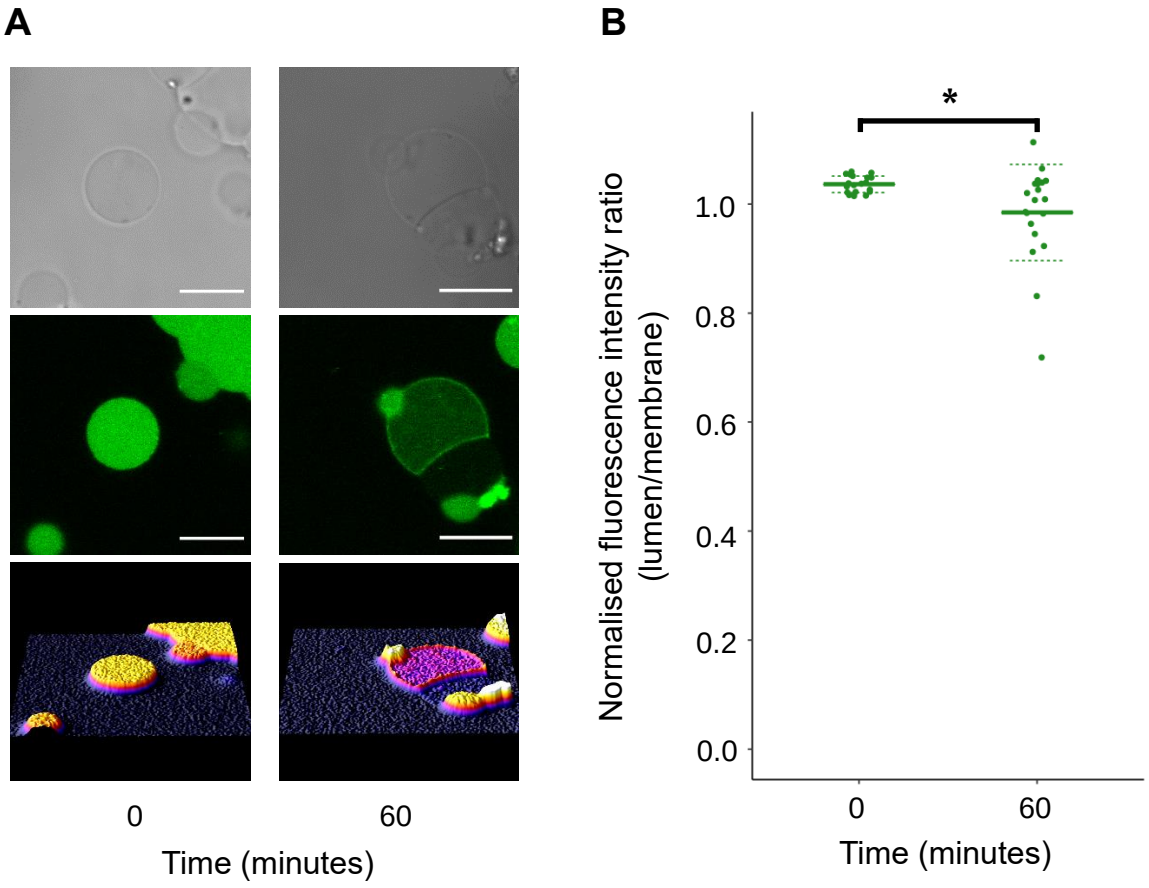


Figure 4.8 Co-translational recruitment of ribosomes to the membrane during translation of GalP.

(A) Representative confocal images of GUVs encapsulating PURE reactions supplemented with 1 μM of Alexa Fluor-488 labelled ribosomes synthesising GalP-TolA-SecM before and after a 60-minute incubation at 37°C. Scale bars are 10 μm. **(B)** Fluorescence emission radial profiles were generated from 20 individual GUVs from each condition. Data presented are individual, mean and standard deviations. * denotes statistically significant difference ($p < 0.001$) as measured by paired, two-tailed student's t-test.

4.2.4 The signal peptide of PR is essential for membrane recruitment of translating ribosomes and protein insertion

All of the data provided up to this point have shown that the spontaneous insertion of membrane proteins PR and GalP into empty GUV membranes is a co-translational process that involves ribosome membrane recruitment due to an innate, protein-driven mechanism.

The N-terminal helix is the first hydrophobic structure to emerge from the ribosome during membrane protein translation, and therefore seemed the logical choice to test for inherent targeting capability. The N-terminus of PR is especially interesting as it codes for a small, cleavable hydrophobic domain that is predicted to form an eighth alpha helix (Figure 4.9 A, B) and has been reported to be recognised by the signal recognition particle (SRP) co-translational targeting mechanism *in vivo* (Soto-Rodriguez and Baneyx, 2019).

To investigate the effects of removal of this domain, the first 18 amino acid residues of PR were removed by PCR to generate the construct PRΔSP (Figure 4.9 C). Bulk PURE reactions were then set up supplemented with 2 mg ml^{-1} POPC LUVs and BODIPY-tagged lysine amino acids for in-gel detection of translation products. Reactions were constructed to synthesise EGFP, full length PR-EGFP and PRΔSP-EGFP. The results of SDS-PAGE analysis are shown in Figure 4.9 D, where translation products were detected corresponding to all of the synthesised proteins (marked with white arrows). The band corresponding to PRΔSP-EGFP was detectable but was markedly weaker than the band corresponding to the full-length protein.

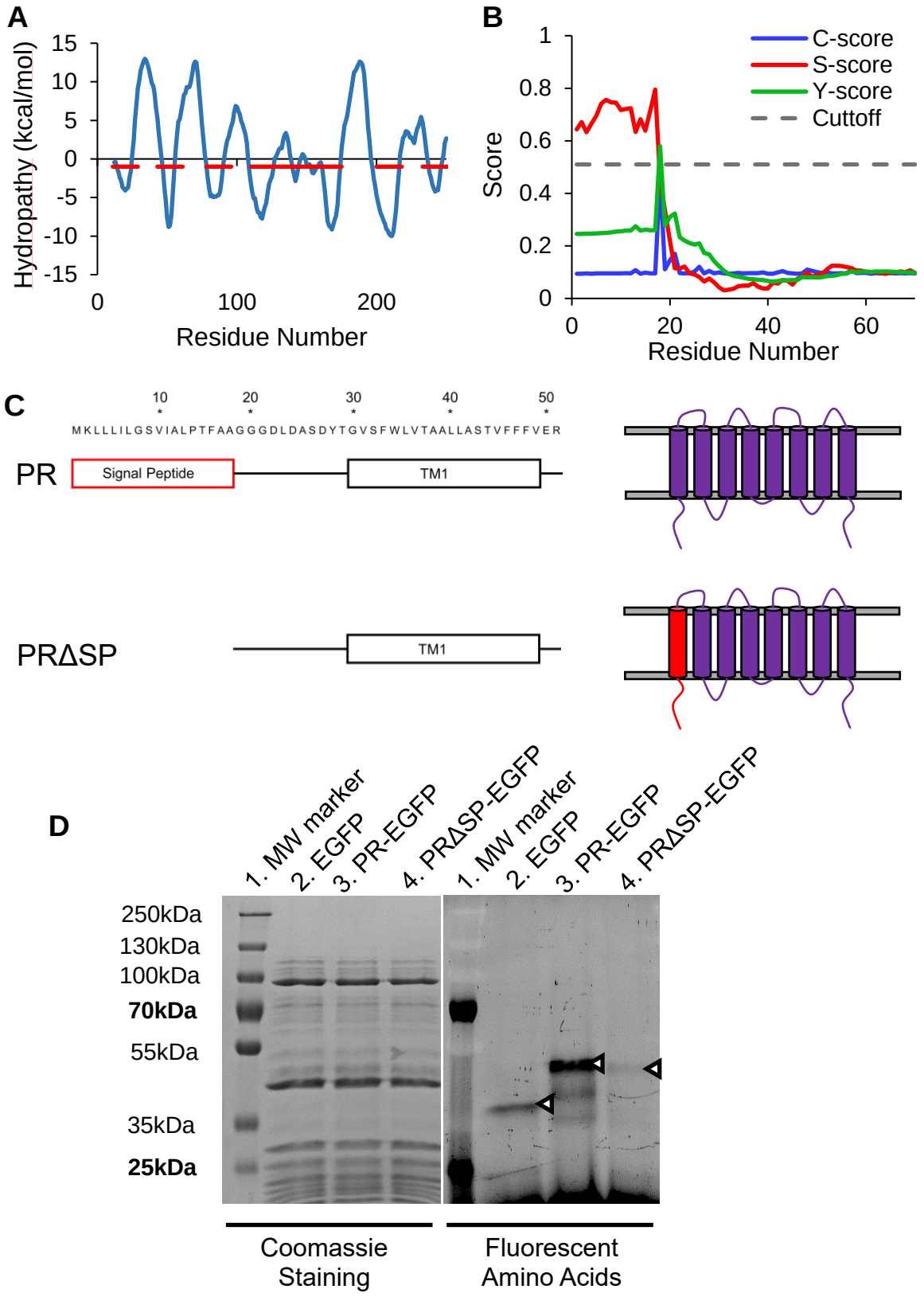


Figure 4.9 PURE synthesis of a truncated form of PR lacking the signal peptide.

See overleaf for figure legend

Figure 4.9 PURE synthesis of a truncated form of PR lacking the signal peptide.

(A) The full-length amino acid sequence of PR was used to generate a hydropathy plot using the MPEx software using the octanol Whimley-White scale (Snider et al., 2009). Red underlines indicate regions predicted to transition from water to the lipid bilayer. **(B)** Signal peptide prediction was performed using the SignalP 4.1 software (Petersen et al., 2011). S-score is signal peptide score, C-score is raw cleavage site score and Y-score is the combined cleavage site score. These combined scores indicate the presence of a signal peptide with a cleavage site at residue 18. **(C)** Schematic representation showing the first 51 amino acid residues of PR indicating the signal peptide and transmembrane helix 1 (TM1). Cartoons show the full-length protein with red portions indicating removed regions. **(D)** SDS-PAGE analysis of bulk PURE reactions supplemented with 20 ng μl^{-1} of pDNA and fluorotect detection reagent. Images are of the same gel imaged for fluorescence and stained with coomassie quick stain. Fluorescent molecular weight bands are labelled in bold.

Synthesis of PRΔSP-EGFP inside POPC GUVs using the PURE system revealed a dramatic loss in membrane localisation of the protein in response to removal of the signal peptide (Figure 4.10). EGFP and PR-EGFP were used as controls for membrane localisation. While the majority of localisation was lost for PRΔSP-EGFP, there was still significantly more than for soluble EGFP indicating that some of the protein was still being recruited to the membrane.

Protein insertion assays were then performed to investigate how this loss of localisation affected the protein's ability to insert into the membrane. Figure 4.11 A and B show that signal peptide removal led to a loss in the membrane lipid order defects that were seen at the outer leaflet for full-length PR. The nile red fluorescence emission ratio for PRΔSP was determined at 2.050 ± 0.118 which was similar to the ratio determined for no protein control GUVs of 2.036 ± 0.105. These ratios were significantly higher than the ratio for full length PR of 1.974 ± 0.130 suggesting a loss in membrane insertion due to removal of the signal peptide.

The reduction of PRΔSP membrane insertion was further confirmed using the HA fluorescence assay. Figure 4.11 C and D show that GUVs synthesising PRΔSP with an HA epitope inserted in the extracellular loop between helices 2 and 3 was unable to bind fluorescently labelled HA, in contrast to the full-length protein containing the same epitope. The ratio of HA present in the extra-vesicular solution and bound to the membrane was determined to be 1.249 ± 0.077 for PRΔSP, significantly higher than the ratio of 0.757 ± 0.267 determined for full-

length PR. The result indicates no binding of fluorescent HA following removal of the signal peptide and thus no translocation of the HA epitope-containing loop region.

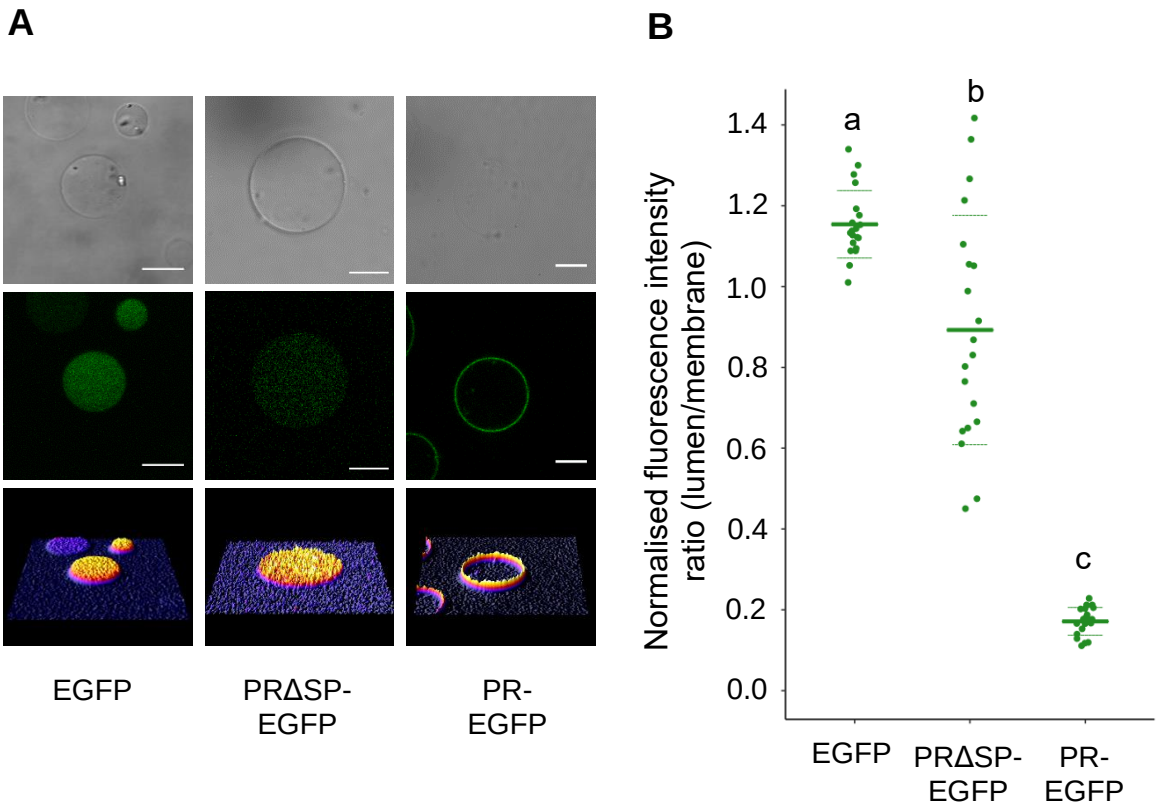


Figure 4.10 PR lacking the signal peptide is unable to localise to GUV membranes.

(A) Representative confocal microscopy images of PURE encapsulating GUVs following synthesis of EGFP, PRΔSP-EGFP and PR-EGFP. Rows of images are; bright field, fluorescence emission and 3-dimensional representations of fluorescence from top to bottom respectively. Scale bars are 10 μm. **(B)** Radial profiles of EGFP fluorescence emission were analysed and are represented as a lumen/membrane fluorescence intensity ratio normalized to background. Data shown are individual values (circles), mean values (solid lines) and standard deviations (dashed lines) from ≥ 20 individual GUVs. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

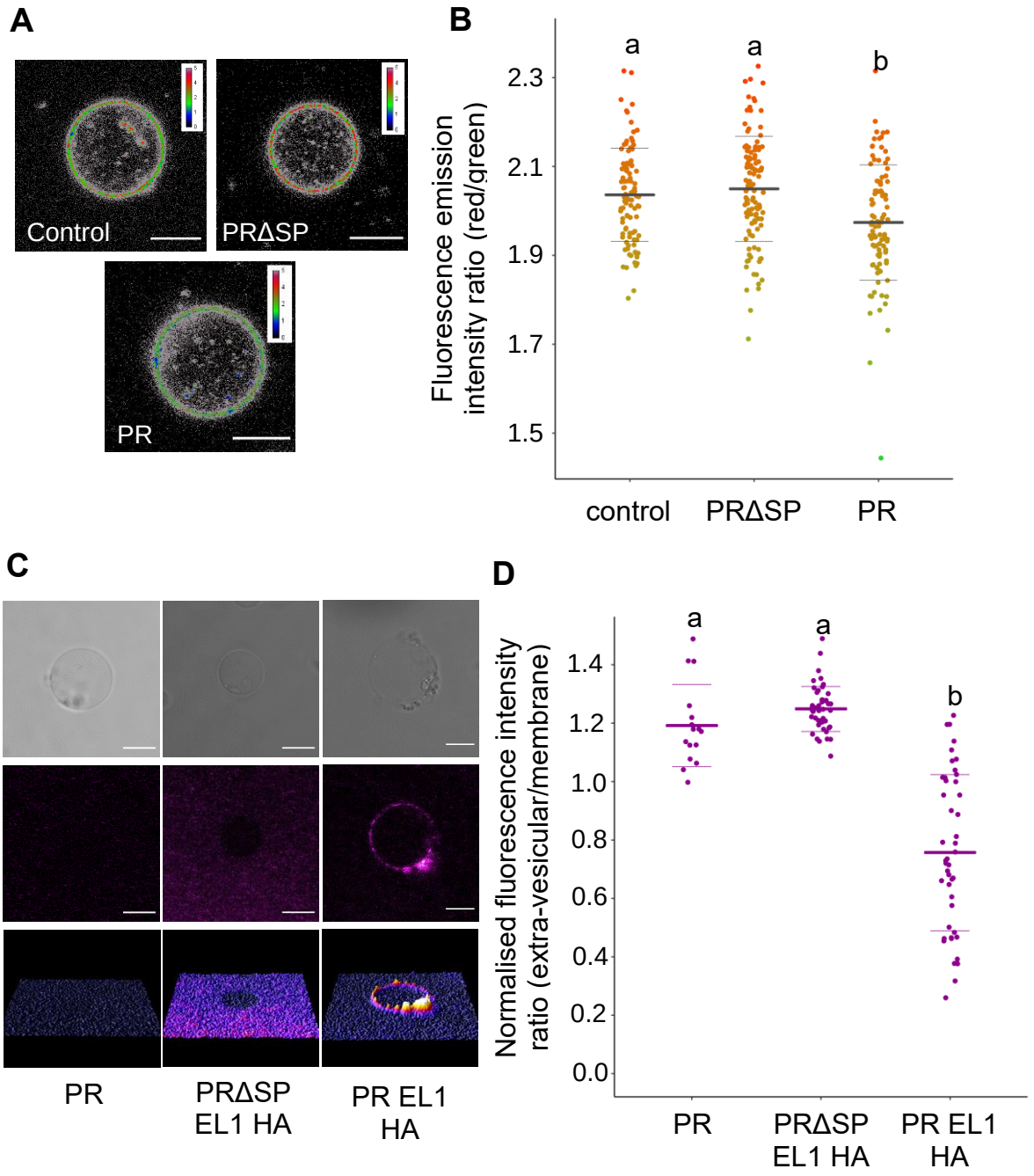


Figure 4.11 PRΔSP is unable to insert into GUV membranes.

(A) Representative confocal images of GUVs containing no protein, PRΔSP and PR treated with 0.1 μM nile red. Fluorescence emission was collected at 510 - 590 nm and from 650 – 750 nm and images were divided following background removal. A rainbow LUT was applied to the resulting composite image to enhance ratio visualisation. Scale bars are 10 μm. **(B)** Quantification of nile red ratio from ≥ 20 GUVs from three independently generated batches. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis. **(C)** Representative confocal images of GUVs following protein synthesis and incubation with Alexa Fluor 647-conjugated HA antibody. Scale bars are 10 μm. **(D)** Quantification of extra-vesicular/membrane fluorescence ratio of Alexa Fluor-647 conjugated HA antibody fluorescence. Values shown are individual, mean and standard deviation of ≥ 25 GUVs. All data were normalised to the

(continued from previous page) lowest ratio observed from 16 GUVs containing PR without an HA epitope. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

4.2.5 The N-terminal helix of GalP is essential for membrane recruitment of ribosomes and protein insertion

Unlike PR, GalP does not contain an N-terminal signal peptide domain. This is in fact the case for the majority of α-helical membrane proteins destined for insertion into the inner membrane of bacteria. For some of these proteins, the first transmembrane helix has been shown to act as a recognition site for chaperones and targeting factors such as the SRP (Guna and Hegde, 2018). Could the first transmembrane helix of GalP also act as a targeting helix in the absence of SRP and other chaperones, in a similar fashion to the signal peptide of PR?

To answer this question a number of truncated GalP constructs were generated via PCR based on the position of transmembrane domains identified by hydropathy analysis (Figure 4.12 A). In the first construct the first transmembrane domain was removed while leaving the soluble N-terminus to be translated (GalP-TM1). In the second construct both the N-terminal region and the first transmembrane domain were removed (GalP-N-TM1). Finally, to test the possibility of hairpin insertion (Lu et al., 2018), a third construct was made in which the first two transmembrane domains were removed along with the N-terminus (GalP-N-TM1+2) (Figure 4.12 B). The truncated proteins were synthesised in bulk PURE reactions and analysed on 10% polyacrylamide gels. Figure 4.12 C shows the resulting SDS-PAGE gel analysis. Synthesis of GalP-mCherry and GalP-TM1-mCherry was clearly visible in lanes 2 and 3 of the resulting gel. Although removal of the N-terminal region resulted in decreased protein yield the protein bands corresponding to the expected molecular weights for GalP-N-TM1-mCherry and GalP-N-TM1+2-mCherry were still detectable following adjustment of image brightness and contrast settings.

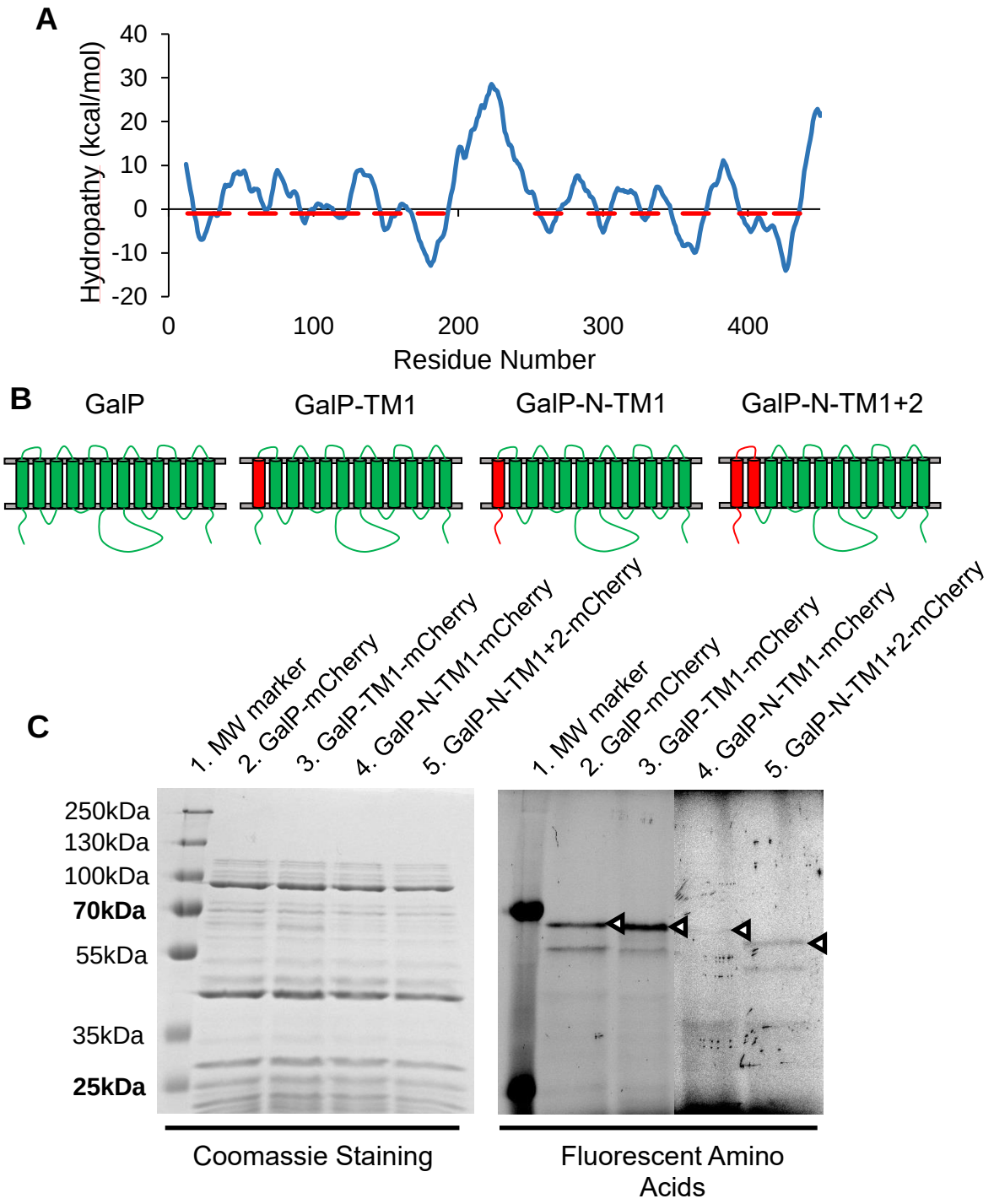


Figure 4.12 Synthesis of GalP N-terminal truncations mutants in bulk.

