# Structure Investigations of Membrane Protein OEP16

by

James Duncan Zook

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2012 by the
Graduate Supervisory Committee:

Petra Fromme, Chair
Julian Chen
Rebekka Wachter

ARIZONA STATE UNIVERSITY

August 2012

ABSTRACT

Membrane protein structure is continuing to be a topic of interest across the scientific community. However, high resolution structural data of these proteins is difficult to obtain. The amino acid transport protein, Outer Envelope Protein, 16kDa (OEP16) is a transmembrane protein channel that allows the passive diffusion of amino acids across the outer chloroplast membrane, and is used as a model protein in order to establish methods that ultimately reveal structural details about membrane proteins using nuclear magnetic resonance (NMR) spectroscopy. Methods include recombinant expression of isotope enriched inclusion bodies, purification and reconstitution in detergent micelles, and pre-characterization techniques including circular dichroism (CD) spectroscopy, dynamic light scattering (DLS), and high pressure liquid chromatography (HPLC). High resolution NMR spectroscopy was able to assign 99% of the amide backbone and the chemical shifts provided detailed secondary structure of OEP16 on a per residue basis using the software TALOS+. Relaxation studies explored the intramolecular dynamics of OEP16 and results strongly support the resonance assignments. Successful titration studies were able to locate residues important for amino acid binding for import into the chloroplast as well as provide information on how the transmembrane helices of OEP16 are packed together. For the first time there is experimental evidence that can assign the location of secondary structure in OEP16 and creates a foundation for a future three dimensional structure.

# DEDICATION

The work presented in this dissertation is dedicated to my daughter, Violet Catherine Zook, born on October 12th, 2011. Your strength helps motivate me to continue learning and growing and never-ever giving up. This work is also dedicated to my loving wife, Elizabeth Wolf Zook. Without her encouragement, I may have not been able to complete this work. You have helped me through so many of life's challenges and I cannot thank you enough for being there for me. Finally, I would also like to dedicate this research to my father, Irvin Lee Zook, who taught me the value of hard work and encouraged me to continue higher education and pursue a doctoral degree.

ACKNOWLEDGMENTS

The results of the research reported in this dissertation has been achieved through several collaborative efforts across the globe. Dynamic light scattering experiments were made available with the help of Stephanie Cope and Danmei Bian through Dr. Sara Viana's lab at Arizona State University (ASU). The magnetic resonance research center (MRRC) at ASU under the supervision of Dr. Brian Cherry has been a great resource for the use of the 800MHz NMR instrument and introduced me to the world of protein NMR. Additional NMR expertise was provided Dr. Michael Brown and his lab at the University of Arizona (UA), Tucson, specifically Dr. Trivikram Moulgu. Use of the cryoprobe 600MHz instrument at UA was provided by Dr. Neil Jacobsen and Dr. Guantian Lin who taught me the finer points of NMR data processing. Finally, the synchrotron radiation source circular dichroism (SRCD) study was made possible through collaborative efforts with Dr. Bonnie Wallace and her lab at Birkbeck College at the University of London. Thank you all for sharing your knowledge and resources with me.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

# 1. Introduction

## 1.1 Drive for Membrane Protein Structure Determination

Protein structure has great importance in many areas of science: biochemistry, biology, medicinal chemistry, materials science and engineering to name a few. The structure of a protein is closely related to the function of the protein; therefore solving the structure of a protein can provide enormous insight into how the protein operates. For example, many medications act as enzyme inhibitors as a way to treat a symptom or combat a disease [1]. If the structure of the enzyme is known, then it can provide details on how ligands interact with the enzyme on an atomic resolution scale. Therefore ligands can be designed and manufactured in order to create the most effective drug possible.

A specific type of proteins that have severe relevance in the field of medicine and energy are the integral, transmembrane (TM) proteins. These proteins are involved in a plethora of cellular functions including inter- and intra- cell signaling, molecule transport and chemical energy conversion [2, 3]. There are estimates that more than one half of all drugs on the market target membrane proteins (as of 2011), and therefore understanding the relationship between membrane protein function and structure can provide enormous insight into the field of drug design [2]. Additionally the major players in respirative chemical energy conversion performed in mitochondria as well as the proteins involved in converting light into chemical energy via photosynthesis are integral transmembrane proteins [3].

Despite the extreme importance of these transmembrane proteins, structural and functional information regarding these proteins are severely underrepresented. According to the protein database (PDB) server, as of March 27, 2012, a total number of 74,460 protein structures have been solved [4]. However as of April 1, 2012, only 980 of these structures are membrane proteins, representing only about 1% of the total solved

structures. Additionally, only 326 (~33%) of the 980 structures are of unique proteins that do not share significant sequence identity. It has been predicted that 20%-30% of proteins in a cell are membrane proteins, and therefore more progress in membrane protein structure determination is needed [5].

This large gap in information is in part due to the difficulty of characterizing membrane proteins. In order to obtain a high resolution structure of a membrane protein, the protein must be purified in its native form from all other proteins in the cell. Although this process can be difficult for even soluble proteins, it becomes very challenging for membrane proteins by comparison. This is largely because an integral transmembrane protein is embedded into the membrane and associates strongly with the hydrophobic alkyl chains of the membrane bilayer [3]. Therefore simply removing the protein from this environment will most likely corrupt the native form of the protein due to unfavorable hydrophobic interactions and force the protein to aggregate and precipitate out of the solution, rendering the protein useless for experimental structure determination.

Therefore the most common method for obtaining sufficient amounts of pure, natively folded membrane protein is to remove the protein from the membrane by embedding it into a detergent micelle. A detergent micelle is formed when sufficient amounts of a detergent molecule is added to a solution so that conditions become energetically favorable for the molecules to aggregate/assemble in a way that minimizes water exposure to the hydrophobic alkyl tails, and maximizes exposure to the hydrophilic head groups. Embedding the membrane protein into a detergent micelle allows the transmembrane region of the protein to be protected by the hydrophobic alkyl chains of the detergent while still remaining soluble in water. There are hundreds of different detergents to choose from, and each one offers different benefits and detriments for

maintaining protein fold. Choosing the detergent to use is a critical step in obtaining correctly folded, purified membrane proteins [6].

Membrane proteins occupy a relatively small volume of a cell compared to the soluble proteins. Conceptually this means that, in general, the amount of protein that can be isolated from the native membranes for an effective concentration to perform a specific task is significantly smaller than the soluble proteins. Therefore a larger density of protein per cell is required to obtain sufficient yields of protein for high resolution structure investigation. Generally to overcome this problem, a very high volume of cells are grown to compensate for the low yields of purified membrane proteins obtained. A second method is to use recombinant DNA techniques, as done in this study, to express large quantities of protein in the form of inclusion bodies that are then slowly solublilized and refolded into a native-like structure that can be used for structure determination. Another method is over-expressing the proteins to a moderate yield and allow the protein to insert into the membrane *in vivo* which then can be extracted with mild detergents.

One way for high resolution structure determination of proteins is through the use of X-ray crystallography. Of the 74,460 protein structures in the PDB, nearly 90% are solved through this method [4]. In short (discussed in further detail in section 2.4) proteins are gently forced into a crystal-lattice formation that is then subjected to a high energy x-ray beam that reflects off of the electrons in the protein to form a diffraction pattern. Due to the crystal-lattice, the reflections add constructively in order to form a detectable signal on the receiver (generally a CCD camera) [7]. Although single molecule diffraction may be possible in the future [8], it is not yet an option for protein structure determination, and therefore the proteins must be assembled into a lattice.

The significant drawback with this method, and probably the narrowest bottleneck for membrane protein structure determination is that protein crystallization is

extremely difficult [6]. Proteins have been evolutionarily designed to resist aggregation in order to maintain a soluble cellular environment. Additionally, there is an enormous amount of variables that must be considered when performing crystallization experiments: precipitate, protein concentration, ionic strength of solution, pH, temperature, humidity, and isotopes used to name a few. The large number of variables create a near infinite number of possible conditions for protein crystallization.

Another method for solving protein structures that circumvents the need for protein crystallization is solution state nuclear magnetic resonance (NMR) spectroscopy. NMR uses a very powerful superconducting magnet to align the nuclei of atoms in a molecule and then uses a transverse radio frequency (rf) pulse to perturb the equilibrium. As the nuclei return to equilibrium, they generate a magnetic signal that can be detected. Analysis of this signal can reveal an enormous amount of information based on how the nuclei is perturbed including structural information, diffusion information, protein dynamics, and functional studies [9, 10]. A more detailed look at NMR theory is provided in section 2.1.4.

NMR spectroscopy has limitations as well, generally revolving around the size of the protein being studied. Larger (>50kDa) and more flexible proteins tend to be more difficult to characterize with NMR than smaller and rigid proteins. This is due to both the greater number of signal peaks in an NMR spectrum that can lead to ambiguous overlapping, as well as due to the slower tumbling rate which results in a broader NMR peak which can also lead to ambiguous peak overlap [11]. However, protein NMR spectroscopy is a developing technology that continually allows the structure determination of larger protein complexes, as can be seen by the solution NMR structure of the integral transmembrane protein diacylglycerol kinase which functions as a 40kDa homotrimer for an effective molecular weight (MW) of 120kDa [12].

Solid state NMR (ssNMR) is a developing technology that may prove very useful for solving the structure of membrane proteins. This is in part due to ssNMR's ability to ignore the slow isotropic motion of larger proteins (because there is no isotropic motion in solids), and also in part that by embedding a membrane protein into a membrane will orient the protein which is a significant benefit when using ssNMR. If the samples are not oriented, then Magic Angle Spinning (MAS) is employed to average out chemical shift anisotropy and dipolar couplings [11]. Increased spectral resolution and sensitivity is improving through the development of stronger magnetic fields. Resolution increases linearly with the strength of $B_0$ and sensitivity increases as $B_0^{3/2}$ where $B_0$ is the strength of the static magnetic field [10]. Sensitivity has also improved several times over through the use of cryo-probe technology [13]. Additionally, the sensitivity of solution NMR experiments can also be increased by several orders of magnitude through the use of high-field dynamic nuclear polarization (DNP) which uses magnetization transfer of electrons to atomic nuclei [14]. This may prove very useful in the future for characterization of membrane proteins through solution NMR.

## 1.2 The Model Membrane Protein - OEP16

The Outer Envelope Protein, 16kDa (OEP16) provides an excellent model for structure exploration using NMR. OEP16 was first reported in 1997 as an amino acid selective channel protein by Pohlmeyer *et al* [15]. As discussed below, structural information regarding this protein has been extremely limited. Over the last 15 years, several crystallization attempts have been made for x-ray diffraction studies with extremely poor results [16]. Only low resolution information has been obtained through hydropathy plots from sequence data, electron microscopy, and circular dichroism of OEP16 in liposomes that result in very low signal to noise. Although methods have been employed to improve the quality of circular dichroism data, NMR proves to be the best

tool for obtaining detailed structural information about this membrane protein. The techniques employed to obtain structural data of OEP16 provide a model system for which future membrane proteins may be studied through NMR.

### 1.2.1 The Importance of OEP16 to Chloroplast Function

The chloroplast is the result of an endosymbiotic event where a cyanobacteria-like organism was engulfed by a ancient, larger, eukaryotic organism. Throughout evolutionary history this ancient progenitor would develop into a wide range of higher level photosynthetic organisms ranging from algae to plants [17, 18]. During this time, the endosymbiotic cell excised or moved the genetic material required for independent survival, and maintained the genetic information for the conversion of light energy to chemical energy [18].

A diagram of a plant cell chloroplast is depicted in figure 1.1. These chloroplasts have a morphology that includes two membranes that separate the inside of the chloroplast from the outside: the inner membrane and the outer membrane. Inside the chloroplast is another membrane system, called the thylakoid membrane system, that houses the protein complexes involved in photosynthesis and provides a barrier to establish a proton gradient across the thylakoid membrane. The thylakoid membrane is divided into disk-like parts that stack upon each other to form grana and more extended regions that form the stroma lamella [3].

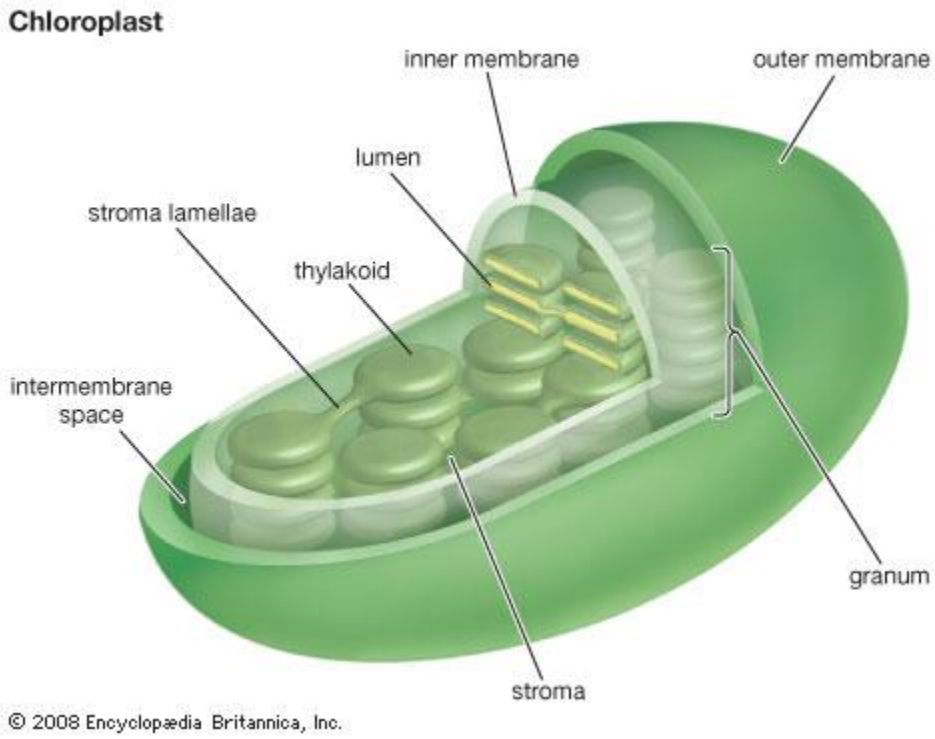


Figure 1.1. A model of a chloroplast [19]. Chloroplasts contain two membranes: the outer and inner. At the very center of chloroplast is the thylakoid membrane that contains PSII, Cyt $b_6f$, PSI, and ATP synthase which use a proton gradient across the membrane between the lumen and stroma as a means for converting light energy into chemical energy in the form of ATP. The thylakoid is arranged as stacks of disks called granum connected between bridges called stroma lamellae.

The major function of the chloroplast is to convert energy provided by the sun into chemical energy that is used by the cell, and all other forms of life. The main players for converting light energy to chemical energy are photosystem II (PSII), photosystem I (PSI), Cytochrome $b_6f$ complex, and the ATP synthase shown in figure 1.2. PS II splits two molecules of water to form molecular oxygen and four protons on the lumen side of the thylakoid membrane. The energy required for this reaction comes from four photons. The electrons are then transferred to the cytochrome $b_6f$ complex through the use of plastoquinones. The $b_6f$ complex pumps protons across the membrane as the electrons are transferred through it to plastocyanin. Photosystem I then uses light energy to drive

the transmembrane electron transfer from plastocyanin at the lumen side to ferrodoxin at the stromal side of the membrane, from where the electrons are transferred to the ferrodoxin-NADP reductase where $NADP^+$ is reduced to NADPH. This process creates a electrochemical gradient that drives ATP synthase to synthesize ATP from ADP + $P_i$ [18].

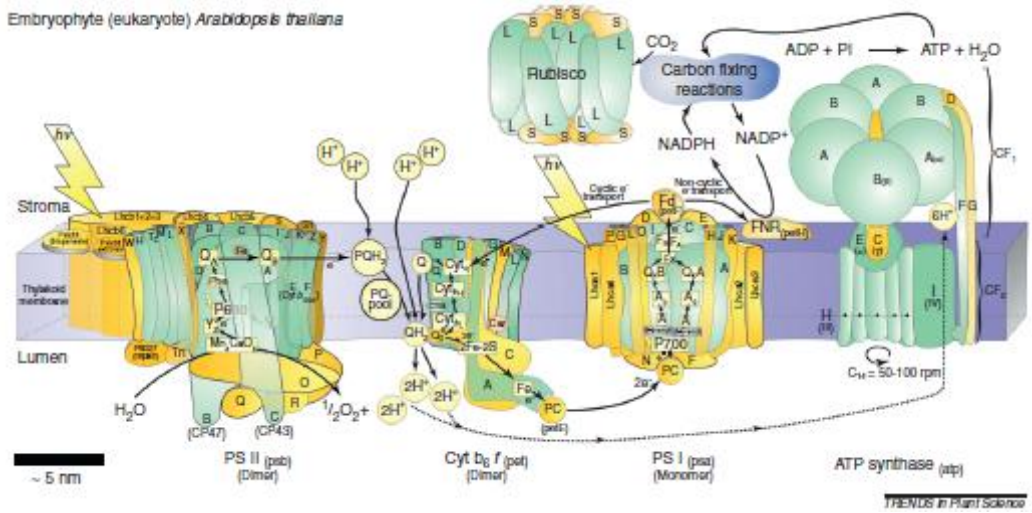


Figure 1.2. The main protein complexes for photosynthesis. Light energy is used to split water at PSII and shuttle the electrons to the cytochrome $b_6f$ complex by way of plastoquinones (PQ); as the electrons pass through the $b_6f$ complex to plastocyanin, protons are pumped from the stroma to the lumen; plastocyanin (PC) carries electrons to PSI where light energy is used to shuttle electrons from the lumen to ferrodoxin (Fd) the stromal side of the membrane. The proton gradient formed by this process is used to drive the ATP synthase to make ATP from ADP +$P_i$. Green protein subunits indicate proteins synthesized in the chloroplast; yellow indicates polypeptides synthesized in cytosol [18]. Reused with permission from Elsevier, April 16, 2012.

Almost all chemical energy on earth originates from this pathway; many of the proteins involved with this process are coded by the chloroplast genome and synthesized by ribosomes within the chloroplast seen in figure 1.2 [18]. This means that chloroplasts contain genetic information independent from the nuclear genome as well as the enzymes for RNA transcrition and protein synthesis. The chloroplast genome, or plastome, does not encode for the biosynthesis of amino acids, however. In order for translation to occur

it is necessary for the chloroplast to import all 20 amino acids for the synthesis of the polypeptide chain by the ribosome. The amino acids are synthesized by enzymes in the cytosol of the plant cell encoded by the nuclear genome or they can be obtained by way of nutrient intake. Either way they must be diffused into the chloroplast.

Due to the zwitterionic, polar nature of amino acids, they cannot diffuse directly through the membrane the way small, nonpolar molecules (like $CO_2$) can [3]. A channel must be present for passive diffusion. The outer envelope protein 16kDa (OEP16) is the protein that provides this channel for selective diffusion of amino acids across the outer membrane of chloroplasts [15]. This function is essential for the function of the chloroplast, as without OEP16, no proteins can be synthesized by the chloroplast.

### 1.2.2 The Current Understanding of Structure and Function of OEP16

Electrophysiological data collected by Pohlmeyer *et al* shows that OEP16 provides a 10Å channel that allows the selective diffusion of amino acids including (but not limited to): Glycine, Valine, Arginine, Lysine, Glutamate, and Glutamine. OEP16 does not provide a diffusion channel for sugars, such as sucrose, fructose, and glucose, nor for other small molecules such as dihydroxyacetone or tetraethylamine [15]. Recent studies have also shown that OEP16 may also be involved with the translocation of NADPH: protochlorophyllide oxidoreductase A (PORA) [20-23].

Structural knowledge of OEP16 is extremely limited. Initial structural insights were obtained largely from hydropathy plot data from the primary structure of OEP16 and CD spectra of OEP16 in liposomes [16]. These studies suggested that OEP16 was a homodimer that consisted of β-sheet structure at the N-terminus of the protein with three transmembrane helices [15]. Linke *et al* provided electron micrographs that led to particle sizes that represented a trimeric state; he additionally provided a new analysis of the hyrdopathy plot data that suggested OEP16 was completely α-helical with four

transmembrane helices [16, 24]. The most recent structural evidence published by Ni and Zook *et al* include a more complete circular dichroism study that agrees with Linke's argument that OEP16 is completely α-helical with size exclusion chromatography data that suggests a higher multimeric state of the protein [25]. Mutational studies performed on OEP16 revealed that the first two transmembrane helices are essential for the function of OEP16, while the last two helices are likely present for structural support [26]. Additionally, OEP16 has two major forms: a Leaf form (OEP16-L) and a Seed form (OEP16-S) which contains an extra domain in the loop region of the protein between helices one and two [27]. All studies performed on OEP16 in this study use OEP16-L.

## 1.3 The Aim of this Study

Even with the technological improvements of NMR, methods for utilizing them have fallen short for membrane protein NMR due to underdeveloped methods for characterization. Therefore, in this study we develop methods for obtaining structural data of membrane proteins using NMR. This includes methods in producing well folded, isotopically enriched protein in detergent micelles, pre-characterization through the use of traditional biochemical methods: HPLC, polyacrylamide gel electrophoresis, light scattering, and circular dichroism. NMR is then used in order to obtain detailed structural information of the membrane protein which will lead to a high resolution structure in the future.

# 2. Theory

## 2.1 Spectroscopic Methods of Characterization

### 2.1.1 Ultraviolet and Visible Spectroscopy

The data obtained by observation of light absorption in the ultraviolet to visible (UV-Vis) wavelength range (180nm to 780nm) can reveal information about organic compounds containing conjugated bond systems [28]. This is particularly useful in protein spectroscopy regarding the peptide bond, aromatic residues, and prosthetic groups containing conjugated systems (e.g. Flavins). Because recombinant $OEP16^{His}$ does not contain any prosthetic groups, only UV-Vis spectroscopy regarding the peptide bond and the aromatic residues will be addressed.

Light absorption occurs when a photon of sufficient energy promotes an electron from a low energy state to a higher energy state with the energy gap equaling the energy of the photon. In general a photon will promote an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) resulting in an exited state of the molecule. For protein studies in the UV-Vis wavelength range of light, this can happen in two ways shown by the energy diagram in figure 2.1. The lower energy absorption results from an event where a nonbinding electron (*n*) is promoted to a $\pi$-antibonding ($\pi^*$) orbital, called a $n \rightarrow \pi^*$ transition or an electric transition. The higher energy absorption event comes from promoting an electron from a $\pi$ bonding molecular orbital to a $\pi^*$ antibonding orbital, referred to as a $\pi \rightarrow \pi^*$or magnetic transition [28]. Both the electric and magnetic transitions of the peptide bond are diagrammed in figure 2.1.

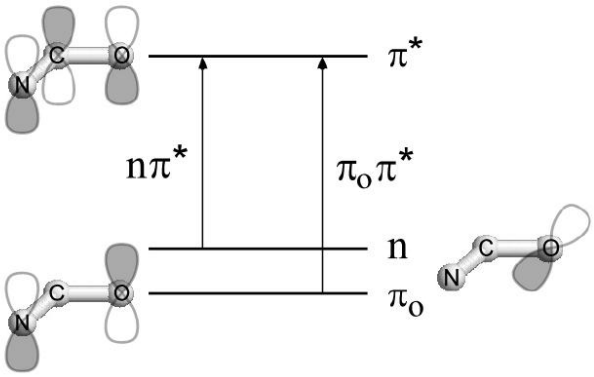


Figure 2.1. The electric and magnetic transitions allowed by a peptide bond. The larger energy transition is seen by the π→π* transition (left) compared to the n→π* transition (right).

The relation of light and the material that absorbs light can be described using the commonly known Beer-Lamberts law.

$$A = \varepsilon l c$$

where A is the absorbance reported by the spectrophotometer. ε is the molar extinction coefficient which is the molecules absorbtivity, (the strength of the molecule's ability to absorb light at the wavelength of light measured). The path length of the material is denoted as *l* and the concentration as c [29].

The Beer-Lambert law is a combination of two laws previously used to describe the interaction of light with matter. The first is Lamberts-Bouguer's Law that states that the power of light is decreased when the sample thickness is increased, described by the following expression:

$$dP = k_1 P_0 dl$$

where the power of light is P, the initial light power is $P_0$, and the thickness of the sample is *l*. The second equation is Beer's Law, that states that the fraction of energy produced

by monochromatic light absorbed by a material is directly proportional to the concentration of the material used to absorb the light expressed as:

$$dP = k_2 P dc$$

where c is the concentration of the sample [29]. These two expressions can be combined to give the following relationship:

$$P = P_0 e^{-k_2 bc}$$

$$P = P_0 \cdot 10^{-\varepsilon lc} \text{ where } \varepsilon = \frac{k_2}{2.303}$$

$$\log\left(\frac{P_0}{P}\right) = A = \varepsilon lc$$

π electrons have a lower excitation energy than σ electrons, and the more delocalized the electrons are, the lower excitation energy is needed, and as a system becomes more conjugated, the wavelengths of absorption therefore become longer [29]. This is seen by the presence of a very broad band in the UV-Vis spectrum of proteins containing aromatic residues (tryptophan and tyrosine especially) at around 280nm. Determination of protein concentration can be achieved with a high level of confidence using UV absorption spectroscopy of these residues.

The most efficient way to calculate protein concentration is to obtain the strength of the absorption signal at 280nm, and use an extinction coefficient calculated by

$$\varepsilon_{protein} = N_Y \cdot \varepsilon_Y + N_W \cdot \varepsilon_W + N_C \cdot \varepsilon_C$$

where $N_Y$ and $\varepsilon_Y$ are the number and extinction coefficient ($1490 M^{-1}cm^{-1}$) of tyrosines respectively; $N_W$ and $\varepsilon_W$ are the number and extinction coefficient ($5500 M^{-1}cm^{-1}$) of tryptophans; $N_C$ and $\varepsilon_C$ are the number and extinction coefficient ($125 M^{-1}cm^{-1}$) of cystines (two cysteine residues connected via disulfide bond [30]. All extinction coefficients listed are for 280nm wavelength only.

The above method, however, has a significant drawback. The wavelength at maximum absorption ($\lambda_{max}$) and therefore extinction coefficient can change depending on the local environment (solvent) of the residue being measured) [29]. This can be especially troublesome for membrane proteins due to the possible presence of hydrophobic interactions of the detergent with tryptophan or tyrosine residues. Therefore an alternative method of measuring protein concentration can be useful. A common method involves using a difference spectra of unfolded protein at neutral pH (7.1) and unfolded protein at high pH (12.5) were tyrosines become deprotonated [31]. By denaturing the protein in 6M guanidinium hydrochloride (GuHCl), the local environment should be ubiquitous throughout the protein. Molar concentration can then be calculated using the following expression

$$[Protein] = \frac{\Delta A_{294}}{(2357 \cdot Y + 830 \cdot W)}$$

where $\Delta A_{294}$ is the difference of absorption between the sample at pH 12.5 and pH 7.1 at 294nm, Y and W are the number of tyrosine and tryptophan residues respectively [31-33]. The coefficients 2357 and 830 relate to the extinction coefficients at pH 7.1 and 12.5 for tyrosine and tryptophan respectively [33].

### 2.1.2 Circular Dichroism Spectroscopy

Overall secondary structure of proteins can be determined quite easily using the difference spectrum of left and right handed circularly polarized light, known as circular dichroism. Circularly polarized light is created by superimposing two sources of polarized light perpendicular to each other, ± 90° out of phase (depending on whether left or right handed circularly polarized light is needed). Superimposing left handed and right handed light of equal intensity and wavelength will re-linearize the light [34].

If this light is then passed through a chiral sample that shows different absorption between left handed and right handed circularly polarized light, a phenomenon called

ellipticity will be observed seen in figure 2.2. Units reported by the instrument are generally in units of degrees or milidegrees. This arises from the angle formed between the major axis of the ellipse and the hypotenuse from the right triangle formed between the major and minor axis [34]. Although the instrument strictly measures intensities between left-handed and right-handed circularly polarized light, the angle is calculated using the following expression:

$$\theta = \tan^{-1}\left(\frac{I_1 + I_2}{I_1 - I_2}\right)$$

where I is the intensity of the right or left handed circularly polarized light.

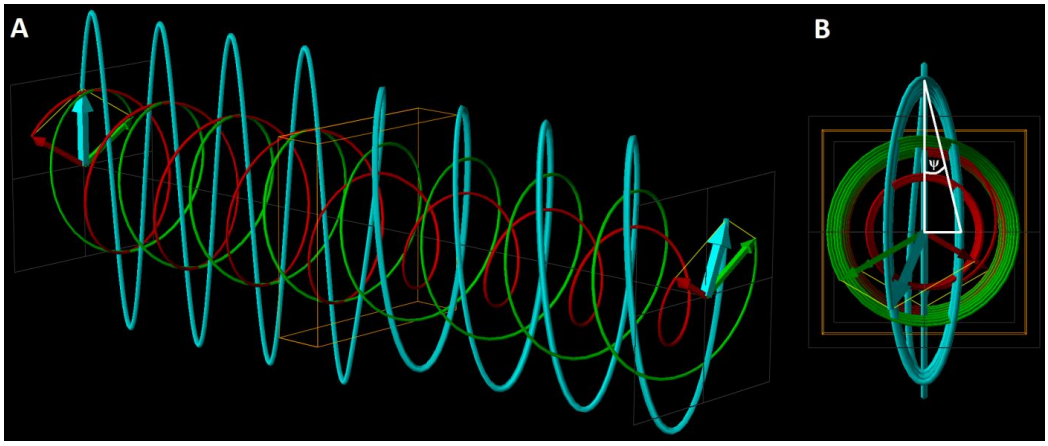


Figure 2.2. Simulation of CD spectroscopy. The sum (blue) of right handed (green) and left handed (red) circularly polarized light at the same wavelength and intensity will re-linearize the light. As the light passes through a material (orange box) that has different extinction coefficients between the left handed and right handed circularly polarized light the sum of the two will no longer re-linearsize, but form an ellipse which is characterized by ψ (B) which is a front-on view of figure a at a different time-point.

Because circular dichroism measures intensities of light passing through a material, the same factors that are addressed in the Beer-Lamberts Law (concentration and pathlength) must be considered. One of the most common units, and the unit used in this manuscript is Delta Epsilon (Δε or dε), and it is defined as:

$$\Delta e = \varepsilon_L - \varepsilon_R = \frac{A_L - A_R}{l \cdot c}$$

where $\varepsilon_L$ and $\varepsilon_R$ are the extinction coefficients for left and right handed circularly polarized light respectively, l is the path length of the cuvette, and c is the molar concentration of the sample. Often (not always) $\Delta\varepsilon$ is defined as per amino acid in order to describe the signal as a sum of contributions of individual amino acid residues contributing part of the CD signal. $\Delta\varepsilon$ has units of $M^{-1}cm^{-1}$ [34].

