The Overexpression, Purification, and Structural Characterization of the *Eshcerichia coli* Integral Membrane Protein Glycerol Facilitator.

By

Darren Manley

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

THE OVEREXPRESSION, PURIFICATION, AND STRUCTURAL CHARACTERIZATION OF THE *ESHCERICHIA COLI* INTEGRAL MEMBRANE PROTEIN GLYCEROL FACILITATOR

BY

DARREN MANLEY

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

To My Mom and Dad

You know a dream is like a river
Ever changin' as it flows
And a dreamer's just a vessel
That must follow where it goes
Trying to learn from what's behind you
And never knowing what's in store
Makes each day a constant battle
Just to stay between the shores.. and

I will sail my vessel
'Til the river runs dry
Like a bird upon the wind
These waters are my sky
I'll never reach my destination
If I never try
So I will sail my vessel
'Til the river runs dry

Garth Brooks

Table of Contents

Acknowledgments

I would first like to thank my supervisor Joe O'Neil for taking a chance on me years ago. It was Joe's ability to see the person behind the transcript that made him feel that I was worthy of such a chance. I only hope that now, after all these years, he doesn't regret his decision. Joe's patient and relaxed approach to life and research is contagious, and makes him an excellent teacher, psychologist, and friend. Thank-you, Joe.

If it were not for my parents, John and Jenny, I would not be here! By this I mean that if it were not for their sacrifice, leaving everything and everybody they knew and loved when they moved us from England to Canada in 1982, I would likely be unemployed and uneducated in England. I would not have had the opportunity to take my education as far as I have, and would have settled for much less in life. I thank them for the many years of support, encouragement, kind (and occasionally harsh) words, and unconditional love. It was my Parents that talked me out of quitting University, after a disastrous first year, and convinced me to give it another year (which I've managed to stretch into 10). Of course, my stepparents, Roger and Aline, were indistinguishable from my parents in their support, criticism, and love over the past 12 years. I would also like to thank the rest of my family, my sister Donna, brother-in-law Gerry, niece Amanda, nephew Adam, Nan Barnfield, Nan Walter, Uncle Paul, Aunt Donna, and my cousins Mark and Colin for their support and love throughout the last billion years that I've been at school. All I can say is thank-you, and I love you.

I can't forget to thank my many friends for their support, criticism, and vast amount of patience over the years. I would first like to thank my best-friend of 17 years, John Kafka, for his support, admiration, and the much needed butt kicking. John is the only person that tells me exactly

what I need to hear, rather than what I want to hear. Thank-you Johann. I thank Rick Oleschuk and Mark McComb for continually trying to convince me that I picked the wrong field and should have gone into Analytical Chemistry, and for keeping me sane throughout the years. I would also like to thank the short-lived "Climbing Club" for taking over where Rick and Mark left off. Of course there are also the many people that I have shared lab space with over the years, Dr. Lynda Donald, Dr. Dave Hosfield, Dr. Ayeda Ayed, Kelvin Luther, Xing Lee, Kelly (Turbo) Witzke, Jamie Galka, and Shaheen Shojana who have taught me more than they know. I thank my "normal" friends, Rob Falconer (R.F.), Samara Goeting, and Michelle McDonald who have wanted nothing to do with University and remind me on a regular basis that there is more to life than academics.

I would like to especially thank Dr. Gillian Henry who is an excellent scientist, remarkable individual, and a good friend who always knew the answer, regardless of the question. It was Gillian who taught me how to do research, to not simply follow a protocol but to understand it, and when necessary, to change it to better suit my needs. Gillian also provided me an outlet to express frustration, anger, bewilderment, or excitement about daily lab or life happenings during our daily coffee breaks and for that I also thank her.

Last, but definitely not least, I would like to thank Dawn, the mother of my child, for making the last year bearable. For many years I purposely kept myself away from the dangers of commitment, convincing myself that it was because I wanted to concentrate on my schoolwork. Dawn reached into my self-imposed prison and showed me just what love is, and that commitment is not something to be afraid of. Now, with Dawn, I have what my education could never give me, a family. Together we have 3 beautiful girls who I love more than I can express, Ashley, Morgan, and our new baby Madison. Now I am happy. I love you Dawn.

Sadly, my stepfather Roger passed away 6 months before completion of my thesis, and two weeks before the birth of my daughter Madison. Roger will never know how much I thank him for his support and love. However, I am thankful that he was with us long enough to know that my schooling was almost at an end, and that I am happy in life. I will always remember you.

To add to the sadness of Rogers passing, my Mother died 4 months after him. She loved Roger more than life itself, and I feel, died of a broken heart. My Mom was one of the people that I learned the most from about life, especially when to be stubborn. I am thankful that before she left us she got to know my Daughter Madison, and she knew that I was happy with the direction my life was going. Mom, I will always remember you with fondness and love. I hope I've made you proud.

List of Figures

List of Tables

List of Abbreviations

2-D : two-dimensional

3-D : three-dimensional

A : absorption

A_{280} : absorbance at 280 nm

AQP : aquaporin

BS : bis(sulfosuccinimidyl) suberate

CCA : Convex Constraint Algorithm

CD : Circular Dichroism

CHIP : Channel-forming Integral Protein

CMC : critical micelle concentration

DM : β-D- dodecylmaltoside

DSG : disuccinimidyl gluterate

DSS : disuccinimidyl suberate

E : energy

ε : molar absorptivity

G3P : glycerol-3-phosphate

GF : glycerol facilitator

GK : glycerol kinase

h : Planck's constant

H_{II} : hexagonal II phase

IMAC : immobilized metal affinity chromatography

I_o :incident radiation

IPTG : isopropyl-β-D-thiogalacto-pyranoside

I_t : transmitted radiation

kDa : kilodalton

K_w : partition coefficient of water

λ : wavelength

L_α : liquid crystalline

L_β : lamellar gel phase

LB : Luria-Bertani

LCPL: left circularly polarized light

LDAO : lauryldimethylamine oxide

MALDI-TOFMS : matrix-assisted laser desorption ionization time-of-flight mass spectrometry

MIP : major intrinsic protein

M_r : relative mass

NHS ester : *N*-hydroxysuccinimide ester

NMR : Nuclear Magnetic Resonance

NPA : Asparagine-Proline-Alanine

NTA : nitrilotriacetic acid

OD : optical density

OG : octyl-β-D-glucopyranoside

P_β : ripple

PCR : polymerase chain reaction

P_{dw} : diffusional water permeability

P_f : osmotic water permeability

P_{fp} : transport rate across a membrane containing n macroscopic pores of radius r

P_{fT} : total flow of water within narrow pores

PU : polyurethane

θ_{obs} : observed ellipticity

RCPL: right circularly polarized light

SDS : sodium dodecyl sulfate

SDS-PAGE : sodium dodecyl sulphate-polyacrylamide gel electrophoresis

TB : terrific broth

T_m : temperature at the transition midpoint

υ : frequency

$[\theta]_\lambda$: molar ellipticity

Abstract

The Major Intrinsic Proteins are found throughout the bacterial, plant, and animal kingdoms and are responsible for the rapid transport of water and other small, polar solutes across membranes. The superfamily includes the aquaporins, the aquaglyceroporins, and the glycerol facilitators.

We have overexpressed and purified the *Escherichia coli* inner membrane glycerol facilitator using a plasmid expression system. Following optimization of expression, approximately 7.0 milligrams of 95% pure protein are obtained from one liter of *Escherichia coli* cells using immobilized metal affinity chromatography. Well-resolved matrix-assisted laser desorption ionization mass spectra were obtained by solubilization of the protein in octyl-β-D-glucopyranoside (M_r = 33,650.3 Da; error ~ 0.4%). The recombinant glycerol facilitator was inserted into the bacterial inner membrane, was functional, and was inhibited by $HgCl_2$. Polyacrylamide gel electrophoresis suggests that the facilitator is predominantly monomeric when solubilized with dodecyl-β-D-maltoside, octyl-β-D-glucopyranoside, and sodium dodecyl sulphate, but that it self-associates forming soluble oligomers when urea is used during extraction. Furthermore, the data suggest that the oligomeric state of glycerol facilitator on polyacrylamide gels is concentration dependent when solubilized in dodecyl-β-D-maltoside and, to a lesser degree in sodium dodecyl sulphate, where monomers, dimers, trimers, and tetramers are observed at high concentrations. Similar oligomeric species have been demonstrated to exist in the bacterial membrane by chemical cross-linking experiments.

Far-UV circular dichroism analysis shows that recombinant glycerol facilitator is predominantly α-helical. Helix content is significantly higher in protein prepared in the absence of urea (40-55%) than in its presence (30%). This observation indicates that glycerol facilitator may be in

a more denatured state after exposure to urea. The helix content of octyl-β-D-glucopyranoside-solubilized glycerol facilitator is intermediate at 42%.

Near-UV circular dichroism analysis showed that dodecyl-β-D-maltoside-solubilized glycerol facilitator adopts a stable tertiary structure. On the other hand, the tertiary structure of sodium dodecyl sulphate-solubilized protein is dynamic. The stability of the protein in the two detergents was examined by thermal denaturation studies where secondary and tertiary structure were monitored by circular dichroism. It is concluded that the protein adopts a molten-globule state in sodium doecyl sulphate.

CHAPTER 1

INTRODUCTION

1.1 The Cell Membrane

1.1.1 The Hydrophobic Effect

"Oil and water don't mix" and "like dissolves like" are widely misunderstood aphorisms. Liquid water consists of a dynamic and disordered network of hydrogen bonded molecules (5-7 kJ/mol), where each water molecule transiently hydrogen bonds to its 4 nearest neighbors [1]. When polar substances dissolve in water, they interfere with this hydrogen bond network such that hydrogen bonds between water molecules are replaced by hydrogen bonds with the solute. When the solute is nonpolar and unable to form hydrogen bonds, the water molecules rearrange to form a "water cage" around the solute to maintain the maximum number of hydrogen bonds [2]. Thus, enthalpy changes accompanying dissolution are minimized at the cost of an entropy decrease due to the ordering of water molecules around the nonpolar solute. In some cases, to minimize the area of ordered water molecules, and therefore the entropy decrease, nonpolar solutes are segregated from bulk water. This separation of nonpolar molecules from water is called the hydrophobic effect, and is the driving force behind many biological mechanisms including protein folding, aggregation, and oligomerization, stacking of polynucleotide bases, and lipid bilayer formation [2].

1.1.2 The Plasma Membrane

The plasma or cell membrane is a selective barrier between the inside and outside of a cell and acts to segregate cellular organelles and compartments, thus defining the struc-

ture and function of the cell. Cellular membranes are made up of a mixture of glycerophospholipids, phosphosphingolipids, glycosphingolipids and sterols (Figure 1-1), amphipathic molecules that self-associate into a two-dimensional array called a bilayer [1].

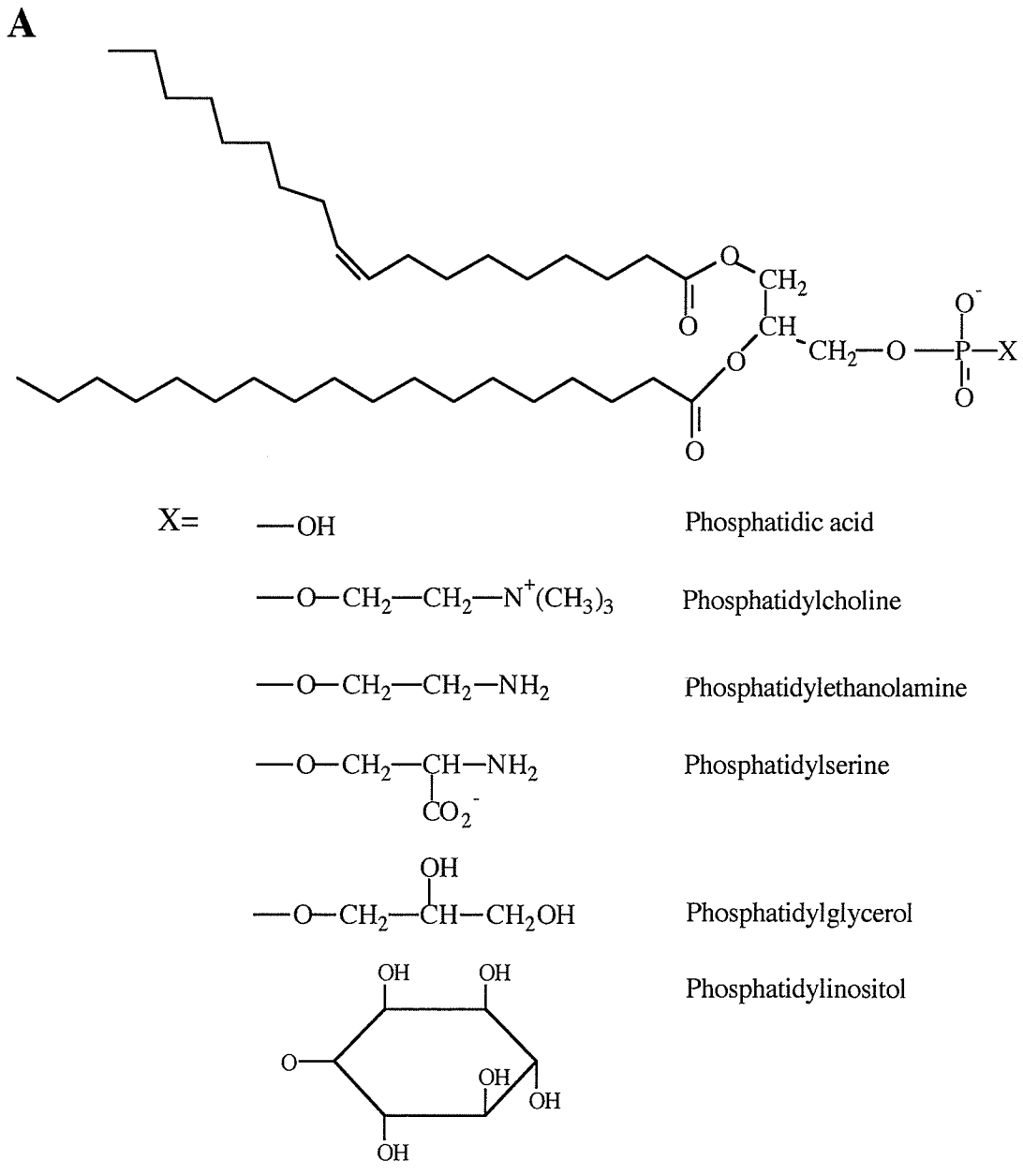
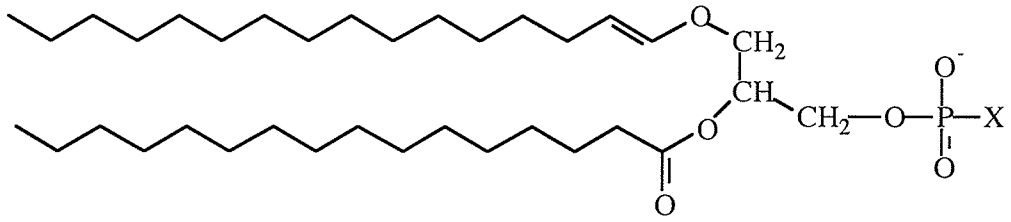


Figure 1-1. Structures of some membrane lipids. **A**, phospholipids.

B



X= —O—CH_2—CH_2—$N^+(CH_3)_3$ Choline plasmalogen

—O—CH_2—CH_2-NH_3^+ Ethanolamine plasmalogen

C

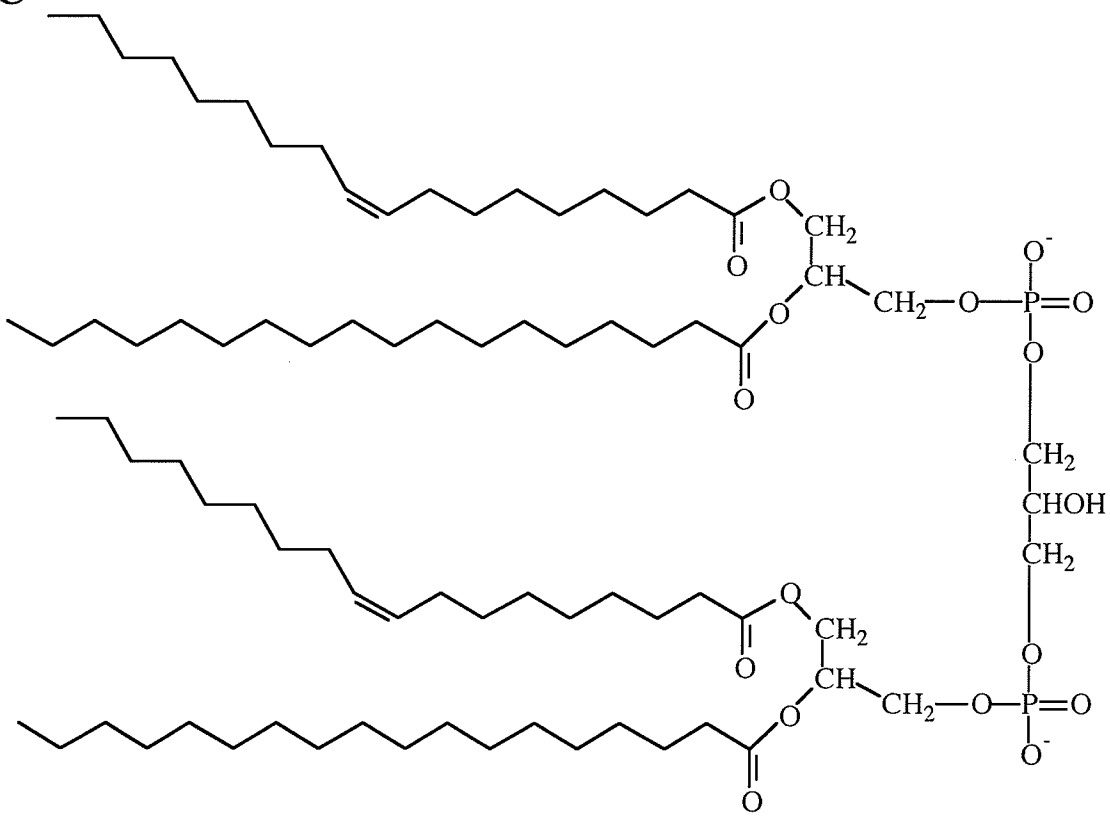
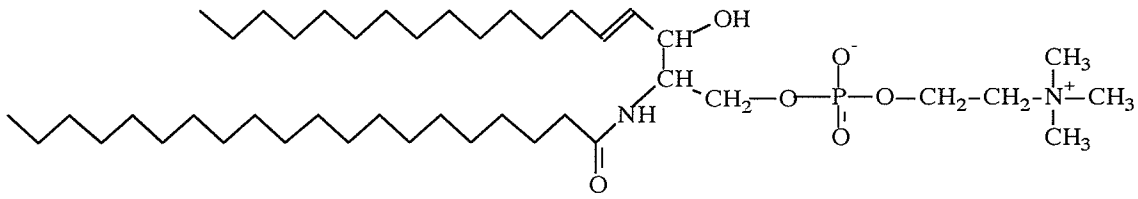


Figure 1-1 (continued). B, plasmalogen; **C,** cardiolipin.

D



E

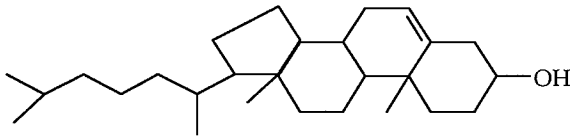


Figure 1-1 (continued). D, sphingomyelin; **E**. cholesterol.

The formation of the membrane bilayer is a direct result of the hydrophobic effect, as the nonpolar acyl chains of the lipids are isolated from the aqueous environment forming a hydrophobic barrier [2] (see Figure 1-2).

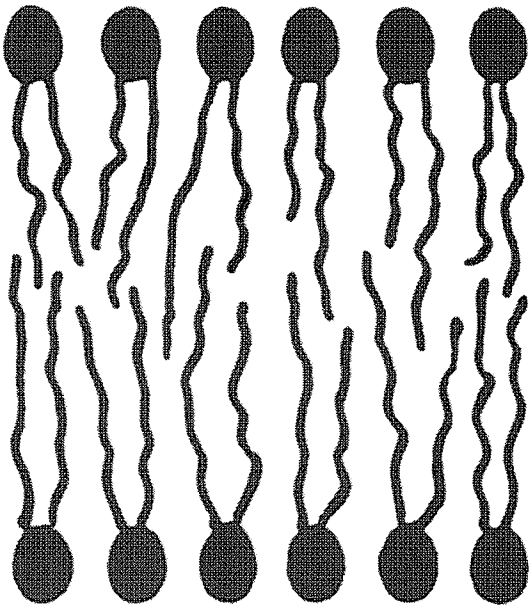


Figure 1-2. A cross-section of a phospholipid bilayer. The round segments represent the polar end of the phospholipid, the acyl chains are drawn as wavy lines.

Although the preferred conformation of phospholipids, both in the anhydrous state and in aqueous dispersions, is the bilayer form shown in Figure 1-2, other phases are possible [3]. The different phases formed by these amphipathic molecules depend on the structure and conformation of the amphipath, the nature of the solvent (whether aqueous or

nonpolar), and on ionic strength, pH, and temperature (see Figure 1-3) [4]. There are three

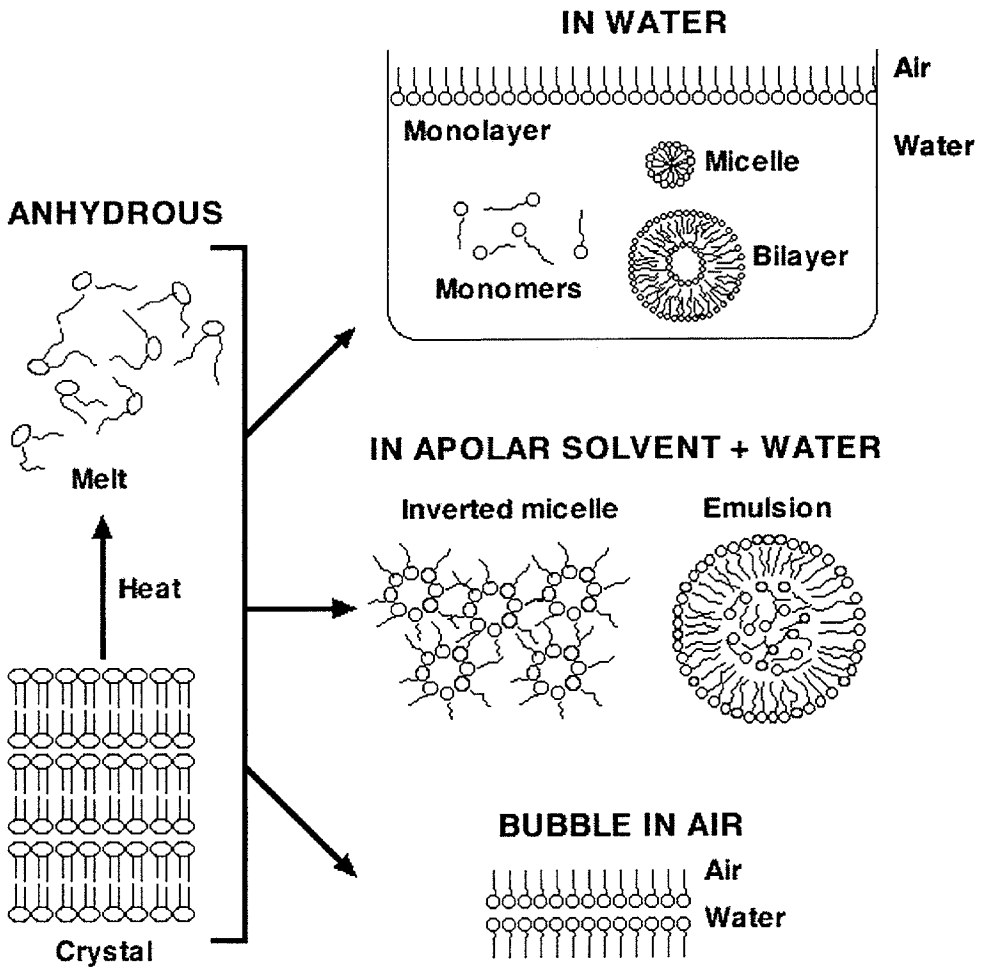


Figure 1-3. Organization of lipid dispersions in air, water and apolar solvents.

main classes of amphipath phases called hexagonal I (micelle), hexagonal II (inverted micelle), and lamellar (bilayer). The organization of each of the phases is shown in figure 1-4.

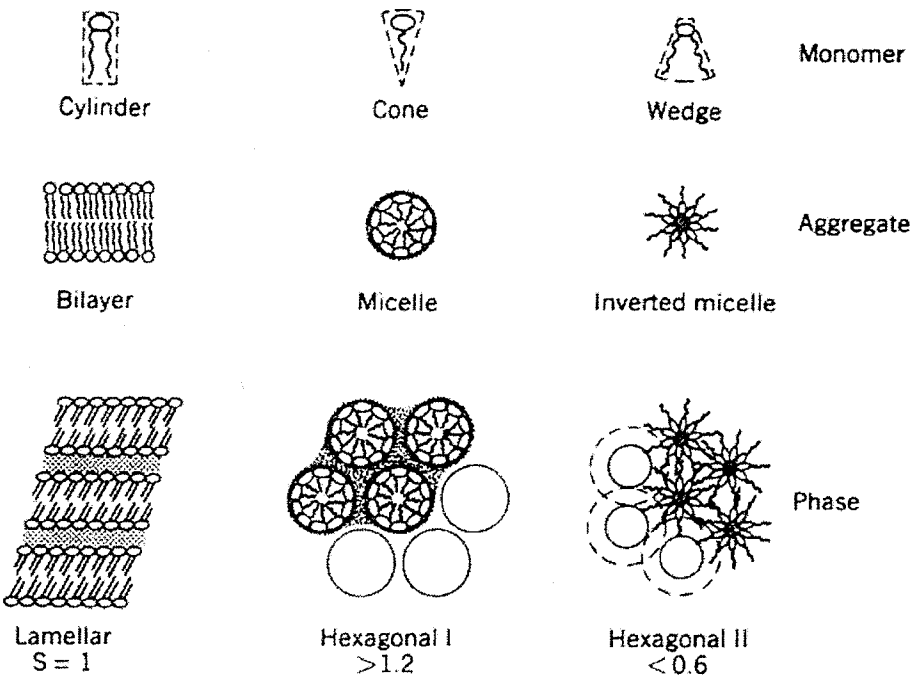


Figure 1-4. The shape of the phospholipid monomer determines the shape of the aggregate and its phase characteristics. The monomers in the inverted micelle and hexagonal II structures are drawn with one, rather than two, lipid tails for simplicity (adapted from reference 4 and used with permission).

The three phases have different macroscopic and molecular properties including osmotic pressure, mobility, X-ray diffraction characteristics arising from structural order and orientation, and the spectral characteristics arising from molecular motion of individual aggregate component molecules [1, 4]. Changes in the state of ionization and hydration of the amphipath contribute to macroscopic changes in the structure and organization of the lipid aggregates [5]. The packing parameter(s) in an aggregate can be defined as:

$$S = \frac{v}{a \bullet l} \tag{1-1}$$

where v is the amphipath hydrophobic volume, a is the area per head group at the interface, and l is the length of the hydrocarbon chain (see Figure 1-4) [4]. The term S is effectively the ratio of the hydrophobic:hydrophilic surface areas of a lipid and indicates how the prop-

erties of the monomers determine the macroscopic structures and properties of the aggregates. In general, for amphipaths in a mono- or bilayer phase, S=1. As the head group area decreases (S > 1) hexagonal II phase is preferred, while a larger head group area (S < 1) promotes micellar organization in hexagonal I phase [1, 4]. For the purpose of this thesis only the bilayer and micellar phases will be considered further.

1.1.2.1 Micelles

Amphipaths with large headgroups (S < 1/2), and (generally) one acyl chain form micelles in aqueous dispersions, where the acyl chains are sequestered in the center and the polar headgroups are solvent exposed (see Figures 1-3 and 1-4). The micelle shape is determined by the monomer and ranges from spheres of approximately 5 nm to cylinders of several hundred nanometers [4]. Micelles are commonly attributed to detergent molecules, but are also formed by lysophospholipids (phospholipids with only one acyl chain) [2]. At very low concentrations detergent monomers sit at the water/air interface so that the acyl chains are presented to the air and the polar headgroups interact with the bulk water[1]. As the detergent concentration increases the hydrophobic effect causes the monomers to aggregate into micelles. The detergent concentration at which the number of detergent monomers in the aqueous phase remain constant and in equilibrium with micelles is called the critical micelle concentration or CMC [6]. Once the CMC has been reached the number of micelles increases with increasing detergent concentration, while the number of free monomers is constant and in equilibrium with the monomers within the micelles [6] (see Figure 1-5).

1. It is this action by which detergent molecules disrupt the strong hydrogen bond network found at the water/air interface and gives detergents the label of surface-active, or surfactant, molecules.

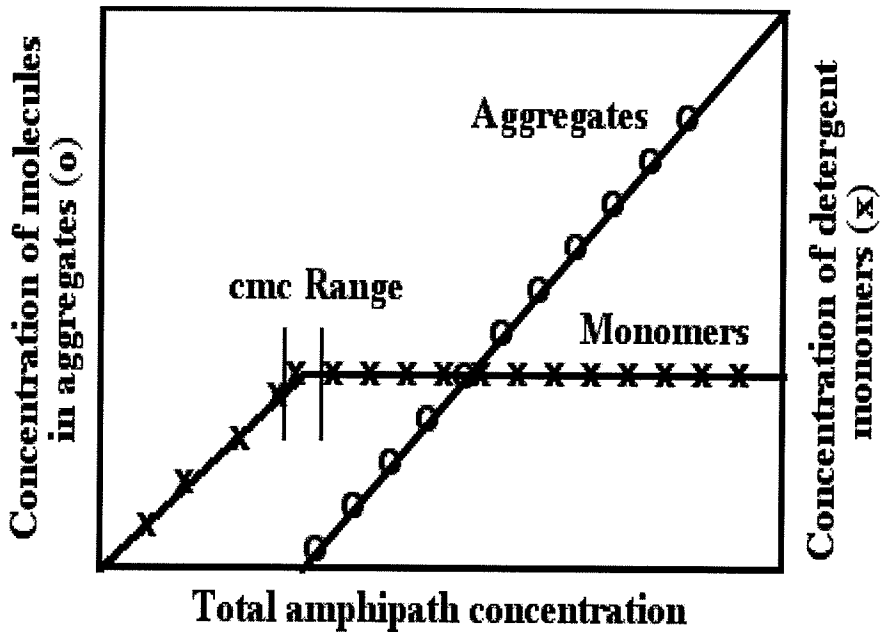


Figure 1-5. Relationship between monomer and micelle concentration as a function of total amphipath concentration (adapted from reference 2).

It can be shown that the CMC is related to the hydrophobic transfer free energy ($\Delta G°_{mic}$), due to exclusion of water from the nonpolar micelle interior, by the following equation:

$$\Delta G°_{mic} = (\mu°_M - \mu°_1) \cong kT \ln X_{CMC} \quad \textbf{(1-2)}$$

where $\mu°_1$ and $\mu°_M$ are respectively the standard state unitary free energies of the amphipath in monomeric form in aqueous solution and with M molecules per aggregate, k is Boltzmann's constant, T is temperature and X_{CMC} is the mole fraction of the amphipath where 50% of the detergent is in the form of micelles [1]. Formula 1-2 indicates that $\Delta G°_{mic}$ becomes more negative as the hydrophobicity of the amphipath increases and the CMC decreases; that is, water expulsion from the micelle is energetically more favorable [2]. Thus, detergent molecules with long aliphatic chains have lower CMCs than detergents with shorter chains.

The properties of the amphipath monomer, such as the size and charge of the head-group and the length of the acyl chain, determine the sizes and shapes of the micelles [7]. For example, the aggregation number (N), or number of monomers per micelle, is approximately 60 for sodium dodecyl sulfate (SDS), 140 for β-D- dodecylmaltoside (DM), and 4 for cholic acid (Figure 1-6) [4]. N is related to the relative standard state chemical potential

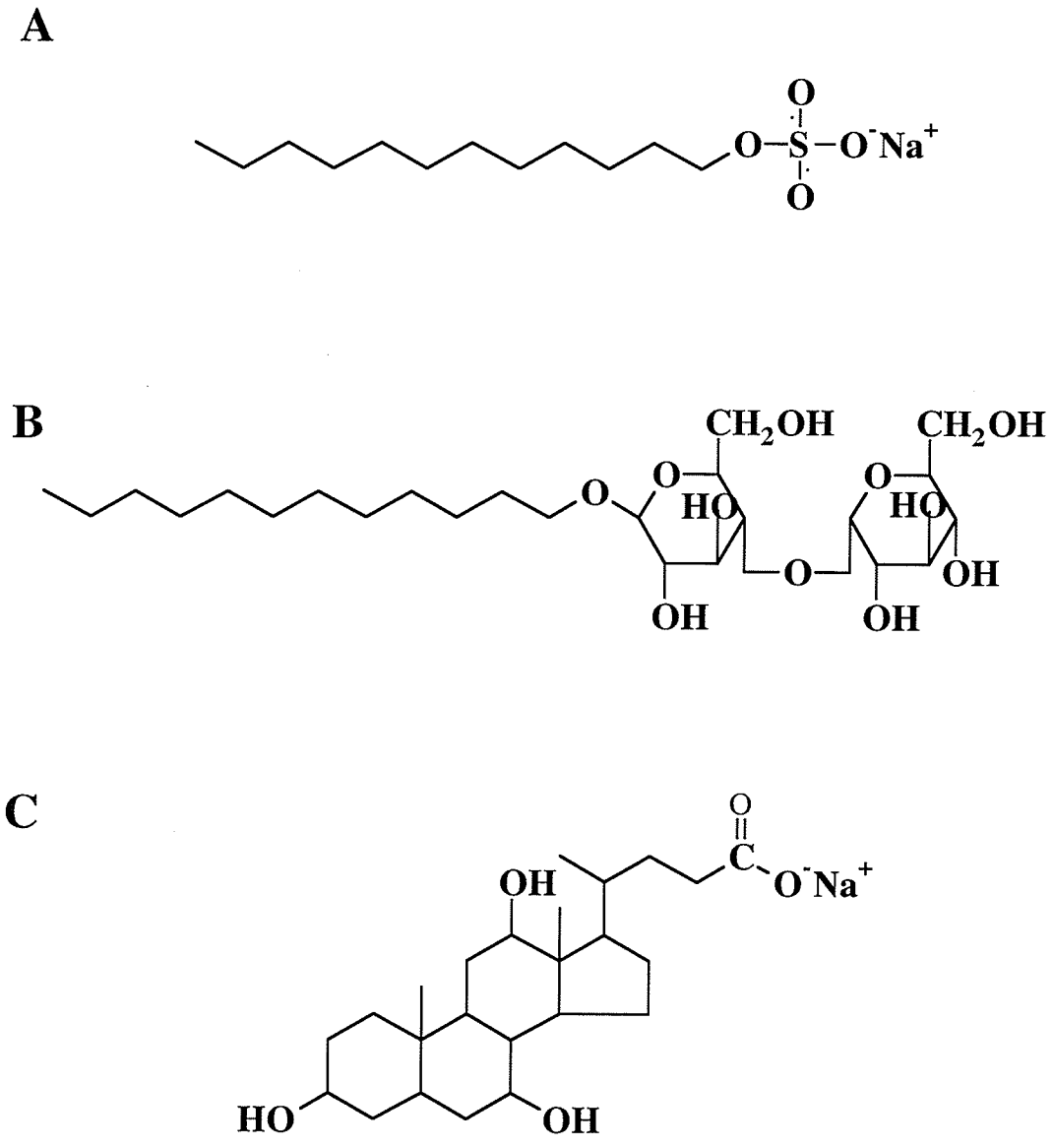


Figure 1-6. The molecular structures of **A**, sodium dodecylsulfate, **B**, β-D-dodecylmaltoside and **C**, cholic acid.

(or mean free energy) of the molecule in various aggregated forms given by μ°_N [4]. The value μ°_N depends on the average molecular surface area and allows the micelle geometry to be rationalized according to the following,

$$\mu^{\circ}_N = \gamma S + C/S + H_N \quad \textbf{(1-3)}$$

where γS and C/S are interfacial attractive and repulsive terms, respectively, and H_N is a "bulk energetic term" representing the free energy associated with the alkyl chains [1]. The γS and C/S terms are contributions from intermolecular interactions at the water-hydrocarbon interface, where γ is the surface tension expressed in units of energy per cm^2 (approximately 50 erg/cm^2), S is the surface area per molecule, and C is a constant representing intermolecular repulsion. C/S includes electrostatic and steric interactions that are repulsive. When molecules are close together (small area per molecule) this term is large. γS combines all the attractive molecular interactions at the interface. The optimal surface area per molecule is then given by minimizing free energy with respect to molecular surface area (i.e. $d\mu^{\circ}_N/dS = 0$) where,

$$S_o = \sqrt{C/\gamma}\,. \quad \textbf{(1-4)}$$

Therefore S_o (the optimal surface area occupied by the molecule at the hydrocarbon interface) is directly proportional to repulsive forces (C) and indirectly proportional to attractive forces (γ) [1]. It is the balance of the two opposing forces, hydrophobic attraction and electrostatic repulsion, that governs the sizes and shapes of detergent micelles [2]. The aggregation number is high for DM because electrostatic repulsion between the headgroups is weak whereas it is lower for SDS because of high repulsion between the negatively charged headgroups.

1.1.2.2 Bilayers and Biological Membranes

Detergent molecules have cone-like molecular shapes making spherical micelles the most thermodynamically stable aggregate as this shape maximizes headgroup and minimizes hydrocarbon interaction with solvent (Figure 1-4) [1, 4]. However, naturally occurring phospholipids (see Figure 1-1A) have a more cylindrical shape making micelle formation thermodynamically unfavorable, due to the "empty areas" that a phospholipid micelle would produce, such that water/hydrocarbon interactions would result (Figure 1-4). Thus, the most thermodynamically stable conformation of aqueous phospholipid dispersions is a lamellar bilayer (see Figures 1-3 and 1-4). An exception to this is an aqueous dispersion of pure phosphatidylethanolamine, where the headgroup surface area is sufficiently small that the molecular shape is an inverted cone making the hexagonal II (H_{II}) phase the thermodynamically stable aggregate (see Figure 1-4 [8]. Similarly, pure aqueous dispersions of phosphatidic acid and phosphatidylserine are fully protonated at acidic pH reducing the repulsive forces (C) between individual molecules, thereby reducing the surface area (see equation 1-4) shifting the preference from lamellar bilayer to H_{II} phase formation [1].

The lamellar bilayer phase can be divided into three distinct phases: liquid crystalline (L_α), ripple (P_β), and lamellar gel phase (L_β) (Figure 1-7) [1, 2, 4]. The liquid crystalline phase has two dimensional order but extensive disorder in the acyl-chains, and represents the state of the majority of lipids in biological membranes. The lamellar gel phase is formed at low temperatures with tighter packing of the phospholipid molecules resulting from a high degree of ordering of the acyl chains. The ripple phase is formed from the lamellar gel phase at intermediate temperatures, and represents the state in transition

from gel to liquid crystalline phase. It is called the ripple phase because of its wave-like appearance in electron micrographs [1, 2, 4].

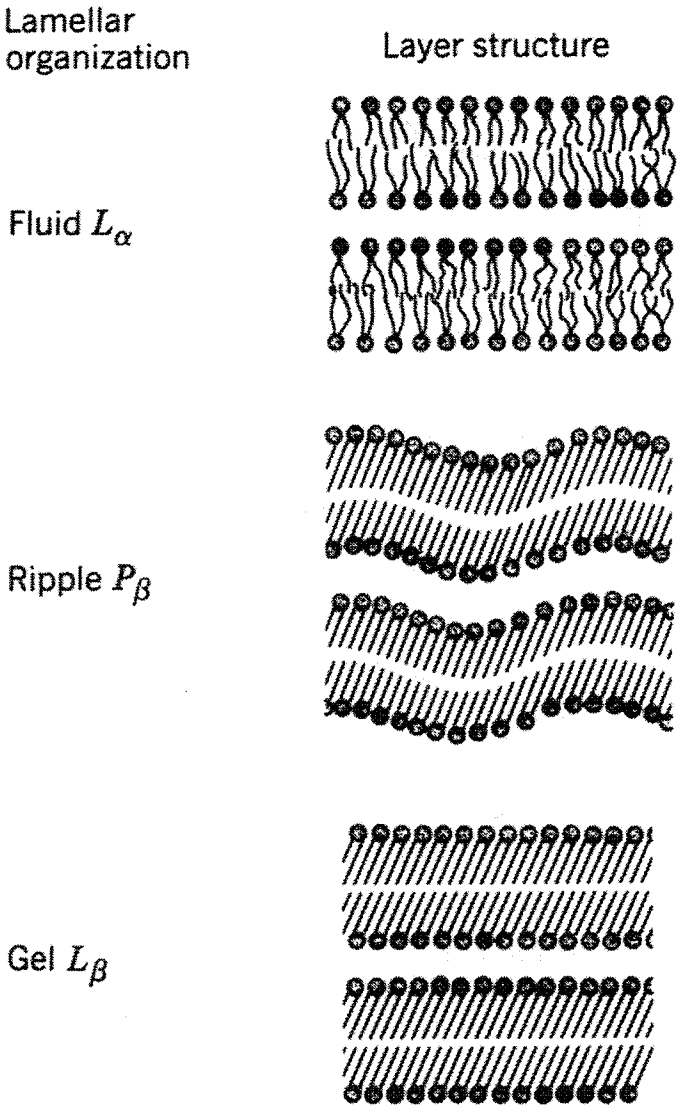


Figure 1-7. Lamellar bilayer phase in fluid, ripple and gel states. (See text for details. Taken from reference 4 and used with permission.)

In 1925 Gorter and Grendel extracted erythrocyte lipids with acetone, resuspended them in water, and found that the surface area occupied by the lipids is twice that of the calculated area of erythrocytes. This led to their suggestion that the erythrocyte membrane

is composed of a "bimolecular leaflet" [9]. Based on surface tension measurements Danielli and Davson [10] adapted the "bimolecular leaflet" model and suggested that the lipid head-groups, and hence the bilayer surfaces, are covered by layers of protein. However, technological advances have facilitated an improvement in the understanding of the membrane and the proteins associated with them. In particular, it was learned that membrane proteins are hydrophobic globular proteins with high degrees of α-helix content embedded in the membrane [11], and that intact biological membranes undergo phase transitions in temperature ranges similar to those of aqueous dispersions of phospholipids, i.e. $L_\alpha \rightarrow L_\beta$ transition [12]. The Singer and Nicolson "Fluid Mosaic Model" takes these facts into account and provides us with our current definition of biological membranes. The "Fluid Mosaic Model" states that the experimental results and thermodynamic considerations (i.e. hydrophobic effect) "fit in with the idea of a mosaic structure for membranes in which globular molecules of integral membrane proteins alternate with sections of phospholipid bilayer in the cross section of the membrane" with the "lipids constituting the matrix" in a "fluid rather than crystalline state" (see Figure 1-8) [13].

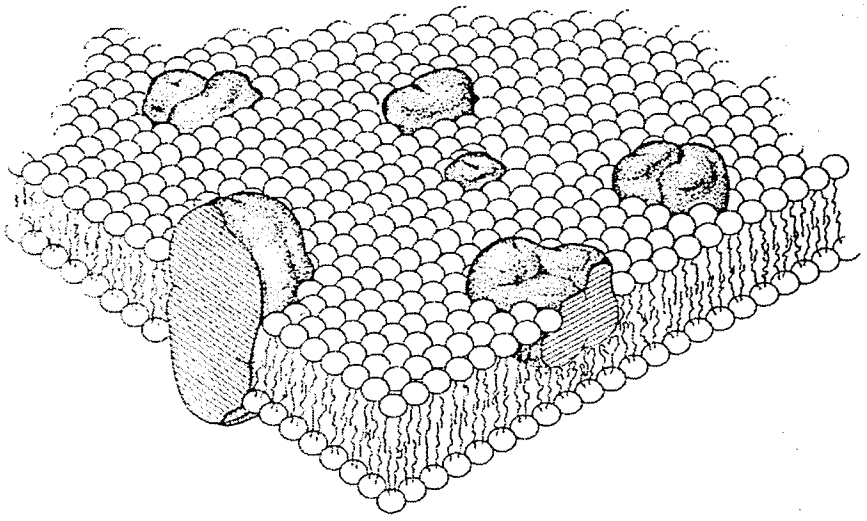


Figure 1-8. The Singer and Nicolson "Fluid Mosaic Model" of the cell membrane. (Taken from reference 13, and used with permission.)

1.1.3 Membrane Permeability

The ability of molecules to passively cross the membrane bilayer is governed by a solubility-diffusion mechanism [14]. The barrier which must be breached in order to cross the bilayer is the hydrocarbon core composed of the phospholipid acyl chains. This hydrocarbon core can be considered to be a simple hydrocarbon (i.e. liquid oil) membrane into which the solute of interest must partition and traverse. Thus, measurements that have been made to determine the free energy of transfer of solutes from aqueous solvents into hydrocarbon solvents, can be considered similar (if not identical) to their free energies of transfer across a lipid bilayer [2]. It can be understood, then, that hydrophilic polar molecules will have low permeabilities and small ions such as Na^+, K^+ and Cl^- will be virtually impermeable, whereas nonpolar and lipophilic molecules will display much higher permeabilities [15].

Since the partition coefficient for water into the membrane phase is small its concentration within the membrane is then very low. Therefore, as water partitions into the bilayer at each interface, and crosses from one side to the other by simple diffusion, it does so one water molecule at a time [14]. Osmosis occurs by a solubility-diffusion mechanism and it can be shown that the solute permeability coefficient for an oil membrane is directly proportional to the partition coefficient of water (K_w; expressed as the equilibrium concentration in the membrane divided by the concentration in the aqueous phase) into and diffusion constant (D_w, is on the order of 2.4×10^{-5} $cm^2 \bullet sec^{-1}$) across the hydrophobic interior of the membrane, and inversely proportional to the thickness of the membrane (δ; cm) as follows,

$$P_f = \frac{D_w \cdot K_w \cdot \overline{V}_w}{\delta \cdot \overline{V}_{oil}} \quad \textbf{(1-5)}$$

where $\overline{V}_{oil}$ and $\overline{V}_w$ are the partial molar volumes of oils and water, respectively. Essentially, we can see that P_f provides a measure of the speed at which an impermeant solute (e.g. water) can cross a membrane in response to an external pressure, either osmotic or hydrostatic [14, 15].

Another measure of a membrane's water permeability is the transport rate of isotopically labeled water (P_{dw}) measured in the absence of any applied osmotic or pressure difference across the membrane. Because both P_f and P_{dw} occur by solubility-diffusion they will be equal as long as the concentration of water in the oil membrane is so low that interactions between water molecules can be ignored [16].

For transport across a membrane containing n macroscopic pores of radius r and length L, P_f is now given by P_{fp},

$$P_{f_p} = \frac{n}{A} \cdot \frac{\pi r^4 RT}{8L\eta \overline{V}_w} \quad \textbf{(1-6)}$$

where η is the viscosity of water [14]. Equation 1-6 assumes that the size of permeant molecule(s) is negligible compared to the dimensions of the pore. As pore dimensions become comparable with that of the permeant molecule, the expressions for P_f must take into account the radii of the permeant molecules, laminar flow generated by pressure differences between the two sides of the membrane, and a diffusional component of flow from the gradient in the chemical potential of water within the pore (assuming exclusion of large impermeant solutes on one or both sides of the membrane) [14]. The total flow of water within narrow pores (P_{fT})is given by,

$$P_{f_T} = \frac{n\pi r^2}{AL} \cdot \left(\frac{r^2 RT}{8\eta \bar{V}_w} \right) + P_{d_w} . \qquad \text{(1-7)}$$

For water transport across a membrane containing macroscopic pores then $P_f > P_{dw}$. This is mainly because laminar flow (P_f) through pores of radius r is proportional to r^4 whereas diffusional flow (P_{dw}) is proportional to r^2.

The relevance of P_f/P_{dw} ratios is apparent in studies on the water transport properties of the red cell membrane where P_f is calculated to be about 2 X 10^{-2} cm/sec [17] whereas that for P_{dw} is about 3 X 10^{-3} cm/sec [18]. These values provided early indications that the red cell membrane contained water channels or pores, as the P_f/P_{dw} ratio is about 6.7, this will be discussed further in section 1.3.1[1].

1.2 Membrane Proteins

1.2.1 Introduction

The lipid to protein (L/P) ratio varies greatly from one biological membrane to another. For example, the L/P ratios for myelin and erythrocytes are approximately 4 and 0.75, respectively [19]. This difference in protein content largely reflects the extent of material transport across the membrane from one side to the other, with a higher protein content reflecting a greater need for transport. The functions of membrane proteins include acting as solute channels and pores, receptors, signal and energy transducers, and enzymes [21, 22, 23, 24]. Membrane proteins are classified into two categories according to their

1. It is important to note that due to the speed at which water crosses the membrane, both in the presence and absence of pores, stopped-flow instrumentation must be employed in its measurement [20].

mode of attachment to the membrane, termed peripheral and integral membrane proteins (see Figure 1-9) [13]. Peripheral proteins are those that are (a) removed from the membrane by disruption of protein-membrane electrostatic interactions or hydrogen bonds by mild treatment such as an increase in ionic strength or the addition of a chelating agent to the medium; (b) dissociate from the membrane free of lipids, and (c) are soluble in aqueous buffers free of detergent or lipid additives. The majority of membrane

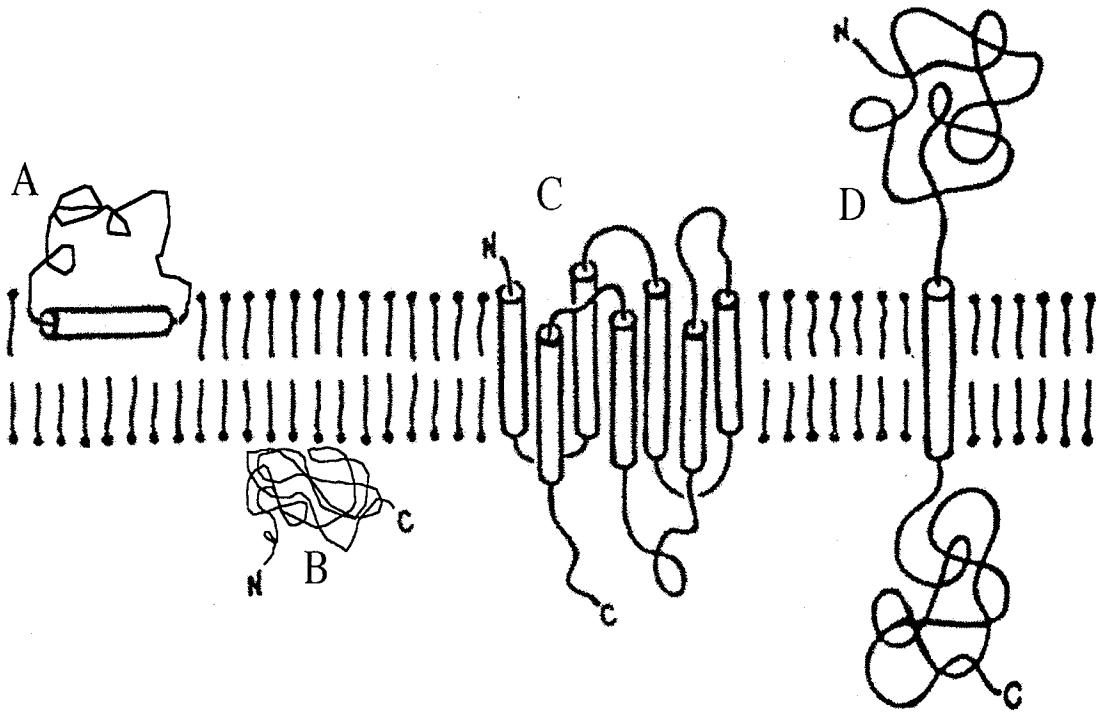


Figure 1-9. Diagrammatic representation of peripheral (B) and integral (A, C and D) membrane proteins. Protein A is an integral membrane protein that is embedded in one leaflet of the membrane bilayer. Protein C is an integral membrane protein that is folded within the membrane. Protein D is an integral membrane protein that has one transmembrane segment connecting domains folded in the aqueous phases on one or both sides of the membrane (adapted from reference 25). To simplify the figure the phospholipids have been drawn with only one, rather than two, lipid tails.

proteins belong to the integral membrane protein category, and have the following properties: (a) They are removed from the membrane only by treatments with detergents, bile salts, protein denaturants, or organic solvents. (b) They remain associated with lipid mole-

cules after extraction from the membrane, and (c) they are soluble in aqueous buffers only in the presence of detergent or lipid additives (see Figure 1-10) [13].

