

Solution State Characterization of the *E. coli* Inner Membrane Protein Glycerol Facilitator

Jamie John Galka

Table of Contents

Solution State Characterization of the <i>E. coli</i> Inner Membrane Protein Glycerol	
Facilitator	i
Acknowledgements	iv
List of Figures	vi
List of Tables	viii
Abbreviations	viii
Abstract	ix
Chapter 1 Introduction	1
1.1 The Cell and Cellular Membranes: A Brief History	1
1.1.1 Lipids	2
1.1.2 Biological Membranes	4
1.1.3 Membrane Proteins	9
1.1.3.1 Membrane-Associated Proteins	10
1.1.3.2 Integral Membrane Proteins	11
1.2 Protein Structure and Folding	13
1.2.1 Protein Folding Models	16
1.2.2 Membrane Protein Folding	22
1.2.3 Membrane Protein Insertion	23
1.2.4 Membrane Protein Recognition	25
1.2.5 Exogenous Membrane Protein Assembly	26
1.2.6 The Energetics of Proteins in the Membrane	28
1.3 Major Intrinsic Protein Superfamily	33
1.3.1 Aquaporins and Aquaglyceroporins	34
1.3.1.1 Structure of the AQP1 Channel	37
1.3.2 Structure of the GlpF Channel	40
1.3.2.1 The Structure of GlpF	41
1.3.2.2 Glycerol Transport Through GlpF	43
1.3.2.3 Folding and Stability of the AQPs and GLPs	45
1.4 Biophysical Techniques	47
1.4.1 Circular Dichroism	47
1.4.1.1 Far UV CD	48
1.4.1.2 Near UV CD	50
1.4.2 Fluorescence Spectroscopy	52
1.5 Purpose of the Research	54
Chapter 2 Materials and Methods	55
2.1 Materials	55
2.2 Methods	56
2.2.1 Glycerol Facilitator Expression and Purification	56
2.2.2 Electrophoresis	57
2.2.3 Circular Dichroism and Fluorescence	58
2.2.4 Curve Fitting	59

Chapter 3 Results	61
Biophysical Characterization of GlpF in Detergent Solutions	61
3.1 Quaternary Structure Analysis of Detergent Solubilized Glycerol Facilitator	61
3.1.1 Glycerol Facilitator Quaternary Structure in Sodium Dodecyl Sulfate Solution	62
3.1.2 Glycerol Facilitator Quaternary Structure in Octyl Glucoside Solution.....	64
3.1.3 Glycerol Facilitator Quaternary Structure in Dodecyl Maltoside Solution	65
3.1.4 Glycerol Facilitator Quaternary Structure in Lyso-Myristoyl Phosphatidylcholine Solution	78
3.2 Tertiary Structure Analysis of Detergent-Solubilized Glycerol Facilitator	83
3.2.1 Glycerol Facilitator Tertiary Structure in Dodecyl Maltoside Solution	83
3.2.2 Glycerol Facilitator Tertiary Structure in Lyso-Myristoyl Phosphatidylcholine Solution	95
3.2.3 Glycerol Facilitator Tertiary Structure in Sodium Dodecyl Sulphate Solution	103
3.2.4 Glycerol Facilitator Tertiary Structure in Octyl Glucoside Solution.....	105
3.3 Secondary Structure Analysis of Detergent-Solubilized Glycerol Facilitator	105
3.3.1 Glycerol Facilitator Secondary Structure in Dodecyl Maltoside Solution	105
3.3.2 Glycerol Facilitator Secondary Structure in Lyso-Myristoyl Phosphatidylcholine Solution	110
3.3.3 Glycerol Facilitator Secondary Structure in Sodium Dodecyl Sulphate Solution	114
Chapter 4 Discussion	117
4.1 Glycerol Facilitator in Dodecyl Maltoside and Lyso-Myristoyl Phosphatidylcholine Solutions at Neutral pH.....	117
4.2 Effects of Heat on α -Helical Membrane Proteins	123
4.2.1 Effects of Heat on Glycerol Facilitator in Dodecyl Maltoside Solution and Lyso-Myristoyl Phosphatidylcholine Solutions at Neutral pH.....	127
4.3 Effects of Urea	134
4.4 Influence of pH on Protein Structure	136
4.4.1 Influence of pH on GlpF Structure	138
4.6 Glycerol facilitator in Sodium Dodecyl Sulfate and Octyl Glucoside solutions ..	140
4.7 Future Work	142
Bibliography	143

Acknowledgements

This section may be one of the most difficult, for without the contributions of many people this work would not have been possible. I am not a person who loves to write, so forgive me if the “thank yous” are brief; it is not for lack of appreciation.

I would like to first thank my family; without the support and love from my dear wife Rolande I would not have been able to succeed in this venture. To my beautiful children Sophie and Elliot, thank you for making me realize (even it was at 3:00 a.m.) that there are more important things in life than “writing and corrections”. To my parents and siblings, thank you for listening and for encouragement over the years.

To everyone in the O’Neil lab during my (many) years, thank you for all the help and useful discussions. A special thanks to Shaheen Shojania for all the encouragement when it meant the most. I very much enjoyed the years spent together and the excellent friendship which has come out of it.

To the entire thesis committee, thank you for all your hard work of the years for annual reviews, etc., and for all the excellent comments on the thesis.

To my supervisor, Joe O'Neil, thank you very much for the continued support over the years; without your guidance and support I would never have been able to complete such an endeavour.

List of Figures

Figure 1: Diagram of a typical Langmuir trough.....	3
Figure 2: Typical lipid bilayer, as described by Gorter and Grendel.....	5
Figure 3: Davson and Danielli membrane model.....	6
Figure 4: The Singer-Nicholson fluid mosaic model of the cell membrane.....	8
Figure 5: Protein folding: prediction from genome to function.....	14
Figure 6: Reaction coordinates for a two-state folding process.....	17
Figure 7: The sigmoidal transition curve characteristic of a two-state equilibrium transition.....	18
Figure 8: Representation of the protein folding funnel.....	21
Figure 9: The two main structural motifs of membrane proteins are the α -helix and the β -barrel.....	23
Figure 10: The assembly of endogenous integral membrane proteins involves a temporary arrangement of the ribosome (with mRNA) and the translocon.....	27
Figure 11: Helix folding, insertion and association: the four conceptual thermodynamic steps.....	29
Figure 12: SDS intercalation of membrane proteins.....	33
Figure 13: Hourglass representation of the AQP1 protein.....	38
Figure 14: The GlpF tetramer.....	41
Figure 15: The hydrogen bonding network of GlpF as glycerol passes through the channel.....	44
Figure 16: Far UV CD of predominantly α -helical, β -sheet and coil protein conformations.....	49
Figure 17: SDS PAGE electrophoregram of GlpF dissolved in 150 mM SDS solution eluted from the Ni-NTA-resin at pH 4.....	63
Figure 18: SDS PAGE electrophoregram of GlpF dissolved in 150 mM SDS solution eluted from the Ni-NTA-resin using imidazole competition.....	64
Figure 19: SDS PAGE electrophoregram of GlpF dissolved in OG solution eluted from the Ni-NTA-resin using low pH.....	65
Figure 20: SDS PAGE electrophoregrams of thermal unfolding of GlpF.....	66
Figure 21: SDS PAGE electrophoregram of a pH titration of DDM-solubilized GlpF.....	69
Figure 22: SDS titration of pH 4 DDM-solubilized GlpF observed by SDS PAGE.....	71
Figure 23: BN PAGE electrophoregram showing the thermal unfolding of imidazole-eluted pH 7 DDM-solubilized GlpF.....	73
Figure 24: BN PAGE electrophoregram showing pH induced unfolding of imidazole-eluted DDM-solubilized GlpF.....	74
Figure 25: SDS PAGE electrophoregram monitoring the effects of urea on pH 7 DDM-solubilized GlpF after 3 weeks.....	76
Figure 26: BN PAGE electrophoregram monitoring the effects of urea on GlpF after 3 weeks.....	77
Figure 27: SDS (A) and BN (B) electrophoregrams showing the thermal denaturation of imidazole-eluted LMPC-solubilized GlpF.....	79

Figure 28: SDS (A) and BN (B) electrophoregrams of pH titrations of imidazole-eluted LMPC-solubilized GlpF.....	81
Figure 29: Near UV CD spectrum of DDM-solubilized GlpF protein (69.4 μM) at pH 7 and 25 $^{\circ}\text{C}$	84
Figure 30: Spectroscopic analysis of the thermal denaturation of imidazole-eluted, pH 7 DDM-solubilized GlpF (69.4 μM) tertiary structure monitored by near UV CD.....	85
Figure 31: Molar ellipticity changes of DDM-solubilized GlpF (69.4 μM) at 268 nm... ..	87
Figure 32: Thermal denaturation of DDM-solubilized GlpF (2.2 μM) monitored by fluorescence spectroscopy ($\lambda_{\text{ex}} = 280 \text{ nm}$).....	88
Figure 33: Spectroscopic analysis of the pH titration of DDM-solubilized GlpF (69.1 μM) tertiary structure monitored by near UV CD.	90
Figure 34: Acid induced unfolding of DDM-solubilized GlpF (2.2 μM) monitored by fluorescence spectroscopy ($\lambda_{\text{ex}} = 280 \text{ nm}$).....	91
Figure 35: Spectroscopic analysis of the urea titration of pH 7 DDM-solubilized GlpF (55.4 μM) tertiary structure monitored by near UV CD.	93
Figure 36: Spectroscopic analysis of the time dependence of 8 M urea on pH 7 DDM-solubilized GlpF (55.4 μM) tertiary structure by near UV CD.....	94
Figure 37: Molar ellipticity changes at 268 nm of DDM-solubilized GlpF (55.4 μM) in 8 M urea as a function of time (closed squares) fit to an exponential decay (Equation 2, Materials and Methods).	95
Figure 38: Near UV CD spectrum of LMPC-solubilized GlpF (25.0 μM) at pH 7 and 25 $^{\circ}\text{C}$	96
Figure 39: LMPC-solubilized GlpF molar ellipticity changes at 268 nm.	98
Figure 40: Thermal denaturation of LMPC-solubilized GlpF (3.4 μM) monitored by fluorescence spectroscopy ($\lambda_{\text{ex}} = 280 \text{ nm}$).....	100
Figure 41: pH titration of LMPC-solubilized GlpF (13.7 μM) tertiary structure monitored by near UV CD.	101
Figure 42: Figure 3-26: Acid induced denaturation of LMPC-solubilized GlpF (3.4 μM) monitored by fluorescence spectroscopy ($\lambda_{\text{ex}} = 280 \text{ nm}$).	102
Figure 43: Near UV CD spectrum of imidazole-eluted, pH 7 SDS-solubilized GlpF (14.0 μM).	104
Figure 44: Far UV CD spectrum of DDM-solubilized GlpF (3.0 μM) at pH 7, and 25 $^{\circ}\text{C}$	106
Figure 45: Thermal denaturation of DDM-solubilized GlpF (3.0 μM) secondary structure monitored by far-UV CD.....	107
Figure 46: DDM solubilized GlpF (3.0 μM) mean residue ellipticity ($\text{deg cm}^2/\text{dmole}$) and fractional helix changes at 209 nm as a function of increasing temperature (60 $^{\circ}\text{C}/\text{h}$) fit to a two-state equilibrium model.....	109
Figure 47: pH titration of DDM-solubilized GlpF (3.0 μM) secondary structure monitored by far-UV CD.....	110
Figure 48: Thermal unfolding of LMPC-solubilized GlpF (1.5 μM) secondary structure monitored by far-UV CD.....	111
Figure 49: LMPC-solubilized GlpF (1.5 μM) mean residue ellipticity ($\text{deg cm}^2/\text{dmole}$) and fractional helix changes at 209 nm as a function of increasing temperature (60 $^{\circ}\text{C}/\text{h}$) fit to a two-state equilibrium model.....	113

Figure 50: pH titration of LMPC-solubilized GlpF (1.5 μ M) secondary structure monitored by far-UV CD.....	114
Figure 51: Far UV CD spectrum of SDS-solubilized GlpF protein (1.5 μ M) at pH 7. .	116
Figure 52: Energetics and mechanism of membrane protein unfolding.....	126

List of Tables

Table 1: List of known mammalian, <i>E. coli</i> , and archaeobacterial aquaporins, and their transport properties.	36
--	----

Abbreviations

AQP, aquaporin; Bis-Tris, Bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)methane; BN, Blue Native; CD, circular dichroism; DTT, dithiothreitol; DDM, dodecyl- β -D-maltoside; EDTA, ethylenediaminetetraacetic acid; GLP, aquaglyceroporins; GlpF, glycerol facilitator; kDa, kilodalton; LMPC, lysomyristoylphosphatidylcholine; MALDI, matrix-assisted laser desorption ionization; MP, membrane proteins; M_r , relative mass; MRE, Mean Residue Ellipticity; NMR, Nuclear Magnetic Resonance; NTA, nitrilotriacetic acid; OG, octyl- β -D-glucopyranoside; PAGE, polyacrylamide gel electrophoresis; PCC, protein-conducting channel; pI, isoelectric point; SDS sodium dodecyl sulphate; SRP, signal-recognition particle; TOFMS, time-of-flight mass spectrometry; Tris, tris(hydroxymethyl)aminomethane; UV, ultra-violet; XRD, X-ray diffraction; Θ , degrees.

Abstract

The Major Intrinsic Proteins are represented in all forms of life; plants, animals, bacteria and recently archaeobacteria have all been shown to express at least one member of this superfamily of integral membrane proteins. We have overexpressed the *E. coli* aquaglyceroporin, glycerol facilitator (GlpF), to use as a model for studying membrane protein structure, folding and stability. Understanding membrane protein folding, stability, and dynamics is required for a molecular explanation of membrane protein function and for the development of interventions for the hundreds of membrane protein folding diseases. X-ray analysis of GlpF crystals shows that the protein exists as a tetramer in the crystallized state [1]. However, preparations of stable aqueous detergent solutions of GlpF in its native oligomeric state have been difficult to make; the protein readily unfolds and forms non-specific aggregates in many detergents. Here, I report the study of the structure and stability of the glycerol facilitator in several detergent solutions by blue native and sodium dodecyl sulphate polyacrylamide gel electrophoresis, circular dichroism, and fluorescence. For the first time, stable protein tetramers were prepared in two different detergent solutions (dodecyl maltoside (DDM) and lyso-myristoyl phosphatidylcholine (LMPC)) at neutral pH. Thermal unfolding experiments show that the protein is slightly more stable in LMPC than in DDM and that the thermal stability of the helical core at 95°C is slightly greater in the former detergent. In addition, tertiary structure unfolds before quaternary and secondary structures in LMPC whereas unfolding is more cooperative in DDM. The high stability of the protein is also evident from the unfolding half-life of 8 days in 8 M urea suggesting that hydrophobic interactions contribute to the stability. The GlpF tetramers are less resistant to acidic conditions;

LMPC-solubilized GlpF shows loss of tertiary and quaternary structure by pH 6, while in DDM the tertiary structure is lost by pH 5, however the tetramer remains mostly intact at pH 4. The implications of thermal and chemical stress on the stability of the detergent-solubilized protein and its *in vivo* folding are discussed.

Chapter 1 Introduction

1.1 The Cell and Cellular Membranes: A Brief History

The cell membrane is one of the most unifying themes in all of biology. With the exception of some viruses, all living organisms are reliant on the cell membrane in some fashion. The cell membrane is a selectively permeable barrier that serves many important functions such as: containing the cell contents so they do not mix with the extracellular environment; facilitating the transport of materials necessary for survival of the cell; separating different metabolic regions within the cell; allowing for the formation of electrochemical gradients used in energy transduction; providing the required scaffolding needed for anchoring cellular receptors responsible for communication between cells and the environment; and much more.

Dr. Robert Hooke is credited with the discovery of the “cell” in experiments observing the structure of cork, and published in the historical book “*Micrographia*” [2]. It is said that he coined the term “cell” because the walls around the cork reminded him of monks' chambers called "*cellula*".

The eukaryotic cell can be sectioned into three main parts: the nucleus, which contains genetic information, the cytoplasm, where most cellular processes take place, and finally,

the cell membrane, which regulates the influx and efflux of different molecules. Within the cytoplasm, other membranous systems also exist (for example the golgi apparatus, endoplasmic reticulum, mitochondria, chloroplasts, *etc.*). We now know a great deal about membranes, but it was not always so.

1.1.1 Lipids

The understanding of cell membranes did not originate with studies on biological systems as might be expected, but with chemistry, and studies on the interaction of lipids with water. It has been suggested that the most likely person to first scientifically study the effects of oil and water was Benjamin Franklin [3] – an unlikely candidate indeed.

During a stay in England in 1774, Franklin conducted studies on the effects of oil on the surface of water, and found that upon the addition of the oil to the water, the oil spread into a very thin film over a large portion of the water prompting Franklin to comment that the pond appeared “*as smooth as a looking glass*” [3].

More than a century passed before Franklin’s experiment was repeated by Lord Rayleigh (born John William Strutt) in 1890 [3]. Using quantitative techniques, he calculated the area that a known volume of oil would occupy, and was therefore also able to calculate the thickness of the oil film [3]. Though his results were published, they received little notice in the broader scientific community. This could have remained a scientific side-

note if not for the communication with a German woman named Agnes Pockels; Pockels had developed (in her kitchen) a device which carefully, and accurately, measured the area of an oil film [3]. This device was later modified and improved by Irving Langmuir, and is commonly referred to as the Langmuir trough. It is still used today by scientists studying surface properties of materials (Figure 1). Langmuir was the first to suggest that fatty acid molecules form a monolayer with the polar carboxyl group interacting with the surface of the water, while the nonpolar hydrocarbon chains orient themselves vertically with respect to the water surface [4]. This observation was paramount in solving the puzzle of lipid bilayers and the basic structure of the cell membrane.

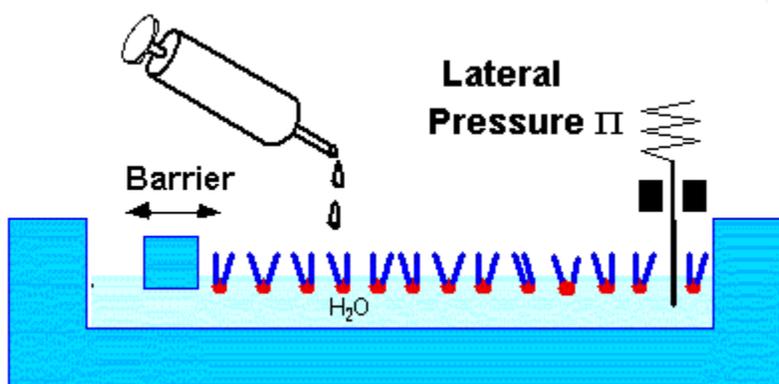


Figure 1: Diagram of a typical Langmuir trough.

The properties of thin films consisting of amphiphilic molecules grafted to the air water interface are investigated with a Langmuir Trough, usually made of Teflon. A moveable barrier seals one side of the water surface, so the molecular area can be varied. Reproduced with permission from [5].

1.1.2 Biological Membranes

At approximately the same time that Raleigh was studying the properties of lipids on the water surface, Charles Ernest Overton was completing a doctoral degree at the University of Zurich [3]. Serendipitously, Overton found that nonpolar molecules would pass rapidly into the cells of the plants he was studying. This is in stark contrast to the then current scientific view that the cell membrane was virtually impervious to all molecules except water [3]. Overton came up with a hypothesis which consisted of two main points: first, that the cell membrane shared similarities to nonpolar molecules such as olive oil, and secondly, that certain molecules were able to pass into the cell by dissolving into the lipidic centre of the cell membrane [3]. At the time, there was considerable opposition to Overton's hypothesis; however, we are now able to appreciate the magnitude of these discoveries.

The first people to purposely study the lipids of the cell membrane were most likely Gorter and Grendel. In their experiments, the cell membrane lipids were extracted with a variety of organic solvents, and were found to be able to form both bilayers, and monolayers [6, 7]. Using a device similar to the Langmuir trough, the surface areas of the lipid monolayers were found to be approximately twice those of the red blood cell membranes from which they were extracted [6, 7]. In more recent years, the experiments of Gorter and Grendel have been found to have suffered from the poor techniques of the time [8]. In fact, Gorter and Grendel's extraction of the red blood cell lipids was incomplete; however, coupled to their underestimation of the area of the red blood cell,

the errors cancelled one another, and the conclusions are basically correct [8]. Based on these observations, Gorter and Grendel were able to devise a plausible model of the cell membrane structure. According to the measurements on red blood cells from many different animals, they concluded that the membrane of these cells must be a bilayer of lipid molecules [6, 7] (Figure 2).

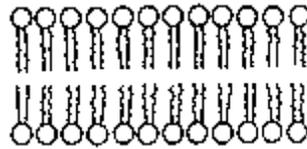


Figure 2: Typical lipid bilayer, as described by Gorter and Grendel.

The lipid bilayer, as described by Gorter and Grendel, consists of two lipid monolayers with the hydrophobic tails pointed inwards towards each other, while the polar head groups remain in contact with the aqueous solvent. Reprinted from [9] with permission.

While the Gorter and Grendel model helped explain the structure of the bilayer, it was not generally accepted as a membrane model, partly due to the decreased water-absorption between that of the synthetic phospholipid bilayers and actual biological membranes [10]. The first model to be well accepted in the scientific community was that of Davson and Danielli [10]. The proposed model was a lipid bilayer, of which both sides were covered with a layer of globular proteins (Figure 3).

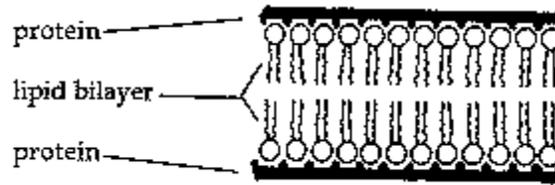


Figure 3: Davson and Danielli membrane model.

The membrane model devised by Davson and Danielli includes the bilayer of Gorter and Grendel, however it also includes layers of proteins on the surfaces of the bilayer. Reprinted from [9] with permission.

It is interesting to note, that no mention of Gorter or Grendel is present in the original model proposed by Davson and Danielli. In the 1950s, electron microscopy allowed scientists to view biological membranes. Cross-sections taken from the electron micrographs showed that the plasma membrane appeared to be 7 to 8 nm in thickness. The Davson-Danielli model predicted a thicker membrane and needed to be revised to incorporate β -strand proteins as opposed to the globular proteins previously proposed to be lining both sides of the membrane. The new version fit the observed thickness of the plasma membrane. In 1957 Robertson proposed a modification of the Davson-Danielli model based on the electron microscopy studies, which was called the “unit membrane” model [11]. Under the high magnification of the electron microscope, images of the heavy metal-stained plasma membrane have a trilaminar appearance with two electron dense layers separated by an electron-transparent region. The appearance of the membrane under the electron microscope further validated the Davson-Danielli model, as most early electron microscopists interpreted the darker regions as adhesion of the stain to the proteins and lipid head groups, while the hydrophobic core of the membrane was expected to be relatively unstained [10]. In a somewhat circular argument, the Davson-

Danielli membrane-model and the electron micrographs were each used as justification for the interpretation of the other. By the 1960s, the Davson-Danielli “*lipo-protein sandwich*” model was widely accepted as the structural model for not only the plasma membrane, but indeed for all cellular membranes.

The success of the model would not last, as scientists realized two main faults with the model; first, the generalization that all membranes were identical was attacked on the grounds that under the electron microscope (the instrument credited for the Davson-Danielli model’s success) inner mitochondrion membranes were significantly thinner than that of the plasma membrane, (6 nm, as opposed to the 7-8 nm of the plasma membrane) and appeared as a row of beads [10]. The second major problem identified with the Davson-Danielli “*protein sandwich*” model was with the proteins. Under this model, the membrane proteins were placed in association with phospholipid head groups and with the aqueous cytosol. However, it was known that membrane proteins were not generally soluble in aqueous solution; they were amphipathic and exhibited structural regions that were hydrophilic, and others which were hydrophobic. If membrane proteins were layered on top of the membrane, the hydrophobic regions would have to be in the aqueous cytosol, or in contact with the polar phospholipid head groups. In addition, the hydrophobic proteins would be separating the phospholipid head groups from the aqueous cytosolic environment.

In 1972 Singer and Nicolson introduced a new membrane model [12]. According to the new model the membrane proteins were integrated into the membrane bilayer (Figure 4), not layered on top as previously described by Davson and Danielli (see Figure 3).

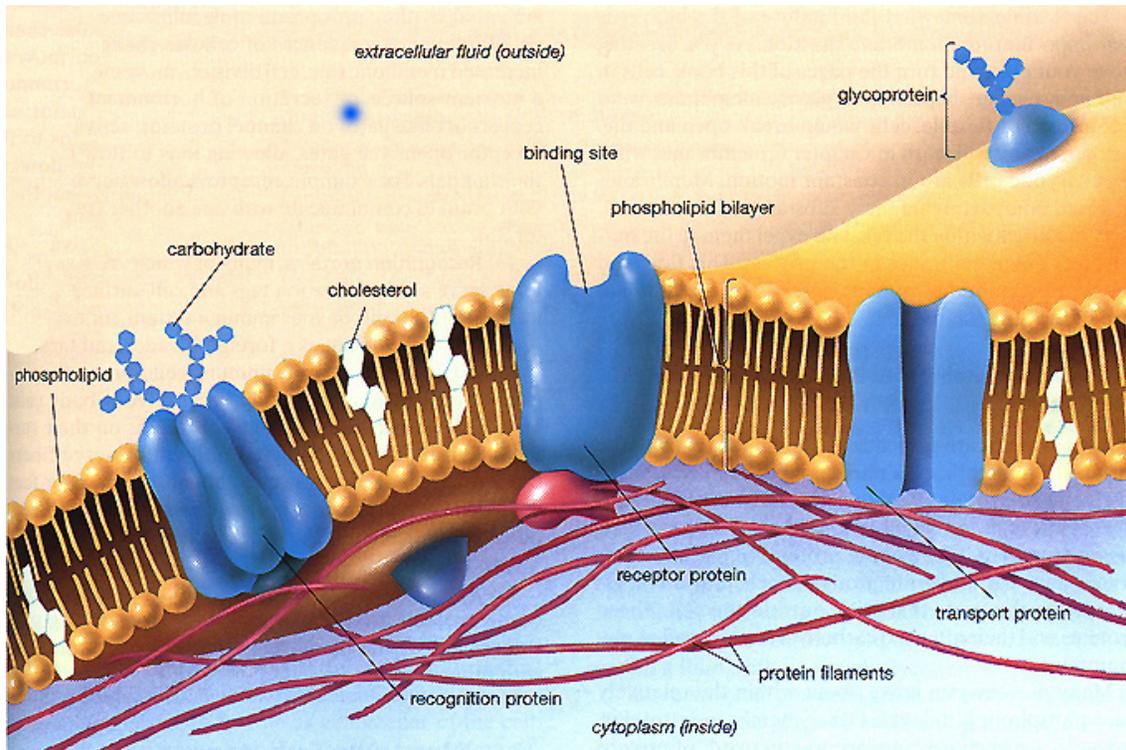


Figure 4: The Singer-Nicolson fluid mosaic model of the cell membrane.

The modern representation of the cell membrane is described by the fluid-mosaic model. This model is composed of the lipid bilayer, which is embedded with integral (for example the transport protein) membrane proteins. Peripheral (not shown) membrane proteins may be associated with the lipid bilayer, or with integral membrane proteins. Reprinted from [13] with permission.

In the new model, the phospholipid bilayer remained much the same as in the Gorter and Grendel, and Davson-Danielli models. As such, the hydrophobic regions of the integral membrane proteins would be in contact with the hydrophobic lipidic centre. Only the hydrophilic regions of the membrane proteins would protrude beyond the lipid tails, and

would then be in contact with the polar phosphate head groups, and the aqueous cytosol. The membrane proteins would appear to dot the membrane in mosaic-like fashion. In addition, the membranes would not be rigid, but would allow lateral movement of the lipids and the integral membrane proteins; hence the term “fluid mosaic”.

Not all membrane proteins need to cross the membrane. Some integral membrane proteins inhabit only one leaflet of the bilayer, for example the cyclooxygenase proteins [14]. In addition, not all membrane proteins are inserted into the bilayer; many proteins exist as modeled by Davson and Danielli; they are globular proteins which are associated with the surface of the membrane bilayer. Due to the increased knowledge on the complexity of the membrane system, membrane proteins are classified by their type of interaction with the membrane.

1.1.3 Membrane Proteins

Approximately 30% of all proteins coded in the genomes of all organisms, both prokaryotic and eukaryotic, contain a hydrophobic amino-acid sequence which suggests that they are membrane proteins [15]. Membrane proteins can be primarily differentiated into two groups based on the protein interactions with the membrane. *Integral membrane proteins* contain at least one transmembrane segment which completely traverses the

membrane, while *membrane associated proteins* interact with the membrane, but do not contain any transmembrane segments and therefore, do not cross the bilayer.

1.1.3.1 Membrane-Associated Proteins

Proteins that are only weakly associated with the membrane are referred to as membrane-associated proteins. One type of membrane associated protein is the peripheral membrane protein. This group of proteins is often easily removed from the membrane, as there are no parts of the protein embedded into the membrane. The proteins remain in contact with the membrane through polar interactions with the phospholipid head groups, or with other proteins inserted in the membrane. Cytochrome c, a small heme-protein, is an example of a peripheral membrane protein [16, 17].

In addition to the peripheral membrane proteins, anchored membrane proteins are considered to be membrane-associated proteins. These proteins are tethered to the membrane *via* fatty acyl, isoprenoid, or glycosylphosphatidylinositol (GPI) lipids [18]. No part of the protein is directly inserted into the membrane; however, the protein is covalently bonded to a molecule that resides, at least partially, in the interior of the membrane phospholipid bilayer. Examples of this group of proteins are the signal transducing guanine nucleotide binding proteins, or “*G-proteins*”. The “*G-proteins*” are divided into large (a heterotrimeric arrangement of $G\alpha$, $G\beta$, and $G\gamma$) and small

(monomeric homologues of $G\alpha$) G-proteins [18]. The protein is tethered to the bilayer through prenylation of the γ -subunit C-terminus, or through palmitoylation or sometimes myristoylation of the α -subunit N-terminus [18].

1.1.3.2 Integral Membrane Proteins

Membrane proteins that have at least some part embedded into the membrane bilayer are considered to be integral membrane proteins. Monotopic integral membrane proteins reside in only one of the two membrane leaflets; they do not span the membrane. Even though monotopic membrane proteins are rare, some very biologically important membrane proteins are monotopic. For example, cyclooxygenase (COX) enzymes COX-1 and COX-2 are monotopic membrane proteins [14, 19-21]. COX-1 is a constitutive form of the enzyme, and helps maintain normal gastric mucosa and kidney function, while COX-2 is induced by inflammation. The proteins pair up into heterodimers which serve as the biologically active form of the enzyme. While both proteins have an equal affinity for their substrate arachidonic acid, from which prostaglandins are formed, only one of the two isoforms is able to process the arachidonic acid at any one time.

Prostaglandins are the lipidic mediators of pain, inflammation and heart attacks. The COX enzymes are the targets of non-steroidal anti-inflammatory drugs (NSAIDs) such as aspirin, ibuprofen and acetaminophen, which inhibit both COX-1 (undesirable) and COX-2 (desirable). More recently, COX-2-specific inhibitors have been developed such as Vioxx™ (rofecoxib), Bextra™ (valdecoxib) and Celebrex™ (celecoxib). While some

of these are considered safe, others (rofecoxib) are implicated in elevated risk of heart and kidney failure (see [22] for review).

Bitopic membrane proteins are single spanning membrane proteins, and often act as signal and antigen receptors. The topology of integral membrane proteins depends on topogenic sequences such as signal, stop-transfer, and signal-anchor sequences [23, 24]. Bitopic membrane proteins are sub-typed based on the orientation of the transmembrane segment, and whether the signal sequence is cleaved or retained after insertion into the bilayer [25]. Type I proteins have a cleavable N-terminal signal sequence which initiates membrane insertion, followed by a stop-transfer sequence which anchors the protein into the membrane with a cytoplasmic C-terminus. Type II and Type III proteins are inserted into the membrane *via* a non-cleavable N-terminal signal-sequence, the signal-anchor sequence, which is responsible for both insertion and anchoring. In Type II proteins, the signal-anchor sequence transfers the C-terminus across the membrane, leaving the N-terminus cytoplasmic. Type III proteins are inserted into the membrane leaving the C-terminus cytoplasmic. One important feature that determines the topology of membrane proteins is the amount and type of charged residues flanking the transmembrane segment. Positive charges, usually from Arg and Lys are most often found cytoplasmic, which has led to the positive-inside rule [26, 27]. Bitopic membrane proteins can be imagined as two polar domains connected by one long non-polar domain. The positive-inside rule states that for both prokaryotic and eukaryotic organisms, the majority of positive charge will remain cytoplasmic, suggesting that the topological selection is encoded through charged residues [24].

Membrane proteins that span the membrane several times are termed polytopic. Ion channels, transporters and facilitators (such as GlpF) are most often polytopic membrane proteins. For polytopic membrane proteins the first hydrophobic segment (which may be a cleavable signal sequence or the first transmembrane segment) usually dictates its own and subsequent transmembrane segments' orientation. However, the process is complicated. The positive-inside rule, length of the transmembrane segment or loop connecting sequential transmembrane segments and interactions between transmembrane segments all influence the final topology of the protein [28, 29].

1.2 Protein Structure and Folding

The ability of a protein to fold into its native three dimensional structure, even with nearly infinite structural possibilities, is still poorly understood despite much research. At nearly the same time as Singer and Nicholson released their fluid-mosaic membrane model [12], Anfinsen realized that the full three dimensional structure of a protein is determined purely by its amino-acid sequence [30]. Furthermore, the structure of a protein defines its function, and by extension, the ability of a protein to function relies on its capability to fold rapidly and reliably to its native state [30]. The ultimate goal in protein folding research is to be able to predict the three-dimensional fold of the protein based solely on the amino-acid sequence (see Figure 5).

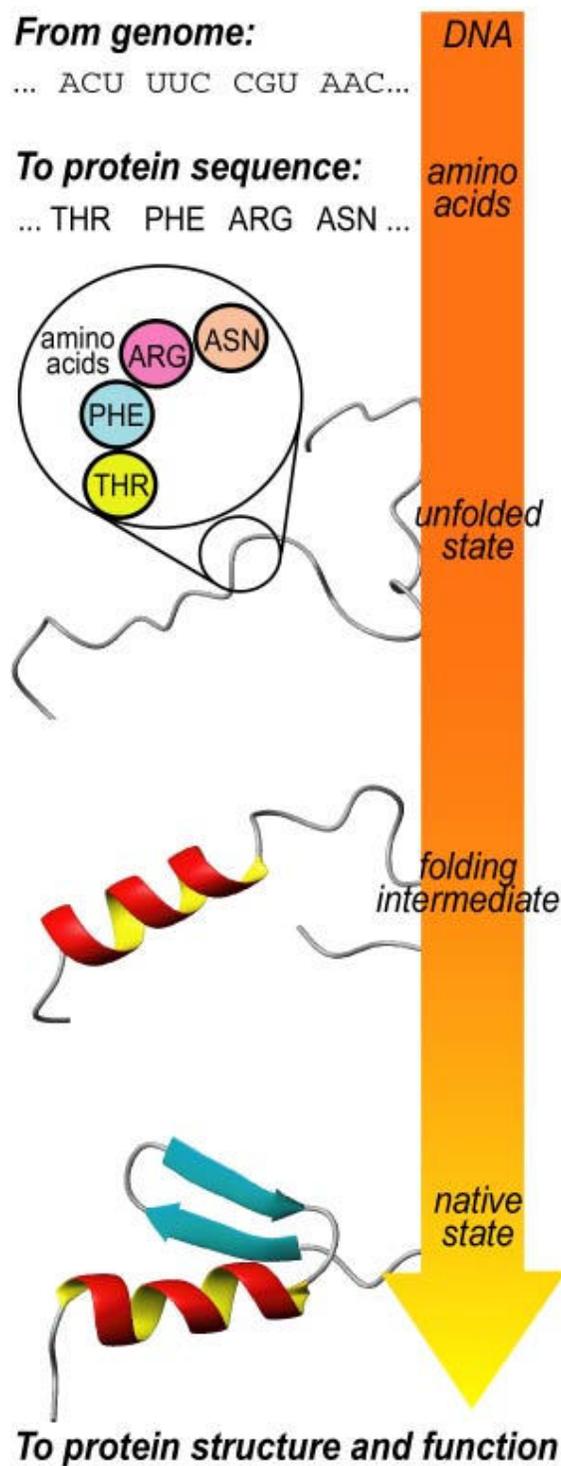


Figure 5: Protein folding: prediction from genome to function.