(A) A hydropathy plot was generated for GalP using MPEx and the octanol Whimley-White scale (Snider et al., 2009). Red underlines indicate predicted transmembrane regions. **(B)** Cartoons showing full length GalP and truncations, with red portions indicating removed regions. **(C)** SDS-PAGE analysis of bulk PURE reactions supplemented with 20 ng μl^{-1} of pDNA and fluorotect detection reagent. Images are of the same gel imaged for fluorescence and stained with coomassie quick stain. Fluorescent molecular weight bands are labelled in bold. The brightness and contrast of lanes 4 and 5 of the right hand gel have been adjusted due to low levels of protein synthesis. Protein bands are marked with white arrows.

Localisation of the various GalP truncations was then investigated in GUVs encapsulating PURE reactions. Figure 4.13 clearly shows that the absence of the first transmembrane helix has a severe effect on protein targeting regardless of the presence or absence of the hydrophilic N-terminal region. Radial fluorescence profiling was used to determine lumen to membrane fluorescence ratios of 0.679 ± 0.142 for GalP-TM1-mCherry; 0.717 ± 0.246 for GalP-N-TM1-mCherry and 0.622 ± 0.170 for GalP-N-TM1+2-mCherry, all of which were significantly higher than the ratio of 0.365 ± 0.082 determined for full length GalP-mCherry.

The loss of targeting was indicative of a loss of downstream transmembrane helix insertion as was the case for PR. Figure 4.14 shows the results for HA loop translocation following the synthesis of GalP truncation mutants with an HA epitope inserted into extracellular loop 3 of the protein. Quantification of fluorescently labelled HA antibody binding resulted in extra-vesicular/membrane fluorescence ratios of 1.052 ± 0.129, 1.155 ± 0.076 and 0.953 ± 0.133 for GalP-TM1, GalP-N-TM1 and GalP-N-TM1+2 respectively, following normalisation to a GalP (no HA) negative control. These results were not significantly different from the negative control ratio of 1.091 ± 0.071 but were significantly different from the positive control ratio of 0.819 ± 0.206 indicating a lack of insertion. The only exception was GalP-N-TM1+2 which was not significantly different ($p \geq 0.001$) from either the positive or negative control suggesting some of this truncated protein was inserted.

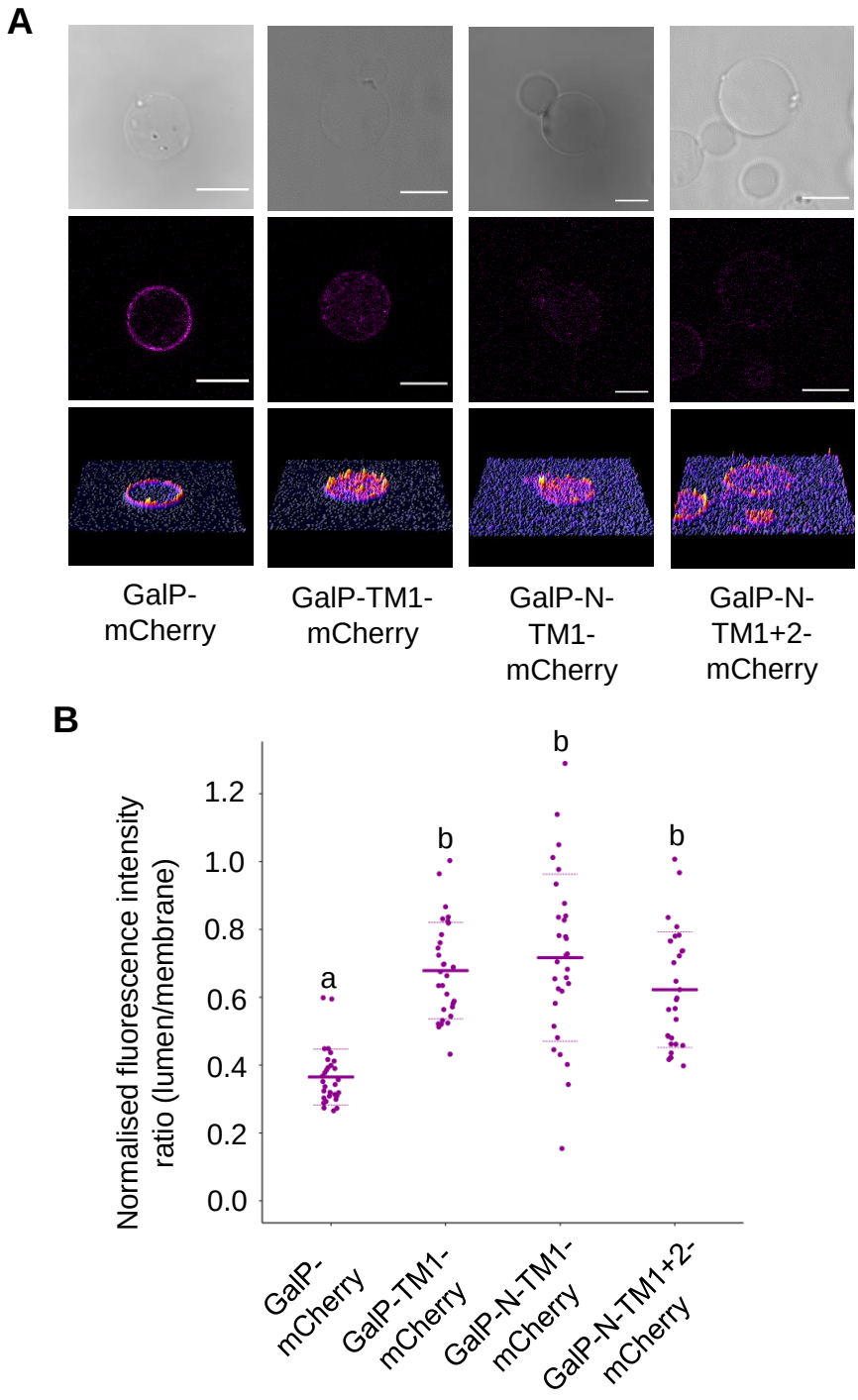


Figure 4.13 Transmembrane helix one of GalP is important for co-translational protein targeting.

(A) Representative confocal microscopy images of PURE encapsulating GUVs following synthesis of GalP-mCherry, GalP-TM1-mCherry, GalP-N-TM1-mCherry and GalP-N-TM1+2-mCherry. Scale bars are 10 μm. **(B)** Radial profiles of mCherry fluorescence emission were analysed and are represented as a lumen/membrane fluorescence intensity ratio normalised to background. Data shown are individual values (circles), mean values (solid lines) and standard deviations (dashed lines) from ≥ 30 individual GUVs. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

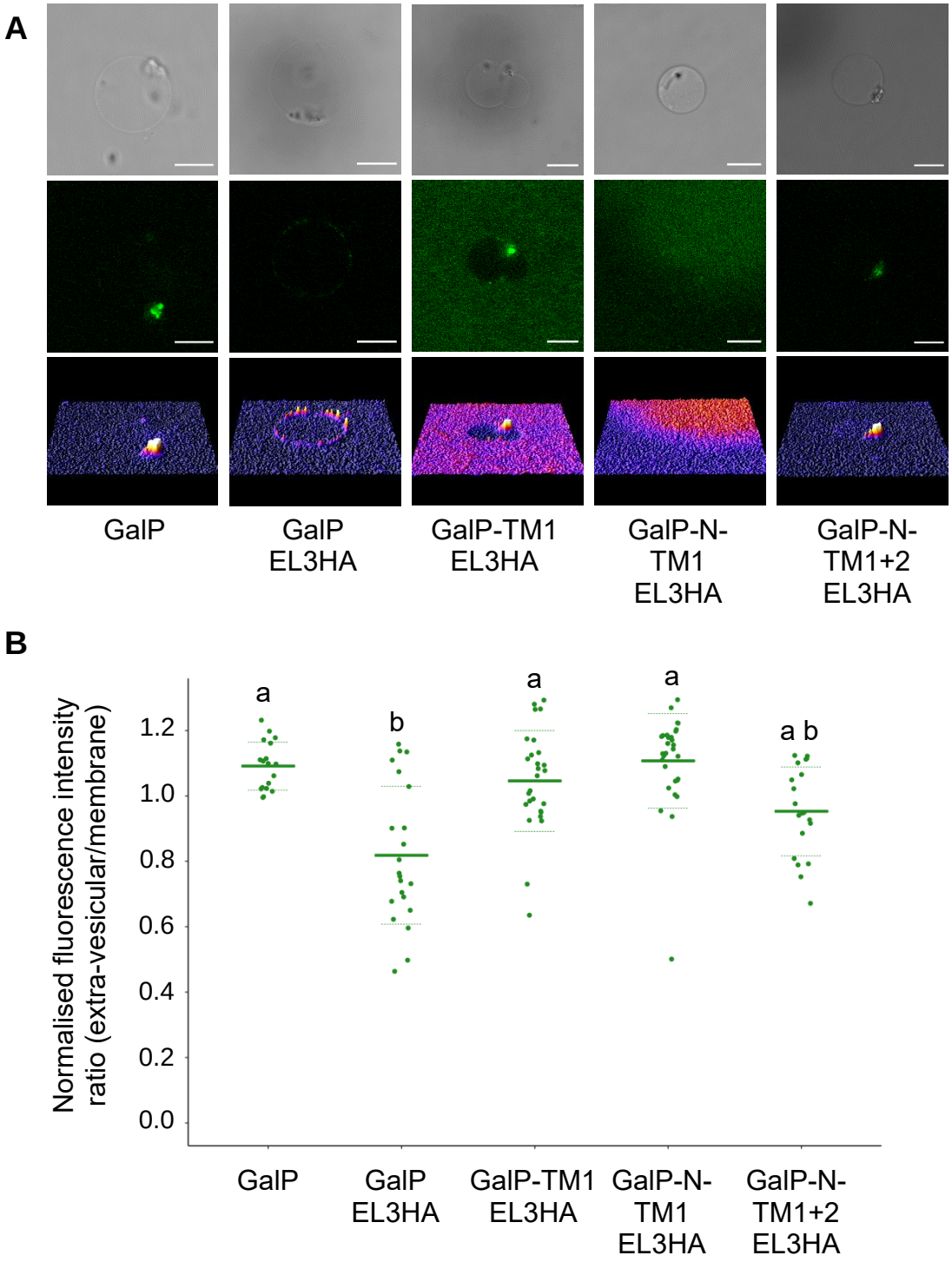


Figure 4.14 Mislocalised GalP mutants are no longer inserted into GUVs.

(A) Representative confocal images of GUVs following protein synthesis and incubation with Alexa Fluor 488-conjugated HA antibody. Scale bars are 10 μm. **(B)** Quantification of extra-vesicular/membrane fluorescence ratio of Alexa Fluor-488 conjugated HA antibody fluorescence. Values shown are individual, mean and standard deviation of ≥ 20 GUVs. All data were normalised to the lowest ratio observed from 16 GUVs containing GalP without an HA epitope. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

4.2.6 The signal peptide of PR has a high affinity for the membrane and guides downstream helix insertion

Previous data had shown that the presence of the signal peptide was essential for the targeting of translating ribosomes to the membrane and the subsequent insertion of downstream helices. To examine this in more detail constructs were generated to express the signal peptide alone, the first TM domain alone and the signal peptide with the first TM domain. Analysis of the localisation of these proteins in GUVs (Figure 4.15 A, B) showed that the signal peptide alone was able to localise to the membrane quite efficiently, with a fluorescence ratio determined at 0.437 ± 0.058. The first TM helix of PR was unable to localise to the membrane despite the lack of the downstream helices resulting in a ratio of 1.002 ± 0.284. Fluorescence yield was also greatly reduced when synthesising PR TM1 in isolation. Synthesis of the signal peptide with TM1 resulted in a recovery of membrane localisation with a fluorescence ratio of 0.424 ± 0.099.

Membrane insertion analysis was then performed to determine whether the signal peptide was able to span the membrane in the absence of downstream helices. Ratiometric fluorescence experiments were performed using the previously established nile red assay. Expression of the signal peptide alone resulted in a ratio of 2.047 ± 0.118, which was not significantly different from control, vesicles with no membrane inserting protein present. Synthesis of the signal peptide and TM1 however resulted in a ratio of 1.894 ± 0.117 which was significantly ($p < 0.001$) blue-shifted compared to control vesicles and signal peptide only vesicles (Figure 4.15 C, D).

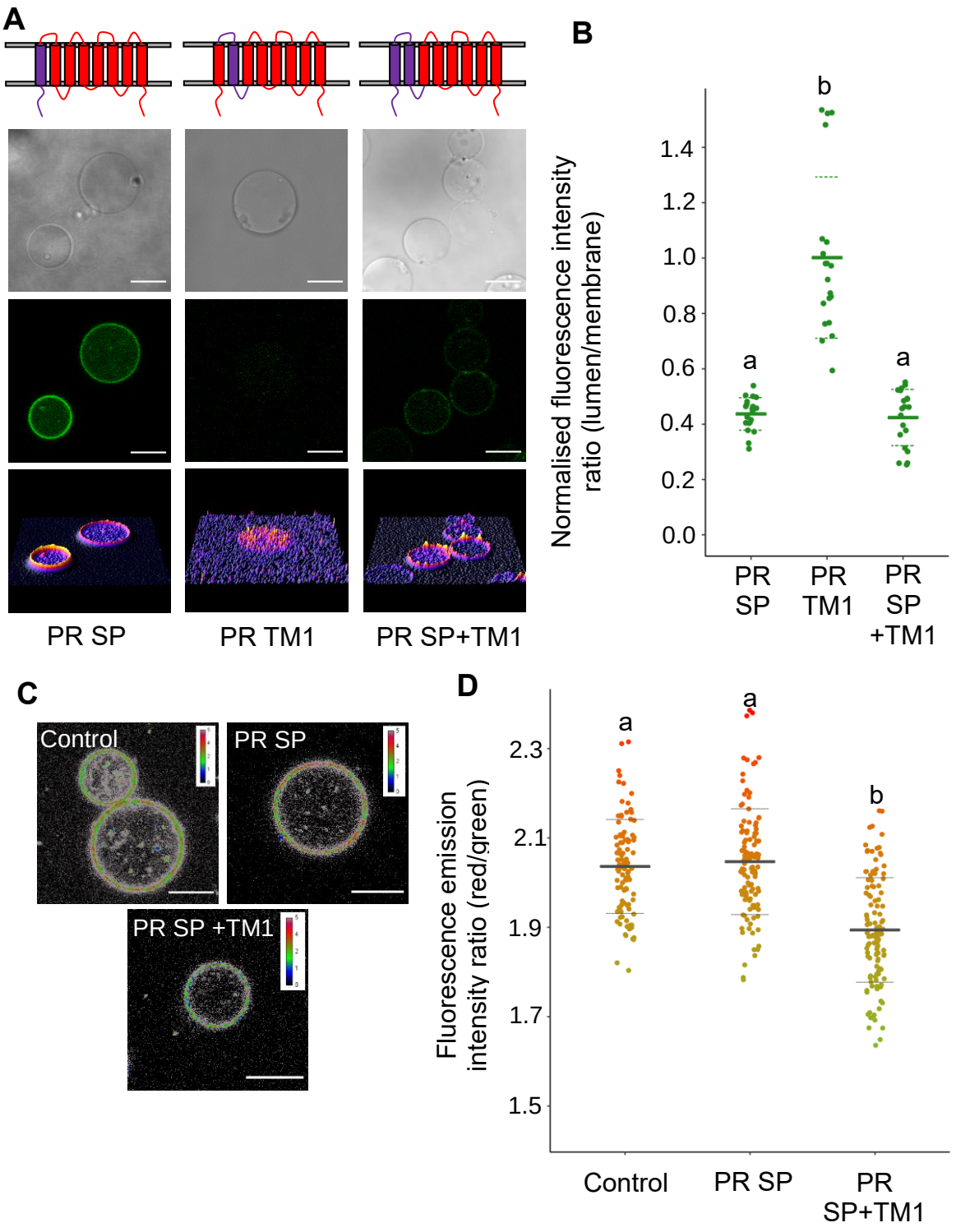


Figure 4.15 The signal peptide of PR is required for localisation and insertion of TM1.

(A) Cartoons representations and respective confocal images of construct where purple sections indicate regions present and red regions that have been removed. Scale bars are 10 µm. **(B)** Radial fluorescence profiles for 20 individual GUVs expressing each construct. Different letters denote statistically significant differences ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis. **(C)** Representative confocal images of GUVs containing no protein, PR SP and PR SP+TM1 treated with 0.1 µM nile red. A rainbow LUT

(*continued from previous page)* was applied to the resulting composite image to enhance ratio visualisation. Scale bars are 10 µm. **(D)** Quantification of nile red fluorescence emission ratio from ≥ 30 individual GUVs from three individually prepared batches. Different letters denote statistically significant ($p < 0.001$) differences as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

4.2.7 Mutational analysis of PR signal peptide and effects on membrane targeting

In order to investigate the physical properties of the PR signal peptide that were responsible for the observed membrane targeting effects, single amino acid truncations were introduced from the N-terminus along the entire length of the signal peptide, while maintaining the presence of the mature protein downstream (Table 4.1).

Table 4.1 Proteorhodopsin signal peptide truncations generated by PCR to remove single amino acid residues from the N-terminus while maintaining the start codon.

Contrsuct	Signal Peptide	Mature Protein
PR	M K L L L I L G S V I A L P T F A A	G G G D L D A S D Y T
PRΔ1	M L L L I L G S V I A L P T F A A	G G G D L D A S D Y T
PRΔ2	M L L I L G S V I A L P T F A A	G G G D L D A S D Y T
PRΔ3	M L I L G S V I A L P T F A A	G G G D L D A S D Y T
PRΔ4	M I L G S V I A L P T F A A	G G G D L D A S D Y T
PRΔ5	M L G S V I A L P T F A A	G G G D L D A S D Y T
PRΔ6	M G S V I A L P T F A A	G G G D L D A S D Y T
PRΔ7	M S V I A L P T F A A	G G G D L D A S D Y T
PRΔ8	M V I A L P T F A A	G G G D L D A S D Y T
PRΔ9	M I A L P T F A A	G G G D L D A S D Y T
PRΔ10	M A L P T F A A	G G G D L D A S D Y T
PRΔ11	M L P T F A A	G G G D L D A S D Y T
PRΔ12	M P T F A A	G G G D L D A S D Y T
PRΔ13	M T F A A	G G G D L D A S D Y T
PRΔ14	M F A A	G G G D L D A S D Y T
PRΔ15	M A A	G G G D L D A S D Y T
PRΔ16	M A	G G G D L D A S D Y T
PRΔSP	M	G G G D L D A S D Y T

The localisation of these various signal peptide truncations was then tracked using a C-terminal EGFP following synthesis within GUVs. Figure 4.16 shows the resulting localisation effects of each of the signal peptide truncations. A significant loss in membrane localisation was seen from PRΔ14 onwards, indicating that these four amino acids were essential for membrane localisation. The addition of upstream amino acids led to varying degrees of localisation recovery. A complete recovery in localisation capacity was not observed until all hydrophobic residues were included (PRΔ1). The positive lysine residue immediately downstream of the start codon methionine did not negatively affect membrane localisation.

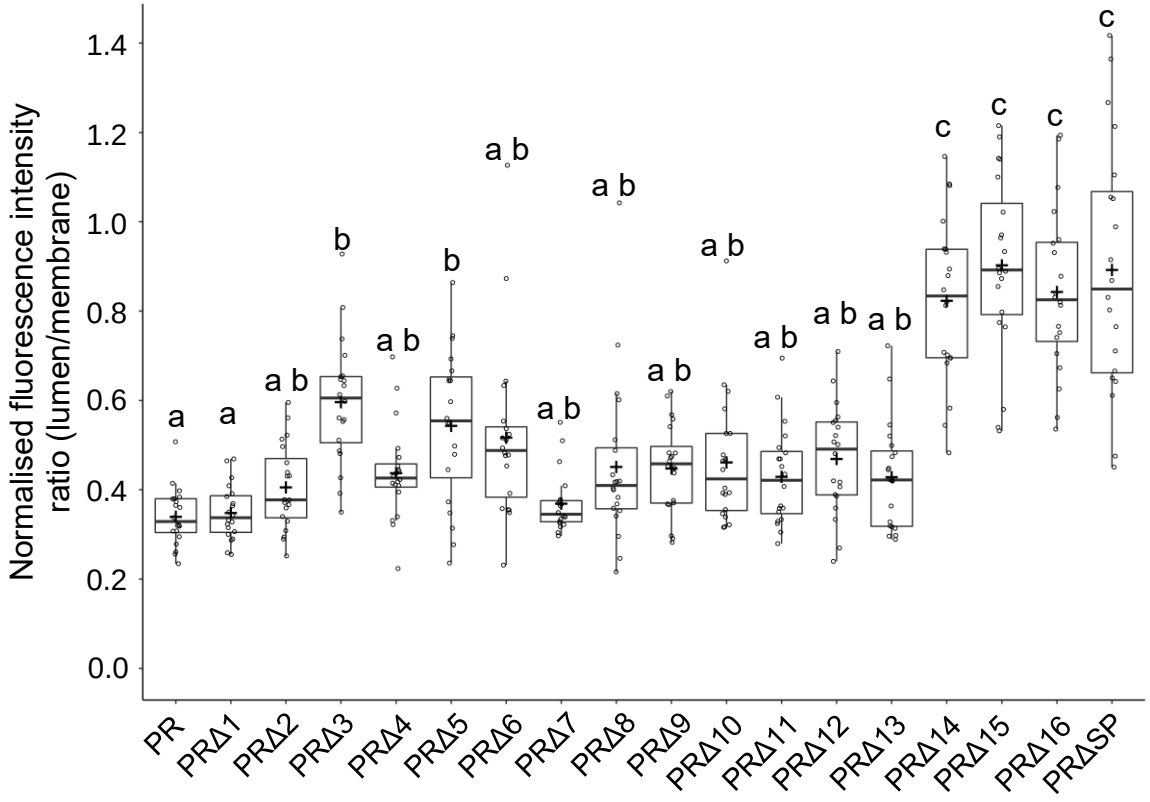


Figure 4.16 Localisation of PR-EGFP signal peptide truncations synthesised inside GUVs.

Localisation of EGFP fluorescence was quantified using radial profiling of confocal microscopy images. 20 individual GUVs for each signal peptide truncation were measured as well as full length protein (PR) and PR with the signal peptide completely removed (PRΔSP). Different letters denote statistically significant ($p < 0.001$) differences as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis. All individual data points are shown along with the mean (+), median and interquartile range. Whiskers extend to 1.5x the interquartile range.

These data implicated a threonine residue at position 15 of the PR signal peptide as being important for membrane localisation. A construct containing a single amino acid substitution was therefore generated by overlapping PCR to produce PR T15A where threonine was replaced by alanine. This point mutant was then synthesised inside GUVs and its localisation was analysed as before. The lumen/membrane fluorescence ratio was measured as 0.556 ± 0.239, significantly higher than the previously determined ratio for full length PR but significantly lower than the ratio for PRΔ14 (Figure 4.17 A).

Protein structural predictions were carried out using the RaptorX software (Kaellberg et al., 2012) where the first 49 amino acids were supplied to give coverage of the signal peptide and the entire TM1 helix of PR. Structures were predicted for native PR (Figure 4.17 B) and PR T15A (Figure 4.17 C). Regions around the mutation site were analysed using ChimeraX visualisation software (Goddard et al., 2018) and predicted hydrogen bonds were included as dashed yellow lines (Figure 4.17 B, C insets).

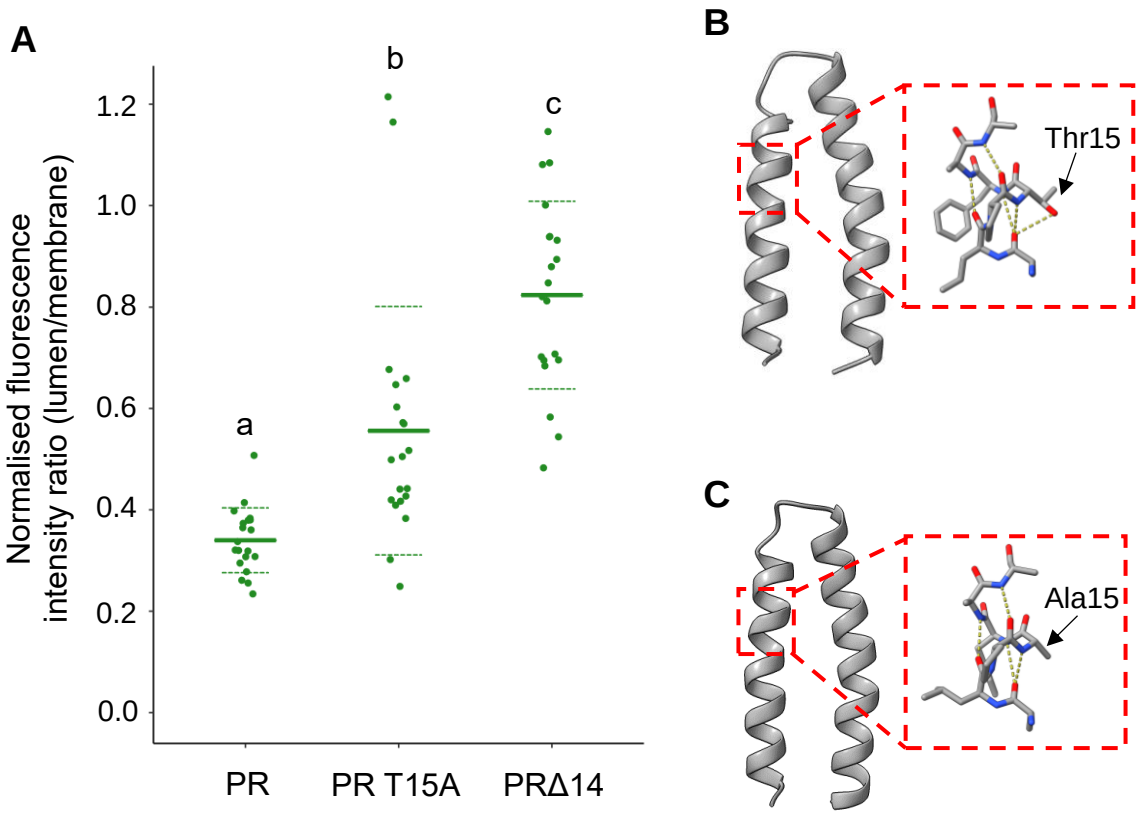


Figure 4.17 The single amino acid substitution T15A affects membrane localisation.