Extraction of integral membrane proteins requires disruption of the native membrane environment (see Figure 1-10) [26, 27]. While peripheral proteins interact with the membrane via electrostatic interactions, integral proteins are embedded within the membrane such that a portion of the protein, the transmembrane segment, crosses the membrane [13]. As a rule the transmembrane segment is an individual α-helix, and integral membrane proteins may have one or more such transmembrane helices (see Figure 1-9) [25]. Porin proteins are an exception to this α-helix rule and fold into a membrane-embedded, 8-18 stranded antiparallel β-barrel [28, 29].

1.2.2 Membrane Protein Structure Determination

Determining the structure of a protein at the atomic level by Nuclear Magnetic Resonance (NMR) Spectroscopy and X-ray diffraction requires that ample quantities of pure protein be readily available. Indeed, these sample requirements are equally important for low resolution structural analysis such as Circular Dichroism (CD) Spectropolarimetry and Mass Spectrometry. Structural studies of membrane proteins are made more difficult than their water soluble counterparts by their naturally low expression levels and the requirement that lipids or detergents be co-purified or added *in vitro* to retain protein structure [25, 27a]. For example, the long sought and recently discovered calcium-release-activated calcium channel protein is estimated to exist in about 100 copies per cell in native resting lymphocytes and in about 10 copies per cell in wild-type CHO-K1 cells [27b]. To date 15385 structures have been entered into the protein data bank [31]. Of these, 12690 structures were solved by X-ray diffraction and other techniques, 2368 by NMR, and 327 by theoret-

ical modeling. Although it is estimated that one-quarter of the proteins in many organisms are membrane proteins, only 246 of the entries are for membrane protein structures. However, there is more than one entry for some proteins, such as the 19 entries

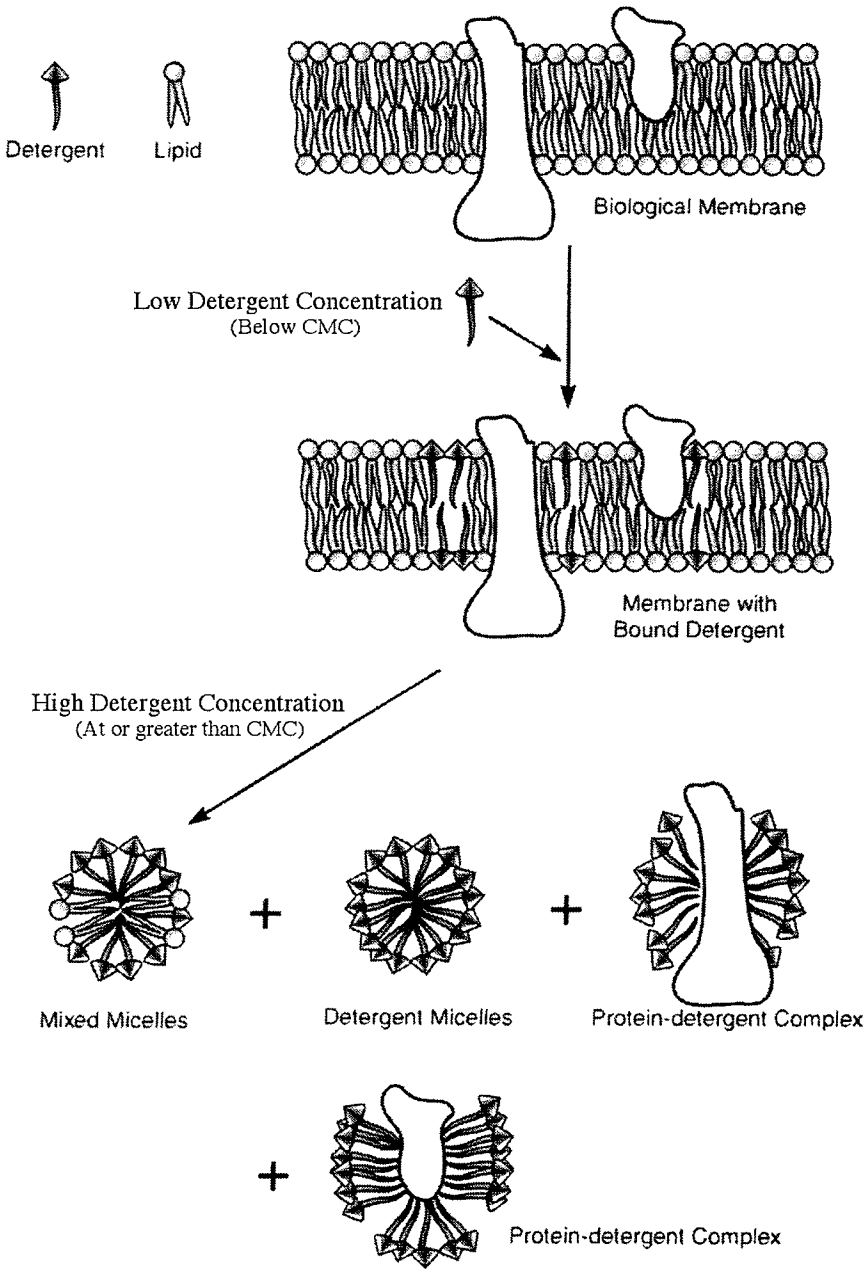


Figure 1-10. Detergent solubilization of cell membranes and extraction of membrane proteins (adapted from A Guide to the Properties and Uses of Detergents in Biology and Biochemistry by Calbiochem [30], and used with permission).

for bacteriorhodopsin, often indicating structures of different protein mutants. Of the 246 membrane proteins there are 31 unique structures solved by NMR, 3 by electron cryocrystallography, 12 by theoretical modeling, and the remaining 200 (not accounting for repeat entries) by X-ray diffraction.

Two-dimensional (2-D) crystals of membrane proteins have proved valuable for the determination of detailed three-dimensional (3-D) structures of many membrane proteins including bacteriorhodopsin [32], the light-harvesting chlorophyll a/b-protein complex [33], and aquaporin-1 [34]. These structures were acquired using the technique called electron cryocrystallography (also known as cryo-electron microscopy). There are several advantages to analyzing 2-D crystals by electron cryocrystallography and image analysis compared with X-ray diffraction analysis of 3-D crystals [35]. First, because images of the crystals are recorded, phases are calculated directly by Fourier transformation. In X-ray crystallographic analysis all phase information is lost and must be retrieved by heavy atom substitution of the protein crystal, which can often lead to cracking and destruction of the protein crystal [36]. Secondly, the amount of sample required for electron cryocrystallography is significantly lower than for X-ray analysis and this is important when the naturally low expression levels of membrane proteins are considered. Further, electron cryocrystallography permits the study of the membrane protein in its native membrane environment, rather than first extracting protein from the membrane and purifying it [35]. This technique is called *in situ* crystallization and can be accomplished by treating the membrane protein in its native membrane with mild detergents to extract lipid molecules and increase the protein/lipid ratio [16].

Electron cryocrystallography involves the use of a transmission electron microscope to acquire both electron-diffraction patterns, similar to X-ray diffraction, and high-resolution images of the macromolecule of interest [38]. Sample preparation involves the embedding of the solubilized or membrane reconstituted protein in a matrix, such as glucose or amorphous ice, that can preserve the high-resolution protein structure at the low temperatures and vacuum of the electron microscope. Glucose is the matrix of choice if high resolution structures are desired, mainly due to crystal surface irregularities produced by ice-embedding and the resultant increase in electron diffraction noise [38, 39]. After embedding the sample it may be negatively stained with heavy-atom salts and adsorbed on a carbon support film. Observation in the electron microscope is carried out under low electron doses and at low temperature (-160°C or less), conditions that reduce the extent of radiation damage of the specimen [39, 40]. Two-dimensional projections are individually obtained by tilting the 2-D crystals with respect to the electron beam. Computer processing is used to obtain the amplitudes of the structure factors from electron diffraction patterns, and to retrieve the corresponding phases from the electron micrographs [38]. The 3-D information can then be reconstructed by combining the data from a set of 2-D projections (see Figure 1-11).

Generally the structural resolution of proteins determined by cryoelectron crystallography is lower than those determined by X-ray diffraction. This is mainly due to the loss of structural data in the direction perpendicular to the direction of the tilt axis, and resolution being limited to the finest detail reproduced in the collected micrographs [34]. The technology used for the collection and processing of electron diffraction data is rapidly

improving and today is able to provide atomic resolution structures [41]. The main difficulty remains the formation of membrane protein crystals.

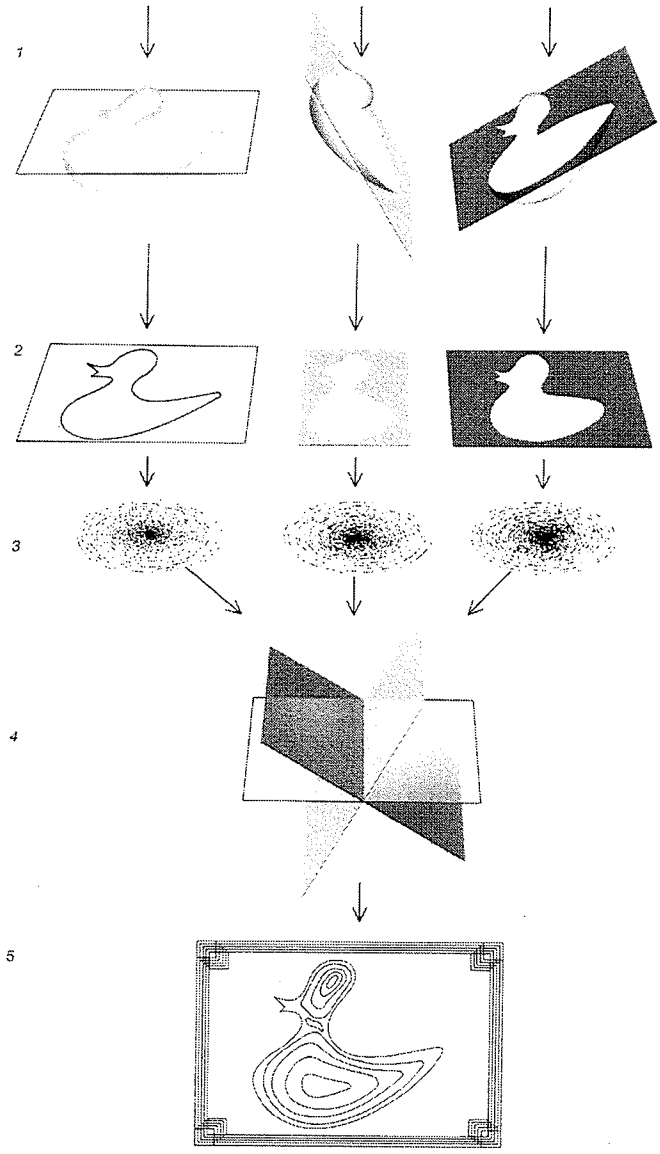


Figure 1-11. The diagrammatic representation of the steps involved in the determination of a three-dimensional structure from two-dimensional samples. (1) Specimens are tilted through a range of angles to the incident beam. (2) The resulting micrographs constitute a set of different projection views. (3) Fourier transforms are computed from the micrographs, each one is equivalent to a different plane through the origin of a three-dimensional Fourier transform. (4) From a large number of such planes the complete three-dimensional transform is computed (only three planes are shown). (5) From the computed three-dimensional transform a number of contour maps are plotted, each showing the density values at a different level in the membrane. When the maps are stacked, the complete three-dimensional structure is revealed. Taken from reference 42 and used with the authors' permission.

1.3 The Aquaporins

1.3.1 Introduction

In early studies on the water permeability of red blood cells it was found that trans-membrane water transport is faster than can be accounted for by simple diffusion [43, 44]. When an osmotic gradient is imposed across a cell membrane the ratio of osmotic water permeability (P_f) to diffusional permeability (P_{dw}) is 1 if there are no water channels (see section 1.1.3) [45]. Early research indicated that in the presence of an osmotic gradient P_f was greater than 0.1 cm/sec, more than 10 times the measured P_{dw} (< 0.01 cm/sec, measured in the absence of an osmotic gradient), suggesting the presence of a specific carrier for water [46]. Koefoed-Johnsen and Ussing [44] found that P_f increased in renal proximal tubules upon the addition of an antidiuretic hormone, which they explained as a widening of pore diameter. The work by Macey and Farmer [43] showed that water and urea cross the plasma membrane by independent pathways, and that P_f is reversibly inhibited in the presence of mercurials, suggesting the presence and inhibition of osmotic water flow, respectively. More recent analysis indicates that the P_f of the red cell membrane is about 2 $x10^{-2}$ cm/sec [47] whereas the P_{dw} is about 3 $x10^{-3}$ cm/sec [48], giving a P_f / P_{dw} ratio of about 6.7.

The first water-filled channel protein was discovered by Denker *et al.* in 1988 [49] as an abundant erythrocyte integral membrane protein of M_r 28,000 (28kDa), with one quarter of the protein N-glycosylated. This 28kDa protein co-purified with the Rh polypeptide and comprised approximately 20% of the protein recovered from initial purifications. Originally considered a degradation product it was shown to be immunologically distinct from the 32kDa Rh polypeptide [49]. A similar 28kDa protein was purified from rat kidney cells, and immunolocalized to the apical brush borders of renal proximal tubules in

human kidneys [49]. Other data suggested that the 28kDa protein exists as a dimer or oligomer in the native lipid membrane, with no intermolecular disulfides (electrophoretic mobility was unaffected by reducing agents) [51].

The cDNA for the 28 kDa protein, initially named *Ch*annel-forming *I*ntegral *P*rotein (CHIP) 28, was isolated from human bone marrow [52] and shown to function as a mercury-sensitive, water-selective, membrane channel by expression in Xenopus oocytes [53]. Mutagenesis experiments determined that of the four cysteines within CHIP28, cysteine 189 (Cys189) is the one that confers mercury-sensitivity. Further, replacing Cys189 with larger amino acids abolished the water permeability of the molecule [54]. Reconstitution of both highly [55] and partially purified [56] CHIP28 into proteoliposomes confirmed its function as a water channel, leading to the renaming of CHIP28 as Aquaporin-1 (AQP1) [57]. The low Arrhenius activation energy (<4 kcal/mol) indicates that AQP1-mediated water transport occurs as a continuous stream with the direction of flow determined by the osmotic gradient [55].

The discovery of AQP1 was the thin edge of the wedge, and subsequently it was found that in addition to animals, both plants and bacteria express membrane channels that are highly selective for water [58]. In *E. coli*, aquaporin-Z (AQPZ) permits the cell to respond to changes in the osmolarity of its environment (Calamita *et al.,* 1995). In mammals, AQPs 0, 1, 2, 4, 5, 6, and 8 easily discriminate between water and urea which have diameters of 2.8Å and 4.0Å, respectively (see Table 1-1) [59, 60]. However, transport assays suggest that the capacity to transport water varies widely among the aquaporins. For example, AQPs 2 and 4 are highly efficient water pores whereas AQP0 barely elevates water transport above passive diffusion in those cells in which it is expressed [59].

Table 1-1. Mammalian localization, molecules transported, presumed function, and mutant phenotype of the ten mammalian aquaporins.

Name (Mw)	Localization	Transport	Function	Mutant Phenotype
AQP0 (28 kDa)[a]	Lens fiber cells	Water	Lens transparency	Cataracts in knock-out mice)
AQP1 (28 kDa)[b]	Red blood cells; choroid plexus; kidney proximal tubule descending limb of Henle; eye; vascular endothelia	Water	Constitutively active water channel	Severe urinary concentrating defect in knockout mice (no clinical phenotype in humans)
AQP2 (28 kDa)[c]	Apical membrane renal collecting duct cells	Water	Vasopressin-regulated water channel	NDI in humans
AQP3 (27 kDa)[d]	Basolateral membrane renal collecting duct cells, trachea, nasopharynx and red blood cells	Water, glycerol and urea	Water and urea efflux pathway collecting duct cells	Unknown
AQP4 (30 kDa)[e]	Basolateral membrane renal collecting duct cells, trachea, nasopharynx, cerebellum and hypothalamus	Water	Water efflux pathway collecting duct cells; spinal fluid reabsorption; osmoreception	Partial NDI in knockout mice
AQP5 (28 kDa)[f]	Salivary and lacrimal glands; cornea; type I pneumocytes; upper airway secretory epithelia	Water	Secretion of tears, saliva and sputum	Hypertonic viscous saliva in knockout mice
AQP6 (30 kDa)[g]	Renal Epithelia	Water	Glomerular filtration; tubular endocytosis; acid-base metabolism	Unknown
AQP7 (~37 kDa)[h]	Testis; adipose tissue; renal proximal tubule	Water and glycerol	sperm cryoprotection	Unknown
AQP8 (34 kDa)[i]	Hepatocytes; pancreas; testis; colon; placenta; heart; salivary glands	Water	Unknown	Unknown
AQP9 (~32 kDa)[j]	Adipose tissue	Water, glycerol and urea	Unknown	Unknown

NDI, Nephrogenic diabetes insipidus

a-71; b-72; c-64; d-73; e-74; f-75.; g-76.; h-68; i-77; j-78;

1.3.2 The Major Intrinsic Protein Superfamily

The glycerol facilitators, aquaporins, and aquaglyceroporins comprise a superfamily of proteins [61] named after the first discovered aquaporin[1], the Major Intrinsic Protein (MIP) of the eye lens fiber [62]. The proteins are distinguishable by their pore specificities [58] amino acid sequences, and the organization of their genes [61]. Hydropathic analysis [63] suggested that the protein fold contains six membrane-spanning helices (1-6) connected by five loops (A-E) (Figure 1-12). Loops B and E contain a characteristic aspar-

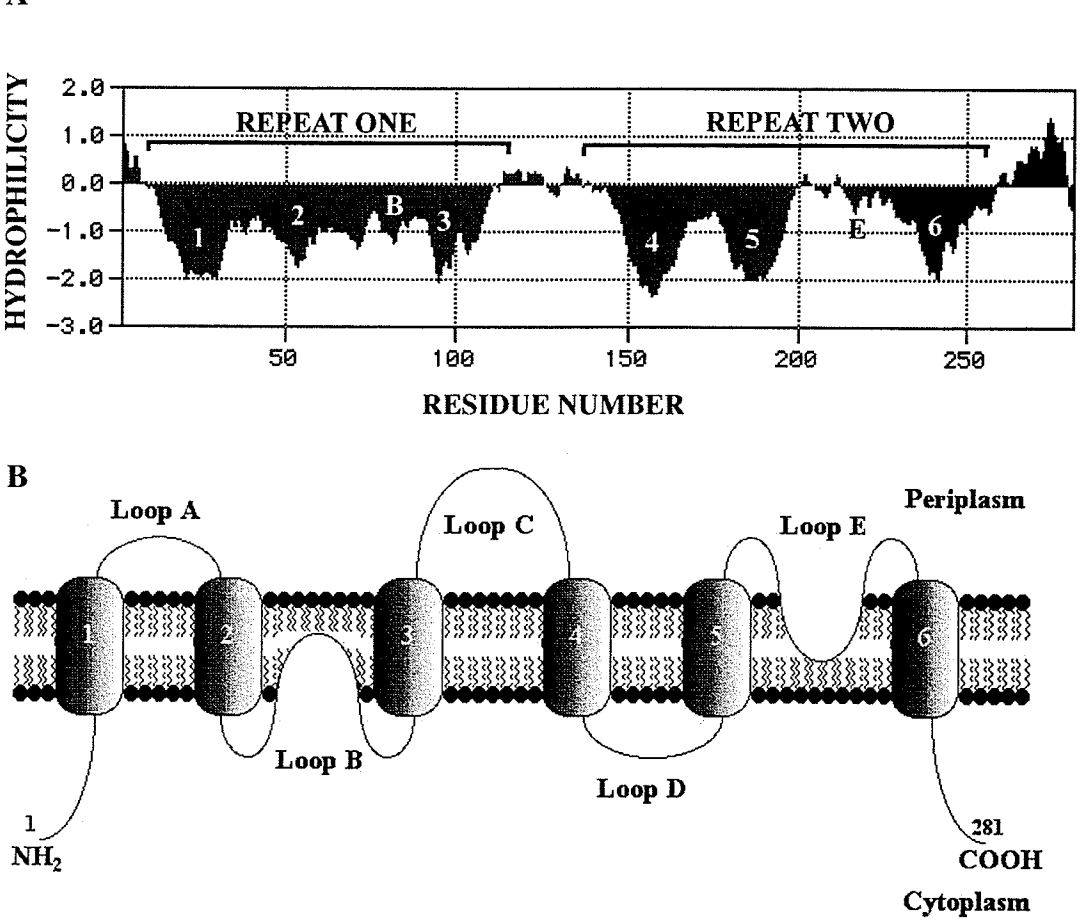


Figure 1-12. A) The Kyte-Doolittle hydrophilic plot (identical to hydropathy plot, except the values on the y-axis are reversed) of the *E. coli* glycerol facilitator. Negative hydrophilicity denotes hydrophobic regions (AQP1 hydropathy plot is similar). The six transmembrane regions are indicated, as are loops B and E and the region of internal homology. **B)** A model representing the topology of both AQP1 and *E. coli* glycerol facilitator. Shown are the six transmembrane segments (1-6) and the 5 loops (A-E).

1. It was not classified as an aquaporin until after the discovery of AQP1

agine-proline-alanine (NPA) motif [64] that is found in every member of the family and forms an important component of the channel pore [65, 66].

In bacteria, glycerol transport is essential for the efficient use of glycerol as a carbon source [67]. For example, *Escherichia coli* will express the inner membrane glycerol facilitator in the absence of glucose, so that glycerol may rapidly enter the cell and be used as a carbon source (see section 1.4) [67]. In mammals, glycerol transport is an essential activity in several tissues including hepatocytes and adipocytes, both of which take up glycerol released by triacylglycerol hydrolysis [68]. Adipocytes also export glycerol following lipolysis. Candidate glycerol transporters have been identified in human adipocytes and rat spermatids (aquaporin-7) [58, 68], in rat hepatocytes (aquaporin-9) [69], and in a variety of tissues that express aquaporin-3 [70]. Aquaporins-3, -7, and -9 are sometimes called aquaglyceroporins reflecting their dual specificities and their phylogenetic homology to the bacterial glycerol facilitators (see Table 1-1) [61].

1.3.3 Aquaporin Molecular Structure

Circular dichroism studies showed that in sonicated membrane suspensions AQP0 contains approximately 58% α-helix and 20% β-sheet [71]. Mutagenic and immunological studies indicated that the amino- and carboxy-termini are located intracellularly [79], and that loops A, C, and E are extracellular, and loops B and D are cytoplasmic. Hydrophobic loops B and E (indicated in Figure 1-12) each contain an NPA motif, thus the amino-terminal NPA motif in loop B is located on the cytoplasmic side and the carboxy-terminal motif in loop E is located on the extracellular side of the membrane [53]. Further, the N- and C-terminal halves of the MIP family display internal homology, centered around the NPA motifs, such that the two halves are orientated 180° to each other within the membrane (See

Figure 1-12) [52, 53, 54, and 81]. This tandem repeat of amino- and carboxy-terminal halves likely arose as a result of an ancient gene duplication [46]

AQP1 is expressed at very high levels in red blood cells (2 x 10^5 copies/cell [82]), making it a very good natural overexpression system. Thus, the majority of structural analyses of the MIP proteins has been performed on AQP1 [45]. Biochemical and freeze-fracture studies indicate that AQP1 exists as a homotetramer in membranes [82, 83], with an extracellular complex glycan attached at Asn-42 to one of the four subunits [82]. Radiation inactivation experiments [84] and co-expression of wild-type AQP1 and a mercury-resistant AQP1 mutant (C189S) [54] indicate that the monomers are the active water-transporters, rather than the center of the tetramer forming the channel. Mutations within loops B and E are not tolerated well in AQP1 [79]. This functional resistance to mutation, and the presence of the mercury-sensitive Cys189 and the NPA motifs within these loops, indicates the importance of loops B and E in channel function. Furthermore, the hydrophobicity of loops B and E suggest they may extend into the bilayer. The 'Hourglass Model' of AQP1 proposed that loops B and E fold into the lipid bilayer, forming an hourglass structure with a single aqueous pathway formed by the overlapping loops (see Figure 1-13).

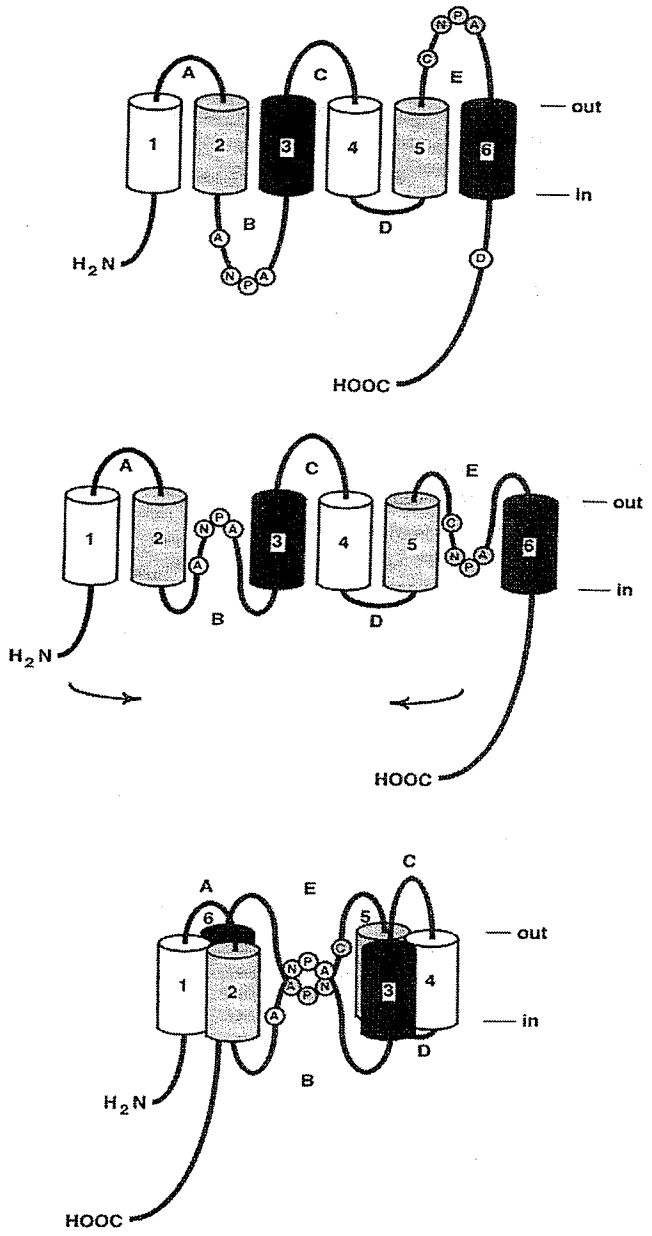


Figure 1-13. The Hourglass Model of Aquaporin-1. The six transmembrane segments shown at the top, with the NPA motif shown. Loop B folds into the cytoplasmic face of the bilayer and loop E into the extracellular face giving the intermediate model (middle). Horizontal arrows indicate the folding together of the amino- and carboxy-terminal halves of the molecule resulting in the hourglass structure with loops B and E potentially forming part of the single aqueous pathway (bottom; taken from reference 81, and used with the author's permission).

In 1994, the freeze-fracture study by Verbavatz *et al.* [83] was sufficient to yield a model for the tetrameric assembly of membrane embedded AQP1. The structural elements responsible for the control of water-selective permeability of AQP1 remained unclear in this low-resolution structure.

Biologically active 2-D crystals of AQP1 purified from red cell membranes were produced by Murata *et al.* [66], by reconstituting octyl-ß-D-glucopyranoside (OG)-solubilized AQP1 into proteoliposomes at a high protein-to-lipid ratio. The AQP1 2-D crystal arrays were shown to have a monomer osmotic water permeability of approximately 3 $x10^9$ water molecules per AQP1 subunit per sec, indistinguishable from native water channels in erythrocyte membranes, that was reduced to less than 10% of the original level upon incubation with submillimolar levels of $HgCl_2$.

The first 3-D map was produced by Li and Jap at 6Å resolution [86], closely followed by the models of Walz *et al.* at 6Å [87] and Cheng *et al.* at 7Å [88]. Subsequent refinements and spectral improvements have provided an AQP1 electron diffraction crystal structure with a resolution of 3.8Å (in the membrane plane, 4.6Å normal to the membrane plane) [66]. The 3.8Å electron density map resolves the amino-acid side-chain protrusions thereby making it possible to fit the aquaporin sequence into the map.

The higher resolution structure showed that AQP1 monomers contain six tilted, membrane-spanning α-helices in a right-handed bundle. The helices are arranged into two repeats that exhibit a pseudo two-fold axis parallel to the membrane plane at the center of the molecule. The two repeats are composed of helices 2-1-3 and 5-4-6 (see Figure 1-14).

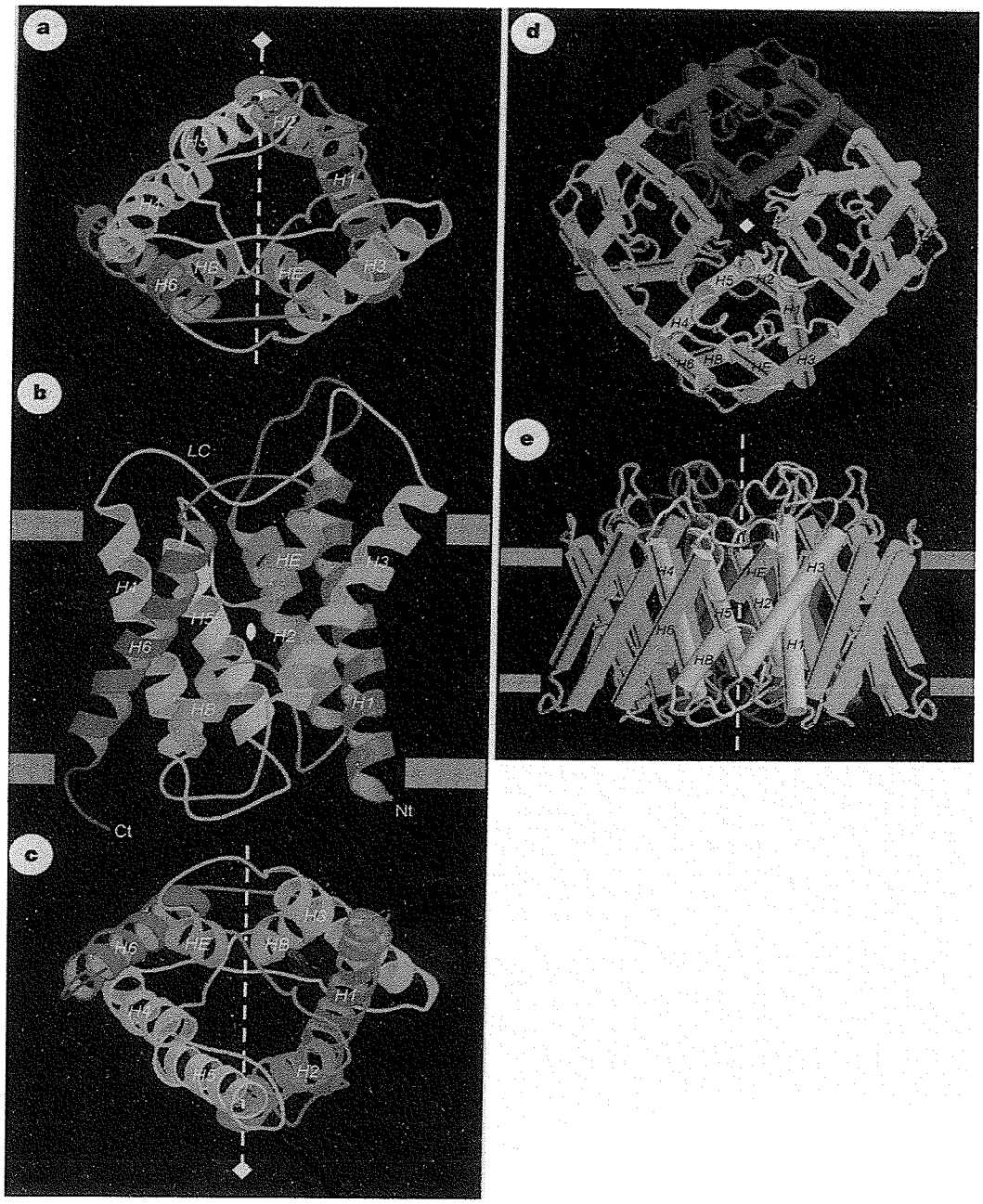


Figure 1-14. Ribbon diagrams of AQP1 showing the six transmembrane helices (H1-H6), and the connecting loops in different colors (loop helices B and E are labelled HB and HE). **A**, End view from the extracellular surface; **B**, side view; **C**, end view from the cytoplasmic surface. **D**, Cylinder model of the AQP1 tetramer; end view from the extracellular surface; **E**, side view. The yellow diamonds and yellow dotted line indicate the four-fold axis of the AQP1 tetramer. The dotted lines and spindle in white show the pseudo two-fold symmetry axis of the AQP1 monomer. The grey bands indicate the surface of the lipid bilayer (taken from reference 66 and used with the authors' permission).

Figure 1-14b illustrates the arrangement of the 6 helices as they cross the membrane from the cytoplasmic to the extracellular side of the membrane (helices 1, 3, 5) and *vice-versa* (helices 2, 4, 6). The helical crossing angles between helices 1 and 3 (36.7°), helices 4 and 6 (40.9°), and helices 2 and 5 (28.5°) stabilize the right-handed coiled-coil interactions [89]. Due to imperfections in the electron-diffraction technique, loops connecting individual helices provided only weak signals in the density map. Loops B and E, which are highly resolved, fold back into the membrane and position the NPA motif found in both loops midway through the membrane, with one NPA motif crossing the other. Following the NPA regions loops B and E form short pore-helices (helix B and helix E) before connecting to the next helix (helix 3 and helix 6 respectively).

The two repeats (helices 2-1-3 and 5-4-6) are locked together by the crossing of helices 2 and 5, by interactions involving highly conserved residues Gly 57 and Gly 173, and by the interactions between pore-helices B and E with helices 6 and 3, respectively.

In the 2-D crystals [66], AQP1 forms a homotetramer in which each monomer interacts with two neighboring monomers through contacts between the membrane-spanning α-helices (see Figure 1-14 D and E). Helix 1 forms a left-handed coiled-coil with helix 4 of the adjacent monomer with a crossing angle of 48.5° and it also interacts with helix 5 at the extracellular surface. Helix 2 forms a left-handed coiled-coil with helix 5 of the adjacent monomer with a crossing angle of 48.0° and interacts with helix 4 at the cytoplasmic surface. All interactions are stabilized by a network of hydrogen bonds, and by interactions between loops A and D which surround the four-fold axis of the tetramer on the extracellular and cytoplasmic surfaces, respectively.

1.3.4 The Structural Basis of Water Selectivity

AQP1 can facilitate the transport of 3 x 10^9 water molecules per second per monomer [66]. Loops B and E interact via their NPA motifs and form part of the aqueous pore, and helices 2 and 5 with the C-terminal halves of helices 1 and 4 form the remaining part (see Figure 1-15). The pore is constricted to 3 Å at the centre of the membrane just beside the region where loops B and E interact. The diameter of a water molecule is 2.8 Å.

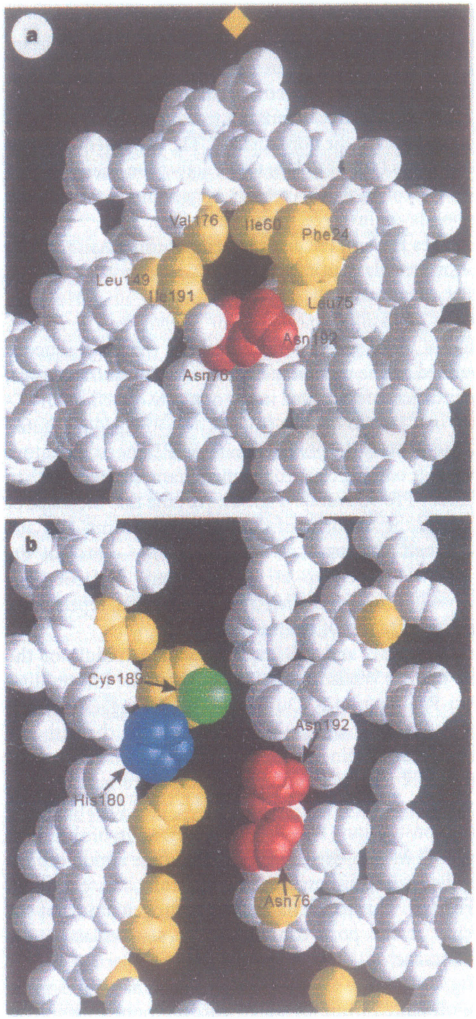


Figure 1-15. A, Section of a space filled model cut parallel to the plane of the membrane through the centre of the AQP1 monomer showing the pore diameter. **B**, Section cut perpendicular to the plane in **A**, showing the side view of the interior of the AQP1 pore. The yellow diamond in **A** indicates the four-fold symmetry axis of the AQP1 tetramer. Hydrophobic side-chains facing the inside of the pore are yellow, Asn 76 and Asn 192 are red, cysteine is green and His 180 is blue (taken from reference 66 and used with permission).

The loop B and E pore helices (HB and HE) are oriented so that the positive ends of their helix dipoles [66] point towards the centre of the membrane and generate a positive electrostatic field there. Thus, water molecules nearing the membrane centre orient so that their electronegative oxygen atoms face the NPA side of the pore (see Figure 1-16a-c).

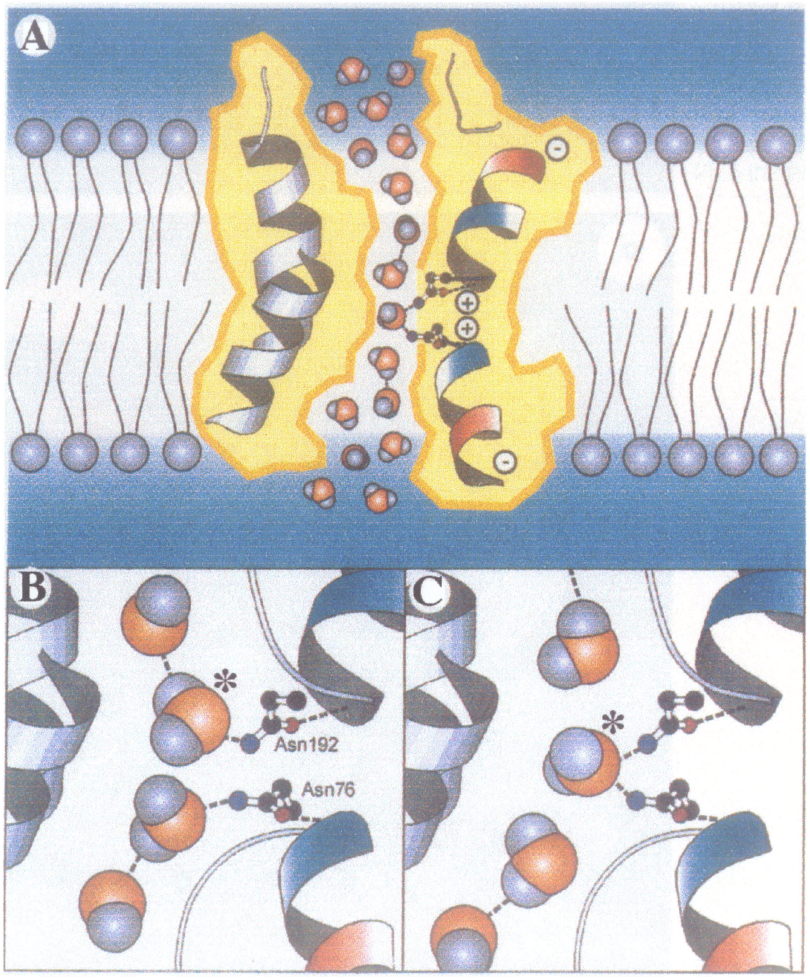


Figure 1-16. Schematic representations explaining the mechanism for water selectivity of AQP1. **A**, A diagram illustrating how partial charges from the helix dipoles, represented by the + and - signs, determine the orientation of the water molecules passing through the constriction of the pore. **B** and **C**, Diagrams illustrating the sequential passage of a water molecule (labeled with *). Water molecules hydrogen bond with Asn 76 and/or Asn 192, which extend their amido groups into the constriction of the pore (taken and adapted from reference 66 and used with permission).

The extension of the NPA motif asparagine side-chains into the pore at the site of the 3 Å constriction places the amido groups in a position to form hydrogen bonds with a water oxygen atom. The pore diameter limits the size of the particles transported to those less than 3 Å, and limits transport through the centre of the pore to one molecule at a time. Thus, only one water molecule at a time crosses the 3 Å constriction and is removed from the bulk hydrogen bond network. Upon formation of these hydrogen bonds the water molecule's two O-H bonds are perpendicular to the channel axis (see Figure 1-16c), these hydrogen atoms at then prevented from hydrogen bonding to adjacent water molecules in the single-file column, and because only one hydrogen bond is affected, there is an energy cost of no more than 3 kcal/mol [66]. The water molecules do not get stuck in the channel because the water hydrogens are presented to the channel face opposing the NPA motif that is lined with hydrophobic residues, i.e. there are no hydrogen-bond acceptors opposite the NPA motif asparagine residues. The structure thus explains how water molecules can permeate the pore with a minimal energy barrier, whereas hydrogen ions are blocked from transfer because of the required hydrogen-bond isolation from bulk water and repulsion by the positive ends of the dipoles in the centre of the membrane.

Thus, cryoelectron microscopy has provided a structure of AQP1 with sufficient resolution to allow identification of the key elements in the water selectivity and transport mechanisms within AQP1. This technique holds great promise for structural analysis of membrane and other proteins that are not amenable to three-dimensional crystal growth and standard X-ray diffraction analysis.

1.4 The *Escherichia coli* Glycerol Facilitator Protein

1.4.1 Introduction

Escherichia coli is a rod-shaped gram-negative bacterium that colonizes the intestinal tract of humans. *E. coli* has simple nutritional requirements and grows well if provided with simple energy sources such as glucose or glycerol, a nitrogen source and a few mineral salts. *E. coli* cytoplasm is packed with ribosomes, DNA, mRNA, tRNA, proteins and, as a prokaryotic cell, is devoid of subcellular organelles [90]. The cell envelope has three layers: the plasma (inner) membrane, the peptidoglycan layer, and the outer membrane (Figure 1-17). The inner membrane consists of phospholipids organized into a bilayer (see section

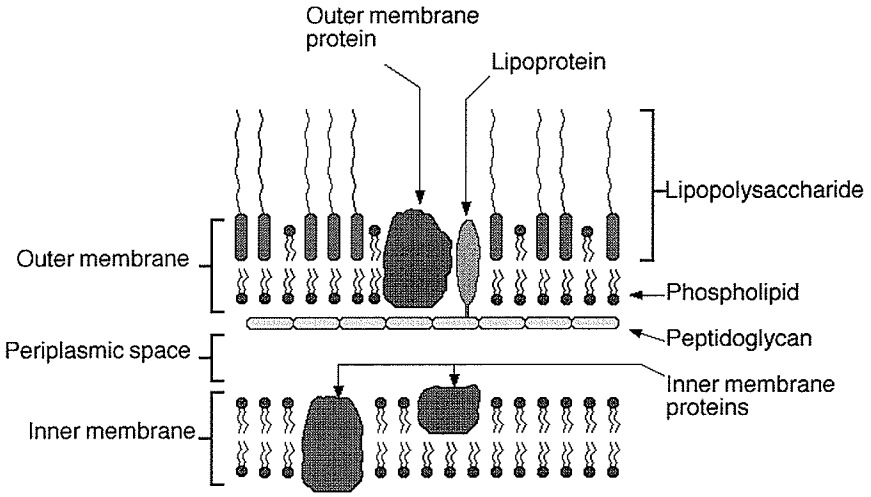


Figure 1-17. The gram-negative cell envelope.

1.1). The peptidoglycan layer is separated from the inner membrane by the periplasmic space and is comprised of glycan chains, made up of disaccharide units (*N*-acetylglucosamine-*N*-acetylmuraminic acid), linked together by peptide bonds. The outer membrane contains phospholipids, proteins and large complex molecules containing both lipid and

carbohydrates called lipopolysaccharides. The peptidoglycan and outer membrane layers together are generally called the cell wall [90].

1.4.2 The Glycerol Catabolic System

In 1898 it was found that crystals of dihydroxyacetone (DHA) could be prepared from the spent culture media of *Acetobacter xylinum* supplied with glycerol as a carbon source [91]. Studies have since shown [67] that glycerol is dissimilated in bacteria by two modes. The first mode begins with dehydrogenation followed by phosphorylation and the second starts with phosphorylation followed by dehydrogenation. The two enzymes responsible for these reactions are glycerol dehydrogenase and the ATP-dependent glycerol kinase (GK). In both cases dihydroxyacetone phosphate (DHAP) is the terminal product which then enters the glycolytic pathway. The reaction intermediates DHA and *sn*-glycerol-3-phosphate (G3P) may also be utilized by these organisms. Figure 1-18 illustrates the bacterial glycerol catabolic system.

That glycerol rapidly crosses the *E. coli* plasma membrane was inferred from the inability to sustain osmotic pressure across the membrane in a hypertonic glycerol environment [93]. Analysis showed that there is no concentrative mechanism for uptake as indicated by the inability of glycerol kinase-negative mutants to accumulate glycerol. Further evidence against a concentrative mechanism for glycerol was provided by comparing the growth K_m (minimal glycerol concentration where growth is half-maximal) with the K_m of glycerol kinase (in this case the glycerol concentration where enzyme activity is half-maximal). The two K_m values are approximately 1 μM indicating that the utilization of glycerol in *E. coli* is dependent on GK [94]. The data thus indicated that glycerol is not transported

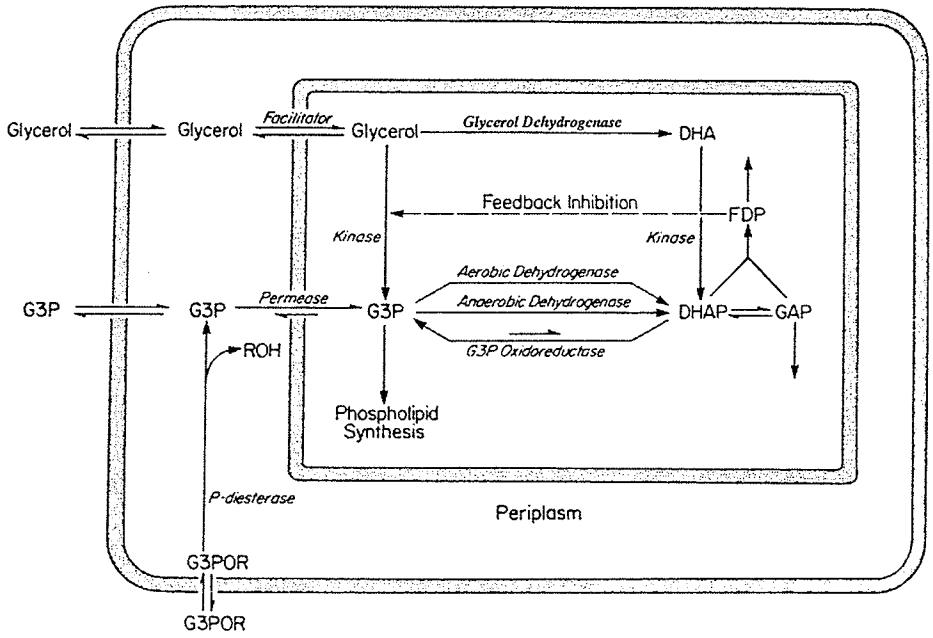


Figure 1-18. Network for the utilization of glycerol, G3P, and glycerophosphodiesters (G3POR) by *E. coli*. Porins in the outer membrane allow free passage of small substrates. Abbreviations not defined in the text: GAP, D-glyceraldehyde 3-phosphate; and FDP, fructose-1,6-bisphosphate (taken from reference 95 and used with permission).

against its concentration gradient and that it is phosphorylated by GK upon entry into the cell.

Evidence for the existence of a specific protein facilitating the equilibration of glycerol across the inner membrane was discovered upon analysis of the changes in the light-scattering properties of *E. coli* in response to osmotic stress caused by the introduction of hypertonic and hypotonic solutions of glycerol. *E. coli* cells that were catabolite-repressed by growth in glucose-containing media had half-times of equilibration of 21 seconds, whereas cells grown in glycerol-containing media or cells made constitutive in the glycerol system by mutagenesis of the glycerol dissimilation system [96, 97], had half-times of equilibration (i.e. mid-point of equilibration of a concentration gradient) of 10 and 2 sec-

onds, respectively. These data suggest that glycerol can enter the cell by nonspecific simple diffusion at a rate much less than that of water and by a specific facilitated transport mechanism at a rate that approaches that of water transport. Thus, when supplied glycerol as a sole carbon source *E. coli* produces the glycerol facilitator (GF) protein providing an energy-independent mechanism for the transport of glycerol across the inner membrane, where upon it is immediately phosphorylated by GK [95]. The glycerol facilitator was initially cloned in 1990 by Sweet *et al.* [98].

Measurement of the growth rate of wild-type *E. coli* cells in relationship to glycerol concentration shows that the half-maximal growth rate occurs at glycerol concentrations of approximately 1 μM (see page 38) [97]. *Shigella flexneri* (strains M4243 and 24570) lack a facilitator but possess a glycerol kinase with a substrate K_m close to that of the *E. coli* enzyme (approximately 1 μM), however the growth K_m is 5-10 mM (when glycerol is the sole carbon source). When *S. flexneri* was transduced with phage P1 raised on wild-type *E. coli*, maximum growth occurred at glycerol concentrations below 1 mM [97]. These data show that, regardless of the high intrinsic permeability of the cell membrane to glycerol in the absence of a facilitator, the rate of diffusion is insufficient for maximum use of this carbon source.