The strength of a CD signal is calculated by the integrated intensity of the band, also called the rotational strength. Rotational strength is experimentally defined as:

$$R = (2.295 \times 10^{-46}) \int \left(\frac{\Delta\varepsilon}{\lambda}\right) d\lambda$$

with units of $J \cdot cm^3$. Rotational strength has also been defined theoretically:

$$R = Im\{\mu \cdot m\}$$

This is the imaginary part of the scalar product of electric (μ) and magnetic (m) dipole transition moments [35]. Using this equation with the wave functions of the ground and excited states it is possible to calculate a rotational strength of a transition:

$$R_{0a} = Im\{\langle 0|\mu|a\rangle \cdot \langle a|m|0\rangle\}$$

μ is the electric dipole moment operator which represents the linear displacement of a charge when excited; m is the magnetic dipole transition moment operator, which describes the circular displacement of electron density when excited. A helical displacement of charge results upon superpositioning μ and m. The helical displacement of charge interacts differently with right and left handed circularly polarized light [35].

Although proteins consist of amino acids that are chiral in nature, proteins will form secondary structures that are chiral themselves. These chiral superstructures in turn modulate the CD spectra of the peptide bonds that can be used to estimate an overall view of the secondary structure. All protein containing α-helices or β-sheet display bands at wavelengths corresponding to $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions. As shown by figure 2.3, the electronic transitions are at lower energies (higher wavelengths) while magnetic

transitions are observed at higher energies (lower wavelengths). Proteins with high β-sheet content shows a minimum at about 215nm for the electronic transition and a maximum at 195nm for the magnetic transition. The CD spectra of β-sheet containing proteins can vary greatly due to the significant differences in bond geometries between parallel and anti-parallel sheets as well as the different twists that are often present. Proteins containing α-helices have two minima at about 222nm and 208nm with a strong maximum at 190nm. The generation of a third band arises from a phenomenon known as exciton coupling [34]. Exciton coupling is observed when an excited state is delocalized among a number of chromophores located in close spatial proximity and orientation. The result is a band at a higher energy, 190nm and a band at a lower energy, 208nm where theses states are linear combinations of the excited state of the individual chromophores [34, 36].

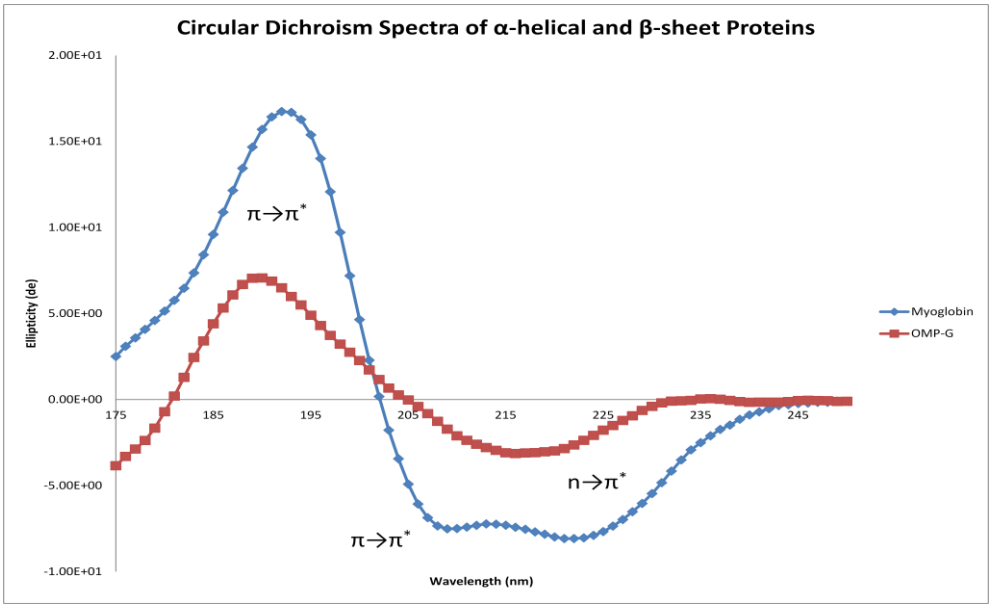


Figure 2.3. CD Spectra of an α-helically rich protein (myoglobin at 73.9%) and a protein high in β-sheet (OMP-G at 67.6%). Both bands corresponding to magnetic ($\pi \rightarrow \pi^*$) and electric ($n \rightarrow \pi^*$) transitions are blue shifted for proteins high in β-sheet . The magnetic transition energy is split in α-helical proteins due to exciton coupling along the helix which results in the negative band at 208nm indicative of α-helical CD spectra. The spectra was constructed using data from Lees *et al* (myoglobin) and Abdul-Gader *et al* (OMP-G) [37, 38].

This splitting event is not observed in exclusively β-sheet proteins because although there are a number of chromophores in close spatial proximity, the orientation is significantly different. For proteins that do not conform to a secondary structure, there is very little ellipticity above 210nm, however the $\pi \rightarrow \pi^*$ transition band is negative and therefore can be easily distinguished from the secondary structure containing counterparts [31].

The further the UV CD data is collected, the more information regarding secondary structure is obtained. However, conventional CD instruments generally use a xenon arc lamp to generate the necessary wavelengths of light. This becomes problematic as the quality of the beam is significantly diminished at higher energies. Light from xenon arc lamps drop in flux by almost two orders of magnitude when going from 280nm to 180nm [34]. This makes data collected at the far UV end of the spectrum less reliable due to a decrease in the signal to noise ratio. Further complications arise in the far UV in that many buffers and other components of a solution begin absorbing light below 200nm.

The best solution to overcome both of these challenges is to use a synchrotron radiation source of photons for CD measurements. In most cases SRCD sources can maintain a constant flux down to below 140nm. Moreover the overall flux of synchrotron radiation is several orders of magnitude greater than the xenon arc lamp at all wavelengths. This higher flux allows an extended wavelength range for performing CD measurements. The increased flux can allows data collection of samples where the buffering solution absorbs. Although the majority of photons are absorbed by the buffer, enough photons can reach the detector to provide a reliable CD measurement [34].

Once CD measurements are performed it is necessary to employ methods to extract structural information from them. A linear combination of secondary structure component spectra, $B_{k\lambda}$, can be expressed as:

$$C_\lambda = \sum f_k B_{k\lambda}$$

in order to estimate the secondary structure content of a measured CD spectrum, $C_\lambda$. This linear combination equation is the basis for analytical methods of protein CD analysis; $f_k$ is the fraction of secondary structure of type k with $B_{k\lambda}$ being a reference for calculation at wavelength $\lambda$ (usually obtained from solved protein structures as protein CD tends to deviate from ideal conformations). A 'basis set' or 'reference set' is a set of proteins that have known CD spectra and known secondary structure, usually obtained from the solved structure of the protein.

With a reference set of N proteins with known CD and secondary structure, many linear equations can be written:

$$[C_\lambda]_N = \left[\sum f_k B_{k\lambda}\right]_N$$

Using a brute-force method of solving the N number of equations simultaneously and minimizing differences between the calculated and experimental CD via least-squares method one can calculate $B_{k\lambda}$, which is the component spectra that contains all of the common elements between the measured CD and the reference set [35].

Although the above method works in theory, this brute-force procedure becomes difficult to handle due to the N number of equations is much greater than the number of unknowns. Therefore a number of algorithms have been developed in order to handle such problems. CONTIN is a program that utilizes ridge regression that can fit the measured CD spectrum as a linear combination of proteins measured in the reference set using the following expression:

$$\sum_{\lambda}\left(C_{\lambda}^{calc}-C_{\lambda}^{obs}\right)^{2}+\alpha^{2}\sum_{j=1}^{N}\left(v_{j}-N^{-1}\right)^{2}$$

Where $C^{calc}$ is the value of the constructed spectrum at wavelength λ, $C^{obs}$ is the measured value at wavelength λ, α is the regularization coefficient to prevent overfitting, N is the number of proteins in the reference set, and $v_j$ is the coefficient of the $j^{th}$ protein in the reference set used for the calculation of $C^{calc}$ [35]. This is then used to determine the secondary structures:

$$f_{k}^{calc}=\sum v_{j}f_{k}^{j}$$

This calculation can also be solved using a matrix based method. Using two matrices: a spectral matrix, **C** that has a dimension of (Number of CD data points) x (Number of proteins in reference set), and a structural matrix **F** with a dimension of (Number of Secondary Structures) x (Number of proteins in reference set). The relationship between these two matrices are as follows:

$$\mathbf{C}=\mathbf{BF}$$

This matrix equation can be solved using a method called singular value decomposition to estimate a value of $f_k$ for any $C_\lambda$ [35].

Although ridge regression allows for the capability for constructing a spectra, not all of the spectra in a reference set are equal in importance for the calculation. Therefore it is common practice to select a limited number of proteins from the reference set to create a subset, which is the basic idea of the variable selection method [35]. Different subsets are formed by sequentially removing a protein from the reference set. By creating a large quantity of reference protein subsets with different numbers of reference proteins in each provides a large number of possible solutions, which in turn becomes computationally difficult. One way to get around this is to limit the number of reference proteins, removal of spectra is performed by excising the spectra that is least similar to

the measured spectra; referred to as local linearization [35]. CONTIN/LL uses ridge regression along with local linearization to construct a spectra that best fits the measured one for extracting secondary structure elements. Subsets of reference sets are also created by selecting a number of proteins randomly chosen from the reference set to create a 'minimal basis' [35]. The number of subsets then provide a large pool of protein combinations from which solutions can be derived, and the process is repeated when enough valid solutions are calculated.

Inclusion of the measured protein within the reference set will improve the quality of $B_{k\lambda}$, but the $f_k$ values are still unknown. The self-consistent method provides a process by guessing the secondary structure of the protein and then finding the protein in the reference set that is most similar to the spectra measured. Solution of the resulting matrix equation reveals a new set of $f_k$ values that are significantly greater than the original guess. This process is then repeated recursively replacing each resulting solution as the new guess until convergence is met [35].

Two common algorithms employ the matrix method of singular value decomposition and the use of the variable selection method, local linearization, and minimal basis: SELCON3 and CDSSTR [35]. These algorithms along with CONTIN/LL are provided by the analysis software package CDPro, which is offered free on the web by the Sreerama lab at Colorado State University. Also included in CDPro is an extensive library of reference sets that can be used for deconvolution of secondary structure from CD data.

### 2.1.3 Light Scattering

Light is described as an oscillating field that has both electric and magnetic components. When this oscillating field interacts with a neutral particle, a dipole is induced onto that particle. This dipole then oscillates as well. The oscillating dipole then

emits radiation and is called 'scattering'. This principle can be used to extract important information on protein macromolecules including size and secondary structure [39].

Light scattering can be divided into three different categories: static light scattering, dynamic light scattering, and Raman scattering. The only technique utilized by this study is dynamic light scattering, therefore it will be the only one addressed. Dynamic light scattering is a useful tool due to its ability to calculate a particle size without intimate knowledge of the system.

Macromolecules in solution undergo Brownian motion, random movement caused by much smaller molecules randomly colliding with the particle to move it slightly in solution. Light that is scattered by different particles in solution phase shift and therefore interfere. The interference of light will either be constructive or destructive, depending on where the particles are located relative to each other. As the particles move through solution, the interference between the scattered light changes, causing a fluctuation in the intensities of the scattered light [39]. This fluctuation therefore reflects the motion of the scattering particles; if the particle is large, the movement is slow, and therefore fluctuation of light intensity is slow as well; if the particle is small, the movement is faster, and therefore the fluctuation of the intensity is fast.

The intensity of light fluctuations is analyzed via a mathematical algorithm called an autocorrelation function. An autocorrelation function describes the relationship of a measured signal with itself by expressing the probability of a signal intensity being equal after a delayed amount of time, $\tau$. At short time intervals between measured signals, the probability is high that the intensities are equal due to the limited amount of time a particle has to travel a distance; the longer the time interval is the less likely the intensity will be equal to that at time point, $\tau = 0$. The autocorrelation function is defined as:

$$g(\tau) = \frac{\langle I(t)I\langle t+\tau\rangle\rangle}{\langle I(t)\rangle^2}$$

Larger particles travel slower, therefore the intensities remain equal for longer time intervals. This allows for a relationship to be developed between the autocorrelation function and the diffusion constant (D). The following expression is used to fit the autocorrelation function of a monodispersed sample:

$$g(t) = 1 + \beta e^{-2D\mu^2\tau}$$

where β is the amplitude of the function, μ is a term to describe the wavelength of light used (λ) and the angle of the detector relative to the incident beam of light (θ): $\mu = \frac{4\pi}{\lambda}\sin\frac{\theta}{2}$ [39].

This translational diffusion coefficient, D, can be used to extract a hydrodynamic radius ($R_h$) of the particle using the Stokes-Einstein relation:

$$R_h = \frac{kT}{6\pi\eta D}$$

where k is Boltzman's constant, T is temperature, and η is the viscosity of the measured solution. Obtaining the $R_h$ of a macromolecule with minimal knowledge of the system is the strength of dynamic light scattering. If assumptions are made regarding the shape of the macromolecule as well as the size of the water shell, and the density of the particle is known, a molecular weight, (M), can be estimated from $R_h$ using geometric volumes of the assumed shape. For a spherical particle:

$$M = (R_h - r_w)^3 N_A \frac{4\pi}{3\nu}$$

where $N_A$ is Avagadro's Number, $r_w$ is the particle radius contributed by water, and ν is the partial specific volume (the inverse of density).

### 2.1.4 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is an extremely powerful technique used for characterizing both protein folding and determining protein structures.

#### 2.1.4.1 Principles of NMR

NMR takes advantage of an intrinsic property of fermions known as 'spin'. Spin is a form of angular momentum that is not produced by the rotation of a particle. All quarks have a spin of 1/2 and of the three quarks in protons and neutrons, two of them align antiparallel to each other leaving the third quark to be responsible for the 1/2 spin that is characteristic of both hadrons [9].

Pauli's exclusion principle applies to spins in the nucleus as well as those in the electron cloud, and a nuclear shell theory comparable to the better known electron shell theory has been developed in an attempt to explain spin systems that are so important for NMR. The heaviest atom considered in this protein study is $^{15}N$, therefore only the first two energy levels are considered for all nuclei. The 1s nuclear shell may only contain maximally two spin states (one up and one down), similar to the 1s electron shell model. However, unlike the electron shell model, there is a 1p shell that can contain six spin states (three up and three down) [40]. For example: $^{1}H$ has only a single proton spin that occupies the 1s proton shell, and therefore has a total spin of +1/2. Protons and Neutrons behave independently in this matter and, therefore, $^{2}H$ has a single proton spin that occupies the 1s proton shell and a single neutron spin that occupies the neutron shell giving a total spin of 1 (1/2 + 1/2). Helium has two protons and two that occupy the proton shell with one spin up (+1/2), and one spin down (-1/2) and the neutron shell with one up spin and one down spin providing a spin state of 0 and is therefore "NMR" silent.

For each spin $I$, the nucleus has a spin degeneracy of $2I + 1$ nuclear states. This means that in the absence of a magnetic field, all of these nuclear states have the same

energy and therefore are indistinguishable from each other. However, when an external magnetic field is applied, each of the nuclear states occupies a slightly different energy level, and therefore the degeneracy is broken. This separation of energy states is known as Zeeman Splitting illustrated in figure 2.4[9]. As the strength of the magnetic field increases, the energy difference between the split states also increases which is what provides the increased resolution of NMR spectra at higher magnetic fields.

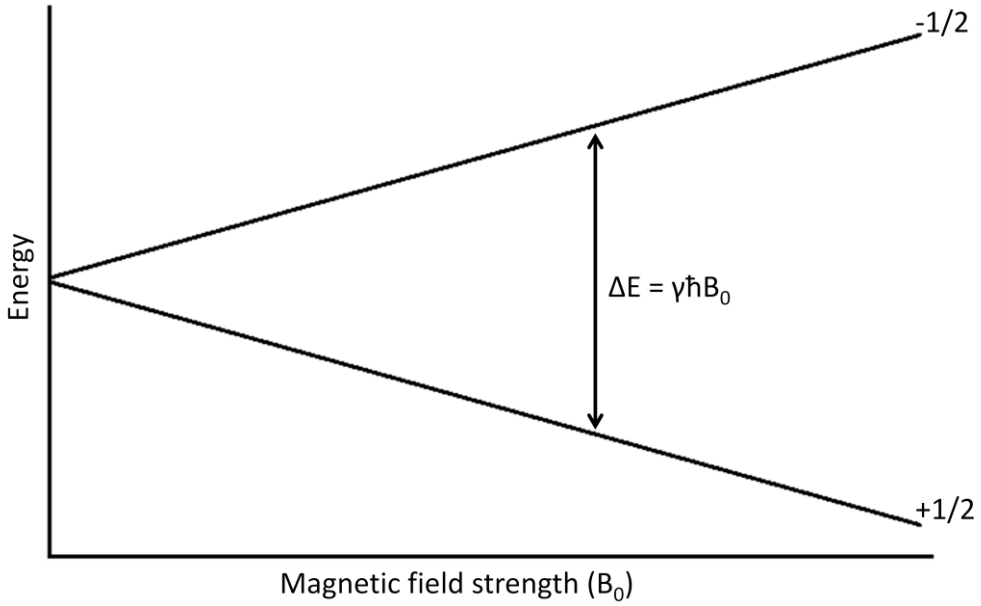


Figure 2.4. Energy diagram of two spin states as a function of magnetic field strength. The spin states become nondegenerate upon application of a magnetic field. The lower energy spin state orientation (+ or -) is determined by the sign of $\gamma$ which can be either positive or negative (for most nuclei $\gamma$ is positive, therefore the lower energy is the + spin).

Nuclei as well as electrons posses 'intrinsic magnetism.' This magnetism is not the result of circulating current as it is for orbital magnetic moments which is what gives rare earth magnets their magnetic capabilities, but rather is simply considered as a characteristic of the nuclei. Also unlike rare earth magnets, nuclei do not align perfectly along a magnetic field the way a compass needle does in the presence of a magnetic field; instead the nuclei precess around the magnetic field vector at a specific angle (shown by

figure 2.5) with a specific frequency defined by the strength of the magnetic field ($B_0$) and the gyromagnetic ratio ($\gamma$) of the isotope in question. This frequency is called the Larmor frequency ($\omega^0$):

$$\omega_0 = -\gamma B_0$$

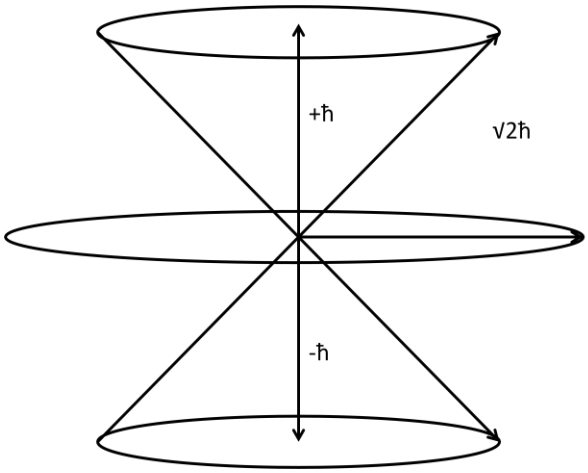


Figure 2.5. Illustration of precession about $B_0$ for a spin 1 system. The magnitude of the z-axis component is known, but the uncertainty principle forbids us to know about the xy plane.

A molecule of interest in a static magnetic field does not provide much information by itself. This is because although precession of the bulk magnetic moment in a static field suggests an induced electromotive force by Faraday's law of induction, the bulk magnitization vector, **M**(t), is colinear with $B_0$ and therefore no signal is observed [10]. Therefore an applied magnetic field in the form of a radiofrequency (rf) pulse is used to disturb the equilibrium. After an rf pulse, **M**(t) is not (in general) perpendicular to the static field; therefore the bulk magnetization will precess around the static field with an angular frequency of $-\gamma B_0$ and generates a signal in the probe coil [10]. If the rf frequency ($\omega_{rf}$) is equal to $\omega_0$ then the rf pulse is said to be *on-resonance*. The rf field is pulsed for a period of time $\tau_p$ (known as the pulse length) to define the flip angle [9, 10]. Rf is used to manipulate the spins in a system in order to obtain specific

information regarding the system. Generally several rf pulses are needed to perform a useful NMR experiment, called a pulse sequence or pulse program.

2.1.4.2 Making NMR Work for Proteins

In order for NMR to be useful for protein studies, multiple dimensional NMR must be employed. The experiments performed in the study of OEP16 include two and three dimensions, although four dimensional experiments are also common. Multiple dimensional experiments all follow a similar general schematic, with the only real difference is the number of evolution periods involved (See Figure 2.6). For example, a two dimensional experiment begins with a preparation pulse sequence followed by a single evolution period which is a varied time delay that allows for coherence transfer between spins (t1) before a mixing pulse sequence is performed and the FID is detected (t2). The Fourier transform is then first performed on t2 axis for each time interval on t1, then a second Fourier transform is performed on the t1 axis which provides the two dimensional plot [10]. To get a third dimension, a second evolution period is added. Two of the most common 2D NMR experiments are the $^{1}H$, $^{15}N$ Heteronuclear Single Quantum Coherence (HSQC) and $^{1}H$, $^{13}C$ HSQC experiments which reveals information about protons that are a directly bonded to the heteroatom (N or C) Therefore in a perfectly resolved $^{1}H$,$^{15}N$-HSQC NMR spectra, there will be a resonance peak for every nitrogen in the protein backbone, which makes this experiment particularly useful. Many three dimensional experiments have proven useful for studying OEP16 as well: HNCA, HNCoCA, HNCACB, CBCAcoNH, HNCO, HNcaCO, TOCSY-HSQC, and NOESY-HSQC. Although most of these experiments can provide some information regarding the tertiary fold of a protein, these are primarily used for resonance assignment.

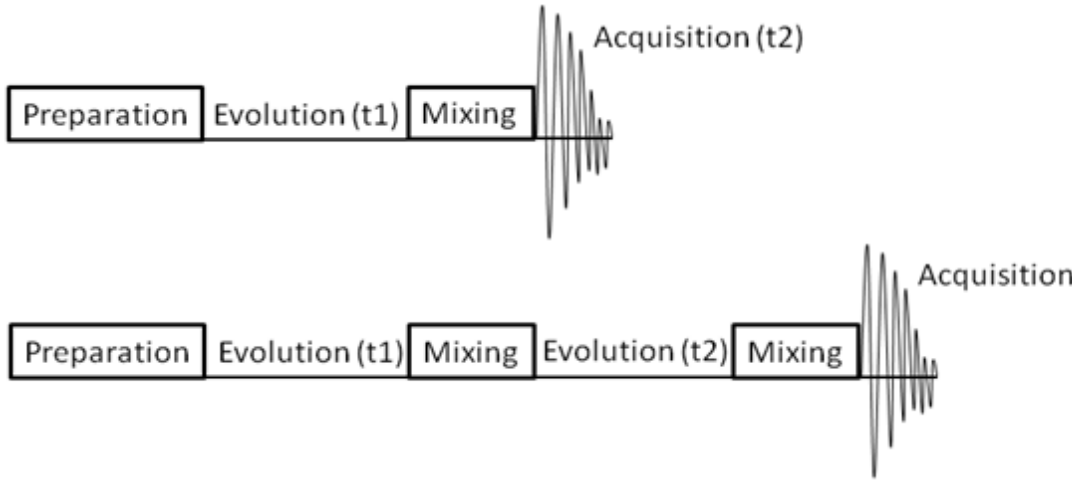


Figure 2.6. Pulse sequence concepts for multiple dimensional NMR. In general the only difference between different dimensions is the number of evolution times used for 2D spectra (top) and 3D spectra (bottom)

2.1.4.3 Resonance Assignment

Arguably the most important element of characterizing a protein via NMR is the assignment of resonances. When an NMR experiment is performed, the end result is a spectrum that has (hopefully) resolved and dispersed peaks. However these peaks mean nothing unless it is known what part of the macromolecule is responsible for creating them. Therefore several experiments have been developed in order to answer this problem.

In general, most of these experiments take advantage of the connectivity of amino acids in a polypeptide chain. The HNCA (Figure 2.7) experiment is a wonderful example to demonstrate this concept. The HNCA experiment transfers magnetization from the amide proton to the bound nitrogen, and then transfers coherence to the α-carbon of its amino acid as well as to the α-carbon of the amino acid before it [10]. This is one way a $^{1}H$, $^{15}N$-HSQC can be pulled out into another dimension, a carbon dimension. So for every nitrogen resonance in the protein backbone, there will be two

carbon peaks, a strong signaled peak from the current amino acid, and a weak signal from the prior amino acid (i-1).

Simply speaking, to find the resonance that represents the prior amino acid, one has to find the strong resonance signal at the same carbon chemical shift, which will in turn have a weaker signal resonance pointing to the residue ahead in the polypeptide chain. This process of following resonances is called 'walking the backbone' and the result is what is called a 'strip plot' which is a convenient way of displaying the connectivity.

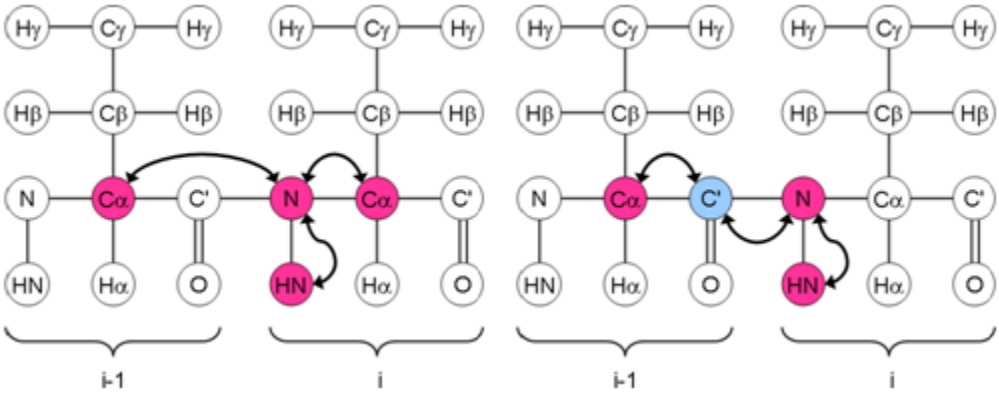


Figure 2.7. Concept for the HNCA (left) and HN(co)CA (right) pulse sequences [41]. The resulting spectra will contain $C_\alpha$ chemical shift values that can be used to assign the resonance peak.

With larger proteins comes more peaks with varying levels of signal quality, and very often the weak resonance peak gets buried in the noise, and therefore the next α-carbon shift in the sequence becomes unknown. A helpful experiment has been developed for use in conjunction with the HNCA called the HN(co)CA (Figure 2.7), which gives strong signals only to the i-1 residue of a poly peptide chain [10]. Therefore if a HNCA signal for the i-1 residue is ambiguous due to poor quality, the HN(co)CA can be used. To determine the identity of an amino acid residue, the chemical shift of the α-carbon (CA) proves useful. The each amino acid has a characteristic α-carbon shift; for

example, alanine has an average CA shift of 53.14 ± 2.12 and valine has an average CA shift of 62.48±2.91 [42].

More often than not, however, the CA shift ranges of amino acids overlap, for example histidine has a CA shift of 56.55±2.92 and the CA shift of glutamate is 57.31±2.18 and thus the identity of the CA shift becomes ambiguous [42]. Looking at the β-carbon provides the ability for a much more characteristic fingerprint to each amino acid. Therefore the HNCACB experiment is a very powerful tool for obtaining residue information. The HNCACB provides the same information an HNCA experiment does (It provides a strong signal for the Cα of the amino acid, and a weak signal for the Cα of the i-1 residue.) as well as provides a strong negative peak for both the current residue's β-carbon and a weak negative peak for the i-1 residue's β-carbon [10]. Not only does this help with the identification of amino acids, it also improves the confidence of the strip plots by using the CA and CB carbons to walk the backbone. Similar to the HNcoCA experiment, there is an experiment designed to only reveal the CA and CB shifts at the i-1 position; this is called CBCA(co)NH, where magnetization is first transferred from the $C_\beta$ to the $C_\alpha$, then to the amide nitrogen through the carbonyl carbon, ultimately transferred to the directly bonded proton for FID detection in the t3 [10].

The HNCA and HNCACB experiments can be used for walking the backbone of a protein, but the pulse sequence tends to be long, and the longer the pulse sequence is, the broader the peaks become. The most sensitive 3D HCN experiment available is the HNCO, which pulls a $^{1}H$, $^{15}N$ HSQC into the third dimension via the carbonyl carbon shift [10]. This particular pulse program transfers magnetization from the amide proton to the nitrogen and then to the directly bonded carbonyl carbon (CO) at the i-1 position. This experiment produces relatively sharp and well dispersed peaks compared to the other 3D experiments. However, to walk the backbone, it is imperative to have 2 peaks

per strip in the strip plot, and therefore the HNcaCO experiment becomes useful. This produces the same peaks seen in the HNCO, but also includes a second, minor peak for the carbonyl carbon of the $i^{th}$ residue by way of allowing magnetization to be transferred through the α-carbon (but the CA shift is not observed in this experiment) [10].

Because OEP16 is a large protein by NMR standards (especially when considering the detergent micelle), more information is needed to confirm the sequence assignment. Therefore a $^{15}$N-edited NOESY experiment in tandem with a $^{15}$N-edited TOCSY experiment, can provide useful information about residue connectivity. NOESY is an acronym for Nuclear Overhauser Effect SpectroscopY, and is used to observe spins that are spatially close (<5 Å). A $^{15}$N-edited NOESY-HSQC experiment first exchanges magnetization between all of the protons via the NOE pulse sequence, then transferred to a nitrogen, and then back to the amide proton for FID detection. This experiment pulls the $^{1}$H, $^{15}$N-HSQC into a third dimension by all protons that are correlated via NOE [10]. This experiment is difficult to interpret by itself, and that is why a $^{15}$N-TOCSY-HSQC is helpful. TOCSY stands for TOtal Correlation SpectroscopY, and is used to observe protons of a residue by transferring magnetization of the protons through the covalent bonds to the nitrogen and finally back to the amide proton for detection [10]. This experiment is necessary to separate the resonances that are due to the NOEs intra-residue (therefore share the same peaks as the TOCSY-HSQC) from the inter-residue NOEs. Although long distant NOEs are difficult to spot in a $^{15}$N-edited NOESY-HSQC, NOEs from the i-1 position are generally very strong and easy to detect. If there is a strong α-helical component, NOEs from the i-4 positions also become apparent. These can be used to create strip plots and walk the backbone in a very different way compared to the $^{13}$C experiments.