Reprinted from [31] with permission (Stanford©).

Many proteins are capable of folding into their native state spontaneously in an aqueous environment. In order for this to be achieved the protein must satisfy both a thermodynamic and a kinetic requirement. The thermodynamic requirement is that the protein must reliably fold into a single, stable conformation. The kinetic requirement is that the protein must fold into this state on an appropriate timescale. The duration of protein folding varies greatly depending on the size and complexity of the protein. Slower folding proteins may require many minutes or hours to fold, however many proteins fold on the millisecond timescale [32]. In fact, the very fastest folding proteins fold in hundreds of nanoseconds, the fastest known presently being the villin headpiece [33-37]. In 1969 Levinthal realised that if proteins sampled all available conformational space sequentially during the folding process, it would take an astronomical amount of time to complete the folding, even if all conformational space was sampled on the nanosecond or picosecond timescales [38]. More recent studies on protein folding have shown that in order for a 100 amino-acid peptide to sample every available conformation in a completely random fashion would take on the order of 10^{36} sec, or 10^{29} years [39]. Given such an unmanageable timescale, a mechanism must exist for proteins to fold in a timely fashion. This means that only a few of the available conformations are sampled therefore minimizing the time spent traversing the conformational space, which implies the existence of kinetic folding pathways as first suggested in 1968 by Levinthal [38]. The difficulty in understanding how proteins are able to rapidly and reliably fold into their native three dimensional structures is commonly referred to as “*the protein folding problem*”. Interestingly, while the proteins themselves fold on a rapid timescale, it takes remarkably long for computers to simulate the folding process. This has led to the

“*folding@home program*” [31], which connects personal computers from around the world to process folding simulations.

1.2.1 Protein Folding Models

Models put forth to explain the phenomenon of a linear random coil amino-acid chain folding into its native tertiary structure must satisfy two important features of the folding process, at least for small, single domain proteins. First, the model must predict a two-state folding process, and secondly the model must predict cooperativity in the folding [40]. These conditions are not mutually exclusive of one another, as the condition of cooperativity leads to an observation of two states: folded, and unfolded. The two-state folding condition is explained simply as an equilibrium condition for which the only available states are: a lower energy folded state, and a higher energy unfolded, or random coil, state separated by a single energy barrier (Figure 6) [40].

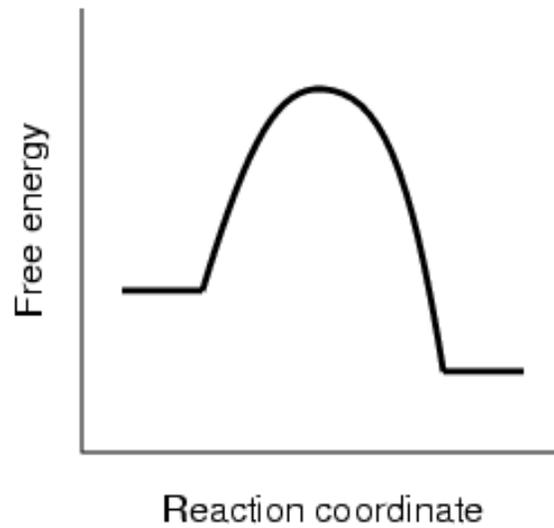


Figure 6: Reaction coordinates for a two-state folding process.

The low energy folded state, and higher energy unfolded state are separated by a single energy barrier.

The second feature of protein folding models, that the folding must be a co-operative process, is shown diagrammatically in Figure 7.

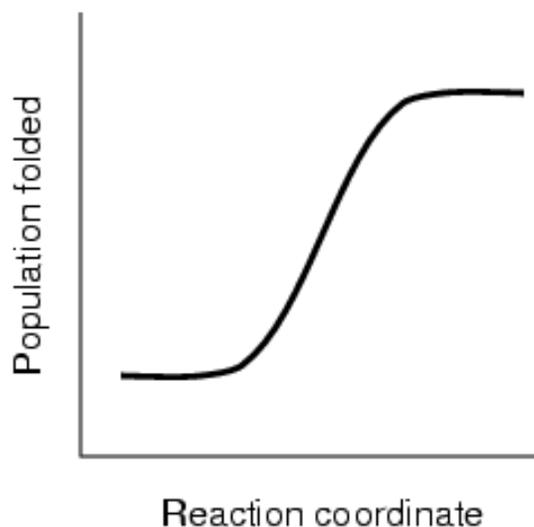


Figure 7: The sigmoidal transition curve characteristic of a two-state equilibrium transition.

The cooperativity of protein folding is exemplified by the sigmoidal curve of the population of protein molecules in the folded state compared to the reaction coordinated. Proteins fold in a cooperative fashion, such that the folding of one part of the molecule influences the folding of the next, making it “*easier*” for that protein molecule to reach the folded state.

Proteins are held together almost exclusively through weak interactions; hydrogen bonds, van der Waals interactions and electrostatic forces make up the bulk of the forces holding the protein structure together. Each of the individual interactions contribute only a small favourable energy to the structure, however the large number of interactions add together to stabilize the three-dimensional structure.

One of the first models to describe co-operative protein folding is the helix-coil theory [41, 42]. This model states that all amino-acids in a peptide chain have a given helix propensity and that each amino-acid exists in one of two states: helical or random coil. However, the probability of an individual amino-acid becoming helical is also dependent on the neighbouring amino-acids; if an amino-acid obtains a helical conformation, the

neighbouring amino-acids now have an increased likelihood of also becoming helical.

The *helix-coil model* can be viewed as a “zipper-like” process, where a single nucleation from coil to helix is introduced into the polypeptide chain, and proliferation of the helix continues as described above [41]. In this model, only conformations such as

...ccchhhhccc... or ...hhhhccccc... are acceptable, while ...hhhhccccchhhhccccc...

is not an acceptable conformation as multiple nucleation points would be required¹.

However, the *helix-coil model* may also be described as a matrix, where nucleation points are free to occur anywhere in the polypeptide chain; the polypeptide is no longer restricted to single helical segments as in the zipper model [41]. In either of these models, helix propagation will continue rapidly, and will become co-operative after only a few initial helix conformational transitions [41, 42].

Many models of protein folding have been proposed since the inception of the *helix-coil model*. The *frame-work model* involves an organized assembly where local secondary structure forms dependent on the primary structure, but independent of the tertiary structure [43, 44]. The secondary structural elements will diffuse until they are close enough together to form the overall tertiary structure. In this manner the protein folding process greatly limits the conformational space required for rapid and reliable folding.

The *nucleation model* proposes that tertiary structure immediately forms as a consequence of the formation of intact secondary structural elements [45, 46]. The nucleation events are a result of small sections of intact secondary structure, which then spread tertiary structure as a direct consequence of their formation, *i.e.* tertiary structure is forming “just behind” the secondary structure. The *hydrophobic collapse model*

¹ h = helix, c = coil

suggests that the intact tertiary structure forms through the rearrangement of an already compact structure, or *molten-globule* [47, 48]. The hydrophobic collapse resulting in the formation of the compact structure or molten-globule must then be an early event along the reaction coordinate of the protein folding pathway. Of these models, both the *framework* and *hydrophobic collapse* models imply that there must be formation of kinetic intermediates, while the nucleation model does not require such an intermediate [49]. However, a modified nucleation model, the nucleation-condensation model, first introduced by Fersht, allows for a loose folding nucleus to be formed, and subsequently the nucleus will condense through a transition-state associated with the tertiary structure formation [49-52]. The *nucleation-condensation model* is supported for several small proteins including the chymotrypsin inhibitor-II [50, 51] and the barstar protein [53]. Interestingly, studies of more than 20 different proteins show that all proteins in the study adopt a molten-globule structure when treated with mild denaturants [54, 55]. The observation of a molten-globule in each case supports the hydrophobic collapse model, and is indeed a model favoured by many. It is probable however that more than one possible folding model is applicable dependent on the protein's properties.

More recently, a more general approach to the protein folding pathways has been described by the so-called protein folding funnel [39, 56-58]. This approach models the energy surface of a protein folding pathway as a funnel (Figure 8).

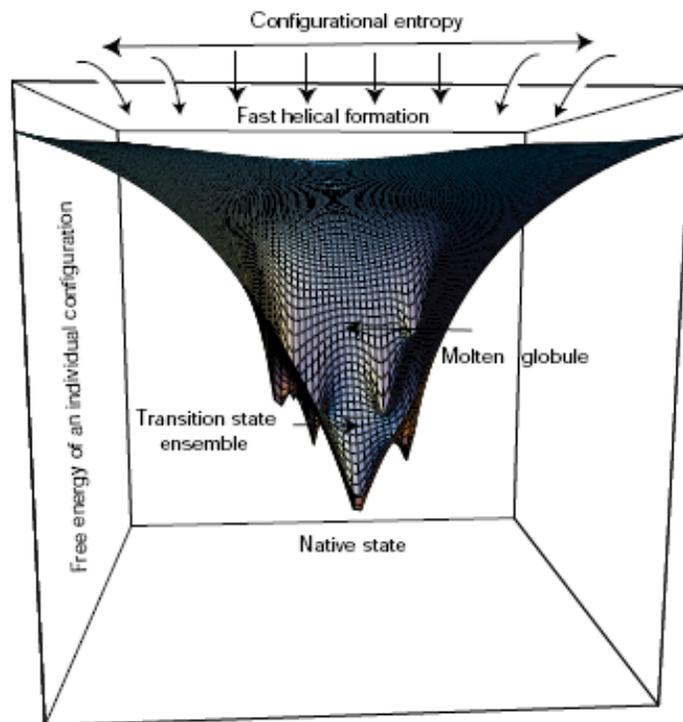


Figure 8: Representation of the protein folding funnel.

Reprinted from [59] with permission.

The rim of the funnel consists of the many unfolded conformations of the protein, representing the high entropy of the unfolded state, while the bottom of the funnel represents the global free energy minima, indicative of the native folded tertiary structure. This allows for the protein folding to be accomplished *via* many different routes; the folding may take the most direct route as would occur by following the steepest points of the funnel. However, the protein may also take a more leisurely folding pathway that involves a shallower descent, and more transition states indicated by the local maxima along the protein folding funnel. Figure 8 shows a theoretical funnel with multiple folding pathways each demonstrating different possible pathways for the folding of a protein [59].

1.2.2 Membrane Protein Folding

The ultimate goal in understanding membrane protein folding is the same as that of their soluble counterparts: to be able to predict the three-dimensional fold of the protein based solely on the amino-acid sequence. One main distinction of membrane proteins is that they must be transferred into the oily hydrocarbon region of the bilayer. The thermodynamic cost of transporting polar or charged residues into the membrane is very high, and as such the majority of amino-acid side chains of transmembrane segments are non-polar; a much higher percentage of Ala, Ile, Leu, Phe and Val are found in the transmembrane segments of membrane proteins than in the remainder of the protein. The second consequence of the high thermodynamic cost of polar groups residing in the lipid bilayer is that the highly polar peptide bond must maximize its hydrogen-bonding in the transmembrane segments so as to lower the energy cost. Two main structural motifs satisfy this condition for maximizing the hydrogen-bonding potential: the α -helix, and the β -sheet, with the latter requiring that the β -strands be in close contact as is found in the β -barrel. Examples of proteins containing these structural motifs, the α -helical bacteriorhodopsin (2BRD) [57, 60], and the β -barrel porin (2POR) [58, 60], are shown in Figure 9.

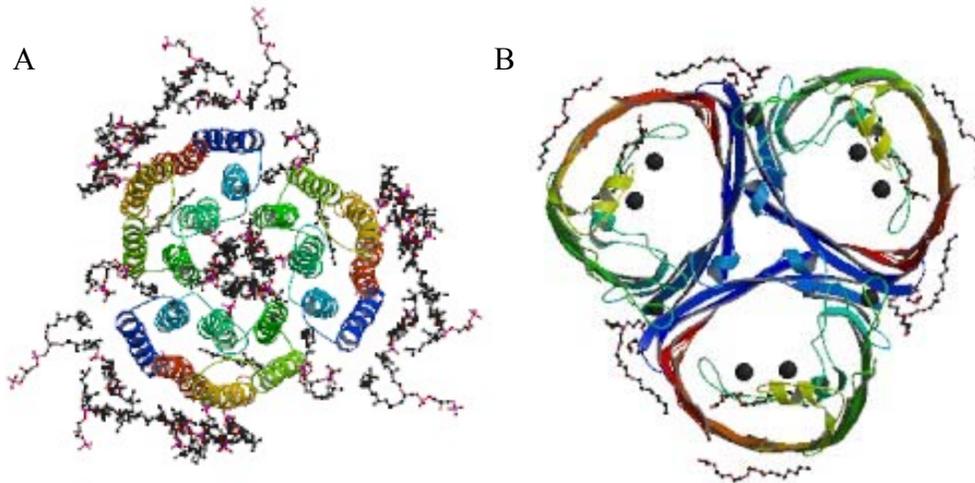


Figure 9: The two main structural motifs of membrane proteins are the α -helix and the β -barrel. Bacteriorhodopsin (A) [57, 60] (reproduced from [60]) exemplifies the helical membrane proteins, while the β -barrel proteins are represented by porin (B) (reproduced from [60] with permission)[58, 60].

These simple thermodynamic principles are sufficient for understanding the basic structural parameters of membrane proteins, however in order to predict the detailed three-dimensional structure one must have a detailed understanding of the thermodynamic and kinetic properties of membrane protein folding.

1.2.3 Membrane Protein Insertion

In order for integral membrane proteins to function, they must first be inserted into the membrane. The mechanism by which membrane proteins are inserted into the membrane is similar to the mechanism used by the cell to transport secretory proteins across the

membrane; in both cases, translocation proceeds through a protein complex which supplies an aqueous channel through the membrane [56, 61]. This protein-conducting channel, or PCC, is made up of a hetero-trimeric arrangement of integral membrane proteins [62]. In eukaryotes, the Sec61 complex makes up the PCC (also referred to as the translocon); in mammals this is comprised of a hetero-trimeric arrangement of Sec61 α , Sec61 β and Sec61 γ , whereas in yeast the Sec61p, Sbh1p and Sss1p comprise the corresponding homologues [63]. The bacterial homologues SecY, SecG and SecE have also been identified [63].

The recent crystal structure of the *Methanococcus jannaschii* translocon has facilitated the understanding of membrane insertion *via* the PCC [64]. The crystal structure shows that the PCC contains only one copy of the SecY complex, and suggests that the single hetero-trimer is sufficient for complete functionality [63, 64]. The structure shows that the SecY complex pore is formed exclusively by the SecY subunit, while SecG and SecE proteins are peripheral to the pore [64]. This structural data is supported by cross-linking studies demonstrating that the movement of the polypeptide is through the centre of the SecY subunit [65]. Subunits SecG and SecE, and the N- and C-terminal domains of SecY are arranged in such a manner, as to allow one side of the PCC to open laterally. This ability to open laterally allows the PCC a mechanism to release the transmembrane segments of the growing polypeptide into the membrane bilayer [63, 64]. In order to allow partitioning of each transmembrane segment to the bilayer, the PCC must open and close many times; recent structural studies have indicated that the PCC may function in a “*breathing-like*” fashion where the open-close cycle is continuous [63].

While the crystal structure addressed many aspects of the PCC, further structural studies have added to the debate on the PCC function. Recently, the structure of the PCC bound to a translating ribosome was solved using cryogenic-electron microscopy [66]. In this structural arrangement, two PCC complexes are involved with their lateral openings facing one another [66]. This new structural information adds many new questions, which will require further investigation into the PCC structure and function.

1.2.4 Membrane Protein Recognition

In order for the cell to target an integral membrane protein to the lipid bilayer, the cell must first recognize that the polypeptide is indeed an integral membrane protein.

Translocation to the membrane may occur in a co-translational or post-translational approach [67]. The post-translational method involves the complete synthesis of the polypeptide and subsequent release from the ribosome prior to membrane insertion, while in the co-translational method the protein is inserted into the membrane concurrent with translation by the ribosome attached to the PCC complex.

Proteins are targeted to the PCC by signal sequences involving stretches of hydrophobic amino-acids [68, 69]. Two key components to the signalling are the signal recognition particle (SRP), and the SRP receptor [70, 71]. The SRP is a cytosolic protein-RNA

complex which binds to the growing hydrophobic polypeptide as it is being translated by the ribosome [72]. The binding of the SRP causes a pause in translation, and targeting of the ribosome to the membrane. The SRP-bound complex is then bound by the SRP receptor which juts from the membrane and serves as a landing pad for the SRP-ribosome conglomeration [73, 74]. In a GTP-dependent fashion, the SRP is released from the polypeptide chain and the ribosome is transferred to the PCC, allowing translation to continue and, for the case of an integral membrane protein, the newly synthesized polypeptide to be translocated into the bilayer [71].

1.2.5 Exogenous Membrane Protein Assembly

As previously mentioned, the translocation of endogenous integral membrane proteins (those required by the cell for regular function) involves a very complicated arrangement of ribosomes transiently attached to the translocon [75, 76] (Figure 10).

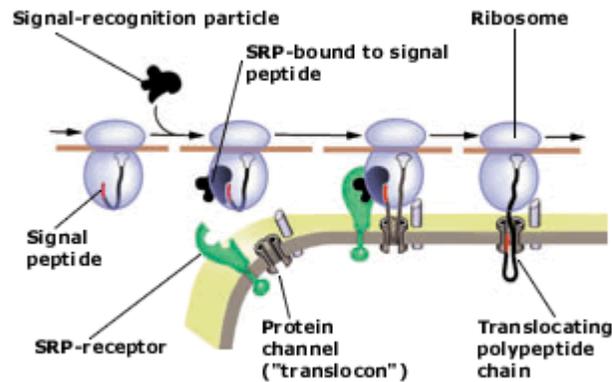


Figure 10: The assembly of endogenous integral membrane proteins involves a temporary arrangement of the ribosome (with mRNA) and the translocon.

The protein is then translated into the cavity of the translocon. Upon completion, the ribosome is removed from the translocon and the membrane protein is released into the membrane bilayer, where it assumes its final three-dimensional structure [75, 76]. Reproduced from [77] with permission.

Despite all of the above information, the mechanism for the folding of endogenous integral membrane proteins after release from the translocon is as of yet, still unknown. Assembly of exogenous membrane proteins (proteinaceous toxins, antimicrobial peptides, *etc.*) differs from that of the constitutive membrane proteins, as they must enter the target cell membrane without the aid of the translocon. The assembly and folding of exogenous membrane proteins in the membrane bilayer is likely guided by the same processes as constitutive membrane proteins, and therefore may serve as a model for all membrane protein folding. Examples of pore-forming toxins have been reviewed by Gouaux [78], while the structure and function of antimicrobial peptides has been reviewed by White *et al.* [79].

1.2.6 The Energetics of Proteins in the Membrane

Exogenous membrane proteins often use simple physiochemical processes for insertion into the membrane, and as such have been used to gain insight into membrane protein folding and stability [79]. The mechanism by which a membrane protein finds its way into a membrane, whether endogenous or exogenous, is irrelevant to the energetics once it is located in the lipid bilayer [80]; there the protein sits in a free energy minimum within the lipid bilayer of the membrane. The sorting out of protein-protein and protein-bilayer energetics that determine the free energy minimum can be done in principle by unfolding and re-folding the protein within the bilayer. However, this is difficult if not impossible to do with whole membrane proteins; they are insoluble in the bilayer in the unfolded state (due to the energetic cost of exposing the peptide bonds to the bilayer hydrocarbon interior), and they are insoluble in the aqueous phase in their folded and unfolded states (due to their hydrophobic nature) [81].

Modelling the energetics of membrane protein folding *via* exogenous peptides allows the separation of the folding into two main categories: first, secondary structure formation and insertion into the membrane bilayer, and secondly, association of the secondary structural components. The entire folding process has been conceptually divided into the four main steps as shown in Figure 11 [80, 82-84].

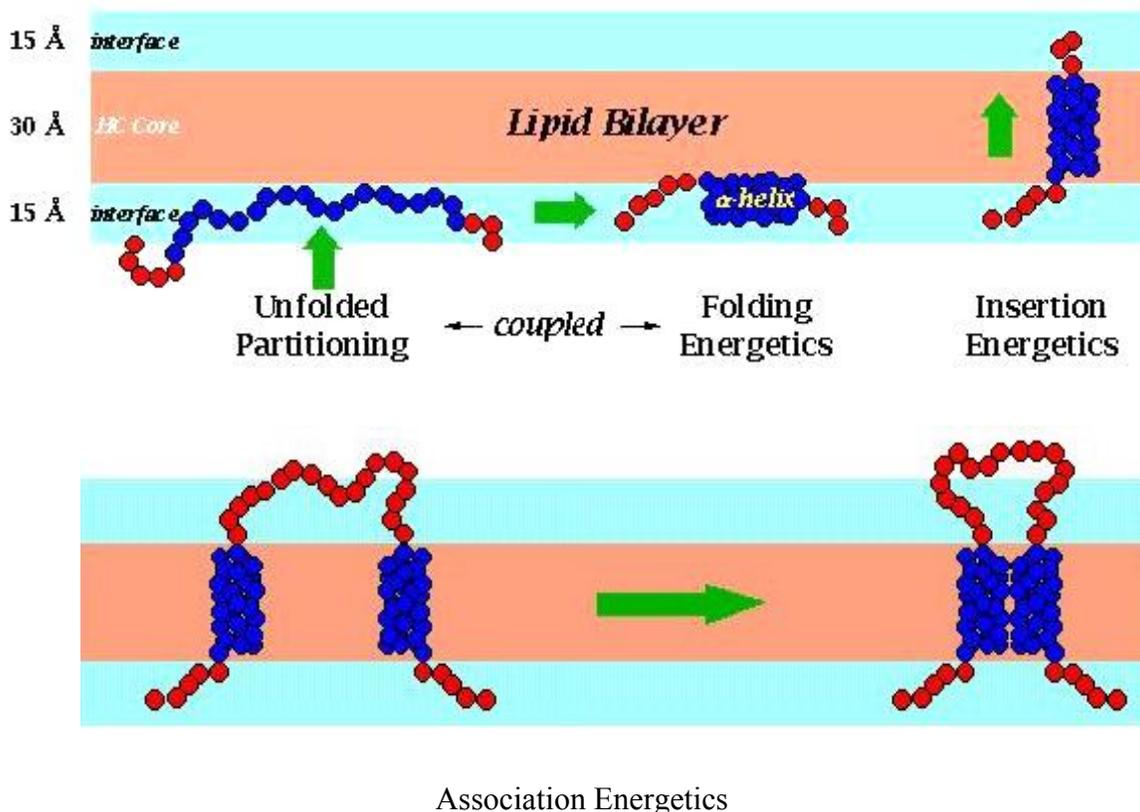


Figure 11: Helix folding, insertion and association: the four conceptual thermodynamic steps.

The reference, or starting state, is taken as an unfolded protein at the interface [85]. Secondary structure formation occurs during the partitioning-coupling phase, where maximizing hydrogen bonding of the peptide bonds minimizes the energetic cost of residing at the interface [86, 87]. Following secondary structural folding is membrane insertion [88-90]. It is likely the least adequately studied, but may be the most important. Finally, helix-helix association occurs within the membrane, which is likely independent of the intra-membrane assembly [83]. Reprinted with modifications and permission from [91].

First, the unfolded peptide must partition from the bulk solution to the bilayer interface, secondly, the unfolded peptide must assume a secondary structure at the interface, thirdly, the secondary structural element must insert into the membrane bilayer and lastly, the secondary structural elements must come together within the bilayer.

The first three steps of the folding model described above are likely impossible to study using whole proteins, and therefore White and Wimley have studied the interactions between the bilayer and small peptides as a model for the thermodynamic processes of an entire membrane protein [82]. This allows membrane proteins to be viewed as assemblies of individual secondary structural components. In order to begin the thermodynamic studies by this model, one must have a completely unfolded state in solution, which will then partition to the bilayer. As this situation likely does not exist in the cell, a virtual reference state has been defined using an experimental interfacial hydrophobicity scale developed from studies on pentapeptides [85]. It is found that aromatic residues are particularly favoured at the interface, while charged residues and the peptide bond are equally disfavoured [85]. This hydrophobicity scale provides a means to calculate the virtual free-energy cost of transferring an unfolded chain from solution to the bilayer interface [85].

Secondary structure formation at the bilayer interface seems to be determined primarily by the reduction in the free-energy of partitioning of the peptide bond. It is reported that the formation of the hydrogen-bonds in β -sheet secondary structural elements is responsible for a 0.5 kcal/mol reduction per peptide bond [87], while hydrogen-bond formation in α -helix secondary structural elements reduces the free energy of partitioning by 0.4 kcal/mol per peptide bond [86]. This may appear modest, but the cumulative effect can be rather large, as evident for the soluble peptide melittin where a total reduction of approximately 5 kcal/mol is observed [86].

The third step in the proposed folding model is of great importance, but is the least understood. This most likely stems from the hydrophobic nature of the materials needed to study the insertion of the peptide into the membrane. Estimates of the free-energy cost of inserting a hydrogen-bonded peptide bond into the bilayer vary from 0 kcal/mol to 1.6 kcal/mol [88-90]. This large variation implies that calculations of inserting peptides as short as 20 amino-acids may be over- or underestimated by as much as 30 kcal/mol. Clearly more research into the insertion of peptides into the bilayer is needed for a clearer understanding of the energetics of insertion.

The last stage of the folding model above is the association of the inserted secondary structural elements. The insertion-association of α -helices has been termed the “*two-stage*” folding model by Popot and Engelman (see Figure 11) [83]. Stage one involves insertion of secondary structural elements, (*i.e.* steps 1-3 as mentioned above) [83]. In stage two, the secondary structural elements, which are in thermodynamic equilibrium with their lipidic and aqueous environments, must pack together to form functional transmembrane structures [83]. In order for the final structure to be obtained, some rearrangements of the helices are to be expected, however, others such as helix flipping, or insertion of new segments, are kinetically forbidden [83]. This is similar to the case of soluble globular proteins, where interplay of kinetic and thermodynamic forces is expected during the packing of preformed secondary structural elements [92]. Experimentally, bacteriorhodopsin (BR) has been shown to refold starting from the completely unfolded polypeptide [93], and from two denatured proteolytic fragments [94]. These fragments have been refolded in lipid bilayers [95] to produce two-

dimensional arrays, which have been studied crystallographically and have been shown to be indistinguishable from native BR [96]. These outcomes are expected when one considers the two-stage model of protein folding. Here, each transmembrane helix behaves as an independent folding domain, where the thermodynamic forces such as helix-helix interactions, polar forces, and packing effects control the final structural arrangement [83].

Although only six helical MP have been refolded from a denatured state [97] a common requirement for successful re-folding is dissolution of a folding-competent protein in an ionic detergent such as SDS or N-lauroylsarcosine. Interestingly, a recent study of the interactions of SDS with membrane proteins [98] suggests that in the presence of SDS, membrane proteins are not fully unfolded but that the SDS molecules intercalate between the protein helices. The membrane proteins have lost the close tertiary contacts due to the insertion of the SDS molecules, but continue to exist as a loosely packed structure (see Figure 12). Ionic detergent micelles encourage helix formation by providing a hydrophobic environment not unlike that of a membrane [99] and discourage aggregation by virtue of micelle charge repulsion. Little is known about the extent of secondary and tertiary structure formation of MP in detergent micelles or about the nature of intermediates on the folding pathway. However, thermodynamic and structural studies of bacteriorhodopsin [100, 101] and glycophorin A [102, 103] suggest that the final stage of MP compaction is facilitated by helix “*knobs-in-holes*” interactions driven by London dispersion forces. Understanding membrane protein folding *in vitro* and how lipids, bilayers, detergents, co-factors, and protein sequence contribute to protein folding and

stability will yield insights into MP folding *in vivo* and, as suggested by White and Wimley [82], is likely to contribute to the elucidation of the principles and forces governing water-soluble protein folding. Other benefits include improving the recovery of MP from inclusion bodies, and accelerating structure determination by NMR spectroscopy and X-ray diffraction.

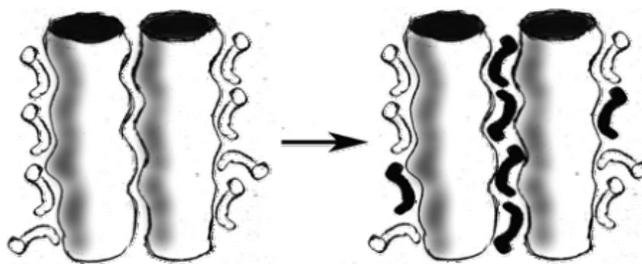


Figure 12: SDS intercalation of membrane proteins.

On the left is shown two interacting transmembrane helices in a detergent which stabilizes the native structure (white). On the right, after addition of SDS (black) some of the SDS intercalates between the helices, thus changing the spectroscopic properties. Reprinted from [98] with permission.

1.3 Major Intrinsic Protein Superfamily

The Major Intrinsic Protein (MIP) superfamily is comprised of all members of aquaporins, glycerol facilitators and aquaglyceroporins (see section 1.3.1) [104]. Named after the first aquaporin discovered, the Major Intrinsic Protein [105] found in the eye lens, the family of proteins all function as membrane channels that selectively transport water, small neutral molecules, or ions out of and between cells, and may play an important role in cellular metabolism. While MIP (since renamed AQP0) gave its name

to this family, the discovery that MIP was indeed a membrane channel was not made until after the discovery of aquaporin1 (AQP1) [106]. The proteins are clustered in the family based on their pore specificities, amino acid sequences and gene organization [104, 107]. Analysis of these proteins using hydropathy plots [108] suggested that the main protein fold consists of six transmembrane helices (1-6) connected by 5 loops (A-E). Characteristic of every member of the MIP superfamily is an asparagine-proline-alanine (NPA) motif² [109] in loops B and E, which form an important aspect of the protein channel [1, 110].

1.3.1 Aquaporins and Aquaglyceroporins

Aquaporins (AQPs) are an important family of proteins that efficiently channel water through cell membranes. Although water can diffuse across biological membranes at measurable rates, in their absence, water barely trickles across the hydrophobic cell membranes. Physiologists had long suspected a mechanism by which water would cross the membrane at a rate much larger than simple diffusion [111, 112]. The first functionally characterized water channel protein, or AQP, was designated AQP1 [113]. With AQP1 present the Arrhenius activation energy for water crossing the bilayer is less than 5 kcal/mol, and as such water can pour through the membrane at the rate of approximately three billion water molecules per AQP1 channel per second [114, 115]. To date thirteen mammalian AQP genes have been identified, and members of all three

² AQPs 11 and 12 have less conserved NPA motifs.

kingdoms of life have been shown to contain at least one AQP gene [107, 116-119], reinforcing the fundamental importance of this family of proteins (see

Table 1).

Name	Transport Properties	
	Water	Glycerol
AQP0	Yes	No
AQP1	Yes	No
AQP2	Yes	No
AQP3	Yes	Yes
AQP4	Yes	No
AQP5	Yes	No
AQP6	Yes	No
AQP7	Yes	Yes
AQP8	Yes	No
AQP9	Yes	Yes
AQP10	Yes	Yes
AQP11 ³	No	No
AQP12 ⁴	Unknown	Unknown
AQPZ	Yes	No
GlpF	Yes	Yes
AQPM	Yes	Yes

Table 1: List of known mammalian, *E. coli*, and archaeobacterial aquaporins, and their transport properties.

To date there are thirteen known mammalian aquaporins (AQP0-12), two known *E. coli* aquaporins (AQPZ and GlpF) and only one known archaeobacterial aquaporin (AQPM). Each is listed above with their transport properties.

³ AQP 11 lacks any apparent function.

⁴ AQP 12 transport properties are as of yet unknown.

1.3.1.1 Structure of the AQP1 Channel

The coding region of the AQP1 cDNA corresponds to a 269 amino acid polypeptide, which is predicted by hydropathy analysis to contain six transmembrane segments (1-6) connected by 5 loops (A-E) [120]. The polypeptide is interestingly arranged into two similar repeats, with the N-terminal half (repeat-1) showing approximately 20% identity to the C-terminal half (repeat-2) [120]. Two loops, B and E, show an even greater similarity, and each contain the AQP signature NPA motif [120]. Water transport across the red cell membrane is known to be inhibited by mercurials, which binds to free sulfhydryl groups of cysteine residues [111]. AQP1 contains four cysteine residues, however only Cys-189 (located in loop E) has been shown to be sensitive to mercurials [121]. Replacement of Cys-189 by Ser abolishes all mercurial sensitivity, however replacing the corresponding residue in loop B (Ala-73 → Cys) re-establishes mercurial sensitivity, whereas substitutions elsewhere in the protein fail to produce this behaviour [121]. In order to account for these observations the AQP1 protein was modeled to have an hourglass-like topology in the membrane bilayer, where the six transmembrane helices surround a central domain formed from the dipping of loop B into the bilayer from the cytoplasmic surface, and the dipping of loop E into the bilayer from the extracellular surface (see Figure 13) [122].

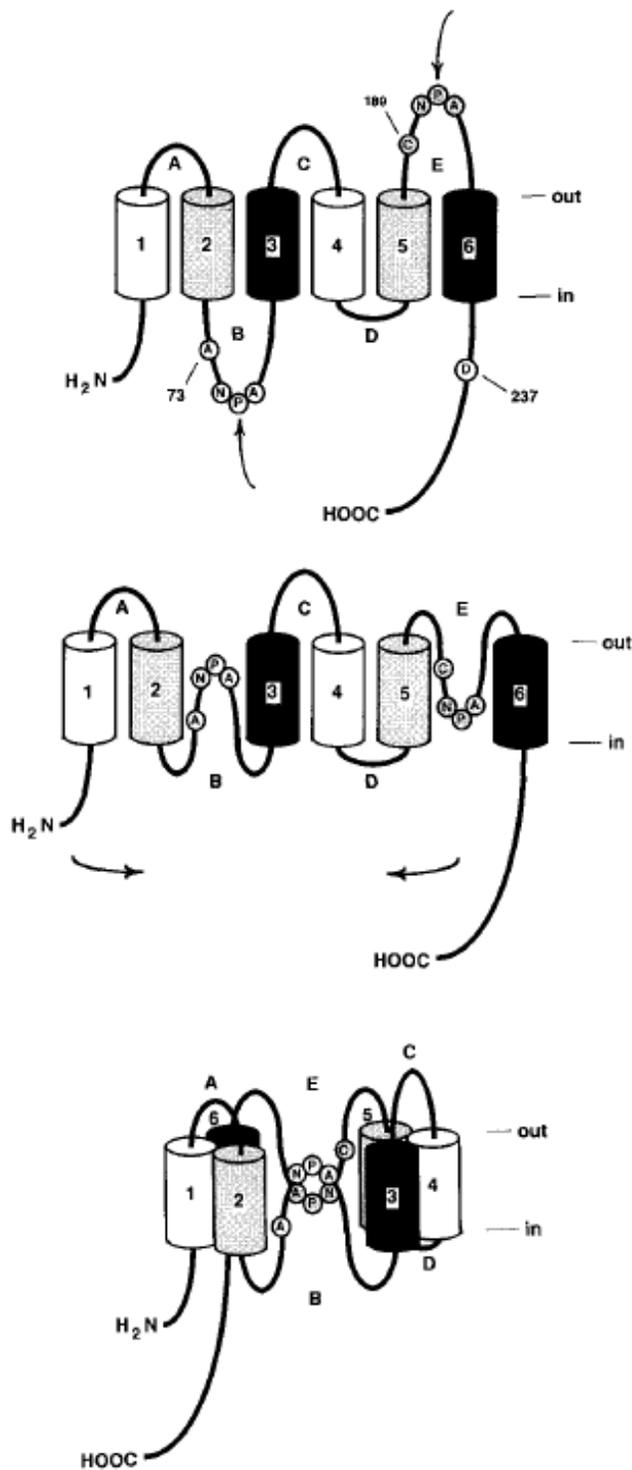


Figure 13: Hourglass representation of the AQP1 protein.

In the top frame, the two NPA motifs are labelled with an arrow. In the middle frame, the NPA motifs are folded into the middle of the bilayer of the membrane. The bottom frame shows the “*collapsing*” of the structure into two sets of three helices surrounding the two NPA motifs. Reprinted from [122] with permission.