1.4.3 The *E. coli glp* Regulon

The glp regulon consists of five operons in three locations on the *E. coli* linkage map. These are the *glpTQ* operon encoding the G3P transport protein and the periplasmic glycerophosphodiesterase; the *glpBA* operon encoding the anaerobic G3P dehydrogenase and its membrane anchor protein; the *glpR* operon encoding the repressor protein; the *glpD* operon encoding aerobic G3P dehydrogenase; and the *glpFKX* operon encoding GF, GK

and a protein of unknown function (see Figure 1-19) [95, 98]. The genes of the *glp* regulon

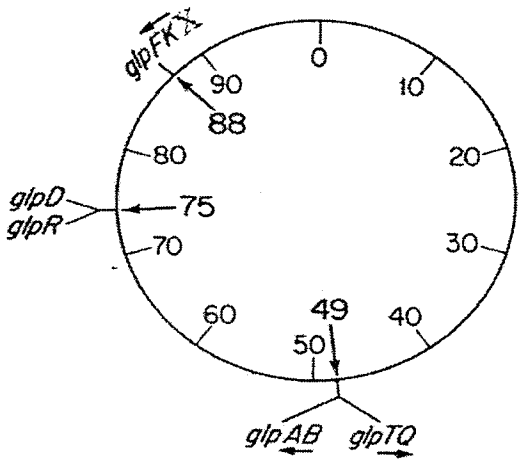


Figure 1-19. Genetic map of the *glp* regulon of *E. coli* (the arrows indicate the direction of expression).

are under negative control of the *glp* repressor, a tetrameric protein encoded by the *glpR* gene, mediated by binding of the *glp* repressor to its operator sites within the *glp* regulon. Affinity of the repressor for its operators is decreased in the presence of G3P, the inducer for the regulon (see Figure 1-20). G3P levels reflect the activity of GK, which is inhibited

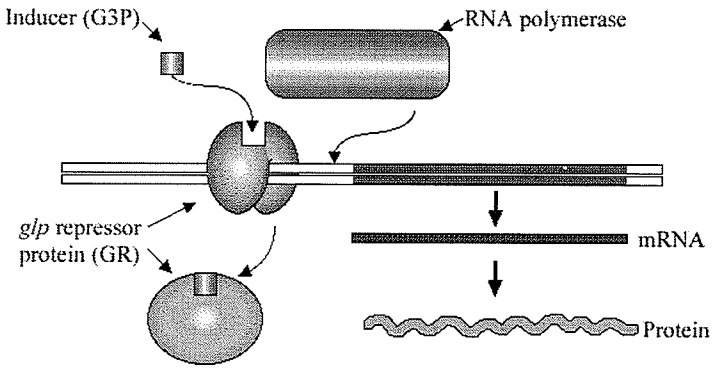


Figure 1-20. Negative control of the glp regulon

by the feedback inhibitor fructose-1,6-bisphosphate (FDP) when glucose levels are suffi-

cient for growth [99]. Thus, when glucose levels are low FDP levels decline and GK activity increases, resulting in the increase in G3P and induction of the *glp* regulon.

There has been considerable speculation and some kinetic evidence that GF exists in close association with GK at the inner membrane and that the kinase may be regulated by interaction with the facilitator [100]. This association has been compared to the interactions of hexokinase and glycerol kinase with the respective mitochondrial porins [101]. GK shows complex regulatory behaviour and is functional in both a dimeric and tetrameric form [102]. Fructose 1,6-bisphosphate and the phosphoenolpyruvate phosphotransferase protein IIA^{Glc} are both allosteric inhibitors of GK that bind at different sites [103, 104]. To inhibit the GK tetramer and lock it in an inactive tetrameric conformation either two molecules of fructose 1,6-bisphosphate bind to the tetramer [103] or a phosphocarrier protein IIA^{Glc} binds to each monomer [105].

1.4.4 Transport Properties

The transport properties of *E. coli* GF have been measured in the bacterium, in *Xenopus* oocytes, and after reconstitution into liposomes [95, 106, 107, 108,109, 110], by measuring the changes in light-scattering properties in relationship to osmotically induced plasmolysis (shrinkage) and deplasmolysis (swelling). GF has been shown to transport D- and L-glyceraldehyde, glycine, urea, ribitol, xylitol, D-arabitol, and D-sorbitol. Not transported are sn-glycerol-1-phosphate, G3P, glycylglycine, inositol, erythrose, D-arabinose, L-arabinose, D-ribose, D-xylose, D-galactose, and D-sorbose [95]. *E. coli* GF is therefore

selective for short chain alditols, with transport efficiency inversely related to alditol chain length (see Figure 1-21).

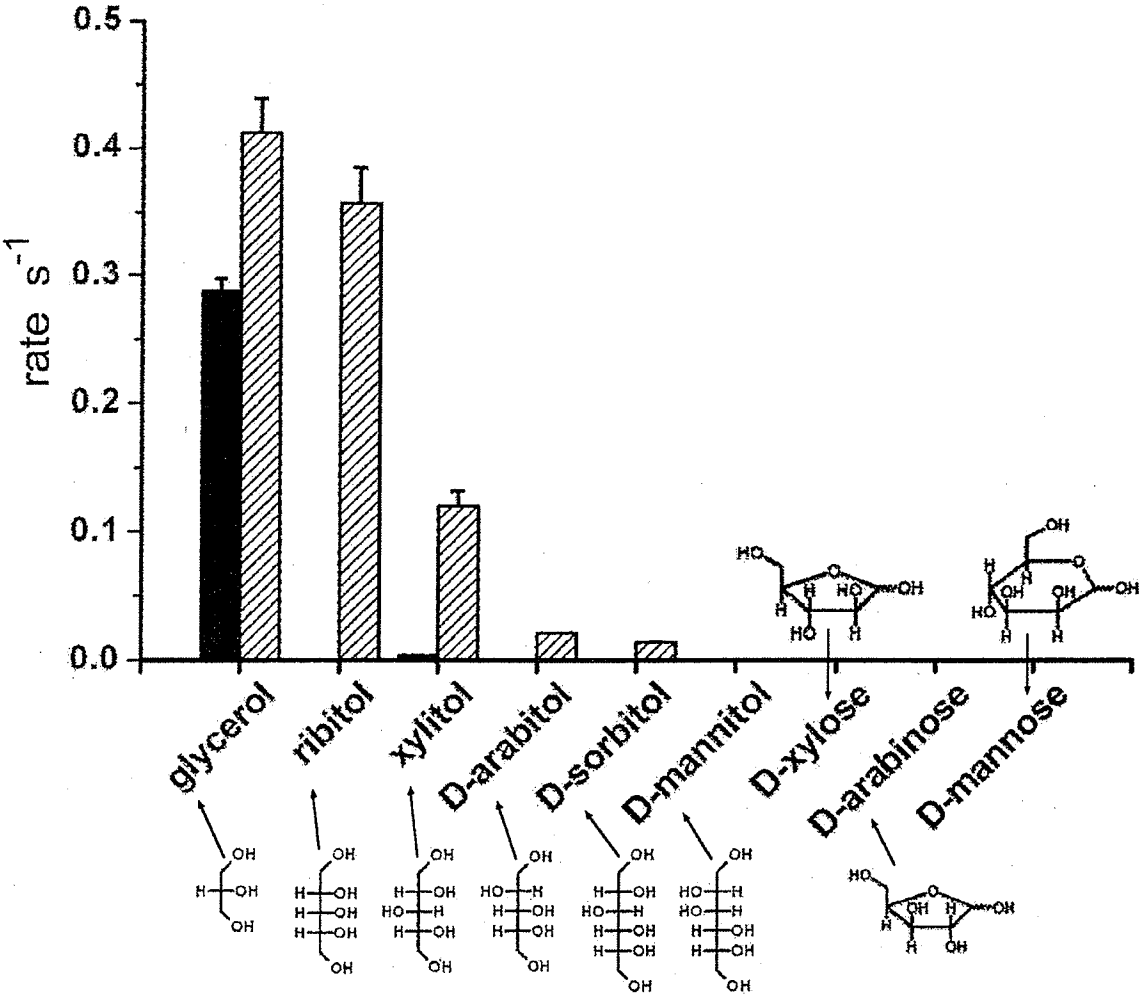


Figure 1-21. Relative rates for transport of a selection of carbohydrates into protein-free liposomes (black bars) and into GF-containing proteoliposomes (hatched bars). Structures are indicated in the Fisher and Haworth diagrams. Error bars indicate the standard deviation from 10 accumulations (data from reference 110).

1.4.5 The GF X-ray Crystal Structure

Three weeks following the publication of a large portion of the material presented in this thesis [111], the X-ray crystal structure of GF was solved at a resolution of 2.2 Å [110]. Structural refinement to this resolution has permitted individual side-chain densities of the protein to be identified and the amino-acid sequence of GF inserted into the electron density map, providing an unambiguous picture of the GF structure (see Figure 1-22). GF crystallizes as a symmetric arrangement of four channels with the plane of the bilayer perpendicular to the four-fold axis (Figure 1-22a). The structure is nearly identical to the electron diffraction structure of AQP-1 [66]. GF is composed of six transmembrane (M1, M2, M4, M5, M6, and M8)[1] and two half-membrane-spanning α-helices (M3 and M7) that form a right-handed helical bundle around each channel. Based on the crystal structure the α-helix content is calculated to be 67% [112]. The helices project outward to form a channel entrance on each surface. Interhelix 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 and M5 and M8 of its neighbor.

The internal sequence homology between the N- and C-terminal halves of GF is preserved in the 3-dimensional structure as the two segments are related by a twofold-symmetry axis that passes through the center of the bilayer (see Figure 1-22b, yellow and blue segments). Similarly to AQP1 (using the same helix and loop labels), the N-terminal segment begins on the cytoplasmic side of the membrane with helix M1 which crosses the membrane and connects to helix M2 via loop A. Helix M2 crosses the membrane and

1. Fu *et al.* [110] have chosen to label the transmembrane and loop helices sequentially. Therefore, homologous helices between GF and AQP1 crystal structures are M1-H1, M2-H2, M3-HB, M4-H3, M5-H4, M6-H5, M7-HE, and M8-H6.

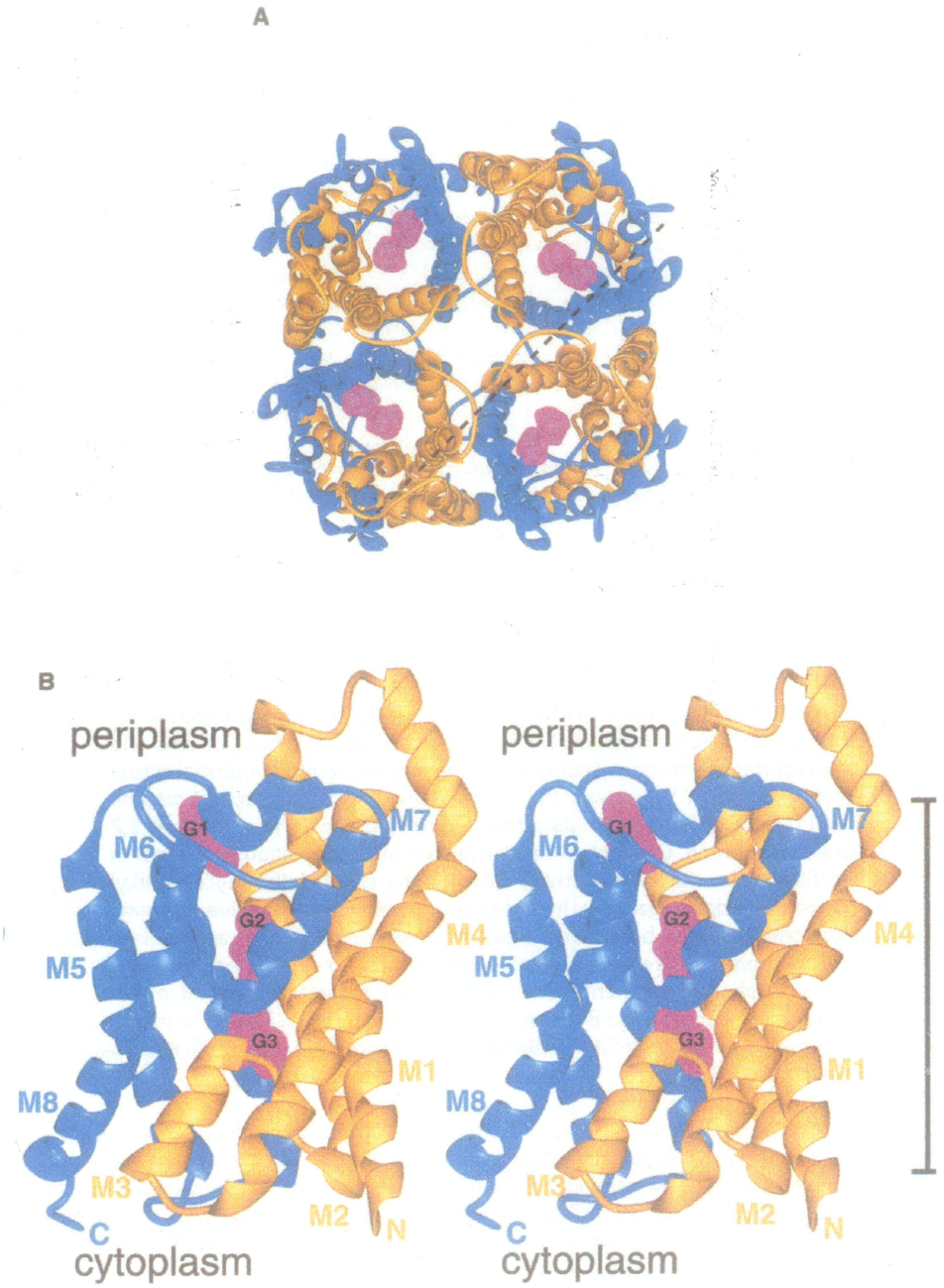


Figure 1-22. The 2.2 Å resolution X-ray crystal structure of *E. coli* GF. The twofold symmetry is displayed by yellow residues (6-108) and blue residues (144-254). **A**, A ribbon diagram of the GF tetramer viewed from the periplasmic side. **B,** A stereo view of the monomeric glycerol channel viewed down the twofold axis. The vertical bar (35 Å) represents the putative position of the membrane, and purple indicates 3 glycerol molecules (taken from reference 110 and with permission).

connects to loop B. Loop B projects into the center of the channel from the cytoplasmic side and forms the half-spanning helix M3, that begins with the conserved NPA motif. Loop B returns to the cytoplasmic side of the membrane and connects to helix M4 which crosses to the periplasmic side. The C-terminal segment is similarly organized starting with helix M5 crossing from the periplasmic to the cytoplasmic side of the membrane. Loop E extends into the center of the channel from the periplasmic side and forms the second half-spanning helix M7, which terminates with the second conserved NPA motif (see Figure 1-22b, and Figure 1-23)

The two half helices formed by helices M3 and M7 meet at their N-terminal ends in the center of the membrane and, similarly to AQP1, largely form the selectivity filter of GF. The NPA motifs are in an orientation where the proline rings of one monomer are in Van der Waals contact nestled between the proline and alanine side chains of the symmetry related helix (see Figure 1-24). The channel has a 15 Å wide vestibule on the periplasmic surface which is constricted to approximately 3.8 Å by 3.4 Å at a distance of 8 Å above the twofold axis. This constriction forms the beginning of a 28 Å long, glycerol selective channel, with a radius of approximately 3.5 Å, that crosses to the cytoplasmic surface (see Figure 1-25).

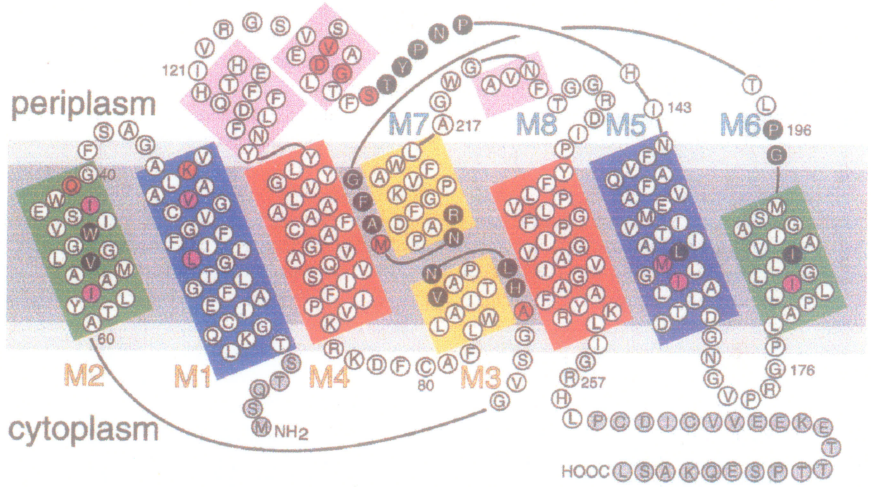


Figure 1-23. The amino acid sequence of GF is arranged topologically as in the structure, with helices viewed as if from inside the channel. The NPA motifs are in the center of the figure. Three-and-one-half helices compose each segment, labeled M1 to M4 and M5 to M8. Homologous helices are boxed in similar colors. Residues in black circles interact with glycerol. Residues in red circles contribute carbonyl oxygens or amide NHs to the channel. Grey circles represent residues that are not seen in the structure. The location of the cell membrane is indicated in grey (taken from reference 110 and used with permission).

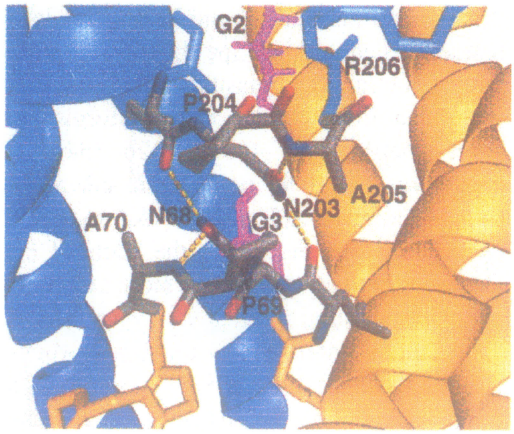


Figure 1-24. The two NPA motifs at their interface where the two half-helices M3 (yellow) and M7 (blue) meet to form a junction. The proline rings from each NPA lie between the proline and alanine side chains of the opposite helix. The asparagine amide groups of one helix form two symmetry-related hydrogen bonds, to the alanine nitrogen of the same half-helix and a carbonyl oxygen of the opposite half-helix. Atoms are colored according to atom type (oxygen, red; carbon, grey; nitrogen, blue; and sulfur, yellow; taken from reference 110 and used with permission).

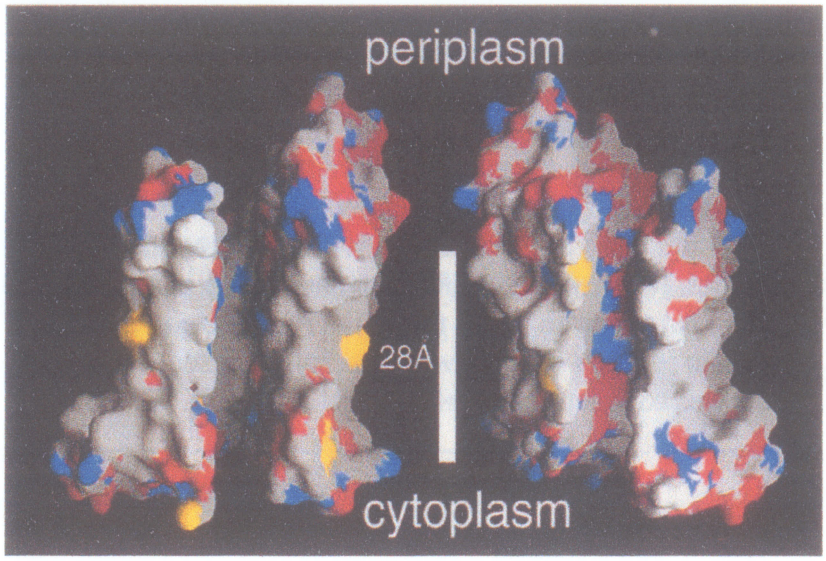


Figure 1-25. The glycerol channel in GF. Atoms are colored according to atom type (oxygen, red; carbon, gray; nitrogen, blue; and sulfur, yellow) showing the hydrophobic surface (left) and the polar surface (right). The length of the narrow (~3.5 Å) part of the channel is indicated by the while line (taken from reference 110 and used with permission).

1.4.6 Structural Basis of Glycerol Transport

The crystal structure shows three glycerol molecules (G1, G2, and G3) within the channel [110]. G1 forms a hydrogen bond with Tyr138 at the periplasmic entry to the vestibule. G2 and G3 are organized into a site, that Fu *et al.* [110] have termed the "selectivity filter", which is large enough in cross-section to accommodate a single CH-OH group (Figure 1-26). The alkyl backbone of G2 is packed tightly against the aromatic rings of Trp48 and Phe200, providing no space for substitutions at the C-H hydrogen positions. As a result of this constriction, CH-OH groups of alditol molecules can only pass through the filter in single file. The presence of a water molecule between G2 and G3 suggests that water may be co-transported. However, as described below, as glycerol is transported across the membrane it becomes progressively dehydrated by the exchange of one set of stringent hydrogen bonds for another. As Figure 1-26 illustrates, the selectivity filter of GF is designed

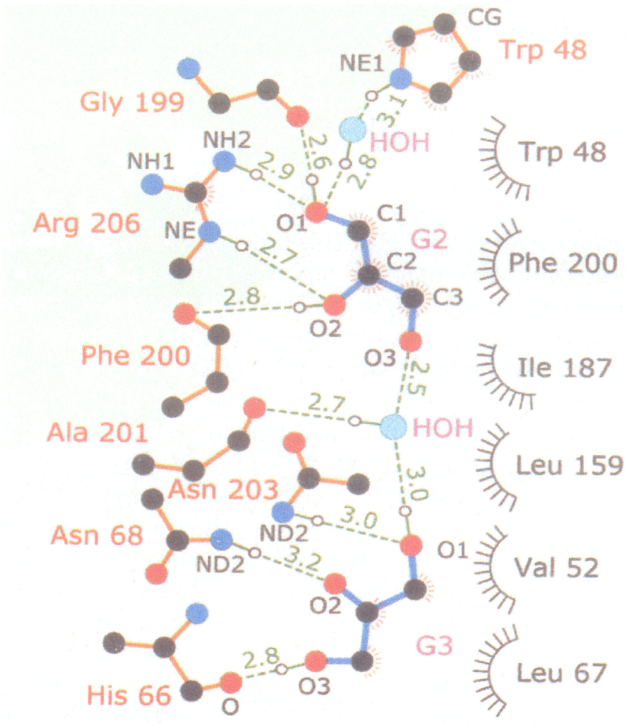


Figure 1-26. The hydrogen bonding network (dotted lines) showing interactions of G2, G3, and associated water molecules with the selectivity filter. Hydrophobic interactions are indicated by radial lines around the residues. O1 and O2 of G2 are hydrogen bond acceptors from successive NHs of the guanidinium group of Arg 206 and hydrogen bond donors to the carbonyl oxygens of Gly199 and Phe200 (respectively). The NH_2s of Asn203 and Asn68 form hydrogen bonds to the O1 and O2 of G3, (respectively) and the carbonyl of His66 is an acceptor for O(3)H of G3. Distances (in angstroms) between heavy atoms are indicated (taken from reference 110 and used with permission).

such that one side of the channel is lined with hydrophobic amino acids (indicated by radial lines in Figure 1-26) and the other with polar residues. As glycerol molecules enter the pore they reorient so that their alkyl backbones face the hydrophobic side of the channel and their polar "hydroxyl sides" face the channel's polar side. Figure 1-26 shows that successive glycerol molecules, G2 and G3, adopt the same orientation as they are transported. Furthermore, the carbonyls of Gly199, Phe200, and Ala201 are oriented on the periplasmic side of the filter by hydrogen bonds to main-chain amides of Phe200, Ala201 and the buried COOH from Glu152. According to Fu *et al.* [110], the "electrostatic triangle" that is formed by the negative charge of Glu152, the amides of Phe200 and Ala201, and the positive

charge of Arg206, polarizes two successive hydroxyl groups on permeating alditols. This is shown in Figure 1-26, where the "electrostatic triangle" polarizes the C1 and C2 hydroxyls on G2. This provides a further selection mechanism requiring that the permeant molecule be polarizable in sections parallel to the plane of the membrane.

The constriction and dual hydrophobic/polar nature of the selectivity filter of GF provides an explanation of its stereoselectivity towards permeant alditols (see Figure 1-21). For example, the hydroxyl groups of ribitol all have the same stereospecific relationship to the carbon backbone, whereas a chiral stereoisomer such as D-arabitol has a mixed arrangement of hydroxyls (differing from ribitol at the orientation of the hydroxyl group at C3). It has been demonstrated that ribitol is transported by GF 10 times faster than D-arabitol [110]. This difference can be explained with Figure 1-26, where we see that two successive -CHOH groups of G2 are oriented so that the carbon backbone is lined up along the channel axis and the C-H hydrogen(s) are in contact with the aromatic ring of Phe200 or Trp48. The successive -CHOH groups of ribitol would be similarly oriented within the channel, whereas D-arabitol would require that the molecule either reorient or suffer the transport consequences of placing the -CHOH group of C3 in an unfavorable environment.

1.5 Biophysical Techniques

1.5.1 Electromagnetic Radiation and Spectroscopy

Light is a form of electromagnetic radiation, and can be considered to be composed of two waves oriented perpendicular to one another (see Figure 1-27).[113]. Electromag-

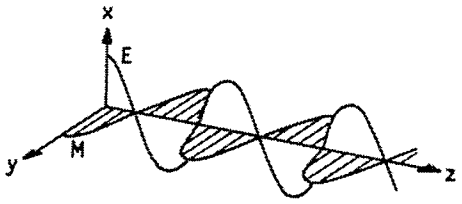


Figure 1-27. Electromagnetic radiation is made up of a magnetic wave (M) and a perpendicular electric wave (E) propagated in the z-direction

netic waves are generated by oscillating electric or magnetic dipoles and are propagated in a vacuum at the speed of light (c) with energy (E) given by the following equations:

$$E = h \cdot \nu \tag{1-8}$$

and

$$\nu = c/\lambda \tag{1-9}$$

so that

$$E = (h \cdot c)/\lambda \tag{1-10}$$

and thus

$$E \propto \frac{\nu}{\lambda} \tag{1-11}$$

where h is Planck's constant (6.63 x 10^{-34} J/s), υ is the frequency and λ the wavelength of the electromagnetic radiation [113].

Electromagnetic radiation can range from low energy radiowaves (wavelength approximately 10 cm) to high energy gamma rays (wavelength approximately 10^{-11}cm).

This variation in energies provides the basis of biological spectroscopic techniques (see Table 1-2) [114]. Generally, when light strikes matter it will either be scattered or absorbed and re-emitted [113]. Table 1-2 list the many spectroscopic techniques and applications that deal with the measurement of these interactions. For our purposes only scattering and absorption in the ultraviolet region will be considered further. In particular the scattering intensity in relation to particle size and density will be considered as will the extinction coefficient of elliptically polarized light at different wavelengths in relation to particle shape and symmetry.

Table 1-2. Biologically useful spectroscopic regions[a]

Typical wavelength (cm)	Approximate energy (kcal/mole)	Spectroscopic region	Spectroscopic techniques and applications
10^{-11}	3×10^{8}	γ-Ray	Mössbauer
10^{-8}	3×10^{5}	X-Ray	X-ray diffraction, scattering
10^{-5}	3×10^{2}	Vacuum UV	Electronic spectra
3×10^{-5}	10^{2}	Near UV	Electronic spectra
6×10^{-5}	5×10^{3}	Visible	Electronic spectra
10^{-3}	3×10^{0}	IR	Vibrational spectra
10^{-2}	3×10^{-1}	Far IR	Vibrational spectra
10^{-1}	3×10^{-2}	Microwave	Rotational spectra
10^{0}	3×10^{-3}	Microwave	Electron paramagnetic resonance
10	3×10^{-4}	Radio frequency	Nuclear magnetic resonance

a. Reference 114.

1.5.2 Light Scattering

1.5.2.1 Particles that are small compared with the incident wavelength

When electromagnetic radiation interacts with a particle that is small compared to the incident wavelength the oscillating electric field induces oscillations of the electrons

within the particle [115] (Figure 1-28). The frequency of the oscillation is governed by the

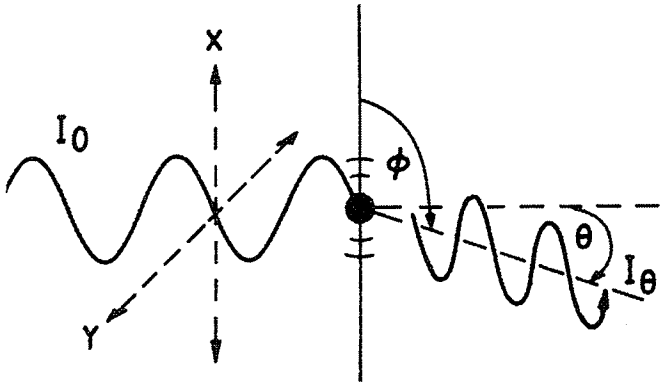


Figure 1-28. Incident radiation (I_o) polarized along the x axis causes the particle to oscillate and emit radiation with intensity I_θ at a distance r and angle ϕ to the direction of the induced dipole, and angle θ to the incident radiation (taken from reference 113 and used with permission)

energy of the system and will depend on how easy it is to displace an electron from the nucleus (molecular polarizability, α) and the frequency of the applied radiation [113, 115]. The dipole moment created by the oscillating electron transmits electromagnetic radiation like an antenna, and will disperse some of the energy (I_θ) in directions other than that of the incident radiation (I_o; see Figure 1-28) [114]. When unpolarized light of intensity I_o interacts with a sphere with a radius (r) very small compared to the wavelength of the light scattering can occur in all directions and the intensity (i) of the light scattered will be given by:

$$i = \frac{8\pi^4 r^6 n_o^4}{R^2 \lambda^4} \cdot \left(\frac{\left(\frac{n}{n_o}\right)^2 - 1}{\left(\frac{n}{n_o}\right)^2 + 2} \right)^2 \cdot (1 + \cos^2\theta) \cdot N I_o \psi \qquad \textbf{(1-12)}$$

λ is the wavelength of light *in vacuo*; n is the index of refraction of the particle; n_o is the index of refraction of the medium; θ is the angle of observation measured from the direction of propagation of the incident beam; R is the distance to the point of observation; N is the number of particles per unit volume; I_o is the intensity of the incident beam of unpolarized

light, and ψ is the volume of the sample illuminated (see Koch [116] for a detailed explanation). It can be shown that the absorbancy (A= $\log_{10}(I_o/I)$ where I is the transmitted light) of a suspension of particles that are much smaller than the wavelength of the incident light is

$$A = \frac{32\pi^3}{2.3 \times 3} \cdot \left[\frac{d_n/d_c}{n_o}\right]^2 \cdot \frac{q^2 \cdot N}{\lambda^4} \qquad \textbf{(1-13)}$$

(where d_n/d_c is the change in refractive index with change in solute concentration and is a constant for the solution [113]). That is, the absorbancy is proportional to the concentration (c) of the particles, to the square of their anhydrous mass (q), and inversely proportional to the fourth power of the wavelength of light incident on the sample. Therefore, in this regime, changes in particle diameter that do not involve a change in the anhydrous mass have no effect on absorbancy [116].

1.5.2.2 Particles that are large compared to the incident wavelength

Scattering from large particles differs from scattering from small particles in that the scattered radiation produced from different points on the same particle can interfere [113, 114, 115]. Figure 1-29 illustrates that scattering can occur from points that are a significant fraction of a wavelength apart, and since the scattering points are fixed with respect

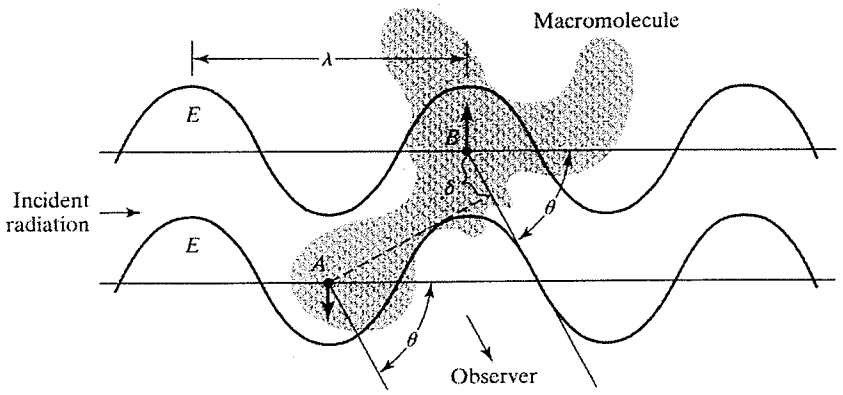


Figure 1-29. Scattering from a macromolecule that is large compared to the wavelength of incident radiation. Two points from which scattering occurs are shown at A and B. The phase of the radiation, and thus the induced dipoles, are different at the two points. (Taken from reference 115 and used with permission).

to each other, scattering interference must be accounted for. The calculation for this includes interference between light scattered from all scattering points on the molecule, with total scattering averaged over all orientations [114, 115, 116]. It is useful to define a function, P(θ), that describes the angular dependence of scattering from a large particle in a way that compares how the scattering at angle θ corresponds to that which the same molecule would have if its dimensions were shrunk to be much smaller than λ while its mass remained constant [113, 115]. P(θ) is defined by

$$P(\theta) = \frac{intensity\ of\ scattering\ by\ real\ particle\ at\ angle\ \theta}{scattering\ by\ hypothetical\ shrunken\ particle\ at\ angle\ \theta}. \quad \textbf{(1-14)}$$

It can be shown that

$$P(\theta) \cong exp\left(\frac{-Q^2 \cdot R^2_G}{3}\right) \quad \textbf{(1-15)}$$

where $Q=(4\pi/\lambda)sin(\theta/2)$ and R_G is the radius of gyration of the particle and is a measure of the dimensions of the particles. For example, for a sphere of radius r, $R_G=(3/5)^{1/2}r$; for a long rod of length L, $R_G=L/(12)^{1/2}$. Therefore, the measured angle of scattering, θ, can give a measure of R_G, thus providing information about the dimensions of the particles or the

particle shape [115]. The radius of gyration can also be defined as the root-mean-square average of the distance of scattering elements from the center of mass of the scattering particle [115].

The paper by A. Koch [116] provides an excellent theoretical approach to the understanding of scattering from large particles of different shapes, such as cells, mitochondria, and liposomes, and provides a much more detailed explanation than presented in this thesis. In summary, for spheres with a diameter (r) on the order of the wavelength of light Koch showed that

$$A = \frac{9\pi}{2.3 \times 3} \cdot \left[\frac{d_n / d_c}{n_o}\right]^2 \cdot \frac{q^2 \cdot N}{r^2 \lambda^2}. \qquad \textbf{(1-16)}$$

Thus, the turbidity or absorbance of a solution of scatterers varies inversely as the square of their radii and as their volumes$^{2/3}$. Koch developed this theory to explain the effect of bacterial cell growth and swelling (deplasmolysis) and cell shrinkage (plasmolysis) on turbidity or absorbance. When cells are induced to plasmolyse[1] the turbidity increases, while deplasmolysis results in a turbidity decrease [117]. This scattering theory is used here to measure the transport activity of the glycerol facilitator in the presence of xylitol (a glycerol analog that differs by the addition of two CHOH groups and is transported at a much slower rate). Another application is the common practice of following the growth of a bacterial culture by measuring the increase in turbidity as the number of cells increases. Here the turbidity increase simply reflects the increase in the number of scattering particles.

1. Increasing the osmotic pressure of the culture medium is an effective and rapid way to induce plasmolysis.

1.5.3 Circular Dichroism

As mentioned, electromagnetic radiation can be considered to be composed of electric (E) and magnetic (M) waves oriented perpendicular to one another (see Figure 1-27 and Section 1.5.1) [113, 118]. In unpolarized light the E-component oscillates in all directions perpendicular to the direction of propagation, i.e. if the direction of propagation is on the z-axis, then unpolarized light oscillates in all directions in the xy-plane [113]. When the oscillating E-component interacts with an electron it induces a displacement of the electron with respect to the nucleus with a magnitude dependent on the ease with which the electron can be displaced from the nucleus and the frequency of radiation. Specifically, the electron may undergo a transition when the frequency of the oscillating E-component matches the difference in energy between the initial and final electronic states [113, 114, 118, 119].

The electronic energy states can be described by molecular orbitals, such that electrons undergoing a transition are transferred from one molecular orbital to another. Molecular orbitals include the low energy σ and π bonding orbitals and high energy σ^* and π^* antibonding orbitals [121]. A further type of molecular orbital is the non-bonding (n) orbital and, along with the π and π^* orbitals, they are associated with unsaturated centers in molecules [121]. The absorption of energy from the applied ultra-violet and visible radiation occurs from the excitation of electrons in only a few atoms termed chromophores, and results in the transfer of the electrons from $\sigma \rightarrow \sigma^*$, $\pi \rightarrow \pi^*$, or $n \rightarrow \pi^*$ [113, 114, 118, 119, 121].

An absorption spectrum of a chromophore is acquired by the measurement of absorption intensity in relation to the change in the wavelength of the applied radiation over a ranges of wavelengths. As implied above, individual chromophores will display different

wavelengths where there is an absorption maximum, a wavelength in which the resonance energy of the electronic transition matches the energy of the applied radiation (equation 1-8). The absorbance (A) is defined as;

$$A = log\left(\frac{I_o}{I_t}\right), \tag{1-17}$$

where I_0 is the incident radiation and I_t is the transmitted radiation [113, 114, 118, 119]. Absorbance is also related to the experimental measure of intensity at a given wavelength, known as molar absorptivity (ε), by the Beer-Lambert law

$$A = \varepsilon c l \tag{1-18}$$

where c is the sample concentration and l is the path length of the sample container [113]. The molar absorptivity reflects the transition probability and is related to the magnitude of the transition dipole moment.

In plane polarized light the E-component oscillates in one plane perpendicular to the direction of propagation; for example if the direction of propagation were along the z-axis then the plane polarized light might oscillate parallel to the x or y-axis [118]. Further, if another wave was oscillating parallel to the y-axis, and was in phase with the x-axis component, then superimposition of the waves would yield a new wave oriented 45° to the x-axis [118]. However, if the two waves do not have the same phase, then the superimposition does not produce a fixed direction for the new wave. If the phase difference between the two waves is 90° (or π/2) then superimposition will produce a wave with a helical path (See

Figure 1-30) [118, 121]. If the amplitudes of the two waves are equal then the superimposed

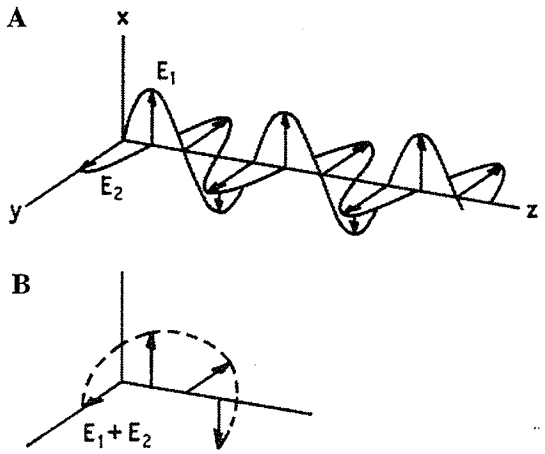


Figure 1-30. A, Waves E_1 and E_2 are polarized along the xz and yz axes respectively, and are 90° out of phase to each other. **B,** The superimposition of E_1 and E_2 produces a left circularly polarized wave (Taken from reference 113 and used with permission).

wave is called circularly polarized. Unequal amplitudes will produce a wave that is elliptically polarized [121, 122]. Further, depending on the phase difference, either left or right circularly polarized light (LCPL; RCPL) can be produced [113, 118, 121, 122]. Consider Figure 1-30 where E_1 is out of phase from E_2 by 90° and the sum $E_1 + E_2$ produces left circularly polarized light (the rule is to point your thumb in the direction of propagation and the hand that has the fingers curled in the direction of the helical wind tells the "handedness" of the wave). If E1 were out of phase by -90° then the sum $E_1 + E_2$ would produce right circularly polarized light (see Figure 1-31) [122].

Optically active chromophores are asymmetric, that is, they have no plane or center of symmetry [113]. These asymmetric chromophores interact with L and RCPL differently giving rise to the phenomena of optical rotatory dispersion (ORD) and circular dichroism (CD) [121, 122]. Analogous to absorption spectroscopy, CD deals with the differential

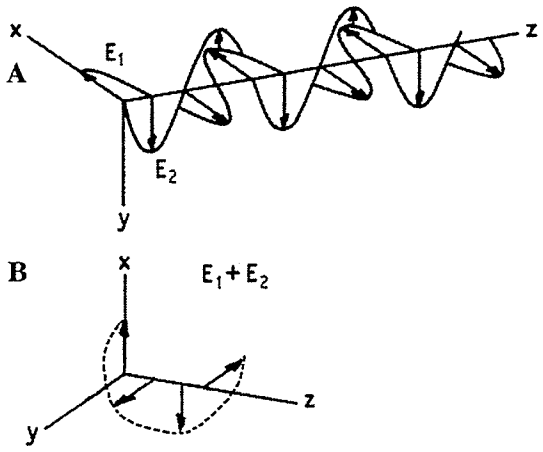


Figure 1-31. The superimposition of two waves (**A**) where one (E_1) is out of phase from the other (E_2) by $-\pi/2$, produces right circularly polarized light, **B** (adapted from reference 113).

absorption of LCPL and RCPL [122]. Since absorption spectroscopy and CD deal with the same electronic transitions then they occur at the same energies and can be dealt with similarly. Indeed, the Beer-Lambert law is obeyed in CD such that

$$\Delta A = A_L - A_R = \varepsilon_L cl - \varepsilon_R cl = \Delta\varepsilon cl \qquad \textbf{(1-19)}$$

where the subscripts indicate LCPL and RCPL. However, unlike absorption bands, CD bands can be positive and negative depending on the sign of $\Delta\varepsilon$, which depends on the wavelength of I_o and the characteristics of the molecule under investigation (see Figure 1-32) [113, 122]. The difference in absorption between LCPL and RCPL is usually very

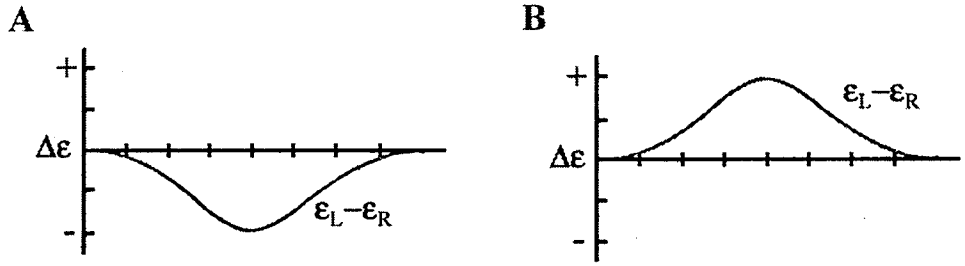


Figure 1-32. CD spectra where **A**, RCPL is absorbed less than LCPL therefore $\Delta\varepsilon$ is negative and **B**, LCPL is absorbed less than RCPL therefore $\Delta\varepsilon$ is positive (adapted from reference 113).

small, about 1 part in 10^3-10^5, and thus requires sensitive instrumentation and pure samples for detection [122].

In CD spectropolarimetry substances are irradiated with plane polarized light, which is equivalent to the superimposition of LCPL and RCPL. The differential absorption of LCPL and RCPL changes the vector sum of the two waves from linearly polarized to elliptically polarized (see Figure 1-33) [113, 118, 122]. Although CD is a differential absorption measurement, in the early days the ellipticity of the transmitted beam was measured [122]. The ellipticity is defined as the tangent of the angle (θ_{obs}) which is the ratio of the minor (b) to the major (a) axes of the ellipse (see Figure 1-33) and it can be shown that

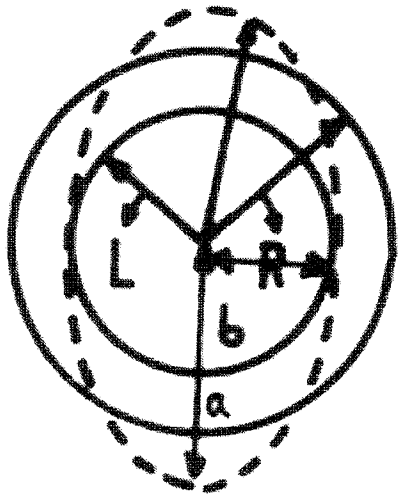


Figure 1-33. Differential absorption of LCPL (L) and RCPL (R) gives an elliptically polarized beam, as indicated by the dashed ellipse, where a and b are the radii of the circles turned by RCPL and LCPL respectively, after passage through the sample (Taken from reference 113 and used with permission).

the ellipticity is directly proportional to CD accordingly:

$$[\theta_{obs}] = 3298\Delta\varepsilon \qquad \textbf{(1-20)}$$

Molar ellipticity ($[\theta]_\lambda$) is most commonly reported and is defined as

$$[\theta]_\lambda = \frac{3298\Delta A}{cl} = \left(\frac{\theta_{obs} M_w}{100Lc}\right) \quad \textbf{(1-21)}$$

where θ_{obs} is the observed ellipticity in degrees, M_w is the molecular weight, and L is the path length in decimeters [113, 121, 122].

Whilst CD deals with the difference in absorption of LCPL and RCPL, ORD is a measurement of changes in the velocity with which LCPL and RCPL traverse the sample [122]. The change in velocity of the incident beams is a result of the differences in the refractive indices (n) for the two circular beams where $n_L \neq n_R$. Thus, the resultant beam remains plane polarized, but is rotated by an angle α (see Figure 1-33)

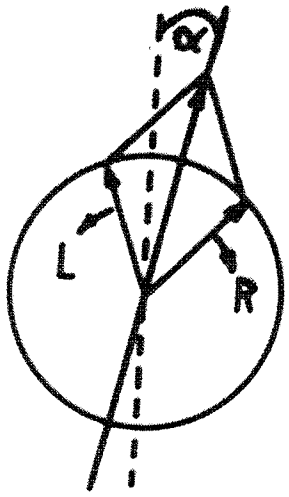


Figure 1-34. The retardation of LCPL with respect to RCPL, due to differences in n_L and n_R, has resulted in the sum LCPL + RCPL producing plane polarized light rotated by an angle α with respect to the incident beam (Taken from reference 113 and used with permission).

The observation that only asymmetric or chiral chromophores display CD and ORD phenomena is a consequence of the displacement of charge upon absorption of energy [121]. Chiral molecules are those that do not display a plane or center of symmetry, i.e. they are asymmetric. As stated above, electrons undergo transitions from low energy to high energy states upon the absorption of radiation. This transition has a linear component called

the electric transition dipole moment (μ) and a circular component called the magnetic dipole moment (m) [113, 121]. The magnetic moment is generated perpendicular to the plane of the rotation. In order for chromophores to be optically active both μ and m must be non-zero, thus resulting in a helical displacement of charge upon transition (see Figure 1-35) [113, 121]. Examples of such transitions are the $n \rightarrow \pi^*$ transition (ranging from 230

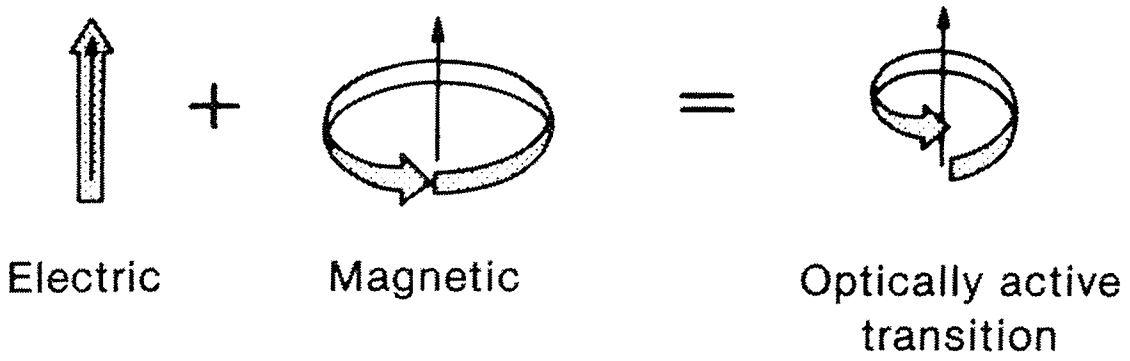


Figure 1-35. Helical displacement of charge can be considered to be the sum of an electric (μ) and magnetic (m) displacement (Taken from reference 113 and used with permission).

to 210 nm) and the $\pi \rightarrow \pi^*$ transitions (190 nm and one at higher energy) in peptide bonds [121]. Symmetrical or achiral molecules may display optical activity due to the effects of their environment. Thus, if the electronic environment of a symmetrical molecule is asymmetric, then that molecule is CD sensitive. This is the case for CD signals of aromatic amino acids, which contain no asymmetric centres but are CD sensitive in a folded protein environment [121].

1.5.3.1 Circular Dichroism and Protein Structure

The understanding of protein structure is fundamental to our understanding of protein function in biological systems. As stated in section 1.2.2, NMR and X-ray or electron diffraction analysis of proteins can provide detailed structural information about proteins at atomic resolution. Circular dichroism does not provide high resolution structural data, however it is a very valuable tool for determining changes in protein structure. Proteins have four levels of structural organization: 1) The primary structure, is the amino acid sequence; 2) Secondary structure is the organization of the polypeptide backbone into regions of regular repeating arrangements such as α-helices and β-sheets (see Figure 1-36);. 3) Tertiary structure is the organization of secondary structural elements into a

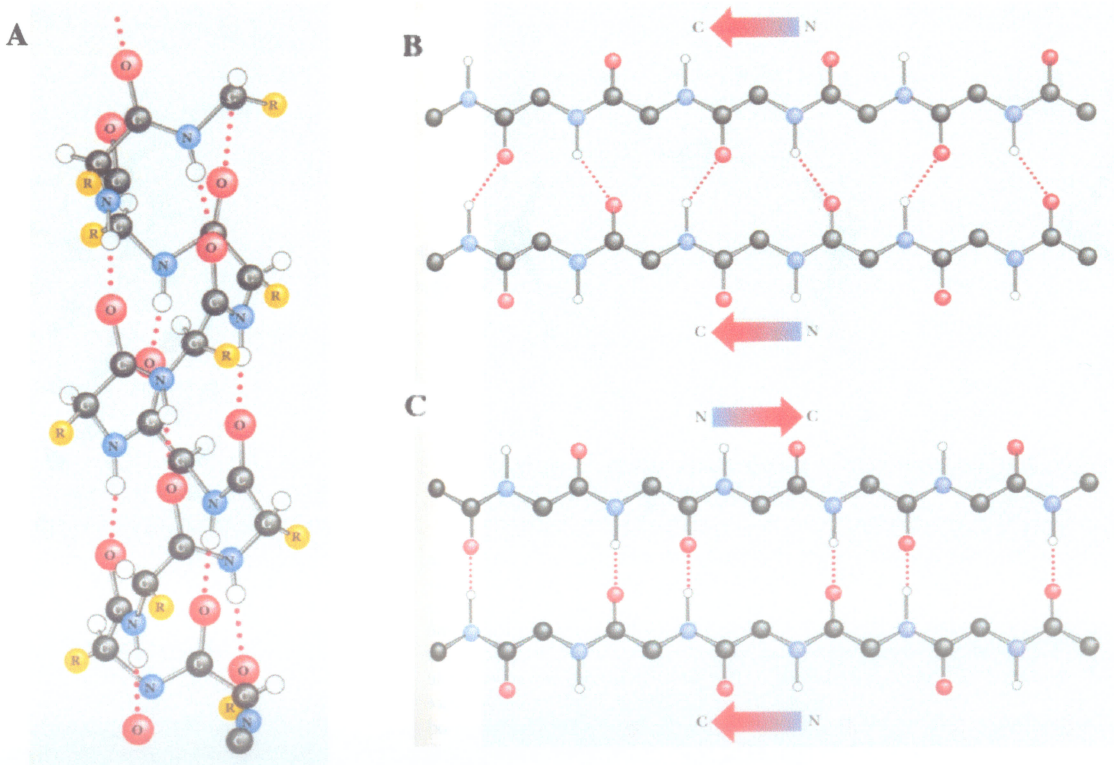


Figure 1-36. Ball and stick models of **A**, an α-helix; **B**, a parallel β-sheet; and **C**, an antiparallel β-sheet (Taken from reference 124 and used with permission).

"folded" conformation; and 4) Quaternary structure is the association of individually folded

polypeptides into a multiprotein complex [132]. CD can be used to provide information about two of these structural levels, namely secondary and tertiary structure [113, 114, 118, 119, 121, 122 and references therein]. It is important to note that while CD is not good at providing high resolution structural details, it is excellent in determining whether structural changes have occured from one sample to another.