2.1.4.4 Constructing a 3D Picture

Although several more NMR experiments are needed in order to obtain a three dimensional structure, it is possible to tease out some structural information from the experiments described previously. As discussed, the $^{15}$N-edited NOESY experiments can provide some information regarding the helical content of a protein, and if long-range NOEs are uncovered, can provide much more detailed information on how the polypeptide chain is oriented in three dimensional space. What is not as obvious, is that chemical shift data can provide information regarding the secondary structure content of proteins by providing dihedral angle restraints [43, 44]. This is because chemical shifts are very sensitive to the local molecular environment, and thus conformation. The characterization of the chemical shifts for $^{1}H_{\alpha}$, $^{13}CO$, $^{13}C_{\alpha}$, and $^{13}C_{\beta}$ are well characterized for the protein backbone. There are two common ways to interpret chemical shift data for use in determining dihedral angles in the protein backbone: the first is to calculate the secondary chemical shift, the other is to compare the observed chemical shift to a database of solved proteins that also have observed chemical shifts; this is done using the program TALOS [43, 44]. The secondary chemical shift is defined as the observed chemical shift value minus the expected chemical shift for the same residue in a random coiled polypeptide; the values for these numbers show characteristic patterns for secondary structure elements[10]. TALOS is a program developed for use in determining dihedral angle restraints via a database of proteins that have known chemical shifts for assigned $^{1}H_{\alpha}$, $^{13}CO$, $^{13}C_{\alpha}$, and $^{13}C_{\beta}$ resonances. The result is a set of calculated dihedral angles and a probability that the residue is part of an α-helix or β-Sheet.

By itself, chemical shift data is not enough for a structure of even a small protein. In most cases, more NMR experimental data is needed, long range NOES and residual dipolar couplings are a few good examples. However, due to large improvements in *ab*

*initio* molecular modeling, there have been some reports of successful protein structures with the chemical shift as the only experimental data using the program CS-Rosetta [45, 46].

2.1.4.5 Exploring Protein Dynamics through NMR

Relaxation experiments can provide a significant amount of information regarding the overall characteristics of a protein. In general relaxation is the study of how a spin perturbed by an rf returns to equilibrium. This return to equilibrium is explained using two systems: spin-lattice relaxation, and spin-spin relaxation [9].

Spin-lattice relaxation also called longitudinal relaxation is the process by which spin realigns with the magnetic field. When a protein sample is suddenly introduced to a magnetic field, all of the atoms do not immediately align with the magnetic field at the same time. Due to thermal fluctuations within the sample and differing electromagnetic environment of each atomic spin, this alignment happens at different speeds for each atom. Therefore this realignment is gradual and by very good estimations can be fit exponentially and characterized by the time constant T1.

$$M_z^{nuc}(t) = M_0^{nuc}\left(1 - e^{\frac{-(t-t_{on})}{T1}}\right)$$

As stated above, the Z direction is defined as being the direction of the magnetic field, therefore the equation describes the magnetization of the nuclear spin ($M_z^{nuc}$) as a function of time, *t*, as it goes to equilibrium, $M_0^{nuc}$, when the external magnetic field is introduced at time, $t_{on}$ [9].

Spin-spin relaxation, also called transverse relaxation is the process by which the atomic spins dephase from each other. This is because each spin experiences a slightly different magnetic field and therefore rotates at slightly different larmor frequencies. This process can also be considered exponential and characterized by a time constant.

However, unlike spin-lattice relaxation, spin-spin relaxation occurs on the xy plane [9, 10].

One of the major mechanisms responsible for relaxation seen in this study is through dipole-dipole interactions with the magnetic field. As a protein moves in solution relative to the static magnetic field (either through tumbling or through internal protein dynamics), the magnetic field exerted by one spin on another spin changes, causing a fluctuating magnetic field which contributes to relaxation. Another major mechanism responsible for relaxation is chemical shift anisotropy. For non-spherical electron density, as a protein moves through tumbling or internal motion, the magnetic field caused by electrons changes orientation relative to the static magnetic field. This changes the magnitude of the magnetic field of the electron, resulting in fluctuation, and therefore relaxation [9].

An important concept for protein relaxation is the spectral density function, J(ω). The spectral density function is defined as twice the Fourier transform of the autocorrelation function [9]:

$$J(\omega) = 2\int_0^\infty G(\tau)\, e^{\{-i\omega\tau\}} d\tau$$

The autocorrelation function was addressed in section 2.1.3 when dynamic light scattering was discussed. In a similar the way photon intensity fluctuates as photons scatter off of a diffusing particle, a fluctuating magnetic field can be treated in a very similar way. However, it is impossible to measure the fluctuating magnetic field of a spin directly, and therefore a lot of effort has been put into relating the spectral density functions to the measurable relaxation values, $T_1$ and $T_2$ [47].

$$\frac{1}{T_1} = \frac{d^2}{4}[J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X)] + c^2 J(\omega_X)$$

$$\frac{1}{T_2} = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H + \omega_X)]$$

$$+ \frac{c^2}{6}[4J(0) + 3J(\omega_X)] + R_{ex}$$

where $d = \frac{\mu_0 \hbar \gamma_H \gamma_X}{4\pi r_{XH}^{-3}}$ and $c = \frac{\omega_X \Delta\sigma}{\sqrt{3}}$, $\mu_0$ is the permeability of free space, $\hbar$ is plank's constant $\gamma_H$, $\gamma_X$ are the gyromagnetic ratios of the $^1$H and X nucleus respectively, $r_{XH}$ is the X-H bond length, $\omega_H$ and $\omega_X$ are the Larmor frequencies $^1$H and X spins and $\Delta\sigma$ is the chemical shift anisotropy of the X spin. The spectral densities can be defined as [48]:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - S^2)\tau_c}{1 + (\omega\tau_c)^2}\right]$$

$S^2$ is called the generalized order parameter, $\tau_m$ is the isotropic correlation time of the protein molecule. $\tau_c$is called the effective correlation time and is defined as:

$$\frac{1}{\tau_c} = \frac{1}{\tau_m} + \frac{1}{\tau_f}$$

Where $\tau_f$ is the correlation time of the fast internal motions. If it is assumed that the internal motions are restricted in amplitude and fast enough to neglect their relaxation contributions, the spectral density function can be simplified to:

$$J(\omega) = \frac{2}{5}S^2\left[\frac{\tau_m}{1 + (\omega\tau_m)^2}\right]$$

Which allows $T_1/T_2$ to be independent of $S^2$ and therefore the average $\tau_m$ for each residue can be calculated explained by Kay *et al* [49] and therefore an average overall correlation time, $\overline{\tau_c}$, can be determined.

The average overall correlation time can be used to estimate a hydrodynamic radius of the particle using the following variant of Stoke's Law [10]:

$$r_H^3 = \frac{4\pi\eta\overline{\tau_c}}{3kT}$$

The faster the protein tumbles, or the internal movement of the protein is the faster the fluctuation of the magnetic field. Therefore looking at the relaxation of a protein's backbone amide can give us information on what part of the backbone is dynamic compared to where the protein structure is static. In general, if an amide is not relaxing via internal motion, it is due to the tumbling of the macromolecule. Therefore a global isotropic rotational correlation time (the time it takes for the protein to tumble one radian, assuming a completely spherical particle) can be estimated by looking at only the relaxation of the stationary amides.

## 2.2 High Pressure Liquid Chromatography

Many methods exist for protein purification and characterization by high performance liquid chromatography (HPLC), but only two will be addressed in this study: Affinity chromatography and size exclusion chromatography (SEC). Affinity chromatography, specifically Ni-NTA affinity chromatography was essential for providing sufficiently purified OEP16 protein for experimental characterization, SEC was used in both a way to improve purification and supply size information on OEP16.

### 2.2.1 Affinity Chromatography by Ni-NTA

Affinity chromatography exploits the ability for some proteins to specifically and reversibly bind competitively to a ligand. For HPLC purification, the ligand is covalently attached to an immobile material (often cross-linked dextrose called sephadex) that is packed into an HPLC column creating the stationary phase. As a protein mixture is passed through the column as the mobile phase, the proteins that bind to the ligand will remain on the column while the rest of the solution flows through unhindered. The bound protein is then eluted from the column usually by adding a competitive ligand molecule which will bind to either the protein or stationary phase, allowing the protein to disassociate from the column material [3, 50].

For nickel affinity chromatography, a chelating agent, nitrilotriacetic acid (NTA) is covalently bound to sephadex. NTA uses four coordination sites to bind $Ni^{2+}$ ions. Histidine residues of proteins in the mobile phase will bind to the $Ni^{2+}$ ions immobilized by the NTA-sephadex on the column. The bound protein is recovered by either adding imidizole as a protein competitor or by lowering pH to protonate the histidine residue, which significantly lowers its affinity to $Ni^{2+}$ [51].

A protein with several histidine residues in tandem will drastically increase the affinity of the polypeptide to $Ni^{2+}$. Tandem histidines are rare and therefore recombinant technologies are employed to add a histidine tail to the protein of interest. Conventionally a tail of six consecutive histidines are used, called a 6xHis tail. This allows purification of proteins to >95% homogeneity in a single HPLC run. Additionally, Ni-NTA affinity chromatography works independent of protein tertiary structure and therefore can be used to purify proteins under strong denaturing conditions necessary to solubilize inclusion bodies [51], which is of particular interest for this project as seen in section 3.2.

### 2.2.2 Size Exclusion Chromatography

SEC, also known as gel filtration chromatography or gel permeation chromatography involves the use of small, porous material packed into the columns and then filled with the mobile phase. The sample is injected over the column and flow through is detected by some spectroscopic method. The injected sample molecules diffuse into pores: the smaller the particle, the deeper it can permeate into the pore as well as permeate into smaller pores while larger particles are unable to penetrate deeply into the pores, or not able to enter at all. Since the effective volume of the column is smaller for the larger particles, they elute first, followed by molecules of decreasing size. This is the major mechanism by which SEC works and is called 'steric exclusion.' Other

mechanisms may be involved as well, such as restricted diffusion and separation by flow [39].

Steric exclusion works because diffusion coefficients are large enough that the time required for particles to diffuse into and out of pores is much shorter than the time for the molecules to stay in a given part of the HPLC column. For Steric exclusion, the elution volume ($V_e$) of a protein can be expressed as:

$$V_e = V_0 + K_d V_i$$

where $V_0$ , or void volume, is the volume of the column outside of the pores, $K_d$ is called the distribution coefficient, and $V_i$ is the volume within the pores [39]. Therefore there exists size limits for SEC that depend on the type of stationary phase used. The highest particle size limit is determined when the particle is too large that it will not diffuse into the largest pore ($V_e = V_0$). The smallest particle size limit is the particle size where it can occupy all of the porous space ($V_e = V_0 + V_i$ or 'total volume'). The distribution coefficient describes the volume fraction of pores that are available to a particle of a given size and is independent of the inner diameter or length of the column [39].

If the time required for particles diffuse in and out of the porous material is closer to the time scale needed for the molecules to stay in the given part of the column, restricted diffusion plays a stronger role in separation. When this is the case the permeation depth of a particle is governed by its diffusion coefficient (and therefore indirectly related to particle size). Larger molecules with slower diffusion times take longer to traverse in and out of the pores, and do not stay in the column area long enough to penetrate into the entire volume allowed by the pore. In size exclusion chromatography this becomes more of an issue for large protein complexes [39].

Separation of flow is considered only in the cases of very high molecular weight particles where they are capable of penetrating the pores. This system then creates a

series of narrow capillaries and the larger particles become statistically closer to the center of the capillaries and thus travel faster due to decreased flow resistance [39]. Solution travels slower on the walls of the capillaries due to friction provided by the walls.

In practice to calculate a particle size a set of well characterized protein standards are used to form a standard curve with a very high molecular weight particle used to calculate the $V_0$ and a very small molecule used to calculate $V_t$. Elution volumes are obtained and a partition coefficient ($K_{av}$) for the standards and unknown is calculated using the following equation:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

$K_{av}$ is by good approximation linearly related to the log(MW), and can be used then to obtain a particle size.

2.3 SDS Polyacrylamide Gel Electrophoresis

In short, SDS polyacrylamide gel electrophoresis (SDS-PAGE) is a technique that is used to separate proteins solely based on the proteins molecular weight. Proteins are introduced to the detergent SDS that bind to the protein to form a negatively charged complex [52]. The larger the protein, the more detergent that is needed, and therefore the more negatively charged. The proteins are generally submitted to detergent concentrations high enough to denature the biomolecules to form a prolate elipsoid or rod that has a length roughly equal to the molecular weight, thus removing any shape contributions. These negatively charged complexes traverse toward the positively charged cathode through a sieve-like matrix of polyacrylamide gel when an electrical field is applied [52]. Similar to SEC, the studied protein is compared to a set of standards and the distance proteins travel is linearly related to the log(MW).

## 2.4 Protein Crystallization and X-Ray Diffraction

Although X-ray crystallography of proteins is not a significant part of this study, protein crystallization was attempted and therefore is addressed briefly in this text. In addition, X-ray diffraction still remains the single most effective way of obtaining high resolution protein structures [4] and therefore also warrants a brief explanation.

Imaging objects, regardless of the size, requires a source of photons that will interact with an object, causing the photons to diffract. In cases where the object is greater than 400nm in length, these photons can then be refocused to a magnified image, using a lens, whether it be from an eye lens or a lens from a microscope or telescope. However, the smaller the object becomes, the harder it is to resolve the image [7].

The Rayleigh criterion of resolution states that two points can be just resolved if they are separated in the image by a distance called 'the radius of the Airy disk' ($r_{Airy}$) which is defined as:

$$r_{Airy} = 0.61\left(\frac{\lambda}{NA}\right)$$

where λ is the wavelength of light used and NA is the numerical aperture of the lens being used to magnify the image [53]. Therefore the resolving capabilities are largely limited by the wavelength of light used. In order to resolve molecules to atomic resolution, high energy photons in the X-ray region of the spectra are needed [7].

This in turn becomes a problem for two reasons: the first is that X-rays cannot be refocused by a lens, and therefore an image cannot be formed directly; the second is that X-rays interact only very weakly with molecules, and therefore the diffracted X-ray photons cannot be detected when a single molecule interacts with the photon beam [7].

To get around the first problem, researchers obtain an image that is formed from the diffraction of photons. Using the diffraction pattern that is formed by the scattered x-rays, well developed mathematical techniques are applied to pull the diffraction pattern

(which has units of inverse space or K-space) into real space, forming an electron density map which can be used to fit the diffracted molecules [7].

The other problem is overcome by the use of assembling the individual molecules into an ordered lattice. The individual molecules are all oriented in the same direction and therefore interact with the x-ray beam in an identical fashion, which allows the detection of the diffracted x-ray photons [7]. However, ordering protein macromolecules into a crystal lattice is an extremely difficult challenge and remains the bottleneck for almost all protein structures [6]. Development of high intensity, high quality x-ray beams is currently an area of intense research in hopes that single molecule diffraction may become a reality in the future [8].

The most common method for protein crystallization is to introduce a protein solution to a precipitant containing solution, such as polyethylene glycerols, (PEGs) or salts that require large amounts of water to maintain a hydration shell. Then water is slowly removed from the system (usually by evaporation techniques) in order to slowly increase both the precipitate concentration and the protein concentration [7]. As the water that is keeping the protein soluble is removed the macromolecules are forced to aggregate together in order to reduce protein surface area.

Crystal formation can best be described using a phase diagram exampled in figure 2.8. Protein crystal formation occurs in two steps. The first step is nucleation, which becomes the seed for which crystals form. Nucleation occurs when the protein and precipitant concentration reaches a threshold where there is not enough water in the solution to keep the aggregated proteins soluble. This forces the protein out of solution as a precipitate. Because protein is being removed from solution, the concentration of soluble protein decreases, which moves the solution into the growth, or metastable zone of the phase curve. During this part of the crystallization experiment, crystal formation

occurs, where the nucleated protein aggregates slowly recruit soluble protein to form a crystal lattice [7]. This continues until the protein concentration reaches an equilibrium between crystal formation and dissolving protein.

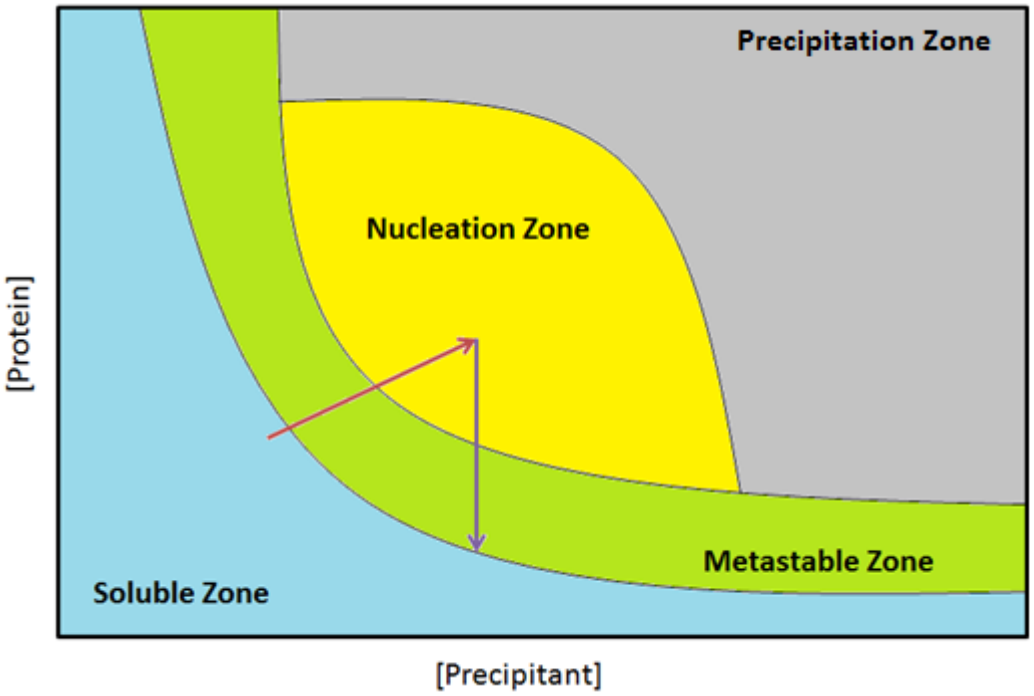


Figure 2.8. Phase diagram for protein crystallization. Arrows indicate a typical vapor diffusion experiment: As water is slowly removed, the protein condition is moved from the soluble zone, through the metastable zone, into the nucleation zone depicted by the red arrow. When nucleation occurs the protein is removed from solution, lowering protein concentration, pushing the solution into the metastable zone where crystal growth occurs (shown by purple arrow) until it meets the interface between the soluble and metastable zones.

More often than not the protein and precipitant concentrations reach a point where many nucleation events occur. The result of this is a shower of small crystals, or even worse, the formation of an amorphous, acrystalline precipitate [7]. If this happens, then precipitation and protein concentration conditions can be altered in order to more slowly introduce the system into the nucleation zone. Other times, the metastable zone can be an area of the phase diagram that is too narrow to produce large crystals, which

can be another problem.  Therefore other precipitation conditions are explored for adequate crystal growth.

# 3. Materials and Methods

## 3.1 Expression of OEP16 from *E. coli*

### 3.1.1 Preparation of *E. coli* from Glycerol Stock

*E. coli* BL21(DE3) cells containing the *oep16pea* gene inserted into a histidine tagged pET21b vector was provided in a 30% glycerol stock by Da Qun Ni . Two plasmid constructs were developed and provided by the group of J. Soll at the University of Munich: one that contained the *oep16pea* gene followed by a stop codon and one without a stop codon. This allowed for the expression of wild-type OEP16 ($OEP16^{wt}$), as well as OEP16 with a C-terminal tail with a 6x histidine tag, $OEP16^{His}$.

Lysogeny broth (LB) agar plates were prepared by dissolving 1g tryptone, 1g NaCl, 1.5g agar and 0.5g yeast extract in 90mL RO water and adjusted to pH 7.0. The solution was brought to 100mL in 100mL graduated cylinder with RO water. The solution was autoclaved for 45min sterilization, 45 minutes exhaust. Similarly, 100mL of liquid media was prepared without agar and was divided evenly across two 100mL fluted culture flasks and autoclaved. The solution was allowed to cool to ~50°C when 5mg ampicillin was added. The solutions were homogenized and the agar containing LB was poured into 2 plates inside a laminar flow hood.

Once the agar plates had cooled and solidified, using aseptic technique, a loop of *E. coli* from each of the provided glycerol stocks were streaked out onto each of the plates. Aseptic technique involved the use of the 'one glove' method. After sterilizing the countertop and gloves with bleach, one glove was removed. All sterile material was handled with the gloved hand, every other item was handled with the ungloved hand. Aseptic technique also required sterilization of all materials used: either by autoclave for media and glassware or flame sterilization for the tungsten loops used for streaking cultures. The plates were then sealed with paraffin and allowed to incubate upside-down

overnight at 37°C. The following morning a single colony from each of the $OEP16^{wt}$ and $OEP16^{His}$ plates were transferred to the liquid media cultures and allowed to incubate, shaking at 200rpm at 37°C overnight. The next day, 500µL aliquots of the cell culture stock samples were added to several 1.5mL eppendorf tubes with 500µL water with 60% glycerol to make 30% glycerol stock growths. These stock cells containing either the $OEP16^{wt}$ or $OEP16^{His}$ plasmid were stored at -80°C and would then be used for all subsequent studies.

3.1.2 Growth of Non-isotope Enriched *E. coli*

LB media for growth of *E. coli* was prepared in one liter volumes by adding 10g tryptone, 10g NaCl, and 5g yeast extract to ~800mL of RO water, stirring constantly in a 1L beaker. The pH was adjusted to 7.0. Subsequently, the volume was brought to 1L in a 1L graduated cylinder with RO water and transferred to a 2L fluted culture flask.

The solution was then sterilized by use of autoclave: sterilizing for 45min at 121°C followed by 45min of cooling and exhaustion. After cooling down to room temperature (using an ice bath if time was an issue), 50mg of sterile filtered ampicillin was added to the solution and mixed to homogeneity. 20mL of the solution was then transferred to an autoclaved culture tube for use as a preculture growth. The remaining LB media could be stored at 4°C until needed (maximum of about five days).

A loop scraped from the -80° *E. coli* 30% glycerol stock was added to the 20mL preculture tube. The solution was allowed to incubate overnight, shaking at 200rpm at 37°C. The following morning the 1L of the LB medium was warmed in a 37°C incubator, shaking at 37°C to aerate the solution to improve growth. After 2 hours the 20mL preculture was added to the 1L medium. At $OD_{600}$ = 0.7, which was generally achieved between 2 and 3 hours of growth. 0.238g of Isopropyl –β-D-1-

thiogalactopyranoside (IPTG) was added to the culture, corresponding to a final concentration of 1mM.

Expression of protein was permitted for 4 hours. The cells were harvested via centrifugation: the culture was divided into four 250mL volumes and centrifuged at 8000 x g for 8 minutes. The cell pellets were then combined in a pre-massed 50mL falcon tube using a scoopula and the wet mass of cells was recorded. The cell pellet was stored at -80°C. In the later part of my graduate school studies, when large amounts of protein was prepare, the cell culture was upscaled to culture volumes of 4L at once.

3.1.3 Growth of Isotopically Enriched *E. coli*

For the growth of isotopically enriched *E. coli*, several different strategies were attempted to produce the highest yields of protein possible. The first method uses M9 minimal media throughout the entire growth process. The second method utilizes LB broth as a preculture, and then inoculated into M9 minimal media. The final method that was developed required *E. coli* cells be adapted to high concentrations of deuterated water ($D_2O$).

For all methods used, M9 minimal media was essential. To prepare M9 minimal media, several stock solutions were needed: 1M $MgSO_4$, 1M $CaCl_2$, and 5x M9 Salts. A 1L 5x M9 salt solution was prepared by adding 65g $Na_2HPO_4$, 15g $KH_2PO_4$, and 2.5g NaCl to 800mL of deionized water (di$H_2O$). The solution was then adjusted to a volume of 1L in a graduated cylinder with di$H_2O$. The stock solutions along with 800mL di$H_2O$ were then autoclaved for 45min with 45min of cooling and exhausting. The solutions were then set to cool to room temperature.

To make a 1L M9 minimal media culture solution: 200mL of 5x M9 salts, 2mL of $MgSO_4$, 100µL of $CaCl_2$, 10mL of 100x Vitamins for Basal Medium Eagle (Modified) (Provided by MP Biomedicals, Catalogue number: 1600449) with contents listed in table

3.1, 1g of $^{15}NH_4Cl$, and 4g $^{13}C$-glucose were added to the 800mL di$H_2O$ in an autoclaved 2L culture flask. pH was adjusted to 7.3 and 50mg of ampicillin was added to culture tube and mixed to homogeneity. 2 x 20mL of solution was transferred to two autoclaved culture tubes for preculture. The 1L M9 minimal media culture solution was kept at 4°C for up to five days.

| Component | Concentration (mM) |
|---|---|
| **Biotin** | 0.41 |
| **Choline Chloride** | 0.72 |
| **D-Calcium Pantothenate** | 0.42 |
| **Folic Acid** | 0.23 |
| **myo-Inositol** | 1.11 |
| **Nicotinamide** | 0.82 |
| **Pyridoxal HCl** | 0.49 |
| **Riboflavin** | 0.03 |
| **Thiamine HCl** | 0.30 |

Table 3.1. Components of 100x Basal Medium Eagle Vitamin (Modified) used for growing isotope enriched *E. coli* for use in NMR studies.

Using aseptic techniques, loops of frozen 30% glycerol stock solution of *E. coli* cells were added to each of the preculture tubes containing M9 minimal media. The tubes were incubated overnight at 37°C shaking at 200rpm. The next morning the 1L culture solution was warmed to room temperature in 37°C incubator, shaking at 200rpm. Both 20mL preculture growths were then added to the 1L medium and cells were allowed to grow to $OD_{600}$ = 0.8. To reach this density, it generally took between 8 and 10 hours. At $OD_{600}$ = 0.8, protein expression was induced by addition of 0.238g IPTG (1mM final concentration) and expression of OEP16 was permitted for 5 hours. Cells were harvested via centrifugation at 8000 x g for 8min at 4°C, similar to the procedure used by non-isotope enriched cell harvesting methods. The mass of the pellet from the 1L culture was determined and the cell pellet was stored at -80°C.

An alternative method to express isotope enriched OEP16 was to first grow a 20mL preculture of cells in LB media overnight, using methods described in section 3.1.2. The next morning 5mL of the preculture was added to the 1L culture of M9 minimal media and allowed to grow to $OD_{600}$ = 0.8. This generally took 6 to 8 hours. Subsequently expression was induced with 1mM IPTG. Harvesting methods were identical to the methods described above.

It was imperative to perdeuterate OEP16 for specific NMR experiments. Therefore it was necessary to grow *E. coli* in high concentrations (80% to 90%) of $D_2O$. To do accomplish this task, it was necessary to adapt OEP16 to these conditions. Sterile 50mL of LB media was prepared along with Sterile 50mL of M9 minimal media with the following concentrations of $D_2O$: 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, and 90%. Each of the 50mL solutions were divided evenly between two sterile culture tubes. Using aseptic techniques, a loop of glycerol stock was added to each of the two LB culture tubes and the cell culture was incubated overnight at 37°C shaking at 200rpm. The next day a loop of the LB culture was aseptically used to inoculate both 20% $D_2O$ M9 minimal media cultures and allowed to incubate overnight at 37°C shaking at 200rpm. The next day a 30% glycerol stock of the cell culture was prepared, and a loop was used to aseptically inoculate the 30% $D_2O$ M9 minimal media solutions and allowed to incubate overnight at 37°C shaking at 200rpm. This was repeated for each increased $D_2O$ culture step until cells were successfully grown in 90% $D_2O$.

To create a triply labeled sample ($^{2}H/^{13}C/^{15}N$ sample), sterile M9 minimal media was prepared as described above with 1g $^{15}NH_4Cl$, 4g $^{13}C$-glucose, and 800mL $D_2O$. The preculture solutions were aseptically inoculated with a loop from the 90% $D_2O$ stock solutions and allowed to incubate overnight at 37°C shaking at 200rpm. The following morning, the precultures were used to inoculate the 1L culture. The culture was

permitted to grow to $OD_{600}$ = 0.8, which took between 11 and 12 hours. 1mM IPTG was then added to the cell culture and OEP16 was expressed for 6 hours until cells were harvested via the methods described above.

### 3.2 Purification of $OEP16^{His}$ from *E. coli*

OEP16 expressed recombinantly resulted in the formation of inclusion bodies, which are insoluble protein aggregates. Two methods were used to obtain correctly folded protein which both followed a similar general strategy: isolation of inclusion bodies, unfolding of the protein in chaotropic agents such as urea or guanidinium hydrochloride (GuHCl), purification of unfolded $OEP16^{His}$, and refolding of $OEP16^{His}$. Depending on the amount of protein needed and the conditions required two different methods, described below, were used.

#### 3.2.1 Original Purification Method

This method (called Method 1) was established and published in Ni and Zook *et al* [25]. An on-the-column refolding method was developed for obtaining sufficient yields of OEP16 in SDS micelles for structural investigation. Although there are modifications to the original purification scheme, the method in this section largely describes the procedure from Ni and Zook *et* al [25]. A schematic diagram is shown in figure 3.1a. The cell pellet with expressed $OEP16^{His}$ was resuspended in 30mL of cold lysis buffer (25% sucrose, 1mM EDTA, 20mM Tris, 10mg/g cell lysozyme), pH 8 and kept at 4°C. In an ice bath, the solution was tip-sonicated at 50% power, 50% pulse for 30 seconds on and 30 seconds off to break the cells. This was repeated five times until solution was homogenous and viscous.

