Toward the Dynamic Structure of the *Escherichia coli* Integral Membrane Protein Glycerol Facilitator

By

Mary D. Hernando

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

University of Manitoba

Winnipeg

Abstract

The aquaporin family of integral membrane proteins (IMPs) are found in all forms of life, from prokaryotes to humans. *Escherichia coli* glycerol facilitator (GF) is a member of the aquaglyceroporin family that allows for the highly selective passive diffusion of its substrate glycerol across the inner membrane of the bacterium. Though the structure of GF was solved in 2000, little is known about the dynamics of GF and the role the dynamics play in the function and stability of the protein. This lack of information pertaining to dynamics is common for many IMPs. Here, preparations of isotope-labelled GF for solution and solid-state nuclear magnetic resonance (NMR) spectroscopy are explored and optimized in order to reveal atomic dynamics of the protein. The stability of the GF homotetramer solubilized in various agents, including detergents, bicelles, lipid nanodiscs, random heteropolymers, and other buffer additives is explored. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), intrinsic tryptophan fluorescence, size exclusion chromatography multi-angle light scattering (SEC-MALS), dynamic light scattering (DLS), and differential scanning calorimetry (DSC) were used to determine the oligomeric forms of GF and probe its stability. Using these techniques, it was found that GF tetramers self-associate to form octamers and higher M_r oligomers 24 hours after solubilization. The protein was found to be most stable in lauryl maltose neopentyl glycol (LMNG), where it existed in a tetramer-octamer equilibrium for 9 days. The structures of the GF tetramer and octamer were explored by negative stain electron microscopy (EM), size-exclusion chromatography small-angle light scattering (SEC-SAXS), and solid-state magic-angle spinning (MAS) NMR spectroscopy. Although NMR

sample preparation still needs optimization for full structure determination, negative

stain EM and SEC-SAXS revealed low-resolution structures of the detergent-solubilized

tetramer and octamer. The octamer forms from the association of the cytoplasmic faces of

two tetramers, the interaction apparently mediated by their disordered N- and C-termini.

Acknowledgements

First and foremost, to Joe: for the years of sage advice and the daily, frequent interruptions to pick your brain. I hope you don't mind that I've borrowed your mantra, "adapt or die"; I promise I've retained much more than that! I truly could not have asked for a better advisor and mentor. Thank you.

To my committee members: Deb Court, Mazdak Khajehpour, and Jörg Stetefeld. Thank you for the input, guidance, and support over the years. To my external examiner, Jan Rainey, for making the trek to Winnipeg for my defense and the fresh perspective on my thesis.

To George Oriss, Fraser Ferens, the Hoenger group at the University of Colorado, the Nieuwkoop group at Rutgers, Ting Xu at Berkeley, Matt McDougall, and Ewan McRae, who collaborated with me on this work.

To the past and present graduate and undergraduate students and volunteers I had the pleasure of working with in the lab; especially Vu To, who took me under his wing as a baby graduate student and taught me how to run a column beside my desk without *technically* breaking any rules.

Dedication

I am lucky to have been shaped by so many people over the years that it would take another chapter to do everyone of import justice, but the trajectory of my life would be completely different had it not been for the following people:

My wonderful husband, Charlie, who I'm not entirely certain was prepared to be married to a lifelong student, but embraced the role nonetheless. Words are not enough to express how much I appreciate you, so I'll say this: you are my heart. I will always be thankful that you've supported my dreams while working hard on ours.

My siblings, Chrissy and Tunes, who never fail to call me out and keep me grounded. Thanks for (still) letting me boss you around.

Finally, all I've accomplished would not have been possible without my parents, who immigrated to a new country in their mid-20s with 6-month-old me in tow. I could write pages and pages looking back at the ways you've impacted my life. I dedicate this to you both: for all the thankless work that went into providing for Chris, Tina, and I on what little support you had. For every sacrifice; for your unconditional dedication and perseverance in the face of hardship. I would never have gotten here without everything you've given. I am beyond thankful and eternally blessed to call you Mom and Dad. *Mahal na mahal ko po kayong dalawa.*

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A(0)	X-ray scattering amplitude
Apol	Amphipol
AQP	Aquaporin
BN-PAGE	Blue native poly-acrylamide gel electrophoresis
CHS	Cholesteryl hemisuccinate
DARR	Dipolar assisted rotational resonance
DDM	Dodecyl-β-D-maltoside
DHPC	1,2-dihexanoyl-sn-glycero-3-phosphocholine
DLS	Dynamic light scattering
D_{max}	Maximum molecule length
DMNG	Decyl maltose neopentyl glycol
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DNAse	Deoxyribonuclease
DSC	Differential scanning calorimetry
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
DTAB	Dodecyltrimethylammonium bromide
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
GF	Glycerol Facilitator
HDX	Hydrogen-deuterium exchange
IMAC	Immobilized metal affinity chromatography
IMP	Integral membrane protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
l(s)	X-ray scattering intensity
LB	Luria broth
LDAO	Lauryldimethylamine-N-oxide
LMNG	Lauryl maltose neopentyl glycol

LMPC	Lysomyristoylphosphatidylcholine
MALS	Multi-angle light scattering
MAS	Magic angle spinning
MBP	Maltose binding protein
MD	Molecular dynamics
MIP	Major intrinsic protein
MS	Mass spectrometry
MSP	Membrane scaffold protein
MWCO	Molecular weight cut off
NMO	Neuromyelitis optica
NMR	Nuclear magnetic resonance
OG	Octyl-β-D-glucopyranoside
$\square$(r)	Pairwise distance distribution function
PMP	Peripheral membrane protein
PMSF	Phenylmethyl-sulfonyl fluoride
$R_{1\rho}$	Rotating frame longitudinal relaxation
RD	Relaxation dispersion
RF	Radio frequency
R_g	Radius of gyration
RHP	Random heteropolymer
RI	Refractive index
rms	Root mean squared
RNAse	Ribonuclease A
s	SAXS scattering vector
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SMALP	Styrene maleic anhydride lipid particle
SUV	Small unilamellar lipid vesicle
TB	Terrific broth

TDM	Tetradecyl-β-D-maltoside
TEV	Tobacco etch virus
TM	Transmembrane
UV	Ultraviolet
V_p	Protein excluded volume
wt	Wild-type

Chapter One: Introduction

1.1   The Cell Membrane

All cells and their organelles are encompassed by a lipid membrane, an effective barrier that separates the inside of the cell or organelle from the outer environment. The cell membrane and its components are responsible for a variety of cellular functions, including: inter- and intracellular signalling, energy transduction, waste removal, and the transport of biologically important molecules[1].

The Fluid Mosaic model of the membrane, developed by Singer and Nicolson in 1972[2], is used to describe most cell membranes. The following sections highlight several historical observations and studies that laid the foundation on which Singer and Nicolson developed the Fluid Mosaic model of the membrane.

1.1.1 Films on Water

The use of oils to calm water surfaces was common knowledge when Gaius Plinius Secundus wrote about it in his encyclopedic series *Naturalis Historia*[3]. He noted that the fishermen in his time would use cooking oil to smooth rough waters to see schools of fish, though this remained purely observation until further experiments carried out by Benjamin Franklin in 1772[4].

Benjamin Franklin is known as the father of electricity; however, he not only experimented with lighting, but with calming waves on a pond in northern England. He added a teaspoonful of oil to the water and, in a letter that was published in Philosophical Transactions in the Royal Society of London[4], noted: "*the oil, though not more than a teaspoonful, produced a calm over a space... perhaps half an acre, smooth as a looking glass.*"

Benjamin Franklin's work was taken a step further by Lord Rayleigh (John William Strutt) in 1890 when he applied a known volume of olive oil onto a known surface area of water. With these parameters, he was able to calculate the molecular size of triolein (16 Å), the main triglyceride in olive oil[5].

Irving Langmuir earned a Nobel prize for his work in air-water interface studies. He repeated Franklin's oil on water experiment and reported the molecular length of triolein to be 13 Å[6]. He found that many amphipathic materials spread on water to form a film one molecule thick, evidenced by a decrease in surface tension once enough material was added to form a monolayer. He argued that since pure saturated hydrocarbons and benzene do not spread on water, and amphipathic molecules with regions that were hydrophilic, or water loving, and hydrophobic, or water averse, do spread, then the hydrophilic regions must be dissolved in water and the hydrophobic parts point into the air (Figure 1)[6]. Though we now know the membrane is a lipid bilayer composed of two

stacked phospholipid monolayers, information about the cell membrane and its

architecture only came to light several decades ago.
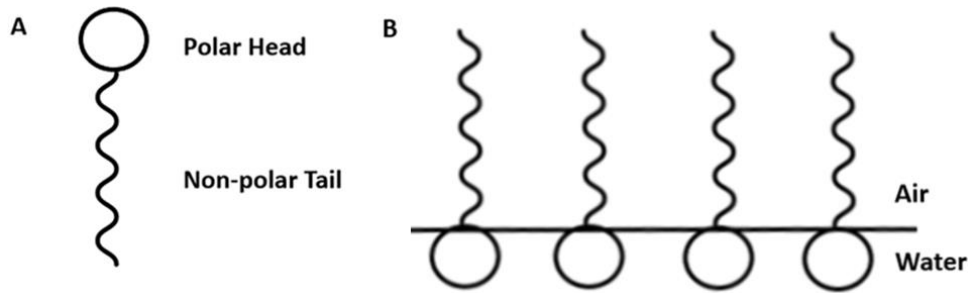
Figure 1. Irving Langmuir's description of surfactant structure and orientation at the air-water interface. (A) Model of an amphipathic surfactant molecule with the polar head (white circle) and nonpolar tail (wavy line). (B) The air-water interface with surfactant polar heads interacting with the water and nonpolar tails oriented into the air. Based on Stillwell, 2016[7].

1.1.2 Building Towards a Membrane Lipid Bilayer Model

Robert Hook studied the structure of cork cells and is credited with coining the term "cell" after observing the cork cell shape, which reminded him of monk's living quarters[8]. However, the earliest studies on cellular membranes were conducted on the plasma membrane. In 1773, William Hewson conducted microscopic studies on the shape of erythrocytes (red blood cells), where he observed osmotic swelling and shrinking in water. Noting the changes in cell shape in the presence and absence of water, he concluded the cells must have some structure encircling their protoplasm[9].

Charles Ernest Overton is considered the first "membranologist" whose experiments on the permeability of different types of cell membranes showed that the membrane was not only permeable to water, but to other molecules as well[10]. In 1899, he showed that the rate at which these molecules traversed the membrane depended on their charge: neutral moieties crossed the membrane more quickly than charged ones. He tested the solubility of these molecules in olive oil and found a correlation between the solute's permeability in the membrane and its solubility in olive oil, concluding that the membrane contains lipids. In addition to this, he also recognized that the solutes he tested were capable of passively diffusing across the membrane, but that other molecules may require assistance crossing the membrane[11].

The first studies on cell membranes that suggested a lipid bilayer were published by Gorter and Grendel in 1925[12]. They extracted erythrocyte membranes in benzene, which

were dried into lipid monolayers. At the time of their work, they knew the number of extracted erythrocytes that would produce a measured monolayer area and the rough surface area of an average erythrocyte. Using these parameters, they found that the monolayer surface area extracted from the erythrocytes was double the surface area of the cells; this led them to conclude that the membrane was a lipid bilayer. Though Gorter and Grendel's experiments correctly suggested that the membrane is a lipid bilayer, their methods were not without limitations[7]: erythrocytes are among the only eukaryotic cells without organelles or other extensive membrane systems that would increase the surface area of the lipid monolayer. In addition, benzene is a poor solvent for extracting lipids; extraction was incomplete (estimated to be approximately 70%), and their calculation that the cell membrane consisted of 100% lipid meant that there was no room for integral membrane proteins (IMPs). Nonetheless, these errors favourably cancelled out and their work set the foundation on which lipid bilayer membrane models would be built.

The late 1920s saw the rest of the pieces of the proverbial cell membrane lipid bilayer puzzle fall into place. In the same year of Gorter and Grendel's lipid bilayer discovery, Leathes and Raper suggested that phospholipids are key structural components in the cell[13], while Hugo Fricke measured the length of the erythrocyte membrane[14]. Interestingly, while the measured 33 Å was double the monolayer length Lord Rayleigh measured (16 Å), he did not acknowledge that the membrane was a lipid bilayer. One year later, urease was isolated by James B. Sumner, the first enzyme ever reported to be isolated[15]. Within the next decade, a membrane lipid bilayer model would emerge.

1.1.3 Early Membrane Models

It is in 1935 that the advances in membrane biology from the previous decade were used to create the first model of the membrane lipid bilayer. At the time, it was accepted that proteins played some role in the structure of the membrane, that there were various types of membranes that carried out different functions, and that membranes were semipermeable[16]. All these notions, however, raised more questions than they revealed answers. While it was accepted that proteins were structurally important in membranes, all proteins studied thus far were water-soluble and did not reflect the functions of membrane proteins. The variety of membranes and their functions raised questions about how different these membranes were, and what common features they might share. Lastly, how did this tie into the permeability of the membrane?

The Pauci-Molecular model introduced by Danielli and Davson[17] is the first representation of the membrane lipid bilayer that was mostly accepted by the scientific community. While it was known that proteins play some role in the structure of the membrane, all proteins studied at that point were globular and water soluble; to account for this, Danielli and Davson included a layer of loosely associated water-soluble protein to both sides of the membrane in their initial proposal (Figure 2A). After recognizing that the membrane was semipermeable, Danielli later modified the model to include peptide-lined pores (Figure 2B).
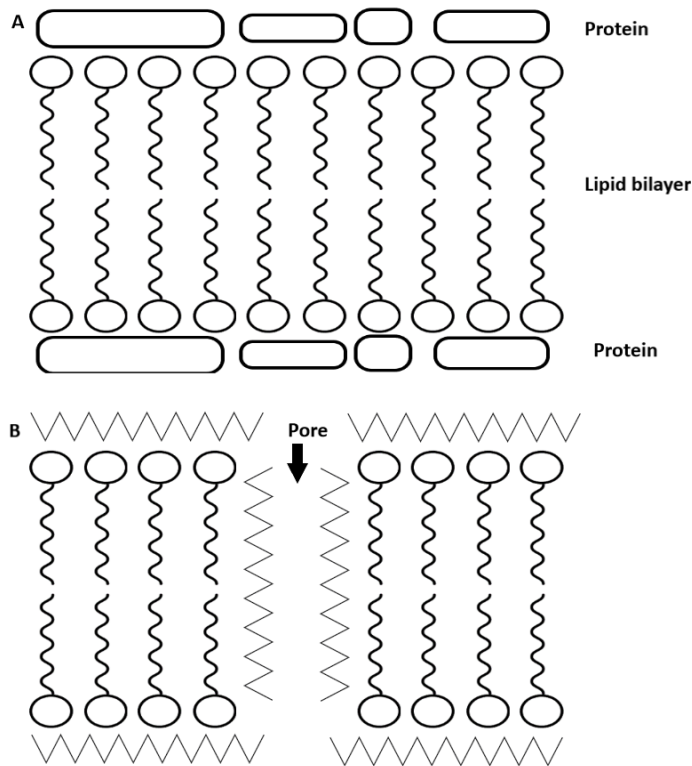
Figure 2. The Pauci-Molecular model of the membrane lipid bilayer by Danielli and Davson. *(A)* Original proposal of the model. Water-soluble proteins (white rectangles) on either side of the membrane with the lipid bilayer (white circles and wavy lines) between the protein layers. *(B)* Modified model of the lipid bilayer containing pores lined by peptide (jagged lines). Based on Stillwell, 2016[7].

This model served as the basis for other modifications and models, though there were still several issues that needed to be addressed. The first and most important is the lack of integral membrane proteins in this model. The Pauci-Molecular model also assumes a static environment; we know now that the membrane is dynamic[2]. Danielli and Davson also failed to address that lipid molecules vary from organism to organism. Lastly, this sandwich model of protein-lipid bilayer-protein ignores favourable interactions of the polar head groups of the phospholipids with water.

Another popular membrane model proposed was the Lipoprotein Subunit model (Figure 3) by Andrew Benson[18] and David Green[19] (known for their work with thylakoids and mitochondria, respectively). In both thylakoid and mitochondrial membranes, the protein-lipid ratio is relatively high compared to the plasma membrane[20,21]. This allowed them to propose the following: a single, lipoprotein layer in which the hydrophobic, nonpolar amino acids of the proteins would form a sheet of laterally associating proteins, with other exposed nonpolar surfaces being covered by the hydrocarbon chains of the phospholipids. The phospholipids would insert themselves into the cavities of the associated protein layer, foregoing the need for a lipid bilayer.
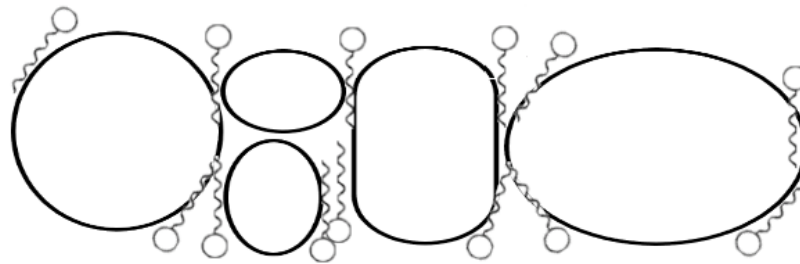
Figure 3. *The Lipoprotein Subunit model. Lipids are shown as small white circles and black wavy lines. Proteins are larger white circles or ovals. Based on Stillwell, 2016[7].*

Benson and Green were both highly regarded in their respective fields, which likely contributed to the attention and consideration this proposal received. This model has obvious shortcomings: a single mutation in any protein could completely disrupt the membrane, and it seemed to ignore growing evidence for membrane lipid bilayers.

1.1.4 The Fluid Mosaic Model

In 1972, Singer and Nicholson described the Fluid Mosaic model of membranes, which is presently the accepted model of the cell membrane. They defined the membrane as "*an oriented, two-dimensional viscous solution of amphipathic proteins (or lipoproteins) and lipids in instantaneous thermodynamic equilibrium*" (Figure 4)[2].
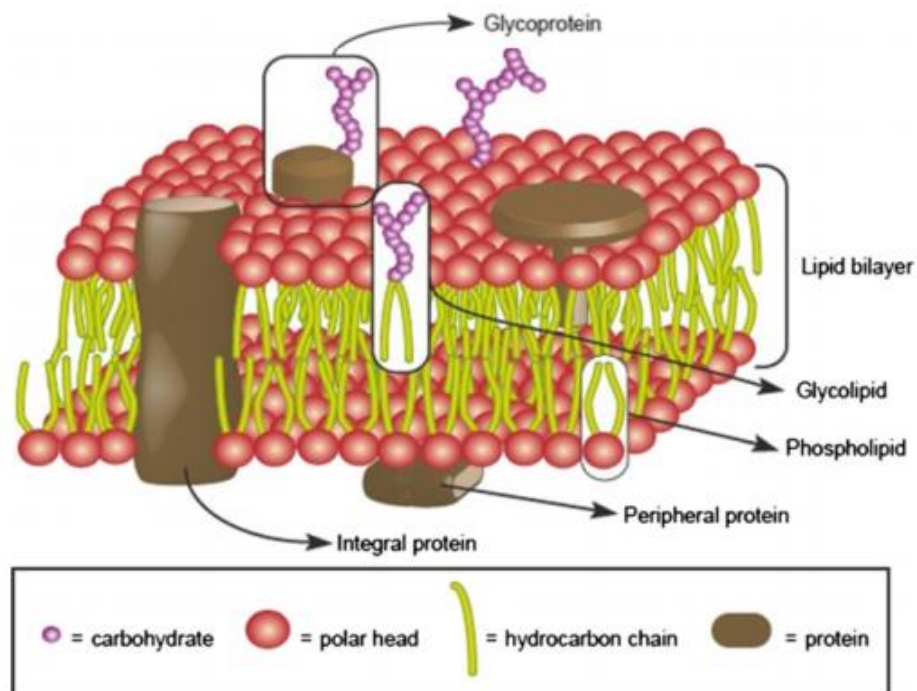
Figure 4. Schematic of the cell membrane as proposed by Singer and Nicholson. The membrane is a phospholipid bilayer with associated peripheral and embedded membrane proteins, and carbohydrates that may interact with the polar lipid head groups, exposed membrane protein surfaces, or water. Reproduced with permission from Lombard, 2014[22].

This model addresses the issues that arose with Danielli and Davson's Pauci-Molecular model. Singer and Nicholson proposed two broad classes of membrane proteins: peripheral membrane proteins, or those loosely associated with the membrane

surface, and integral membrane proteins, those that contain segments that are embedded and thus anchored within the lipid bilayer. On the outer membrane surface, carbohydrates are attached to proteins (glycoproteins) and lipids (glycolipids). Unlike the Pauci-Molecular model, the Fluid Mosaic model also addresses the heterogeneity of membranes by showing membrane proteins, associated components, and lipids asymmetrically distributed within the inner and outer leaflets of the lipid bilayer.

Singer and Nicholson also addressed the issue of dynamics in their earliest depictions of the Fluid Mosaic model, where they proposed free lateral diffusion of all membrane components. This was modified to accommodate the limited movement of proteins that interact with the stationary cytoskeleton within the cytoplasm. The following sections will detail the two main components of the cell membrane: the lipid bilayer and membrane proteins.

1.1.5 The Lipid Bilayer

The early work of surface film pioneers such as Lord Rayleigh and Irving Langmiur, in addition to early membrane studies by William Hewson and Charles Overton, contributed to Gorter and Grendel's proposal that the membrane was composed of lipids that form a bilayer. Singer and Nicholson's Fluid Mosaic model suggests that the lipid bilayer is composed of amphipathic molecules called phospholipids, which form the main structural component of the membrane. This section explores the forces that drive phospholipids to form lipid bilayers in the aqueous environment.

Water Molecules and Hydrogen Bonding

Water is central to life on Earth, and in our explorations of the solar system and universe, most researchers assume the presence of water is necessary for the presence extraterrestrial life[23]. Water is arguably the most important solvent in the world and the membrane lipid bilayer would not exist without the special properties of water, the characteristics of water molecules, and the interactions of water molecules with each other and other moieties that are dissolved in it.

Water molecules are polar and possess a partial negative charge about the oxygen. The partially positively charged hydrogens are oriented in a tetrahedral configuration with respect to the oxygen (Figure 5). The most important feature of water is its ability to

hydrogen bond (H-bond) with itself and other polar molecules. This hydrogen bonding

affinity is responsible for water's unique characteristics as a solvent and the hydrophobic

effect.

Averaged over time in liquid water, each molecule makes approximately 3.4 H-

bonds, the densest H-bonding system in any solvent[7]. Water molecules can make up to

four full H-bonds, which is seen in ice (Figure 5). While the covalent bond between the

oxygen and hydrogen atoms is stronger than the H-bonds to adjacent water molecules

(approximately 492 kJ/mol and 23.3 kJ/mol, respectively), the H-bonds are still nearly 5

times stronger than thermal collision fluctuation energy and van der Waals forces at

25°C[7]. At 40°C, near physiological temperature, water molecules still retain their extended

H-bond systems; it is only in the gaseous phase that the H-bonds disappear.
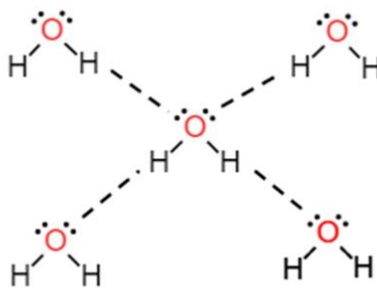
Figure 5. *The structure of water and the four hydrogen bonds that water molecules make in ice. In liquid water, water molecules make an average of 3.4 H-bonds. Dashed lines are hydrogen bonds, filled lines are covalent bonds. The image was made on ChemDoodle: Chemical Publishing Software version 10. Based on Stillwell, 2016.[7]*

Surface tension is a measure of the strength of the water surface film. Due to the extended H-bonding system in liquid water, the surface tension of water is 72.8 dyne·cm^{-1} at 25°C, second only to mercury[24]. In comparison, a weaker H-bonding system such as ethanol has a surface tension of only 22 dyne·cm^{-1}. While water forms approximately 3.4 H-bonds with its neighbours in bulk liquid, molecules positioned at the air-water or oil-water interface are only able to form two H-bonds (Figure 6). To maximize the number of H-bonds in the system and increase favourable interactions, water molecules at the interface are pulled by cohesive forces into the bulk solution. This minimizes the surface area of water at the interface and reduces the number of water molecules that form only two H-bonds. This force is so strong that, in the absence of gravitational forces, raindrops would be spherical[25].

The Hydrophobic Effect

The hydrophobic effect is a major driving force in biology. It contributes to driving water-soluble proteins to fold in solution, to the assembly of the membrane phospholipid bilayer, and the formation of the DNA double helix. In his book, The Hydrophobic Effect: Formation of Micelles and Biological Membranes[26], Charles Tanford discusses in detail the aversion of hydrophobic moieties to water, or the hydrophobic effect, *via* thermodynamic experiments and principles. As outlined earlier, water molecules can make approximately 3.4 H-bonds with adjacent water molecules in the bulk liquid; water molecules that line the surface at the air-water or oil-water interface are only able to H-

bond to two other water molecules, decreasing the thermodynamic stability of the H-bonding system.

If a highly nonpolar liquid such as hexane is mixed with water, two of the four sites for water H-bonds will instead interact less favourably with a hexane molecule (Figure 6). Molecular dynamics simulations have shown that, in the absence of gravity, hexanes dissolved in water form a perfect sphere, minimizing the surface area of the hexane-water interface[25]. This result is somewhat surprising, as it suggests that the interaction of the water-water H-bonds is stronger than the entropy of dispersing the hexane molecules within the water. This also suggests that the H-bonding system in the bulk water is stronger than the weak van der Waals forces between the hydrocarbon and water molecules. It is therefore more energetically favourable for hexane molecules to remain in an entropically unfavourable spherical conglomeration rather than to disperse, as this maximizes the number of water molecules forming 3.4 H-bonds and minimizes the number of ordered water molecules. For a comprehensive review on the hydrophobic effect, see Chandler, 2005[27].
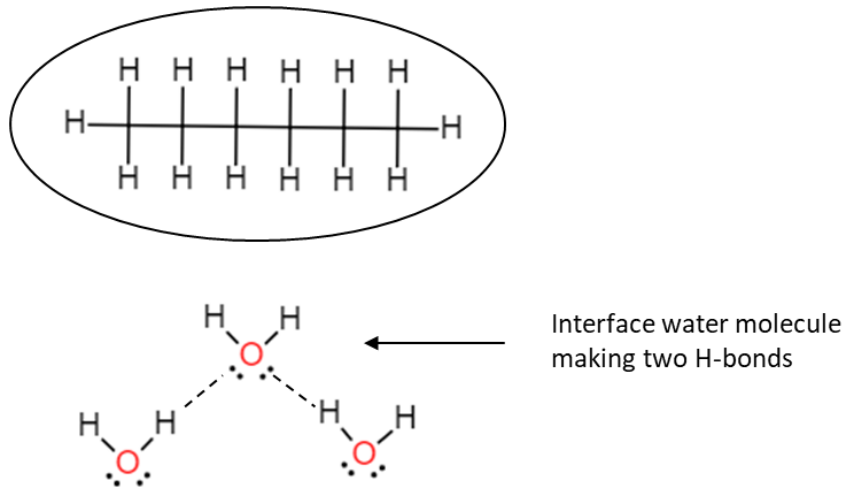
Figure 6. Hexane molecules interacting with liquid water molecules causes loss of two water-water H-bonds at the interface.

Phospholipids

Phospholipids (PLs) are amphipathic molecules that contain a glycerol molecule bound to two fatty acid hydrocarbon chains and a polar phosphate head group (Figure 7). At physiological pH, the phosphate is negatively charged and the PL head groups can be zwitterionic or anionic. In addition to providing a negative charge at the membrane surface, PLs are considered the scaffold of the membrane. Table 1 outlines the seven phospholipids found in mammals.
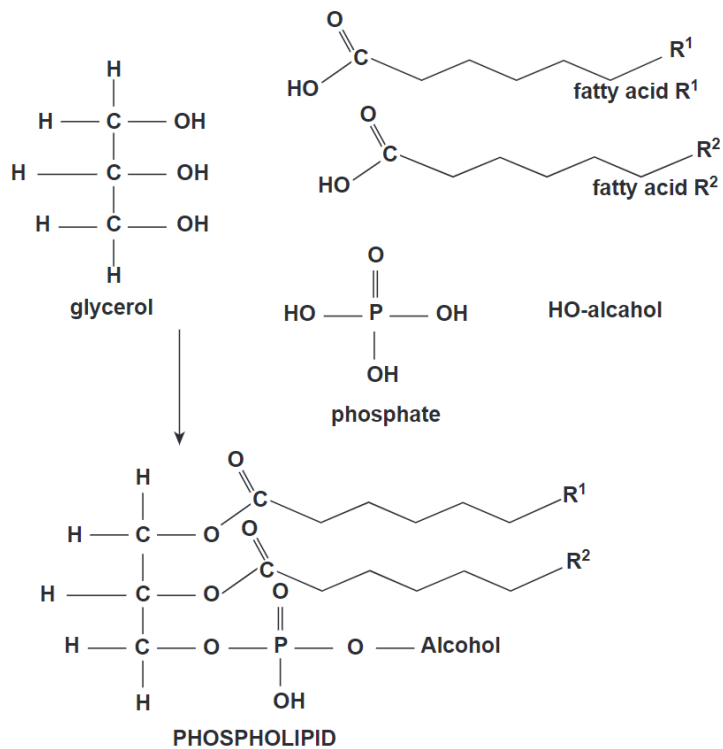
Figure 7. The structure of phospholipids. Top: components of a model phospholipid. Bottom: model phospholipid. Lipid tails are normally 14 to 24 carbons long, unbranched, and may be saturated or unsaturated, with zero to six cis nonconjugated double bonds. Reproduced with permission from Stillwell, 2016[7].

Table 1. *The seven phospholipids found in mammals.*

Phospholipid	Abbreviation	Head Group
Phosphatidic Acid	PA	None
Phosphatidylethanolamine*	PE	Ethanolamine
Phosphatidylcholine	PC	Choline
Phosphatidylserine	PS	Serine
Phosphatidylinositol	PI	Inositol
Phosphatidylglycerol*	PG	Glycerol
Cardiolipin*	CL	Phosphatidylglycerol

*The most common PLs in the *E. coli* inner membrane.

In *Escherichia coli* (*E. coli*), there are three main classes of phospholipids that make up the bacterial inner membrane (see Figure 8 for examples): phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and cardiolipin (CL). PE and PG head groups interact at the water-phospholipid interface *via* water bridging and direct H-bonding, and to a lesser degree at physiological concentrations, by ion bridging[28]. In molecular dynamics simulations (see below), a simplified model for the *E. coli* inner membrane is often used consisting of PE and PG.
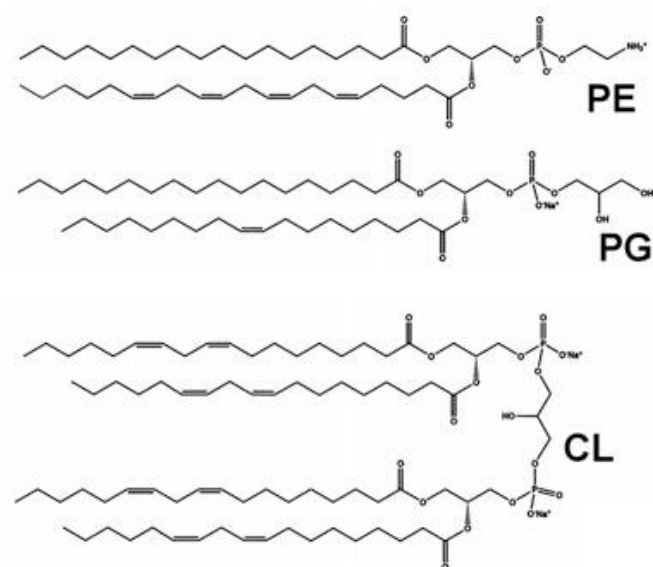
Figure 8. *The three major PLs in E. coli membranes. PE is phosphatidylethanolamine, PG is phosphatidylglycerol, and CL is cardiolipin. Reproduced with permission from Popot, 2018[29].*

Bacterial membranes are composed of 70-80% PE, where PE's function in the membrane is largely structural[30]. PE is considered zwitterionic but carries a slight net negative charge at physiological pH. Though PE preferentially H-bonds with PG, PE also forms H-bonds between other PE molecules, which creates strong, membrane-stabilizing interlipid contacts. Several studies probing the asymmetric distribution of phospholipids in the inner and outer layers of the bilayer exploit the chemically reactive head group of PE and show that certain phospholipids accumulate preferentially in the inner or outer leaflet of the bilayer[31]. Indeed, PE has been shown to be commonly found in the inner layer of the membrane[7].

PG is anionic and accounts for ~10-20% of the phospholipids in the *E. coli* inner membrane[29]. Studies suggest that enrichment of PG and other anionic fatty acids in the

cell membrane are required for protein translocation processes. SecA is a water-soluble

protein that is part of the Sec secretion group of proteins and requires the presence of

anionic lipids to successfully and efficiently translocate precursor proteins across the *E.*

coli inner membrane[29,32]. Anionic lipids like PG are also implicated in the membrane

protein folding process. Leader peptidase (Lep) is an elongation factor in prokaryotic cells

and eukaryotic mitochondria. It has a large periplasmic domain, two transmembrane

domains and a small cytoplasmic loop; it requires the Sec machinery for incorporation

into the membrane. Studies in which the negative charges of the anionic phospholipids

were shielded *in vivo* and in testing the system in vesicles that had low amounts of

anionic phospholipids all showed a decrease in efficiency of protein translocation across

the lipid bilayer[33].

Like PG, CL is anionic at physiological pH, but makes up less than 5% of the

bacterial cell inner membrane[34]. The structure of CL is unusual as it is composed of two

phosphatidic acid molecules bound to a glycerol, which results in two negative charges,

as well as four nonpolar tails. CL is found in larger amounts (approximately 20%) in

mitochondrial membranes. Studies on the function of CL in mitochondrial membranes

allude to raft-like microdomains (defined in the next section) in the membrane. These

microdomains have been implicated in the activation of apoptosis[35]. CL also plays a role

in the electron transport chain (ETC), stabilizing important protein components of the

ETC such as cytochrome c oxidase and NADH dehydrogenase[36]. First isolated in 1947[37],

the amount of CL has since been linked to a number of diseases, including Tangier's

disease[38], Alzheimer's disease[39], and Parkinson's disease[40].

Lipid Rafts and Raft-like Microdomains in the Bacterial Membrane

Eukaryotic cells have a diverse phospholipid profile compared to prokaryotes. The

advantages of having such a variety of phospholipids was studied as early as 1977[41], but it

was not until the late 1990s that the idea of lipid rafts gained traction in membrane

biology. The so-called lipid rafts have been observed in the outer leaflet of the plasma

membrane, where areas enriched in sphingolipids (Figure 9) and cholesterol group

together with membrane proteins and form short-range ordered domains. These domains

carry out important biological processes, which can range from highly specialized

functions such as signal transduction[42] to organization of the cytoskeleton[43]. The

presence of lipid rafts has been debated, despite a growing amount of evidence in the

literature. Methods to detect lipid rafts, such as resistance to solubilization in Triton X-

100[44] and sensitivity to cholesterol depletion, are indirect and cannot be clearly

interpreted. Lipid rafts have also been difficult to visualize in the cell, and when evidence

of visualization emerged, characteristics of the rafts varied widely, possibly due to the

range of structures and their biological roles[45–49].
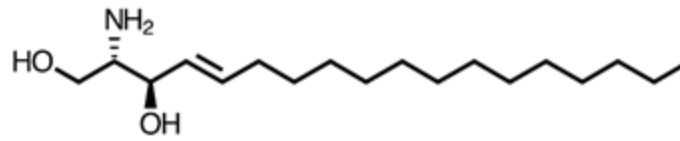
Figure 9. *The structure of sphingosine, a derivative of serine. The derivatization of the hydroxyl group at C-1 gives rise to the various sphingolipids found in the membrane. Reproduced with permission from Stillwell, 2016[7].*

There is also evidence of raft-like domains in bacteria. *E. coli* membranes have been shown to have pre-existing PE- and PG-enriched domains using a fluorescent label. PE-enriched domains have been observed at the septum of a dividing cell; FtsZ, which has an analogous role to tubulin in the cell division process, requires the presence of PE[50]. 10-nonyl acridine orange is a CL-specific dye that allows for the visualization of CL in the membrane. CL has also been shown to be enriched at polar and septal regions of the *E. coli* inner membrane, where CL introduces curvature of the membrane and is linked to high curvature sites in the cell division process[51]. Anionic PLs such as PE and CL introduce high negative curvature in the membrane and thus are found in areas such as the poles of the cell.

Assembly of the Cell Membrane

When membrane phospholipids are added to water, a decreased surface tension (approximately 7-15 dyne·cm^{-1}, much lower than 72.8 dyne·cm^{-1} in liquid water) is observed[52]. This effect is similar to the observed effect of surfactants and suggests that the

partitioning of phospholipids into the air-water interface, and by extension, at the membrane-water interface, is energetically favourable[53]. Similar to the hexane-water simulations, phospholipids in water segregate their lipid tails, adopting a tail-to-tail orientation away from water molecules to minimize the disruption of the H-bonding system in the bulk water[7]. While the hydrophobic effect is the largest contributor to the favourable formation of membrane lipid bilayers, other stabilizing forces include polar head group interactions with water and each other. van der Waals forces also play a role in the formation of the membrane through the interaction of the lipid tails[54]. The lipid bilayer self-assembles in solution and will spontaneously close upon itself to shield all exposed nonpolar moieties[55].

Bacterial cell membranes

Most bacteria can be divided into two groups based on their cell envelope: Gram negative and Gram positive. The terminology stems from the Gram staining test, developed by Hans Christian Gram in 1884[56]. Gram positive cells have a single lipid bilayer that envelopes the cytoplasm of the cell. These cells contain a thick peptidoglycan layer that reacts with the stain. Gram negative cells, such as *E. coli,* have outer and inner membranes which are separated by the periplasmic space. Their outer membrane is surrounded by a thinner peptidoglycan layer that does not retain the Gram stain as well. Glycerol facilitator (GF) is an integral membrane protein found in the inner membrane of *E. coli* and the protein of interest in this work.

1.1.6 Membrane Proteins

Membrane proteins and membrane-associated proteins make up the second largest component of biological membranes. Approximately 30% of protein-encoding genes in the human genome encode membrane proteins[57] that can be broadly classified based on their relationship with the membrane. Peripheral membrane proteins are proteins that can transiently associate with the membrane and other fixed moieties within the membrane. Integral membrane proteins are proteins that contain a segment that is embedded within the membrane bilayer.

Peripheral Membrane Proteins

Peripheral membrane proteins (PMPs) carry out their functions at the membrane surface and are localized there through electrostatic interactions or H-bonding with either the polar surfaces of transmembrane proteins or the head groups of the phospholipids. Alternatively, they may also interact *via* an anchor that links the PMP to the membrane[58]. They are responsible for a number of cellular processes, such as the well-studied PMP cytochrome c (cyt c), which plays a role in the mitochondrial electron transport chain[59], and myelin binding protein, which is important in the myelination of nerves in the central nervous system[60]. While many PMP functions have been identified, there is much still unknown about their interactions at the membrane interface due to

the transient nature of their interactions; trapping PMPs in their functionally active form is particularly difficult.

Integral Membrane Proteins

Integral membrane proteins (IMPs) are classified by their secondary structures. The majority of membrane proteins are in the α-helical conformation; investigations of the human genome suggest that as high as 30% of protein-encoding genes encode α-helical membrane proteins, and in prokaryotes, surveys suggest approximately 10-15% of the genome encodes α-helical membrane proteins[61]. Membrane proteins can also form β-pleated sheets to make a β-barrel, noticeably to a lesser degree: in prokaryotes, less than 3% of genes encode β-barrel membrane proteins[62].

β-barrel Integral Membrane Proteins

Gram-negative bacteria like *E. coli* and organelles such as mitochondria and chloroplasts in eukaryotic cells possess two envelopes: the outer membrane (OM) and inner membrane, which are separated by the periplasmic space. IMPs found in the OM are called outer membrane proteins (OMPs). OMPs function in a variety of ways, from maintaining membrane asymmetry[63] to substrate transport across the OM[64].

Despite their variety of functions, OMPs generally adopt a β-barrel conformation, though some OMPs span the OM with a single α-helix[65,66] and, recently, were found to form α-helical channels[67]. β-strand secondary structure conformations are extended relative to α-helices, containing approximately 2 residues per turn (Figure 10). Arranged side-by-side, β-strands can form β-sheets, with adjacent backbones of each strand H-bonding with their neighbour to form the sheet. β-sheets are also referred to as pleated, as the consecutive peptide planes regularly alternate their direction above and below the plane of the sheet. β-sheets may adopt parallel or anti-parallel conformations, though β-barrels almost always consist of an even number of antiparallel strands that twist and coil to allow H-bonding of the first and last strands, forming a single enclosed pore in the membrane (Figure 11)[68].
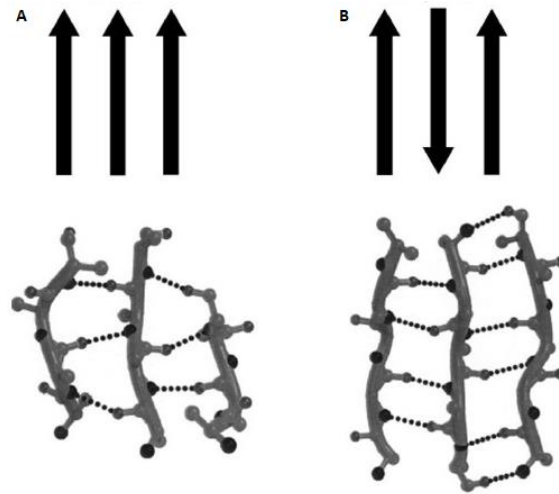
Figure 10. β-strand secondary structure of OMPs. *(A)* β-strands oriented in parallel; the tip of the black arrow denotes the C-terminus of the strand. Bottom depicts the same polypeptide in β-sheet parallel conformation; the backbone is depicted in light grey, side chains of the residues alternating from above and below the plane of the sheet are shown as grey and black spheres, and the dotted black lines denote the H-bonds made between adjacent β-strands to form the β-sheet. *(B)* β-strands oriented in the anti-parallel conformation. Bottom are three β-strands forming an anti-parallel β-sheet. Reproduced with permission from Stillwell, 2016[7].
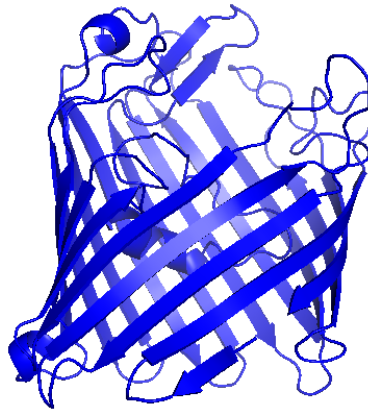
Figure 11. E. coli outer membrane protein OmpF. OmpF secondary structure shown as a cartoon. Arrows denote the β-strands in anti-parallel formation. (PDB ID: 2OMF)

α-helical Integral Membrane Proteins

Spectroscopic studies using circular dichroism (CD), infrared spectroscopy, and optical rotatory dispersion[69,70] on isolated membranes in the mid-1960s suggested that there were proteins with α-helical character in the plasma membrane. In 1975, Henderson and Unwin determined the approximately 7 Å low resolution structure of the *Halobacterium halobium* proton pump bacteriorhodopsin and demonstrated the α-helical character of transmembrane proteins[71]. A decade later, the atomic-resolution structure of the photosynthetic reaction centre of *Rhodopseunomonas viridis* would be solved by Deisenhofer, Michel, and Huber, confirming the presence of transmembrane α-helices[72].

α-helical transmembrane proteins are found in the inner or plasma membranes of prokaryotic cells and eukaryotic mitochondria and can consist of 1-14 α-helices[73]. Single pass or bitopic IMPs contain a single transmembrane helix and a large, soluble, extramembranous domain. Bitopic IMPs are the most numerous of all IMPs and account for approximately 45% of all α-helical IMPs[74]. They are found in all forms of life; in humans, in particular, they play major roles as signal and antigen receptors, among other important cell functions[75]. Bitopic IMP function was commonly attributed to the extramembranous domain but recent evidence has shown that the transmembrane segment may also play a role in the functionality of the protein[76].

Polytopic α-helical IMPs are those that comprise two or more membrane-spanning α-helices. Between 10-15% of IMPs are 7 transmembrane (TM) helical IMPs in eukaryotes; in prokaryotes, 6- and 12-TM IMPs are more common[73]. This motif is observed in a variety of IMPs with important biological roles, including signalling cascades[77], a variety of channels[78], and in olfactory[79] and visual[80] receptors. Of particular interest is the G-protein coupled receptor (GPCR) superfamily, receptors for important hormones like dopamine and epinephrine[77]. 7-TM IMPs are observed in all cells and across many families of α-helical IMPs; because of their major role in cell health, 7-TM IMPs are coined the "magnificent seven"[81].

In contrast to bitopic proteins, whose function is often tied to their large, extramembranous domains, 7-TM IMPs function through their transmembrane domain. They will often form pores or active sites within the membrane which allow the passive passage or active transport of the substrates that are usually excluded from the membrane, or contain binding sites for ligands of signalling cascades. These small helix bundles are functionally versatile and important in the maintenance of cell health.

1.1.7 Major Intrinsic Protein Family and Aquaporins

The Major Intrinsic Protein (MIP) superfamily are 7-TM IMPs. MIPs form channels within the membrane that allow water and some non-polar solutes to traverse the membrane. They are found highly conserved in pro- and eukaryotes, with over 100 members of the MIP superfamily identified; phylogenic analyses indicate that known MIPs originate from two divergent bacterial MIPs[82]. These broadly classify MIP family members depending on their sequence similarity to aquaporins (AQPs), which are water channels, or aquaglyceroporins, which conduct glycerol and other less polar substrates.

The first water channel discovered was AQP1 in 1992[83]. For many years, it was believed that water permeability was intrinsic to the lipid membrane, negating the need for (and therefore studies in search of) water-specific membrane protein channels. It became evident, however, that certain cells, such as red blood cells where AQP1 was discovered, transported water at rates higher than the water molecules could physically permeate across the lipid membrane.

AQPs are well-documented water and glycerol channels, allowing for the highly-selective and rapid passive diffusion of their substrate through the pore. The pores are so highly selective that moieties smaller than water, such as protons and hydroxide ions, are excluded from crossing the membrane. Other molecules such as urea, silicon, metalloids, and even gasses such as carbon dioxide and ammonia have been proposed as substrates for various members of this family[84].

1.1.8 Glycerol Facilitator

Glycerol facilitator (GF) is an aquaglyceroporin found in the inner membrane of *E. coli*. Its crystal structure was solved in 2000[85], which supplied valuable insight into the selectivity of the GF pore. GF remains one of the best studied AQPs to date and is the protein of interest in this work.

In general, AQP monomers share a conserved structure (Figure 12a, b): they are small helix bundles that consist of six TM helices and two half-helices. Each monomer contains two subdomains consisting of three TM helices and one half-helix each, related by a quasi two-fold axis of symmetry and are oriented 180° relative to each other[86] (see Figure 12). The subdomains appear to be the product of an ancient tandem gene duplication[87].

The half-helices are oriented with their N-termini towards each other within the membrane to form a seventh membrane-spanning pseudo-helix. Each half-helix contains a highly conserved Asn-Pro-Ala (NPA) motif on its N-terminus that plays an important role in the selectivity of the pore[88]. AQPs are categorized into three subgroups based on conserved amino acids that flank the NPA motifs: conventional AQPs or water channels, glycerol and small, polar molecule conducting channels, and unorthodox channels that do not exhibit high sequence conservation about the NPA motifs.
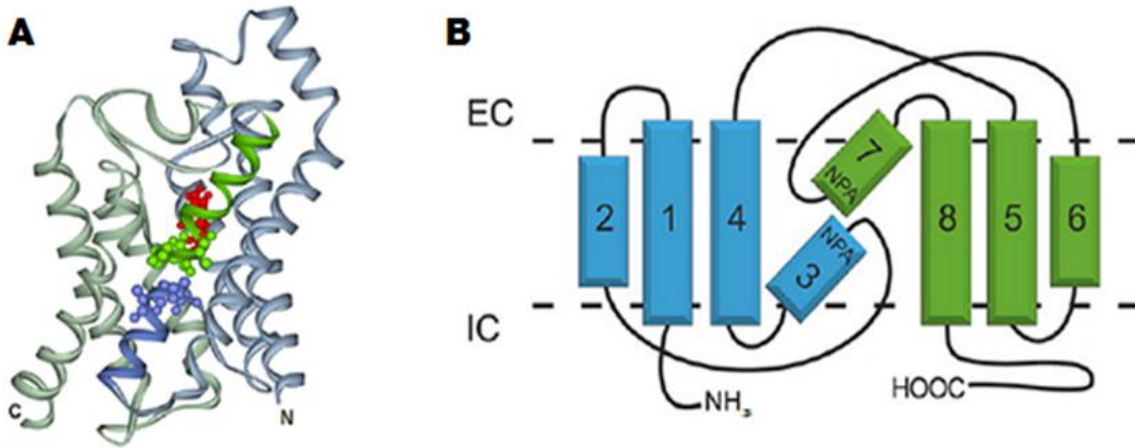
Figure 12. The secondary structure of the E. coli inner membrane protein glycerol facilitator monomer. (A) Ribbon diagram of the GF monomer. Light blue and grey helices are TM helices 1, 2, and 4, and 5, 6, and 8, respectively. Green and dark blue are half-helices 7 and 3. Residues of the NPA motifs are green spheres (half-helix 7) and blue spheres (half helix 3). Red spheres correspond to the aromatic/arginine (ar/R) selectivity filter (SF). Both the N- and C- termini are found in the cytoplasm. (B) The GF monomer contains two subdomains, related via 180°. Each domain contains three TM helices and a half-helix. Adapted with permission from Klein et al., 2014[86].

The atomic resolution crystal structure (2.2 Å) of GF illuminated the physical and chemical filters of the GF pore that account for the high selectivity of its substrate glycerol. The GF pore resembles the shape of an hourglass and is amphipathic in nature. There is a physical constriction in which the pore narrows to 3.4 Å from a 15 Å water-filled vestibule above the extracellular opening. This forms the narrowest part of the pore, which also contains the aromatic/arginine selectivity filter (ar/R SF). In addition to physically restricting the size of molecule that may traverse the pore, the ar/R SF forms an intricate H-bond system that only glycerol, and occasionally, water molecules may fulfill (Figure 13). Hydronium and hydroxide ions are unable to satisfy the H-bond requirements and are excluded. Proton passage *via* the Grotthus mechanism[89] is additionally blocked due to electrostatic repulsion. Positive electrostatic potential is concentrated at the center

of the pore by the ar/R SF and the two half helices, where the positive N-termini of their

helical macrodipoles meet at the center of the pore. In addition, one side of the channel is

lined with hydrophobic residues, which interact with the carbon backbone of glycerol,

contributing to the specificity of the pore to glycerol over water (Figure 14). Final

contributions to GF pore specificity are from desolvation penalties and configurational

barriers. Water and glycerol pass through the pore in single-file; solvent-substrate H-

bonds are replaced with substrate-channel H-bonds, which lowers the energy barrier and

allows quick (ns timescale) transport of substrates[90]. Hydronium and hydroxide ions are

destabilized from the surrounding aqueous environment, as the proper H-bonding

system is unfulfilled, rendering the passage of these ions unfavourable[91,92].
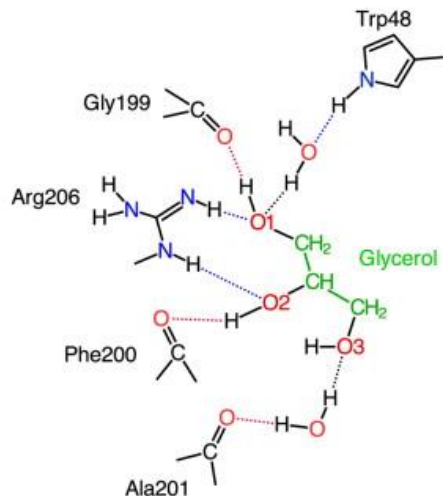
Figure 13. The ar/R SF and some of the residues that donate and accept hydrogen bonds to create an intricate system that contributes to the specificity of the GF pore. Adapted with permission from Klein et al., 2014[86].
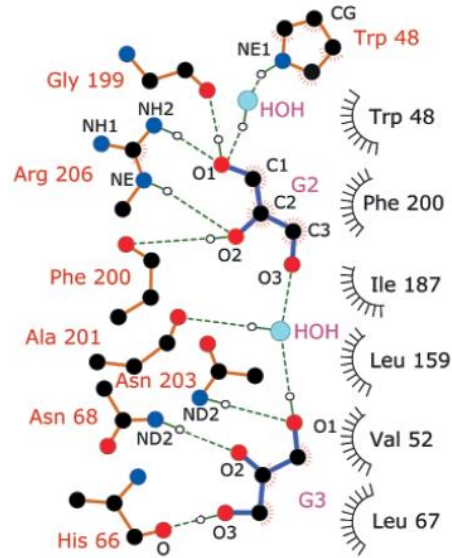
Figure 14. *The polar and non-polar residues that line the GF pore. Pictured are R206 and W48 of the ar/R SF, as well as two glycerol (G2 and G3) and two water molecules. Also depicted is the amphipathic nature of the GF pore. On the left are residues that interact through H-bonding interactions (labelled red), on the right in black are non-polar residues. H-bonds are shown as dotted lines. Red and black radial lines show hydrophobic interactions. Reproduced with permission from Fu et al., 2000[85].*

1.1.9 AQPs in Health and Disease

In vivo, GF and other AQPs form homotetramers that contain four independently

working pores (Figure 15)[85]. Water channels allow the passage of water through the pore

down an osmotic gradient and can function to maintain osmotic pressure within the cell.