The overlapping of loops B and E was predicted to form a central pore through the centre of the protein, with the NPA motifs juxtaposed and flanked by the mercurial inhibitory site [122]. Subsequent high-resolution structural models of AQP1 (cryo-electron microscopy, human AQP1, 3.8 Å [110] and X-ray diffraction (XRD), bovine red cell AQP1, 2.2 Å [123]) showed this hourglass model to be basically correct, with loops B and E forming non-bilayer spanning helices at the centre of the protein. These observations permitted molecular dynamics simulations on AQP1, which subsequently led to the understanding of how an open channel can allow for rapid flow of water molecules across the membrane, but retain complete exclusion to proton conductance [124, 125].

The hourglass structure of AQP1 contains two vestibules, one intracellular the other extracellular, where water is in bulk solution. The vestibules are separated from each other by a span of approximately 20 Å, which is so narrow (2.8 Å at its narrowest constriction) it can only accommodate water molecules in single file [123]. At this narrowest constriction is the side chain of a conserved arginine (R195) following the NPA motif of loop E, which provides a positive charge, and a conserved histidine (H180) side chain on the opposite wall, which provides a partial positive charge [123]. Together, they serve to repel the positively charged protons [124]. Further down the channel, a second mechanism exists, which prevents the leakage of protons. As the water molecules reach the juxtaposed NPA motifs they form hydrogen bonds with the side chains of the two asparagine residues of the NPA motifs (N192 and N76) [124]. This, in combination

with the two partial positive charges at the centre of the channel, which result from the two non-spanning helices of loops B and E, forces a transient reorientation of the water molecule's dipole, which prevents proton conduction *via* the Grotthus effect [124]. This example demonstrates how biology and chemistry are intimately intertwined, and how biology uses chemistry to perform fundamental cellular functions.

1.3.2 Structure of the GlpF Channel

Many AQPs transport water and glycerol (as well as other small non-ionic solutes) across biological membranes, and are termed the aquaglyceroporins (GLPs). Interestingly *E. coli* contains one member of each of the classical AQPs (AQPZ [126]) and the GLPs (GlpF [127]). The *E. coli* GlpF is encoded by the first gene in the *glpFKX* operon, which also encodes glycerol kinase (GlpK) and a protein of unknown function (GlpX) [127]. The *glpF* gene encodes a 281 amino acid polypeptide, which is predicted by hydrophathy analysis to contain six transmembrane segments (1-6) connected by 5 loops (A-E), similar to AQP1. The GlpF polypeptide is also arranged into two similar repeats and contains the two NPA motifs (residues 68-70 and 203-205).

1.3.2.1 The Structure of GlpF

Fu *et al.* [1] resolved the crystal structure of GlpF to 2.2 Å. Structural refinement to this level permits the observation of individual side-chain electron densities, and thus allows for the fitting of the amino-acid sequence into the electron density map to form an unambiguous structural model of the GlpF protein (see Figure 14).

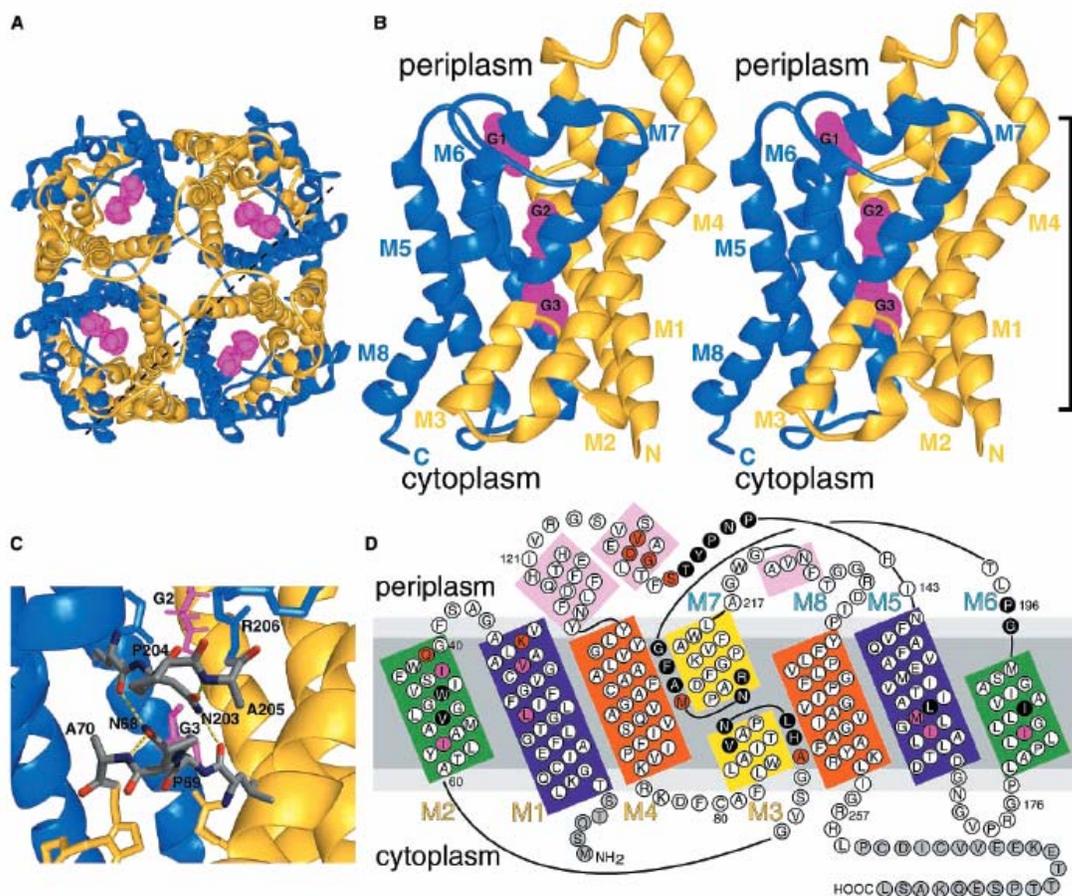


Figure 14: The GlpF tetramer.

A) The crystal structure of the GlpF tetramer. B) Stereoview of one GlpF monomer. C) Close-up showing the interactions of the two NPA motifs. D) Schematic diagram of the folding of the GlpF protein. Reprinted from [1] with permission.

The GlpF protein crystallizes as a symmetric tetramer of channels (see Figure 14A), with the plane of the bilayer perpendicular to the four-fold axis (see Figure 14B). The structure is very similar to that of the AQP1 [123]. GlpF has six transmembrane spanning helices (M1, M2, M4, M5, M6 and M8) and two half spanning helices (M3 and M7) which fold into a right handed helix bundle around each channel. Based on the crystal structure, 189 of the 281⁵ residues (67%) are helical. Inter-helix packing angles within the monomers range from approximately +35° to +40°, with intermonomer helix angles of approximately -20° both between M1 and M2 of one monomer with M5 and M8 of its neighbour [1]. The internal sequence homology of GlpF is shown in Figure 14 (yellow and blue segments). Similar to the other AQPs, the N-terminal segment originates on the cytoplasmic side of the membrane. Helix M1 crosses the membrane and is followed by loop A, which connects it to helix M2. Helix M2 spans the membrane and connects to loop B. Loop B reinserts into the centre of the pore from the cytoplasmic side and forms the half-spanning helix M3, which originates with the conserved NPA motif at residues 68-70. Loop B returns to the cytoplasmic side of the membrane and connects to helix M4, which crosses to the periplasmic side of the membrane. The C-terminal segment is arranged much as the N-terminal segment and is connected to it by loop C beginning with helix M5 spanning the membrane to the cytoplasmic side, where it is connected to helix M6 via loop D. Helix M6 is followed by loop E which reinserts into the centre of the pore from the periplasmic side of the membrane and forms the half-spanning helix M7, which originates with the conserved NPA motif at residues 203-205.

⁵ Fu *et al.* analyzed a 254 residue (6-259) GlpF polypeptide with N- and C-terminal segments truncated. The N- and C-terminal segments are assumed to be unstructured.

Loop E returns to the periplasmic side of the membrane and is followed by helix M8 which spans the membrane leaving the C-terminus cytoplasmic (see Figure 14D).

The two half helices M3 and M7 meet with their N-terminal ends near the centre of the pore similar to AQP1, and is largely the selectivity filter of GlpF. The NPA motifs are arranged so that the proline ring is nestled between, and in van der Waals contact with, the proline and alanine side chains of the other NPA motif (see Figure 14C).

1.3.2.2 Glycerol Transport Through GlpF

The GlpF crystal structure shows three glycerol molecules (G1-G3) within the channel [1]. G1 is found in the periplasmic vestibule of GlpF where it makes a hydrogen bond with Y138 [1]. G2 and G3 are located deep within the pore in what Fu *et al.* [1] call the “*selectivity filter site*”, where the channel is only large enough to house one CH-OH group (see Figure 15). This also explains why GlpF transports long chain polyols but not sugars.

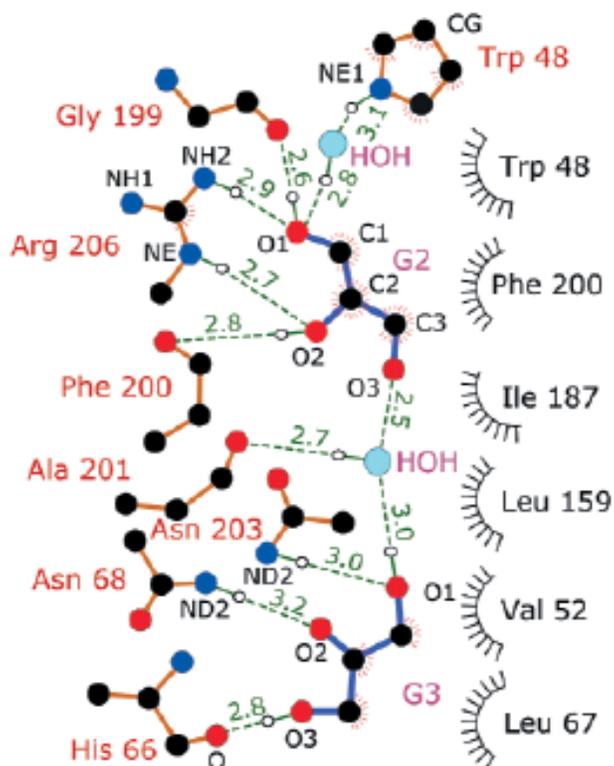


Figure 15: The hydrogen bonding network of GlpF as glycerol passes through the channel.

The hydrogen bonding network (dotted lines, distances in Angstroms) showing interactions of G2, G3 and associated water molecules with the selectivity filter. Hydrophobic contacts are indicated as radial lines around atoms or residues. O1 and O2 are hydrogen bond donors to successive NHs of Arg 206, and an acceptor from carbonyl oxygens of Gly 199 and Phe 200 (respectively). The NH2s of Asn 203 and Asn 68 form hydrogen bonds with O1 and O2 of G3 respectively, while O3 forms a hydrogen bond with His 66.

The alkyl backbone of G2 is tightly packed into the aromatic rings of W48 and F200, which provide part of a “greasy slide” for the glycerol backbone and offer no space for substitutions of the CH hydrogen atom. As a result of this constriction, the CH-OH groups of glycerol must pass through the selectivity filter in single file. Glycerol may be co-transported with water as evident from the water molecule between G2 and G3; however as the glycerol molecule is transported it is substantially dehydrated via the exchange of one set of stringent hydrogen bonds for another. As is shown in Figure 15,

the selectivity filter is designed with two “*sides*”: one side is polar, with contributions from both side chain and backbone atoms while the other side is lined with hydrophobic amino acids [1]. As a glycerol molecule enters the pore it is forced to reorient such that its alkyl backbone can form van der Waals interactions with, and “*slide down*”, the hydrophobic side, while the polar hydroxyl groups are forced to face the polar side of the selectivity filter. Fu *et al.* [1] suggest that an “*electrostatic triangle*” is achieved by the protein in order to polarize two successive hydroxyl groups of the glycerol, and is shown in Figure 15. The carbonyls of G199, F200 and A201 are oriented on the periplasmic side of the filter *via* hydrogen bonds to main chain amides of F200, A201 and the side chain carbonyl of E152. The negative charge of E152, the amides of F200 and A201, and the positive charge of R206 form the “*electrostatic triangle*”. This provides yet another filter, where the molecule filling the pore must be polarizable in sections parallel to the plane of the membrane.

1.3.2.3 Folding and Stability of the AQPs and GLPs

The quaternary structure of the GlpF protein, and indeed the whole AQP family of proteins has been under debate since their initial characterization in 1992 [128]. While there is a general consensus in the current literature that the AQPs are tetrameric in both the membrane and in detergent solution [110, 124, 129-136] the oligomeric state of the GLPs in solution and in the membrane remains a point of interest. Electron microscopy studies on the oligomeric state of the GlpF protein have led to ambiguous conclusions.

Both negative staining and cryo-electron microscopy of 2-D GlpF crystals have suggested, based on the size of the observed particles, that the GlpF protein is a homotetramer similar in size to AQP1 [137, 138]. In addition, it was also found that in the two-dimensional crystal arrays, the unit cell was an octamer comprised of two tetramers in side-on association [137]. The tetrameric assembly analysis of GlpF by electron microscopy agrees well with the model of a GlpF tetramer directly interacting with the Glycerol Kinase tetramer in *E. coli* as proposed by Voegelé *et al.* [139]. In contrast, freeze-fracture electron microscopy studies of GlpF inserted into oocyte membranes suggested that the protein is monomeric in the membrane based on particle size analysis [130]. However, we must consider that the heterologous system may induce non-native effects on the structure of the protein owing to difference in lipid composition between the oocyte membrane and the native *E. coli* membranes. The crystallization of GlpF from octyl glucoside (OG)-solution, and subsequent analysis by x-ray diffraction [1], indicates that in 3-D crystals the GlpF protein exists as a tetramer, and is suggested to be the biologically active form of the protein.

Solution studies of the GlpF protein have also led to ambiguous results concerning the oligomeric state of the protein. The current literature suggests essentially two possibilities: first that the GLPs are inherently weakly associated tetramers which exist predominantly as a monomers in the membrane and in solution [130-134, 140], and secondly that the GLPs exist as stable tetramers in the membrane, but are reduced to monomeric or weakly associated multimeric species in detergent solution [1, 129, 137, 138, 141]. In addition to using SDS PAGE for monitoring the coupled events of

membrane protein unfolding/dissociation, SDS PAGE may also furnish qualitative information regarding the oligomeric stability in that most membrane proteins do not dissociate in the presence of SDS unless heated to high temperatures [142], *e.g.* aquaporins are tetramers on SDS-PAGE and in the crystal, while GlpF is a monomer or mixture of oligomers on SDS-PAGE, but tetrameric in the crystal. In addition to using SDS PAGE, native PAGE has been used successfully in determining the oligomerization state of many membrane proteins [143].

1.4 Biophysical Techniques

1.4.1 Circular Dichroism

One method for monitoring the different aspects of protein structure is circular dichroism (CD) spectroscopy. CD is a spectroscopic method that depends upon the differential absorption of left- and right-circularly polarized light by optically active molecules or groups and is a standard tool in biochemistry and biophysics where it is used to characterize biopolymers such as proteins and nucleic acids (see [144], and references therein for a review). CD is invariably bound to the interactions among the groups being measured, and is therefore a measure of molecular geometry [145]. When CD is measured, a characteristic curve results (for example, see Figure 16) from the rotary strength of the chromophore, which is called a Cotton effect. Cotton effects arise when the electronic motions of a transition are such that parallel components of electronic

and/or magnetic moments are generated. Theoretical models [145] show that for simplified systems Cotton effects are generated by one of three mechanisms: (1) Both electronic and magnetic transitions are on the same chromophores (the “*one-electron*” theory). (2) Both chromophores have electronic transitions, and due to their geometry, these transitions couple to produce a Cotton effect (the Kuhn-Kirkwood mechanism, or “*exciton*” theory). (3) One chromophore has an electronic transition coupled to a magnetic transition on a neighboring chromophore (the “ μ -*m*” mechanism). For a more detailed description, see [145], and references therein. While the structural information obtained from CD is limited compared to that obtained by XRD, and NMR spectroscopy, it has many advantages. One such advantage is the wide range of solution conditions and temperatures which can be examined. In addition, data collection is rapid and only small amounts of sample are required for CD. And, as with other spectroscopic techniques, the sample is not inherently destroyed by CD. Consequently CD is a valuable addition to other structural biology techniques.

1.4.1.1 Far UV CD

In the far UV region of the spectrum, protein absorbance is dominated by the $\pi \rightarrow \pi^*$ transition at 190 nm and the $n \rightarrow \pi^*$ transition near 220 nm [146]. In typical proteins the most important contribution to the far UV signal is the orientation of the peptide bonds with respect to one another. It is therefore possible to approximate the far UV spectrum as a linear combination of signals from α -helical, β -sheet, and “random coil” secondary structural elements (see Figure 16), as well as from aromatic contribution.

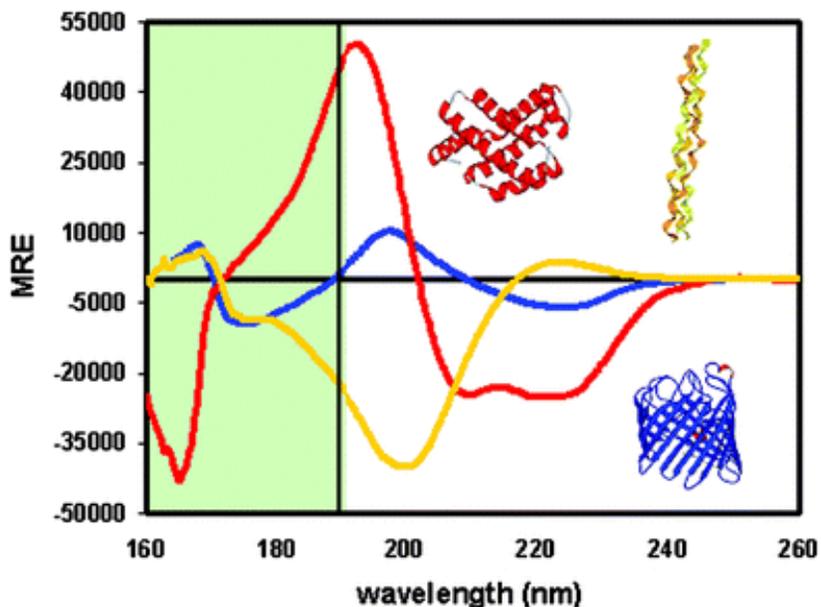


Figure 16: Far UV CD of predominantly α -helical, β -sheet and coil protein conformations.

The three main secondary structural elements are represented: α -helix (red) β -sheet (blue) and coil structures (yellow). Data below 190 nm (shaded green) is not attainable via conventional CD instrumentation, and requires the use of a synchrotron radiation source. Reproduced from [147] with permission.

By comparing the far UV CD spectrum of a protein of unknown secondary structure to measured far UV CD spectra of proteins with known secondary structure, an estimate of the content of α -helical, β -sheet and “disordered” structure can be obtained. This has recently become very accessible due to the advent on the internet of web pages dedicated to protein structure, such as DICHROWEB [148] and the Protein Data Bank [60], and to algorithms for protein secondary structure analysis (for example CDSSTR [149]). In addition, the use of synchrotron radiation allows the collection of data to much lower wavelengths than possible on conventional instrumentation [147]. However, the most useful application of far UV CD may be the detection of changes in the secondary structure of proteins. Since the detection of changes at the secondary structural level

must involve many residues, these changes are considered to be large scale, and not individual residue perturbations.

1.4.1.2 Near UV CD

The study of the tertiary structure of proteins by CD centres on the electronic transitions of the aromatic chromophores (*i.e.* side chains of phenylalanine, tyrosine and tryptophan) in the near UV region of the spectrum (320-260 nm) [150]. Because both the occurrence of these residues, and their molar absorptivities in the near UV region are low (relative to the peptide bond in the far-UV region), the near UV CD absorption and resulting CD bands are much weaker than in the far UV CD [151]. As such, much higher concentrations of protein are required in near UV CD analyses [150].

In proteins, only the peptide (exploited in far UV CD studies) and disulfide bonds are intrinsically optically active [150]. Because the aromatic chromophores are not intrinsically optically active, it is their asymmetrically organized electronic environment which bestows upon them transitions measurable by CD spectroscopy. These transitions rely on their local environment, and on whether there is any freedom of rotation around the C_{α} - C_{β} or C_{β} - C_{γ} bonds, thus their optical activity is largely determined by the folded nature of the protein. Therefore, a near UV CD spectrum is considered to be indicative of the overall protein tertiary structure, with any tertiary structural changes being reflected in the near UV CD spectrum [152].

The aromatic transitions are of the $\pi \rightarrow \pi^*$ type and are labeled 1L_a and 1L_b based on the Platt notation; these are perpendicular to one another in the plane of the π -bonding system [150]. The transitions occur at different wavelengths for each of the aromatic chromophores. However, the intense bands produced by the overlapping of 1L_a (275 nm) and 1L_b (280-290 doublet) tryptophan transitions tends to dominate the near UV CD spectra of folded proteins due to its large (with respect to phenylalanine and tyrosine) molar absorptivity [152].

Near UV CD in the 320 nm to 260 nm range results from absorbance of aromatic residue side chains held rigidly in asymmetric environments [150]. Although near UV CD can not quantitatively assess tertiary structure, the absorption bands may be qualitatively analyzed as arising from aromatic side chains held in rigid positions and occupying asymmetric environments in folded proteins. In a tertiary unfolded protein, the aromatic side chain residues will have no rigid positions or asymmetric environments and therefore no preference to absorb left or right circularly polarized light, which will result in a lack of intensity throughout the near UV region. Therefore, any decrease in near UV CD intensity may be interpreted as a decrease in the overall tertiary structure of the protein.

1.4.2 Fluorescence Spectroscopy

Fluorescence from the amino acid tryptophan has long been known to be sensitive to the polarity of its local environment [153-159], and is an inviting candidate for such a probe, except that the required microscopic information has not been conveniently accessible.

Fluorescence spectroscopy has many advantages in studying protein tertiary structure and folding; fluorescence has high sensitivity, is non-destructive, fast, and is sensitive to the environment of the fluorophore. When studying protein tertiary structure using fluorescence spectroscopy it is the indole-ring of tryptophan which is usually monitored due to the large molar absorptivity and high quantum yield of the tryptophan residue and because the excitation and emission wavelengths have a relatively large separation. In folded globular proteins tyrosine emission is often highly quenched by radiationless energy transfer to tryptophan as well as by interactions with other protein groups [155]. About 300 papers per year abstracted in Biological Abstracts report work that exploits or studies tryptophan (Trp) fluorescence in proteins. Among the properties used are changes in the fluorescence intensity, wavelength maximum (λ_{max}), band shape, anisotropy, fluorescence lifetimes, and energy transfer. They are applied to folding/unfolding, substrate binding, external quencher accessibility, and other questions or areas of investigation.

The unfolding of a protein may be followed by measuring protein tryptophan fluorescence during thermal or chemical denaturations, or during unfolding of the protein due to extremes of pH. The average tryptophan fluorescence reflects the change in the

local environment around these residues, and therefore can be used as a direct means of detecting the unfolding of proteins in solution [160]. Tryptophan λ_{\max} is quite sensitive to its local environment, ranging from ~ 308 nm (azurin) to ~ 355 nm (*e.g.*, glucagon) and roughly correlates with the degree of solvent exposure of the chromophore. However, emission maximum for tryptophan in proteins is sensitive to both the polarity and the dynamics of the environment surrounding the side-chain, and is blue-shifted in environments of low polarity such as the hydrophobic interior of a protein or in a detergent micelle. In 1967, Konev and colleagues put forward the hypothesis of the existence of two main classes of tryptophan residues in proteins [161]. More recently (2001) Burstein and colleagues have extended the classification of Trp fluorescence in proteins into 5 classes depending on the polarity and dynamics of the Trp environments [162]. First, in Class A ($\lambda_{\max} = 308$ nm) the tryptophans are buried and do not form hydrogen-bound complexes in the excited state (exciplexes [163]) with solvent or neighbouring protein groups. Secondly, Class S ($\lambda_{\max} = 316$ nm) includes the buried tryptophan residues that can form exciplexes with 1:1 stoichiometry. Third is Class I ($\lambda_{\max} = 330\text{--}332$ nm) which represent the buried fluorophores that can form the exciplexes with 2:1 stoichiometry. Fourth is Class II ($\lambda_{\max} = 340\text{--}342$ nm) which represent the fluorophores exposed to the bound water possessing very long dipole relaxation time, which precludes completing the relaxation-induced spectral shift during the excited-state lifetime. Lastly, is Class III ($\lambda_{\max} = 350\text{--}353$ nm) which contains rather fully exposed tryptophans surrounded by highly mobile water completely relaxing during the excitation lifetime, which makes their spectra almost coinciding with those of free aqueous tryptophan.

1.5 Purpose of the Research

In light of the paucity of information on membrane protein structure and folding, and specifically the conflicting evidence of the structural characterization of members of the MIP superfamily in solution and in the solid state, I propose to study the structure of the *Escherichia coli* integral membrane protein glycerol facilitator in solution. I propose to study the secondary, tertiary and quaternary levels of structure in neutral, zwitterionic and negatively charged membrane mimetic environments. The stability of the protein will be measured using heat, urea and changes in pH. Fluorescence and circular dichroism will be used to monitor secondary and tertiary structure, while quaternary structure will be monitored *via* SDS- and BN-PAGE.

Chapter 2 Materials and Methods

2.1 Materials

Dodecyl- β -D-maltoside was purchased from Anatrace (Maumee, OH) or Sigma (St. Louis, MO). Sodium dodecyl sulphate, octyl- β -D-glucopyranoside, Coomassie Brilliant Blue R-250, β -mercaptoethanol, and urea were obtained from Sigma. Bis-Tris, Coomassie Brilliant Blue G-250, glycerol, and Tris, were from Fisher Scientific (Fairlawn, NJ). Bromophenol blue was from BDH Inc. (Toronto, ON). Dithiothreitol was from Aldrich Chemical Co. (Milwaukee, WI) and imidazole was from Fluka (Switzerland). 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphocholine was from Avanti Polar Lipids (Alabaster, AL). Nickel-nitrilotriacetic acid (NTA) resin was from Qiagen (Toronto, ON). SDS-PAGE molecular weight markers were from Fermentas Life Sciences (Burlington, ON) and bovine serum albumin, the molecular weight standard for BN-PAGE was from Sigma. D₂O and ¹⁵N ammonium chloride were purchased from Cambridge Isotope Laboratories (Andover, MA). All other materials were of the highest purity commercially available.

2.2 Methods

2.2.1 Glycerol Facilitator Expression and Purification.

E. coli glycerol facilitator was expressed in BL21(DE3)pLysS cells [164] or C43(DE3) cells [165] from a pET28b(+) plasmid (Novagen) encoding an N-terminal His₆ purification tag and T7 epitope, as described previously [126, 141, 166]. The M_r of the GlpF including the N-terminal fusion tags is 33,505 kDa and was confirmed by mass spectrometry [141]; its calculated pI is 7.2 (MacVector, Oxford Molecular). For the purposes of purification of the glycerol facilitator, *E. coli* cells were incubated with DNase, RNase and lysozyme (1 mg each per litre of cells harvested) for 30 minutes, or mechanically sheared by brief sonication on ice. Unbroken cells were removed by centrifugation, and membranes were prepared by ultracentrifugation at 100,000xg for 60 min, at 4°C, in a Beckman SW28 rotor and either a Beckman L8 70M or Beckman-Coulter LE 80k centrifuge [167]. The membranes were then solubilized in 25 mM sodium phosphate buffer pH 7.6, containing 200 mM NaCl, 2 mM β-mercaptoethanol and one of DDM (30 mM), OG (50 mM), SDS (150 mM), or LMPC (20 mM), and the protein was purified using immobilized metal chelate chromatography. Protein purification was done by washing the resin with 15 ml of 25 mM sodium phosphate buffer (pH 7.5) containing 200 mM NaCl, 30 mM imidazole, and detergent (3 mM DDM, 25 mM OG, or 150 mM SDS) or lyso-lipid (2 mM LMPC). The resin was next washed with the same buffer containing 50 mM imidazole until the A₂₈₀ was below 0.01. GlpF

was eluted from the Ni²⁺-NTA column using the same buffer containing 250 mM imidazole. In the case of SDS-solubilized protein, the resin was washed with 5 mM imidazole and the protein eluted using 50 mM imidazole or elution was done by lowering the pH as described previously [141]. For the far UV CD experiments, GlpF samples were dialyzed twice against buffer containing no imidazole. Adjustment of the pH was done by adding glacial acetic acid (up to 50 mM) followed by concentrated HCl, as necessary. Protein purity was analysed by SDS-PAGE [168] and MALDI-Mass Spectrometry [141].

2.2.2 Electrophoresis

Electrophoresis was done using the Hoefer® Mighty Small II SE 250 mini-vertical gel electrophoresis unit.

For the SDS denaturing gels, protein samples were incubated in 50 mM Tris-HCl pH 6.8, 2% SDS, 0.1% Bromophenol blue, 10% glycerol, and 1% (V/V) β-mercaptoethanol for 10 minutes at 25°C, prior to gel loading. Separation was by SDS-PAGE in Laemmli discontinuous gels [168] composed of a 4% acrylamide stacking gel and a 10% resolving gel. After electrophoresis, proteins were visualized by staining with Coomassie Brilliant Blue R250.

For the Blue Native gels [169] protein was incubated in 80 mM BisTris-HCl, pH 7.0, 400 mM 6-amino-caproic acid, and Brilliant Blue G-250 (1/100 of detergent) for 10 minutes at 25°C prior to gel loading. BN PAGE was done using step (4%-10% polyacrylamide) or continuous (4%-15%) gradients at 5°C. Electrophoresis was started at 50 Volts and continued until all the protein entered the gel after which the voltage was increased to 200 Volts for 3-4 hours.

For pH titrations *via* both SDS PAGE and BN PAGE, sample pH was sequentially adjusted, with aliquots removed at each pH increment.

2.2.3 Circular Dichroism and Fluorescence.

CD and fluorescence spectra were acquired with a Jasco 810 spectropolarimeter/fluorometer calibrated with camphorsulfonic acid. Sample temperature was controlled by a Julabo F25 circulating water bath or with a built-in Pelletier device. Briefly, for acquisition of far UV CD spectra, protein solutions were placed in a quartz cuvette with a 0.1 cm path length and spectra were collected at 10 or 20 nm/min between 250-185 nm with a response time of 16 or 8 seconds respectively, and data pitch of 0.1 nm. For near-UV CD spectra the samples were placed in a quartz cuvette with a 1.0 cm path length and spectra were collected at 5 or 20 nm/min between 340-250 nm with a response time of 32 or 8 seconds respectively, and data pitch of 0.1 nm. Baselines were collected in the same

fashion on buffer solutions and spectra were baseline corrected. Protein concentrations were determined using the calculated molar absorptivity of 38,305 M⁻¹ cm⁻¹ and are indicated in the Figure legends. The CD intensity and wavelength of the spectropolarimeter were calibrated using solutions of *d*-10-camphorsulphonic acid [170]. Mean Residue Ellipticities ($\theta \times 10^{-3}$ deg•cm²•dmole⁻¹•residue⁻¹) were calculated using the equation: $[\theta]_M = M\theta/(10)(l)(c)(n)$ where M is 33,505 grams per mole, θ is the measured ellipticity in millidegrees, *l* is the cell path length, *c* is the protein concentration in g/L, and *n* = 315 peptide bonds (due to the 316 residues). Deconvolution of the CD spectra into pure component spectra was performed using the algorithm CDSSTR [149] accessed through Dichroweb [148]. For calculations of helix content we have assumed that none of the 35 amino acids in the His₆-T7 tag are helical. Molar ellipticities (deg•cm²•dmole⁻¹) were calculated using: $[\theta] = \theta/10(c)(l)$ where θ is the measured ellipticity in millidegrees, *l* is the cell path length in cm and *c* is the protein concentration in mol/L. Fluorescence spectra were collected in a 1 cm² rectangular quartz cell with a 0.5 s response time, a data pitch of 1 nm, a scanning speed of 100 nm/min, and excitation and emission bandwidths of 8 nm and 10 nm, respectively.

2.2.4 Curve Fitting

Near UV and far UV CD temperature titrations were done at a rate of 60 °C/hr, and

fitted to Equation 1, where Y_{obs} is the observed ellipticity, Y_f and Y_u are the ellipticities of the low (folded) and high (unfolded) temperature forms of the protein respectively, m_f and m_u are the slopes of the curves at low and high temperatures respectively, R is the gas constant, and ΔH and ΔS are the enthalpy and entropy of unfolding [171, 172].

$$\text{Equation 1: } Y_{obs} = \frac{Y_f + m_f T + (Y_u + m_u T) \bullet e^{(-\Delta H + T\Delta S)/(RT)}}{1 + e^{(-\Delta H + T\Delta S)/(RT)}}$$

Errors were calculated using the Student's t test. The temperature gradient of 60 °C/hr was chosen so as to be consistent within experiments, and with current literature [173].

Equation 1 fits the thermal dependence of the ellipticity to a two-state unfolding transition yielding a value for $T_m = \Delta H / \Delta S$, the midpoint of the transition where the concentrations of folded and unfolded protein are equal. As the unfolding reactions are irreversible the fits were used only to extract T_m values as quantitative indicators of the thermal stability of the protein. The cooperativity of the transitions was quantified by calculating the range of temperatures (ΔT) over which 80% of the unfolding occurs and which corresponds to a K_{eq} range of 0.25 to 4.0. The urea unfolding rate constant was obtained by fitting the time-dependence of the near UV CD ellipticity at 268 nm to Equation 2.

$$\text{Equation 2: } |Y_{obs}| = A \bullet e^{(-k \bullet t)}$$

All the fits were done using the non-linear least-squares fitting program in *Mathematica*TM 5.1 [174].

Chapter 3 Results

Biophysical Characterization of GlpF in Detergent Solutions

Detergent plays an important role in determining the folded states of membrane proteins and GlpF is no exception. We have prepared GlpF in a variety of detergents including a denaturing ionic detergent (SDS), non-denaturing detergents (OG and DDM) and a lyso-phospholipid (LMPC). These were chosen as they represent a variety of polar head groups and carbon chain lengths in the lipidic tails, are relatively inexpensive, and are commonly used detergents/lyso-lipids in membrane protein research. Thermal unfolding of GlpF in these detergent solutions provides an assessment of protein stability by using the mid-point of unfolding. In addition, lowering the pH of detergent-solubilized GlpF induces an acid-unfolded “molten globule-like” structure in DDM and LMPC; OG-dissolved GlpF proved to be too unstable for such analysis. In order to characterize the effects of these thermal and pH changes we have studied the quaternary, tertiary and secondary structural elements of GlpF in these detergent solutions.

3.1 Quaternary Structure Analysis of Detergent Solubilized Glycerol Facilitator

The GlpF quaternary structure in each of the detergent solutions was studied using SDS PAGE and BN PAGE. Most water soluble proteins are unfolded by SDS, which binds to

them at a ratio of about 1 SDS per 2 amino acids [175]. Thus, SDS PAGE is considered a denaturing technique which gives reasonably accurate protein molecular weights based on similar charge-to-mass ratios. By comparison, for membrane proteins BN PAGE is a native electrophoretic method used to observe membrane protein complexes (quaternary structure) when non-denaturing detergents are used to stabilize the protein [169].

3.1.1 Glycerol Facilitator Quaternary Structure in Sodium Dodecyl Sulfate Solution

The SDS electrophoregrams in Figure 17 and Figure 18 show that when GlpF is solubilized in SDS solution, it exists as a mixture of oligomers whether eluted from Ni-NTA-resin using a pH step-gradient (Figure 17) or using imidazole competition at pH 7.8 (Figure 18).

The electrophoregram of GlpF dissolved in 150 mM SDS and eluted from Ni-NTA-resin using a pH gradient (Figure 17), shows the existence of a mixture of oligomers. These include monomer, just below 36 kDa (lactate dehydrogenase molecular weight standard) dimer, which runs between 47 and 85 kDa (ovalbumin and bovine serum albumin molecular weight standards respectively), trimer near 85 kDa and tetramer near 118 kDa (β -galactosidase molecular weight standard), with dimer representing the dominant species (see Figure 17).

SDS solubilized GlpF eluted from Ni-NTA-resin using imidazole competitive binding exists as a mixture of oligomers as observed by SDS PAGE (see Figure 18). Monomers (M), dimers (D) and trimers (Tr) are observed in Figure 18, however the predominant species appears to be monomeric under the neutral pH imidazole elution conditions.