1.5.3.2 Far-UV circular dichroism and protein secondary structure

The study of secondary structure of proteins by CD depends on the electronic transitions of the amide chromophore in the far-UV region (260-180 nm) [121]. The amide group has two $\pi \rightarrow \pi^*$ transitions, one near 190 nm with an ε_{max} of approximately 10^4 M^{-1} cm^{-1}, and another at higher energy (<190 nm). The carbonyl oxygen has two lone pairs of electrons (*n*), one in a pure 2p orbital oriented parallel to the amide plane and perpendicular to the carbonyl bond, the other having 2s and 2p character with its axis directed parallel to the carbonyl bond. The $n \rightarrow \pi^*$ transition occurs near 220 nm, is electrically but not magnetically forbidden[1], and has an ε_{max} around 100 M^{-1} cm^{-1} (see Figure 1-37) [121]. Thus, the $\pi \rightarrow \pi^*$ transition of a peptide group is more intense that the $n \rightarrow \pi^*$ one.

Since the conformation of the polypeptide backbone is different in different secondary structural elements (see Figure 1-36), there is a difference in the electronic environment around the amide chromophores and hence the characteristic CD spectrum of each element is different [122]. Early studies concentrated on individual secondary structural elements that adopt a specific conformation under given conditions (see Figure 1-38). For example,

1. Electronic transitions are forbidden when there is no linear displacement of charge.

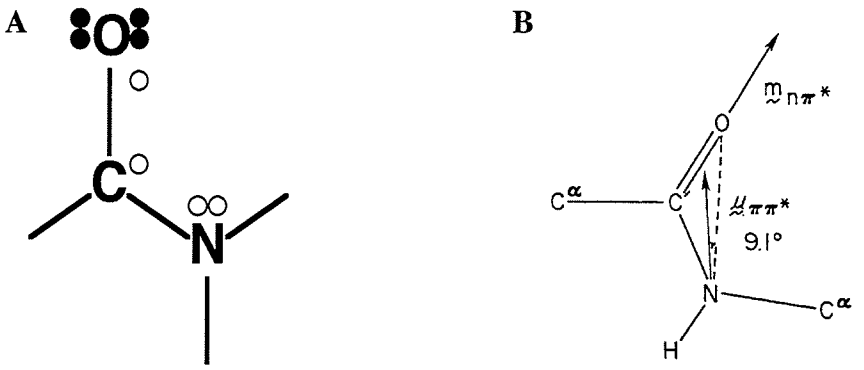


Figure 1-37. A, The amide chromophore with the electrons in the π bonds shown as open circles and the electrons in the nonbonding orbitals shown as filled circles (Taken from reference 113 and used with permission). **B**, The electric transition dipole moment of the amide π→ π* transition ($\mu_{\pi\pi^*}$) and the magnetic transition dipole moment of the amide n → π* transition ($m_{n\pi^*}$) (Taken from reference 122 and used with permission)

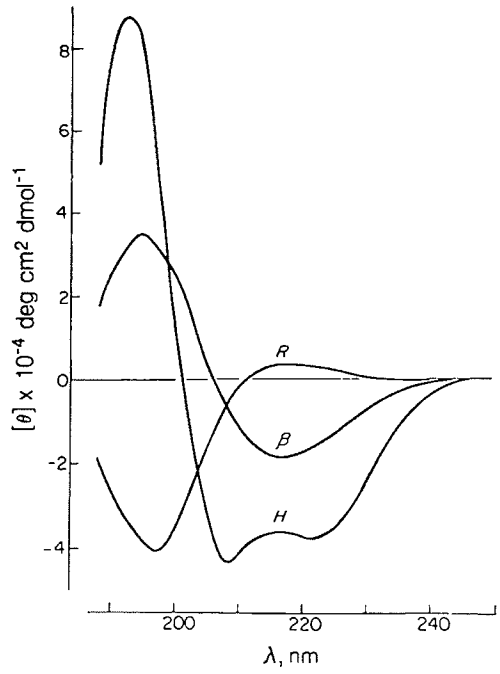


Figure 1-38. CD spectra of the helix (H), b-form, and unordered form of poly-lysine (see text for details; taken from reference 125 and used with permission).

it was found that poly-lysine adopts an α-helical conformation at alkaline pH and shows negative CD bands at 222 nm ($n \rightarrow \pi^*$) and 208-210 nm ($\pi \rightarrow \pi^*_{//}$) and a positive band at approximately 190 nm ($\pi \rightarrow \pi^*_{\perp}$) [126]. The $\pi \rightarrow \pi^*$ transition is split into two components, one polarized parallel to the helix axis (denoted $\pi \rightarrow \pi^*_{//}$) and the other polarized perpendicular to the axis (denoted $\pi \rightarrow \pi^*_{\perp}$) [126]. It was also found that poly-lysine adopts a β-sheet conformation in an alkaline sodium dodecyl sulfate solution [127] after heating to 52°C and cooling to 25°C [128], and it shows a negative CD band near 216-218 nm ($n \rightarrow \pi^*$) and a positive band between 195 and 200 nm ($\pi \rightarrow \pi^*$). Parallel and antiparallel β-sheets appear to have qualitatively similar CD spectra [125]. Further, poly-lysine at neutral pH adopts an unordered conformation with a strong negative CD band near 200 nm ($\pi \rightarrow \pi^*$) and a weak band near 220 nm ($n \rightarrow \pi^*$) which can be either positive or negative. According to Park *et al.*[129], transmembrane α-helices are different from helices in water soluble proteins in that they are approximately twice the length and they are immersed in a hydrophobic environment. The net result of these differences is that the electronic transitions can be somewhat red shifted[1], and their spectra are more intense due to the increased length of the membrane spanning helix [130].

The identification of the secondary structural elements is, in general, far from routine [122], because the CD spectrum ($f(\lambda)$) is a weighted contribution from all secondary structural elements present in a protein[2], accordingly:

$$f(\lambda) = \sum_{i=1}^{P} w_i g_i(\lambda) \qquad \textbf{(1-22)}$$

1. A decrease in the dielectric constant effectively reduces the energy difference between the excited and ground states, thus transitions occur at lower energies, i.e. they are red shifted [115]
2. This reaffirms the necessity for pure protein samples.

where $f(\lambda)$ is the weighted sum of the structural elements, P is the number of possible conformational components, $g_i(\lambda)$ is the ellipticity value of the ith component at wavelength λ, and w_i is the percentage of the ith pure component of the protein [131]. The constraints on $f(\lambda)$ would then be

$$\sum_{i=1}^{P} f(\lambda) = 1 \qquad \textbf{(1-23)}$$

and

$$f(\lambda) \geq 0. \qquad \textbf{(1-24)}$$

One method of analysis is to adjust the weights attributed to the various pure spectra (see Figure 1-38) until the experimental spectrum is calculated. However, this approach does not take into account the effects of unknown factors such as solvent or chain length on the spectra of proteins compared to the known values obtained from model polypeptide chains which themselves differ for different polypeptides [121, 125, 132].

A different approach to the deconvolution of CD spectra is the replacement of model polypeptides with measured protein CD spectra. This has become possible because much CD data are available on both water soluble and membrane proteins, many of which also have corresponding X-ray crystal structures with which to compare the CD-derived secondary structural weights [131]. There are a variety of programs available via the internet that compare a user-supplied CD spectrum with CD spectra of proteins of known crystal structure, and hence known secondary structural components. This set of proteins will be referred to as the basis set. One such program is the Convex Constraint Algorithm (CCA) written by Gerald Fasman's group [131]. The CCA program operates on the basis set of water soluble or membrane protein CD spectra, between 190 and 260 nm, and extracts the

common features of the basis set in terms of spectra and conformational weights for the corresponding spectra. The user appends the experimental CD spectrum to be deconvoluted to the provided data set and selects the number of pure components to be extracted. The CCA program satisfies equation 1-29 while maintaining the limits imposed by equations 1-30 and 1-31, such that all conformational weights are positive and the weighted sum has a value of 1. Sometimes the extracted curves do not agree with the CD spectra of model polypeptides or the weighted sum does not reliably reconstruct the user-appended spectrum. These errors indicate that the deconvolution may be invalid and that the selection of a different number of pure components is necessary. The deconvolution may also be improved by removal of one or more of the basis sets. Perzcel *et al.* [131] have shown that one or more of the basis set proteins may influence the deconvolution, such as the β-sheet content being overestimated in the deconvolution using the membrane protein basis set due to inclusion of porin protein data [131].

Along with conformational weights for the user appended experimental spectrum, CCA generates weights for the remainder of the proteins in the data set. This allows comparisons of their newly deconvoluted weights to those derived from crystal structures, providing a means of checking reliability, or "goodness of fit", of the deconvolution. Comparison of soluble protein secondary structural components determined by X-ray crystallography and by the CCA method gave a root-mean-square deviation of 0.05 for α-helix and 0.06 for β-sheet for a set of 10 reference proteins. [121]. For example, the secondary structure of the reaction center from *Rhodopseudomonas sphaeroides* was determined to be 17% α-helix, 34% α_T-helix (transmembrane α-helix), 15% β-sheet, and 34% unordered by X-ray crystallography. The values determined by use of the CCA program provided struc-

tural weights of 20% α-helix, 19% α_T-helix, 24% β-sheet, and 30% unordered, values that are considered to be in good agreement with the X-ray data [121]. Therefore, while the determination of protein structure and structural weights by X-ray crystallographic analysis provides information with an accuracy that far surpasses that acquired by CD, the use of CD and the CCA program provides a means for rapid and accurate determination of structural information, including data pertaining to protein structural changes, on protein samples not immediately amenable to X-ray analysis.

1.5.3.3 Near-UV circular dichroism and changes in protein tertiary structure

The study of protein tertiary structure by CD centres on the electron transitions of the aromatic chromophores in the near-UV region (320-260 nm). Specifically, these chromophores are the side chain residues of phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) [121]. Since the occurrence of these residues in proteins and their molar absorptivities are both low, the near-UV absorption and CD bands are much weaker than in the far-UV region (see Table 1-3). Thus, much higher concentrations of protein are required for near-UV studies [121].

Table 1-3. Spectroscopic Properties and Occurrence of the Aromatic Amino Acids[a].

	Absorbance Maximum (nm)	**Molar Absorbance ($M^{-1}cm^{-1}$)**	**Occurrence in Proteins (%)**
Phenylalanine	257.4	197	3.5
Tyrosine	274.6	1420	3.5
Tryptophan	279.8	5600	1.1

a. Reference 132

In proteins it is only the peptide and disulfide bonds that are intrinsically optically active [113]. The aromatic chromophores are not intrinsically optically active, and it is their electronic environment that is asymmetrically organized around them that induces their CD

measurable transitions. Thus, because the CD of these bands relies on their environment and on whether there is any freedom of rotation about C_α - C_β, or C_β - C_γ bonds, their optical activity is largely determined by the folded structure of the protein. Therefore, near-UV CD protein spectra are considered to be a "fingerprint" of the overall protein tertiary structure, with any structural changes being reflected in the near-UV CD spectrum [119].

The aromatic transitions are $\pi \rightarrow \pi^*$ transitions labeled ^1L_a and ^1L_b, and are perpendicular to each other in the plane of the π bonding system [119, 121] (see Figure 1-39). In

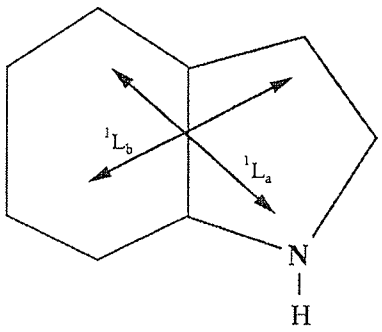


Figure 1-39. Schematic of the tryptophan indole structure with the transition moments of the ^1L_a and the ^1L_b transitions shown [119].

Phe the aromatic transitions occur near 210 nm (^1L_a) and 260 nm (^1L_b) [121, 133] whereas in Tyr they occur at 245 nm and 295 nm [119, 121]. However, intense bands produced by the overlapping ^1L_a and ^1L_b transitions of Trp at 280 nm tend to overwhelm near UV CD spectra of proteins because of the larger molar absorptivity of Trp [119, 134]. The two transitions in Trp tend to localize hydrogen bond stabilized negative charge density at the nitrogen for the ^1L_a transition and in the hydrophobic region of the indole ring (to the left in Figure 1-39) for the ^1L_b transition. The ^1L_a transition may then be red or blue shifted, relative to the ^1L_b transition, by as much as 12 nm due to hydrogen bonding. [134]. Tyr is also

capable of hydrogen bonding, resulting in a shift of around 4 nm in its transition [133]. Thus, the CD sensitive transitions in Trp and Tyr are very sensitive to their electronic environments. Stabilization of partial charges by hydrogen bonding, or burying hydrophobic regions, i.e. changing the electronic environments of the aromatic residues, is reflected by dramatic changes in the aromatic CD spectrum. On the other hand, Phe does not form hydrogen bonds and is much less sensitive to its environment [133].

Near UV CD is only useful if the protein is in a unique, folded, and stable conformation [121]. Conditions where the protein does not adopt one conformation, for example unfolded proteins that adopt a largely random conformation, will yield a protein with no aromatic CD signal. Absence of signal may also result if several conformations are present, as rotations about the C_α-C_β bond average the electronic environment of the aromatic moiety leading to the loss of asymmetry imposed by the folded structure [119, 121]. Thus, a loss of CD signal reflects the loss of protein structure, and conversely the acquisition of signal reflects protein folding [121]. The former approach is taken in this thesis, where the loss of GF near-UV CD signal is observed upon temperature denaturation of the sample (see sections 3.5.2 and 3.5.3).

1.6 Proposed Experiments

The study of membrane protein structure is central to the understanding of biological processes at an atomic level, such as solute translocation. However, as noted earlier, the study of membrane proteins lags far behind the study of water-soluble proteins. Less than 2% of all the protein structures entered into the protein databank represent membrane proteins which are thought to comprise 25% of many genomes [136]. The discrepancy between the study of water soluble and membrane proteins reflects the difficulties encoun-

tered during the cloning, expression, solubilization, and purification of membrane proteins. The majority of membrane proteins are present in small quantities, and cannot be purified from their natural expression site in quantities sufficient for structural study. Membrane proteins are never soluble in water and many, once solubilized in detergent, can rapidly denature or aggregate [137]. Furthermore, attempts at forming 3-D crystals for crystallographic analysis have often resulted in protein precipitation [138]. The large mass of the membrane protein/detergent complexes has made NMR analysis problematic [139].

Until the mid 1990s, the atomic structures of only 8 membrane proteins were solved, including bacteriorhodopsin [140], a photosynthetic reaction centre [141], and a light harvesting complex [142]. All are naturally highly expressed and purified in large quantities from natural sources. One of the most important advances in the study of membrane proteins was the application of protein-affinity tags for purification purposes [143]. Since the overexpression of membrane proteins using various expression systems made it less problematic to obtain them in milligram quantities, it was a natural evolution to affix affinity tags to the proteins to facilitate their rapid purification, thus minimizing protein aggregation and/or denaturation [137, 143]. However, to this day, few membrane proteins have been overexpressed, possibly because their overexpression is toxic to the cell [137].

We set out to develop methods to study membrane protein structure and dynamics in solution. To this end we have cloned, overexpressed, and affinity-purified the *E. coli* GF in detergent. We report secondary, tertiary, and quaternary structure analysis of the recombinant protein obtained from mass spectrometry, circular dichroism, and chemical cross-linking experiments.

CHAPTER 2

MATERIALS & METHODS

2.1 Materials

Lauryldimethylamine oxide was obtained from Calbiochem (San Diego, CA). Dodecyl-β-D-maltoside was purchased from Anatrace (Maumee, OH). Sodium dodecyl sulphate, octyl-β-D-glucopyranoside, N-Laurylsarcosine, Triton X-100, Tween 80, rifampicin, sodium deoxycholate, diatomaceous earth, urea, sinapinic acid, nitro blue tetrazolium, 5-bromo-4-chloro-3-indolylphosphate, equine heart myoglobin (16,951 Da), and bovine insulin (5,733 Da) were obtained from Sigma (St. Louis, MO). DNA amplification was performed using the Expand High Fidelity PCR system and dNTPs from Boehringer Mannheim (Laval, PQ) which was also the source of the 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside. PCR primers were made in the Department of Microbiology, University of Manitoba. Prep-A-Gene DNA purification kit was purchased from Bio-Rad Laboratories (Mississauga, ON). Mutagenesis primers, restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, and the Klenow fragment of DNA polymerase I were purchased from Life Technologies (Rockville, MD). Helper phage R408 was obtained from Promega (Madison, WI). Ready-to-go pUC18, pre-cut at the SmaI site, was obtained from Pharmacia Biotech (Baie d'Urfé, PQ). The pET28b(+) expression vector, anti-T7 antibody-alkaline phosphatase conjugate, and the bacterial strains Novablue, BL21(DE3), and BL21(DE3)pLysS were obtained from Novagen (Madison, WI). Nickel-nitrilotriacetic acid (NTA) resin was from Qiagen (Toronto, ON). Disuccinimidyl suberate, disuccinimidyl glutarate, and bis(sulfosuccinimidyl) suberate were purchased from Pierce (Rockford,

Il). HPLC-grade acetonitrile and methanol were from Fisher Scientific (Fair Lawn, NJ). Deionized water was prepared with a Milli-Q plus-TOC water purification system (Millipore, Bedford, MA. The non-porous ether-type PU membrane, 50 μm in thickness (XPR625-FS), was supplied by Stevens Elastomerics (Northampton, MA). The membrane was washed with water and methanol prior to use in order to remove polar and non-polar surface contaminants. Glutaraldehyde, osmium tetroxide, uranyl acetate, lead citrate, and Spurr resin were purchased from Marivac (Halifax, NS). All other materials were of the highest purity commercially available.

2.2 Methods

2.2.1 Expression vector construction.

E. coli genomic DNA was prepared using the methods described in Saito and Miura [144]. *E. coli glpF* DNA was amplified from genomic DNA using the polymerase chain reaction (PCR) [145] and the Expand High Fidelity PCR system from Boehringer Mannheim. The forward PCR primer oligonucleotide was identical in sequence to residues 188-218 ($^{5'}$CATTAACTCTTCA**GGATCC**GATTATGAGTC$^{3'}$; see Figure 2-1) of the published *glpF* gene [146a] and its upstream region which encodes a naturally-occurring *BamHI* restriction site (shown above in bold type). The reverse primer was complementary to residues 1043-1081 of *glpF* and its downstream region except that a non-complementary XhoI restriction site was incorporated from residues 1070-1075 ($^{5'}$ATGTTT**CTCGAG**CCC- GTAGTCATATTACAGCGAAGCTTT$^{3'}$, XhoI site shown in bold type). The

amplified DNA was purified by agarose gel electrophoresis [146b], the *glpF* DNA band

Figure 2-1. The expected DNA sequence of the glpF PCR product. The glpF start and stop codons and the restriction sites used for cloning purposes are indicated. Also indicated are the locations of the forward PCR primer (blue underline) and the reverse primer (red underline).

was excised and the DNA was extracted from the agarose using the Prep-A-Gene DNA purification kit [147]. The 3' adenine overhangs introduced by *Taq* DNA polymerase during PCR amplification were filled in by the Klenow fragment of DNA polymerase I

[148] to produce a DNA fragment with two "blunt ends". The 5' DNA termini were phosphorylated with T4 polynucleotide kinase [146b] and the DNA was inserted into the "blunt-end" SmaI site of pUC18 using T4 DNA ligase [146b]. The resultant plasmid is designated pUC18-*glpF*.

Competent Novablue cells (Novagen) were transformed with pUC18-*glpF* [149] and transformants were selected by growth on plates containing Luria-Bertani (LB) medium [146b] supplemented with 12.5 μg/ml tetracycline, 25 μg/ml ampicillin, 30 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, and 30 μg/ml isopropylthio-β-D-galactoside (IPTG)[1]. Plasmids containing *glpF* DNA were purified using the Prep-A-Gene purification kit [147]. The DNA encoding GF was excised from pUC18 using the BamHI site and XhoI restriction sites. The DNA was purified by agarose gel electrophoresis, recovered using the Prep-A-Gene kit, and inserted into a similarly digested pET28b(+) expression plasmid. The resulting construct is designated pET28g*lpF* (see Figure 2-2).

The pET28b(+) vector [150] is an IPTG-inducible expression vector designed to incorporate into the expressed protein an amino-terminal His_6 tag for rapid protein purification by immobilized-metal affinity chromatography [151, 152], a thrombin cleavage site for removal of the His_6 segment, and an 11-residue T7-epitope for Western immunoblot detection.

1. The pUC vectors carry the amino-terminal fragment of the β-galactosidase gene. Upon expression in the presence of IPTG the pUC encoded fragment complements the defective β-galactosidase encoded by the host cell (α-complementation). Bacteria expressing both fragments of the enzyme form blue colonies in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside. Successful insertion of foreign DNA into the SmaI site of pUC18 results in the inactivation of the amino-terminal portion of the β-galactosidase gene and the production of white colonies [146b].

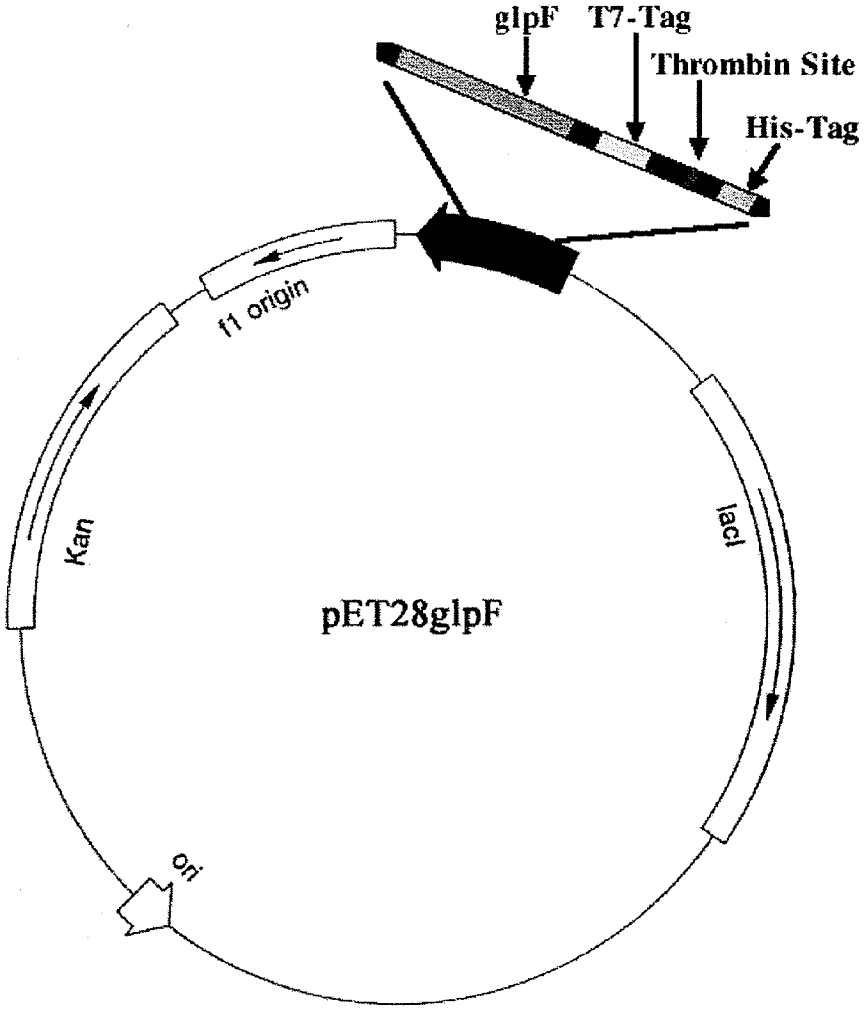


Figure 2-2. Circular map of the pET28glpF expression vector. The f1 origin of replication (used to produce single-stranded DNA with the use of f1 helper phage), kanamycin resistance gene (Kan), the gene encoding the lac operon repressor (lacI), and the origin of replication (ori) are shown.

The expected molecular weight (M_r) of the expressed glycerol facilitator including the N-terminal fusion tags is 33,505 kDa and its predicted pI is 7.56 (calculated using the program MacVector 6.5, designed by Oxford Molecular). The *E. coli* strains BL21(DE3) and BL21(DE3)pLysS [153] were used as expression hosts. DE3 indicates that the cell is a λDE3 lysogen and contains the gene for T7 RNA polymerase under *lacUV5* control. Therefore, these cells are induced to express T7 RNA polymerase upon the addition of IPTG, thereby resulting in the expression of the recombinant genes from the T7 promoter (Novagen). The pLysS-containing cells express small amounts of T7 lysozyme, a natural inhibitor of T7 RNA polymerase. This is useful for reducing basal protein expression in the absence of inducer and can improve the viability of cells expressing toxic proteins (137, 150). T7 lysozyme also accelerates cell lysis by hydrolysis of *E. coli* cell wall peptidoglycan, which it can access after the cells have been frozen and thawed.

Transformants containing the pET28*glpF* construct were selected on LB-agar plates supplemented with the 34 μg/ml chloramphenicol and 30 μg/ml kanamycin.

2.2.2 DNA Isolation and Sequencing

Plasmid DNA was isolated from 2 ml *E. coli* cultures using a method adapted from that of Kim and Pallaghy [154]. Briefly, cells from an overnight culture are harvested by centrifugation and resuspended in 500 μl of suspension solution (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 100 μg/ml DNase-free RNaseA). The cells are lysed by addition of 500 μl of 0.2 M NaOH containing 1% SDS followed by incubation for 5 minutes at room temperature. The mixture is neutralized by 500 μl of 4 M potassium-acetate, pH 4.8. Precipitated protein, genomic DNA, and cellular debris are pelleted by centrifugation and the DNA-containing supernatant is transferred to a microcentrifuge tube containing an equal

volume of diatomaceous earth solution[1]. An equal volume of 6 M guanidine hydrochloride (Gdn-HCl) is added to the DNA/diatomaceous earth mixture, it is mixed gently, and the sediment is pelleted by centrifugation. The supernatant is carefully removed and discarded. The Gdn-HCl wash is repeated twice, followed by 3-6 washes with 20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 0.4 M NaCl, 50% ethanol. The pellet is air-dried, and the pure plasmid DNA is eluted by resuspending the pellet in 100-200 μl of water or TE buffer (20 mM Tris-HCl, pH 8.0, 1 mM EDTA) and removed from the diatomaceous earth by centrifugation and retention of the supernatant. The purity and yield of the DNA is checked by agarose gel electrophoresis, and then it is stored at -20°C.

DNA for sequencing was prepared by the modified alkaline lysis method suggested by Perkin-Elmer [http://www.ucalgary.ca/~dnalab/Plasmidprep.html]. *GlpF* DNA was sequenced (see Figure 2-3) on an ABI 373A automated sequencer at the University of Calgary Core DNA Services and two PCR-induced mutations, Asp168Val and a silent mutation for valine (GTG -> GTA), were discovered. The D168V mutation was reverted to wild type by Kunkel mutagenesis [155] using f1 helper phage R408 and the f1 origin of replication on the pET28b(+) vector to produce the single-stranded plasmid DNA [156, 157]. The expression vector will be referred to as pET28*glpF* (see Figure 2-2).

1. Sigma's normal quality diatomaceous earth is pure enough for most applications. It is suspended in double deionized water at 50 mg/ml, and left to sediment for one day. The water and milky suspension above the sediment is carefully removed and replaced with an equal volume of fresh water, and the sediment is resuspended. This process is repeated until the water above the sediment remains clear and suspension free, at which point the diatomaceous earth solution is ready for use (adapted from reference 154).

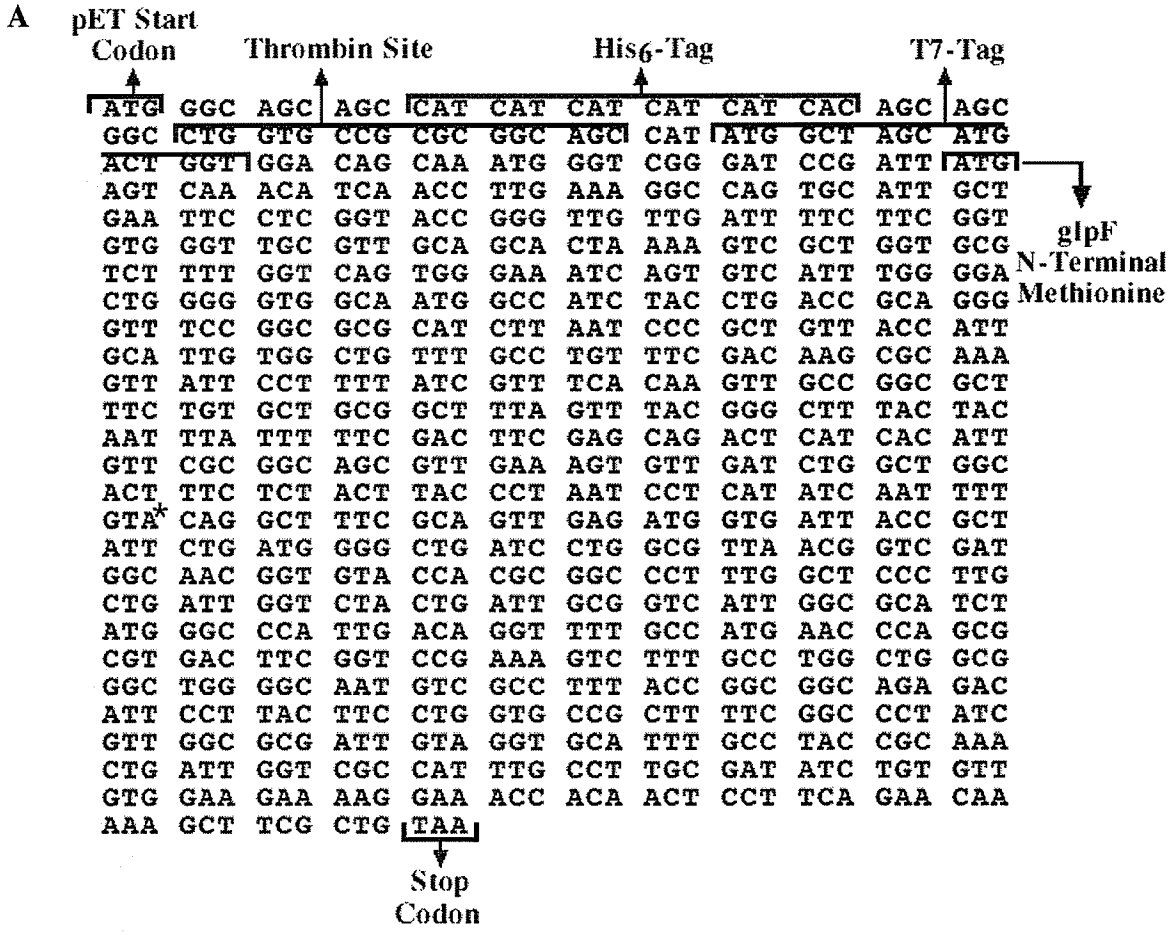


Figure 2-3. A, The DNA sequence of the recombinant *glpF*-pET expression product. Indicated are the segments coding for the recombinant start codon, His_6-tag, thrombin cleavage site, T7-tag, and the *glpF* wildtype N-terminal methionine. The asterisk indicates the site of the silent mutation coding for valine; the wildtype codon is GTG. **B**, The amino acid sequence of the recombinant *glpF*-pET expression product. Indicated are the segments coding for the recombinant start codon, His_6-tag, thrombin cleavage site, T7-tag. The *glpF* wildtype N-terminal methionine is boxed, and the two NPA motifs are underlined.

2.2.3 Solulization of Glycerol Facilitator

The detergents sodium dodecyl sulphate (with and without 4 M urea), octyl-β-D-glucopyranoside, N-laurylsarcosine, Triton X-100, Tween 80, dodecyl-β-D-maltoside, sodium deoxycholate, and lauryldimethylamine oxide were tested for their ability to extract GF from insoluble cell fractions. Cells induced to express GF (see section 2.2.4) were pelleted, flash frozen in liquid nitrogen, resuspended in 500 μl of buffer A (50 mM sodium phosphate pH 7.6, 50 mM NaCl, and 10 mM β-mercaptoethanol) plus 1% detergent, and incubated with frequent agitation for 1 hour. The suspensions were centrifuged and the supernatants were analyzed for the presence of GF by SDS-PAGE [158] and Western immunoblotting.

2.2.4 Glycerol Facilitator Overexpression and Purification

BL21(DE3)pLysS cells containing pET28*glpF* were grown in antibiotic-supplemented Terrific Broth[1] (TB) [146b] at 37°C. Glycerol Facilitator expression was induced with the addition of 1mM IPTG when the culture reached an optical density measured at 600 nm (OD_{600}) of 0.6-0.7. The bacterial RNA polymerase was arrested by the addition of rifampicin (200μg/ml) at 1 hour following induction, and cells were harvested after a further 1 hour of growth [160, 161]. The cells were chilled on ice for 5 minutes and centrifuged at 4,000g for 10 minutes at 4°C using a Sorvall RC-5B Refrigerated Superspeed Centrifuge and the GSA rotor. The cell pellets were flash frozen in liquid nitrogen and stored at -20°C. Prior to assay the cell pellets were defrosted and re-suspended in buffer A, and the cells

1. Terrific Broth is a rich growth medium that permits a higher cell density than LB, and was designed as an improved media for the preparation of plasmid DNA [159].

were then lysed by passage through 18- and 22-gauge needles. DNA was sheared by sonicating (Fisher Sonic Dismembrator Model 300) the lysed cells on ice three times for 30 seconds at 30% output. The ruptured cells were centrifuged at 4°C for 20 minutes at 2,800g using the Sorvall GSA rotor. The resulting pellet was re-suspended in buffer A containing a detergent, and stirred at room temperature for 1 hour. The concentrations of the detergents were 3 mM DM, 150 mM SDS, or 150 mM SDS and 4 M urea. For solubilization with DM it was necessary to extend the incubation time a further 30 minutes, or until the majority of the visible cellular material had dissolved. Insoluble material was removed by centrifugation at 4,000g for 10 minutes at 4°C using the Sorvall GSA rotor. In the case of SDS or SDS/urea solubilized material centrifugation was performed at room temperature using a Dynac bench top centrifuge for 20 minutes set at the maximum speed (i.e. 90). The supernatant was added to 1 ml (bed volume) Ni^{2+}-NTA Sepharose pre-equilibrated in detergent solution, and the slurry was stirred at room temperature for 1 hour. Prior to being poured into a column and washed with the detergent solution the Ni^{2+}-NTA Sepharose was pelleted by low speed centrifugation using a Dynac bench top centrifuge for 5 minutes at a speed setting of 40-50. The column washes were monitored by measuring the absorbance of the eluate at 280 nm. Elution was begun once the A_{280} was below 0.1. Prior to elution, in the case of extraction with urea, the urea was slowly removed using a 100 ml urea gradient composed of 50ml of 4M urea and 50ml of buffer A. The urea gradient was followed by washing the column with several volumes of buffer A plus 150 mM SDS at pH 7.6, followed by a second wash at pH 7.2. For the DM and OG preparations the column was washed in several column-volumes of buffer A plus 3 mM DM at pH 6.5. For the OG preparation the DM was exchanged for OG by washing with several column volume of buffer

A plus 100 mM OG. Facilitator was then eluted in buffer A plus 150 mM SDS at pH 6.5, or 30 mM DM at pH 4.0, or 100 mM OG at pH 4.0. All samples were analyzed by the SDS-PAGE and Western immunoblotting techniques described below.

2.2.5 Electrophoresis and Immunoblotting

All SDS-PAGE samples were prepared by incubating them in 50 mM Tris (pH 6.8), 2% SDS, 0.1% Bromophenol blue, 10% glycerol and 100 mM DTT for 10 minutes at 37°C prior to gel loading. Boiling of GF samples is avoided as this results in protein aggregation to such a degree that a large portion is retained at the interface between the stacking and resolving gel and in the sample well itself. This is a problem commonly associated with boiling of membrane proteins [162]. Separation was achieved by SDS-PAGE in Laemmli discontinuous gels [158] composed of a 2.5% acrylamide stacking gel and a 10% resolving gel (unless otherwise noted). After electrophoresis, proteins were visualized either by Coomassie staining or by immunodetection after electrophoretic transfer to a nitrocellulose membrane using the Mini Trans-Blot® Cell from Bio-Rad. Briefly, the SDS gel and nitrocellulose were equilibrated in transfer buffer (25mM Tris, 192mM glycine, and 20% methanol) for 10-15 minutes at 4°C. The gel holder was then loaded from the cathode to the anode in the following order: pre-soaked sponge, pre-soaked filter paper, SDS gel, nitrocellulose, pre-soaked filter paper, pre-soaked sponge. The "sandwich" was placed into the buffer tank filled with 4°C transfer buffer, an ice pack, and a magnetic stirbar. After the lid was placed on the tank, the entire unit was packed in ice and water in a Styrofoam container which was then placed on a magnetic stirrer. The leads were connected to a power pack set at 100-110 volts (constant voltage) and transfer proceeded for 2 hours. Following the protocol described by Novagen [152], the nitrocellulose was incubated in anti-T7 antibody-

alkaline phosphatase conjugate which binds to the 11 amino-acid T7 epitope on the amino-terminus of the recombinant GF. Immunoreactive proteins were visualized by color development in nitro blue tetrazolium (334 μg/ml) and 5-bromo-4-chloro-3-indolylphosphate (175 μg/ml) in a buffer composed of 100mM Tris-HCl, pH9.5, 100mM NaCl and 1mM $MgCl_2$.

2.2.6 Sub-cellular Localization of overexpressed GF

BL21(DE3) cells containing pET28*glpF* were grown in antibiotic-supplemented LB [146b] at 37°C until the culture reached an OD_{600} of 0.7. GF expression was induced by the addition of IPTG to a final concentration of 1 mM. After 3 hours the cells were chilled on ice for 5 minutes and harvested by centrifugation at 4,000g for 10 minutes using a Sorvall RC-5B Refrigerated Superspeed Centrifuge and the GSA rotor. The cell pellets were flash frozen in liquid nitrogen, defrosted, and re-suspended in buffer A containing 1 mM PMSF. The cells were then lysed by passing them through 18- and 22-gauge needles. DNA was sheared by sonicating the ruptured cells on ice three times for 30 seconds. An insoluble fraction, possibly containing inclusion bodies, was obtained by centrifuging the ruptured cells at 4°C for 15 minutes at 17,000g. The supernatant was then centrifuged at 4°C for 2 hours at 465,000g (Beckman L8 70M Preparative Ultracentrifuge) yielding a pellet composed of crude membranes [163]. The presence of GF in the 3 cell fractions was determined with Coomassie Brilliant Blue R-250-stained SDS polyacrylamide (SDS-PAGE) gels and by Western immunoblot analysis of protein electrophoretically transferred to nitrocellulose membranes (as above).

2.2.7 Transmission Electron Microscopy

BL21(DE3) cells overexpressing GF, and non-induced control cells, were harvested 3 hours after induction and fixed immediately with glutaraldehyde to a final concentration of 2% (v/v). Agarose cores were prepared using 3% agarose at 45°C-50°C. The agarose cores were post-fixed in 2% (w/v) osmium tetroxide in 200mM cacodylate buffer (pH 7.2), en bloc stained in 1% (w/v) uranyl acetate, dehydrated through a graded series of ethanol, and embedded in Spurr resin. Ultra-thin sections were cut on the Reichert ultracut ultramicrotome, mounted on copper grids, and post-stained with uranyl acetate and lead citrate. Prepared sections were examined with a Hitachi H-7000 transmission electron microscope (EM studies were done by Lynn Burton, Department of Botany, University of Manitoba).

2.2.8 Mass Spectrometry

Purified membrane protein was prepared for mass spectrometry by exchanging SDS for 70 mM octyl-β-D-glucopyranoside by washing Ni^{2+}-NTA-bound GF with a gradient of decreasing concentrations of SDS and increasing concentrations of OG. GF was eluted from the column in OG as described above. Samples were prepared for MALDI-TOFMS using a polyurethane (PU) membrane as a sample support [164]. Briefly, 2 μl of OG solubilized GF was placed on the polyurethane membrane and allowed to dry slowly. Methanol (2 μL) was then added and also allowed to dry. Water was added, in 20 μL aliquots, in intervals of one minute to the sample spot, and removed prior to the addition of matrix. The sinapinic acid matrix solution (2 μL) was then added to the washed sample and allowed to crystallize slowly. After the matrix had crystallized, the membrane was placed on a silver disk which had been coated with a thin layer of adhesive (Spraymount, 3M).

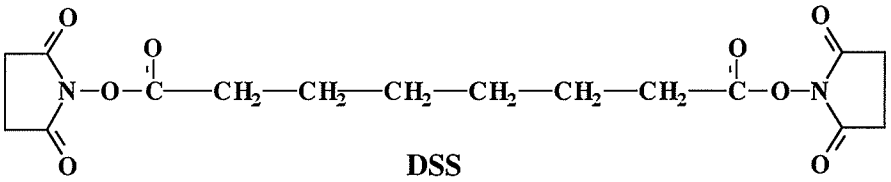
Excess membrane was trimmed from the disk and the disk was placed into the MALDI probe.

MALDI-TOFMS was performed in the linear mode on a reflecting time-of-flight mass spectrometer, an instrument built in the Time-of-Flight Lab, University of Manitoba [165]. An accelerating potential of 25 kV was used. Spectra were obtained using a nitrogen laser (337 nm) (VSL 337 ND, Laser Science Inc., Cambridge, MA, USA) with the fluence adjusted slightly above threshold. To avoid saturation of the detector by low mass matrix ions, the detector was pulsed on for approximately 19,000 ns after each laser shot. External calibrations for measurements using the PU membrane were performed with protein standards (bovine insulin, equine myoglobin) prepared on similar membrane targets (MS analysis was performed by Mark McComb, Department of Chemistry, Univeristy of Manitoba).

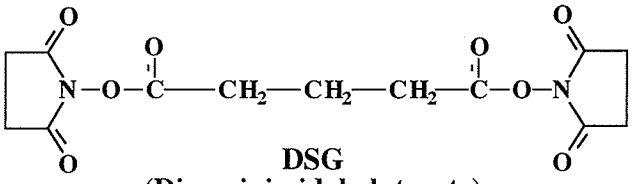
2.2.9 Chemical Cross-Linking

E. coli membranes were isolated by centrifuging lysed GF expressing cells (see section 2.2.4) at 4°C for 10 minutes at 10,000g. The supernatant was recovered and centrifuged at 4°C for 1 hour at 180,000g in a Beckman L8-70M Preparative Ultracentrifuge. The resulting pellet, containing crude *E. coli* membranes [166], was resuspended in buffer A (minus β-mercaptoethanol) to yield a 10% (w/v) suspension. The chemical cross linkers disuccinimidyl suberate (DSS), disuccinimidyl glutarate (DSG), and bis(sulfosuccinimidyl) suberate (BS) are *N*-hydroxysuccinimide esters (NHS-esters). They react with primary amines such as the ε-amino of accessible lysine side-chains to form covalent cross links (see Figures 2-4 and 2-5). For cross-linking, 6 μl of 50 mM DSS or DSG in dimethylsulfoxide or 6 μl of 50 mM BS in 20 mM sodium phosphate, pH 7.5, 150 mM NaCl were mixed with 100 μl of a 2% membrane suspension, giving a final concentration of 2.8 mM

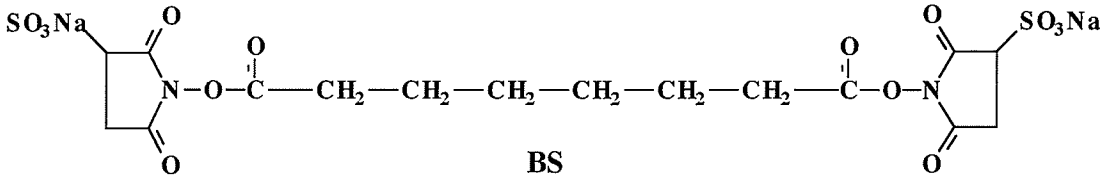
cross linker. Incubations were carried out at 4°C for various periods of time ranging from 30 minutes to 24 hours. Cross linking was quenched with the addition of 2.65 μL of a 2mM Tris-HCl pH 8.0 solution followed by a further 15 minute incubation at 4°C. All samples were analyzed by SDS-PAGE and Western immunoblotting.



DSS
(Disuccinimidyl suberate)
Mw 368.35 Da
Spacer Arm Length 11.4Å



DSG
(Disuccinimidyl glutarate)
Mw 326.26 Da
Spacer Arm Length 7.72Å



BS
(Bis(Sulfosuccinimidyl) suberate)
Mw 572.43 Da
Spacer Arm Length 11.4Å

Figure 2-4. Structures of the *N*-hydroxysuccinimide cross linkers disuccinimidyl suberate, disuccinimidyl glutarate, and bis(sulfosuccinimidyl) suberate.

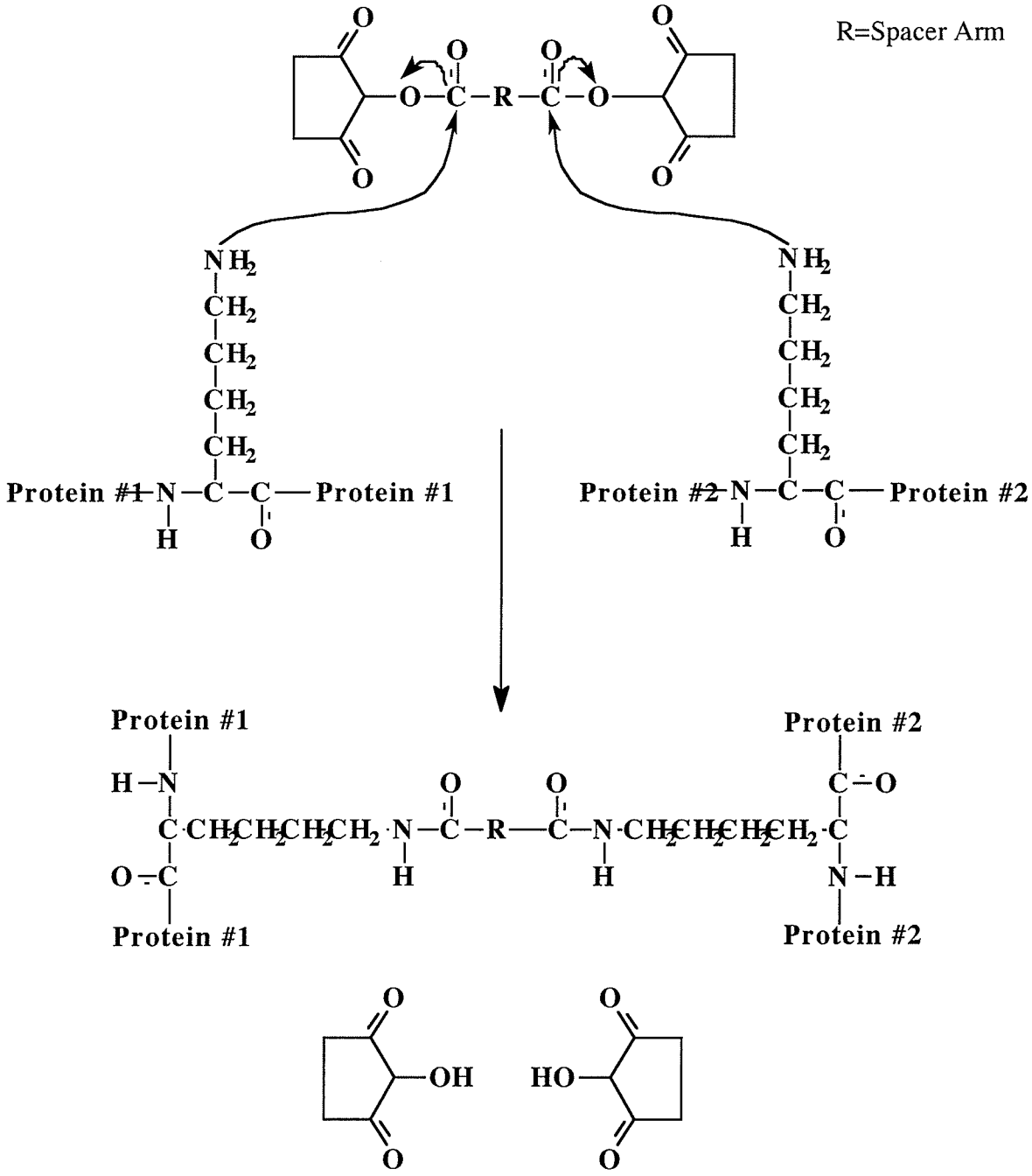


Figure 2-5. Reaction scheme for the NHS-ester mediated crosslinking of two protein segments via the ε-amino of lysine residues [167].

2.2.10 Xylitol Transport Assay

Transport assays were performed as described by Heller *et al.*[106]. Briefly, induced and non-induced cells were harvested by centrifugation at 4°C for 10 minutes at 2,800g. A concentrated cell suspension was prepared by resuspending the cell pellet in buffer A at one third the original volume. The concentrated cells were then diluted into buffer A containing 0 or 250 mM xylitol to a final optical density at 600 nm (OD_{600}) of 1.0, and any changes in OD_{600} resulting from osmotically-induced plasmolysis (shrinking) and deplasmolysis (swelling) were monitored continuously for 5 minutes (see section 1.5.2.2). Transport activity was blocked with the addition of 1.0 mM $HgCl_2$ to the *E. coli* cells 5 minutes prior to dilution into xylitol containing buffer [107].

2.2.11 Circular Dichroism Analysis and Deconvolution

All CD experiments were acquired on a Jasco-500A CD spectropolarimeter interfaced to a personal computer, with sample temperature controlled by a Haake G circulating water bath. The software used for data collection (written by Dr. Kruczynski at the University of Manitoba) collects the CD signals in millivolts. The CD intensity and wavelength of the spectropolarimeter was routinely calibrated using solutions of *d*-10-camphorsulfonic acid [168]. For accurate structural determination from far-UV CD spectra it is necessary to acquire data into the vacuum-UV region (below 190 nm), or to the lowest wavelength possible with the available instrumentation, so that the 190 nm $\pi \rightarrow \pi^*_{\perp}$ transition is recorded [169]. Figure 1-38 illustrates the importance of recording this transition where α–helix, β-form, and unordered form of poly-lysine all show different intensities and signs in this wavelength region. The Jasco 500A spectropolarimeter can provide good quality spectra down to 200 nm, below which atmospheric oxygen absorption increases and spectral qual-

ity rapidly declines. To reduce the degree of spectral loss below 200nm the $N_2(g)$[1] flow through the instrument is increased from 5 L/min to 25 L/min to flush the oxygen from the sample chamber. The increase in $N_2(g)$ flow thus facilitated the acquisition of CD spectra down to 190 nm with no apparent loss of spectral quality. Lastly, to improve spectral quality below 200 nm the scan rate was reduced from 2 nm/min with a time constant of 8 seconds to 0.5 nm/min. Spectra at wavelengths between 210-185 nm were obtained with a time constant of 32 seconds. This reduction in scan rate and increase in time constant facilitated the collection of more photons between 210-185 and thus decreased the statistical noise (see above).