(A) Quantification of membrane localised fluorescence signal using radial profiling of confocal images. Fluorescence was measured from 20 individual GUVs expressing PR T15A and results were compared to previously determined ratios for PR and PRΔ14. Different letters denote statistically significant ($p < 0.01$) differences as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis. **(B)** Protein structure for native PR was predicted using RaptorX software (Kaellberg et al., 2012) and visualised using ChimeraX (Goddard et al., 2018). The left helix corresponds to the signal peptide and the right helix to TM1. Inset shows the hydrogen bonding environment immediately surrounding Thr15. **(C)** Protein structure for PR T15A was predicted and visualised as in (B). The inset again shows the hydrogen bonding environment around the mutated Ala15 residue.

4.3 Discussion

Membrane protein targeting to the bacterial plasma membrane and the eukaryotic ER has been extensively studied *in vivo*. The major conserved pathway for the co-translational targeting of membrane proteins is the SRP pathway (Nagai et al., 2003). The ribonecleoprotein SRP recognises and binds to hydrophobic segments as they appear from the ribosome exit tunnel, and facilitates the recruitment of the translating ribosome to the translocon complex for subsequent insertion (Angelini et al., 2005). A growing number of membrane proteins have now been shown to spontaneously integrate into lipid bilayers without the assistance of the SRP for targeting or the translocon for insertion (Harris et al., 2017; Iwamoto et al., 2018; Berhanu et al., 2019). Very little is known about which physical properties of the proteins make this possible. In the absence of the SRP or other chaperones, the hydrophobicity must be the main driver for the proteins to partition into the membrane environment (Kiefer and Kuhn, 1999), but it also increases the danger for aggregation and misfolding if the nascent protein is exposed to the aqueous phase. The research described in this chapter has therefore focussed on identifying the fundamental protein features that drive the targeting and integration process for PR and GalP. The results obtained drew particular interest to the first hydrophobic region appearing from the ribosome during translation, the signal peptide of PR and the first transmembrane spanning domain (TM1) of GalP.

As mentioned previously, membrane protein targeting and insertion are expected to be co-translational processes (Cymer and von Heijne, 2013). However, it was not known whether some post-translational membrane targeting of aggregates was occurring which could explain the high levels of membrane association reported for PR and GalP when using the PURE system in GUVs. The addition of a bilayer environment after protein synthesis, and the subsequent lack of membrane bound signal for both PR and GalP ruled out post-translational targeting or insertion as the prevailing mechanism (Figures 4.1 and 4.2), thus lending support to the co-translational hypothesis in this minimal environment.

Association of membrane protein-coding mRNAs with the membrane has been observed *in vivo* in both prokaryotes and eukaryotes (Nevo-Dinur et al., 2011; Holt and Bullock, 2009). It is anticipated that this association is due to a proteinaceous receptor-mRNA complex, but the intrinsic properties of the mRNA

should also be considered. RNA can indeed adopt structures that lead to membrane association in the absence of any receptor proteins (Khvorova et al., 1999; Janas and Yarus, 2003). Localisation of PR and GalP encoding mRNA transcripts was analysed here using the aptamer Spinach2 along with the conditionally fluorescent dye DHFBI (Strack et al., 2013). Figures 4.3 and 4.4 show the results for PR and GalP respectively. In both cases, bulk reactions without ribosomes showed increasing levels of DHFBI fluorescence indicating the synthesis of mRNA-containing Spinach2. Fluorescence of GalP mRNA was stable throughout the 180 minute experiment reaching a plateau at ~120 minutes whereas PR mRNA fluorescence decreased slightly after reaching a peak at ~80 minutes. It is possible that PR mRNA is less stable than GalP mRNA but further experiments would need to be performed to confirm this trend. The experiment clearly showed that Spinach2-containing mRNA was being synthesised and that DHFBI could be used to fluorescently label it. Thus, mRNA localisation was tracked following synthesis in GUVs. Membrane localisation was not observed for either PR-mRNA or GalP-mRNA suggesting that the preferential mRNA localisation at the membrane was not responsible for the observed protein localisation.

To further confirm that localisation of PR and GalP was co-translational and involved the targeting of translating ribosomes, an assay was developed to track stalled RNCs. The hypothesis was that membrane protein synthesising RNCs would accumulate at the membrane. Stalled RNCs have been used before to probe co-translational folding and translocation events (Cymer and von Heijne, 2013; Ismail et al., 2015) and to analyse the early stages of co-translational protein folding (Nilsson et al., 2015). The best studied stalling peptide is a 17 amino acid sequence taken from the *E. coli* SecM protein. It stalls ribosomes following the translation of its C-terminal proline residue which alters the shape of the peptidyl transferase centre of the ribosome (Bhushan et al., 2011). DNA encoding this sequence was attached to the 3' ends of PR and GalP following an 84 residue unstructured linker sequence from *E. coli* TolA. Bulk reactions were then performed and SDS-PAGE analysis was used to confirm stalling. Figures 4.5 and 4.7 show stalling results for PR and GalP respectively. Both proteins showed stalling upon introduction of SecM, reflected in reduced protein synthesis in the SecM containing samples. A protein band was still readily detectable in GalP-

TolA-SecM samples, but the band intensity was lower than for GalP alone despite the higher number of labelling sites due to the presence of TolA. Localisation analysis in GUVs (Figures 4.6 and 4.8) clearly showed that production of stalled RNCs led to the accumulation of fluorescently labelled ribosomes at the membrane of GUVs.

Taken together, the data showed that the observed protein localisation and insertion was due to the co-translational recruitment of ribosomes to the membrane by nascent chain-inherent features. The question remained, however, what were these features?

Since the PURE system lacks any chaperoning elements such as the SRP, and the liposomes do not contain any membrane bound receptors or translocon apparatus, the hydrophobic effect must have been the driving force behind the targeting of PR and GalP. However, given the large aqueous volume of the GUVs and the propensity for membrane helical aggregation in aqueous environments, the efficient localisation was surprising. This led to the hypothesis that the first hydrophobic helix to emerge from the ribosome was important and sufficient for co-translational targeting. PR is unusual for an inner membrane protein in that it contains an N-terminal hydrophobic signal sequence (Figure 4.9). *In vivo* evidence from *E. coli* has shown that the signal sequence is important for protein yield in the membrane fraction, but the mechanism of action remained elusive (Soto-Rodriguez and Baneyx, 2019). Section 4.2.4 shows the effects of signal peptide removal on the localisation and insertion of PR in GUVs. Signal peptide removal led to a drop in protein synthesis yield in the presence of LUVs and when synthesised in GUVs there was a complete loss of localisation to the membrane. There was also a complete loss in detectable levels of insertion by either the nile red assay or the HA epitope translocation assay. These results indicated that the protein was being synthesised but was unable to insert into the membrane, likely due to aggregation in the vesicle lumen. This was surprising since the signal peptide is less hydrophobic than the majority of the other helices. It seemed likely that the signal peptide had a higher affinity for the lipid environment and probably the interface region than for itself, whereas the downstream helices were more likely to self-aggregate than target to the interface. This fundamental, N-terminal helical targeting mechanism could have evolved to lower the risk of aggregation of polytopic membrane proteins.

The data on PR led to the question of whether or not the first transmembrane helix of GalP plays a similar role in targeting for efficient insertion of downstream helices. Section 4.2.5 shows the effects of TM1 removal on GalP localisation and insertion. Because of a long soluble N-terminal domain in GalP two constructs were generated to remove TM1, one maintaining the N-terminal soluble domain (GalP-TM1) and one removing the N-terminus together with TM1 (GalP-N-TM1). Since insertion of helical hairpins could not be ruled out (Lu et al., 2018) a further construct was generated removing both TM1 and TM2 as well as the N-terminus (GalP-N-TM1+2). Protein yield was only decreased upon the removal of the soluble N-terminal domain, however, localisation was severely impacted in all of the truncation mutants. This loss of protein localisation was again followed by a loss of protein insertion as measured by the HA epitope translocation assay. There was a slight improvement in detectable levels of insertion when both TM1 and TM2 were removed suggesting that some of this protein could insert properly through the insertion of pairs of helical hairpins (Engelman and Steitz, 1981). The observed effect of GalP-N-TM1+2 was not significantly different from either negative or positive controls suggesting that a larger sample size would be required to resolve differences. The combined data suggest that the N-terminal helix of GalP plays an important fundamental role in co-translational targeting of this protein to the membrane, similar to that of the signal peptide of PR.

Together the results indicate a role for the N-terminal helices of various membrane proteins in efficient membrane recruitment of translating ribosomes in the absence of SRP-based targeting or any other chaperoning factors. It is possible that this inherent targeting ability is also present *in vivo* and that other targeting pathways such as the SRP evolved to further aid the process and to introduce targeting specificity in increasingly complex environments (Bohnsack and Schleiff, 2010).

The signal peptide of PR was then analysed in more detail to determine how it is able to aid membrane targeting. First, constructs were generated to analyse the signal peptide in isolation, TM1 in isolation, and the signal peptide plus TM1. The signal peptide was found to localise to the membrane even in the absence of any downstream helices, while TM1 was unable to localise to the membrane. The re-attachment of the signal peptide to TM1 led to a recovery of membrane

localisation (Figure 4.15 A, B). Insertion was probed using the nile red assay and results indicated that the signal peptide alone was unable to perturb the outer leaflet, however when attached to TM1 perturbation did occur. These data together confirmed the aggregation propensity of PR TM1 and the targeting effects of the signal peptide, while also suggesting that the signal peptide is localised mainly to the interface of the bilayer and does not deeply penetrate into the hydrophobic core in the absence of a downstream helix. Previous research has shown the lipid affinity of various signal peptides (Briggs et al., 1985; Hoyt and Gierasch, 1991), but this feature was never considered from a targeting perspective and went relatively unstudied following the discovery of the Sec translocon. More experiments and simulation studies are now required to further elucidate the bilayer interaction of the PR signal peptide.

A truncation analysis was then performed to analyse the effects of amino acid removal on localisation of PR. Table 4.1 shows the sequence of each of the truncations that were generated by PCR. Figure 4.16 then shows how these truncations affected protein localisation. Removing the single positively charged residue from the N-terminus had no effect on protein localisation, however a potential role in guiding topology would require further testing of insertion, since single charge alterations have been shown to affect membrane protein topology (Seppala et al., 2010). Removing N-terminal hydrophobicity had modest effects on targeting up until the removal of Thr15 (PRΔ14) where major localisation defects were seen. To test the importance of this residue in the whole protein context, a single point mutation was introduced into the native PR sequence to replace Thr15 with Ala. This point mutant was also defective in membrane localisation albeit to a lesser extent than observed for PRΔ14. What was clear was that the four amino acids beyond Pro14 were essential for the observed targeting effect. Threonine is able to form hydrogen bonds and may stabilise structure formation (Figure 4.17), but may also have an anchoring effect through hydrogen bonding directly to lipid headgroups (Bondar and White, 2012). Immediately downstream of this threonine is a phenylalanine which may help by anchoring the translating ribosome to the interface through formation of a cation-π interaction between the aromatic ring of phenylalanine and the positively charged quaternary amine present in the choline molecule of POPC lipids (Broemstrup and Reuter, 2010).

In conclusion, this chapter has shown that targeting and insertion of PURE synthesised PR and GalP into GUV membranes occurs in a co-translational manner and is controlled by a protein-inherent mechanism. The first hydrophobic domain to emerge from the ribosome can facilitate the membrane recruitment of ribosomes synthesising both of these proteins in the absence of any targeting chaperones recognising these sequences. In-depth analysis of the signal peptide of PR identified the final four residues as essential for the targeting mechanism possibly due to interactions with lipid headgroups or other protein regions.

5. Membrane tethered ribosomes rescue aberrant membrane protein localisation and insertion

5.1 Introduction

The recruitment of ribosomes to the membrane is of critical importance *in vivo* for efficient biogenesis of membrane proteins. As mentioned in previous chapters, this is a highly controlled process involving various proteinaceous chaperoning components (Alder and Johnson, 2004; Luirink et al., 2012). Examining this process *in vivo* is challenging due to the complex responses mounted by both eukaryotic and prokaryotic cells in response to the removal of targeting machineries. For example, Wickstrom et al. (2011) reported that depletion of the *E. coli* SRP homolog Ffh resulted in an induction of the σ^{32} stress response due to the formation of cytoplasmic protein aggregates and led to a subsequent increase in the levels of cytoplasmic chaperones DnaK and GroEL/ES. Mutka and Walter (2001) found that upon depletion of SRP54 in *S. cerevisiae* there was enhanced expression of a number of genes encoding heat shock chaperones and concomitant repression of genes involved in protein synthesis.

A potential solution to the targeting problem is to permanently attach ribosomes to the membrane, although this would compromise the high rates of cytoplasmic protein synthesis that are required by most cells. An *in vivo* example of this strategy can be found in the mitochondria of yeast. Mitochondria maintain a small genome encoding a small subset of highly hydrophobic, polytopic membrane proteins of the respiratory electron transport chain (Solieri, 2010). Mitochondrial ribosomes are permanently attached to the inner mitochondrial membrane via interactions with the YidC homolog Oxa1 (Bonnefoy et al., 2009). Furthermore, lack of a Sec-type translocon in the mitochondrial inner membrane (Glick and VonHeijne, 1996), suggests that Oxa1 alone directs the insertion of membrane proteins encoded by the mitochondrial genome. Upon depletion of Oxa1 from yeast mitochondria, the insertion of subunit 2 of the cytochrome oxidase complex (Cox2) was severely affected, although other mitochondrially encoded proteins such as cytochrome b (Cytb) and subunit 6 of ATP synthase (Atp6p) were still able to insert relatively efficiently (Hell et al., 2001). Similar insertion defects were observed when only the C-terminal ribosome binding

domain of Oxa1 was removed, illustrating the importance of ribosome membrane tethering to this process (Jia et al., 2003).

An *in vitro* investigation into the importance of ribosome membrane binding has not been performed until now, although evidence has emerged suggesting that an increasing aqueous volume reduces membrane protein insertion in a minimal system. Soga et al. (2014) showed by using the PURE system encapsulated in GUVs, that as vesicle volume increased the proportion of EmrE inserted into the membrane was reduced. Attempts have been made to reconstruct the membrane protein insertion process through *de novo* synthesis and assembly of SecYEG and YidC in liposomes, albeit with limited success (Matsubayashi et al., 2014; Ohta et al., 2016). However, the possible effects of ribosome localisation on protein insertion were not considered in these studies. Such analysis would be of great value not only for understanding the fundamental determinants of membrane protein integration but also for constructing an artificial cell model which would require the synthesis of several aggregation-prone membrane proteins.

The results presented in Chapter 4 have shown that removal of the PR signal peptide or of GalP TM1 negatively impacts on protein localisation and insertion. Can these defects be rescued by synthetically tethering ribosomes to the inner leaflet of GUV membranes? This chapter focuses on answering this question based on experiments in which His-tagged ribosomes were artificially attached to the inner leaflet of GUV membranes using NTA(Ni) affinity for histidine. Confocal microscopy analyses of individual vesicles was carried out to visualise ribosome complex localisation and to analyse the localisation and insertion of truncated version of PR and GalP.

5.2 Results

5.2.1 Tethering of EGFP to GUV membranes

The co-translational targeting and insertion of membrane proteins PR and GalP was assessed in Chapter 4 and results indicated that the N-terminal hydrophobic helix of both proteins was essential for localisation and insertion. The localisation effect was surprising considering the presence of multiple downstream hydrophobic helices. The co-translational nature of N-terminal helix-driven targeting suggested that ribosomes were actively localised to the membrane following the emergence of the first N-terminal helix from the exit tunnel. If this was the case, artificial membrane attachment of ribosomes should rescue the localisation of proteins in which the N-terminal helix was removed.

To test this hypothesis, a method was developed to attach functional ribosomes to the inner leaflet of GUV membranes by taking advantage of Nitrilotriacetic acid (NTA) affinity for histidine. The synthetic lipid 1,2-dioleoyl-*sn*-glycero-3-[(N-(5-amino-1-carboxypentyl) iminodiacetic acid)succinyl] (nickel salt) (DGS-NTA(Ni)) was used to generate GUVs containing surface exposed NTA(Ni) moieties at concentrations of up to 3 mol% with POPC making up the remainder of lipids.

In order to test the ability of the NTA(Ni)-functionalised GUVs to bind His-tagged protein, a C-terminal His-tagged version of EGFP (EGFP-His) was overexpressed and purified from BL21(DE3) *E. coli* cells using standard affinity chromatography techniques (see section 2.2.2). Following extensive dialysis to remove residual imidazole, EGFP-His was included at 1 μM in the inner phase solution for droplet transfer generation of GUVs. Lipid solutions were prepared containing POPC and various mol% concentrations of DGS-NTA(Ni) (0%, 1%, 2% and 3%). Following collection, GUVs were analysed using confocal microscopy to determine the binding of EGFP-His to DGS-NTA(Ni). Figure 5.1 shows that increasing concentrations of DGS-NTA(Ni) resulted in increased membrane localisation of EGFP-His, indicating that the protein was successfully binding to the NTA(Ni) moiety of the lipid.

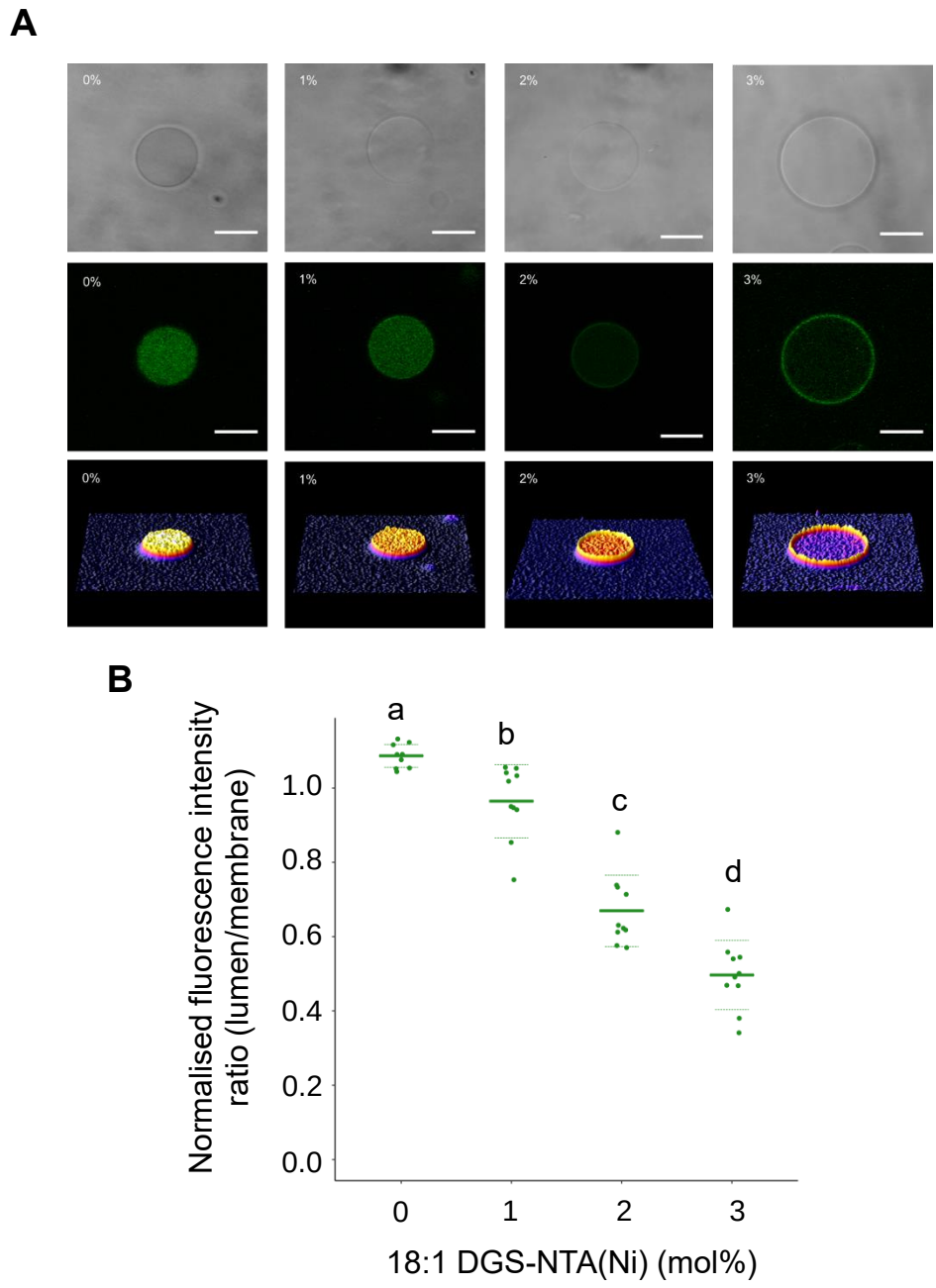


Figure 5.1 Attachment of His-tagged EGFP to the inner leaflet of GUVs.

(A) Representative confocal images of GUVs composed of POPC and 0, 1, 2 and 3 mol% of DGS-NTA(Ni) encapsulating 1 μM of His-tagged EGFP. Scale bars are 10 μm. **(B)** Quantification of EGFP fluorescence using radial profiles from 10 individual GUVs. Different letters denotes statistically significant ($p < 0.05$) differences as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

5.2.2 Purification of functional histidine-tagged 70S ribosomes

Previously, an *E. coli* strain had been engineered to produce 70S ribosomes with hexa-histidine tags on the four L7/12 stalk proteins of the large ribosomal subunit (Ederth et al., 2009)(Figure 5.2), and was kindly provided by Professor Suparna Sanyal (Uppsala, Sweden). These tags not only allowed a simplified purification protocol to be used (see section 2.2.2) but should allow the attachment of ribosomes to GUV membranes containing DGS-NTA(Ni). The tagged ribosomal proteins were not located in close proximity to the exit tunnel of the ribosome (Figure 5.2) and were therefore unlikely to affect the properties of the membrane protein nascent chain as it emerged.

Ribosomes were purified using standard NTA(Ni) columns before being subjected to extensive dialysis to remove residual imidazole from the preparation according to the previously established protocol (Ederth et al., 2009). Following this, 70S ribosomes were further purified by ultracentrifugation through a 30% sucrose cushion to remove any contaminating soluble proteins. These imidazole-free 70S ribosomes were then added to ribosome-free PURE reactions at a concentration of 1 μM. Figure 5.3 shows the resulting SDS-PAGE gel following the bulk synthesis of α-hemolysin in the presence of fluorescent lysine amino acids as a positive control. Synthesis only occurred in the presence of ribosomes and the band intensity of translation products was similar for PURE-supplied ribosomes and purified His-tagged ribosomes. This indicated that the purification protocol produced ribosomes of sufficiently high quality for protein synthesis in PURE reactions.

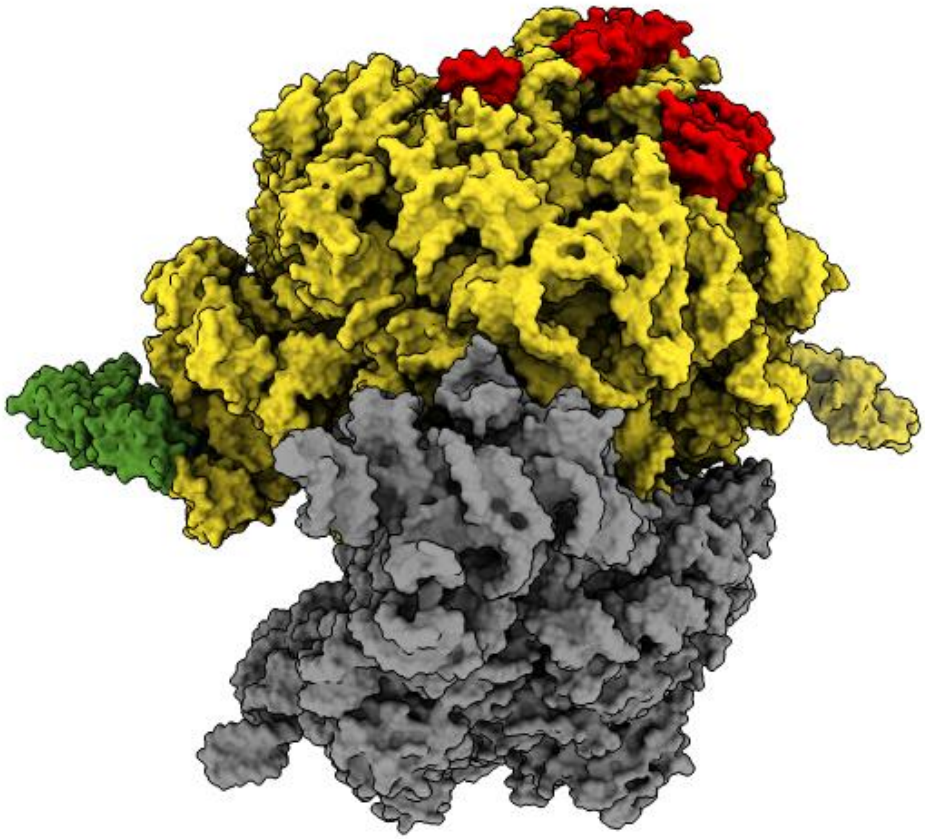


Figure 5.2 Structure of the bacterial 70S ribosome.