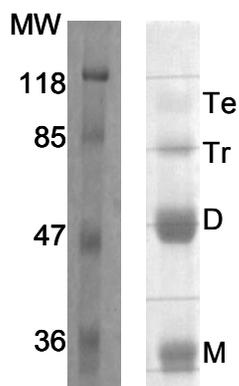


Figure 17: SDS PAGE electrophoregram of GlpF dissolved in 150 mM SDS solution eluted from the Ni-NTA-resin at pH 4.

Monomer (M), dimer (D), trimer (Tr) and tetramer (Te) of the GlpF protein are all observed in the right lane. Molecular weight standards are shown in the left lane.

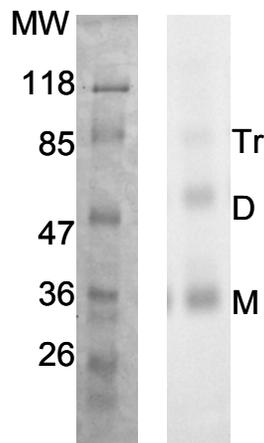


Figure 18: SDS PAGE electrophoregram of GlpF dissolved in 150 mM SDS solution eluted from the Ni-NTA-resin using imidazole competition.

Monomeric (M), dimeric (D) and trimeric (Tr) protein. Molecular weight standards are shown in the left lane, GlpF is shown in the right lane.

3.1.2 Glycerol Facilitator Quaternary Structure in Octyl Glucoside Solution

An SDS PAGE electrophoregram of GlpF solubilized in OG shows that it exists as a mixture of oligomers when eluted from NTA-resin with low pH. Significant amounts of monomer, dimer, trimer, tetramer, pentamer and hexamer can be identified on the electrophoregram (see Figure 19 right lane). Low pH-eluted GlpF solubilized in OG appears to be almost equal parts monomer and dimer, with lesser amounts of the other oligomers. In addition, Figure 19 shows that considerable amounts of high molecular weight oligomers (‡) are present at the top of the SDS PAGE gel, but are not resolved under these electrophoresis conditions. The observation that OG-dissolved GlpF exists as a mixture of oligomers is in good agreement with much of what is seen in the current

literature [129, 133, 134]. However, OG has been used to prepare GlpF protein that crystallizes as a tetramer [1].

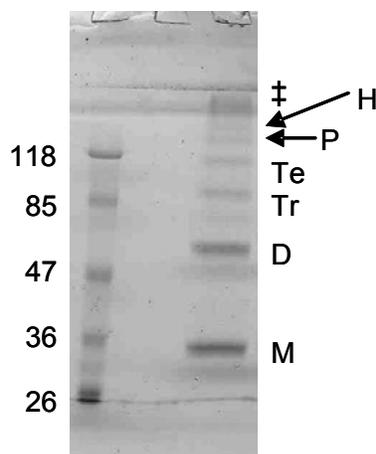


Figure 19: SDS PAGE electrophoregram of GlpF dissolved in OG solution eluted from the Ni-NTA-resin using low pH.

Monomer (M), dimer (D), trimer (Tr), tetramer (Te), pentamer (P) and hexamer (H) are observed, as well as other larger molecular weight aggregates (‡) that can not be resolved under these conditions.

3.1.3 Glycerol Facilitator Quaternary Structure in Dodecyl Maltoside Solution

In DDM solution, GlpF electrophoreses predominantly as a monomer with only small amounts of dimer, trimer and tetramer observed when eluted from the Ni-NTA-resin using a pH gradient (Figure 20A lane 1). However, when eluted from the Ni-NTA-resin using an imidazole gradient (Figure 20B lane 1) the GlpF protein is almost exclusively tetrameric with only very small amounts of monomer, dimer and trimer and occasionally octamer observed, as shown in Figure 20 lane 1. This represents the first preparation

where detergent-solubilized tetrameric GlpF protein is observed *via* SDS-PAGE analysis. X-ray diffraction (XRD) analysis of the GlpF protein has shown that, similar to other members of the AQP family [123, 176], GlpF exists as a tetramer [1] and that the tetramer is suggested to be the biologically active form of the GlpF protein. The small amounts of monomer, dimer, and trimer may be caused by the presence of SDS during the electrophoresis. SDS may cause small amounts of the GlpF protein to denature, causing both dissociation and partial aggregation of the tetramer into these other species. The new extraction and purification preparation that I implemented has resulted in greater protein purity and a decrease in the amount of monomer, dimer and trimer observed on SDS electrophoregrams when non-denaturing detergents are chosen.

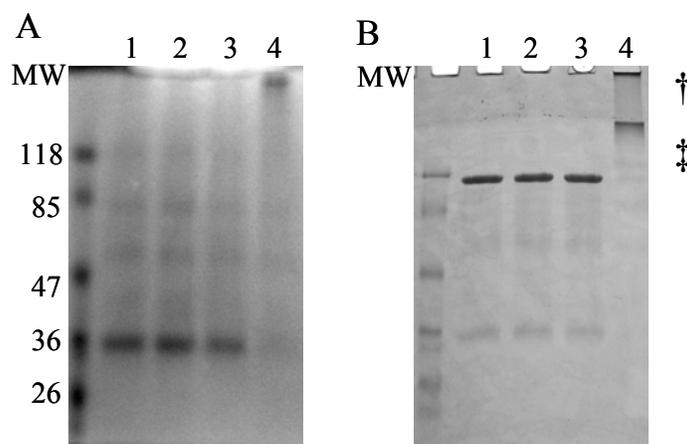


Figure 20: SDS PAGE electrophoregrams of thermal unfolding of GlpF.

The protein was solubilized in DDM and eluted from the Ni-NTA-resin using (A) pH gradient or (B) imidazole competition. (A) and (B): lane 1, 20 °C; lane 2, 40 °C; lane 3, 60 °C; lane 4, 80 °C. † Very high molecular weight oligomers only able to enter the stacking gel. ‡ High molecular weight oligomers able to just enter the separating gel.

Thermal unfolding of proteins can be used to estimate their stability (see [142] and references therein). If this process is not a reversible one no thermodynamic information can be extracted, however useful information about the stability of the protein in solution can be deduced by studying the midpoint of the unfolding transition [142]. The transition midpoint is taken as the point where the number of protein molecules that are folded is equal to the number of unfolded protein molecules.

Thermal unfolding of the low-pH eluted GlpF, solubilized using DDM (Figure 20A), shows an introduction of higher molecular weight oligomers accompanied by a diminishing of the monomeric species. At 20 °C (Figure 20A lane 1) the GlpF protein is predominantly monomeric, while small amounts of dimer, trimer and tetramer are also observed. This pattern is virtually identical at 40 °C (Figure 20A lane 2). A slight decrease in overall protein intensity is observed once the temperature is elevated to 60 °C (Figure 20A lane 3) although no large molecular weight aggregates can be seen. Upon reaching 80 °C (Figure 18A lane 4) only small amounts of monomer, dimer, trimer and tetramer can be observed, with the majority of the GlpF protein being present as very high molecular weight aggregates that barely enter the separating gel. The transition midpoint for denaturation of the monomer in DDM solution at low pH appears to be above 60 °C judging from the intensities of the bands on the gel.

A similar pattern is observed for the imidazole-eluted DDM-solubilized GlpF (Figure 20B) when thermally unfolded. At 20 °C (Figure 20B lane 1) GlpF is predominantly tetrameric on an SDS PAGE gel when dissolved in DDM. Very small amounts of

monomer, dimer and trimer can also be observed. This pattern is maintained on the gel through 40 °C (Figure 20B lane 2) and 60 °C (Figure 20B lane 3) with little or no change in the distribution of species. When the temperature is raised to 80 °C (Figure 20B lane 4) the tetrameric GlpF protein completely disappears from the gel and is now observed as a multitude of oligomers with most of the GlpF protein existing as high molecular weight species. Some of these high molecular weight aggregates are too large to enter the separating gel and can only enter the stacking gel (†), while others enter the top of the stacking gel only (‡). These high molecular weight aggregates are likely formed when the thermally denatured GlpF tetramers dissociate into monomers. These thermally denatured monomers then aggregate into larger oligomers. However, the presence of SDS maintains a small amount of the smaller oligomers in solution, hence the observation of dimer, trimer, tetramer and pentamer, while a large amount of the protein exists as soluble high molecular weight aggregates seen at the interface between the stacking gel and the running gel (‡). The large amount of protein that just enters the top of gel most likely represents insoluble high molecular weight aggregates of the protein (†).

In order to probe the effect of pH on GlpF quaternary structure, imidazole-eluted GlpF at pH 7.8 was subjected to a pH titration (see Figure 21). Figure 21 lane 1 shows that at pH 7.8, DDM-solubilized GlpF runs predominantly as a tetramer on an SDS PAGE gel (as seen previously in Figure 20), with minor amounts of monomeric species evident. Lowering the pH to 7.0 (Figure 21 lane 2) resulted in no observable change in the DDM-solubilized GlpF oligomerization state. However, when the pH is lowered to 6.0 (Figure

21 lane 3) an increased amount of monomeric DDM-solubilized GlpF is detected and a corresponding decrease in the amount of tetrameric DDM-solubilized GlpF is observed on the gel. By pH 5.0 (Figure 21 lane 4), the DDM-solubilized GlpF tetramer is no longer observed, and only the monomeric species is evident. Lowering the pH to 4.0 causes no additional change to the oligomerization state of the protein (Figure 21 lane 5).

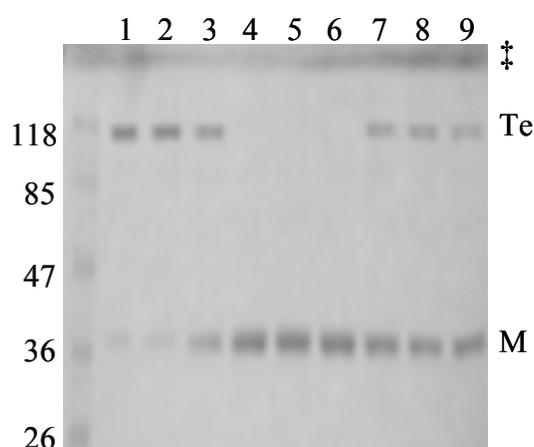


Figure 21: SDS PAGE electrophoregram of a pH titration of DDM-solubilized GlpF.

The GlpF protein was eluted from the Ni-NTA-resin using imidazole competition showing acid induced denaturation and possible renaturation of the GlpF tetramer upon restoration of the pH to near neutral. Lane 1, pH 7.8; lane 2, pH 7; lane 3, pH 6; lane 4, pH 5; lane 5, pH 4; lane 6, pH 5; lane 7 pH 6; lane 8, pH 7; lane 9, pH 7.8. Monomer (M), tetramer (Te) and protein aggregates (‡) are labelled for clarity.

In order to determine if the acid-induced unfolding of the tetramer could be reversed, the pH of the same protein preparation was elevated and the refolding monitored by SDS-PAGE. As the pH is incrementally raised back to 7.8 from 4.0, a slightly different pattern emerges. When the pH is raised to 5.0 (Figure 21 lane 6) only monomeric DDM-solubilized GlpF is observed, as is seen in Figure 21 lanes 4 and 5 (pH 5.0 and 4.0 respectively). A notable difference is first observed when the pH is returned to 6.0

(Figure 21 lane 7), with a re-appearance of the tetrameric DDM-solubilized GlpF. Compared to Figure 21 lane 3 (pH 6.0), there appears to be significantly more monomeric GlpF protein and less tetrameric GlpF protein. Of even greater difference is the protein oligomerization when the pH is raised to pH 7 and pH 7.8 (Figure 21 lanes 8 and 9 respectively). Here the tetramer comprises an even smaller amount of the total protein observed, and monomeric GlpF makes up a much greater fraction of protein on the gel. This result apparently shows that the acid-induced unfolding of the GlpF tetramer is partially reversible. By eye, approximately 30 % of the protein is restored to the tetramer at pH 6, with no further increase in tetramer when the pH is raised to 7 or 7.8.

Figure 21 lane 5 shows that at pH 4, DDM-solubilized GlpF runs as a monomer on SDS PAGE. However, when the pH is raised back to pH 7 (Figure 21 lane 8) a significant amount of tetramer is reformed. During each SDS PAGE experiment proteins are subjected to approximately 60 mM SDS from the sample treatment buffer. In order to separate the effects of SDS and low pH on tetramer formation, an SDS titration was carried out on DDM-solubilized GlpF at pH 4. If the DDM-solubilized GlpF is not exposed to SDS at pH 4 (Figure 22 lane 1), when returned to pH 7 it runs predominantly as a tetramer with very little monomer observed, but noticeable amounts of higher molecular weight oligomers are present (‡). These oligomers appear to be octameric, and are also observed in AQP0 preparations [177]. Little to no change is seen if the pH 4 DDM-solubilized GlpF experiences only 2 mM SDS before raising the pH back to 7 (Figure 22 lane 2). However once the SDS concentration at pH 4 reaches 5 mM (Figure 22 lane 3), the introduction of monomeric species is observed. As the SDS concentration

is further raised to 10 mM (Figure 22 lane 4), the majority of tetrameric and octameric GlpF is lost and significant amounts of both monomer and dimer are observed. Further increases in SDS concentration to 20 mM, 40 mM and 60 mM (Figure 22 lanes 5, 6, and 7 respectively) continue to decrease the tetrameric and octameric species while increasing the monomeric and dimeric forms of the pH 4 DDM-solubilized GlpF.

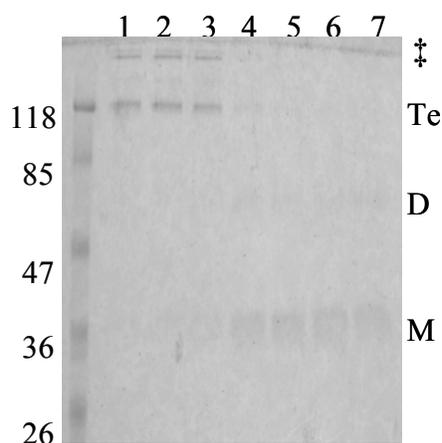


Figure 22: SDS titration of pH 4 DDM-solubilized GlpF observed by SDS PAGE.

The sample pH was restored to pH 7 in all cases prior to electrophoresis. Lane 1, 0 mM SDS; lane 2, 2 mM SDS; lane 3, 5 mM SDS; lane 4, 10 mM SDS; lane 5, 20 mM SDS; lane 6, 40 mM SDS; lane 7, 60 mM SDS. Monomer (M), dimer (D), tetramer (Te) and oligomers (‡) are labelled for clarity.

These experiments suggest that at pH 4, DDM-solubilized GlpF is only slightly unfolded at the quaternary level. However, if the pH 4 protein is subjected to any amount of SDS greater than 10 mM, the protein is converted to a monomeric species which is only partially able to restore its quaternary contacts when the pH is raised back to neutrality. This result suggests that the stability of the DDM-solubilized GlpF tetramer at pH 4 is very low, as only 10 mM SDS is needed to dissociate the tetramer. However, it also

raises the question whether in the absence of SDS the tetramer dissociates at low pH and reforms at pH 7, as suggested in Figure 21, or if the tetramer remains intact at pH 4.

To investigate this further, the quaternary structure of DDM-solubilized GlpF at neutral pH was examined using BN PAGE. Thermal stability was assessed by a temperature denaturation and was monitored using BN PAGE. At 20 °C (Figure 23 lane 1) DDM-solubilized GlpF exists exclusively as a tetramer (Te) when observed by BN PAGE; no amount of monomer, dimer, trimer, or other oligomers are detectable. This indicates that the monomeric, dimeric and trimeric species observed on SDS PAGE for DDM-solubilized GlpF (as seen in Figure 20B) are induced by the presence of SDS required to run the SDS PAGE electrophoregrams. As the temperature is raised the tetramer is maintained until 60 °C as was seen in the SDS PAGE experiment. As the temperature is raised to 80 °C and above, the tetrameric species disappears from the gel and is replaced with higher molecular weight species too large to enter the pores of the gel. This is likely the result of the GlpF tetramer dissociating to monomeric species with the increase in temperature. Unlike SDS PAGE, where the presence of SDS helps maintain soluble monomers and small oligomers, BN PAGE appears unable to solubilize the thermally unfolded monomers. Monomers are not observed because they are prone to aggregation and give rise to very large oligomers.

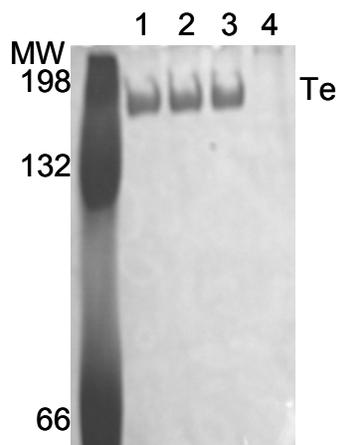


Figure 23: BN PAGE electrophoregram showing the thermal unfolding of imidazole-eluted pH 7 DDM-solubilized GlpF.

Lane 1, 20 °C; lane 2, 40 °C; lane 3, 60 °C; lane 4, 80 °C.

A pH titration of DDM-solubilized GlpF was also studied by BN PAGE in order to observe the effect of acidic pH on the GlpF tetramer under native electrophoresis conditions. At pH 7 (Figure 24 lane 1) DDM-solubilized GlpF is observed as a tetramer (Te) on the BN PAGE electrophoregram with a small amount of octamer (O), and high molecular weight aggregates (\ddagger) at the interface between the stacking and running gels are also observed. As the pH is lowered to pH 6 (Figure 24 lane 2) the intensity of the observed octamer is decreased coinciding with an increase in the higher molecular weight aggregates. In addition, the appearance of very high molecular weight aggregates (\dagger) at the top of the gel is observed. However, no apparent large decrease in the tetramer intensity is observed. As the pH is lowered to 5 (Figure 24 lane 3) this pattern is continued and the octamer now has almost completely disappeared, again with little to no

change in the intensity of the tetramer. However, when the pH is lowered to 4 (Figure 24 lane 4) the tetramer is greatly diminished with a corresponding increase in very large oligomers that are unable to enter the gel. The large amount of high molecular weight aggregates of DDM-solubilized GlpF at pH 4 suggest that the protein is now an unstable tetramer, prone to aggregation. This is very different from the SDS PAGE situation where DDM-solubilized GlpF at pH 4 is exclusively monomeric (Figure 21 lane 5). In DDM, only at pH 3 (Figure 24 lane 5) does the tetramer completely disappear from the gel and is at the same time replaced by a monomeric (M) GlpF species.

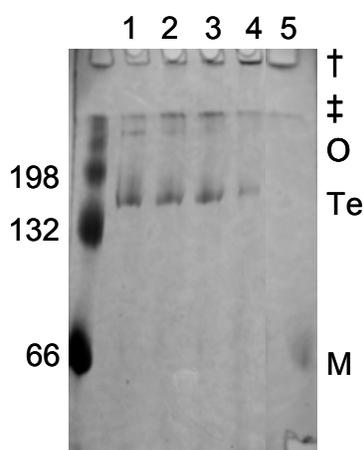


Figure 24: BN PAGE electrophoregram showing pH induced unfolding of imidazole-eluted DDM-solubilized GlpF.

Lane 1, pH 7; lane 2, pH 6; lane 3, pH 5; lane 4, pH 4; lane 5, pH 3.

Since the calculated pI of the GlpF protein is 7.56 [141], lowering the pH should increase the net charge on GlpF; increased charge repulsion should make the protein less prone to aggregation, unless the protein is unfolded by these effects. These results suggest that the

octamer observed between pH 7 and pH 5 is an unstable oligomer, prone to aggregation with even the slightest decrease in pH. This is in contrast to the tetramer, as it appears stable over this pH range. As the pH is decreased and charge repulsion builds, the GlpF tetramer is prone to aggregation near pH 4. The absence of monomeric GlpF above pH 3 suggests that either the tetramers remain intact and aggregate as groups of tetramers, or that the charge repulsion is insufficient to prevent individual monomers from aggregating into very large species. As the pH is lowered to 3, we see the appearance on the BN-gels of monomeric GlpF. Either the tetramers are dissociating to soluble monomeric species or the charge repulsion is now great enough to prevent some monomers from aggregating, and are therefore observed on the BN PAGE electrophoregram.

Furthermore, although the ionization state of Coomassie Blue G250 is not expected to change in the pH range studied here, it is conceivable that the interactions between the dye and protein change under these conditions resulting in some of the observations. It is also conceivable that the change in ionization state of the aminocaproic acid at low pH plays a role in these observations.

Urea is a common protein denaturant used to measure the stability of many water-soluble proteins [178]. Initial experiments on DDM-solubilized GlpF indicated a time dependence of the effects of urea (see Figure 3-20) and in order to permit the establishment of equilibrium conditions the protein was incubated in various concentrations of urea for 3 weeks. Figure 25 and Figure 26 show the effects of additions of urea on the DDM-solubilized protein quaternary structure. According to SDS PAGE, low concentrations of urea (2-4 M; Figure 25 lanes 2-3) have a minor effect on the

amount of tetramer (Te) present in the solution. Small increases in monomer (M), dimer (D) and trimer (Tr) are observed but the most notable change is the significant loss of a high molecular species that appears to be an octamer (O) that is sometimes observed in the preparations. After incubation for 3 weeks in 8 M urea (Figure 25, lane 5), the tetramer is nearly absent, monomer, dimer, and trimer have increased substantially, and the octamer is greatly reduced. In addition, protein oligomers electrophoresing between the tetramer and octamer appear that are likely non-specific aggregates soluble in high concentrations of SDS and urea.

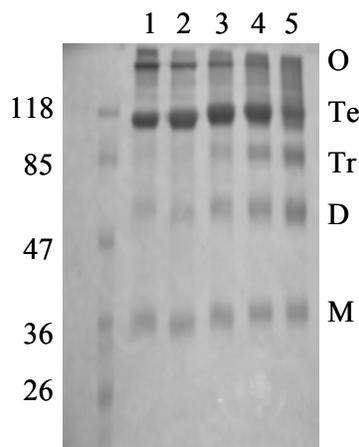


Figure 25: SDS PAGE electrophoregram monitoring the effects of urea on pH 7 DDM-solubilized GlpF after 3 weeks.

Lane 1, 0M urea; lane 2, 2M urea; lane 3, 4M urea; lane 4, 6M urea; lane 5, 8M urea. Monomer (M), dimer (D), trimer (Tr), tetramer (Te) and Octamer (O) are labelled on the electrophoregram.

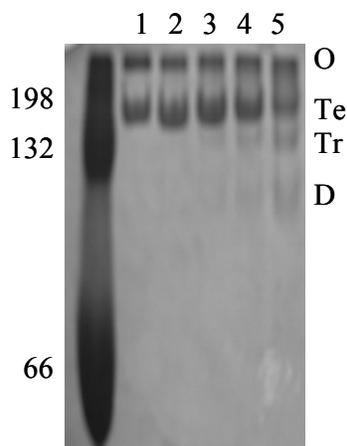


Figure 26: BN PAGE electrophoregram monitoring the effects of urea on GlpF after 3 weeks.

Lane 1, 0M urea; lane 2, 2M urea; lane 3, 4M urea; lane 4, 6M urea; lane 5, 8M urea. Dimer (D), trimer (Tr), tetramer (Te) and Octamer (O) are labelled on the electrophoregram.

The preceding experiment indicates the high stability of the GlpF protein with respect to urea in 60 mM SDS. In order to explore the effects of urea in the absence of SDS, BN PAGE was used. BN PAGE of the same samples is shown in Figure 26. The BN gels appear to be less sensitive to the unfolding induced by urea at low concentrations because in the absence of SDS the low molecular weight species are not observable on the gels. Presumably any monomer that forms, aggregates into high molecular weight aggregates that are too large to enter the gel. It is also possible that in the absence of SDS the tetramer is resistant to dissociation into monomer, dimer and trimer in low concentrations of urea. However, at 6 M and 8 M urea the unfolding of the tetramer (Te) and octamer (O) are evident from their diminishing bands on the BN gel and from the very faint bands that appear to be trimer (Tr) and dimer (D) that appear. The latter are likely soluble in small amounts because of the high concentrations of urea present, whereas in the absence of urea at pH 7 and 20 °C they are never observed.

3.1.4 Glycerol Facilitator Quaternary Structure in Lyso-Myristoyl Phosphatidylcholine Solution

Both SDS and BN PAGE show that in LMPC solution GlpF exists predominantly as a tetramer when eluted from Ni-NTA-resin using an imidazole gradient (Figure 27 A and B respectively, lanes 4). Heating of the imidazole-eluted GlpF protein in LMPC solution results in a similar pattern to that observed for the DDM-solubilized GlpF. The SDS PAGE electrophoregram shows that LMPC-solubilized GlpF is predominantly tetrameric (Te) at 20 °C (Figure 27 A lane 4) with some monomer (M), and very small amounts of dimer (D) and trimer (Tr) observed. This pattern is maintained on the gel through 40 °C (Figure 27 A lane 3) and 60 °C (Figure 27 A lane 2) with little or no change. As the temperature is raised to 80 °C (Figure 27 A lane 1) the GlpF tetramer is reduced in intensity, with a resultant increase in the monomeric, dimeric and trimeric species observed, and in addition high molecular weight aggregates (‡) are introduced. However, the majority of the LMPC-solubilized GlpF still exists as a tetramer at 80 °C. As was seen with the thermal denaturation of DDM-solubilized GlpF (Figure 20 B), some of the observed high molecular weight aggregates only enter the top of the polyacrylamide stacking gel while others remain at the interface between the stacking gel and separating gel.

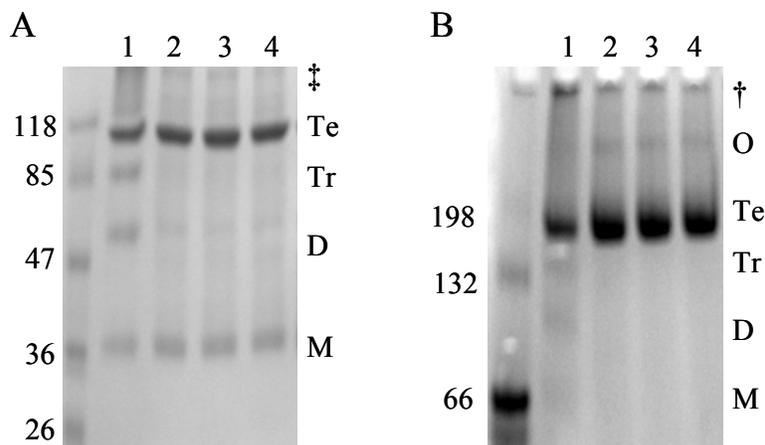


Figure 27: SDS (A) and BN (B) electrophoregrams showing the thermal denaturation of imidazole-eluted LMPC-solubilized GlpF.

(A) and (B): lane 4, 20 °C; lane 3, 40 °C; lane 2, 60 °C; lane 1, 80 °C.

At 20 °C, LMPC-solubilized GlpF exists almost exclusively as a tetramer (Te) when observed on a BN PAGE electrophoregram (Figure 27B lane 4) with no trace of any low molecular weight species. However, as is sometimes seen, the existence of very small amounts of an octameric (O) GlpF species is evident. The GlpF tetramer is again maintained through 40 °C (Figure 27 B lane 3) and 60 °C (Figure 27B lane 2) as was observed on the SDS PAGE electrophoregram (Figure 27A). As the temperature is raised to 80 °C (Figure 27B lane 1), the tetrameric species is slightly reduced in intensity and is replaced by some higher molecular weight species (†) as evident from the protein species too large to enter the polyacrylamide separating gel, in addition to small amounts of monomeric, dimeric and trimeric species of the GlpF protein.

Of particular note in the thermal denaturation of LMPC-solubilized GlpF is the retention of the majority of the tetrameric structure on both SDS and BN PAGE electrophoregrams

even at temperatures as high as 80 °C (Figure 27A lane 1 and Figure 27B lane 1 respectively). This is in contrast to the DDM-solubilized GlpF, which at temperatures of 80 °C have almost no tetrameric GlpF present when observed with either SDS or BN PAGE (Figure 20B lane 4 and Figure 23 lane 4 respectively), suggesting that the phosphatidylcholine head group, or the longer myristoyl carbon chain, endows the GlpF protein with greater thermal stability.

In order to probe the effect of acidic conditions on LMPC-solubilized GlpF quaternary structure, the LMPC-solubilized GlpF protein eluted from Ni-NTA-resin using imidazole (pH 7.2) was subjected to a pH titration and monitored using both denaturing SDS PAGE, and non-denaturing BN PAGE. Denaturing SDS PAGE shows that in LMPC solution, GlpF is predominantly tetrameric (Te) near neutral pH (Figure 28A lane 5). Once the pH is lowered to 6.0 however (Figure 28A lane 4), increased amounts of monomer (M), dimer (D) and trimer (Tr) are detected coinciding with a decrease in the amount of tetramer observed. By pH 5.0 (Figure 28A lane 3), the LMPC-solubilized GlpF tetramer has almost completely disappeared from the SDS PAGE gel with the GlpF protein now existing as a mixture of oligomers (monomer, dimer, trimer and tetramer), but is predominantly monomeric. This pattern is retained at pH 4.0 (Figure 28A lane 2) and pH 3.0 (Figure 28A lane 1) on the SDS gel.

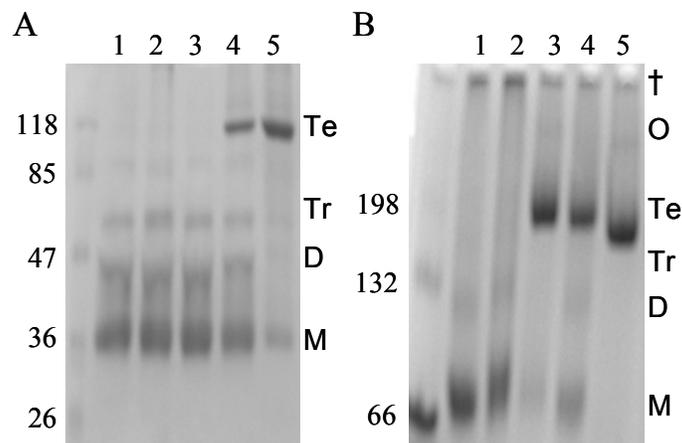


Figure 28: SDS (A) and BN (B) electrophoregrams of pH titrations of imidazole-eluted LMPC-solubilized GlpF.

(A) and (B): lane 5, pH 7.2; lane 4, pH 6.0; lane 3, pH 5.0; lane 2, pH 4.0; lane 1, pH 3.0. Monomer (M), dimer (D), trimer (Tr), tetramer (Te) and octamer (O) are labelled for clarity.

A pH titration of LMPC-solubilized GlpF was also studied by BN PAGE in order to investigate the role of pH in the absence of SDS (Figure 28B). At pH 7.2 (Figure 28B lane 5), LMPC-solubilized GlpF is predominantly tetrameric (Te) on the BN gel with only a small amount of the octameric species (O) present, as expected based on the SDS gel pattern. When the pH is lowered to 6.0 (Figure 28B lane 4) a decrease in the tetrameric GlpF protein structure is observed with a corresponding increase in monomeric (M) and dimeric (D) species. Little change is observed at pH 5.0 (Figure 28B lane 3). Once the pH is lowered to 4.0 (Figure 28B lane 2) the tetrameric species has almost completely disappeared from the gel. In its place are low molecular weight species (monomer, dimer and trimer) and very high molecular weight species (†), which are too large to enter the pores of the separating gel. This pattern is maintained at pH 3 (Figure 28B lane 1). Of particular interest is the intolerance of the GlpF tetramer to pH changes when dissolved in LMPC solution, compared to that of GlpF in DDM solution.

The LMPC-solubilized GlpF tetramer starts to unfold at pH 6, and is completely unfolded by pH 4 (Figure 28B), while DDM-solubilized tetrameric GlpF appears to be stable to pH 5, with only partial unfolding at pH 4 (Figure 24). It is possible that the zwitterionic nature of the phosphatidylcholine head group ($pK_a = 0.8$ [179]) could be responsible for this sensitivity to low pH.

The combined results of the quaternary structure analysis lead to the conclusion that only when the GlpF protein is eluted from Ni-NTA-resin using an imidazole competition at or near neutral pH, can the GlpF protein exist as a stable tetramer with appropriate quaternary contacts in detergent solution. In addition, only DDM- and LMPC-solubilized GlpF exhibited stable tetramer quaternary contacts.

3.2 Tertiary Structure Analysis of Detergent-Solubilized Glycerol Facilitator

3.2.1 Glycerol Facilitator Tertiary Structure in Dodecyl Maltoside Solution

When the GlpF protein is dissolved in DDM solution at room temperature and neutral pH, the near UV CD spectrum is suggestive of a well folded protein, *i.e.* there is significant intensity across the near UV region (Figure 29). This agrees with the quaternary structure analysis, which shows the protein exists as a folded tetramer (see section 3.1). Phenylalanine would be expected to contribute the least to the near UV CD spectrum due to its high structural symmetry and the low sensitivity of its transitions to changes in solvent polarizability [152]. However, due to their high abundance in the GlpF protein the 21 phenylalanines could contribute significantly to the spectral intensity, and are the likely origin of the band at approximately 260 nm. Tryptophan residues usually dominate the near UV CD spectra of proteins because of their large molar absorptivity, and the bands between 265 nm and 290 nm are attributable to the 5 tryptophan residues; the Trp 1L_a band is usually near 275 nm and the 1L_b doublet usually falls between 280 and 290 nm [150].

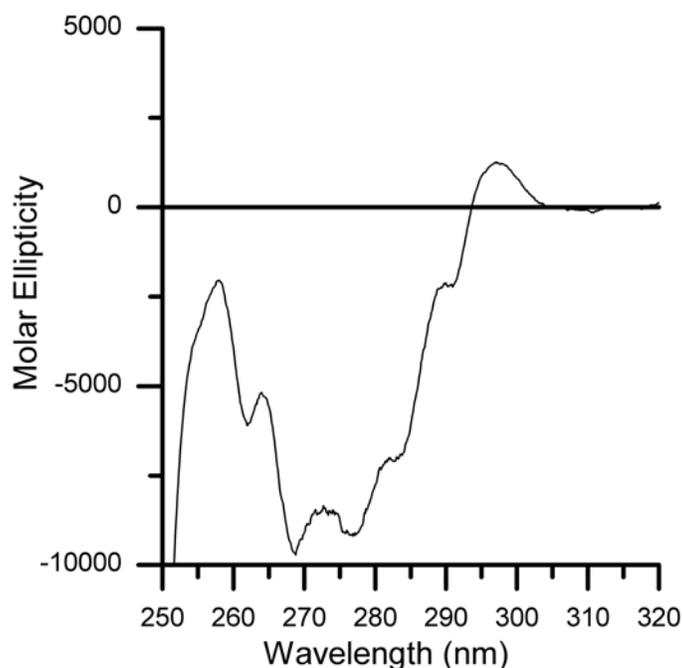


Figure 29: Near UV CD spectrum of DDM-solubilized GlpF protein (69.4 μM) at pH 7 and 25 $^{\circ}\text{C}$.

In order to analyze the thermal stability of the GlpF tertiary structure in DDM solution, the DDM-solubilized GlpF was subjected to a thermal denaturation. Figure 30 clearly shows that as the temperature is raised from 20 $^{\circ}\text{C}$ to 95 $^{\circ}\text{C}$ the near UV CD intensity diminishes to nearly zero, indicating a total loss of tertiary structure. As the temperature is raised from 20 $^{\circ}\text{C}$ to 65 $^{\circ}\text{C}$ (dotted line), the GlpF spectra show a gradual loss of intensity, indicating partial unfolding of the DDM-solubilized GlpF tertiary structure. Continued heating to 70 $^{\circ}\text{C}$ reduces the intensity to nearly zero, which represents a total loss of the GlpF tertiary structure.

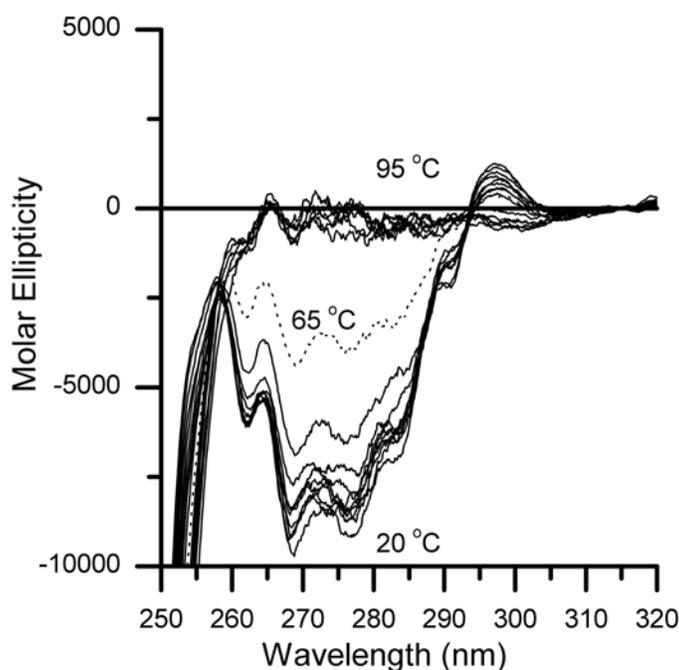


Figure 30: Spectroscopic analysis of the thermal denaturation of imidazole-eluted, pH 7 DDM-solubilized GlpF (69.4 μM) tertiary structure monitored by near UV CD.