CD spectra of purified GF were acquired utilizing the CD optimization parameters of Hennessey and Johnson [170]. To this end the first two mirrors in the Jasco 500A CD spectropolarimeter were replaced, thus limiting the loss of signal intensity due to ozone damaged mirrors[2]. Total absorbance, including the cell, solvent, and sample, was kept below 1.0 so that at least 10% of the light is transmitted [122]. To reduce the loss of spectral quality due to solvent absorption, a cylindrical cell with a 0.05 cm and 0.5 cm pathlength was used during far and near-UV data acquisition, respectively. For far-UV CD acquisition the 0.05 cm cell reduces the affects of solvent absorption by a factor of 10 compared to a cell of path length 1 cm. However, this does require that the sample concentration be increased by a factor of 10. For accuracy the spectropolarimeter was calibrated using an aqueous solution of (+)-10-camphorsulphonic acid [171]. Lastly, the statistical noise was reduced by increasing the number of photons collected by increasing the time constant and

1. Nitrogen gas is used to flush oxygen from both the optics of the spectropolarimeter and the sample chamber.

2. The high intensity light source rapidly converts oxygen to ozone, which damages the optics rapidly [169].

decreasing the scan speed (to allow the instrument to respond to signal changes at the higher time constant) [122, 170]. As the time constant is increased and the scanning rate is slowed, the improvements in spectral quality diminish quickly. Therefore Hennessey and Johnson [170] proposed that at a given scan rate the optimum time constant be determined by

$$\left(\frac{\text{Time Constant (s)}}{60(\text{s})}\right) \cdot \text{Scan Rate (nm/min)} \leq 0.33 \qquad \textbf{(2-1)}$$

During spectral acquisition the instrument was purged with $N_2(g)$ at 5 L/min for wavelengths greater than 210 nm and at 25 L/min below 210nm. For CD experiments, GF was eluted from the Ni^{2+}-NTA column in a 10 mM sodium phosphate buffer containing 10 mM NaCl and detergent (30 mM DM at pH 4.0 or 100 mM OG at pH 4.0, or 150 mM SDS at pH 6.5 or 150 mM SDS with 4 M urea at pH 6.5). Adjustment of pH was done by adding detergent-buffer solutions at the appropriate pH. Total absorption was always below 1.0 above 200nm. For far-UV analysis the samples were placed in a water-jacketed quartz cuvette with a 0.05 cm path length and CD spectra were collected at 2 nm/min between 260-205 nm with a time constant of 8 seconds, and at 0.5 nm/min between 210-185 nm with a time constant of 32 seconds, at 25°C. For near-UV analysis the samples were placed in a quartz cuvette with a 0.5 cm path length and the CD spectra were collected at 2 nm/min between 320-260 nm with a time constant of 8 seconds. Baselines were collected in the same fashion on buffer solutions and spectra were baseline corrected. The protein concentrations were determined using the calculated molar absorptivity at 280 nm of 38,305 M^{-1} cm^{-1} [111]. Ellipticity (θ) was calculated using the equation:

$$\theta = m^{\circ}(mV \cdot C) \quad \textbf{(2-2)}$$

where m° is the sensitivity setting of the spectropolarimeter in millidegrees per cm, mV is the baseline-corrected signal in millivolts, and C (cm/millivolt) is an experimentally determined correction factor[1]. Mean Residue Ellipticities ($[\theta]$ x10^{-3} deg•cm^2•dmole^{-1}) were calculated using the equation [122]:

$$[\theta]_M = \frac{M}{n} \cdot \theta/(10 \cdot l) \cdot c \quad \textbf{(2-3)}$$

where M is 33,505 grams per mole, n is the number of amino acid residues in the protein, θ is the measured ellipticity in millidegrees, l is either 0.05cm or 0.5cm, and c is the protein concentration in g/L. Deconvolution of the far-UV CD spectra, into pure component spectra, was performed using the Convex Constraint Algorithm (CCA) written by Perczel *et al.* [131] (see section 1.5.3.1).

2.2.12 Curve Fitting of CD Temperature Titration Data

Near- and far-UV CD temperature titrations were individually acquired for GF solubilized in 150 mM SDS and 30 mM DM. The titration data were fitted to the equation

$$\theta_{obs} = \frac{\theta_{fold} + \theta_{unfold} \cdot 10^{\frac{2(T - T_m)}{\Delta T}}}{1 + 10^{\frac{2(T - T_m)}{\Delta T}}} \quad \textbf{(2-4)}$$

where θ_{obs} is the observed ellipticity, θ_{fold} the ellipticity of fully folded protein, and θ_{unfold} the ellipticity of unfolded protein [172]. Equation 2-4 was used to fit the thermal dependence of the ellipticity to a two state unfolding transition, and yields a value for T_m, the midpoint of the transition where the concentration of folded protein equals that of unfolded.

1. The correction factor is obtained from comparing the experimental intensities at 192.5 nm and 290.5 nm obtained from a spectrum of 1.0 mg/ml *d*-10-camphorsulfonic acid to the published values; a ratio of -2.0 indicates a calibrated instrument [168]. The correction factor is necessary to correct for loss of signal intensity due to an aging deuterium lamp, oxidized mirrors, and other instrumental artifacts.

The fit also yields ΔT, the width of the transition which is the temperature range from log K_{eq} = +1 to log K_{eq} =-1 and covers the central 80% of the transition. The assumption is that there is a linear dependence of log K_{eq} on temperature with a slope of $2/\Delta T$ and an intercept of $-2T_m/\Delta T$. That is, log $K_{eq} = 2(T-Tm)/\Delta T$. The values θ_{fold} and θ_{unfold} represent the ellipticity at 208 nm or 273 nm (far- and near-UV respectively) for folded and unfolded GF and were estimated from the denaturation curves (see figure legends for Figures 3-30 to 3-32). A non-linear least squares fitting algorithm in the program *Mathematica*™ was used to determine the T_m and ΔT values for GF unfolding.

CHAPTER 3

RESULTS

3.1 Cloning the Glycerol Facilitator

The gene encoding the *E. coli* glycerol facilitator (GF) was amplified using the polymerase chain reaction [145] with *E. coli* genomic DNA as the template. Attempts to insert the GF PCR product directly into the expression vector were unsuccessful due to incomplete or inefficient restriction enzyme digestion of the PCR product or inefficient ligation of GF DNA to the expression vector. Thus, to insert GF DNA into the expression vector it was necessary to clone it first into the vector pUC18 and subsequently subclone it into pET28b(+) (see Section 2.2.1).

Upon sequencing the pET28*glpF* expression construct two PCR-induced mutations were discovered; Asp168Val and Val146Val (GTG -> GTA). The latter mutation was ignored as it is a silent mutation and does not change the identity of the amino acid in the translated product, however it was necessary to repair the former. The Val168Asp "back-mutation" of GF was accomplished by first producing a single-stranded DNA template using f1 helper phage R408 and the f1 origin of replication on the pET28b(+) expression vector [156, 157]. The single-stranded pET28*glpF* DNA was mutagenized using the Kunkel mutagenesis technique [155] and re-sequenced to confirm the reversion.

3.2 Optimization of GF Expression

To optimize GF expression levels the *E. coli* expression strain, culture temperature, growth medium, cell density at the time of induction, and time of harvest after induction were varied. Initially the growth conditions suggested by Novagen were used such that

cells are induced when they are in mid-log phase (OD_{600} approximately 0.6 to 0.7) and harvested 3 hours later [152]. Using these parameters, growth curves (not shown) indicated that *E. coli* induced to overexpress the glycerol facilitator grow normally after induction by 1 mM IPTG.

The first step in expression optimization dealt with determining the best time, or cell density, for the induction of protein expression. A series of experiments were performed where expression was induced at different points throughout the mid-log growth phase (OD_{600} = 0.4 – 1.4). The BL21(DE3) strain of *E. coli* cells were induced at the following optical densities: 0.4, 0.5, 0.7, 0.8, 0.9, 1.0, 1.2, and 1.4. The cells were grown for 3 hours at 37°C, 1 ml aliquots were harvested and pelleted, and the whole cells were then electrophoresed. Subsequent Western immunoblot analysis of the SDS-PAGE gels (Figure 3-1) indicated that cells induced at the lowest cellular density, OD_{600} = 0.4, had significantly more (approximately double the amount) GF monomer and dimer than cells induced at higher cell densities. Estimation of GF levels using Western immunoblot analysis was

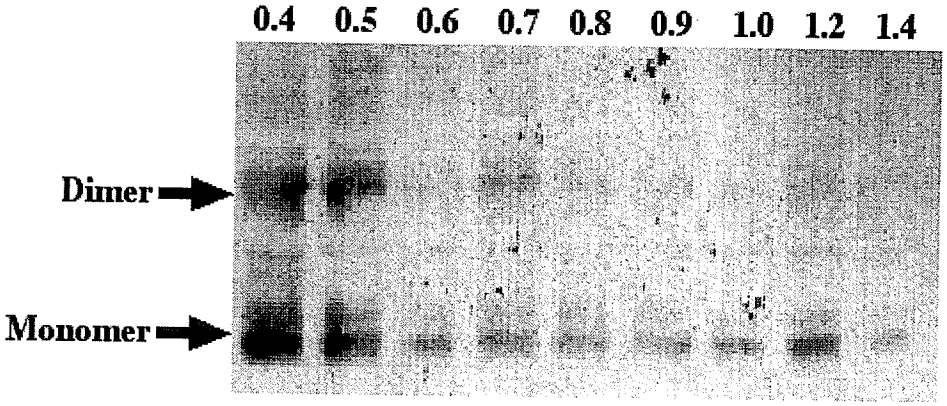


Figure 3-1. Western blot displaying GF and the optimization of the cell density at the time of induction of expression (monomer, dimer and OD_{600} are indicated, see text for details).

necessary since crude cell extracts were used. A duplicate SDS-PAGE gel stained with Coomassie blue indicated that the individual lanes were loaded with approximately equal amounts of cells, because the intensities of Coomassie-stained protein bands were the same (not shown). Thus, the elevated levels of immuno-sensitive material in the Western blot in cells induced during early log phase are not a result of gel overloading, but likely reflect an elevation in GF content.

Another series of experiments were designed to investigate the effects of the length of the interval between induction of GF expression and the harvesting of cells. Cells were induced to express GF at $OD_{600} = 0.7$ (mid-log phase of growth) and harvested at the following times (in hours) after induction: 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 12, >24. Again, 1 ml aliquots were harvested and pelleted. It was estimated (visually) that the size of the cell pellets increased with time, until 5 hours post induction, indicating that cell number steadily increases from 0.5 to 5 hours. Furthermore, SDS-PAGE gels and Western immunoblot analysis show that the GF levels peak at 1.5 - 2 hours (not shown), and that the levels of GF expression decline markedly by 4 hours post-induction, relative to total cell protein. Thus, the optimum time for harvesting GF expressing cells is approximately 2 hours post induction. These data suggest that GF may be proteolytically degraded *in vivo*, and that harvesting the cells at an earlier time provides an elevated level of GF.

It has been shown that the temperature during expression appears to affect the protein expression levels in bacterial systems [137]. Comparing the GF expression levels in cultures grown at 25°C, 30°C, and 37°C (see Figure 3-2) indicates that optimum GF expres-

sion is attained at 30°C [148b]. The exceptional increase in expression levels seen at 30°C,

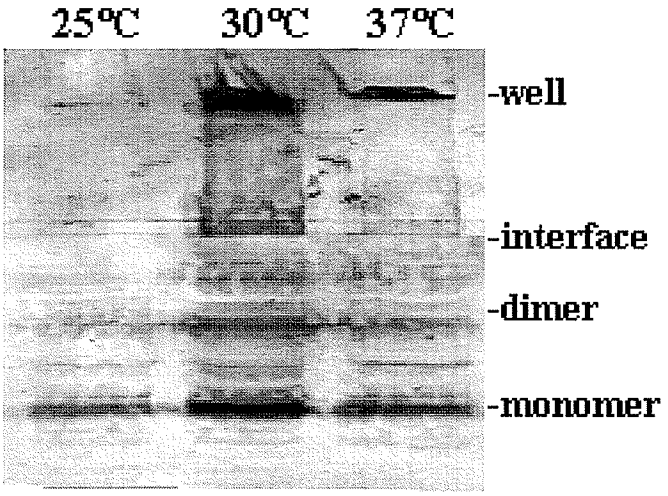


Figure 3-2. Western blot displaying the optimization of culture temperature during expression (temperatures are indicated). The position of the SDS-PAGE sample well and stacking/resolving gel interface, along with the position of GF monomer and dimer, are indicated. Immunolocalization of GF in the sample well and interface indicates extensive aggregation (see section 3.5.2 for details).

compared to 37°C, may indicate that the insertion of GF into the membrane, rather than its expression, is rate limiting. Thus, growth at lower temperatures may allow expressed GF to be inserted into the membrane rather than proteolytically degraded [137]. The low levels of GF expressed at 25°C most likely reflects the reduced growth rate of *E. coli* at room temperature, i.e. a lower number of *E. coli* cells expressing GF.

The cell line used for GF expression was also analyzed for expression efficiency. Expression of GF in the cell lines BL21(DE3) and BL21(DE3)pLysS indicated that the strain harboring the pLysS plasmid is a better expression host, approximately doubling GF expression levels (determined visually by Western analysis, data not shown). The pLysS plasmid contains the gene encoding T7 lysozyme so that cells containing this plasmid accumulate low levels of T7 lysozyme. The advantage gained by the use of pLysS plasmids is

three-fold. First, the plasmid confers resistance to the antibiotic chloramphenicol. A second advantage is that T7 lysozyme is a natural inhibitor of T7 RNA polymerase, the IPTG-inducible RNA polymerase responsible for expression from the pET28glpF expression vector [173, 174]. The presence of T7 lysozyme increases the tolerance of cells to plasmids with toxic inserts [152] since expression from "leaky" promoters (i.e. expression from a promoter in the absence of induction) is essentially nonexistent due to the inhibition of T7 RNA polymerase. Therefore, there is no background expression from expression vectors such as pET28glpF containing the T7lac promoter [152]. The observation that BL21(DE3)pLysS is a better expression host than BL21(DE3) suggests that overexpression of GF may be toxic to the cell. Thus, the low levels of T7 lysozyme expressed by the pLysS plasmid likely reduced background GF expression and therefore increased cell viability before induction.

A third advantage of using BL21(DE3)pLysS for expression is in the preparation of cell extracts. During growth of the pLysS harboring strain, the T7 lysozyme is retained in the cytoplasm [124]. However, after the cells are harvested, one freeze-thaw cycle causes a disruption in the cell membrane and provides a means for the T7 lysozyme to access the peptidoglycan layer, resulting in cell lysis [152].

The pET expression vectors utilize a T7 RNA polymerase/promoter system, such that transcription of the T7 RNA polymerase gene is directed from a lac UV promoter, and is inducible by the addition of IPTG. This is an advantage for overexpression since only T7 RNA polymerase, and not *E. coli* RNA polymerase, can trancribe from the T7 promoter. One advantage is that toxic protein expression is limited to times after induction with IPTG. A second advantage is that the pET system facilitates the use of rifampicin to inhibit *E. coli*

RNA polymerase, thereby halting bacterial protein synthesis at transcription [161, 175]. Rifampicin is a inhibitor of bacterial RNA polymerase. Inhibition is complete 10 minutes after rifampicin addition, provided an excess of rifampicin is used (200μg/ml is commonly used) [161, 175]. Since rifampicin has no effect on T7 RNA polymerase only those genes under the control of a T7 promoter are expressed upon its addition to a growing and induced culture. The idea is that upon addition of rifampicin, bacterial protein synthesis is halted whereas plasmid protein synthesis continues. The best time for stopping transcription of *E. coli* proteins must be determined so that expression of the recombinant protein is not adversely affected by an non-viable host. To determine the optimum time for rifampicin addition after induction, rifampicin was added at 1 hour and 2 hours post induction, with the cells induced at OD_{600} of 0.4, 0.6, or 1.0 -- early, mid, and late-log phase, respectively. The optimum time for addition of rifampicin was determined to be at 1 hour following induction, with GF expression levels remaining relatively constant between 1-3 hours after rifampicin addition, and showing no preference for an optimum cell density at induction (see Figure 3-3). Further, cell cultures grown at 30°C and 37°C appear to have similar expression levels when rifampicin is added to the culture. Thus, for experimental simplicity, it was decided that for GF expression cells should be incubated at 37°C and induced at an OD_{600} of 0.7 (i.e. mid-log phase), followed by rifampicin addition at one hour post-induction, and cell harvesting 2 hours after rifampicin addition. The addition of rifampicin increased GF yield 2-3 fold (Figure 3-3).

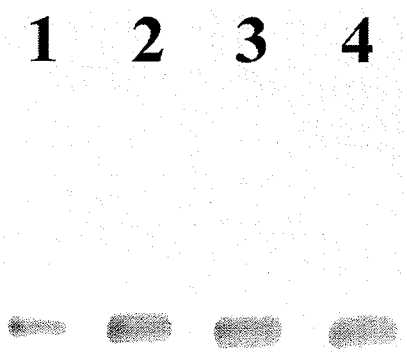


Figure 3-3. Western blot displaying the optimization of time of rifampicin addition (200μg/ml) after induction of expression of GF from the T7 promoter of pET28glpF. All cultures were incubated at 37°C, induced with 1 mM IPTG, and harvested 1 hour after rifampicin addition. Lane 1, induced at OD_{600} = 0.4 with no rifampicin added; lane 2, induced at OD_{600} = 0.6 and rifampicin added 1 hour post-induction; lane 3, induced at OD_{600} = 0.6 and rifampicin added 2 hour post-induction; lane 3, induced at OD_{600} = 1.0 and rifampicin added 1 hour post-induction;

A final refinement to growth and expression conditions was the choice of bacterial media. The medium commonly chosen for bacterial cell growth and protein expression is the Luria-Bertani (LB) medium which contains 5 g/L yeast extract, and 10 g/L peptone and sodium chloride [148b]. However, for growth of *E. coli* destined for plasmid DNA preparations a growth medium designed by Tartoff and Hobbs [159] provided a much higher cell density than LB, and thus increased the plasmid yields. This medium, named Terrific Broth (TB), contains 12 g/L peptone, 24 g/L yeast extract, 4 ml/L glycerol, and 100 mM potassium phosphate (pH 7.4). As observed by Tartoff and Hobbs [159], we find that cells growing in TB medium produce higher yields of GF than those grown in LB suggesting that TB provides higher cell densities than LB.

In summary, for optimum GF expression levels, the *E. coli* cells are incubated at 37°C in TB broth, induced with 1 mM IPTG at an OD_{600} of 0.7, and bacterial protein synthesis is halted at one hour post-induction with the addition of 200μg/ml rifampicin, with the cells harvested two hours after rifampicin addition.

3.3 Sub-cellular Localization of GF

When proteins are overexpressed in bacterial cells, it is common for the cells to store the extra protein as insoluble granules called inclusion bodies [137]. In the overexpression of membrane proteins in *E. coli* inclusion bodies may be produced or extra membranes may be produced by the cell to provide a place for the extra protein to be stored. In *E. coli* this extra membrane might be visible in electron micrographs as invaginations or convolutions of the cell surface [137]. To determine whether overexpressed GF is being stored as inclusion bodies or placed in the membranes, we acquired electron micrographs to search for morphological changes in GF expressing cells. However, Figure 3-4 shows that transmission electron micrographs of uranyl acetate-stained cells over expressing GF display normal cell morphology and organelle content when compared with non-induced controls (Figure 3-4). The electron micrographs provide no evidence for the accumulation of inclusion bodies or for additional membranes in the cells[1] thus suggesting that GF expression is not affecting cell morphology in any way.

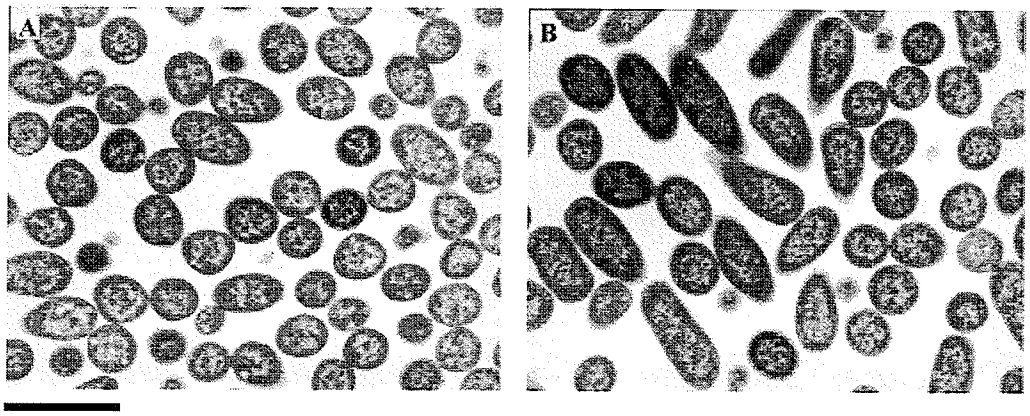


Figure 3-4. Transmission electron micrographs of the *E. coli* BL21(DE3) harboring the pET28*glpF* vector at 3 hours post-induction. Comparison of **A**, non-induced and **B**, induced cells indicates no morphological changes or accumulation of inclusion bodies in response to GF overexpression.

1. For a detailed review see Grisshammer and Tate [137].

Figure 3-5 shows a Coomassie-stained SDS-PAGE electrophoregram of IPTG-induced and non-induced cellular fractions produced as described in Section 2.2.6. No

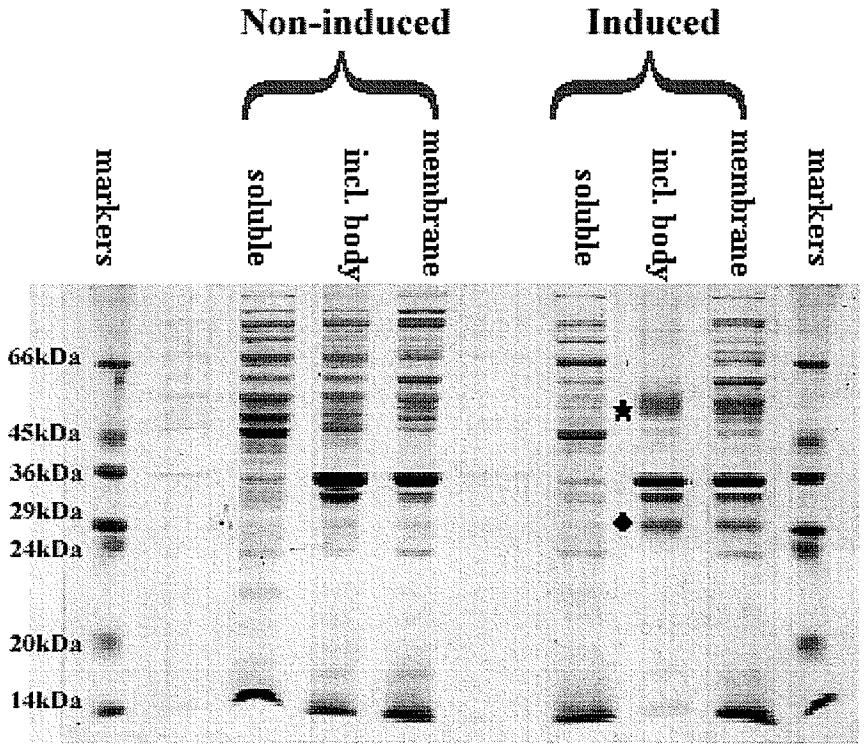


Figure 3-5. Coomasie stained gel (12%) displaying expression of pET28*glpF* in the *E. coli* strain BL21(DE3). Indicated are proteins expressed without (Lanes 2-4) and with (lanes 5-7) IPTG induction. Lanes 1 and 8 are molecular weight markers (66, 45, 36, 29, 24, 20, 14kDa), lanes 2 and 5 are soluble fractions, lanes 3 and 6 are inclusion body fractions and lanes 4 and 7 are membrane fractions. Putative GF is indicated by a diamond (monomer) and a star (dimer).

glycerol facilitator is detectable by Western analysis in the non-induced cells or in the soluble fraction of the induced cells (data not shown). The Coomassie-stained gel shows that IPTG stimulates the production of a 27kDa protein that is approximately equally distributed between the membrane and insoluble inclusion body fractions, results that are corroborated by Western analysis (Figure 3-6). This suggests that some of the protein is found in inclusion bodies as well as membranes. However, electron microscopic

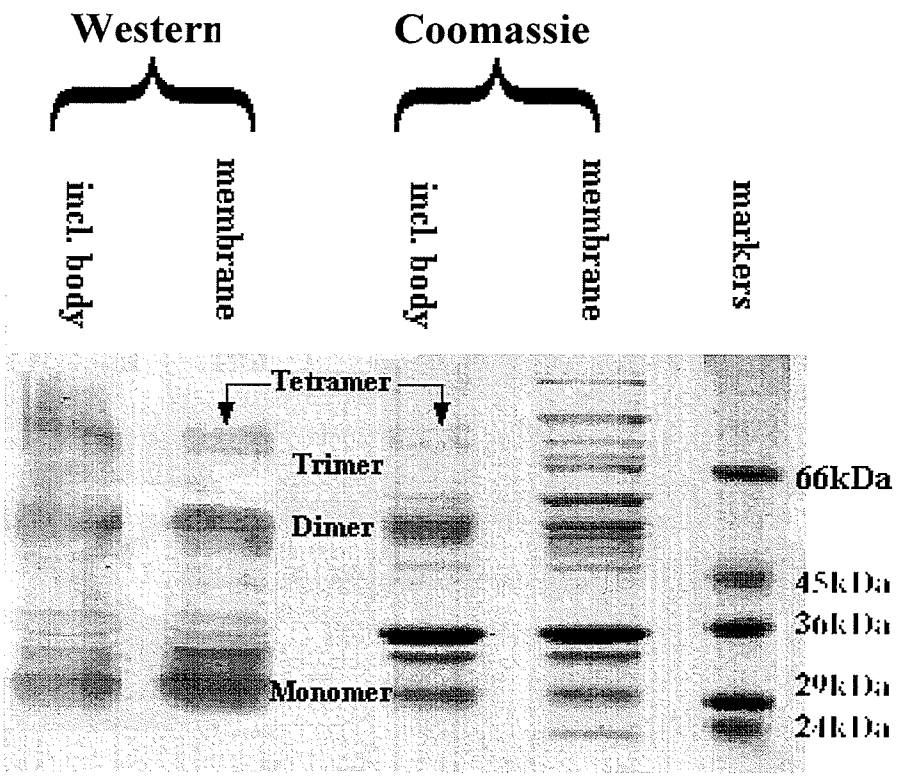


Figure 3-6. Analysis of the inclusion body and membrane fractions from the IPTG-induced expression of *glpF* in BL21(DE3). This suggests that GF is localized in the insoluble cellular fractions, i.e. membranes and possible inclusion bodies

data does not reveal the existence of inclusion bodies in the overexpressing cell. Furthermore, the crude methods used to prepare the inclusion bodies resemble those used in the preparation of *E. coli* membranes for cross-linking studies (see Section 2.2.7). It is likely, therefore, that the immuno-sensitive GF in the lanes labeled "inclusion bodies" may be due to membrane contamination. Thus, for the purposes of protein purification, and to recover as much GF containing material as possible during purification, the GF expressing cells are lysed and the total insoluble debris is collected by centrifugation for 20 minutes at 2,800g in a GSA rotor.

3.4 Detergent Extraction and Purification of GF

3.4.1 Detergent Solubilization of GF.

The detergents sodium dodecyl sulphate (with and without 4 M urea), octyl-β-D-glucopyranoside, N-laurylsarcosine, Triton X-100, Tween 80, dodecyl-β-D-maltoside, sodium deoxycholate, and lauryldimethylamine oxide (LDAO) were tested for their ability to extract GF from insoluble cell fractions. The detergents were chosen based on their availability and their prevalence in published membrane protein purification protocols [176, 177, 178, 179a]. Cells expressing GF are harvested and incubated in detergent solutions as described in Section 2.2.3. The effectiveness of the extractions was determined based on the intensities of Coomassie-stained electrophoregrams. The most effective detergents for GF extraction were dodecyl-β-D-maltoside and LDAO (see Table 3-1). Owing to the cost

Table 3-1. Determination of the Optimum Detergent for GF Extraction

Detergent	Extraction Efficiency[a]
Dodecyl-β-D-maltoside	>70%
Lauryldimethylamine oxide	approximately 70%
Sodium deoxycholate	>50%
N-laurylsarcosine	approximately 50%
Octyl-β-D-glucopyranoside	approximately 50%
Triton X-100	<<50%
Sodium dodecyl sulphate	<<50%
Tween 80	none extracted

a. Based on visual estimates from SDS-PAGE and Western blots comparing the amount of GF extracted to the amount of GF in whole cells.

of the former we continued to develop the purification protocol with LDAO. However LDAO caused some experimental difficulties. The most severe was the aggregation of purified GF into very high molecular weight complexes to such a degree that the majority of the protein remained in the SDS-PAGE sample wells and the stacking/resolving gel inter-

face (see Figure 3-7). In these early experiments, SDS-PAGE sample preparation involved

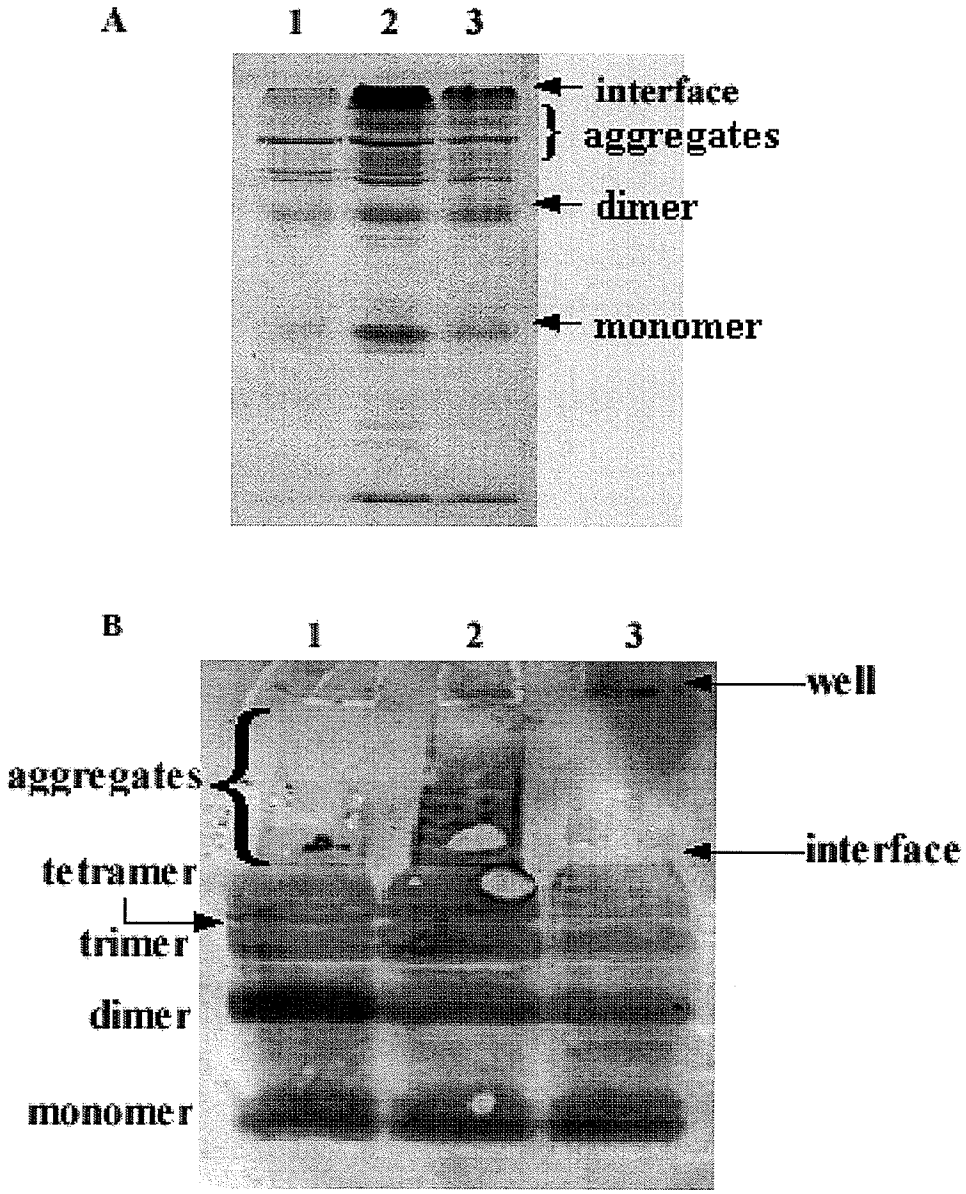


Figure 3-7. Initial purification of LDAO-extracted GF using Ni^{2+}-Nitrilotriacetic acid-Sepharose (i.e. Ni^{2+}-NTA-Sepharose): **A**, Coomassie stained SDS-PAGE and **B**, Western blot analysis of 3 GF fractions. The GF monomer (approximately 29kDa) and dimer (approximately 50-52kDa) are visible in the SDS gel (**A**). Higher molecular weight species are also visible and are probably trimer and tetramer (**B**). Aggregates, too large to enter the gel, are seen at the stacking/resolving gel interface in both **A** and **B**, and in the sample well in **B**. Lanes 1, 2 and 3 indicate successive GF containing fractions as they are eluted from the Ni^{2+}-NTA-Sepharose column.

a 5 minute incubation in boiling water [158]. It was soon discovered that this causes aggre-

gation of detergent-extracted protein [162], which was particularly severe in the LDAO-extracted protein. It is possible that the solubility of the LDAO-GF complex is more temperature sensitive than that for the other detergent extracts. Further, the LDAO gel samples were viscous and contained particulates even before heating, a phenomenon that was not observed with the other detergents.

Due to difficulties encountered with LDAO, the addition of 4 M urea to 150 mM SDS was tested for protein extraction. Although Table 3-1 indicates that the initial tests with SDS proved it to be one of the least effective extraction agents, incubating the insoluble cell fraction in SDS/urea at room temperature for 60-90 minutes did facilitate nearly complete solubilization. During the course of this research similar improved extraction efficiencies were also found for GF if longer incubation periods were used for 150 mM SDS (minus urea) and 30 mM DM all at room temperature (data not shown).

3.4.2 Purification of Detergent Solubilized GF.

The purification of detergent-solubilized GF is accomplished by taking advantage of the interaction between the recombinant N-terminal His_6-tag and the divalent cation Ni^{2+} using the method of immobilized-metal affinity chromatography (IMAC)[179b]. IMAC involves mixing the solubilized His_6-tag containing protein with Ni^{2+}-NTA Sepharose at a pH above the pKa of histidine, i.e. above pH 6. The non-protonated histidines in the His_6-tag chelate with the immobilized Ni^{2+} and allow the resin to be washed with buffer, thereby removing all non-bound proteins and impurities. The NTA resin-bound protein is then eluted from the column by either protonating the His_6-tag by lowering the pH, or by adding a histidine analog such as imidazole to compete for Ni^{2+}. However, it was found that the high concentrations of imidazole (>250 mM) necessary for elution inter-

fered with subsequent analysis of GF by absorption spectroscopy and related techniques. Therefore, a reduction of pH was used to elicit GF elution from the Ni^{2+}-NTA column.

The extraction of GF from the insoluble cellular lysate, with a mixture of 150 mM SDS and 4 M urea, and purification by IMAC using Ni^{2+}-NTA resin is shown in Figure 3-8. IPTG stimulates the production of a protein that migrates with an apparent M_r of approx-

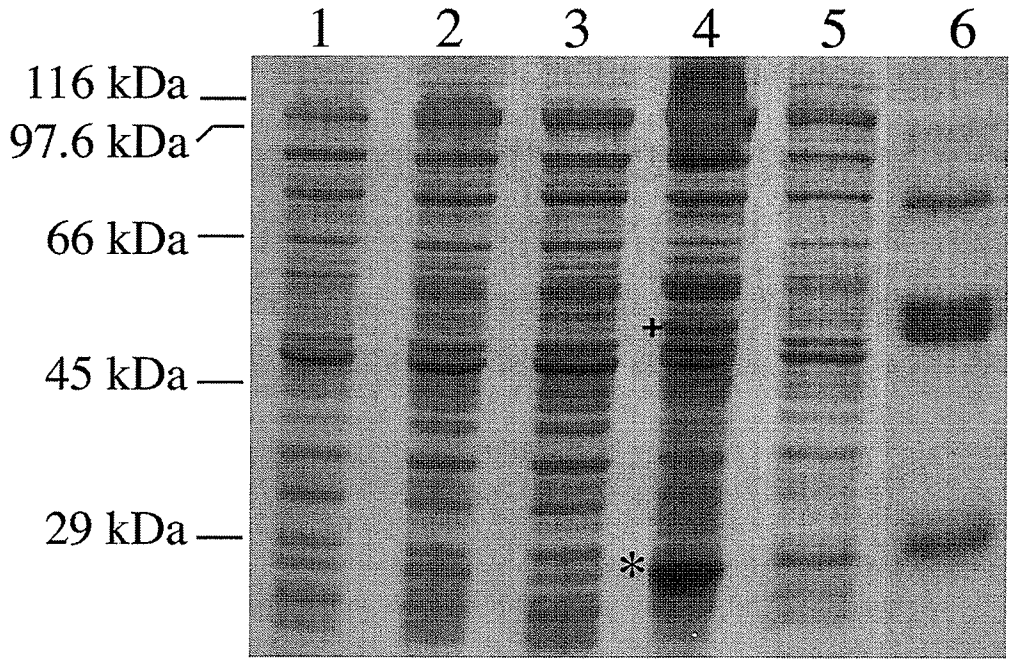


Figure 3-8. Coomassie-stained gel of SDS/urea-extracted GF purified using a Ni^{2+}-nitrilotriacetic acid column. Lane 1, pre-induction cells; Lane 2, post-induction cells; Lane 3, soluble cell lysate; Lane 4, insoluble cell lysate; Lane 5, column flow-through; Lane 6, GF eluted from a Ni^{2+}-nitrilotriacetic acid column by lowering the pH. Note the presence of GF monomer, dimer, trimer and tetramer in Lane 6.

imately 29kDa and is most evident in the insoluble cell lysate dissolved in SDS/urea (marked by an asterisk in Lane 4). Also apparent in both Coomassie-stained gels and Western immunoblots (marked by a plus sign in Lane 4, also see Lane 6) is a protein with an apparent mass of 52 kDa. There is no recombinant protein detectable in the non-induced

cells (Lane 1) or in the soluble fraction of the induced cells (Lane 3), results confirmed by Western immunoblotting (not shown). Lane 6 shows that elution of GF from the Ni^{2+}-NTA resin, upon reducing the pH, yields proteins that electrophorese at positions corresponding to masses of approximately 29 kDa, 52 kDa, 75 kDa, and 100 kDa. All four recombinant molecules are observable in anti-T7 Western immunoblots of the polyacrylamide gels suggesting that they are composed of GF species ranging from monomer to tetramer (see section 3.5). Because other membrane proteins retain their native oligomeric state during SDS-PAGE [180, 181, 182, 183], this apparent oligomerization of GF suggests that the protein may be completely unfolded in SDS. The presence of a doublet band at approximately 52 kDa in GF samples extracted using SDS/urea suggests the existence of two distinct species of GF dimer, one less mobile than the other possibly due to different partially unfolded states of the protein.

GF prepared by extraction of cells with SDS/urea never showed aggregates above the tetramer whereas heat-treated protein formed very high molecular weight aggregates perhaps indicating denaturation. Indeed, because SDS is known to precipitate at temperatures below 15°C [186], GF samples extracted with SDS containing buffers were left on the bench-top at room temperature to limit any temperature induced SDS/protein precipitation. To explore the nature of the oligomers observed in Lane 6 of Figure 3-8 the dithiothreitol concentration was increased in the SDS-PAGE gel sample preparation buffer to reduce any possible disulfide bridges [132]. However, no change in GF oligomerization was observed (data not shown). Further, adding EDTA to the SDS-PAGE sample buffer to chelate any Ni^{2+} that may have leeched from the Ni^{2+}-NTA column, thus causing aggregation of GF

through their His_6 tags and free Ni^{2+} in solution [Hom and Volkman, 1998], had no effect on aggregate distribution (data not shown).

Extraction and purification of GF were also carried out with DM and SDS (without urea), using essentially the same protocol developed for SDS/urea (see section 2.2.4). In both cases the purifications yielded approximately 14 mg of 95% pure GF from 9.5 grams of cells recovered from 2 litres of *E. coli* cell culture. Large scale extraction and purification of GF using octyl glucoside (OG) was not attempted due to the high CMC and cost of OG. Detergent concentrations for extraction of membrane proteins should be above the CMC, so that plenty of micelles are present to solubilize the protein (see section 1.2.1)[2, 13, 137]. With a CMC of 20-25 mM, 100-200 mM OG is required to effectively solubilize GF, but at more than $80 per gram this would not be cost effective. Thus to prepare facilitator in OG we chose to initially extract and purify GF using DM, then exchange DM for OG while GF remained bound to the Ni^{2+}-NTA column (see section 2.2.4). Figure 3-9 compares the electrophoretic mobilities of purified recombinant GF dissolved in SDS/urea, SDS, DM, and OG. Predominantly monomeric protein is observed in each of the three detergents (Lanes 2-4) except when the protein is extracted with SDS/urea (Lane 1) where monomer, dimer, trimer, and tetramer can been seen. However, Western immunoblots reveal small amounts of oligomeric recombinant protein in all of these preparations (not shown). All attempts to raise the pH of the OG preparation from the elution pH of 4.0 to a pH of 7.6 resulted in the precipitation of the protein, producing visible oligomers ranging from monomer to heptamer and high molecular weight species stuck at the stacking/resolving gel interface (see Figure 3-10).

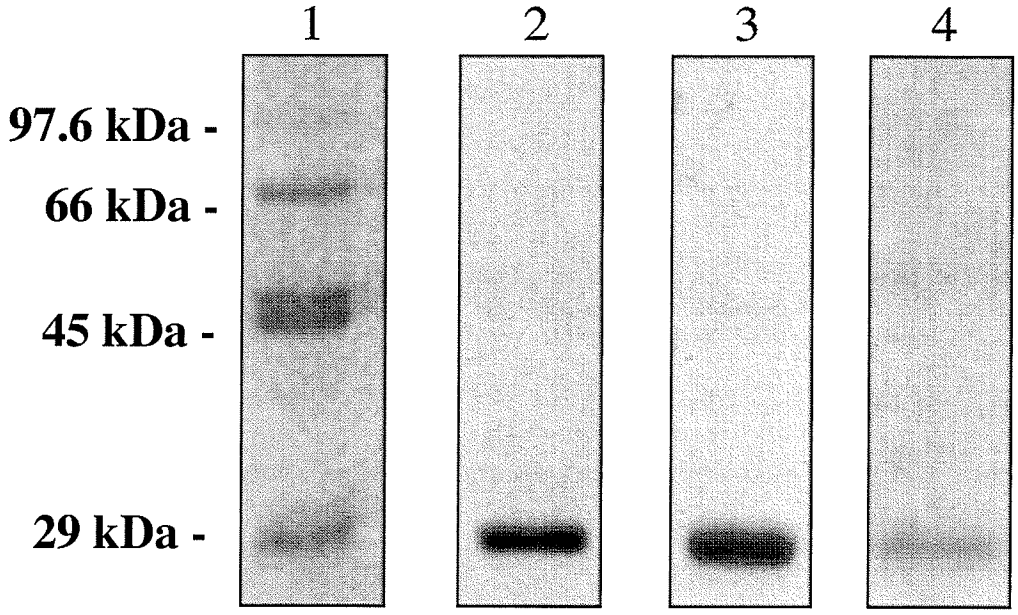


Figure 3-9. Coomassie-stained SDS-PAGE gel of GF extracted with: lane 1, SDS/urea; lane 2, SDS; lane 3, DM; lane 4, OG. Monomer, dimer, trimer, and tetramer are visible in the SDS/urea sample (all samples were purified using the protocol outlined in section 2.2.4).

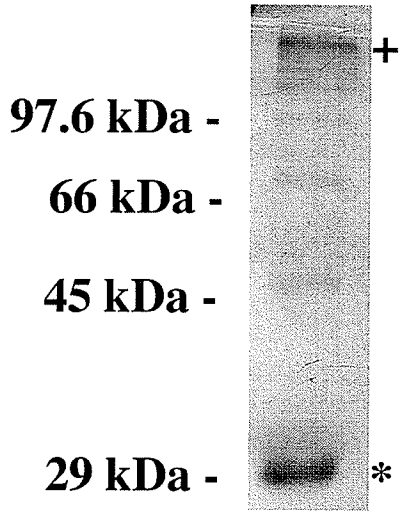


Figure 3-10. Coomassie-stained SDS-PAGE gel of pH induced GF aggregation in OG. Note the appearance of monomer (marked by an asterisk), dimer, trimer, tetramer, pentamer, hexamer, heptamer, and high molecular weight aggregates at the stacking/ resolving gel interface (marked by a plus sign).

Since pH-induced precipitation could not be reproduced with GF purified in DM, SDS, or SDS/urea, this suggests that OG-solubilized GF is only marginally stable.

3.5 Biophysical Characterization of GF

3.5.1 Oligomeric State of GF *in vivo*.

There is considerable doubt whether the native state of GF in a membrane is a tetramer similar to AQP1 [87] or a monomer as suggested by Lagree *et al.* [185]. To probe the oligomeric state of native membrane-embedded GF, chemical cross-linking was employed using disuccinimidyl suberate (DSS), disuccinimidyl glutarate (DSG), and bis(sulfosuccinimidyl) suberate (BS). DSS and DSG are lipophilic primary diamines that can penetrate the cell membrane and react with the amino groups of lysines to form a covalent cross-link. Similarly, BS is a hydrophilic primary diamine that reacts with the solvent-exposed amino groups of lysines and forms covalent cross-links. Western immunoblots of cell membrane extracts, in the absence of cross-linker, suggest that GF exists in the membrane predominantly as a monomer (see Figure 3-11, Lane 1). However, GF is evidently

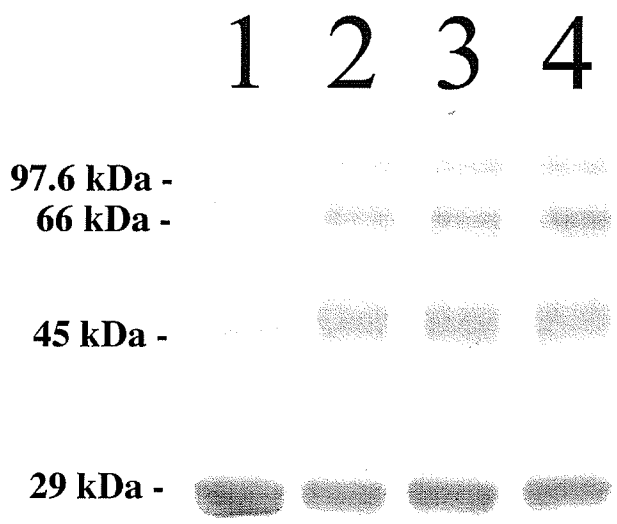


Figure 3-11. Western immunoblot of DSS cross-linked membranes from GF expressing cells. Lane 1, non-cross-linked membranes; lane 2, membranes cross-linked for 30 minutes; lane 3, membranes cross-linked for 1 hour; lane 4, membranes cross-linked for 2 hours.

susceptible to oligomerization *in vivo* as the proportion of high M_r protein present in mem-

brane increases in the presence of cross-linker as compared to the non-cross-linked control (see Figure 3-11, Lanes 2-4). Furthermore, the species produced by cross-linking are identical in electrophoretic mobility to the species present in the purified, concentrated protein (see Figure 3-9, Lane 1). Only results for DSS are shown in Figure 3-11, however similar results were observed with the cross-linkers DSG and BS (data not shown). Cross-linking with DSS for extended periods of time does not produce any observable species higher in molecular weight than the apparent tetramer (data not shown). The absence of cross-linked GF aggregates larger in mass than the apparent tetramer, even after cross-linking for 24 hours, and the low background in the Western blots, suggests that the cross-linking of GF is specific. These data therefore indicate that GF is capable of forming tetramers in the membrane.

3.5.2 Oligomeric State of GF *in vitro*.

Structural analysis by NMR spectroscopy and X-ray diffraction requires highly purified, homogeneous protein preparations. In the previous section it was shown that GF can be cross-linked to form homo-oligomers in its native membrane state. In Figures 3-8 (lane 6) and 3-10 it is apparent that extraction and purification of GF with an SDS/urea buffer promote the formation of GF oligomers ranging from monomer to tetramer, and that pH-adjusted GF solubilized in OG rapidly aggregates into high molecular weight species. To explore the possible role of urea in oligomerization, GF purified individually in SDS and DM was incubated with 2-6.5 M urea. Figure 3-12 shows that in the SDS-purified GF samples the addition of urea did not affect the oligomeric state, as a small amount of dimer

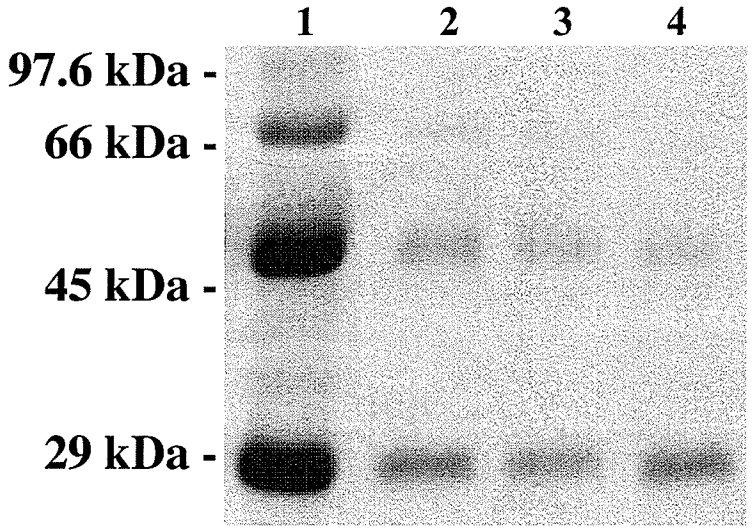


Figure 3-12. Coomassie-stained SDS-PAGE showing the effects of incubating SDS-purified GF in urea. Lane 1, 0 M urea; lane 2, 2 M urea; lane 3, 4 M urea; lane 4, 6 M urea. GF monomer, dimer, trimer and tetramer are in the concentrated SDS sample in lane 1 (see text for details). Dimer and trimer are visible in all samples containing urea with no shift in oligomeric state noticeable with increasing urea concentrations.

and trimer are present at all concentrations of urea. In contrast, Figure 3-13 shows that DM-purified GF is predominantly monomeric (lane 1) and the addition of urea resulted in the formation of dimer, trimer, and tetramer at urea concentrations above 2 M (lanes 2-4).

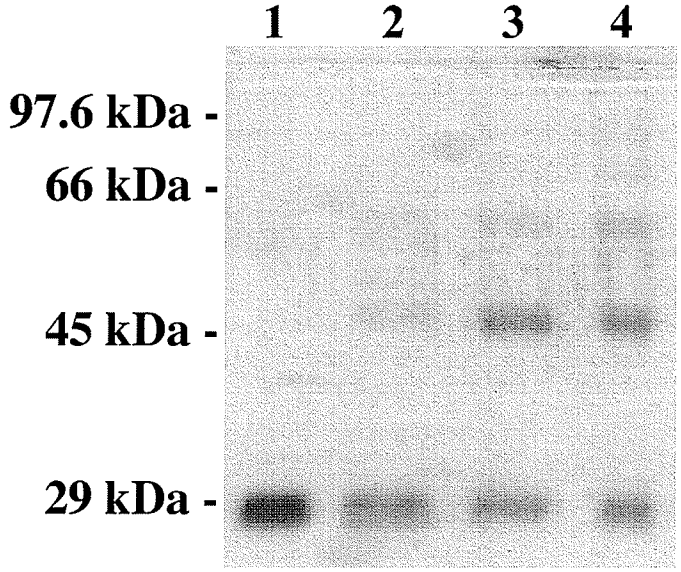


Figure 3-13. Coomassie-stained SDS-PAGE showing urea-induced aggregation of DM-purified GF. Lane 1, 0 M urea; lane 2, 2 M urea; lane 3, 4 M urea; lane 4, 6.5 M urea. GF dimer is visible in all samples containing urea; dimer and trimer are visible at higher urea concentrations in lanes 3 and 4. High molecular weight aggregates appear at the stacking/resolving gel interface in the 6.5 M urea sample.

Indeed, traces of dimeric GF appear when 2 M urea is present (see Figure 3-13, lane 2). At 6.5 M urea, dimer, trimer, and tetramer are visible in the Coomassie-stained polyacrylamide gels (see Figure 3-13, lane 4), suggesting that urea is promoting oligomerization of the DM purified protein.