In *E. coli*, glycerol uptake occurs in the absence of glucose, where glycerol is used by the

cell as an alternative carbon source[86]. In humans, 13 AQPs have been identified to date

(Table 2). Human AQPs are expressed throughout the body but are expressed and

function within a localized tissue, though the cell localization and functions of AQPs 11

and 12 are still unknown. The three best studied human aquaglyceroporins are AQP3,

AQP7, and AQP9. AQP10 is also an aquaglyceroporin and is expressed in the

gastrointestinal tract; there is limited information on its physiological function[93,94].
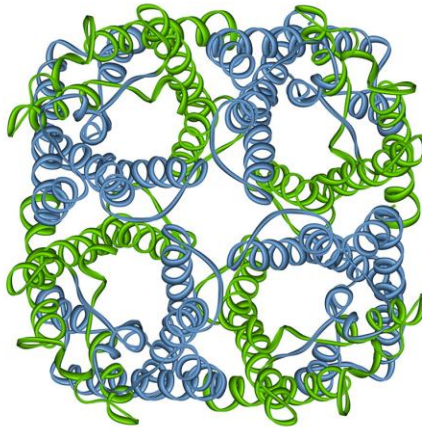
Figure 15. *The structure of the GF homotetramer. GF forms a homotetramer in vivo with four independently working pores. The monomers are represented as blue and green ribbons, where the helices are coloured to represent the two subdomains in the monomer. Reproduced with permission from Klein et al., 2014[86].*

Table 2. *A summary of human AQPs, sites of expression, and known functions.*

Aquaporin	Major sites of expression	Function
Aquaporin-0	Eye: lens fibre cells	Fluid balance within the lens
Aquaporin-1	Red blood cells	Osmotic protection
	Kidney: proximal tubules	Concentration of urine
	Eye: ciliary epithelium	Production of aqueous humor
	Brain: choroid plexus	Production of cerebrospinal fluid
	Lung: alveolar epithelial cells	Alveolar hydration state
Aquaporin-2	Kidney: collecting ducts	Mediates antidiuretic hormone activity
Aquaporin-3	Kidney: collecting ducts	Reabsorption of water into blood
	Trachea: epithelial cells	Secretion of water into trachea
Aquaporin-4	Kidney: collecting ducts	Reabsorption of water
	Brain: ependymal cells	Cerebrospinal fluid balance
	Brain: hypothalamus	Osmosensing function
	Lung: bronchial epithelium	Bronchial fluid secretion
Aquaporin-5	Salivary glands	Production of saliva
	Lacrimal glands	Production of tears
Aquaporin-6	Kidney	Unknown
Aquaporin-7	Fat cells	Transports glycerol out of adipocytes
	Testis and sperm	Unknown
Aquaporin-8	Testis, pancreas, liver and others	Unknown
Aquaporin-9	Leukocytes	Unknown
Aquaporin-10	Gastrointestinal tract	Unknown

With permission from Verkman *et al.*, 2014[95].

AQP3 is expressed in the cornea, colon, and in the epidermis. AQP3-mediated glycerol transport is required for maintaining skin hydration; AQP3-deficient mice exhibit reduced skin elasticity and hydration[96]. Interestingly, AQP3 has also been implicated in cell proliferation, where AQP3-deficient mice manifest with slow skin and cornea wound healing and colon cell regeneration. AQP3-*null* mice display resistance to skin tumour formation, prompting proposals that AQP3-mediated glycerol transport is key to the mechanism of cell proliferation in these tissues[97]. Exploring AQP3 inhibition may yield insight into potential prevention or treatment of certain tumours.

AQP7 is expressed in the plasma membrane of adipocytes (fat storage cells) and has been connected to obesity. AQP7 is important in the facilitation of glycerol out of the cell. AQP7-*null* mice show increased amounts of fat mass and adipocyte enlargement due to the accumulation of glycerol and triglycerides, an effect that is more pronounced as the mice age[98]. Some researchers suggest that, by regulating AQP7, fat mass can be altered by modulating adipocyte size and glycerol storage[99].

AQP9 is expressed in hepatic (liver) cells and is considered solely responsible for glycerol uptake into these cells[100]. AQP7 and AQP9 were also shown to allow arsenic trioxide to cross their pores, which may be the mechanism for arsenic uptake and poisoning in the liver[101]. Some studies also suggest coordination between the glycerol channels in adipocyte (AQP7) and liver (AQP9) cells that may play a role in glycerol and glucose metabolism[102].

There are several known AQP-related human disease states, or aquaporinopathies. Humans lacking functional AQP1, which is expressed in the kidneys, show a decreased ability to concentrate urine[103]. AQP0 is expressed in the lens fibre and is responsible for water permeation in the eye lens and cell-cell adhesion[104]. Subjects with AQP0 mutations develop congenital cataracts as a result of decreased cell-cell adhesion rather than impaired water permeation[105,106]. Nephrogenic diabetes insipidus (NDI) is a hereditary disease caused by loss-of-function mutations in AQP2[99,107,108]. NDI patients suffer polyuria (large amounts of dilute urine) and polydipsia (abnormal great thirst) leading to dehydration and cognitive defects.

A notable aquaporinopathy is the autoimmune disease neuromyelitis optica (NMO), a rare form of multiple sclerosis that affects the spinal cord and optic nerve[109]. An NMO diagnosis has a poor prognosis, as the disease causes blindness, paralysis, and death. The presence of autoantibody AQP4-IgG is the cause of NMO in most cases. This antibody binds to epitopes of AQP4 in the central nervous system, causing an inflammatory response that can result in demyelination and neuron death. NMO can be treated with aggressive immunosuppression, but therapeutic agents targeting AQP4, AQP4-IgG, and mediators of the inflammatory response are being explored[110] and clinical trials have been conducted with promising results[111].

The previous section highlighted the variety of roles of AQPs in the body and several aquaporinopathies, such as NMO, a severe aquaporinopathy for which AQP4 is a direct drug target. Mutations in and misassembly of IMPs has been implicated in and shown to be the direct cause of a variety of disease states[112,113]. Because of this, there is significant interest in studying IMP structures, the mechanisms through which they function, and elucidating information about their structures that can be exploited for drug delivery and design.

Approximately 60% of drug targets are IMPs and about 50% of small molecule drugs target IMPs[114]. G-protein coupled receptors (GPCRs) account for about a third of small-molecule drug targets[115]; they are attractive targets because of their role in signal transduction, activation, and inhibition. Many efforts over the last decade to find new drug targets and develop new classes of drugs have been dedicated to targeting the water-filled vestibules of cell-surface IMPs[116,117]. Another avenue of exploration is targeting TM domains, which are not only structurally important for the IMP but also make functionally important interactions with other IMPS and the membrane itself[118]. Drugs targeting TM domains endeavour to disrupt both TM domain-TM domain and TM domain-membrane interactions and conformational changes.

1.2 Protein Dynamics

Section 1.1.8 outlined in detail the information gleaned from solving the structure of GF and illustrated the relationship between the protein's structure and its function. While knowing the structure of a protein can yield valuable information, one must also remember that proteins are not static structures, as they are often depicted, and do not exist in static environments. Protein dynamics relates functionally important conformational changes over different timescales. Probing these motions at the atomic level, as well as their amplitude and direction, is critical to fully understanding the mechanisms by which proteins carry out their functions.

Solution NMR spectroscopy has been useful in discovering the dynamic properties of certain proteins, such as calmodulin (CaM). CaM is a small, cytoplasmic protein responsible for calcium sensing and signal transduction. In conjunction with X-ray diffraction and molecular dynamics simulations, CaM was found to adopt at least 40 different conformations in response to the calcium concentration in the cell. Different binding partners appear to bind only to a limited subset of these conformations of CaM. This allows a single protein to translate a complex, time-varying signal to multiple possible outcomes, and is subtle compared to a binary on-off signaling system[119,120]. The CaM molecule is only one example of the importance of understanding protein dynamics as they relate to protein function.

1.2.1 Protein Dynamics: Motions and Timescales

The motions of atoms in a protein can range from timescales on the order of femtoseconds (fs) to seconds (s) and longer[121]. Shorter timescales convey smaller motions, such as bond vibration on the fs timescale. Other relatively larger, local motions, including side chain rotations and loop motions occur from the picosecond (ps) (Tier 2, Figure 16) to nanosecond (ns) (Tier 1, Figure 16) timescale[121].

Biologically relevant dynamics generally take place on the microsecond (μs) to millisecond (ms) timescale (Tier 0, Figure 16)[121]. These are rarer collective motions that may recruit several domains to carry out functions such as enzyme catalysis, signal transduction, and other protein-protein interactions.

1.2.2 The Free Energy Landscape

The free energy landscape (FEL) model is used to describe protein folding and the

relationship between all possible conformations of a protein. The lowest energy

conformation, or ground state, is the basin of the FEL. For complex systems such as

proteins, a number of thermodynamic conformations at the basin are sampled over time;

exploration of the different conformations is also dependent on environmental factors

such as temperature or pH[122,123]. Some proteins can slowly interconvert between several

low energy states.
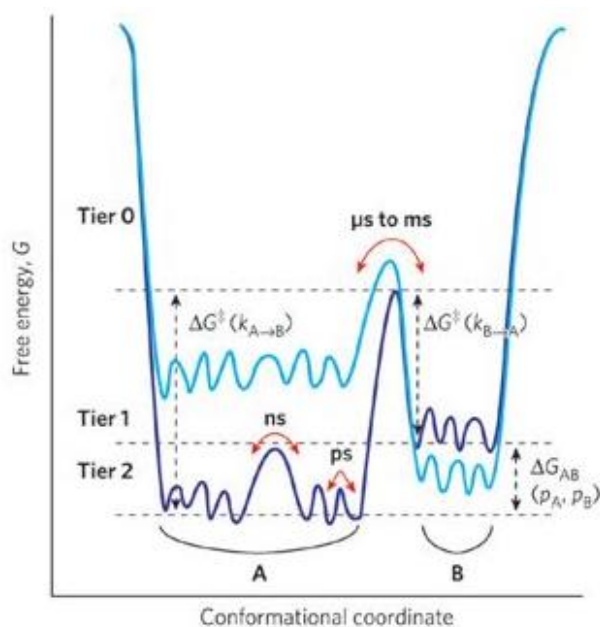
Figure 16. *The free energy landscape (FEL) of a protein. Tier 0 motions (light blue curve) are large, longer timescale (μs-ms), biologically relevant motions separated by large kinetic energy barriers. Tier 1 (ns) and Tier 2 (ps) (dark blue) motions are small, shorter timescale motions of proteins separated by small kinetic energy barriers and shallower minima, allowing for the continuous sampling of substates. A and B are different biologically relevant conformations of the protein. Reproduced with permission from Henzler-Wildman et al., 2007[121].*

The low energy basin of the FEL is made up of local minima separated by small kinetic energy barriers, and thus appears rugged. These minima represent the small conformational changes (Tier 1 or 2 dynamics) that the protein samples over time, which manifests as an averaged low energy structure that can be observed by biophysical methods. Larger kinetic energy barriers (Tier 0 dynamics) separate the ground state from excited state(s). FELs are multidimensional; both amplitude and timescale of the motions of all the atoms in the protein contribute to and affect the landscape[121]. It is also important to remember that an FEL is representative of the protein in a particular environment; it will change when conditions like temperature, pressure, or binding partner are altered (see A (blue) and B (cyan) in Figure 16).

1.2.3 Probing IMP Structure and Dynamics at the Atomic Level

IMPs comprise an estimated 23% of the human proteome; however, a severe bottleneck to determine IMP structure appears, as approximately 1% of unique structures deposited into the PDB are IMPs[124]. Deposited structures and dynamics studies of IMPs lag behind the number of water-soluble protein structures because of the difficulty in their purification in the folded state and complexity of their native lipid environment.

There are many commonly used methods in the literature to probe protein dynamics. This section will highlight the three methods that are used to probe protein structure and dynamics with atomic resolution: X-ray crystallography, molecular dynamics (MD) simulations, and nuclear magnetic resonance (NMR) spectroscopy[121].

X-ray crystallography

X-ray crystallography is a powerful tool to obtain the atomic-resolution structures of proteins and has virtually no particle size limitations. X-rays are scattered by protein crystals kept at cryogenic temperatures to minimize atomic motion. Electron density maps are calculated from the diffraction patterns of the scattered x-rays; from these maps, 3-dimensional structures are built and rigorously refined. For protein X-ray crystallographers, the biggest challenge is often producing well-diffracting crystals[125].

Even at temperatures near absolute zero, flexible regions of proteins can still be indirectly detected through X-ray diffraction. For example, highly dynamic regions of a protein may lack electron density or not be well fitted in electron density maps; similarly, residues that can assume different conformations may appear with partial occupancy in the structure. X-ray B-factors, which convey the extent to which the electron density is spread out in a crystal, are often interpreted as indicating the amplitude of motion of the atoms in the crystal, indicative of protein flexibility[122]. While static disorder can also increase the B-factor, a frequent cause of static disorder is the co-existence of multiple conformations in the crystal, also an indication of the inherent dynamics of the protein. In some studies, proteins have been captured in various intermediate states that help to elucidate their mechanism. For example, the nucleoside transporter from *Neisseria wadsworthii* was crystallised in five different conformations representing different steps during the nucleoside transport cycle. An elevator-like mechanism for transport was proposed based on the X-ray diffraction structures observed[126].

Molecular Dynamics (MD) Simulations

MD simulations of proteins are an increasingly powerful tool for probing the dynamics of IMPs. In MD simulations, Newton's equations of motion are solved for a system of atoms that interact with their neighbours, resulting in a simulation of the trajectories of each of the atoms in the system over the time of the calculation. For large molecules such as proteins, interaction forces such as covalent bonds, electrostatic

44

interactions, H-bonds, and van der Waal's interactions are modelled by classical

molecular force fields or empirical potentials as opposed to quantum mechanical force

fields[127]. This classical treatment cannot simulate chemical reactions, but conformational

changes and even protein folding events have been successfully simulated using this

approach[128].

MD simulations can be used for probing dynamics at high (atomistic) and low

(coarse-grain) resolution. However, IMPs and their native environments are complex and,

despite advances that have increased computational power[129], larger systems such as an

IMP surrounded by the lipid bilayer are presently challenging. Obtaining simulations at

the biologically relevant time scale (milliseconds and longer) is difficult due to the sheer

calculation load[130]. While MD simulations can simulate a protein in a bilayer, the

complexity of even the *E. coli* inner membrane, which contains more than 29 different

PLs[131], makes simulating the lipid membrane taxing. To increase the efficiency of

computation, some strategies employ a mixture of atomistic and coarse-grain

simulations[58,132,133].

NMR spectroscopy has been fruitful in determining the structures of many proteins that are difficult to crystallize, including proteins with flexible loops and termini and highly glycosylated proteins[134]. NMR spectroscopy is also the leading method to study protein dynamics at atomic resolution.

Early protein dynamics studies using solution NMR spectroscopy were limited to small proteins (approximately 50 kDa), though the structures of several larger detergent-solubilized[135,136] and lipid bilayer nanodiscs reconstituted[137–139] IMPs have been solved. Significant advances in solid-state NMR (ssNMR) spectroscopy have made NMR spectroscopy amenable to larger, slow-tumbling systems, such as detergent-solubilized, oligomeric IMPs[140]. NMR studies collect dynamic information on ensemble properties such as the nuclear Overhauser effect (NOEs), paramagnetic relaxation enhancement (PRE), three bond scalar couplings, trans-hydrogen bond scalar couplings, chemical shifts, and residual dipolar coupling[122,141,142]. These parameters as an ensemble yield insight into the minute details in the chemical environments of the atoms over various timescales, which can be used to inform on the three-dimensional structure and structural fluctuations that are essential for the biological function of a protein.

NMR spectroscopy is based on the inherent property of "spin" of an atomic nucleus. Atoms with spin have their own magnetic moment vector, and when placed in

an external magnetic field, will align with or against the field which are the low and high energy levels, respectively. NMR spectroscopy is the signal detected following the excitation of spins from the lower to the upper energy level. On a macroscopic scale, the NMR signal is proportional to the population difference between the high and low energy states. Predictably, the high energy state is less populated than the low energy state.

NMR probes structural changes by monitoring nuclear spin interactions; the shielding of nuclei by electrons gives information about torsion angles, hydrogen bonding and molecular packing, in addition to the chemical environment[143]. This means that chemical shifts are sensitive to slight changes in protein structure. The proximity of nuclei probed by protein NMR spectroscopy (^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P) can be measured by their dipolar coupling to approximately 0.1 Å within 5 Å and to 1 Å between 5 to 15 Å[143].

A novel approach is to use a combined room temperature X-ray diffraction/NMR spectroscopy method, which utilizes data from X-ray diffraction (B-factors) and NMR order parameters that yield information about the timescales of motion. Together with the qFit method, which creates a small ensemble of conformations where each residue is locally fitted within the electron density map, the library of conformers can be examined for possible dynamic processes[144]. Though this has only been applied to dihydrofolate reductase, a soluble enzyme, further method development could allow for IMP dynamics to be explored this way.

In solids, or for large, slowly tumbling molecules in liquids, NMR resonances are broadened because of orientation-dependent interactions such as dipole-dipole coupling. Methods to narrow the resonances, such as magic angle spinning, can be applied to solids, solutions, and proteins embedded in lipid vesicles, and are classified as solid-state NMR spectroscopy. Compared to solution NMR spectroscopy, ssNMR spectroscopy allows for IMPs to be studied in native or membrane-like environments, such as bicelles[145]. Though considerable progress has been made in the field, challenges with studying the dynamics of membrane proteins *via* ssNMR spectroscopy remain. The sensitivity of the method is an issue: the high and low nuclear energy levels are close together and the populations of the two states are very similar[146]. Because of this low sensitivity, high concentrations of isotope-labelled protein are required. This is often a challenge, as membrane proteins express at low levels and can be toxic when overexpressed in a cell. While this can be offset by increasing the volume of the cell culture expressing the protein, preparations requiring large amounts of isotope-labelled media are often economically unfeasible.

For structural studies, IMPs require solubilization by detergents, bicelles, or nanodiscs, resulting in increased hydrodynamic diameters, longer rotational correlation times and, consequently, increased solution NMR linewidths[147]. Dynamic exchange of detergent molecules may also contribute to line-broadening in solution NMR spectra.

As a result, most solution NMR studies of membrane proteins have been limited to small transmembrane helices dissolved in detergent micelles[148] or more recently in nanodiscs[149]. GF has a tetramer M_r of 130 kDa; even without any associated detergent, the GF tetramer would require TROSY-type experiments[150] and deuteration for the application of solution NMR methods although, in principle, highly flexible portions of the protein might be observable using more traditional approaches to protein NMR analysis.

In the past 10 years, significant progress has been made in solid-state NMR spectroscopy that is permitting its application to IMPs solubilized with a variety of agents, even including native cell membranes[143,151,152]. These advances are permitting high-resolution structural information to be obtained from a variety of non-crystalline biological and non-biological materials and, for example, led to the first high-resolution structures of amyloid fibrils[153]. Three major advances in solid-state NMR spectroscopy have led to the current state of the method. Improved magnet technology has led to high-field magnets ranging from 21.1 – 25.9 Tesla (900 MHz – 1.1 GHz proton frequency) that have greatly increased the sensitivity and resolution of both solid- and liquid-NMR spectroscopy and reduced the amount of isotope-labelled protein needed for measurements[154].

The development of faster Magic Angle Spinning (MAS) rotors, up to 110 kHz, has permitted the detection of backbone and side-chain proton resonances in

multidimensional experiments[155]. Spinning the sample rotor at 54.7° at very high speeds simulates isotropic Brownian dynamics and reduces ^1H-^1H- dipolar coupling and orientation-dependent interactions and proton linewidths. Besides the structural information available in proton resonances, the high magnetogyric ratio of protons greatly increases the sensitivity of solid-state NMR compared to ^{13}C- and ^{15}N-detected NMR. It also appears that at very high spinning speeds, protein deuteration is no longer needed for line-narrowing[152]. One of the consequences of faster spinning rotors is the requirement of smaller rotors and a potential loss of sensitivity owing to the reduced rotor volumes. It turns out, however, that the sensitivity losses are partially compensated by the increased detection sensitivity of the smaller receiver coils. In addition, coherence lifetimes are significantly longer at higher MAS speeds, leading to improved sensitivity in multidimensional correlation experiments[155]. Other advances in coil design have improved radio-frequency (RF) homogeneity and reduced the RF-induced heating of hydrated protein samples[154].

Finally, a suite of multidimensional pulse sequences has been developed to take advantage of the aforementioned technical advances in solid-state NMR spectroscopy[152,156]. Because chemical shift dispersion in the proton dimension is narrow, the significantly greater dispersion of ^{13}C and ^{15}N resonances in proteins are of great value in improving the resolution of resonances in multidimensional spectra. The multidimensional solid-state experiments now permit rapid assignment of backbone and side-chain ^1H, ^{13}C, and ^{15}N resonances[155]. The power of solid-state NMR comes partly from

the sensitivity of chemical shifts to local structure including H-bonding and bond torsion angles. In addition, dipolar coupling can be used to measure interatomic distances to within ±0.1-Å for distances up to 5 Å and with a tolerance of ±1 Å for distances up to 15 Å[143]. Importantly, solid-state NMR pulse sequences have been developed to measure dynamics of proteins over timescales ranging from ps – ms[157]. Multidimensional rotating frame ($R_{1\rho}$) relaxation dispersion (RD) experiments have also been developed to measure conformational dynamics on the ms – μs timescales where much biological activity occurs[158].

A few examples will illustrate the power of solid-state NMR to shed light on the structures and functions of IMPs. Most importantly, solid-state NMR can be used to measure protonation states of polar residues in IMPs. This has already been used to discover the proton conduction and gating mechanism in the influenza M2 proton channel[159]. Key protonation states in the prokaryotic K^+ channel KcsA were also measured by ssNMR[160]. Chemical shift measurements by ssNMR in KcsA also revealed the conformational landscape of the protein and how it is affected by pH and potassium ion concentrations[143,160]. In addition, ssNMR can detect structural disorder at the atomic level in IMPs[161]; in the M2 proton channel, disordered regions have been shown to influence the conformational ensemble of the TM region as well as determine the protonation state of the channel gate[162].

One of the remaining challenges for the application of solid-state NMR to IMPs is in sample preparation. Solid-state NMR has been successfully applied to IMPs dissolved with the help of detergents, nanodiscs[163] and, importantly, lipid vesicles[164]. Reconstituting IMPs into lipid vesicles is usually done by exposing detergent-solubilized protein to lipid, removing the detergent by extensive dialysis in the presence of detergent-binding materials such as Biobeads or cyclodextrin, followed by concentration of the protein[165]. This process typically takes on the order of days to weeks but can be accelerated with the use of sucrose/cyclodextrin/lipid gradient centrifugation[166]. Inserting the very low volume samples into the small rotors (0.7 mm) can also be challenging, but it is now possible to sediment the protein directly into the rotor using specially-designed ultracentrifuge device kits (*e.g.* spiNpack by Giottobiotech)[167]. However, finding the correct lipid:protein ratio is still a trial-and-error process. It appears that the lipid:protein ratio subtly affects the static protein conformational heterogeneity that gives rise to line-broadening in spectra[155,165]. Optimization of lipid:protein ratios can be done by observing when resolved glycine backbone or tryptophan side-chain resonances appear in 2D ^{1}H-^{15}N correlation spectra.

Despite these challenges, numerous studies using ssNMR spectroscopy on IMPs have provided insight into their structure and dynamics. ssNMR spectroscopy has been used to probe smaller IMPs, like the 100-residue influenza virus M2 protein, a proton conducting channel whose gating and conductance mechanisms were shown to be tied to

dynamics within the single TM helix and contain intrinsically disordered extramembranous domains[159,161,168].

IMPs with larger TM domains and smaller extramembranous regions, such as members of the 7-TM helix family, have also been explored, though most studies have been dedicated to members of the GPCR or microbial rhodopsin families. The *Anabaena* sensory rhodopsin (ASR) is a 7-TM helix photoreceptor whose structure was solved *via* ssNMR spectroscopy[169]. Human aquaporin-1 (AQP1), the aquaporin responsible for water homeostasis in red blood cells, was expressed in mg quantities in yeast *Pichia Pastoris*. AQP1 was incorporated into proteoliposomes and good quality ssNMR spectra were obtained. It is likely that these approaches can be optimized and applied to other IMPs[170].

1.2.4 Glycerol Facilitator Dynamics

Though IMPs such as GF exist in a dynamic environment, GF and other pore-forming IMPs are considered relatively rigid; that is, their tertiary structure is thought to remain largely intact, even while carrying out their function. However, GF and similar IMPs must undergo some conformational change to allow for the passage of their substrates, as a rigid structure would likely result in substrates being semi-permanently bound to the pore.

Studies probing the dynamics of GF consist of short molecular dynamics simulations and hydrogen-deuterium exchange studies. de Groot *et al.* were able to simulate AQP1 and GF in a bilayer for 10 ns[171]. They showed that the rate of permeation of water molecules in the simulation was comparable to observed experimental rates. They also noted that water permeation was slower in GF than in AQP1, due to a constriction at the conserved NPA motifs in the absence of glycerol. Jensen *et al.* supported this by modeling GF in a bilayer and calculating the width of the pore in the presence and absence of glycerol[172]. MD simulations using the X-ray crystal structures of GF with and without glycerol in the pore (GF+g, GF-g, respectively) compared to a simulated GF model without glycerol in the pore were studied. When comparing the X-ray crystal structures of GF+g and GF-g, the width of the GF pore narrowed approximately 0.5 Å in the absence of glycerol. The simulated GF without glycerol in the pore showed a narrowing of 1 Å compared to the GF+g pore. These results suggested that the glycerol channel opens in response to interactions with glycerol.

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) has been used to probe the dynamics of membrane proteins in detergents[173–175] and other membrane mimetics such as bicelles[176] and nanodiscs[177,178]. HDX-MS monitors the exchange of backbone amide hydrogens with deuterium in solution[173]. After the protein is diluted into D_2O, exchange is quenched at different time points to measure conformational fluctuations on different timescales. Conformational changes relate to rates of exchange; regions exposed to solution and those that are poorly H-bonded will exchange rapidly compared to regions that are not accessible to the solvent or form strong H-bonds.

Pan *et al.* found that solvent-exposed regions of detergent-solubilized GF, such as the loops that protrude extra- and intracellularly, exchanged quickly and completely, suggesting that they are disordered[91]. TM helices 1 and 5, which are responsible, in part, for stabilizing monomer-monomer interactions, exhibited very little exchange and thus are fairly rigid. Interestingly, TM7, the half-helix closer to the periplasm, also exchanges quickly and exhibits HDX characteristics similar to the flexible loops. One of the conserved NPA motifs is found in the loop connecting it to the helix which suggests that TM7 dynamics may play a role in the diffusion of glycerol and water through the GF pore.

In the following sections, I review the methods applied in this thesis to the study of *E. coli* glycerol facilitator.

1.3   Light Scattering Techniques

1.3.1 Dynamic Light Scattering

Dynamic light scattering (DLS) measures the time-dependent fluctuations in light scattering intensity of particles in solution[179]. Fluctuations in scattering intensity are caused by random Brownian motion of particles in solution; Brownian motion is affected by the size and shape of the particles, as well as the temperature and viscosity of the solution. By measuring the scattering fluctuations of a laser at some angle to the incident light, the translational diffusion coefficient (D_T) can be obtained and information about the size distribution and the size of the particles is gathered. However, proteins and other macromolecules in aqueous solutions are solvated and are not always spherical[180]. By monitoring changes in the intensity of scattered light, DLS provides a mechanism by which the hydrodynamic diameter, in this case, of the protein and associated water, lipid, or detergent molecules, can be calculated[179]. The measured time-dependent changes in light scattering are correlated against short decay intervals (τ) and the intensity autocorrelation function can be obtained for mono-disperse samples (Equation 1)[181]. The translational diffusion coefficient (D_T) and particle radius (R_h) for an ideal sphere are related by the Stokes-Einstein equation (Equation 2); this equation can be used to calculate the particle's hydrodynamic radius. Figure 17 shows a schematic of the typical DLS instrumentation.

Equation 1. $G(\tau) = 1 + be^{-2D_T q^2 \tau}$

where b is a constant dependent on the instrumentation, D_T is the translational diffusion coefficient, and q is the scattering vector.

Equation 2. $R_h = \dfrac{kT}{3\pi\eta D_T}$

where R_h is the hydrodynamic diameter (m), k is Boltzmann's constant $(m^2 \cdot kg/K \cdot s^2)$, T is the absolute temperature (K), η is the viscosity (Pa·s), and D_T is the translational diffusion coefficient (m/s^2). In DLS, the measured time-dependent changes in light scattering are fit to an autocorrelation function from which the diffusion coefficient is determined, and the Stokes-Einstein equation is used to calculate the particle hydrodynamic radius.
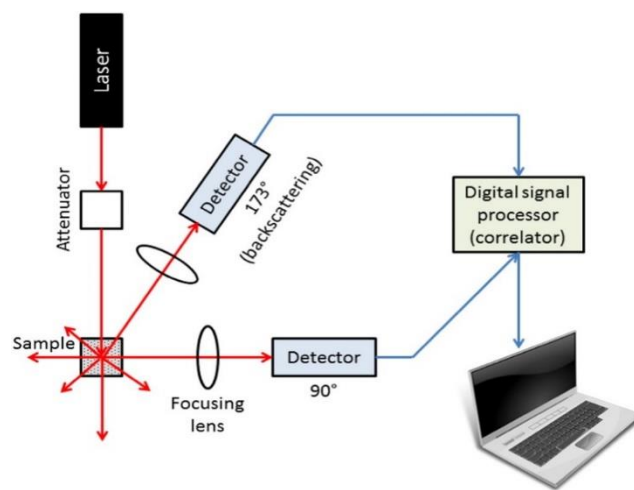
Figure 17. *Schematic of a typical DSL instrument. Reproduced with permission from Bhattacharjee, 2016*[181].

1.3.2 Size-Exclusion Chromatography Multi-Angle Light Scattering

Size-exclusion chromatography multi-angle light scattering (SEC-MALS) can be used to obtain the absolute molecular weight and hydrodynamic radius of molecules in solution[29,182]. SEC-MALS is a powerful technique that utilizes size-exclusion chromatography (SEC) as a fractionation step that prior to light scattering measurement. Performing SEC on a sample before measuring light scattering allows for the separation of different sized species, and with adequate peak separation, light scattering can be measured on individually fractionated populations of molecules.

SEC-MALS measures static light scattering (SLS), which measures the time-averaged light scattering intensity of a population of particles[29]. The advantage of measuring scattering over several angles as compared to a single angle is that the radius of gyration of the molecule can be calculated and this can be used to determine the molar mass of the particle scattering the light. In addition to SLS, SEC-MALS also employs a refractive index and UV absorbance detector for concentration determination. SLS yields information about the weight-averaged molecular mass of the particle, which makes SEC an integral step to assure that homogenous samples are being measured. In series, these three detectors elucidate information about hydrodynamic radius, molecular weight, and oligomerization, if present.

For SEC-MALS studies of IMPs, the membrane mimetic that the protein is solubilized in must be considered. Molecules used to solubilize IMPs such as detergents

will form a detergent shell around the protein. Because of this, solubilized IMPs can appear to have a larger M_r after analysis of the LS/RI/UV data. To address this issue, a $\left(\frac{dn}{dc}\right)$ refractive index calibration is applied, which relates changes in the RI to changes in protein concentration, with the latter being monitored by UV absorbance at 280 nm. $\left(\frac{dn}{dc}\right)$ calibrations are carried out in the sample buffer containing the IMP solubilization agent in the presence of standard proteins. From this, the light scattering caused by the detergent shell is separated from the light scattering of the protein; the protein's molecular mass can then be accurately determined using the following relationships[183].

Equation 3. $MM_{pp} = k_1 \dfrac{(LS)}{\left(\frac{dn}{dc}\right)_{app}(RI)}$

Equation 4. $\left(\dfrac{dn}{dc}\right)_{app} = k_2 A \dfrac{(RI)}{(UV)}$

where MM_{pp} = the molar mass of the protein, k_1 and k_2 = calibration constants, LS = the signal from the light scattering detector, RI = the signal from the refractive index detector, $\left(\frac{dn}{dc}\right)_{app}$ = the apparent refractive index increment of the protein-solubilizing agent complex, A = the extinction coefficient of the protein, and UV = the signal from the UV absorbance detector.

By combining equations 2 and 3, the molar mass of the protein surrounded by solubilization agent (MM_{pp}), as well as its oligomeric state (N) can be calculated.

Equation 5. $MM_{pp} = \frac{K(LS)(UV)}{A(RI)^2}$

Equation 6. $N = \frac{MM_{pp}}{MM}$

where K = a calibration constant, and MM = the molar mass calculated using the protein sequence. Knowledge of the protein sequence is required for calculating N and the extinction coefficient, A^{184} in equation 4.

From $\left(\frac{dn}{dc}\right)_{app}$ the quantity of solubilizing agent (δ) surrounding the protein can be approximated as follows:

Equation 7. $\delta = \frac{\left(\frac{dn}{dc}\right)_{app} - \left(\frac{dn}{dc}\right)_{pp}}{\left(\frac{dn}{dc}\right)_{detergent}}$

Where δ = the mass (g) of associated solubilization agents in the detergent shell per gram of protein, $\left(\frac{dn}{dc}\right)_{pp}$ = the specific refractive index of the proteins in the sample buffer (calculated from a calibration step), $\left(\frac{dn}{dc}\right)_{detergent}$ = the specific refractive index of the solubilizing agent associated with the protein.

The mass of the protein/detergent complex ($MM_{complex}$) can be determined with Equation 7.

Equation 8. $MM_{complex} = MM_{pp}(1 + \delta)$

1.4 Small Angle X-Ray Scattering

Similar to the SEC-MALS experiments described in Section 1.3.2, small angle X-ray scattering (SAXS) experiments are often done in-line with a pre-separation SEC step in order to separate and collect X-ray scattering data from differently sized species in a sample. Samples are irradiated with highly collimated x-rays, ideally from a synchrotron source, and the scattering intensity (I) is collected at multiple small angles[185].

For data presentation, the scattering intensity I(s) of the protein is usually plotted against the "momentum transfer" or scattering vector, (s) where $s = 4\pi\sin(\frac{\theta}{\lambda})$, 2θ is the scattering angle and λ is the x-ray wavelength[186]. The one-dimensional scattering curves contain low resolution information about the three-dimensional structure of a particle because the particles are undergoing rapid Brownian dynamics. The scattering curve is dependent on the distribution of electrons in the molecule averaged over all orientations but because there are no x-ray lenses to refocus the scattered x-rays, the scattered x-rays cannot be used to directly reconstruct an image of the protein. Instead, the scattering curves are fitted to low-resolution *ab initio* protein models generated by algorithms with the use of beads that simulate the scattering of the protein and the solvent. The low-resolution images reliably report on the size and shape of the molecule, yielding information such as molecular weight, maximum molecular dimension (D_{max}), and radius of gyration (R_g). The low-resolution SAXS models can also be fit with high resolution structures obtained *via* X-ray crystallography or NMR spectroscopy where available[187].

AQP0 has been modeled with the detergent halo using SAXS and these results have been

helpful for the analysis of several other membrane proteins[188,189].

1.5 Negative-Stain Transmission Electron Microscopy

Negative-stain electron microscopy (negative stain EM) is often used to quickly screen samples for their suitability for cryo-electron microscopy (cryo-EM) studies but is also valuable for examining protein homogeneity, oligomerization, and structure[190,191]. Protein samples are adsorbed onto an EM grid, enveloped by a layer of electron-dense heavy metal solution such as uranyl acetate, and dried. This results in a high contrast between the electron-dense background and the particle. The background scatters electrons more intensely and so particles appear lighter in the images as they have lower electron scattering power. Stain may also accumulate in exposed grooves within the particle, causing irregular contrast and the ability to visualize some particle details.

Subsequent to image acquisition, negative stain image classification can be conducted on the set of collected images using mathematical algorithms that compare the pixel intensities of the images[192]. Image classification can be used to compare the acquired images to projections of high-resolution structures. Image classification and class averaging also increases the signal-to-noise ratio of the images and permits finer details to be observed in the images. It also is the only way to assess the heterogeneity of samples that can arise from a number of sources, including conformational differences, multiple oligomers, and differences in orientation on the grid. In some cases, the class averages can be used to construct low resolution (approximately 20 Å) 3D models of a protein or complex, as long as the sample is monodisperse[193]. 3D model generation is possible when

particles orient randomly on the grids, providing a variety of views and negating the need

for additional imaging at different angles.

1.6 Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) is a biophysical technique that monitors the heat flux, or power, to a sample as a function of time and temperature[194]. As temperature is changed, the DSC apparatus measures the heat either absorbed or emitted from the sample which is based on the temperature difference between the sample and a reference cell.

There are two classes of DSC: heat-flux and power compensated. Heat-flux DSCs have a sample cell and an empty reference cell that are placed on a thermoelectric disc and surrounded by a furnace[195]. As the furnace heats, the heat will transfer through the thermoelectric disc at equal rates to the reference cell and the cell containing the sample[195]. However, the temperature difference between the cells due to the heat capacity of the sample, which can be measured by area thermocouples. The heat flow resulting from this can be measured by $q = \frac{\Delta T}{R}$ where q is the sample heat flow, T is the difference in temperature between the sample and the reference, and R is the resistance of the thermoelectric disc[196]. Power compensated DSCs contain sample and reference cells that are heated by separate furnaces. Both cells are kept at the same temperature, and differences in the thermal power required to maintain the same temperature in the cells are plotted as a function of temperature or time[197].

DSC is used regularly to measure the thermal stability of proteins and explore protein folding[198–200]. For IMP studies, DSC can be used to test the thermal stability of proteins of interest in different detergents and other solubilizing agents, where a higher thermal melting point is indicative the superior stabilization of one solubilizing agent over another[201,202]. DSC can be used to test the stability of membrane mimetics and their potential in IMP studies[203]. In particular, it was found that lipid nanodiscs (see Section 1.8.6) retained the thermodynamic characteristics of a lipid bilayer and may be more suitable for solubilizing IMPs than lipid vesicles. Here, the thermal melting points of GF in a variety of solubilizing agents were measured *via* power-compensated DSC.

1.7 Studier's Autoinduction Method

The following section describes the background and methods used to overexpress GF in *E. coli* cells.

In 2005, William Studier developed a set of defined media that can be used to optimize the screening process for proteins of interest[204]. Screening processes commonly require parallel cultures to indicate which target proteins, if any, express well and would be useful to scale up to multi-milligram amounts for structural studies such as X-ray crystallography and NMR spectroscopy. Parallel studies can be laborious when growth rates, and therefore induction times, vary. Studier discovered a cause of unintended induction by trace amounts of lactose observed in complex media and optimized defined media for parallel studies that would reliably induce the cells for overexpression of the target protein at high cell densities.

The *lac* operon is commonly used in the overexpression of recombinant proteins in *E. coli*[205]. In the absence of glucose, *E. coli* can metabolize lactose as a carbon source. When taken up into the cell, lactose is metabolized to allolactose which activates transcription of the *lac* operon, which contains the genes that transcribe the lactose-metabolizing machinery. To begin transcription, the lac promoter (*lacUV5*) recruits RNA polymerase for transcription of the *lacZ*, *Y*, and *A* genes. The *lac* promoter is used in the inducible T7 RNA polymerase (T7 pol) system in recombinant cells for overexpression of target proteins in *E. coli*. The *lacUV5* promoter regulates transcription of T7 pol in the *E.*

coli chromosome. T7 pol is an extremely active polymerase with high fidelity for its promoter. In the absence of lactose, the endogenous *lacI* repressor will bind to the *lacUV5* promoter upstream of the gene of a target protein inserted into an expression plasmid, which is downstream of the T7 pol promoter. Cell cultures are typically induced in the mid-log phase with a molecular mimic of allolactose, isopropyl β-D-1-thiogalactopyranoside (IPTG); after induction, cell growth slows and overexpression of the protein of interest commences[206].

Studier developed both non-inducing and auto-inducing media. Autoinduction circumvents the need to induce cells with IPTG in the log phase. Studier's medium contains glucose, lactose, and glycerol. Enough glucose is provided to the cells to grow to a high density, but, when the glucose runs out, lactose (present in small amounts) will enter the cell to be metabolized as a carbon source. This induces transcription, and expression of the target protein takes precedence over cell growth. Glycerol is provided as a carbon source to the cells to support growth after induction.

Studier's autoinduction media was shown to increase protein yields compared to IPTG-induced media. It is also useful in high-throughput testing for various proteins for structural studies, as well as the overexpression of generally poorly expressing membrane proteins. Studier also described other media designed to overexpress target proteins with selenomethionine or isotope labels incorporated in them for X-ray crystallography or NMR, spectroscopy respectively[207].

1.8 Detergents and Other Membrane Mimetics

The first sections of this chapter outlined the importance of the cell membrane and its function, as well as the two largest components that make up the membrane: phospholipids and IMPs. Section 1.1.5 touched upon the variable composition of membranes in prokaryotes and eukaryotes, the phospholipids that comprise them, and the membrane proteins whose functions are tied to their localization in the cell or organelle. Previous sections also outlined the importance of studying membrane protein structure and dynamics to glean insight into the diverse functions of IMPs responsible for maintaining cell health and their roles in a variety of disease states.

To study membrane proteins, they must be extracted from the membrane; because of their unique native lipid environment, IMPs have highly hydrophobic surfaces and therefore require a solubilizing agent to remain in solution[29]. To this end, amphipathic molecules such as detergents have been used to disrupt and extract IMPs from the lipid membrane and solubilize them for structural and dynamics studies. The following sections will give a brief overview of the types of agents used in this work and those commonly found in the literature to extract and solubilize IMPs.

1.8.1 Detergents

Detergents are amphipathic molecules with a hydrophilic head group and hydrophobic tail; their structure allows them to insert into and extract IMPs from the lipid membrane bilayer[29]. Historically, detergents have been used to extract and solubilize IMPs, and they remain the most common solubilizing agent used in IMP studies (Figure 18). Detergents can be broadly classified as non-polar, zwitterionic, and ionic, where non-polar detergents are considered gentle in that they disrupt protein-lipid and lipid-lipid interactions but do not disturb protein-protein interactions, essentially maintaining the native state of the protein[208].
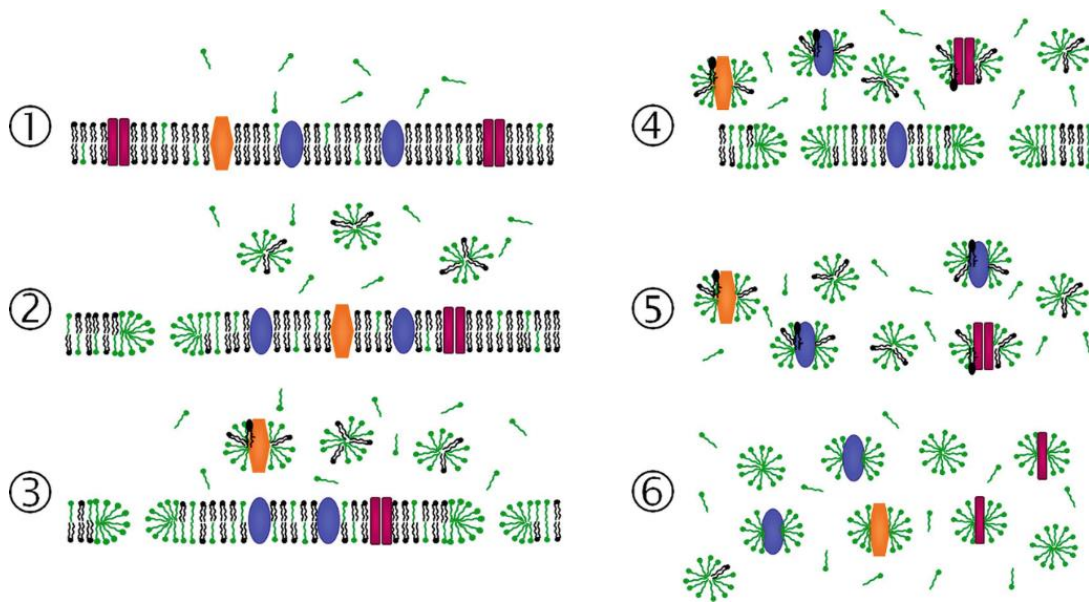
Figure 18. Detergent extraction and solubilization of IMPs from the lipid membrane. 1-4 show the effect of increasing the concentration of a detergent (green) to initiate the extraction of three different IMPs (red, blue, and orange) from the lipid membrane (black). Steps 5 and 6 show the extracted IMPs solubilized in detergent micelles. Reproduced with permission from Popot, 2018[29].

Detergent structures vary widely, but detergents can be characterized by a number of parameters. The critical micellar concentration (CMC) of a detergent is the concentration at which detergents spontaneously form micelles[7], provided that the sample is above the critical micellar temperature (CMT). Micelles are pseudo-spherical clusters of detergent monomers where the detergent head groups form a hydrophilic outer surface and sequester their hydrophobic tails within[29]. The CMC is the parameter used when selecting detergents for extraction and purification, though the number of molecules making up the detergent micelle, or aggregation number N, and micelle size (usually reported as M_w) are also taken into consideration[208]. The detergents and their properties used in this work are listed in Table 3.

Table 3. The detergents explored in this work and their properties.

Detergent	CMC (mM)	Aggregation Number	Micelle Size (kDa)
n-Dodecyl-β-D-Maltopyranoside (DDM)	0.17	79-149	72
n-Octyl-βD-Glucopyranoside (OG)	18-20	27-100	25
Sodium Dodecyl Sulfate (SDS)[a]	6-8	62	18
n-Dodecyl-N-N-Dimethylamine-N-Oxide (LDAO)	1-2	76	17-22
Lauryl Maltose Neopentyl Glycol (LMNG)	0.01	Unknown	91
Decyl Maltose Neopentyl Glycol (DMNG)	0.036	Unknown	Unknown
Dodecyltrimethylammonium Bromide (DTAB)[b]	16	48	Unknown

[1]Data taken from Anatrace website with the exception of [a] ThermoFisher Scientific and [b] Moulik, *et al.*, 1995[209].

It should be noted, however, that successfully extracting and solubilizing IMPs using detergents does not necessarily guarantee that the protein remains natively folded and maintains its function. In fact, it has been shown that lipid-protein contacts are important in a number of IMPs with varying roles in cells[210–213]. Detergent solubilization frequently strips away natively interacting lipids, which may also compromise the native structure based on changes in the lateral pressure and hydrophobic mismatch[208]. Though detergents have been used extensively in IMP research, other solubilizing agents that might provide an environment that better parallels the native lipid bilayer are being explored and developed.

1.8.2 Styrene Maleic Anhydride Lipid Particles

Styrene maleic anhydride lipid particles (SMALPs, Figure 19) are novel solubilization agents that are touted as a detergent-free method[214]. The SMA polymer is an amphipathic molecule that has been shown to insert into lysed cell membranes and encircle the lipids surrounding IMPs, solubilizing and stabilizing proteins in their native lipid environment[214], which eliminates the need for extraction from the membrane and solubilization *via* a detergent. SMALPs have been used to prepare membrane proteins to probe conformational dynamics *via* electron paramagnetic resonance (EPR) spectroscopy[215], structural assays *via* negative stain EM[216], solid-state NMR[217], and for sub-nm modeling of IMPs using single-particle cryo-EM[218]. Recently, three different proteins were extracted from the *E. coli* inner membrane and the phospholipids associated with each protein was analyzed[219]. The authors showed that each protein was removed from the membrane with a distinct phospholipid profile, whereas the SMALPs show no preference for any of the lipids. This supports the view that integral membrane proteins preferentially interact with specific lipids that may be important for maintaining them in their native conformation.
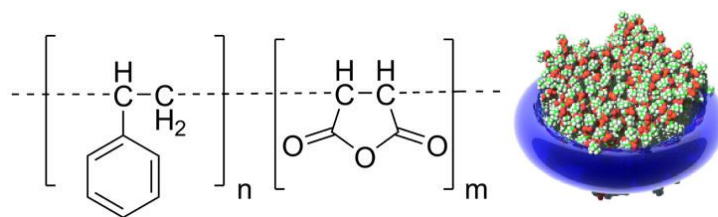
Figure 19. *Styrene maleic anhydride lipid particles (SMALPs) are used to solubilize IMPs in their native lipid environment.* **Left:** *the structure of the SMA co-polymer.* **Right:** *a SMALP, where the SMA co-polymer is shown as a blue band and the lipids in the center are shown as space filling models. Reproduced with permission from Jamshad et al., 2015[220].*

1.8.3 Liposomes

Liposomes are composed of hydrated lipids that can form one or more bilayers encircling a small volume of water (Figure 20)[221]. The diameters of these artificial spherical vesicles can be tuned to range from nanometers to several micrometers, depending on the preparation. They have been used to solubilize many membrane proteins in their active conformations, including members of the aquaporin family. *E. coli* aquaporin Z has even been inserted into liposomes that were subsequently immobilized on a polydopamine membrane to create a novel water-purification system[222]. Liposomes are a suitable bilayer mimetic, but are large and viscous, and are not amenable to solution NMR spectroscopy, X-ray diffraction and some other structural tools. Recently, solid-state NMR methods have been used to probe the structure and dynamics of liposome-embedded proteins and this is dealt with in further detail in Section 1.2.3.
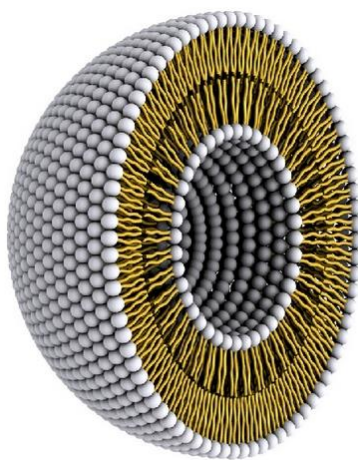
Figure 20. *Structure of a model liposome. White spheres are the hydrophilic head groups of the phospholipids and yellow lines are the hydrophobic tails. Adapted from Bitounis, et al., 2012[223].*

1.8.4 Bicelles

Bilayered discoidal mixed micelles, or bicelles, were first described in 1995 as a potential solubilization method for IMP structural studies *via* NMR spectroscopy[224]. As indicated in Figure 21, mixtures of long and short chain lipid, or detergents and lipids, do not always form the classical micelle structure but instead the long-chain lipids form small lipid bilayers surrounded by the detergent or short-chain lipid. The size and shape of the bicelles depends on the lipid:detergent ratio, or q ratio, in addition to other physical parameters such as temperature, hydration level, and ionic strength[29,225]. Isotropic bicelles can be used in liquid NMR studies[224], and magnetically oriented bicelles are amenable to solid-state NMR studies[145,226]. The Protein Data Bank contains several X-ray diffraction structures of membrane proteins crystallized from bicelles including bacteriorhodopsin[227], LeuT[228], and the Voltage-Dependent Anion Channel-2[229]. Bicelles provide a physiologically relevant environment compared to detergent micelles and may be superior in retaining the native structure of IMPs.
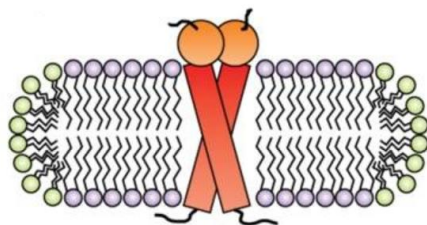
Figure 21. Representation of an IMP (red, orange) solubilized by a classical bicelle. Long-chain lipids (grey circles and black lines) form a bilayer surrounded by short-chain lipid or detergent (green circles and black lines). Adapted from Piai, et al., 2017[230].

1.8.5 Amphipathic Polymers

Amphipathic polymers, or amphipols, were shown to solubilize and reasonably stabilize four different IMPs in the absence of detergent[231]. They were initially used in a number of attempts to study membrane proteins by solution NMR spectroscopy[232], but have recently found success in the preparation of proteins for structure determination by cryo-electron microscopy[233]. Amphipol-solubilized enzymes have been shown to retain their enzymatic activity, though in some cases activity appears to be dampened compared to detergent-solubilized or bicelle-incorporated IMP[234,235].

Amphipols are relatively short, linear, amphipathic polymers that contain randomly placed hydrophobic and hydrophilic units (Figure 22)[236]. These polymers are highly soluble, negatively charged, and encircle IMPs to maximize interactions between the hydrophobic subunits of the polymer and the membrane-interacting (hydrophobic) surface of the IMP (Figure 23). The hydrophilic moieties in the polymer interact with the aqueous solution to keep the Amphipol-IMP complex soluble.

The most commonly used amphipol, A8-35, is a polyacrylate chain containing randomly-incorporated isopropylamine and octylamine side-chains. Each chain contains approximately 35 subunits. About 1/3 of the subunits are unreacted and bear a negatively charged carboxylate at pH 7. The chains self assemble in water to form micellar aggregates containing about 9 A8-35 chains[237].
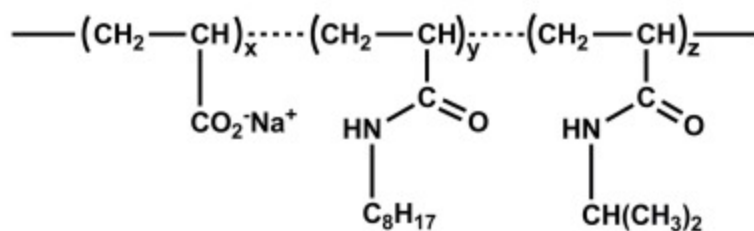
Figure 22. Structure of amphipol A8-35. The polymer is comprised of randomly distributed groups where the molar percentages of x = 35%, y = 25%, and z = 40%. Reproduced with permission from Le Bon, et al., 2018[238].
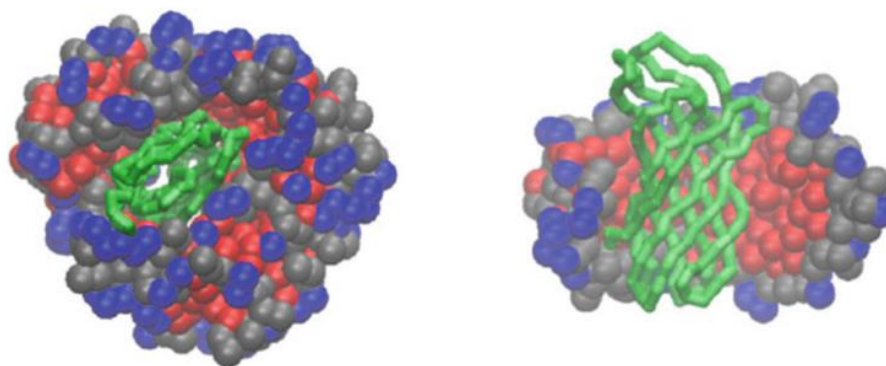
Figure 23. IMP OmpX solubilized by amphipol A8-35. Amphipols form a belt around the exposed hydrophobic surface area of IMPs to solubilize them. OmpX is shown in green; red spheres are the octyl chains of the amphipol, grey spheres are isopropyl chains, and blue spheres are the carboxylate moieties of A8-35. Reproduced with permission from Popot, 2018[29].

1.8.6 Lipid-Protein Nanodiscs

Lipid-protein nanodiscs (LPNs) are similar to bicelles in that they are comprised of a solubilized lipid bilayer but instead are encircled by two copies of a membrane scaffold protein (MSP), which are amphipathic α-helical proteins that shield the hydrophobic tails of the lipid molecules from the aqueous solution[239] (Figure 24). MSPs are derived from human serum apolipoprotein A1 (Apo-A1), the main constituent of High Density Lipoproteins (HDL) responsible for the transport of triglycerides, phospholipids and cholesterol in serum[239]. LPNs are a particularly intriguing membrane mimetic in that the size of the solubilized lipid bilayer can be adjusted by modifying the length of the MSP used[240].
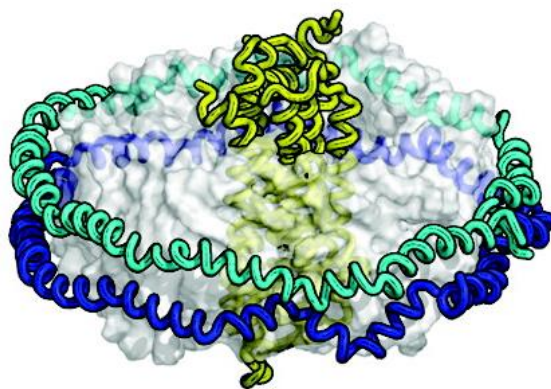
Figure 24. *Lipid-protein nanodisc-solubilized bacteriorhodopsin (lime green), a 7-TM IMP. The lipid bilayer is shown as light grey. Two MSP monomers are shown in ribbon format in cyan and dark blue. With permission from Nath, et al., 2007*[239].