The suspension was then centrifuged at 10,000 x g for 20 minutes at 4°C. The supernatant was discarded and the pellet containing the inclusion bodies and cell debris was recovered. This pellet was then washed to remove lipids and any other cell debris

that would hinder purification or protein folding. This was achieved using two different detergent solutions. The first solution contained 1% deoxycholate, 1% Triton X-100, 1mM EDTA, 10mM β-mercaptoethanol (β-ME), 200mM NaCl and 20mM Tris, pH 7.5. 10 mL of this solution was used to resuspend the pellet through manual agitation with a spatula as well as vortexing. The suspension was then centrifuged at 12,000 x g for 10min at 4°C. The supernatant was discarded and pellet recovered. This was repeated twice. Afterwards a second detergent buffer was used to exchange the detergent. This buffer was the same as the first, only replacing the deoxycholate and Triton X-100 with 0.5% Mega-9. Where the first solution was used to remove lipids and DNA debris, this significantly less harsh detergent was used to remove the residual deoxycholate and Triton X-100. The pellet was resuspended in 10mL of the second detergent buffer and centrifuged as described for the first washing step and repeated once. The supernatant was discarded and the pellet was recovered.

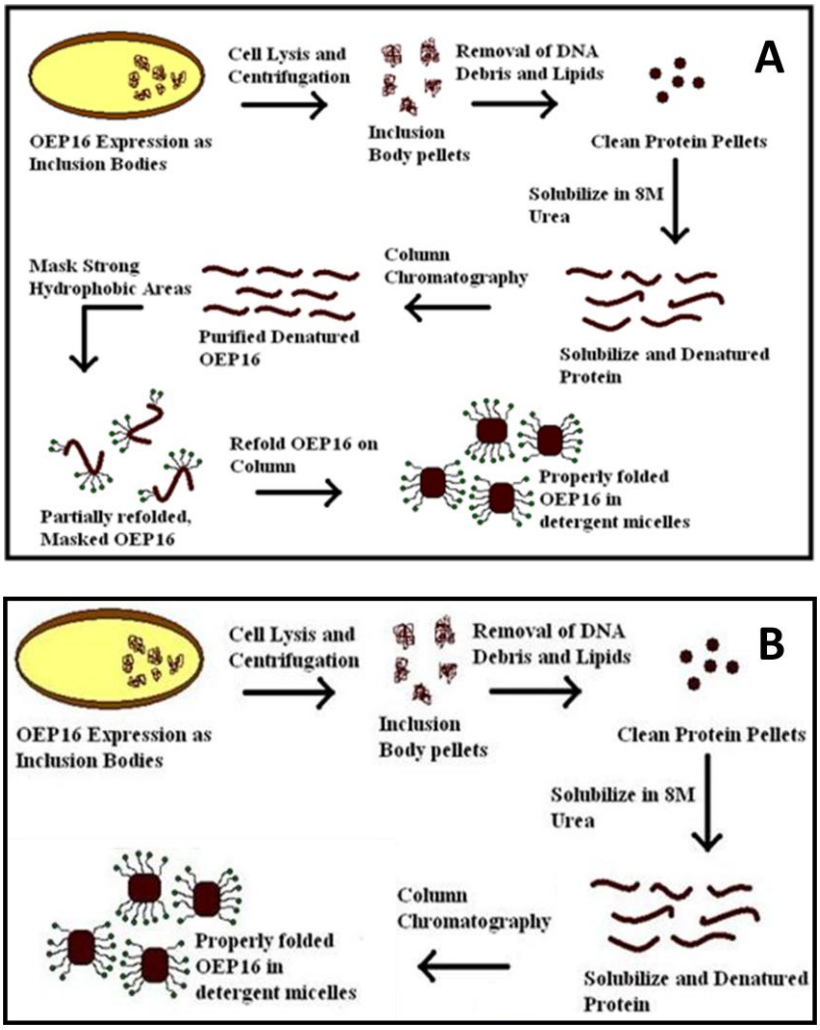


Figure 3.1. Purification schemes for both Method 1 and Method 2. Original purification scheme provided by Ni and Zook *et al* (A). After protein is expressed as inclusion bodies the cell is lysed and the inclusion bodies cleaned with detergent to remove cell debris; the inclusion bodies are then solubilized with 8M urea or 6M GuHCl; soluble, unfolded OEP16 is then purified by affinity chromatography; afterwards the strong hydrophobic regions of the protein eluant is masked with SDS while 8M urea is diluted to 3M; the protein is then refolded on the HPLC column and eluted with imidizole. The new purification method, designed for mild detergents, (B) removed the necessity of masking the protein via slow dilution which as hypothesized for being a significant source of protein loss. This was done by combining the two HPLC steps and masking the protein directly on the column.

To unfold the protein contained in the inclusion body pellet, a strong chaotropic urea buffer (buffer A) was used containing 8M Urea, 100mM $NaH_2PO_4$, and 10mM Tris,

pH 8.0. 20mL of denaturing buffer was used to resuspend the protein from the pellet. The suspension was allowed to incubate, shaking at 100rpm overnight at 4°C. The next morning the protein-urea solution was centrifuged at 12,000 x g for 30 minutes to remove any material that remained insoluble after the incubation. The pellet was discarded and the supernatant was used for further purification of $OEP16^{His}$.

The use of an Äkta HPLC system was incorporated in the purification procedure of $OEP16^{His}$ which differs slightly from the original gravity-driven affinity chromatography. Using an Äkta explorer HPLC (GE Healthcare), a 12mL column containing Ni-NTA superflow resin (Qiagen) was equilibrated with buffer A, the same Urea buffer, pH 8.0 used to solubilize OEP16 from the inclusion bodies. After equilibration the 20mL sample was loaded onto the column. Non-binding flow through eluted from the column, then the pH was then adjusted to about 6.5 by the increase of buffer B (the same 8M urea denaturing buffer, pH 4.0) to 50%. This allowed for the elution of nonspecific binding protein to elute from the Ni-NTA column. Finally, 100% buffer B was added to the column to elute $OEP16^{His}$. At this point $OEP16^{His}$ had been successfully purified in an unfolded state. To make any use of the purified protein, it must be in a folded state. The process for folding the protein is described below.

To prevent hydrophobic interactions of $OEP16^{His}$ with the column resin during the on-the-column refolding, a masking step was introduced as described by Ni and Zook *et al* [25]. This was done through a dilution method where the protein solution was diluted to 3M Urea with a reconstitution buffer containing 10% glycerol, 100mM NaCl, 20mM Tris, pH 8.0 and 1% SDS. The amount of solution used to dilute depended on the total volume of eluted protein but total volume of diluted protein was usually between 25 and 35mL. This dilution was done slowly using the Äkta explorer HPLC to have a highly controlled flow rate that would slowly drop-wise dilute the solution over one hour.

To refold the protein, the Äkta explorer HPLC at a flow rate of 1.0mL/min was used to equilibrate the same 12mL Ni-NTA resin with buffer A, which contained 3M urea, 10% glycerol, 100mM NaCl 20mM Tris, pH 8.0, and 1% SDS. The pH of the protein solution was increased to 8.0 with concentrated NaOH and the protein was loaded onto the equilibrated column. The HPLC method is summarized in appendix table A.1. When the flow through eluted from the column, buffer B (no urea, 10% glycerol, 100mM NaCl, 20mM Tris, pH 8.0, and 0.6% SDS) was increased from 0% to 34% over 30 seconds. After 25 minutes at 34%, buffer B was increased again to 50% over 30 seconds. This was repeated for 66%, 84%, and 100% buffer B. After 25 minutes of equilibraion of the column with 100% buffer B, the protein was eluted from the column using an elution buffer containing 0.4% SDS, 10% glycerol, 100mM NaCl, 20mM Tris, pH 8.0, and 300mM.

### 3.2.2 The New Purification Method

Although it is reported by Ni and Zook *et al* that Method 1 could be used with a selection of very few other detergents; the yield was reported to be very low [25]. I was only been able to get $OEP16^{His}$ in appreciable amounts with the detergent SDS using Method 1. Therefore a new method was developed in order to fold $OEP16^{His}$ and elute the protein in high amounts in different detergents that might be more suitable for structure determination.

The overall strategy of the new method (Method 2) was to avoid the preincubation step and instead mask the protein while on the column with a harsh detergent to prevent non-specific, hydrophobic interactions with the column resin. I determined that preincubation with a mild detergent via dilution as done in the original method can lead to protein aggregation before the solution is even applied to the column.

Secondly, I reasoned that mild detergents may not be sufficient to mask the strongly hydrophobic regions from interacting with the column resin.

This new protocol is summarized in figure 3.1b, and the cornerstone of this new method is to use a single HPLC experiment to both purify and refold the protein. The strategy is to equilibrate the HPLC with the solution used to solubilize the protein from the inclusion bodies, inject the protein, and switch to a chaotropic buffer that also contains a high concentration of harsh detergent to prevent the protein from nonspecifically interacting with Ni-NTA resin. Afterward, the chaotropic, harsh detergent buffer is slowly removed and simultaneously replaced with a milder detergent buffer more suitable for tertiary folded membrane proteins.

The cell lysis and inclusion body washing steps remained largely unchanged for the new purification method. However, instead of 8M urea for solubilization of protein from inclusion bodies, a Guanidinium Hydrochloride(GuHCl) buffer was used instead (6M GuHCl, 100mM NaCl, 1mM β-ME, and 20mM Tris, pH 8.0). The Äkta explorer HPLC Ni-NTA column was then equilibrated with the same denaturing 6M GuHCl buffer. After the protein-GuHCl solution was injected onto the column, A buffer containing 8M urea, 10% glycerol, 100mM NaCl and 20mM Tris pH 8.0 was applied over the column to replace the GuHCl as chaotrope. This was necessary because sodium dodecyl sulfate (SDS) is insoluble in the presence of GuHCl and is needed in subsequent steps.

After binding of $OEP16^{His}$ to the column, buffer A (8M urea,10% glycerol, 100mM NaCl, 1% SDS, and 20mM Tris, pH 8.0) was added to the column and ran for two column volumes. In this step the protein folds in SDS micelles. Then every 10 minutes a 10% increase of buffer B was applied to the running buffer of the column. Buffer B contains 2x cmc milder detergent with no urea to exchange SDS. 100% buffer

B was applied to the column for two column volumes to remove any residual urea and SDS. Finally an elution buffer containing the same 2x cmc detergent as well as 300mM imidizole was used to elute the protein from the column. A detailed table for the HPLC method is described by Table A.2 in appendix A.

The new method established allowed for folding of $OEP16^{His}$ and elution in high yields with a large set of ionic and non-ionic detergents: the anionic SDS; the non-ionic detergents Cymal-4 (Cy4), β-D-dodecylmaltopyranoside (β-DDM), α-D-dodecylmaltopyranoside (α-DDM), and β-octylglucopyranoside (β-OG); the zwitterionic Sulfobetaine-12 (SB12), 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) and dodecylphosphocholine (DPC). The method can therefore be used for detergents with low and high cmc values (see table 3.2).

| Detergent | Purpose | cmc | Micelle MW | Ref |
|---|---|---|---|---|
| Deoxycholate | Wash inclusion bodies | 2.4mM | 13kDa | [54] |
| Mega-9 | Wash inclusion bodies | 16mM | - | [55] |
| Triton X-100 | Wash inclusion bodies and Refold Protein | 0.014% | 47kDa | [56] and [57] |
| SDS | Refold Protein | 8.3mM | 17kDa | [54] and [58] |
| β-DDM | Refold Protein | 0.165mM | 50kDa-76kDa | [59] |
| SB12 | Refold Protein | 2.8mM | 19kDa | [60] and [61] |
| Cymal-4 | Refold Protein | 7.3mM | 12kDa | [62] |
| β-OG | Refold Protein | 23.4mM | 8kDa-29kDa | [63] and [64] |
| CHAPS | Refold Protein | 8mM | 6kDa | [65] and [66] |
| DPC | Refold Protein | 1.5mM | 19kDa | [67] |

Table 3.2. Detergents used in the purification and refolding of OEP16. For more information regarding folding membrane proteins in detergent micelles see Wiener *et al.* [6] and Vergis *et al* [68]. Table adapted from Ni and Zook *et al* [25].

## 3.3 Characterization of OEP16

The isolated $OEP16^{His}$ was characterized through multiple biophysical methods including SDS polyacrylamide gel electrophoresis (SDS-PAGE), gel filtration chromatography, light scattering, circular dichroism (CD, nuclear magnetic resonance (NMR) spectroscopy, and X-Ray crystallography. Each of these particular experiments provides a certain level of insight into the homogeneity and structure of OEP16.

### 3.3.1 SDS-PAGE

To determine the purity of OEP16, the quickest and most direct way to measure is through SDS-PAGE. The gel was prepared in a vertical slab apparatus provided by Bio-Rad using a strategy adapted from Schägger [69] . First the 12% resolving gel was prepared by adding 1.5mL of 40% acrylamide:bis solution, 2.65mL of 3x tris stock solution (0.3% SDS, 3.0M Tris, pH 8.45) and 0.83mL glycerol to a 15mL falcon tube. The solution was mixed to homogeneity before adding 10µL of 40% ammonium persulfate and 5µL TEMED. The final solution was inverted three to four times before using a syringe to add the polymerizing solution to the vertical gel slab apparatus. An additional 0.25mL of water was gently added to the top of the solution to remove bubbles that may have formed in between the plates and hinder the gel surface to dry out. The water was removed before the stacking gel solution was applied.

After the 12% gel was fully polymerized, a 5% stacking gel was prepared in a similar fashion. 0.313mL of 40% acrylamide, 0.625mL 3x tris buffer, 1.55mL $H_2O$ (instead of glycerol) was added to a 15mL falcon tube, with the 10µL ammonium persulfate and 5µL TEMED added last to prevent gel formation outside of the gel mold plates. A 10-well comb was inserted into the top of the gel to provide sample reservoirs.

Once the entire gel was finished polymerizing, the gel plate was placed in a Bio-Rad bracket that creates a sealed internal compartment separate from the outside basin.

Inside running buffer (0.1M tricine, 0.1% SDS, and 0.1M Tris base, pH 8.9) was used to fill the inside reservoir completely. The outside basin was filled with roughly 500mL of outside running buffer that contained 0.1M Tris, pH 8.9. The comb from the gel plate was removed to allow for sample to be added.

For sample preparation 10µL of 4x SDS-PAGE sample buffer (26.7% SDS, 17.3% β-mercaptoethanol, 26.7% Glycerol, 133mM Bromophenol Blue, and 133mM Tris, pH 8.0) was added to 30uL of protein sample containing between 10µg and 20µg of protein. This solution was mixed via pipetting and was incubated for 5 minutes at 95°C. For a 10-well comb, roughly 30uL of sample was gently added to each of the wells. Before the gel is run, 10µL of a pre-prepared Precision Plus Protein™ Unstained Standard (Bio-Rad #161-0363) protein ladder (Bio-Rad) was added to slot one of the gel plate. To separate the proteins via gel electrophoresis a voltage of 173V was applied to the gel until the bromophenol blue dye reached the resolving gel, then voltage was increased to 245V. The experiment was stopped once the dye front reached the bottom of the gel plate.

The gel was removed and dyed using Coomassie Blue (0.1% CBB G250, 30% methanol, 10% acetic acid) stain for 5 minutes. Afterwards a harsh destaining solution (50% methanol, 10% acetic acid) was applied to the stained gel for 1 to 5 minutes to remove the majority of the unbound stain. The gel was then incubated in a less harsh destaining solution (5% methanol, 10% acetic acid) overnight. Gels were imaged with a Canon A590 PowerShot camera.

### 3.3.2 Size Exclusion Chromatography

All gel filtration chromatography experiments were performed on a 24mL Superdex 200 10/300 GL column (GE Healthcare). An Äkta explorer HPLC system was used for all gelfiltration runs with a 0.5mL/min flowrate. For analytical runs a 100µL

sample of the protein elutate from the Ni-NTA purification (200µM and 500µM) was injected onto the column. In the chromatography experiment absorption of protein elution was monitored at 280nm. Chromatography experiments were allowed to run for 30mL to ensure that all material flowed through the column.

A standard curve was prepared using both high molecular weight (#17-0441-01 and #28-4038-42) and low molecular weight (#28-4038-41) protein standards provided by GE healthcare. A high ionic strength buffer was prepared containing 100mM sodium phosphate, pH 7.2 with 100mM NaCl. Protein standards were prepared in two separate mixtures, A and B with concentrations ranging between 0.3mg/mL and and 5mg/mL described in table 3.3. The Superdex 200 10/300 GL gelfiltration column was equilibrated with 50mL of the high ionic strength buffer. 100µL of protein mix solution was injected onto the column at a flowrate of 0.5mL/min. Absorbance was measured at 280nm.

| Protein | MW (kDa) | $R_H$ (nm) | Concentration (mg/mL) | Kit |
|---|---|---|---|---|
| Aprotinin | 6.5 | - | 3.0 | LMW |
| Ribonuclease A | 13.7 | 1.64 | 3.0 | LMW |
| Carbonic Anhydrase | 29.0 | - | 3.0 | LMW |
| Ovalbumin | 44.0 | 3.05 | 4.0 | LMW/HMW* |
| Conalbumin | 75.0 | - | 3.0 | LMW/HMW* |
| Aldolase | 158.0 | 4.81 | 4.0 | HMW |
| Catalase | 232.0 | 5.12 | 2.0 | HMW** |
| Ferritin | 440.0 | 6.10 | 0.3 | HMW |
| Thyroglobulin | 669.0 | 8.50 | 5.0 | HMW |

Table 3.3. Proteins standards used for size determination by SEC. * Indicates proteins used in only in kit #28-4038-42, ** refers to the protein used only in kit #17-0441-01.

### 3.3.3 Circular Dichroism Spectroscopy

Circular Dichroism was performed on a home-source for circularly polarized light using a Jasco J-710 circular dichroism spectropolarimeter as well as circularly

polarized light generated by synchrotron radiation. The latter experiments were performed in the form of a collaboration with Dr. Bonnie Wallace from the Institute of Structural and Molecular Biology at Birkbeck College University of London. The requirements of the protein sample for synchrotron radiation circular dichroism (SRCD) differs significantly from that of home-source circular dichroism, as well as the experimental setup. Therefore they are discussed independently.

3.3.3.1 Conventional CD Spectroscopy

It is not possible to perform CD spectroscopy on $OEP16^{His}$ directly eluted off of the Ni-NTA column due to the strong absorption of light by imidizole, glycerol, and NaCl in the wavelength range (240nm to 190nm) required for circular dichroism spectroscopy. Therefore it was necessary to perform a buffer exchange with a solution more amiable for CD measurements.

Buffer exchange was performed using a 'quick dialysis method' I developed in order to speed the dialysis process and reduce the buffer volume needed for it. Protein eluted from the Ni-NTA purification and refolding step was concentrated to ~5mg/mL using appropriately sized ultrafiltration devices in order to concentrate the $OEP16^{His}$-detergent micelles without co-concentration of the empty detergent micelles (Filter cut offs were 100kDa for Triton X-100 and 30kDa for the rest). Concentration of the samples was performed at low centrifugation speeds between 1,000 and 2,000 x g for 2 minute increments and mixed via pipetting in between concentration steps in order to maintain a homogenous protein concentration within the ultrafiltration device with the aim to avoid aggregation of the protein during the ultrafiltration step.

The concentrated protein was then diluted 100x with a CD buffer containing 100mM NaF, 1mM β-ME, 2xcmc detergent (for all detergents tested with the exception of SB12 which was at concentrations of 10x cmc to maintain solubility) [31]. The

protein solution was re-concentrated using the same method as described above. This procedure was repeated once more to remove any traces of chemicals that interfere with the CD meeasurement. The final solution was diluted to a concentration of about 0.2mg/mL [31]. 200µL of the sample was placed in a 1mm quartz CD spectrophotometer cuvette. CD spectra were performed on a Jasco J-710 spectropolarimeter, monitoring CD at wavelengths from 190 to 260nm with a 1nm bandwidth, averaging over five scans at 1nm increments.

Circular dichroism spectra was also measured at a single wavelength as a function of temperature in order to collect data on the thermal stability of OEP16 in various detergent conditions. Samples were prepared in a similar fashion and concentrations as used for near UV CD measurements. CD was measured at 222nm and the temperature was increased linearly. Temperature was increased at a rate of 1°C/minute starting at 20°C and data was collected to about 105°C.

3.3.3.2 Synchrotron Radiation Source CD

SRCD was measured at beamline CD1 on ASTRID at the Institute for Storage Ring Facilities located at Aarhus University in Denmark in collaboration with Bonnie Wallace's group. As explained in section 2.1.2, SRCD is not as sensitive to light absorbing buffers compared to home-source CD [34]. Therefore sample preparation of $OEP16^{His}$ for SRCD measurements was similar to that of near UV CD measurements described above, with solutions containing 1mg/mL $OEP16^{His}$, and refolding buffer, pH 7.2. $OEP16^{His}$ CD spectra were measured with a pathlength of $5x10^{-3}$cm and a bandwidth of 0.5nm scanning between 175 and 280nm. Not all $OEP16^{His}$-detergent micelle conditions were available at the time of measurement, therefore only $OEP16^{His}$ in SDS, DPC, and SB12 were used.

3.3.3.3 CD Data Analysis for Secondary Structure

Secondary structure of the protein was predicted based on the circular dichroism data using the programs Contin/LL, SELCON 3, and CDSSTR provided by the CDPro software package made available by Narasimha Sreerama at Colorado State University. Reference protein sets used for Jasco J-710 CD measurements included SP43 (43 soluble proteins), SDP48 (SP43 + five denatured proteins), and SMP56 (SDP48 + 13 Membrane proteins). More basis sets were available for SRCD data due to the very far UV wavelengths used in the CD measurements. The very far UV spectrum can provide more detailed information regarding the secondary structure of proteins, including 3-10 and polyproline helices [34]. These include the same three basis sets used for home-source CD measurements as well as SP37 and SP37A which both have 37 soluble proteins, SDP42 (SP37 and five denatured proteins, and SMP50 (SP37 + 13 membrane proteins) [35, 70-86].

Algorithms provided by CDpro were written for DOS based computers between 15 and 20 years ago. These programs could take several minutes to calculate a result, depending on the quality of CD data; this was especially true for the CDSSTR algorithm. The availability of a large amount of data from home-source CD measurements combined with the desire to obtain secondary structure calculations using all possible basis sets demanded a more automated and user friendly interface was needed.

To handle the high volume of CD data collected in this study, I developed and programmed a Graphical User Interface (GUI) software program, CDProg Fit, that was capable of interfacing with the FORTRAN encoded SELCON3, CONTIN/LL, and CDSSTR deconvolution algorithms used to predict secondary structure. This was accomplished by using the programming language Python to create a single window environment where all three programs were ran at once and iteratively ran through all

basis sets provided by CDPro (Figure 3.2). No modifications to the original programs were made. Python was capable of communicating (or PIPEing) to other programs, and this was how CDProg Fit interfaced with the other executable files in the CDPro package. CDProg Fit configured the CD data measured into a format that was read by the deconvolution algorithms, then executed them and read the output. The output was then collected by CDProg Fit and displayed as a fraction of α-helices, β-sheets, β-turns, and random coils along with providing a confidence level reported by the deconvolution program in the form of RMSD or NRMSD.

To operate CDProg Fit, the CD data in a spreadsheet with the first column being the wavelength (1nm increments) and the second column being the CD measurement, was cut and pasted into the field provided. Another tool that made CDProg Fit very useful was that CD data could be analyzed with the input data being in input units other than Δe or Molar Ellipticity. As long as the protein concentration, cell path length, and number of amino acid residues was known and entered into the fields provided, raw CD data measured directly in millidegrees or degrees could be used as well. Lastly the basis set is chosen (usually selecting to calculate using all sets) and then the command button Run Fit was used to execute the program.

Although no changes were made to improve the results returned by the deconvolution algorithms, a higher throughput method was possible using CDProg Fit by providing a GUI for a straight forward, simple method that was able to screen a measurement against all basis sets. This allowed the program to execute the deconvolution algorithms (sometimes totaling 30 minutes of calculations) while the analyst performed other duties.

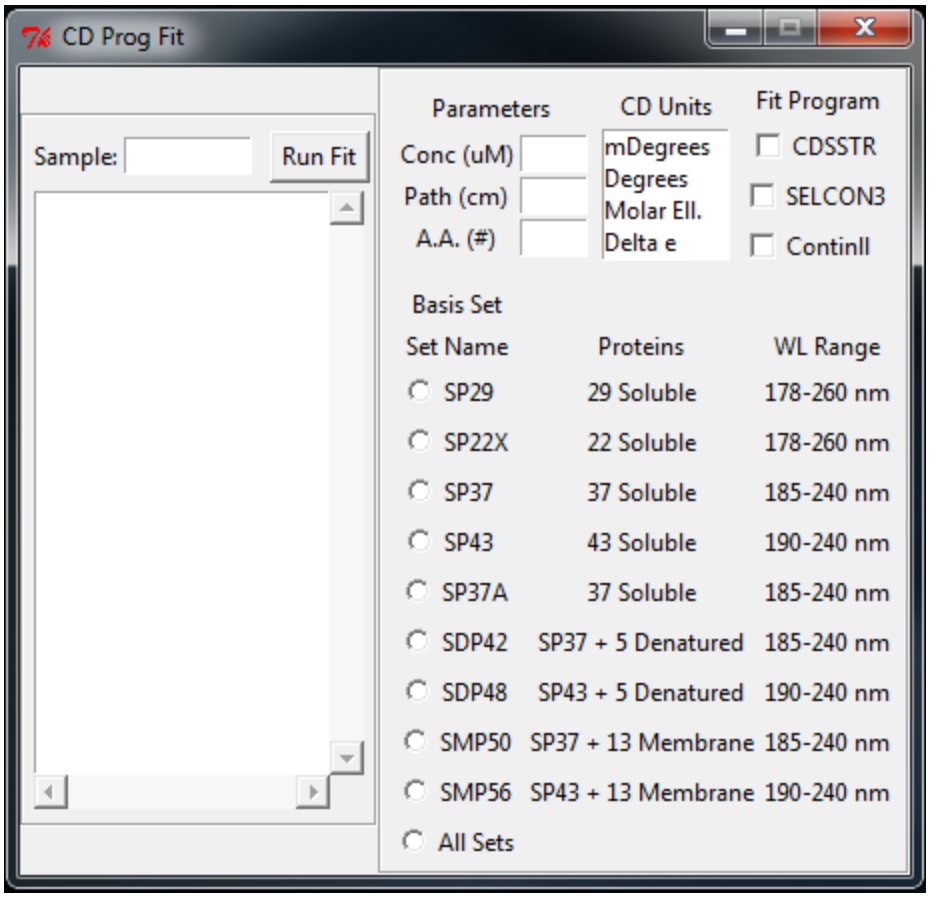


Figure 3.2. The GUI interface for CD data deconvolution. Named CD Prog Fit, was developed for more efficient deconvolution of CD data. This program was designed to interface with the DOS based fortran CDPro developed by Sreerama *et al*. The CD data was put the large field on left while necessary parameters (concentration, cuvette pathlength, number of amino acids) were input along with CD units used. Additionally the deconvolution program and reference sets are chosen for deconvolution. When all inputs are entered, the button 'Run Fit' is pressed to execute the deconvolution programs selected and report the secondary structure prediction.

### 3.3.4 Dynamic Light Scattering Spectroscopy

To observe particle size of $OEP16^{His}$-detergent micelles, dynamic light scattering (DLS) was employed to obtain a diffusion coefficient in order to elucidate a hydrodynamic radius ($R_H$); the theory of which is discussed in detail in section 2.1.3. Dynamic light scattering experiments were performed on either a Wyatt Dynapro

NanoStar operating at a laser wavelength of 658nm or a Molecular Dimensions SpectroSize 302 with a 785nm Laser.

For the Wyatt instrument, 100µL samples were prepared at 1mg/mL using the refolding buffer used to fold $OEP16^{His}$. The refolding buffers used containing no $OEP16^{His}$ were also measured to compare sizes and signal strengths of the empty detergent micelles. Scattered light was measured with a detector at a 90° angle. Ten five-second measurements were made and repeated twice.

The Molecular Dimensions instrument had a unique setup with the ability to measure DLS in small protein solution volumes between 1µL and 5µL directly in a hanging drop of a crystallization plate, shown in figure 3.3. Protein samples ranging between 1mg/mL to 10mg/mL were pipetted onto a glass cover slide at volumes between 3µL and 5µL. Reservoirs contained the identical buffer conditions as the protein solution. An angle of 150° was used to detect the scattered light. Samples were measured ten times with 20 second measurements.

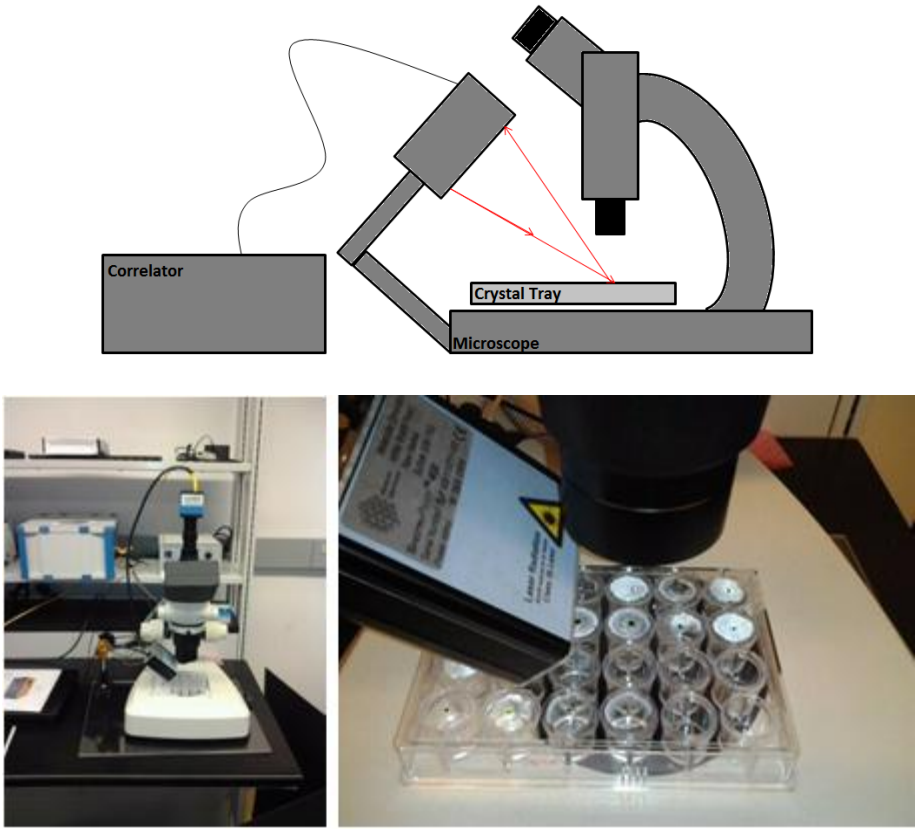


Figure 3.3. Molecular Dimensions DLS setup. Shown both as a schematic (top) and pictorially (bottom). This allowed DLS to be performed directly onto a crystallization droplet with volumes down to 1µL. The laser beam was aligned onto the droplet using an IR camera attached to a microscope, the scattered light from the laser (red arrow) was sent to the correlator via a fiber optic cable.