Schematic representation of the bacterial 70S ribosome complex (PDB entry 4V4P). Yellow and grey surfaces represent the 50S and 30S subunits respectively; green and red surfaces represent the histidine-tagged L7/12 proteins and exit tunnel proteins respectively.

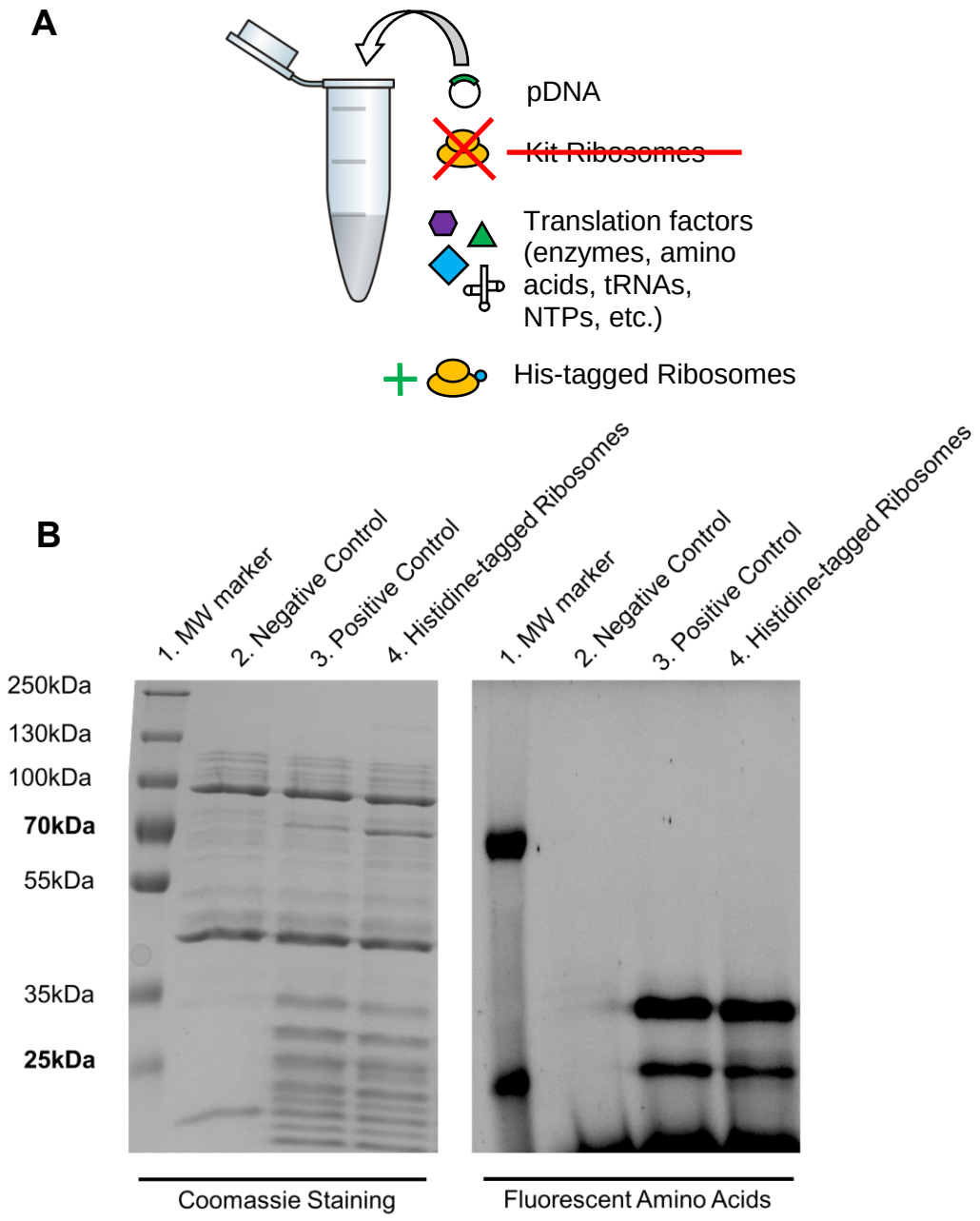


Figure 5.3 Purified, histidine tagged ribosomes are functional in the PURE system.

(A) Schematic showing the assembly of bulk PURE reactions for protein synthesis using His-tagged ribosomes. **(B)** SDS-PAGE analysis of a bulk PURE reactions supplemented with 20ng μl^{-1} pDNA encoding α-hemolysin and fluorotect in-gel detection reagent. Images show the same gel stained with coomassie quick stain (left) and analysed for fluorescence (right). Lane 1 contains PageRuler Plus pre-stained molecular weight marker, lane 2 contains a PURE reaction without ribosomes, lane 3 contains a standard PURE reaction with manufacturer-supplied ribosomes and lane 4 contains a PURE reaction supplemented with 1 μM His-tagged ribosomes. Fluorescent molecular weight bands are labelled in bold.

5.2.3 Tethering 70S ribosome complexes to GUV membranes

To visualise ribosome localisation within GUVs, propidium iodide (PI) was used to selectively label ribosomal RNA (rRNA) for fluorescence analysis. Figure 5.4 shows that, following the inclusion of PI at a concentration of 100 μg ml^{-1} inside GUVs, mean fluorescence intensity from the GUV lumen significantly increased from 34.526 ± 0.184 to 171.562 ± 4.419 in the presence of ribosomes at 1 μM. No DNA or non-ribosomal proteins were added to the inner solution. These data suggested that PI could be used to analyse the localisation of ribosomes inside GUVs.

Vesicles were therefore prepared with and without 2.7 mol% DGS-NTA(Ni) and encapsulating non-tagged and His-tagged ribosomes along with 100 μg ml^{-1} PI, 100 mM potassium glutamate and 18 mM magnesium acetate. 2.7 mol% was chosen due to the stability of DGS-NTA(Ni) containing GUVs in PURE buffer conditions at this concentration. Following radial profile analysis of PI fluorescence emission, it was clear that His-tagged ribosomes were localised to the membrane in the presence of DGS-NTA(Ni) but not in any other conditions. A mean lumen/membrane fluorescence ratio was determined at 0.567 ± 0.182 for DGS-NTA(Ni) containing GUVs encapsulating His-tagged ribosomes. This value was significantly lower than the mean values determined for DGS-NTA(Ni) containing vesicles encapsulating non-tagged ribosomes (1.05 ± 0.029), POPC vesicles encapsulating His-tagged ribosomes (1.074 ± 0.031) and POPC vesicles encapsulating non-tagged ribosomes (1.083 ± 0.046) (Figure 5.5).

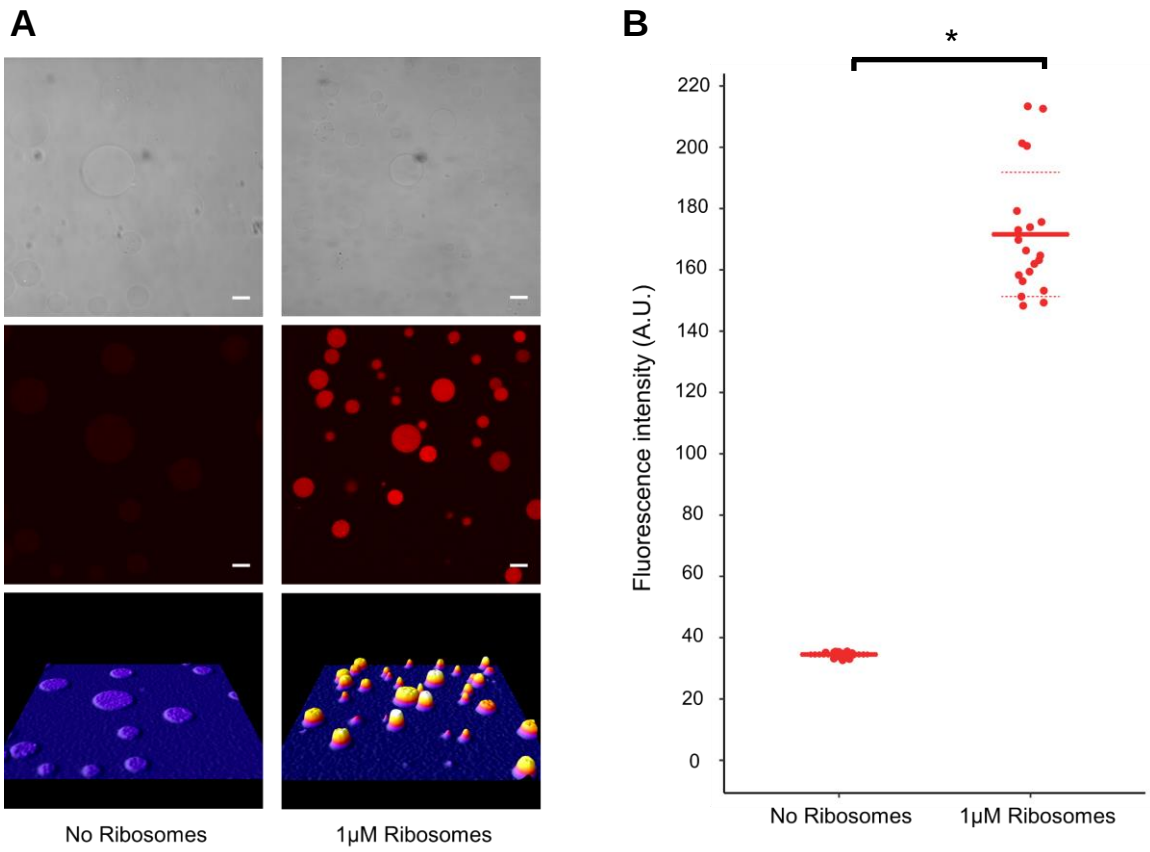


Figure 5.4 Propidium iodide staining of rRNA encapsulated within GUVs.

(A) Representative confocal microscopy images of GUVs encapsulating 100 µg ml^{-1} propidium iodide and either without or with 1 µM of 70S ribosomes. **(B)** Quantification of fluorescence emission intensity from propidium iodide within 20 individual GUVs from each condition normalized to background fluorescence. * indicates significant difference ($p < 0.001$) as measured by Wilcoxon rank sum test.

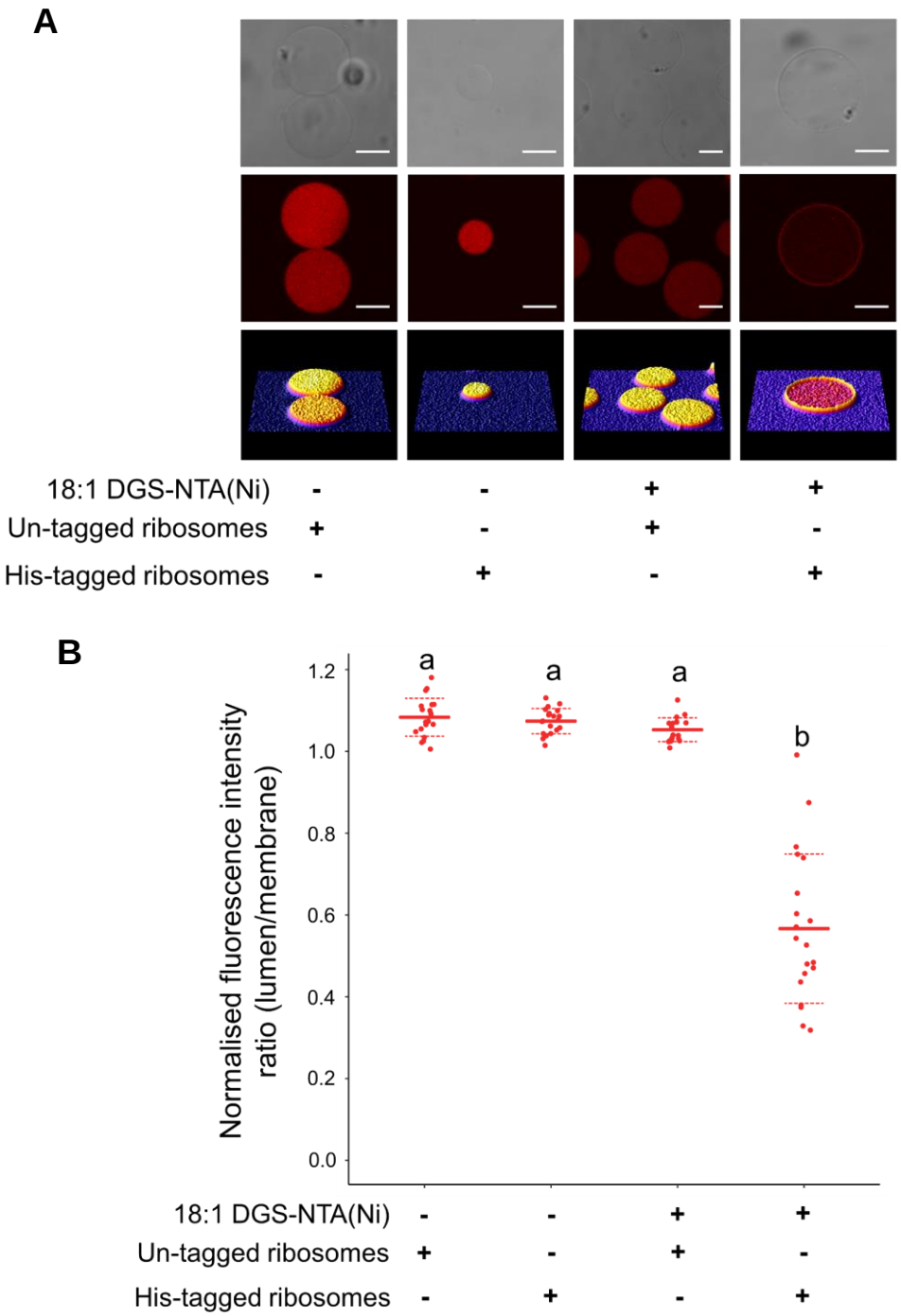


Figure 5.5 Synthetic attachment of histidine-tagged 70S ribosome complexes to GUV membranes.

(A) Representative confocal images of GUVs encapsulating 1 µM 70S ribosomes stained with 100 µg ml^{-1} propidium iodide. GUVs were generated with and without 2.7 mol% DGS-NTA(Ni) in the membrane. The internal solution contained 100 µM potassium glutamate and 18 µM magnesium acetate to mimic the ionic concentrations of the PURE system and to maintain ribosome complex integrity. Scale bars are 10 µm. **(B)** Propidium iodide fluorescence emission was localised using radial profiling for 20 individual GUVs from each condition. Different letters denote statistical significance ($p < 0.001$) as measured by one-way ANOVA using Tukey (HSD) post-hoc analysis.

5.2.4 Investigating the effects of pre-translational ribosome tethering on membrane protein localisation and insertion

Previous results indicated that at 2.7 mol% of NTA(Ni) containing lipid, functional His-tagged ribosome complexes were preferentially tethered to the membrane. Thus, experiments were performed to analyse the effects on protein localisation and insertion.

PURE reactions supplemented with 1 μM His-tagged ribosomes and 20 ng μl^{-1} pDNA encoding PRΔSP-EGFP were encapsulated within GUVs with and without 2.7 mol% DGS-NTA(Ni). Following protein synthesis at 37°C, localisation of the EGFP signal was analysed by confocal microscopy using radial profiling (Figure 5.6). Without DGS-NTA(Ni) present, the mean lumen/membrane ratio from 20 individual GUVs was 0.692 ± 0.258. When DGS-NTA(Ni) was present at 2.7 mol% to induce ribosome membrane binding, the mean ratio was 0.377 ± 0.206, significantly lower than when ribosomes were not tethered.

Insertion of PRΔSP synthesised with membrane-bound ribosomes was then analysed using PRΔSP-EL1HA containing an HA epitope on the first extracellular loop of the protein (see section 3.2.4). The binding of Alexa Fluor-647 conjugated HA antibody was analysed as before. Results presented in Figure 5.7 show that without DGS-NTA(Ni), the normalised mean ratio of fluorescence was 1.188 ± 0.076 indicating no translocation of the HA epitope containing region. This was significantly reduced upon the addition of DGS-NTA(Ni) to 0.840 ± 0.193 indicating that protein insertion was rescued when ribosomes were tethered to the membrane.

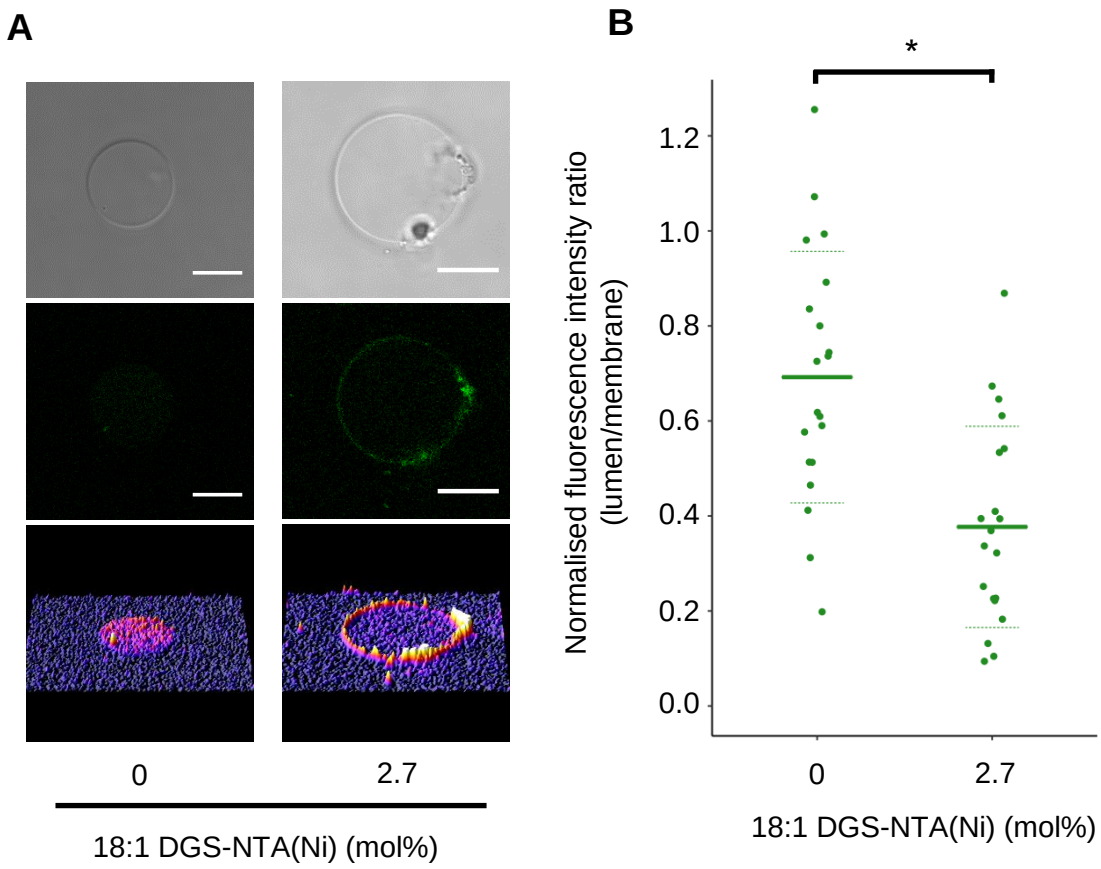


Figure 5.6 Membrane attachment of ribosomes leads to increased membrane localisation of PRΔSP.

(A) Representative confocal images of GUVs encapsulating PURE reactions following the synthesis of PRΔSP-EGFP with his-tagged ribosomes. Left column shows GUVs composed from 100 mol% POPC while the right column shows GUVs generated with 2.7 mol% DGS-NTA(Ni). Scale bars are 10 µm. **(B)** Quantification of membrane localisation of EGFP. 20 individual GUVs were analysed by radial profiling with and without DGS-NTA(Ni). * denotes statistical significance ($p < 0.001$) as determined by two-tailed, unpaired student's t-test.

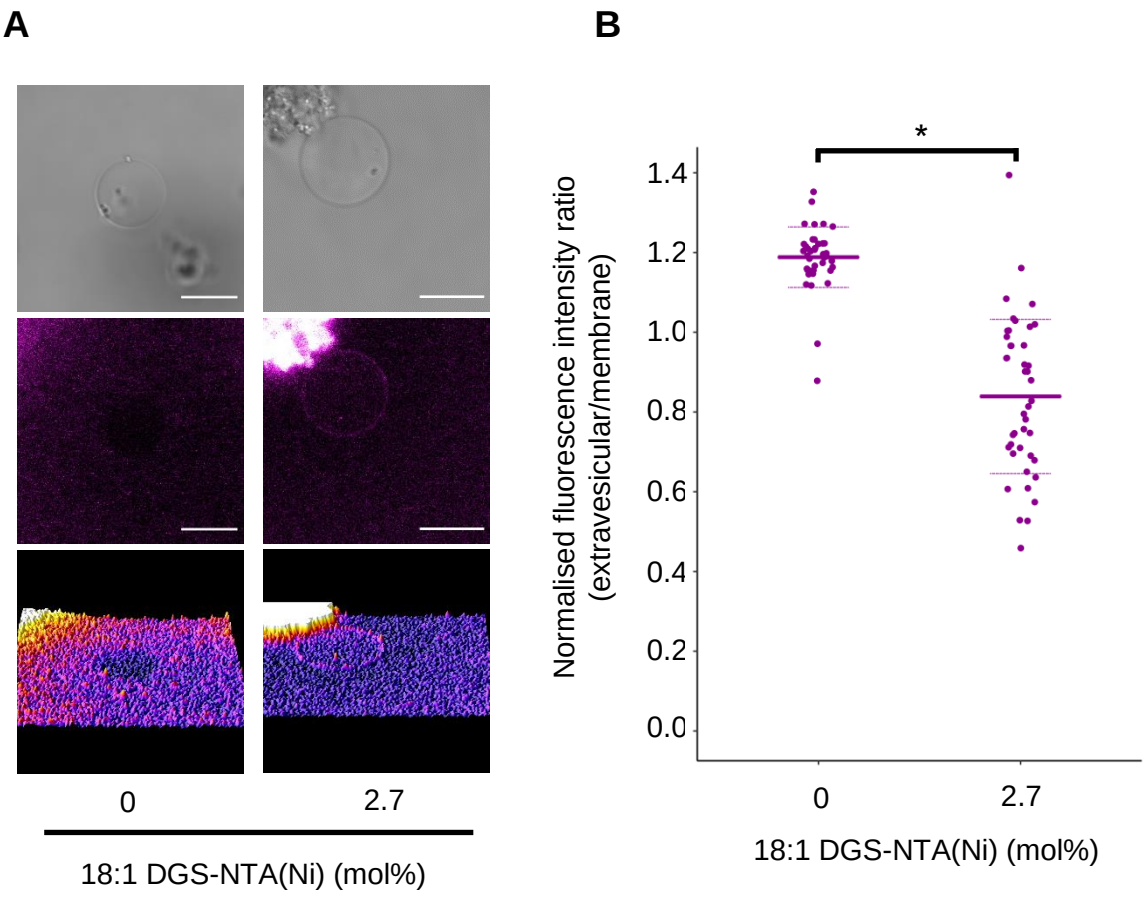


Figure 5.7 Membrane attachment of ribosomes rescues insertion of PRΔSP.

(A) Representative confocal images of GUVs encapsulating PURE reactions following the synthesis of PRΔSP-EL1HA with His-tagged ribosomes and subsequent staining with Alexa Fluor 647-conjugated HA antibody. Left column shows GUVs composed from 100 mol% POPC while the right column shows GUVs generated with 2.7 mol% DGS-NTA(Ni). Scale bars are 10 µm. **(B)** Quantification of membrane bound fluorescence from Alexa Fluor 647-conjugated HA antibody. Radial profiles were calculated from ≥ 43 individual GUVs from each lipid condition. * denotes statistical significance ($p < 0.001$) as determined by two-tailed, unpaired student's t-test.

To confirm that the signal from Alexa Fluor-647 conjugated HA antibody was not due to direct binding of the antibody to DGS-NTA(Ni), a control experiment was performed. A PURE reaction supplemented with His-tagged ribosomes, but no pDNA, was encapsulated within GUVs containing 2.7 mol% DGS-NTA(Ni) and incubated for three hours at 37°C as before. GUVs were then treated with fluorescent HA antibody followed by washing in fresh outer buffer. Vesicles were analysed by confocal microscopy and these DGS-NTA(Ni) containing GUVs exhibited no membrane localised HA signal as evident from the mean ratio of

1.212 ± 0.111 which was no different to negative control experiments containing PR without an HA tag (Figure 5.8). These results indicated that HA antibody was binding directly to the HA epitope present in PRΔSP-EL1HA and not to the NTA(Ni)-containing lipid.