The image of a nanodisc shown in Figure 24 is based on extensive characterization by DSC[203,241], SAXS[242], MD simulations[243], solid-state NMR spectroscopy[244,245], and mass spectrometry[246]. The structure and dynamics of membrane proteins solubilized in nanodiscs were probed by hydrogen-deuterium exchange mass spectrometry[247], cryo-EM[243], solution NMR spectroscopy[248,249], SAXS[250], and solid-state NMR spectroscopy[251]. Although bacteriorhodopsin could be crystallized directly from nanodiscs for structure determination[252], X-ray diffraction has not been widely applied to the structure determination of membrane proteins solubilized with the use of lipid nanodiscs. One limitation of this technology is that the sizes of the nanodiscs range from 7 – 13 nm; large or oligomeric proteins may not be amenable to solubilization in these materials.

1.8.7 Random Heteropolymers

In 2018, Panganiban *et al.* reported a novel polymer that was used in place of detergent to promote the folding of AQP1 during cell-free protein synthesis[253]. The expression and correct folding of AQP1 was monitored by the fluorescence of green fluorescent protein (GFP), which was fused to AQP1. They found an increase in expression and stabilization of AQP1 when their novel polymers were added to the cell-free synthesis mixture.

To develop these polymers, Panganiban *et al.* analysed the surfaces of water-soluble proteins and determined the size distributions and distances between charged and hydrophobic patches. The polymers consist of four monomers with varying hydrophobic character; the monomers are incorporated into the polymer with a fixed distribution based on the average surface characteristics of water-soluble proteins. The four monomers are methyl methacrylate, oligo(ethylene glycol) methacrylate, 2-ethylhexyl methacrylate and 3-sulfopropyl methacrylate (Figure 25). Though the overall ratio of monomers in the polymer is known, the sequence in which they are added is random and varies from preparation to preparation; thus, the polymers are termed random heteropolymers (RHPs). RHPs were designed with the characteristics of intrinsically disordered proteins in mind: high flexibility within the chain to allow for changes in conformation to maximize favourable interactions between the polymer subunits and the hydrophobic and charged patches on the protein of interest's surface. It

was envisioned that these polymers would be able to solubilize water-soluble proteins in organic solvents and IMPs in water.
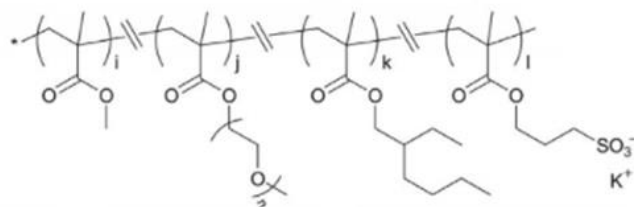
Figure 25. Structure of random heteropolymers (RHPs). The molecular structure of the four polymer subunits (i-l) of varying hydrophobicity in RHPs, which are added to the polymer with fixed distributions and random sequences. Reproduced with permission from Panganiban et al., 2018[253].

1.9 Purpose of the Research

Despite the many roles that IMPs play in the cell, there is still much to be learned about the mechanisms by which they function. The importance of elucidating atomic motions that enable their biological activity has been highlighted, as IMPs have been implicated in a number of human diseases. There are, however, reasons that only 1% of unique structures deposited into the PDB are IMPs and why so little is known about their dynamics.

First, most endogenous IMPs are expressed in low amounts. In some cases, overexpression in a host cell is toxic[254]. To probe structure and dynamics at the atomic level, common methods require mg amounts of protein. Methods such as NMR spectroscopy require expensive isotope labels which are added to the growth media. This can be cost prohibitive if mg quantities of target protein can only be obtained by growing multiple litres of cell culture.

Second, IMPs require a membrane-like environment. In the absence of a solubilizing agent such as a detergent, IMPs will irreversibly aggregate and precipitate due to the large exposed hydrophobic surface and the hydrophobic effect. In some cases, IMPs will even remain unfolded in an attempt to maximize the number of favourable interactions between hydrophilic residues[255]. Some studies suggest that although IMPs are properly folded when solubilized by detergent micelles, they may not retain their *in vivo* biological activity[225].

Third, and alluded to in the section above, is the tendency of IMPs to aggregate irreversibly in solution. While novel solubilizing agents and membrane mimetics have been developed, IMPs behave differently in the presence of different solubilizing agents; their solubility often requires empirical determination and is context-dependent[114]. Owing to the lack of a true mimetic of the native lipid membrane environment, the hydrophobic surface area can be improperly shielded, encouraging interaction of the exposed hydrophobic areas.

The overall purpose of this work was to probe the dynamics with atomic resolution of *E. coli* glycerol facilitator using liquid or solid-state NMR spectroscopy. The work presented will outline the methods used to optimize GF expression, purification, solubilization, and characterization of its oligomeric state for NMR spectroscopy studies as well as present the characterization of the GF octamer and preliminary 2D DARR NMR spectra of LMNG-solubilized GF. It is our hope that studying the dynamics of GF, a bacterial protein expressed in a system that allows for high yields, easy mutation, and requires no post-translational modifications, will lay the foundation for studying the structure and dynamics of analogous proteins and other IMPs *via* solid or liquid NMR spectroscopy.

Chapter 2: Materials and Methods

2.1 Materials

Dodecyl-β-D-maltoside (DDM), tetradecyl-β-D-maltoside (TDM), imidazole, and isopropyl β-D-1-thiogalactopyranoside (IPTG) were purchased from Gold Bio (St. Louis, MO). Octyl-β-D-glucopyranoside (OG), sodium dodecyl sulfate (SDS), 40% acrylamide/bisacrylamide, poly(ethylene) glycol 2000 (PEG 2000), Triton X-100, lysozyme, RNAse A, DNAse I, and α-lactose monohydrate were purchased from Sigma (St. Louis, MO). BL21(DE3) competent cells were purchased from New England Biolabs (Ipswich, MA). Top10 *E. coli* storage cells were purchased from Invitrogen (Waltham, MA). Isotopically labelled amino acids, glucose, ammonium chloride, glycerol, and deuterium oxide were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). 1000x metals solution was purchased from Teknova (Hollister, CA). Amphipol A8-35, PMAL C-8, lauryl maltose neopentyl glycol (LMNG), decyl maltose neopentyl glycol (DMNG), lysomyristoylphosphatidylcholine (LMPC), dodecyltrimethylammonium bromide (DTAB), and the 10:1 LMNG:CHS (cholesteryl hemisuccinate) mixture were purchased from Anatrace (Maumee, OH). Lauryldimethylamine-N-oxide (LDAO), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC), and *E. coli* polar lipid extract were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL). Ni-NTA agarose resin was purchased from Qiagen (Venlo, Netherlands). Sodium chloride, sodium azide, sodium phosphate, Amicon® Ultra-4

Centrifugal Filter Units, Amicon® Ultra-15 Centrifugal Filter Units, Coomassie Brilliant Blue R-250, and Coomassie Brilliant Blue G-250 were purchased from MilliporeSigma Canada (Oakville, ON). Bio-Beads SM2 were purchased from Bio-Rad (Hercules, CA). Phenylmethyl-sulfonyl fluoride and sodium cholate were purchased from ThermoFisher Scientific (Waltham, MA). The plasmids for the nanodisc MSPs (Membrane Scaffold Proteins) were a gift from Stephen Sligar (pMSP1E3D1: Addgene plasmid # 20066; http://n2t.net/addgene:20066; RRID:Addgene_20066, pMSP1D1: Addgene plasmid # 20061; http://n2t.net/addgene:20061; RRID:Addgene_20061, pMSP2N2: Addgene plasmid # 29520; http://n2t.net/addgene:29520; RRID:Addgene_29520.) C43 (Walker) cells were kindly provided by Fraser Ferens and Deb Court (Department of Microbiology, University of Manitoba). Random heteropolymers (RHP) were provided by Professor Ting Xu (University of California, Berkley).

2.2 Methods

2.2.1 Glycerol Facilitator Expression

Initial work was done with GF expressed in *E. coli* BL21(DE3) pLysS cells containing a pET28b(+) vector encoding the endogenous GF gene with an N-terminal His$_6$ tag, T7 epitope, and thrombin cleavage site (thrombin-plasmid)[256]. We subsequently discovered that a mutation had been introduced into the gene sequence, changing aspartic acid to valine at position 168 in the protein. A new codon-optimized gene was created (GenScript, Nanjing, China) to correct the mutation and add an N-terminal His$_6$ purification tag and TEV cleavage site to the codon optimized, mutation-free GF gene. The gene was ligated into a pET28b(+) vector (TEV-plasmid) that was transformed into *E. coli* BL21(DE3) cells. The M_r of this GF construct is 32 523 Da; its pI is 6.64 (both calculated by (https://web.expasy.org/compute_pi/)). No differences in behaviour between the mutated and native protein were observed in the course of this work.

Twenty-five mL overnight pre-cultures were grown in lysogeny broth (LB) or terrific broth (TB) at 37°C and shaking at 300 rpm. Studier's autoinduction system[207] was used to maximize GF expression in the final cultures. *E. coli* cultures were grown in Studier LS-5052 (for preparations of ^{15}N and ^{13}C uniformly labelled GF) or C-501750 (for optimized incorporation of ^{13}C for uniform labelling). All culture media were titrated to pH 7 before sterilization. Final cultures were inoculated with 5 mL of overnight pre-culture per litre of final culture and shaken at 300 rpm for 48 hours at 20 °C. Cells were

harvested by centrifugation at 5000 x g for 15 minutes at 4°C. Harvested cells were lysed

via 3 rounds of freeze-thaw cycles and incubated with lysozyme (150 mg per litre of

culture) for 30 minutes, and DNase (Deoxyribonuclease I) and RNase (Ribonuclease A) (1

mg each per litre of cells) for 10 minutes, with gentle stirring. The pH of the lysate was

monitored between freeze-thaw cycles and maintained at approximately pH 7. Membrane

pucks were prepared by layering the lysate over 70% w/v sucrose and centrifuging at 22

000 rpm for 60 minutes at 4°C in a Beckman SW 32 TI rotor in a Beckman Coulter

Optima LE-80K Ultracentrifuge. The supernatant was discarded, and the membrane

pucks were stored on top of the sucrose layer at -80 °C. Protein preparation is described

below.

To prepare soluble GF that is expressed in the cytoplasm, a pET28b(+) vector

containing the sequence to express a fusion protein encoding the codon-optimized *E.

coli* GF gene with an N-terminal TEV cleavage site, His_6 tag, and maltose binding protein

(MBP), and C-terminal apolipoprotein AI*3 (GF-Fus plasmid) (GenScript, Nanjing, China)

was transformed into BL21(DE3) cells. The GF-Fus-containing cells were pre-cultured and

final cultures were grown in LB at 37°C, shaking at 300 rpm. Final cultures were

inoculated with the overnight pre-culture, grown to OD approximately 0.6 and induced

with 1 mM IPTG, then allowed to grow overnight. Harvesting the cells and lysis by freeze-

thaw cycles were carried out as described above, and the lysate was centrifuged at 12 000

rpm for 1 hr at 18°C in a Thermo Scientific F14-14x50cy rotor and Thermo Scientific Sorvall

Lynx 6000 centrifuge or a Thermo Scientific Sorvall Legend XTR Centrifuge using a F15S-8x50cy rotor.

To explore the effect of the disordered N- and C-termini on the oligomeric state of GF, a pET28b(+) vector containing *E. coli* glycerol facilitator with a 19 amino acid truncation at the C-terminus (CT-) in addition to an N-terminal His$_6$ tag and TEV cleavage site (GenScript, Nanjing, China) was expressed in Bl21(C43) cells. Overexpression and membrane puck formation followed the same steps outlined for the Bl21(DE3) cells containing the TEV plasmid above.

2.2.2 GF Purification

GF Extraction and Purification

GF was extracted from frozen membrane pucks in 25 mM sodium phosphate buffer pH 7.4, containing 150 mM NaCl, 0.04% w/v sodium azide (Buffer A), and one of 30 mM DDM, 20 mM SDS, 20 mM DTAB, 20 mM OG, or 0.05% w/v LMNG and 0.005% w/v CHS for three hours at room temperature with gentle stirring. Insoluble matter was removed by centrifugation at 12 000 rpm for 1 hr at 18°C in a Thermo Scientific F14-14x50cy rotor and Thermo Scientific Sorvall Lynx 6000 centrifuge or a Thermo Scientific Sorvall Legend XTR Centrifuge using a F15S-8x50cy rotor. The supernatant containing detergent-solubilized GF was purified using immobilized metal (nickel) affinity chromatography (IMAC)[257]. For studies carried out on DDM-solubilized GF, the resin was first washed with 30 mL of Buffer A containing 0.5 mM DDM (Buffer A1). The resin was then washed with the same buffer containing 60 mM imidazole (Buffer A2) until the A_{280} was below 0.01. The protein was eluted from the IMAC resin using Buffer A1 containing 250 mM imidazole (Elution Buffer). 1.5 mL fractions were collected until the A_{280} was below 0.01.

To extract GF from the membranes using SMA 2K[214], the membrane pucks were first homogenized with a dounce homogenizer in buffer containing 50 mM phosphate buffer pH 8, 500 mM NaCl, and 10% v/v glycerol (SMA solubilization buffer) at a membrane concentration of 40 mg/mL. SMA 2K was then applied to the homogenized

membranes at 2.5% w/v, gently inverted to mix, and incubated for 2 hours at room temperature with gentle agitation. The mixture was centrifuged for 45 minutes at 110 000 x g at 4°C in a Beckman SW 32 TI rotor and Beckman Coulter Optima LE-80K Ultracentrifuge. The supernatant was applied to the IMAC column pre-equilibrated with SMA solubilization buffer. The column was washed with SMA solubilization buffer after the supernatant was applied, and a subsequent wash using the same buffer containing 60 mM imidazole was done. The SMALPs containing GF were eluted from the resin with SMA solubilization buffer containing 250 mM imidazole at pH 7.5 based on procedures outlined by Lee *et al.*[214].

SEC-MALS, negative stain EM, and SAXS studies were done on GF first extracted from the membrane with DDM as described above. Detergent exchange occurred on the IMAC column, where Buffers A1 and A2 and Elution Buffer containing 0.05% w/v LMNG instead of DDM were used. Buffer A1 was used to wash the resin and one resin volume of buffer was left in the column for one hour at room temperature to equilibrate and allow for detergent exchange. The same washing and elution protocols were followed after equilibration using Buffer A2 and Elution Buffer containing 0.05% w/v LMNG.

In screening other detergents and solubilization agents except for LMNG, detergent exchange occurred at the final elution step in the IMAC column. To obtain LMNG-solubilized GF, the Buffers A1 and A2 and Elution Buffer contained 0.05% w/v LMNG. Thirty mL of Buffer A1 containing LMNG was used to wash the column and

remove impurities. After that wash, another column volume of Buffer A1 containing LMNG was applied to the column and equilibrated for one hour at room temperature to allow for detergent exchange. LMNG-solubilized GF was then purified as described above. For all other detergents, the column was washed with Buffers A1 and A2 containing DDM and subsequently washed with one column volume of detergent-free Buffer A to reduce excess DDM. One resin volume of Elution Buffer containing the appropriate detergent was applied and allowed to equilibrate for one hour before elution from the column. For all GF purifications, protein yield was calculated with Beer's Law using absorbance at 280 nm, measured using a Thermo Scientific NanoDrop 2000c Spectrophotometer, and $\varepsilon = 38\ 305\ M^{-1} cm^{-1}$ (calculated from https://web.expasy.org/compute_pi/).

In many cases, GF was further purified by Size-Exclusion Chromatography (SEC) or purified and characterized by SEC-Multi-Angle Light Scattering (MALS) and this is described in detail in Section 2.2.3.

GF in Liposomes

Preparation of Mixed Micelles

An *E. coli* polar lipid extract (Avanti Polar Lipids) was used to form mixed micelles according to the method described by Doerrler *et al.*[258]. Thin films were formed by roto-evaporation; lipids were then sonicated for 5 minutes in liposome buffer containing 50

mM HEPES pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, and 2 mM β-mercaptoethanol where the lipid concentration was 25 mg/mL. The lipids were then frozen in an ethanol/dry ice bath and thawed slowly to room temperature. Lipid vesicles were homogenized by 5 passages through a 27-gauge needle and diluted in the liposome buffer to 4 mg/mL. 1 μmol of DDM per 1 mg of lipid was added to solubilize the lipids and generate mixed micelles.

GF Incorporation into Liposomes

DDM-solubilized GF was added to the mixed micelles where the protein:lipid ratio was 1:100 (w/w), then the mixture was incubated for 30 minutes at room temperature with gentle agitation. 80 mg/mL of Bio-beads were added to the mixture to remove detergent and the mixture was incubated for 2 hours at 4°C. Fresh Bio-beads were added after 2 hours and 16 hours at 4°C. Proteoliposomes were isolated by centrifugation at 125 000 x g for 1 hour at 4°C in a Beckman SW 32 TI rotor and Beckman Coulter Optima LE-80K Ultracentrifuge. The proteoliposomes were resuspended at 0.25 mg/mL of GF in liposome buffer containing 1 mM β-mercaptoethanol.

Preparation of Bicelles

Bicelles were prepared according to the method described by Morrison *et al.*, with a q ratio (DMPC/DHPC), of 0.5 or 4[259]. Dried lipids were first hydrated with bicelle buffer containing 10 mM sodium phosphate pH 7.5, 0.15 mM sodium azide, and 7% D_2O (99.9%) for 2-3 hours. After hydration, 3-5 freeze-thaw cycles with vortexing between cycles were performed, and the bicelles were stored at -80°C.

Incorporation of GF into Bicelles

Pre-formed bicelles were thawed at room temperature until the mixture resembled a gel; they were then vortexed briefly to encourage a homogenous phase and kept on ice. DDM-solubilized GF was added to the bicelles at approximately 5 mg/mL with a bicelle to protein ratio of 1:4 (v/v). The mixture was gently pipetted and incubated on ice for 30 minutes to allow for reconstitution. Successful reconstitution was checked by measuring the melting temperature by differential scanning calorimetry.

Preparation of Empty Nanodiscs

Two different membrane scaffold proteins (MSPs) were used that can form two different sized nanodiscs. MSP1E3D1 and MSP1D1 form nanodiscs with diameters of 12 nm and 9.5 nm, respectively[240]. MSPs were expressed in BL21(DE3) cells containing a pET28a(+) plasmid encoding the MSP gene and a His_6-tag. Pre-cultures and final cell cultures were grown in TB at 37°C with shaking at 300 rpm. Final cultures were inoculated with an overnight pre-culture and allowed to grow to OD approximately 0.6, then induced with IPTG. After induction, the cells were grown for another 2.5 hours. The cells were then harvested by centrifugation at 5000 x g for 15 minutes at 4°C. After harvesting, the cells were resuspended in a buffer containing 20 mM sodium phosphate pH 7.4, 1% Triton X-100, 0.01% sodium azide, 1 mM phenylmethyl-sulfonyl fluoride (PMSF) to discourage proteolytic degradation, and 5 mg of DNAse I. The cells were lysed *via* three 1-minute rounds of sonication; the lysate was then clarified by centrifuging for 30 minutes at 30 000 x g at room temperature.

MSPs were purified from the lysate by IMAC. The nickel resin was first equilibrated with buffer containing 40 mM Tris/HCl pH 8, 0.3 M NaCl, and 1% Triton X-100 (MSP buffer A), and the lysate was applied to the column. The resin was washed with a buffer containing 40 mM Tris/HCl pH 8, 0.3 M NaCl, 50 mM sodium-cholate, and 20

mM imidazole, then a subsequent wash with a buffer containing 40 mM Tris/HCl buffer pH 8, 0.3 M NaCl, and 50 mM imidazole. MSPs were eluted from the IMAC column with a buffer containing 40 mM Tris/HCl buffer pH 8, 0.3 M NaCl, 0.4 M imidazole, and 0.01% sodium azide. Fractions containing protein were verified using SDS-PAGE, and those fractions were pooled and dialyzed against a buffer containing 20 mM Tris/HCl pH 7.4, 0.1 M NaCl, 0.5 mM EDTA, and 0.01% sodium azide. Dialysis was done over 24 hours, and fresh dialysis buffer was used after the initial four hours of dialysis. The protein yield per purification was calculated using absorbance at 280 nm and Beer's law (ε = 21 430 $M^{-1}\,cm^{-1}$ for MSP1D1 and 29 910 $M^{-1}\,cm^{-1}$ for MSP1E3D1)[240]. Dialysed samples were concentrated to 1 mg/mL using an Amicon Ultra centrifugal filter (10 kDa MWCO, Millipore Sigma) and stored at 4°C.

Empty nanodiscs were initially formed with MSP1E3D1 as the surface area of the nanodisc (8 900 Å^2) is potentially suitable for incorporation of a GF tetramer. The surface area of the GF tetramer is approximately 3 920 Å^2 based on the surface area of a single transmembrane ⍺-helix (140 Å^2)[240], and this agrees well with the surface area of the GF tetramer based on the structure deposited in the Protein Data Bank, PDB ID: 1FX8[85]. Nanodiscs were prepared by mixing DMPC with the MSP at a 160:1 ratio, in addition to 20 mM cholate in dialysis buffer with incubation for 1 hour. Bio-beads were added to remove the detergent and the mixture was allowed to incubate for 18 hours with gentle stirring to ensure removal of the detergent. Lipid-filled nanodiscs were obtained after removal of the Bio-beads.

After elution from the IMAC column, the most concentrated fraction of DDM-solubilized GF based on absorbance at 280 nm was added to the MSP/lipid/cholate mixture, where the number of nanodiscs was in excess of the number of GF tetramers and the mixture was equilibrated for 1 hour before the addition of Bio-beads. The mixture was equilibrated for 18 hours before removal of the Bio-beads, as described above. GF incorporation was verified by SDS-PAGE and DSC.

Purification of the MBP-GF-ApoA1* Fusion Protein

After centrifugation, the supernatant containing the GF fusion protein was applied to the IMAC column. Buffers A1 and A2 and Elution Buffer were prepared without detergent, and the IMAC column wash and elution steps were followed as described above for GF.

2.2.3 Characterization of the Stability and the Oligomeric State of Solubilized GF

The oligomeric state of detergent-solubilized GF was determined by size-exclusion chromatography multi-angle light scattering (SEC-MALS), SDS-PAGE (poly-acrylamide gel electrophoresis), blue native (BN) PAGE, dynamic light scattering (DLS), and differential scanning calorimetry (DSC).

Size Exclusion Chromatography-Multi-angle Light Scattering (SEC-MALS)

SEC-MALS measurements were carried out on the fraction eluted from the IMAC column containing the largest amount of protein based on absorbance at 280 nm. Samples were centrifuged for 5 minutes at 15 000 rpm in an Eppendorf centrifuge 5424 with rotor FA-45-24-11 to remove large aggregates. SEC-MALS measurements first gel filtered samples using a Superdex 200 Increase 10/300 column on an AKTA Pure 25 chromatography system. Gel filtration was done with Buffer A1 pre-filtered through a 0.22 μm filter. The gel filtered sample then passed through the DAWN Heleos-II light scattering (LS), OptiLab T-rEX refractive index (RI), and UV absorbance detectors in series. Protein conjugate analysis was performed using the ASTRA software (Wyatt Technology) (See Section 1.3.2 of the Introduction).

Denaturing gels were run on a Hoefer® Mighty Small II SE 250 mini-vertical gel electrophoresis unit. Native gels were run on the Mighty Small II or an Amersham ECL gel box.

SDS-PAGE gel samples were incubated for 10 minutes at room temperature at a 1:5 sample buffer to sample ratio. The sample loading buffer consisted of 50 mM Tris-Cl buffer pH 6.8, 2% w/v SDS, 0.1% w/v Bromophenol blue, 10% w/v glycerol, and 1% v/v β-mercaptoethanol. Gels were composed of a 5% acrylamide stacking gel and 8% resolving gel. The gel running buffer contained 25 mM Tris/HCl buffer pH 8.3, 125 mM glycine, and 0.5% w/v SDS. Electrophoresis started at 80 volts until all protein entered the resolving gel; the voltage was then increased to 150 volts. Following electrophoresis, gels were fixed with 25% isopropyl alcohol and 10% acetic acid for 20 minutes before staining with Coomassie Brilliant Blue R-250 to visualize protein bands.

Blue Native gel samples were incubated in 12.5 mM Tris buffer pH 7, 20% w/v glycerol and 0.02% w/v Coomassie Brilliant Blue G-250, for 10 minutes with a 1:1 buffer to sample ratio. Pre-cast 4%-12% continuous gels were purchased from GE Healthcare Life Sciences (used with the Amersham gel box) or GenScript (SurePAGE™ 10-well plates, used with the Mighty Small II box). Gels were run at 50 volts until all the protein entered

the gel. The voltage was increased to 200 volts and electrophoresis continued for a further 1-1.5 hours.

Intrinsic Tryptophan Fluorescence

Intrinsic tryptophan fluoresence spectra were measured on a Jasco 810 spectropolarimiter/fluorometer. Sample temperature was maintained at 20°C with a built-in Pelltier device. GF samples were placed in a 1 cm path length quartz fluorescence cuvette and spectra were acquired from 300 to 450 nm using a 1 s response time, a 1 nm data pitch, a scanning speed of 10 nm/min, a bandwidth of 5 nm, and an excitation wavelength of 280 nm. Samples were equilibrated at room temperature at the new pH for 10 minutes after titration before measuring spectra.

Dynamic Light Scattering

Experiments testing the effect of storage temperature and protein concentration on the oligomeric state of GF were conducted with a Nano-S Dynamic Light Scattering system (Malvern Instruments Ltd., Malvern, UK) using 45 µL quartz cuvettes. The pathlength of the cell is 3 nm. Backscatter detection was measured at an angle of 173° with a laser at 633 nm. All samples were equilibrated for 10 minutes at 20°C before DLS measurements were taken. DLS samples were measured in triplicate and averaged.

For samples used to explore the effect of storage temperature the concentration of GF in Elution buffer containing 0.5 mM DDM was 97 µM; equal volumes of the sample were used for storage at room temperature, at 4°C, and -80°C. To test the effect of protein concentration on the stability and oligomerization of GF, a sample containing Elution Buffer, 0.5 mM DDM, and 112 µM GF was separated into three aliquots. One aliquot was not diluted, another was diluted 1:10, and the third diluted 1:100 in Buffer A1 containing 0.5 mM DDM. Half of the volume of each of the three initial aliquots was stored at 4°C to monitor the effects of dilution and storage temperature. The other halves were kept at room temperature. All samples were mixed before aliquoting to ensure the even distribution of GF. DLS measurements were taken as described above.

Thermal Melting of GF in Detergents and Other Membrane Mimetics

The melting temperature of GF solubilized by detergents, bicelles, nanodiscs, and RHPs was conducted on a Nano DSC (TA Instruments, New Castle, DE). Samples were scanned from 20°C to 90°C at a rate of 1 °C/minute in 0.3 mL capillary cells. NanoAnalyze software (TA Instruments) was used to process the data.

Samples were measured after being eluted from the IMAC column in Elution Buffer and either 0.5 mM DDM or 0.05% w/v LMNG against a buffer blank of Elution Buffer and the detergent used. To measure the melting temperature of GF incorporated into bicelles, the sample was first equilibrated to room temperature after incubation with

empty bicelles, then measured against a buffer blank containing Elution Buffer and 0.5 mM DDM. The melting temperature of GF in nanodiscs was measured after the final dialysis step. The buffer blank used was the dialysis buffer outlined above. RHP-solubilized GF samples were measured after purification from the IMAC column against a buffer blank of Buffer A and 2 mg/mL of 36 kDa RHP.

2.2.4 Characterizing the GF Octamer

Negative Stain Electron Microscopy

Sample Preparation

DDM-solubilized GF samples were eluted from the IMAC column as outlined above. The most concentrated fraction eluted from the IMAC column based on absorbance at 280 nm was used to prepare negative stain grids. LMNG-solubilized GF samples were eluted from the IMAC column and fractions containing GF based on absorbance at 280 nm were pooled and dialysed overnight using 3500 MWCO dialysis tubing (Fisher Scientific) to remove imidazole. The pooled sample was concentrated using an Amicon Ultra centrifugal filter (100 kDa MWCO, Millipore Sigma) using a Sorvall Legend XTR centrifuge (ThermoFisher Scientific) and a TX-750 4 swing out rotor. Samples were concentrated to approximately 300 µL and gel filtered on the SEC-MALS as described above with buffer containing 0.05% w/v LMNG to remove aggregates and assess the quality of the protein. Fractions containing GF tetramers and octamers were identified by the peak elution position and collected.

Electron microscopy

Negative stain electron microscopy was carried out by Cynthia Lee Page at the Boulder Electron Microscopy Services in the Department of Molecular, Cellular, and Developmental Biology at the University of Colorado under the direction of Professor Andy Hoenger. Approximately 4 μl of GF diluted 1:100 with SEC-MALS buffer were adsorbed to a glow discharged continuous carbon copper grid (Electron Microscopy Sciences, Hatfield, PA). The sample was allowed to adhere for 30 seconds, blotted off, washed twice with water, then stained with 2% aqueous uranyl acetate for 30 seconds before blotting dry.

Image data were collected on an FEI Tecnai F20 FEG transmission electron microscope (FEI-Company, Eindhoven, The Netherlands) operating at 200 kV. Images were acquired at a magnification of 62 000x and a defocus of -0.6 μm using a total dose of approximately 15 to 30 electrons/Å^2. Images were recorded binned by two on a 4K × 4K Gatan Ultrascan 895 CCD camera (Gatan Inc., Pleasanton, CA). The resulting pixel size corresponds to 7.3 Å on the specimen. SerialEM software[260] was used to automate the data acquisition. Images were processed and 2D averages were computed using Relion[261].

Size Exclusion Chromatography-Small -Angle X-ray Scattering

Sample Preparation

LMNG-solubilized GF samples were prepared similarly to LMNG-solubilized GF negative stain EM samples. After dialysis, samples were concentrated with an Amicon Ultra centrifugal filter (100 kDa MWCO, Millipore Sigma) using a Sorvall Legend XTR centrifuge (ThermoFisher Scientific) and TX-750 4 swing out rotor and a second gel filtration step was done. After the second gel filtration step, fractions containing GF tetramers and octamers were collected based on peak elution position, pooled, and concentrated using an Amicon Ultra centrifugal filter (100 kDa MWCO, Millipore Sigma) using a Sorvall Legend XTR centrifuge (ThermoFisher Scientific) and TX-750 4 swing out rotor to approximately 8 mg/mL.

SAXS Data Acquisition, Processing, and Modeling

In-line SEC-SAXS data were collected by Ewan McRae and Matthew McDougall (Department of Chemistry, University of Manitoba) at Beamline 21, Diamond Light Source (Chilton, Oxfordshire, UK). 50 µL containing 8 mg/mL of LMNG-solubilized GF was passed over a 2.4 mL Superdex 200 Increase column equilibrated with Buffer A1 containing 0.01% w/v LMNG using an Agilent 1200 high performance liquid chromatography (HPLC) system. The eluted column fractions were injected directly into

the SAXS beam. Further information about the system can be found here:

https://www.diamond.ac.uk/Instruments/Soft-Condensed-Matter/small-

angle/B21/description.html.

ScÅtter (http://www.bioisis.net/users/sign_in/) and the ATSAS suite[262] including

Crysol[263], Gnom[264], and Sasref[265] software were used to process the SAXS data; models of

the GF homotetramer surrounded by the DDM detergent shell were constructed using

the GF crystal structure (PDB ID: 1FX8)[85] and the Memprot algorithm[266]. The GF octamer

model was created using Pymol (https://www.schrodinger.com/pymol) to generate

symmetry mates.

The LMNG-solubilized GF octamer was built as follows: the CHARMM-GUI

Micelle Builder algorithm[267] was used to construct an LMNG micelle around the protein

tetramer. This detergent-belted tetramer was superimposed on each half of the X-ray

diffraction octamer model shown in Figure 57 using Pymol[268]. A SAXS scattering curve

was generated using the CRYSOL algorithm[263] in the ATSAS package. The resulting

scattering curves had χ^2 values of 9.1. Next, the SASREF algorithm[269] in ATSAS was used

to search for conformations of the octamer that best fit the measured scattering curve by

varying the positions and orientations of each tetramer with its detergent belt. This

reduced the χ^2 values to 6.8. The best model had the tetramers and their detergent belts

tilted in a manner similar to that observed in the negative stain EM images shown in

Figure 60. In the final step, models of the protein termini were added. Since the SASREF

algorithm[269] is limited to a maximum of 10 subunits, each of the 8 N-and C-termini were

combined into one continuous amino acid sequence adopting a 3_{10} helical conformation

(to reduce the size of the water box) built using Chimera[270] and energy-minimized with

the GROMACS 5.1.4 molecular dynamics simulation package[271]. The simulations were

carried out under periodic boundary conditions at constant temperature (T = 310 K) and

pressure (P = 1 bar). A 50 ns MD simulation was then performed on the energy-minimized

structure and 8 conformations were chosen at random from the last 10 ns of the

simulation to represent different conformations of the presumably disordered termini.

The 8 terminal conformations were added to the best fitting SAXS model as described

above. The termini were added either to the interface between the tetramers or to the

surfaces of the octamers in the case where the tetramers interacted through their

periplasmic surfaces.

2.2.5 Solution and Solid-state NMR Spectroscopy of GF

Solution NMR Sample Preparation

^{15}N isotope-labelled GF was extracted from *E. coli* membranes with extraction buffer containing 20 mM SDS followed by purification *via* IMAC, as described above. The most concentrated fraction eluted from the column based on absorbance at 280 nm was added to an NMR tube with 7% D_2O and 75 μM 2,3-dimethyl-2-silapentane-5-sulfonate (DSS); the tube sealed with Teflon tape.

Solution NMR Data Collection and Processing

Solution NMR spectra were acquired on approximately 100 μM SDS-solubilized GF by Vu To with an Agilent/Varian INOVA 600 MHz NMR spectrometer using a triple-resonance probehead at 25°C at the University of Manitoba. Heteronuclear single quantum coherence (HSQC) NMR spectra were acquired over 10 hours with a spectral width of 16x32 ppm ($^1Hx^{15}N$) with 128 scans and 128 increments in the ^{15}N dimension. The standard Agilent/Varian BioPack pulse sequence was used[272].

Solid-state NMR Sample Preparation

After elution from the IMAC column, samples containing ^{15}N and ^{13}C isotope-labelled GF, based on absorbance at 280 nm, were pooled and dialysed against a buffer containing 25 mM sodium phosphate pH 7.5, 150 mM sodium chloride, 0.04% sodium azide, and 0.01% w/v LMNG overnight to remove imidazole and reduce the concentration of LMNG to 0.01% w/v. After dialysis, the pooled samples were concentrated to approximately 75 µL using a centrifugal concentrator. This sample was then dialysed against the same dialysis buffer above containing 50% w/v PEG 2000 to further reduce the sample volume. Dialysis was stopped at an approximate sample volume of 20 µL and a protein concentration of approximately 32 mg/mL.

ssNMR Spectroscopy Data Collection and Processing

Approximately 8 mg of ^{13}C-, ^{15}N-labelled LMNG-solubilized GF were packed into a 1.6 mm ssNMR rotor. All NMR spectra were acquired on a 600 MHz Bruker Avance III spectrometer (Bruker Biospin Ltd.) equipped with a 1.6 mm 40 kHz MAS (Magic Angle Spinning) probe (Phoenix NMR) at 0 °C. For all experiments, the ^{1}H 90 degree pulse was 161.3 kHz at 77.5 W, the ^{13}C 90 degree pulse was 125 kHz at 315 W, and the ^{15}N 90 degree pulse was 93.6 kHz at 400 W. The MAS rate was 13.333 kHz. The ^{1}H transmitter offset was set to 2 kHz. Spinal-64 ^{1}H decoupling at 110 kHz was used during all acquisition periods. 1D hN and hC CP-MAS spectra were obtained with 256 scans and processed with 60 Hz

line broadening. The CP contact time for the hN experiment was 1.5 ms. The ^{15}N transmitter offset was 7.6 kHz and the SW was 39.7 kHz. The ^{1}H pulse was rectangular with a 23 W power level. The ^{15}N pulse was a 90%-100% linear ramp about 200.6 W. The CP conditions for the hC experiment were a contact time of 1 ms, a rectangular 28 W ^{1}H pulse, and a ^{13}C 90%-100% tangent ramp centered at 218 W. The transmitter offset was 15.1 kHz and the SW was 81.5 kHz. 2D DARR (Dipolar Assisted Rotational Resonance) NMR spectra[273] were acquired with States-TPPI in the indirect dimension over 4.5 days with a 10 ms mixing time. The transmitter frequency offset was 100 ppm in both directions; the spectra shown are a summation of four blocks of the experiment with 32 scans and 768 rows. The same hC CP conditions stated above were used. The SW was 81.5 kHz in T_2 and 80 kHz in T_1. The DARR field on the ^{1}H channel was 13.333 kHz. Acquisition times were 24.6 ms in T_2 and 4.6 ms in T_1. NMR spectra were processed using NMRPipe[274] with an exponential apodization function of 80 Hz and a 150 Hz Gaussian; the data were zero-filled in the direct dimension to 8192 points and in the indirect dimension to 4096 points. The first three points were corrected with linear prediction using 12 coefficients. In the DARR experiment, magnetization is transferred from protons to 1-bond-coupled ^{13}C and then transferred through space to nearby ^{13}C. At short mixing times nearby intra-residue connections are observed but for long mixing times inter-residue connections can be observed and used for resonance assignment. T_2 measurements were done with a pseudo-2D spin-echo experiment. The ^{13}C transmitter offset was centered in the CA region at 52 ppm. The refocusing pulse was an Rsnob at 10.15 dB with a width of 637 uS to decouple

the neighbouring COs and CBs. There was a total of 8 rows acquired, each with 16 scans. The direct dimension was processed using the same parameters as the hC spectrum. T_2 relaxation times were fitted using Topspin Dynamics Center (Bruker Biopsin Ltd.) with an integral region of 23 to 81 ppm and a calculated T_2 of 1.8 ms.

Chapter 3: Results

3.1 Optimizing GF Overexpression

Initial protein expression experiments used BL21(DE3) pLyS cells containing the thrombin-plasmid. These cells were grown in rich (LB or TB) and M9 minimal media. Solubilization and purification in DDM resulted in protein yields of approximately 5 mg of DDM-solubilized GF per litre of cell culture in rich media and 3 mg/L in minimal media. To improve protein yields, the pLysS cell glycerol stocks were refreshed by re-plating and preparing new glycerol stocks with well-growing isolated colonies. The thrombin-plasmid was also purified from the pLysS cells and transformed into BL21(DE3) cells. Protein yields remained the same after refreshing the pLysS glycerol stocks, clonal selection, and transformation into BL21(DE3) cells.

The thrombin-plasmid contains the endogenous GF gene, which was sequenced and found to encode a valine instead of an aspartic acid at position 168[275]. A new construct was created with the correct sequence (V168D), an N-terminal His$_6$ tag, a TEV protease cleavage site, and was codon optimized. Overexpression of GF with the new TEV-plasmid in BL21(DE3) cells (TEV-cells) in rich media increased protein yields to approximately 20 mg/L.

In this work, Studier media LS-5052 (uniform labelling) and C-750501 (uniform labelling optimized for maximum incorporation of ^{13}C and additional labelling with ^{15}N)[204] were screened for production of GF amenable to NMR spectroscopy measurements. All media were adjusted to pH 7 before sterilization; this was done to prevent acidification of the cell culture from excreted protons and metabolic waste. TEV-cells grown in LS-5052 and C-750501 reproducibly yielded 30 to 40 mg of GF per litre of cell culture, enough protein for characterization and NMR studies.

3.2 Purification of DDM-solubilized, Tetrameric GF

The SEC-MALS elution profile of DDM-solubilized GF shown in Figure 26 shows that GF exists as a mixture of about equal amounts of tetramer and octamer, and some larger aggregates, shortly after purification from the IMAC column. Storing the sample for 24 hours at room temperature resulted in visible precipitate. NMR spectroscopy measurements require samples to remain in solution and stable for several days and up to one week, in addition to being highly concentrated and isotopically-labelled; a polydisperse sample such as the one shown in Figure 26 would not likely be amenable to NMR spectroscopy.
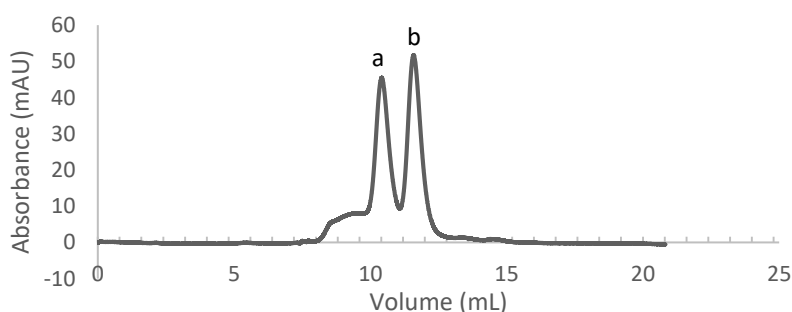
Figure 26. The SEC-MALS elution profile of DDM-solubilized GF freshly purified from TEV-cells grown in rich media. The y-axis is absorbance at 280 nm. The major x-axis markers represent the volume eluted from the SEC column. The minor x-axis makers denote the volumes of fractions that were collected. Peak (a) corresponds to the GF octamer. Peak (b) is the GF tetramer.

A variety of buffer conditions were screened to discourage GF octamer formation and precipitation. The following sections describe the optimization of DDM-solubilized GF homotetramer preparations of protein expressed from TEV-cells grown in LB rich medium.

3.2.1 Buffer Environment

We optimized GF solubility using a buffer containing 250 mM sodium phosphate at pH 7.5 including 25 mM sodium chloride. Based on the polydispersity observed in the SEC-MALS elution profile in Figure 26, further optimization was required to obtain a monodisperse DDM-solubilized GF tetramer.

Phosphate Buffer Concentration

Optimization of phosphate buffer concentration and pH considered the range of pHs required to keep GF soluble and the temperature range over which the buffer is soluble. GF was previously shown to aggregate irreversibly in solution at low and high pH[276]; a pH range of approximately 6-8 is required for formation of soluble GF tetramers and to reduce visible aggregates in the sample to a minimum. Phosphate buffer has a buffering capacity within this pH range and is physiologically relevant[277]. Buffers used in the laboratory were stored at 4°C to reduce observed microorganism growth in the buffer. High concentrations of phosphate buffer precipitate at low temperatures, which reduces the concentration of buffer in solution. Phosphate buffer concentration was optimized to moderate the pH drop observed in the freeze-thaw cycles during the purification process, as well as to inhibit precipitation of phosphate buffer at 4°C. Phosphate buffer concentration was lowered to 25 mM; purification with the new concentration suggested adequate buffering power during the freeze-thaw cycles with little pH change and no

visible precipitate after storage at 4°C. Sodium azide (0.04% w/v) was also added to the buffer to discourage microorganism growth, which also allowed for buffer storage at room temperature.

Effect of pH

Previous pH titrations of GF followed by SDS-PAGE appeared to show that GF dissociates into monomers, dimers, and trimers at low pH[276]; raising the pH to within the physiological range resulted in partial reassembly of the tetramer. At pH values greater than pH 8, visible aggregation was observed. However, the oligomeric state of GF was monitored *via* SDS-PAGE, which contains high concentrations of SDS that denatures proteins and might have contributed to the observed oligomeric state of GF. We therefore decided to use intrinsic protein fluorescence to monitor the effects of pH on GF tertiary structure as well as the folding and refolding of the tetramer in the absence of denaturing SDS.

GF monomers have 5 tryptophan residues (Figure 27) that can be used to monitor intrinsic GF tryptophan fluorescence over a pH range *via* excitation at 280 nm. Figure 28 shows the fluorescence intensity of DDM-solubilized GF at the starting, low, and ending pH's of the titration (7.4, 3.7, and 7.5 respectively). At pH 7.4, the tryptophan fluorescence maximum is at 335 nm and contains shoulders at 319 nm and 309 nm at the beginning of the titration. At the lowest pH (3.7), the shoulders have disappeared, the fluorescence

intensity is enhanced, and the main peak has shifted to 341 nm. The red shift suggests that there are more tryptophan residues exposed to water and that the GF tetramer has dissociated at low pH. This interpretation agrees with the SDS-PAGE results that show the protein tetramer dissociating below pH 6[278]. When the sample is titrated back to the starting pH (7.5), the shoulders at 309 and 319 nm are restored and the main peak shifts back to 335 nm, but the fluorescence is slightly more quenched than it was at the start of the titration. These results support the earlier SDS-PAGE results[278] suggesting that much of the protein can be restored to its neutral pH tetrameric state following acidification. Unfortunately, the protein is not very stable at acidic pH, forming irreversible precipitates so that NMR studies on the putative monomer at pH 3.7 were not possible. Note also that during the titration the amount of Rayleigh scattering at the excitation wavelength increases, suggesting increasing amounts of aggregated protein.
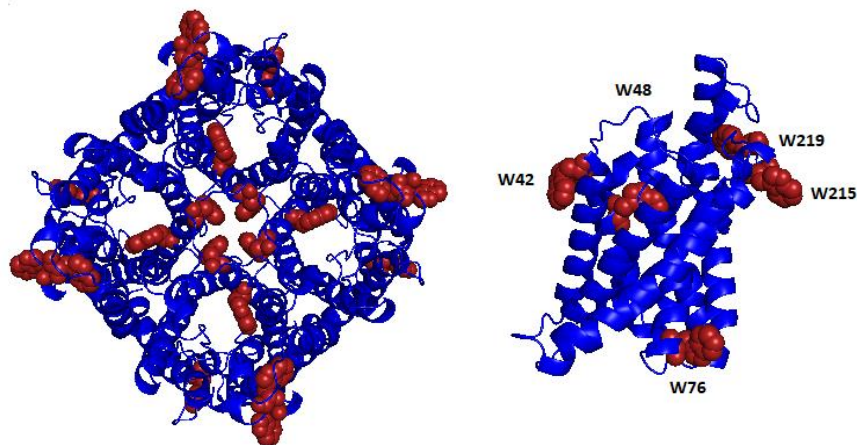
Figure 27. Ribbon diagram of the GF crystal structure. Left: the GF homotetramer, viewing from the periplasm-facing surface. Right: the GF monomer with the periplasmic surface at the top and the cytoplasmic surface at the bottom. Blue ribbon: transmembrane α-helices and disordered loops, red spheres: tryptophan residues. (PDB ID: 1FX8).
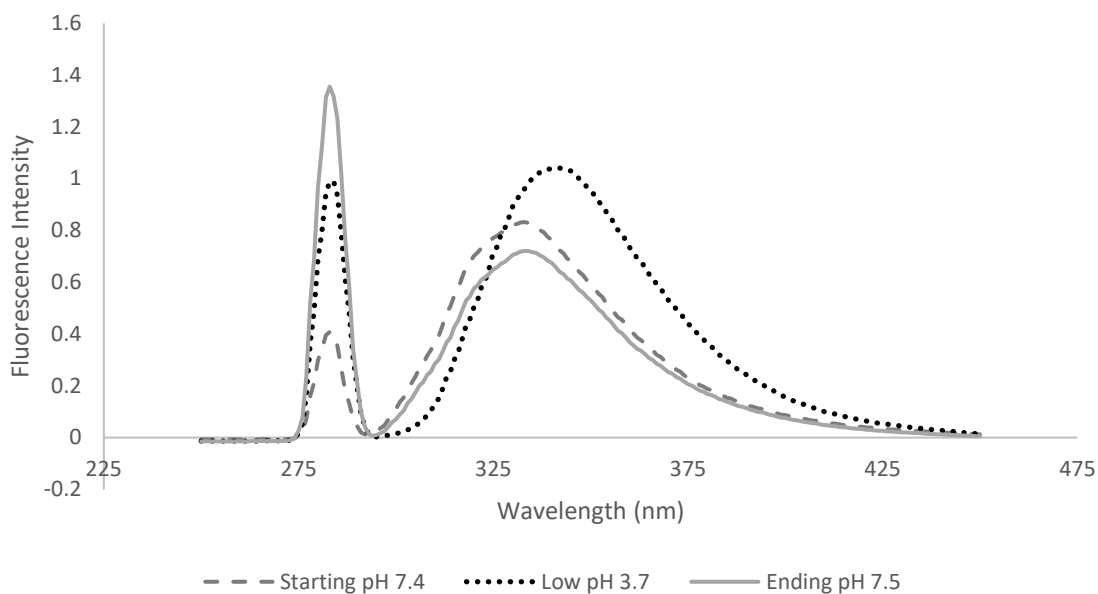
Figure 28. *A pH titration of GF tryptophan fluorescence was conducted with excitation at 280 nm. Spectra were measured at pH 7.4 (grey dashed line), pH 3.7 (dotted line) and following restoration to pH 7.5 (solid grey line). At pH 7.4 (grey dashed line) the spectrum shows a peak at 335 nm and shoulders at 319 nm and 309 nm. At pH 3.7 the fluorescence intensity is enhanced, the shoulders have disappeared, and the main peak has shifted to 341 nm, suggesting greater exposure to water of some of the Trp. Upon restoration of the pH to 7.5 (solid grey line), the shoulders at 319 nm and 309 nm reappear, the main peak returns to 335 nm, but the intensity is quenched more than what was observed prior to acidification.*

Sodium Chloride Concentration

SEC-MALS elution profiles of DDM-solubilized GF suggested that the sample was polydisperse, despite optimization of the phosphate buffer concentration and exploring the effect of pH on the oligomeric state of GF. Initial GF experiments used 25 mM NaCl, which is added to promote protein solubility. Several concentrations of NaCl were explored and optimization led to an NaCl concentration of 150 mM, which also provides an osmolarity that is a reasonably physiological[279]. The newly optimized conditions led to SEC-MALS elution profiles of DDM-solubilized GF that contained a single peak

117

corresponding to the GF tetramer (Figure 29, top), which, after protein conjugate analysis, reproducibly yielded an approximate M_r of 150 kDa, within acceptable error limits of the true M_r of 128 kDa. Unfortunately, 24 hours later, the SEC-MALS elution profile of the same sample stored at room temperature showed two peaks corresponding to the octamer (calculated M_r approximately 290 kDa) and the tetramer (Figure 29, bottom).
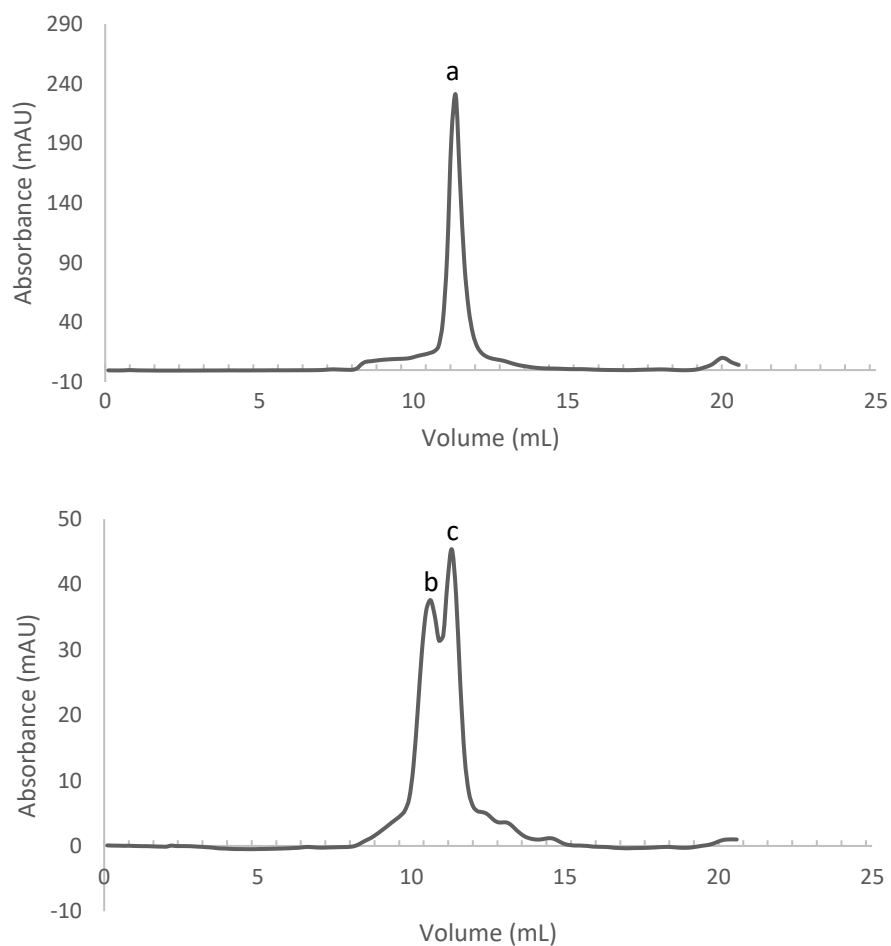
*Figure 29. Optimizing the buffer environment for DDM-solubilized GF. **Top:** the SEC-MALS elution profile showing a pure homotetramer of DDM-solubilized GF, freshly purified from TEV-cells grown in rich media, where peak (a) corresponds to the GF homotetramer. **Bottom:** the sample from (a) stored for 24 hours at room temperature shows two peaks corresponding to the octamer (peak b) and tetramer (peak c).*

Imidazole Concentration

Poly-histidine tags are used to purify recombinant proteins on IMAC columns as the imidazole ring in histidine interacts strongly with the immobilized transition metal ions[280]. The histidine analog imidazole is used to elute proteins of interest bound to the resin. However, imidazole also contributes to the salt concentration of the buffer and we wondered if the imidazole was a factor in the long-term instability of detergent-solubilized GF. The 250 mM imidazole was removed by gel filtration with a running buffer containing optimized Buffer A1 (25 mM phosphate buffer pH 7.5, 150 mM NaCl, 0.04% w/v azide) and 0.5 mM DDM or by overnight dialysis against the same buffer. DLS was used to compare samples stored with imidazole in the buffer with those that were desalted *via* SEC. Although DLS is less informative than SEC-MALS, it has the advantage that measurements can be taken more quickly and does not involve diluting the samples, which could alter protein self-association. Panel A in Figure 30 shows a DLS spectrum of Elution Buffer containing 0.5 mM DDM. The peak with an average particle diameter of 6.5 nm is in good agreement with the measured diameter of DDM micelles of 8.1 nm at 22°C, with the known size of the detergent[281], and with earlier measurements in our group[282]. The peak at about 615 nm likely corresponds to a rod-like DDM micelle. Rod-like micelles appear under a variety of conditions including increasing temperature[282], salt concentration, and surfactant concentration[26,283-285]. Note that larger particles scatter more intensely than smaller particles (Rayleigh scattering scales to the sixth power of the diameter) and, therefore, the intensities of the peaks do not directly report the relative

amounts of the spherical and micellar aggregates[179]. Indeed, the number of rod-like

micelles is likely to be very small and the reported relative number of 615 nm diameter

particles is 0%.
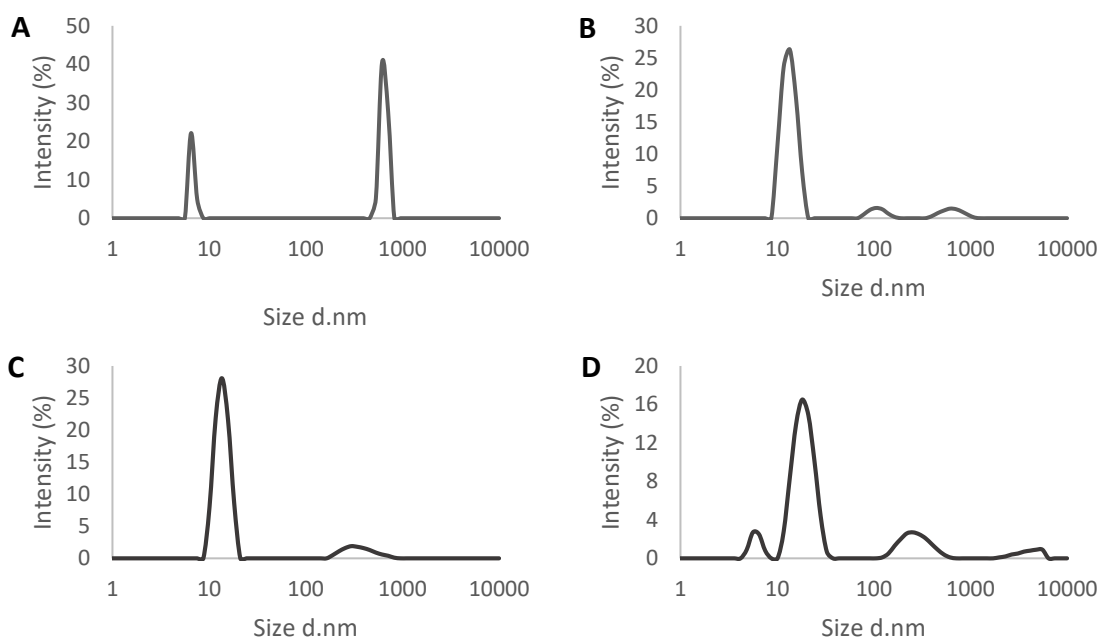
*Figure 30. The effect of salt concentration on DDM-solubilized GF self-association. DLS measurements show the hydrodynamic size of particles distributed by intensity. (**A**) Elution Buffer containing 0.5 mM DDM. (**B**) Freshly prepared DDM-solubilized GF in buffer A containing 250 mM imidazole prior to desalting by gel filtration. (**C**) SEC sample (**B**) suggests a single, homogenous sample. (**D**) Sample (**C**) stored for 5 days at room temperature. (**D**) suggests that desalted DDM-solubilized GF remains mostly homogenous with small amounts of higher M_r oligomers.*

Panel B in Figure 30 shows the DLS spectrum of DDM-solubilized GF upon elution

from the IMAC resin. In excellent agreement with previous results[282] and the known size

of the GF tetramer[85], the major peak appears at about 13.5 nm. The peak at about 712 nm

likely represents rod-like empty micelles and is greatly diminished compared to that

observed in Panel A, presumably because most of the detergent is associated with GF,

although we cannot rule out the possibility that this is a GF-DDM aggregate. A new peak

observed at about 106 nm likely represents a small amount of aggregated GF, which is also observed in the SEC-MALS elution profiles (Figure 29). SEC-MALS purification of the sample in B results in the removal of the aggregated protein giving rise to a preparation that appears to contain DDM-solubilized tetramers (13.5 nm), some GF-DDM aggregate (342 nm), and, possibly, some rod-like micelles (615 nm, Figure 30C). Storage of the protein for 5 days at room temperature results in a broadening of the protein peak that moves to a higher average hydrodynamic radius of about 18 nm (Figure 30D), suggesting the appearance of slightly larger protein oligomers, likely to be octamers, judging from the previous SEC-MALS results (Figure 29, bottom). This sample also contains more aggregate with average diameter of 255 nm. Interestingly, storing the sample for 5 days appears to release the DDM, which reforms 6.5 nm diameter micelles. Overall, it appeared from these experiments that immediately removing imidazole from the detergent-solubilized sample by gel filtration gave rise to less self-association over time (Figure 30). However, it cannot be ruled out that diluting the protein by gel filtration may also have contributed to the reduced self-association.