The oligomeric state of GF also displays a protein-concentration dependence in both SDS and DM purified samples. The electrophoregram in Figure 3-14 A shows successive fractions of SDS-solubilized GF eluted from a Ni^{2+}-NTA column. GF monomer, dimer, trimer, and tetramer are observable in Coomassie-stained gels in the most concentrated solutions (greater than 28 μM, lanes 2-4; also Figure 3-12, lane 1), whereas in more dilute solutions (less than 28 μM) only monomer is detected (Figure 3-14 A, lanes 1 and 5). This effect of protein concentration on the oligomeric state of GF is more pronounced in DM solubilized samples (Figure 3-14 B) where dimer is observed at concentrations much lower than in the SDS solubilized protein. Thus, GF dimer is apparent at protein concentrations above 7 μM in DM (lanes 3-7 in Figure 3-14 B), whereas in SDS dimer is barely visible at GF concentrations of 20-30 μM (lanes 1and 4 in Figure 3-14 A). This suggests that, unlike SDS, DM is less capable of disaggregating/denaturing GF and keeping it in a monomeric state at high protein concentration. This likely reflects the more aggressive, strongly denaturing, nature of SDS in the solubilization of proteins [186]. However, even at a high concentration of SDS (150 mM) a significant amount of GF remains in an oligomeric state at high protein concentrations (Figure 3-12, lane 1 and Figure 3-14 A, lanes 2-4), possibly in a partially denatured state. Thus the data suggest a protein concentration dependence on the oligomerization of GF.

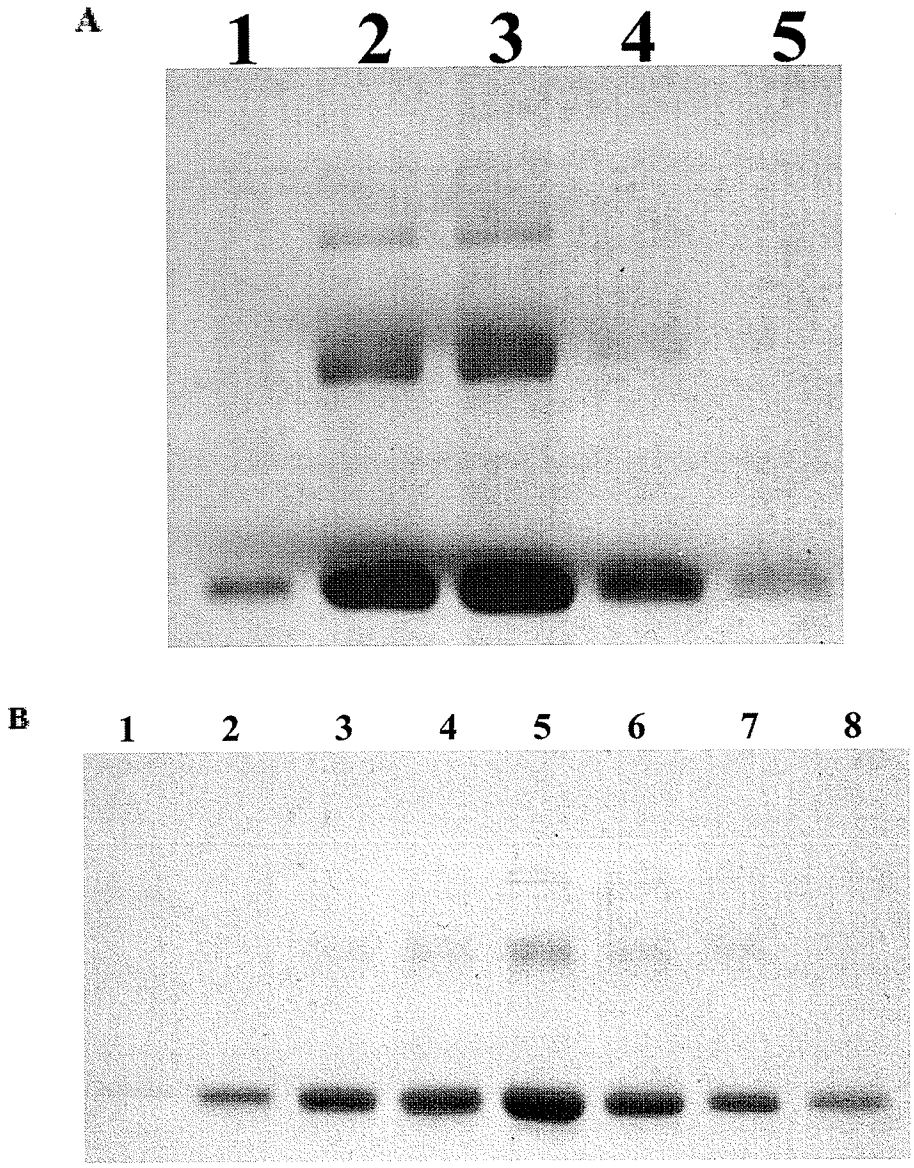


Figure 3-14. Protein concentration dependence of GF oligomer formation. **A**, SDS solubilized GF: lane 1, 21 μM GF; lane 2, 71 μM GF; lane 3, 91.2 μM GF; lane 4, 28.7 μM GF; lane 5, 8.95 μM GF. SDS-purified GF is monomeric at low protein concentrations (below approximately 21 μM GF, lanes 1 and 5), whereas dimer, trimer and tetramer are evident at higher concentrations (lanes 2-4). **B**, DM solubilized GF: lane 1, 1.1 μM GF; lane 2, 6.3 μM GF; lane 3, 15 μM GF; lane 4, 19 μM GF; lane 5, 33 μM GF; lane 6, 15 μM GF; lane 7, 7 μM GF; lane 8, 4.9 μM GF. DM-purified facilitator is monomeric at protein concentrations below 6.3 μM (lanes 1, 2 and 8), whereas dimer is evident at higher concentrations (lanes 3-7).

All the SDS-PAGE samples were prepared by heating at 37°C for 10 minutes, rather than following the standard protocol of boiling them for 5 minutes [148b, 158]. This was done because boiling the samples results in aggregation of GF to such a degree that a large portion is retained at the interface between the stacking and resolving gels and in the sample well itself (see Figure 3-15 A and B). This effect of heat-induced aggregation of GF

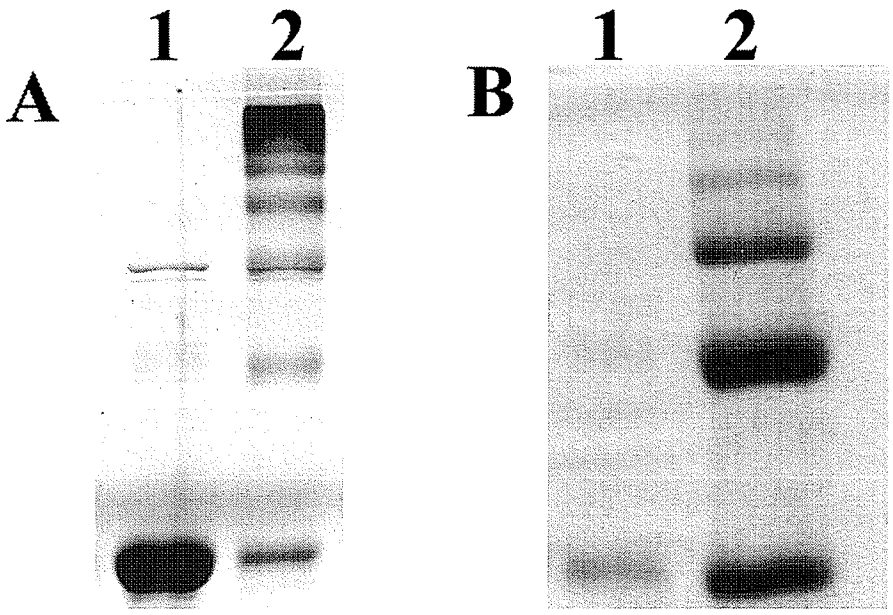


Figure 3-15. Temperature induced aggregation of GF. **A**, DM solubilized GF (77 μM): lane 1, prepared by heating at 37°C for 10 minutes; lane 2, incubated at 80°C for 1 hour, before preparing sample as in lane 1. **B**, SDS solubilized GF, with samples in lanes 1 (15 μM) and 2 (91.2 μM) prepared in the same manner as in **A**.

is a problem commonly associated with boiling membrane proteins. Boiling causes membrane proteins to denature and aggregate via the exposed hydrophobic segments [187]. The results shown in Figure 3-15 are a further illustration that SDS is more capable than DM of disaggregating GF and keeping it in a lower oligomeric state.

3.5.3 Activity of GF in Membranes

The transport properties of the GF have been studied extensively over the last 30 years (see section 1.4.4) [94-96, 106, 107, 110, 188, 189]. Heller *et al.* [106] reported that GF will equilibrate a glycerol gradient across the cytoplasmic membrane of *E. coli* within 30 seconds in uninduced cells (i.e. not expressing GF) and in less than 10 seconds in cells expressing GF, such that equilibrium is reached within the time required for manual mixing and obtaining the first reading. A half-time of equilibration of less than 10 seconds is much too fast to be followed by standard spectrometric methodologies, thus stopped-flow techniques have been developed to measure glycerol transport rates indirectly [16]. Glycerol transport induces changes in light scattering in response to changes in cellular dimensions [107, 188]. It is also known that GF facilitates the transport of other low molecular weight polyols, such as xylitol, but at a much slower rate than glycerol. Heller *et al.* [106] reported that xylitol transport is approximately 7 times slower than glycerol transport, showing practically no entry in 5 minutes in uninduced cells and a half-time of equilibration of 40-45 seconds in induced cells. This transport is slow enough to measure using standard spectrometric methodologies. To determine if recombinant facilitator is functional we measured xylitol transport in induced and noninduced cells (see Figure 3-16). Figure 3-16 illustrates the xylitol transport activity of *E. coli* cells expressing GF and non-induced controls. Both the induced and noninduced cells rapidly plasmolyse (shrink) when placed in a hypertonic xylitol environment due to the sudden increase in external osmotic pressure. This cellular plasmolysis is observed as a rapid increase in OD_{600} as soon as the uninduced and induced cells are placed in the xylitol-containing buffer (▲→◆ and ■→●, respectively). Cells induced to express GF (●) have the ability to transport xylitol across the membrane and into the cell, and can therefore deplasmolyse (re-swell).

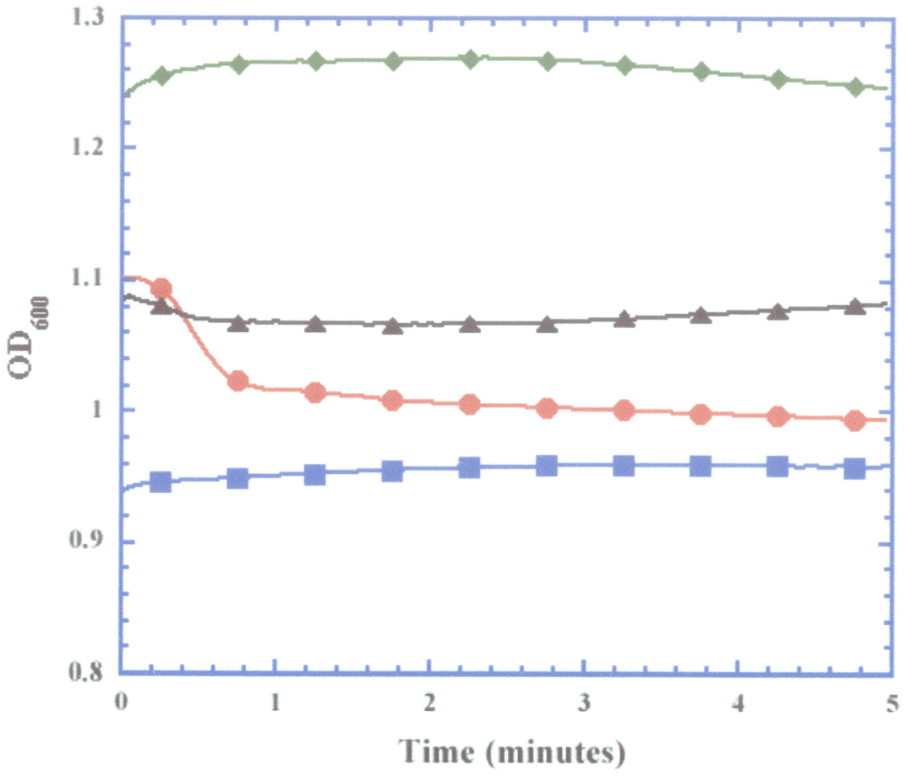


Figure 3-16. Optical changes associated with osmotically induced plasmolysis and deplasmolysis of *E. coli* cells. Both induced (●) and noninduced (◆) cells plasmolyse rapidly when placed in xylitol-containing buffer. This is observed as an increase in OD_{600} at t=0 (■→● and ▲→◆). Cells diluted in xylitol-free buffer do not shrink (induced (■) and noninduced (▲) cells, respectively). Only those cells induced to express GF, and therefore transport xylitol, show re-swelling, observable as a decrease in OD_{600} (●).

The deplasmolysis is observed as a return of the OD_{600} (●) to nearly that of the induced cells in isotonic buffer, i.e. buffer without xylitol (■). However, cells not induced to express GF show no indications of deplasmolysis, and remain in a plasmolysed state (◆). The half-time of equilibration of xylitol across the membrane of GF expressing cells is 30 - 40 seconds, similar to the results of Heller *et al.* [106] who reported a half-time of 42 seconds. Also, the addition of mercuric ions (Hg^{2+}) to GF-expressing cells, prior to the addition of xylitol, blocks xylitol transport (data not shown). Inhibition of GF by Hg^{2+} ions was reported by Maurel *et al.* [107]. Thus, the transport data acquired from GF expressing *E. coli* cells, indicate the presence of a membrane channel with the transport properties that

have previously been reported for GF [106, 107]. Further, this shows that the His_6-tag and T7-epitope that were added to the N-terminal end of GF do not affect the measured transport properties of GF.

3.5.4 Mass Spectrometric Analysis of Detergent-Extracted GF.

As discussed in section 1.2.1, detergents are necessary for membrane protein extraction and solubilization. However, mass spectral analysis of membrane proteins is hampered by the necessary presence of detergent which degrades spectral quality by decreasing signal-to-noise (S/N) ratio and mass resolution [190], due to formation of detergent/protein adducts. Further, in the absence of detergents membrane protein mass spectra are degraded due to protein self-aggregation and consequent loss of sample for desorption [191]. Progress in this area has been reported. For example, MALDI-MS analysis of membrane proteins solubilized in ionic detergents at concentrations well above their CMCs was reported with good S/N ratios and mass resolution [191, 192]. The authors suggest that improved solubility of the membrane proteins at increased detergent concentrations provides an increased level of protein at the surface layers of matrix crystals available for desorption [192]. We attempted but could not reproduce these results with GF solubilized in any concentration of SDS. We therefore turned to the polyurethane sample support MALDI-MS sample preparation protocol developed in the Department of Chemistry (University of Manitoba) for mass spectral analysis of GF [164]. Following the suggestion of Rosinke *et al.* [191] SDS was exchanged for OG by on-column detergent exchange to pro-

vide increased S/N ratios and mass resolution. Figure 3-17 shows the MALDI-TOF mass

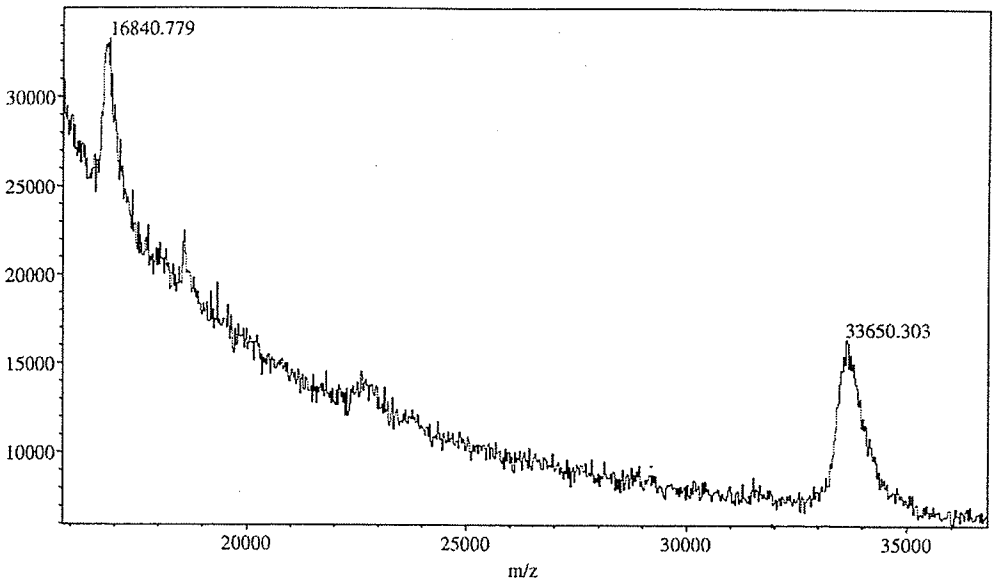


Figure 3-17. MALDI-TOF mass spectrum of detergent-solubilized GF. GF was solubilized in 70 mM octyl-β-D-glucopyranoside, deposited onto a polyurethane membrane, and washed following the protocol discussed in section 2.2.6. $[M+H\}^+$ = 33 650.303 Da (0.4% error); the theoretical molecular mass = 33 505 Da. $[M+2H]^{2+}$ = 16 840.779 (0.5% error); the theoretical molecular mass = 16 752.5 Da.

spectrum of GF obtained from a solution containing 70 mM octyl-β-D-glucopyranoside-solubilized GF. A peak corresponding to a singly-charged, $[M+H]^+$ GF monomer is present with a mass-to-charge ratio (m/z) of 33,650.3. A peak corresponding to a doubly-charged molecule, $[M+2H]^{2+}$, is observed with an m/z of 16,840.8 Da, giving a monomer mass of 33,681.6 Da. The observed masses compare well to the theoretical mass of recombinant GF which is 33,505 Da. The $[M+H]^+$ mass is within approximately 0.4% and the $[M+2H]^{2+}$ within approximately 0.5% of the theoretical mass of the protein.

No oligomeric GF species were observed in the MALDI mass spectra possibly due to precipitation of high molecular weight species during the MALDI-TOFMS sample preparation. Alternatively, the result indicates that no GF oligomers were present in the sample.

Based on the high tendency of the OG-solubilized protein to oligomerize described in section 3.4.2, the former cause seems more probable. Thus, the apparent lack of GF oligomeric species in the MALDI-TOF mass spectrum may be due to sample precipitation reducing the quantity that co-crystallizes with the MALDI matrix. This would result in a loss of analyte available for desorption and consequently a loss of signal [191, 192]. A final possibility is that GF oligomers disaggregate in the mass spectrometer, since non-covalent interactions are not always maintained in MS [193].

3.6 Circular Dichroism Analysis of GF

The development of the overexpression protocol for *E. coli* GF, and the yields and purity attained have made it is feasible to begin structural studies of GF. Whilst NMR and crystallographic techniques can provide structural data at atomic resolution, circular dichroism can be utilized to provide a "global" analysis of protein structural properties such as the overall secondary structure and the tertiary structural characteristics [121, 122]. With the availability of CD deconvolution programs, and powerful desktop computers, the deconvolution of secondary structural spectra into their component spectra has become routine. To this end, far-UV CD analysis was performed to provide information about the secondary structural organization of GF solubilized in 30 mM DM, 100 mM OG, 150 mM SDS, and 150 mM SDS with 4 M urea, with subsequent deconvolution of the spectra into their individual structural components (see section 3.5.1). Near-UV CD analysis was also performed on GF solubilized in both 30 mM DM and 150 mM SDS, to extract data on the tertiary structure of GF in detergent (see section 3.5.2). Further, temperature denaturation of GF solubilized in both 30 mM DM and 150 mM SDS, as monitored by near- and far-UV

CD, was done to investigate the stability of the secondary and tertiary structure of GF (see section 3.5.3)

3.6.1 Determination of the Secondary Structure of GF by Far-UV CD.

CD spectra of purified GF were acquired into the far UV region utilizing the CD optimization parameters of Hennessey and Johnson [170] (see section 2.2.9). Figure 3-18 A shows the spectrum of SDS/urea extracted GF dissolved in SDS at pH 7.6. Figures 3-18 B and C show the spectra of GF dissolved in SDS at pH 4.2 and 7.5, respectively. Figure 3-18 D, E and F shows the spectra of GF dissolved in OG at pH 4.0 and DM at pH 4.4 and 7.0. A spectrum of OG at an elevated pH could not be acquired because any attempt to adjust the pH resulted in instant precipitation of the protein (see section 3.4.2). The CD spectra of GF in all detergents display the general characteristics of a predominantly α-helical protein with negative bands at 218-222 nm and 209-212 nm, positive bands at 193-194 nm, and cross-over points at 201-203 nm (see Figure 3-18 A-F). The ideal α-helical values are 222, 208, 190-195 and 205 nm respectively [121]. Indeed, comparison of these spectra to those of the α-helical form of polylysine provides a visual confirmation of the high helical nature of GF in the detergents used (see Figure 1-38).

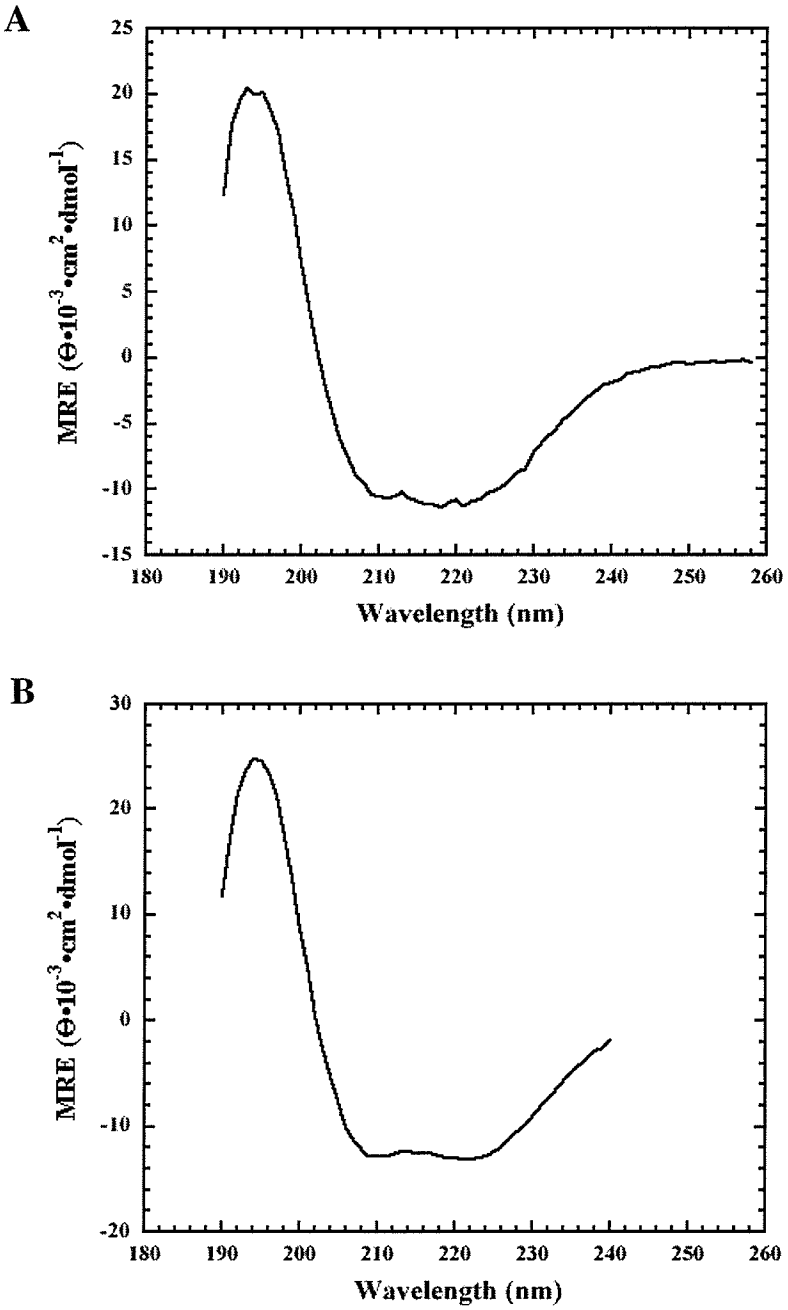


Figure 3-18. CD spectra of GF solubilized in 10 mM sodium phosphate buffer containing 10 mM sodium chloride and detergent at 25°C. **A**, SDS/urea-extracted GF dissolved in SDS at pH 7.6; **B**, SDS-extracted GF dissolved in SDS at pH 4.2.

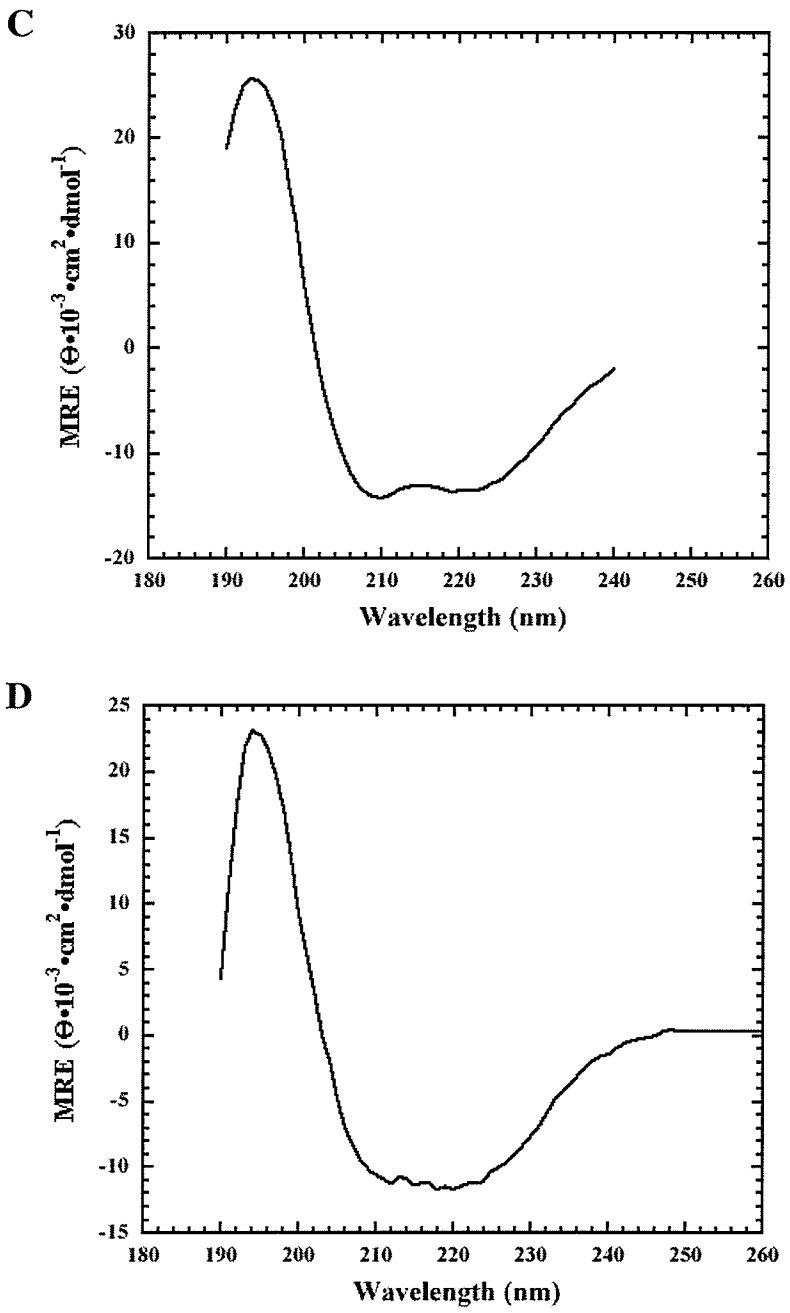


Figure 3-18 (continued). C, SDS-extracted GF dissolved in SDS at pH 7.5; **D**, DM-extracted GF dissolved in OG at pH 4.0

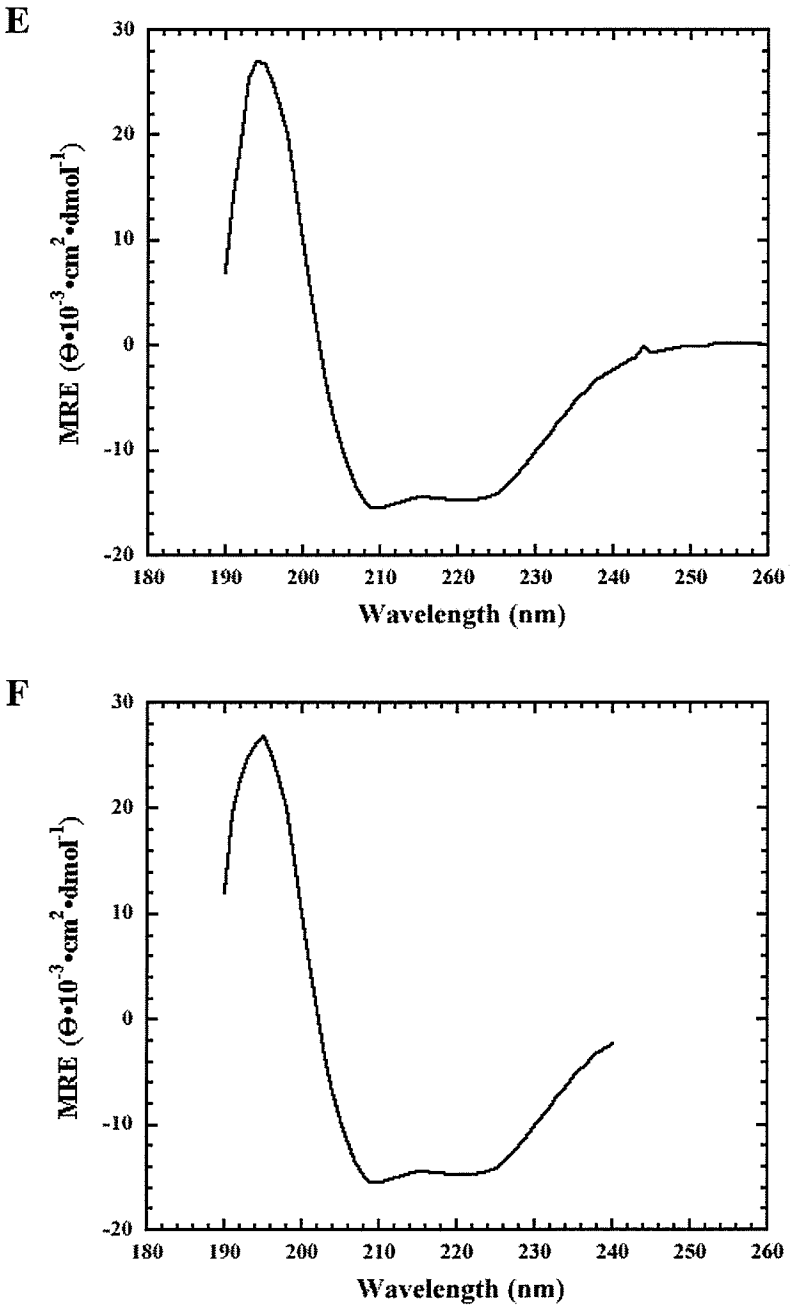


Figure 3-18 (continued). **E**, DM-extracted GF dissolved in DM at pH 4.4; **F**, DM-extracted GF dissolved in DM at pH 7.0

3.6.1.1 Deconvolution of GF Secondary Structure by Convex Constraint Analysis

As illustrated in section 1.5.3.1, the far-UV CD spectrum of a protein is a weighted sum of all the 2° structural elements present in the sample (see equations 1-29, 1-30, and 1-31). One method of determining the weights of these secondary structural elements is to select a set of weights (w_i) of the pure component spectra that minimizes the difference between the measured spectrum ($f(\lambda)$) and the spectrum that results from the weighted sum. However, this method has fallen into disfavor due to inconsistencies between the pure component CD spectra of model compounds and those found in real proteins (see Section 1.5.3.1). An alternative approach to the determination of protein secondary structure from CD spectra uses a basis set of pure components derived from the CD spectra of proteins with known three-dimensional structures [121]. A variety of programs are available that can perform this type of analysis including Singular Value Decomposition [170], Ridge Regression [194], and Convex Constraint Analysis (CCA) [129]. In this thesis CCA was chosen because a basis set of CD spectra of membrane proteins was available for use with it.

The GF CD spectra in all detergents were deconvoluted into their pure structural components (α-helix, parallel and anti-parallel β-sheet, β-turn, and unordered) with CCA using two different basis sets. The first basis set was composed of the CD spectra of 30 water-soluble globular proteins including myoglobin, α-lactalbumin, lysozyme, papain, insulin, and subtilisin [131]. The second basis set contained the CD spectra of 24 membrane proteins including the photosynthetic reaction centers from *Rhodobacter sphaeroides* and *Rhodopseudomonas viridis*, cytochrome C oxidase, and bacteriorhodopsin from *Halobacterium halobium* [129]. The membrane protein basis set also contained the CD spectra of the porin protein OmpF and the porin from *Rb. capsulatu,* two proteins that have a high

content of β-sheet and unordered structure and very little α-helix [131]. As mentioned in section 1.5.3.1, CCA can only fit a protein spectrum to the basis spectra contained in the proteins in the basis set. Conversely, proteins with a large fraction of one secondary structure component can skew the deconvolution [131]. Thus, the presence of the porin proteins in the membrane protein basis set yielded unreasonably large β-sheet components for detergent solubilized GF. The porin data were therefore removed from the basis set in the deconvolution reported in the following Tables and Figures.

The results of the CCA deconvolution using the soluble protein basis set are shown in Figures 3-19A to 3-24A, and the CCA derived pure components are listed in Table 3-2. Also shown are the weighted sum spectra, calculated using equation 1-29 (dashed line with ◆), and the original GF CD spectra (solid line with ▲; see Figures 3-19B to 3-24B). It is apparent that the weighted sums of the pure components reconstruct the experimental spectra quite well in all cases. The deconvolution shows that in all the protein preparations, regardless of detergent and pH, the α-helix is the largest secondary structural component present. The α-helix content is highest in DM and SDS (48-55%), lower in OG (42%), and is lowest is SDS/urea-extracted protein (32%).

CCA deconvolution using the membrane protein data set, with 2-5 pure component deconvolutions, proved to be unsatisfactory. The two criteria for determining the

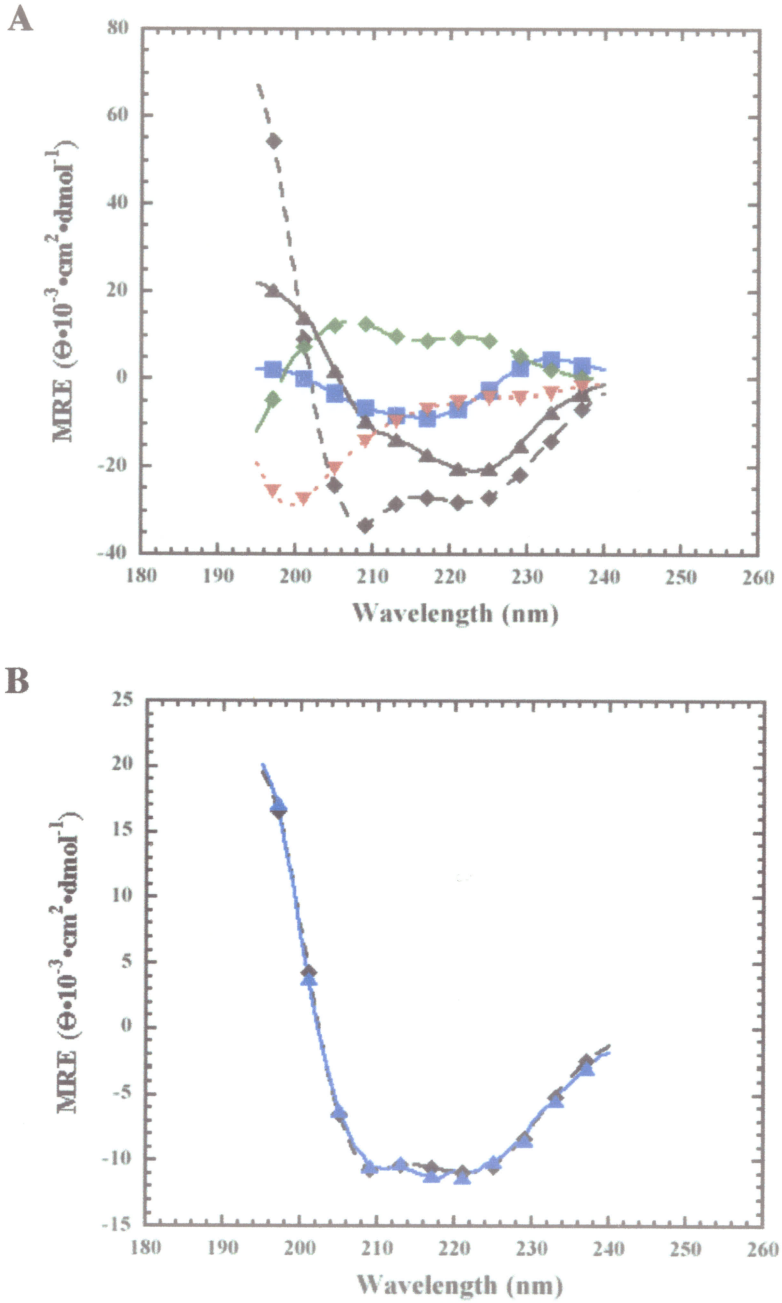


Figure 3-19. CCA deconvolution of SDS/urea extracted GF solubilized in SDS at pH 7.6. **A**, The pure component curves derived from a 5-component deconvolution. The component weights are 32% α-helix (◆), 30% β-sheet (◆), 18% β-turn (▲), 12% unordered (▼), and 8% aromatic (■). **B**, The fitted spectrum, a weighted sum of the pure components (dashed line with ◆), and original CD spectrum (solid line with ▲).

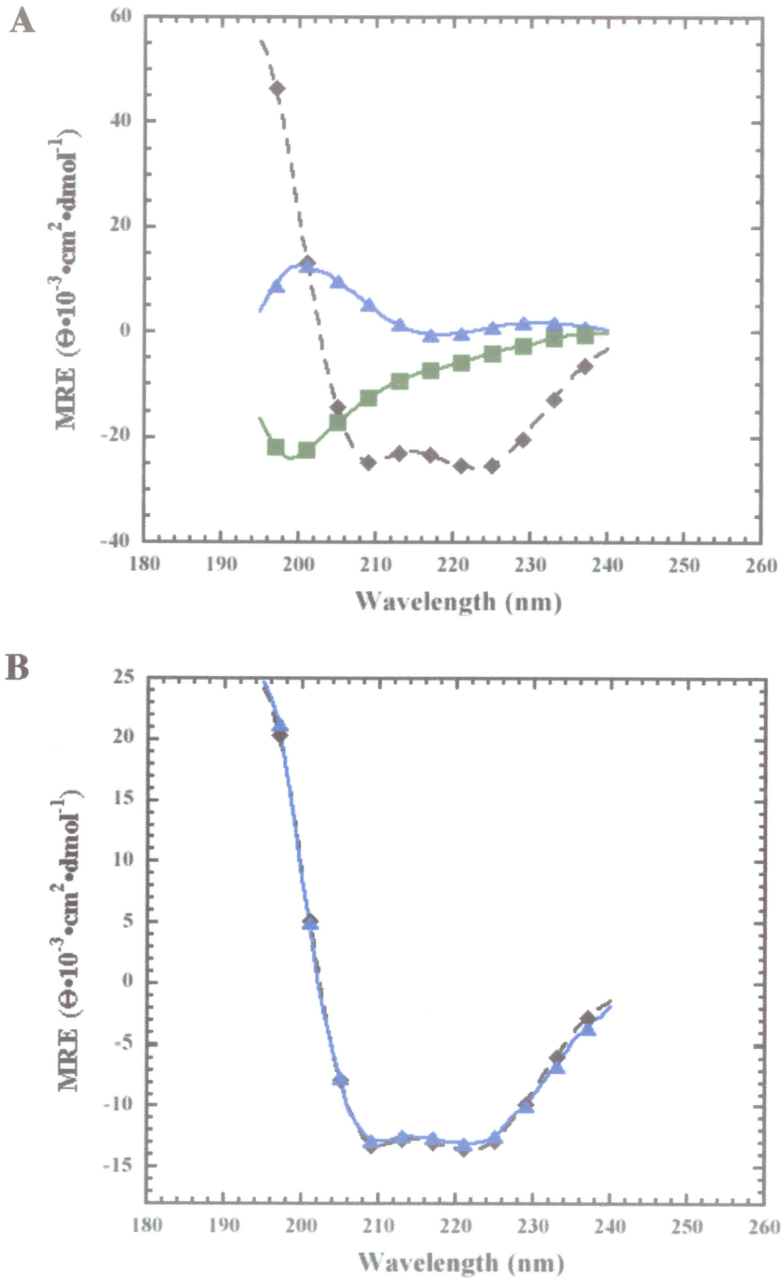


Figure 3-20. CCA deconvolution of SDS extracted GF solubilized in SDS at pH 4.2. **A**, The pure component curves derived from a 3-component deconvolution. The component weights are 48% α-helix (◆), 30% β-sheet (▲), and 22% unordered (■). **B**, The fitted spectrum, a weighted sum of the pure components (dashed line with ◆), and original CD spectrum (solid line with ▲).

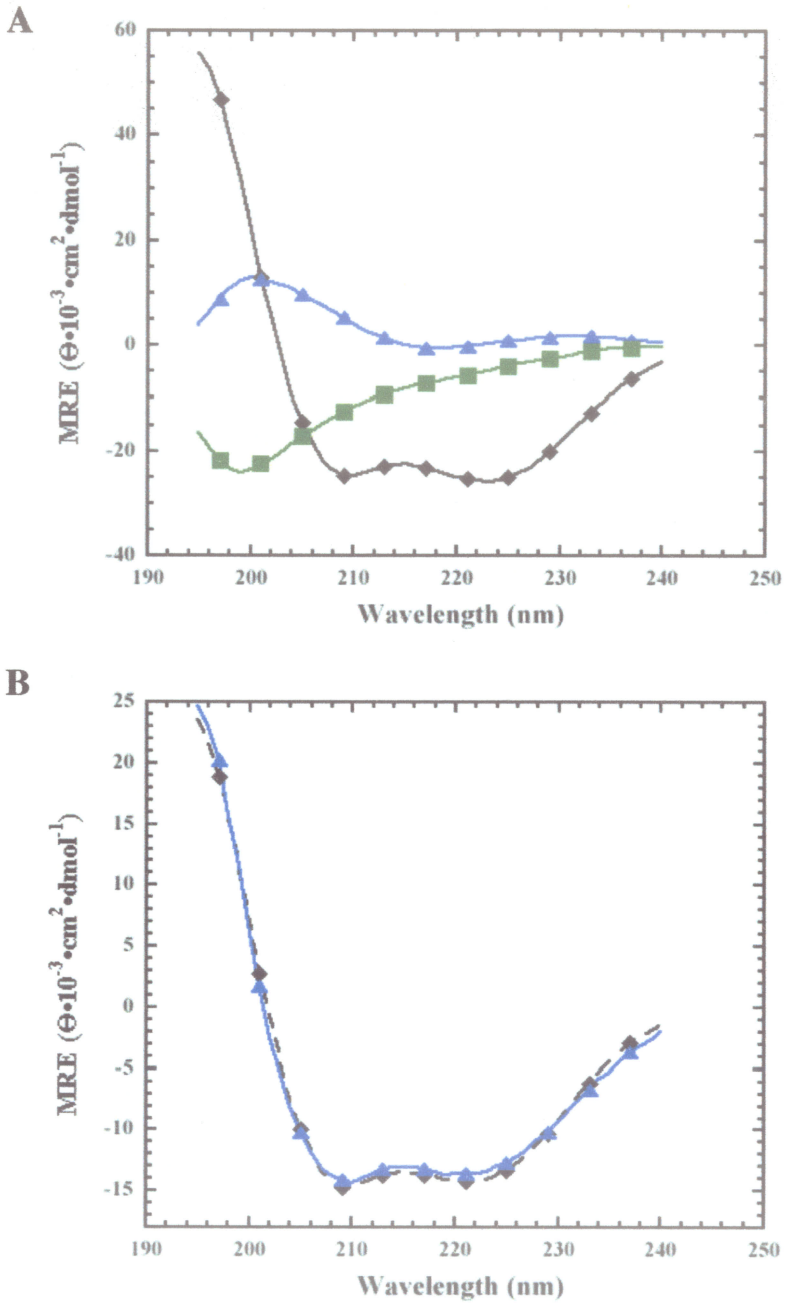


Figure 3-21. CCA deconvolution of SDS extracted GF solubilized in SDS at pH 7.5. **A**, The pure component curves derived from a 3-component deconvolution. The component weights are 49% α-helix (◆), 22% β-sheet (▲), and 29% unordered (■). **B**, The fitted spectrum, a weighted sum of the pure components (dashed line with ◆), and original CD spectrum (solid line with ▲).

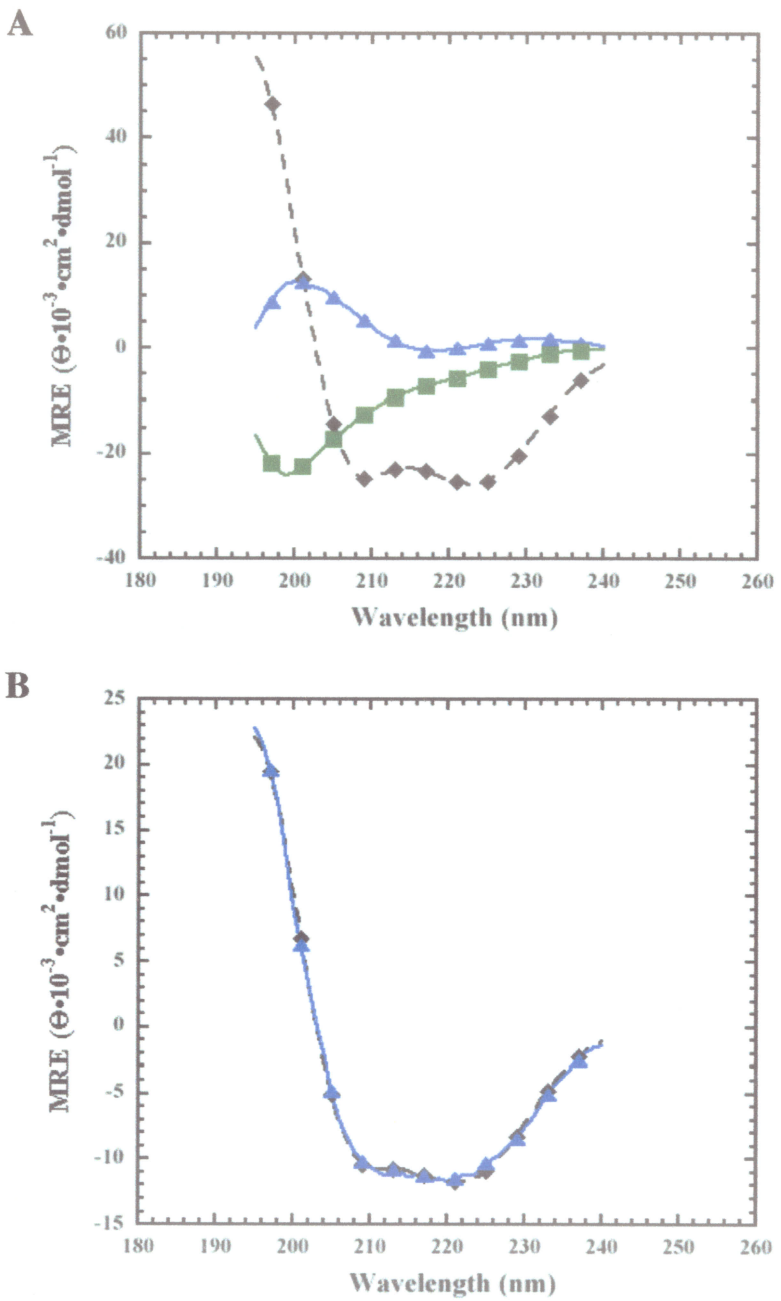


Figure 3-22. CCA deconvolution of DM extracted GF solubilized in OG at pH 4.0. **A**, The pure component curves derived from a 3-component deconvolution. The component weights are 42% α-helix (◆), 41% β-sheet (▲), and 17% unordered (■). **B**, The fitted spectrum, a weighted sum of the pure components (dashed line with ◆), and original CD spectrum (solid line with ▲).

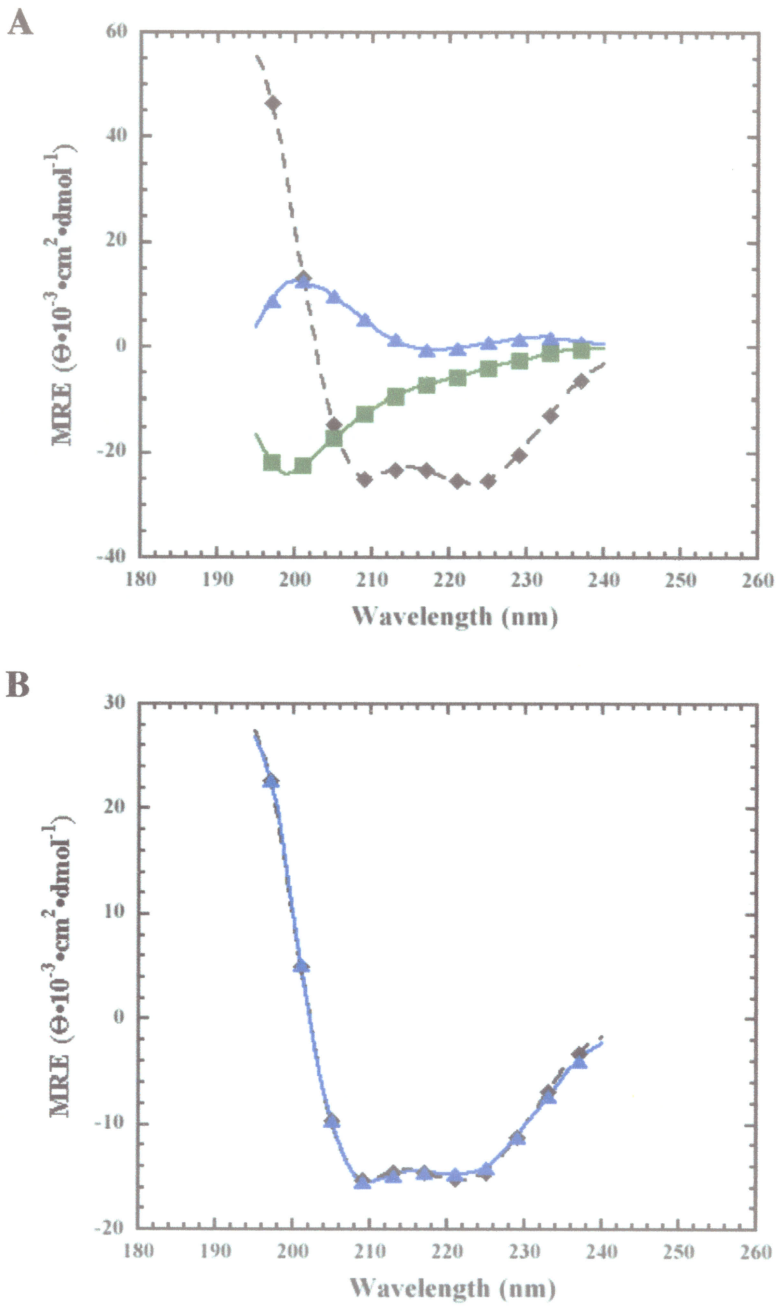


Figure 3-23. CCA deconvolution of DM extracted GF solubilized in DM at pH 4.4. **A**, The pure component curves derived from a 3-component deconvolution. The component weights are 55% α-helix (◆), 22% β-sheet (▲), and 23% unordered (■). **B**, The fitted spectrum, a weighted sum of the pure components (dashed line with ◆), and original CD spectrum (solid line with ▲).