The spectra at 20°C, 65°C, and 95°C are labelled for clarity.

Figure 31 shows the thermal unfolding of DDM-solubilized GlpF by following the near UV CD intensity at 268 nm, which is predominantly due to tryptophan. As the temperature is raised from 20°C to approximately 60°C small linear decreases in the negative ellipticity of the near UV CD signal at 268 nm (from approximately -9500 deg cm^2/dmol at 20°C to approximately -8000 deg cm^2/dmol at 60°C) are observed. Thus, at 60°C the protein retains approximately 80 % of its 20°C near UV CD intensity. Only as the temperature reaches 65°C does the loss of intensity become non-linear, as the intensity decreases to approximately -7000 deg cm^2/dmol at 268 nm. At 70°C the near UV CD intensity at 268 nm further decreases to approximately -4800 deg cm^2/dmol . At 75°C the near UV CD intensity at 268 nm is nearly zero indicating a total loss of tertiary

structure. As the temperature is further increased, the intensity becomes again slightly negative. This may indicate misfolding and aggregation of the protein with a slight propensity for certain structured local environments about the tryptophans. Following other regions of the near UV CD spectra results in a similar mid-point of unfolding, suggesting that the process occurs simultaneously across all areas of the protein. Both Figure 30 and Figure 31 show a cooperative unfolding transition between 60-70°C. Analysis of the thermal denaturation (Figure 31) using the program *Mathematica*TM 5.1 to fit the denaturation to a 2-state process (see Equation 1 in Materials and Methods) shows that the DDM-solubilized GlpF near-UV CD ellipticity is eliminated with a mid-point of unfolding (T_m) of $65.5 \pm 0.8^\circ\text{C}$ and a transition width (ΔT), of 7.5°C . The narrow width of the transition suggests a high level of cooperativity of the transition. Because of the slow scan rate (5 nm/min) and long response time (32 sec), the heating rate for these experiments was $0.28^\circ\text{C}/\text{min}$. When the scan rate and response time were adjusted (20 nm/min and 8 seconds respectively) to allow a rate of heating to $1^\circ\text{C}/\text{min}$, the melting point increased to $71 \pm 2^\circ\text{C}$ (data not shown). This result is in good agreement with those of Sehgal *et al.* [173], who measured the T_m dependence of adhesion involved in diffusion adherence (AIDA) protein. The T_m of AIDA is elevated by 6°C when the scan rate is increased from 0.25 to $1^\circ\text{C}/\text{min}$. However, the T_m is more sensitive to heating rates below $0.66^\circ\text{C}/\text{min}$, and less sensitive to those above.

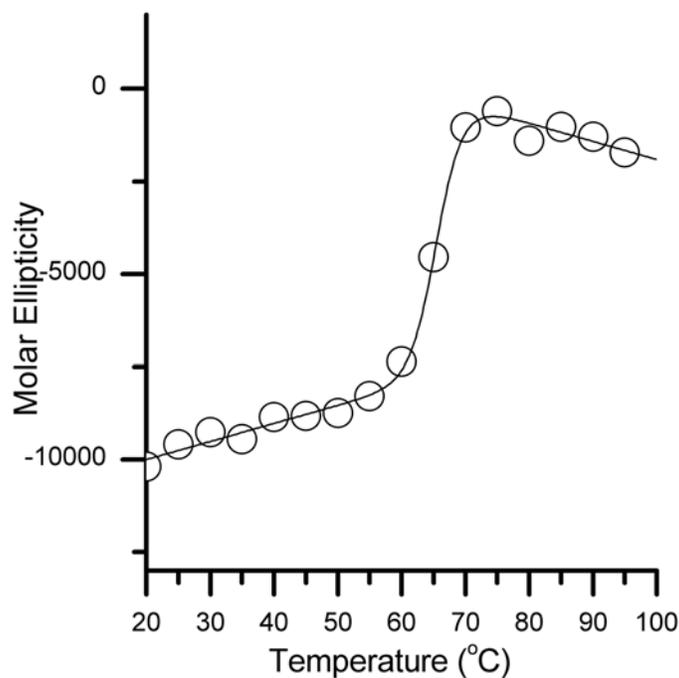


Figure 31: Molar ellipticity changes of DDM-solubilized GlpF (69.4 μ M) at 268 nm.

The changes in molar ellipticity at 268 nm (open circles) of DDM- solubilized GlpF as a function of increasing temperature (60°C/h) fit to a two state equilibrium model (see Materials and Methods Equation 1). $T_m = 65.5 \pm 0.8^\circ\text{C}$ and $\Delta T = 7.5^\circ\text{C}$.

Finally, fluorescence spectroscopy was also used to study the thermal sensitivity of the DDM-solubilized GlpF protein tertiary structure by monitoring Trp fluorescence. Figure 32 shows that elevated temperatures progressively quench the Trp fluorescence from a relative intensity of approximately 0.35 at 20°C to approximately 0.13 at 90°C. In addition, a small red shift from 329 to 333 nm is detected in the spectrum at temperatures of 70°C and above. This indicates that as the tertiary structure unfolds there is only a slight increase in accessibility of the Trp residues to water. This result is not unexpected; as the protein unfolds in detergent solution, there may not be much greater solvent (water) accessibility due to the interactions of the unfolded protein with the detergent

present in solution. Based on the XRD structure [1] the most likely candidate for the slight shift is Trp-42, located at the interface between the monomers.

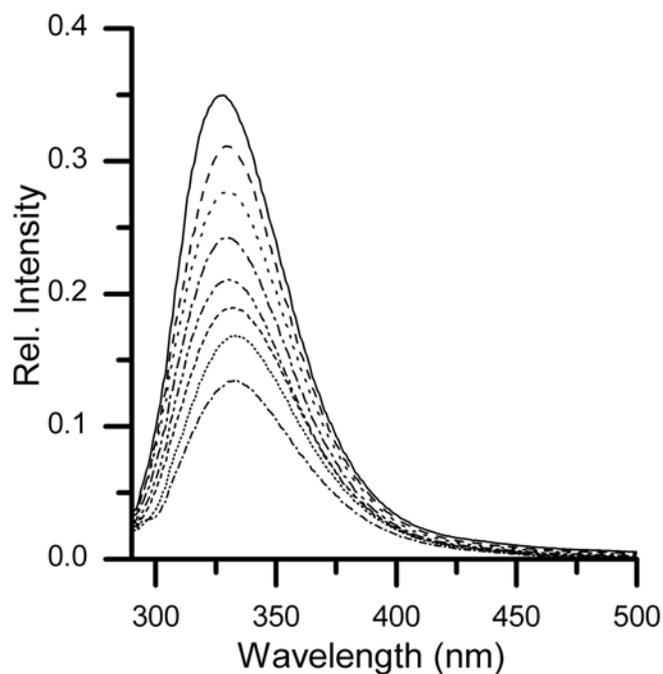


Figure 32: Thermal denaturation of DDM-solubilized GlpF (2.2 μ M) monitored by fluorescence spectroscopy ($\lambda_{ex} = 280$ nm).

Solid, 20°C; dash, 30°C; dot, 40°C; dash-dot, 50°C; dash-dot-dot, 60°C; short dash, 70°C; short dot, 80°C; short dash-dot, 90°C.

This red-shift in Trp fluorescence at and above 70°C is consistent with the thermal unfolding of GlpF tertiary structure when monitored by near UV CD analysis, which shows that the DDM-solubilized GlpF tertiary structure thermally denatures with a $T_m = 65.5 \pm 0.8^\circ\text{C}$.

In order to determine the effects of acidic pH on DDM-solubilized GlpF tertiary structure the protein was subjected to a pH titration and monitored by near UV CD spectroscopy (Figure 33). The spectrum at pH 7 indicates that the DDM-solubilized protein is well folded, with near UV CD intensity at 268 nm of approximately $-9000 \text{ deg cm}^2/\text{dmol}$, similar to that observed in Figure 30. The peak positions and intensities of the DDM-solubilized protein near UV CD spectra are retained as the pH is lowered to pH 6 and to pH 5, with little to no change. This result is in good agreement with that of the BN PAGE analysis of quaternary structure during acid unfolding (Figure 24 lane 3). Not until the pH is lowered to pH 4 does the intensity of the near UV CD signal diminish to approximately $-6500 \text{ deg cm}^2/\text{dmol}$ at 268 nm, indicating a loss of approximately 30 % of the tertiary structure of the DDM-solubilized GlpF protein. This result is also in good agreement with the BN PAGE analysis of the quaternary structure during acid unfolding of DDM-solubilized GlpF (Figure 24 lane 4) where, according to the gels the GlpF tetramer is starting to unfold. As the pH is lowered to three, an even greater loss of the near UV CD intensity is observed; the near UV CD intensity at 268 nm is approximately $-4000 \text{ deg cm}^2/\text{dmol}$ indicating a loss of approximately 60 % of the tertiary structure of the DDM-solubilized GlpF protein at pH 3. However, while the DDM-solubilized GlpF protein shows a complete loss of tetrameric structure at pH 3 as observed by BN PAGE (Figure 24 lane 5), the pH 3 near UV CD spectrum shows at least some tertiary structure is retained (Figure 33). The dissociation of the GlpF tetramer, as evident from the gels, combined with the retention of near UV CD intensity leads to the suggestion that at pH 3 the GlpF protein exists as a monomeric species in DDM solution, with at least some of its tertiary structure intact.

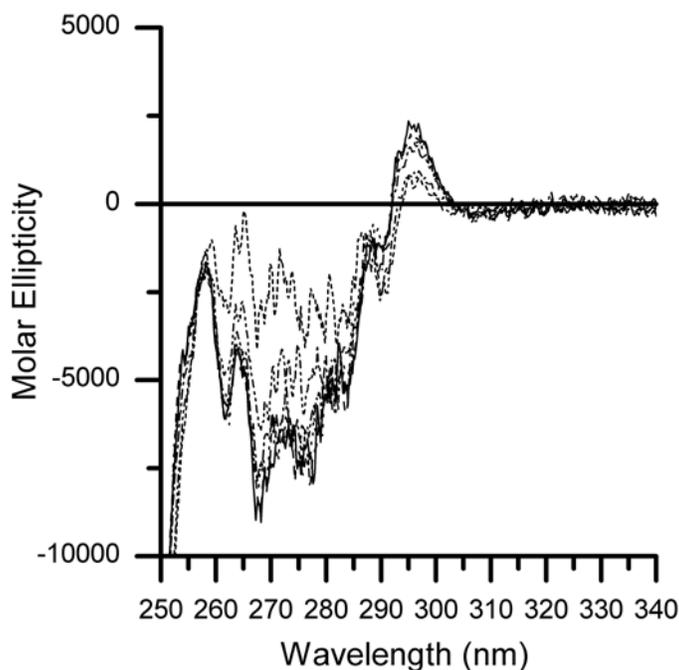


Figure 33: Spectroscopic analysis of the pH titration of DDM-solubilized GlpF (69.1 μM) tertiary structure monitored by near UV CD.

Solid, pH 7.0; dash, pH 6.0; dash-dot, pH 5.0; dash-dot-dot, pH 4.0; dot pH 3.0.

Fluorescence spectroscopy was also used to monitor the tertiary structure unfolding of DDM-solubilized GlpF by acidic pH and the results are shown in Figure 34. Between pH 7 and pH 5 the Trp emission is quenched by about 15% but there is no detectable change in the position of the emission maximum at 329 nm. However, between pH 5 and pH 4 the quantum yield is restored to slightly more than its original value and a shift in the peak maximum occurs by about 4 nm to the red. Between pH 4 and 3 the fluorescence is quenched by about 30% but the emission maximum remains at 333 nm.

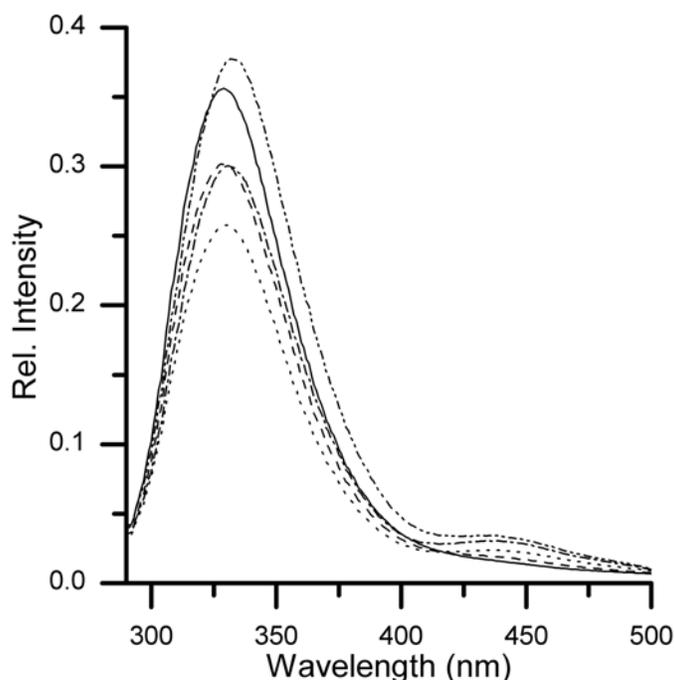


Figure 34: Acid induced unfolding of DDM-solubilized GlpF (2.2 μ M) monitored by fluorescence spectroscopy ($\lambda_{\text{ex}} = 280$ nm).

Solid, pH 7; dash, pH 6; dot, pH 5; dash-dot, pH 4; dash-dot-dot, pH 3.

These changes likely reflect direct effects of the protonation of different side-chains in the local environments of the five Trp residues as well as indirect effects of the unfolding of the tertiary structure and possible oligomerization as the pH is lowered. The shift of the emission maximum from 229 nm to 333 nm was also observed in the thermal unfolding experiments and therefore suggests that the protein begins to unfold below pH 5.0, which agrees with near UV CD, and SDS and BN gel analyses. Overall, the results are compatible with the concept that the pH 3 form of the protein in DDM contains little tertiary structure, and may exist as a monomer and/or high molecular weight oligomer.

To study the stability of the DDM-solubilized GlpF protein to chemical denaturants, the protein was subjected to a urea titration. Urea titration monitored by near UV CD spectroscopy of DDM-solubilized GlpF at pH 7 found that concentrations as high as 8 M urea had little effect on the tertiary structure of GlpF protein after incubation for 1 hour (Figure 35). This indicates that the protein's tertiary structure is highly stable at pH 7 in DDM solution, even at very high concentrations of chemical denaturants. However, a time dependence in the effects of urea was noticed. To investigate the urea unfolding kinetics the pH 7 DDM-solubilized GlpF in 8M urea was monitored over the course of three weeks (Figure 36). After exposure to 8M urea for 1 hour (Figure 36 solid line), the pH 7 DDM-solubilized GlpF near UV CD spectrum shows little difference from that of 0M urea (Figure 35 solid line), as was shown in Figure 35. The high stability of the DDM-solubilized GlpF protein to chemical denaturants further supports the conclusion that in DDM solution, GlpF exists as a well-folded, stable protein. After a period of 24 hrs however, the near UV CD intensity is somewhat diminished, indicating some loss of tertiary structure; the intensity at 268 nm is reduced from approximately $-9000 \text{ deg cm}^2/\text{dmol}$ after 1 hour (Figure 36 solid line) to less than $-5000 \text{ deg cm}^2/\text{dmol}$ after 24 hrs (Figure 36 dash line). Further observation of the pH 7 DDM-solubilized GlpF in 8M urea near UV CD signal intensity after 7 days (Figure 36 dot line) indicates greater loss of tertiary structure, as the intensity at 268 nm is further reduced to approximately $-2000 \text{ deg cm}^2/\text{dmol}$. After an incubation time of 21 days in 8M urea (Figure 36 dash-dot line), the pH 7 DDM-solubilized GlpF near UV CD intensity is almost completely eliminated with the intensity at 268 nm being approximately $-1000 \text{ deg cm}^2/\text{dmol}$.

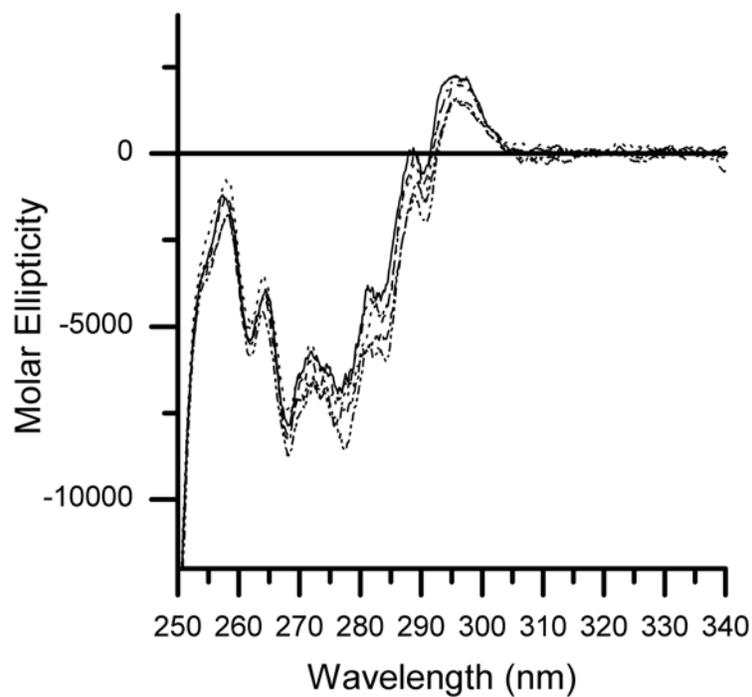


Figure 35: Spectroscopic analysis of the urea titration of pH 7 DDM-solubilized GIpF (55.4 μ M) tertiary structure monitored by near UV CD.

Solid, 0 M urea; dash, 2 M urea; dot, 4 M urea; dash-dot, 6 M urea; dash-dot-dot, 8 M urea.

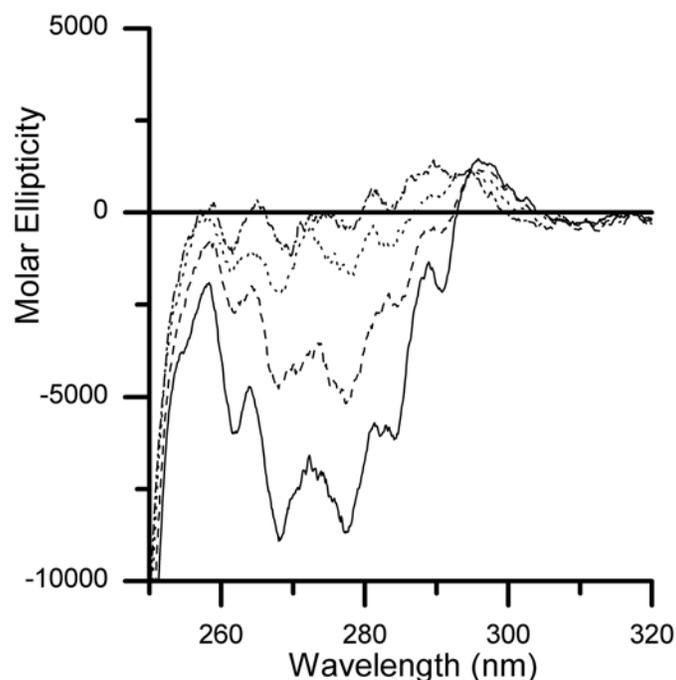


Figure 36: Spectroscopic analysis of the time dependence of 8 M urea on pH 7 DDM-solubilized GlpF (55.4 μM) tertiary structure by near UV CD.

Solid, 1 hr; dash, 24 hrs; dot, 7 days; dash dot, 21 days.

When the intensity at 268 nm is fitted to an exponential decay, (Figure 37) the added urea reduces the tertiary structure with an unfolding rate constant of $4 \pm 1 \times 10^{-6} \text{ s}^{-1}$. Although only four points were collected, it appears from Figure 37 that they may fit better to a two-component decay reflecting an early, fast unfolding event, and a later, slower unfolding. However, the small number of points collected would make the fit to biexponential decay statistically meaningless. Nevertheless, the high concentrations of urea required to unfold the protein and the slow unfolding rate both suggest a high thermodynamic stability of the present preparation.

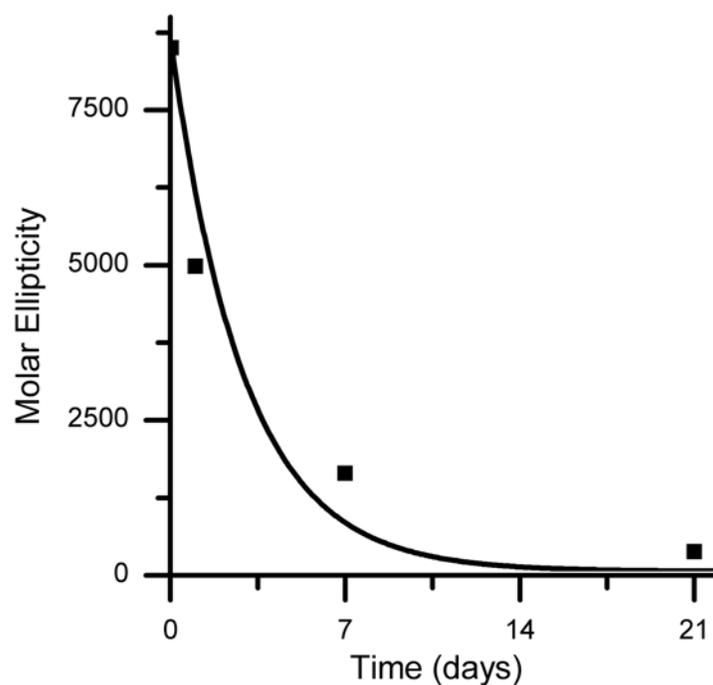


Figure 37: Molar ellipticity changes at 268 nm of DDM-solubilized GlpF (55.4 μM) in 8 M urea as a function of time (closed squares) fit to an exponential decay (Equation 2, Materials and Methods).

The exponential fit (solid line) of the molar ellipticity changes show an unfolding rate constant for DDM-solubilized GlpF in 8 M urea of $4 \pm 1 \times 10^{-6} \text{ s}^{-1}$.

3.2.2 Glycerol Facilitator Tertiary Structure in Lyso-Myristoyl Phosphatidylcholine Solution

When the tertiary structure of GlpF dissolved in LMPC at room temperature and neutral pH is studied using near UV CD spectral features characteristic of a well-folded protein with intense bands across the entire near UV range are observed (Figure 38). The 21 phenylalanines are the likely origin of the band at approximately 260 nm [180], and the

bands between 265 nm and 290 nm are attributable to the 5 tryptophan residues; the Trp 1L_a band is usually near 275 nm and the 1L_b doublet usually falls between 280 and 290 nm [150].

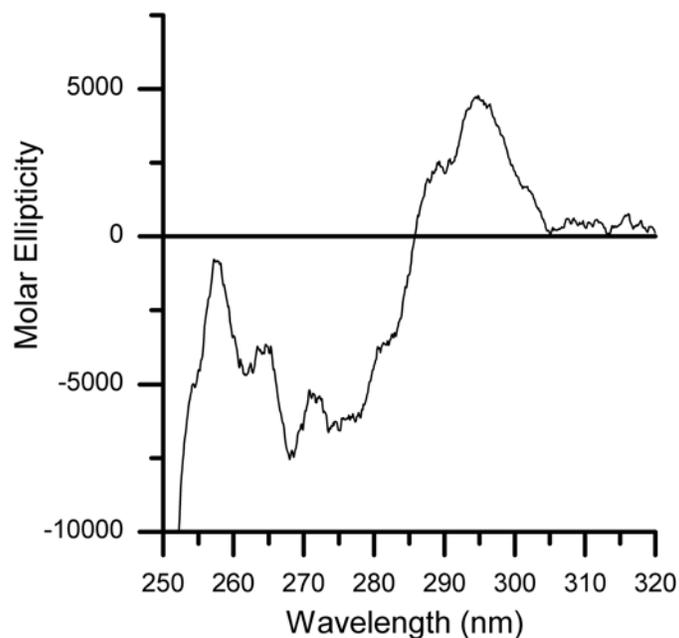


Figure 38: Near UV CD spectrum of LMPC-solubilized GlpF (25.0 μ M) at pH 7 and 25 $^{\circ}$ C.

When Figure 38 is compared to Figure 29, subtle differences in tertiary structure of the GlpF protein dissolved in different non-denaturing solutions are evident. While the overall shapes of the spectra and band positions are nearly identical some differences exist, especially around 290 nm. Figure 29 shows that in DDM-solution, the near UV CD intensity near 295 nm is approximately 2000 deg cm^2/dmol , while the same near UV CD region in LMPC-solution gives rise to an intensity of nearly 5000 deg cm^2/dmol . Conversely, the intensity at 268 nm (the region where tryptophans are active) is more

intense in DDM-solution (approximately $-10000 \text{ deg cm}^2/\text{dmol}$) than in LMPC-solution (approximately $-7500 \text{ deg cm}^2/\text{dmol}$). This suggests a subtle change in the local environment of the tryptophans. This may be interpreted as a decrease in absorption at 268 nm and an increase at 295 nm, or that the tryptophans' absorption in both regions is more positive in LMPC- than in DDM-solution leading to the less negative band at 268 nm, and the more positive band at 295 nm.

In order to study the stability of the GlpF protein tertiary structure in LMPC solution, a thermal denaturation of the protein was monitored using near UV CD spectroscopy. The thermal denaturation was monitored by following the intensity of the tryptophan near UV CD band at 268 nm, (Figure 39), as was done for GlpF in DDM solution. The non-linear least squares fit of the thermal denaturation in LMPC solution to a two-state unfolding transition using the program *Mathematica*TM 5.1 (see Materials and Methods Equation 1) is shown in Figure 39 (solid line) and determined the T_m to be $74.9 \pm 1.45^\circ\text{C}$ and $\Delta T = 9.2^\circ\text{C}$.

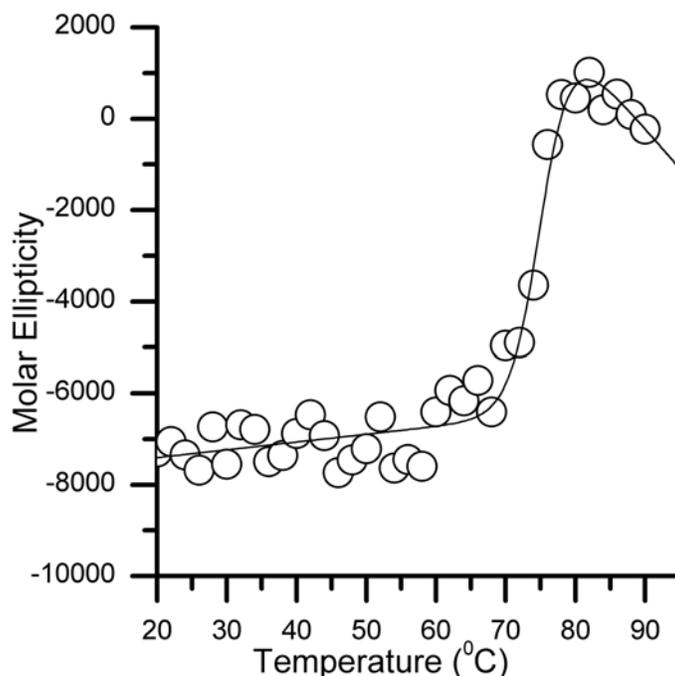


Figure 39: LMPC-solubilized GlpF molar ellipticity changes at 268 nm.

The molar ellipticity changes of GlpF (25.0 μ M) in LMPC solution (open circles) as a function of increasing temperature (60 $^{\circ}$ C/h) fit to a two state equilibrium model (solid line) (see Materials and Methods Equation 1). $T_m = 74.9 \pm 1.5^{\circ}$ C and $\Delta T = 9.2^{\circ}$ C.

This result shows that the GlpF tertiary structure has a much greater stability to thermal denaturation in LMPC-solution compared to DDM-solution; the mid-point of unfolding is 9.4 $^{\circ}$ C greater in LMPC-solution than for DDM-solubilized GlpF. In addition, inspection of the pre-unfolding regions of Figure 39 and Figure 31 shows that the slope of the pre-unfolding regions is significantly shallower in LMPC-solution than in DDM-solution, which may further indicate that the GlpF protein tertiary structure has a greater thermal stability in LMPC solution than when it is dissolved in DDM solution. These results agree very well with the DDM- and LMPC-solubilized GlpF quaternary structure thermal denaturations (Figure 23 and Figure 27 B respectively); both protein solutions were found

to be tetrameric up to 60°C however LMPC-solubilized GlpF retained a greater amount of the tetramer at a temperature of 80°C than did the DDM-solubilized protein. It is interesting to note that the post-denaturation pattern for each of the DDM- and LMPC-solubilized GlpF proteins is similar (Figure 31, and Figure 39 respectively); as the protein unfolds thermally, the near UV CD intensity approaches zero, but regains a limited amount of negative intensity at the highest temperatures. This observation is difficult to interpret as the protein is aggregating at these temperatures (observed as precipitate in solution). One interpretation is that as the GlpF protein thermally denatures, there is an interaction in the aggregation of the unfolded monomers leading to a slightly structured local environment around the aromatic residues.

Fluorescence spectroscopy was also used to study the thermal stability of the GlpF protein tertiary structure by monitoring Trp fluorescence. Figure 40 shows that elevated temperatures progressively quench the Trp fluorescence from a relative intensity of approximately 0.7 at 20°C to approximately 0.2 at 90°C. In addition, a small red shift from 329 to 333 nm is detected in the spectrum only at temperatures above 70°C. This indicates that as the tertiary structure unfolds there is only a slight increase in accessibility of the Trp residues to water. The most likely candidate is Trp-42, located at the interface between the monomers. This result is consistent with the near UV CD analysis which showed that the LMPC-solubilized GlpF tertiary structure thermally denatures with a $T_m = 74.9 \pm 9.2^\circ\text{C}$, as the fluorescence red shift occurs between 70°C and 80°C. In comparison to GlpF dissolved in DDM, this result also shows the greater thermal stability of the GlpF protein tertiary structure in LMPC-solution.

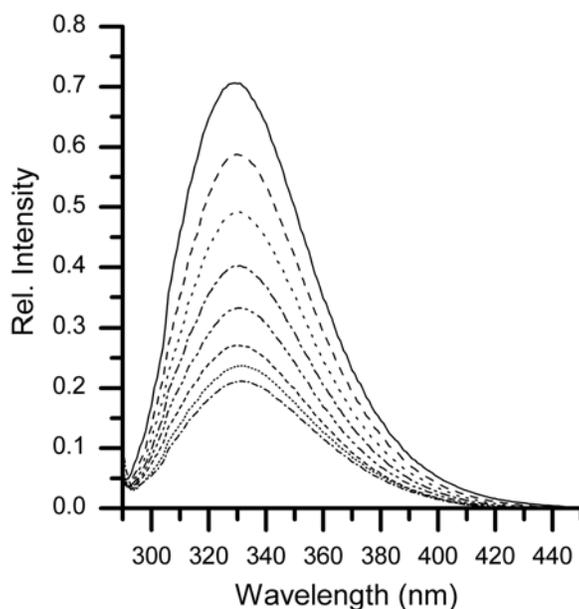


Figure 40: Thermal denaturation of LMPC-solubilized GlpF (3.4 μ M) monitored by fluorescence spectroscopy ($\lambda_{\text{ex}} = 280$ nm).

Solid, 20°C; dash, 30°C; dot, 40°C; dash-dot, 50°C; dash-dot-dot, 60°C; short dash, 70°C; short dot, 80°C; short dash-dot, 90°C.

In order to determine the effects of acidic pH on LMPC-solubilized GlpF tertiary structure the protein was subjected to a pH titration and monitored by near UV CD (Figure 41). During the pH titration in LMPC solution, there is little to no change to the band at 295 nm, which is most likely attributable to the 1L_b transitions of the 7 tyrosine residues, suggesting that this region of the protein's tertiary structure does not undergo major structural changes when subjected to acidic conditions. Alternatively, the signals from tyrosine may be less sensitive to the structural changes induced by the change in pH. In addition, the intensity of the near UV CD signal as a whole does not decrease only to zero but crosses over into positive ellipticity. This result suggests that when

dissolved in LMPC-solution, the GlpF protein's tertiary structure unfolds due to lowering the pH in a different manner than in DDM, which might be explained by the zwitterionic head-group of LMPC, compared to the neutral head-group of DDM. This is in good agreement with the quaternary structure analysis in LMPC and DDM solutions (Figure 28A and B, and Figure 21 and Figure 24 respectively), where the quaternary structure appears less sensitive at lower pH in DDM-solution, than in LMPC-solution.

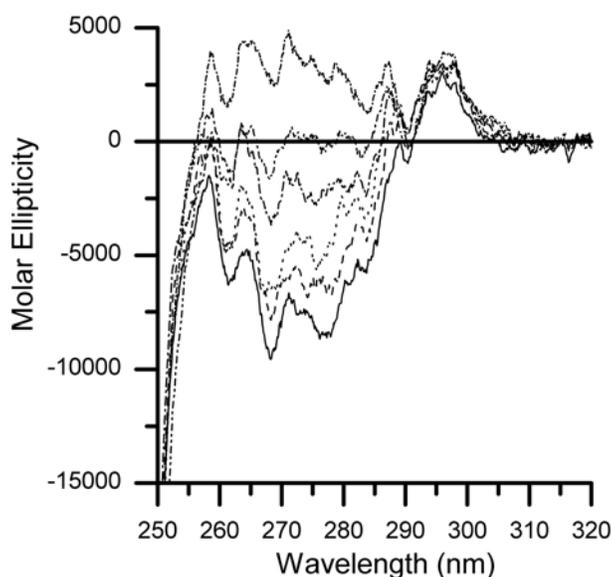


Figure 41: pH titration of LMPC-solubilized GlpF (13.7 μ M) tertiary structure monitored by near UV CD.

Solid, pH 7.6; dash, pH 7.0; dot, pH 6.0; dash-dot, pH 5.0; dash-dot-dot, pH 4.0; short dash-dot, pH 3.0.

The effects of lowering the pH on tertiary structure were also monitored using Trp fluorescence and are shown in Figure 42. Between pH 7 and pH 6 the Trp emission is quenched by about 10% and there is a red shift in the position of the emission maximum

by about 2 nm. Between pH 6 and pH 5 the quantum yield is further quenched by approximately 15% of the original emission, and another shift in the peak maximum occurs by about 2 nm to the red. Between pH 5 and 4 the fluorescence is quenched by about 20% but the emission maximum remains at 333 nm. The fluorescence at pH 3 is almost identical to that of pH 4, suggesting no additional changes to the tertiary structure when the pH is lowered from 4 to 3.

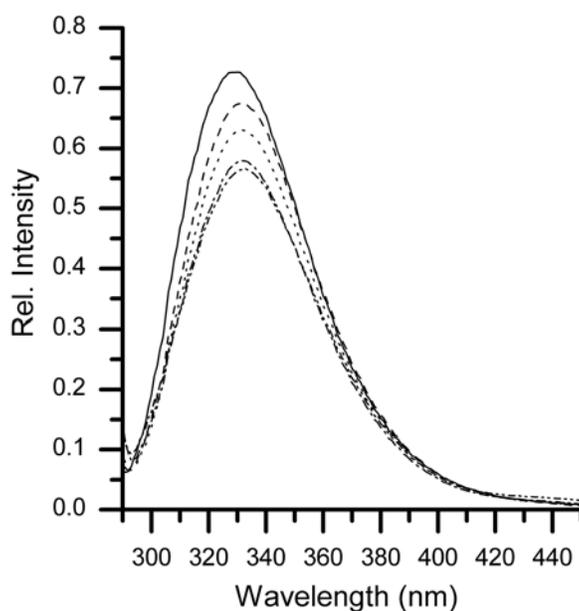


Figure 42: Figure 3-26: Acid induced denaturation of LMPC-solubilized GlpF (3.4 μ M) monitored by fluorescence spectroscopy ($\lambda_{\text{ex}} = 280$ nm).