### 3.3.5 Tryptophan Fluorescence Microscopy

During the study of OEP16$^{His}$ formation of very thin crystalline material was observed. One of the ways to characterize the composition of the material was to use tryptophan fluorescence coupled to a microscope. This would reveal if the material was protein by exciting the tryptophan residues at 280nm and look for a fluorescent image at emission wavelength 360nm. If there was a fluorescent image present, the crystalline material was likely protein.

Typically the formation of precipitate would occur during buffer dialysis necessary for CD and NMR spectroscopy studies. This happened in both conventional

dialysis (observed specifically with β-OG detergent buffers) or during the 'quick dialysis' method using ultrafiltration membranes, observed with both DPC and Cymal-4 containing detergents. When the formation of precipitate was observed, 5µL of the solution was pipetted onto the cover slide and placed upside-down in a Limbo 24-well plate containing 500µL reservior solution (dialysis buffer). Drop was imaged using a JANSi UVEX UV microscope with CrystalDetect software, exciting at 280nm and monitoring at 360nm.

### 3.3.6 Nuclear Magnetic Resonance Spectroscopy

#### 3.3.6.1 Protein Conditions for NMR Analysis

Isotope-labeled protein samples were concentrated to concentrations between 0.8mM and 1.2mM and dialyzed against an NMR buffer via the 'quick dialysis' method described above. Many different buffer conditions were attempted in order to maintain protein stability; this involved various salt conditions, differing pH, and other additives. The most reliable NMR buffer contained 100mM NaCl to help with the highly cationic protein, 20mM sodium phosphate, pH 7.2, 1mM β-ME to ensure reduced cysteine residues, 10% $D_2O$ for NMR deuterium lock needed for NMR experiments, 10% Glycerol for improved stability of protein discovered by Ni and Zook, *et al* [25], 2x cmc detergent to maintain micelle formation, 1mM EDTA to remove paramagnetic ions (such as $Ni^{2+}$ that leeched from the Ni-NTA column), 1mM 2,2-Dimethyl-2-silapentane-5-sulfonic acid (DSS) for NMR referencing and 0.02% sodium azide to prevent bacterial growth.

#### 3.3.6.2 NMR Experiments

Most of the NMR spectra of $OEP16^{His}$ in SDS were collected at the University of Arizona, Tucson through a collaborative effort from Dr. Michael Brown and the UA NMR facility on a Varian Inova 600 using a cryoprobe. Samples include a 1.5mM $^{15}N$

labeled sample as well as a triply labeled ($^{13}C/^{15}N/^{2}H$) sample at 1mM (Refer to table 3.4 for complete list of samples and experiments). All experiments were performed at 37°C unless otherwise stated. Experiments performed on the $^{15}N$-labeled samples were: a two dimensional $^{15}N,^{1}H$-HSQC, a $T_1$ relaxation experiment on 2D $^{15}N$-HSQC at times sampled at 10ms, 20ms, 30ms, 50ms, 100ms, 200ms, 300ms, 500ms, 800ms, 1000ms, 1500ms and 2000ms, a $T_2$ relaxation experiment on 2D $^{15}N$-HSQC at times sampled at 10ms and every 20ms after until 190ms, and the three dimensional experiments NOESY-$^{15}N$-HSQC, and TOCSY-$^{15}N$-HSQC [87, 88]. The relaxation experiments were also repeated at 25°C. Experiments performed on the triply labeled sample include HNCO, HN(ca)CO, HNCA, HN(co)CA, HNCACB, and CBCA(co)NH [89-93].

| | SDS | | DPC | β-DDM | | α-DDM | SB12 |
|---|---|---|---|---|---|---|---|
| NMR Experiment | $^{15}N$ | $^{15}N/^{13}C/^{2}H$ | $^{15}N$ | $^{15}N$ | $^{15}N/^{13}C$ | $^{15}N$ | $^{15}N$ |
| 2D NHSQC | X | X | X | X | X | X | X |
| 2D NHSQC T1 | X | | | | | | |
| 2D NHSQC T2 | X | | | | | | |
| 2D Arg Titration | X | | | | | | |
| 2D $D_2O$ Titration | X | | | | | | |
| 3D HNCO | | X | | | | | |
| 3D HN(CA)CO | | X | | | | | |
| 3D HNCA | | X | | | X | | |
| 3D HN(CO)CA | | X | | | | | |
| 3D HNCACB | | X | | | | | |
| 3D CBCA(CO)NH | | X | | | | | |
| 3D TOCSY-NHSQC | X | | | | | | |
| 3D NOESY-NHSQC | X | | | | | | |

Table 3.4. NMR experiments performed for different OEP16-micelle conditions and isotopes. The X marker indicates experiment performed for the specific sample listed.

Several other $^{15}N$-HSQC experiments were performed on OEP16 in other detergent conditions using a 800MHz Ultra-Shield Varian NMR at the Magnetic

Resonance Research Center (MRRC) located at Arizona State University with the collaborative efforts of Dr. Jeff Yarger and Dr. Brian Cherry. $^{15}$N-HSQC experiments were performed on OEP16$^{His}$ in SB12, DPC, SDS, α-DDM and β-DDM. Experiments were performed at either 37°C or 45°C in the case of α-DDM and β-DDM. Measurements performed at ASU did not involve the use of a cryoprobe. The $^{15}$N-HSQC NMR experiment was performed as a way to screen for possible mild-detergent conditions where more in-depth 3D NMR experiments could be performed.

All data collected was converted into NMRPipe format in order to perform the necessary mathematical transforms which included solvent suppression, apodization by application of a sine-bell window, zero filling, appropriate phase shifting, linear prediction, and complex fourier transforms [94]. The linear predicted and fourier transformed data was then exported into SPARKY format for peak picking, strip plotting, and resonance assignment. SPARKY was also used to calculate the relaxation times of the individual resonances of the $^{15}$N-HSQC spectra [95].

Resonance assignment was performed using chemical shift statistics gathered by the Biological Magnetic Resonance data Bank (BMRB) at the University of Wisconsin [42]. An amino acid was selected as a candidate for a resonance peak when it fit within one standard deviation of the mean for $C_\alpha$, $C_\beta$, Co, and N chemical shifts. Areas of ambiguity were resolved using strip plots to determine sequential amino acid resonances.

Chemical shift data from the three dimensional experiments were used by TALOS in an attempt to obtain phi-psi torsion angles for secondary structure analysis, estimation of order parameters ($S^2$) and for future use in three dimensional structure determination [43, 44]. TOCSY-$^{15}$N-HSQC data was in-part used to identify the $H_\alpha$ chemical shift for TALOS calculations. NOESY-HSQC data was also used to walk the

protein backbone and assign $^{15}N$ resonances where $^{13}C$ chemical shift peaks were ambiguous.

Titration experiments were also performed to investigate the structural and ligand binding properties of OEP16 in SDS micelles. For these titration experiments, A 100% $D_2O$ solution containing 100mM arginine was titrated directly into the NMR tube. 2D $^{15}N$-HSQC experiments were performed for arginine concentrations of 0mM, 0.5mM, 1.5mM, 7.7mM, and 12mM. $D_2O$ concentrations increased for each addition of solution leading to concentrations of 10%, 10.4%, 11.2%, 13.3%, 16.1%, and 19.2% $D_2O$ in the NMR $OEP16^{His}$ sample.

### 3.3.7 Crystallization Attempts on OEP16

Hanging drop and sitting drop experiments using the vapor diffusion method were performed on $OEP16^{His}$ in various detergents in an attempt at crystallization for X-Ray diffraction studies. A 96-well sitting drop method was used for broad, commercial based screening, while a 24-well hanging drop method was used for fine screening. Protein concentrations between 10mg/mL and 15mg/mL were used in these studies.

Several commercial screens were utilized in an attempt for crystallization. These kits include: Nextal's Classics, MbClass I, MbClass II, and PEGs suites as well as Precipitant Synergy 64 crystallization screen provided by Emerald Biosystems. All crystallization trays were incubated in the dark at 20°C. Several fine screening attempts were made on conditions where amorphous precipitate was formed using the Make Tray web resource provided by Hampton Research.

### 3.3.8 Sequencing $OEP16^{His}$

The *oep16pea* gene in the pET21b plasmid vector was sequenced for primary structure in order to obtain information on the C-terminal tail, this was necessary because the primary sequence is necessary for assignment of resonances. 3mL cultures of BL21

(DE3) cells containing the plasmid construct were grown overnight selected by 100ug/mL ampicillin. Plasmid was isolated using the Biorad miniprep kit. Quantity and purity was assessed using 260/280nm absorption. This ratio was useful because it revealed how much DNA is present compared to the presence of protein. 300ng of DNA was submitted to the DNA sequencing facility at the School of Life Sciences, Arizona State University, Tempe. T7 universal primers were used for sequencing both in forward and reverse directions. Obtained sequences from the facility were used to elucidate the amino acid sequence of OEP16 discussed in the results section.

# 4. Experimental Results

## 4.1 Sequencing Results

The correct primary structure of a protein is extremely important in the resonance assignment of proteins by NMR spectroscopy. The sequence of $OEP16^{wt}$ has been known for decades, but the sequence of $OEP16^{His}$ has never been published. In order to assign the C-terminal tail residues along with the 6xhistidine tag, sequencing the plasmid for $OEP16^{His}$'s primary structure was necessary. This was done using the modern Sanger technique involving the use of fluorescent dyes (for more details see materials and methods section 3.3.8). The results of plasmid sequencing revealed the full primary structure of the $OEP16^{His}$ construct that includes the C-terminal tail and 6x histidine tag:

MPRSSFSGSLSSPKLDVVIDMGNPFLNLTVDGFLKIGAVAATRSVAEDTFHIIRKG
SISSNDFEKSLKKMCKEGAYWGAIAGVYVGMEYGVERIRGTRDWKNAMFGGA
VTGALVSAASNNKKDKIAVDAITGAAIATAAEFINYLT-LEHHHHHH

The dash in between T146 and L147 separates the final residue of wild type OEP16 and the C-terminal tail and histidine tag. Primary structure analysis calculates the molecular weight of the protein to be 16555Da, a pI of 8.69, and an extinction coefficient ($\lambda$ = 280nm) of $16960 M^{-1}cm^{-1}$. These numbers were used in all calculations from UV-Vis spectral studies and CD spectroscopy.

## 4.2 Purification of OEP16

The original purification and refolding method (Method 1) established by Ni and Zook *et al* was optimized for obtaining high yields of $OEP16^{His}$ in SDS micelles [25]. The method (described schematically in methods figure 3.1a) involved expressing $OEP16^{His}$ as inclusion bodies; washing the inclusion body pellets using detergents to remove cell debris; solubilizing the inclusion bodies; purifying the inclusion bodies using HPLC driven Ni-NTA chromatography (seen in figure 4.1a); masking the hydrophobic

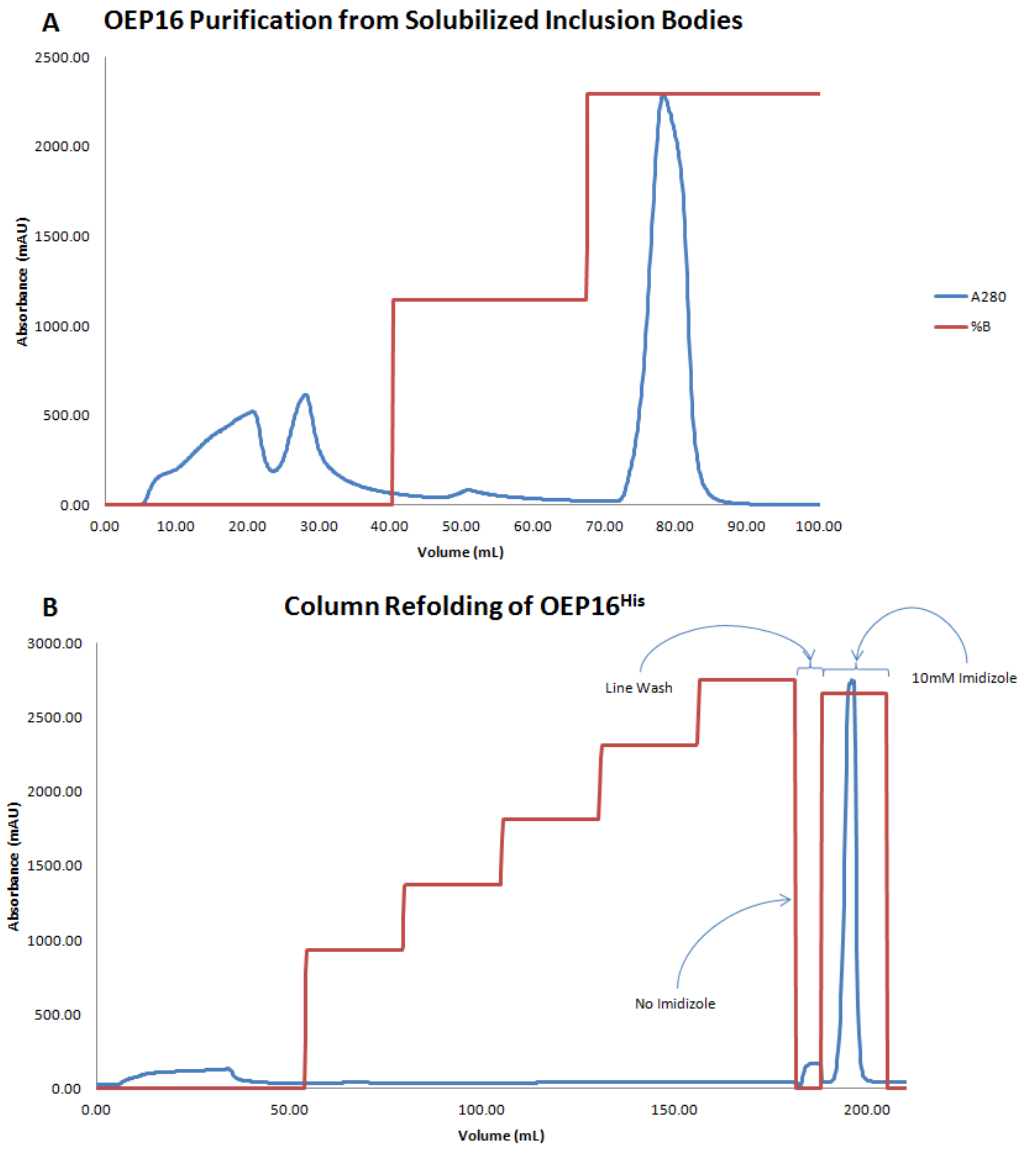


Figure 4.1. HPLC Method 1 used for isolating and purifying OEP16 in SDS micelles. (A) Solubilized inclusion bodies are injected onto an Ni-NTA column equilibrated with Buffer A contianing 8M Urea at pH 8.0; after nonbinding protein flows through, 50% of Buffer B containing 8M Urea at pH 4.0 is added to the column (red line) to remove non-specific binding protein at a pH of 7.2; elution of pure $OEP16^{His}$ is obtained when %B is increased to 100%. (B) Refolding of masked OEP16 on the Ni-NTA column. The strongly hydrophobic regions of OEP16 eluted from A were masked and partially refolded by dilution to 3M Urea and adding 1% SDS; the protein was then injected onto the Ni-NTA column equilibrated with Buffer A containing 3M Urea and 1% SDS; a stepwise gradient was then used to slowly remove the urea and lower the SDS concentration to 0.6%. Elution of protein was performed by adding an imidizole buffer containing 0.4% SDS in a stepwise fashion: 10mM imidazole and then 300mM imidazole. Absorbance was monitored at 280nm (blue line). Surprisingly, all $OEP16^{His}$ eluted at 10mM Imidizole.

regions of the inclusion bodies using a dilution method; and then refolding the protein directly onto the Ni-NTA column as seen in figure 4.1b [25]. Surprisingly, only 10mM imidazole was needed to elute $OEP16^{His}$ completely from the column. Average yields of reconstituted OEP16 in SDS using Method 1 were 25mg per one gram cell pellet or about 50mg of protein per 1L cell culture. This provided enough material for very good signal to noise ratio for NMR experiments (see section 4.6).

However, in order to support a tertiarily folded OEP16 protein in SDS micelles, a method was needed to obtain high yields of OEP16 in milder detergents traditionally more suited to study membrane proteins. Method 1 was used as a foundation to develop a method for purifying and refolding OEP16 in a wider range of detergent conditions (Method 2) shown schematically in methods figure 3.1b . Similar to Method 1, $OEP16^{His}$ was expressed as inclusion bodies, washed, and then solubilized using chaotropic buffers [25]. However, unlike Method 1, the protein was purified and refolded using a single HPLC run shown in figure 4.2. After OEP16 in 6M GuHCl was bound to the Ni-NTA column, the protein was washed using 8M Urea and then masked using SDS in a 1% SDS and 8M Urea buffer. The protein was then slowly refolded using a mild detergent buffer containing no urea. The protein was finally eluted with 300mM imidizole with 2x cmc mild detergent. Successful purification attempts were made using anionic detergent SDS, zwitterionic detergent DPC, and non-ionic detergents Cymal-4, β-DDM, α-DDM, and Triton X-100. Yields of protein were calculated by measuring absorbance of eluant at 280nm and then multiplied across entire volume collected. The reported yields are normalized by cell pellet mass used for the protein preparation shown in table inset in figure 4.2. This procedure allowed for the purification and reconstitution of OEP16 in a wide range of detergents; however, the purification using the detergents SB12, CHAPS,

and β-OG were considered unsuccessful and is discussed in detail in discussion section 5.4.

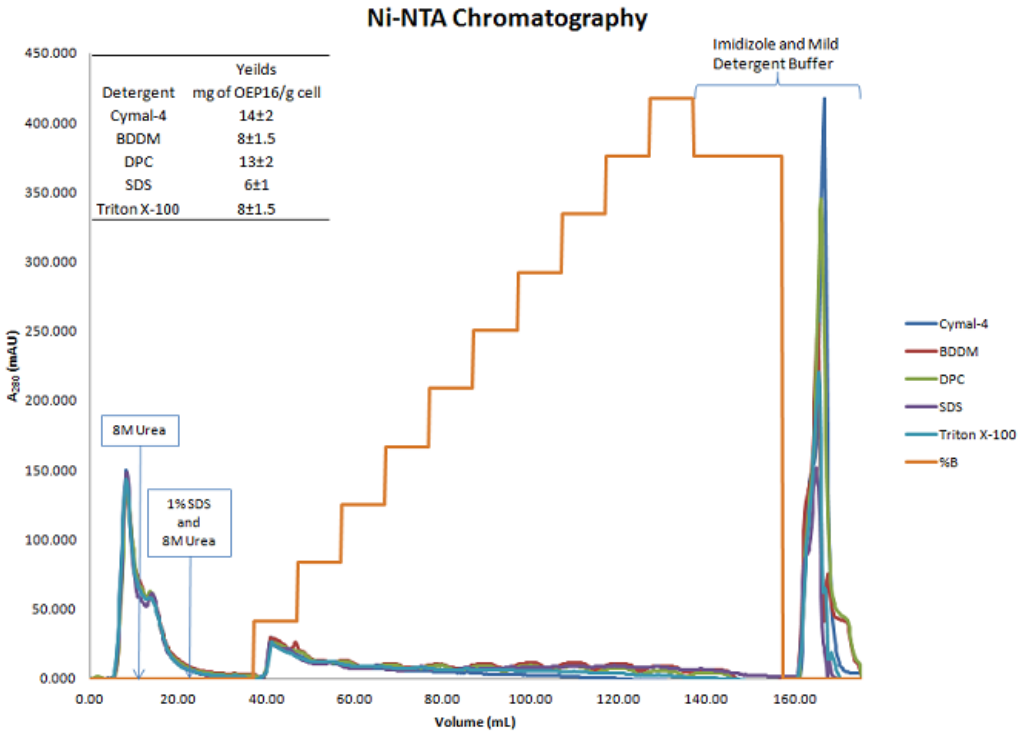


Figure 4.2. HPLC for Method 2. Ni-NTA chromatogram for purifying and refolding OEP16 in various detergents. Three peaks are generally seen: One for unbound flow through, another for nonspecifically bound protein eluted with SDS was introduced, and the final peak when 300mM Imidizole is added. Included are typical yield results for a 1g cell pellet. A 1L prep usually generates between 2.5g and 3g of cell pellet, allowing for yields of protein ranging between 17mg and 42mg. Refer to materials and methods 3.2.2 for buffer contents.

## 4.3 Circular Dichroism Spectroscopy Results

### 4.3.1 CD Results for OEP16 in Different Detergent Conditions

One of the most used ways of obtaining information about the secondary structure composition of a protein is to use circular dichrism (CD). This can provide the overall secondary structure content of OEP16 in different detergents. OEP16 displayed a considerable variety of CD spectra depending on the detergent that was used, as shown in figure 4.3. For all detergents, the spectra depicted two negative bands: one at about

220nm and the other at 208nm. Along with the negative bands, one positive band at 192nm was also observed indicative of α-helices. Although the spectra maxima and minima shifted slightly depending on the detergent, the significant difference was in the intensities of these bands which varied between the detergents.

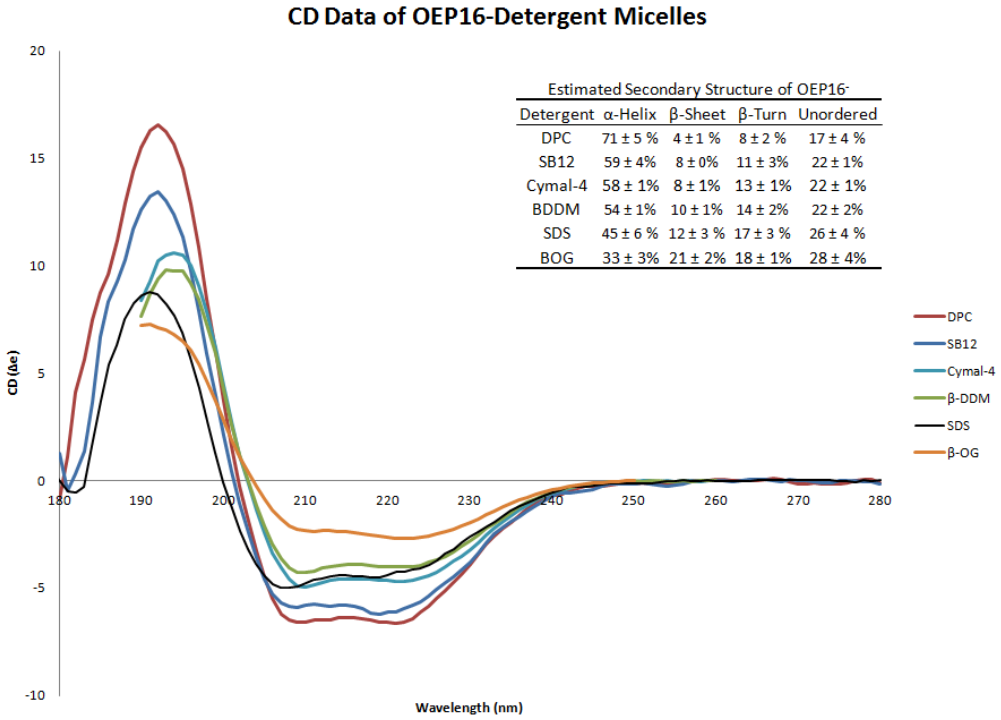


Figure 4.3. Circular dichroism spectra of OEP16 in various detergents. All display the double minima characteristics of an α-helical protein with intensities that vary depending on the detergent used to refold OEP16. The table included reports the average secondary structure content calculated by CDSSTR and CONTIN/LL using appropriate basis sets for the data collected. DPC, SB12, and SDS were collected using synchrotron radiation at CD1 on ASTRID. Cymal-4, β-DDM, and β-OG were collected using the in-house Jasco J-710.

CD spectra of OEP16 was collected using a home source spectropolarimeter (Jasco J-710) for all detergents investigated. However, with the collaborative effort from Bonnie Wallace's group at Birkbeck College, an opportunity to use synchrotron radiation circular dichroism (SRCD) became available for OEP16 in several detergents (DPC, SB12 and SDS). SRCD data collected was reliable to about 180nm (compared to the

~190nm using the Jasco J-710) which provided a wider range of available reference sets for use in deconvolution (SP37, SP37A, SDP42, and SMP50 described in method section 3.3.3.3).

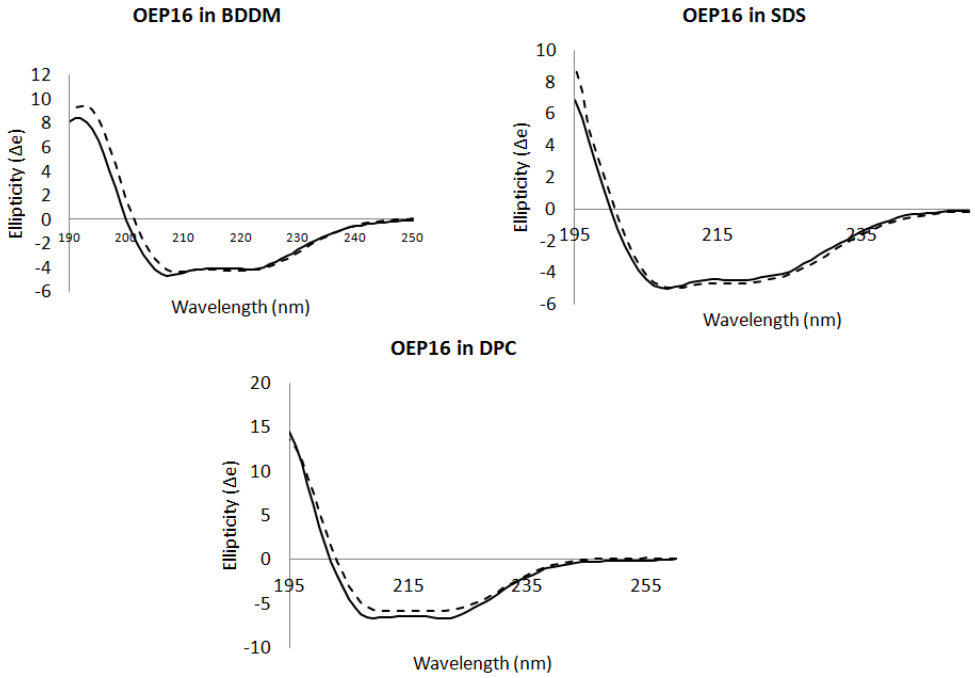


Figure 4.4. Comparison of SRCD data to in-house data. Circular Dichroism spectra taken from home source (dashed) and synchrotron source (solid) for OEP16 in three different detergents. Overall there is little difference between the spectra when the regions of high beam quality are considered.