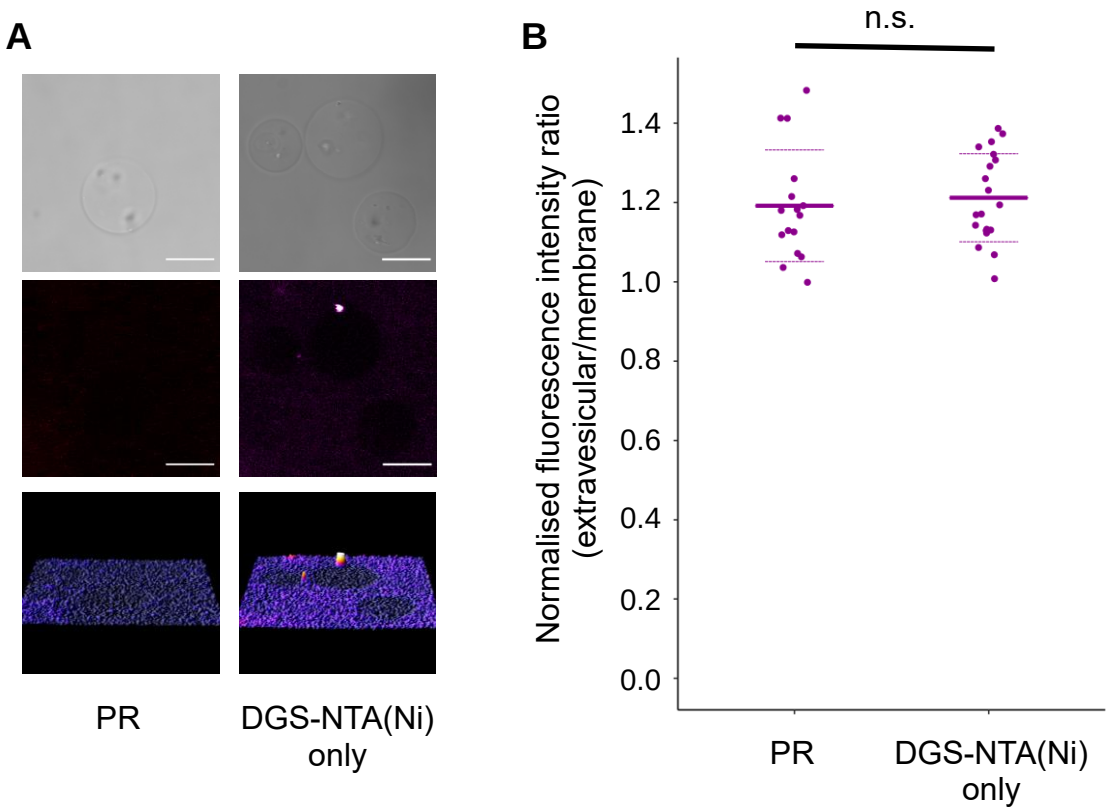


Figure 5.8 Alexa Fluor-647 conjugated HA antibody does not bind directly to DGS-NTA(Ni).

(A) Representative confocal images of GUVs encapsulating PURE reactions and His-tagged ribosomes following protein synthesis and staining with Alexa Fluor 647-conjugated HA antibody. Left column shows GUVs composed from 100 mol% POPC synthesising PR-EGFP while the right column shows GUVs generated with 2.7 mol% DGS-NTA(Ni). Scale bars are 10 µm. **(B)** Quantification of membrane bound fluorescence from Alexa Fluor 647-conjugated HA antibody. n.s. indicates non-significant differences ($p = 0.644$) as measured using a student's unpaired, two sample t-test.

Following the successful rescue of PRΔSP localisation and insertion by membrane tethering of ribosomes, localisation of GalP truncations was analysed. Figure 5.9 shows the results of protein localisation experiments for each of the truncated versions of GalP. Results are paired to show differences for each truncation, with or without DGS-NTA(Ni) in the GUV membrane. No statistically significant

differences were detected indicating that ribosome tethering was not able to rescue the localisation defects caused by removal of GalP TM1.

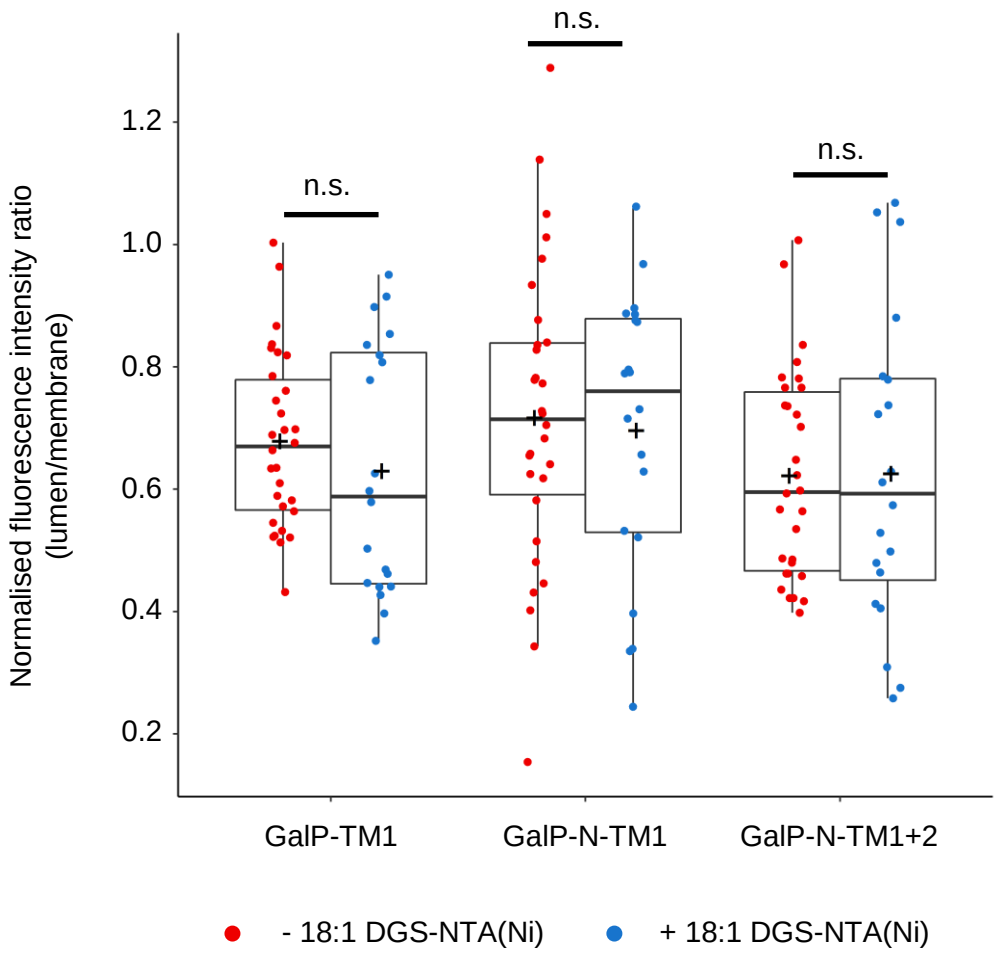


Figure 5.9 Localisation of TM1 truncated versions of GalP with membrane bound ribosomes.

PURE reactions were encapsulated within GUVs with and without DGS-NTA(Ni) for the synthesis of GalP truncations -TM1, -N-TM1 and -N-TM1+2. Radial profiling was used to quantify membrane localisation of mCherry for ≥ 20 individual GUVs. Red data points indicate individual ratios for GUVs composed from 100 mol% POPC, while blue data points indicate ratios for GUVs containing 2.7 mol% DGS-NTA(Ni) in the membrane. n.s. indicates non-significant differences ($p \geq 0.360$) as measured using student's unpaired, two sample t-tests for each GalP truncation.

5.3 Discussion

The recruitment of translating ribosomes to the membrane for the efficient co-translational insertion of polytopic membrane proteins is a process ubiquitous throughout nature (Sommer and Schleiff, 2014; Steinberg et al., 2018). This can be achieved through targeting by proteinaceous chaperones such as the SRP *in vivo* (Akopian et al., 2013b), or, as shown in Chapter 4, through direct nascent chain-lipid interactions *in vitro*. The two proteins tested, PR and GalP, both exhibited membrane recruitment via the physical characteristics of their N-terminal helices. The question therefore arose, could the mislocalisation exhibited by these truncated proteins be rescued through tethering of ribosomes to the membrane?

To answer the above question, this chapter has focused on the development of a novel technique to investigate the effects of ribosome membrane tethering on membrane protein localisation and insertion in a minimal system. In Chapter 4 it was shown that a truncated version of PR (PRΔSP) and several truncated versions of GalP did not effectively localise to the membrane. Here, the same proteins were synthesised in GUVs with membrane bound ribosomes to test whether ribosome localisation was sufficient for rescuing their membrane localisation. Results showed that PRΔSP localisation and insertion was indeed rescued by ribosome tethering, whereas the results for GalP were not conclusive.

The synthetic lipid DGS-NTA(Ni) has been used previously for the attachment of His-tagged fluorescent proteins to the inner leaflet of GUV membranes (Peters et al., 2015). It has also been used to chaperone and aid the co-translational insertion of membrane protein Cx43 into empty liposomes. Ando et al. (2018) showed that N-terminally His-tagged Cx43, synthesised in the presence of LUVs by the PURE system, was able to insert more efficiently in the presence of DGS-NTA(Ni). These results suggested that the emergence of a nascent chain with high affinity for the membrane (in this case histidine affinity for DGS-NTA(Ni)) aided spontaneous membrane insertion. However, inclusion of a soluble chaperone enhanced the insertion further, suggesting that some nascent chains were unable to reach a membrane prior to their misfolding. It was therefore decided to use this NTA-containing lipid to attach ribosomes to the membrane of GUVs in the hopes of aiding co-translational protein insertion without the need

for soluble chaperones or targeting machineries. Following initial confirmation that DGS-NTA(Ni) could be used for the attachment of His-tagged EGFP to the inner leaflet of GUVs (Figure 5.1), 70S ribosomes containing His-tagged L7/L12 proteins of the large ribosomal subunit were purified using simple NTA(Ni) affinity chromatography (Ederth et al., 2009). These His-tagged ribosomes were then successfully attached to the inner leaflet of GUV membranes when DGS-NTA(Ni) was included in the lipid solution (Figure 5.5) thus providing an excellent model system to test the effects of ribosome tethering on membrane protein co-translational localisation and insertion.

PR is an excellent model to test these effects due to the presence and function of the N-terminal signal peptide. This signal peptide has been shown to be non-essential for protein function *in vivo*, indicating that correctly folded protein can still be formed in its absence (Soto-Rodriguez and Baneyx, 2019). In this situation translating ribosomes are likely recruited to the membrane via the SRP pathway or via other soluble chaperones, despite the absence of the signal peptide. However, *in vitro* data presented in Chapter 4 have shown that, in the absence of these targeting chaperones, a truncated version of PR (PRΔSP) fails to be targeted to the membrane of GUVs for insertion. Previous data using the PURE system in GUVs has also shown that as vesicle volume increases the levels of insertion of another polytopic membrane protein, EmrE, are reduced (Soga et al., 2014). These data all imply that ribosome localisation may play a key role in determining membrane protein fate in a minimal system devoid of any targeting chaperones.

The effects of ribosome membrane tethering were therefore tested by synthesising PRΔSP using the PURE system inside GUVs with membrane attached ribosomes. Localisation data showed that PRΔSP-EGFP was more associated with the membrane when ribosomes were tethered, as might be expected considering the localisation of translation (Figure 5.6). Protein insertion was also completely recovered following ribosome tethering. Without DGS-NTA(Ni), no insertion was detectable using fluorescent HA antibody and PRΔSP-EL1HA, but when DGS-NTA(Ni) was included in vesicles at 2.7 mol% to induce ribosomal tethering, insertion of PR was rescued as was evident from the significant drop in fluorescence ratio indicating HA antibody binding to surface exposed HA epitopes of PRΔSP-EL1HA (Figure 5.7).

Previous data had also indicated a role for the N-terminal helix of GalP in the targeting of RNCs to the membrane. Therefore, the effects of ribosome tethering on the localisation of GalP TM1 truncations was also analysed. The localisation of GalP TM1 truncations, unlike PRΔSP was unaffected by the membrane tethering of ribosomes (Figure 5.9). This was surprising given the localisation of translation to the membrane and suggests that TM1 of GalP is essential for initially anchoring the protein to the membrane and for avoiding irreversible aggregation. The lower hydrophobicity of TMs 2-4 of GalP (Figure 4.12) may require anchoring of the nascent chain to the membrane via TM1 to avoid unwanted aggregation and this cannot simply be compensated for by positioning ribosomes close to the membrane. The thermodynamic drive for membrane partitioning of these less hydrophobic helices as well as downstream folding events may also require the adoption of specific structure within TM1 as has been shown to be the case for the α-helical membrane embedded protease GlpG (Paslawski et al., 2015).

In conclusion, the work carried out in this chapter has exhibited the efficacy of using the synthetic lipid DGS-NTA(Ni) to tether functional 70S ribosome complexes to GUV membranes for enhancing the localisation and insertion of mislocalised membrane proteins in a minimal system devoid of chaperoning components. The enhancing effects of ribosome tethering on membrane protein insertion were clearly exhibited for a truncated version of PR (PRΔSP) that was unable to insert when ribosomes were soluble and not membrane bound. These data address for the first time the effects of ribosome localisation on membrane protein insertion in a minimal cell-free system and highlight the importance of co-translational membrane protein targeting even in a minimal *in vitro* system. The presented approach also paves the way for future studies investigating the synthesis of aggregation prone membrane proteins inside GUVs as a chassis for artificial cell generation.

6. General discussion

6.1 Minimal requirements for recruitment and insertion of membrane proteins into lipid membranes

Investigating the minimal requirements for membrane protein targeting and insertion in cell-mimicking model environments has led to remarkable discoveries that have begun to address some of the questions arising when we think about the evolution of cellular membranes. For example, the requirement for a Sec-type translocon and/or insertases for membrane protein insertion and assembly leads to an obvious 'chicken and egg' paradox. These proteins are themselves membrane-embedded proteins and therefore, early in evolutionary history, simple cellular lifeforms must have been able to integrate proteins into their membranes without the help of this machinery. Nevertheless, given the ubiquity of translocons and insertases in modern cells they must have provided a fitness advantage and subsequently become indispensable for the insertion and assembly of the complex membrane environment we see in cells today. To understand the exact functions of proteins and pathways it is worthwhile retracing their evolution and interrogating which processes did or did not require their functions. Artificial cells offer excellent opportunities to do just this; and they are now being used to better understand the minimal requirements for insertion of membrane proteins into lipid bilayers. Such research has already provided evidence that translocons and insertases are not an absolute requirement for the insertion and folding of many polytopic membrane proteins, at least not in simplified cellular environments (Kuruma and Ueda, 2015; Harris et al., 2017). This thesis has added another important piece of the puzzle by showing that the co-translational recruitment of two unrelated membrane proteins can also be achieved in an artificial cell-size mimicking system that is devoid of translocon, insertase or chaperoning proteins.

The co-translational recruitment of membrane protein nascent chains to the translocon via the signal recognition particle (SRP) or other soluble chaperones is an essential aspect of membrane protein biogenesis *in vivo*. However, it is likely that early cells were able to insert membrane proteins without the aid of this specific pathway (Pohlschroder et al., 2005). It is not known why, or at which point in time the SRP pathway has become essential for cellular membrane

biogenesis, but it's emergence likely reduced the risk of hydrophobic protein aggregation in increasingly large and complex early cells by allowing them to efficiently target translating ribosomes to the membrane. Also, specific targeting to particular membrane regions or compartments could have been needed for enabling the complex sub-cellular organisation of modern cells (see below). Based on the evidence that many membrane proteins can indeed insert into the membranes of artificial cells without translocons or insertases, a logical follow-on question was whether they also maintain inherent mechanisms to aid co-translational membrane targeting.

The experimental work presented in this thesis consisted of the development of an experimental system allowing the interrogation of individual mechanisms underpinning the targeting and insertion of membrane proteins. A bottom-up approach was employed using the PURE cell-free protein synthesis system and giant unilamellar vesicles (GUVs) to examine these fundamental processes in an environment devoid of cellular complexity and specific targeting and chaperoning pathways, while mimicking cell size. The results presented here have uncovered the essential nature of the first hydrophobic region of both proteorhodopsin (PR) and galactose permease (GalP) to protein-inherent membrane targeting, and have shown the importance of ribosome membrane localisation for the insertion process in the absence of such targeting. Figure 6.1 summarises the evidence obtained from experimental work.

Results presented in Chapter 3 showed that when full-length PR and galP were synthesised inside cell-size mimicking GUVs, they were able to localise and insert spontaneously into the membrane despite the lack of any targeting or chaperoning proteins in the recombinant PURE system (Figure 6.1, Scheme 1). Chapter 4 then expanded on these results and showed that protein localisation and insertion was a co-translational process involving ribosome membrane recruitment driven by the N-terminal hydrophobic domains of both PR and GalP. These domains were essential for the observed localisation and insertion and their removal meant that proteins were no longer recruited to the membrane and inserted, likely due to increased aggregation within the aqueous vesicle interior (Figure 6.1, Scheme 2). The results presented in Chapter 5 showed that when ribosomes were tethered to the membrane, localisation and insertion of the N-terminally truncated version of PR was rescued (Figure 6.1, Scheme 3).

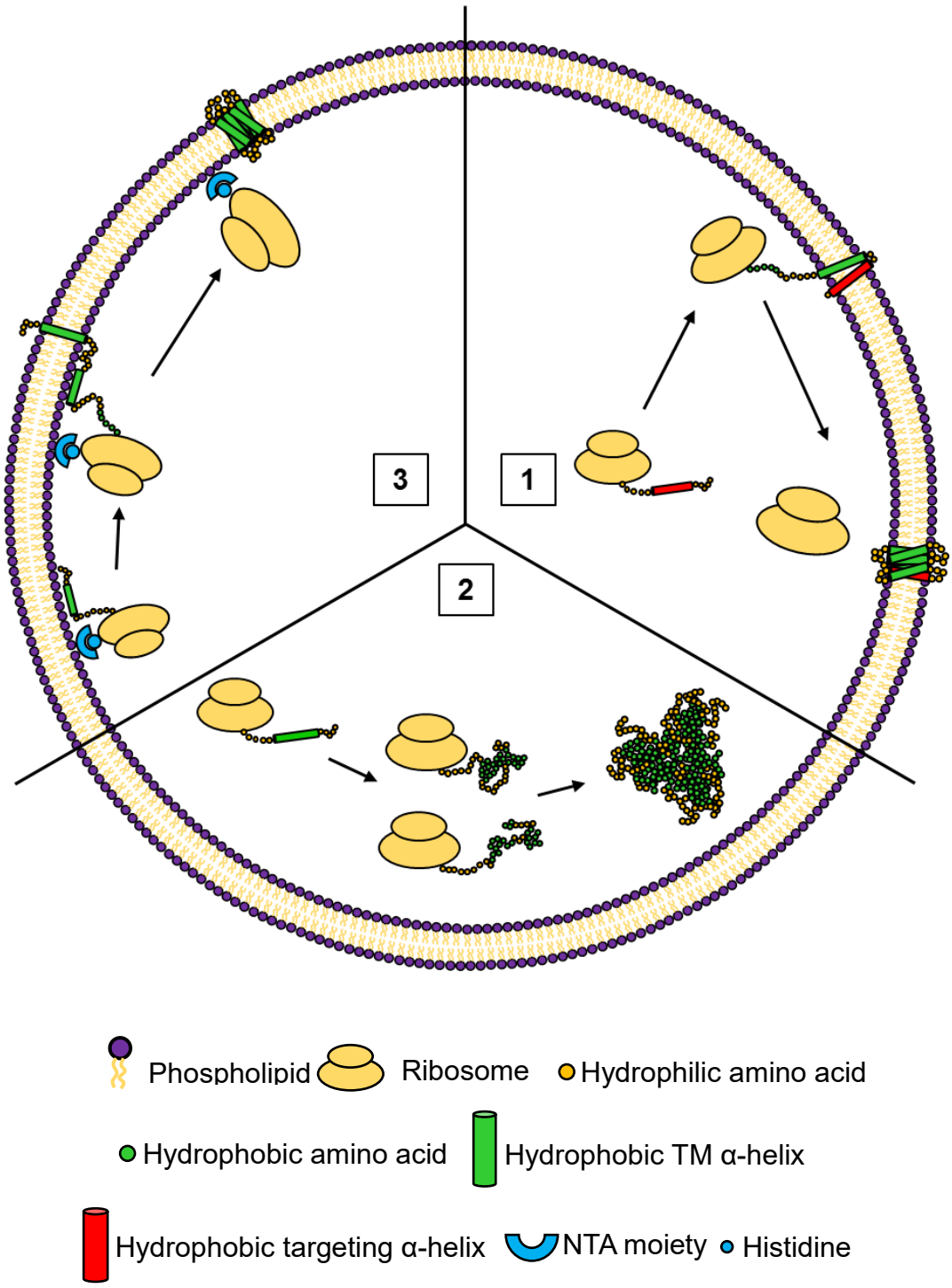


Figure 6.1 Schematic model of the importance of co-translational membrane protein targeting in a simplified cell-mimicking system.

(Scheme 1) The first hydrophobic region to emerge from the ribosome actively localises translation to the membrane and thus facilitates the direct insertion of proteins in a cell-size mimicking system devoid of targeting and insertion mediators. **(Scheme 2)** In the absence of these targeting domains, proteins have a higher propensity for self-aggregation in the aqueous environment and are thus unable to directly insert into cell-sized liposomes. **(Scheme 3)** When ribosomes are tethered to the membrane using NTA(Ni)-histidine affinity, protein insertion is able to be rescued.

The latter results stressed the importance of ribosome localisation for the spontaneous insertion process. Interestingly, the localisation of N-terminally truncated versions of GalP was not rescued by the tethering of ribosomes to the membrane. This suggests that this N-terminal helix is essential for both ribosome recruitment to the membrane and for the insertion process itself. It is possible that the formation of structure within this helix is required for subsequent structure formation steps, and that without this folding intermediate, the protein is highly susceptible to aggregation and misfolding despite being synthesised in close proximity to the membrane. Data for the membrane protease GlpG has suggested a similar requirement for structure formation of its N-terminal helix to aid downstream folding events (Paslawski et al., 2015).

Considering all of the presented data, it seems plausible that N-terminal hydrophobic regions have evolved to have high affinity for the membrane, or at least for its interfacial region, and relatively low self-affinity in order to limit protein aggregation in the aqueous environment. This simplistic targeting mechanism may pre-date the evolution of more specific mechanisms such as the SRP, although firm evidence for this is still lacking.

6.2 Why did evolving cells require more complex targeting machineries?

Evidence clearly points to the fact that many ‘modern’ polytopic membrane proteins are able to spontaneously insert into empty lipid bilayer membranes despite the absence of targeting and chaperoning systems. Why then did cells evolve more specific pathways such as the SRP system?

There are several complexity related factors, beyond simply increasing cell size, that likely contributed to the requirement for more regulated targeting in early cells. The most obvious example would be for specific targeting to different membranes within the same cell, for example one might think of the thylakoid membrane of an ancient cyanobacteria or the mitochondria of a more ‘modern’ eukaryote. Indeed, a recent report has shown that depletion of SRP in yeast caused the mistargeting of many SRP substrate proteins to the mitochondria instead of the ER (Costa et al., 2018). This fascinating report showed that without the specificity provided by SRP targeting, these proteins reverted to a more rudimentary form of targeting, probably based on membrane environmental factors. Clearly, in a relatively complex, multi-membrane organism, targeting controls became increasingly important to increase the efficiency of membrane biogenesis.

However, even before the evolution of multi-membrane organisms, the biogenesis of membrane proteins in an increasingly complex membrane environment may well have required the evolution of specific targeting and chaperoning systems. For example, it is easy to imagine that in an increasingly crowded and diverse intracellular environment there was an increased propensity for unwanted and detrimental aggregation of highly hydrophobic proteins that necessitated the evolution of targeting mechanisms such as SRP-based targeting or possibly mRNA targeting. It has been proposed that the spontaneous insertion of membrane proteins was the norm in early cells (Pohlschroder et al., 2005). Phylogenetic evidence suggests that YidC was the first insertase protein to evolve in order to assist membrane protein insertion and folding (Bohnsack and Schleiff, 2010). The subsequent evolution of the Sec translocon then allowed more complex organisms to evolve containing proteins with large soluble domains requiring efficient export from the cytoplasm

(Bohnsack and Schleiff, 2010). SRP, due to its RNA-based structure, is thought to have evolved early to help RNA-based ribosomes to deal with hydrophobic polypeptides (Walter et al., 2000). However, direct evidence for this is lacking and even a 'more evolved' SRP such as that found in eukaryotes still maintains an RNA-based structure (Akopian et al., 2013b).