Effect of Osmolytes

Glycerol is an osmolyte that moderately increases protein stability[286] in addition to being the substrate of GF. Glycerol was screened for its potential in stabilizing DDM-solubilized GF. The SEC-MALS elution profile of DDM-solubilized GF with 5% w/v

glycerol added (Figure 31) shows multiple peaks, corresponding to tetramer, octamer, and aggregates, indicating a polydisperse sample in the presence of glycerol. Thus, it appears that glycerol increased the protein's self-association compared to its absence (Figure 29, top).
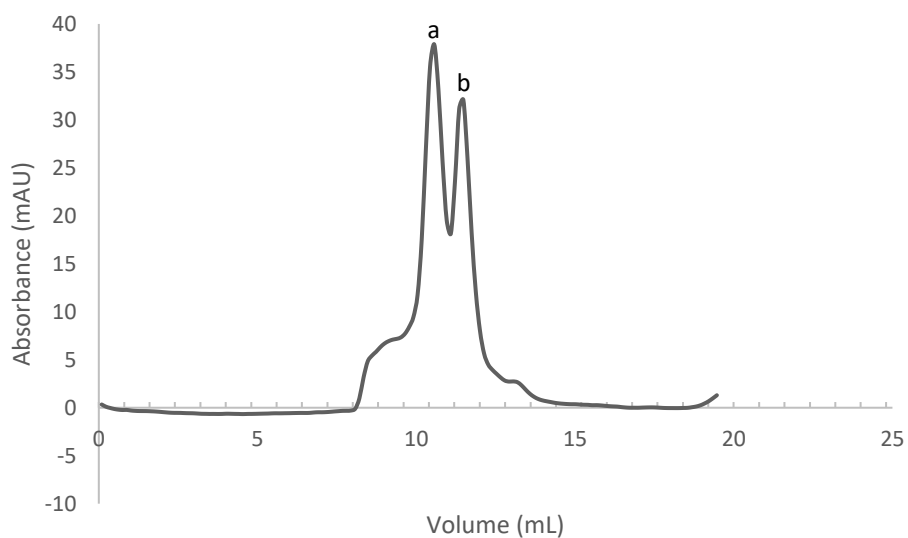
Figure 31. *The effect of the substrate and osmolyte glycerol on the oligomeric state of GF. The SEC-MALS elution profile of DDM-solubilized GF in Buffer A1 and 5% w/v glycerol shows multiple peaks and suggests that the addition of glycerol does not encourage a monodisperse sample, yielding octamer (peak a) and tetramer (peak b) and some larger aggregates.*

Other Buffer Additives

Baynes *et al.* proposed the "*gap effect theory*" which may explain why some buffer additives appear to improve protein solubility and stability[287]. The gap effect theory suggests that these additives increase the free energy of protein-protein interactions and discourage aggregation. In their work, 0.5 M arginine was shown to slow the aggregation

of protein-folding intermediates as well as reduce the amount of recombinant proteins being incorporated into inclusion bodies during overexpression. For DDM-solubilized GF in the presence of 0.5 M arginine, the SEC-MALS elution profile showed that GF elutes as a single tetrameric peak (Figure 32, top). The sample was stored for three days at room temperature and gel filtered, and the SEC-MALS elution profile showed multiple peaks corresponding to higher M_r oligomers and possibly a small amount of monomeric protein (Figure 32, bottom). Thus, 0.5 M arginine was not effective at preventing GF self-association.
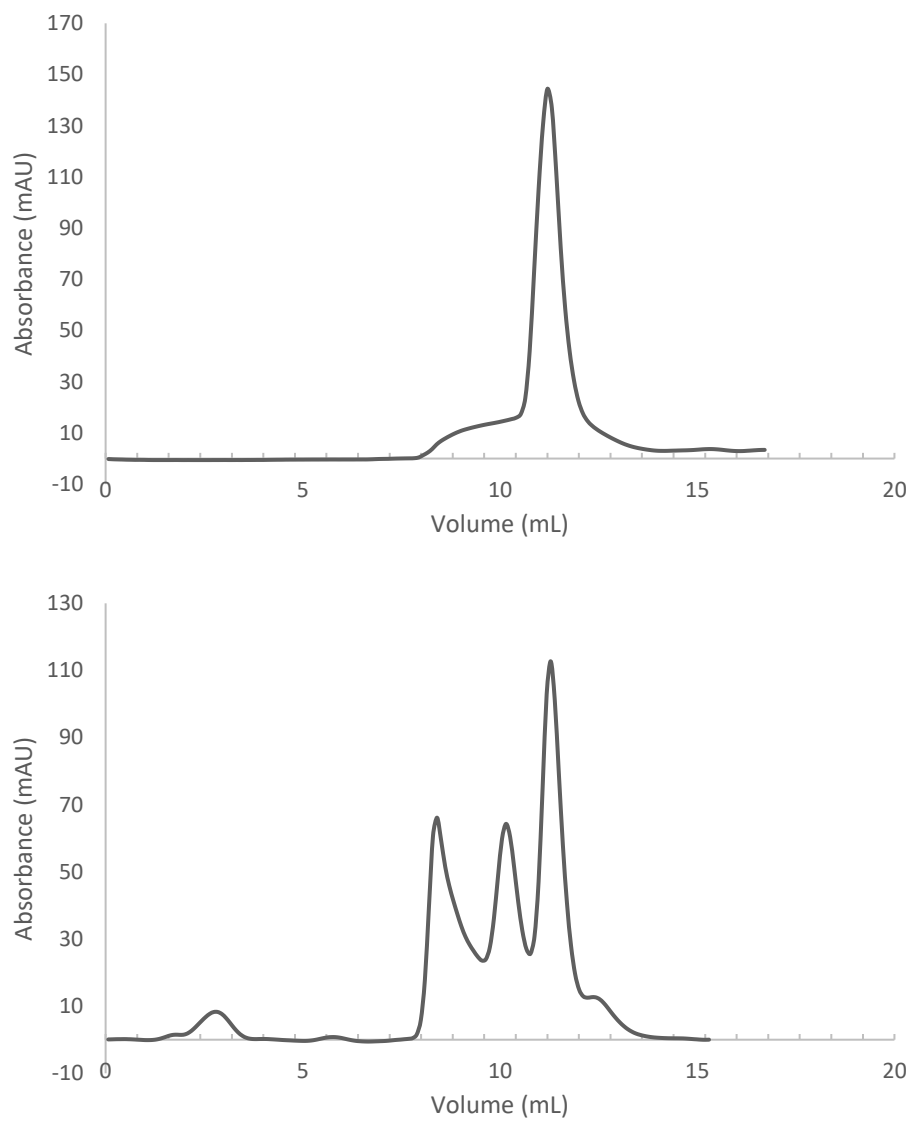
Figure 32. *The effect of 0.5 M arginine on the oligomeric state of GF. The SEC-MALS elution profile of DDM-solubilized GF in Buffer A1 and 0.5 M arginine.* **Top:** *the most concentrated GF fraction freshly purified by IMAC elutes as a single, tetrameric peak.* **Bottom:** *the sample from (top) stored at room temperature for 3 days and gel filtered. Based on the multiple peaks, the elution profile shows that the sample has associated to larger oligomers and aggregates and possibly a small amount of monomeric protein.*

Eliezer *et al.* were able to stabilize the equilibrium intermediate of a small water-soluble protein for solution NMR studies by adding 10% v/v ethanol (EtOH) to their sample[288]. We wondered if addition of ethanol might reduce GF hydrophobicity-mediated self-association and added 10% EtOH to the GF fractions immediately upon elution from the IMAC column. Shortly afterwards, visible precipitation was observed in the sample. When we reduced the EtOH concentration to 5% v/v, no visible precipitation was observed, and DLS measurements showed that DDM-solubilized GF appeared to remain predominantly tetrameric (13.5 nm) (Figure 33, top) accompanied by small amounts of 91 nm and 1718 nm diameter aggregates. However, 24 hours later (Figure 33, bottom), visible precipitate was observed, the smallest diameter particle had increased to 15.7 nm, and a significant amount of 68 nm aggregate was present, suggesting that ethanol was unable to prevent GF self-association.
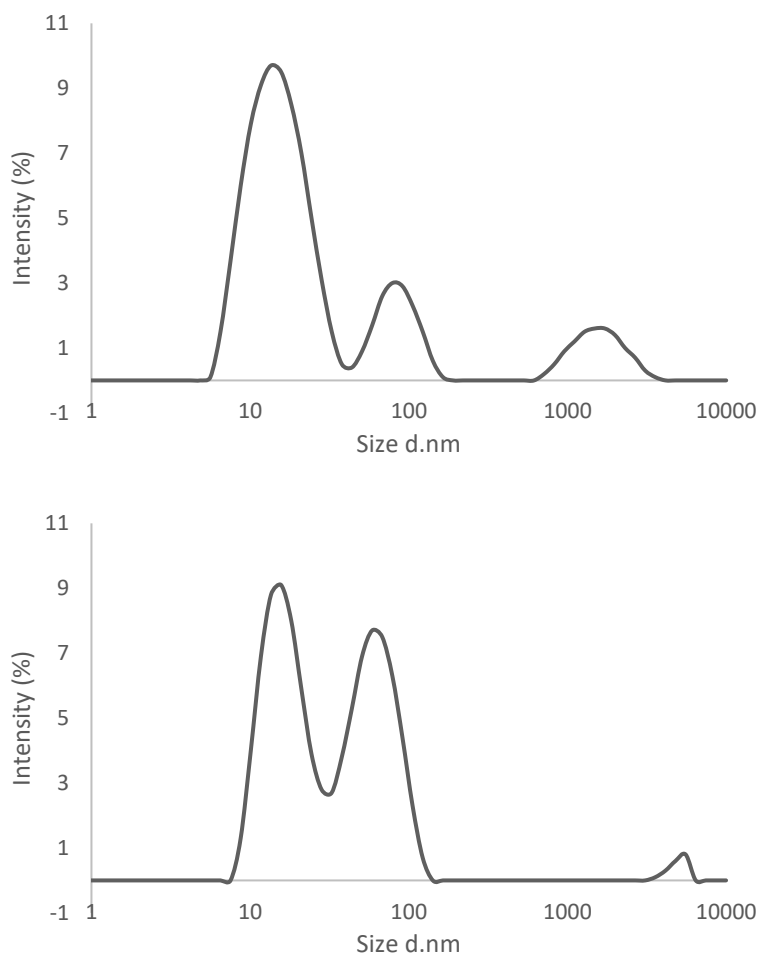
Figure 33. The effect of 5% v/v ethanol on the oligomeric state of GF. DLS measurements show the hydrodynamic size of particles distributed by intensity of DDM-solubilized GF in Buffer A1 and 5% EtOH v/v. **Top:** DDM-solubilized GF in Elution Buffer and 5% EtOH showing the presence of three different sized species, though most of the particles appear to be GF tetramer. **Bottom:** the sample from (top) was stored for 24 hours at room temperature, resulting in a greater abundance of aggregates and the appearance of visible precipitate in the sample.

Adenosine triphosphate (ATP) is a putative biological hydrotrope[289]. Hydrotropes

are small amphiphilic molecules that work in molar concentrations to solubilize

hydrophobic molecules[289]. Addition of 12 mM ATP to the SEC-MALS elution buffer

resulted in a major peak corresponding to LMNG-solubilized GF tetramer, but it was not

well resolved from a broad peak arising from multiple GF higher molecular weight

oligomers (Figure 34). Elution profiles of the protein dissolved in LMNG in the absence

of ATP are shown below.
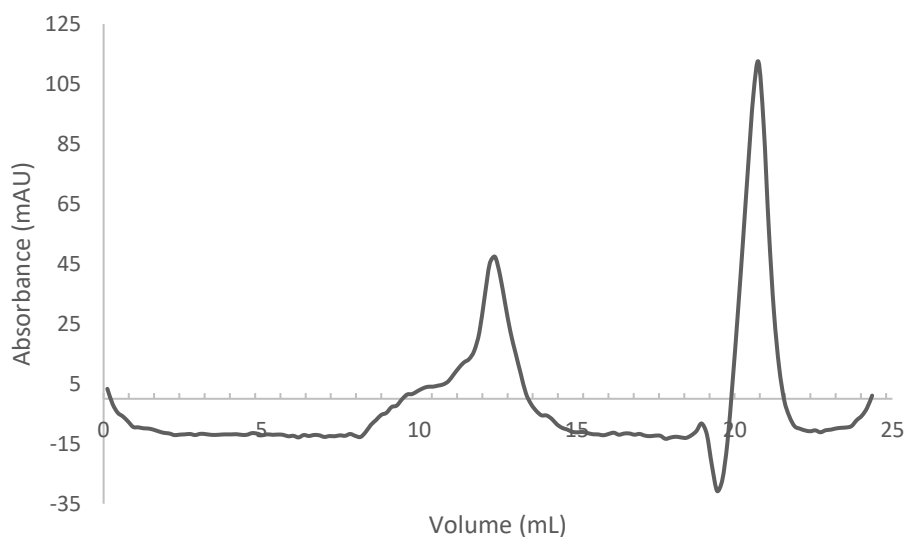
Figure 34. *The effect of 12 mM ATP on the oligomeric state of GF. The SEC-MALS elution profile of LMNG-solubilized GF in Buffer A1 and 12 mM ATP suggests a small tetrameric peak and higher M_r oligomer shoulder. The peak eluting at 21 mL is likely ATP.*

Nickel is easily stripped from the IMAC column resin with ethylenediaminetetraacetic acid (EDTA) when the resin is being regenerated. It is possible that some nickel could elute with GF as it is being eluted from the IMAC column. This might cause nonspecific aggregation of GF, possibly by association of the His_6 tags on the protein. Thus, 1 mM EDTA was added to the IMAC-eluted protein samples to remove any nickel bound to the His_6 tag and reduce putative tetramer-tetramer association and further aggregation. The SEC-MALS elution profile in Figure 35 shows a single tetrameric peak for freshly prepared GF; however, high molecular weight aggregates and some dissociated species are observed in the SEC-MALS elution profile of the same sample 24 hours later (Figure 35 top and bottom, respectively). The GF tetramer contains two putative cation binding sites at the 4-fold symmetry axis in the centre of the tetramer[85] and it is conceivable that removal of bound metal ions by EDTA enhances the dissociation of the tetramer.
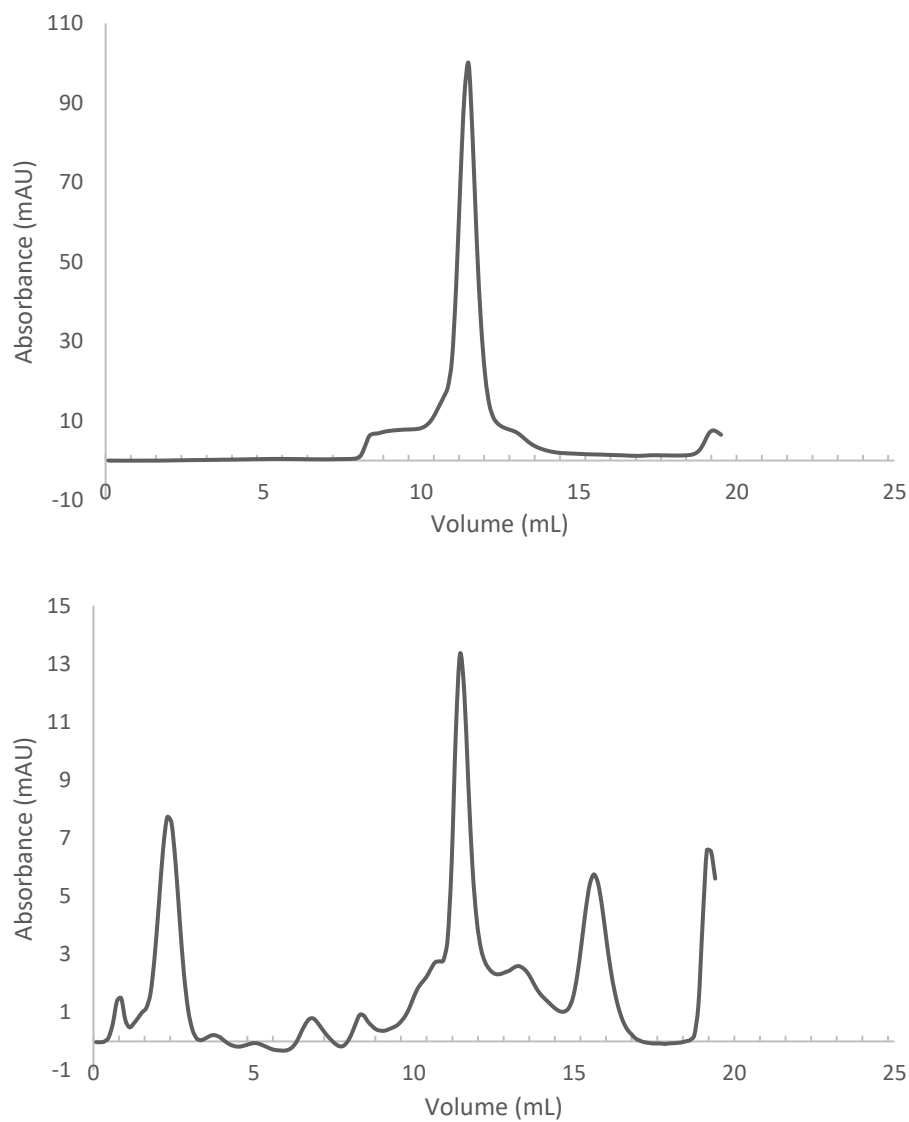
Figure 35. *The effect of 1 mM EDTA on the oligomeric state of GF. The SEC-MALS elution profiles of DDM-solubilized GF in Buffer A1 and 1 mM EDTA.* **Top:** *the most concentrated GF fraction freshly purified by IMAC elutes as a single, tetrameric peak.* **Bottom:** *the sample from (top) stored at room temperature for 1 day and gel filtered. Based on the multiple peaks, the elution profile shows that the sample has associated to form higher M_r oligomers and aggregates as well as less associated forms such as monomer and dimer.*

Affinity tags like poly-His tags used in the purification of proteins are often cleaved in a separate purification step. The cysteine protease tobacco etch virus (TEV) protease cleaves its target amino acid sequence Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser) between the Gln and Gly/Ser residues with high specificity[290]. Thrombin is a serine protease that preferentially cleaves Arg/Lys-Xaa in amino acid sequences A-B-Pro-Arg/Lys-A'-B', where A and B are hydrophobic amino acids and A', B' are nonacidic amino acids and A-Arg/Lys-B, where A or B are Gly[291]. Poly-His tags on target proteins are often followed by a TEV or thrombin cleavage site to enable specific removal of the affinity tag. TEV is favoured due to its high specificity for its amino acid sequence but has been shown to be less active in the presence of detergents, which are necessary when studying IMPs. Thrombin has been shown to remain active in the presence of 94 detergents and would be amenable to His-tag cleavage in GF, given that there is only one cleavage site accessible in the sequence[292]. However, cleavage of affinity tags on membrane proteins often fail when the tag is close to the membrane-spanning domain of the protein[293].

In early work with a thrombin-cleavable His$_6$-tag, no attempt was made to remove the tag owing to the low protein yields and the fact that others in the group had tried and failed previously. The codon-optimized GF construct containing a TEV-removable His$_6$-tag gave significantly increased yields and attempts were made to remove the His$_6$ tag to reduce self-association and aggregation. However, His$_6$-tag cleavage was unsuccessful

possibly due to the low activity of TEV protease in the presence of detergents or owing to the inaccessibility of the cleavage site.

Sample Storage

Temperature

To explore the effects of temperature on GF oligomerization, DDM-solubilized GF was stored at room temperature, at 4°C, or at -80°C. The polydispersity of samples was monitored over time *via* DLS (Figures 36 and 37). After thawing, the samples stored at -80°C showed visible aggregate and no further measurements were made. DLS measurements on all remaining samples revealed some level of soluble aggregate, but suggested that storing GF at 4°C slightly improved GF stability by slowing tetramer-tetramer association over a period of 10 days.
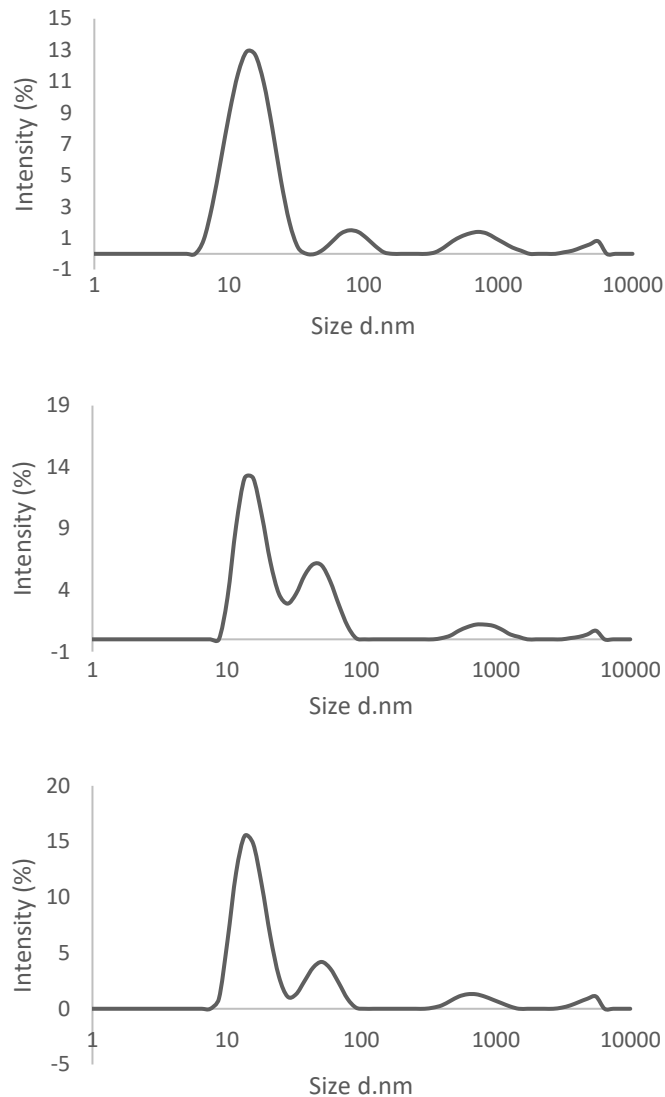
Figure 36. *The effect of storage temperature on the oligomeric state of GF after 24 hours.* **Top:** *freshly purified DDM-solubilized GF in Elution Buffer showing three different-sized species in solution, where the most intense peak just above 13.5nm is likely the GF tetramer, the peak at 78.8 nm is an aggregate, and the peak at 712 nm is likely rod-like DDM micelles.* **Middle:** *the same sample from (top) stored for 24 hours at room temperature; visible precipitate is observed in the sample and the fraction of aggregated species larger than the tetramer has significantly increased.* **Bottom:** *the same sample from (top) stored for 24 hours at 4°C. The sample shows only a small increase in higher diameter particles compared to the freshly prepared sample from (top).*
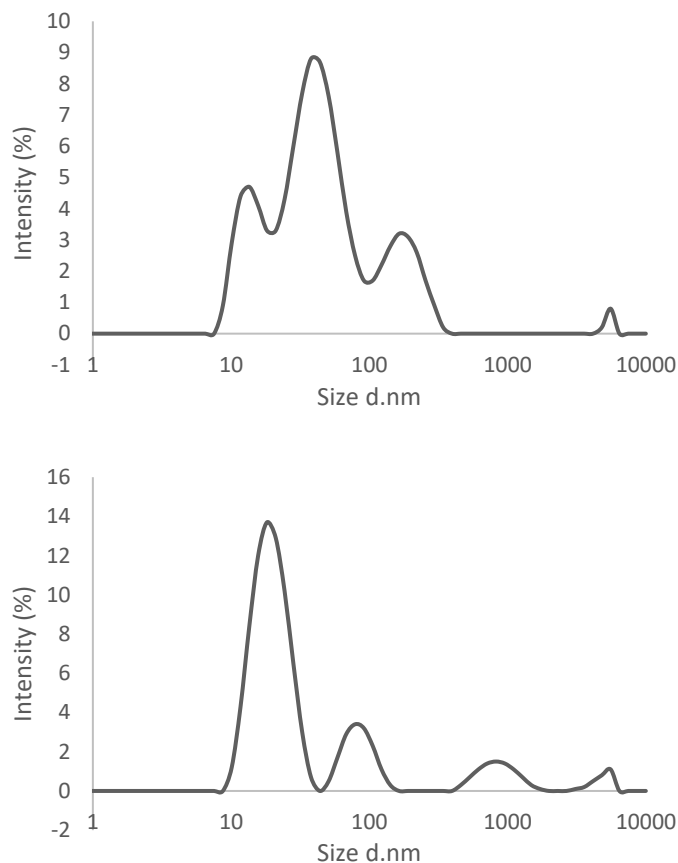
Figure 37. *The effect of storage temperature on the oligomeric state of GF after 10 days. DLS measurements show the hydrodynamic size of particles distributed by intensity.* **Top:** *compared to Figure 36 (top), DDM-solubilized GF stored at room temperature for 10 days shows a diminished tetramer peak (13.5 nm) and two peaks at 43.8 nm and 190.1 nm with significantly elevated intensity at higher apparent diameters, likely corresponding to different higher M_r oligomers.* **Bottom:** *the sample was stored for 10 days at 4°C. The size of the most intense peak (18.2 nm) suggests significant GF self-association although the most abundant aggregate is significantly smaller than that observed in Figure 36 (top), suggesting that perhaps storage at 4°C leads to reduced GF aggregation compared to that at room temperature.*

Highly concentrated DDM-solubilized GF samples (approximately 100 μM or more) were observed to visibly precipitate one hour after elution from the IMAC column. Swift aggregation suggested that there was insufficient detergent in the sample to adequately solubilize GF, leading to self-association and formation of octamers and higher M_r oligomers. To combat this, samples were diluted 1/10 with Buffer A1 (no imidazole) and 0.5 mM DDM and stored at room temperature (Figure 38) and 4°C (Figure 39). DLS measurements of samples diluted 1/10 showed no visible aggregate if the sample was diluted immediately following elution from the IMAC column, compared to undiluted samples, where slow self-association to octamer and higher M_r oligomers over several days were observed (Figure 39C). In this experiment, based on the DLS measurements, storing the diluted samples at 4°C (Figure 39) showed little improvement of GF stability and solubility compared to storage at room temperature (Figure 38).
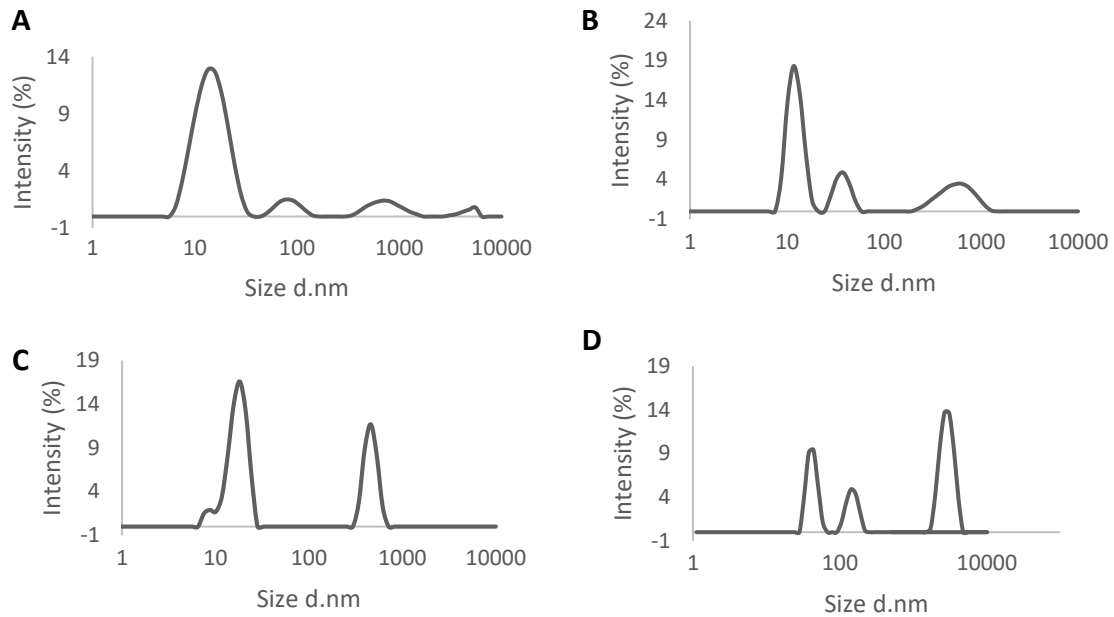
Figure 38. *The effect of protein concentration on the oligomeric state of DDM-solubilized GF in Elution Buffer or diluted 1 in 10 with Buffer A1 and stored at room temperature.* **(A)** *Freshly purified DDM-solubilized GF (110 μM) shows a sample that is predominantly tetrameric GF (13.5 nm) with a small amount of aggregate (91 nm) and some rod-like micelles (712 nm).* **(B)** *Sample (A) diluted 1/10 with Buffer A1 at room temperature contains a higher detergent:protein ratio, more rod-like micelles at 615 nm, and a slightly more intense aggregate peak at 38 nm.* **(C)** *Sample (A) stored for 24 hours at room temperature showing no aggregate peaks at 50 – 100 nm but a significant increase in the intensity of the presumptive rod-like micelle peak (459 nm). It is impossible to discount, however, that this peak reflects a GF-DDM aggregate. The small shoulder at 8.7 nm may reflect empty DDM micelles.* **(D)** *Sample (A) after storage for 6 days at room temperature. Interestingly, the tetramer has completely disappeared and is replaced by apparent 43 nm, 164 nm, and 3100 nm aggregates.*
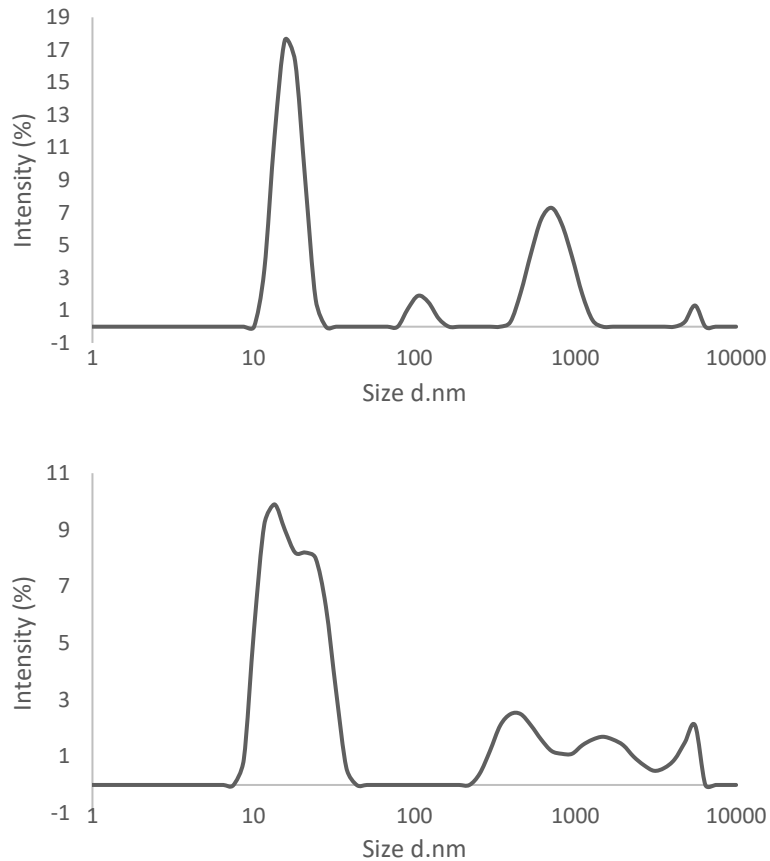
Figure 39. *The effect of protein dilution and storage at 4°C on the oligomeric state of GF. DLS measurements showing hydrodynamic size of particles distributed by intensity of DDM-solubilized GF in Elution Buffer, diluted 1 in 10 with Buffer A1, and stored at 4°C.* **Top:** *an aliquot of the sample from Figure 38B stored at 4°C for 24 hours shows the tetramer to have been replaced by a particle with an average apparent diameter of 15.7 nm, a small amount of 106 nm particles, and some particles with average diameters of 712 nm. There is an increase in aggregate peaks compared to Figure 38B.* **Bottom:** *the same sample from (top) stored for 6 days at 4°C; the relative intensity of the tetramer peak (13.5 nm) has diminished and the intensity of the aggregate peaks at 24 nm, 525 nm, and 1700 nm have increased compared to 38A and 37B.*

3.2.2 Stabilizing GF in Aqueous Solution

Historically, detergents have been used to extract, solubilize, and stabilize IMPs[294]. However, detergents that are proficient at extracting IMPs from the membrane may not necessarily be amenable to keeping the IMP of interest soluble for an extended period of time. Many studies published in the literature report extraction with one detergent and exchange into another detergent or other membrane mimetics such as liposomes, bicelles, nanodiscs, or polymers[295]. The following sections outline the various detergents used to extract GF from the membrane, and the subsequent detergents and other membrane mimetics tested to improve the solubility and stability of the GF tetramer.

GF Extraction

Octyl-glucoside (OG)

The GF homotetramer crystal structure was solved in octyl-β-D-glucoside[85], a non-ionic detergent with a monosaccharide head group and 8-carbon chain. OG is one of the more commonly used detergents used to extract IMPs from membranes[208]. However, in my hands, extraction of *E. coli* membranes with 20 mM OG for 3 hours at room temperature yielded insufficient amounts of purified GF for characterization studies.

SDS is a denaturing, anionic detergent with a 12-carbon tail that is commonly used to denature proteins for polyacrylamide gel electrophoresis. DDM-extracted GF tetramer is relatively resistant to dissociation by SDS, with most tetramer dissociating to monomers only above 40% w/v SDS[276]. I used 0.6% w/v SDS to extract GF from the *E. coli* membrane. The oligomeric state of SDS-solubilized GF was monitored by SDS-PAGE, suggesting that, in SDS, GF exists in mostly monomeric form, with minimal tetramer, trimer, and dimer bands. However, after 24 hours at room temperature, the samples developed a visible precipitate.

We speculated that GF monomers might be suitable for dynamics measurements *via* solution NMR spectroscopy as they are approximately 30 kDa, close to the traditional protein size limit for solution NMR spectroscopy[296]. Uniformly [15]N- labelled GF was prepared using LS-5052 to grow the cells. GF was extracted from the membranes and solubilized with SDS, then purified by IMAC. Heteronuclear single quantum coherence (HSQC) spectra were acquired on the SDS-solubilized GF monomers (Figure 40). The spectrum shows on the order of about 16 intense cross-peaks, another 10 that are weaker and a broad mass of peaks that are poorly resolved. Visible in the spectra are peaks corresponding to Asn, Gln and Arg side-chains. Two likely Gly backbone amide resonances are observed as well as 3 or 4 resonances that are likely Ser/Thr. About 11 other backbone amide resonances are resolved but not identifiable. The resolved backbone amide resonances likely belong to the flexible N- and C-termini of the protein

that are expected to have rapid rotational correlation times owing to a lack of structure.

In the X-ray diffraction structure[85], 22 residues in the C-terminus and 5 in the N-terminus

are not observed, presumably owing to conformational averaging. In addition to these, we

might also expect to observe 19 residues in the disordered N-terminal His_6-tag. Altogether

about 20 resonances from the disordered termini are not observed and these may

constitute the remainder of the very broad resonances that are observed in the spectrum.

The rest of the protein resonances are not observed, likely because of a combination of

line broadening owing to conformational exchange, detergent exchange, hydrogen

exchange, and oligomer association. After 10 hours of acquisition, precipitate was visible

in the sample and no further attempts at solution NMR spectroscopy were made.
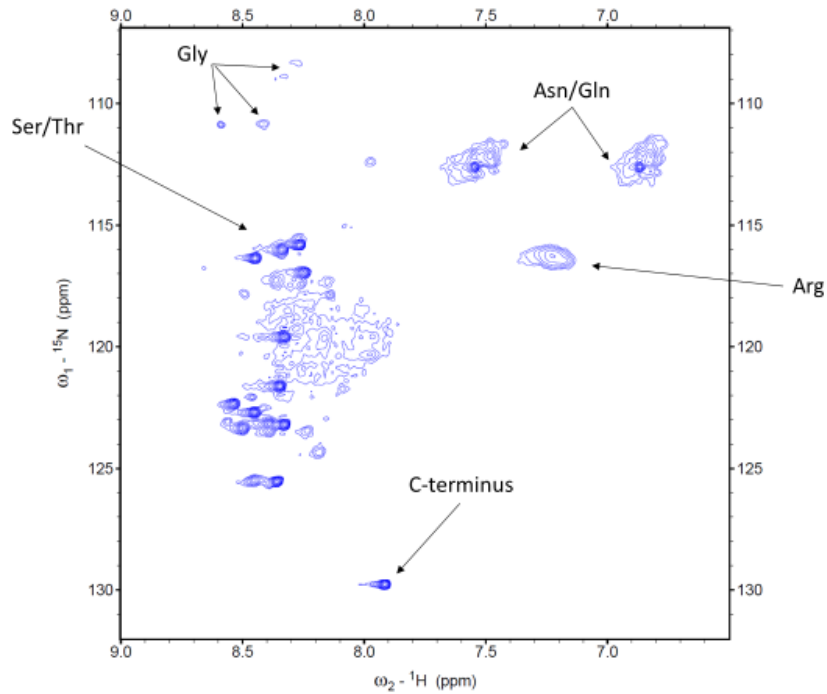
Figure 40. *HSQC spectrum of SDS-solubilized ^{15}N-labelled GF monomers with several residue-type assignments. The narrow chemical shift range suggests a disordered conformation for the observable resonances.*

Dodecyltrimethylammonium Bromide (DTAB)

DTAB is a cationic detergent with a positively charged head group and a 12-carbon tail. Twenty mM DTAB was used to extract GF from the membrane but it resulted in poor protein yields and no further characterization was done.

Lauryl Maltose Neopentyl Glycol:Cholesteryl Hemisuccinate (LMNG:CHS)

Recent publications[297,298] reported using a mixture of LMNG and CHS, a cholesterol mimetic, to extract and purify an ion channel and GPCR and solve their structures *via* cryo-EM. Extraction of GF from the membrane using 5% LMNG and 0.5%

CHS resulted in poor protein yields and no further characterization of the protein was done.

Dodecyl Maltoside (DDM)

DDM is the most common detergent used to extract IMPs; approximately 30% of unique IMP structures deposited into the PDB in 2016, 2017, and 2018 were extracted from the membrane using DDM[124]. DDM is a non-ionic and gentle detergent, which consists of a maltose head group and 12-carbon tail. DDM has been used throughout this work and, compared to other detergents tested here, was observed to be the best detergent to use to successfully extract GF from the membrane of *E. coli*. Protein yields of DDM-solubilized GF were consistently 30-40 mg/L of cell culture in Studier's autoinduction media.

SMA 2K

The SMA 2K co-polymer (Figure 41) comprises a 2:1 ratio of styrene to maleic acid and is the recommended starting SMA polymer to use for initial compatibility tests with the protein of interest. Purification follows *via* the affinity tag on the IMP of interest. SMALP (Styrene Maleic Acid Lipid Particle)-protein solubilization is indicated when the initially turbid SMA/membrane mixture becomes translucent. Multiple extractions and solubilization of GF using SMA 2K were attempted, with no visible change in sample turbidity; incubation times, SMA:membrane (w/w) ratios, and other parameters, such as

different buffer conditions, were adjusted to encourage GF solubilization *via* SMALPs

with little success.
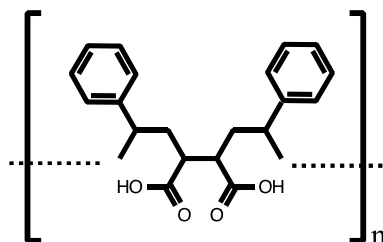
Figure 41. *The structure of the styrene maleic anhydride co-polymer. Styrene maleic anhydride lipid particles (SMALPs) are used to solubilize IMPs in their native lipid environment*[214].

After extraction from the membrane, GF was solubilized in a variety of detergents and other membrane mimetics to improve the stability of the GF tetramer and slow the time-dependent self-association forming octamer and higher M_r oligomers. Most detergent exchange was done at the IMAC step in the purification process, where GF was incubated with the new detergent while bound to the resin before elution. In some cases, detergent exchange was done during SEC. In the case of bicelles, liposomes, and nanodiscs, GF was eluted from the resin, incubated with the respective membrane mimetic, then tested for incorporation. In the following sections, GF was first extracted from the membrane with DDM and exchanged into another detergent or solubilizing agent.

Detergents

DDM

Compared to other detergents tested here, GF extraction from the membrane with DDM yielded the most protein. In addition to successfully extracting GF with DDM, GF solubilized in DDM was optimized and we obtained a pure, homogenous tetramer after elution from the IMAC column (Figure 29, top). Though DDM appeared to be a suitable detergent, DDM-solubilized GF samples showed visible precipitate over hours to several days after purification, and SEC-MALS elution profiles showed octamer formation after

one day (Figure 29, bottom). To improve the solubility and stability of GF, several detergents and membrane mimetics were explored.

Lauryldimethylamine-N-oxide (LDAO)

LDAO is a zwitterionic detergent with a charged head group and a 12-carbon chain. LDAO was used to solubilize the photosynthetic reaction center in the first membrane protein crystal structure ever solved[72]. GF was extracted with DDM and purified on an IMAC resin followed by exchange of DDM for LDAO during SEC-MALS. After exchange of LDAO for DDM, visible precipitate was observed in the sample and it was not further characterized.

Lysomyristoylphosphatidylcholine (LMPC)

LMPC is a zwitterionic lysophospholipid. It was previously shown that LMPC-solubilized GF is more thermally stable than DDM-solubilized GF[276]. The SEC-MALS elution profile of GF extracted with DDM and solubilized with LMPC shows a single peak corresponding to tetrameric GF (Figure 42, top). Following storage at room temperature for three days, the LMPC-solubilized GF sample showed visible precipitate and the SEC-MALS elution profile showed peaks corresponding to tetramer, octamer and higher M_r oligomers (Figure 42, bottom).
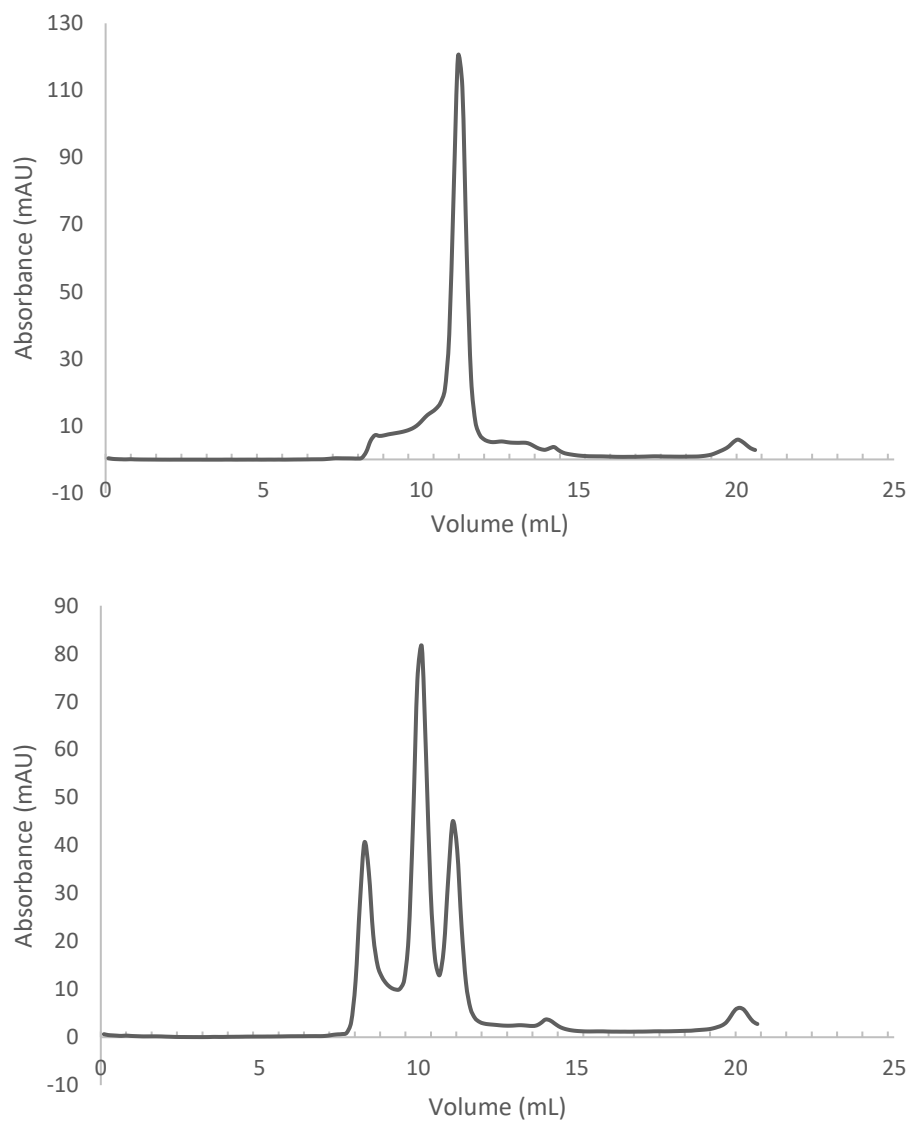
Figure 42. The effect of lysophospholipid LMPC on the oligomeric state of GF. The SEC-MALS elution profile of GF extracted with DDM and exchanged into 5 mM LMPC. **Top:** freshly prepared LMPC-solubilized GF elutes as a tetramer. **Bottom:** the sample stored at room temperature for 3 days shows tetramer, octamer and larger aggregates.

Dodecyltrimethylammonium Bromide (DTAB)

DTAB has a positively charged head group and has a carbon chain the same length as DDM and was chosen to monitor the effect of a cationic detergent on the self-association of GF tetramers. GF required a larger volume of DTAB-containing elution buffer to elute completely from the IMAC column, suggesting that GF is poorly solubilized in DTAB than in DDM. The SEC-MALS elution profile in Figure 43 suggests that DTAB-solubilized GF exists as a mixture of dissociated monomers, dimers and trimers.
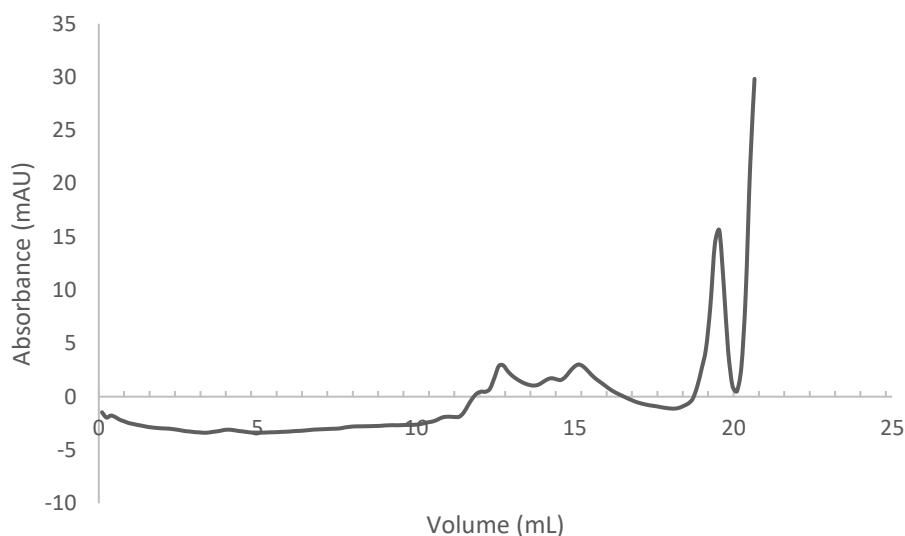
Figure 43. The effect of the cationic detergent DTAB on GF solubility. The SEC-MALS elution profile of GF extracted with DDM and exchanged into 3 mM DTAB suggests that DTAB-solubilized GF exists as a mixture of dissociated monomers, dimers and trimers.

Lauryl Maltose Neopentyl Glycol (LMNG)

LMNG is a member of the MNG (maltose neopentyl glycol) class of surfactants (Figure 44)[299]. MNGs have a different structure from detergents in that they contain two hydrophilic heads and two hydrophobic tails grafted onto a neopentyl glycol core. NMGs form large micelles and have relatively low CMCs. Freshly purified LMNG-solubilized GF elutes from the SEC-MALS column predominantly as a tetramer with a small octamer shoulder and a small amount of higher M_r aggregates (Figure 45, top). This sample was stored for 9 days at room temperature; no visible precipitate was observed, and the SEC-MALS elution profile of the stored sample (Figure 45, bottom) is remarkably similar to that of the freshly purified GF sample, showing a small increase in the fraction of octamer (Figure 45, top). This sample was by far the most stable of all of my GF preparations, and these results were highly reproducible over many repetitions. However, although the protein in LMNG is highly soluble over long periods of time, the existence of two oligomeric forms might be problematic for NMR analysis.
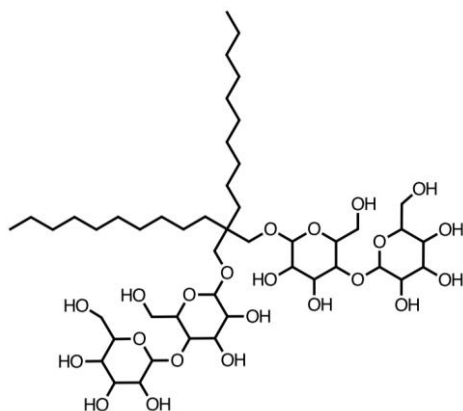
Figure 44. *The structure of LMNG*[299].

Figure 45. The effect of LMNG on the stability and solubility of GF. The SEC-MALS elution profile of GF extracted with DDM and exchanged into 0.05% w/v LMNG. **Top:** freshly prepared LMNG-solubilized GF shows a mostly tetrameric protein (peak b) with a small octamer shoulder (a). Note that the elution position of the GF oligomers is not the same as that observed in DDM. The identity of the octamer and tetramer peaks was established by other methods and is described below. **Bottom:** the sample from (top) stored for 9 days at room temperature shows a slightly increased ratio of octamer (peak c) to tetramer (peak d), suggesting that LMNG stabilizes GF better in solution than DDM, although in DDM the tetramer slowly converts to octamer.