The in-house Jasco J-710 CD data was comparable with the SRCD data of the protein samples measured using both methods shown in figure 4.4. The significant difference is that SRCD measurements were much more reliable and allowed for data collection to 180nm whereas the in-house measurements were reliable only to ~195nm. Deconvolution results are summarized in the table inset on figure 4.3 as an average of the deconvolution analysis from both Contin/LL and CDSSTR among all basis sets used (a comprehensive list can be seen in table 4.1) SELCON3 data reported similar results to

| Sample | Method | Basis | α-Helix | β-Sheet | β-Turn | Unordered | RMSD | NRMSD |
|---|---|---|---|---|---|---|---|---|
| SB12 | CDSSTR | SP37 | 0.637 | 0.082 | 0.105 | 0.18 | 0.156 | 0.023 |
| SB12 | Contnll | SP37 | 0.553 | 0.074 | 0.139 | 0.234 | 0.161 | 0.024 |
| SB12 | CDSSTR | SP43 | 0.631 | 0.071 | 0.091 | 0.21 | 0.171 | 0.027 |
| SB12 | Contnll | SP43 | 0.55 | 0.078 | 0.143 | 0.23 | 0.102 | 0.016 |
| SB12 | CDSSTR | SP37A | 0.642 | 0.038 | 0.109 | 0.181 | 0.221 | 0.033 |
| SB12 | Contnll | SP37A | 0.538 | 0.064 | 0.104 | 0.273 | 0.179 | 0.027 |
| SB12 | CDSSTR | SDP42 | 0.658 | 0.082 | 0.116 | 0.144 | 0.162 | 0.024 |
| SB12 | Contnll | SDP42 | 0.583 | 0.063 | 0.166 | 0.189 | 0.184 | 0.028 |
| SB12 | CDSSTR | SDP48 | 0.629 | 0.066 | 0.091 | 0.218 | 0.16 | 0.025 |
| SB12 | Contnll | SDP48 | 0.554 | 0.073 | 0.147 | 0.227 | 0.101 | 0.016 |
| SB12 | CDSSTR | SMP50 | 0.606 | 0.093 | 0.113 | 0.193 | 0.262 | 0.039 |
| SB12 | Contnll | SMP50 | 0.522 | 0.095 | 0.141 | 0.242 | 0.141 | 0.021 |
| SB12 | CDSSTR | SMP56 | 0.626 | 0.068 | 0.087 | 0.221 | 0.312 | 0.049 |
| SB12 | Contnll | SMP56 | 0.567 | 0.071 | 0.129 | 0.233 | 0.106 | 0.017 |
| DPC | CDSSTR | SP37 | 0.702 | 0.061 | 0.066 | 0.175 | 0.13 | 0.016 |
| DPC | Contnll | SP37 | 0.663 | 0.043 | 0.085 | 0.209 | 0.173 | 0.022 |
| DPC | CDSSTR | SP43 | 0.704 | 0.05 | 0.06 | 0.185 | 0.15 | 0.02 |
| DPC | Contnll | SP43 | 0.66 | 0.045 | 0.086 | 0.21 | 0.129 | 0.017 |
| DPC | CDSSTR | SP37A | 0.698 | 0.05 | 0.081 | 0.138 | 0.261 | 0.033 |
| DPC | Contnll | SP37A | 0.682 | 0.044 | 0.069 | 0.171 | 0.188 | 0.024 |
| DPC | CDSSTR | SDP42 | 0.806 | 0.053 | 0.058 | 0.08 | 0.144 | 0.018 |
| DPC | Contnll | SDP42 | 0.685 | 0.047 | 0.131 | 0.137 | 0.167 | 0.021 |
| DPC | CDSSTR | SDP48 | 0.787 | 0.048 | 0.066 | 0.097 | 0.18 | 0.024 |
| DPC | Contnll | SDP48 | 0.658 | 0.045 | 0.1 | 0.198 | 0.118 | 0.016 |
| DPC | CDSSTR | SMP50 | 0.76 | 0.032 | 0.048 | 0.155 | 0.246 | 0.031 |
| DPC | Contnll | SMP50 | 0.664 | 0.042 | 0.087 | 0.208 | 0.121 | 0.015 |
| DPC | CDSSTR | SMP56 | 0.766 | 0.024 | 0.041 | 0.164 | 0.303 | 0.04 |
| DPC | Contnll | SMP56 | 0.682 | 0.03 | 0.084 | 0.204 | 0.096 | 0.013 |
| SDS | CDSSTR | SP37 | 0.524 | 0.085 | 0.158 | 0.231 | 0.229 | 0.049 |
| SDS | Contnll | SP37 | 0.395 | 0.139 | 0.203 | 0.264 | 0.174 | 0.038 |
| SDS | CDSSTR | SP43 | 0.484 | 0.113 | 0.149 | 0.251 | 0.211 | 0.048 |
| SDS | Contnll | SP43 | 0.415 | 0.127 | 0.185 | 0.273 | 0.083 | 0.019 |
| SDS | CDSSTR | SP37A | 0.524 | 0.059 | 0.14 | 0.238 | 0.274 | 0.059 |
| SDS | Contnll | SP37A | 0.367 | 0.112 | 0.127 | 0.363 | 0.173 | 0.037 |
| SDS | CDSSTR | SDP42 | 0.513 | 0.122 | 0.173 | 0.191 | 0.155 | 0.034 |
| SDS | Contnll | SDP42 | 0.414 | 0.142 | 0.218 | 0.225 | 0.183 | 0.039 |
| SDS | CDSSTR | SDP48 | 0.483 | 0.111 | 0.125 | 0.285 | 0.167 | 0.038 |
| SDS | Contnll | SDP48 | 0.417 | 0.116 | 0.177 | 0.29 | 0.083 | 0.019 |
| SDS | CDSSTR | SMP50 | 0.463 | 0.14 | 0.176 | 0.227 | 0.233 | 0.05 |
| SDS | Contnll | SMP50 | 0.353 | 0.163 | 0.217 | 0.266 | 0.145 | 0.031 |
| SDS | CDSSTR | SMP56 | 0.482 | 0.118 | 0.139 | 0.266 | 0.234 | 0.053 |
| SDS | Contnll | SMP56 | 0.432 | 0.112 | 0.18 | 0.276 | 0.067 | 0.015 |
| Cymal-4 | CDSSTR | SMP56 | 0.595 | 0.059 | 0.123 | 0.222 | 0.104 | 0.02 |
| Cymal-4 | Contnll | SMP56 | 0.591 | 0.051 | 0.128 | 0.23 | 0.072 | 0.014 |
| Cymal-4 | CDSSTR | SDP48 | 0.58 | 0.091 | 0.148 | 0.18 | 0.113 | 0.021 |
| Cymal-4 | Contnll | SDP48 | 0.563 | 0.086 | 0.132 | 0.219 | 0.08 | 0.015 |
| Cymal-4 | CDSSTR | SP43 | 0.582 | 0.078 | 0.124 | 0.217 | 0.133 | 0.025 |
| Cymal-4 | Contnll | SP43 | 0.564 | 0.084 | 0.128 | 0.224 | 0.079 | 0.015 |
| β-OG | CDSSTR | SMP56 | 0.366 | 0.172 | 0.187 | 0.27 | 0.234 | 0.07 |
| β-OG | Contnll | SMP56 | 0.341 | 0.22 | 0.181 | 0.258 | 0.05 | 0.015 |
| β-OG | CDSSTR | SDP48 | 0.333 | 0.237 | 0.205 | 0.233 | 0.177 | 0.053 |
| β-OG | Contnll | SDP48 | 0.301 | 0.21 | 0.17 | 0.318 | 0.056 | 0.017 |
| β-OG | CDSSTR | SP43 | 0.342 | 0.216 | 0.18 | 0.262 | 0.164 | 0.049 |
| β-OG | Contnll | SP43 | 0.303 | 0.195 | 0.177 | 0.325 | 0.055 | 0.017 |
| β-DDM | CDSSTR | SP43 | 0.528 | 0.099 | 0.141 | 0.229 | 0.138 | 0.029 |
| β-DDM | Contnll | SP43 | 0.532 | 0.103 | 0.135 | 0.23 | 0.082 | 0.017 |
| β-DDM | CDSSTR | SDP48 | 0.526 | 0.114 | 0.165 | 0.188 | 0.137 | 0.028 |
| β-DDM | Contnll | SDP48 | 0.529 | 0.108 | 0.138 | 0.226 | 0.083 | 0.017 |
| β-DDM | CDSSTR | SMP56 | 0.56 | 0.096 | 0.118 | 0.222 | 0.125 | 0.026 |

Table 4.1. Total deconvolution results used to estimate secondary structure from CD data for OEP16 refolded in different detergents [96]. The method column indicates what algorithm was used (CONTIN/LL or CDSSTR), the basis column refers to the reference protein set used. Regardless of the algorithm or basis set used, the α-helix is the major secondary structure component.

CDSSTR and CONTIN/LL but reported a high RMSD and is therefore not considered in the deconvolution analysis. The differences between the deconvolution algorithms are described in theory section 2.1.2).

The α-helical content was lowest (33%) for β-OG and highest (71%) for DPC. β-sheet content ranged from between 4% in DPC and 21% in β-OG. Although the zwitterionic DPC reported the highest amount of α- helical content, the β-DDM sample also displayed a large amount of α-helices.

4.3.2 Melting Studies of OEP16

The stability of a protein was a high priority where NMR is concerned. NMR experiments can take days or weeks to complete and it was necessary that they remain in the same folded and multimeric state the entire time. Additionally, increasing the temperature during an NMR experiment can reduce the global isotropic correlation time which improves spectra line widths. Monitoring CD as a function of temperature can reveal information regarding the stability of a protein. I monitored CD changes at 222nm as a function of temperature for OEP16 in the following detergents: Cymal-4, β-DDM, β-OG, DPC and SB12. The results are shown in figure 4.5. The melting profiles of OEP16 in different detergents show a range of apparent $T_m$'s between 60°C and 70°C, with the exception of SDS which resisted denaturation at temperatures well above 100°C. The apparent $T_m$'s are summarized in the inset table. Normalization was performed on all non-SDS containing samples using the following expression [97]:

$$Norm(\theta_t) = 1 - \frac{(\theta_t - \theta_u)}{(\theta_f - \theta_u)}$$

Where $\theta_t$ is the CD measured at 222nm at temperature t, $\theta_f$ is the CD measured at the initial temperature, and $\theta_u$ is the CD at the final temperature. Secondary structure could not be restored upon cooling of the samples and precipitate accumulated in the CD

cuvette indicating the "melting" of OEP16 in detergent is an irreversible process. However, in the case of SDS, even in the presence of 5.8M urea, OEP16 did not precipitate and resisted denaturing up to 100°C as seen in the figure 4.5 inset. By visual estimate, the $T_m$ was estimated to be around 120°C, although mathematical extrapolation by nonlinear fitting could not be achieved. OEP16 in SDS without urea shows little change in CD measurements (data not shown) suggesting a very high level of thermal stability.

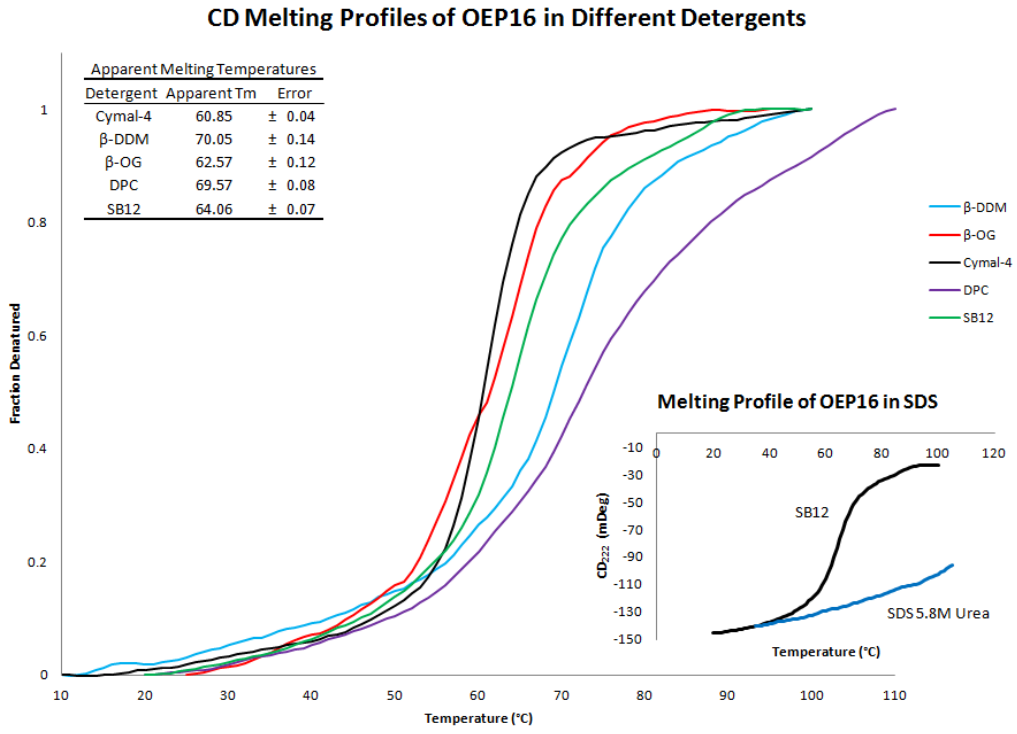


Figure 4.5. Melting profiles of OEP16 in different detergents. This process was irreversible with apparent $T_m$'s ranging between 60°C and 70°C shown in the table. The inset displays the melting profile of OEP16 in SDS and 5.8M urea (blue line), which resisted denaturing up to >100°C. Included is a melt profile for OEP16 in SB12 (black line) for use as a comparison.

## 4.4 Dynamic Light Scattering Results

For NMR spectroscopy and protein crystallography a monodispersed sample is important. This ensures that the NMR signal propagates from a single conformational

and multimeric form. Additional information regarding the size of the protein micelle can provide insight into the multimeric state of the protein if sufficient knowledge of the solution system is known. DLS provides a simple method for determining monodispersity and provide a size estimate. DLS measurements for OEP16 in β-DDM, Cymal-4, DPC, and SDS are shown in figure 4.6 and are summarized in table 4.2. The evaluation of the data using the Stokes-Einstein equation (described in 2.1.3) provided by the instrument software revealed Stokes radii of OEP16 in Cymal-4, DPC, SDS, and β-DDM which ranged between 3.7nm and 4.3nm. Data from SB12 is not shown due to the strong scattering signal from the empty detergent micelle at 1% concentration. Radii calculated from DLS estimated molecular weights between 88kDa and 114kDa assuming the protein-micelle assumed a spherical shape and a that may suggest higher level multimers assuming that all the assumptions for calculating a molecular weight and radii from diffusion data are correct; this includes viscosity which is significantly altered when glycerol and temperature are considered.

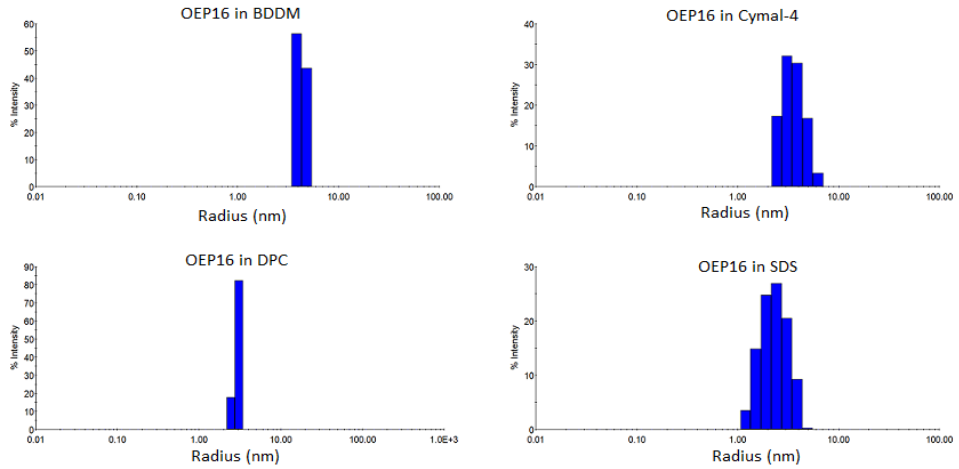


Figure 4.6. DLS results of OEP16 in β-DDM, Cymal-4, DPC, and SDS. Spectra are more monodispersed for lower cmc detergents: DPC and β-DDM at 0.01% and 0.05% respectively compared to Cymal-4 and SDS with 0.37% and 0.4% respectively. This is likely do to mixed scattering from both protein-micelle complexes and empty detergent micelles. Overall the histograms show fairly well monodispersed samples. All data was collected at protein concentrations of 1mg/mL with 100mM NaCl, 20mM $NaH_2PO_4$, pH 7 with 2x cmc detergent.

| Detergent | Radius | Empty Micelle | Calculated MW | Multimer |
|---|---|---|---|---|
| DPC | 3.8nm | 19.0kDa | 96.2kDa | 5.8 |
| SDS | 3.7nm | 17.9kDa | 87.6kDa | 5.3 |
| β-DDM | 4.3nm | 58.0kDa | 114.4kDa | 6.9 |
| Cymal-4 | 3.7nm | 12.0kDa | 93.5kDa | 5.6 |

Table 4.2. Results of DLS calculations. Molecular weights estimate that OEP16 forms a multimer in the listed detergents. Calculated MW results from the MW determined from total complex subtracted from the MW of the empty micelle.

## 4.5 Size Exclusion Chromatography Results

In addition to dynamic light scattering, size exclusion chromatography (SEC) can provide details into the size and multimeric states of OEP16 in different detergents as well as some insight into whether the protein is isolated in a monodispersed state. As described in section 3.3.2, a 100uL of protein solution was injected onto a Sephadex 200 10/300 GL column at a flow rate of 0.5mL/min. The protein traverses the column at a rate dependant on the size described in the theory section 2.2.2. The retention values of OEP16 in different detergents were used to estimate a molecular weight using a standard curve of well characterized, soluble proteins shown in figure 4.7.

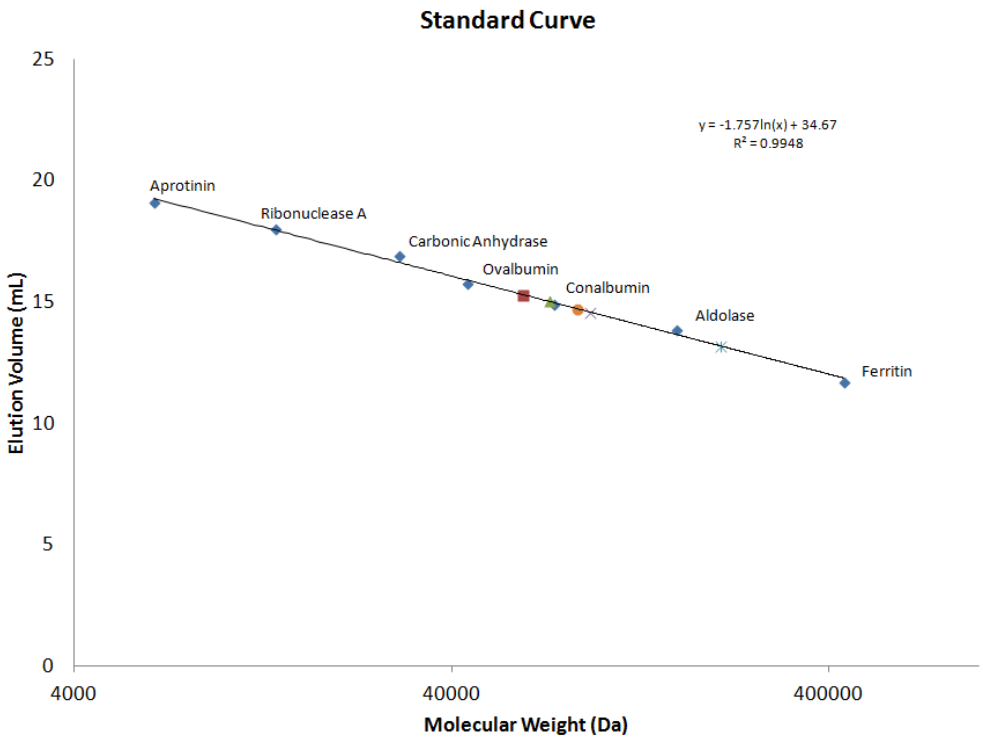


Figure 4.7. Standard curve for SEC. Used for calculating the MW of OEP16 in different detergent micelles. Blue diamonds represent the standard protein used. Red square is OEP16 in Cymal-4; Green triangle is OEP16 in DPC; Purple X is OEP16 in SDS; Orange circle is OEP16 in β-DDM; and Blue star is OEP16 in SB12.

All of the SEC results shown in figure 4.8 displayed one major peak at retention volumes ranging between 14.4mL and 15.2mL, with the exception of SB12 with a retention volume of 13.0mL. OEP16 in SDS micelles reported two other minor peaks at lower retention volumes (one similar to SB12) that suggest aggregate and/or multimer formation. From the retention volumes and a standard curve, a molecular weight was calculated. This was done by fitting a logarithmic line through the standard curve and using the resulting equation (shown in figure 4.7) to obtain a molecular weight. Molecular weights had a wide range of values summarized in table 4.3, from 63kDa for Cymal-4 to 207kDa for SB12.

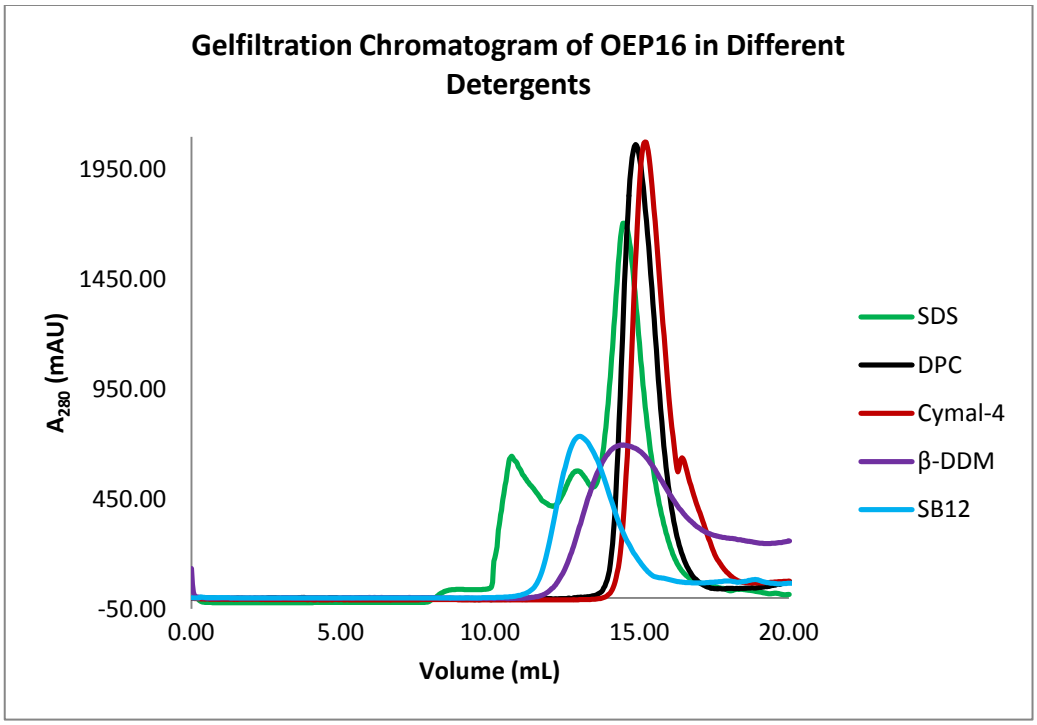


Figure 4.8. SEC results for OEP16 in different detergent conditions. A standard curve was used to calculate a molecular weight of the protein from each retention time. SDS displays three peaks that suggests aggregate formation. A wide range of MW were calculated, between 62kDa for Cymal-4 and 207kDa for SB12.

The width of the elution peaks can also reveal information about the distribution of sizes. Therefore I fit each peak in the chromatogram to a Gaussian curve in order to include the distribution of sizes (within one standard deviation) for each molecular weight calculated which is included in table 4.3. Also included in table 4.3 is the range of oligomeric states possible for the size exclusion measurement which was calculated using the following expression:

$$O_{low} = \frac{M_{SEC} - M_{EM} - M_D}{M_{OEP16}}$$

and

$$O_{high} = \frac{M_{SEC} - M_{EM} + M_D}{M_{OEP16}}$$

where $O_{low}$ and $O_{high}$ are the low and high range of the oligomeric states, respectively. $M_{SEC}$ is the molecular weight of the OEP16-micelle complex from the size exclusion chromatography experiments; $M_{EM}$ is the molecular weight of the empty micelle from table 3.2; $M_D$ is the molecular weight distribution number and $M_{OEP16}$ is the molecular weight of an OEP16 monomer without a micelle (16.6kDa).

OEP16 in β-DDM micelles display a relatively wide range of molecular weight values (between 63kDa and 155kDa) which provides little information regarding the actual size of OEP16 in β-DDM micelles. The overall size distribution across all size exclusion experiments does not allow a precise measurement of OEP16 to be made, but the data does strongly suggest the presence of multimers.

Additionally SDS displayed three peaks corresponding to molecular weight values of 120kDa, 281kDa and 914kDa (which suggest the presence of aggregates). Therefore before any spectroscopic studies were performed on OEP16 in SDS micelles, SEC was performed to obtain only one oligomeric form. Despite the different multimeric states, CD spectroscopy performed on the different multimeric forms revealed identical secondary structure components (data not shown). Other than SDS, lower retention volume peaks were not observed in any of the other SEC HPLC experiments.

| Detergent | OEP16-Micelle MW from SEC (kDa) | Distribution (kDa) | Empty Micelle (kDa) | Range of Oligomeric States | |
|---|---|---|---|---|---|
| SDS | 93 | 19 | 17 | 4 | 6 |
| DPC | 73 | 14 | 19 | 3 | 4 |
| Cymal-4 | 62 | 12 | 12 | 2 | 4 |
| β-DDM | 86 | 39 | 63 | -1 | 4 |
| SB12 | 207 | 64 | 19 | 8 | 16 |

Table 4.3. Size and oligomeric state results of OEP16 in different detergents through SEC. The range of oligomeric states was calculated by using the difference of the OEP16-Micelle MW from SEC and the empty micelle, then adding (for high end) or subtracting (for low end) the distribution of MW and then dividing by the MW of OEP16 (16.6kDa). Although an exact MW measurement is not possible, there is strong evidence of a multimeric state.

## 4.6 OEP16 Crystallization Results

Obtaining well ordered crystals would be an incredible leap towards the structure determination of OEP16. Crystallization of OEP16 has been attempted in the past by Linke *et al* [16] but crystals grown only diffracted to a resolution of 20Å. Despite over a decade of failed crystallization attempts, the benefit of obtaining well diffracting crystals would provide a plethora of information from X-Ray diffraction. Crystallization experiments have been performed under various conditions for OEP16 in different detergents using vapor diffusion methods as described in methods section 3.3.8. However, as expected, OEP16 strongly resisted crystal formation for all detergents in this study. The crystals that did form seemed to appear very close in morphology to those described by Linke *et al* [16] in that they were very flat and seemed to be very poor quality. Presence of crystalline material was confirmed using tryptophan fluorescent microscopy as seen in figure 4.9. The precipitate that formed seemed less crystalline and looked more like 2D lamellar sheets.

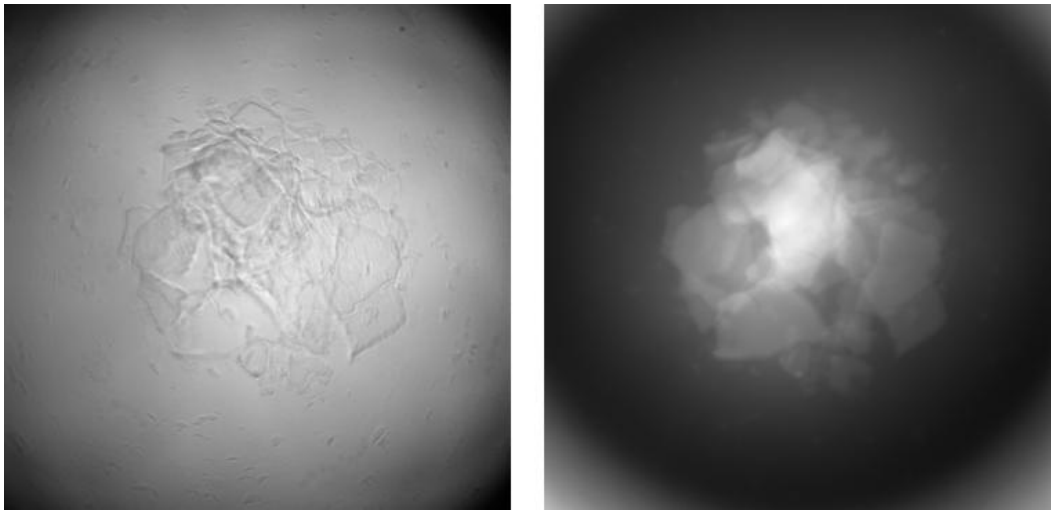

Figure 4.9. Bright field and fluorescent images of OEP16 crystals. OEP16 in DPC and Cymal-4 produce crystal-like precipitate that can be observed by both light (left) and UV fluorescence (right) microscopy. No diffraction patterns could be detected using home source X-Rays and suggest a very poor crystal quality. Solid State NMR may be a viable option for this precipitate.

This lamellar sheet formation occurred consistently when OEP16 was concentrated by ultrafiltration in the detergents DPC and Cymal-4. Traditional crystallization experiments using vapor diffusion on in 96 and 24 well crystallization trays did not yield any discernible protein crystals. However, upon concentration via ultrafiltration using 30kDa MWCO filters, sheet-like precipitate would form in the concentration tube. When the tube was inverted, the precipitate would suspend in the solution. With the naked eye, it was possible to observe the precipitate as very thin sheets that were flexible. However, diffraction patterns could not be observed using a home source X-Ray. Even though these crystals would be of little use for X-ray diffraction studies, there is a possibility that they could be used in future solid state NMR studies.

## 4.7 Nuclear Magnetic Resonance Results

### 4.7.1 NMR on OEP16 in Mild Detergents

The inability to grow well diffracting crystals of OEP16 forced me to turn to other methods for structure determination. NMR spectroscopy was performed in order to obtain detailed structural information. Preliminary NMR $^{15}$N-HSQC experiments on OEP16 in SDS and SLS micelles have been previously reported by Ni and Zook *et al* [25]. NMR experiments of OEP16 in mild detergents were performed in an attempt to provide comparative structural information between detergents. Two dimensional $^{15}$N-HSQC experiments were performed on OEP16 in mild detergents as a rapid way to qualitatively evaluate the possibility of obtaining adequately resolved resonance peaks in the more time intensive three dimensional experiments. NMR experiments were attempted for OEP16 in β-DDM, α-DDM, SB12 and DPC due to either the stability the detergent provided (as in β-DDM and α-DDM) or their ability to form a small micelle

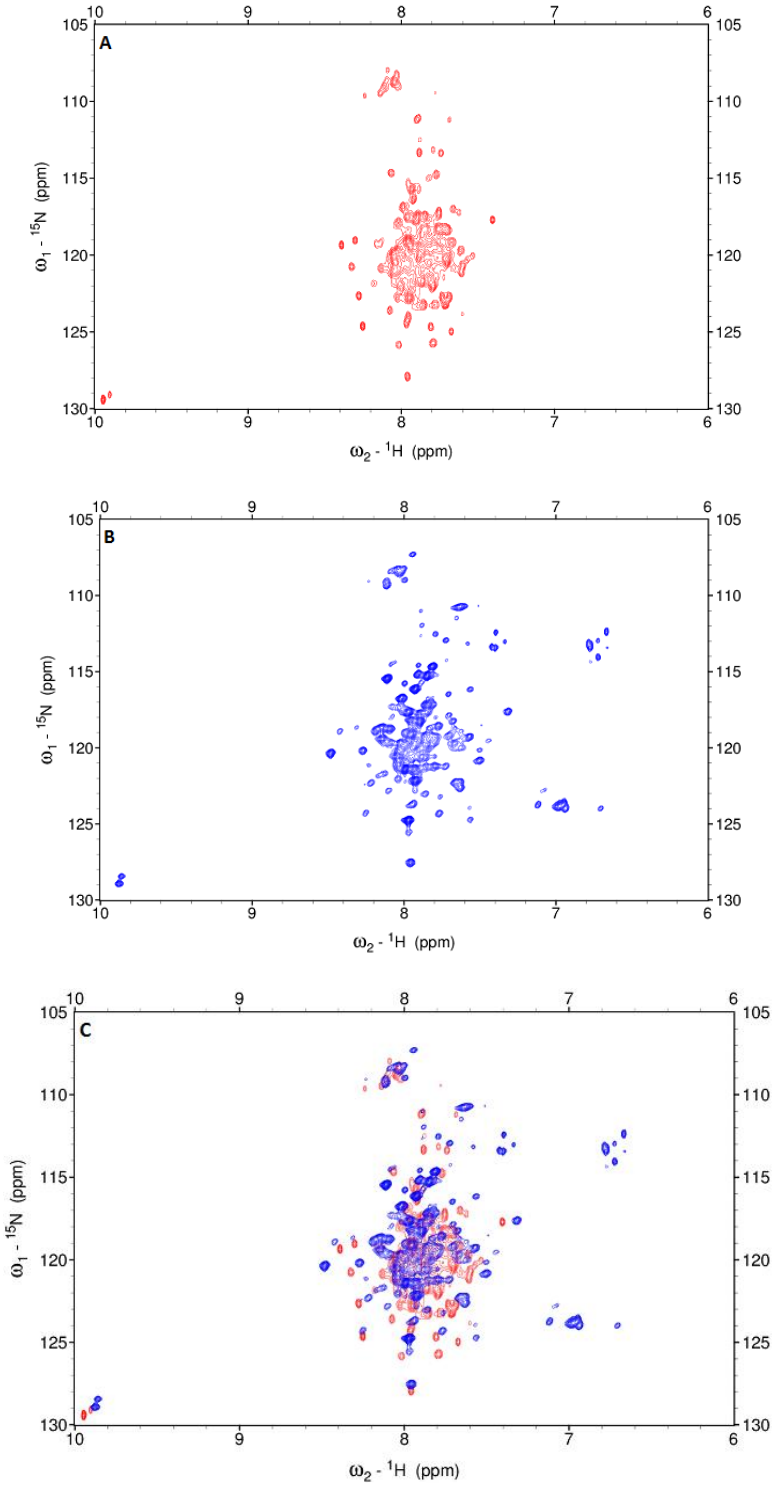


Figure 4.10. $^{15}$N-HSQC spectra of OEP16 in (A) α-DDM (red), (B) β-DDM (Blue) and (C) as an overlay of the two spectra. Both spectra were collected under similar conditions and display similar peak dispersion and peak widths for nearly all backbone amides observed. The β-DDM spectra also displays the side chain amides from asparagine and arginine residues seen mostly between 6ppm and 7ppm.

and maintain high solubility of protein (as seen in DPC and SB12). The NMR spectra of OEP16 in α-DDM and β-DDM were obtained in nearly identical solution conditions.