What seems clear is that increasing cell size, in conjunction with increasing diversity and complexity of membrane embedded components and cytoplasmic environments, necessitated the evolution of precise targeting mechanisms. Subsequently, the evolution of a YidC insertase and a Sec translocon vastly increased the potential functions of membrane proteins and led to the wide variety of structures we see today. When we think about reconstructing a cell from the bottom-up it will be important to determine how increases in the number and complexity of membrane proteins being synthesised affects their biogenesis in simplified membranes, and how the addition of targeting chaperones affects this in systems mimicking realistic cell size and volume such as GUVs.

6.3 Advantages and limitations of encapsulated membrane protein synthesis in GUVs

The size and encapsulation characteristics of GUVs make them excellent tools to investigate fundamental processes such as membrane protein targeting and insertion using a bottom-up approach, as has been shown throughout this thesis. Their size makes them amenable to optical microscopy-based applications and fluorescent assisted cell sorting (FACS), which allow single vesicle analysis to be performed, and to uncover details that are masked in population-level studies (Apellaniz et al., 2010; Soga et al., 2014). GUVs have also been used for electrophysiological studies from single-channel patch clamp recordings to whole-GUV voltage clamp experiments (Berrier et al., 2011; Garten et al., 2017). Table 6.1 provides a comprehensive list of studies that have utilised GUVs and cell-free reactions for membrane protein synthesis along with the techniques used for their analysis.

Despite these advantages there are also a number of well-established techniques to which this system is not amenable. For example, the analysis of membrane protein insertion and topology using SDS-PAGE-based methods was not possible due to the low yield of GUVs, and therefore protein. LUVs or SUVs provide a superior liposome model for such experiments and have been used here in conjunction with GUV-based experiments. Alterations to membrane composition are also difficult to achieve using the presented GUV formation method. For example, high quantities of non-bilayer forming lipids such as phosphoethanolamine (PE)-based lipids is not possible, and high concentrations (> 10 mol%) of anionic phospholipids led to vesicle instability when combined with cell-free protein synthesis solutions. The droplet transfer approach also produces a heterogeneous size mixture of GUVs that can lead to high levels of experimental variation, thus masking small changes in factors such as membrane localisation. Thus, in future, superior artificial cells need to be developed to combine the advantages of GUVs with respect to size and tractability with novel opportunities for consistent, large-scale production and enhanced stability. One emerging technology in this field is droplet microfluidics for the generation of homogeneous GUVs (Martino and deMello, 2016).

Table 6.1 Cell-free membrane protein synthesis using GUVs

Protein	Techniques	Microscopy analysis	Refs
MscL	patch clamp	-	(Berrier et al., 2004; Yoshimura et al., 2008; Berrier et al., 2011)
CYb5	confocal microscopy	localisation	(Nomura et al., 2008)
hVDAC1	patch clamp	-	(Deniaud et al., 2010)
Cx43	confocal microscopy	localisation and leakage	(Liu et al., 2013)
EmrE	confocal microscopy and FACS	leakage and insertion	(Soga et al., 2014; Uyeda et al., 2016)
BmOR1 and BmOrco	patch clamp and confocal microscopy	localisation	(Hamada et al., 2014)
GPAT and LPAAT	mass spectrometry	-	(Scott et al., 2016)
9 IMPs and 2 OMPs *	FACS	-	(Ohta et al., 2016)
Letm1	FACS	-	(Okamura et al., 2019)
BR	confocal microscopy	fluorescent protein production	(Berhanu et al., 2019)
PR and GalP	confocal microscopy	localisation and insertion	*This study*

***** IMPs - EmrE, SugE, MdtI, MdtJ, sdhD, EptB, YdcX, YeaQ, YdcZ, YfdY
OMPs - PhoE, TolC

6.4 GUV solute leakage upon membrane protein insertion

While the results presented in this thesis have focussed on the localisation and insertion of membrane proteins into GUVs and not on their functionality post-insertion, a number of experiments were carried out assess the functionality of both PR and GalP. The difficulties encountered with the use of the PURE system and GUVs for functional analysis of membrane proteins will therefore be discussed briefly.

Several assays were used to investigate the function of PR by assessing pH changes and membrane potential changes upon light illumination. The major difficulty in measuring bulk pH changes in PURE encapsulating GUVs was the high concentration of buffer, and thus high proton buffering capacity, of the PURE solution required for protein synthesis. It may be possible to detect local changes in the vicinity of the membrane using labelled lipids such as 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-(carboxyfluorescein), which contains a pH sensitive dye attached to the headgroup, although this was attempted unsuccessfully with no detectable levels of proton transport being observed. Another assay used to investigate PR function employed fluorescent membrane potential indicators. This was also unsuccessful with large intra and inter-experimental variation likely masking any positive responses.

Another major problem with the measurement of ionic transport events using the PURE/GUV system was noted following attempts to monitor the monosaccharide transport function of GalP. Initial experiments using the fluorescent glucose analogue 2-(N-(7-Nitrobenz-2-oxa-1,3-diazol-4-yl)Amino)-2-Deoxyglucose (2-NBDG) showed that even a fairly large hydrophilic molecule such as this was able to readily traverse the GUV membrane when either GalP or PR were synthesised inside using the PURE system. It should also be noted that this leakage effect was observed when an *E. coli* extract-based cell-free system was used. This suggested that insertion, and possibly misfolding of protein within the bilayer, was enough to disrupt the bilayer structure significantly and allow water-soluble, hydrophilic molecules to pass through. This leakage issue was probably also to blame for the inability to detect proton transport in PR containing GUVs.

For GalP this problem could be overcome by examining the leakage characteristics of two water soluble dyes and determining changes in their leakage kinetics when GalP was present. Preliminary results using 2-NBDG and the spectrally distinct soluble dye rhodamine B, indicated that GalP was able to facilitate the downhill transport of 2-NBDG although replication would be required to confirm these results. Uphill transport of 2-NBDG was not observed, although this may again have been due to fast dissipation of a proton gradient through membrane leakage or the proteins inverted topology discussed in chapter 4.

6.5 Future directions

One of the major questions to arise from the present study is whether the protein-inherent targeting mechanism identified here also plays a role *in vivo*. PR is an excellent model to address this question since it is known that the signal peptide aids the biogenesis process *in vivo,* but is not essential, and that cleavage of the signal is not required for protein function (Soto-Rodriguez and Baneyx, 2019). It should therefore be possible to use the truncations generated here to interrogate their effects on the biogenesis of PR in *E. coli*. Conditional knock-out strains could be used to examine this biogenesis process when factors such as Ffh or SecE are eliminated. If the signal peptide were able to enhance inserted protein yields even in an Ffh depleted strain, this would be excellent evidence that the inherent affinity of the signal peptide for the bilayer also plays a significant role in membrane protein biogenesis *in vivo*.

In terms of the development of a bottom-up artificial cell, the next step would be to synthesise both PR and GalP in the same GUV and analyse glucose uptake rates in response to illumination. Proton pumping by PR has been observed following reconstitution into liposomes (Tunuguntla et al., 2013) and native structure is known to be adopted following *de novo* synthesis using cell-free kits (Gourdon et al., 2008; Reckel et al., 2011). The size of GUVs together with the high buffer capacity required by the PURE system has meant that the direct observation of pH changes has so far remained elusive. Since GalP is known to transport glucose both passively and actively in the presence of a proton gradient it should be possible to assess the functionality of PR indirectly through sugar uptake rates. The uptake of glucose by GUVs, either passively or actively, could also be utilised for the generation of ATP via the glycolysis pathway. Enzymes required for this process could be purified and subsequently encapsulated, or synthesised directly by the PURE system alongside GalP and PR. This would be an important step given that to date ATP generation in artificial cells has only been shown using pre-constructed ATP synthase. The *de novo* synthesis and assembly of this multi-subunit protein complex has so far not been possible using minimal cell-free protein synthesis. Therefore, the use of the glycolysis pathway is a promising alternative for ATP production.

While truncation analysis has identified four essential amino acids that guide membrane recruitment of translating ribosomes and aid the insertion of PR, the exact function of these residues remains unclear. At this stage it is not clear whether any combination of four residues with similar overall hydrophobity can achieve ribosome recruitment or whether the identified threonine residue at position 15 is in fact critical for this function. Further mutational analysis therefore needs to be performed in order to investigate the necessity for each of these four residues at the given position.

Performing molecular dynamics simulations on the signal peptide of PR may also provide important information regarding the observed membrane targeting effect. For example, previous studies have shown that structure formation occurs initially at the interfacial region of the bilayer (Ulmschneider et al., 2011). Does the signal peptide of PR penetrate the membrane or remain in the interfacial region when coming into contact with the membrane? And how does this compare with the interaction of the first transmembrane helix with the bilayer?

In summary, the technology developed during this PhD project has enabled the identification of minimal and fundamental mechanisms responsible for the membrane targeting and insertion of two important polytopic α-helical membrane proteins in cell-sized lipid vesicles. The knowledge obtained and the tools generated provide a starting point for further investigations into the more complex mechanisms operating *in vivo*, and for more detailed analysis of the biophysical processes underpinning membrane recruitment processes. This thesis also presents an important step towards the generation of artificial cells capable of effectively synthesising membrane proteins for self-sufficiency of energy production. Artificial cells have many potential applications in areas such as medicine and biotechnology, but also offer an exciting opportunity to enhance our understanding of the origins of life.

References

Akopian, D., Dalal, K., Shen, K., Duong, F. & Shan, S.-o. 2013a. SecYEG activates GTPases to drive the completion of cotranslational protein targeting. *Journal of Cell Biology,* 200(4), pp 397-405.

Akopian, D., Shen, K., Zhang, X. & Shan, S.-o. 2013b. Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annual Review of Biochemistry,* 82, pp 693-721.

Alder, N. N. & Johnson, A. E. 2004. Cotranslational membrane protein biogenesis at the endoplasmic reticulum. *Journal of Biological Chemistry,* 279(22), pp 22787-22790.

Almeida, P. F., Ladokhin, A. S. & White, S. H. 2012. Hydrogen-bond energetics drive helix formation in membrane interfaces. *Biochimica Et Biophysica Acta-Biomembranes,* 1818(2), pp 178-182.

Altamura, E., Milano, F., Tangorra, R. R., Trotta, M., Omar, O. H., Stano, P. & Mavelli, F. 2017. Highly oriented photosynthetic reaction centers generate a proton gradient in synthetic protocells. *Proceedings of the National Academy of Sciences of the United States of America,* 114(15), pp 3837-3842.

Altrichter, S., Haase, M., Loh, B., Kuhn, A. & Leptihn, S. 2017. Mechanism of the Spontaneous and Directional Membrane Insertion of a 2-Transmembrane Ion Channel. *Acs Chemical Biology,* 12(2), pp 380-388.

Amstutz, P., Pelletier, J. N., Guggisberg, A., Jermutus, L., Cesaro-Tadic, S., Zahnd, C. & Pluckthun, A. 2002. In vitro selection for catalytic activity with ribosome display. *Journal of the American Chemical Society,* 124(32), pp 9396-9403.

Angelini, S., Deitermann, S. & Koch, H. G. 2005. FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. *Embo Reports,* 6(5), pp 476-481.

Anghel, S. A., McGilvray, P. T., Hegde, R. S. & Keenan, R. J. 2017. Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. *Cell Reports,* 21(13), pp 3708-3716.

Apellaniz, B., Nieva, J. L., Schwille, P. & Garcia-Saez, A. J. 2010. All-or-None versus Graded Single-Vesicle Analysis Reveals Lipid Composition Effects on Membrane Permeabilization. *Biophysical Journal,* 99(11), pp 3619-3628.

Awayda, M. S., Ismailov, I. I., Berdiev, B. K. & Benos, D. J. 1995. A cloned renal epithelial Na+ channel protein displays stretch activation in planar lipid bilayers. *American Journal of Physiology,* 268(6 pt 1), pp C1450-C1459.

Baars, L., Wagner, S., Wickstrom, D., Klepsch, M., Ytterberg, A. J., van Wijk, K. J. & de Gier, J.-W. 2008. Effects of SecE depletion on the inner and outer membrane proteomes of Escherichia coli. *Journal of Bacteriology,* 190(10), pp 3505-3525.

Batey, R. T., Rambo, R. P., Lucast, L., Rha, B. & Doudna, J. A. 2000. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science,* 287(5456), pp 1232-1239.

Bauer, M., Behrens, M., Esser, K., Michaelis, G. & Pratje, E. 1994. PET1402, a nuclear gene required for proteolytic processing of cytochrome oxidase subunit 2 in yeast. *Molecular & General Genetics,* 245(3), pp 272-278.

Becker, M., Gzyl, K. E., Altamirano, A. M., Vuong, A., Urbahn, K. & Wieden, H.-J. 2012. The 70S ribosome modulates the ATPase activity of Escherichia coli YchF. *Rna Biology,* 9(10), pp 1288-1301.

Beja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., Jovanovich, S., Gates, C. M., Feldman, R. A., Spudich, J. L., Spudich, E. N. & DeLong, E. F. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science,* 289(5486), pp 1902-1906.

BenTal, N., Sitkoff, D., Topol, I. A., Yang, A. S., Burt, S. K. & Honig, B. 1997. Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution. *Journal of Physical Chemistry B,* 101(3), pp 450-457.

Berhanu, S., Ueda, T. & Kuruma, Y. 2019. Artificial photosynthetic cell producing energy for protein synthesis. *Nature Communications,* 10(1325), pp 1-10.

Bernard, S. & Papineau, D. 2014. Graphitic Carbons and Biosignatures. *Elements,* 10(6), pp 435-440.

Berrier, C., Guilvout, I., Bayan, N., Park, K.-H., Mesneau, A., Chami, M., Pugsley, A. P. & Ghazi, A. 2011. Coupled cell-free synthesis and lipid vesicle insertion of a functional oligomeric channel MscL MscL does not need the insertase YidC for insertion in vitro. *Biochimica Et Biophysica Acta-Biomembranes,* 1808(1), pp 41-46.

Berrier, C., Park, K. H., Abes, S., Bibonne, A., Betton, J. M. & Ghazi, A. 2004. Cell-free synthesis of a functional ion channel in the absence of a membrane and in the presence of detergent. *Biochemistry,* 43(39), pp 12585-12591.

Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D. N. & Beckmann, R. 2011. SecM-Stalled Ribosomes Adopt an Altered Geometry at the Peptidyl Transferase Center. *Plos Biology,* 9(1), pp e1000581.

Blobel, G. & Dobberstein, B. 1975. Transfer of proteins across membranes. 1. Presence of proteolytically processed and unprocessed nascent immunoglobulin light-chains on membrane-bound ribosomes of murine myeloma. *Journal of Cell Biology,* 67(3), pp 835-851.

Bochkareva, E., Seluanov, A., Bibi, E. & Girshovich, A. 1996. Chaperonin-promoted post-translational membrane insertion of a multispanning membrane protein lactose permease. *Journal of Biological Chemistry,* 271(36), pp 22256-22261.

Bogdanov, M., Dowhan, W. & Vitrac, H. 2014. Lipids and topological rules governing membrane protein assembly. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1843(8), pp 1475-1488.

Bogdanov, M., Xie, J., Heacock, P. & Dowhan, W. 2008. To flip or not to flip: lipid-protein charge interactions are a determinant of final membrane protein topology. *Journal of Cell Biology,* 182(5), pp 925-935.

Bohnsack, M. T. & Schleiff, E. 2010. The evolution of protein targeting and translocation systems. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1803(10), pp 1115-1130.

Bondar, A.-N. & White, S. H. 2012. Hydrogen bond dynamics in membrane protein function. *Biochimica Et Biophysica Acta-Biomembranes,* 1818(4), pp 942-950.

Bonnefoy, N., Chalvet, F., Hamel, P., Slonimski, P. P. & Dujardin, G. 1994. OXA1, a Saccharomyces cerevisiae nuclear gene whose sequence is conserved from prokaryotes to eukaryotes controls cytochrome oxidase biogenesis. *Journal of Molecular Biology,* 239(2), pp 201-212.

Bonnefoy, N., Fiumera, H. L., Dujardin, G. & Fox, T. D. 2009. Roles of Oxa1-related inner-membrane translocases in assembly of respiratory chain complexes. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1793(1), pp 60-70.

Bormann, B. J., Knowles, W. J. & Marchesi, V. T. 1989. Synthetic peptides mimic the assembly of transmembrane glycoproteins. *Journal of Biological Chemistry,* 264(7), pp 4033-4037.

Borowska, M. T., Dominik, P. K., Anghel, S. A., Kossiakoff, A. A. & Keenan, R. J. 2015. A YidC-like Protein in the Archaeal Plasma Membrane. *Structure,* 23(9), pp 1715-1724.

Botelho, A. V., Huber, T., Sakmar, T. P. & Brown, M. F. 2006. Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes. *Biophysical Journal,* 91(12), pp 4464-4477.

Braun, S., Pokorna, S., Sachl, R., Hof, M., Heerklotz, H. & Hoernke, M. 2018. Biomembrane Permeabilization: Statistics of Individual Leakage Events Harmonize the Interpretation of Vesicle Leakage. *Acs Nano,* 12(1), pp 813-819.

Breyton, C., Haase, W., Rapoport, T. A., Kuhlbrandt, W. & Collinson, I. 2002. Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature,* 418(6898), pp 662-665.

Briggs, M. S., Gierasch, L. M., Zlotnick, A., Lear, J. D. & Degrado, W. F. 1985. In Vivo Function And Membrane-Binding Properties Are Correlated For Escherichia-Coli Lamb Signal Peptides. *Science,* 228(4703), pp 1096-1099.

Broemstrup, T. & Reuter, N. 2010. How does Proteinase 3 interact with lipid bilayers? *Physical Chemistry Chemical Physics,* 12(27), pp 7487-7496.

Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. 2012. Cell-free protein synthesis: Applications come of age. *Biotechnology Advances,* 30(5), pp 1185-1194.

Chartron, J. W., Hunt, K. C. L. & Frydman, J. 2016. Cotranslational signal-independent SRP preloading during membrane targeting. *Nature,* 536(7615), pp 224-228.

Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S. & Hegde, R. S. 2018. EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell,* 175(6), pp 1507-1519.

Choi, H. J. & Montemagno, C. D. 2005. Artificial organelle: ATP synthesis from cellular mimetic polymersomes. *Nano Letters,* 5(12), pp 2538-2542.

Cohen, C. & Parry, D. A. D. 1990. Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. *Proteins-Structure Function and Genetics,* 7(1), pp 1-15.

Copley, S. D., Smith, E. & Morowitz, H. J. 2007. The origin of the RNA world: Co-evolution of genes and metabolism. *Bioorganic Chemistry,* 35(6), pp 430-443.

Costa, E. A., Subramanian, K., Nunnari, J. & Weissman, J. S. 2018. Defining the physiological role of SRP in protein-targeting efficiency and specificity. *Science,* 359(6376), pp 689-692.

Cymer, F. & von Heijne, G. 2013. Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proceedings of the National Academy of Sciences of the United States of America,* 110(36), pp 14640-14645.

Cymer, F., von Heijne, G. & White, S. H. 2015. Mechanisms of Integral Membrane Protein Insertion and Folding. *Journal of Molecular Biology,* 427(5), pp 999-1022.

De Marothy, M. T. & Elofsson, A. 2015. Marginally hydrophobic transmembrane alpha-helices shaping membrane protein folding. *Protein Science,* 24(7), pp 1057-1074.

Deaton, J., Sun, J., Holzenburg, A., Struck, D. K., Berry, J. & Young, R. 2004. Functional bacteriorhodopsin is efficiently solubilized and delivered to membranes by the chaperonin GroEL. *Proceedings of the National Academy of Sciences of the United States of America,* 101(8), pp 2281-2286.

del Alamo, M., Hogan, D. J., Pechmann, S., Albanese, V., Brown, P. O. & Frydman, J. 2011. Defining the Specificity of Cotranslationally Acting Chaperones by Systematic Analysis of mRNAs Associated with Ribosome-Nascent Chain Complexes. *Plos Biology,* 9(7), pp.

Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., Ren, W., Hirata, K., Yamamoto, M., Fan, S. & Yan, N. 2015. Molecular basis of ligand recognition and transport by glucose transporters. *Nature,* 526(7573), pp 391-396.

Deniaud, A., Liguori, L., Blesneac, I., Lenormand, J. L. & Pebay-Peyroula, E. 2010. Crystallization of the membrane protein hVDAC1 produced in cell-free system. *Biochimica Et Biophysica Acta-Biomembranes,* 1798(8), pp 1540-1546.

Deshaies, R. J., Sanders, S. L., Feldheim, D. A. & Schekman, R. 1991. Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature,* 349(6312), pp 806-808.

Do, H., Falcone, D., Lin, J. L., Andrews, D. W. & Johnson, A. E. 1996. The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. *Cell,* 85(3), pp 369-378.

Dodd, M. S., Papineau, D., Grenne, T., Slack, J. F., Rittner, M., Pirajno, F., O'Neil, J. & Little, C. T. S. 2017. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature,* 543(7643), pp 60-64.

Driessen, A. J. M. & Nouwen, N. 2008. Protein translocation across the bacterial cytoplasmic membrane. *Annual Review of Biochemistry,* 77, pp 643-667.

du Plessis, D. J. F., Berrelkamp, G., Nouwen, N. & Driessen, A. J. M. 2009. The Lateral Gate of SecYEG Opens during Protein Translocation. *Journal of Biological Chemistry,* 284(23), pp 15805-15814.

Du Plessis, D. J. F., Nouwen, N. & Driessen, A. J. M. 2011. The Sec translocase. *Biochimica Et Biophysica Acta-Biomembranes,* 1808(3), pp 851-865.

Ederth, J., Mandava, C. S., Dasgupta, S. & Sanyal, S. 2009. A single-step method for purification of active His-tagged ribosomes from a genetically engineered Escherichia coli. *Nucleic Acids Research,* 37(2), pp e15.

Egea, P. F., Shan, S. O., Napetschnig, J., Savage, D. F., Walter, P. & Stroud, R. M. 2004. Substrate twinning activates the signal recognition particle and its receptor. *Nature,* 427(6971), pp 215-221.

Elbaz, Y., Steiner-Mordoch, S., Danieli, T. & Schuldiner, S. 2004. In vitro synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *Proceedings of the National Academy of Sciences of the United States of America,* 101(6), pp 1519-1524.

Engelman, D. M., Chen, Y., Chin, C. N., Curran, A. R., Dixon, A. M., Dupuy, A. D., Lee, A. S., Lehnert, U., Matthews, E. E., Reshetnyak, Y. K., Senes, A. & Popot, J. L. 2003. Membrane protein folding: beyond the two stage model. *Febs Letters,* 555(1), pp 122-125.

Engelman, D. M. & Steitz, T. A. 1981. The Spontaneous Insertion of Proteins Into and Across Membranes - The Helical Hairpin Hypothesis. *Cell,* 23(2), pp 411-422.

Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science,* 255(5041), pp 178-183.

Ernst, O. P., Lodowski, D. T., Elstner, M., Hegemann, P., Brown, L. S. & Kandori, H. 2014. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chemical Reviews,* 114(1), pp 126-163.

Falk, M. M., Buehler, L. K., Kumar, N. M. & Gilula, N. B. 1997. Cell-free synthesis and assembly of connexins into functional gap junction membrane channels. *Embo Journal,* 16(10), pp 2703-2716.

Fernandez-Vidal, M., White, S. H. & Ladokhin, A. S. 2011. Membrane Partitioning: "Classical" and "Nonclassical" Hydrophobic Effects. *Journal of Membrane Biology*, 239(1-2), pp 5-14.

Findlay, H. E., Harris, N. J. & Booth, P. J. 2016. In vitro synthesis of a Major Facilitator Transporter for specific active transport across Droplet Interface Bilayers. *Scientific Reports,* 6, pp 39349.

Findlay, H. E., Rutherford, N. G., Henderson, P. J. F. & Booth, P. J. 2010. Unfolding free energy of a two-domain transmembrane sugar transport protein. *Proceedings of the National Academy of Sciences of the United States of America,* 107(43), pp 18451-18456.

Froderberg, L., Houben, E., Samuelson, J. C., Chen, M. Y., Park, S. K., Phillips, G. J., Dalbey, R., Luirink, J. & de Gier, J. W. L. 2003. Versatility of inner membrane protein biogenesis in Escherichia coli. *Molecular Microbiology,* 47(4), pp 1015-1027.

Fujii, S., Matsuura, T. & Yomo, T. 2015. Membrane Curvature Affects the Formation of a-Hemolysin Nanopores. *Acs Chemical Biology,* 10(7), pp 1694-1701.

Furusato, T., Horie, F., Matsubayashi, H. T., Amikura, K., Kuruma, Y. & Ueda, T. 2018. De Novo Synthesis of Basal Bacterial Cell Division Proteins FtsZ, FtsA, and ZipA Inside Giant Vesicles. *Acs Synthetic Biology,* 7(4), pp 953-961.

Gao, T., Petrlova, J., He, W., Huser, T., Kudlick, W., Voss, J. & Coleman, M. A. 2012. Characterization of De Novo Synthesized GPCRs Supported in Nanolipoprotein Discs. *Plos One,* 7(9), pp e44911.

Garavito, R. M. & Ferguson-Miller, S. 2001. Detergents as tools in membrane biochemistry. *Journal of Biological Chemistry,* 276(35), pp 32403-32406.

Garten, M., Mosgaard, L. D., Bornschloegl, T., Dieudonne, S., Bassereau, P. & Toombes, G. E. S. 2017. Whole-GUV patch-clamping. *Proceedings of the National Academy of Sciences of the United States of America,* 114(2), pp 328-333.