To further probe the enhanced stability of LMNG-solubilized GF, differential scanning calorimetry (DSC) was used to measure melting temperatures. DSC measures the change in heat released and absorbed in a sample system relative to a buffer blank[300]. DSC measurements suggest that the melting temperature of DDM-solubilized GF is approximately 76°C, while LMNG-solubilized GF melts at 84°C (Figure 46 top and

148

bottom, respectively). The higher melting temperature observed in LMNG further supports the superior stabilizing effect of LMNG over DDM.
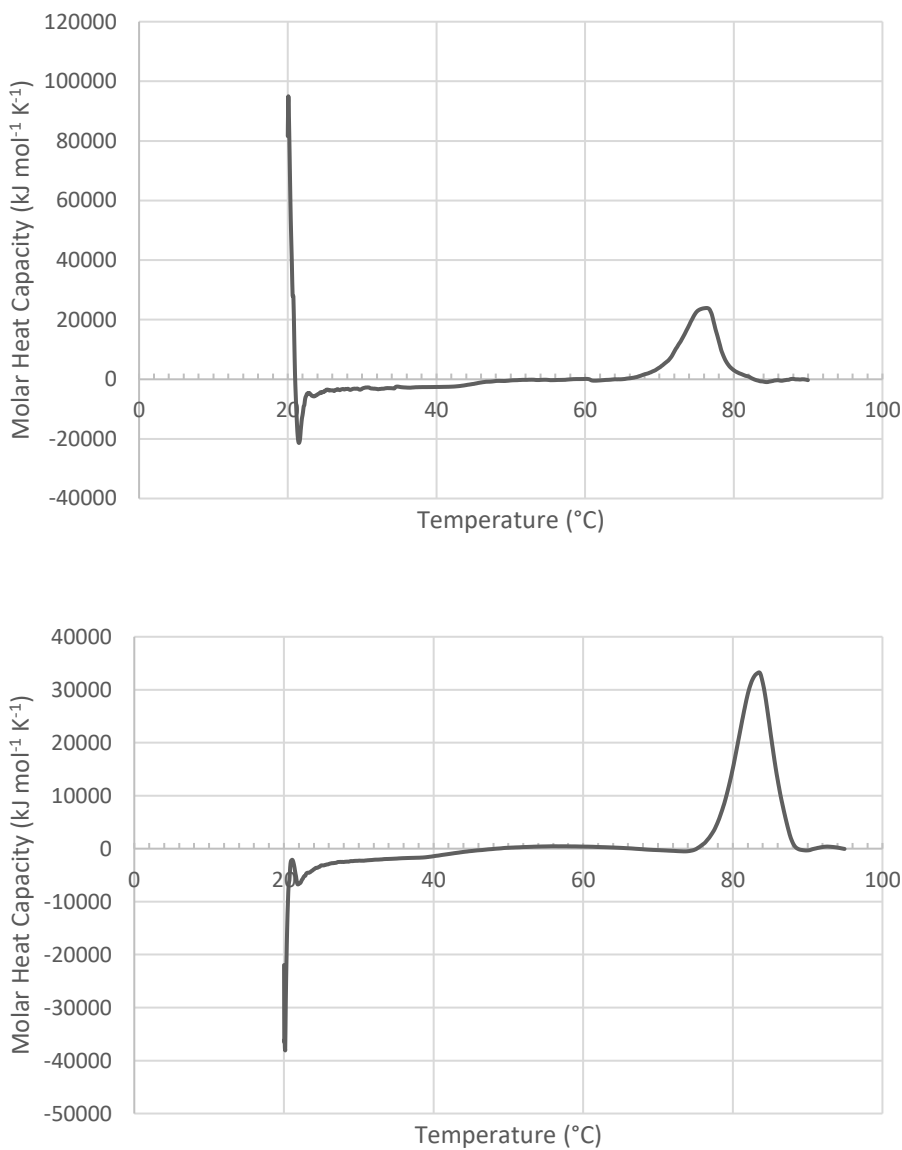
Figure 46. *The effect of LMNG on the melting temperature of GF. Differential scanning calorimetry (DSC) measures the thermal melting of proteins.* **Top:** *the baseline-corrected thermogram of DDM-solubilized GF reveals a melting temperature of approximately 76°C.* **Bottom:** *the baseline-corrected thermogram of LMNG-solubilized GF shows a melting temperature of approximately 84°C, suggesting that GF is more stable when solubilized in LMNG than in DDM.*

Initial extraction using LMNG:CHS and subsequent purification of GF resulted in a single tetramer peak on the SEC-MALS elution profile, though poor extraction efficiency was noted as protein yields were only approximately 4 mg/L. Extracting first with DDM and exchanging to LMNG:CHS reproducibly yielded a large amount of LMNG:CHS-solubilized GF. These samples, however, were polydisperse as indicated by SEC-MALS (Figure 47). Since no improvement was observed from adding CHS, it was omitted for succeeding preparations.
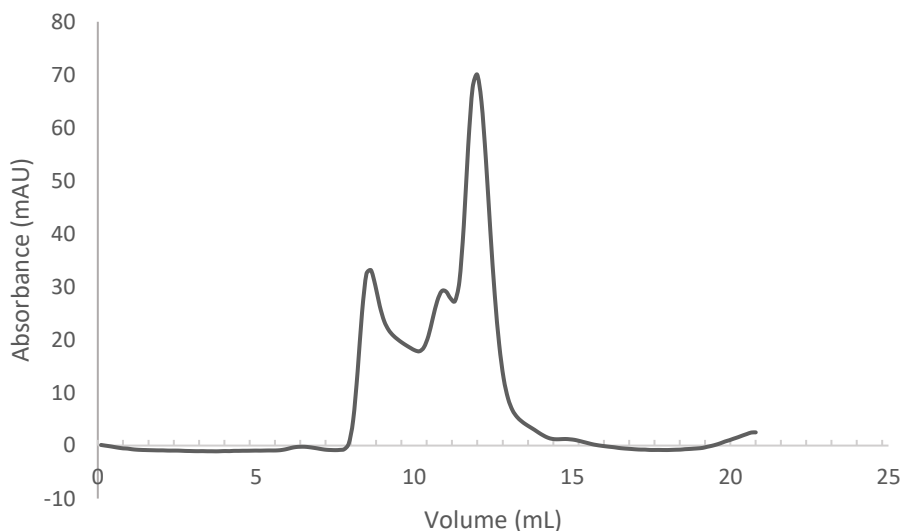
Figure 47. *The effect of LMNG:CHS on the solubility of GF. The SEC-MALS elution profile of GF extracted with DDM and solubilized in 0.05% LMNG:0.005% CHS w/v indicates that the sample is polydisperse containing GF tetramer, octamer and other high M_r aggregates.*

Decyl Maltose Neopentyl Glycol (DMNG)

DMNG is also a member of the MNG class of detergents. DMNG carbon chains are two carbons shorter than those in LMNG, and DMNG has been shown to be more effective than LMNG at extracting IMPs[301]. The SEC-MALS elution profile of GF solubilized with 0.5 mM DMNG showed a major peak for tetramer but in addition showed multiple higher and lower M_r peaks and therefore was not considered suitable for further characterization (Figure 48).

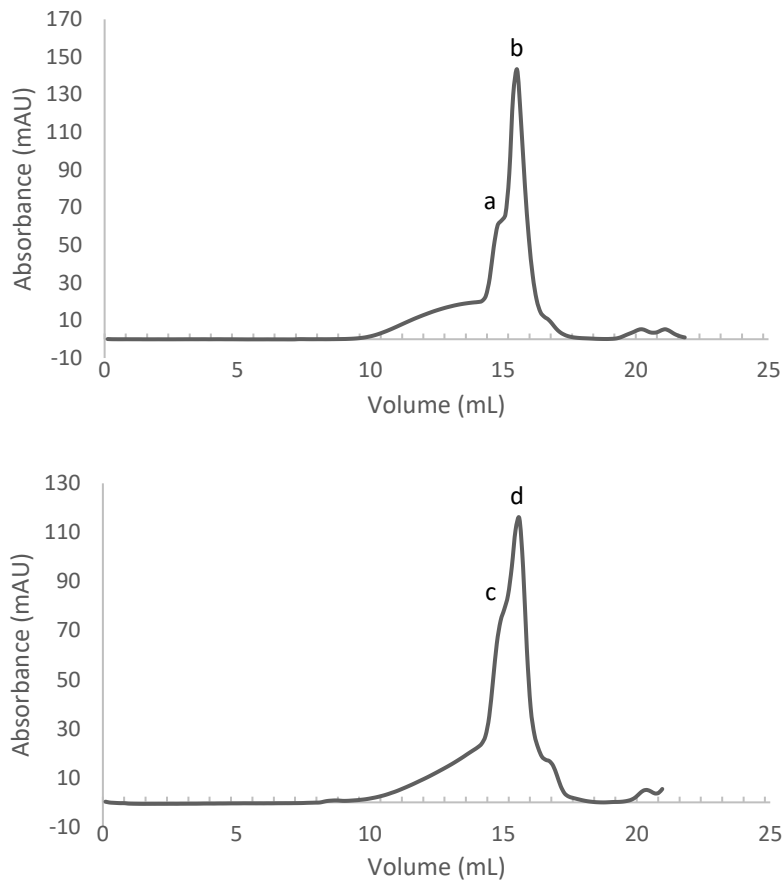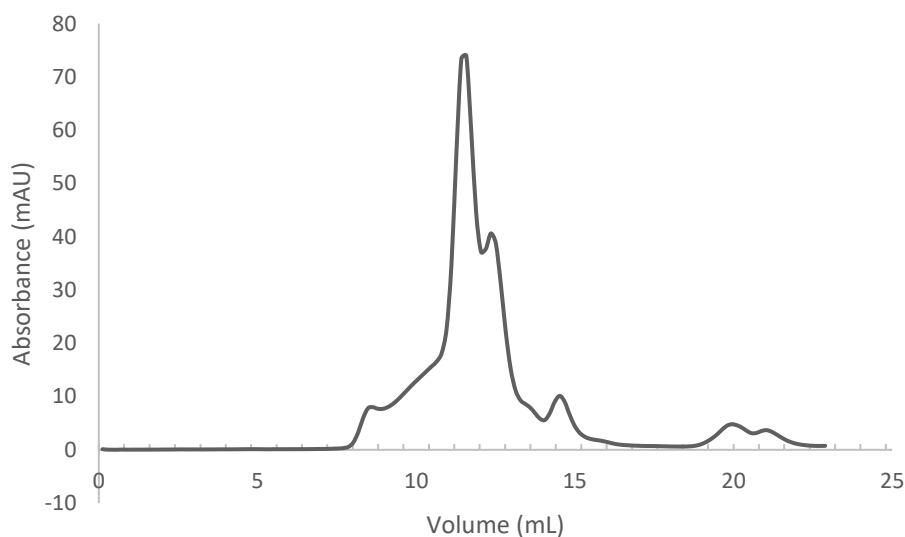*Figure 48.* *The effect of DMNG on the solubility and stability of GF. The SEC-MALS elution profile of GF extracted with DDM and solubilized in 0.5 mM DMNG shows multiple peaks indicative of a polydisperse sample.*

Other Membrane Mimetics

The use of detergents and their capacity to provide a physiologically relevant environment for IMPs has always been a concern for membrane protein researchers. Though the ease with which detergents extract and solubilize IMPs from the membrane makes them advantageous, the environmental disparity between a detergent micelle *in vitro* and the native lipid membrane *in vivo* may greatly affect IMP folding and stability, and thus the overall structure and function of proteins of interest. The following sections report the results of my attempts to incorporate GF into several membrane mimetic systems used commonly in and emerging from the literature.

Liposomes

Liposomes are an attractive membrane mimetic as they provide a lipid bilayer environment that can comprise a similar mixture of phospholipids as the native environment for the protein of interest. Due to their large size and viscosity, liposomes are not amenable to solution NMR spectroscopy. However, significant progress in solid-state NMR spectroscopy[146] makes them a viable membrane mimetic for IMP studies. A native *E. coli* lipid mix (Avanti Polar Lipids) was used here to form liposomes. DDM-solubilized GF was reconstituted into small, unilamellar lipid vesicles (SUVs) composed of native *E. coli* lipids. SDS-PAGE suggested that GF tetramers, octamers and a higher M_r aggregates were incorporated into or associated with the SUV bilayers. No further attempts to characterize the proteoliposomes due to the presence of large M_r aggregates.

Like liposomes, bicelles are an appealing membrane mimetic as they also provide a bilayer environment for IMPs extracted from the native lipid membrane[226]. Bicelles differ from liposomes in that they comprise a mixture of short-chain and long-chain lipids. The long-chain lipids self-associate in aqueous solution to form a bilayer disc, which is solubilized by a short-chain belt covering the exposed lipid hydrocarbon. The ratio of long to short chain lipid is the q ratio, and is used to classify the bicelles[230]. In addition, the radius of the bilayer disc can be directly calculated from the q. Larger q ratios correspond to larger, anisotropic bicelles and are generally considered more biologically relevant. A q ratio of 4 has been suggested as optimal for ssNMR spectroscopy[302]. Small q ratios yield isotropic, fast tumbling discs that are better suited for small IMPs studied by solution NMR and a q = 0.5 has been suggested to be optimal for that purpose[230].

DDM-solubilized GF was incubated with empty preformed DMPC/DHPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine/1,2-dihexanoyl-*sn*-glycero-3-phosphocholine) bicelles with q ratios of 4 and 0.5 to test for their suitability for solid-state and solution NMR measurements, respectively. Incorporation of DDM-solubilized GF into bicelles with q ratios of 4 and 0.5 both resulted in visible aggregate after incubation for 30 minutes on ice. These aggregates might be suitable for ssNMR spectroscopy, but they would not be suitable for solution NMR studies and were not further characterized.

Amphipols (APols) are effective at solubilizing IMPs but cannot be used to extract IMPs from the membrane[238]. In the absence of detergent, they bind tightly to IMPs in aqueous buffers. Above their relatively low CMC (0.002% w/v for A8-35), APols form micelles, though in the presence of an IMP most amphipols will be bound to it. Amphipol A8-35 was tested for its ability to solubilize DDM-extracted GF. SEC-MALS measurements were made using a running buffer without amphipol or detergent. The elution profile shown in Figure 49 shows two maxima, suggesting that at least two species are present in the amphipol-solubilized mixture. Although the SEC column was not calibrated in the presence of amphipol, the peak positions suggest a mixture of unresolved GF octamer and tetramer as well as a higher M_r oligomer. The breadth of the peak may indicate multimer equilibrium or that the column was overloaded. Since the SEC-MALS elution profile of the amphipol-solubilized GF (Figure 49) was difficult to interpret, SDS-PAGE was used to assess the GF oligomer state. Figure 50 shows an SDS-PAGE electrophoregram of one-week-old DDM- and amphipol-solubilized GF. It indicates that the amphipol-solubilized samples exhibit higher amounts of tetramer than octamer, which supports the above interpretation of the SEC-MALS elution profile. In contrast, the DDM-solubilized protein shows higher amounts of octamer than tetramer. High M_r aggregates that did not enter the stacking gel are not shown. After one week at room temperature, amphipol-solubilized samples did not show any visible precipitate.
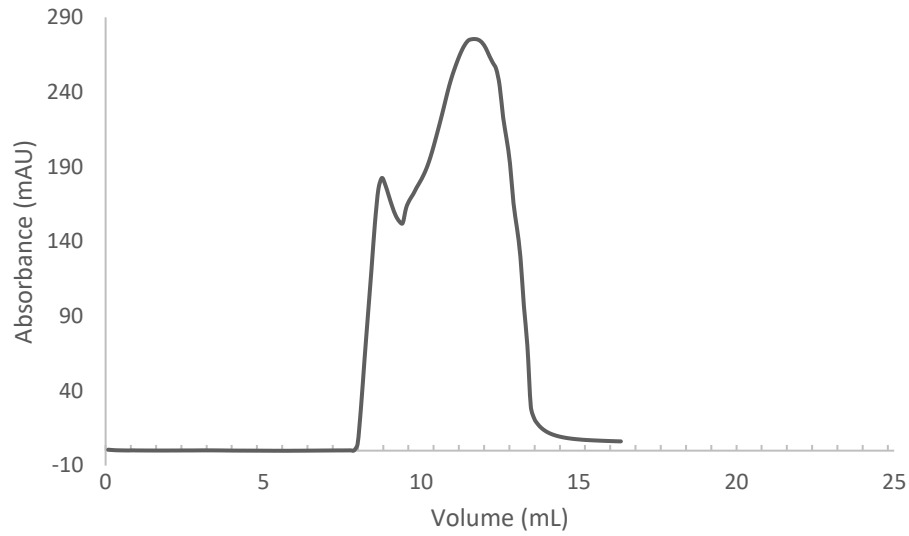
Figure 49. *The effect of amphipols on the solubility and stability of GF. The SEC-MALS elution profile of GF extracted from DDM and exchanged into amphipol A8-35 at 1.5 mg amphipol:1 mg GF. The broad peak observed suggests multimer formation during elution.*
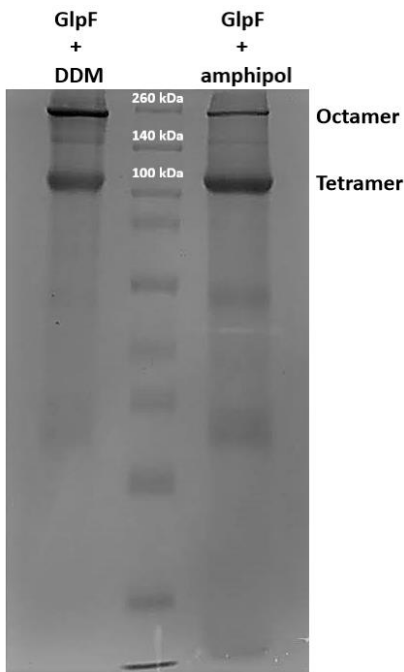
Figure 50. *SDS-PAGE gel electrophoregram of 7-day old samples of DDM- and amphipol-solubilized GF. Both protein samples show some octamer and tetramer, though based on the intensity and thickness of the bands, amphipol-solubilized GF appears to remain mostly tetrameric over one week. DDM-solubilized GF appears to associate to octamers over the same period. Not shown are the bands corresponding to larger aggregates that did not enter the running gel.*

155

Poly (Maleic Anhydride-alt-1-Decene) substituted with 3-(Dimethylamino) Propylamine (PMAL)

PMALs are a zwitterionic class of amphipol that differ from the classic amphipols by consisting of alternating polar/nonpolar subunits consisting of carboxyl, ammoniumamidate and C8, C12, or C16 alkyl chains (Figure 51). PMAL C-8 contains C8 alkyl chains. PMALs have been shown to solubilize enzymatic IMPs that retain their enzymatic activity, whereas the same enzyme solubilized in amphipol A8-35 had not been explicitly demonstrated to retain enzymatic activity[235]. PMALs have also been shown to be able to insert IMPs into a membrane without disrupting it[234]. In this work, PMAL C-8 was used to solubilize GF. I was unable to characterize PMAL-solubilized GF by SDS-PAGE, as the PMAL appeared to bind to the Coomassie present in the SDS-PAGE staining solution used to visualize protein bands in the gel (Figure 52). PMAL also appeared to absorb at 280 nm, preventing accurate determination of protein yield. SEC-MALS was not attempted with the PMAL-solubilized GF sample due to the polymer absorbance at 280 nm.
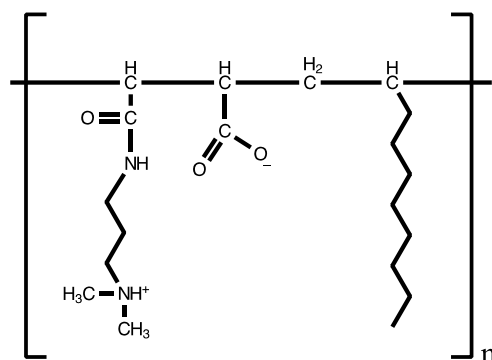
Figure 51. *Structure of PMAL B-100. PMAL B-100 is comprised of zwitterionic subunits with varying chain lengths. PMAL C-8 denotes PMAL B-100 subunits with an 8-carbon chain. With permission from Gorzelle et al.*[235]
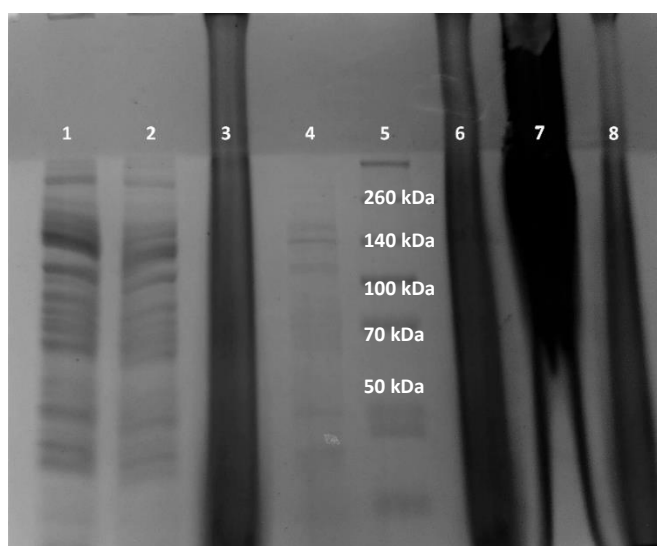
Figure 52. *SDS-PAGE electrophoregram of the different purification steps for PMAL-solubilized GF. Lane 1: DDM-solubilized E. coli membranes. Lane 2: IMAC column eluate during sample loading. Lane 3: Initial PMAL wash and detergent exchange step on the IMAC column (Buffer A1 containing 25 mg/mL PMAL to give a 5:1 w/w PMAL:GF ratio). Lane 4: DDM wash of the IMAC column, to remove loosely-bound contaminants (occurs before PMAL wash) Lane 5: Protein standard M_R ladder. Lane 6: PMAL-solubilized GF dialysed overnight to remove imidazole. Lane 7: Sample from Lane 6, concentrated. Lane 8: Highest absorbing fraction from IMAC.*

Nanodiscs were first described in the mid-2000s by Stephen Sligar *et al.*[239] Nanodiscs comprise a lipid bilayer encircled and solubilized by a membrane scaffold protein (MSP), an amphipathic helical protein. Sligar *et al.* were initially studying high-density lipoprotein (HDL) and its role in artherosclerosis[303] and engineered MSPs from human serum apolipoprotein A1, which were observed to form lipo-protein particles. The engineered MSPs wrap around a lipid bilayer in a belt-like configuration, forming a small, discoidal structure. The size of the disc can be adjusted by modifying the length of the MSP used to solubilize the bilayer. Recently, ssNMR spectroscopy was used to successfully distinguish small conformational differences between detergent-solubilized and nanodisc-solubilized IMPs[304,305].

Two different MSPs were tested for incorporation of GF into nanodiscs. Similar to the bicelle preparations of GF, incorporation was unsuccessful, as measured by DSC. Incorporation of GF into nanodiscs was measured by DSC on the expectation that nanodisc-incorporated GF would have an elevated thermal melting point. However, the DSC thermogram in Figure 53 shows that the melting temperature of GF in the presence of nanodiscs (MSP1E3D1) is the same as that measured in DDM. The fact that the melting points of GF and the MSP are the same in the nanodisc preparation for the individual proteins suggests that GF was not incorporated into the nanodiscs. It is possible that the GF tetramer may be too large to be incorporated into even the largest nanodisc, and

certainly the GF octamers are too large. The DSC thermogram of the empty MSP1E3D1

nanodisc is shown in Figure 54.
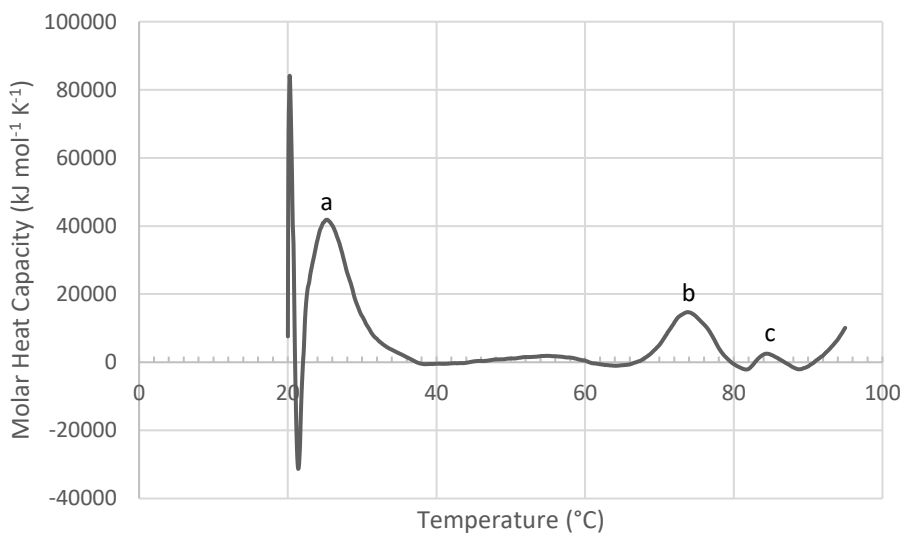
Figure 53. *Monitoring the incorporation of GF into nanodiscs. A DSC thermogram of DDM-solubilized GF after incubation in the presence of nanodiscs containing DHPC and cholate shows three peaks corresponding to the thermal melting points of the lipid (peak a, 26°C), DDM-solubilized GF (peak b, 75°C), and nanodisc scaffold protein MSP1E3D1 (peak c, 86°C). The unchanged melting point of GF suggests that incorporation of GF into the nanodiscs was unsuccessful.*
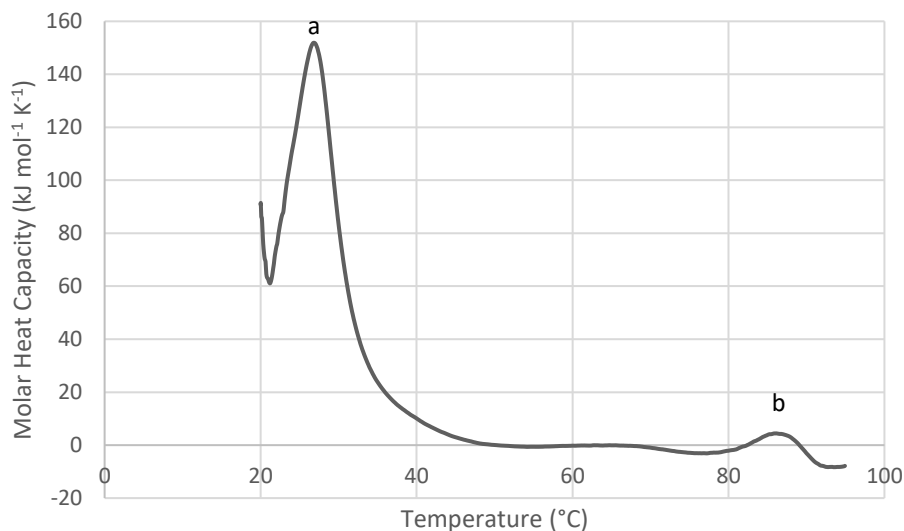
Figure 54. *DSC thermogram of empty nanodiscs formed with MSP1E3D1. Peak a is the thermal melting point of the lipid at approximately 26°C. Peak b is the melting point of MSP1E3D3 at approximately 86°C.*

Random Heteropolymers (RHP)

DDM-solubilized GF was incubated with 36 kDa average molecular weight RHPs. Enough material was available for exchange at the final IMAC elution step, DSC measurements, and SDS-PAGE. SDS-PAGE suggested that the 36 kDa RHPs solubilize GF, but GF appears to be present in solution as a tetramer, octamer, and higher M_r oligomers (Figure 55). After one week, RHP-solubilized GF samples show no visible aggregate. DSC measurements of RHPs in buffer A and RHP-solubilized GF were complex and inconclusive.
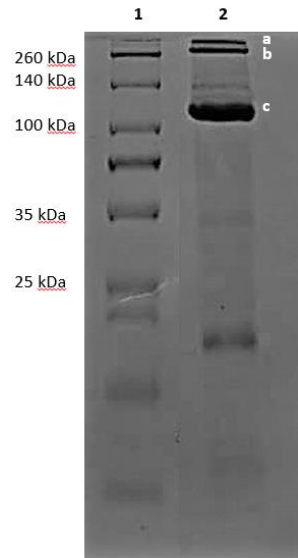
Figure 55. *Using RHPs to solubilize GF. SDS-PAGE electrophoregram of 36 kDa RHP-solubilized GF. Lane one is the standard protein ladder. Lane two is 36 kDa RHP-solubilized GF, which shows that the sample runs mostly as a tetramer (band c), though some octamer (band b) and higher M_r oligomers (band c) are also present.*

Mizrachi *et al.* were able to express and purify DsbB, a small, membrane-bound enzyme that catalyzes cysteine disulfide bonds in *E. coli* without added detergents or other solubilizing agents[306]. They expressed DsbB as a fusion protein with a highly soluble cytoplasmic protein, maltose binding protein (MBP), fused to the N-terminus of DsbB. On the C-terminus, a truncated version of human apolipoprotein AI (ApoAI*) was fused. MBP is a decoy protein, where expression of MBP preceding DsbB directs expression of the fusion protein into the cytoplasm, instead of directing its incorporation into the membrane. ApoAI* is a version of the amphipathic helix that is also used to solubilize nanodiscs; it shields the hydrophobic parts of DsbB from water and promotes solubilization. When expressed together, a water-soluble version of DsbB is produced in the cytoplasm. Some enzymatic activity was observed in DsbB after purification and cleavage of the N-terminal MBP.

I designed and expressed a pET28b(+) vector encoding a MBP-GF-ApoAI* fusion protein that was expressed in BL21(DE3) *E. coli* cells. SDS-PAGE analysis of IMAC eluted protein showed low M_R bands indicative of proteolytic cleavage (Figure 56). Addition of proteolysis inhibitors to the cell culture and buffers in the purification process may aid in reducing proteolytic cleavage, but no further optimization of this approach was undertaken.
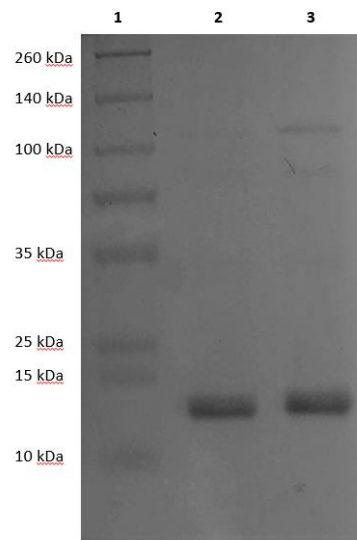
Figure 56. *The electrophoregram of the MBP-GF-ApoAI* fusion protein shows that the protein is likely proteolytically cleaved during the purification process. Lane one is the standard protein ladder. Lanes two and three are the two highest absorbing fractions at 280 nm.*

3.2.3 Structural Characterization of the GF Octamer

Following the exploration of solubilizing agents, temperature, pH, solubilizing conditions and protein constructs, it was determined that extraction of GF with DDM and solubilization of GF with LMNG in concentrations close to its CMC provided protein preparations that were the most stable and the least likely to precipitate even at high concentrations of protein. However, in LMNG, GF exists as a mixture of octamers and tetramers, so I next sought to characterize the octamer structure to identify the cause of the aggregation. Knowing which parts of the protein are interacting to form the octamer might enable the introduction of mutations that would eliminate the self-association.

Examination of the GF crystal structure symmetry mates reveals that the cytoplasm-facing surfaces of the tetramers face each other, suggesting that the interface between the tetramers is likely occupied by the disordered N- and C-termini (Figure 57). Although the disordered termini are not observed in the crystal structure, it is possible that they may be interacting to promote octamer formation. Negative stain Electron Microscopy (EM) and Small Angle X-ray Scattering (SAXS) were employed to elucidate the solution structure of the octamer.
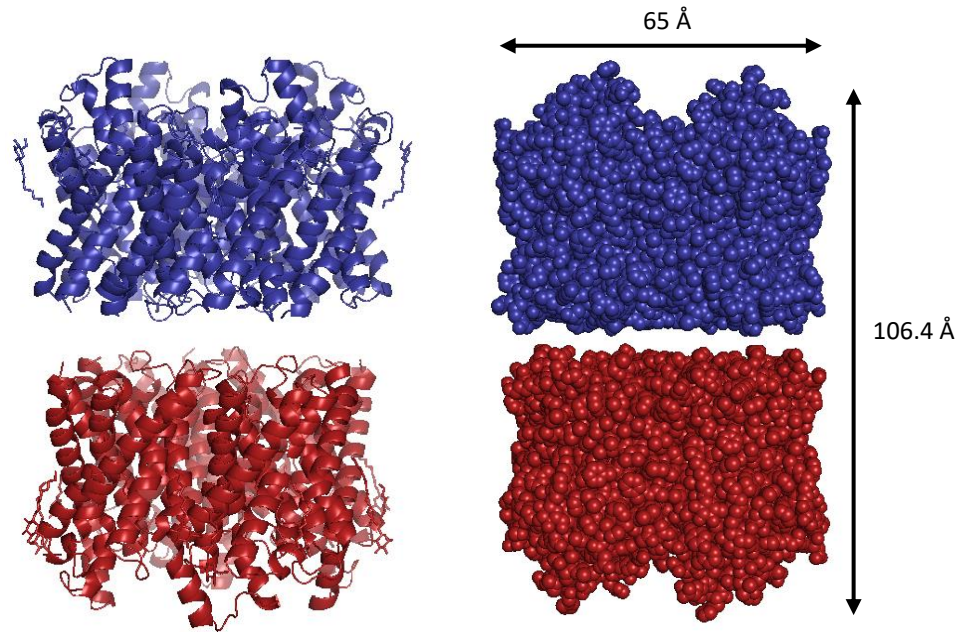
Figure 57. *The crystal structure of GF symmetry mates reveals the potential structure of the GF octamer. The cytoplasm-facing surface of the tetramers are at the interface. The disordered N- and C-termini (not seen in the crystal structure) protrude from the cytoplasm-facing surfaces and may occupy the interface, possibly causing the observed self-association of the tetramers. Figure 61 shows the GF tetramer and disordered N- and C-termini. GF is shown in ribbon (left) and space-filling (right) models; blue and red denote different tetramers. (PDB ID: 1FX8).*

Negative stain EM was conducted on tetrameric DDM-solubilized GF and on tetrameric and octameric LMNG-solubilized GF. The micrographs collected for DDM-solubilized GF tetramers were of poor quality (Figure 58) and could not be used to calculate 2D class averages. LMNG-solubilized GF tetramers and octamers were separated by SEC-MALS and octamer and tetramer fractions were collected based on their elution positions. During SEC, the LMNG concentration was lowered from 0.05% to 0.01% w/v to reduce to a minimum the number of empty micelles. Figure 59 shows one of the 400 electron micrographs taken of the LMNG-solubilized GF octamer preparation. Both the LMNG-solubilized GF octamer and tetramer preparations contained a variety of structures that might reflect the presence of LMNG spherical micelles and worm-like structures that have been observed previously[307]. Empty LMNG micelles have been reported to have diameters of about 5 nm by negative stain EM[307]. Both sets of images also contain an abundance of particles that appear to be octamer and tetramer. This suggests that, following their separation by SEC, the octamers and tetramers re-equilibrate to form a mixture.

Figure 58. *Sample of a negative stain electron micrograph of DDM-solubilized GF tetramer. The length of the scale bar is 100 nm.*
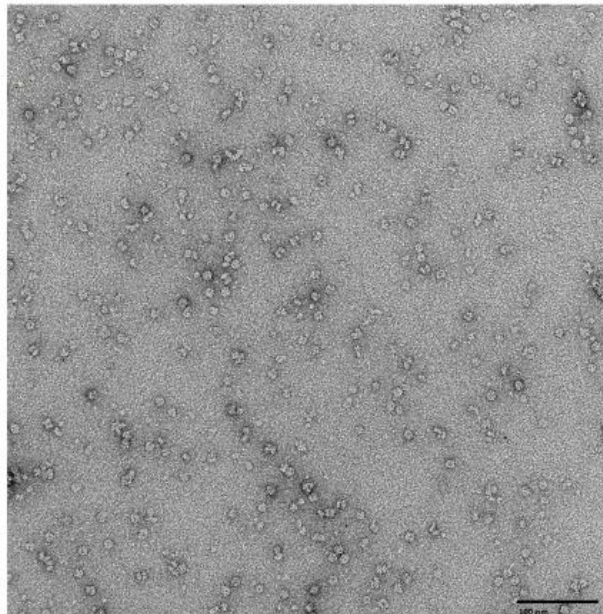
Figure 59. *Sample negative stain electron micrograph of LMNG-solubilized GF octamer used to calculate the 2D class averages in Figure 60. The length of the scale bar is 100 nm.*

The individual images were processed using Relion[261] to create 2D class averages. The three largest classes of particles based on 2D averages from the octamer sample are shown in Figure 60. Care should be taken in interpreting negative stain EM images of membrane proteins in detergent owing to artifacts that can arise from images of the detergent even below the CMC[308]. For example, the "halo" that appears to surround each image in Figure 60 is likely an artefact of image averaging that was used to create the class averages. This sort of artefact is commonly described in the literature[307]. However, when a crystal structure is known, the proteins can be confidently identified when the dimensions of the crystal structure are in agreement with the observed dimensions of the negatively-stained EM particles. Keeping in mind that negative stain EM images the surfaces of particles, the largest 2D class image (Figure 60A-1) shows what appears to be a GF octamer in which the tetramers are stacked in a manner that appears very similar to what is observed in the crystal structure (Figure 57), with the main difference being that the orientations of the two tetramers are tilted with respect to each other. Figure 60B shows three sets of less abundant octamers obtained from the same data set. The longest distance across the octamer in the crystal structure is about 106 Å, as shown in Figure 57, which is in reasonably good agreement with the EM images shown in Figure 60. Also in good agreement are the widths of each tetramer, which in the crystal structure[85] is about 65 Å. Between the two tetramers, electron density is observed that might arise from the disordered N- and C-termini. In the crystal structure[85], the two tetramers are separated by about 10 Å at their closest approach. However, the cytoplasmic surfaces contain a number of cavities that could accommodate the disordered termini (Figure 61A).
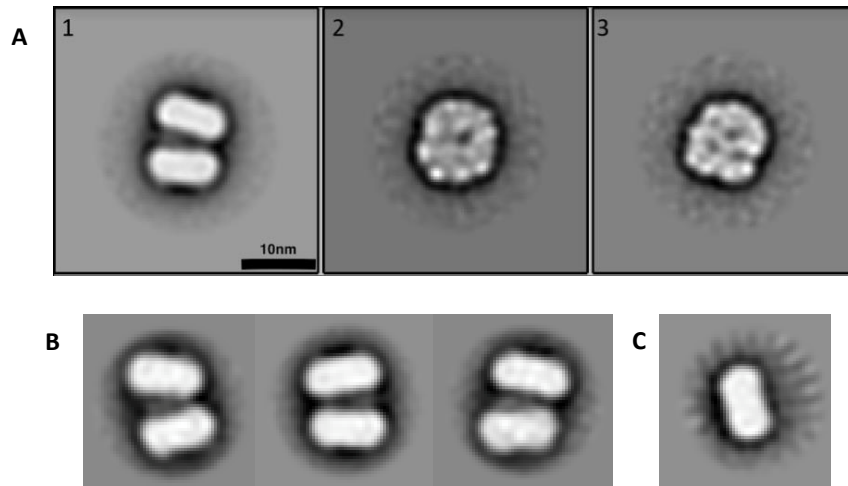
Figure 60. (A) 2D class averages of negative stain EM images of LMNG-solubilized GF octamer. The three most populated class averages are shown where (1) used 5899 particles to form the image, (2) used 5079 particles, and in (3), 4618 particles were used. The scale bar = 100 Å. (B) The less populated 2D class averages of octamers. (C) 2D class average of a tetramer from the octamer preparation.

One difference between the X-ray and EM structures is that the two tetramers are not parallel to one another in the EM octamer images. A possible explanation for the tilt of the structures is that the EM images captured a solution state in which multiple orientations of the tetramers are possible in the octamer, whereas the crystal lattice can only accommodate one orientation. The N-terminus of our GF construct contains an additional 20 residues from the His$_6$-tag that might contribute to the differences observed between the crystal structure and the EM images. There is no mention of whether or not a purification tag was present on the GF protein that produced the X-ray diffraction structure in Figure 57[85]. Figure 61A highlights the N- and C-terminal residues of each monomer in the crystal structure. In each case, the termini emerge near the periphery of the protein, suggesting the possibility that they could escape the tetramer interface altogether. In contrast, the images in Figure 60A1 and Figure60B suggest that the lowest

free energy state of the termini in the EM structures is occupying the space between the tetramers and perhaps their interaction explains the octamer formation. The second and third most populous 2D classes appear to be views of the octamer or tetramer looking down on the periplasmic or cytoplasmic surfaces. Figure 61B shows a model of the GF tetramer surrounded by an LMNG detergent belt built as described in the Methods section and discussed in more detail below.
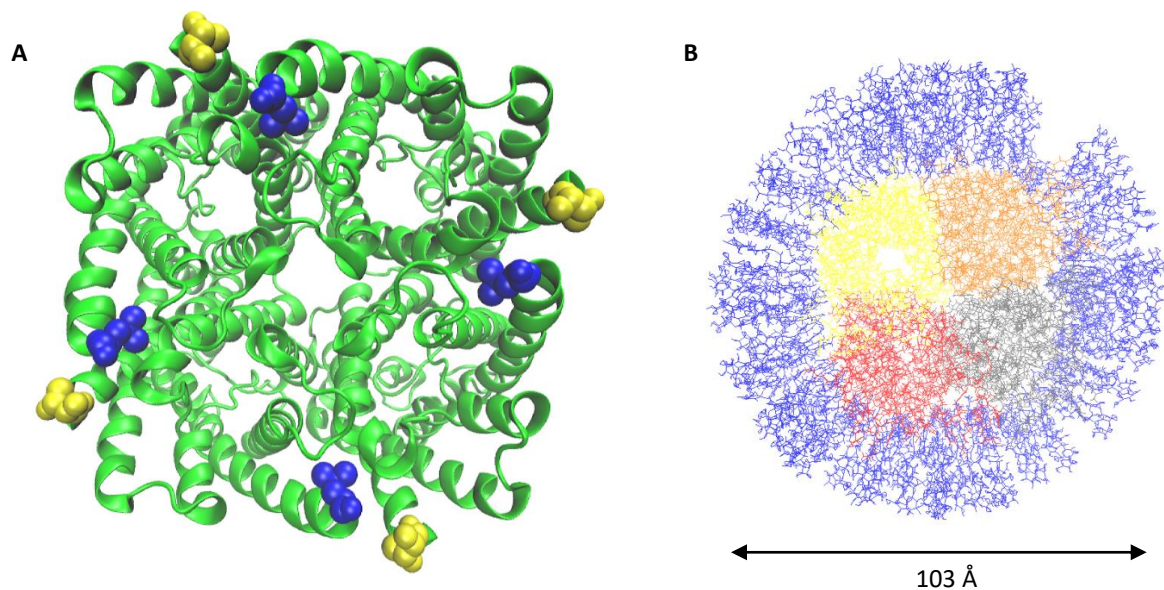
Figure 61. (A) The crystal structure of the GF tetramer (green ribbon) with the N- and C-terminal residues of each monomer shown as yellow and blue spheres, respectively. The location of the terminal residues in the monomers suggests that the termini can escape the interface of the tetramers in the octamer. (PDB ID: 1FX8). (B) A model of the GF tetramer surrounded by an LMNG detergent belt built as described in the Methods section. The LMNG molecules are shown in blue and the monomers in the GF tetramer are individually coloured yellow, orange, red, and grey.

The diameter of the detergent belt circumscribing each tetramer is about 103 Å and this agrees well with the dimensions of the images shown in Figures 60A-2 and 60A-3. It is difficult to determine if the image surfaces are periplasmic or cytoplasmic. On the

periplasmic surface one might expect to observe 4 "towers" protruding from the tetramer that project above the membrane. These are shown most clearly in the space filling model in Figure 57 (right). The highest point of each tower comprises the C-terminus of helix-4, the N-terminus of helix-5, and a connecting loop. Interestingly, four lightly stained areas are observed on one "face" of each of the tetramers in the octamer 2D class averages (Figure 62), which may be these towers. On the other hand, if the surfaces are cytoplasmic, one might expect to observe 4 amino and 4 carboxy termini projecting from the surfaces. Bear in mind, however, that the resolution of the images is on the order of 17 Å and that the termini are expected to be disordered. Finally, that the SEC-purified octamer preparation contains tetrameric GF is evidenced by the 2D class average images shown in Figure 60C. Both dimensions of this particle are in good agreement with the X-ray diffraction structure shown in Figure 57.
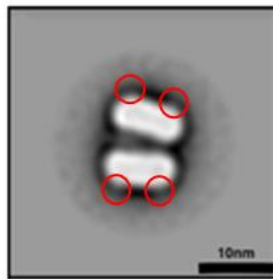
Figure 62. The potential sites of the "towers" of the GF monomers. The most populated 2D class average of the GF octamer sample shows four more lightly stained areas around the perimeter of the molecule where the "towers" may be.

SEC-SAXS was conducted on LMNG-solubilized GF to investigate the shape and size of the self-associated octamers and to further explore the nature of the octamer configuration observed *via* negative stain EM. Significantly higher protein concentrations are required for SAXS than for negative stain EM, where, generally, concentrations of 5 mg/mL are acceptable for SAXS, and for negative stain EM, sample concentrations are on the order of 0.01 mg/mL.

GF octamers and tetramers were separated before SAXS measurements *via* in-line SEC. Figure 63A shows the elution profile from the SEC column. Integrated scattering intensity (grey) is plotted against the time intervals over which SAXS measurements were made (Frame Number). As an indication of particle size, the radius of gyration (R_g) calculated from the SAXS measurements of the eluted fractions are plotted in red. The elution profile shows a plateau of high M_r aggregates, followed by octamer and tetramer peaks. The R_g values measured for the octamer and tetramer, as described below, agree well with the particle sizes identified in the SEC-MALS measurements reported above.

A SAXS scattering curve averaged over several frames for the tetramer is shown in Figure 63B (See Section 1.4). Figure 63C shows a classical Guinier plot of the scattering data, $\ln[I(s)]\ vs.\ s^2$, that can be used to provide a model-free estimate of the R_g of the protein from the slope at low angles. The Guinier approximation assumes that for low

scattering angles, $\left(s < \frac{1}{R_g}\right), I(s) \sim I(0)\exp[-\frac{(sR_g)^2}{3}]$, where I(0) is the forward scattering at

zero angle[187]. The radius of gyration is the average root mean squared (rms) distance of

the electrons to the centre of density in the protein. For the GF tetramer, the scattering

data gave an R_g of 49.9 Å, in reasonably good agreement with the dimensions of the

protein[85]. The intercept of the Guinier plot can also be used to obtain an estimate of the

M_r of the protein from the known concentration and I(0); this was determined from

Figure 63C to be 368.2 kDa for GF[309,310]. Subtracting the mass of the protein tetramer from

this estimate and dividing by the molecular weight of an LMNG molecule suggests that

about 236 LMNG molecules help solvate the GF tetramer.
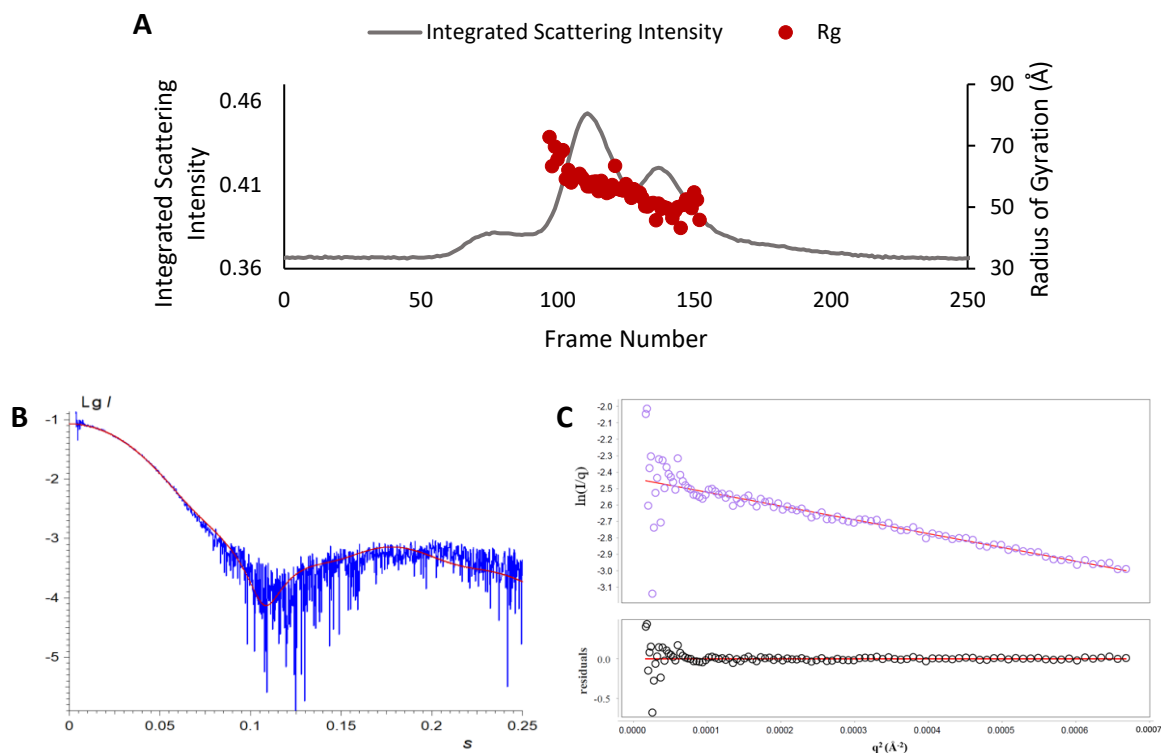
*Figure 63. LMNG-solubilized GF tetramer SAXS scattering data. **A.** The SEC elution profile of LMNG-solubilized GF at a concentration of 8 mg/mL showing peaks of tetramer, octamer, and a small amount of high M_r oligomers. **B.** The baseline-subtracted tetramer SAXS scattering curve (blue) and the fit of the model calculated using the Memprot algorithm[266] (red). The y-axis is the scattering intensity I on a logarithmic scale. The x-axis, s in units of nm⁻¹, is the scattering vector where $s = 4\pi sin\theta/\lambda$. **C.** A Guinier plot of the data in B.*

The spherically averaged scattering intensity I(s) is the square of the scattering amplitude, A(s), as follows: $I(s) = <I(s)>_\Omega = <A(s)A^*(s)>_\Omega$. The scattering amplitude is the Fourier transformation, F , of the difference in electron density, $\Delta\rho(r)$, between the protein $\rho(r)_p$ and the solvent $\rho(r)_s$ as follows: $A(s) = F[\Delta\rho(r)]$[187]. A convenient way to analyze the scattering data represents the scattering particle as a pairwise distance distribution function $\rho(r)$, which is essentially a histogram of distances between every pair of electrons within the particle. The $\rho(r)$ function is generated by Fourier transformation of the scattering intensity as follows: $\rho(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{s^2 l(s) sin(sr)}{sr} ds$. Graphs

of the pairwise difference distribution function $\rho(r)$ versus r are useful for detecting

conformational changes. The $\rho(r)$ and its plots can also be used to determine the

maximum length of the molecule, D_{max}, and the R_g. The D_{max} is the point on the graph

where the curve reaches the horizontal axis and the R_g can be calculated from $\rho(r)$. For

GF, the pairwise difference distribution graph in Figure 64A yielded a D_{max} of 97 Å and a

R_g of 55.6 Å. Integrating the scattering intensity provides an estimate of the protein

excluded volume, V_p. For GF, the volume determined is 271 294 Å^3. By comparison, the

volume of the GF tetramer crystal structure calculated using the VADAR 1.8 server[311] is 145

146.5 Å^3, in reasonably good agreement with the known dimensions of the protein

tetramer with additional mass from the solubilizing detergent.