The two detergents display spectra that are very similar and are shown in figure 4.10, and displays broad resonance peaks within the narrow dispersion proton range (between 6.5ppm and 8.5ppm). The narrow proton dispersion suggests a strong presence of α-helices and very little β-sheet, as is suggested by CD spectroscopy.

β-DDM appears to be the superior detergent in this case due to its ability to resolve most of the peaks of the α-DDM spectra and a few peaks absent in the α-DDM. The β-DDM spectra provides increased dispersion which implies that future 3D experiments would allow for easier interpretation. Therefore a three dimensional HNCA and HNCACB was performed on $^{13}C/^{15}N$-labeled OEP16 in β-DDM in an attempt to assign resonances. Both experiments displayed very broad, unresolved resonances which could not be assigned, very likely due to the slow rotational isotropic motion provided by the large micelle to increase relaxation times. The NMR experiments involving SB12 seen in figure 4.11 displayed extremely broad and poorly dispersed peaks that suggested SB12 was not a detergent that could be practically used for NMR studies.

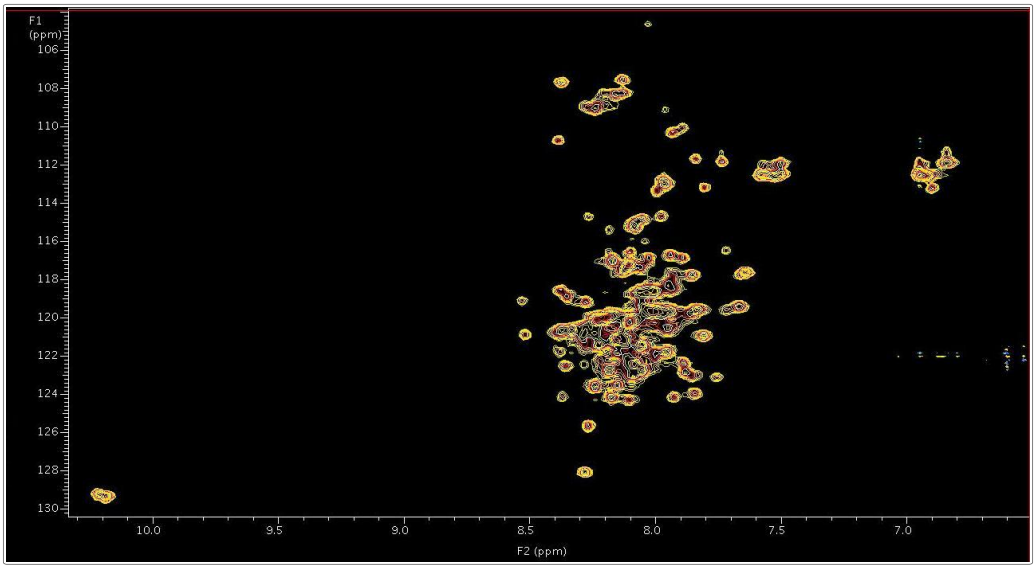


Figure 4.11. 2D $^{15}$N-HSQC of OEP16 in SB12 micelles. NMR data was processed and viewed using the Varian supplied software VNMJ. Very few resolved peaks are observed in the spectra due to low peak dispersion and broad peak widths. Therefore SB12 was abandoned as a candidate for NMR studies.

OEP16 in DPC, however, provides a very different $^{15}$N-HSQC spectra seen in figure 4.12. DPC shows a wide range of peak widths as well as a much wider dispersion of the amide backbone resonances in the proton dimension, between 6.5ppm and 9.5ppm. At some point during the 20 minute $^{15}$N-HSQC experiment, the protein denatured and an extremely turbid solution was left in the NMR tube. The wide dispersion of resonances in the proton dimension suggests the presence of β-sheets, which was not observed in the CD studies. A likely reason for this difference in secondary structure is that β-sheet formation occurred while the protein denatured, very similar to the precipitate formed by amyloid fibrils. The instability of OEP16 in DPC at high concentrations hindered further solution NMR studies. In an attempt to keep OEP16 in DPC micelles stable, a wide range of solution conditions were attempted to maintain structural integrity. These conditions include varying pH, ionic strength, glycerol content, and temperature. Additionally the method for concentrating OEP16 in DPC micelles was varied as well in

order to have a carefully controlled, slow concentration of sample. However, despite my every effort, precipitation could not be avoided.

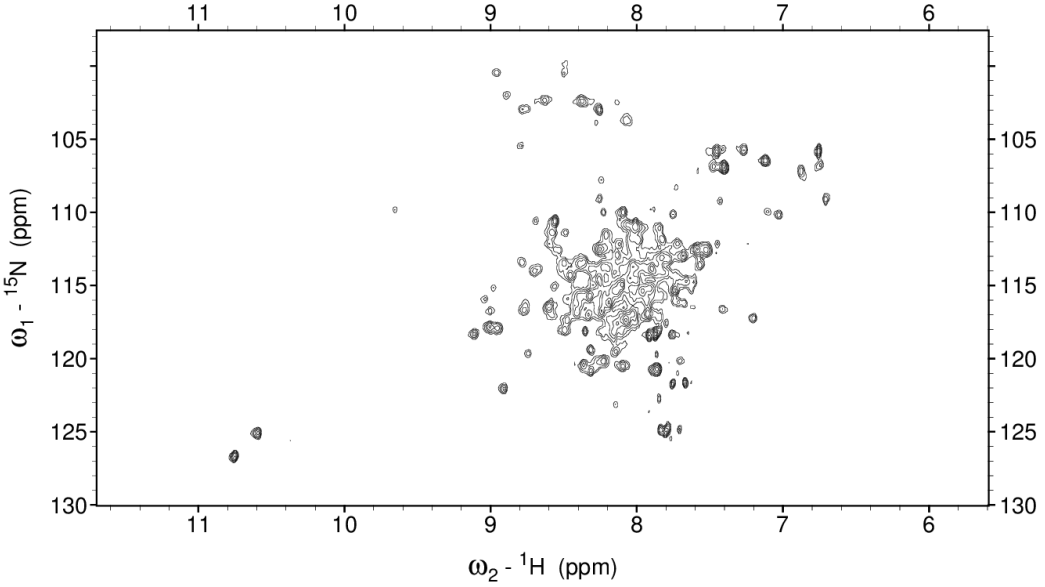


Figure 4.12. 2D $^{15}$N-HSQC of OEP16 in DPC. It displays a spectra very different to OEP16 in α-DDM or β-DDM micelles. The wide dispersion in the protein dimension may suggest the presence of β-sheets. The protein denatured and precipitated during the experiment, which may suggest the formation of amyloid-like, β-sheet aggregates.

DPC is a favored detergent for membrane protein studies due to its ability to display well defined peaks in gel filtration data and clear circular dichroism spectra. However, only one protein (OmpF) has been successfully crystallized using DPC. In 2010, OmpF was accidently crystallized while trying to express, purify and crystallize the integral membrane protein KdpD [98]. Therefore it is possible that DPC does not behave well at high protein concentrations; this is discussed in further detail in the discussion section.

Due to the poor quality of NMR spectra in milder detergents, OEP16 in SDS was chosen for further structural investigations. As seen in figure 4.13, the $^{15}$N-HSQC spectra of OEP16 in SDS micelles is remarkably similar to the OEP16-βDDM spectra. For nearly all peaks in the β-DDM spectra, there is a matching peak in the SDS spectra. This

suggests that the amide residues of the protein backbone are in a similar structural conformation, which provides strong evidence that OEP16 in SDS is similar to OEP16 in β-DDM as well as α-DDM (due to the similarity between the β-DDM and α-DDM spectra) and folds into a three dimensional structure. α-DDM is an extremely mild detergent that is capable of solubilizing intact photosystem supercomplexes [99]. To have similar spectra between α-DDM and SDS strongly suggests a tertiarily folded protein. Additionally, the SDS resonance peaks are much more narrow, which is caused by a significantly smaller detergent-micelle complex. This peak narrowing provides much more resolution in the central part of the spectra which suggests that three dimensional experiments could lead to successful resonance assignment.

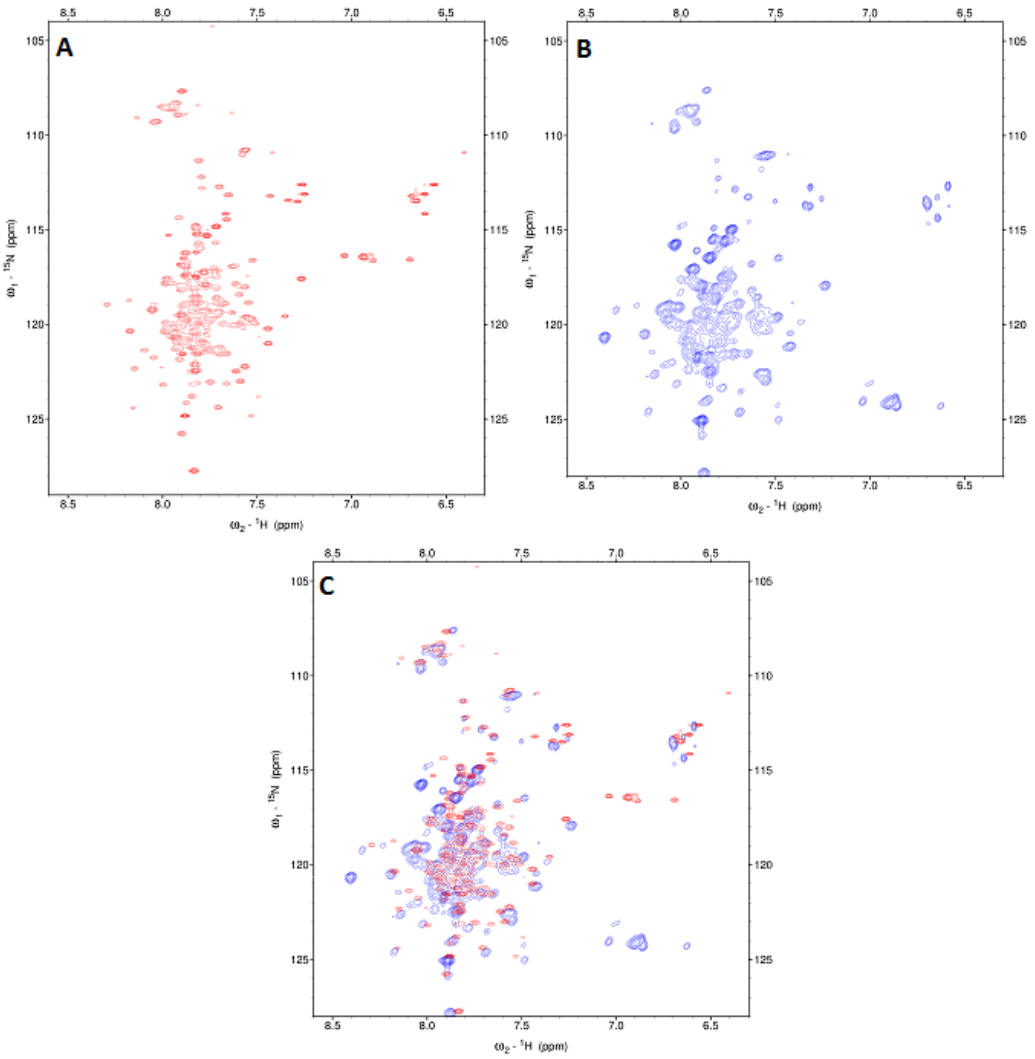


Figure 4.13. 2D $^{15}$N-HSQC to compare OEP16 in SDS to OEP16 in β-DDM. (A) OEP16 in SDS micelles display a spectra (Red) very similar to (B) OEP16 in β-DDM micelles (Blue) with much narrower line-widths. This is largely due to the faster global isotropic rotational correlation times provided by the much smaller OEP16-SDS complexes. (C) an overlay of the two spectra shows that even though there is a slightly greater peak dispersion in the β-DDM micelles, the smaller protein-micelle complexes provide better resolution due to narrow line-widths.

### 4.7.2 Successful Assignment of OEP16 Backbone

By far the most successful spectroscopic results were obtained from the NMR studies of OEP16 in SDS micelles. However, before any structural information can be obtained by NMR, assignment of the resonances is necessary. The very sensitive

HN(ca)CO and HNCO (figure 4.14) experiments were used to walk the backbone of OEP16 in SDS micelles to aid in resonance assignment. HN(ca)CO displays a chemical shift of CO for the $i^{th}$ residue, and both display chemical shifts at the $i\text{-}1^{th}$ residue. The CO chemical shift is not particularly useful for identifying amino acids, but rather identifies resonances that are sequential to each other.

The experiments that provided the bulk of the amino acid identification were the HNCACB and the CBCA(co)NH experiments (see figure 4.15). These experiments use the $C_\alpha$ and $C_\beta$ chemical shift values to identify both the $i^{th}$ residue (HNCACB) and the $i\text{-}1^{th}$ residue (HNCACB and CBCA(co)NH using statistics provided by the Biological Magnetic Resonance Bank (BMRB). In the HNCACB spectra, $C_\alpha$ chemical shifts are positive values and $C_\beta$ chemical shifts are negative values (in the CBCA(co)NH all values are positive). HNCA and HN(co)CA were used to resolve $C_\alpha$ chemical shifts if the HNCACB and/or the CBCA(co)NH peaks were not resolvable [10, 42].

Another set of experiments that were useful were the TOCSY-$^{15}$N-HSQC (figure 4.16a) which transfers magnetization of $^{1}$H spins through the bonds of an amino acid residue then to the amide hydrogen and the NOESY-$^{15}$N-HSQC (figure 4.16b) which transfers magnetization of $^{1}$H spins through space then to the amide hydrogen [10]. These two experiments allowed me to walk the backbone of a protein by way of spatial relationships, rather than through-bond connectivity. Roughly 95% of the protein was assigned using $^{13}$C chemical shifts from the HNcaCO, HNCO, HNCA, HNcoCA, HNCACB, and CBCAcoNH experiments which transfer magnetization through covalent bonds. Examples of strip plots taken from the spectra are illustrated in figure 4.17 (HNcaCO and HNCO) and figure 4.18 (HNCACB and CBCAcoNH) . The remaining assignments were possible using a combined TOCSY-$^{15}$N-HSQC and NOESY-$^{15}$N-HSQC experiments (figure 4.19). This provided amino acid sequence information by

looking at short range NOEs between amide protons, $H_\alpha$, and $H_\beta$ protons. Therefore the backbone to be 'walked' by through-space magnetization transfers, which is a significantly different experiment compared to the $^{13}C$ experiments.

Nearly complete assignment was achieved; only the assignments for residues M1 and the $C_\beta$ of Y144 remain unknown. However, due to significant peak overlap in the 2D $^{15}N$ HSQC, only 123 resonances could be resolved seen in figure 4.20 which corresponds to 80% of the backbone of OEP16. Additionally all seven asparagine side chain amides are resolved as well as five of the six arginine side chain amides, but remain unassigned. Also, interestingly, seven residues (F6, S7, G8, S44, L67, G74, A75) displayed two assignment peaks in the 2D $^{15}N$ HSQC seen in figure 4.21, but maintain the same CO, $C_\alpha$, and $C_\beta$ and $H_\alpha$ chemical shifts, which strongly suggests that OEP16 may exist in two different conformations.

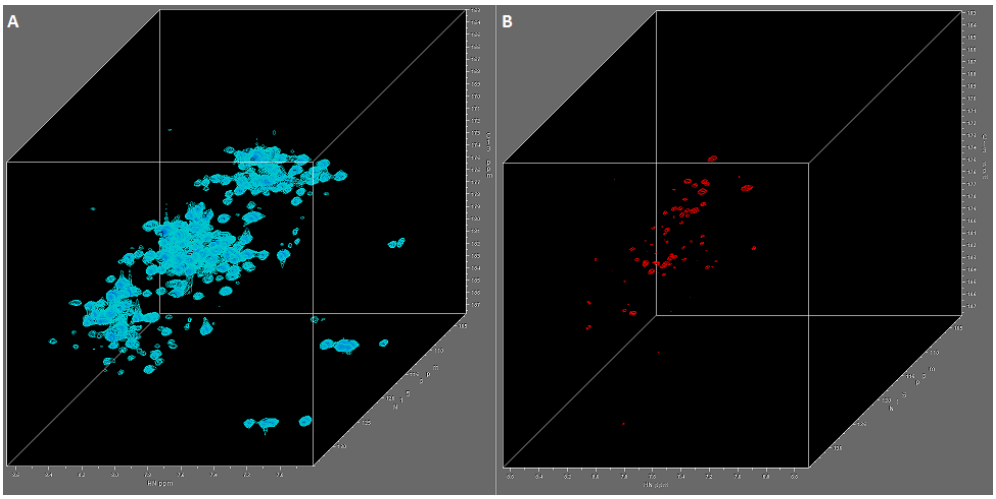


Figure 4.14. 3D views of (A) HNCO and (B) HN(ca)CO experiments. These experiments observe CO chemical shifts for the $i^{th}$ residue (B) and the $i\text{-}1^{th}$ residue (A and B) in order to obtain sequence information about the protein. This is very useful for the amino acid assignment, even though the CO chemical shift reveals little detail about the identity of a specific amino acid.

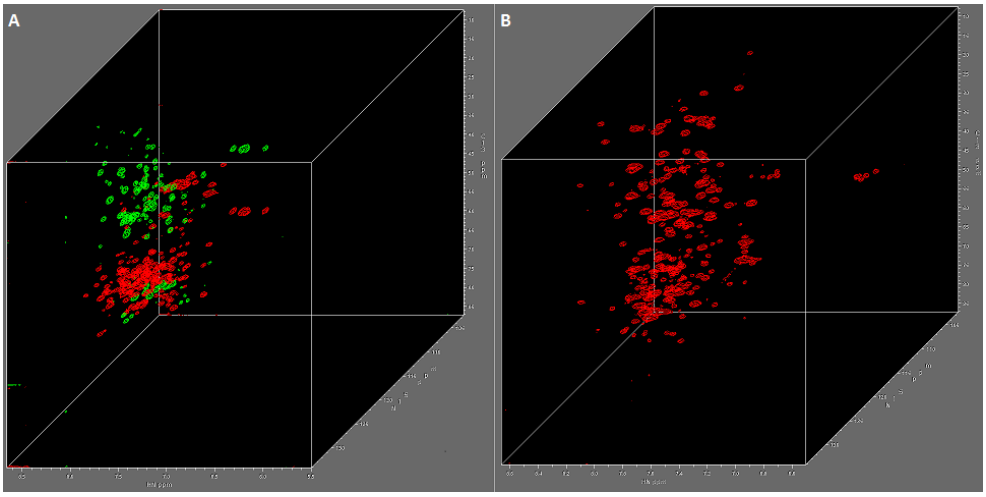


Figure 4.15. 3D views of (A) HNCACB and (B) CBCA(co)NH experiments. They provide the bulk of the information regarding the identities of specific resonances. The red resonances are positive values which correspond to $C_\alpha$ chemical shifts of the $i^{th}$ and $i\text{-}1^{th}$ residue in the HNCACB spectra and both $C_\alpha$ and $C_\beta$ chemical shifts of the $i\text{-}1^{th}$ residue in the CBCA(co)NH spectra. Green resonance peaks are $C_\beta$ chemical shifts of the $i^{th}$ residue and $i\text{-}1^{th}$ residue in the HNCACB spectra. The $C_\alpha$ and $C_\beta$ chemical shifts can usually provide a unique fingerprint for the identity of an amino acid using the chemical shift statistics in the BMRB [42].

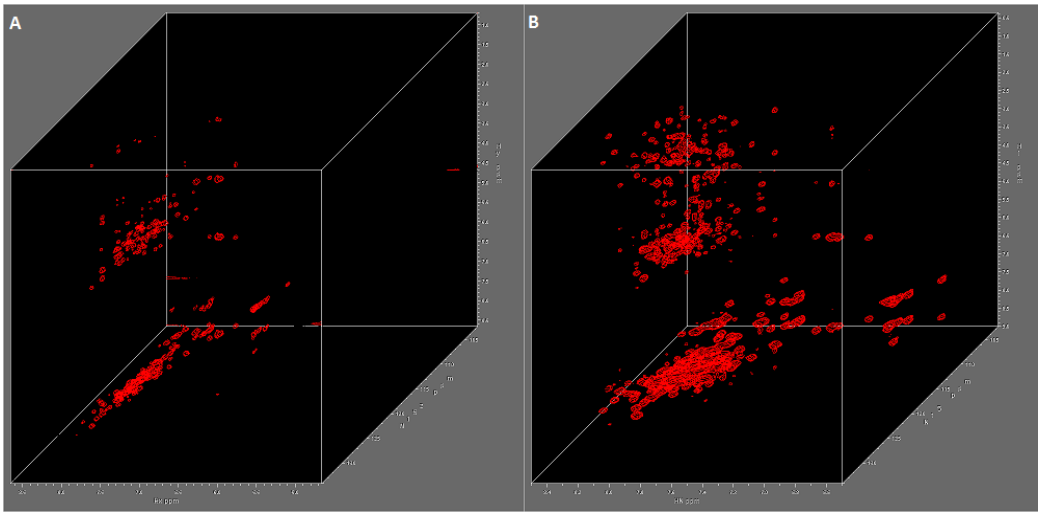


Figure 4.16. 3D views of (A) TOCSY-$^{15}$N-HSQC and (B) NOESY-$^{15}$N-HSQC for resonance assignment. These sets of experiments provide a way to walk the backbone using short range NOEs to observe the sequence of the protein via through space connectivity. The information obtained from the NOESY experiment can therefore aid in resonance assignment. Additionally the $H_\alpha$ chemical shift observed in the TOCSY experiment provides information that is used by TALOS to calculate secondary structure elements.

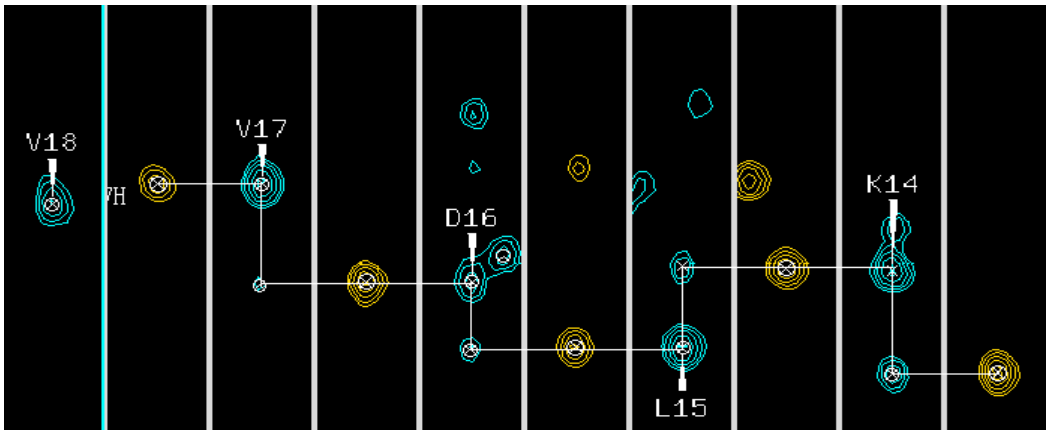


Figure 4.17. Example of strip plot from HN(ca)CO (Blue) spectra and HNCO (yellow) spectra. The chemical shifts for these experiments do not provide much for identifying amino acids, but the sensitivity of the experiments (specifically the HNCO) can provide information on how the resonances are connected through the protein backbone. However, The chemical shifts for these spectra are used for TALOS calculations, which is a powerful tool for estimating secondary structure and torsion bond angles.

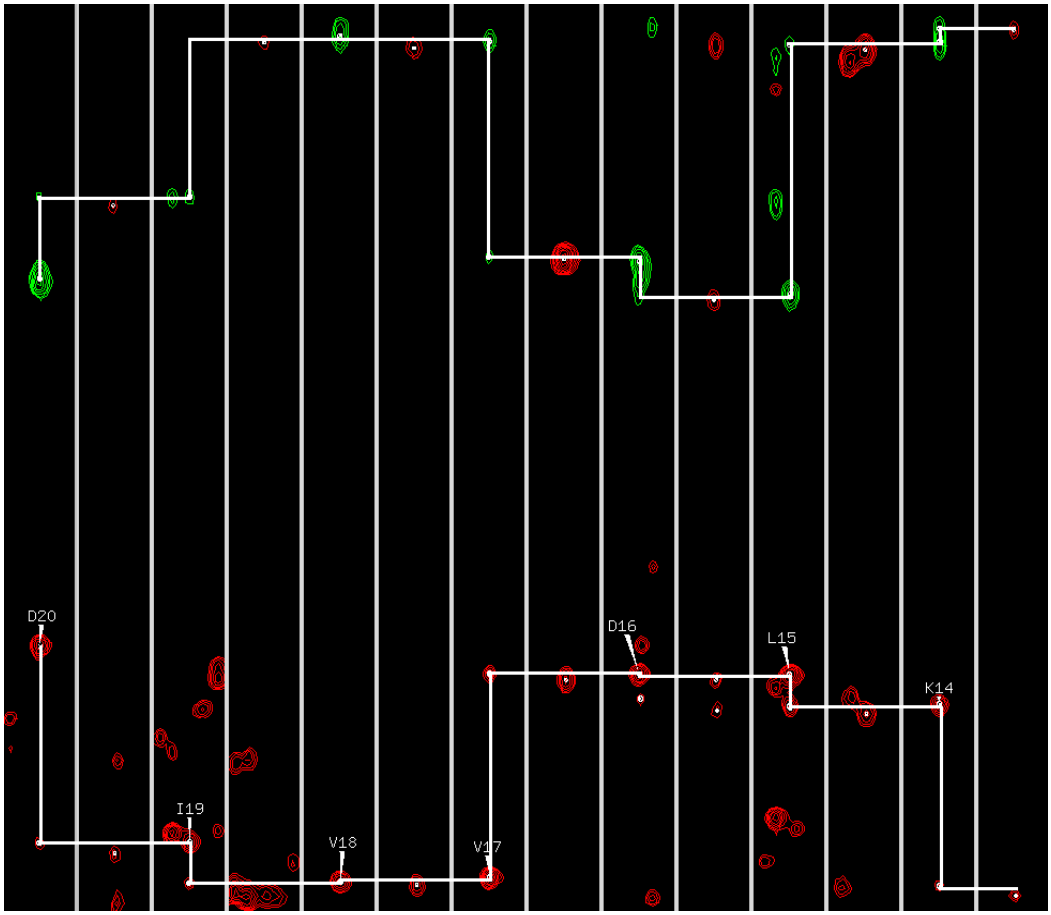


Figure 4.18. Strip plot example for HNCACB (green peak-containing strips) and CBCA(co)NH. Chemical shifts of the alpha and beta carbons are used for identification of specific amino acids. Because the i and i-1 residue can be observed (i-1 only for CBCA(co)NH), walking the backbone is also used to aid in residue identification. TALOS uses these chemical shifts for calculation of secondary structure and estimation of torsion angles.