Solid, pH 7; dash, pH 6; dot, pH 5; dash-dot, pH 4; dash-dot-dot, pH 3.

As was the case for the DDM-solubilized GlpF, these changes likely reflect direct effects of the protonation of different side-chains in the local environments of the five Trp residues as well as indirect effects of the unfolding of the tertiary structure and

oligomerization as the pH is lowered. The shift of the emission maximum between pH 7.0 and pH 6.0 suggests a significant unfolding event that confirms the sensitivity of the LMPC-dissolved GlpF to low pH as was shown by SDS-PAGE (Figure 28 A). The shift in λ_{max} to 333 nm below pH 6.0 was also observed in the thermal unfolding experiments and indicates that in LMPC-solution, the protein continues to unfold, which agrees with near UV CD (Figure 41), and the SDS and BN gel (Figure 28) analysis. Overall, the results are compatible with the concept that at neutral pH the LMPC-solubilized GlpF protein is well-folded, is more sensitive to acid than the DDM-solubilized protein, but that the low pH form of the protein in LMPC may contain residual tertiary structure that is different to that detected in DDM solution.

3.2.3 Glycerol Facilitator Tertiary Structure in Sodium Dodecyl Sulphate Solution

Figure 43 shows the near UV CD spectrum of the GlpF protein dissolved in SDS solution at pH 7.0. The SDS-solubilized GlpF shows only positive ellipticity throughout the near UV range with only poorly resolved bands. When compared to both the DDM- and LMPC-solubilized GlpF protein (Figure 29 and Figure 38 respectively) one immediately notices a large difference between the near UV CD spectra. It has been shown previously [181] that thermal changes induce no change in the near UV CD signal, which indicates that the SDS-solubilized GlpF protein exists in a state with few native tertiary structure contacts.

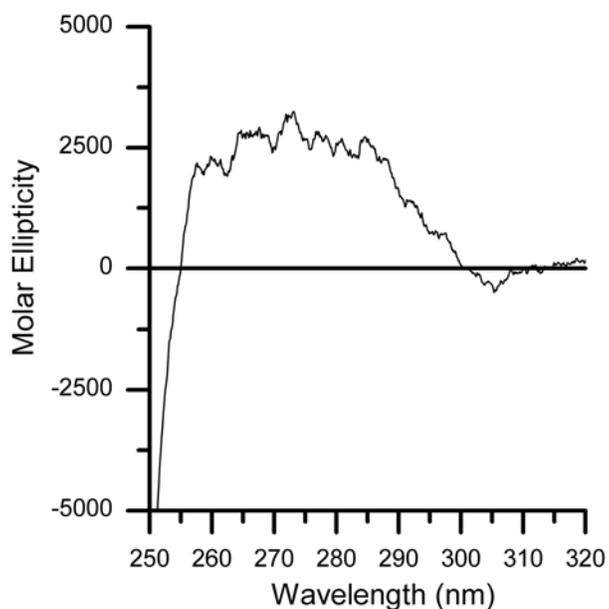


Figure 43: Near UV CD spectrum of imidazole-eluted, pH 7 SDS-solubilized GlpF (14.0 μM).

The near UV CD results suggest that in SDS detergent solution at neutral pH, the GlpF protein's tertiary structure is unfolded, and the protein resides in a state characterized by the positive ellipticity observed in the near UV region. It is interesting to note that the acid-induced structural changes in LMPC solution also result in positive ellipticity throughout the regions of the near UV CD spectrum (Figure 41).

3.2.4 Glycerol Facilitator Tertiary Structure in Octyl Glucoside Solution

Due to the low stability of the GlpF protein when dissolved in OG solution, the concentrations required for near UV CD were not attainable and thus, we were unable to perform tertiary structural analysis of OG-dissolved GlpF.

3.3 Secondary Structure Analysis of Detergent-Solubilized Glycerol Facilitator

In order to study the secondary structure of GlpF in the different detergent solutions far UV CD was used. CD in the far UV region is primarily due to the amide bond, and may be analyzed as originating from secondary structural components of the protein (see section 1.5).

3.3.1 Glycerol Facilitator Secondary Structure in Dodecyl Maltoside Solution

The far UV CD spectrum of DDM-solubilized GlpF at 20°C pH 7, (see Figure 44), displays the characteristics of a predominantly α -helical protein with negative bands at 219-222 nm and 209-211 nm, a positive band at 193-194 nm, and a cross-over point at 200-201 nm. The ideal α -helical values are 222, 208, 190-195 and 205 nm respectively. Deconvolution of the far UV CD spectra measured at 20°C using the CDSSTR algorithm

[149] accessed through the DichroWeb website [148] yields 53% α -helix, 17% β -strand, 15% turns, and 15% unordered for DDM-solubilized GlpF. On the assumption that the 35-residue His₆-T7 tag is disordered, the deconvolution indicates that in DDM solution 60% of the GlpF residues are helical. This is in close agreement with the secondary structure of the tetrameric GlpF observed by X-ray diffraction [1] where 64% of the residues are α -helical, 3% are 3_{10} helical, and the balance are in turns or irregular structure.

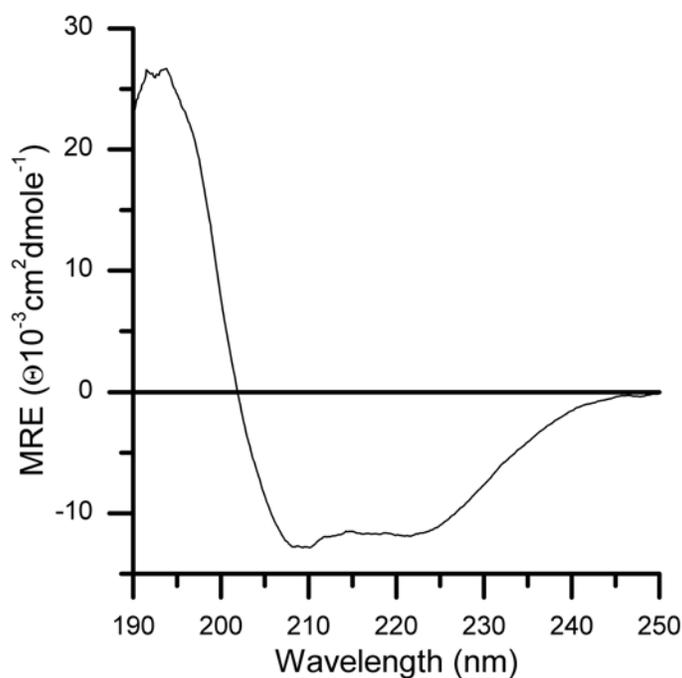


Figure 44: Far UV CD spectrum of DDM-solubilized GlpF (3.0 μ M) at pH 7, and 25 °C.

In order to measure the stability of the GlpF secondary structure in pH 7 DDM solution, the DDM-solubilized GlpF was subjected to a thermal denaturation. Figure 45 shows

that as the temperature is raised from 20°C to 95°C the far UV CD intensity diminishes, indicating a loss of secondary structure. The dashed line at 65°C indicates the start of the unfolding transition. The unfolding is continued at 70°C (dotted line), and is nearly complete at temperatures above 70°C.

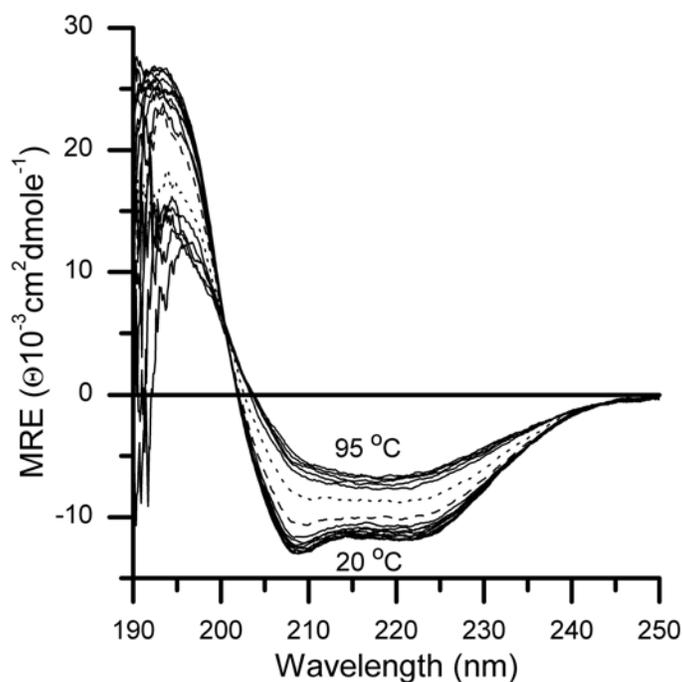


Figure 45: Thermal denaturation of DDM-solubilized GlpF (3.0 μ M) secondary structure monitored by far-UV CD.

The dashed and dotted lines indicate the spectra at 65 °C and 70 °C respectively.

A plot of the temperature dependence of the ellipticity at 209 nm fit to a 2-state unfolding transition is shown in Figure 46. The plot indicates that the DDM-solubilized protein secondary structure is stable up to 60°C following which increasing temperature reduces the fractional helicity, as deduced from far UV CD, from 60% to 20% with a T_m of

69.7±0.7°C and ΔT of 9.8°C. When the scan rate and response time are adjusted (from 10 nm/min and 16 seconds to 20 nm/min and 8 seconds) the rate of heating was increased from 0.43 °C/min to 1 °C/min. This results in an increase of the T_m to 75±1.1 °C (data not shown). Recall that the DDM GlpF near UV CD ellipticity is relatively unperturbed up to 60°C (Figure 30 and Figure 31) and is eliminated with a T_m (71±2°C) that is only slightly lower than that measured by far UV CD. This indicates that the thermal transition involves a cooperative loss of both secondary and tertiary structure above 60°C in DDM solution, which is in excellent agreement with the SDS-PAGE and BN-PAGE analysis, where quaternary structure is not lost until temperatures above 60°C (Figure 20 B, and Figure 23 respectively).

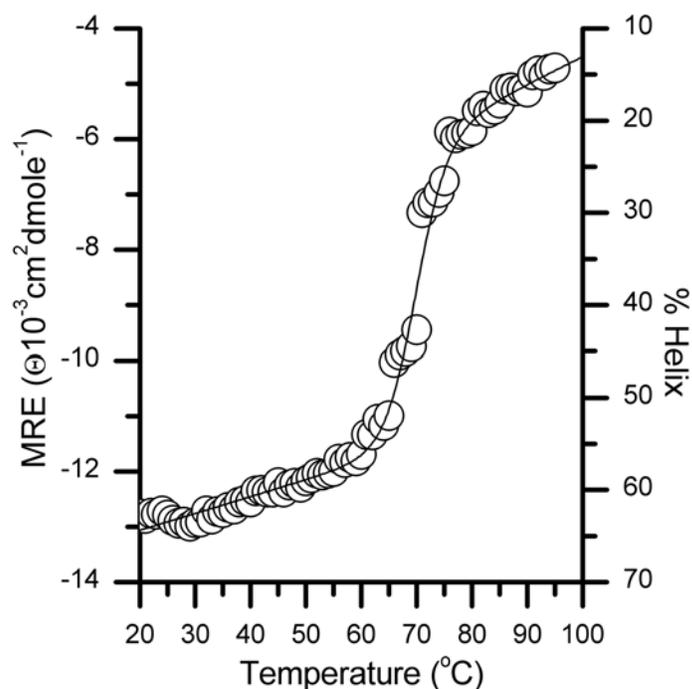


Figure 46: DDM solubilized GlpF (3.0 μ M) mean residue ellipticity ($\text{deg cm}^2/\text{dmole}$) and fractional helix changes at 209 nm as a function of increasing temperature (60°C/h) fit to a two-state equilibrium model.

$T_m = 69.7 \pm 0.7^\circ\text{C}$ and $\Delta T = 9.8^\circ\text{C}$. The % helix was determined using the CDSSTR algorithm [149] accessed through the DichroWeb website [148].

In order to determine the effects of acidic pH on DDM-solubilized GlpF secondary structure, the protein was subjected to a pH titration and monitored by far UV CD. The results in Figure 47 show that the spectra change very little over the pH range studied. The helix content of DDM-solubilized GlpF protein at low pH is only slightly lower than that of the neutral pH GlpF tetramer. Thus, the secondary structure of the protein is remarkably insensitive to pH and the low pH form of the protein in DDM solution contains native secondary structure, little tertiary structure (Figure 33), and may exist as a high molecular weight oligomer (Figure 21 and Figure 24).

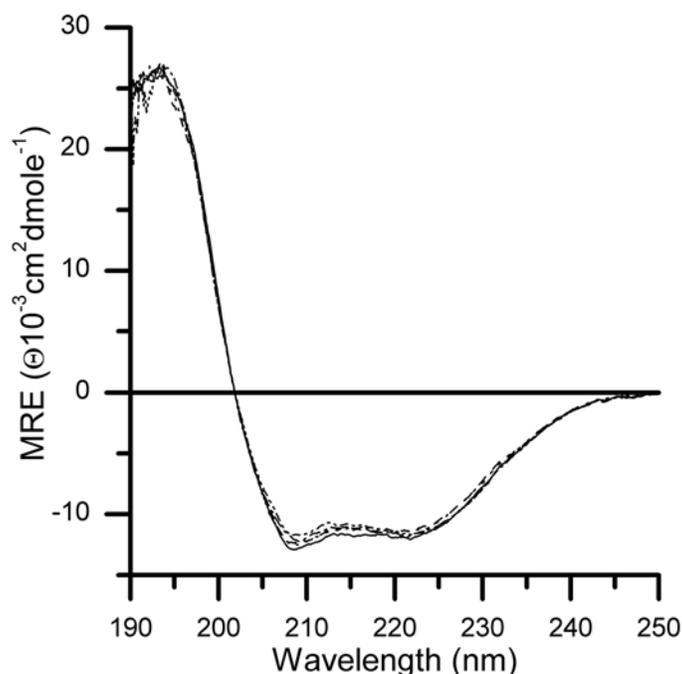


Figure 47: pH titration of DDM-solubilized GlpF (3.0 μ M) secondary structure monitored by far-UV CD.

Solid, pH 7.0; dash, pH 6.0; dot, pH 5.0; dash-dot, pH 4.0; dash-dot-dot pH 3.0.

3.3.2 Glycerol Facilitator Secondary Structure in Lyso-Myristoyl Phosphatidylcholine Solution

The far UV CD spectrum of LMPC-solubilized GlpF at 20°C, (see Figure 48 solid line), also displays the characteristic features of an α -helical protein with negative bands at 219-222 nm and 207-209 nm, a positive band at 192-193 nm, and a cross-over point at 201-202 nm. The ideal α -helical values are 222, 208, 190-195 and 205 nm respectively. Deconvolution of the LMPC-solubilized far UV CD spectra measured at 20°C using the CDSSTR algorithm [149] yields 53% α -helix, 10% β -strand, 15% turns, and 22%

unordered for LMPC-solubilized GlpF. Assuming that the 35-residue His₆-T7 tag is disordered the deconvolution indicates that in LMPC solution 60% of the GlpF residues are helical. This agrees closely with the secondary structure observed by X-ray diffraction [1] for the tetrameric protein.

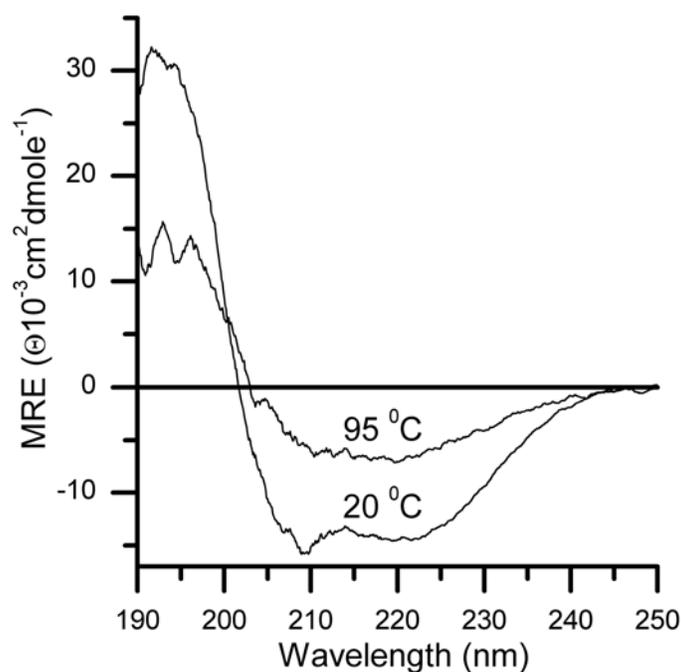


Figure 48: Thermal unfolding of LMPC-solubilized GlpF (1.5 μ M) secondary structure monitored by far-UV CD.

In order to measure the stability of GlpF prepared in pH 7 LMPC, the effects of temperature were examined. The thermal unfolding of secondary structure was followed by far UV CD spectropolarimetry and the spectra at 20°C and 95°C are shown in Figure 48. Plots of the temperature dependence of the ellipticity at 209 nm fit to a 2-state unfolding transition are shown in Figure 49. The temperature dependence of the thermal

unfolding indicates that the LMPC-solubilized GlpF protein secondary structure is stable up to 70°C following which increasing temperature reduces the fractional helicity as deduced from far UV CD from 60% to 18% with a T_m of $86.9 \pm 11.5^\circ\text{C}$ and ΔT of 15.1°C . Figure 39 shows that in LMPC solution the GlpF tertiary structure is also relatively unperturbed up to 70°C (as determined by near UV CD ellipticity) and is eliminated with a T_m of $74.9 \pm 1.45^\circ\text{C}$ and ΔT of 9.2°C , which are only slightly lower than those measured by far UV CD. This indicates that the thermal transition involves a cooperative loss of both secondary and tertiary structure above 70 °C in LMPC.

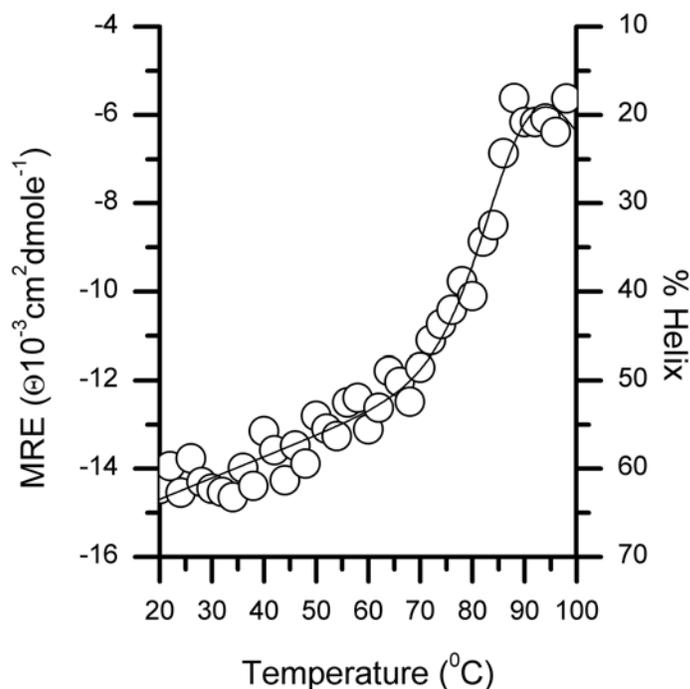


Figure 49: LMPC-solubilized GlpF (1.5 μM) mean residue ellipticity ($\text{deg cm}^2/\text{dmole}$) and fractional helix changes at 209 nm as a function of increasing temperature ($60^\circ\text{C}/\text{h}$) fit to a two-state equilibrium model.

$T_m = 86.9 \pm 11.5^\circ\text{C}$ and $\Delta T = 15.1^\circ\text{C}$. The % helix was determined using CDSSTR algorithm [149] accessed through the DichroWeb website [148].

In order to determine the effects of acidic pH on LMPC-solubilized GlpF secondary structure the protein was subjected to a pH titration and monitored by far UV CD and the results are shown in Figure 50. The pH titration shows that in LMPC solution the spectra change very little over the pH range studied, as was also observed for the protein in DDM solution. The far UV CD spectrum of LMPC-solubilized GlpF protein at low pH is virtually identical to that of the neutral pH GlpF tetramer. This result is compatible with the concept that the low pH form of the protein in LMPC solution contains native secondary structure (Figure 50), little tertiary structure (Figure 41), and may exist as a high molecular weight oligomer (Figure 28).

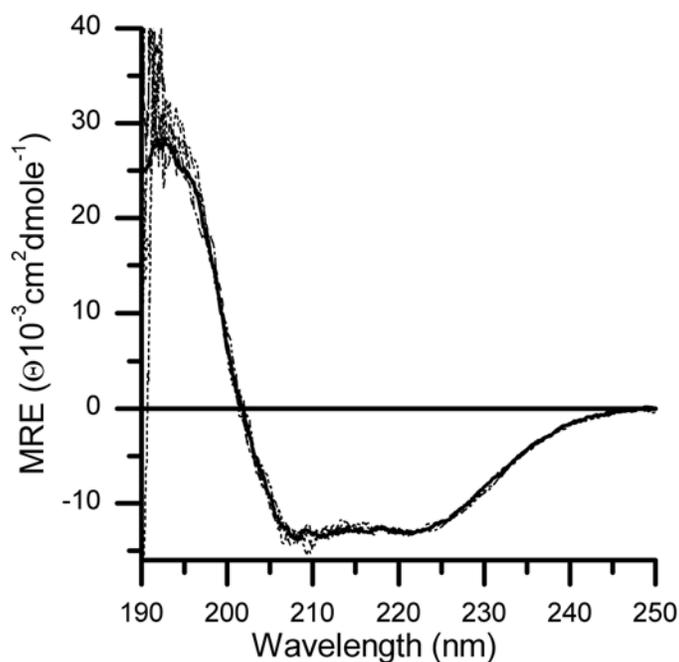


Figure 50: pH titration of LMPC-solubilized GlpF (1.5 μ M) secondary structure monitored by far-UV CD.

Solid, pH 7.0; dash, pH 6.0; dot, pH 5.0; dash-dot, pH 4.0; dash-dot-dot pH 3.0.

3.3.3 Glycerol Facilitator Secondary Structure in Sodium Dodecyl Sulphate Solution

The far UV CD spectrum of SDS-solubilized GlpF eluted from the column *via* imidazole competition at 20°C pH 7, (see Figure 51), displays the characteristics of an α -helical protein with negative bands at 219-222 nm and 209-211 nm, a positive band at 193-194 nm, and a cross-over point at 202-203 nm. The ideal α -helical values are 222, 208, 190-

195 and 205 nm respectively. Deconvolution of the far UV CD spectrum measured at 20°C using the CDSSTR algorithm [149] accessed through the DichroWeb website [148] yields 18 % α -helix, 32 % β -strand, 22 % turns, and 27 % unordered for SDS-solubilized GlpF. On the assumption that the 35-residue His₆-T7 tag is disordered the deconvolution indicates that in SDS solution only 20 % of the GlpF residues are helical. This differs significantly from previous preparations of SDS-solubilized GlpF in our lab [181]. A possible explanation for the discrepancy is that the use of imidazole to remove the protein from the column in the presence of SDS results in impurities in the preparation as detected in the SDS-PAGE electrophoregrams. As previously discussed, the secondary structure observed by X-ray diffraction [1] indicates that 61 % of the residues are α -helical.

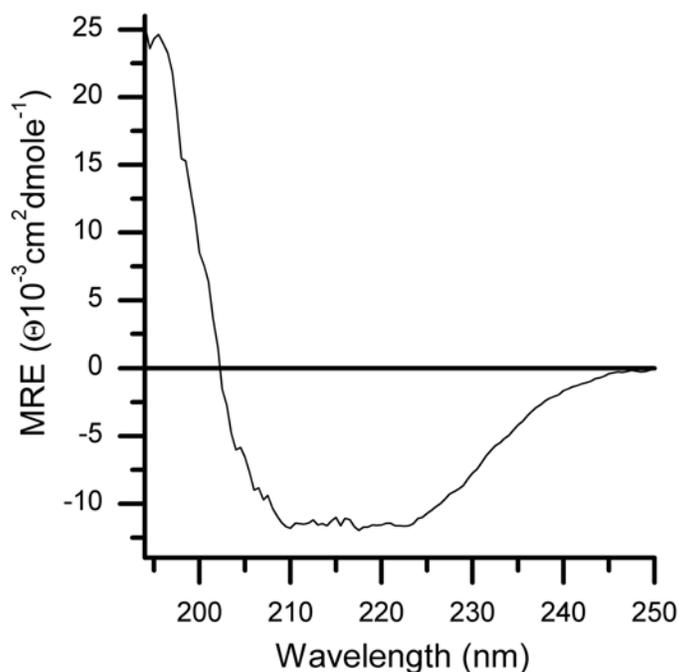


Figure 51: Far UV CD spectrum of SDS-solubilized GlpF protein (1.5 μM) at pH 7.

This result suggests that in the denaturing SDS solution, GlpF is unable to achieve its proper secondary fold. However, one third of the residues retain an α -helical conformation even in the presence of high concentrations of SDS in the absence of folding-promoting detergents. Combined with the results of Figure 43 and Figure 18, the GlpF protein dissolved in SDS-solution at pH 7, shows evidence of being only partially folded at the levels of quaternary (Figure 18), tertiary (Figure 43) and secondary structure (Figure 51).

Chapter 4 Discussion

4.1 Glycerol Facilitator in Dodecyl Maltoside and Lyso-Myristoyl Phosphatidylcholine Solutions at Neutral pH

In all previous attempts to prepare detergent solutions of glycerol facilitator the protein appeared predominantly monomeric by several assays, including velocity sedimentation analysis [129, 132-134], freeze-fracture particle size analysis [130], and SDS-PAGE/Western Blots [129, 130, 132-134, 141] although SDS-PAGE sometimes showed small amounts of dimer, trimer, tetramer, and higher molecular weight species [129, 141]. These observations have led to the suggestion that, in contrast to the aquaporins, the glycerol facilitators exist as monomers or weakly associating tetramers in membranes [129, 130, 132-134, 141]. In solution-state studies that suggest the GlpF protein has only weak homo-oligomeric associations, the GlpF protein is found to exist predominantly as a monomer, or a mixture of oligomers when observed using SDS-PAGE analysis [129]. One common feature of each of these reports is that the GlpF protein was solubilized in OG detergent. A possible explanation is that the OG detergent destabilizes the folded conformation of GlpF leading to the formation of non-specific aggregates of unfolded protein. This is in agreement with multiple observations of the sensitivity of membrane protein conformation to the properties of the detergent/detergents used for their solubilizations [81, 142, 182, 183]. For example, the Ca^{2+} -ATPase is an integral membrane protein responsible for the coupling of energy production and transport of Ca^{2+} ions across the membranes of muscle cells. Ca^{2+} -ATPase has been shown to have a

variety of activity levels that are dependent on lipid chain length, with a maximum activity at a chain length of 16 carbons [184].

OG has a molecular dynamics calculated monomer length of 14.6 – 14.8 Å, with the hydrocarbon tail having an average length of 8.2 – 8.3 Å [185]. X-ray diffraction [1] and molecular dynamics calculations [186] estimate that the hydrophobic thickness of the GlpF protein in a lipid bilayer is approximately 25 Å, which is similar to that of other helical proteins such as the Ca²⁺-ATPase from muscle sarcoplasmic reticulum [186, 187]. The OG molecular dynamics simulations [185] also show that each carbon in the detergent tail contributes approximately one Å to the length of the detergent. Since the hydrophobic region of GlpF is approximately 25 Å in length [186], two molecules of OG would only interact with approximately 64% of the 25 Å long hydrophobic surface of GlpF. However, the GlpF hydrophobic region will be well matched by two lipid/detergent carbon tails of approximately 12-13 atoms. This is in excellent agreement with our results, which show a marked increase in the stability of the GlpF protein's quaternary, tertiary and secondary structures in detergent solutions with 12 (DDM) and 14 (LMPC) carbon atoms in the detergent tail whereas the protein is unstable in the 8-carbon detergent OG. The molecular dynamics simulations [185] suggest that OG would be a poor detergent for the GlpF protein, as the shorter carbon chains (8.2 Å) would be unable to fully match the 25 Å hydrophobic region of the GlpF protein. This inability to match the hydrophobic region likely contributes to the underlying explanation for why many membrane proteins are successfully crystallized from shorter detergents such as

octyl- and nonyl glucoside [1, 123, 188-191]; partial instability in detergent solution will promote protein-protein interactions, resulting in crystallization.

Based on the hydrophobic mismatching of GlpF and OG, it is not surprising that the GlpF protein exists in OG as a mixture of oligomers. Our results on the oligomerization state of the GlpF protein when solubilized in OG-solution (Figure 3-3) show that the protein exists as a mixture of oligomers, with the monomeric and dimeric species being of similar abundance, which is in excellent agreement with what is observed in the literature [129-134, 141]. The results presented here show for the first time that DDM- and LMPC-solubilized recombinant GlpF prepared by elution with imidazole from an immobilized Ni²⁺ resin at pH 7.6 exists as a tetramer by electrophoresis on Blue Native polyacrylamide gels (Figure 23 and Figure 27B). Even under very harsh conditions, such as those encountered during SDS PAGE (Figure 20B lane 1 and Figure 27A lane 4 respectively), the GlpF protein remains predominantly tetrameric which suggests that the GlpF protein is not inherently weakly associated. Thus, given the proper conditions, GlpF can exist as a stable tetramer in solution, as well as in the crystal-state [1, 138]. The BN gels indicate that GlpF monomer, dimer and trimer are virtually undetectable whereas the SDS gels reveal small amounts of the alternate species. This suggests that the SDS present during electrophoresis on SDS-PAGE can cause a small amount of dissociation of the tetramer and that extraction and purification of the protein at neutral pH in DDM and LMPC yields a tetramer that is highly resistant to dissociation even by SDS. The stability of the oligomeric state of GlpF to SDS-detergent is then similar to the properties of many membrane proteins, including the aquaporins [129, 130, 133, 134, 142, 192].

In addition to the observation of tetramer, an octameric species is observed in both DDM- and LMPC-solutions using both SDS and BN electrophoresis techniques. An octameric species of GlpF is also found in electron microscopy studies of two-dimensional crystal arrays of the protein, where the unit cell was found to consist of an octamer comprised of two tetramers in side-on association [137]. The only other member of the AQPs to have shown association between tetramers is AQP0, where a 7.0 Å crystal structure shows that the AQP0 tetramers are associated in a head-to-head fashion, along their four-fold axes [177]. AQP0 is also a GLP, as it has been demonstrated to transport both water [193] and glycerol [194], and is the most abundant protein in the plasma membrane of the eye lens [195]. The observed association of the AQP0 tetramers [177] supports the possibility that AQP0 is involved in the extensive intercellular coupling of the eye lens [196]. In GlpF, the observation of a tetramer supports a “*coupling*” of the GlpF to the GlpK (which is active as both a dimer and tetramer) as a mechanism to channel glycerol into the cell, which could activate the GlpK [139] to phosphorylate the glycerol, thus preventing diffusion of the glycerol back across the membrane. A 1:1 association between the two protein tetramers would provide the most efficient pathway for transferring glycerol from the facilitator to the kinase preventing loss from the cytoplasm of membrane-permeable glycerol [1, 197]. This coupling of sugar permeation to kinase-catalyzed sugar phosphorylation and control of transport is known to occur for both glucose and fructose sugars in yeast [198, 199].

CD and fluorescence spectroscopies were used to characterize the secondary and tertiary structures of the glycerol facilitator protein prepared in neutral DDM- and LMPC-solutions. The far UV CD spectra of DDM- and LMPC-solubilized GlpF at 20°C pH 7 (see Figures 3-28 and 3-32), display the characteristics of a predominantly α -helical protein with minima, crossovers, and maxima similar to ideal α -helices [200].

Deconvolution of the CD spectra measured at 20°C using the CDSSTR algorithm [149] yields 53% α -helix for DDM-solubilized GlpF, and 53% α -helix for LMPC-solubilized GlpF. On the assumption that the 35-residue His₆-T7 tag is disordered the deconvolution indicates that in DDM and LMPC solution 60% of the GlpF residues are helical. This is in excellent agreement with the secondary structure observed by X-ray diffraction where 64% of the residues are α -helical, 3% are 3_{10} helical, and the balance are in turns or irregular structure [1]. These results also support the observation of increased stability of the GlpF protein in detergent solutions whose hydrocarbon chains are well matched to the hydrophobic region of the protein as these measurements were not possible in the shorter, eight carbon tail OG.

The 90°C DDM-solubilized GlpF fluorescence spectrum shown in Figure 32 (short dash-dot) has a small shoulder at 300 nm and a peak at 329 nm. The shoulder at 300 nm could arise from very weak emission from the tyrosine residues. In folded globular proteins tyrosine emission is often highly quenched by radiationless energy transfer to tryptophan as well as by interactions with other protein groups [155]. However, the peak may be due to scattering from the aggregation and precipitation of the thermally denatured GlpF from DDM solution. This shoulder is not observable in the LMPC-solubilized fluorescence

spectra shown in Figure 40. The emission maximum for tryptophan in proteins is sensitive to both the polarity and the dynamics of the environment surrounding the side-chain, and is blue-shifted in environments of low polarity such as the hydrophobic interior of a protein or in a detergent micelle. Burstein and colleagues have classified tryptophan fluorescence in proteins into 5 classes depending on the polarity and dynamics of the tryptophan environments [162]. According to their scheme, the dominant fluorescence at 329 nm in GlpF corresponds closely to a class of tryptophan side chains that are in a relatively non-polar environment and H-bonded in a 2:1 exciplex that fluoresces at about 331 nm [162]. Inspection of the 3D structure of GlpF (shown in Figure 14) [1] indicates that Tryptophan-42 is well buried in the intermonomer interface and Tryptophan-219 is buried but close to the surface of the protein. Two tryptophan side-chains (76 and 215) project from the lateral surfaces of the protein and likely contact detergent but might also interact with water. In the crystal structure, Tryptophan-48 forms part of the lining of the glycerol channel [1] and appears as if it would be highly exposed to bulk water in a membrane or in a detergent micelle. Emission from water-exposed tryptophan residues is usually observed at 350 nm [162] but the fluorescence of tryptophan-48 is not distinguishable from the major band at 329 nm in Figures 3-16 or 3-24. This may indicate that its fluorescence is quenched or that it interacts with detergent molecules that shift its fluorescence to lower wavelengths.

In summary, the DDM- and LMPC-solubilized GlpF protein fluorescence results are compatible with the observations of native secondary, tertiary, and quaternary structure described above and with the 3D structure determined by X-ray diffraction [1].

Near-UV CD spectra of glycerol facilitator tetramer solubilized in DDM and LMPC at neutral pH and 20°C are shown in Figure 29 and Figure 38 respectively. These spectra are usually interpreted as indicating the presence of tertiary structure in proteins as only aromatic residues held rigidly in an asymmetric environment will display near UV CD bands [150]. The observation of well-defined CD absorbance bands suggests that the aromatic residues are fixed firmly in the structure in asymmetric environments and is a strong indication of a stable tertiary fold.

4.2 Effects of Heat on α -Helical Membrane Proteins

The scarcity of membrane protein thermodynamic data in the literature reflects the difficulties and challenges associated with obtaining accurate estimates of the energetic parameters characterizing the thermal unfolding of membrane proteins. For a review see [142]. The results of compiled unfolding data suggest that a fundamental difference exists between the unfolding of soluble and membrane proteins [201]. In particular, the intramembranous secondary structural core of membrane proteins appears to be highly stable to thermal denaturation, while the extramembranous loops behave similarly to soluble proteins under thermal stress. Although many water-soluble proteins are now known to undergo incomplete thermal and chemical unfolding they contain only residual structure in their unfolded states (for example see [202, 203]). In contrast, the observed

reduced unfolding enthalpies measured for membrane proteins compared to soluble proteins suggest an incomplete thermal unfolding in which the core of the protein remains intact at elevated temperatures [201].