Glick, B. S. & VonHeijne, G. 1996. Saccharomyces cerevisiae mitochondria lack a bacterial-type Sec machinery. *Protein Science,* 5(12), pp 2651-2652.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. & Ferrin, T. E. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science,* 27(1), pp 14-25.

Gourdon, P., Alfredsson, A., Pedersen, A., Mahnerberg, E., Nyblom, M., Widell, M., Berntsson, R., Pinhassi, J., Braiman, M., Hansson, O., Bonander, N.,

Karlsson, G. & Neutze, R. 2008. Optimized in vitro and in vivo expression of proteorhodopsin: A seven-transmembrane proton pump. *Protein Expression and Purification,* 58(1), pp 103-113.

Groves, J. D., Wang, L. & Tanner, M. J. A. 1998. Functional reassembly of the anion transport domain of human red cell band 3 (AE1) from multiple and non-complementary fragments. *FEBS Letters,* 433(3), pp 223-227.

Gumbart, J. C., Ulmschneider, M. B., Hazel, A., White, S. H. & Ulmschneider, J. P. 2018. Computed Free Energies of Peptide Insertion into Bilayers are Independent of Computational Method. *Journal of Membrane Biology,* 251(3), pp 345-356.

Guna, A. & Hegde, R. S. 2018. Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Current Biology,* 28(8), pp R498-R511.

Gushchin, I. & Gordeliy, V. 2018. Microbial Rhodopsins. *Membrane Protein Complexes: Structure and Function,* 87, pp 19-56.

Halic, M., Becker, T., Pool, M. R., Spahn, C. M. T., Grassucci, R. A., Frank, J. & Beckmann, R. 2004. Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature,* 427(6977), pp 808-814.

Hamada, S., Tabuchi, M., Toyota, T., Sakurai, T., Hosoi, T., Nomoto, T., Nakatani, K., Fujinami, M. & Kanzaki, R. 2014. Giant vesicles functionally expressing membrane receptors for an insect pheromone. *Chemical Communications,* 50(22), pp 2958-2961.

Harris, N. J. & Booth, P. J. 2012. Folding and stability of membrane transport proteins in vitro. *Biochimica Et Biophysica Acta-Biomembranes,* 1818(4), pp 1055-1066.

Harris, N. J., Reading, E., Ataka, K., Grzegorzewski, L., Charalambous, K., Liu, X., Schlesinger, R., Heberle, J. & Booth, P. J. 2017. Structure formation during translocon-unassisted co-translational membrane protein folding. *Scientific Reports,* 7, pp 8021.

He, M. M. & Kaback, H. R. 1998. In vitro folding of a membrane protein: effect of denaturation and renaturation on substrate binding by the lactose permease of Escherichia coli. *Molecular Membrane Biology,* 15(1), pp 15-20.

Hell, K., Neupert, W. & Stuart, R. A. 2001. Oxa1p acts as a general membrane insertion machinery for proteins encoded by mitochondrial DNA. *Embo Journal,* 20(6), pp 1281-1288.

Henderson, P. J. F. & Maiden, M. C. J. 1990. Homologous sugar transport proteins in Escherichia coli and their relatives in both prokaryotes and eukaryotes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences,* 326(1236), pp 391-410.

Hodgman, C. E. & Jewett, M. C. 2012. Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering,* 14(3), pp 261-269.

Holt, C. E. & Bullock, S. L. 2009. Subcellular mRNA Localization in Animal Cells and Why It Matters. *Science,* 326(5957), pp 1212-1216.

Hoyt, D. W. & Gierasch, L. M. 1991. Hydrophobic Content and Lipid Interactions of Wild-Type and Mutant OmpA Signal Peptides Correlate With Their In Vivo Function. *Biochemistry,* 30(42), pp 10155-10163.

Huang, K. S., Bayley, H., Liao, M. J., London, E. & Khorana, H. G. 1981. Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *Journal of Biological Chemistry,* 256(8), pp 3802-3809.

Huppa, J. B. & Ploegh, H. L. 1997. In vitro translation and assembly of a complete T cell receptor-CD3 complex. *Journal of Experimental Medicine,* 186(3), pp 393-403.

Hutchison, C. A., III, Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z.-Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., Glass, J. I., Merryman, C., Gibson, D. G. & Venter, J. C. 2016. Design and synthesis of a minimal bacterial genome. *Science,* 351(6280), pp aad6253-1-aad6253-11.

Ishihara, G., Goto, M., Saeki, M., Ito, K., Hori, T., Kigawa, T., Shirouzu, M. & Yokoyama, S. 2005. Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors. *Protein Expression and Purification,* 41(1), pp 27-37.

Ismail, N., Hedman, R., Linden, M. & von Heijne, G. 2015. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nature Structural & Molecular Biology,* 22(2), pp 145-149.

Ito, K., Wittekind, M., Nomura, M., Shiba, K., Yura, T., Miura, A. & Nashimoto, H. 1983. A temperature-sensitive mutant of E. coli exhibiting slow processing of exported proteins. *Cell,* 32(3), pp 789-797.

Iwamoto, M., Elfaramawy, M. A., Yamatake, M., Matsuura, T. & Oiki, S. 2018. Concurrent In Vitro Synthesis and Functional Detection of Nascent Activity of the

KcsA Channel under a Membrane Potential. *Acs Synthetic Biology,* 7(4), pp 1004-1011.

Jacobs, R. E. & White, S. H. 1989. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. *Biochemistry,* 28(8), pp 3421-3437.

Jagannathan, S., Hsu, J. C. C., Reid, D. W., Chen, Q., Thompson, W. J., Moseley, A. M. & Nicchitta, C. V. 2014. Multifunctional Roles for the Protein Translocation Machinery in RNA Anchoring to the Endoplasmic Reticulum. *Journal of Biological Chemistry,* 289(37), pp 25907-25924.

Janas, T. & Yarus, M. 2003. Visualization of membrane RNAs. *RNA-a Publication of the RNA Society,* 9(11), pp 1353-1361.

Jewett, M. C. & Forster, A. C. 2010. Update on designing and building minimal cells. *Current Opinion in Biotechnology,* 21(5), pp 697-703.

Jia, H. & Schwille, P. 2019. Bottom-up synthetic biology: reconstitution in space and time. *Current opinion in biotechnology,* 60, pp 179-187.

Jia, L. X., Dienhart, M., Schramp, M., McCauley, M., Hell, K. & Stuart, R. A. 2003. Yeast Oxa1 interacts with mitochondrial ribosomes: the importance of the C-terminal region of Oxa1. *Embo Journal,* 22(24), pp 6438-6447.

Joseph, S. K., Boehning, D., Pierson, S. & Nicchitta, C. V. 1997. Membrane insertion, glycosylation, and oligomerization of inositol trisphosphate receptors in a cell-free translation system. *Journal of Biological Chemistry,* 272(3), pp 1579-1588.

Kaellberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H. & Xu, J. 2012. Template-based protein structure modeling using the RaptorX web server. *Nature Protocols,* 7(8), pp 1511-1522.

Kalmbach, R., Chizhov, I., Schumacher, M. C., Friedrich, T., Bamberg, E. & Engelhard, M. 2007. Functional cell-free synthesis of a seven helix membrane protein: In situ insertion of bacteriorhodopsin into liposomes. *Journal of Molecular Biology,* 371(3), pp 639-648.

Khvorova, A., Kwak, Y. G., Tamkun, M., Majerfeld, I. & Yarus, M. 1999. RNAs that bind and change the permeability of phospholipid membranes. *Proceedings of the National Academy of Sciences of the United States of America,* 96(19), pp 10649-10654.

Kiefer, D. & Kuhn, A. 1999. Hydrophobic forces drive spontaneous membrane insertion of the bacteriophage Pf3 coat protein without topological control. *Embo Journal,* 18(22), pp 6299-6306.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. & Yokoyama, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *Febs Letters,* 442(1), pp 15-19.

Klammt, C., Lohr, F., Schafer, B., Haase, W., Dotsch, V., Ruterjans, H., Glaubitz, C. & Bernhard, F. 2004. High level cell-free expression and specific labeling of integral membrane proteins. *European Journal of Biochemistry,* 271(3), pp 568-580.

Klostermann, E., Helling, I. D., Carde, J. P. & Schunemann, D. 2002. The thylakoid membrane protein ALB3 associates with the cpSecY-translocase in Arabidopsis thaliana. *Biochemical Journal,* 368, pp 777-781.

Kobilka, B. K. 1990. The role of cytosolic and membrane factors in processing of the human beta-2 adrenergic receptor following translocation and glycosylation in a cell-free system. *Journal of Biological Chemistry,* 265(13), pp 7610-7618.

Kobilka, B. K., Kobilka, T. S., Daniel, K., Regan, J. W., Caron, M. G. & Lefkowitz, R. J. 1988. Chimeric alpha 2-,beta 2-adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science,* 240(4857), pp 1310-1316.

Komar, J., Alvira, S., Schulze, R. J., Martin, R., Nijeholt, J. A. L. A., Lee, S. C., Dafforn, T. R., Deckers-Hebestreit, G., Berger, I., Schaffitzel, C. & Collinson, I. 2016. Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Biochemical Journal,* 473, pp 3341-3354.

Krolkiewicz, S., Sanger, H. L. & Niesbachklosgen, U. 1994. Structural and functional characterisation of the signal recognition particle-specific 54 kDa protein (SRP54) of tomato. *Molecular & General Genetics,* 245(5), pp 565-576.

Kucherak, O. A., Oncul, S., Darwich, Z., Yushchenko, D. A., Arntz, Y., Didier, P., Mely, Y. & Klymchenko, A. S. 2010. Switchable Nile Red-Based Probe for Cholesterol and Lipid Order at the Outer Leaflet of Biomembranes. *Journal of the American Chemical Society,* 132(13), pp 4907-4916.

Kuhn, A. & Kiefer, D. 2017. Membrane protein insertase YidC in bacteria and archaea. *Molecular Microbiology,* 103(4), pp 590-594.

Kumar, H., Kasho, V., Smirnova, I., Finer-Moore, J. S., Kaback, H. R. & Stroud, R. M. 2014. Structure of sugar-bound LacY. *Proceedings of the National Academy of Sciences of the United States of America,* 111(5), pp 1784-1788.

Kuruma, Y., Nishiyama, K., Shimizu, Y., Muller, M. & Ueda, T. 2005. Development of a minimal cell-free translation system for the synthesis of presecretory and integral membrane proteins. *Biotechnology Progress,* 21(4), pp 1243-1251.

Kuruma, Y. & Ueda, T. 2015. The PURE system for the cell-free synthesis of membrane proteins. *Nature Protocols,* 10(9), pp 1328-1344.

Ladokhin, A. S. & White, S. H. 1999. Folding of amphipathic alpha-helices on membranes: Energetics of helix formation by melittin. *Journal of Molecular Biology,* 285(4), pp 1363-1369.

Lane, N. & Martin, W. F. 2012. The Origin of Membrane Bioenergetics. *Cell,* 151(7), pp 1406-1416.

LeBarron, J. & London, E. 2016. Highly Hydrophilic Segments Attached to Hydrophobic Peptides Translocate Rapidly across Membranes. *Langmuir,* 32(41), pp 10752-10760.

Lee, H. C. & Bernstein, H. D. 2001. The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proceedings of the National Academy of Sciences of the United States of America,* 98(6), pp 3471-3476.

Lemmon, M. A. & Engelman, D. M. 1994. Specificity and promiscuity in membrane helix interactions. *Quarterly Reviews of Biophysics,* 27(2), pp 157-218.

Li, M., Khan, S., Rong, H., Tuma, R., Hatzakis, N. S. & Jeuken, L. J. C. 2017. Effects of membrane curvature and pH on proton pumping activity of single cytochrome bo(3) enzymes. *Biochimica Et Biophysica Acta-Bioenergetics,* 1858(9), pp 763-770.

Lind, C., Hojeberg, B. & Khorana, H. G. 1981. Reconstitution of delipidated bacteriorhodopsin with endogenous polar lipids. *Journal of Biological Chemistry,* 256(16), pp 8298-8305.

Liu, Y.-J., Hansen, G. P. R., Venancio-Marques, A. & Baigl, D. 2013. Cell-Free Preparation of Functional and Triggerable Giant Proteoliposomes. *Chembiochem,* 14(17), pp 2243-2247.

London, E. & Khorana, H. G. 1982. Denaturation and renaturation of bacteriorhodopsin in detergents and lipid-detergent mixtures. *Journal of Biological Chemistry,* 257(12), pp 7003-7011.

Long, A. R., O'Brien, C. C. & Alder, N. N. 2012. The Cell-Free Integration of a Polytopic Mitochondrial Membrane Protein into Liposomes Occurs Cotranslationally and in a Lipid-Dependent Manner. *Plos One,* 7(9), pp e46332.

Lu, W., Schafer, N. P. & Wolynes, P. G. 2018. Energy landscape underlying spontaneous insertion and folding of an alpha-helical transmembrane protein into a bilayer. *Nature communications,* 9(1), pp 4949-4949.

Luirink, J., High, S., Wood, H., Giner, A., Tollervey, D. & Dobberstein, B. 1992. Signal-sequence recognition by an Escherichia coli ribonucleoprotein complex. *Nature,* 359(6397), pp 741-743.

Luirink, J., Yu, Z., Wagner, S. & de Gier, J.-W. 2012. Biogenesis of inner membrane proteins in Escherichia coli. *Biochimica Et Biophysica Acta-Bioenergetics,* 1817(6), pp 965-976.

Lyford, L. K. & Rosenberg, R. L. 1999. Cell-free expression and functional reconstitution of homo-oligomeric alpha 7 nicotinic acetylcholine receptors into planar lipid bilayers. *Journal of Biological Chemistry,* 274(36), pp 25675-25681.

Ma, Y., Muench, D., Schneider, T., Sahl, H.-G., Bouhss, A., Ghoshdastider, U., Wang, J., Doetsch, V., Wang, X. & Bernhard, F. 2011. Preparative Scale Cell-free Production and Quality Optimization of MraY Homologues in Different Expression Modes. *Journal of Biological Chemistry,* 286(45), pp 38844-38853.

MacKenzie, K. R. 2006. Folding and stability of alpha-helical integral membrane proteins. *Chemical Reviews,* 106(5), pp 1931-1977.

Maggio, R., Vogel, Z. & Wess, J. 1993. Reconstitution of functional muscarinic receptors by co-expression of amino- and carboxyl-terminal receptor fragments. *Febs Letters,* 319(1-2), pp 195-200.

Marger, M. D. & Saier, M. H. 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends in Biochemical Sciences,* 18(1), pp 13-20.

Marsh, D. 2010. Liquid-ordered phases induced by cholesterol: A compendium of binary phase diagrams. *Biochimica Et Biophysica Acta-Biomembranes,* 1798(3), pp 688-699.

Martino, C. & deMello, A. J. 2016. Droplet-based microfluidics for artificial cell generation: a brief review. *Interface Focus,* 6(4), pp 20160011.

Martino, C., Kim, S. H., Horsfall, L., Abbaspourrad, A., Rosser, S. J., Cooper, J. & Weitz, D. A. 2012. Protein Expression, Aggregation, and Triggered Release

from Polymersomes as Artificial Cell-like Structures. *Angewandte Chemie-International Edition,* 51(26), pp 6416-6420.

Martyna, A., Bahsoun, B., Badham, M. D., Srinivasan, S., Howard, M. J. & Rossman, J. S. 2017. Membrane remodeling by the M2 amphipathic helix drives influenza virus membrane scission. *Scientific Reports,* 7, pp 44695.

Matsubayashi, H., Kuruma, Y. & Ueda, T. 2014. In Vitro Synthesis of the E. coli Sec Translocon from DNA. *Angewandte Chemie-International Edition,* 53(29), pp 7535-7538.

Matthies, D., Haberstock, S., Joos, F., Doetsch, V., Vonck, J., Bernhard, F. & Meier, T. 2011. Cell-Free Expression and Assembly of ATP Synthase. *Journal of Molecular Biology,* 413(3), pp 593-603.

Milano, F., Trotta, M., Dorogi, M., Fischer, B., Giotta, L., Agostiano, A., Maroti, P., Kalman, L. & Nagy, L. 2012. Light induced transmembrane proton gradient in artificial lipid vesicles reconstituted with photosynthetic reaction centers. *Journal of Bioenergetics and Biomembranes,* 44(3), pp 373-384.

Moon, C. P. & Fleming, K. G. 2011. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proceedings of the National Academy of Sciences of the United States of America,* 108(25), pp 10174-10177.

Moore, M., Harrison, M. S., Peterson, E. C. & Henry, R. 2000. Chloroplast Oxa1p homolog albino3 is required for post-translational integration off the light harvesting chlorophyll-binding protein into thylakoid membranes. *Journal of Biological Chemistry,* 275(3), pp 1529-1532.

Moritani, Y., Nomura, S.-i. M., Morita, I. & Akiyoshi, K. 2010. Direct integration of cell-free-synthesized connexin-43 into liposomes and hemichannel formation. *Febs Journal,* 277(16), pp 3343-3352.

Morowitz, H. J., Heinz, B. & Deamer, D. W. 1988. The chemical logic of a minimum protocell. *Origins of Life and Evolution of the Biosphere,* 18(3), pp 281-287.

Mravic, M., Thomaston, J. L., Tucker, M., Solomon, P. E., Liu, L. & DeGrado, W. F. 2019. Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science,* 363(6434), pp 1418-1423.

Mukherjee, S., Raghuraman, H. & Chattopadhyay, A. 2007. Membrane localization and dynamics of Nile Red: Effect of cholesterol. *Biochimica Et Biophysica Acta-Biomembranes,* 1768(1), pp 59-66.

Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C. & Jovine, L. 2003. Structure, function and evolution of the signal recognition particle. *Embo Journal,* 22(14), pp 3479-3485.

Nagamori, S., Smirnova, I. N. & Kaback, H. R. 2004. Role of YidC in folding of polytopic membrane proteins. *Journal of Cell Biology,* 165(1), pp 53-62.

Nallani, M., Andreasson-Ochsner, M., Tan, C.-W. D., Sinner, E.-K., Wisantoso, Y., Geifman-Shochat, S. & Hunziker, W. 2011. Proteopolymersomes: In vitro production of a membrane protein in polymersome membranes. *Biointerphases,* 6(4), pp 153-157.

Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S. & Amster-Choder, O. 2011. Translation-Independent Localization of mRNA in E. coli. *Science,* 331(6020), pp 1081-1084.

Nilsson, O. B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Mueller-Lucks, A., Trovato, F., Puglisi, J. D., O'Brien, E. P., Beckmann, R. & von Heijne, G. 2015. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Reports,* 12(10), pp 1533-1540.

Nishiyama, K., Hanada, M. & Tokuda, H. 1994. Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of Escherichia coli at low temperature. *Embo Journal,* 13(14), pp 3272-3277.

Noireaux, V. & Libchaber, A. 2004. A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences of the United States of America,* 101(51), pp 17669-17674.

Nomura, S.-i. M., Kondoh, S., Asayama, W., Asada, A., Nishikawa, S. & Akiyoshi, K. 2008. Direct preparation of giant proteo-liposomes by in vitro membrane protein synthesis. *Journal of Biotechnology,* 133(2), pp 190-195.

O'Neil, J., Carlson, R. W., Paquette, J.-L. & Francis, D. 2012. Formation age and metamorphic history of the Nuvvuagittuq Greenstone Belt. *Precambrian Research,* 220, pp 23-44.

Oesterhelt, D. & Stoeckenius, W. 1973. Functions of a new photoreceptor membrane. *Proceedings of the National Academy of Sciences of the United States of America,* 70(10), pp 2853-2857.

Ohta, N., Kato, Y., Watanabe, H., Mori, H. & Matsuura, T. 2016. In vitro membrane protein synthesis inside Sec translocon-reconstituted cell-sized liposomes. *Scientific Reports,* 6, pp 36466.

Okamura, K., Matsushita, S., Kato, Y., Watanabe, H., Matsui, A., Oka, T. & Matsuura, T. 2019. In vitro synthesis of the human calcium transporter Letm1 within cell-sized liposomes and investigation of its lipid dependency. *Journal of Bioscience and Bioengineering,* 127(5), pp 544-548.

Palczewski, K. 2006. G protein-coupled receptor rhodopsin. *Annual Review of Biochemistry,* 75, pp 743-767.

Pao, S. S., Paulsen, I. T. & Saier, M. H. 1998. Major facilitator superfamily. *Microbiology and Molecular Biology Reviews,* 62(1), pp 1-34.

Park, E., Menetret, J.-F., Gumbart, J. C., Ludtke, S. J., Li, W., Whynot, A., Rapoport, T. A. & Akey, C. W. 2014. Structure of the SecY channel during initiation of protein translocation. *Nature,* 506(7486), pp 102-106.

Paslawski, W., Lillelund, O. K., Kristensen, J. V., Schafer, N. P., Baker, R. P., Urban, S. & Otzen, D. E. 2015. Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proceedings of the National Academy of Sciences of the United States of America,* 112(26), pp 7978-7983.

Pellowe, G. A. & Booth, P. J. 2019. Structural insight into co-translational membrane protein folding. *Biochimica et biophysica acta. Biomembranes,* (In Press)

Peters, R. J. R. W., Nijemeisland, M. & van Hest, J. C. M. 2015. Reversibly Triggered Protein-Ligand Assemblies in Giant Vesicles. *Angewandte Chemie-International Edition,* 54(33), pp 9614-9617.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods,* 8(10), pp 785-786.

Petriman, N.-A., Jauss, B., Hufnagel, A., Franz, L., Sachelaru, I., Drepper, F., Warscheid, B. & Koch, H.-G. 2018. The interaction network of the YidC insertase with the SecYEG translocon, SRP and the SRP receptor FtsY. *Scientific Reports,* 8, pp 578.

Pohlschroder, M., Hartmann, E., Hand, N. J., Dilks, K. & Haddad, A. 2005. Diversity and evolution of protein translocation. *Annual Review of Microbiology,* 59, pp 91-111.

Popot, J. L. & Engelman, D. M. 1990. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry,* 29(17), pp 4031-4037.

Popot, J. L. & Engelman, D. M. 2000. Helical membrane protein folding, stability, and evolution. *Annual Review of Biochemistry,* 69, pp 881-922.

Popot, J. L., Gerchman, S. E. & Engelman, D. M. 1987. Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. *Journal of Molecular Biology,* 198(4), pp 655-676.

Popot, J. L., Trewhella, J. & Engelman, D. M. 1986. Reformation of crystalline purple membrane from purified bacteriorhodopsin fragments. *Embo Journal,* 5(11), pp 3039-3044.

Powers, T. & Walter, P. 1997. Co-translational protein targeting catalyzed by the Escherichia coli signal recognition particle and its receptor. *Embo Journal,* 16(16), pp 4880-4886.

Preuss, M., Ott, M., Funes, S., Luirink, J. & Herrmann, J. M. 2005. Evolution of mitochondrial Oxa proteins from bacterial YidC. *Journal of Biological Chemistry,* 280(13), pp 13004-13011.

Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., Kuehl, J. V., Melnyk, R. A., Lamson, J. S., Suh, Y., Carlson, H. K., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Zane, G. M., Rubin, B. E., Wall, J. D., Visel, A., Bristow, J., Blow, M. J., Arkin, A. P. & Deutschbauer, A. M. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature,* 557(7706), pp 503-509.

Quistgaard, E. M., Low, C., Guettou, F. & Nordlund, P. 2016. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology,* 17(2), pp 123-132.

Quistgaard, E. M., Low, C., Moberg, P., Tresaugues, L. & Nordlund, P. 2013. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nature Structural & Molecular Biology,* 20(6), pp 766-768.

Rapoport, T. A. 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature,* 450(7170), pp 663-669.

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America,* 106(6), pp 1760-1765.

Reckel, S., Gottstein, D., Stehle, J., Loehr, F., Verhoefen, M.-K., Takeda, M., Silvers, R., Kainosho, M., Glaubitz, C., Wachtveitl, J., Bernhard, F., Schwalbe,

H., Guentert, P. & Doetsch, V. 2011. Solution NMR Structure of Proteorhodopsin. *Angewandte Chemie-International Edition,* 50(50), pp 11942-11946.

Reid, D. W. & Nicchitta, C. V. 2012. Primary Role for Endoplasmic Reticulum-bound Ribosomes in Cellular Translation Identified by Ribosome Profiling. *Journal of Biological Chemistry,* 287(8), pp 5518-5527.

Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., Lee, D., Huang, C. L. V., Chandrasegaran, S., Cai, Y., Boeke, J. D. & Bader, J. S. 2017. Design of a synthetic yeast genome. *Science,* 355(6329), pp 1040-1044.

Ridder, A., van de Hoef, W., Stam, J., Kuhn, A., de Kruijff, B. & Killian, J. A. 2002. Importance of hydrophobic matching for spontaneous insertion of a single-spanning membrane protein. *Biochemistry,* 41(15), pp 4946-4952.

Robinson, P. J. & Woolhead, C. A. 2013. Post-translational membrane insertion of an endogenous YidC substrate. *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1833(12), pp 2781-2788.

Roos, C., Kai, L., Proverbio, D., Ghoshdastider, U., Filipek, S., Doetsch, V. & Bernhard, F. 2013. Co-translational association of cell-free expressed membrane proteins with supplied lipid bilayers. *Molecular Membrane Biology,* 30(1), pp 75-89.

Rosenberg, R. L. & East, J. E. 1992. Cell-free expression of functional Shaker potassium channels. *Nature,* 360(6400), pp 166-169.