Kratky plots are graphs of $s^2I(s)$ *vs.* s and are helpful in diagnosing disorder in a

protein. For example, a distinct peak in the graph indicates a folded conformation and a

plateau is diagnostic of a disordered chain. Normalized or dimensionless Kratky plots are

graphs of $\frac{(sR_g)^2 I(s)}{I(0)}$ *vs.* sR_g. For globular proteins, a peak should appear at $\sqrt{3}$ and

deviations from this suggest asymmetry or flexibility in the protein[312]. Normalized Kratky

plots are normalized for protein concentration, volume, and R_g, and are used to detect

changes in conformation or oligomerization. Figures 64B and 64C show the Kratky plot

and normalized Kratky plot of the GF tetramer SAXS data. They show a peak at 1.70 and

confirm that the protein is globular and not very flexible.
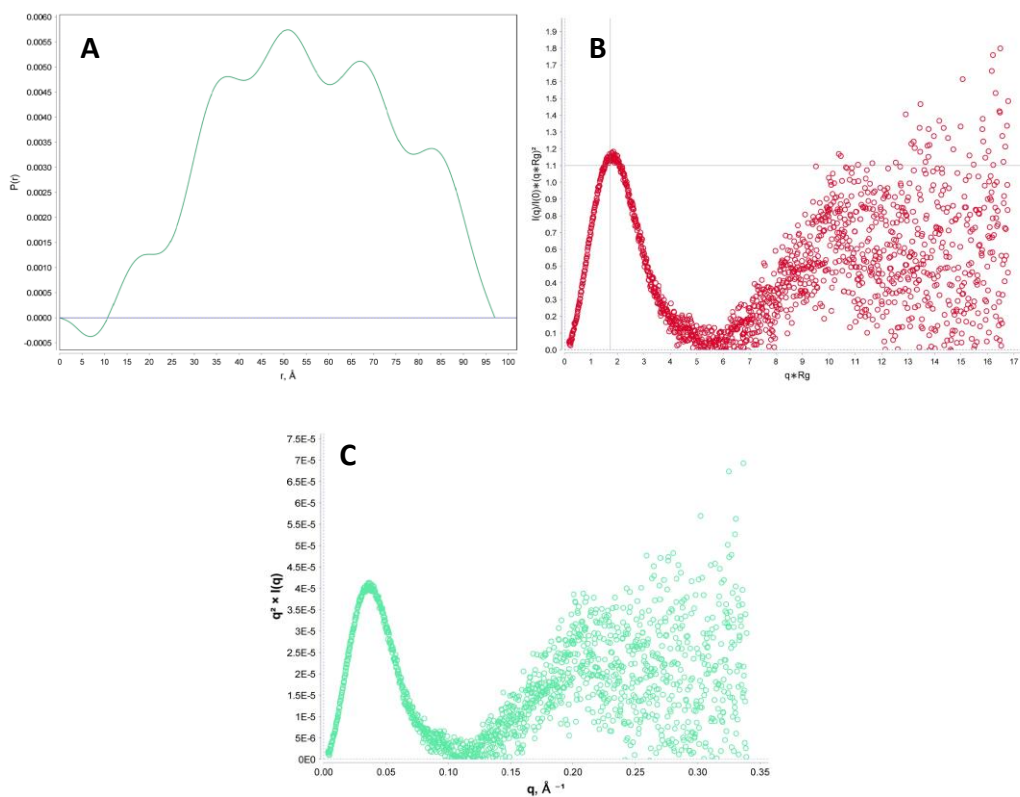
Figure 64. *LMNG-solubilized SAXS scattering data. (A) An electron pair distribution function graph of the GF scattering data. (B) A Kratky plot of the GF SAXS data. (C) A normalized Kratky plot of the GF tetramer SAXS data.*

The scattering data in Figure 63B had sufficient signal to allowed us to prepare an

ab initio 3D reconstruction of the tetramer surrounded by a DDM detergent shell using

the Memprot algorithm[266] and the X-ray diffraction structure of the GF tetramer[85] (Figure

65). Memprot could not fit the protein to an LMNG shell owing to the lack of electron

density information for LMNG. The most important measure of how well the model

structure agrees with the measured data is how well the measured scattering curve fits

the theoretical scattering curve calculated from the model[313]. The simulated SAXS

scattering curve in Figure 63B fits the measured data very well, with the χ^2 difference

between the two curves being 1.489. One thing to keep in mind is that both the N-

terminal His$_6$-tag, the first 5 N-terminal residues of the protein, and the C-terminal 22 residues are missing from the structure[85] and are likely the origin of the deviations of the theoretical curve from the measured one.
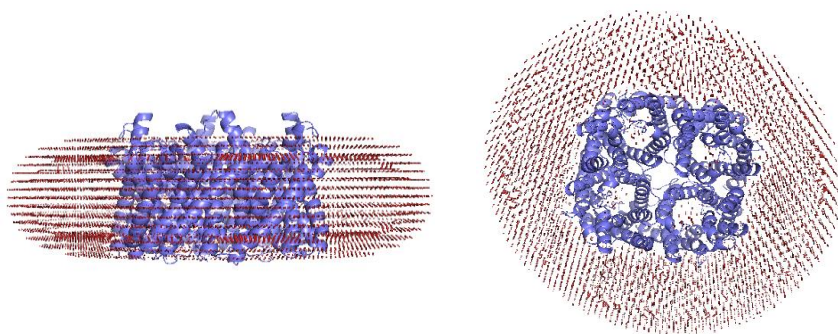
Figure 65. *3D reconstruction of LMNG-solubilized GF using SEC-SAXS data. The GF homotetramer is shown as a cartoon (blue) surrounded by a shell of detergent molecules (red) using the electron density for DDM. Note that the model shows that GF recruits detergent molecules only to the exposed hydrophobic surfaces to form a detergent belt. Modeled with the Memprot algorithm[266] and the GF crystal structure (PDB ID: 1FX8).*

Analysis of the SAXS data from the GF octamer was more problematic. For one thing, the Memprot algorithm[266] is not configured to construct a detergent belt around each of the tetramers in the octamer. We therefore constructed models of the octamer as described in the Methods section. A SAXS curve of the octamer calculated using CRYSOL[263] from the crystal structure[85] shown in Figure 57 deviated from the measured octamer SAXS curve (see Figure 66) with a χ^2 value of 115. Superimposing the DDM-tetramer model generated with Memprot (Figure 65) on the octamer shown in Figure 57 resulted in a model with an improved χ^2 goodness of fit of 24.7. A further improvement was made by building a model of the GF octamer with LMNG detergent belts around each tetramer and superimposing it on the X-ray diffraction structure shown in Figure 57. The

SAXS scattering curve (Figure 66B) generated from this model shows an improved fit with a χ^2 value of 9.1. Optimization of the orientations of each tetramer and its affiliated detergents belt using SASREF[269] resulted in a model (Figure 66C) with an improved fit to the measured scattering curve (Figure 66D) with a χ^2 of 6.8. Here, the model shows the tetramers and their detergent belts tilted with respect to each other in a manner that is reminiscent of that observed in the negative-stained EM image of the octamer (Figure 60A-1). Note, however, that the periplasmic faces of the tetramers are facing each other, not the cytoplasmic faces as in the crystal structure.
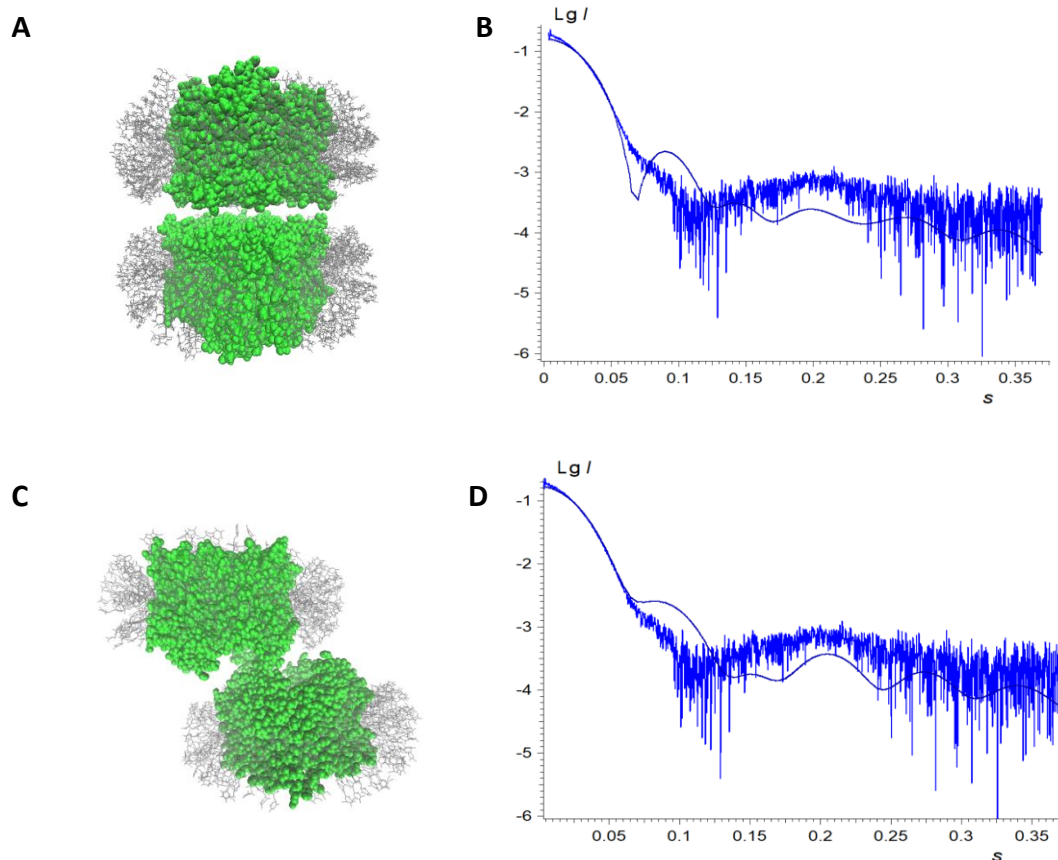
Figure 66. *GF octamer models and their SAXS scattering curves. (A) A space-filling model of the GF octamer (green) with LMNG detergent belts (grey lines) around each tetramer that was built from the X-ray diffraction structure[85] shown in Figure 57. (B) The SAXS scattering curve (dark blue) of the structure in (A) generated with CRYSOL[263]and compared to the measured octamer SAXS curve (light blue). (C) A model of the GF octamer with LMNG detergent belts around each tetramer in which the tetramer orientations were optimized using SASREF[269]. (D) The SAXS scattering curve (dark blue) of the structure in (C) generated with CRYSOL[263] and compared to the measured octamer SAXS curve (light blue).*

Further improvements in the fitting of the octamer SAXS data were achieved by adding models of the amino and carboxy termini (see the Methods section) to the best fit octamer models oriented either with their cytoplasmic surfaces interacting (Figure 67A) or their periplasmic surfaces interacting (Figure 67C). These models gave the closest agreement with the measured SAXS scattering curves with χ^2 values of 3.9 and 2.4,

respectively. Although the model with the tetramers interacting through their cytoplasmic surfaces had the closest fit to the measured SAXS curve, the difference in the χ^2 values between the models in Figure 67A and 66C are too small to allow for the discrimination between the models. The number of LMNG molecules surrounding each GF tetramer is 100, the maximum permitted by the Micelle Builder algorithm[267]. This is compared to the 236 LMNG molecules that were estimated to associate with each tetramer based on the tetramer SAXS data above. Observation of the LMNG-belted octamer models shows that some regions of the presumed membrane-interacting surfaces on the octamer do not interact with detergent, suggesting that the models could be improved further. However, overall there is good agreement between the SAXS data and the SAXS curves generated from the model octamers, suggesting that the models are a valid depiction of the structure of the octamer.
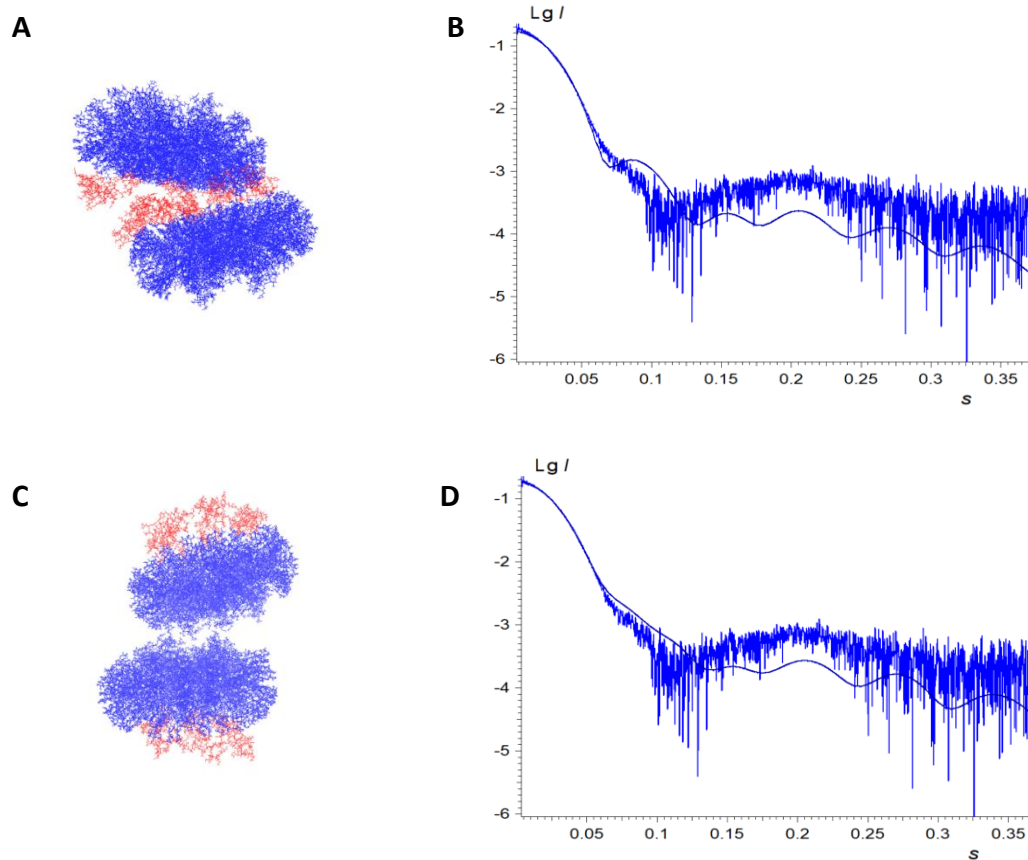
Figure 67. *GF octamer models and their SAXS scattering curves. **(A)** A model of the GF octamer with LMNG detergent belts around each tetramer (blue) with cytoplasmic surfaces facing each other and N- and C-termini (red) modelled as described in the Methods section. **(B)** The SAXS scattering curve (dark blue) of the structure in (A) generated with CRYSOL[263] and compared to the measured octamer SAXS curve (light blue). **(C)** A model of the GF octamer with LMNG detergent belts around each tetramer with periplasmic surfaces facing each other and N- and C-termini (red) modelled as described in the Methods sections. **(D)** The SAXS scattering curve (dark blue) of the structure in (C) generated with CRYSOL[263] and compared to the measured octamer SAXS curve (light blue).*

C-terminally Truncated GF

To discern whether the disordered GF N- and/or C-termini are indeed causing the observed self-association of GF, a truncated GF construct was created with the last 19 amino acids, corresponding to the C-terminus, removed. The expression plasmids (GF-CT-) were transformed into *E. coli* C43 cells and grown in LS-5052. After harvesting the cells, the cell pellet mass was similar to that for preparations expressing the full GF protein. Following the purification steps used for all our preparations, no CT- GF could be isolated from these cells. This may indicate that the C-terminus is needed for proper processing of the protein and insertion into the plasma membrane or that the protein missing its C-terminus is degraded by the cell.

3.3 Solid-State NMR Spectroscopy of LMNG-solubilized GF

Studier's autoinducing medium, LS-5052 were used for uniform labelling of GF with ^{13}C and ^{15}N using cells containing the TEV plasmid. This produced protein yields of 30 – 40 mg of ^{15}N and ^{13}C-labelled GF per litre of cell culture. In Studier's autoinduction media, the main source of carbon is glycerol. However, additional sources of unlabelled carbon, such as succinate, are also in the medium that could dilute the ^{13}C. To ensure high levels of ^{13}C incorporation in the solid-state NMR samples, C-750501 was optimized to produce isotope-labelled GF. It minimizes the presence of other carbon sources and can also be used to incorporate and uniformly label proteins with ^{15}N.

To enhance GF stability, LMNG-solubilized GF was dialysed to remove imidazole. The sample was then gel filtered to remove high M_r soluble aggregates, empty micelles, and free monomers by reducing the concentration of LMNG to its CMC (0.001% w/v). Reducing the concentration of LMNG to its CMC drives any detergent left in the sample to associate with and solubilize GF.

Magic-angle spinning at high speeds necessitates the use of small sample volumes and highly concentrated protein for ssNMR measurements. For example, the new Bruker 111 kHz MAS probe uses a 0.7 mm diameter rotor with a sample volume of 0.5 μL requiring on the order of 1 mg of protein at a concentration of 2 g/mL. The 1.6 mm rotor used in this work has a volume of 8 L. For comparison, SAXS samples were concentrated to approximately 8 mg/mL.

183

Figure 68 shows an elution profile of LMNG-solubilized GF concentrated to approximately 25 mg/mL prior to gel filtration, showing the presence of high levels of soluble aggregate. For the ssNMR experiments, the fractions containing octamer and tetramer (Figure 68, peaks b and c) were pooled and concentrated using Amicon spin filters, achieving concentrations of 32 mg/mL in approximately 80 µL. Microdialysis with the use of PEG-2000 to draw water out of the sample resulted in an NMR sample of approximately 13 mg of ^{13}C/^{15}N-labelled GF in 20 µL (650 mg/mL).
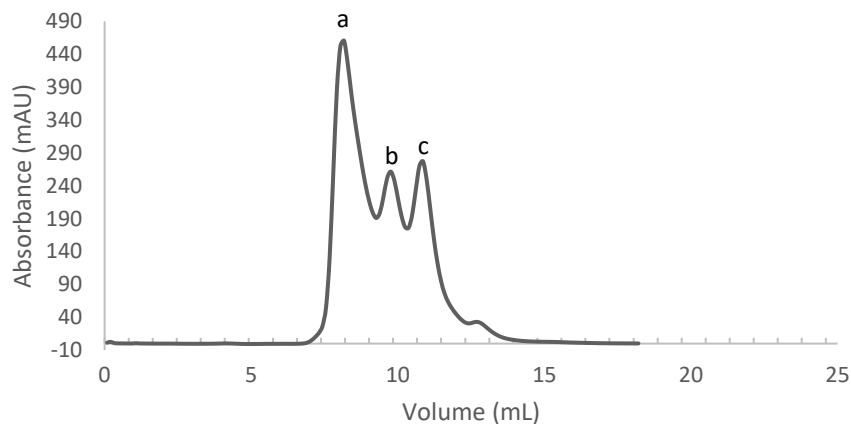
Figure 68. Concentrated, LMNG-solubilized GF forms soluble, high M_r aggregates. The SEC-MALS elution profile of LMNG-solubilized GF dialysed overnight to remove imidazole and concentrated to approximately 25 mg/mL shows the formation of soluble, high M_r aggregates (peak a). Peaks b and c are GF octamer and tetramer, respectively.

One-dimensional ^{15}N (Figure 69, top) and ^{13}C spectra (Figure 69, bottom) of the above-described sample gave high NMR signal intensity indicating that adequate amounts of ^{13}C- and ^{15}N-labelled protein were present in the rotor. The observed chemical shift range shows that the protein is folded and not denatured.
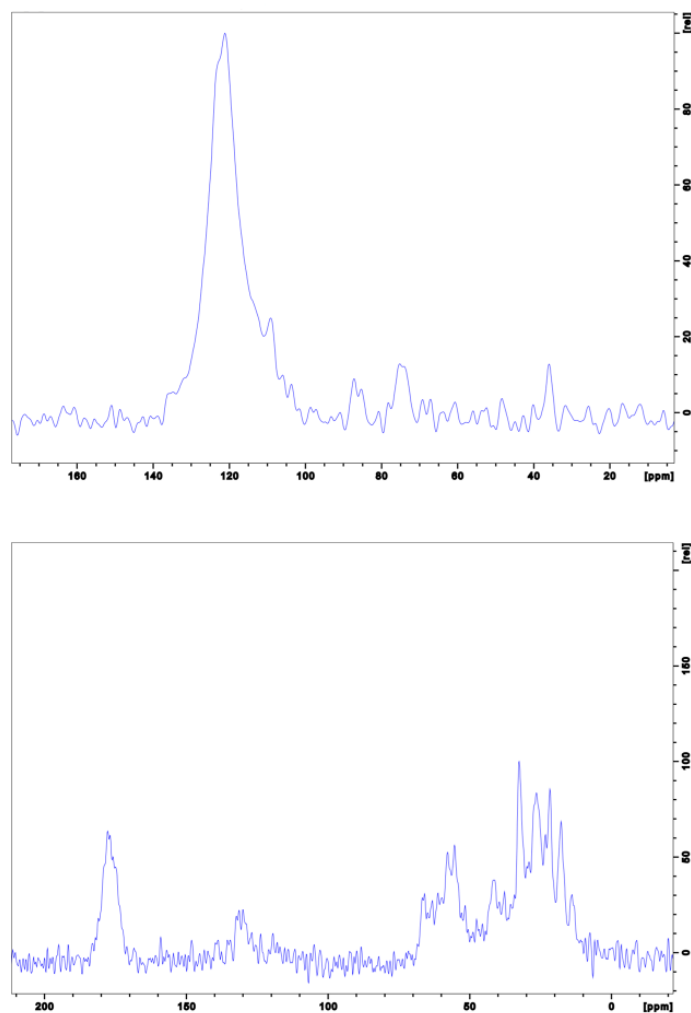
Figure 69. The 1D ^{13}C- and ^{15}N-spectra of LMNG-solubilized GF indicate adequate amounts ^{13}C- and ^{15}N-labelled protein are present in the rotor and that the protein is not denatured. **Top:** 1D hN ^{15}N-spectrum of LMNG-solubilized GF. **Bottom:** 1D hC ^{13}C-spectrum of LMNG-solubilized GF.

Preliminary carbon-carbon 2D NMR spectra from the ^{13}C/^{15}N labelled LMNG-solubilized GF sample. Figure 70 shows 2D DARR (Dipolar Assisted Rotational Resonance) spectra of LMNG-solubilized GF. Figure 70 (top) shows the full spectrum from the methyl region to the carbonyl region whereas Figure 70 (bottom) shows an expansion of the high-field region. Several resonances have been identified in Figure 70

(bottom) based on known unique ^{13}C chemical shift ranges in folded proteins[314]. For example, Ser and Thr Cβ resonances are always the most downfield shifted in proteins and Ala Cβ are always the most upfield shifted. Other unique shifts giving rise to identifiable peaks are the Pro-Cα-Cβ and Ile-Cγ-Cδ.
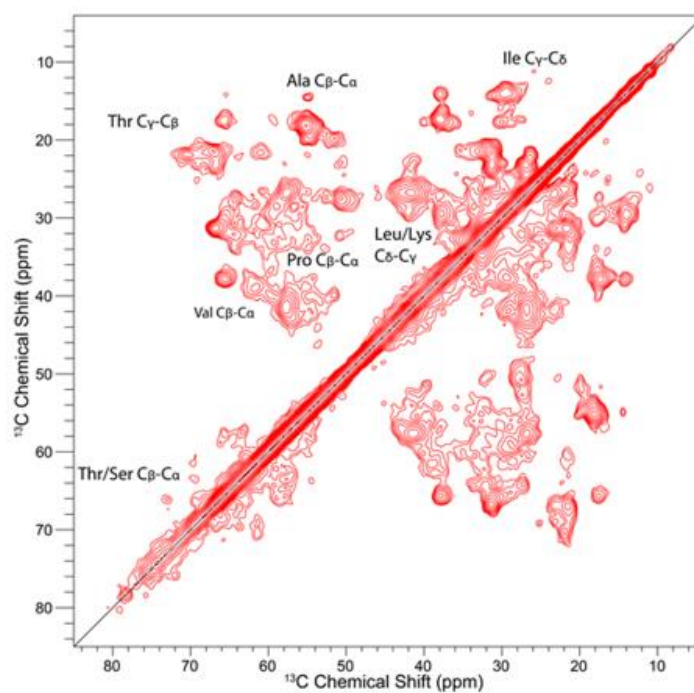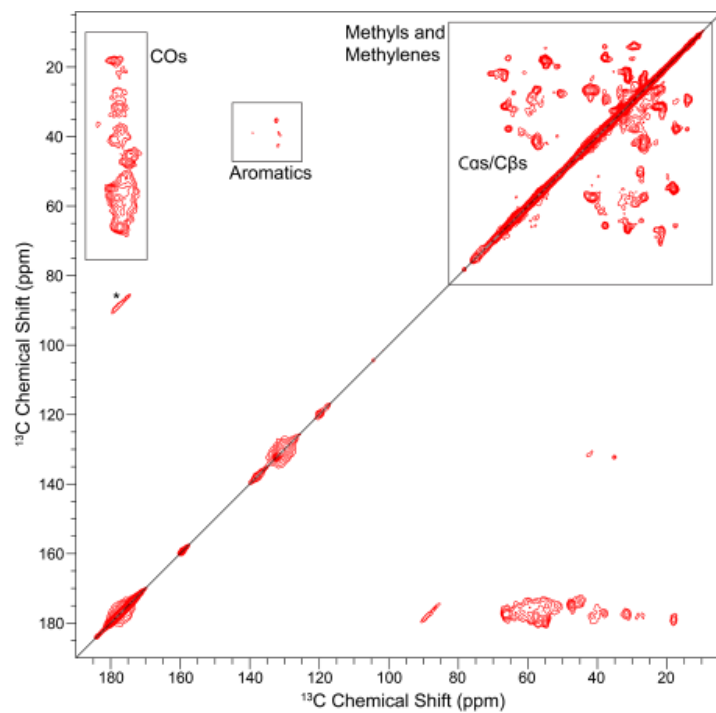
Figure 70. *2D DARR (Dipolar Assisted Rotational Resonance) spectra of LMNG-solubilized GF.* **Top:** *the full 2D DARR spectrum is shown with the ¹³C-methyl and methylene, Cα and Cβ and carbonyl regions of the spectrum identified.* **Bottom:** *the high-field region of the spectrum shown in (top). Several crosspeaks in the spectrum are identified based on known unique chemical shift ranges.*

The average ^{13}C T_2 relaxation time of the upfield region of the spectrum from 23-81 ppm was measured by 1D NMR as described in the Methods sections yielding a T_2 of 1.79 ms with a standard deviation of 0.04. The relaxation decay is shown in Figure 71. This rapid transverse relaxation precluded the measurement of 3D spectra necessary for observation and assignment of all the resonances in the protein.
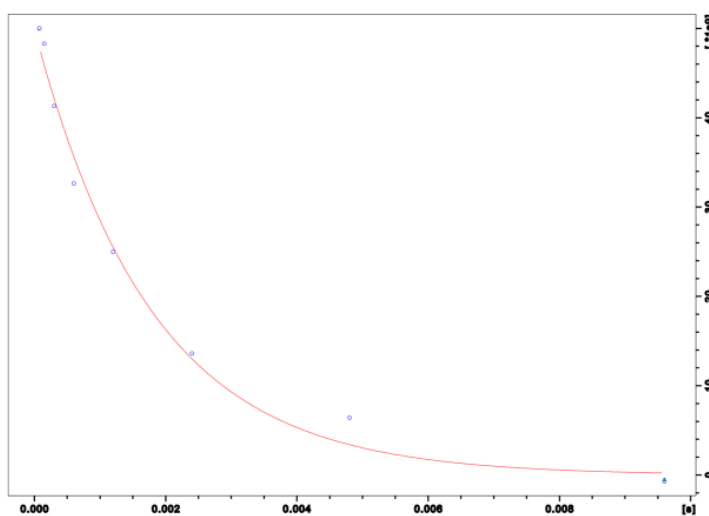
Figure 71. *The average ^{13}C T_2 relaxation time of resonances in the upfield region from 23-81 ppm of LMNG-solubilized GF is 1.79 ms with a standard deviation of 0.04.*

The short T_2s suggest exchange-broadening that might arise from several sources. Although a simple tetramer-octamer equilibrium could broaden many of the resonances at the interface between the tetramers, it seems unlikely that this would affect the entire protein. Another possibility is that the protein is undergoing conformational exchange on the ms-µs timescale that significantly changes the chemical shifts of most of the resonances in the protein. This would be somewhat surprising because no such conformational change has been reported in GF or in any aquaporin. Furthermore, it is

188

difficult to imagine the protein retaining its exquisite substrate specificity if it were to undergo very large conformational fluctuations because the substrate-specificity clearly depends, to a large extent, on the pore maintaining its size within very narrow limits. However, it is certainly possible that the detergent environment has led to conformational flexibility in the protein that is not present in the native bacterial membrane. In addition, it's also possible that detergent exchange could lead to line-broadening of many resonances in the protein, especially those near the surface and interacting with the protein. Besides all these possible sources of dynamic exchange, it is also possible that the NMR resonance in the protein dissolved in LMNG at high concentrations are broadened owing to static heterogeneity. Recall that in bilayer-dissolved IMP's the lipid:protein ratio can affect the static protein conformational heterogeneity and lead to line-broadened spectra[155,165]. This problem has been overcome in some cases by careful exploration of lipid:protein ratios to discover the right combination that gives rise to one predominant conformation. Perhaps membrane proteins, especially those that do not bind a prosthetic group, are more conformationally flexible than we realize.

Chapter 4: Discussion

Integral membrane proteins (IMPs) play important roles in many aspects of cellular function and homeostasis. Despite being encoded by approximately 30% of the genome[57], knowledge about the structure and dynamics of IMPs lags significantly behind soluble proteins; this bottleneck is generally attributed to the IMPs native lipid environment, which is required for the protein to remain stable and solubilized. The lipid environment presents a unique challenge for techniques such as X-ray diffraction and NMR spectroscopy that are used to obtain high-resolution structural and dynamics information. To circumvent this issue, researchers have used detergents and developed membrane mimetics to encourage IMP solubility and stability. Substantial developments in fields like ssNMR spectroscopy[146] and cryo-electron microscopy (cryo-EM)[315] now allow the use of these techniques to explore the structures of IMPs .

Little is known about the dynamics of GF and IMPs in general. The work presented here was aimed at the preparation of an isotope-labelled GF sample to study its protein dynamics *via* NMR spectroscopy. To this end, preliminary 2D carbon-carbon ssNMR spectra were collected on LMNG-solubilized ^{13}C- and ^{15}N-labelled GF. The bulk of the work was dedicated to preparing an isotopically-labelled, homogenous, detergent-solubilized GF sample that was stable for one week for NMR spectroscopy studies. This included optimization of the buffer environment and solubilizing agent. Here, I purified homogenous tetrameric DDM-solubilized GF, which aggregated after 24 hours, and an

LMNG-solubilized GF octamer-tetramer equilibrium that was stable for 9 days. The octamer was characterized *via* size-exclusion chromatography multiangle light scattering (SEC-MALS), negative stain electron microscopy (EM), and size-exclusion chromatography-small angle X-ray scattering (SEC-SAXS). Finally, methods to discourage self-association to obtain pure GF homotetramer in LMNG were explored.

4.1 GF Overexpression

NMR spectroscopy experiments often require milligram amounts of isotopically labelled protein. Studying IMP dynamics *via* NMR spectroscopy is challenging in part due to low levels of protein expression in native host cells; heterologous overexpression of IMPs often slows cell growth and may be toxic[316]. Initially, BL21(DE3)-pLysS cells were used to overexpress GF which resulted in poor protein yield. These cells contain an additional plasmid that expresses T7 lysozyme to reduce basal level protein expression for the gene of interest[317], which makes them suitable for expression of otherwise toxic genes. It has been shown, however, that overexpression of IMPs in pLysS cells can result in cytoplasmic aggregates containing the overexpressed IMP, as well as chaperones, proteases, and other cytoplasmic proteins[254]. The pLysS cells also conferred ampicillin resistance, whereas the pET28b(+) plasmid encoding the recombinant GF gene conferred kanamycin resistance; using two antibiotics that may be metabolized at different rates presented additional complexity to the overexpression media and growth protocol. To minimize the variables in the overexpression protocol, it was decided to use cells lacking the pLysS plasmid and to design a new expression plasmid (see below).

During the course of my work, we discovered a mutation at position 168 from an aspartic acid to a valine in the GF gene sequence used in the pLysS cells. A new plasmid containing a codon optimized, mutation-free GF gene sequence was transformed into BL21(DE3) cells. Cells are biased towards certain codons during overexpression, and in some cases, specific codons are avoided entirely[318]. Codon optimization adjusts the

codons used in a target sequence to those that are used preferentially in the host cell. This has been shown to increase protein overexpression in *E. coli*[319]. BL21(DE3) cells are used routinely to overexpress recombinant protein[320] but are often not suitable for IMP overexpression[321] likely due to overloading the Sec translocase, which is required for membrane protein biogenesis[322]. Despite this, the BL21(DE3) cells used to overexpress GF in this work grew robustly in rich media and expressed abundant protein that was inserted in the *E. coli* inner membrane. This new expression system allowed enough protein expression for the characterization of solubilized GF. However, rich media such as LB or TB cannot be used to incorporate isotope labels for NMR spectroscopy studies.

M9 minimal medium[323] is widely used to prepare ^{13}C- and ^{15}N- uniformly labelled and site-specific labelled protein. Protein yields are often significantly lower in this medium compared to rich media, likely due to slower cell growth in the log phase leading to lower cell density. To combat this, cells were grown to high cell density in a rich medium and allowed to equilibrate with the minimal medium before induction[324]. GF protein yields were still too low with this strategy, and growing cells in large volumes of isotope-containing M9 minimal media would be cost prohibitive and impractical. A more reasonable approach was to explore different labelling media to prepare isotope-labelled GF.

W. F. Studier's autoinduction media[204] were routinely used to grow isotope-labelled GF. The media used here are LS-5052 and C-750501 for ^{13}C- and ^{15}N- uniform

labelling. Using these media without further adjustments resulted in nearly 7-fold higher GF protein yields compared to the M9 minimal media; minor modifications such as titrating the final culture to pH 7 before inoculation further increased the yield of protein. Excretion of protons and metabolic waste products such as lactic acid[325] acidifies the growth medium as the cells grow. *E. coli* grow optimally in a pH range of 5.5 to 8[326] and titrating to pH 7 enhanced cell growth and overexpression of GF. LS-5052 was initially used to prepare uniform-labelled GF but contains 20 mM succinate, which can act as an alternate carbon source, potentially diluting the incorporation of isotope labels. C-750501, which was optimized for maximum ^{13}C incorporation, was used instead of LS-5052. C-750501 contains glucose and glycerol, which can be purchased relatively inexpensively isotopically labelled with ^{13}C, as the only carbon sources available to the cells; this ensures high incorporation of the label into the recombinant protein.

A stark difference in protein yield was observed between GF purified from M9 minimal media and Studier's autoinduction media though both media provide sufficient nutrients for cell growth. It is possible that the growth conditions used for these media contribute to the disparity. For M9 minimal media, cells are grown at 37°C for approximately 10 hours. For final cultures in Studier's autoinduction media, growth at 20°C is recommended over 48 hours. The cells in M9 minimal media are likely expressing large amounts of GF quickly, which can slow cell growth or be detrimental to the cell. At 20°C, growth and overexpression are likely progressing slowly compared to 37°C, which allows the cells to better adapt to the presence of large amounts of GF. Allowing the cells

194

to grow for longer periods of time allows for similar cell densities to cultures grown at 37°C. Another important contribution to the difference in GF yield is the higher cell densities achievable in Studier's media compared to LB and M9.

4.2 Extracting GF from the Membrane

After successfully overexpressing the IMP of interest, the membrane-embedded protein is usually extracted from the native lipid environment for structural or functional studies. Though other novel surfactants and membrane mimetics have been developed, detergents are commonly used to extract IMPs from the membrane[327].

I explored several extraction agents for the removal of GF from the *E. coli* membrane and found that in general, detergents were the most effective in terms of the amount of protein that could be extracted. Dodecyl maltoside (DDM), which is the most commonly used detergent in the purification of IMPs[327], was the most effective at extracting large amounts of GF from the native lipid bilayer compared to other detergents and extraction agents that were explored.

Octyl glucoside (OG) has a shorter carbon chain length than DDM. OG was the detergent used to solubilize GF when its crystal structure was solved[85]; however in my hands, yields of OG-extracted and solubilized GF were significantly lower than of DDM-extracted GF. Furthermore, characterization of the sample revealed a mixture of tetramer, octamer, and higher MW aggregates, and after 24 hours, the OG-solubilized sample contained visible precipitate. While both OG and DDM are nonionic detergents, OG is considered a harsher, more denaturing detergent than DDM due to its shorter carbon chain[294]. Shorter carbon chains may not adequately shield the hydrophobic areas of the IMP from the aqueous environment, causing association and aggregation.

Ionic detergents, which are considered harsh and denaturing, were also tested in the extraction step. SDS and Dodecyltrimethylammonium bromide (DTAB), like DDM, have 12-carbon chains; SDS has a negatively charged head group while DTAB's head group is positively charged. It is possible that solubilizing GF in an ionic detergent would discourage aggregation by charge repulsion of the micelles. Extracting GF from the membrane using SDS and DTAB resulted in poorer yields than in OG and DDM, with DTAB less effective for extracting GF from the membrane than SDS.

LMNG and LMNG:CHS have been shown to extract delicate IMPs from the membrane. For example, LMNG was used to extract STRA6, a receptor responsible for the internalization of retinol[328]; other IMPs, such as GPCRs, have been extracted using LMNG:CHS[329,330]. CHS is a cholesteryl molecule that is often used in conjunction with detergents to solubilize and stabilize eukaryotic IMPs[331], and has been also used to extract IMPs when mixed with DDM[332]. CHS enhances the stabilization of solubilized IMPs because cholesterol is present in the lipid bilayer of eukaryotic cells[333]. CHS likely provides an effective mimic of the cholesterol in the native eukaryotic lipid environment and encourages proper folding. LMNG and LMNG:CHS did not extract GF from the membrane as effectively as DDM or OG. LMNG is much larger that DDM and may be a poor extraction agent for GF due to its relatively large structure, which may be sterically hindered from penetrating and disrupting the membrane. In this work, I extracted the membranes with LMNG for the same period of time as for DDM and longer extraction times might increase the amount of GF extracted from the membrane. The SEC-MALS

elution profile of GF extracted and solubilized in LMNG indicated a pure homotetramer suggesting that future studies optimizing membrane extraction with LMNG may be warranted.

DDM and OG are relatively inexpensive detergents; however, frequent membrane extractions using high concentrations of detergent may not be economical. SMA co-polymers theoretically allow for a detergent-free method to solubilize IMPs in their native lipid environment[334]. However, in my hands, SMA 2k was not an effective solubilizing agent for GF. I made no attempt to optimize the extraction conditions and used conditions that were optimized for other membrane proteins[214]. The buffers that are suggested for optimal SMA use fall at the upper limit of the pH range amenable to GF without causing aggregation[278], and their compatibility with GF is unknown. From the results sections above, the optimal buffer environment of GF, and likely IMPs in general, is protein dependent. Varying buffer type, pH, salt concentration and storage temperature even slightly may cause self-association and aggregation and likely contributed to the poor extraction of GF from the membrane. It is possible that better optimization of these variables might lead to higher yields of soluble, homogenous GF.

4.3 Osmolytes and Other Additives

After observing GF self-association to octamers in DDM and LMNG, different additives and sample storage conditions were explored to discourage octamer formation in LMNG and further aggregation in DDM. Methods such as removing the His_6 affinity tag to reduce any affinity tag-mediated aggregation were impeded by low activity of the TEV protease in the presence of detergents[292]. A thrombin cleavage tag could be used instead, as thrombin has been shown to be active in the presence of a variety of detergents[292].

Osmolytes are molecules found intracellularly that are known to stabilize proteins; they play a role in regulating cell volume under water stress and protect proteins from other stressors such as temperature and pH fluctuations[335]. The mechanism by which osmolytes interact with and stabilize proteins is still incompletely understood, but it has been hypothesized that they (1) increase the hydration of proteins, encouraging them to fold more compactly and (2) interact unfavourably with the peptide backbone, increasing the free energy of the unfolded state and strengthening the protein secondary structure[336]. Glycerol was added to the detergent-solubilized GF sample after purification for two reasons: as an osmolyte, it may improve protein stability[286], and as the substrate, glycerol may induce small conformational changes that aid in stabilizing the protein[337]. SEC-MALS elution profiles (Figure 31) indicated a 1:1 ratio of octamer to tetramer in DDM-solubilized GF after addition of glycerol, so other additives that were proposed in the literature to aid in solubilizing IMPs or improve and promote proper protein folding

were explored. The osmolytes TMAO[336], ATP[289], and arginine[287] initially showed promising results, but the samples revealed similar self-association to the DDM-solubilized sample after 24 hours.

Because of the extremely variable native environments of IMPs, the solubilizing agent used to solubilize an IMP is often optimized for the IMP of interest. This is time-intensive and tedious, though the development of high-throughput screening methods for overexpression and solubilization of IMPs is progressing[338,339]. This aids in identifying a point from which to begin optimization. Often, once a buffer has been chosen, minor adjustments to pH, temperature, ionic strength, and detergent concentration are required.

Phosphate buffer was used here because the optimal buffer range encompasses physiological pH and because this buffer is amenable to NMR measurements in that it does not interfere with the measurements[340]. Previous work showed that GF is susceptible to changes in pH[278]; at low pH (less than pH 6), GF dissociated to monomers, and at high pH (greater than pH 8), visible aggregation was observed. Titrating DDM-solubilized GF (Section 3.2, Effect of pH), to high (pH 8.5) and low (pH 3.7) pH showed that the GF tetramer would dissociate at low pH and reform after being titrated back to pH 7.5. However, based on the fluorescence intensity of GF, it appears that, while some GF tetramers are recovered, not all the protein is retained in tetrameric form. Because of this,

and to remain in a physiologically relevant environment, all experiments were carried out at pH 7.4.

Purification of GF was conducted at room temperature; samples were stored at 4°C and -80°C to observe the effect of temperature on the oligomeric state of the sample. Based on results obtained *via* DLS, there was little observable difference between storing the sample at room temperature and at 4°C. At low temperatures, higher concentrations of phosphate buffer are insoluble, which may affect the buffering capacity of the buffer, and thus, the solubility of GF. When GF was frozen for storage, the thawed sample showed visible precipitate after warming to room temperature. Working at room temperature was both the most convenient and physiologically relevant temperature to store GF.

Adjusting the ionic strength of the buffer allowed for the preparation of purely tetrameric, DDM-solubilized GF. Ionic strength does not affect nonionic detergents as much as ionic detergents[341], but is still required in the environment to maintain close to physiologically relevant conditions[340]. SEC elution profiles suggested that initial preparations of DDM-solubilized GF in buffer containing 250 mM NaCl were purified as both octamer and tetramer. Homogenous DDM-solubilized GF tetramer was obtained in buffer containing 150 mM NaCl, which is close to the physiological concentration and the concentration of saline used for medical purposes[342]. While ionic strength does not affect nonionic detergents such as DDM and LMNG, it certainly will interact with the protein

and may thereby alter the solubility and self-association of the protein. Whereas the salt concentration was optimized for DDM solubilization, the salt concentrations used for LMNG solubilization were the same as those used for DDM. It is conceivable that further optimization of the salt concentration in LMNG may result in more homogenous samples. However, the similarity in the nature of the detergent molecules suggests that this is unlikely.

4.4 Solubilizing GF with Detergents, Membrane Mimetics, and Other Agents

Despite the development of other detergents and novel solubilization methods, DDM is still the most commonly used detergent initially explored for solubilization and stabilization of IMPs[341]. Its nonionic and relatively gentle character likely allows it to be amenable to a large variety of IMPs. Sugar-based detergents have been and continue to be popular in IMP studies; global analysis of detergents used to extract and solubilize IMPs for structural and dynamics studies found that sugar-based detergents are used in approximately 75% of the studies analysed[341]. DDM and OG, among other sugar-based detergents, are used based on their popularity, proven usefulness, and for economical reproducibility.

Though sugar-based detergents are used commonly and have remained popular since their introduction into IMP studies in the mid-1980s, their compatibility with IMPs is not necessarily guaranteed. It is common to extract the IMP of interest with one detergent and exchange it into another detergent or into a membrane mimetic. Depending on the IMP of interest, other agents such as lipid vesicles, bicelles, nanodiscs, or other types of detergents, such as zwitterionic or ionic detergents, may be superior to the extracting detergent for stabilizing the IMP. Reconstitution into lipid bilayers such as vesicles, bicelles, and nanodiscs present a physiologically relevant environment. This calls into question the relevance of structures and other characterizations gleaned from detergent-solubilized IMPs. It may be fruitful to examine the structure of an IMP

solubilized in detergent and in a lipid bilayer, such as bicelles or nanodiscs, to see what, if any, differences there may be.

In my work with freshly prepared DDM-solubilized GF, the protein eluted from the SEC column as a homotetramer but, over 24 hours, the protein self-associated to form octamers and higher MW oligomers, leading to visible aggregate. Solubilizing GF in different detergents with varying chain length, head group, and charge, allowed for possible insight into optimizing solubilization of GF and exploration of the processes leading to aggregation of detergent-solubilized IMPs.

For GF extracted with DDM and exchanged with another solubilizing agent, efforts were made to reduce the amount of DDM in the sample and encourage complete detergent exchange. Nevertheless, trace amounts of DDM may still be present, resulting in a lack of homogeneity both in the detergent composition and protein conformation. Additionally, endogenous membrane lipids may have remained bound to GF during the extraction process and detergent exchange[343]. Some of the inhomogeneity of the GF protein preparations may have been the result of incomplete detergent and phospholipid exchange and further purification steps such as ion exchange chromatography may be used to ensure that GF is solubilized in a homogenous environment.

On the other hand, solubilizing GF *via* a mixed micelle or a bicelle may be what is needed to enhance its stability outside of the membrane. Historically, any q ratio was

thought to form the bicelle architecture as described in Section 1.8.4 above. Instead, it

was shown that at q < 0.5, mixed micelles are formed instead of the classical bicelle

structure and an ellipsoidal micelle is formed with bicelle-like properties for 0.5 < q < 1[225].

The dynamics of the *E. coli* outer membrane protein OmpX were examined in detergent

micelles, bicelles (q = 0.5), and nanodiscs[344]. The dynamic variability of OmpX was

compromised in micelles compared to the OmpX in the lipid bilayer environment of

bicelles and nanodiscs, which shows that the dynamics of IMPs are affected by the

detergent micelle. The β_2-adrenergic receptor (B2AR) was reported to be stable in both

micelles and nanodiscs. Micelle-solubilized B2AR was observed to retain, in higher

proportions, the inactive form of the receptor in micelles. When reconstituted into

nanodiscs, B2AR retained the active conformation[345]. This suggests that, while micelles

may be amenable to solubilizing IMPs, the architecture of micelles compared to bicelles,

nanodiscs, and other lipid bilayer environments may cause subtle differences in IMP

structure and stability.

     Amphipols, SMALPS, and RHPs have been developed to provide alternatives to

detergent solubilization of IMPs, though the use of detergents still predominates in IMP

studies. Amphipols (APols) have been shown to solubilize a variety of IMPs[346], though

characterizing the APol/IMP complex is still ongoing[347]. APol-solubilized

bacteriorhodopsin (BR) was correctly folded and retained functionality, but manifested

slower photocycling rates compared to rates in the native membrane[348]. Interestingly,

Apol-solubilized BR photocycling kinetics were more similar to the rates seen for

membrane-embedded BR compared to detergent-solubilized BR, which suggests that Apols may stabilize IMPs better than detergents. Other studies of APol-solubilized IMPs, such as the sarcoplasmic reticulum Ca^{2+}-ATPase, suggest that APols constrain the inherent dynamics of the protein, resulting in slow hydrolytic activity and Ca^{2+} dissociation[349,350].

SMALPs have also been used to solubilize rhomboid proteases with activity in the SMALPs more similar to that of the protease in the membrane than in detergents[351]. One drawback, however, is that only half of the protein was incorporated into SMALPs. Sensory rhodopsin II reconstituted into SMALPs showed similar constraints on conformational dynamics when compared to the dynamics of the protein in liposomes or nanodiscs[215].

RHPs are the most recent polymer developed and designed to solubilize IMPs in aqueous solutions and retain the tertiary structure of water-soluble proteins in organic solvents[253]. RHPs have been shown to solubilize AqpZ during cell-free synthesis. My results above in Section 3.2.2 show that GF appears to be soluble in RHPs.

All three systems have great potential, though there is still much to be learned about the complexes they form with IMPs. Based on my results, solubilization *via* Apols appears to be the most promising system, though it was not pursued in this work due to the success of LMNG-solubilization. RHPs also show promise but rigorous

characterization studies were not feasible due to the amount of RHP available for our experiments. Though SMALPs have been developed over the past decade[214], characterizing and optimizing the SMA copolymer and the lipid particles devoid of IMP themselves is still ongoing[352,353]. SMALPs were not further pursued here as the system likely requires further optimization; at the very least, detergent-solubilized GF should be characterized in the SMA-solubilization buffer, as it was shown in my results that GF is polydisperse as a result of its buffer environment (Section 3.2.1). Preliminary work here suggests that additional characterization of Apol/GF and RHP/GF complexes may be useful for discouraging aggregation and improving the stability of GF.

It was previously shown that high concentrations of SDS dissociate GF tetramers into monomers, but that the protein retains most of its α-helical content[278]. In this study, SDS was used to solubilize ^{15}N-labelled GF to determine the feasibility of studying GF monomer dynamics *via* solution NMR spectroscopy. Only a few very broad NMR resonances were observed (Figure 40), suggesting that most of the protein is in dynamic conformational exchange on the millisecond to microsecond timescale. While SDS is a strong solubilizing detergent, visible precipitate appeared in the samples after 24 hours, possibly because the solubilized monomers do not retain their native structure in the presence of high concentrations of SDS. Instead, they may be solubilized, poorly folded molten globules[354]. It is of note that GF does not immediately unfold and aggregate in SDS. Some cooperativity in the folding of GF has been shown where monomers associate within the membrane, forming tetramers and assuming their native conformation[86].

Because of this cooperativity, it is likely that any dynamics information obtained on SDS-solubilized GF monomers would not be representative of the individual monomers in the GF homotetramer.

LMNG has been shown to stabilize delicate membrane proteins such as human rhodopsin, a GPCR responsible for light perception[355]. LMNG and LMNG:CHS mixtures were not as successful as DDM or OG at extracting GF from the membrane; however, monitoring the stability and characterizing the oligomeric state of freshly prepared and aged samples of LMNG-solubilized GF suggested that LMNG is better at stabilizing GF than DDM (Figures 45 and 46). GF samples did not show any observable preference to solubilization in LMNG or LMNG:CHS. CHS is a detergent that is a cholesterol-like mimetic; it is added to other detergents to form a mixture used to solubilize and stabilize IMPs, notably GPCRs and other eukaryotic IMPs[356,357]. *E. coli* is prokaryotic and does not contain any cholesterol, so it is unsurprising that CHS does not contribute to the stabilization of detergent-solubilized GF. It is of note, however, that some Gram negative bacteria produce hopanoids, which are structurally homologous to sterols[358]. Hopanoids have been observed primarily in the outer membrane of gram negative bacteria where they play a role constraining the lipid bilayer and forming a liquid ordered phase[359], as well as a potential role in antibiotic resistance[358].

The observed overall stabilization of LMNG-solubilized GF compared to other detergents and solubilization agents may be attributable to the structure of LMNG.

Compared to DDM and OG, LMNG molecules are larger, containing two hydrocarbon tails and two polar head groups. The two hydrocarbon tails of LMNG may be better at shielding the hydrophobic surface of GF than DDM. Despite the observed stabilizing effects, LMNG-solubilized GF appears to be stabilized in an octamer-tetramer equilibrium, based on SEC-MALS elution profiles and protein conjugate analysis (Figure 45). The CMC of LMNG is low compared to DDM, approximately 0.01 mM[299] for LMNG compared to 0.17 mM[360] for DDM. LMNG micelles are relatively stable, with off-rates that are four orders of magnitude lower than DDM[361]. During preparation of buffer containing LMNG, the detergent required several hours to dissolve at room temperature, which was accelerated by vigorous vortexing or by applying heat to the buffer. The strong self-association of LMNG may indicate a high affinity for GF, explaining its superior solubilizing power. Nevertheless, the observation of GF octamers and aggregates in LMNG suggests that some hydrophobic areas in the protein are exposed, giving rise to the higher molecular weight aggregates. It's possible that octamers form during the detergent exchange process and that the slow dissociation of LMNG from its micelles leaves hydrophobic patches on GF exposed leading to aggregation. However, as discussed below, it is also possible that octamer formation is mediated by the relatively hydrophilic N- and C- termini of the protein.

4.5 Characterizing the GF Octamer

Protein aggregation and disease are closely associated; this is seen in over 40 amyloidopathies such as Alzheimer's disease[362], Creutzfeldt-Jakob disease[363], and Type II diabetes[364]. Although amyloids are usually formed from water-soluble proteins[365], those studying IMPs must also contend with the tendency of IMPs to oligomerize and aggregate.

I optimized conditions where DDM-solubilized GF could be purified as a homotetramer but associated to octamers and higher MW oligomers after 24 hours. By exploring other detergents, I found that LMNG-solubilized GF remained stable for more than one week at room temperature, but appeared to be solubilized in a tetramer-octamer equilibrium. Characterizing the GF octamer was done to identify methods to discourage GF self-association in solution, so we first examined the crystal structure symmetry mates to identify potential GF octamer models. We found that in the crystal, two tetramers stack to form an octamer (see Section 3.2.3) and used negative stain EM and size-exclusion chromatography-small-angle X-ray scattering to gain insight into the LMNG-solubilized octamers.

With resolution limits of approximately 18 Å[366], negative stain EM is often employed as a preliminary step to screen samples for cryo-EM studies. We first attempted to collect images of negatively stained, freshly purified DDM-solubilized GF, which oligomerizes to form octamers and higher M_r aggregates over 24 hours. Negative stain EM

was attempted on a DDM-solubilized GF sample aged for 3 days at room temperature. Since this sample was expected to oligomerize and aggregate in these conditions, we sought to discern whether the precipitates observed were either structured, insoluble, high M_r oligomers of folded GF, or unfolded, aggregated GF oligomers. Unfortunately, the negative stain EM images of this sample did not provide any insight into the nature of the precipitated protein.

Although we were unable to collect useful images of the DDM-solubilized GF aggregates, we were able to collect excellent images of both tetrameric and octameric LMNG-solubilized GF (Figure 59). Through image processing, 2D class averages were constructed, showing views of octamer and tetramer looking down on the periplasmic or cytoplasmic surfaces and the lateral membrane-interacting surface. These views of the protein were verified by the close similarity of features of the images to those of the X-ray diffraction structure and of detergent-solubilized models that we built based on the crystal structure (Figure 57). The 2D class averages of the LMNG-solubilized GF octamer sample revealed the octamer is composed of two stacked tetramers where the interaction surfaces are not parallel. Because of the low resolution of negative stain EM, it was difficult to determine if the tetramers interact *via* their cytoplasmic or periplasmic surfaces. However, electron density between the tetramers in the 2D class averages and the appearance of the "towers" in the images (Figure 62) suggested that the interfaces are the cytoplasmic faces of the tetramers, where the disordered N- and C-termini protrude. This is in line with the model created from the symmetry mates in the crystal structure

and suggests that the polar, disordered termini may be responsible for octamer formation. An explanation for the tilting of the tetramers with respect to each other is not obvious when each tetramer is composed of identical monomers and leads to the suggestion that the orientation of the tetramers in the octamers is an artefact of the protein's interaction with the EM grid or the staining procedure.

While the 2D class average images were very high quality (17 Å resolution), they did not permit a reconstruction of a 3D detergent-solubilized protein structure. This may be explained by a lack of images representing all particle orientations. This in turn may be owing to the protein having a preferred surface for interacting with the EM grid. It is interesting to note that the most recognizable features of the protein are of the membrane-embedded lateral surfaces that interact with lipid in the membrane and detergent in the images. This suggests that the detergent belt may in some way aid in the uptake of stain by the protein.

Size exclusion chromatography-small angle X-ray scattering (SEC-SAXS) is a powerful tool used to probe the low-resolution (between 10-50 Å) structures of biological macromolecules and complexes in solution[185]. SEC-SAXS has also been used to characterize a few integral membrane proteins including photosystem I (PS I) of *Tricondyloides elongatus* where the β-decylmaltoside detergent belt shields the hydrophobic surface of the monomer and homotrimer[367]. Here, SEC-SAXS was used to probe the structure of the LMNG-solubilized GF octamer. We successfully modeled the

GF tetramer and DDM detergent halo using the scattering data collected and the Memprot algorithm[266], confirming the tetramer solution structure. However, the GF octamer and LMNG detergent belt was more difficult to model despite the higher concentrations of protein octamer. A major problem was the inability of Memprot to model two detergent belts, but it's also possible that, unlike the tetramer, the SEC could not adequately separate the octamer from higher M_r aggregates and/or tetramers (Figure 63A). Other attempts to obtain better signal from the GF octamer included increasing the concentration of the LMNG-solubilized GF sample but this resulted in larger amounts of high M_r GF oligomers and did not increase the concentration of octameric GF.

Despite these problems, by building GF octamer models with an LMNG detergent belt using the Micelle builder algorithm in CHARMM-GUI[267], we were able to calculate hypothetical scattering curves based on the models and compare the fits of the hypothetical and experimental scattering curves. It's interesting to note that the best fitting octamer models were comprised of two tetramers with cytoplasmic or periplasmic surfaces tilted with respect to one another in a manner highly reminiscent to what is observed in the 2D class average negative stain EM images (Figure 60). However, the SEC-SAXS octamer data could not support a cytoplasmic over periplasmic surface interaction. One obvious area for improving the models is to build detergent belts with greater amounts of detergent. Recall that the very successful models of the tetramer contained about twice the mass of detergent per protein molecule as in the octamer models. It is of note also that the direct calculation of electron density from SAXS scattering data, DENSS

(<u>https://www.tdgrant.com/denss/</u>) was recently expanded to be able to calculate the structures of membrane proteins with detergent belts, and could be used here to better model the detergent belts around the tetramers in the octamer structure. Additionally, the calculation the R_g value, which gives the spherical average of the particle causing the x-ray scattering[368] should be considered as both the octamer and tetramer models of LMNG-solubilized GF are not perfectly spherical. Nevertheless, based on the structure of the model created from the crystal structure symmetry mates and the negative stain EM 2D class averages, it appears that in solution, the octamers are formed by the GF tetramers with their cytoplasm-facing surfaces at the interface.