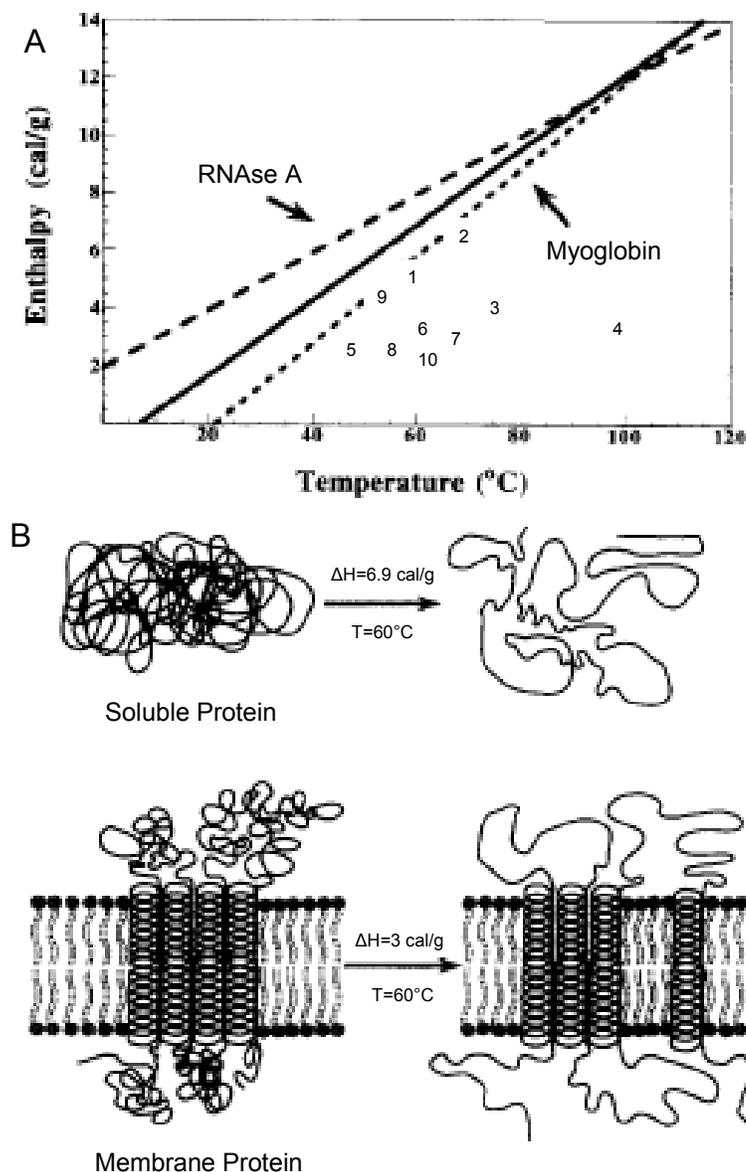
Bacteriorhodopsin is the prototypical seven transmembrane α -helical membrane protein with a stable monomer as the active unit, and trimeric oligomerization in the membrane. Bacteriorhodopsin is the sole protein component in *Halobacterium salinarum* purple membrane, and in fact, the purple membrane consists of a two-dimensional crystalline array of bacteriorhodopsin trimers [204]. The conformational stability of bacteriorhodopsin has been studied calorimetrically in native purple-membranes [205-207], and in lipid-detergent micelles at neutral pH [208], and is characterized by a high temperature thermal transition in purple membranes ($T_m \sim 100^\circ\text{C}$) and a low unfolding enthalpy ($\Delta H \sim 110$ kcal/mol, or 3.7 cal/g) compared to the average value of ~ 12 cal/g for soluble proteins at this temperature [201]. The thermal transition of monomeric bacteriorhodopsin in lipid-detergent micelles is significantly lower ($T_m \sim 65^\circ\text{C}$), but remarkably the thermal unfolding enthalpy is similar to that measured for the trimer [208, 209]. Subsequent studies revealed irreversibility to the thermal unfolding and scan-rate dependence for the differential scanning calorimetry endotherms [210]. Although the thermodynamics analysis is compromised it is worth noting that the magnitude of the unfolding enthalpy is significantly lower than that of the soluble proteins.

Cytochrome-*c* oxidase is a multi-subunit α -helical membrane protein which has been extensively studied thermodynamically. The results of several investigations show that

the magnitudes of the unfolding enthalpies vary with *in vitro* reconstitution conditions [211-217]. For example, beef oxidase reconstituted in dimyristoyl phosphatidylcholine exhibits a biphasic thermal transition ($T_m = 52$ and 64°C) with the lower temperature transition involving the thermal denaturation of subunit III and a corresponding enthalpy of 4.8 cal/g, while subunits I and III unfold by a complex path at higher temperatures [214]. When reconstituted into detergent, the biphasic denaturation simplifies to a broad single transition with a mid-point of unfolding of $\sim 56^\circ\text{C}$ and an enthalpy of ~ 2.7 cal/g, which is significantly lower than the $\Delta H \sim 6.9$ cal/g expected for soluble proteins at this temperature [214]. Yeast cytochrome oxidase thermally unfolds with an enthalpy of ~ 2.4 cal/g when reconstituted in lipids, and exhibits a similar biphasic thermal transition to that of beef oxidase under low ionic strength [212].

Photosystem II is a multidomain α -helical membrane protein assembly, which has also been thermodynamically analyzed in order to determine its stability [218, 219], and although it has a very complex thermogram with many transitions (due to the many subunits) it is reported to unfold with an enthalpy of $\sim 5-6$ cal/g [219]. Here too, the thermal unfolding is extremely sensitive to *in vitro* conditions [219], which is corroborated by studies suggesting that the ratio and nature of the surfactants play a critical role in the stability of membrane proteins [182].

Based on the above examples (and others), Haltia and Freire [201] proposed a possible explanation for the observed low enthalpies and high thermal stabilities of helical membrane proteins when compared to soluble proteins (see Figure 52).



[98].

Figure 52: Energetics and mechanism of membrane protein unfolding.

(a) Comparison of the calorimetric enthalpies (cal/g) of select membrane proteins with an average soluble protein (solid line), with a protein having a highly hydrophobic core (myoglobin, dotted-line) and with a protein having a highly hydrophilic core (RNase A, dashed-line). Membrane proteins: Photosystem II (1) [219]; the band 3 protein (2) [220, 221]; monomeric bacteriorhodopsin (3) [208]; trimeric bacteriorhodopsin in the purple membrane (4) [205-207]; *Paracoccus denitrificans* cytochrome-*c* oxidase subunit III (5) [211]; *Paracoccus denitrificans* cytochrome-*c* oxidase wild-type (weighted average of two transition temperatures) (6) [211]; *Paracoccus denitrificans* cytochrome oxidase purified from the subunit III-minus mutant (7) [211]; bovine cytochrome-*c* oxidase (8) [214]; bovine cytochrome-*c* oxidase subunit III (9) [214]; cytochrome-*c* oxidase from yeast (10) [212]. Note the low enthalpies of denaturation of most membrane proteins. (b) A schematic model of the thermal denaturation of a soluble and an α -helical membrane protein. (Reproduced with modifications and permission from [201]).

It was suggested that the likely explanation for the low enthalpy changes and high thermal stability of membrane proteins is that, unlike their soluble protein counterparts, they do not completely unfold under the thermal stress. Only loop and extramembranous regions unfold, whereas the transmembrane α -helices remain intact, but may lose packing interactions [81]. This hypothesis has been supported by subsequent studies on membrane protein unfolding

4.2.1 Effects of Heat on Glycerol Facilitator in Dodecyl Maltoside Solution and Lyso-Myristoyl Phosphatidylcholine Solutions at Neutral pH

In order to measure the thermal stability of GlpF prepared in pH 7 DDM- and pH 7 LMPC-solutions, spectroscopic and by electrophoretic methods were used. The thermal unfolding of secondary structure was followed by far UV CD spectroscopy and the spectra are shown in Figure 45 (DDM) and Figure 48 (LMPC). Plots of the temperature dependence of the ellipticity at 209 nm fit to a 2-state unfolding transition are shown in Figure 46 (DDM) and Figure 49 (LMPC). The spectra indicate that the DDM-solubilized protein secondary structure is relatively stable up to 60°C following which increasing temperature reduces the fractional helicity as deduced from far UV CD from ~ 60% to ~ 20% with a T_m of $75 \pm 1.1^\circ\text{C}$. The DDM-solubilized GlpF near UV CD ellipticity is relatively unperturbed up to 60°C (Figures 3-14 and 3-15) and is reduced with a T_m of $71 \pm 2^\circ\text{C}$ that is only slightly lower than those measured by far UV CD. This suggests that the thermal transition involves a cooperative loss of both secondary and tertiary structure above 60°C in DDM. The secondary structure of LMPC-solubilized protein is stable up

to 70°C following which increasing temperature reduces the fractional helicity as deduced from far UV CD from ~ 60% to ~ 18% with a T_m of $86.9 \pm 11.5^\circ\text{C}$ and ΔT of 15.1°C . Similarly for LMPC-solubilized GlpF the near UV CD ellipticity is relatively unperturbed up to 70°C (Figure 3-23) and is eliminated with a T_m of $74.9 \pm 1.5^\circ\text{C}$ and a ΔT of 9.2°C , which are lower than those measured by far UV CD. The broad transition range of the secondary structure unfolding and the separation of the secondary and tertiary unfolding transitions suggest a different thermal unfolding mechanism and stability of the protein in LMPC-solution, compared to DDM-solution. The broad transition of the secondary structure thermal unfolding is indicative of only partial unfolding of helices through the temperature range of tertiary structure unfolding. At 78.8°C , 90% of the tertiary structure in LMPC-solution has unfolded, while approximately 75% of the secondary structure remains intact and, when 50% of the tertiary structure has unfolded only 10% of the secondary structure has unfolded in LMPC-solution. By comparison, GlpF dissolved in DDM-solution has 90% of the tertiary structure unfolded at approximately 69.1°C , and approximately only 55% of the secondary structure remains intact, and when 50% of the tertiary structure is unfolded approximately 20% of the secondary structure has unfolded in DDM-solution. In LMPC, all levels of structure are more thermally stable than in DDM and the unfolding appears less cooperative: the melting points of the secondary structure ($T_m = 87^\circ\text{C}$) and quaternary structure ($T_m > 80^\circ\text{C}$) are similar and significantly higher than that of the tertiary structure ($T_m = 75^\circ\text{C}$), however the secondary structure unfolding transition ($\Delta T = 15^\circ\text{C}$) is broad suggesting helices with varying stability. The lower stability of the protein in DDM is also indicated by the lower thermal-stable helix content remaining at

95°C in DDM (14%) compared to LMPC (22%). The loss of tertiary structure before secondary structure in LMPC and the presence of a core of thermal-stable helices in both detergents support the standard view that the connecting loops are more thermal labile than the α -helices. However, the cooperative unfolding of much of the helix content along with the tertiary and quaternary structure in DDM indicate the influence that detergent can have on the unfolding pathway.

Finally, the thermal stability of the GlpF protein tertiary structure in DDM- and LMPC solutions was monitored by tryptophan fluorescence (Figure 32 and Figure 40). Elevated temperatures progressively quench the tryptophan fluorescence and a small red shift from 329 to 333 nm is detected in the spectrum at temperatures above 60°C in concert with the changes observed in the near UV CD spectra. This indicates that as the tertiary structure unfolds there is a slight increase in accessibility of one or more tryptophan residues to the aqueous environment. The most likely candidate is tryptophan-42, located at the interface between the monomers, and is supported by observations of thermal unfolding of the quaternary structure discussed below.

To measure the thermal stability of the quaternary structure in DDM- and LMPC solutions, SDS- and BN-PAGE were used (Figure 20B, Figure 23, Figure 27A and Figure 27B respectively). SDS-PAGE indicates that in DDM-solution the tetramer is stable up to 60°C but at higher temperatures the protein unfolds forming at least two high molecular weight complexes; one aggregate enters both the stacking and running gel and the other just enters the stacking gel (Figure 20B). The results of BN PAGE of the heated

protein in DDM-solution are shown in Figure 23. The unfolding and aggregation of GlpF are evidenced only by the disappearance of the tetrameric protein from the gels suggesting that the high M_r oligomers are too large to enter the cross-linked polyacrylamide. It seems likely that the oligomers observed on the SDS gel in Figure 20B underestimate the actual oligomer size formed at high temperature in DDM because of the ability of SDS to dissociate them. Most of the time, the oligomers are soluble complexes however, judging from the only very faint development of turbidity in the samples in the absence of SDS.

Similarly for the GlpF protein dissolved in LMPC-solution, SDS-PAGE shows that the tetramer is stable to 60°C with increasing amounts of higher molecular weight oligomers appearing as the temperature is raised further (Figure 27A). However, in DDM-solution the tetramer is completely dissociated and aggregated by 80°C (Figure 20B lane-4), whereas in LMPC solution a large fraction of the tetramer is still intact at 80°C (Figure 27A lane-1). This greater stability to thermal denaturation in LMPC-solution is likely attributable to a combination of the longer detergent carbon-tail length and the phosphatidylcholine head group. Molecular dynamics simulations [186] and magnetic resonance studies [222] indicate that matching the hydrophobic region of the protein to that of the solvent (lipid, detergent, etc.) dramatically affects the lipid ordering at the surface of the membrane protein, thereby affecting the protein's structure and function. In addition, Jensen and Mouritsen [186] show that when lipids with carbon chains longer than sixteen interact with the GlpF protein, they show a decrease in the order parameters, suggesting that lipids with sixteen carbons are an excellent hydrophobic match to the

GlpF protein. This supports the result that in LMPC-solution GlpF will be more stable to thermal denaturation than in DDM-solution, as the myristoyl carbon tail of LMPC will be better matched to the hydrophobic region of GlpF. A possible explanation of the role of the detergent head group in the higher thermal stability of the GlpF in LMPC compared to DDM is that the phosphocholine head group provides an electrostatic environment that is better matched to GlpF at pH 7 than the neutral maltoside. This seems likely in light of the high abundance (over 80 mole per cent) of phosphatidylethanolamine in the *E. coli* inner membrane which is isocoulombic with phosphatidylcholine. In LMPC, the quaternary structure of the GlpF appears to be more thermally stable than the tertiary structure. This suggests that, *in vivo*, the tetramer may form by the association of molten-globule-like structures that contain native α -helices but no fixed tertiary structure and that compact tertiary folds form only in the assembled tetramer. This pathway would assign a chaperone function to each of the monomers and highlights the importance of the tetramer in stabilization of the protein tertiary structure. Such a mechanism may be widespread as many membrane proteins are oligomeric and may require multiple protein:protein contacts for formation and stabilization of a well packed tertiary cores. Dimeric procaspase-3, a water soluble protein, has been shown to unfold to an associated dimer intermediate and the stability of the dimer contributes a significant fraction of the conformational free energy of the protein [223].

The results of BN PAGE of the heated protein in LMPC-solution are shown in Figure 27B. They confirm that in LMPC-solution the GlpF protein retains a large proportion of

its tetrameric structure at 80°C, as evident from the intensity of the band on the BN-electrophoregram. The unfolding and aggregation of GlpF at higher temperatures are evidenced by the slight disappearance of the tetrameric protein from the gel and the appearance of high M_r oligomers that just enter the cross-linked polyacrylamide. In addition, a small amount of monomer, dimer and trimer are observed suggesting that LMPC has a greater ability than DDM to partially dissociate the large oligomers and keep them in solution, although not to the extent that SDS can.

The temperature dependence of the loss of quaternary structure in both detergent solutions agrees well with the observed loss of tertiary and secondary structures shown in Figure 31 and Figure 46 (DDM tertiary and secondary structure thermal unfolding fits respectively) and Figure 39 and Figure 49 (LMPC tertiary and secondary structure thermal unfolding fits respectively). Although the temperature resolution of the quaternary structure thermal denaturations is too low to determine mid-points of unfolding, the results are consistent with those obtained using spectroscopic techniques for secondary and tertiary structures. Taken together, these results suggest that the major thermal transitions measured at ~65°C in DDM, and at 75°C in LMPC involve a cooperative loss of quaternary, tertiary and secondary structure in DDM, with a less cooperative unfolding pathway in LMPC. These results also explain the small red shift in Tryptophan emission under thermal denaturation in both detergent-solutions as most of the Tryptophan side-chains are likely buried in hydrophobic environments in the aggregates. The high melting points for the GlpF structure and its resistance to unfolding

by SDS indicate the high stability of the tetramer under the conditions of its preparation in both DDM- and LMPC solutions.

The thermal denaturations of detergent-solubilized GlpF are irreversible and as such a thermodynamic analysis is not valid; however, when used for comparison purposes only, we find that the GlpF protein in DDM- and LMPC-solutions exhibit characteristics similar to other membrane proteins. In LMPC-solution, the calculated enthalpy of unfolding of the tertiary structure ($T_m = 74.9 \pm 1.5^\circ\text{C}$) is 3.4 cal/g. This is well within the expected range for membrane proteins based on collected data [205-208, 211, 212, 214, 219-221], and well below the average of 8.5 cal/g at this temperature reported for soluble proteins [214]. The calculated enthalpy of unfolding for the secondary structure in LMPC solution is 1.4 cal/g, or 0.15 kcal/mol-residue, which is also much lower than the average soluble protein. The unfolding enthalpy values for GlpF in DDM-solution are 2.4 cal/g and 2.5 cal/g for tertiary and secondary structure respectively. It is interesting to speculate on the cooperativity of the thermal unfolding based on these enthalpies. In DDM-solution, both near and far UV CD have similar thermal midpoints, and similar values for the enthalpy of unfolding, suggesting a single transition. However, in LMPC both the midpoints and the enthalpies of unfolding are significantly different for the tertiary and secondary structure suggesting that the protein may unfold in two discrete stages with tertiary structure unfolding before secondary structure.

4.3 Effects of Urea

Urea is a common protein denaturant used to measure the stability of many water-soluble proteins and has recently been shown to reversibly unfold a β -barrel protein [178]. Initial experiments on DDM-solubilized GlpF indicated a time dependence of the effects of urea (see Figure 36) and in order to permit the establishment of equilibrium conditions the protein was incubated in various concentrations of urea for 3 weeks. Figure 25 and Figure 26 show the effects of additions of urea on the DDM-solubilized protein quaternary structure. According to SDS-PAGE low concentrations of urea (2-4 M; lanes 2-3 Figure 25) have a minor effect on the amount of tetramer present in the solution. Small increases in monomer, dimer and trimer are observed but the most notable change is the significant loss of a high molecular species that appears to be an octamer that is sometimes observed in the preparations. After incubation for 3 weeks in 8 M urea (Figure 25, lane 5), the tetramer is nearly absent, monomer, dimer, and trimer have increased substantially, and the octamer is greatly reduced. In addition, protein oligomers electrophoresing between the tetramer and octamer appear that are likely non-specific aggregates of unfolded protein, which are soluble in high concentrations of SDS and urea. BN-PAGE of the same samples is shown in Figure 26. The BN gels appear to be less sensitive to the unfolding induced by urea at low concentrations because in the absence of SDS the low molecular weight species are not observable on the gels presumably because they form aggregates too large to enter the gel. It is also possible that in the absence of SDS the tetramer is resistant to dissociation in low concentrations of urea. However, at 6 M and 8 M urea (Figure 26 lanes 4 and 5 respectively) the unfolding of the tetramer and octamer are evident from their diminishing bands on the

BN gel and from the very faint bands that appear to be trimer and dimer that appear. The latter are likely soluble in small amounts because of the high concentrations of urea present.

Figure 36 shows another indication of the resistance of the protein to denaturation by urea. Incubation of the DDM-solubilized GlpF in 8 M urea for 1 day causes little change in the near UV CD spectrum. However, over the course of 21 days, the added urea reduces the tertiary structure with an exponential unfolding rate constant of $4 \pm 1 \times 10^{-6} \text{ s}^{-1}$ (Figure 37). The high concentrations of urea required to unfold the protein and the slow unfolding rate both suggest a high thermodynamic stability of the protein dissolved in DDM-solution. The four-helix bundle RNA-binding protein Rop dimer displays similar stability. It unfolds with a guanidine-HCl concentration midpoint of 3.23 M, a thermal unfolding midpoint of 70°C , and an unfolding rate constant of $0.7 \times 10^{-6} \text{ s}^{-1}$ [224]. A tetrameric mutant is even more stable with a thermal unfolding mid-point of 101°C [225]. Intersubunit interactions make significant contributions to the stability of the Rop fold and the same is likely to be the case for the GlpF tetramer.

The mechanism by which urea disrupts protein structure is still not completely understood. Urea has long been thought to disrupt hydrophobic interactions within proteins [226-228], however more recent studies suggest that the disruption is not in the hydrophobic domains, but of hydrogen-bonds in the hydrophilic regions of the protein [229]. If we presume that urea does not disrupt hydrophobic interactions both between the protein and detergent and within the protein [229], and that urea can interact with the

water-soluble loop regions of the protein, this suggests that hydrophobic interactions between the protein monomers play a major role in stabilizing the protein structure and explain why high concentrations of urea are required to unfold the GlpF protein.

4.4 Influence of pH on Protein Structure

The recent interest in partially folded states of proteins is in part due to advances in NMR spectroscopy that permit observations of structural details. However, partially unfolded proteins also provide insight into protein folding and protein stability [230-233]. It has been demonstrated that some partially folded states may be sampled by the linear polypeptide during the folding process prior to the appearance of native structure [47, 234, 235]. The formation of partially folded states, called molten globules, early in the protein folding pathway leads to a dramatic decrease in the conformational space sampled by the polypeptide during the folding process, and could therefore simplify the protein folding problem [44, 231, 236]. The consensus characteristics of the molten globule are that it must have a significant amount of native-like secondary structure, a compact fold which lacks strong tertiary interactions, enhanced accessibility to the solvent and low cooperativity to thermal unfolding [54, 232, 236-239].

A mutant of the four helix bundle Interleukin-6 lacking the first 22 residues, and one disulfide bond exhibits structural features similar to that of the native protein at neutral

pH [240] and that of a molten globule when studied at low pH [241]. Far UV CD analysis determined that even at pH values as low as 2, the interleukin-6 mutant retains the secondary structural features of the native protein exhibiting minima at 208 and 222 nm [241]. However, the effects of pH 2 are more dramatic on the tertiary structure of the protein. At pH 2, the intensity of the near UV CD is significantly lower than that of the native protein suggesting a loss of tertiary structure [241]. With intensities being only approximately 50 % of those obtained from the native protein [240, 241], the addition of acid has likely induced a looser structure, resulting in flexible environments near the aromatic groups [242]. Fluorescence studies on the acidified state of interleukin-6 also indicate a loss of tertiary structure [241] as changes in both the intensity and the emission maximum of the tryptophan fluorescence are observed as the pH is lowered.

Measurements conducted on model compounds showed that upon acidification there is a decrease in tryptophan quantum yield, while the emission maximum remains constant [241]. This suggested that the blue shift in the fluorescence maximum at pH 2 of the interleukin-6 protein reports changes in the polarity of the microenvironment near the tryptophan residue [243]. However, the decrease in fluorescence intensity at pH 2 is likely primarily due to an acid quenching [241]. In addition, the intensity of tyrosine fluorescence near 303 nm increases upon acidification, which is interpreted as originating from a weakening of the tyrosine-to-tryptophan energy transfer.

4.4.1 Influence of pH on GlpF Structure

Conditions of acidic pH (3 or less) lead to a dissociation of the GlpF tetramer in both DDM and LMPC as observed by both SDS and BN PAGE. This result is similar to the effects of acidic pH on the oligomerization state of AQPZ, where the protein must be acidified in order to disrupt the stable tetramer [192]. The electrophoregrams in Figure 21 and Figure 28A show the influence of pH in combination with SDS on the quaternary structure of GlpF in DDM and LMPC respectively. Figure 21, lanes 1-2 show that the DDM-solubilized tetramer is stable between pH 7.8 and pH 7.0 on SDS-PAGE. At pH 6.0, (Figure 21, lane 3) the tetramer band is diminished and the monomer band increased, indicating partial dissociation of the tetramer into monomer; dissociation appears to be complete at pH 5.0 (Figure 21, lane 4). It is worth noting that the lanes containing monomer show very little dimeric or trimeric protein suggesting that the monomer could be in an unaggregated and partially folded state in the mixture of SDS and DDM. Figure 28A shows that the LMPC-solubilized GlpF quaternary structure is more pH sensitive than the DDM-solubilized GlpF. Electrophoresis of the LMPC-solubilized GlpF in the presence of SDS suggests that the tetrameric structure near pH 6 is disrupted, and there is an increased amount of dimeric and trimeric protein observed, which are likely due to oligomerization of partially unfolded monomeric GlpF. Electrophoresis of the same preparations on BN polyacrylamide gels shows the sensitivity of the protein to pH in the absence of SDS (Figure 24 and Figure 28B). According to BN-PAGE, the protein is predominantly a tetramer at pH 5.0 in both DDM (Figure 24 lane 3) and LMPC (Figure 28B) lane 3, however it must be less stable than at neutral pH as it is sensitive to dissociation by SDS (Figure 24 lane 3 and Figure 28A lane 3). At pH 4.0 in DDM

solution, BN-PAGE shows that the tetramer has nearly completely disappeared even in the absence of SDS (Figure 24, lane 4). In its place a high molecular weight oligomer appears at the top of the gel. At pH 4.0 in LMPC solution (Figure 28B lane 2), BN-Page shows that the tetramer is completely disrupted, with the appearance of monomeric GlpF and high molecular weight oligomers.

The changes in secondary and tertiary structure brought about by lowering the pH were investigated spectroscopically. pH titrations by far UV CD in DDM solution (Figure 47), and in LMPC solution (Figure 50) show that the helix content of the protein at low pH is nearly identical to that of the neutral pH tetramer. However, near UV CD shows that at pH 3.0, the DDM-solubilized protein has lost approximately 80% of its near UV signal and tertiary structure (Figure 33). At pH 3.0 in LMPC solution (Figure 41), the near UV CD signal intensity has passed through zero intensity, into positive ellipticity. This has interesting similarities to the shape and intensity of the spectrum of the protein dissolved in SDS solution (Figure 43). Recall that under the conditions of SDS solution, it has been proposed [98] that membrane proteins are only slightly unfolded by intercalation of the SDS monomers into the protein, disrupting some of the close contacts. Low pH may induce similar effects on GlpF, where the close contacts of the helices are again only partially disrupted and the helices themselves are not unfolded. The effects of pH on the fluorescence of the DDM-solubilized protein were also measured (Figure 34). Between pH 7 and pH 5 the tryptophan emission is quenched by about 15% but there is no detectable change in the position of the emission maximum. However, between pH 5 and pH 4 the quantum yield is restored to slightly more than its original value and a shift in

the peak maximum occurs by about 4 nm to the red. Between pH 4 and 3 the fluorescence is quenched by about 30% but the emission maximum remains at 333 nm. These changes likely reflect direct effects of the protonation of different side-chains in the local environments of the five Tryptophan residues as well as indirect effects of the unfolding of the tertiary structure and oligomerization as the pH is lowered. The shift of the emission maximum to 333 nm was also observed in the thermal unfolding experiments and indicates that the protein unfolds below pH 5.0, which agrees with near UV CD, and SDS and BN gel analysis.

Overall, the results are compatible with the concept that the low pH form of the protein in DDM and LMPC is a molten globule that contains native secondary structure, little tertiary structure, and may exist as a high molecular weight oligomer. However, the addition of SDS and elevation of the pH to 7 (for example see Figure 22) appear to dissociate the oligomer leading to the formation of a monomeric molten globule. Furthermore, the lack of non-specific aggregates (dimer, trimer) on the SDS gels suggests that the high molecular weight oligomer is not a non-specific aggregate but may be an association of molten globule tetramers that are dissociable by SDS or raising the pH.

4.6 Glycerol facilitator in Sodium Dodecyl Sulfate and Octyl Glucoside solutions

We have previously shown that GlpF prepared in SDS solution at pH 4.2 electrophoreses predominantly as a monomeric protein on SDS-PAGE and contains an α -helix content of

about 48% (53% if the affinity tag is ignored) [141]. Attempts to elute the SDS-solubilized protein from the Ni²⁺ resin at neutral pH with imidazole yielded mostly monomeric protein that eluted in low concentrations of imidazole but also contained a low background of protein impurities (not shown), which is likely due to the effective low pH microenvironments induced by SDS [244]. The α -helix in low pH-eluted, SDS-solubilized protein is thermally stable up to 50°C and at 70°C the protein still retains about 40% of its helix content (data not shown). Thus, in SDS the α -helices are thermally very stable. Near UV CD spectra of the protein dissolved in SDS at various temperatures have been shown to exhibit a weak positive ellipticity and lack well-defined absorption bands, which suggest some transient asymmetry in the environments of the aromatic residues [181]. However, increasing the temperature does not eliminate this spectrum and there is therefore no evidence of a two-state unfolding transition. The protein in SDS thus appears to reside in a state retaining much of its secondary structure and in addition having retained some tertiary contacts but no quaternary structure. This result supports a recent proposal of SDS binding to membrane proteins [98] shown in Figure 12, where the SDS monomers intercalate between the helices, disrupting the majority of tertiary contacts but not destroying secondary structural elements.

GlpF was also prepared in OG by elution from Ni²⁺ resin at low pH [141] or with 250 mM imidazole. Dilute solutions could be studied by far UV CD, but both forms of the protein readily precipitate from solution at the concentrations required for near UV CD analysis. Although the preparations of GlpF in DDM solution at neutral pH remain in solution for several months, far UV CD shows a slow loss of structure over this time

course. The improved stability of the protein dissolved with the aid of 12 carbon detergents compared to 8 carbon detergents, and the further increase in stability with the use of detergents containing 14 carbons shows that significantly improved conditions for stabilizing membrane proteins in solution have been achieved.

4.7 Future Work

The present results demonstrate that conformationally stable solutions of membrane proteins can be prepared as long as the correct solution conditions such as pH, temperature, and detergent are found. In the case of the GlpF, this stability is in good agreement with the observations of a tightly packed 3D structure and low temperature B-factors by X-ray diffraction [1]. The results also agree with calculations of root mean-squared deviations of C^α atoms during molecular dynamics simulations that indicate only very minor conformational changes during glycerol passage through the pore [124, 245]. The molecular dynamics simulations indicate that there is no cooperativity between the subunits of the tetramer during glycerol passage [245] and this suggests that the tetramers play a structural role stabilizing the monomer conformation in the hydrophobic membrane. The molecular dynamics simulations also point out that the stability of the pore conformation is critical both for the conduction of glycerol and for the selectivity of the pore for glycerol over water [245]. Kinetically and thermodynamically stable preparations of membrane proteins in detergent may be amenable to structure and

dynamics analysis by multinuclear NMR TROSY experiments [246] if sufficient deuterium-labelled protein can be prepared at concentrations of 200-300 μM .

The equilibrium molten globule is speculated to be an analogue of an important kinetic intermediate in the folding of many water-soluble proteins [247]. It has also been shown to be an important intermediate on the membrane insertion pathway of the colicin Ia channel domain [248]. I have shown here that dissociation of the GlpF tetramer leads to non-specific aggregation or molten globule formation, suggesting that GlpF may exist as a molten globule during its insertion into the membrane by the Sec translocase and while in the membrane before it forms the tetramer. More detailed structural analysis of the acid-unfolded molten globule described here may be possible using hydrogen-deuterium exchange and NMR spectroscopy as has been done for several water-soluble proteins [247].

Bibliography

1. Fu, D., et al., *Structure of a glycerol-conducting channel and the basis for its selectivity*. Science, 2000. **290**(5491): p. 481-6.
2. Hooke, R., *Micrographia: or, Some physiological descriptions of minute bodies made by magnifying glasses*. Vol. First. 1665, London: J. Matyn and J. Allestry.
3. Tanford, C., *Ben Franklin Stilled the Waves: An Informal History of Pouring Oil on Water with Reflections on the Ups and Downs of Scientific Life in General*. 1989, Duke University Press: Durham, NC.
4. Langmuir, I., *The constitution and fundamental properties of solids and liquids. II. Liquids*. J Am Chem Soc, 1917. **39**: p. 1848-1906.
5. Ahrens, H. *Langmuir Trough*. 2001 [cited; Available from: <http://www3.physik.uni-greifswald.de/method/filmwaage/efilmwag.htm>].
6. Gorter, E. and F. Grendel, *On bimolecular layers of lipoids on the chromocytes of blood*. J. Exp. Med., 1925. **41**(439-443).
7. Hendler, R.W., *Biological membrane ultrastructure*. Physiol Rev, 1971. **51**(1): p. 66-97.

8. Sadava, D.E., *Cell Biology: Organelle Structure and Function*. 1993, Boston: Jones and Bartlett.
9. Eichman, P. *From the Lipid Bilayer to the Fluid Mosaic: A Brief History of Membrane Models*. [cited; Available from: <http://www1.umn.edu/ships/9-2/membrane.htm>.
10. Campbell, N.A., *Biology*. Third ed. 1993: The Benjamin/Cummins Publishing Company, Inc.
11. Robertson, J.D., *New observations on the ultrastructure of the membranes of frog peripheral nerve fibers*. *J Biophys Biochem Cytol*, 1957. **3**(6): p. 1043-8.
12. Singer, S.J. and G.L. Nicolson, *The fluid mosaic model of the structure of cell membranes*. *Science*, 1972. **175**(23): p. 720-31.
13. Geibel, G. *The Cell Membrane*. 2004 [cited; Available from: http://sun.menloschool.org/~cweaver/cells/c/cell_membrane/.
14. Spencer, A.G., et al., *Subcellular localization of prostaglandin endoperoxide H synthases-1 and -2 by immunoelectron microscopy*. *J Biol Chem*, 1998. **273**(16): p. 9886-93.
15. Wallin, E. and G. von Heijne, *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms*. *Protein Sci*, 1998. **7**(4): p. 1029-38.
16. de Kruijff, B. and P.R. Cullis, *Cytochrome c specifically induces non-bilayer structures in cardiolipin-containing model membranes*. *Biochim Biophys Acta*, 1980. **602**(3): p. 477-90.
17. Tuominen, E.K., C.J. Wallace, and P.K. Kinnunen, *Phospholipid-cytochrome c interaction: evidence for the extended lipid anchorage*. *J Biol Chem*, 2002. **277**(11): p. 8822-6.
18. Voet, D., J.G. Voet, and C.W. Pratt, *Fundamentals of Biochemistry*. 1998: John Wiley & Sons, Inc.
19. Kurumbail, R.G., et al., *Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents*. *Nature*, 1996. **384**(6610): p. 644-8.
20. Picot, D. and R.M. Garavito, *Prostaglandin H synthase: implications for membrane structure*. *FEBS Lett*, 1994. **346**(1): p. 21-5.
21. Picot, D., P.J. Loll, and R.M. Garavito, *The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1*. *Nature*, 1994. **367**(6460): p. 243-9.
22. Zhang, J., E.L. Ding, and Y. Song, *Adverse Effects of Cyclooxygenase 2 Inhibitors on Renal and Arrhythmia Events*. *JAMA*, 2006. **296**(13): p. E1-E14.
23. Blobel, G., *Intracellular protein topogenesis*. *Proc Natl Acad Sci U S A*, 1980. **77**(3): p. 1496-500.
24. von Heijne, G. and Y. Gavel, *Topogenic signals in integral membrane proteins*. *Eur J Biochem*, 1988. **174**(4): p. 671-8.
25. Spiess, M., *Heads or tails--what determines the orientation of proteins in the membrane*. *FEBS Lett*, 1995. **369**(1): p. 76-9.
26. Gafvelin, G., et al., *Topological rules for membrane protein assembly in eukaryotic cells*. *J Biol Chem*, 1997. **272**(10): p. 6119-27.