Rues, R.-B., Doetsch, V. & Bernhard, F. 2016. Co-translational formation and pharmacological characterization of beta1-adrenergic receptor/nanodisc complexes with different lipid environments. *Biochimica Et Biophysica Acta-Biomembranes,* 1858(6), pp 1306-1316.

Russ, W. P. & Engelman, D. M. 2000. The GxxxG motif: A framework for transmembrane helix-helix association. *Journal of Molecular Biology,* 296(3), pp 911-919.

Sahin-Toth, M., Lawrence, M. C., Nishio, T. & Kaback, H. R. 2001. The C-4 hydroxyl group of galactopyranosides is the major determinant for ligand recognition by the lactose permease of Escherichia coli. *Biochemistry,* 40(43), pp 13015-13019.

Sardis, M. F. & Economou, A. 2010. SecA: a tale of two protomers. *Molecular Microbiology,* 76(5), pp 1070-1081.

Schibich, D., Gloge, F., Poehner, I., Bjorkholm, P., Wade, R. C., von Heijne, G., Bukau, B. & Kramer, G. 2016. Global profiling of SRP interaction with nascent polypeptides. *Nature,* 536(7615), pp 219-223.

Schopf, J. W. 1993. Microfossils of the Early Archean Apex Chert - New Evidence of the Antiquity of Life *Science,* 260(5108), pp 640-646.

Schopf, J. W., Kitajima, K., Spicuzza, M. J., Kudryavtsev, A. B. & Valley, J. W. 2018. SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proceedings of the National Academy of Sciences of the United States of America,* 115(1), pp 53-58.

Scott, A., Noga, M. J., de Graaf, P., Westerlaken, I., Yildirim, E. & Danelon, C. 2016. Cell-Free Phospholipid Biosynthesis by Gene-Encoded Enzymes Reconstituted in Liposomes. *Plos One,* 11(10), pp.

Scotti, P. A., Urbanus, M. L., Brunner, J., de Gier, J. W. L., von Heijne, G., van der Does, C., Driessen, A. J. M., Oudega, B. & Luirink, J. 2000. YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *Embo Journal,* 19(4), pp 542-549.

Seddon, A. M., Curnow, P. & Booth, P. J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica Et Biophysica Acta-Biomembranes,* 1666(1-2), pp 105-117.

Seelig, J. & Ganz, P. 1991. Nonclassical Hydrophobic Effect In Membrane-Binding Equilibria. *Biochemistry,* 30(38), pp 9354-9359.

Segre, D., Ben-Eli, D., Deamer, D. W. & Lancet, D. 2001. The lipid world. *Origins of Life and Evolution of the Biosphere,* 31(1-2), pp 119-145.

Seppala, S., Slusky, J. S., Lloris-Garcera, P., Rapp, M. & von Heijne, G. 2010. Control of Membrane Protein Topology by a Single C-Terminal Residue. *Science,* 328(5986), pp 1698-1700.

Shanmugaw, S. K. & Dalbey, R. E. 2019. The Conserved Role of YidC in Membrane Protein Biogenesis. *Microbiology Spectrum,* 7(1), pp.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. & Ueda, T. 2001. Cell-free translation reconstituted with purified components. *Nature Biotechnology,* 19(8), pp 751-755.

Shin, J., Jardine, P. & Noireaux, V. 2012. Genome Replication, Synthesis, and Assembly of the Bacteriophage T7 in a Single Cell-Free Reaction. *Acs Synthetic Biology,* 1(9), pp 408-413.

Snider, C., Jayasinghe, S., Hristova, K. & White, S. H. 2009. MPEx: A tool for exploring membrane proteins. *Protein Science,* 18(12), pp 2624-2628.

Soga, H., Fujii, S., Yomo, T., Kato, Y., Watanabe, H. & Matsuura, T. 2014. In Vitro Membrane Protein Synthesis Inside Cell-Sized Vesicles Reveals the Dependence of Membrane Protein Integration on Vesicle Volume. *Acs Synthetic Biology,* 3(6), pp 372-379.

Solieri, L. 2010. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. *Trends in Microbiology,* 18(11), pp 521-530.

Sommer, M. S. & Schleiff, E. 2014. Protein Targeting and Transport as a Necessary Consequence of Increased Cellular Complexity. *Cold Spring Harbor Perspectives in Biology,* 6(8), pp.

Song, W. Q., Raden, D., Mandon, E. C. & Gilmore, R. 2000. Role of Sec61 alpha in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel. *Cell,* 100(3), pp 333-343.

Soto-Rodriguez, J. & Baneyx, F. 2019. Role of the signal sequence in proteorhodopsin biogenesis in E-coli. *Biotechnology and Bioengineering,* 116(4), pp 912-918.

Steinberg-Yfrach, G., Rigaud, J. L., Durantini, E. N., Moore, A. L., Gust, D. & Moore, T. A. 1998. Light-driven production of ATP catalysed by F0F1-ATP synthase in an artificial photosynthetic membrane. *Nature,* 392(6675), pp 479-482.

Steinberg, R., Knuepffer, L., Origi, A., Asti, R. & Koch, H.-G. 2018. Co-translational protein targeting in bacteria. *Fems Microbiology Letters,* 365(11), pp.

Stirling, C. J., Rothblatt, J., Hosobuchi, M., Deshaies, R. & Schekman, R. 1992. Protein translocation mutants defective in the insertion of integral membrane proteins into the endoplasmic reticulum. *Molecular Biology of the Cell,* 3(2), pp 129-142.

Strack, R. L., Disney, M. D. & Jaffrey, S. R. 2013. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nature Methods,* 10(12), pp 1219-1224.

Szyrach, G., Ott, M., Bonnefoy, N., Neupert, W. & Herrmann, J. M. 2003. Ribosome binding to the Oxa1 complex facilitates co-translational protein insertion in mitochondria. *Embo Journal,* 22(24), pp 6448-6457.

Tanaka, Y., Sugano, Y., Takemoto, M., Mori, T., Furukawa, A., Kusakizako, T., Kumazaki, K., Kashima, A., Ishitani, R., Sugita, Y., Nureki, O. & Tsukazaki, T. 2015. Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. *Cell Reports,* 13(8), pp 1561-1568.

Tangorra, R. R., Operamolla, A., Milano, F., Omar, O. H., Henrard, J., Comparelli, R., Italiano, F., Agostiano, A., De Leo, V., Marotta, R., Falqui, A., Farinola, G. M. & Trotta, M. 2015. Assembly of a photosynthetic reaction center with ABA tri-block polymersomes: highlights on protein localization. *Photochemical & Photobiological Sciences,* 14(10), pp 1844-1852.

Tsukazaki, T., Mori, H., Echizen, Y., Ishitani, R., Fukai, S., Tanaka, T., Perederina, A., Vassylyev, D. G., Kohno, T., Maturana, A. D., Ito, K. & Nureki, O. 2011. Structure and function of a membrane component SecDF that enhances protein export. *Nature,* 474(7350), pp 235-238.

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(7215), pp 988-991.

Tuckey, C., Asahara, H., Zhou, Y. & Chong, S. 2014. Protein synthesis using a reconstituted cell-free system. *Current protocols in molecular biology,* 108(16.31), pp 1-22.

Tunuguntla, R., Bangar, M., Kim, K., Stroeve, P., Ajo-Franklin, C. M. & Noy, A. 2013. Lipid Bilayer Composition Can Influence the Orientation of Proteorhodopsin in Artificial Membranes. *Biophysical Journal,* 105(6), pp 1388-1396.

Ueno, T., Kaneko, K., Sata, T., Hattori, S. & Ogawa-Goto, K. 2012. Regulation of polysome assembly on the endoplasmic reticulum by a coiled-coil protein, p180. *Nucleic Acids Research,* 40(7), pp 3006-3017.

Ulmschneider, J. P., Smith, J. C., White, S. H. & Ulmschneider, M. B. 2011. In Silico Partitioning and Transmembrane Insertion of Hydrophobic Peptides under Equilibrium Conditions. *Journal of the American Chemical Society,* 133(39), pp 15487-15495.

Ulmschneider, J. P., Smith, J. C., White, S. H. & Ulmschneider, M. B. 2018. The importance of the membrane interface as the reference state for membrane protein stability. *Biochimica Et Biophysica Acta-Biomembranes,* 1860(12), pp 2539-2548.

Ulmschneider, M. B., Ulmschneider, J. P., Freites, J. A., von Heijne, G., Tobias, D. J. & White, S. H. 2017. Transmembrane helices containing a charged arginine

are thermodynamically stable. *European Biophysics Journal with Biophysics Letters,* 46(7), pp 627-637.

Ulmschneider, M. B., Ulmschneider, J. P., Schiller, N., Wallace, B. A., von Heijne, G. & White, S. H. 2014. Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. *Nature Communications,* 5, pp 4863.

Uyeda, A., Nakayama, S., Kato, Y., Watanabe, H. & Matsuura, T. 2016. Construction of an in Vitro Gene Screening System of the E. coli EmrE Transporter Using Liposome Display. *Analytical Chemistry,* 88(24), pp 12028-12035.

Valent, Q. A., Scotti, P. A., High, S., de Gier, J. W. L., von Heijne, G., Lentzen, G., Wintermeyer, W., Oudega, B. & Luirink, J. 1998. The Escherichia coli SRP and SecB targeting pathways converge at the translocon. *Embo Journal,* 17(9), pp 2504-2512.

van Dalen, A., van der Laan, M., Driessen, A. J. M., Killian, J. A. & de Kruijffa, B. 2002. Components required for membrane assembly of newly synthesized K+ channel KcsA. *Febs Letters,* 511(1-3), pp 51-58.

van den Berg, B., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. 2004. X-ray structure of a protein-conducting channel. *Nature,* 427(6969), pp 36-44.

van der Laan, M., Nouwen, N. P. & Driessen, A. J. M. 2005. YidC - an evolutionary conserved device for the assembly of energy-transducing membrane protein complexes. *Current Opinion in Microbiology,* 8(2), pp 182-187.

van Geest, M. & Lolkema, J. S. 2000. Membrane topology and insertion of membrane proteins: Search for topogenic signals. *Microbiology and Molecular Biology Reviews,* 64(1), pp 13-33.

vanKlompenburg, W., Nilsson, I., vonHeijne, G. & deKruijff, B. 1997. Anionic phospholipids are determinants of membrane protein topology. *Embo Journal,* 16(14), pp 4261-4266.

Vitrac, H., Bogdanov, M. & Dowhan, W. 2013. In vitro reconstitution of lipid-dependent dual topology and postassembly topological switching of a membrane protein. *Proceedings of the National Academy of Sciences of the United States of America,* 110(23), pp 9338-9343.

Vitrac, H., MacLean, D. M., Jayaraman, V., Bogdanov, M. & Dowhan, W. 2015. Dynamic membrane protein topological switching upon changes in phospholipid

environment. *Proceedings of the National Academy of Sciences of the United States of America,* 112(45), pp 13874-13879.

Von Heijne, G. 1992. Membrane protein structure prediction hydrophobicity analysis and the positive-inside rule. *Journal of Molecular Biology,* 225(2), pp 487-494.

Von Heijne, G. & Blomberg, C. 1979. Trans-membrane translocation of proteins. The direct transfer model. *European Journal of Biochemistry,* 97(1), pp 175-182.

Walde, P., Cosentino, K., Engel, H. & Stano, P. 2010. Giant Vesicles: Preparations and Applications. *Chembiochem,* 11(7), pp 848-865.

Wallin, E. & von Heijne, G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science,* 7(4), pp 1029-1038.

Walter, J. M., Greenfield, D., Bustamante, C. & Liphardt, J. 2007. Light-powering Escherichia coli with proteorhodopsin. *Proceedings of the National Academy of Sciences of the United States of America,* 104(7), pp 2408-2412.

Walter, P., Keenan, R. & Schmitz, U. 2000. Structural biology - SRP - Where the RNA and membrane worlds meet. *Science,* 287(5456), pp 1212-1213.

White, S. H., Ladokhin, A. S., Jayasinghe, S. & Hristova, K. 2001. How membranes shape protein structure. *Journal of Biological Chemistry,* 276(35), pp 32395-32398.

White, S. H. & Wimley, W. C. 1998. Hydrophobic interactions of peptides with membrane interfaces. *Biochimica Et Biophysica Acta-Reviews on Biomembranes,* 1376(3), pp 339-352.

White, S. H. & Wimley, W. C. 1999. Membrane protein folding and stability: Physical principles. *Annual Review of Biophysics and Biomolecular Structure,* 28, pp 319-365.

Wickstrom, D., Wagner, S., Baars, L., Ytterberg, A. J., Klepsch, M., van Wijk, K. J., Luirink, J. & de Gier, J.-W. 2011. Consequences of Depletion of the Signal Recognition Particle in Escherichia coli. *Journal of Biological Chemistry,* 286(6), pp 4598-4609.

Wiener, M. C. & White, S. H. 1992. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophysical Journal,* 61(2), pp 434-447.

Wilkinson, B. M., Esnault, Y., Craven, R. A., Skiba, F., Fieschi, J., Kepes, F. & Stirling, C. J. 1997. Molecular architecture of the ER translocase probed by chemical crosslinking of Sss1p to complementary fragments of Sec61p. *Embo Journal,* 16(15), pp 4549-4559.

Willmund, F., del Alamo, M., Pechmann, S., Chen, T., Albanese, V., Dammer, E. B., Peng, J. & Frydman, J. 2013. The Cotranslational Function of Ribosome-Associated Hsp70 in Eukaryotic Protein Homeostasis. *Cell,* 152(1-2), pp 196-209.

Wimley, W. C. & White, S. H. 1992. Partitioning of tryptophan side-chain analogs between water and cyclohexane. *Biochemistry*, 31(31), pp 12813-12818.

Wimley, W. C. & White, S. H. 1993. Membrane partitioning - distinguishing bilayer effects from the hydrophobic effect. *Biochemistry*, 32(25), pp 6307-6312.

Wimley, W. C. & White, S. H. 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology,* 3(10), pp 842-848.

Wimley, W. C. & White, S. H. 2000. Designing transmembrane alpha-helices that insert spontaneously. *Biochemistry,* 39(15), pp 4432-4442.

Winterstein, L.-M., Kukovetz, K., Rauh, O., Turman, D. L., Braun, C., Moroni, A., Schroeder, I. & Thiel, G. 2018. Reconstitution and functional characterization of ion channels from nanodiscs in lipid bilayers. *Journal of General Physiology,* 150(4), pp 637-646.

Wrubel, W., Stochaj, U. & Ehring, R. 1994. Construction and in vivo analysis of new split lactose permeases. *Febs Letters,* 349(3), pp 433-438.

Xu, C., Hu, S. & Chen, X. 2016. Artificial cells: from basic science to applications. *Materials Today,* 19(9), pp 516-532.

Yoshimura, K., Usukura, J. & Sokabe, M. 2008. Gating-associated conformational changes in the mechanosensitive channel MscL. *Proceedings of the National Academy of Sciences of the United States of America,* 105(10), pp 4033-4038.

Yu, H., Siewny, M. G. W., Edwards, D. T., Sanders, A. W. & Perkins, T. T. 2017. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science,* 355(6328), pp 945-949.

Yu, W., Sato, K., Wakabayashi, M., Nakaishi, T., Ko-Mitamura, E. P., Shima, Y., Urabe, I. & Yomo, T. 2001. Synthesis of functional protein in liposome. *Journal of Bioscience and Bioengineering,* 92(6), pp 590-593.

Zhang, J., Pan, X., Yan, K., Sun, S., Gao, N. & Sui, S.-F. 2015. Mechanisms of ribosome stalling by SecM at multiple elongation steps. *Elife,* 4, pp e09684.

Zheng, H., Taraska, J., Merz, A. J. & Gonen, T. 2010. The Prototypical H+/Galactose Symporter GalP Assembles into Functional Trimers. *Journal of Molecular Biology,* 396(3), pp 593-601.

Zhou, F. X., Merianos, H. J., Brunger, A. T. & Engelman, D. M. 2001. Polar residues drive association of polyleucine transmembrane helices. *Proceedings of the National Academy of Sciences of the United States of America,* 98(5), pp 2250-2255.

Appendix I

The following tables show primer sequences used for all cloning reactions. Restriction sites are underlined, stock codons are in bold and start codons ar in bold and italicised.

Table A1 Primers used to generate cell-free expression constructs used in Chapter 3

Construct	Primer Name	Primer Sequence (5' to 3')
Vector Linearisation	KpnI_RP	GATCGGTACCTATATCTCCTTCTTAAAGTT
	SpeI_RP	GATCACTAGTTATATCTCCTTCTTAAAGTT
	XhoI_FP	GATCCTCGAGTAAGGTTAACCTGC
	BamHI_FP	GATCGGATCCGGTTAACCTGCAGG
	SphI_FP	GATCGCATGCGGTTAACCTGCAGG
EGFP	EGFP_FP	GATCGGTACC***ATG***GTGAGCAAGGGCGAGG
	EGFP_RP	GATCCTCGAGTCATTACTTGTACAGCTCGTCCATGCC
PR	PR_FP	GATCGGTACC***ATG***GGTAAATTATTACTGATATTAGGTAGTG
	PR_RP	GATCCTCGAGATTATTAGCATTAGAAGATTCTTTAACAGCAAC
PR-EGFP	PR_RP	GATCGGATCCAGCATTAGAAGATTCTTTAACAGCAAC
	EGFP_FP	AATAGGATCC*GGCTCG*ATGGTGAGCAAGGGCGAGG
GalP	GalP_FP	CATGACTAGT***ATG***CCTGACGCTAAAAAACAGG
	GalP_RP	CATGCTCGAGTTAATCGTGAGCGCCTATTTCG
GalP-mCherry	GalP_RP	GATCGCATGCATCGTGAGCGCCTATTTCGC
	mCherry_FP	CATGGCATGC*GGTAGT*ATGGTGAGCAAGGGCGAGGAGG
	mCherry_RP	CATGCTCGAGTTACTTGTACAGCTCGTCCATGCCG
PR-EL1HA	PREL1_FP	ATGATGTGCCGGATTATGCGACTGTATTTAGATACATTGATTGGT
	PREL1_RP	TCCGGCACATCATACGGATAAATCCATACCCCTCTCATGT
	EL1HA_FP	ACATGAGAGGGGTATGGATTTATCCGTATGATGTGCCGGATTATGCG
	EL1HA_RP	TCAATGTATCTAAATACAGTCGCATAATCCGGCACATCATACGGATA
PR-IL3HA	PRIL3_FP	ATGATGTGCCGGATTATGCGGCTGTGCAATCAGCTTACAA
	PRIL3_RP	TCCGGCACATCATACGGATATTTTCCTTCTCCAGCCCATA
	PRIL3HA_FP	TATGGGCTGGAGAAGGAAAATATCCGTATGATGTGCCGGATTATGCG
	PRIL3HA_RP	TTGTAAGCTGATTGCACAGCCGCATAATCCGGCACATCATACGGATA
GalP-EL3HA	GalPEL3_FP	ATGATGTGCCGGATTATGCGACCGGTGCATGGCGCTGGAT
	GalPEL3_RP	TCCGGCACATCATACGGATAATCAGAAAGATAAGCACCGAGGATCCC
	GalPEL3HA_FP	TCGGTGCTTATCTTTCTGATTATCCGTATGATGTGCCGGATTATGCG
	GalPEL3HA_rP	ATCCAGCGCCATGCACCGGTCGCATAATCCGGCACATCATACGGATA
GalP-IL3HA	GalPIL3_FP	ATGATGTGCCGGATTATGCGGAAGCGAAACGCGAACTGGA
	GalPIL3_RP	TCCGGCACATCATACGGATAGTCACGCAGGCGTAGCAGCA
	GalPIL3HA_FP	TGCTGCTACGCCTGCGTGACTATCCGTATGATGTGCCGGATTATGCG
	GalPIL3HA_rP	TCCAGTTCGCGTTTCGCTTCCGCATAATCCGGCACATCATACGGATA

Table A2 Primers used to generate cell-free expression constructs used in Chapter 4

Construct	Primer Name	Primer Sequence (5' to 3')
Vector Linearisation	SalI_FP	GATCGGATCCGGTTAACCTGCAGG
PR-Spinach2	PRSp2_FP	GATCGGATCCGATGTAACTGAATGAAATGGTGAAGG
	PRSp2_RP	GATCCTCGAGGATGTAACTAGTTACGGAGCTCAC
GalP-Spinach2	GalPSp2_FP	GATCGTCGACGATGTAACTGAATGAAATGGTGAAG
	GalPSp2_RP	Same as PRSp2_RP
PR-TolA	TolA_FP	CATGGGATCCCAGAAGCAAGCTGAAGAGGCGG
	TolA_RP	CATGCTCGAGTTCAGTTGCCGCTTTCTTTCTTGCTTC
PR-TolA-SecM	SecM_FP	GATCGCATGCTTCAGCACGCCCGTCTGGATAAG
	SecM_RP	GATCCTCGAGAGGGCCAGCACGGATGCC
	PR-TolA_RP	CATGGCATGCTTCAGTTGCCGCTTTCTTTCTTGC
GalP-TolA-SecM	TolA_FP	CATGGAGCTCCAGAAGCAAGCTGAAGAGGC
	GalP_RP	CATGGAGCTCTGCATCGTGAGCGCCTATTTCG
PRΔSP	PRΔSP_FP	GATCGGTACC***ATG***GGTGGTGGTGACCTTGATG
GalP-TM1	GalP-TM1_FP	GATCGTCGACGCTGGCGCACTGCCGTTTATTG
GalP-N-TM1	GalP-N-TM1_FP	GATCGTCGAC***ATG***GCTGGCGCACTGCCGTTTATTG
GalP-N-TM1+2	GalP-N-TM1+2_FP	GATCGTCGAC***ATG***AAACTCGGGCGCAAAAAGAGC
PRSP	PRSP_FP	GATCGGTACC***ATG***GGTAAATTATTACTGATATTAGGTAGTG
	PRSP_RP	CATGCTCGAGTTATGCAGCAAATGTAGGAAGTGC
PRSP-EGFP	PRSPEGFP_FP	Same as PRSP_FP
	PRSPEGFP_RP	CATGGGATCCTGCAGCAAATGTAGGAAGTGC
PRTM1-EGFP	PRTM1_FP	GATCGGTACC***ATG***GGTGTTTCTTTTTGGTTAGTTAC
	PRTM1_RP	CATGGGATCCAACTCTATCTCTTTCAACAAAGAAAAATACAG
PRSP+TM1	PRSP+TM1_FP	Same as PRSP_FP
	PRSP+TM1_RP	CATGCTCGAGTTAAACTCTATCTCTTTCAACAAAGAAAAATACAG
PRSP+TM1-EGFP	SPTM1_FP	Same as PRSP_FP
	SPTM1_RP	Same as PRTM1_RP
PRΔ1	PRΔ1_FP	CATGGGTACC***ATG***TTATTACTGATATTAGGTAGTGTTATTGCACTTC
PRΔ2	PRΔ2_FP	CATGGGTACC***ATG***TTACTGATATTAGGTAGTGTTATTGCACTTCC
PRΔ3	PRΔ3_FP	GATCGGTACC***ATG***CTGATATTAGGTAGTGTTATTGCACTTCCTAC
PRΔ4	PRΔ4_FP	GATCGGTACC***ATG***ATATTAGGTAGTGTTATTGCACTTCCTACATTTG
PRΔ5	PRΔ5_FP	GATCGGTACC***ATG***TTAGGTAGTGTTATTGCACTTCCTACATTTG
PRΔ6	PRΔ6_FP	GATCGGTACC***ATG***GGTAGTGTTATTGCACTTCCTACATTTGC
PRΔ7	PRΔ7_FP	GATCGGTACC***ATG***AGTGTTATTGCACTTCCTACATTTGC
PRΔ8	PRΔ8_FP	GATCGGTACC***ATG***GTTATTGCACTTCCTACATTTGCTGC
PRΔ9	PRΔ9_FP	GATCGGTACC***ATG***ATTGCACTTCCTACATTTGCTGCAGG
PRΔ10	PRΔ10_FP	GATCGGTACC***ATG***GCACTTCCTACATTTGCTGCAGGTGG
PRΔ11	PRΔ11_FP	GATCGGTACC***ATG***CTTCCTACATTTGCTGCAGGTGG
PRΔ12	PRΔ12_FP	GATCGGTACC***ATG***CCTACATTTGCTGCAGGTGGTGG
PRΔ13	PRΔ13_FP	GATCGGTACC***ATG***ACATTTGCTGCAGGTGGTGG
PRΔ14	PRΔ14_FP	GATCGGTACC***ATG***TTTGCTGCAGGTGGTGGTGAC

PRΔ15	PRΔ15_FP	GATCGGTACC***ATG***GCTGCAGGTGGTGGTGACC
PRΔ16	PRΔ16_FP	GATCGGTACC***ATG***GCAGGTGGTGGTGACCTTG
PR T15A	PR T15A_FP	GTGTTATTGCACTTCCTGCATTTGCTGCAGGTGGT
	PR T15A_RP	ACCACCTGCAGCAAATGCAGGAAGTGCAATAACAC

Table A3 Primers used to generate cell-free expression constructs used in Chapter 5

Construct	Primer Name	Primer Sequence (5' to 3')
α-Hemolysin	HL_FP	GATCACTAGT***ATG***AAAACACGTATAGTCAGCTC
	HL_RP	GATCCTCGAGTCATTTTTCTTTTTCCCAATCG