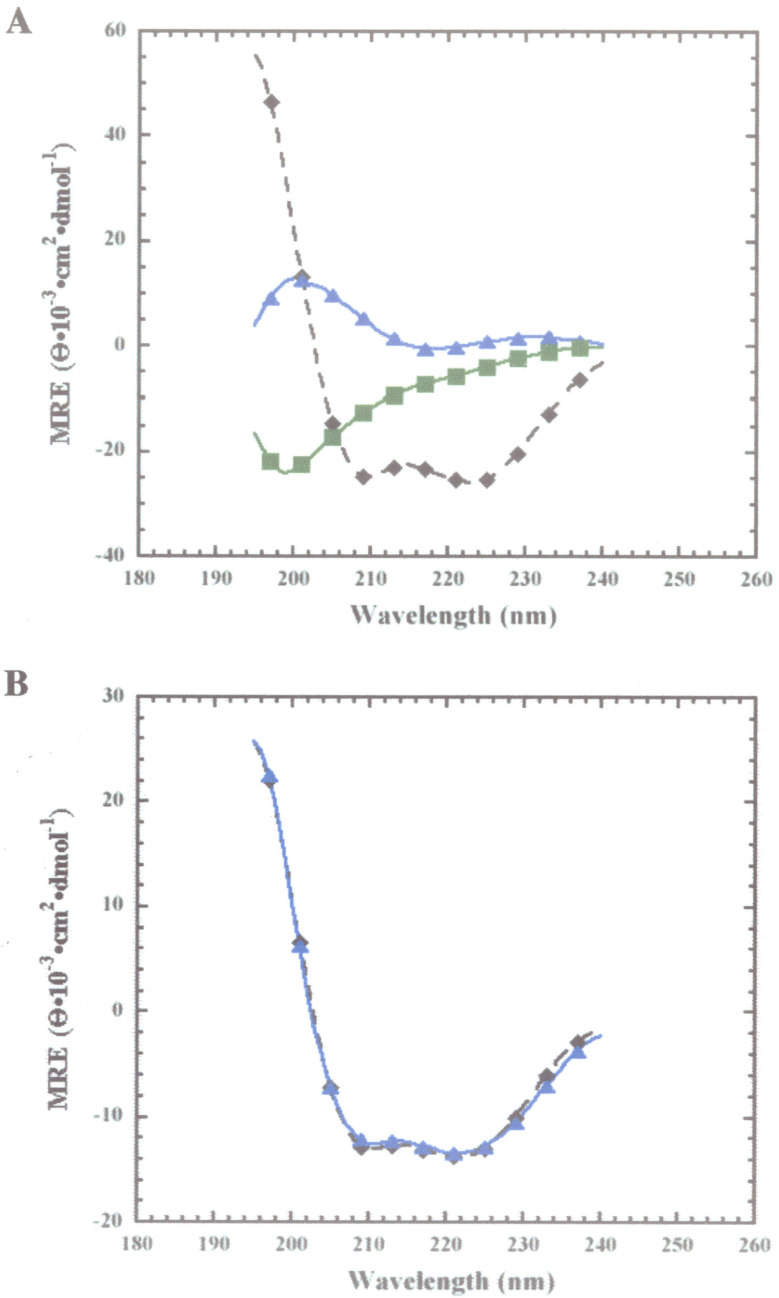


Figure 3-24. CCA deconvolution of DM extracted GF solubilized in DM at pH 7.0. **A**, The pure component curves derived from a 3-component deconvolution. The component weights are 50% α-helix (◆), 32% β-sheet (▲), and 18% unordered (■). **B**, The fitted spectrum, a weighted sum of the pure components (dashed line with ◆), and original CD spectrum (solid line with ▲).

Table 3-2. Weights of Pure Structural Components from CCA Deconvolution of the CD Spectra of GF Solubilized in Different Detergent Solutions

	SDS/urea, pH 7.6	SDS, pH 4.2	SDS, pH 7.5	OG, pH 4.0	DM, pH 4.4	DM, pH 7.0
% α-helix	32	48	49	42	55	50
% β-sheet	30	30	22	41	22	32
% β-turn	18	---	---	---	---	---
% unordered	12	22	29	17	23	18
% aromatic/ disulfide	8	---	---	---	---	---

usefulness of the CCA results are a) that the deconvoluted pure components represent real structural components found in proteins that can be assigned to a known secondary structure, and b) that the fitted spectrum reconstructs the experimental spectrum. Neither of these criteria were met when CCA was performed using the membrane protein basis set, except for the SDS/urea extracted GF sample. The fits yielded either straight lines and curves that could not be identified as any known secondary structural component, i.e. incorrect intensities such as the 208 nm transition of an α-helix curve being more intense than the 190-195 nm transition, or real secondary structural curves that had very low weights (e.g. <1%). For these reasons the CCA deconvolution using the membrane protein basis set was not found useful.

3.6.2 Tertiary Structure of GF

The analysis of protein tertiary structure was performed by monitoring CD in the near-UV region from 320-260 nm. In this wavelength range it is the CD signal originating from the aromatic residues that is being measured (Section 1.5.3.2). Since the occurrence

of these residues in proteins and their molar absorptivities are both low, the CD bands are much weaker than the corresponding bands for the peptide bond absorption measured in the far-UV region (see Table 1-3). Thus, in near-UV CD much higher protein concentrations are required, compared to far-UV studies [121]. However, at high concentrations GF is prone to aggregate in OG and in SDS/urea preparations. Because of aggregation GF concentrations high enough for near-UV CD analysis were only attainable in SDS and DM extracted and purified GF (91.2 μM and 77 μM, respectively).

The near-UV CD spectra of GF, solubilized in SDS and DM are shown in Figure 3-25. The spectrum of GF in SDS shows weak positive ellipticity whereas that in DM shows

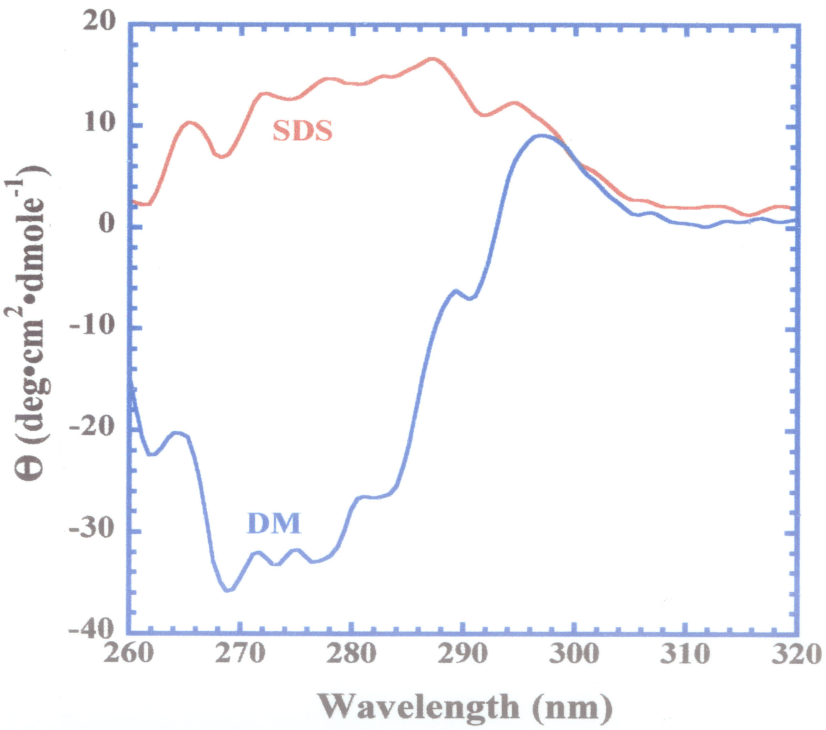


Figure 3-25. Near-UV CD spectra of SDS (91.2 μM) and DM (77 μM) solubilized GF (see text for details).

stronger, predominantly negative ellipticity. The spectra are quite complicated because of the 5 Trp, 7 Tyr, and 21 Phe found in GF. The Phe residues should contribute very little to a CD spectrum due to their high intrinsic symmetry and the low sensitivity of their transitions to changes in solvent polarizability [119]. Nevertheless, the 21 Phe residues in GF may contribute significantly to the CD spectrum in the 260-270 nm region, specifically to the band observed at approximately 265 nm in both detergents (Figure 3-25). The shoulders seen in both detergents at approximately 295 nm are most likely due to the 7 Tyr L_b transitions [121]. It is the 5 Trp residues, that each have an L_a band at approximately 275 nm and L_b doublet between 280 and 290 nm, that produce large overlapping CD signals in the 270-290 nm region, and provide the largest measure of difference between the two detergent solubilized GF samples. It is in this region where the intensity and sign of the CD signals differ the most, and is the best measure of structural difference between the two detergent solubilized states of GF.

The actual contributions of the individual aromatic residues cannot be determined by any deconvolution techniques, or other forms of post-acquisition processing. The only method available for determining the aromatic residue contributions in a near-UV CD spectrum is to change the residues one at a time by mutagenic techniques, and assess the affect on the CD spectrum. However, what the near-UV CD spectrum can indicate is the degree of structure present in the sample, since only aromatic residues held rigidly in an asymmetric environment will display CD bands [119]. Thus the comparison of the SDS and DM solubilized GF near-UV CD spectra indicates a difference in the tertiary structure in the two detergents. As noted above, the facilitator CD spectrum in DM is much more intense, and opposite in sign, to the GF spectrum in SDS.

3.6.3 GF Folding/Unfolding

The loss of protein tertiary and secondary structure, i.e. unfolding or denaturation, can be induced by chemical treatment or by temperature changes, and can be monitored by changes in CD measurements. Changes in tertiary structure are often monitored as changes in the aromatic CD spectral intensity at 273 nm and reflect the change in the degree of freedom of individual amino acids as the protein unfolds (see Section 1.5.3.2) [121]. Depending on the extent of structural loss the CD intensity at 273 nm is either greatly reduced, i.e. partial denaturation, or it disappears. Similarly, the loss of secondary structure can be monitored as a loss of ellipticity values in the peptide bond absorption region, 190-240 nm, as the structure becomes random and the CD signals average to zero [121].

The thermal denaturation of SDS- and DM-solubilized GF in the near- and far-UV region was achieved by increasing the temperature of the CD cell using the Haake circulating water bath, with subsequent acquisition of the CD spectra in both the near- and far-UV regions (see Figures 3-26 to 3-29). The near-UV CD spectra of SDS- and DM-solubilized GF at temperatures ranging from 20-80°C are shown in Figure 3-26 A and B, respectively. They illustrate that not only are the tertiary structures of GF different in the two detergents (see also Figure 3-25 and section 3.5.2), but that they unfold differently. Figure 26 A and B show that in the 20°C to 80 °C temperature range the intensity of the CD spectra of SDS solubilized GF changes very little, whereas the intensity of the DM solubilized GF spectra is greatly reduced at temperatures above 50°C. Monitoring the temperature dependent changes in ellipticity at 273 nm illustrates that the DM solubilized protein displays a sigmoidal unfolding curve whereas SDS-solubilized GF shows little evidence of unfolding at all temperatures (see Figure 3-27; discussed further in Chapter 4).

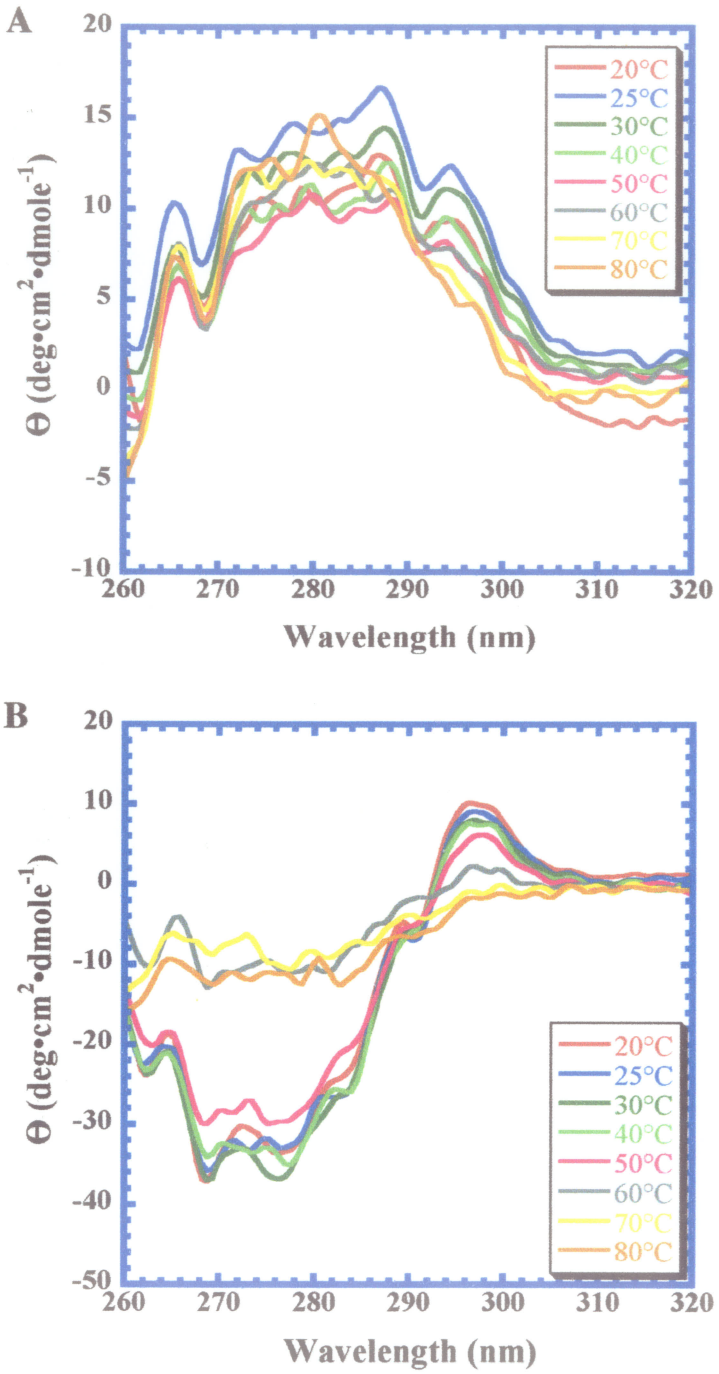


Figure 3-26. Thermal denaturation of **A**, SDS- and **B**, DM-solubilized GF, monitored by near-UV CD.

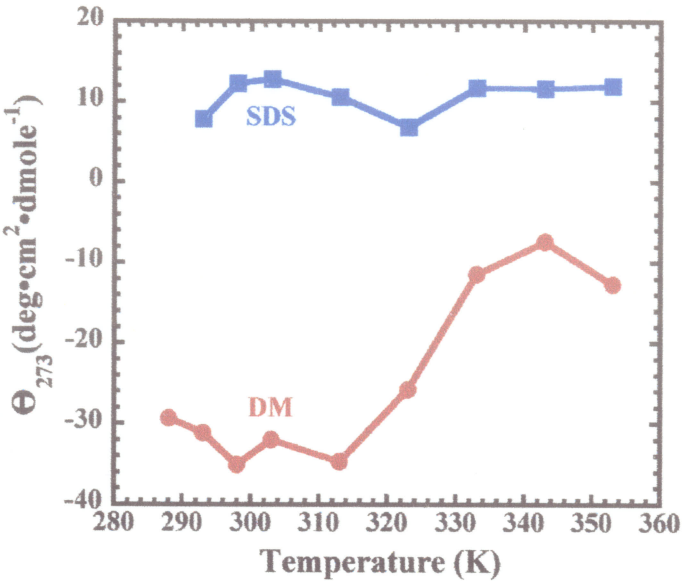


Figure 3-27. Thermal denaturation of SDS (■) and DM (●) solubilized GF monitored by changes in ellipticity at 273 nm in the near-UV CD region.

In separate experiments the thermal dependence of GF secondary structure in the two detergent preparations was determined by monitoring changes in ellipticity in the far-UV region. Figures 3-28 A and B illustrate that as temperature increases the characteristic α-helix ellipticities (approximately 220 nm and 208 nm) of both SDS- and DM-solubilized GF reduce in intensity. Thus, it is apparent that GF secondary structure is lost in both detergents upon temperature increase, with both showing sigmoidal unfolding curves (see Figure 3-29).

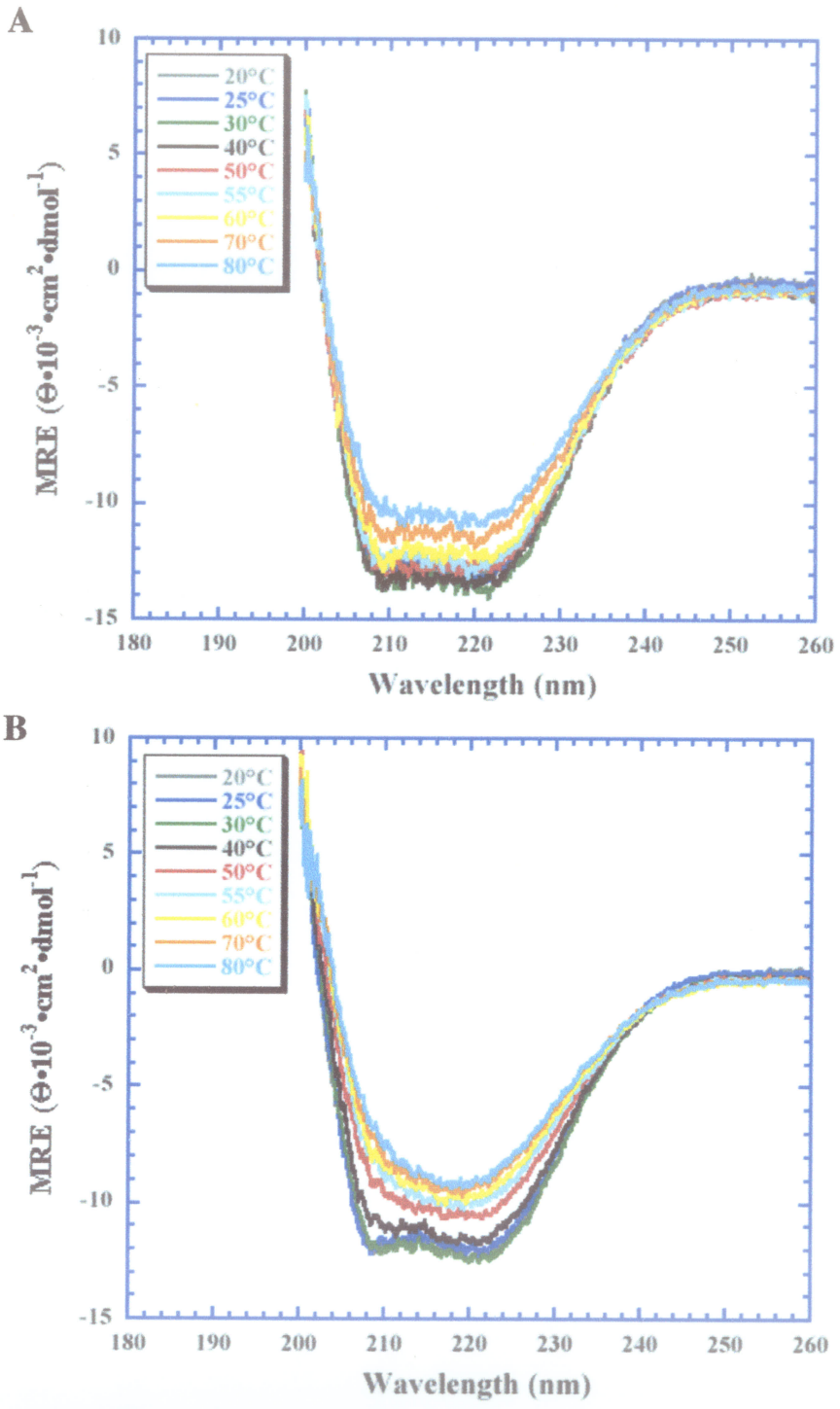


Figure 3-28. Thermal denaturation of A, SDS- (7 μM) and B, DM-solubilized GF (7 μM), monitored by far-UV CD.

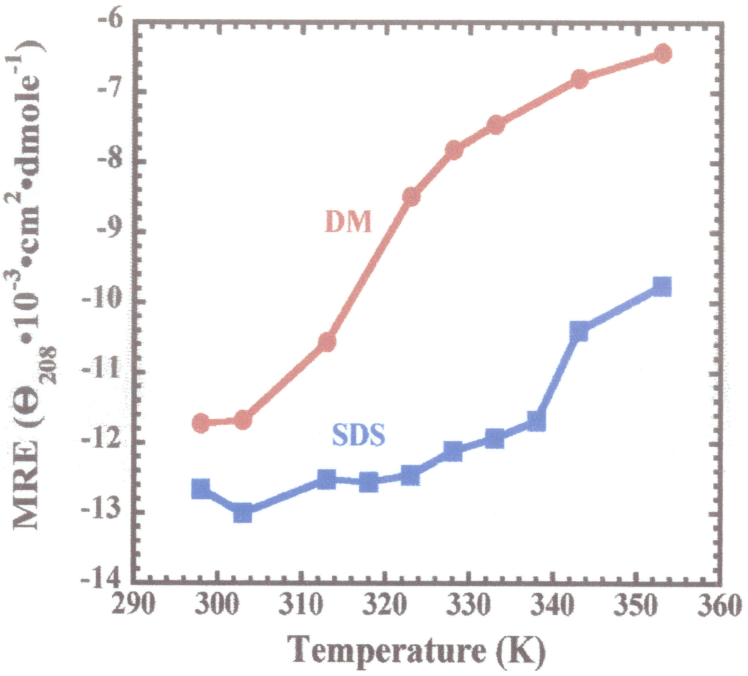


Figure 3-29. Thermal denaturation of SDS (■) and DM (●) solubilized GF monitored by changes in ellipticity at 208 nm, in the far-UV CD region.

The data in Figures 3-27 and 3-29 were fitted to equation 2-4 using the program Mathematica, and provided a means of determining the midpoints (T_m) and widths (ΔT) of three of the four unfolding transitions. The fitted curves are shown in Figures 3-30 to 3-32, and indicate that the thermal denaturation of GF in detergent is truly sigmoidal. This method assumes that the protein exists in two states, helix and random coil in the far-UV region, and folded and unfolded in the near-UV region. The values of T_m and ΔT are given in Table 3-3. Also given in Table 3-3 is the value d χ^2, a parameter that indicated the goodness of fit, *i.e.* small χ^2 values indicate a good fit of the curve to the data. The implications of the data presented in Table 3-3 will be discussed in Chapter 4.

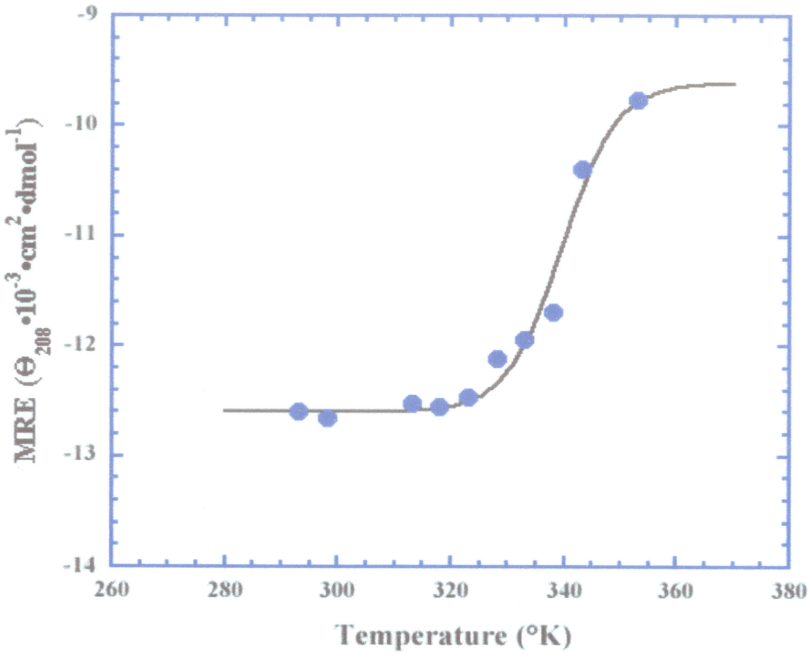


Figure 3-30. Thermal denaturation of SDS solubilized GF monitored by ellipticity changes at 208 nm (●), plotted against the *Mathematica* output of the fitting procedure (solid line). θ_{fold} = -12552, θ_{unfold} = -9498.

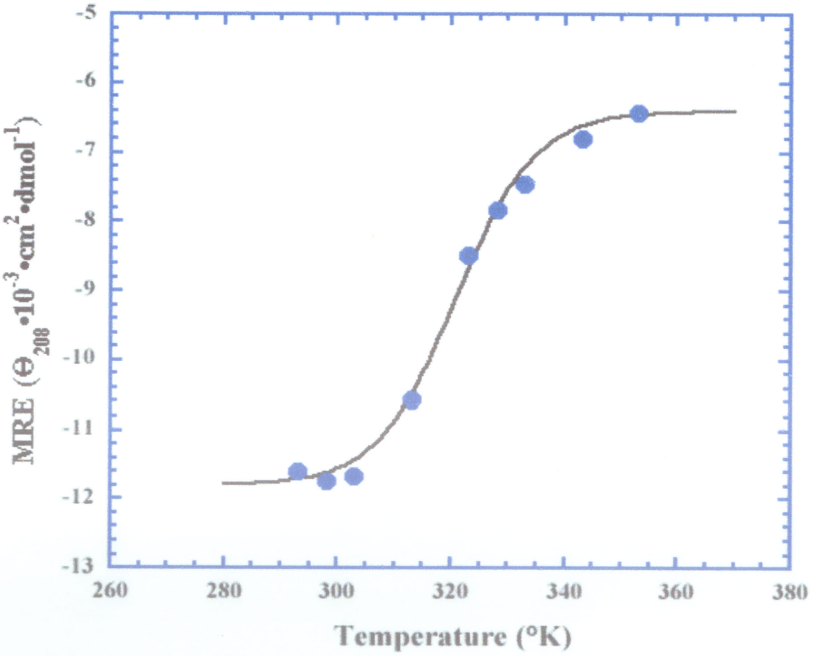


Figure 3-31. Thermal denaturation of DM solubilized GF monitored by ellipticity changes at 208 nm (●), plotted against the *Mathematica* output of the fitting procedure (solid line). θ_{fold} = -11869, θ_{unfold} = -6558.

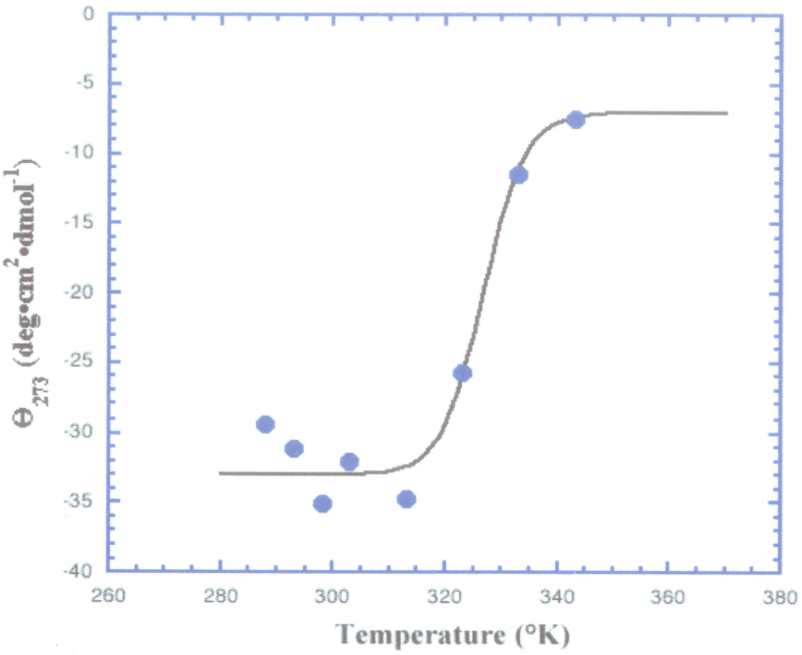


Figure 3-32. Thermal denaturation of DM solubilized GF monitored by ellipticity changes at 273 nm (●), plotted against the *Mathematica* output of the fitting procedure (solid line). θ_{fold} = -32.5 , θ_{unfold} = -7.5.

Table 3-3. Fitted values for SDS and DM solubilized GF

	T_m, °K	ΔT, °K	χ^2
SDS, Far-UV	340.3	22.2	0.224
SDS, Near-UV	---	---	---
DM, Far-UV	320.2	30.7	0.234
DM, Near-UV	326.8	15.9	28.2

CHAPTER 4

DISCUSSION

4.1 GF Overexpression and Purification

One of the major impediments to structural analysis of membrane proteins is the small number of molecules that are typically synthesized by the cell. The overexpression of membrane proteins poses a potential problem as the presence of an abnormally large number of membrane proteins can be toxic to the cell simply because they are associated with the membrane [137]. For example, the increased production of functional tetracycline/ H^+ antiporter resulted in cell death [195]. Many factors influence the (over-) expression levels of membrane proteins in *E. coli* including the strength of the promoter on the expression vector, the proteolytic stability of the recombinant protein, and its translation efficiency. Environmental parameters also play an important role in the level of expression and state of the membrane protein. For example, overexpression of functional rat neurotensin receptor was critically dependent on a temperature of 24°C [196]. Thus, the optimal *E. coli* expression conditions must be determined by a trial-and-error approach [137].

Over the past two decades many systems have been developed to increase the expression levels of both water soluble and membrane proteins [137, 146b, 174] including bacterial [174], insect [197], and mammalian expression systems [198]. Methods used today for the overexpression and purification of both water-soluble and membrane proteins rely on molecular biological manipulation of cellular processes to promote the over-production of recombinant proteins [174], and require a great deal of knowledge about the molecular biology of the host cell line. Fortunately, *E. coli* is one of the most widely used

and well characterized expression systems in molecular biology, -- the "work-horse" of molecular biology.

4.1.1 Advantages of the pET Expression System

Our choice of expression system was initially based on the ability to affix an N-terminal polyhistidine tag, which has been successful for the purification of membrane proteins. For example, the glucose transporter of *E. coli* was overexpressed and purified in milligram quantities by IMAC [178]. The pET expression system adopted for the overexpression of GF adds the desired N-terminal polyhistidine tag and also takes advantage of the strong bacteriophage T7 promoter incorporated into the vector. Among the advantages of this system over other methods, is that the expression of GF can take place only in an expression host cell-line, *i.e.* DE3 lysogens. DE3 lysogens carry a plasmid containing a chromosomal copy of the gene for T7 RNA polymerase, with its expression under control of the *lac* promoter. Thus, the lac repressor protein is inactivated upon the addition of IPTG [199], which leads to the expression of T7 RNA polymerase and, subsequently, the genes inserted into the pET vector that are under control of the T7 promotor. This tightly controlled expression system provides a means for genetic manipulation of the expression vector in non-expression cell-lines, *i.e.* non DE3 lysogens, without background expression of the recombinant protein, advantageous if the overexpressed protein is toxic to the cell. A further advantage gained from the pET expression system relates to the selective inhibition of bacterial RNA polymerase, but not T7 RNA polymerase, by the antibiotic rifampicin. The addition of rifampicin to a growing culture provides a mechanism for shutting down native protein synthesis without any negative effects on expression from the pET

vector [160, 161]. In some cases the recombinant protein represents as much as 50% of total cellular protein. [161].

The pET28*glpF* expression construct yields recombinant protein with an N-terminal hexahistidine (His_6) segment and T7-epitope for protein purification and immunodetection purposes, respectively. The use of polyhistidine segments for protein purification by IMAC has become a widely used technique in the last 4-5 years and in most cases there is little or no effect of the His segment on protein function or structure [200-202]. The addition of the His_6-segment to the N-terminus of GF provides a means of binding the protein to Ni^{2+}-NTA Sepharose resin so that non-bound impurities can be removed. Elution of GF from the Ni^{2+}-NTA resin can be achieved either by washing the column with imidazole which competes with GF for the binding sites, or by reducing the buffer pH to protonate the His_6 segment. The latter method was chosen for GF purification, as the former required further sample processing due to imidazole interference during GF spectroscopic analysis. As shown in Figure 3-9, purification of GF by IMAC provided a means of attaining highly pure GF in one purification step.

4.1.2 Optimization of expression

To optimize the expression levels of GF, growth conditions were varied and the effects on expression were analyzed. It was empirically determined that, for optimum expression, a) the cell line BL21(DE3)pLysS should be used, b) the culture should be grown in TB broth at 37°C, c) IPTG induction should be at an OD_{600} of approximately 0.7, d) rifampicin should be added at one hour post induction, and c) the cells should be harvested two hours after the addition of rifampicin. These parameters provide further evidence that high cell densities are required for maximum GF expression levels, as TB is a

rich medium that provides higher cell densities than the commonly used LB medium [159], and induction at the mid-log phase of growth ensures a relatively high cell density during expression of the recombinant protein. Optimizing expression using the above parameters provides approximately 14 mg of 95% pure GF from 9.5 grams of cells recovered from 2 litres of *E. coli* cell culture.

To maximize GF expression, after stopping native protein synthesis, the optimum time for the addition of rifampicin was determined (see Section 3.2). The results indicate that cell density at the time of rifampicin addition is not critical, in the OD_{600} range of 0.6 - 1.0, since there is no substantial difference in expression levels detected within the limits of Western immunoblot analysis. Thus, induction of expression at mid-log phase (OD_{600} = 0.7), with the rifampicin added 1 hour later, provides cell densities high enough for GF-expression with yields over two-fold higher than the previously optimized GF expression protocol discussed below (see Figure 3-3). However, increasing the growth time after rifampicin addition does not significantly increase GF yield. This observation may indicate that the cellular capacity to express recombinant GF from the T7-promoter is limited when native bacterial protein synthesis is inhibited.

The protocol established above was not the first developed for GF overexpression. Prior to discovering the advantage gained from the use of rifampicin, the protocol involved growth at 30°C, induction of expression at an OD_{600} of approximately 0.4, and cells were harvesting no later than two hours post-induction. Although the *E. coli* strain BL21(DE3)pLysS lacks the *lon* and *ompT* proteases, the advantage observed for growth at the lower temperature was considered to reflect either reduced degradation of GF by other proteases present in *E. coli* [203] or reduced RNase degradation of recombinant GF mRNA

[204]. Grisshammer and Tate [137] suggested that an improvement of membrane protein expression levels with reduced culture temperature indicates that the insertion of protein into the membrane may be rate limiting and that slowing the rate of protein synthesis by reducing the temperature may reduce the amount of membrane-free GF available for proteolytic degradation. They further state that reducing the rate of protein synthesis may also result in less GF mRNA left unbound to ribosomes and therefore less protected from RNase degradation. Thus, it is possible that rifampicin aids GF expression not only by freeing cellular resources for GF expression, but also by shutting down the production of the offending bacterial proteases and RNases, thereby indirectly increasing the survival rate of both GF mRNA and protein.

As mentioned, approximately 14 mg of 95% pure GF are recovered from 2 litres of *E. coli* cell culture, but it may possible to increase these levels to approximately 10 mg per litre (discussed further in Chapter 5). These levels are adequate for electron or X-ray crystallography, but for protein levels adequate for ^{1}H NMR studies, it will be necessary to prepare at least 3 to 4 litres of *E. coli*. Furthermore, for NMR analysis of isotopically labelled GF it will be necessary to use minimal media, and since cell densities are lower in minimal media [146b] it will also be necessary to prepare 4-6 litres of cells.

4.1.3 Detergent Extraction and Purification of GF

Eight detergents were tested for their ability to extract GF from insoluble cell fractions (see Table 3-1). Initially, LDAO was chosen as the best detergent because it gives a high yield of extracted GF, it dissolves the insoluble cell fraction quickly, and is inexpensive. However, the difficulties encountered in the protein purification steps, and the severe aggregation of GF in LDAO (see Figure 3-7) necessitated consideration of an alternative

detergent. The 150 mM SDS and 4 M urea mixture proved to be a highly effective extraction agent for GF as long as the sample temperature was kept near 25°C. Extraction below 15°C was not possible due to SDS crystallization [186]. In the absence of urea, SDS was found to be a successful extraction agent only when the incubation time was increased by approximately one-half hour, to allow for complete solubilization of the water-insoluble cell fraction. However, the non-ionic saccharide detergent β-D-dodecylmaltoside was found to be just as effective as SDS and has the added advantage of causing significantly less denaturation [186] (see Table 3-1). Similar to SDS extraction, large scale extraction of GF using DM was only successful when the extraction time was increased by one-half hour.

No protocols have been published, other than the paper by Manley *et al.* [111], that demonstrate the use of Ni^{2+}-NTA resin for the purification of SDS-solubilized membrane proteins. Indeed, the Qiagen IMAC manual recommends against the use of SDS, yet provides no explanation. The purification of SDS-solubilized GF, by IMAC on a Ni^{2+}-NTA column, proved to be no more difficult than the standard protocol for water-soluble protein suggested by Qiagen [162b], or the purification of DM-solubilized GF. However, it was necessary to adjust the purification protocol based on the detergent used for solubilization. It was found that GF solubilized in DM and OG (independently) elute from the Ni^{2+}-NTA resin at a pH just below the pKa of histidine (*i.e.* below pH 6.0) as expected, while SDS-solubilized GF eluted at a much higher pH (above 7.0). This difference in the elution pH between the non-ionic and ionic detergents may be due to the effects that detergent headgroups can have on the local pH around the micelle, as the strong anionic headgroup of SDS (see Figure 1-6) induces H_3O^+ ions to concentrate around the sulfate headgroup so that the

pH at the micelle surface is lower than that of the bulk solution [206]. Thus, the pH near the His_6-tag may be lower than the pH recorded by the pH meter, causing SDS-extracted GF to elute at an apparently higher pH than expected. However, with careful optimization of the low pH washing steps, SDS-solubilized GF is routinely purified to greater than 95% by Ni^{2+}-NTA IMAC in one step (see Figure 3-9, lane 2).

The extraction and purification of GF with OG was not attempted. The high CMC of OG (20-25 mM) would require the use of large quantities during extraction and purification, and at a cost of more than $80 per gram this was considered to be too expensive. Therefore, to produce OG-solubilized GF, the protein was extracted and purified in DM, and the DM was then exchanged for OG while GF remained bound to the Ni^{2+}-NTA column. Pure OG-solubilized GF was obtained (see Figure 3-9, lane 4). However the protein proved to be quite unstable and attempts to adjust solution pH, after elution from the Ni^{2+}-NTA column, resulted in immediate and severe oligomerization. GF aggregates were produced with molecular weights high enough for a retention of a large fraction of the sample at the stacking/resolving gel interface (see Figure 3-10). It is likely that GF is positively charged at low pH and therefore does not aggregate as easily, though increasing the pH closer to the isoelectric point may make aggregation easier. The OG-solubilized protein was used in an attempt to prepare GF proteoliposomes. However, dialysis of an OG-solubilized GF and *E. coli* phospholipid vesicle mixture also resulted in severe aggregation of facilitator (data not shown). The inability of OG micelles to maintain GF in a soluble state is likely a consequence of the short hydrocarbon chain in the OG monomer. This short carbon chain provides OG with a high CMC (see section 1.1.2.1), which is useful for lipid reconstitution studies, but also produces micelles with small hydrocarbon core diameters

[1, 2]. Evidently, OG micelles do not provide a large enough hydrophobic surface for stable solubilization of GF in solution and perturbations result in the immediate aggregation of GF. Since this is not observed in SDS and DM solubilized GF, and since these two detergents both have 12-carbon acyl chains, it appears that GF solubilization efficiency and stability may increase with increasing detergent acyl chain length.

4.1.4 Is Overexpressed GF Properly Folded?

To determine if recombinant glycerol facilitator is properly folded, inserted into the *E. coli* inner membrane, and active *in vivo*, a xylitol transport assay was conducted. The xylitol transport activity of recombinant GF, expressed from the pET28*glpF* plasmid in *E. coli*, was found to be virtually identical to that previously reported for native GF (see Figure 3-16) [106, 107, 188]. For example, it was reported by Heller *et al.* [106] that the half-time of equilibration of a xylitol gradient across the membrane of GF expressing cells was 0.7 minutes, whereas we measured a half-time of equilibration of approximately 0.6 minutes for recombinant GF. These results suggest that functional protein is inserted into the *E. coli* inner membrane and that the addition of the amino-terminal His_6-T7-tag has little or no effect on the structure and activity of recombinant GF. That the amino-terminal tag has no effect on GF structure/function is in accord with the crystal structure [110] which indicates that the amino-terminus is extracellular, is distant from the pore, and likely has no impact on GF structure or function (see Figure 1-22).

The xylitol-transport study also provides evidence that (at least some of) the protein is being inserted into the bacterial membrane in a functional state, rather than being sequestered in inclusion bodies. Electron micrographs do not show any indication of either extra membranous regions for the insertion of overexpressed GF or dark granular GF-inclusion

bodies (see Figure 3-4). This absence of any visible effect of GF overexpression may be an indication that the expression levels of GF can be improved upon, and that the overexpressing cells are far from saturated with GF. However, this depends on whether the changes that membrane protein overexpression cause on cell morphology is a rule or simply an effect observed in many cases of membrane protein overepression [137].

4.2 Spectroscopic Analysis of GF Structure

4.2.1 GF Primary Structure

MALDI-TOFMS is a powerful analytical tool for accurate mass determination of biomolecules [192]. However, the presence of detergents and lipids in membrane protein preparations poses a challenge for mass spectrometry due to the reduction in the signal-to-noise ratio and mass resolution resulting from detergent or lipid adducts to the protein [190]. The removal of detergent prior to analysis has been demonstrated to degrade spectral quality because of protein precipitation and the consequent loss of sample for desorption [191]. However, it was also shown that the use of detergents at and above their CMC improves the solubility of membrane proteins and increases their concentration at the surface layers of the matrix crystal, providing more analyte for desoption [191, 192]. This observation could not be confirmed with GF solubilized in SDS or DM, therefore the polyurethane (PU) membrane method was used for glycerol facilitator MALDI-TOFMS analysis [164]. The PU method provides a means for extensive washing of samples to remove salts, detergents, and other buffer components prior to desorption and significantly increased the signal-to-noise ratio and mass resolution of GF spectra compared to those of unwashed samples. We reasoned that OG might not interact as strongly with GF and therefore might be a better candidate for the preparation of protein for mass-spectral analysis.

We therefore prepared GF in OG by Ni^{2+}-NTA on column exchange of SDS for OG (see section 3.3.2). The OG solubilized protein yielded a well resolved mass spectrum of monomeric GF with the highest signal above the noise of all the spectra produced (Figure 3-17).

The detergent-exchanged protein was oligomeric according to SDS-PAGE, however MALDI-TOFMS did not detect the presence of the oligomeric species. This suggests that the GF oligomer either dissociates in the mass spectrometer or that not enough oligomer was present for detection. The former is possible since although MALDI-TOFMS is a relatively soft-ionization technique most non-covalent associations do not survive this procedure [207]. However, the latter is more likely since we have observed that the OG-solubilized protein is prone to aggregate, and aggregated membrane proteins are difficult to desorb. A further possibility is that the oligomers observed in SDS-PAGE gels form in the gel during electrophoresis. However, if oligomers were formed during electrophoresis, the GF bands would not be well resolved and would appear very diffuse (smeared) in the Coomassie stained gel, but this is not the case (see Figures 3-9 and 3-10). Therefore, it is most likely that oligomeric GF was not detected by MALDI-TOFMS because the OG solubilized protein aggregated and precipitated upon the addition of the acidic MALDI matrix, and was then not available in the matrix crystal for desorption [191, 192].

MALDI-TOFMS analysis of recombinant GF solubilized in OG indicates that the observed mass of 33,650 Da is within 0.4% of the expected GF monomer mass of 33,505 Da. The difference of 145 Da is not likely due to a bound OG monomer, since the mass difference would then be 454.5 Da, larger than the error from the measurement. The additional mass might be explained by the addition of several sodium or phosphate ions (23 and 95 Da, respectively) from the sample buffer.

4.2.2 GF Secondary Structure

Far-UV circular dichroism analysis of proteins in solution is useful for quantifying secondary structural elements. The occurrence of CD bands at 222 nm and 208 nm in the CD spectra of GF solubilized in SDS/urea, SDS, DM, and OG indicates that the prevalent secondary structure is the α-helix (see section 3.5.1 and Figure 3-19 A-F). Deconvoluting the spectra using the CCA program confirmed this and provided statistical weights for the structural elements of GF in the different detergents (see Table 3-2) [131]. The analysis indicates that GF has the highest α-helix content (50-55%) when it is solubilized in the non-denaturing detergent DM. A slightly lower α-helix fraction is observed for the protein in SDS (48-49%). Significantly lower helix content was observed for GF extracted with SDS/ urea and solubilized in SDS (32%), and for GF solubilized in OG (42%). The difference in helix content between GF extracted in SDS with and without urea likely reflects the denaturing effects of urea on protein structure [186], and suggests that GF is in a more denatured state after exposure to urea. Similarly, the slightly lower helical content of GF solubilized in SDS without exposure to urea, compared to GF solubilized in DM, suggests that GF may be in a slightly more unfolded state in SDS than in DM. This is in accord with the view that SDS is a more denaturing detergent than DM. The even lower helix content measured in OG suggests partial denaturation of the protein in that detergent as well (see section 3.3.2 and Figure 3-10). Recall that the solubility of GF in OG is low and that the protein oligomerizes in this partially unfolded state [208]. The absence of oligomerization of the protein in SDS, even though CD suggests that the protein may be partially unfolded, may be due to charge repulsion between the anionic headgroups of SDS that coat the protein.

Sequence alignments of the *E. coli* GF with the aquaporins [58], hydropathy analysis [63], and the electron diffraction structures of AQP1 [87, 88] suggested that about 47% of GF is composed of 6 transmembrane α-helices. More recent electron and X-ray diffraction data [66, 209, 209a] also indicated that half of loops B and E form short helical segments that form part of the pore of the protein, elevating the helix content to approximately 55%. Several weeks after the publication of a large portion of the material presented in this thesis [111], the X-ray crystal structure of GF was published [110]. The X-ray data show that the GF protein has secondary structural features that are similar to those displayed in the AQP1 cryoelectron crystal structure [66] including the 6 transmembrane α-helices and the two short helices formed by loops B and E[1]. The GF crystal structure also shows the presence of 3 very short helical segments containing 10, 9, and 4 residues formed in loops C and E (see Figure 1-23). Thus, the crystal structure indicates that GF is up to 67% helical. The CCA-derived helical contents of DM, SDS, OG, and SDS/urea solubilized GF are 55%, 49%, 42%, and 32%, respectively [see Table 3-2]. However, these weights were determined for the recombinant form of GF which contains an additional 35 residues relative to wildtype GF. If we assume that no amino acids in the His_6-T7 tag are helical and recalculate the helix contents for the rest of the molecules the helicity rises to 62%, 55%, 47%, and 37% for GF solubilized in DM, SDS, OG, and SDS/urea, respectively. Therefore, the CCA-derived helical contents of DM-solubilized GF are in good agreement with the AQP1 and GF diffraction data, suggesting that GF is in a native-like state in DM. Indeed, that SDS-solubilized GF retains 55% helix suggests that GF remains highly folded in this detergent. One possibility is that in SDS the 3 short loop helices are denatured whereas the

1. I have adopted the nomenclature used by Murata *et al.* [66], *i.e.* the helix and loop nomenclature is the same as for AQP1 (see Sections 1.3.3 and 1.4.5)

transmembrane helices are retained. This would provide an expected helical content of approximately 58%.

4.2.3 GF Tertiary Structure

CD analysis of proteins in the near-UV region reveals tertiary organization. The CD measurable transitions arise from the aromatic amino-acid side chains that are being held in a rigid, asymmetric, electronic environment. However, because there are no deconvolution techniques available to extract detailed structural information from near-UV protein CD spectra, analysis is limited to the determination of spectral, and therefore structural, differences from sample to sample. The simple presence of a measurable near-UV CD spectrum is an immediate indication that a protein sample contains some organized tertiary structure [121]. Judging from the near-UV spectra of GF solubilized in SDS and in DM the protein has elements of tertiary structure in both detergents (see Figure 3-25). Comparison of the near-UV CD spectra of GF also indicates that its tertiary structure is different in the two detergents. The spectrum of DM-solubilized GF has more than twice the intensity of, and is for the most part opposite in sign to, the SDS-solubilized spectrum. This suggests that there are more side-chains fixed in a stable environment in the DM-solubilized protein than in the SDS-solubilized protein. Furthermore since the measured CD spectrum is a statistical weight of all the components present the spectra suggest that the SDS-solubilized GF is dynamic. The weights of all the conformations present in the sample effectively reduce the intensity of the observed spectrum. Whether or not the DM-solubilized spectrum represents native GF tertiary structure cannot be conclusively determined from the spectra, and further analysis is required (see section 4.2.4). Nevertheless, the intensity of the DM-

solubilized GF spectrum, compared to that in SDS, indicates that it is the most likely candidate for native tertiary structure.

One interpretation of the far- and near-UV CD results is that DM-solubilized GF adopts one tertiary structure, while the sample in SDS is more flexible resulting in the averaging of the CD signal to close to zero[1]. It may be that SDS-solubilized GF adopts a molten-globule like state. This is a compact, partly folded protein state with few tertiary interactions and some secondary structure [209b]. This would yield protein with a measurable far-UV CD spectrum (see Figures 3-20 and 3-21) and negligible near-UV CD spectrum (see Figure 3-25).

4.2.4 Thermal Denaturation of GF

The regular arrangement of structural elements gives rise to the distinct CD bands in both the near- and far-UV regions. Thus the loss of protein structure upon unfolding can be followed by monitoring the changes in the CD bands that represent distinct structural elements. For example, changes in secondary structure are monitored by changes in the α-helix intensity at 208 nm and the tertiary structure is measured by following the Trp band at 273 nm. The thermal unfolding of SDS and DM solubilized GF was monitored in this manner (see Figures 3-27 and 3-29) in order to determine if there are any differences in the stability of the molecules in the two environments.

Plotting the change in spectral intensity at 273 nm against the sample temperature (see Figure 3-27) shows that only small changes in the SDS-solubilized sample occur over the course of the experiment. On the other hand the DM-solubilized sample displays a sig-

1. Equation 1-29 is true for both far- and near-UV CD.

moidal unfolding curve (see Figure 3-27). This suggests that the residual near-UV CD spectrum in SDS is not reflecting an organized tertiary fold of the protein that can be denatured by an elevation in temperature. In contrast, the DM-solubilized GF spectra show dramatic changes with increasing temperature and this suggests that GF loses tertiary structure and unfolds at the elevated temperatures.

The effects of elevated temperature on GF-solubilized in SDS and DM, as monitored by far-UV CD, are shown in Figure 3-28. Plotting the change in intensity at 208 nm as a function of temperature (see Figure 3-29) indicates that both SDS- and DM-solubilized GF display sigmoidal unfolding curves (see Figure 3-30 and 3-31) with T_m's of 340 K and 321 K, respectively. Thus GF secondary structural elements are lost in both detergents when the temperature is increased, as a consequence of thermal denaturation. That the midpoint of the thermal denaturation of SDS-solubilized GF is 19 K higher than DM solubilized protein indicates that secondary structure of GF solubilized in SDS is more stable and may be more protected from temperature changes than when it is solubilized in DM.

Therefore, far-UV CD indicates that the GF structural elements in SDS and DM are similar (see Table 3-2) but that the protein unfolds in the detergents in a slightly different manner, reflected by the different T_m values. Indeed, upon thermal denaturation the helical contents in SDS- and DM-solubilized GF are reduced from 49% to 39% in SDS and from 55% to 32% in DM (determined by CCA analysis). Evidently, the protein in DM loses twice as much secondary structure as that lost in SDS. In both cases, the residual helix content may reflect a stable core in the molecule that is more resistant to denaturation. Alternatively, aggregation of the protein (see below) at higher temperatures may complicate the

interpretation of this result. Aggregation could protect some of the folded proteins from unfolding.