27. Heijne, G.V., *The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology*. *Embo J*, 1986. **5**(11): p. 3021-3027.
28. Turner, R.J., *Understanding the biogenesis of polytopic integral membrane proteins*. *J Membr Biol*, 2003. **192**(3): p. 149-57.
29. Wahlberg, J.M. and M. Spiess, *Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain*. *J Cell Biol*, 1997. **137**(3): p. 555-62.
30. Anfinsen, C.B., *Principles that govern the folding of protein chains*. *Science*, 1973. **181**(96): p. 223-30.
31. Pande, V. *Folding@home distributed computing*. 2000 [cited; Available from: <http://folding.stanford.edu>].
32. Kubelka, J., J. Hofrichter, and W.A. Eaton, *The protein folding 'speed limit'*. *Curr Opin Struct Biol*, 2004. **14**(1): p. 76-88.
33. McKnight, C.J., P.T. Matsudaira, and P.S. Kim, *NMR structure of the 35-residue villin headpiece subdomain*. *Nat Struct Biol*, 1997. **4**(3): p. 180-4.
34. McKnight, C.J., et al., *A thermostable 35-residue subdomain within villin headpiece*. *J Mol Biol*, 1996. **260**(2): p. 126-34.
35. Kubelka, J., W.A. Eaton, and J. Hofrichter, *Experimental tests of villin subdomain folding simulations*. *J Mol Biol*, 2003. **329**(4): p. 625-30.
36. Duan, Y. and P.A. Kollman, *Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution*. *Science*, 1998. **282**(5389): p. 740-4.
37. Duan, Y., L. Wang, and P.A. Kollman, *The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation*. *Proc Natl Acad Sci U S A*, 1998. **95**(17): p. 9897-902.
38. Levinthal, C., *Are there pathways for protein folding?* *J. Chim. Phys.*, 1968. **65**: p. 44-45.
39. Dobson, C.M. and M. Karplus, *Protein Folding: A Perspective from Theory and Experiment*. *Angew. Chem. Int. Ed.*, 1998. **37**(7): p. 868-893.
40. Chan, H.S., S. Bromberg, and K.A. Dill, *Models of cooperativity in protein folding*. *Philos Trans R Soc Lond B Biol Sci*, 1995. **348**(1323): p. 61-70.
41. Cantor, C.R. and P.R. Schimmel, *Biophysical Chemistry Part III: The behaviour of biological macromolecules*. 1980, New York: W. H. Freeman and Company.
42. Lifson, S. and A. Roig, *On the theory of helix-coil transition in polypeptides*. *J. Chem. Phys.*, 1961. **34**: p. 1963-74.
43. Ptitsyn, O.B., *[Stages in the mechanism of self-organization of protein molecules]*. *Dokl Akad Nauk SSSR*, 1973. **210**(5): p. 1213-5.
44. Kim, P.S. and R.L. Baldwin, *Intermediates in the folding reactions of small proteins*. *Annu Rev Biochem*, 1990. **59**: p. 631-60.
45. Abkevich, V.I., A.M. Gutin, and E.I. Shakhnovich, *Specific nucleus as the transition state for protein folding: evidence from the lattice model*. *Biochemistry*, 1994. **33**(33): p. 10026-36.
46. Wetlaufer, D.B., *Nucleation, rapid folding, and globular intrachain regions in proteins*. *Proc Natl Acad Sci U S A*, 1973. **70**(3): p. 697-701.

47. Ptitsyn, O.B., et al., *Evidence for a molten globule state as a general intermediate in protein folding*. FEBS Lett, 1990. **262**(1): p. 20-4.
48. Ptitsyn, O., *How molten is the molten globule?* Nat Struct Biol, 1996. **3**(6): p. 488-90.
49. Fersht, A.R., *Nucleation mechanisms in protein folding*. Curr Opin Struct Biol, 1997. **7**(1): p. 3-9.
50. Jackson, S.E. and A.R. Fersht, *Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition*. Biochemistry, 1991. **30**(43): p. 10428-35.
51. Jackson, S.E. and A.R. Fersht, *Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding*. Biochemistry, 1991. **30**(43): p. 10436-43.
52. Fersht, A.R., *Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications*. Proc Natl Acad Sci U S A, 1995. **92**(24): p. 10869-73.
53. Nolting, B., et al., *The folding pathway of a protein at high resolution from microseconds to seconds*. Proc Natl Acad Sci U S A, 1997. **94**(3): p. 826-30.
54. Bychkova, V.E., et al., *Retinol-binding protein is in the molten globule state at low pH*. Biochemistry, 1992. **31**(33): p. 7566-71.
55. Bychkova, V.E. and O.B. Ptitsyn, [*The state of unfolded globules of protein molecules is more quickly becoming a rule, rather than an exception*]. Biofizika, 1993. **38**(1): p. 58-66.
56. Crowley, K.S., G.D. Reinhart, and A.E. Johnson, *The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation*. Cell, 1993. **73**(6): p. 1101-15.
57. Grigorieff, N., et al., *Electron-crystallographic refinement of the structure of bacteriorhodopsin*. J Mol Biol, 1996. **259**(3): p. 393-421.
58. Weiss, M.S. and G.E. Schulz, *Structure of porin refined at 1.8 Å resolution*. J Mol Biol, 1992. **227**(2): p. 493-509.
59. Onuchic, J.N., et al., *Protein folding funnels: the nature of the transition state ensemble*. Fold Des, 1996. **1**(6): p. 441-450.
60. Berman, H.M., et al., *The protein data bank*. Nucleic Acids Res, 2000. **28**: p. 235-242.
61. Simon, S.M. and G. Blobel, *A protein-conducting channel in the endoplasmic reticulum*. Cell, 1991. **65**(3): p. 371-80.
62. Johnson, A.E. and M.A. van Waes, *The translocon: a dynamic gateway at the ER membrane*. Annu Rev Cell Dev Biol, 1999. **15**: p. 799-842.
63. Osborne, A.R., T.A. Rapoport, and B. van den Berg, *Protein translocation by the Sec61/SecY channel*. Annu Rev Cell Dev Biol, 2005. **21**: p. 529-50.
64. Van den Berg, B., et al., *X-ray structure of a protein-conducting channel*. Nature, 2004. **427**(6969): p. 36-44.
65. Cannon, K.S., et al., *Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY*. J Cell Biol, 2005. **169**(2): p. 219-25.
66. Mitra, K., et al., *Structure of the E. coli protein-conducting channel bound to a translating ribosome*. Nature, 2005. **438**(7066): p. 318-24.

67. Rapoport, T.A., B. Jungnickel, and U. Kutay, *Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes*. *Annu Rev Biochem*, 1996. **65**: p. 271-303.
68. Blobel, G. and B. Dobberstein, *Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma*. *J Cell Biol*, 1975. **67**(3): p. 835-51.
69. von Heijne, G., *The signal peptide*. *J. Membr. Biol.*, 1990. **115**(3): p. 195-201.
70. Walter, P., I. Ibrahimi, and G. Blobel, *Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein*. *J Cell Biol*, 1981. **91**(2 Pt 1): p. 545-50.
71. Walter, P. and A.E. Johnson, *Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane*. *Annu Rev Cell Biol*, 1994. **10**: p. 87-119.
72. Walter, P. and G. Blobel, *Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum*. *Nature*, 1982. **299**(5885): p. 691-8.
73. Walter, P. and G. Blobel, *Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes*. *J Cell Biol*, 1981. **91**(2 Pt 1): p. 557-61.
74. Walter, P. and G. Blobel, *Translocation of proteins across the endoplasmic reticulum. II. Signal recognition protein (SRP) mediates the selective binding to microsomal membranes of in-vitro-assembled polysomes synthesizing secretory protein*. *J Cell Biol*, 1981. **91**(2 Pt 1): p. 551-6.
75. Borel, A.C. and S.M. Simon, *Biogenesis of polytopic membrane proteins: membrane segments assemble within translocation channels prior to membrane integration*. *Cell*, 1996. **85**(3): p. 379-89.
76. Do, H., et al., *The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process*. *Cell*, 1996. **85**(3): p. 369-78.
77. Blobel, G. *Protein Targeting*. 1999 [cited; Available from: <http://nobelprize.org>].
78. Gouaux, E., *Channel-forming toxins: tales of transformation*. *Curr Opin Struct Biol*, 1997. **7**(4): p. 566-73.
79. White, S.H., W.C. Wimley, and M.E. Selsted, *Structure, function, and membrane integration of defensins*. *Curr Opin Struct Biol*, 1995. **5**(4): p. 521-7.
80. Jacobs, R.E. and S.H. White, *The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices*. *Biochemistry*, 1989. **28**(8): p. 3421-37.
81. Haltia, T. and E. Freire, *Forces and factors that contribute to the structural stability of membrane proteins*. *Biochim Biophys Acta*, 1995. **1228**(1): p. 1-27.
82. White, S.H. and W.C. Wimley, *Membrane protein folding and stability: physical principles*. *Annu Rev Biophys Biomol Struct*, 1999. **28**: p. 319-65.
83. Popot, J.L. and D.M. Engelman, *Membrane protein folding and oligomerization: the two-stage model*. *Biochemistry*, 1990. **29**(17): p. 4031-7.
84. White, S.H. and W.C. Wimley, *Hydrophobic interactions of peptides with membrane interfaces*. *Biochim Biophys Acta*, 1998. **1376**(3): p. 339-52.

85. Wimley, W.C. and S.H. White, *Experimentally determined hydrophobicity scale for proteins at membrane interfaces*. Nat Struct Biol, 1996. **3**(10): p. 842-8.
86. Ladokhin, A.S. and S.H. White, *Folding of amphipathic alpha-helices on membranes: energetics of helix formation by melittin*. J Mol Biol, 1999. **285**(4): p. 1363-9.
87. Wimley, W.C., et al., *Folding of beta-sheet membrane proteins: a hydrophobic hexapeptide model*. J Mol Biol, 1998. **277**(5): p. 1091-110.
88. Ben-Tal, N., et al., *Free-energy determinants of alpha-helix insertion into lipid bilayers*. Biophys J, 1996. **70**(4): p. 1803-12.
89. Engelman, D.M., T.A. Steitz, and A. Goldman, *Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins*. Annu Rev Biophys Chem, 1986. **15**: p. 321-53.
90. Roseman, M.A., *Hydrophobicity of the peptide C=O...H-N hydrogen-bonded group*. J Mol Biol, 1988. **201**(3): p. 621-3.
91. White, S.H. *General principles of membrane protein folding and stability*. 1998 [cited; Available from: http://blanco.biomol.uci.edu/mp_assembly.html].
92. Finkelstein, A.V. and O.B. Ptitsyn, *Why do globular proteins fit the limited set of folding patterns?* Prog Biophys Mol Biol, 1987. **50**(3): p. 171-90.
93. Huang, K.S., et al., *Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments*. J Biol Chem, 1981. **256**(8): p. 3802-9.
94. Liao, M.J., E. London, and H.G. Khorana, *Regeneration of the native bacteriorhodopsin structure from two chymotryptic fragments*. J Biol Chem, 1983. **258**(16): p. 9949-55.
95. Popot, J.L., S.E. Gerchman, and D.M. Engelman, *Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process*. J Mol Biol, 1987. **198**(4): p. 655-76.
96. Popot, J.L., J. Trehwella, and D.M. Engelman, *Reformation of crystalline purple membrane from purified bacteriorhodopsin fragments*. Embo J, 1986. **5**(11): p. 3039-44.
97. Kiefer, H., *In vitro folding of alpha-helical membrane proteins*. Biochim Biophys Acta, 2003. **1610**(1): p. 57-62.
98. Renthal, R., *An unfolding story of helical transmembrane proteins*. Biochemistry, 2006. **45**(49): p. 14559-66.
99. Fernandez, C., et al., *Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR spectroscopy*. Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13533-7.
100. Kahn, T.W., J.M. Sturtevant, and D.M. Engelman, *Thermodynamic measurements of the contributions of helix-connecting loops and of retinal to the stability of bacteriorhodopsin*. Biochemistry, 1992. **31**(37): p. 8829-39.
101. Pebay-Peyroula, E., et al., *X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases*. Science, 1997. **277**(5332): p. 1676-81.
102. Lemmon, M.A. and D.M. Engelman, *Specificity and promiscuity in membrane helix interactions*. FEBS Lett, 1994. **346**(1): p. 17-20.

103. MacKenzie, K.R., J.H. Prestegard, and D.M. Engelman, *A transmembrane helix dimer: structure and implications*. *Science*, 1997. **276**(5309): p. 131-3.
104. Park, J.H. and M.H. Saier, Jr., *Phylogenetic characterization of the MIP family of transmembrane channel proteins*. *J Membr Biol*, 1996. **153**(3): p. 171-80.
105. Gorin, M.B., et al., *The major intrinsic protein (MIP) of the bovine lens fiber membrane: characterization and structure based on cDNA cloning*. *Cell*, 1984. **39**(1): p. 49-59.
106. Baker, M.E. and M.H. Saier, Jr., *A common ancestor for bovine lens fiber major intrinsic protein, soybean nodulin-26 protein, and E. coli glycerol facilitator*. *Cell*, 1990. **60**(2): p. 185-6.
107. Borgnia, M., et al., *Cellular and molecular biology of the aquaporin water channels*. *Annu Rev Biochem*, 1999. **68**: p. 425-58.
108. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydrophobic character of a protein*. *J Mol Biol*, 1982. **157**(1): p. 105-32.
109. Lee, M.D., L.S. King, and P. Agre, *The aquaporin family of water channel proteins in clinical medicine*. *Medicine (Baltimore)*, 1997. **76**(3): p. 141-56.
110. Murata, K., et al., *Structural determinants of water permeation through aquaporin-1*. *Nature*, 2000. **407**(6804): p. 599-605.
111. Macey, R.I. and R.E. Farmer, *Inhibition of water and solute permeability in human red cells*. *Biochim Biophys Acta*, 1970. **211**(1): p. 104-6.
112. Koefoed-Johnsen, V. and H.H. Ussing, *The contributions of diffusion and flow to the passage of D₂O through living membranes; effect of neurohypophyseal hormone on isolated anuran skin*. *Acta Physiol Scand*, 1953. **28**(1): p. 60-76.
113. Agre, P., S. Sasaki, and M.J. Chrispeels, *Aquaporins: a family of water channel proteins*. *Am J Physiol*, 1993. **265**(3 Pt 2): p. F461.
114. Zeidel, M.L., et al., *Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein*. *Biochemistry*, 1992. **31**(33): p. 7436-40.
115. Zeidel, M.L., et al., *Ultrastructure, pharmacologic inhibition, and transport selectivity of aquaporin channel-forming integral protein in proteoliposomes*. *Biochemistry*, 1994. **33**(6): p. 1606-15.
116. Kozono, D., et al., *Functional expression and characterization of an archaeal aquaporin. AqpM from methanothermobacter marburgensis*. *J Biol Chem*, 2003. **278**(12): p. 10649-56.
117. Hatakeyama, S., et al., *Cloning of a new aquaporin (AQP10) abundantly expressed in duodenum and jejunum*. *Biochem Biophys Res Commun*, 2001. **287**(4): p. 814-9.
118. Morishita, Y., et al., *Molecular mechanisms and drug development in aquaporin water channel diseases: aquaporin superfamily (superaquaporins): expansion of aquaporins restricted to multicellular organisms*. *J Pharmacol Sci*, 2004. **96**(3): p. 276-9.
119. Itoh, T., et al., *Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells*. *Biochem Biophys Res Commun*, 2005. **330**(3): p. 832-8.
120. Preston, G.M. and P. Agre, *Isolation of the cDNA for erythrocyte integral membrane protein of 28 kilodaltons: member of an ancient channel family*. *Proc Natl Acad Sci U S A*, 1991. **88**(24): p. 11110-4.

121. Preston, G.M., et al., *The mercury-sensitive residue at cysteine 189 in the CHIP28 water channel*. J Biol Chem, 1993. **268**(1): p. 17-20.
122. Jung, J.S., et al., *Molecular structure of the water channel through aquaporin CHIP. The hourglass model*. J Biol Chem, 1994. **269**(20): p. 14648-54.
123. Sui, H., et al., *Structural basis of water-specific transport through the AQP1 water channel*. Nature, 2001. **414**(6866): p. 872-8.
124. deGroot, B.L. and H. Grubmuller, *Water permeation across biological membranes: Mechanism and dynamics of aquaporin-1 and GlpF*. Science, 2001. **294**(5550): p. 2353-2357.
125. Tajkhorshid, E., et al., *Control of the selectivity of the aquaporin water channel family by global orientational tuning*. Science, 2002. **296**(5567): p. 525-30.
126. Calamita, G., et al., *Molecular cloning and characterization of AqpZ, a water channel from Escherichia coli*. J Biol Chem, 1995. **270**(49): p. 29063-6.
127. Sweet, G., et al., *Glycerol facilitator of Escherichia coli: cloning of glpF and identification of the glpF product*. J Bacteriol, 1990. **172**(1): p. 424-30.
128. Preston, G.M., et al., *Appearance of water channels in Xenopus oocytes expressing red cell CHIP28 protein*. Science, 1992. **256**(5055): p. 385-7.
129. Borgnia, M.J. and P. Agre, *Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli*. Proc Natl Acad Sci U S A, 2001. **98**(5): p. 2888-93.
130. Bron, P., et al., *Oligomerization state of MIP proteins expressed in Xenopus oocytes as revealed by freeze-fracture electron-microscopy analysis*. J Struct Biol, 1999. **128**(3): p. 287-96.
131. Duchesne, L., et al., *Oligomerization of water and solute channels of the major intrinsic protein (MIP) family*. Kidney Int, 2001. **60**(2): p. 422-6.
132. Duchesne, L., et al., *Role of C-terminal domain and transmembrane helices 5 and 6 in function and quaternary structure of major intrinsic proteins: analysis of aquaporin/glycerol facilitator chimeric proteins*. J Biol Chem, 2002. **277**(23): p. 20598-604.
133. Lagree, V., et al., *Switch from an aquaporin to a glycerol channel by two amino acids substitution*. J Biol Chem, 1999. **274**(11): p. 6817-9.
134. Lagree, V., et al., *Oligomerization state of water channels and glycerol facilitators. Involvement of loop E*. J Biol Chem, 1998. **273**(51): p. 33949-53.
135. Ringler, P., et al., *Structure of the water channel AqpZ from Escherichia coli revealed by electron crystallography*. J Mol Biol, 1999. **291**(5): p. 1181-90.
136. Verbavatz, J.M., et al., *Tetrameric assembly of CHIP28 water channels in liposomes and cell membranes: a freeze-fracture study*. J Cell Biol, 1993. **123**(3): p. 605-18.
137. Braun, T., et al., *The 3.7 Å projection map of the glycerol facilitator GlpF: a variant of the aquaporin tetramer*. EMBO Rep, 2000. **1**(2): p. 183-9.
138. Stahlberg, H., et al., *The 6.9-Å structure of GlpF: a basis for homology modeling of the glycerol channel from Escherichia coli*. J Struct Biol, 2000. **132**(2): p. 133-41.
139. Voegelé, R.T., G.D. Sweet, and W. Boos, *Glycerol kinase of Escherichia coli is activated by interaction with the glycerol facilitator*. J Bacteriol, 1993. **175**(4): p. 1087-94.

140. Roudier, N., et al., *Erythroid expression and oligomeric state of the AQP3 protein*. J Biol Chem, 2002. **277**(10): p. 7664-9.
141. Manley, D.M., et al., *Secondary structure and oligomerization of the E. coli glycerol facilitator*. Biochemistry, 2000. **39**(40): p. 12303-11.
142. Minetti, C.A. and D.P. Remeta, *Energetics of membrane protein folding and stability*. Arch Biochem Biophys, 2006. **453**(1): p. 32-53.
143. DeGrado, W.F., H. Gratkowski, and J.D. Lear, *How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles*. Protein Sci, 2003. **12**(4): p. 647-65.
144. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism*. Biochim Biophys Acta, 2005. **1751**(2): p. 119-39.
145. Schellman, J.A., *Symmetry Rules for Optical Rotation*. Account. Otem. Res., 1968. **1**: p. 144-151.
146. Cantor, C.R. and P.R. Schimmel, *Biophysical Chemistry Part I*. 1980, New York: W. H. Freeman and Company.
147. Miles, A.J. and B.A. Wallace, *Synchrotron radiation circular dichroism spectroscopy of proteins and applications in structural and functional genomics*. Chem. Soc. Rev., 2006. **35**: p. 39-51.
148. Whitmore, L. and B.A. Wallace, *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W668-73.
149. Compton, L.A. and W.C. Johnson, Jr., *Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication*. Anal Biochem, 1986. **155**(1): p. 155-67.
150. Fasman, G.D., *Circular dichroism and the conformational analysis of biomolecules*. 1996, New York, N.Y.: Plenum. ix, 738p.
151. Creighton, T.E., *Proteins: Structure and Molecular Properties*. 1992, New York: W. H. Freeman and Company.
152. Havel, H.A., *Spectroscopic methods for determining protein structure in solution*. 1996, New York ; Weinheim: VCH. xii, 250p.
153. Beechem, J.M. and L. Brand, *Time-resolved fluorescence of proteins*. Annu Rev Biochem, 1985. **54**: p. 43-71.
154. Demchenko, A.P., *Fluorescence analysis of protein dynamics*. Essays Biochem, 1986. **22**: p. 120-57.
155. Eftink, M.R., *Fluorescence Techniques for Studying Protein-Structure*. Methods of Biochemical Analysis, 1991. **35**: p. 127-205.
156. Farrelly, J.G., J.W. Longworth, and M.P. Stulberg, *The interaction of phenylalanyl transfer ribonucleic acid synthetase and phenylalanine transfer ribonucleic acid. Complex formation and resulting fluorescence quenching*. J Biol Chem, 1971. **246**(5): p. 1266-70.
157. Lakowicz, J.R., et al., *Anisotropy-based sensing with reference fluorophores*. Anal Biochem, 1999. **267**(2): p. 397-405.
158. Weber, G., *Fluorescence-polarization spectrum and electronic-energy transfer in proteins*. Biochem J, 1960. **75**: p. 345-52.
159. Weber, G., *Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds*. Biochem J, 1960. **75**: p. 335-45.

160. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*. 1983, New York: Kluwer Academic.
161. Volotovskii, I.D., S.V. Konev, and E.A. Chernitskii, [*Kinetics of low temperature after luminescence of tryptophan and proteins*]. *Biofizika*, 1967. **12**(3): p. 421-6.
162. Reshetnyak, Y.K., Y. Koshevnik, and E.A. Burstein, *Decomposition of protein tryptophan fluorescence spectra into log-normal components. III. Correlation between fluorescence and microenvironment parameters of individual tryptophan residues*. *Biophysical Journal*, 2001. **81**(3): p. 1735-1758.
163. Hershberger, M.V., R. Lumry, and R. Verrall, *The 3-methylindole/n-butanol exiplexes evidence for two exiplex sites in indole compounds*. *Photochem. Photobiol.*, 1981. **33**: p. 609-617.
164. Studier, F.W. and B.A. Moffatt, *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes*. *J Mol Biol*, 1986. **189**(1): p. 113-30.
165. Miroux, B. and J.E. Walker, *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**(3): p. 289-98.
166. Manley, D. and J.D. O'Neil, *Preparation of glycerol facilitator for protein structure and folding studies in solution*. *Methods Mol Biol*, 2003. **228**: p. 89-101.
167. Morein, S., D. Henricson, and L. Rilfors, *Separation of Inner and Outer-Membrane Vesicles from Escherichia-Coli in Self-Generating Percoll Gradients*. *Analytical Biochemistry*, 1994. **216**(1): p. 47-51.
168. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. *Nature*, 1970. **227**(5259): p. 680-5.
169. Schagger, H. and G. von Jagow, *Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form*. *Anal Biochem*, 1991. **199**(2): p. 223-31.
170. Wu, C.S. and G.C. Chen, *Adsorption of proteins onto glass surfaces and its effect on the intensity of circular dichroism spectra*. *Anal Biochem*, 1989. **177**(1): p. 178-82.
171. Robertson, A.D. and K.P. Murphy, *Protein structure and the energetics of protein stability*. *Chemical Reviews*, 1997. **97**(5): p. 1251-1267.
172. Lopez-Llano, J., et al., *The long and short flavodoxins - II. The role of the differentiating loop in apoflavodoxin stability and folding mechanism*. *Journal of Biological Chemistry*, 2004. **279**(45): p. 47184-47191.
173. Sehgal, P., J.E. Mogensen, and D.E. Otzen, *Using micellar mole fractions to assess membrane protein stability in mixed micelles*. *Biochim Biophys Acta*, 2005. **1716**(1): p. 59-68.
174. Wolfram, S., *The mathematica book*. 5th ed ed. 2004, Champaign, Ill. ; London: Wolfram Media : Turnaround. 1464 p.
175. Sambrook, J. and D.W. Russel, *Molecular Cloning, A laboratory manual*. Vol. 3. 2001, Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
176. Jiang, J., B.V. Daniels, and D. Fu, *Crystal structure of AqpZ tetramer reveals two distinct Arg-189 conformations associated with water permeation through the narrowest constriction of the water-conducting channel*. *J Biol Chem*, 2006. **281**(1): p. 454-60.

177. Palanivelu, D.V., et al., *Co-axial association of recombinant eye lens aquaporin-0 observed in loosely packed 3D crystals*. J Mol Biol, 2006. **355**(4): p. 605-11.
178. Hong, H. and L.K. Tamm, *Elastic coupling of integral membrane protein stability to lipid bilayer forces*. Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4065-70.
179. Moncelli, M.R., L. Becucci, and R. Guidelli, *The intrinsic pKa values for phosphatidylcholine, phosphatidylethanolamine, and phosphatidylserine in monolayers deposited on mercury electrodes*. Biophys J, 1994. **66**(6): p. 1969-80.
180. Fasman, G.D., *Distinguishing transmembrane helices from peripheral helices by circular dichroism*. Biotechnol Appl Biochem, 1993. **18 (Pt 2)**: p. 111-38.
181. Manley, D., *PhD Thesis: The overexpression, purification, and structural characterization of the Escherichia Coli integral membrane protein glycerol facilitator.*, in Chemistry. 2002, University of Manitoba: Winnipeg.
182. Kleinschmidt, J.H., *Membrane protein folding on the example of outer membrane protein A of Escherichia coli*. Cell Mol Life Sci, 2003. **60**(8): p. 1547-58.
183. Kleinschmidt, J.H., M.C. Wiener, and L.K. Tamm, *Outer membrane protein A of E. coli folds into detergent micelles, but not in the presence of monomeric detergent*. Protein Sci, 1999. **8**(10): p. 2065-71.
184. Lee, A.G., *Lipid-protein interactions in biological membranes: a structural perspective*. Biochim Biophys Acta, 2003. **1612**(1): p. 1-40.
185. Konidala, P., L. He, and B. Niemeyer, *Molecular dynamics characterization of n-octyl-beta-d-glucopyranoside micelle structure in aqueous solution*. J Mol Graph Model, 2006. **25**(1): p. 77-86.
186. Jensen, M.O. and O.G. Mouritsen, *Lipids do influence protein function-the hydrophobic matching hypothesis revisited*. Biochim Biophys Acta, 2004. **1666**(1-2): p. 205-26.
187. Tieleman, D.P., et al., *Lipid properties and the orientation of aromatic residues in OmpF, influenza M2, and alamethicin systems: molecular dynamics simulations*. Biochemistry, 1998. **37**(50): p. 17554-61.
188. Harries, W.E., et al., *The channel architecture of aquaporin 0 at a 2.2-A resolution*. Proc Natl Acad Sci U S A, 2004. **101**(39): p. 14045-50.
189. Lee, J.K., et al., *Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 A*. Proc Natl Acad Sci U S A, 2005. **102**(52): p. 18932-7.
190. Savage, D.F. and R.M. Stroud, *Structural basis of aquaporin inhibition by mercury*. J Mol Biol, 2007. **368**(3): p. 607-17.
191. Wiener, M.C., *When worlds collide*. Protein Sci, 2006. **15**(12): p. 2679-81.
192. Borgnia, M.J., et al., *Functional reconstitution and characterization of AqpZ, the E. coli water channel protein*. J Mol Biol, 1999. **291**(5): p. 1169-79.
193. Chandy, G., et al., *Comparison of the water transporting properties of MIP and AQP1*. J Membr Biol, 1997. **159**(1): p. 29-39.
194. Varadaraj, K., et al., *The role of MIP in lens fiber cell membrane transport*. J Membr Biol, 1999. **170**(3): p. 191-203.
195. Broekhuysse, R.M. and E.D. Kuhlmann, *Lens membranes. IV. Preparative isolation and characterization of membranes and various membrane proteins from calf lens*. Exp Eye Res, 1978. **26**(3): p. 305-20.
196. Kistler, J. and S. Bullivant, *Lens gap junctions and orthogonal arrays are unrelated*. FEBS Lett, 1980. **111**(1): p. 73-8.

197. Hill, W.G., R.L. Rivers, and M.L. Zeidel, *Role of leaflet asymmetry in the permeability of model biological membranes to protons, solutes, and gases*. J Gen Physiol, 1999. **114**(3): p. 405-14.
198. Jaspers, H.T. and J. van Steveninck, *Transport-associated phosphorylation of 2-deoxy-D-glucose in Saccharomyces fragilis*. Biochim Biophys Acta, 1975. **406**(3): p. 370-85.
199. Bisson, L.F. and D.G. Fraenkel, *Involvement of kinases in glucose and fructose uptake by Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A, 1983. **80**(6): p. 1730-4.
200. Fasman, G.D., *Distinguishing transmembrane helices from peripheral helices by circular dichroism*. Biotechnol Appl Biochem, 1993. **18 (Pt 2)**: p. 111-38.
201. Haltia, T. and E. Freire, *Forces and factors that contribute to the structural stability of membrane proteins*. Biochim Biophys Acta, 1995. **1241**(2): p. 295-322.
202. Schwalbe, H., et al., *Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea*. Biochemistry, 1997. **36**(29): p. 8977-91.
203. Bond, C.J., et al., *Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway*. Proc Natl Acad Sci U S A, 1997. **94**(25): p. 13409-13.
204. Subramaniam, S., et al., *Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin*. Embo J, 1993. **12**(1): p. 1-8.
205. Jackson, M.B. and J.M. Sturtevant, *Phase transitions of the purple membranes of Halobacterium halobium*. Biochemistry, 1978. **17**(5): p. 911-5.
206. Brouillette, C.G., D.D. Muccio, and T.K. Finney, *pH dependence of bacteriorhodopsin thermal unfolding*. Biochemistry, 1987. **26**(23): p. 7431-8.
207. Shnyrov, V.L. and P.L. Mateo, *Thermal transitions in the purple membrane from Halobacterium halobium*. FEBS Lett, 1993. **324**(2): p. 237-40.
208. Brouillette, C.G., et al., *Structure and thermal stability of monomeric bacteriorhodopsin in mixed phospholipid/detergent micelles*. Proteins, 1989. **5**(1): p. 38-46.
209. Heyes, C.D. and M.A. El-Sayed, *The role of the native lipids and lattice structure in bacteriorhodopsin protein conformation and stability as studied by temperature-dependent Fourier transform-infrared spectroscopy*. J Biol Chem, 2002. **277**(33): p. 29437-43.
210. Landin, J.S., M. Katragadda, and A.D. Albert, *Thermal destabilization of rhodopsin and opsin by proteolytic cleavage in bovine rod outer segment disk membranes*. Biochemistry, 2001. **40**(37): p. 11176-83.
211. Haltia, T., et al., *Thermodynamic and structural stability of cytochrome c oxidase from Paracoccus denitrificans*. Biochemistry, 1994. **33**(32): p. 9731-40.
212. Morin, P.E., D. Diggs, and E. Freire, *Thermal stability of membrane-reconstituted yeast cytochrome c oxidase*. Biochemistry, 1990. **29**(3): p. 781-8.
213. Morin, P.E. and E. Freire, *Direct calorimetric analysis of the enzymatic activity of yeast cytochrome c oxidase*. Biochemistry, 1991. **30**(34): p. 8494-500.

214. Rigell, C.W., C. de Saussure, and E. Freire, *Protein and lipid structural transitions in cytochrome c oxidase-dimyristoylphosphatidylcholine reconstitutions*. *Biochemistry*, 1985. **24**(20): p. 5638-46.
215. Rigell, C.W. and E. Freire, *Differential detergent solubility investigation of thermally induced transitions in cytochrome c oxidase*. *Biochemistry*, 1987. **26**(14): p. 4366-71.
216. Semin, B.K., M. Saraste, and M. Wikstrom, *Calorimetric studies of cytochrome oxidase-phospholipid interactions*. *Biochim Biophys Acta*, 1984. **769**(1): p. 15-22.
217. Yu, C.A., S.H. Gwak, and L. Yu, *Studies on protein-lipid interactions in cytochrome c oxidase by differential scanning calorimetry*. *Biochim Biophys Acta*, 1985. **812**(3): p. 656-64.
218. Thompson, L.K., et al., *Molecular basis of the heat denaturation of photosystem II*. *Biochemistry*, 1989. **28**(16): p. 6686-95.
219. Thompson, L.K., J.M. Sturtevant, and G.W. Brudvig, *Differential scanning calorimetric studies of photosystem II: evidence for a structural role for cytochrome b559 in the oxygen-evolving complex*. *Biochemistry*, 1986. **25**(20): p. 6161-9.
220. Davio, S.R. and P.S. Low, *Characterization of the calorimetric C transition of the human erythrocyte membrane*. *Biochemistry*, 1982. **21**(15): p. 3585-93.
221. Snow, J.W., J. Vincentelli, and J.F. Brandts, *A relationship between anion transport and a structural transition of the human erythrocyte membrane*. *Biochim Biophys Acta*, 1981. **642**(2): p. 418-28.
222. Marsh, D., *Lipid interactions with transmembrane proteins*. *Cell Mol Life Sci*, 2003. **60**(8): p. 1575-80.
223. Bose, K. and A.C. Clark, *Dimeric procaspase-3 unfolds via a four-state equilibrium process*. *Biochemistry*, 2001. **40**(47): p. 14236-42.
224. Rosengarh, A., J. Rosgen, and H.J. Hinz, *Slow unfolding and refolding kinetics of the mesophilic Rop wild-type protein in the transition range*. *European Journal of Biochemistry*, 1999. **264**(3): p. 989-995.
225. Lassalle, M.W. and H.J. Hinz, *Unfolding of the tetrameric loop deletion mutant of ROP protein is a second-order reaction*. *Biochemistry*, 1998. **37**(23): p. 8465-8472.
226. Nozaki, Y. and C. Tanford, *The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions*. *J Biol Chem*, 1963. **238**: p. 4074-81.
227. Nozaki, Y. and C. Tanford, *The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale*. *J Biol Chem*, 1971. **246**(7): p. 2211-7.
228. Prakash, V., et al., *Interactions of proteins with solvent components in 8 M urea*. *Arch Biochem Biophys*, 1981. **210**(2): p. 455-64.
229. O'Brien, E.P., et al., *Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism*. *J Am Chem Soc*, 2007. **129**(23): p. 7346-53.
230. Dobson, C.M., *Resting places on folding pathways*. *Curr Biol*, 1992. **2**(7): p. 343-5.

231. Dobson, C.M., *Protein folding. Solid evidence for molten globules*. *Curr Biol*, 1994. **4**(7): p. 636-40.
232. Ptitsyn, O.B., *Molten globule and protein folding*. *Adv Protein Chem*, 1995. **47**: p. 83-229.
233. Ptitsyn, O.B., *A determinable but unresolved problem*. *Faseb J*, 1996. **10**(1): p. 3-4.
234. Balbach, J., et al., *Following protein folding in real time using NMR spectroscopy*. *Nat Struct Biol*, 1995. **2**(10): p. 865-70.
235. Jennings, P.A. and P.E. Wright, *Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin*. *Science*, 1993. **262**(5135): p. 892-6.
236. Peng, Z.Y. and P.S. Kim, *A protein dissection study of a molten globule*. *Biochemistry*, 1994. **33**(8): p. 2136-41.
237. Haynie, D.T. and E. Freire, *Structural energetics of the molten globule state*. *Proteins*, 1993. **16**(2): p. 115-40.
238. Kuwajima, K., *The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure*. *Proteins*, 1989. **6**(2): p. 87-103.
239. Uversky, V.N., et al., *'All-or-none' mechanism of the molten globule unfolding*. *FEBS Lett*, 1992. **314**(1): p. 89-92.
240. Breton, J., et al., *Structure, stability and biological properties of a N-terminally truncated form of recombinant human interleukin-6 containing a single disulfide bond*. *Eur J Biochem*, 1995. **227**(1-2): p. 573-81.
241. De Filippis, V., et al., *Acid-induced molten globule state of a fully active mutant of human interleukin-6*. *Biochemistry*, 1996. **35**(35): p. 11503-11.
242. Strickland, E.H., *Aromatic contributions to circular dichroism spectra of proteins*. *CRC Crit Rev Biochem*, 1974. **2**(1): p. 113-75.
243. Lakowicz, J.R., *Fluorescence studies of structural fluctuations in macromolecules as observed by fluorescence spectroscopy in the time, lifetime, and frequency domains*. *Methods Enzymol*, 1986. **131**: p. 518-67.
244. O'Neil, J.D. and B.D. Sykes, *NMR studies of the influence of dodecyl sulfate on the amide hydrogen exchange kinetics of a micelle-solubilized hydrophobic tripeptide*. *Biochemistry*, 1989. **28**(2): p. 699-707.
245. Jensen, M.O., E. Tajkhorshid, and K. Schulten, *The mechanism of glycerol conduction in aquaglyceroporins*. *Structure*, 2001. **9**(11): p. 1083-1093.
246. Salzman, M., et al., *TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins*. *Proceedings of the National Academy of Sciences of the United States of America*, 1998. **95**(23): p. 13585-13590.
247. Redfield, C., *Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins*. *Methods*, 2004. **34**(1): p. 121-32.
248. Luo, W.B., X.L. Yao, and M. Hong, *Large structure rearrangement of colicin la channel domain after membrane binding from 2D C-13 spin diffusion NMR*. *Journal of the American Chemical Society*, 2005. **127**(17): p. 6402-6408.