In characterizing the GF octamer, we hoped to identify the cause of the self-association which might give insight into methods that would discourage the formation of the octamer. We initially surveyed the GF crystal lattice searching for potential residues that could interact with each other and cause the self-association. It was our thought that by introducing point mutations at sites likely to interact at the interface, the octamer formation could be prevented. We found, however, that in the crystal structures the lateral, lipid-interacting surfaces of the tetramers were too far apart to interact; no residues were found at the interfaces that might interact with the tetramer opposite of it. The closest approach of two tetramers in the crystal lattice is between the cytoplasmic surfaces (Figure 57). However, no interacting residues are observable, presumably because the disordered termini that protrude from the cytoplasmic surfaces are not visible in the structure. This further supports the suggestion that the disordered N- and

C-termini play a role in the formation of the octamer. It is possible to speculate that the elongated N-terminus containing the His$_6$ tag, TEV cleavage site, and linker region causes the tetramers to "tilt" towards each other on one end to accommodate the extra residues. In our preparations, in contrast to the parallel orientation observed in the crystal structure, it's difficult to propose a mechanism for building such an asymmetric structure from the symmetric homotetramer. If the disordered termini are responsible, it might be possible to replicate this in MD simulations of tetramers that include the disordered termini and learn the atomic structural basis of the asymmetry.

Crystallographers often truncate the N- and C-termini of their target proteins to encourage crystallization[369]. A C-terminal truncation was explored here to determine whether the disordered termini were interacting and causing tetramer association. The C-terminus of the GF monomer was chosen as it is longer than the N-terminus and attempts to remove the N-terminal His$_6$ tag *via* TEV cleavage were unsuccessful. BL21(DE3) cells containing the plasmid coding for the C-terminal truncated GF gene grew poorly in rich media; when the plasmid was transformed into C43 cells, robust growth was observed. C41 and C43 strains (Walker strains) are derivatives of BL21(DE3) cells that were selected by exhibiting robust growth despite overexpressing the bovine oxoglutarate-malate transport protein (OMTP)[370]. The C43 cells are C41 cells that tolerate overexpressing OMTP and *E. coli* F-type ATPase subunit b. Though the C43 cells with the plasmid containing the GF C-terminal truncation grew robustly, no protein could be purified. It may be that the C-terminus is required for proper insertion into the

membrane or for proper folding of GF, or that truncation leads to proteolytic

degradation[327,371]; if GF is improperly folded, it may be forming inclusion bodies instead[372].

4.6 ssNMR Spectra of LMNG-solubilized GF

Significant progress in ssNMR spectroscopy methods and instrumentation[146] has allowed for the structures of lipid bilayer-embedded IMPs as large as 350 residues to be solved[373]. Using the solid-state regime, IMPs that were previously too large to study with liquid NMR spectroscopy can now be probed for their structure and dynamics[140]. Here, preliminary MAS ssNMR spectra of LMNG-solubilized GF were obtained and several resonances were identified (Section 3.3). The 2D ^{13}C-^{13}C dipolar assisted rotational resonance (DARR) spectra show potential for measuring GF dynamics via ssNMR spectroscopy, but the measured ^{13}C transverse relaxation rates (T_2) suggest that the LMNG-solubilized sample requires more optimization to obtain a homogenous sample. Thus, despite the extensive screening of detergents, solubilizing agents, and membrane mimetics presented here, LMNG may be a relatively poor mimic of the native lipid bilayer. Hydrophobic mismatch and lipid acyl chain ordering (and the presence of lipids in general) have been discussed extensively in the context of IMP folding[374], native self-association within the lipid bilayer[375], and the stability of IMPs in general[141,376,377]. It is well known that bilayer-dissolved IMPs require careful optimization of the lipid:protein ratio to prevent static protein conformational heterogeneity and line-broadened spectra[378,379]. While it is conceivable that LMNG-solubilized GF might be further optimized for the measurement of ssNMR spectra, the physiological relevance of a bilayer environment and the success in obtaining ssNMR spectra of other IMPS in bilayers suggest that the best avenue forward will be solubilization of GF in lipid vesicles, bicelles, or nanodiscs.

2D DARR was used in part to report the complete MAS ssNMR resonance assignment of microcrystalline YadA membrane protein. YadA is a 105 residue trimeric protein involved in the infection pathway of some food-borne diseases caused by *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*[379]. High quality 2D DARR spectra were also obtained for *Bacillus subtilis* BmrA reconstituted into liposomes, a 130 kDa homodimeric membrane protein that plays a role in multidrug resistance[380]. The newly developed TIDE ($\underline{T}_2$* weighted $\underline{de}$convolution) method has been shown to attain higher resolution and sensitivity when used for measuring 2D DARR spectra on small, bitopic IMPs and may be applicable to larger IMPs, like the 6TM *E. coli* acetate-succinate helicase[381]. These studies support the idea that studies probing the dynamics of IMPs such as GF are imminent.

4.7 Future Work and Conclusions

The ability to relatively inexpensively make large amounts of ^{13}C-, ^{15}N-labelled LMNG-solubilized GF in combination with further developments in ssNMR spectroscopy will, in all likelihood, lead to high quality ssNMR spectra and the eventual resonance assignment of the dynamic structure of GF. As described in Section 4.6, high quality spectra of BmrA were obtained of the protein reconstituted in proteoliposomes[380]; it would be prudent, therefore, to further explore the reconstitution of GF in bicelles, liposomes, or nanodiscs, perhaps by first solubilizing in LMNG to maintain the stability of GF during the reconstitution process. The benefit in studying bicelle- or liposome-solubilized GF is two-fold: based on the success of ssNMR spectroscopy of other IMPs, these environments may be more amenable to acquiring higher quality ssNMR spectra, and an environment that closely mimics the IMP's native environment is expected to support the retention of the IMPs native conformation and prevent aggregation.

The information obtained from comparing the structures of detergent- and lipid bilayer-solubilized GF samples would also be beneficial. It has been suggested that differences between the environments provided by detergent compared to lipid bilayer-solubilization can impact factors such as lateral pressure and hydrophobic mismatch[58,344,376] and cause conformational changes that translate to a folded, detergent-solubilized but inactive protein[210,211,355]. Understanding at the molecular level the differences in the structure and dynamics of IMPs in detergents and bilayers would contribute significantly to our understanding of membrane protein folding and might

219

have practical benefits, such as improving the preparation of IMPs for structure and dynamics studies. Accompanying this are outstanding questions about the formation of the GF octamer and the underlying causes for association of the tetramers and formation of higher M_r aggregates. Though attempts to increase the concentration of octamer resulted in higher than usual amounts of soluble aggregate, it would be valuable to further investigate the nature of octamer formation and what role, if any, its formation plays in GF aggregation.

The overall goal of this work was to probe the dynamics of *E. coli* integral membrane protein glycerol facilitator. Here, we presented the overexpression and purification of ^{13}C, ^{15}N-labelled LMNG-solubilized GF for MAS ssNMR spectroscopy. We were able to obtain 2D DARR spectra where multiple resonances could be identified. We also presented the characterization of the GF octamer and explored the potential cause of the observed self-association of the tetramers. To this end, both negative stain EM 2D class averages and SEC-SAXS models supported the original octamer model proposed: the octamers are offset, stacked tetramers that are tilted towards each other on one end of the interface with what is likely their cytoplasmic-facing surfaces, and both the disordered N- and C-termini protrude into the space between the tetramers.

(1)   Alberts, B.; Johnson, A.; Lewis, J.; Al, E. Membrane Proteins. In *Molecular Biology of the Cell*; 2002.

(2)   Singer, A. S. J.; Nicolson, G. L. The Fluid Mosaic Model of the Structure of Cell Membranes. *Science (80-.).* **1972**, *175* (4023), 720–731.

(3)   Secundus Plinus, G. *The Natural History of Pliny.*, 6th ed.; Bostock, J., Riley, H., Eds.; London, 77AD.

(4)   Iv, X. L.; Franklin, B.; Brownrigg, F. R. S. W.; Farifti, M. Of the Stilling of Waves by means of oil.

(5)   Ortega Gómez, A.; Varela, L.; Bermúdez, B.; Lopez, S.; J.G. Muriana, F.; Abia, R. Nutrigenomics and Atherosclerosis: The Postprandial and Long-Term Effects of Virgin Olive Oil Ingestion; 2012. https://doi.org/10.5772/26141.

(6)   Langmuir, I. The Constitution and Fundamental Properties of Solids and Liquids II. *J. Am. Chem. Soc.* **1917**, *39*, 1848–1906.

(7)   Stillwell, W. *An Introduction to Biological Membranes*, 2nd ed.; Elsevier B.V.: Oxford, 2016.

(8)   Hooke, R. *Microphagia, or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses, with Observations and Inquiries Thereupon.*, First.; Matyn, J., Allestry, J., Eds.; London, 1665.

(9)   Kleinzeller, A. William Hewson's Studies of Red Blood Corpuscles and the Evolving Concept of a Cell Membrane. *Am. J. Physiol. Physiol.* **1996**, *271* (1), C1–C8. https://doi.org/10.1152/ajpcell.1996.271.1.c1.

(10)   Kleinzeller, A. Ernest Overton's Contribution to the Cell Membrane Concept: A Centennial Appreciation. *Physiology* **1997**, *12* (1), 49–53. https://doi.org/10.1152/physiologyonline.1997.12.1.49.

(11)   Kleinzeller, A. Charles Ernest Overton's Concept of a Cell Membrane. In *Current Topics in Membranes*; Deanmer, D., Kleinzeller, A., Fambrough, D., Eds.; 1999; pp 1–22.

(12)   Gorter, E.; Grendel, F. On Bimolecular Layers of Lipoids on the Chromocytes of the Blood. *J. Exp. Med.* **1925**, *41*, 439–443.

(13)   Leathes, J. B.; Raper, H. S. *The Fats*; 1925.

(14)   Fricke, H. The Electrical Capacity of Suspensions with Special Reference to Blood. *J. Gen. Physiol.* **1925**, *9*, 137–152.

(15)   Sumner, J. B. The Isolation and Crystallization of the Enzyme Urease. *J. Biol. Chem.* **1926**, *69*, 435–441.

(16)   De Weer, P. A Century Thinking about Cell Membranes. *Annu. Rev. Physiol.* **2000**, *62* (919–926).

(17)   Danielli, J.; Davson, H. A Contribution to the Theory of Permeability of Thin Films. *J. Cell. Comp. Physiol.* **1935**, *5* (495–508).

(18)   Benson, A. A. On the Orientation of Lipids in Chloroplast and Cell Membranes. *J. Am. Oil Chem. Soc.* **1966**, *43* (5), 265–270. https://doi.org/10.1007/BF02609671.

(19)   Vanderkooi, G.; Green, D. E. Biological Membrane Structure, I. The Protein Crystal Model for Membranes. *Proc. Natl. Acad. Sci.* **1970**, *66* (3), 615 LP – 621. https://doi.org/10.1073/pnas.66.3.615.

(20)   Kühlbrandt, W. Structure and Function of Mitochondrial Membrane Protein Complexes. *BMC Biol.* **2015**, *13* (1), 89. https://doi.org/10.1186/s12915-015-0201-x.

(21)   Ongun, A.; Thomson, W. W.; Mudd, J. B. Lipid Composition of Chloroplasts Isolated by Aqueous and Nonaqueous Techniques. *J. Lipid Res.* **1968**, *9* (4), 409–415.

(22)   Lombard, J. Once upon a Time the Cell Membranes: 175 Years of Cell Boundary Research. *Biol. Direct* **2014**, *9*, 32. https://doi.org/10.1186/s13062-014-0032-7.

(23)   Encrenaz, T. Water in the Solar System. *Annu. Rev. Astron. Astrophys.* **2008**, *46*, 57–87.

(24)   Bartell, F. E.; Case, L. O.; Brown, H. The Surface Tension of Mercury and of Water in Contact with Saturated Vapors of Organic Liquids1. *J. Am. Chem. Soc.* **1933**, *55* (7), 2769–2776. https://doi.org/10.1021/ja01334a021.

(25)   Carpenter, I.; Hehre, W. A Molecular Dynamics Study of the Hexane/Water Interface. *J. Phys. Chem.* **1990**, 531–536.

(26)   Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Molecules.*; John Wiley & Sons: New York, 1973.

(27)   Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640–647.

(28)   Murzyn, K.; Rog, T.; Pasenkiewicz-gierula, M. Phosphatidylethanolamine-Phosphatidylglycerol Bilayer as a Model of the Inner Bacterial Membrane. *Biophys. J.* **2005**, *88*, 1091–1103. https://doi.org/10.1529/biophysj.104.048835.

(29)   Popot, J.-L. *Membrane Proteins in Aqueous Solutions*; 2018. https://doi.org/10.1007/978-3-319-73148-3.

(30)   Vance, J. E.; Tasseva, G. Formation and Function of Phosphatidylserine and Phosphatidylethanolamine in Mammalian Cells. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **2013**, *1831* (3), 543–554. https://doi.org/10.1016/j.bbalip.2012.08.016.

(31)   Vanounou, S.; Parola, A. H.; Fishov, I. Phosphatidylethanolamine and Phosphatidylglycerol Are Segregated into Different Domains in Bacterial Membrane . A Study with Pyrene-Labelled Phospholipids. **2003**, *49*, 1067–1079. https://doi.org/10.1046/j.1365-2958.2003.03614.x.

(32) Hendrick, J. P.; Wickner, W. SecA Protein Needs Both Acidic Phospholipids and SecY/E Protein for Functional High-Affinity Binding to the Escherichia Coli Plasma Membrane. *J. Biol. Chem.* **1991**, *266* (36), 24596–24600.

(33) Van Klompenburg, W.; De Kruijff, B. The Role of Anionic Lipids in Protein Insertion and Translocation in Bacterial Membranes. *J. Membr. Biol.* **1998**, *162* (1), 1–7. https://doi.org/10.1007/s002329900336.

(34) Shokri, A.; Larsson, G. Characterisation of the Escherichia Coli Membrane Structure and Function during Fedbatch Cultivation. *Microb. Cell Fact.* **2004**, *3*, 1–12. https://doi.org/10.1186/1475-2859-3-9.

(35) Sorice, M.; Manganelli, V.; Matarrese, P.; Tinari, A.; Misasi, R.; Malorni, W.; Garofalo, T. Cardiolipin-Enriched Raft-like Microdomains Are Essential Activating Platforms for Apoptotic Signals on Mitochondria. *FEBS Lett.* **2009**, *583* (15), 2447–2450. https://doi.org/10.1016/j.febslet.2009.07.018.

(36) Paradies, G.; Paradies, V.; De Benedictis, V.; Ruggiero, F.; Petrosillo, G. Functional Role of Cardiolipin in Mitochondrial Bioenergetics. *Biochim. Biophys. Acta - Bioenerg.* **2014**, *1837*, 408–417.

(37) Pangborn, M. The Composition of Cardiolipin. *J. Biol. Chem.* **1947**, 351–361.

(38) Fobker, M.; Voss, R.; Reinecke, H.; Crone, C.; Assmann, G.; Walter, M. Accumulation of Cardiolipin and Lysocardiolipin in Fibroblasts from Tangier Disease Subjects. *FEBS Lett.* **2001**, *500*, 157–162.

(39) Montiero-Cardoso, V. F.; Oliviera, M. M.; Melo, T.; Domingues, M.; Moriera, P.; Ferreiro, E.; Piexoto, F.; Videira, R. Cardiolipin Profile Changes Are Associated to the Early Synaptic Mitochondrial Dysfunction in Alzheimer's Disease. *J. Alzheimer's Dis.* **2015**, *43*, 1375–1392.

(40) Houtkooper, R. H.; Vaz, F. M. Cardiolipin, the Heart of Mitochondrial Metabolism. *Cell. Mol. Life Sci.* **2008**, *65* (16), 2493–2506. https://doi.org/10.1007/s00018-008-8030-5.

(41) Jain, M. K.; Wu, N. M. Effect of Small Molecules on the Dipalmitoyl Lecithin Liposomal Bilayer: III. Phase Transition in Lipid Bilayer. *J. Membr. Biol.* **1977**, *34* (1), 157–201.

(42) Brown, D. A.; London, E. Functions of Lipid Rafts in Biological Membranes. *Annu. Rev. Cell Dev. Biol.* **1998**, *14*, 111–136.

(43) Harris, T.; Siu, C.-H. Reciprocal Raft-Receptor Interactions and the Assembly of Adhesion Complexes. *BioEssays* **2002**, *24*, 996–1003.

(44) Heerklotz, H. Triton Promotes Domain Formation in Lipid Raft Mixtures. *Biophys. J.* **2002**, *83*, 2693–2701.

(45) Munro, S. Lipid Rafts : Elusive or Illusive ? Review. **2003**, *115*, 377–388.

(46)   Cheng, X.; Smith, J. C. Biological Membrane Organization and Cellular Signaling. *Chem. Rev.* **2019**, *119* (9), 5849–5880. https://doi.org/10.1021/acs.chemrev.8b00439.

(47)   Levental, I.; Veatch, S. L. The Continuing Mystery of Lipid Rafts. *J. Mol. Biol.* **2016**, *428* (24), 4749–4764. https://doi.org/10.1016/j.jmb.2016.08.022.

(48)   Nickels, J. D.; Cheng, X.; Mostofian, B.; Stanley, C.; Lindner, B.; Heberle, F. A.; Perticaroli, S.; Feygenson, M.; Egami, T.; Standaert, R. F.; et al. Mechanical Properties of Nanoscopic Lipid Domains. *J. Am. Chem. Soc.* **2015**, *137* (50), 15772–15780. https://doi.org/10.1021/jacs.5b08894.

(49)   Nickels, J. D.; Chatterjee, S.; Stanley, C. B.; Qian, S.; Cheng, X.; Myles, D. A. A.; Standaert, R. F.; Elkins, J. G.; Katsaras, J. The in Vivo Structure of Biological Membranes and Evidence for Lipid Domains. *PLoS Biol.* **2017**, *15* (5), 1–22. https://doi.org/10.1371/journal.pbio.2002214.

(50)   Mileykovskaya, E.; Sun, Q.; Margolin, W.; Dowhan, W. Localization and Function of Early Cell Division Proteins in Filamentous Escherichia Coli Cells Lacking Phosphatidyl Ethanolamine. *J. Bacteriol.* **1998**, *180*, 4252–4257.

(51)   Mileykovskaya, E.; Dowhan, W. Visualization of Phospholipid Domains in Escherichia Coli by Using the Cardiolipin-Specific Fluorescent Dye 10- N -Nonyl Acridine Orange. **2000**, *182* (4), 1172–1175.

(52)   Scott, H. L.; Lee, C. Y. The Surface Tension of Lipid Water Interfaces: Monte Carlo Simulation. *J. Chem. Phys.* **1980**, *73* (10), 5351–5353. https://doi.org/10.1063/1.439965.

(53)   Söderlund, T.; Alakoskela, J. M. I.; Pakkanen, A. L.; Kinnunen, P. K. J. Comparison of the Effects of Surface Tension and Osmotic Pressure on the Interfacial Hydration of a Fluid Phospholipid Bilayer. *Biophys. J.* **2003**, *85* (4), 2333–2341. https://doi.org/10.1016/S0006-3495(03)74657-8.

(54)   Dickson, C. J.; Lin, C.; Madej, B. D.; Teigen, K.; Gould, I. R.; Walker, R. C.; Skjevik, Å. A. Simulation of Lipid Bilayer Self-Assembly Using All-Atom Lipid Force Fields. *Phys. Chem. Chem. Phys.* **2016**, *18* (15), 10573–10584. https://doi.org/10.1039/c5cp07379k.

(55)   Alberts, B.; Johnson, A.; Lewis, J. The Lipid Bilayer. In *Molecular Biology of the Cell*; Garland Science: New York, 2002.

(56)   BARTHOLOMEW, J. W.; MITTWER, T. The Gram Stain. *Bacteriol. Rev.* **1952**, *16* (1), 1–29.

(57)   Fagerberg, L.; Jonasson, K.; von Heijne, G.; Uhlen, M.; Berglund, L. Prediction of the Human Membrane Proteome. *Proteomics* **2010**, *10*, 1141–1149.

(58)   Enkavi, G.; Javanainen, M.; Kulig, W.; Rog, T.; Vattulainen, I. Multiscale Simulations of Biological Membranes : The Challenge to Understand Biological Phenomena in a Living Substance. *Chem. Rev.* **2018**.

https://doi.org/10.1021/acs.chemrev.8b00538.

(59) Garber, E.; Margoliash, E. Interaction of Cytochrome c with Cytochrome c Oxidase: An Understanding of the High - to - Low Affinity Transition. *Biochim. Biophys. Acta* **1990**, No. 279–287.

(60) Boggs, J. *Myelin Basic Protein.*; Nova Scientific Publishers: Hauppauge, 2008.

(61) Vit, O.; Petrak, J. Integral Membrane Proteins in Proteomics. How to Break Open the Black Box? *J. Proteomics* **2017**, *153*, 8–20.

(62) Garrow, A.; Agnew, A.; Westhead, D. TMB-HUNT: An Amino Acid Composition Based Method to Screen Proteomes for Beta-Barrel Transmembrane PRoteins. *BMC Bioinformatics* **2005**, *6*.

(63) Wu, T.; McCandlish, A.; Gronenberg, L.; Chng, S.; Silhavy, T.; Kahne, D. Identification of a Protein Complex That Assembles Lipopolysaccharide in the Outer Membrane of Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *31*, 11754–11759.

(64) Benz, R.; Schmid, A.; Van der Ley, P.; Tommassen, J. Molecular Basis of Porin Selectivity: Membrane Experiments with OmpC-PhoE and OmpF-PhoE Hybrid Proteins of Escherichia Coli K-12. *Biochim. Biophys. Acta* **1989**, 8–14.

(65) Saitoh, T.; Igura, M.; Obita, T.; Ose, T.; Kojima, R.; Maenaka, K.; Endo, T.; Kohda, D. Tom20 Recognizes Mitochondrial Presequences through Dynamic Equilibrium among Multiple Bound States. *EMBO J.* **2007**, *26* (22), 4777–4787. https://doi.org/10.1038/sj.emboj.7601888.

(66) Bausewein, T.; Mills, D. J.; Langer, J. D.; Nitschke, B.; Nussberger, S.; Kühlbrandt, W. Cryo-EM Structure of the TOM Core Complex from Neurospora Crassa. *Cell* **2017**, *170* (4), 693-700.e7. https://doi.org/https://doi.org/10.1016/j.cell.2017.07.012.

(67) Krüger, V.; Becker, T.; Becker, L.; Montilla-Martinez, M.; Ellenrieder, L.; Vögtle, F.-N.; Meyer, H. E.; Ryan, M. T.; Wiedemann, N.; Warscheid, B.; et al. Identification of New Channels by Systematic Analysis of the Mitochondrial Outer Membrane. *J. Cell Biol.* **2017**, *216* (11), 3485 LP – 3495. https://doi.org/10.1083/jcb.201706043.

(68) Koebnik, R.; Locher, K.; Van Gelder, P. Structure and Function of Bacterial Outer Membrane Proteins: Barrels in a Nutshell. *Moleular Microbiol.* **2000**, *37* (2), 239–253.

(69) Maddy, A. H.; Malcolm, B. R. Protein Conformations in the Plasma Membrane. *Science (80-.).* **1965**, *150* (3703), 1616–1618.

(70) Lenard, J.; Singer, S. J. Protein Conformation in Cell Membrane Preparations as Studied by Optical Rotatory Dispersion and Circular Dichroism. *PNAS* **1966**, *56* (6), 1828–1835.

(71) Henderson, R.; Unwin, P. N. T. Three-Dimensional Model of Purple Membrane

Obtained by Electron Microscopy. *Nature* **1975**, *257* (5521), 28–32. https://doi.org/10.1038/257028a0.

(72)   Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. Structure of the Protein Subunits in the Photosynthetic Reaction Centre of Rhodopseudomonas Viridis at 3A Resolution. *Nature* **1985**, 618–624.

(73)   Dobson, L.; Remenyi, I.; Tusnady, G. The Human Transmembrane Proteome. *Biol. Direct* **2015**, *10*.

(74)   Lomize, A.; Lomize, M.; Krolicki, S.; Pogozheva, I. Membranome: A Database for Proteome-Wide Analysis of Single-Pass Membrane Proteins. *Biochim. Biophys. Acta* **2017**, D250–D255.

(75)   Hubert, P.; Sawma, P.; Duneau, J.; Khao, J. Single-Spanning Transmembrane Domains in Cell Growth and Cell-Cell Interactions. *Cell Adhes. Migr.* **2010**, *4*, 313–324.

(76)   Zviling, M.; Kochva, U.; Arkin, I. T. How Important Are Transmembrane Helices of Bitopic Membrane Proteins ? **2007**, *1768*, 387–392. https://doi.org/10.1016/j.bbamem.2006.11.019.

(77)   Hilger, D.; Masureel, M.; Kobilka, B. Structure and Dynamics of GPCR Signaling Complexes. *Nat. Struct. Mol. Biol.* **2018**, 4–12.

(78)   Agre, P.; Kozono, D. Aquaporin Water Channels: Molecular Mechanisms for Human Diseases. *FEBS Lett.* **2003**, 72–78.

(79)   Malnic, B.; Godfrey, P. A.; Buck, L. B. The Human Olfactory Receptor Gene Family. *Proc. Natl. Acad. Sci.* **2004**, *101* (8), 2584–2589. https://doi.org/10.1073/PNAS.0307882100.

(80)   Palczewski, K. G-Protein Coupled Receptor Rhosdopsin. *Annu. Rev. Biochem.* **2006**, *75*, 743–767.

(81)   Lefkowitz, R. A Magnificent Time With the "Magnificent Seven" Transmembrane Spanning Receptors. *Circ. Res.* **2003**, *92* (4), 342–344. https://doi.org/10.1161/01.RES.0000061771.61280.B8.

(82)   Danielson, J. A.; Johanson, U. Phylogeny of Major Intrinsic Proteins.

(83)   Preston, G.; Carroll, T.; Guggino, W.; Agre, P. Appearance of Water Channels in Xenopus Oocytes Expressing Red Cell CHIP28 Protein. *Science (80-.).* **1992**, *256*, 385–387.

(84)   Sasaki, S. Expanding the World of Aquaporins: New Members and New Functions. *Eur. J. Physiol.* **2008**, 647–649.

(85)   Fu, D.; Libson, A.; Miercke, L. J.; Weitzman, C.; Nollert, P.; Krucinski, J.; Stroud, R. M. Structure of a Glycerol-Conducting Channel and the Basis for Its Selectivity.

Science (80-.). **2000**, *290* (5491), 481–486. https://doi.org/8914 [pii].

(86) Klein, N.; Neumann, J.; O'Neil, J. D.; Schneider, D. Folding and Stability of the Aquaglyceroporin GlpF: Implications for Human Aqua (Glycero) Porin Diseases. *Biochim. Biophys. Acta (BBA)-Biomembranes* **2014**, *1848*, 622–633.

(87) Neumann, J.; Klein, N.; Schneider, D. Membranproteinfaltung: Wie Falten Membranproteine? - Schritt Für Schritt? *BioSpektrum* **2012**, *18* (2), 146–148. https://doi.org/10.1007/s12268-012-0154-3.

(88) Hub, J.; Grubmuller, H.; de Groot, B. Dynamic and Energetics of Permeation through Aquaporins. What Do We Learn from Molecular Dynamics Simulations? *Handb. Exp. Pharmacol.* **2009**, *190*, 57–76.

(89) Agmon, N. The Grotthuss Mechanism. *Chem. Phys. Lett.* **1995**, *244*, 456–462.

(90) Kong, Y.; Ma, J. Dynamic Mechanisms of the Membrane Water Channel Aquaporin-1 (AQP1). *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 14345–14349.

(91) Pan, Y.; Piyadasa, H.; Neil, J. D. O.; Konermann, L.; Al., E. Conformational Dynamics of a Membrane Transport Protein Probed by H/D Exchange and Covalent Labeling: The Glycerol Facilitator. *J. Mol. Biol.* **2012**, *416* (3), 400–413. https://doi.org/10.1016/j.jmb.2011.12.052.

(92) Lu, D.; Grayson, P. Glycerol Conductance and Physical Asymmetry of the Escherichia Coli Glycerol Facilitator GlpF. *Biophys. J.* **2003**, *85*, 2977–2987.

(93) Ishibashi, K.; Kuwahara, M.; Sasaki, S. Molecular Biology of Aquaporins. In *Reviews of Physiology Biochemistry and Pharmacology: Special Issue on Water Transport Across Biological Membranes*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2000; pp 1–32. https://doi.org/10.1007/BFb0119576.

(94) Verkman, A. S. Aquaporins at a Glance. *J. Cell Sci.* **2011**, *124* (13), 2107–2112. https://doi.org/10.1242/jcs.079467.

(95) Verkman, A. S.; Anderson, M.; Papadopoulos, M. Aquaporins: Important but Elusive Drug Targets. *Nat. Rev.* **2014**, *13* (259–277).

(96) Hara-Chikuma, M.; Verkman, A. Roles of Aquaporin-3 in Epidermis. *J. Invest. Dermatol.* **2008**, *128*, 2145–2151.

(97) Ma, T.; Hara, M.; Sougrat, R.; Verbavatz, J.; Verkman, A. Impaired Stratum Corneum Hydration in Mice Lacking Epidermal Water Channel Aquaporin-3. *J. Biol. Chem.* **2002**, *277*, 17147–17153.

(98) Sohara, E.; Uchida, S.; Sasaki, S. Function of Aquaporin-7 in the Kidney and the Male Reproductive System. *Handb. Exp. Pharmacol.* **2009**, No. 190, 219–231. https://doi.org/10.1007/978-3-540-79885-9_11.

(99) Noda, Y.; Sohara, E.; Ohta, E.; Sasaki, S. Aquaporins in Kidney Pathophysiology.

Nat. Rev. Nephrol. **2010**, *6*, 168–178.

(100) Carbrey, J.; Gorelick-Feldman, D.; Kozono, D.; Praetorius, J. .; Nielsen, S.; Agre, P. Aquaglyceroporin AQP9: Solute Permeation and Metabolic Control of Expression in Liver. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2945–2950.

(101) Carbrey, J.; Song, L.; Zhou, Y.; Yoshinaga, M.; Rojek, A.; Wang, Y.; Liu, Y.; Lujan, H.; DiCarlo, S.; Nielsen, S. Reduced Arsenic Clearance and Increased Toxicity in Aquaglyceroporin-9-Null Mice. *Proc. Natl. Acad. Sci. United States Am.* **2009**, *106*, 15956–15960.

(102) Maeda, N.; Hibuse, T.; Funahashi, T. Role of Aquaporin-7 and Aquaporin-9 in Glycerol Metabolism; Involvement in Obesity. *Handb. Exp. Pharmacol.* **2009**, *190*, 233–249. https://doi.org/10.1007/978-3-540-79885-9_12.

(103) King, L.; Nielsen, S.; Agre, P.; Brown, R. Decreased Pulmonary Cascular Permeability in Aquaporin-1-Null Humans. *PNAS* **2002**, *99*, 1059–1063.

(104) Engel, A.; Fujiyoshi, Y.; Gonen, T.; Walz, T. Junction-Forming Aquaporins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 229–235.

(105) Berry, V.; Francis, P.; Kaushal, S.; Moore, A.; Bhattacharya, S. Missense Mutations in MIP Underlie Autosomal Dominant "polymorphic" and Lamellar Cataracts Linked to 12q. *Nat. Genet.* **2000**, *25*, 15–17.

(106) Chepelinsky, A. Structural Function of MIP/Aquaporin 0 in the Eye Lens; Genetic Defects Lead to Congenital Inherited Cataracts. *Handb. Exp. Pharmacol.* **2009**, *190*, 265–297.

(107) Bichet, D. Hereditary Polyuric Disorders: New Concepts and Differential Diagnosis. *Semin. Nephrol.* **2006**, *26*, 224–233.

(108) Nedvetsky, P.; Tamma, G.; Beulshausen, S.; Valenti, G.; Rosenthal, W.; Klussman, E. Regulation of Aquaporin-2 Trafficking. *Handb. Exp. Pharmacol.* **2009**, *190*, 133–157.

(109) Lennon, V.; Kryzer, T.; Pittock, S.; Verkman, A.; Hinson, S. IgG Marker of Optic-Spinal Multiple Sclerosis Binds to the Aquaporin-4 Water Channel. *J. Exp. Med.* **2005**, *202*, 473–477.

(110) Wingerchuk, D.; Lennon, V.; Lucchinetti, C.; Pittock, S.; Weinshenker, B. The Spectrum of Neuromyelitis Optica. *Lancet Neurol.* **2007**, *6*, 805–815.

(111) Pittock, S. J.; Lennon, V. A.; McKeon, A.; Mandrekar, J.; Weinshenker, B. G.; Lucchinetti, C. F.; O'Toole, O.; Wingerchuk, D. M. Eculizumab in AQP4-IgG-Positive Relapsing Neuromyelitis Optica Spectrum Disorders: An Open-Label Pilot Study. *Lancet Neurol.* **2013**, *12* (6), 554–562. https://doi.org/https://doi.org/10.1016/S1474-4422(13)70076-0.

(112) Molnar, J.; Szakacs, G.; Tusnady, G. Characterization of Disease-Associated Mutations in Human Transmembrane Proteins. *PLoS One* **2016**, *11*.

(113) Ng, D.; Poulsen, B.; Deber, C. Membrane Protein Misassembly in Disease. *Biochim. Biophys. Acta* **2012**, *1818*, 1115–1122.

(114) Yin, H.; Flynn, A. D. Drugging Membrane Protein Interactions. *Annu. Rev. Biomed. Eng.* **2016**, *18* (1), 51–76. https://doi.org/10.1146/annurev-bioeng-092115-025322.

(115) Lu, S.; Zhang, J. Small Molecule Allosteric Modulators of G-Protein-Coupled Receptors: Drug-Target Interactions. *J. Med. Chem.* **2018**.

(116) Bätz, J.; Müller, A.; Tränkle, C.; De Amici, M.; Schrage, R.; Dallanoce, C.; Schmitz, J.; Bock, A.; Hoffmann, C.; Schrobang-Ley, J.; et al. The Allosteric Vestibule of a Seven Transmembrane Helical Receptor Controls G-Protein Coupling. *Nat. Commun.* **2012**, *3* (1). https://doi.org/10.1038/ncomms2028.

(117) Christopoulos, A.; Kenakin, T. G Protein-Coupled Receptor Allosterism and Complexing. *Pharmacol. Rev.* **2002**, *54*, 323–374.

(118) Gong, J.; Chen, Y.; Pu, F.; Sun, P.; He, F.; Zhang, L.; Li, Y.; Ma, Z.; Wang, H. Understanding Membrane Protein Drug Targets in Computational Perspective. *Curr. Drug Targets* **2019**, *20*, 551–564.

(119) Vetter, S.; Leclerc, E. Novel Aspects of Calmodulin Target Recognition and Activation. *Eur. J. Biochem.* **2003**, 404–414.

(120) Liddington, R. A Molecular Full Nelson. *Nature* **2002**, 373–374.

(121) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450* (7172), 964–972. https://doi.org/10.1038/nature06522.

(122) Orozco, M. A Theoretical View of Protein Dynamics. *Chem. Soc. Rev.* **2014**, *43* (14), 5051–5066. https://doi.org/10.1039/c3cs60474h.

(123) Buldyrev, S. V.; Chen, S.-H.; Corsaro, C.; Baglioni, P.; Mallamace, F.; Mallamace, D.; Vasi, C.; Vasi, S.; Stanley, H. E. Energy Landscape in Protein Folding and Unfolding. *Proc. Natl. Acad. Sci.* **2016**, *113* (12), 3159–3163. https://doi.org/10.1073/pnas.1524864113.

(124) White, S. Membrane proteins of known structure. http://blanco.biomol.uci.edu/mpstruc/.

(125) Ilari, A.; Savino, C. Protein Structure Determination by X-Ray Crystallography. *Methods Mol. Biol.* **2008**, *452*, 63–87.

(126) Hirschi, M.; Johnson, Z.; Lee, S. Visualizing Multistep Elevator-like Transitions of a Nucleoside Transporter. *Nature* **2017**, *545*, 66–70.

(127) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.

(128) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci. U. S. A.* **2018**,

115 (21), E4758–E4766. https://doi.org/10.1073/pnas.1800690115.

(129) Sansom, M.; Scott, K.; Bond, P. Coarse-Grained Simulation: A High-Throughput Computational Approach to Membrane Proteins. *Biochem. Soc. Trans.* **2008**, *36*, 27–32.

(130) Goossens, K.; De Winter, H. Molecular Dynamics Simulations of Membrane Proteins: An Overview. *J. Chem. Inf. Model.* **2018**, *58* (11), 2193–2202. https://doi.org/10.1021/acs.jcim.8b00639.

(131) Gidden, J.; Denson, J.; Liyanage, R.; Ivey, D.; Lay, J. Lipid Compositions in Escherichia Coli and Bacillus Subtilis during Growth as Determined by MALDI-TOF and TOF/TOF Mass Spectrometry. *Int. J. Mass Spectrom.* **2009**, *283*, 178–184.

(132) Shi, Q.; Izvekov, S.; Both, G. Mixed Atomistic and Coarse-Grained Molecular Dynamics: Simulation of a Membrane-Bound Ion Channel. *J. Phys. Chem. B* **2006**, *110* (31), 15045–15048.

(133) Liao, C.; Zhao, X.; Liu, J.; Schneebeli, S.; Shelley, J.; Li, J. Capturing the Multiscale Dynamics of Membrane Protein Complexes with All-Atom, Mixed-Resolution, and Coarse-Grained Models. *Phys. Chem. Chem. Physics2* **2017**, *19*, 9181–9188.

(134) Kosol, S.; Contreras-martos, S.; Cedeno, C.; Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **2013**, *18*, 10802–10828.

(135) Hiller, S.; Garces, R. G.; Malia, T. J.; Orekhov, V. Y.; Colombini, M.; Wagner, G. Solution Structure of the Integral Human Membrane Protein VDAC-1 in Detergent Micelles. *Science (80-.).* **2008**, *321* (5893), 1206 LP – 1210. https://doi.org/10.1126/science.1161302.

(136) Gautier, A.; Mott, H. R.; Bostock, M. J.; Kirkpatrick, J. P.; Nietlispach, D. Structure Determination of the Seven-Helix Transmembrane Receptor Sensory Rhodopsin II by Solution NMR Spectroscopy. *Nat. Struct. Mol. Biol.* **2010**, *17* (6), 768–774. https://doi.org/10.1038/nsmb.1807.

(137) Sušac, L.; Horst, R.; Wüthrich, K. Solution-NMR Characterization of Outer-Membrane Protein A from E. Coli in Lipid Bilayer Nanodiscs and Detergent Micelles. *ChemBioChem* **2014**, *15* (7), 995–1000. https://doi.org/10.1002/cbic.201300729.

(138) Hagn, F.; Etzkorn, M.; Raschle, T.; Wagner, G. Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *J. Am. Chem. Soc.* **2013**, *135* (5), 1919–1925. https://doi.org/10.1021/ja310901f.

(139) Dutta, S. K.; Yao, Y.; Marassi, F. M. Structural Insights into the Yersinia Pestis Outer Membrane Protein Ail in Lipid Bilayers. *J. Phys. Chem. B* **2017**, *121* (32), 7561–7570. https://doi.org/10.1021/acs.jpcb.7b03941.

(140) Baldwin, A. J.; Kay, L. E. NMR Spectroscopy Brings Invisible Protein States into Focus. *Nat. Chem. Biol.* **2009**, *5* (11), 808–814. https://doi.org/10.1038/nchembio.238.

(141) Im, W.; Jo, S.; Kim, T. An Ensemble Dynamics Approach to Decipher Solid-State NMR Observables of Membrane Proteins. *Biochim. Biophys. Acta - Biomembr.* **2012**, *1818* (2), 252–262. https://doi.org/10.1016/j.bbamem.2011.07.048.

(142) Ángyán, A. F.; Gáspári, Z. Ensemble-Based Interpretations of NMR Structural Data to Describe Protein Internal Dynamics. *Molecules* **2013**, *18* (9), 10548–10567. https://doi.org/10.3390/molecules180910548.

(143) Mandala, V. S.; Williams, J. K.; Hong, M. Structure and Dynamics of Membrane Proteins from Solid-State NMR. *Annu. Rev. Biophys.* **2018**, *47* (1), 201–222. https://doi.org/10.1146/annurev-biophys-070816-033712.

(144) Fenwick, R.; van den Bedem, H.; Fraser, J.; Wright, P. Integrated Description of Protein Dynamics from Room-Temperature X-Ray Crystallography and NMR. *PNAS* **2013**, 445–454.

(145) De Angelis, A. A.; Opella, S. J. Bicelle Samples for Solid-State NMR of Membrane Proteins. *Nat. Protoc.* **2007**, *2*, 2332.

(146) Opella, S. J. Solid-State NMR and Membrane Proteins. *J. Magn. Reson.* **2015**, *253*, 129–137. https://doi.org/10.1016/j.jmr.2014.11.015.

(147) Rule, G.; Hitchens, K. *Fundamentals of Protein NMR Spectroscopy*; 2006. https://doi.org/10.1007/1-4020-3500-4.

(148) Li, Q.; Ng, H. Q.; Kang, C. Secondary Structure and Topology of the Transmembrane Domain of Syndecan-2 in Detergent Micelles. *FEBS Lett.* **2019**, *593* (5), 554–561. https://doi.org/10.1002/1873-3468.13335.

(149) Bibow, S.; Hiller, S. A Guide to Quantifying Membrane Protein Dynamics in Lipids and Other Native-like Environments by Solution-State NMR Spectroscopy. *FEBS J.* **2019**, *286* (9), 1610–1623. https://doi.org/10.1111/febs.14639.

(150) Hutchison, J. M.; Lu, Z.; Li, G. C.; Travis, B.; Mittal, R.; Deatherage, C. L.; Sanders, C. R. Dodecyl-β-Melibioside Detergent Micelles as a Medium for Membrane Proteins. *Biochemistry* **2017**, *56* (41), 5481–5484. https://doi.org/10.1021/acs.biochem.7b00810.

(151) Miao, Y.; Qin, H.; Fu, R.; Sharma, M.; Can, T. V; Hung, I.; Luca, S.; Gor'kov, P. L.; Brey, W. W.; Cross, T. A. M2 Proton Channel Structural Validation from Full-Length Protein Samples in Synthetic Bilayers and E. Coli Membranes. *Angew. Chemie Int. Ed.* **2012**, *51* (33), 8383–8386. https://doi.org/10.1002/anie.201204666.

(152) Zhang, R.; Mroue, K. H.; Ramamoorthy, A. Proton-Based Ultrafast Magic Angle Spinning Solid-State NMR Spectroscopy. *Acc. Chem. Res.* **2017**, *50* (4), 1105–1113. https://doi.org/10.1021/acs.accounts.7b00082.

(153) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. A Structural Model for Alzheimer's β-Amyloid Fibrils Based on Experimental Constraints from Solid State NMR. *Proc. Natl. Acad. Sci.* **2002**, *99* (26), 16742 LP – 16747. https://doi.org/10.1073/pnas.262663499.

(154) van der Wel, P. C. A. New Applications of Solid-State NMR in Structural Biology. *Emerg. Top. Life Sci.* **2018**, ETLS20170088. https://doi.org/10.1042/ETLS20170088.

(155) Schubeis, T.; Le Marchand, T.; Andreas, L. B.; Pintacuda, G. 1H Magic-Angle Spinning NMR Evolves as a Powerful New Tool for Membrane Proteins. *J. Magn. Reson.* **2018**, *287*, 140–152. https://doi.org/https://doi.org/10.1016/j.jmr.2017.11.014.

(156) Higman, V. A. Solid-State MAS NMR Resonance Assignment Methods for Proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *106–107*, 37–65. https://doi.org/https://doi.org/10.1016/j.pnmrs.2018.04.002.

(157) Lamley, J. M.; Lougher, M. J.; Sass, H. J.; Rogowski, M.; Grzesiek, S.; Lewandowski, J. R. Unraveling the Complexity of Protein Backbone Dynamics with Combined 13C and 15N Solid-State NMR Relaxation Measurements. *Phys. Chem. Chem. Phys.* **2015**, *17* (34), 21997–22008. https://doi.org/10.1039/C5CP03484A.

(158) Rovó, P.; Linser, R. Microsecond Timescale Protein Dynamics: A Combined Solid-State NMR Approach. *ChemPhysChem* **2018**, *19* (1), 34–39. https://doi.org/10.1002/cphc.201701238.

(159) Liao, S.; Yang, Y.; Tietze, D.; Hong, M. The Influenza M2 Cytoplasmic Tail Changes the Proton-Exchange Equilibria and the Backbone Conformation of the Transmembrane Histidine Residue to Facilitate Proton Conduction. *J. Am. Chem. Soc.* **2015**, *137*, 6067–6077.

(160) Bhate, M. P.; McDermott, A. E. Protonation State of E71 in KcsA and Its Role for Channel Collapse and Inactivation. *Proc. Natl. Acad. Sci.* **2012**, *109* (38), 15265 LP – 15270. https://doi.org/10.1073/pnas.1211900109.

(161) Liao, S.; Fritzsching, K.; Hong, M. Conformational Analysis of the Full-Length M2 Protein of the Influenza A Virus Using Solid-State NMR. *Protein Sci.* **2013**, *22*, 1623–1628.

(162) Kwon, B.; Hong, M. The Influenza M2 Ectodomain Regulates the Conformational Equilibria of the Transmembrane Proton Channel: Insights from Solid-State Nuclear Magnetic Resonance. *Biochemistry* **2016**, *55* (38), 5387–5397. https://doi.org/10.1021/acs.biochem.6b00727.

(163) Ding, Y.; Fujimoto, L. M.; Yao, Y.; Marassi, F. M. Solid-State NMR of the Yersinia Pestis Outer Membrane Protein Ail in Lipid Bilayer Nanodiscs Sedimented by Ultracentrifugation. *J. Biomol. NMR* **2015**, *61* (3), 275–286. https://doi.org/10.1007/s10858-014-9893-4.

(164) Barnaba, C.; Ramamoorthy, A. Picturing the Membrane-Assisted Choreography of

Cytochrome P450 with Lipid Nanodiscs. *ChemPhysChem* **2018**, *19* (20), 2603–2613. https://doi.org/10.1002/cphc.201800444.

(165) Lacabanne, D.; Fogeron, M.-L.; Wiegand, T.; Cadalbert, R.; Meier, B. H.; Böckmann, A. Protein Sample Preparation for Solid-State NMR Investigations. *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *110*, 20–33. https://doi.org/https://doi.org/10.1016/j.pnmrs.2019.01.001.

(166) Lacabanne, D.; Lends, A.; Danis, C.; Kunert, B.; Fogeron, M.-L.; Jirasko, V.; Chuilon, C.; Lecoq, L.; Orelle, C.; Chaptal, V.; et al. Gradient Reconstitution of Membrane Proteins for Solid-State NMR Studies. *J. Biomol. NMR* **2017**, *69* (2), 81–91. https://doi.org/10.1007/s10858-017-0135-4.

(167) Bertini, I.; Engelke, F.; Gonnelli, L.; Knott, B.; Luchinat, C.; Osen, D.; Ravera, E. On the Use of Ultracentrifugal Devices for Sedimented Solute NMR. *J. Biomol. NMR* **2012**, *54* (2), 123–127. https://doi.org/10.1007/s10858-012-9657-y.

(168) Pinto, L.; Lamb, R. The M2 Proton Channels of Influenza A and B Viruses. *J. Biol. Chem.* **2006**, *281*, 8997–9000.

(169) Wang, S.; Munro, R.; Shi, L.; Kawamura, I.; Okitsu, T. Solid-State NMR Spectroscopy Structure Determination of a Lipid-Embedded Heptahelical Membrane Protein. *Nat. Methods* **2013**, *10*, 1007–1012.

(170) Emami, S.; Fan, Y.; Munro, R.; Ladizhansky, V.; Brown, L. Yeast-Expressed Human Membrane Protein Aquaporin-1 Yields Excellent Resolution of Solid-State MAS NMR Spectra. *J. Biomol. NMR* **2013**, *55*, 147–155.

(171) de Groot, B.; Grubmuller, H. Water Permeation across Biological Membranes: Mechanism and Dynamics of Aquaporin-1 and GlpF. *Science (80-.).* **2001**, *294*, 2353–2357.

(172) Jensen, M.; Park, S.; Tajkhorshid, E.; Schulten, K. Energetics of Glycerol Conduction through Aquaglyceroporin GlpF. *PNAS* **2002**, *99*, 6731–6736.

(173) Pan, Y.; Brown, L.; Konermann, L. Hydrogen/Deuterium Exchange Mass Spectrometry and Optical Spectroscopy as Complementary Tools for Studying the Structure and Dynamics of a Membrane Protein. *Int. J. Mass Spectrom.* **2010**.

(174) Shukla, A.; Westfield, G. H.; Xiao, L.; Reis, R.; Huang, L.; Tripathi-Shukla, P.; Qian, J.; Li, S.; Blanc, A.; A., O. Visualization of Arrestin Recruitment by a G Protein-Coupled Receptor. *Nature* **2014**, *512*, 218–222.

(175) Orban, T.; Jastrzebska, B.; Gupta, S.; Wang, B.; Miyagi, M.; Chance, M.; Palczewski, K. Conformational Dynamics of Activation for the Pentameric Complex of Dimeric G Protein-Coupled Receptor and Heterotrimeric G Protein. *Structure* **2012**, *20*, 826–840.

(176) Duc, N.; Du, Y.; Thorsen, T.; Lee, S.; Zhang, C.; Kato, H.; Kobilka, B.; Chung, K.

Effective Application of Bicelles for Conformational Analysis of G Protein-Coupled Receptors by Hydrogen/Deuterium Exchange Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 808–817.

(177) Morgan, C.; Hebling, C.; Rand, K.; Stafford, D.; Jorgenson, J.; Engen, J. R. Conformational Transitions in the Membrane Scaffold Protein of Phospholipid Bilayer Nanodiscs. *Mol. Cell Proteomics* **2011**, *10*.

(178) Hebling, C.; Morgan, C.; Stafford, D.; Jorgenson, J.; Rand, K.; Engen, J. Conformational Analysis of Membrane Proteins in Phospholipid Bilayer Nanodiscs by Hydrogen Exchange Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 5415–5419.

(179) Stetefeld, J.; McKenna, S. A.; Patel, T. R. Dynamic Light Scattering: A Practical Guide and Applications in Biomedical Sciences. *Biophys. Rev.* **2016**, *8* (4), 409–427. https://doi.org/10.1007/s12551-016-0218-6.

(180) Kwan, O. T.; Reis, R.; Siligardi, G.; Hussain, R.; Cheruvara, H.; Moraes, I. Selection of Biophysical Methods for Characterisation of Membrane Proteins. *International Journal of Molecular Sciences* . 2019. https://doi.org/10.3390/ijms20102605.

(181) Bhattacharjee, S. DLS and Zeta Potential - What They Are and What They Are Not? *J. Control. Release* **2016**, *235*, 337–351. https://doi.org/10.1016/j.jconrel.2016.06.017.

(182) Gimpl, K.; Klement, J.; Keller, S. Characterising Protein/Detergent Complexes by Triple-Detection Size-Exclusion Chromatography. *Biol. Proced. Online* **2016**, *18* (1), 1–18. https://doi.org/10.1186/s12575-015-0031-9.

(183) Slotboom, D. J.; Duurkens, R. H.; Olieman, K.; Erkens, G. B. Static Light Scattering to Characterize Membrane Proteins in Detergent Solution. *Methods* **2008**, *46* (2), 73–82. https://doi.org/10.1016/j.ymeth.2008.06.012.

(184) Pace, C.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to Measure and Predict the Molar Absorption Coefficient of a Protein. *Protein Sci.* **1995**, *11*, 2411–2423.

(185) Skou, S.; Gillilan, R.; Ando, N. Synchrotron-Based Small-Angle X-Ray Scattering of Proteins in Solution. *Nat. Protoc.* **2014**, *9*, 1727–1739.

(186) Meisburger, S. P.; Thomas, W. C.; Watkins, M. B.; Ando, N. X-Ray Scattering Studies of Protein Structural Dynamics. *Chem. Rev.* **2017**, *117* (12), 7615–7672. https://doi.org/10.1021/acs.chemrev.6b00790.

(187) Kikhney, A. G.; Svergun, D. I. A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589* (19), 2570–2577. https://doi.org/10.1016/j.febslet.2015.08.027.

(188) Berthaud, A.; Manzi, J.; Perez, J.; Mangenot, S. Modeling Detergent Organization around Aquaporin-0 Using Small-Angle x-Ray Scattering. *J. Am. Chem. Soc.* **2012**, *134*, 10080–10088.

(189) Koutsioubas, A.; Berthaud, A.; Mangenot, S.; Perez, J. Ab Initio and All-Atom

Modeling of Detergent Organization around Aquaporin-0 Based on SAXS Data. *J. Phys. Chem. B* **2013**, *117*, 13588–13594.

(190) Boekema, E. Negative Staining of Integral Membrane Proteins. *Micron Microsc. Acta* **1991**, *22* (4), 361–369.

(191) Vahedi-Faridi, A.; Jastrzebska, B.; Palczewski, K.; Engel, A. 3D Imaging and Quantitative Analysis of Small Solubilized Membrane Proteins and Their Complexes by Transmission Electron Microscopy. *J. Electron Microsc. (Tokyo).* **2013**, *62* (1), 95–107. https://doi.org/10.1093/jmicro/dfs091.

(192) Ohi, M.; Li, Y.; Cheng, Y.; Walz, T. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol. Proced. Online* **2004**, *6*, 23–34. https://doi.org/10.1251/bpo70.

(193) Ohi, M.; Li, Y.; Cheng, Y.; Walz, T. Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy. *Biol. Proced. Online* **2004**, *24* (2), 168–172. https://doi.org/10.1002/jemt.1070240208.

(194) Gill, P.; Moghadam, T. T.; Ranjbar, B. Differential Scanning Calorimetry Techniques: Applications in Biology and Nanoscience. *J. Biomol. Tech.* **2010**, *21* (4), 167–193.

(195) Danley, R. L. New Heat Flux DSC Measurement Technique. *Thermochim. Acta* **2002**, *395* (1), 201–208. https://doi.org/https://doi.org/10.1016/S0040-6031(02)00212-5.

(196) Zucca, N.; Erriu, G.; Onnis, S.; Longoni, A. An Analytical Expression of the Output of a Power-Compensated DSC in a Wide Temperature Range. *Thermochim. Acta* **2004**, *413* (1), 117–125. https://doi.org/https://doi.org/10.1016/j.tca.2003.10.006.