4.3 Implications of GF Structure in Detergent and *in vivo*

4.3.1 Detergent Solubilized GF

There is considerable evidence that AQP1 [82, 83, 87, 88, 209a], AQP4 [210], and AQPZ [58] are tetrameric molecules in membranes. Furthermore, purification and concentration of AQP1 [49], AQPZ [166], and AQP0 (MIP-26) [71] results in the production of oligomers (dimer, trimer, tetramer, and sometimes higher oligomers) observable in SDS-PAGE. In fact, several other membrane proteins have been demonstrated to partially or completely retain their membranous oligomeric state during SDS-PAGE including M13 coat protein [180], glycophorin A [181], the fusion domain of the HIV-1 envelope glycoprotein (gp120-gp41) [182], and phospholamban [183]. In addition, differences observed in the electrophoretic mobilities of urea-unfolded and detergent-solubilized mitochondrial voltage-dependent anion channel porin on SDS-PAGE strongly suggest that elements of the tertiary structure of this β-barrel membrane protein are preserved during SDS-PAGE [211]. The observation that GF has only slightly less helical content in SDS than DM supports the idea that membrane proteins dissolved in SDS retain some elements of native structure during polyacrylamide-gel electrophoresis. On the other hand, oligomers of GF in SDS-PAGE are observed after the following denaturing treatments: heating in SDS, extraction of SDS-soluble protein with urea, addition of urea to DM-solubilized protein, elevation of pH in OG-solubilized protein, dialysis of OG-solubilized protein, and extraction of GF with LDAO. That heating and urea can induce oligomerization suggests that

these oligomers arise following partial or complete denaturation of the protein and this data is supported by the observation of lower helix content in these samples, i.e. SDS/urea extracted GF has 32% helix and heat denatured GF has 31-38% helix. However, dimeric facilitator is observed to form in the absence of any denaturing agents (SDS, urea, heating) when the protein is solubilized in DM and is particularly abundant in highly concentrated solutions (see Figure 3-14B).

The strongly denaturing effects of SDS [1, 2, 186] have made it useful in the electrophoresis of proteins. A widely accepted model states that SDS disrupts tertiary interactions and increases helical content of water-soluble, globular proteins yielding rod-like proteins coated with SDS at a 1.4:1.0 ratio (w/w) [146b, 158]. The idea that GF secondary structure is retained when solubilized in SDS may be surprising. However, SDS is not a universal denaturant [180-182], and the retention of secondary structure in SDS solutions has been observed for other membrane proteins including neuraminidase of influenza virus [212], alkaline phosphatase in liver cell plasma membranes [213], and phospholipase A1 of *E. coli* cell membranes [214]. In the light of these observations then is is not too surprising that SDS does not significantly denature GF secondary structure to any greater extent than DM. Indeed, comparison of the experimental helical contents of 53-62% (adjusted to account for the N-terminal His_6-T7 tag), to the expected helix content of 67% based on the crystal structure [110], suggests that GF may be in a native-like state in both detergents.

The far-UV CD spectra of SDS- and DM-solubilized GF show that the heat denatured proteins have a relatively featureless spectrum in the 240-200 nm region having lost the distinct 208 and 222 nm bands (Figure 3-29A and B, 80°C spectrum). The similarities between these CD spectra and those of SDS/urea (see Figure 3-19B) and OG (Figure 3-

22B) solubilized GF suggest that SDS/urea- and OG-solubilized GF are not in a native conformation and are likely partially unfolded. The finding that urea can induce protein denaturation is expected [186], and the dual action of SDS and urea would be expected to denature most proteins. Furthermore, the inability of OG to maintain GF in a native conformation is, as mentioned above, a likely consequence of the short 8 carbon acyl chain that constitutes the hydrophobic tail of OG (Section 4.1.3). That OG-solubilized GF forms high molecular weight aggregates (Figure 3-10), whilst in SDS/urea extracted GF the highest molecular weight species is a tetramer (see Figure 3-9, lane 1), may be a consequence of the highly denaturing and disaggregating effects of SDS and urea and reflects the greater ability of SDS in maintaining GF in a soluble and (in this case) denatured state.

The thermal denaturation of SDS- and DM-solubilized GF, monitored by CD, also illustrates the different influence of these detergents on GF structure. In essence, DM appears to promote the retention of both 2° and 3° structure when it is used for extraction and purification of GF, whereas the SDS extracted and purified protein retains only secondary structural elements. That GF retains secondary but not tertiary structure in SDS is an indication that the protein may adopt a molten globule-like state. The mid-point of the thermally induced unfolding transition (T_m) for GF (321 K), determined by far-UV CD analysis, suggests that DM-solubilized GF loses secondary structure at a lower temperature than in SDS (340 K) even though it may contain greater tertiary structure. This may be due to an increase in the "protective" or stabilizing effect of SDS on GF secondary structure. Since transmembrane helices may fold independently of one another, the higher T_m in SDS may also indicate that, relative to DM-solubilized GF, GF in SDS is already partially denatured. So weakly stable secondary structure in DM unfolds at lower temperatures but is absent in

SDS. This would explain the lower helix content of GF solubilized in SDS (48-49%) compared to DM (50-55%) as a partial loss of native helix content.

One model of protein folding suggests that the different levels of protein structure form in a cooperative manner with secondary structure forming first, followed by tertiary and quaternary structure [132]. Protein unfolding can then be expected to proceed in a reverse, but also cooperative, manner. The thermal denaturation analysis of GF in DM indicates that the loss of secondary structure has a T_m of 320 K and tertiary structure unfolds with a T_m of 327 K. The implication is that secondary structure is lost before tertiary structure, which is highly unlikely. What the CD spectra, both near- and far-UV, do not indicate is the presence or absence of GF oligomers (dimer, trimer, and tetramer) in the protein preparation. Furthermore, the two sets of spectra were acquired on samples that differed in concentration by approximately 10 fold. High GF concentrations have been shown to induce GF oligomerization when it is solubilized in DM (see Figure 3-14, lane 5). Therefore, the elevation in the T_m of unfolding as determined by near-UV CD, compared to far-UV CD, may reflect a higher proportion of oligomerized protein at the higher concentration. Oligomerization may prevent disruption of tertiary structure until the oligomers have dissociated into monomers thereby elevating the T_m for unfolding compared to preparations of monomeric (i.e. less concentrated) protein. It must also be pointed out that owing to a lack of points at higher temperatures the goodness of fit of the near-UV CD data is poorer than for the far-UV CD data and therefore the measured difference in T_m may be smaller than suggested.

The measured transition of GF from "folded" to "unfolded" states resulted in a majority of data points within the transition region that are at lower temperatures for both

SDS solubilized GF measured by far-UV CD (see Figures 3-30) and DM-solubilized GF measured by near-UV (see Figure 3-32). This may indicate that the measured transition for SDS-solubilized GF is not a cooperative unfolding, but that the ellipticity loss reflects aggregation of GF at the elevated temperature (see Figure 3-15), or that the loss of tertiary structure of GF in DM is abrupt.

The structural stability of GF in any of the detergents tested is not high. This relative instability may reflect the removal of a lipid or protein "co-factor" present in intact *E. coli* cells but lost upon extraction and purification of the facilitator, thus leading to a reduction in the structural stability of detergent-solubilized GF. The data indicate that GF may not be fully folded when solubilized in DM or that in DM the protein is less stable than in a bilayer. Indeed, the apparent absence of GF oligomers in *Xenopus* oocyte membranes [215] suggests that oligomerization is weak and raises the possibility that lipid or other components of the *E. coli* membrane may play a role in the stabilization of the protein's conformation as suggested by Truniger and Boos [216]. Further, the fraction of glycerol facilitator present as oligomer in SDS-PAGE is smaller than that reported for AQPZ [166], suggesting that the oligomerization of the bacterial facilitator is weaker than that of the bacterial water channel. That GF can be visualized, *via* SDS-PAGE, in a monomeric form at low protein concentrations in all detergents, in the absence of denaturant, agrees with the work of Lagree and colleagues who show that GF is monomeric in detergent [185, 215]. Furthermore, Borgnia and Agre [217] recently showed that GF can be isolated in monomeric form by sedimentation of GF solubilized in OG directly from bacterial membranes, whereas GF behaves as a multisubunit oligomer when studied by selective adsorption and elution from Ni^{2+}-NTA. They conclude that GF exists in multiple oligomeric states, that GF tetramers

are stabilized while in the membrane and during affinity purification, and that the affinity of the GF monomer-dimer-tetramer association is lower than that for AQPZ. Thus, our observations of the appearance of GF oligomers under a variety of conditions, with tetramer representing the highest molecular weight species in non-denaturing situations, suggests that it is not only the detergent that reduces GF structural stability, but that GF tends to form weakly associated oligomers with the predominant species being tetrameric.

4.3.2 Membrane Embedded GF

We originally proposed that the state of GF *in vivo* is tetrameric, based on our observations of tetramers on SDS-PAGE and on the data acquired on AQP1 [83]. In order to investigate the possibility that the native state of GF is tetrameric, a cross-linking analysis experiment was performed. Cross-linking was achieved using the lipophilic cross linkers DSS and DSG, and the hydrophilic cross-linker BS. Each cross-linking study indicated the appearance of dimer, trimer, and tetramer within the first 30 minutes of cross-linking (see Figure 3-11). The rapid appearance of GF tetramer and the absence of higher-molecular weight species, even after 24 hours, is an indication that cross-linking is specific [218]. If cross-linking were non-specific the immuno-detected monomer, dimer, trimer, and tetramer bands would either not have been well resolved and appeared as multiplets, or would not have been differentiated from the background, due to the cross-linking of molecules of various sizes to GF. The cross-linking data, therefore, suggest that GF exists as a dimer or tetramer in its native membrane state. This conclusion is in good agreement with the recently published paper by Fu *et al.* [110] which shows that GF crystallizes as a tetramer. It also agrees Borgnia and Agre [217] who showed that GF can be extracted from the membrane in a tetrameric state.

4.3.3 Is GF Structure and Function Related to External Glycerol Concentrations and Glycerol Kinase Activity?

The data presented above suggest that the oligomeric state of GF is a weakly associated tetramer in its native membrane environment. What might be the evolutionary reason for this? It has been proposed that oligomerization of porin proteins as trimers [211] permits the formation of a large inter-molecular hydrophobic core that stabilizes the protein fold and this advantage may also be applicable to the GF tetramer. However, if the protein is more stable as a tetramer than a monomer, why has it not evolved into a stable tetrameric structure? Why is it advantageous for GF to form weakly associated tetramers? The answer may be related to an association between the glycerol facilitator and its genetic neighbour, glycerol kinase (GK). After entry into the *E. coli* cytoplasm, glycerol is phosphorylated by the glycerol kinase (GK) [95] to *sn*-glycerol-3-phosphate (G3P) and is then available for use as a carbon source. GK shows complex regulatory behaviour and is functional in both a dimeric and tetrameric form [102]. Fructose 1,6-bisphosphate and the phosphoenolpyruvate phosphotransferase protein IIA^{Glc} are both allosteric inhibitors of GK, and bind at different sites [103, 104]. Two molecules of fructose 1,6-bisphosphate bind to the GK tetramer locking it in the tetrameric conformation in an inactive state [103], while the phosphocarrier protein IIA^{Glc} binds to each monomer [105]. There has been considerable speculation and some kinetic evidence that GK exists in close association with GF at the inner membrane and that the kinase may be regulated by interaction with the facilitator (see Section 1.4.2) [100]. This has been compared to the interactions of hexokinase and glycerol kinase with their respective mitochondrial porins [101]. If the facilitator and kinase do interact at the membrane, then the changing oligomeric state of GK could have an influence on the oligomeric state of the facilitator, or conversely GF may exert some influence on GK

oligomeric state. Assuming a 1:1 interaction between the two proteins, then both the dimer and tetramer states of GK could have access to glycerol from the associated GF dimer and tetramer, respectively (see Figure 4-1). This would not be possible if GF formed very

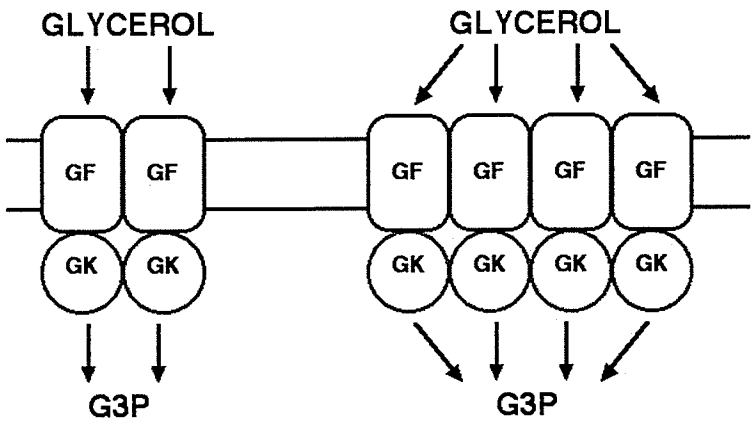


Figure 4-1. Diagrammatic representation of the possible GK-GF tetramer-dimer association.

strong tetramers or if GF could not oligomerize at all, and may explain why GF oligomers are only observed in cross-linking studies or at high protein concentrations. Thus, the weakness of the self-association of GF, in comparison to some of the aquaporins, may reflect the fact that some oligomerization determinants (such as GK or native membrane lipids as mentioned above) are absent from detergent preparations.

CHAPTER 5

FUTURE WORK

Ideally, the structural analysis of GF would be performed using techniques that can provide information at atomic resolution. Techniques such as NMR and X-ray diffraction are commonly used to provide this level of information for water-soluble proteins, but are more difficult to perform on membrane proteins. Detergent micelles often lead to line-width broadening and reduction of spectral resolution in NMR analysis [139], and in crystallography many attempts to form 3-dimensional crystals of membrane proteins have resulted in protein precipitation [27]. One goal of this work was to purify glycerol facilitator for NMR analysis, however early results suggested that GF tended to form high molecular weight oligomers which would compound the line-width problem in NMR analysis. Thus, MALDI-TOFMS, circular dichroism, and chemical cross-linking were chosen as initial, low-resolution, biophysical analytical techniques.

The NMR analysis of detergent-solubilized GF, with the goal of structure determination at atomic resolution, is not feasible at this point in the research project. The oligomeric state of GF, and conditions affecting it, are not yet fully understood. What is known is that conditions such as increasing the sample temperature (see Figure 3-16) or the addition of urea (see Figures 3-13 and 3-14) cause immediate denaturation and aggregation. Also evident is that the choice of detergent affects the oligomeric state, such that OG-solubilized GF aggregates upon perturbation (see Figure 3-11), GF is in a molten globule-like state in SDS, and DM-solubilized GF is in a native state. However, GF does form oligomers at the protein concentrations required for NMR analysis when solubilized in both SDS and

DM (see Figure 3-15). Therefore, the conditions promoting stable monomer formation must be determined and understood before NMR structural analysis can be started.

5.1 Suggestions for Future Work

There are many avenues open to advance this research project. Initially, steps to improve the expression levels should be examined. Evidence exists that upon induction of protein overexpression in the pET system, the mRNA levels of the recombinant protein increase at a faster rate than native mRNA. As a result, the newly transcribed mRNA is often left unprotected by ribosomes, and is rapidly degraded by endonuclease RNaseE, indirectly reducing expression levels [204]. A new strain of *E. coli* has recently been made available from Invitrogen that has a mutation in the *rne* gene, that encodes RNaseE. The strain, BL21 Star™(DE3) with and without pLysS, has been shown to increase protein yields from two- to ten-fold for 70% of the proteins tested [220]. Whether this would increase the yield of GF is not known. It is quite possible that the addition of rifampicin to the growing culture in the optimized protocol is indirectly increasing the level of successful translation by stopping production of native RNaseE. However, since increasing the incubation time after rifampicin addition does not increase GF yield, inhibiting bacterial protein synthesis may be detrimental to GF expression. This limitation might be removed using the BL21 Star™(DE3)pLysS strain, leading to further simplification and enhancement of the GF expression protocol.

There is also some room for improvement in the procedure for extraction of GF from the *E. coli* membrane. The results have shown that the length of the detergent acyl chain influences the stability of GF in solution. My data suggest that the 8 carbon chain of OG is too short to keep GF from denaturing and aggregating even under very gentle con-

ditions, i.e. slow dialysis to remove OG in the presence of phospholipid vesicles. This reflects the lack of a sufficiently hydrophobic environment in which the solubilized GF can be maintained. We found GF to be stable in DM but we did not explore detergents with longer acyl chains, such as tetradecyl-maltoside (TM), for the extraction and structural analysis of GF. An additional benefit of this detergent may be a slight reduction in the micelle size and hence the micelle weight, since the CMC for TM is one-tenth that of DM and the monomer mass difference is only 28 Da and the aggregation number and micelle mass would thus be lower (see Section 1.1.2.1). This would be advantageous for NMR analysis, as the weight reduction of the protein/detergent complex would improve spectral resolution (by decreasing the correlation time) [221].

The xylitol-transport assay of whole cells induced to express GF has shown that the recombinant protein is functionally identical to native GF. It is not known, however, whether the detergent extracted and IMAC purified GF remains in a functional state. To establish that detergent purified GF is in a native state, and that the structural data presented in this thesis represents the native structure, it is necessary to reconstitute the purified protein into liposomes and assay the ability of the proteoliposomes to equilibrate a xylitol gradient. The detergents preferred for such studies are those with high CMCs, as complete removal of the detergent by dialysis would be possible. The production of GF proteoliposomes was attempted using OG solubilized GF, but as described earlier GF is not very stable when solubilized in OG and rapidly aggregates. However, another detergent often used for membrane protein purification and reconstitution is the bile acid sodium cholate [222]. The purification and reconstitution of GF should be attempted in sodium cholate and those preparations used to reconstitute the protein into proteoliposomes, in an effort to

show that GF can be purified in its native state. Acquisition of far- and near-UV CD spectra of GF in proteoliposomes would aid assessment of the state of the protein in DM and other detergents.

Recombinant GF has been demonstrated to be functionally identical to that of wild-type GF. Furthermore, the quaternary structure of native, membrane-embedded GF has been shown to be tetrameric. There is considerable evidence that the AQP1 tetramer represents the functional unit of the protein [82-84, 223], and also that the activity of GK is related to GF, but it is not known if the functional state of GF is the tetramer. Because cross-linking GF in the native membrane has provided the evidence that GF can form a tetrameric species, further analysis of the cross-linked protein may provide an insight into the quaternary state of GF, indicating whether GF is still functional after the tetramer has been fixed by cross-linking. In other words, it may be possible to prepare membrane vesicles containing GF from crude membrane preparations, to cross-link the GF, and to assay the ability of the vesicles to equilibrate a xylitol gradient.

A further extension to the cross-linking studies would be to cross-link GF in the detergent micelle. This could provide a view into the quaternary state of the protein in detergent. Thus far, all inferences about the quaternary state have been based on SDS-PAGE analysis and cross-linking in the native membrane, and have provided evidence that GF can form tetramers in detergent and is tetrameric in the membrane. The SDS-PAGE gels reveal what the state of GF is after the gel sample has been prepared and electrophoresed in the presence of SDS. Cross-linking GF in the detergent micelle prior to electrophoresis, in the absence of any denaturing agent, may provide an indication of the quaternary state of detergent-solubilized GF. These cross-linking studies combined with a gel filtration

study should allow the separation of micelles or proteoliposomes containing high molecular weight forms of GF from monomers.

The secondary structure of GF has been determined using circular dichroism, and the results agree well with the secondary structural elements observed in the X-ray crystal structure [110]. It would be advantageous to confirm our assessments of secondary structure by CD with those obtained by infrared (IR) spectroscopy. There are well characterized relationships between the position of the amide I bands in an IR spectrum and the type of secondary structure [119] making IR analysis an attractive technique for confirmation of GF secondary structure in detergent.

It is interesting to note that neither the CD analysis of GF nor the crosslinking was performed in the presence of glycerol. As the natural substrate of GF, it would seem reasonable that the addition of glycerol to the buffer during structural analysis may have an affect on monomer structure, and may provide a means of stabilizing GF structure. It is doubtful, however, that glycerol would have any affect on GF quaternary structure, as the functional unit of GF is the monomer.

There have been several advances in the study of membrane proteins by NMR in recent years. Advances in ^{15}N, ^{13}C and ^{2}H labeling provide hope for the analysis of larger proteins and protein complexes. The development of 3-dimensional and 4-dimensional heteronuclear NMR methods have extended the mass range for structure determination of proteins by NMR spectroscopy to 48 kDa [224-226]. Random fractional deuteration of proteins uniformly labelled with ^{15}N and ^{13}C reduces heteronuclear relaxation rates so that high quality spectra can be obtained for proteins in the range of 30-40- kDa [139, 227]. This mass range is sufficient for rapid determination of the overall protein folding, and is a rea-

sonable approach for the NMR analysis of the 33 kDa GF monomer in detergent micelles. Thus, in principle, NMR methods are available for the study of detergent solubilized GF. However, it is important that the propensity for GF to form different oligomeric species be addressed first (see above).

It may also be possible to take advantage of the rifampicin inhibition of bacterial protein synthesis for the preparation of isotopically labeled GF. Labelling recombinant proteins with ^{15}N or ^{13}C can be costly, since the isotopes must be added to the growing culture in the form of expensive isotopically labelled nutrients such as ^{13}C-glucose or ^{15}N-glycine. Further the addition of the isotopically labelled nutrients results in the incorporation of the isotopes into the recombinant protein, but does not stop their incorporation into bacterial proteins. The addition of the labelled nutrients after the inhibition of bacterial protein synthesis with rifampicin may provide a means of specifically targeting the recombinant protein for isotope labelling.

The solution of the AQP1 and GF crystal structures, by cryoelectron and X-ray crystallography respectively, has shed light on the selectivity and transport mechanisms in the MIP family [66, 110, 209a]. It is somewhat surprising, however, that their channel structures are amazingly similar, although the channel specificities are quite different. Both structures show the half-helices HB and HE[1] meeting in the center of the channel at a point of constriction (3 Å in AQP1 and 3.5 Å in GF), and that the helix dipoles generate a positive electrostatic field because the N-termini of both helices are in the center of the channel. In both cases the channels present hydrogen bond donors and acceptors on one side and hydrophobic residues on the other, forcing strict orientation of the permeant molecules, *i.e.*

1. I have chosen to use the nomenclature from Murata *et al.* [66]

the oxygens of the water molecules face the hydrogen bond donors and the water hydrogens face the hydrophobic wall, and the hydroxyls of the glycerol molecules face the hydrogen bonding residues forcing the alkyl backbone to contact the hydrophobic wall. Early work on the pore specificities of the MIP superfamily suggested that the pore diameter may be the major selective mechanism employed by the family [70], and since the AQP1 [66] and GF [110] crystal structures show no significant differences other than their pore diameters, this suggestion has become an attractive theory for further investigation. As the pore diameter of GF is greater than that of AQP1, by 0.5 Å, it may be possible to switch the pore specificity from glycerol to water by increasing the size of the residues flanking the GF pore. This would provide evidence that the pore diameter, rather than the specific residues lining it, is responsible for selectivity in the MIP family of proteins.

References

[1] Gennis, R. B., (1989) *Biomembranes: Molecular Structure and Function*, Springer-Verlag, New York.

[2] Tanford, C. (1973) *Hydrophobic Effect*, Wiley, New York.

[3] Hauser, H., Pascher, L., Pearson, R.H., Sundell, S., (1981) *Biochim. Biophys. Acta. 650*, 21-51

[4] Jain, M.K., (1988) *Introduction to Biological Membranes*, Wiley, New York.

[5] Trauble, H., and Eibl, H, (1974) *Proc. Natl. Acad. Sci USA 71*, 214-219.

[6] Tanford; Dill, K.A., and Flory, P.J, (1981) *Proc. Natl. Acad. Sci USA 78*, 676-680.

[7] Israelachvili, J.N., Marcelja, S., and Horn, R.G., (1980) *Quart Rev. Biophys. 13*, 121-200.

[8] Hauser, H., Pascher, I., and Sundell, S., (1981) *Biochim. Biophys. Acta 650*, 21-51.

[9] Gorter, E, and Grendel, F., (1925) *J. Exp. Med. 41*, 439-443.

[10] Danielli, H., and Davson, J.F., (1935) *J. Cell. Comp. Physiol. 5*, 595-610.

[11] Persson, B., and Argos, P., (1996) *Protein Science 5*, 363-371.

[12] Melchoir, D.L., (1970) *Biochim. Biophys. Acta. 219*, 114.

[13] Singer, S.J., and Nicolson, G.L. (1972) *Science 175*, 720-731.

[14] Finkelstein, A. (1987) *Water movement through lipid bilayers, pores, and plasma membranes*. Wiley, New York.

[15] Cantor, C.R., and Schimmel, P.R., (1995) *Biophysical Chemistry, Part III: The Behavior of Biological Macromolecules*, W. H. Freeman & Company, New York.

[16] Agre P., Mathai, J.C., Smith, B.L., Preston, G.M., (1999) *Methods Enzymol. 294*, 550-572.

[17] Moura, T.F., Macey, R.I., Chien, D.Y., Karan, D., Santos, H., (1984), *J. Memb. Biol 81*, 105-111.

[18] Braum, J., (1982). *J. Gen. Physiol. 79*, 791-819.

[19] Devaux, P.F., and Seigneuret, M., (1985) *Biochim. Biophys. Acta. 822*, 63-125

[20] Harris, H.W. Jr., Zeidel, M.L., Jo. I., Hammond. T.G., (1994) *J. Biol. Chem. 269*, 11993-20000.

[21] Clapham, D.E., (1999) *Nat Struct Biol. 6*, 807-810.

[22] Unger, V.M., (2000) *Nat. Struct. Biol. 7,* 1082-1084.

[23] Brown, E.M., (2000) *Cell Biochem Biophys. 33*, 63-95.

[24] Musatov, A., Ortega-Lopez, J., Robinson, N.C., (2000) *Biochemistry 39*, 12996-13004.

[25] Hendrickson, W.A., (1996) *J. Bioenergetics and Biomembranes, 28*, 35-39.

[26] Helenius, A., and Simons, K., (1975) *Biochim. Biophys. Acta. 415*, 29-79.

[27a]Hasler, L., Heymann, J.B., Engel, A., Kistler, Joerg, K., and Walz, T., (1998) *J. Struc. Biol. 121,* 162-171.

[27b] Yue, L., Peng, J-B., Hediger, M.A., Clapham, D.E., (2001) *Nature 410*, 705-709.

[28] Cowan, S.W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R.A., Jansonius, J.N., and Rosenbusch, J.P., (1992) *Nature 358*, 727-733.

[29] Schirmer, T., Keller, T.A., Wang, Y-F., and Rosenbusch, J.P., (1995) *Science 267,* 512-514.

[30] Bhairi, S, M., (2001) *Detergents: A Guide to the Properties and Uses of Detergents in Biological systems*, Calbiochem.

[31] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., (2000) *Nucleic Acids Research 28*, 235-242.

[32] Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H., (1990) *J. Mol. Biol. 213,* 899-929.

[33] Jap, B.K., Walian, P.J., and Gehring, K., (1991) *Nature 350*, 167-170.

[34] Kühlbrandt, W., Wang, D.N., and Fujiyoshi, Y., (1994) *Nature 367,* 614-621.

[35] Yeager, M., Unger, V.M., Mitra, A.K., (1999) *Methods Enzymol. 294*,135-80.

[36] Hammond, C., (2001) *The Basics of Crystallography and Diffraction,* Oxford University Press, New York.

[38] Chiu, W., and Schmid, M.F., (1992) *Curr. Op. Biotech. 4*, 397-402.

[39] Dorset, D., (1996) *Acta. Cryst. B52*, 753-769.

[40] Brisson, A., Olofsson, A., Ringler, P., Schmutz, M., and Stoylova, S., (1994) *Biology of the Cell 80*, 221-228.

[41] Kistler, J and Walz, T., (1998) *J. Struc. Biol., 121*,162-171.

[42] Unwin, N., and Henderson, R., (1984) *Sci Am. 250*, 78-94.

[43] Macey, R.I., and Farmer, R.E.L., (1970) *Biochimica Biophysica Acta 211*, 104-106.

[44] Koefoed-Johnsen, V., and Ussing, H.H., (1953) *Acta Physiol. Scand., 28*, 60-76.

[45] Wintour, E.M., (1997) *Clinical and Experimental Pharmacology and Physiology 24,* 1-9.

[46] Lee, M.D., King, L.S., and Agre, P., (1997) *Medicine 746*, 141-156.

[47] Moura et al., (1984) *J. Memb. Biol 81,* 105-111.

[48] Braum, J., (1982) *J. Gen. Physiol. 79*, 791-819

[49] Denker, B.M., Smith, B.L., Kuhajda, F.P., and Agre P., (1988) *J. Biol. Chem. 263*, 15634-15642.

[51] Smith, B.L., and Agre, P., (1991) *J Biol Chem 266*, 6407-6415.

[52] Preston, G.M., and Agre, P., (1991) *Proc. Natl. Acad. Sci USA 88*, 11110-11114.

[53] Preston, G.M., Carroll, T.P., Guggino, W.B., and Agre, P., (1992) *Science 256,* 385-387.

[54] Preston, G.M., Jung, J.S., Guggino, W.B., and Agre, P., (1993) *J. Biol. Chem. 268* : 17-20.

[55] Zeidel, M.L., Ambudkar, S.V., Smith, B.L., and Agre, P., (1992) *Biochemistry 31*, 7436-7440.

[56] Van Hoek, A.N., and Verkman, A.S., (1992) *J. Biol. Chem. 267*, 18267-18269.

[57] Agre, P., Sasaki, S., and Chrispeels, M.J., (1993) *Am J. Physiology 265*, F461.

[58] Borgnia, M., Nielson, S., Engel, A., and Agre P. (1999) *Ann. Rev. Biochem. 68*, 425-458.

[59] Yang. B. and Verkman, A. S. (1997) *J. Biol. Chem. 272*, 16 410-16416.

[60] Yamamoto, T. and Sasaki, S. (1998) *Kidney Int. 54*, 1041-1051.

[61] Park, J.H., and Saier, M.H. Jr., (1996) *J. Memb. Biol. 153*, 171-180.

[62] Gorin, M.B., Yancey, S.B., Cline, J., Revel, J-P., and Horwitz, J., (1984) *Cell 39*, 49-59.

[63] Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105-132.

[64] Lee, M.D., King, L.S., and Agre, P., (1997) *Medicine 746,* 141-156.

[65] Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J., Stroud, R.M., (2000) *Science 290*, 481-486.

[66] Murat,a K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., Fujiyoshi, Y., (2000) *Nature 407*, 599-605.

[67] Lin E.C.C., (1976) *Ann. Rev. Microbiol 30*, 535-578.

[68] Ishibashi, K., Kuwahara, M., Gu, Y., Kageyama, Y., and Tohsaka, A. (1997) *J. Biol. Chem. 272*, 20782-20786.

[69] Tsukaguchi, H., Shayakul, C., Berger, U. V., Mackenzie, B., Devidas, S., Guggino, W. B., van Hoek, A. N., and Hediger, M. A. (1998) *J. Biol. Chem. 273*, 24737-24743.

[70] Meinild, A-K., Klaerke, D. A., and Zeuthen, T. (1998) *J. Biol. Chem. 273*, 32446-32451.

[71] Horwitz, J. and Bok, D. (1987) *Biochemistry 26*, 8092-8098.

[72] Maunsbach, A.B., Marples, D., Chin, E., Ning, G., Bondy, C., Agre, P., and Nielen, S., (1997) *J. Am. Soc. Neph. 8,* 1-14.

[73] Ishibashi, K., Sasaki, S., Fushimi, K., Uchida, S., Kuwahara, M., Saito, H., Furukawa, T., Nakajima, K., Yamaguchi, Y., Gojobori, T., and Marumo, F., (1994) *Proc. Natl. Acad. Sci USA 91*, 6269-6273.

[74] Frigeri, A., Gropper, M.A., Turck, C.W., and Verkman, A.S., (1995) *Proc. Natl. Acad. Sci USA 92,* 4328-4331.

[75] Ma, T., Song, Y., Gillespie, A., Carlson, E.J., Epstein, C.J., and VBerkman, A.S., (1999) *J. Biol. Chem. 274,* 20071-20074.

[76] Yasui, M., Kwon, T-H., Knepper, M.A., Nielson, S., and Agre, P., (1999) *Proc. Natl. Acad. Sci USA 96*, 5808-5813.

[77] Garcia, F., Kierbel, A., Larocca, M.C., Gradilone, S.A., Splinter, P., LaRusso, N.F., Marinelli, R.A., (2001) *J Biol Chem 276*, 12147-12152.

[78] Ko, S.B.H., Uchida, S., Naruse, S., Kuwahara, M., Ishibashi, K., Marumo, F., Hayakawa, T., and Sasaki, S., (1999) *Biochem. and Mol. Biol. Int. 47,* 309-318.

[79] Preston, G.M., Jung, J.S., Guggino, W.B., and Agre, P., (1994) *J Biol Chem 269*, 1668-1673.

[80] Preston, G.M., Carroll, T.P., Guggino, W.B., and Agre, P., (1992) *Science 256*, 385-387.

[81] Jung, J.S., Preston, G.M., Smith, B.L., Guggino, W.B., and Agre, P., (1994) *J Biol Chem 269*, 14648-14654.

[82] Smith, B.L., and Agre, P., (1991) *J Biol Chem 266*, 6407-6415.

[83] Verbavatz, J-M., Brown, D., Sabolic, I., Valenti, G., Ausiello, D.A., Van Hoek, A.N., Ma, T., and Verkman, A.S.,(1994) *J. Cell. Biol. 123*, 605-618.

[84] Van Hoek, A.N., Hom, M.L., Luthjens, L.H., de Jong, M.D., Dempster, J.A., and Van Os, C.H., (1991) *J Biol Chem 266*, 16633-16635.

[86] Li, H., Lee, S., and Jap, B.K., (1997) *Nature Struct. Biol. 4*, 263-265.

[87] Walz, T., Hirai, T., Murata, K., Heymann, J.B., Mitsuoka, K., Fujiyoshi, Y., Smith, B.L., Agre, P., and Engel, A., (1997) *Nature 387,* 624-627.

[88] Cheng, A., Van Hoek, A.N., Yeager, M., Verkman, A.S., and Mitra, A.K., (1997) *Nature 387*, 627-630.

[89] MacKenzie, K.R., and Engelman, D.M., (1998) *Proc. Natl. Acad. Sci USA 95,* 3583-3590.

[90] Nanninga, N., (1998) *Microbiol Mol Biol Rev. 62*, 110-29.

[91] Bertrand, G. (1898). *C.R. Acad. Sci.* 126:984-986.

[93] Bovell, C.R., Packer, L., Helgerson, R. (1963) *Biochim. Biophys. Acta.* 75:257-266.

[94] Hayashi, S.-I., and Lin, E.C.C. (1965) *Biochim. Biophys. Acta.* 94:479-487.

[95] Lin, E.C.C., (1984) *In* E. Haber (ed), The cell membrane. Its role in interaction with the outside world. Plenum Publishing Corp., New York.

[96] Sanno, Y., Wilson, T.H., and Lin, E.C.C., (1968) *Biochem. Biophys. Res. Commun.,* 32:344-349.

[97] Richey, D.P. and Lin, E.C.C. (1972) *J. Bact.* 112:784-790.

[98] Sweet, G., Gandor, C., Voegele, R., Wittekindt, N., Beuerle, J., Truniger, V., Lin, E.C., and Boos, W., (1990) *J Bacteriol 172*, 424-430.

[99] Lin, E.C.C., and Luchi, S., (1991) *Ann. Rev. Genet.* 35:361-387

[100] Voegele, R. T., Sweet, G. D., and Boos, W. (1993) *J. Bacteriol. 175*, 1087-1094.

[101] Brdiczka, D. (1990) *Experientia 46*, 161-167.

[102] De Riel, J. K., and Paulus, H. (1978) *Biochemistry 17*, 5141-5145.

[103] Ormo, M., Bystom, C. E., and Remington, S. J. (1998) *Biochemistry 37*, 16565-16572.

[104] Bystrom, C. E., Pettigrew, D. W., Branchaud, B. P., O'Brien, P., and Remington, S. J. (1999) *Biochemistry 38*, 3508-3518.

[105] Hurley, J. H., Faber, H. R., Worthylake, D., Meadow, N. D., Roseman, S., Pettigrew, D., and Remington, S. J., (1993) *Science 259*, 673-677.

[106] Heller, K.B., Lin, E.C.C., and Wilson, T.H. (1980) *J. Bact. 144*, 274-278.

[107] Maurel, C., Reizer, J., Schroeder, J.I., Chrispeels, M.J., and Saier, M.H. (1994) *J. Biol. Chem. 269*, 11869-11872.

[108] Calamita, G., Bishai, W.R., Preston, G.M., Giggino, W.B., and Agre, P. (1995) *J. Biol. Chem. 270*, 29063-29066.

[109] Sanders, O.I., Rensing, C., Kuroda, M., Mitra, B., and Rosen, B. (1997) *J. Bact. 179*, 3365-3367.

[110] Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski. J., Stroud, R.M., (2000) *Science 290*, 481-486.

[111] Manley, D.M., McComb, M.E., Perreault, H., Donald, L.J., Duckworth, H.W., O'Neil, J.D., (2000) *Biochemistry 39*, 12303-12311.

[112] Kabsch, W., and Sander, C., (1983) *Biopolymers 22*, 2577-2637

[113] Campbell, I.D., and Dwek, R.A., (1984) *Biological Spectroscopy*, The Benjamin/ Cummings Publishing Company, Inc., California.

[114] Cantor, C.R., and Schimmel, P.R., (1995) *Biophysical Chemistry, Part II: Techniques for the Study of Biological Structural and Function*, W. H. Freeman & Company, New York.

[115] Van Holde, K.E., Johnson, W.C., Ho, P.S., (1998) *Physical Biochemistry*, Prentice Hall, New Jersey.

[116] Koch A.L., (1961) *Biochemica et biophysica acta 51*, 429-441.

[117] Koch & Yoshikawa, W., Akutsu, H., and Kyogoku, Y., (1983) *Biochemica et biophysica acta 735*, 397-406

[118] Freifelder, D., (1976) *Physical Biochemistry: Applications to Biochemistry and Molecular Biology*, W. H. Freeman & Company, San Francisco.

[119] Havel, H.A., (Ed), (1996) *Spectroscopic Methods for Determining Protein Structure in Solution*, VCH Publishers, New York.

[121] Fasman, G.D., (Ed.), (1996) *Circular Dichroism and the Conformationsl Anaylsis of Biomolecules*, Plenum Press, New York.

[122] Johnson, W.C. Jr., (1985) *Methods Biochem Anal. 31*, 61-163.

[124] Garret, C.M., and Grisham, R.H., (1999) *Biochemistry*, Saunders College Publishing, New York.

[125] Yang, J.T., Wu, C.S., Martinez, H.M., (1986) *Methods Enzymol. 130*, 208-269.

[126] Greenfield, N.J., and Fasman, G.D., (1969) *Biopolymers 7*, 595-610.

[127] Sarkar, P.K., and Doty, P., (1966) *Proc. Natl. Acad. Sci USA 55*, 981-989.

[128] Townend, R., Kumosinki, T.F., Timasheff, S.N., Fasman, G.D., and Davidson, B., (1966) *Biochem. Biophys. Res. Commun. 23,* 163-169.

[129] Park, K., Perczel, A., and Fasman, G.D., (1992) *Protein Science 1*, 1031-1049

[130] Tinoco, I., Woody. R.W., and Bradley, D.F., (1963) *J. Chem. Phys. 38*, 1317-1325.

[131] Perczel, A., Hollósi, M., Tusnady, G., and Fasman, G. D., (1991) *Protein Eng. 4*, 669-679.

[132] Creighton, T.E., (1992) *Proteins : Structures and Molecular Properties*, W. H. Freeman & Company, New York.

[133] Horwitz, J., Strickland, E.H., and Billups, C., (1969) *J. Am. Chem. Soc. 91*, 184-190.

[134] Strickland, E.H., (1972) *Biochemistry 11*, 3465-3474.

[136] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E., (2000) *Nucleic Acids Research 28,* 235-242

[137] Grisshammer, R., and Tate, C.G., (1995) *Quarterly Rev. Biophysics 28*, 315-422.

[138] Garavito, R.M., Picot, D., and Loll, P.J., (1996) *J. Bioenergetics and Biomembranes 28*, 13-27.

[139] Kay, L.E., (1997) *Biochem. Cell Biol. 75*, 1-15.

[140] Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H., (1990) *J. Mol. Biol. 213,* 899-929.

[141] Ermler, U., Fritzsch, G., Buchanan, S.K., and Michel, H., (1994) *Structure 2*, 925-936.

[142] Freer, A., Prince, S., Sauer, K., Papiz, M., Hawthornthwaite-Lawless, A., McDermott, G., Cogdell, R., and Isaacs, N.W., (1995) *Structure 4*, 449-462.

[143] Privé, G.G., and Kaback, H.R., (1996) *J. Bioenerg. and Biomemb. 28*, 29-34.

[144] Saito, H. and Miura, K.-I. (1963) *Biochim. Biophys. Acta* 72, 619-29.

[145] Mullis, K. and Faloona, F. (1987) *Meth. Enzymol. 155*, 335-50.

[146a]Muramatsu, S. and Mizuno, T. (1989) *Nucl. Acids Res. 17,* 4378.

[146b]Sambrook, J., Fritsh, E.F., and Maniatis, T., (1989) *Molecular Cloning: A Laboratory Manual*, Coldspring Harbor Press.

[147] Willis, E.H., Mardis, E.R., Jones, W.L., and Little, M.C., (1990) *Biotechniques 9*, 92-99.

[148] Clark, J. M. (1988). *Nucl. Acids Res. 16*, 9677-9686.

[149] Pope, B., and Kent, H. M., (1996) *Nucl. Acid Res. 24*, 536-537.

[150] Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Meth. Enzymol. 185*, 60-89.

[151] Hochuli, E., Dobeli, H., and Schacher, A. (1987) *J. Chromatography 411*, 177-184.

[152] Novagen (1997) *pET system manual*, 7th edition.

[153] Studier, F. W., and Moffatt, B. A., (1986). *J. Mol. Biol. 189*, 113-130.

[154] Kim, K-S., and Pallaghy, C. K. (1996) U.S. Dept. Commerce / NOAA / NMFS / NWFSC / Molecular Biology Protocols (http://research.nwfsc.noaa.gov/protocols/dna-prep.html.).

[155] Kunkel, T. A., Bebenek, K., and McClary, J. (1991) *Meth. Enzymol. 204*, 125-139.

[156] Promega (1996) *Protocols and Applications Guide*, 3rd edition.

[157] Russel, M., Kidd, S., and Kelley, M. R., (1986) *Gene 45*, 333-338.

[158] Laemmli, U. K. (1970) *Nature 227*, 680-685.

[159] Tartof, K.D., and Hobbs, C.A., (1987) *Bethesda Res. Lab. Focus 9,* 12.

[160] Kurucz, I., Jost, C.R, George, A.J., Andrew, S.M, and Segal, D.M. (1993) *Proc. Natl. Acad. Sci USA 90*, 3830-3834.

[161] Kuderova, A., Nanak, E., Truksa, M., Brzobohaty, B. (1999) *Protein Expr. Purif. 16*, 405-409.

[162] Sagne, C., Isambert, M-F., Henry, J-P., and Gasnier, B. (1996) *Biochem. J. 316*, 825-831.

[163] Fiermonte, G., Walker, J.E., and Palmieri, F., (1993) *Biochem. J. 294,* 293-299.

[164] McComb, M. E., Oleschuk, R. D., Manley, D. M., Donald, L., Chow, A., O'Neil, J. D., Duckworth, H. W., Ens, W., Standing, K. G., and Perreault, H., (1997) *Rapid Commun. Mass Spectrom. 11*, 1716-1722.

[165] Tang, X. Beavis, R. C., Ens, W., Lafortune, F., Schueler, B., and Standing, K. G. (1988) *Int. J. Mass Spec. Ion Proc. 85*, 43-49.

[166] Borgnia, M. J., Kozono, D., Calamita, G., Maloney, P. C., and Agre P. (1999) *J. Mol. Biol. 291*, 1169-1179.

[167] Wong, S.S., (1991) *Chemistry of Protein Conjugation and Cross-linking*, CRC Press.

[168] Wu, C.-S.C., and Chen, G.C., (1989) *Analytical Biochemistry 177*, 178-182.

[169] Johnson, W.C., Jr., (1990) *Proteins: Structure, Function, Genetics 7*, 205-214

[170] Hennessey, J. P. Jr., and Johnson, W. C., Jr., (1982) *Anal. Biochem. 125*, 177-188.

[171] Chen, G.C., and Yang, J.T., (1977) *Anal. Lett. 10*, 1195-1207.

[172] Shalongo, W., Dugad, L., and Stellwagen, E., (1994) *J. Am. Chem. Soc. 116*, 2500-2507

[173] Moffatt, B.A., and Studier, F.W., (1987) *Cell 49*, 221-227.

[174] Studier, F.W., (1991) *J. Mol. Biol. 219*, 37-44.

[175] Werhli, W., Knüsel, F., Schmid, K and Staehelin, M., (1968) *Proc. Natl. Acad. Sci USA . 61*, 667-673.

[176] Kaplan, R. S., J. (1996) *Bioenergetics and Biomembranes 28,* 41-47.

[177] Loddenkotter, B., (1993) *Proc. Natl. Acad. Sci USA 90*, 2155-2159

[178] Waeber, U., Buhr, A., Schunk, T., and Erni, B., (1993) *FEBS. 324*, 109-112.

[179a]Koppel, D.A., Kinnally, K.W., Masters, P., Forte, M., Blachly-Dyson, E., and Mannella, C.A., (1998) *J. Biol. Chem. 273*, 13794-13800.

[179b]Qiagen Inc. (2000) The QIAexpressionist 4th Edition: *A Handbook for High-Level Expression and Purification of 6xHis-Tagged Proteins.*

[180] Li, Z., and Deber, C. M., (1991) *Biochem. Biophys. Res. Commun. 180*, 687-693.

[181] Mingarro, I., Whitley, P., Lemmon, M. A., and Heijne, G. V. (1996) *Protein Science 5*, 1339-1341.

[182] Prikster, M., Rucker, J., Hoffman, T. L., Doms, R. W., and Shai, Y. (1999) *Biochemistry 38*, 11359-11371.

[183] Adams, P. D., Engelman, D. M., and Brünger, A. T. (1996) *Proteins: Structure, Function, and Genetics 26*, 257-261.

[185] Lagrée, V., Froger, A., Deschamps, S., Isabelle, P., Delamarche, C., Bonnec, G., Gouranton, J., Thomas, D., and Hubert, J-H. (1998) *J. Biol. Chem. 273*, 33949-33953.

[186] Helenius, A., and Simons, K., (1975) *Biochim. et Biophys. Acta 415*, 29-79

[187] Sagne, C., Isambert, M-F., Henry, J-P., and Gasnier, B. (1996) *Biochem. J. 316*, 825-831.

[188] Alemmohamad, M. M., and Knowles, C. J. (1974) *J. Gen. Micro. 82*, 125-142.

[189] Lin, E. C. C. (1976) *Annu. Rev. Microbiol. 30*, 535-578.

[190] Cohen, S. L. and Chait, B.T., (1996) *Anal. Chem. V68*, 31-37

[191] Rosinke, B., Strupat, K., Hillenkamp, F., Rosenbusch, J., Dencher, N., Krüger, U., and Galla, H-J. (1995) *J. Mass Spec. 30*, 1462-1468.

[192] Amado, F. M. L., Santana-Marques, M. G., Ferrer-Corria, A. J., and Tomer, K. B. (1997) *Anal. Chem. 69*, 1102-1106.

[193] Chapman, J.R., (Ed.), (2000) *Mass Spectrometry of Proteins and Peptides*, Humana Press, New York.

[194] Provencher, S.W., and Glockner J., (1981) *Biochem. 20,* 33-37.

[195] Eckert, B., and Beck, C.F., (1989) *J. Bacteriol. 171*, 3557-3559.

[196] Grisshammer, R., Duckworth, R., and Henderson, R., (1993) *Biochem. J. 295*, 571-576.

[197] Boubik, Y., DiBonito, P., and Jones, I.M., (1995) *Bio/Technology 13*,1079-1084.

[198] Shoji, I., Aizaki, T., Ishii, K., Chiba, T., Saito, I., Miyamura, T., and Matsuura, Y., (1997) *J. Gen. Virol. 78*, 2657-2664.

[199] Amann, E., and Brosius, J., (1985) *Gene 40*, 183-190.

[200] Knol, J., Veenhoff, L., Liang, W-J., Henderson, P. J. F., Leblanc, G., and Poolman, B., (1996) *J. Biol. Chem.. 271,* 15358-15366.

[201] Patzlaff, J. S., Moeller, J. A., Barry, B. A., and Brooker, R. J. (1998) *Biochemistry 37*, 15363-15375.

[202] Hom. L. G. and Volkman, L. E. (1998) *Biotechniques 25*, 20-22.

[203] Grodberg, J. and Dunn, J. J. (1988) *J. Bacteriol. 170*, 1245-1253.

[204] Makarova, O.V., Makarov, E.M., Sousa, R., and Dreyfus, M., (1995) *Proc. Natl. Acad. Sci USA 92*, 12250-12254.

[206] O'Neil, J.D., and Sykes, B.D., (1989) *Biochemistry 28*, 6736-6745

[207] Kaufmann, R., J. (1995) *Biotechnol. 41,* 155-175

[208] Popot, J-L., (1993) *Curr. Opin. Struct. Biol. 3*, 535-540.

[209] Mitsouka, K., Murata, K., Walz, T., Hirai, T., Agre, P., Heymann, J.B., Engel, A., and Fujiyoshi, Y., (1999) *J. Struc. Biol. 128,* 34-43.

[209a]Sui, H., Han, B-G., Lee, J.K., Wallan, P., and Jap, B.K., (2001) *Nature 414*, 872-878.

[209b]Fersht, A., (1999) *Structure and Mechanism in Protein Science : A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman and Company.

[210] Neeley, J. D., Christensen, B. M., Nielsen, S., and Agre, P. (1999) *Biochemistry 38*, 11156-11163.

[211] Jap, B.K., and Walian, P.J., (1990) *Q. Rev. Biophys. 23*, 367-403.

[212] Chen, Y.S. and Hubbel, W.L., (1973) *Exp. Eye Res. 17,* 517-532.

[213] Chabre, M., Cavaggioni, A., Osborne, H.B., and Gulik-Krzywicki, T. (1972) *FEBS Lett. 26*, 197-202.

[214] London, Y., Demel, R.A., Geurts van Kessel, W.S.M., Vossenberg, F.G.A., and Van Deenen, L.L.M., (1973) *Biochim. et Biophys. Acta 311*, 520-530.

[215] Bron, P., Lagree, V., Froger, A., Rolland, J.-P., Hubert, J.-F., Delamarche, C., Deschamps, S., Pellerin, I., Thomas, D., and Haase, W. (1999) *J. Struct. Biol. 128*, 287-296.

[216] Truniger, V. and Boos, B. (1993) *Res. Microbiol. 144*, 565-574.

[217] Borgnia, M.J., Agre, P., (2001) *Proc. Natl. Acad. Sci. USA. 98*, 2888-2893.

[218] Ji, T. H., and Middaugh, C. R., (1980) *Biochim. et Biophys. Acta 603*, 371-374.

[220] Invitrogen™ Life Technologies (2000) *One Shot® BL21Star™ (DE3) & One Shot® BL21Star™ (DE3)pLysS Competent Cells*, Version B, 010402, 25-0402.

[221] Wuthrich, K., (1998) *Nature Struc. Biol. NMR Suppliment,* 492-495.

[222] Brown, G.C., and Chris E. Cooper, C.E., (Eds.), (1995) *Bioenergetics: A Practical Approach*, Oxford University Press.

[223] Walz, T., Smith, A.L., Zeidel, M.L., Engel, A., and Agre, P., (1994) *J. Biol. Chem. 269*, 1583-1586..

[224] Goto, N.K., and Kay, L.E., (2000) *Curr. Opin. Struct. Biol. 10*, 585-592

[225] Clore, G.M., and Gronenborn, A.M., (1998) *Curr Opin Chem Biol. 2*, 564-570.

[226] Grzesiek, S., and Bax, A., (1993) *J Biomol NMR 3*, 185-204.

[227] Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C.B., and Bax, A., (1995) *Nat. Struct. Biol. 2*, 768-776.