(197) Brown, M.; Gallagher, P. *Handbook of Thermal Analysis and Calorimetry*, 5th ed.; Elsevier, 2007.

(198) Conejero-Lara, F.; Sanchez-Ruiz, J. M.; Mateo, P. L.; Burgos, F. J.; Vendrell, J.; Aviles, F. X. Differential Scanning Calorimetric Study of Carboxypeptidase B, Procarboxypeptidase B and Its Globular Activation Domain. *Eur. J. Biochem.* **1991**, *200* (3), 663–670. https://doi.org/10.1111/j.1432-1033.1991.tb16230.x.

(199) Vogl, T.; Jatzke, C.; Hinz, H.-J.; Benz, J.; Huber, R. Thermodynamic Stability of Annexin V E17G: Equilibrium Parameters from an Irreversible Unfolding Reaction,. *Biochemistry* **1997**, *36* (7), 1657–1668. https://doi.org/10.1021/bi962163z.

(200) Cueto, M.; Dorta, M. J.; Munguı́a, O.; Llabrés, M. New Approach to Stability Assessment of Protein Solution Formulations by Differential Scanning Calorimetry. *Int. J. Pharm.* **2003**, *252* (1), 159–166. https://doi.org/https://doi.org/10.1016/S0378-5173(02)00627-0.

(201) Yang, Z.; Wang, C.; Zhou, Q.; An, J.; Hildebrandt, E.; Aleksandrov, L. A.; Kappes, J.

C.; DeLucas, L. J.; Riordan, J. R.; Urbatsch, I. L.; et al. Membrane Protein Stability Can Be Compromised by Detergent Interactions with the Extramembranous Soluble Domains. *Protein Sci.* **2014**, *23* (6), 769–789. https://doi.org/10.1002/pro.2460.

(202) Yang, Z.; Brouillette, C. G. *A Guide to Differential Scanning Calorimetry of Membrane and Soluble Proteins in Detergents*, 1st ed.; Elsevier Inc., 2016; Vol. 567. https://doi.org/10.1016/bs.mie.2015.08.014.

(203) Shaw, A. W.; McLean, M. A.; Sligar, S. G. Phospholipid Phase Transitions in Homogeneous Nanometer Scale Bilayer Discs. *FEBS Lett.* **2004**, *556* (1–3), 260–264. https://doi.org/10.1016/S0014-5793(03)01400-5.

(204) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, *41* (1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016.

(205) Swint-Kruse, L.; Matthews, K. S. Lac Operon. *Encycl. Biol. Chem. Second Ed.* **2013**, 694–700. https://doi.org/10.1016/B978-0-12-378630-2.00249-8.

(206) Marbach, A.; Bettenbrock, K. Lac Operon Induction in Escherichia Coli: Systematic Comparison of IPTG and TMG Induction and Influence of the Transacetylase LacA. *J. Biotechnol.* **2012**, *157*, 82–88.

(207) Studier, F. W. Studier Recipe. **2005**, *234*.

(208) Stetsenko, A.; Guskov, A. An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization. *Crystals* **2017**, *7* (7), 197. https://doi.org/10.3390/cryst7070197.

(209) Moulik, S. P.; Haque, M. E.; Jana, P. K.; Das, A. R. Micellar Properties of Cationic Surfactants in Pure and Mixed States. *J. Phys. Chem.* **1996**, *100* (2), 701–708. https://doi.org/10.1021/jp9506494.

(210) Kern, J.; Guskov, A. Lipids in Photosystem II: Multifunctional Cofactors. *J. Photochem. Photobiol. B Biol.* **2011**, *104* (1), 19–34. https://doi.org/https://doi.org/10.1016/j.jphotobiol.2011.02.025.

(211) Hunte, C.; Richers, S. Lipids and Membrane Protein Structures. *Curr. Opin. Struct. Biol.* **2008**, *18* (4), 406–411. https://doi.org/https://doi.org/10.1016/j.sbi.2008.03.008.

(212) Lemieux, M. J.; Song, J.; Kim, M. J.; Huang, Y.; Villa, A.; Auer, M.; Li, X.-D.; Wang, D.-N. Three-Dimensional Crystallization of the Escherichia Coli Glycerol-3-Phosphate Transporter: A Member of the Major Facilitator Superfamily. *Protein Sci.* **2003**, *12* (12), 2748–2756. https://doi.org/10.1110/ps.03276603.

(213) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; et al. High-Resolution Crystal Structure of an Engineered Human β₂-Adrenergic G Protein–

Coupled Receptor. *Science (80-.).* **2007**, *318* (5854), 1258 LP – 1265. https://doi.org/10.1126/science.1150577.

(214) Lee, S. C.; Knowles, T. J.; Postis, V. L. G.; Jamshad, M.; Parslow, R. A.; Lin, Y. P.; Goldman, A.; Sridhar, P.; Overduin, M.; Muench, S. P.; et al. A Method for Detergent-Free Isolation of Membrane Proteins in Their Local Lipid Environment. *Nat. Protoc.* **2016**, *11* (7), 1149–1162. https://doi.org/10.1038/nprot.2016.070.

(215) Mosslehy, W.; Voskoboynikova, N.; Colbasevici, A.; Ricke, A.; Klose, D.; Klare, J. P.; Mulkidjanian, A. Y.; Steinhoff, H.-J. Conformational Dynamics of Sensory Rhodopsin II in Nanolipoprotein and Styrene–Maleic Acid Lipid Particles. *Photochem. Photobiol.* **2019**, *0* (0). https://doi.org/10.1111/php.13096.

(216) Postis, V.; Rawson, S.; Mitchell, J. K.; Lee, S. C.; Parslow, R. A.; Dafforn, T. R.; Baldwin, S. A.; Muench, S. P. The Use of SMALPs as a Novel Membrane Protein Scaffold for Structure Study by Negative Stain Electron Microscopy. *Biochim. Biophys. Acta* **2015**, *1848* (2), 496–501. https://doi.org/10.1016/j.bbamem.2014.10.018.

(217) Radoicic, J.; Park, S. H.; Opella, S. J. Macrodiscs Comprising SMALPs for Oriented Sample Solid-State NMR Spectroscopy of Membrane Proteins. *Biophys. J.* **2018**, *115* (1), 22–25. https://doi.org/10.1016/j.bpj.2018.05.024.

(218) Parmar, M.; Rawson, S.; Scarff, C. A.; Goldman, A.; Dafforn, T. R.; Muench, S. P.; Postis, V. L. G. Using a SMALP Platform to Determine a Sub-Nm Single Particle Cryo-EM Membrane Protein Structure. *Biochim. Biophys. Acta - Biomembr.* **2018**, *1860* (2), 378–383. https://doi.org/https://doi.org/10.1016/j.bbamem.2017.10.005.

(219) Teo, A. C. K.; Lee, S. C.; Pollock, N. L.; Stroud, Z.; Hall, S.; Thakker, A.; Pitt, A. R.; Dafforn, T. R.; Spickett, C. M.; Roper, D. I. Analysis of SMALP Co-Extracted Phospholipids Shows Distinct Membrane Environments for Three Classes of Bacterial Membrane Protein. *Sci. Rep.* **2019**, *9* (1), 1813. https://doi.org/10.1038/s41598-018-37962-0.

(220) Jamshad, M.; Grimard, V.; Idini, I.; Knowles, T. J.; Dowle, M. R.; Schofield, N.; Sridhar, P.; Lin, Y.; Finka, R.; Wheatley, M.; et al. Structural Analysis of a Nanoparticle Containing a Lipid Bilayer Used for Detergent-Free Extraction of Membrane Proteins. *Nano Res.* **2015**, *8* (3), 774–789. https://doi.org/10.1007/s12274-014-0560-6.

(221) Akbarzadeh, A.; Rezaei-Sadabady, R.; Davaran, S.; Joo, S. W.; Zarghami, N.; Hanifehpour, Y.; Samiei, M.; Kouhi, M.; Nejati-Koshki, K. Liposome: Classification, Preparation, and Applications. *Nanoscale Res. Lett.* **2013**, *8* (1), 1–8. https://doi.org/10.1186/1556-276X-8-102.

(222) Sun, G.; Chung, T.-S.; Jeyaseelan, K.; Armugam, A. Stabilization and Immobilization of Aquaporin Reconstituted Lipid Vesicles for Water Purification. *Colloids Surfaces B Biointerfaces* **2013**, *102*, 466–471. https://doi.org/https://doi.org/10.1016/j.colsurfb.2012.08.009.

(223)   Bitounis, D.; Fanciullino, R.; Iliadis, A.; Ciccolini, J. Optimizing Druggability through Liposomal Formulations: New Approaches to an Old Concept. *ISRN Pharm.* **2012**, *2012*, 738432. https://doi.org/10.5402/2012/738432.

(224)   Sanders, C. R.; Landis, G. C. Reconstitution of Membrane Proteins into Lipid-Rich Bilayered Mixed Micelles for NMR Studies. *Biochemistry* **1995**, *34* (12), 4030–4040. https://doi.org/10.1021/bi00012a022.

(225)   Caldwell, T.; Baoukina, S.; Brock, A.; Oliver, R.; Root, K.; Kreuger, J.; Glover, K.; Tieleman, D.; Columbus, L. Low-q Bicelles Are Mixed Micelles. *J. Phys. Chem. Lett.* **2018**, *9*, 4469–4473.

(226)   Du, U. H. N.; Gildenberg, M.; Ramamoorthy, A. The Magic of Bicelles Lights up Membrane Protein Structure. *Am. Chem. Soc.* **2012**, *112*, 6054–6074.

(227)   Faham, S.; Bowie, J. U. Bicelle Crystallization: A New Method for Crystallizing Membrane Proteins Yields a Monomeric Bacteriorhodopsin Structure11Edited by D. Rees. *J. Mol. Biol.* **2002**, *316* (1), 1–6. https://doi.org/https://doi.org/10.1006/jmbi.2001.5295.

(228)   Wang, H.; Elferich, J.; Gouaux, E. Structures of LeuT in Bicelles Define Conformation and Substrate Binding in a Membrane-like Context. *Nat. Struct. Mol. Biol.* **2012**, *19* (2), 212–219. https://doi.org/10.1038/nsmb.2215.

(229)   Schredelseker, J.; Paz, A.; López, C. J.; Altenbach, C.; Leung, C. S.; Drexler, M. K.; Chen, J. N.; Hubbell, W. L.; Abramson, J. High Resolution Structure and Double Electron-Electron Resonance of the Zebrafish Voltage-Dependent Anion Channel 2 Reveal an Oligomeric Population. *J. Biol. Chem.* **2014**, *289* (18), 12566–12577. https://doi.org/10.1074/jbc.M113.497438.

(230)   Piai, A.; Fu, Q.; Dev, J.; Chou, J. J. Optimal Bicelle q for Solution NMR Studies of Protein Transmembrane Partition. *Chesmistry* **2017**, *23* (6), 1361–1367. https://doi.org/10.1186/s40945-017-0033-9.Using.

(231)   Tribet, C.; Audebert, R.; Popot, J.-L. Amphipols: Polymers That Keep Membrane Proteins Soluble in Aqueous Solutions. *Proc. Natl. Acad. Sci.* **1996**, *93* (26), 15047 LP – 15050. https://doi.org/10.1073/pnas.93.26.15047.

(232)   Elter, S.; Raschle, T.; Arens, S.; Viegas, A.; Gelev, V.; Etzkorn, M.; Wagner, G. The Use of Amphipols for NMR Structural Characterization of 7-TM Proteins. *J. Membr. Biol.* **2014**, *247* (9), 957–964. https://doi.org/10.1007/s00232-014-9669-5.

(233)   Singh, A. K.; Saotome, K.; McGoldrick, L. L.; Sobolevsky, A. I. Structural Bases of TRP Channel TRPV6 Allosteric Modulation by 2-APB. *Nat. Commun.* **2018**, *9* (1), 2465. https://doi.org/10.1038/s41467-018-04828-y.

(234)   Nagy, J. K.; Hoffmann, A. K.; Keyes, M. H.; Gray, D. N.; Oxenoid, K.; Sanders, C. R. Use of Amphipathic Polymers to Deliver a Membrane Protein to Lipid Bilayers. *FEBS Lett.* **2001**, *501* (2–3), 115–120. https://doi.org/10.1016/S0014-5793(01)02627-8.

(235) Gorzelle, B. M.; Hoffman, A. K.; Keyes, M. H.; Gray, D. N.; Ray, D. G.; Sanders, C. R. Amphipols Can Support the Activity of a Membrane Enzyme. *J. Am. Chem. Soc.* **2002**, *124* (39), 11594–11595. https://doi.org/10.1021/ja027051b.

(236) Tribet, C.; Audebert, R.; Popot, J.-L. Amphipols: Polymers That Keep Membrane Proteins Soluble in Aqueous Solutions. *Proc. Natl. Acad. Sci.* **2002**, *93* (26), 15047–15050. https://doi.org/10.1073/pnas.93.26.15047.

(237) Gohon, Y.; Giusti, F.; Prata, C.; Charvolin, D.; Timmins, P.; Ebel, C.; Tribet, C.; Popot, J.-L. Well-Defined Nanoparticles Formed by Hydrophobic Assembly of a Short and Polydisperse Random Terpolymer, Amphipol A8-35. *Langmuir* **2006**, *22* (3), 1281–1290. https://doi.org/10.1021/la052243g.

(238) Le Bon, C.; Marconnet, A.; Masscheleyn, S.; Popot, J. L.; Zoonens, M. Folding and Stabilizing Membrane Proteins in Amphipol A8-35. *Methods* **2018**, *147* (December 2017), 95–105. https://doi.org/10.1016/j.ymeth.2018.04.012.

(239) Nath, A.; Atkins, W. M.; Sligar, S. G. Applications of Phospholipid Bilayer Nanodiscs in the Study of Membranes and Membrane Proteins. *Biochemistry* **2007**, *46* (8), 2059–2069. https://doi.org/10.1021/bi602371n.

(240) Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G. *Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs*; Elsevier Masson SAS, 2009; Vol. 464. https://doi.org/10.1016/S0076-6879(09)64011-8.

(241) Denisov, I. G.; McLean, M. A.; Shaw, A. W.; Grinkova, Y. V; Sligar, S. G. Thermotropic Phase Transition in Soluble Nanoscale Lipid Bilayers. *J. Phys. Chem. B* **2005**, *109* (32), 15580–15588. https://doi.org/10.1021/jp051385g.

(242) Denisov, I. G.; Grinkova, Y. V; Lazarides, A. A.; Sligar, S. G. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* **2004**, *126* (11), 3477–3487. https://doi.org/10.1021/ja0393574.

(243) López, C. A.; Swift, M. F.; Xu, X.-P.; Hanein, D.; Volkmann, N.; Gnanakaran, S. Biophysical Characterization of a Nanodisc with and without BAX: An Integrative Study Using Molecular Dynamics Simulations and Cryo-EM. *Structure* **2019**, *27* (6), 988-999.e4. https://doi.org/https://doi.org/10.1016/j.str.2019.03.013.

(244) Li, Y.; Kijac, A. Z.; Sligar, S. G.; Rienstra, C. M. Structural Analysis of Nanoscale Self-Assembled Discoidal Lipid Bilayers by Solid-State NMR Spectroscopy. *Biophys. J.* **2006**, *91* (10), 3819–3828. https://doi.org/https://doi.org/10.1529/biophysj.106.087072.

(245) Martinez, D.; Decossas, M.; Kowal, J.; Frey, L.; Stahlberg, H.; Dufourc, E. J.; Riek, R.; Habenstein, B.; Bibow, S.; Loquet, A. Lipid Internal Dynamics Probed in Nanodiscs. *ChemPhysChem* **2017**, *18* (19), 2651–2657. https://doi.org/10.1002/cphc.201700450.

(246) Marty, M. T.; Zhang, H.; Cui, W.; Gross, M. L.; Sligar, S. G. Interpretation and

Deconvolution of Nanodisc Native Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (2), 269–277. https://doi.org/10.1007/s13361-013-0782-y.

(247) Redhair, M.; Clouser, A. F.; Atkins, W. M. Hydrogen-Deuterium Exchange Mass Spectrometry of Membrane Proteins in Lipid Nanodiscs. *Chem. Phys. Lipids* **2019**, *220*, 14–22. https://doi.org/https://doi.org/10.1016/j.chemphyslip.2019.02.007.

(248) Yokogawa, M.; Fukuda, M.; Osawa, M. Nanodiscs for Structural Biology in a Membranous Environment. *Chem. Pharm. Bull.* **2019**, *67* (4), 321–326. https://doi.org/10.1248/cpb.c18-00941.

(249) Hagn, F.; Nasr, M. L.; Wagner, G. Assembly of Phospholipid Nanodiscs of Controlled Size for Structural Studies of Membrane Proteins by NMR. *Nat. Protoc.* **2017**, *13*, 79.

(250) Josts, I.; Nitsche, J.; Maric, S.; Mertens, H. D.; Moulin, M.; Haertlein, M.; Prevost, S.; Svergun, D. I.; Busch, S.; Forsyth, V. T.; et al. Conformational States of ABC Transporter MsbA in a Lipid Environment Investigated by Small-Angle Scattering Using Stealth Carrier Nanodiscs. *Structure* **2018**, *26* (8), 1072-1079.e4. https://doi.org/https://doi.org/10.1016/j.str.2018.05.007.

(251) Lakomek, N.-A.; Frey, L.; Bibow, S.; Böckmann, A.; Riek, R.; Meier, B. H. Proton-Detected NMR Spectroscopy of Nanodisc-Embedded Membrane Proteins: MAS Solid-State vs Solution-State Methods. *J. Phys. Chem. B* **2017**, *121* (32), 7671–7680. https://doi.org/10.1021/acs.jpcb.7b06944.

(252) Nikolaev, M.; Round, E.; Gushchin, I.; Polovinkin, V.; Balandin, T.; Kuzmichev, P.; Shevchenko, V.; Borshchevskiy, V.; Kuklin, A.; Round, A.; et al. Integral Membrane Proteins Can Be Crystallized Directly from Nanodiscs. *Cryst. Growth Des.* **2017**, *17* (3), 945–948. https://doi.org/10.1021/acs.cgd.6b01631.

(253) Panganiban, B.; Qiao, B.; Jiang, T.; Delre, C.; Obadia, M. M.; Nguyen, T. D.; Smith, A. A. A.; Hall, A.; Sit, I.; Crosby, M. G.; et al. Random Heteropolymers Preserve Protein Function in Foreign Environments. *Science (80-.).* **2018**, *1243* (March), 1239–1243. https://doi.org/10.1126/science.aao0335.

(254) Wagner, S.; Baars, L.; Ytterberg, A. J.; Klussmeier, A.; Wagner, C. S.; Nord, O.; Nygren, P.-Å.; van Wijk, K. J.; de Gier, J.-W. Consequences of Membrane Protein Overexpression in Escherichia Coli . *Mol. Cell. Proteomics* **2007**, *6* (9), 1527–1550. https://doi.org/10.1074/mcp.m600431-mcp200.

(255) Singh, A.; Upadhyay, V.; Upadhyay, A.; Singh, S.; Panda, A. Protein Recovery from Inclusion Bodies of Escherichia Coli Using Mild Solubilization Process. *Microb. Cell Fact.* **2015**, *14*.

(256) Manley, D.; O' Neil, J. D. Preparation of Glycerol Facilitator for Protein Structure and Folding Studies in Solution BT - Membrane Protein Protocols: Expression, Purification, and Characterization; Selinsky, B. S., Ed.; Humana Press: Totowa, NJ,

2003; pp 89–101. https://doi.org/10.1385/1-59259-400-X:89.

(257) Block, H.; Maertens, B.; Spriestersbach, A.; Brinker, N.; Kubicek, J.; Fabis, R.; Labahn, J.; Schäfer, F. Immobilized-Metal Affinity Chromatography (IMAC). A Review. *Methods Enzymol.* **2009**, *463* (C), 439–473. https://doi.org/10.1016/S0076-6879(09)63027-5.

(258) Doerrler, W.; Raetz, C. ATPase Activity of the MsbA Lipid Flippase of Escherichia Coli. *J. Biol. Chem.* **2002**, *277*, 36697–36705.

(259) Morrison, E. A.; Henzler-Wildman, K. A. Reconstitution of Integral Membrane Proteins into Isotropic Bicelles with Improved Sample Stability and Expanded Lipid Composition Profile. *Biochim. Biophys. Acta - Biomembr.* **2012**, *1818* (3), 814–820. https://doi.org/10.1016/j.bbamem.2011.12.020.

(260) Mastronarde, D. N. Automated Electron Microscope Tomography Using Robust Prediction of Specimen Movements. *J. Struct. Biol.* **2005**, *152* (1), 36–51. https://doi.org/https://doi.org/10.1016/j.jsb.2005.07.007.

(261) Scheres, S. H. W. RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination. *J. Struct. Biol.* **2012**, *180* (3), 519–530. https://doi.org/10.1016/j.jsb.2012.09.006.

(262) Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; et al. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225. https://doi.org/10.1107/S1600576717007786.

(263) Svergun, D.; Barberato, C.; Koch, M. H. J. CRYSOL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **2017**, *32* (2), vi. https://doi.org/10.1016/s0167-8116(15)00050-6.

(264) Svergun, D. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503.

(265) Petoukhov, M. V.; Vicente, J. B.; Crowley, P. B.; Carrondo, M. A.; Teixeira, M.; Svergun, D. I. Quaternary Structure of Flavorubredoxin as Revealed by Synchrotron Radiation Small-Angle X-Ray Scattering. *Structure* **2008**, *16* (9), 1428–1436. https://doi.org/10.1016/j.str.2008.06.009.

(266) Pérez, J.; Koutsioubas, A. Memprot: A Program to Model the Detergent Corona around a Membrane Protein Based on SEC-SAXS Data. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *71*, 86–93. https://doi.org/10.1107/S1399004714016678.

(267) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865. https://doi.org/10.1002/jcc.20945.

(268) The PyMol Molecular Graphics System, Version 2.3.0, Schrodinger, LLC.

(269) Petoukhov, M. V.; Svergun, D. I. Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys. J.* **2005**, *89* (2), 1237–1250. https://doi.org/10.1529/biophysj.105.064154.

(270) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(271) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1), 43–56. https://doi.org/https://doi.org/10.1016/0010-4655(95)00042-E.

(272) Ikura, M.; Kay, L. E.; Bax, A. A Novel Approach for Sequential Assignment of Proton, Carbon-13, and Nitrogen-15 Spectra of Larger Proteins: Heteronuclear Triple-Resonance Three-Dimensional NMR Spectroscopy. Application to Calmodulin. *Biochemistry* **1990**, *29* (19), 4659–4667. https://doi.org/10.1021/bi00471a022.

(273) Takegoshi, K.; Nakamura, S.; Terao, T. 13C–1H Dipolar-Driven 13C–13C Recoupling without 13C Rf Irradiation in Nuclear Magnetic Resonance of Rotating Solids. *J. Chem. Phys.* **2003**, *118* (5), 2325–2341. https://doi.org/10.1063/1.1534105.

(274) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6* (3), 277–293. https://doi.org/10.1007/BF00197809.

(275) Consortium, T. U. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2018**, *47* (D1), D506–D515. https://doi.org/10.1093/nar/gky1049.

(276) Galka, J. J.; Baturin, S. J.; Manley, D. M.; Kehler, A. J.; O'Neil, J. D. Stability of the Glycerol Facilitator in Detergent Solutions. *Biochemistry* **2008**, *47* (11), 3513–3524. https://doi.org/10.1021/bi7021409.

(277) Takeda, E.; Taketani, Y.; Sawada, N.; Sato, T.; Yamamoto, H. The Regulation and Function of Phosphate in the Human Body. *BioFactors* **2004**, *21*, 345–355.

(278) Galka, J.; Baturin, S.; Manley, D.; Kehler, A.; O'Neil, J. D. Stability of the Glycerol Facilitator in Detergent Solutions. *Biochemistry* **2008**, *47*, 3513–3524.

(279) Lodish, H.; Berk, A.; Zipursky, S. Section 15.4 Intracellular Ion Environment and Membrane Electric Potential. In *Molecular Cell Biology*; W.H. Freeman: New York, 2000.

(280) Bornhorst, J. A.; Falke, J. J. Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol.* **2000**, *326*, 245–254.

(281) Heerklotz, H.; Tsamaloukas, A.; Kita-Tokarczyk, K.; Strunz, P.; Gutberlet, T.

Structural, Volumetric, and Thermodynamic Characterization of a Micellar Sphere-to-Rod Transition. *J. Am. Chem. Soc.* **2004**, *126* (50), 16544–16552. https://doi.org/10.1021/ja045525w.

(282) Baturin, S.; Galka, J. J.; Piyadasa, H.; Gajjeraman, S.; O'Neil, J. D. The Effects of a Protein Osmolyte on the Stability of the Integral Membrane Protein Glycerol Facilitator. *Biochem. Cell Biol.* **2014**, *92* (6), 564–575. https://doi.org/10.1139/bcb-2014-0076.

(283) von Berlepsch, H.; Dautzenberg, H.; Rother, G.; Jäger, J. An Investigation of the Micellar Phase of Sodium Sulfopropyl Octadecyl Maleate in Aqueous Sodium Chloride Solutions by Light Scattering Techniques. Evidence of Nearly Rodlike Micelles. *Langmuir* **1996**, *12* (15), 3613–3625. https://doi.org/10.1021/la960036y.

(284) Reiss-Husson, F.; Luzzati, V. The Structure of the Micellar Solutions of Some Amphiphilic Compounds in Pure Water as Determined by Absolute Small-Angle X-Ray Scattering Techniques. *J. Phys. Chem.* **1964**, *68* (12), 3504–3511. https://doi.org/10.1021/j100794a011.

(285) Zhang, J.; Ge, Z.; Jiang, X.; Hassan, P. A.; Liu, S. Stopped-Flow Kinetic Studies of Sphere-to-Rod Transitions of Sodium Alkyl Sulfate Micelles Induced by Hydrotropic Salt. *J. Colloid Interface Sci.* **2007**, *316* (2), 796–802. https://doi.org/https://doi.org/10.1016/j.jcis.2007.08.067.

(286) Street, T.; Bolen, D.; Rose, G. A Molecular Mechanism for Osmolyte-Induced Protein Stability. *PNAS* **2006**, *103*, 13997–14002.

(287) Baynes, B. M.; Wang, D. I. C.; Trout, B. L. Role of Arginine in the Stabilization of Proteins against Aggregation. *Biochemistry* **2005**, *44* (12), 4919–4925. https://doi.org/10.1021/bi047528r.

(288) Eliezer, D.; Jennings, P. A.; Dyson, H. J.; Wright, P. E. Populating the Equilibrium Molten Globule State of Apomyoglobin under Conditions Suitable for Structural Characterization by NMR. *FEBS Lett.* **1997**, *417* (1), 92–96. https://doi.org/10.1016/S0014-5793(97)01256-8.

(289) Patel, A.; Malinovska, L.; Saha, S.; Wang, J.; Alberti, S.; Krishnan, Y.; Hyman, A. A. ATP as a Biological Hydrotrope. *Science (80-.).* **2017**, *356* (6339), 753–756. https://doi.org/10.1126/science.aaf6846.

(290) Kapust, R. B.; Toözseór, J.; Copeland, T. D.; Waugh, D. S. The P1′ Specificity of Tobacco Etch Virus Protease. *Biochem. Biophys. Res. Commun.* **2002**, *294* (5), 949–955. https://doi.org/10.1016/S0006-291X(02)00574-0.

(291) Guillin, M. C.; Bezeaud, A.; Bouton, M. C.; Jandrot-Perrus, M. Thrombin Specificity. *Thromb. Haemost.* **1995**, *74* (1), 129–133. https://doi.org/10.1111/j.1749-6632.1981.tb29757.x.

(292) Vergis, J. M.; Wiener, M. C. The Variable Detergent Sensitivity of Proteases That

Are Utilized for Recombinant Protein Affinity Tag Removal. *Protein Expr. Purif.* **2011**, *78* (2), 139–142. https://doi.org/10.1016/j.pep.2011.04.011.

(293)   Dang, B.; Mravic, M.; Hu, H.; Schmidt, N.; Mensa, B.; DeGrado, W. F. SNAC-Tag for Sequence-Specific Chemical Protein Cleavage. *Nat. Methods* **2019**, *16* (4), 319–322. https://doi.org/10.1038/s41592-019-0357-3.

(294)   Maire, M. Interaction of Membrane Proteins and Lipids with Solubilizing Detergents. *Biochim. Biophys. Acta* **2000**, *1508*, 86–111.

(295)   Arnold, T.; Linke, D. The Use of Detergents to Purify Membrane Proteins. *Curr. Protoc. Protein Sci.* **2008**, 4.8.1-4.8.35. https://doi.org/10.1002/0471140864.ps0408s84.

(296)   Frueh, D. P.; Goodrich, A. C.; Mishra, S. H.; Nichols, S. R. NMR Methods for Structural Studies of Large Monomeric and Multimeric Proteins. *Curr. Opin. Struct. Biol.* **2013**, *23* (5), 734–739. https://doi.org/10.1016/j.sbi.2013.06.016.

(297)   James, Z. M.; Borst, A. J.; Haitin, Y.; Frenz, B.; DiMaio, F.; Zagotta, W. N.; Veesler, D. CryoEM Structure of a Prokaryotic Cyclic Nucleotide-Gated Ion Channel. *Proc. Natl. Acad. Sci.* **2017**, *114* (17), 4430–4435. https://doi.org/10.1073/pnas.1700248114.

(298)   Liang, Y. L.; Khoshouei, M.; Radjainia, M.; Zhang, Y.; Glukhova, A.; Tarrasch, J.; Thal, D. M.; Furness, S. G. B.; Christopoulos, G.; Coudrat, T.; et al. Phase-Plate Cryo-EM Structure of a Class B GPCR-G-Protein Complex. *Nature* **2017**, *546* (7656), 118–123. https://doi.org/10.1038/nature22327.

(299)   Chae, P. S.; Rasmussen, S. G. F.; Rana, R.; Gotfryd, K.; Chandra, R.; Goren, M. A.; Kruse, A. C.; Nurva, S.; Loland, C. J.; Pierre, Y.; et al. MNG Amphiphiles for Solubilization, Stabilization and Crystalization of Membrane Proteins. *Nat Methods* **2010**, *7* (12), 1003–1008. https://doi.org/10.1038/nmeth.1526.Maltose-neopentyl.

(300)   Johnson, C. M. Differential Scanning Calorimetry as a Tool for Protein Folding and Stability. *Arch. Biochem. Biophys.* **2013**, *531* (1–2), 100–109. https://doi.org/10.1016/j.abb.2012.09.008.

(301)   Lantez, V.; Nikolaidis, I.; Rechenmann, M.; Vernet, T.; Noirclerc-Savoye, M. Rapid Automated Detergent Screening for the Solubilization and Purification of Membrane Proteins and Complexes. *Eng. Life Sci.* **2015**, *15* (1), 39–50. https://doi.org/10.1002/elsc.201400187.

(302)   Triba, M. N.; Zoonens, M.; Popot, J. L.; Devaux, P. F.; Warschawski, D. E. Reconstitution and Alignment by a Magnetic Field of a β-Barrel Membrane Protein in Bicelles. *Eur. Biophys. J.* **2006**, *35* (3), 268–275. https://doi.org/10.1007/s00249-005-0014-x.

(303)   Denisov, I. G.; Sligar, S. G. Nanodiscs for Structural and Functional Studies of Membrane Proteins. *Nat. Struct. Mol. Biol.* **2016**, *23* (6), 481–486. https://doi.org/10.1038/nsmb.3195.

(304) Hagn, F.; Wagner, G. Structure Refinement and Membrane Positioning of Selectively Labeled OmpX in Phospholipid Nanodiscs. *J. Biomol. NMR* **2015**, *61* (3–4), 249–260. https://doi.org/10.1007/s10858-014-9883-6.

(305) Mazhab-Jafari, M. T.; Marshall, C. B.; Smith, M. J.; Gasmi-Seabrook, G. M. C.; Stathopulos, P. B.; Inagaki, F.; Kay, L. E.; Neel, B. G.; Ikura, M. Oncogenic and RASopathy-Associated K-RAS Mutations Relieve Membrane-Dependent Occlusion of the Effector-Binding Site. *Proc. Natl. Acad. Sci.* **2015**, *112* (21), 6625–6630. https://doi.org/10.1073/pnas.1419895112.

(306) Mizrachi, D.; Robinson, M.; Ren, G.; Ke, N.; Berkmen, M. A Water-Soluble DsbB Variant That Catalyzes Disulfide Bond Formation in Vivo. *Nat. Chem. Biol.* **2017**, *13* (9), 1022–1028. https://doi.org/10.1038/nchembio.2409.A.

(307) Aduri, N. G.; Ernst, H. A.; Prabhala, B. K.; Bhatt, S.; Boesen, T.; Gajhede, M.; Mirza, O. Human Proton Coupled Folic Acid Transporter Is a Monodisperse Oligomer in the Lauryl Maltose Neopentyl Glycol Solubilized State. *Biochem. Biophys. Res. Commun.* **2017**, *495* (2), 1738–1743. https://doi.org/10.1016/j.bbrc.2017.12.008.

(308) Gewering, T.; Januliene, D.; Ries, A. B.; Moeller, A. Know Your Detergents: A Case Study on Detergent Background in Negative Stain Electron Microscopy. *J. Struct. Biol.* **2018**, *203* (3), 242–246. https://doi.org/10.1016/j.jsb.2018.05.008.

(309) Fischer, H.; De Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F. Determination of the Molecular Weight of Proteins in Solution from a Single Small-Angle X-Ray Scattering Measurement on a Relative Scale. *J. Appl. Crystallogr.* **2010**, *43* (1), 101–109. https://doi.org/10.1107/S0021889809043076.

(310) Piiadov, V.; Ares de Araújo, E.; Oliveira Neto, M.; Craievich, A. F.; Polikarpov, I. SAXSMoW 2.0: Online Calculator of the Molecular Weight of Proteins in Dilute Solution from Experimental SAXS Data Measured on a Relative Scale. *Protein Sci.* **2019**, *28* (2), 454–463. https://doi.org/10.1002/pro.3528.

(311) Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. VADAR: A Web Server for Quantitative Evaluation of Protein Structure Quality. *Nucleic Acids Res.* **2003**, *31* (13), 3316–3319. https://doi.org/10.1093/nar/gkg565.

(312) Rambo, R.; Tainer, J. Accurate Assessment of Mass, Models, and Resolution by Small-Angle Scattering. *Nature* **2013**, *496*, 477–481. https://doi.org/10.1016/j.physbeh.2017.03.040.

(313) Korasick, D. A.; Tanner, J. J. Determination of Protein Oligomeric Structure from Small-Angle X-Ray Scattering. *Protein Sci.* **2018**, *27* (4), 814–824. https://doi.org/10.1002/pro.3376.

(314) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; et al. BioMagResBank. *Nucleic Acids Res.*

2007, *36* (suppl_1), D402–D408. https://doi.org/10.1093/nar/gkm957.

(315) Cheng, Y. Membrane Protein Structural Biology in the Era of Single Particle Cryo-EM. *Curr. Opin. Struct. Biol.* **2018**, *52*, 58–63. https://doi.org/https://doi.org/10.1016/j.sbi.2018.08.008.

(316) Kwon, S. K.; Kim, S. K.; Lee, D. H.; Kim, J. F. Comparative Genomics and Experimental Evolution of Escherichia Coli BL21(DE3) Strains Reveal the Landscape of Toxicity Escape from Membrane Protein Overproduction. *Sci. Rep.* **2015**, *5* (May), 16076. https://doi.org/10.1038/srep16076.

(317) Studier, F. W. Use of Bacteriophage T7 Lysozyme to Improve an Inducible T7 Expression System. *J. Mol. Biol.* **1991**, *219* (1), 37–44. https://doi.org/10.1016/0022-2836(91)90855-Z.

(318) van der Spoel, E.; Rozing, M. P.; Houwing-Duistermaat, J. J.; Eline Slagboom, P.; Beekman, M.; de Craen, A. J. M.; Westendorp, R. G. J.; van Heemst, D. *Heterologous Expression of Membrane Proteins*; 2015; Vol. 7. https://doi.org/10.1017/CBO9781107415324.004.

(319) Burgess-Brown, N. A.; Sharma, S.; Sobott, F.; Loenarz, C.; Oppermann, U.; Gileadi, O. Codon Optimization Can Improve Expression of Human Genes in Escherichia Coli: A Multi-Gene Study. *Protein Expr. Purif.* **2008**, *59* (1), 94–102. https://doi.org/10.1016/j.pep.2008.01.008.

(320) Schlegel, S.; Klepsch, M.; Gialama, D.; Wickström, D.; Slotboom, D. J.; De Gier, J. W. Revolutionizing Membrane Protein Overexpression in Bacteria. *Microb. Biotechnol.* **2010**, *3* (4), 403–411. https://doi.org/10.1111/j.1751-7915.2009.00148.x.

(321) Schlegel, S.; Hjelm, A.; Baumgarten, T.; Vikström, D.; De Gier, J. W. Bacterial-Based Membrane Protein Production. *Biochim. Biophys. Acta - Mol. Cell Res.* **2014**, *1843* (8), 1739–1749. https://doi.org/10.1016/j.bbamcr.2013.10.023.

(322) Luirink, J.; Yu, Z.; Wagner, S.; De Gier, J. W. Biogenesis of Inner Membrane Proteins in Escherichia Coli. *Biochim. Biophys. Acta - Bioenerg.* **2012**, *1817* (6), 965–976. https://doi.org/10.1016/j.bbabio.2011.12.006.

(323) Sambrook, J.; Russell, D. *Molecular Cloning: A Laboratory Manual 3rd Edition.*, 3rd ed.; Cold Spring Harbor Laboratory Press, 2001.

(324) Lian, J.; Ding, S.; Cai, J. Improving Aquaporin Z Expression in Escherichia Coli by Fusion Partners and Subsequent Condition Optimization. **2009**, 463–470. https://doi.org/10.1007/s00253-008-1774-x.

(325) Jurtshuk, P. Medical Microbiology. 4th Edition. In *Medical Microbiology*; Baron, S., Ed.; University of Texas Medical Branch at Galveston: Galveston, 1996.

(326) Presser, K.; Ratkowsky, D.; Ross, T. Modelling the Growth Rate of Escherichia Coli as a Function of PH and Lactic Acid Concentration. *Appl. Environ. Microbiol.* **1997**,

63 (6), 2355–2360.

(327) Arachea, B. T.; Sun, Z.; Potente, N.; Malik, R.; Isailovic, D.; Viola, R. E. Detergent Selection for Enhanced Extraction of Membrane Proteins. *Protein Expr. Purif.* **2012**, *86* (1), 12–20. https://doi.org/10.1016/j.pep.2012.08.016.

(328) Chen, Y.; Clarke, O. B.; Kim, J.; Stowe, S.; Kim, Y.-K.; Assur, Z.; Cavalier, M.; Godoy-Ruiz, R.; von Alpen, D. C.; Manzini, C.; et al. Structure of the STRA6 Receptor for Retinol Uptake. *Science (80-.).* **2016**, *353* (6302), aad8266. https://doi.org/10.1126/science.aad8266.

(329) White, J. F.; Noinaj, N.; Shibata, Y.; Love, J.; Kloss, B.; Xu, F.; Gvozdenovic-Jeremic, J.; Shah, P.; Shiloach, J.; Tate, C. G.; et al. Structure of the Agonist-Bound Neurotensin Receptor. *Nature* **2012**, *490*, 508.

(330) Kang, Y.; Zhou, X. E.; Gao, X.; He, Y.; Liu, W.; Ishchenko, A.; Barty, A.; White, T. A.; Yefanov, O.; Han, G. W.; et al. Crystal Structure of Rhodopsin Bound to Arrestin by Femtosecond X-Ray Laser. *Nature* **2015**, *523*, 561.

(331) Thompson, A. A.; Liu, J. J.; Chun, E.; Wacker, D.; Wu, H.; Stevens, R. C. GPCR Stabilization Using the Bicelle-like Architecture of Mixed Sterol-Detergent Micelles. *Methods* **2011**, *55* (4), 310–317. https://doi.org/10.1016/j.ymeth.2011.10.011.GPCR.

(332) Chen, Y.; Clarke, O. B.; Kim, J.; Stowe, S.; Kim, Y. K.; Assur, Z.; Cavalier, M.; Godoy-Ruiz, R.; Von Alpen, D. C.; Manzini, C.; et al. Structure of the STRA6 Receptor for Retinol Uptake. *Science (80-.).* **2016**, *353* (6302). https://doi.org/10.1126/science.aad8266.

(333) Grouleff, J.; Irudayam, S. J.; Skeby, K. K.; Schiøtt, B. The Influence of Cholesterol on Membrane Protein Structure, Function, and Dynamics Studied by Molecular Dynamics Simulations. *Biochim. Biophys. Acta 1848* **2015**, 1783–1795. https://doi.org/10.1016/j.bbamem.2015.03.029.

(334) Wheatley, M.; Charlton, J.; Jamshad, M.; Routledge, S. J.; Bailey, S.; La-Borde, P. J.; Azam, M. T.; Logan, R. T.; Bill, R. M.; Dafforn, T. R.; et al. GPCR–Styrene Maleic Acid Lipid Particles (GPCR–SMALPs): Their Nature and Potential. *Biochem. Soc. Trans.* **2016**, *44* (2), 619–623. https://doi.org/10.1042/BST20150284.

(335) Yancey, P. Water Stress, Osmolytes and Proteins. *Integr. Comp. Biol.* **2001**, *41*, 699–709.

(336) Judy, E.; Kishore, N. Biological Wonders of Osmolytes: The Need to Know More. *Biochem. Anal. Biochem.* **2016**, *05* (04). https://doi.org/10.4172/2161-1009.1000304.

(337) Jensen, M. Ø.; Tajkhorshid, E.; Schulten, K. Electrostatic Tuning of Permeation and Selectivity in Aquaporin Water Channels. **2003**, *85* (November), 2884–2899.

(338) Broecker, J.; Klingel, V.; Ou, W.; Balo, A. R.; Kissick, D. J.; Ogata, C. M.; Kuo, A.;

Ernst, O. P. A Versatile System for High-Throughput In Situ X - Ray Screening and Data Collection of Soluble and Membrane-Protein Crystals. **2016**. https://doi.org/10.1021/acs.cgd.6b00950.

(339) Baldock, D.; Sheldon, J.; Tailor, R.; Green, K.; Ray, J. An Improved Method for the Expression Screening of Membrane Protein-GFP Fusions in Yeast. **2017**.

(340) Kelly, A. E.; Ou, H. D.; Withers, R.; Do, V.; Corp, B. B.; Hcl, T. Low-Conductivity Buffers for High-Sensitivity NMR Measurements Copy Is a Relatively Insensitive Method , Requiring Concentrations. **2002**, No. 14, 12013–12019. https://doi.org/10.1021/ja026121b.

(341) Seddon, A. M.; Curnow, P.; Booth, P. J. Membrane Proteins, Lipids and Detergents: Not Just a Soap Opera. *Biochim. Biophys. Acta* **2004**, 105–117. https://doi.org/10.1016/j.bbamem.2004.04.011.

(342) Awad, S.; Allison, S.; Lobo, D. The History of 0.9% Saline. *Clin. Nutr.* **2008**, *27*, 179–188.

(343) Lund, S.; Orlowski, S.; De Foresta, B.; Champeil, P.; Le Maire, M.; Moller, J. V. Detergent Structure and Associated Lipid as Determinants in the Stabilization of Solubilized Ca2+-ATPase from Sarcoplasmic Reticulum. *J. Biol. Chem.* **1989**, *264* (9), 4907–4915.

(344) Frey, L.; Lakomek, N. A.; Riek, R.; Bibow, S. Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone Dynamics. *Angew. Chemie - Int. Ed.* **2017**, *56* (1), 380–383. https://doi.org/10.1002/anie.201608246.

(345) Kofuku, Y.; Ueda, T.; Okude, J.; Shiraishi, Y.; Kondo, K.; Mizumura, T.; Suzuki, S.; Shimada, I. Functional Dynamics of Deuterated B2-Adrenergic Receptor in Lipid Bilayers Revealed by NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **2014**, *53* (49), 13376–13379. https://doi.org/10.1002/anie.201406603.

(346) Pocanschi, C. L.; Dahmane, T.; Gohon, Y.; Rappaport, F.; Apell, H.-J.; Kleinschmidt, J. H.; Popot, J.-L. Amphipathic Polymers: Tools To Fold Integral Membrane Proteins to Their Active Form. *Biochemistry* **2006**, *45* (47), 13954–13961. https://doi.org/10.1021/bio616706.

(347) Zoonens, M.; Popot, J.-L. Amphipols for Each Season. *2014 247*, 759–796. https://doi.org/10.1038/nn.3871.Locus-Specific.

(348) Gohon, Y.; Dahmane, T.; Ruigrok, R. W. H.; Schuck, P.; Charvolin, D.; Rappaport, F.; Timmins, P.; Engelman, D. M.; Tribet, C.; Popot, J.-L.; et al. Bacteriorhodopsin/Amphipol Complexes: Structural and Functional Properties. *Biophys. J.* **2008**, *94* (9), 3523–3537. https://doi.org/https://doi.org/10.1529/biophysj.107.121848.

(349) Picard, M.; Dahmane, T.; Garrigos, M.; Gauron, C.; Giusti, F.; le Maire, M.; Popot, J.-

L.; Champeil, P. Protective and Inhibitory Effects of Various Types of Amphipols on the Ca2+-ATPase from Sarcoplasmic Reticulum: A Comparative Study. *Biochemistry* **2006**, *45* (6), 1861–1869. https://doi.org/10.1021/bi051954a.

(350) Toyoshima, C.; Nomura, H. Structural Changes in the Calcium Pump Accompanying the Dissociation of Calcium. *Nature* **2002**, *418* (6898), 605–611. https://doi.org/10.1038/nature00944.

(351) Barniol-Xicota, M.; Verhelst, S. H. L. Stable and Functional Rhomboid Proteases in Lipid Nanodiscs by Using Diisobutylene/Maleic Acid Copolymers. *J. Am. Chem. Soc.* **2018**, *140* (44), 14557–14561. https://doi.org/10.1021/jacs.8b08441.

(352) Kopf, A. H.; Koorengevel, M. C.; van Walree, C. A.; Dafforn, T. R.; Killian, J. A. A Simple and Convenient Method for the Hydrolysis of Styrene-Maleic Anhydride Copolymers to Styrene-Maleic Acid Copolymers. *Chem. Phys. Lipids* **2019**, *218*, 85–90. https://doi.org/https://doi.org/10.1016/j.chemphyslip.2018.11.011.

(353) Danielczak, B.; Meister, A.; Keller, S. Influence of Mg2+ and Ca2+ on Nanodisc Formation by Diisobutylene/Maleic Acid (DIBMA) Copolymer. *Chem. Phys. Lipids* **2019**, *221*, 30–38. https://doi.org/https://doi.org/10.1016/j.chemphyslip.2019.03.004.

(354) Christensen, H.; Pain, R. Molten Globule Intermediates and Protein Folding. *Eur. Biophys. J.* **1991**, *19* (5), 221–229.

(355) Kang, Y.; Zhou, X. E.; Gao, X.; He, Y.; Liu, W.; Ishchenko, A.; Barty, A.; White, T. A.; Yefanov, O.; Han, G. W.; et al. Crystal Structure of Rhodopsin Bound to Arrestin by Femtosecond X-Ray Laser. *Nature* **2015**, *523* (7562), 561–567. https://doi.org/10.1038/nature14656.

(356) Markwick, P. R. L.; Malliavin, T.; Nilges, M.; DePristo, M. A.; Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M.; Dobson, C. M.; Buldyrev, S. V.; Chen, S.-H.; et al. Yeast-Expressed Human Membrane Protein Aquaporin-1 Yields Excellent Resolution of Solid-State MAS NMR Spectra. *Biophys. J.* **2014**, *18* (1), 147–155. https://doi.org/10.1039/c3cs60474h.

(357) O'Malley, M. A.; Helgeson, M. E.; Wagner, N. J.; Robinson, A. S. The Morphology and Composition of Cholesterol-Rich Micellar Nanostructures Determine Transmembrane Protein (GPCR) Activity. *Biophys. J.* **2011**, *100* (2), L11–L13. https://doi.org/10.1016/j.bpj.2010.12.3698.

(358) Sáenz, J. P.; Grosser, D.; Bradley, A. S.; Lagny, T. J.; Lavrynenko, O.; Broda, M. Hopanoids as Functional Analogues of Cholesterol in Bacterial Membranes. *PNAS* **2015**, *112* (38), 11971–11976. https://doi.org/10.1073/pnas.1515607112.

(359) Sáenz, J. P.; Sezgin, E.; Schwille, P.; Simons, K. Functional Convergence of Hopanoids and Sterols in Membrane Ordering. *Proc. Natl. Acad. Sci.* **2012**, *109* (35), 14236 LP – 14240. https://doi.org/10.1073/pnas.1212141109.

(360) VanAken, T.; Foxall-VanAken, S.; Castleman, S.; Ferguson-Miller, S. Alkyl Glycoside

Detergents: Synthesis and Applications to the Study of Membrane Proteins. *Methods Enzymol.* **1986**, *125*, 27–35.

(361) Chung, K. Y.; Kim, T. H.; Manglik, A.; Alvares, R.; Kobilka, B. K.; Prosser, R. S. Role of Detergents in Conformational Exchange of a G Protein-Coupled Receptor. *J. Biol. Chem.* **2012**, *287* (43), 36305–36311. https://doi.org/10.1074/jbc.M112.406371.

(362) Glenner, G. G.; Wong, C. W. Alzheimer's Disease: Initial Report of the Purification and Characterization of a Novel Cerebrovascular Amyloid Protein. *Biochem. Biophys. Res. Commun.* **1984**, *120* (3), 885–890. https://doi.org/https://doi.org/10.1016/S0006-291X(84)80190-4.

(363) Come, J. H.; Fraser, P. E.; Lansbury Jr, P. T. A Kinetic Model for Amyloid Formation in the Prion Diseases: Importance of Seeding. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (13), 5959–5963. https://doi.org/10.1073/pnas.90.13.5959.

(364) Marzban, L.; Park, K.; Verchere, C. B. Islet Amyloid Polypeptide and Type 2 Diabetes. *Exp. Gerontol.* **2003**, *38* (4), 347–351. https://doi.org/https://doi.org/10.1016/S0531-5565(03)00004-4.

(365) Rambaran, R. N.; Serpell, L. C. Amyloid Fibrils: Abnormal Protein Assembly. *Prion* **2008**, *2* (3), 112–117. https://doi.org/10.4161/pri.2.3.7488.

(366) Burgess, S. A.; Walker, M. L.; Thirumurugan, K.; Trinick, J.; Knight, P. J. Use of Negative Stain and Single-Particle Image Processing to Explore Dynamic Properties of Flexible Macromolecules. *J. Struct. Biol.* **2004**, *147* (3), 247–258. https://doi.org/https://doi.org/10.1016/j.jsb.2004.04.004.

(367) Golub, M.; Hejazi, M.; Kölsch, A.; Lokstein, H.; Wieland, D. C. F.; Zouni, A.; Pieper, J. Solution Structure of Monomeric and Trimeric Photosystem I of Thermosynechococcus Elongatus Investigated by Small-Angle X-Ray Scattering. *Photosynth. Res.* **2017**, *133* (1–3), 163–173. https://doi.org/10.1007/s11120-017-0342-6.

(368) Kikhney, A. G.; Svergun, D. I. A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589* (19), 2570–2577. https://doi.org/10.1016/j.febslet.2015.08.027.

(369) Drenth, J. Practical Protein Crystallization. In *Principles of X-Ray Protein Crystallography*; Groningen, 2007; p 303.

(370) Miroux, B.; Walker, J. Over-Production of Proteins in Escherichia Coli: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* **1996**, *260*, 289–298. https://doi.org/10.1006/jmbi.1996.0399.

(371) Varshavsky, A. N-Degron and C-Degron Pathways of Protein Degradation. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (2), 358–366. https://doi.org/10.1073/pnas.1816596116.

(372) García-Fruitós, E.; Villaverde, A. Friendly Production of Bacterial Inclusion Bodies. *Korean J. Chem. Eng.* **2010**, *27* (2), 385–389. https://doi.org/10.2478/s11814-010-0161-3.

(373) Park, S. H.; Das, B. B.; Casagrande, F.; Tian, Y.; Nothnagel, H. J.; Chu, M.; Kiefer, H.; Maier, K.; De Angelis, A. A.; Marassi, F. M.; et al. Structure of the Chemokine Receptor CXCR1 in Phospholipid Bilayers. *Nature* **2012**, *491*, 779.

(374) Booth, P. J.; Curnow, P. Membrane Proteins Shape up: Understanding in Vitro Folding. *Curr. Opin. Struct. Biol.* **2006**, *16* (4), 480–488. https://doi.org/10.1016/j.sbi.2006.06.004.

(375) Anbazhagan, V.; Schneider, D. The Membrane Environment Modulates Self-Association of the Human GpA TM Domain-Implications for Membrane Protein Folding and Transmembrane Signaling. *Biochim. Biophys. Acta - Biomembr.* **2010**, *1798* (10), 1899–1907. https://doi.org/10.1016/j.bbamem.2010.06.027.

(376) Yeagle, P. L. Non-Covalent Binding of Membrane Lipids to Membrane Proteins ☆. *BBA - Biomembr.* **2014**, *1838* (6), 1548–1559. https://doi.org/10.1016/j.bbamem.2013.11.009.

(377) Whiles, J. A.; Glover, K. J.; Vold, R. R.; Komives, E. A. Methods for Studying Transmembrane Peptides in Bicelles: Consequences of Hydrophobic Mismatch and Peptide Sequence. *J. Magn. Reson.* **2002**, *158* (1–2), 149–156. https://doi.org/10.1016/S1090-7807(02)00068-X.

(378) Schubeis, T.; Le Marchand, T.; Andreas, L. B.; Pintacuda, G. 1H Magic-Angle Spinning NMR Evolves as a Powerful New Tool for Membrane Proteins. *J. Magn. Reson.* **2018**, *287*, 140–152. https://doi.org/10.1016/j.jmr.2017.11.014.

(379) Shahid, S. A.; Markovic, S.; Linke, D.; Van Rossum, B. J. Assignment and Secondary Structure of the YadA Membrane Protein by Solid-State MAS NMR. *Sci. Rep.* **2012**, *2*, 1–10. https://doi.org/10.1038/srep00803.

(380) Kunert, B.; Gardiennet, C.; Lacabanne, D.; Calles-Garcia, D.; Falson, P.; Jault, J.-M.; Meier, B. H.; Penin, F.; BÃ¶ckmann, A. Efficient and Stable Reconstitution of the ABC Transporter BmrA for Solid-State NMR Studies. *Front. Mol. Biosci.* **2014**, *1* (June), 1–11. https://doi.org/10.3389/fmolb.2014.00005.

(381) V. S., M.; Gopinath, T.; Wang, S.; Veglia, G. T2* Weighted Deconvolution of NMR Spectra: Application to 2D Homonuclear MAS Solid-State NMR of Membrane Proteins. *Sci. Rep.* **2019**, *9* (1), 8225. https://doi.org/10.1038/s41598-019-44461-3.