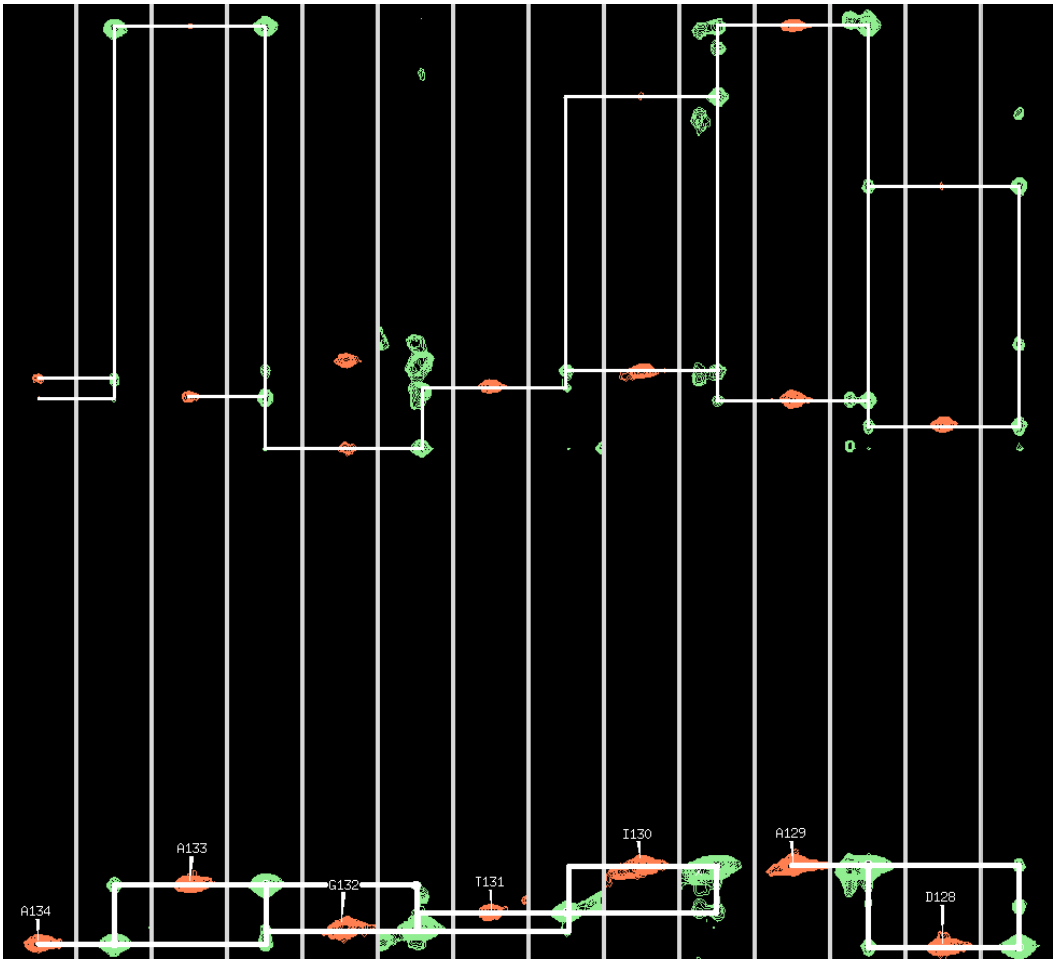


Figure 4.19. Strip plot example for TOCSY-$^{15}$N-HSQC (orange) and NOESY-$^{15}$N-HSQC (green). Used for identification of short range NOE's which are also used to walk the backbone using through space connectivity rather than through bond done in the $^{13}$C experiments. TOCSY is used to discern protons that are covalently bonded from protons that are spatially close. Amide proton NOEs are observed at the bottom of the strip plots, between ~7ppm and ~8.5ppm, $H_\alpha$ NOEs are seen in the center of the strip plots, at around 4ppm, $H_\beta$ NOEs are located around the 2ppm or 3ppm region at the top of the strip plots. The $H_\alpha$ chemical shifts identified by the TOCSY-HSQC are used in TALOS calculations for estimation of secondary structure by torsion angle.

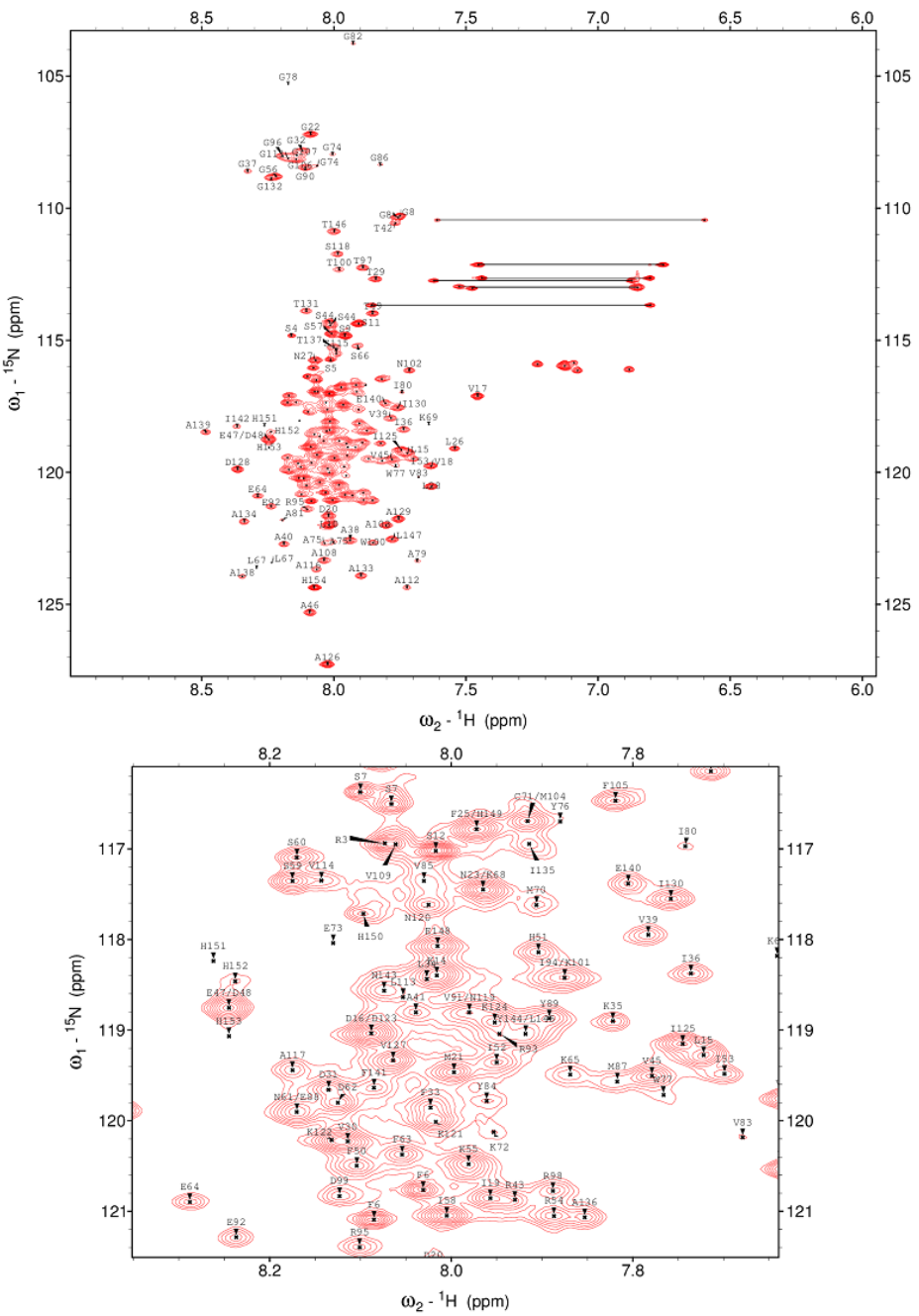


Figure 4.20. Assigned 2D $^{15}$N-HSQC from all 3D strip plot data. This spectra is used for calculation of relaxation data, $D_2O$ titration, and arginine titration analysis. Due to significant peak overlap, only about 80% of the identified peaks could be resolved. All of the side chain $NH_2$ peaks from asparagine are present but as of yet unidentified shown by the split peaks connected by the black line. Additionally several of the arginine residues are present due to aliasing done by the NMR experiment shown in the ~7.0ppm region of the spectra.

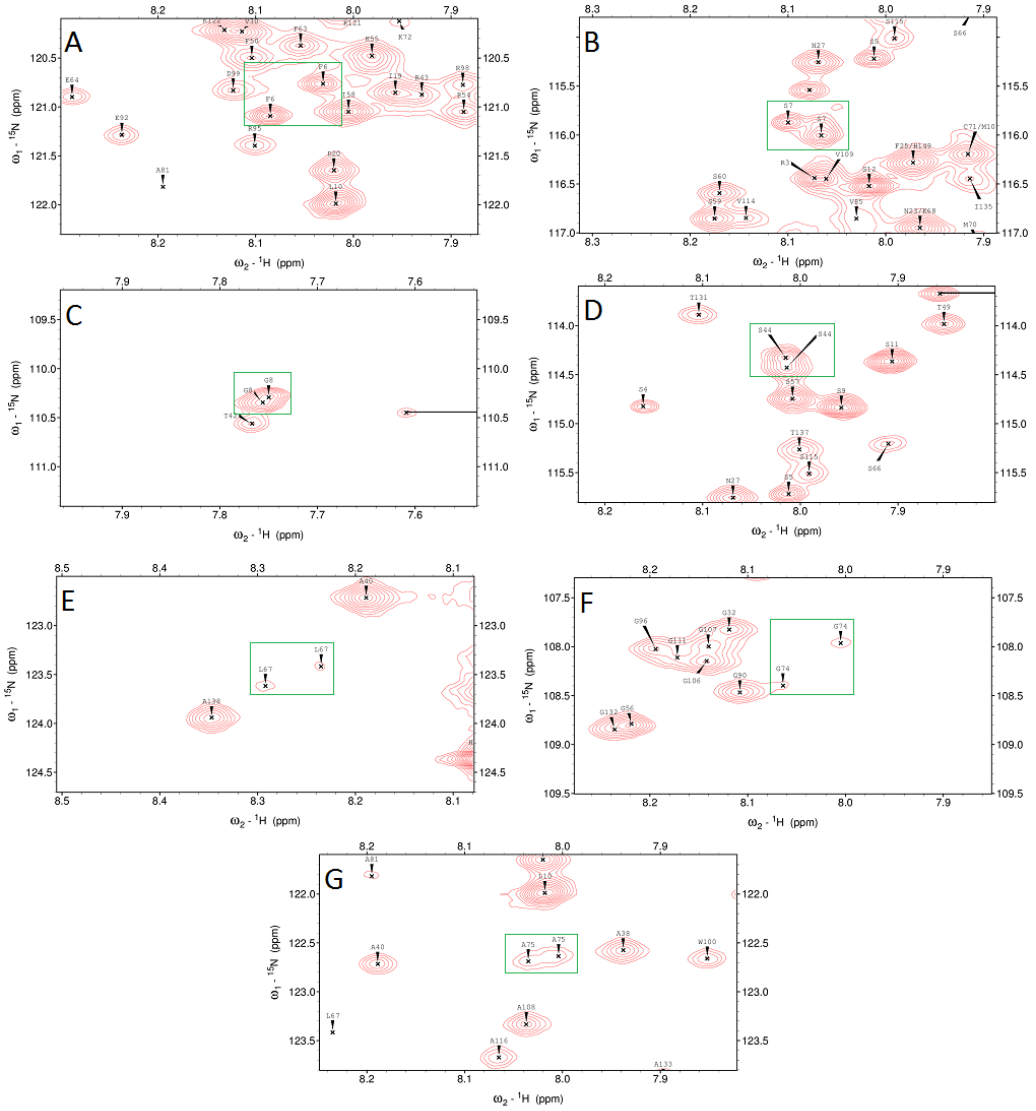


Figure 4.21. The seven residues that show two peaks in the 2D $^{15}$N-HSQC. Despite different $^{1}$H and $^{15}$N chemical shifts, the two peaks share identical $^{13}$CO, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and $^{1}H_{\alpha}$ chemical shifts. This might suggest the presence of two protein conformations.

### 4.7.3 Secondary Structure and Order Parameters by TALOS

Once the protein backbone is assigned, TALOS is a powerful tool that can check the assignments for any possible errors as well as use the secondary chemical shift values to estimate the φ and ψ torsion angles of the peptide backbone to provide secondary structure information on a per residue basis. A secondary chemical shift is the difference

between the measured shift and the shift obtained from a random coil, and is highly correlated with aspects of secondary structure. This is done by comparing the secondary chemical shift to a database of 200 solved proteins. This provides much more structural information compared to CD studies as it can predict secondary structure on a per residue basis whereas CD can only provide the overall secondary structure of a protein. TALOS has been improved to an upgraded version, TALOS+, which uses the larger database (200 proteins), provides a more efficient data mining algorithm, and can provide order parameters ($S^2$) using a similar method described by Berjanskii and Wishart [100].

TALOS+ predictions provided secondary structure estimations that revealed a protein that was exclusively α-helical and contained more flexible loops, with no β-sheet containing residues. The result is shown in figure 4.22a. Helices are particularly apparent in four segments of the protein sequence, suggesting the location of the transmembrane helices of the protein. Order parameters estimated by TALOS additionally provide insight into the flexible regions of the protein, which are primarily at the protein termini and within a loop region between the first and second transmembrane helix illustrated as $1-S^2$ (in order to emphasize flexibility) in figure 4.22b.

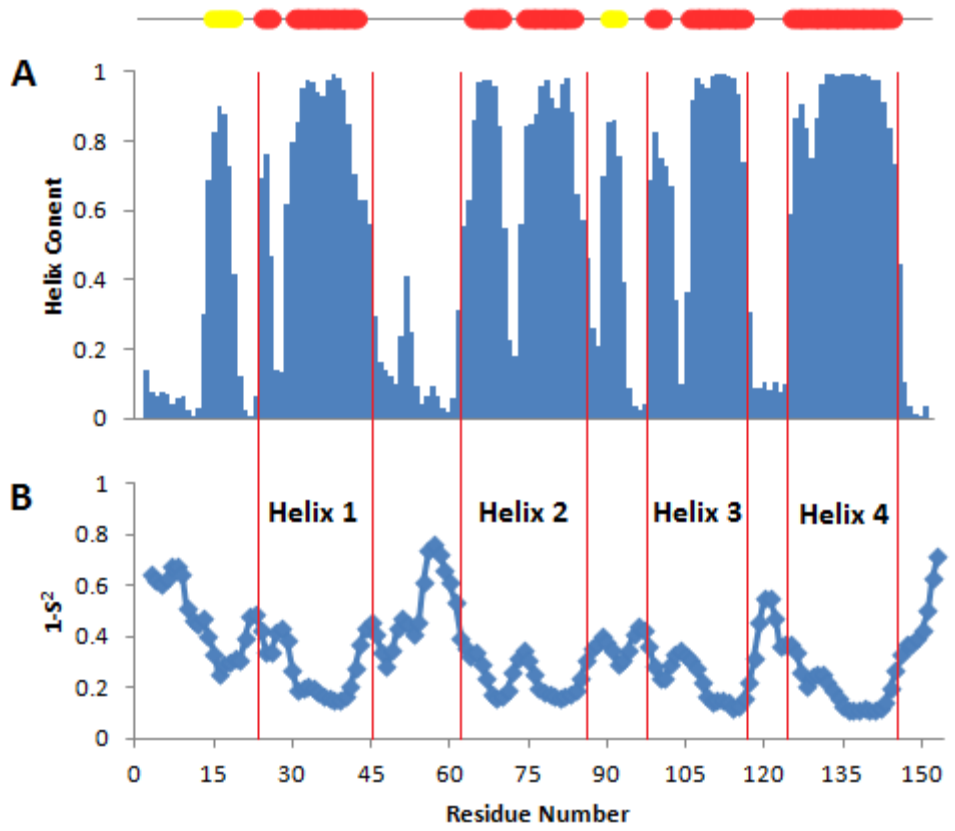


Figure 4.22. Structural analysis of OEP16 from chemical shift data. (A) TALOS uses the measured chemical shifts and analyses it against a database of solved proteins with known chemical shifts. TALOS predicts a protein secondary structure fold that is exclusively α-helices connected to flexible loop regions. (B) TALOS also estimated order parameters (shown as $1\text{-}S^2$ to emphasize flexible regions) that suggest extremely flexible termini and a flexible loop region between helix one and two. Red lines represent the transmembrane helices of OEP16. There may also be small solvent exposed helices at the N terminus, between helix one and two and between helix two and three shown by the yellow bars above.

### 4.7.4 Relaxation Analysis of OEP16

NMR relaxation experiments at 37°C reveal information about the intramolecular movements of a protein as well as provide insights on the overall protein size. Additionally, relaxation measurements can provide insights about the chemical exchange processes happening to a protein. Chemical exchange is defined as the "motional process that involves the making and breaking of chemical bonds" [9]. In the case of OEP16, this involves a conformational change in the peptide bond, which is a relatively rigid entity.

Arguably the most important values used to describe relaxation are the spin-lattice relaxation, $T_1$, and the spin-spin relaxation, $T_2$, values. $T_1$ is related to the time a particle's magnetization aligns with the static magnetic field. $T_2$, which is related to the time the particle's spins depolarize relative to each other. $T_1$ and $T_2$ data were obtained on all resolved and assigned peaks in the 2D $^{15}$N HSQC. Individual spin-lattice and spin-spin relaxation with corresponding correlation times for each residue is shown in figure 4.23 as $R_1$ and $R_2$ values. The $\tau_m$ for each residue can be calculated using methods described in theory section 2.1.4.5. Short $\tau_m$'s suggests areas of rapid magnetic field fluctuation indicating dynamic and flexible domains of the protein, such as those seen at the protein termini as well as between residues 45 and 65. The overall correlation time, $\bar{\tau_c}$, was calculated to be 12.0ns which can be used to calculate a hydrodynamic radius using:

$$r_H^3 = \frac{4\pi\eta\bar{\tau_c}}{3kT}$$

Where η is the viscosity of the solution, k is Boltzmann's constant, and T is the temperature. The viscosity of the 10% glycerol buffer was estimated from parameterization reported by Cheng [101] (addressed in appendix B). This calculated a hydrodynamic radius of 2.36nm. A hydrodynamic radius can be used to estimate a molecular weight ($M_r$) with the relationship [10]:

$$r_H = \sqrt[3]{\frac{3\bar{V}M_r}{4\pi N_A} + r_w}$$

The $M_r$ was then calculated to be 31.8kDa, where $N_A$ is Avogadro's number; $r_w$ is the radius contributed by the water shell, 1.6Å, from Venable and Pastor [102]; $\bar{V}$ is the specific volume of the protein (inverse of density) which for the OEP16-SDS micelle complex was estimated to be 0.847cm$^3$/g using the density of SDS (1.01g/cm$^3$) and

protein density of 1.37g/cm$^3$ reported by Squire and Himmel [103] as well as Gekko and Noguchi [104]. Additionally, the protein is assumed to be completely spherical. Despite the assumptions that needed to be made for the calculation, the resulting molecular weight is almost equal to the sum of an OEP16$^{His}$ monomer and an empty SDS micelle with an aggregation number of 65 which is very close to the published value of 62 [58] (the aggregation number is the number of individual detergent molecules that make up the micelle). However, due to the complex nature of how a protein interacts with the micelle, it is not possible to definitively conclude a monomeric state without additional data.

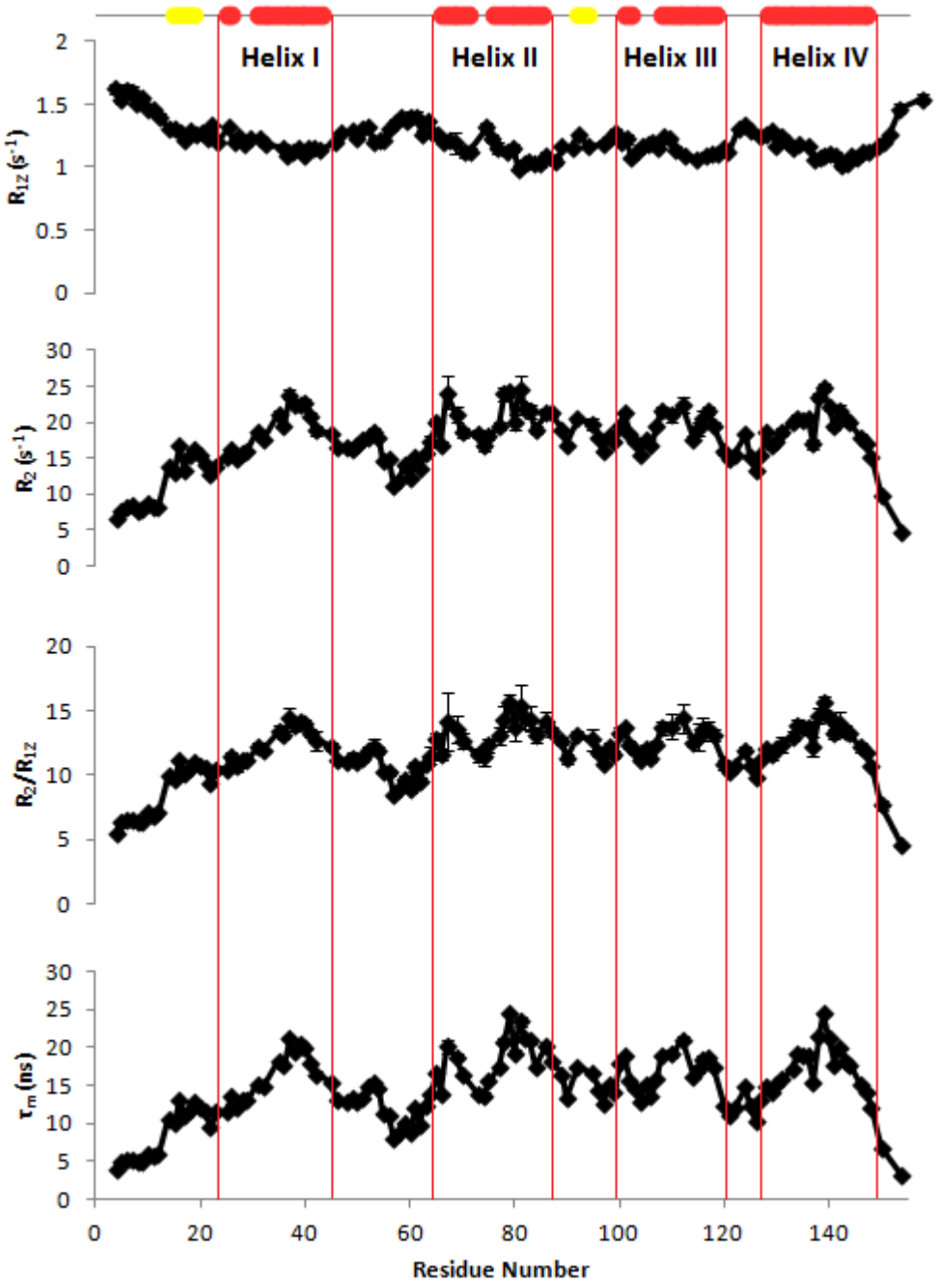


Figure 4.23. Relaxation data shown in units of frequency. This allowed for the calculation of a correlation time per residue, $\tau_m$, which is an indicator on how flexible that region of the protein is. According to the $R_2$ values reported, the most flexible regions of the protein are the termini and between residues 45 and 65.

More information about the internal dynamics of OEP16 can be obtained from additional relaxation experiments at different magnetic field strengths. With the current

collected data, many assumptions are made in order to calculate the $\tau_m$ which may be incorrect this includes the isotropic distribution of the NH bond vectors, the rigidity of the protein (which is likely incorrect due to the flexible regions reported by the $S^2$ values calculated by TALOS), and the stochastic tumbling motion of the protein-micelle complex. These are further examined in the discussion section as well as in the future outlook.

The relaxation experiments performed at 25°C could not be analyzed to the same amount of detail compared to the 37°C data due to the significantly poorer resolution of data. However, to estimate the hydrodynamic radius and molecular weight of the OEP16-SDS micelles, the $T_1$ and $T_2$ values of A139 were calculated to be 1348ms and 28.36ms respectively, using an adapted equation from Kay *et al* [49]:

$$\tau_c \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2} - 7}$$

Where $\nu_N$ is the $^{15}N$ frequency (60.8μs for a 600MHz magnet). This equation only takes into consideration the spectral density terms J(0) and J(ωN)while neglecting the higher frequency terms (for an explanation of spectral density functions see section 2.1.4.5). The $\tau_c$ for A139 at 25°C is estimated to be 21.8ns, which is a significant increase that leads to line broadening, and therefore poorly dispersed spectra. However, due to the decreased temperature and the increased viscosity of the 10% glycerol solution at 25°C (1.22cP at 25°C compared to 0.931cP at 37°C), the MW can be estimated to be around 36kDa which agrees well with the results obtained at 37°C.

4.7.5 Arginine Titration Result

Information about how OEP16 binds the amino acids for diffusion into the chloroplast intermembrane space was obtained via arginine titration. Increasing amounts of arginine were added to $^{15}N$-labeled OEP16 in SDS. Thereafter a 2D $^{15}N$ HSQC was

collected and the resulting chemical shifts were recorded for perturbation analysis. See Materials and Methods section 3.3.6.2 for more detailed information.

Chemical shift perturbations were plotted in units of ppm as a function of arginine concentration for each resolved residue in the $^{15}$N-HSQC. Most plots demonstrated a linear relationship between distance and arginine concentration, which suggests that the interactions between protein and arginine are nonspecific and therefore not of functional importance. However, several residues demonstrated a non-linear relationship which therefore suggested a specific binding of the ligand to the protein. A relationship between the effect of ligand binding as a function of ligand concentration can be described by the following function:

$$y = \frac{B_{max}x}{K_d + x} + N_b x$$

Where the parameters that define specific binding are $B_{max}$, which is the saturation concentration where all ligand binding sites are filled, and $K_d$, which is the concentration at which half of the binding sites are filled. $N_b$ is a constant that describes the linear, non-specific binding component of the binding curve. This means that for specific binding, $B_{max}$ will be a number significantly larger than zero and a relatively low $K_d$ value that does not force the specific binding component of the function to be negligible. Conversely, non-specific binding will fit to the data in such a way that either $B_{max}$ is a number very close to zero or a very large $K_d$ value that makes $N_b$ the directing parameter of the binding function. NMR ligand binding studies can provide significant insight into the functional mechanism of the protein as well as provide qualitative evidence that the protein is well-folded.

Despite the minute chemical shift perturbations, several residues displayed significant specific binding values and are shown in figures 4.24 to 4.29. These residues are S59, E64, L67, V83, G86, and K124 (figures 4.24, 4.25, 4.26, 4.27, 4.28, and 4.29

respectively). S59 is located in the dynamic loop region between the first and second transmembrane helix. E64 and L67 are located at the N-Terminal side of the second transmembrane helix before the helix enters the membrane. V83 and G86 are located within the second transmembrane helix; K124 is on the same side of the membrane as S59, E64, and L67, between helix three and four. All demonstrate good fits with the ligand-binding function, with $R^2$ values ranging between 0.9637 and 0.9982. For comparison, the binding profile for residue S4 is shown in figure 4.30 and shows a very good linear fit which strongly suggests that any binding that occurs is strictly non-specific.

From the data $K_d$ values for each of these residues can be estimated, ranging between 8.4mM for L67 to 2.4mM for V83 with an average $K_d$ of 5.8mM shown in table 4.4. This $K_d$ is expected for protein channels that use selective diffusion. If the $K_d$ values were lower, the ligand would not be able to disassociate from the protein, making a poor channel. This is also seen by the ScrY protein which is a sugar selective pore channel protein that has $K_d$ values in the low millimolar range [105].

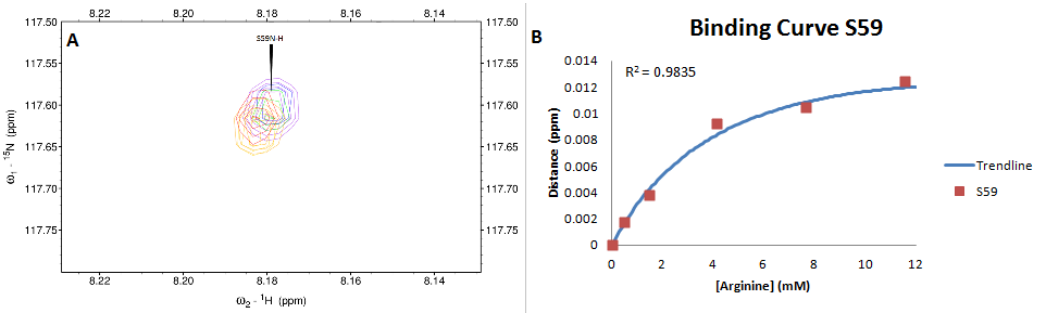


Figure 4.24. Arginine binding to residue S59. (A) The chemical shift perturbation distances measured upon arginine titration. (B) Collected data was used to estimate $B_{max}$ and $K_d$ values of 0.023ppm and 5.8mM respectively using the ligand binding function. These numbers strongly suggest the presence of specific binding of arginine to OEP16.

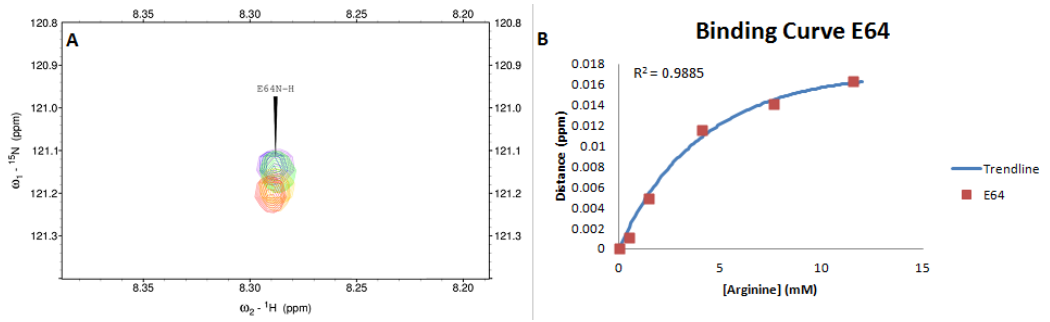


Figure 4.25. Arginine binding to residue E64. (A) The chemical shift perturbation distances measured upon arginine titration. (B) Collected data was used to estimate $B_{max}$ and $K_d$ values of 0.036ppm and 6.9mM respectively using the ligand binding function. These numbers strongly suggest the presence of specific binding of arginine to OEP16.

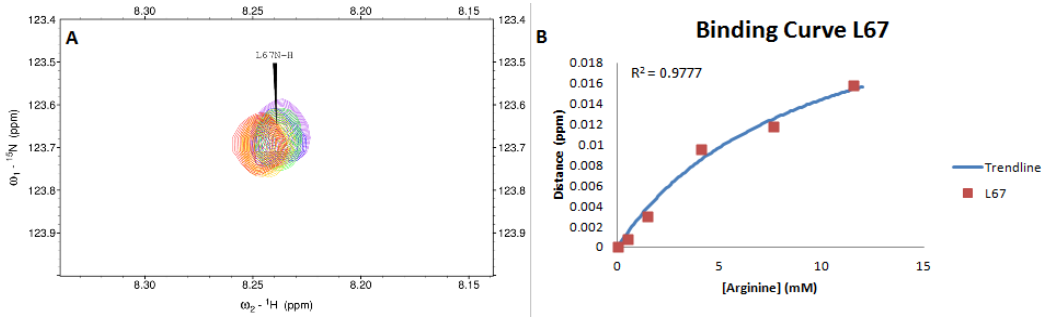


Figure 4.26. Arginine binding to residue L67. (A) The chemical shift perturbation distances measured upon arginine titration. (B) Collected data was used to estimate $B_{max}$ and $K_d$ values of 0.025ppm and 8.4mM respectively using the ligand binding function. These numbers strongly suggest the presence of specific binding of arginine to OEP16.

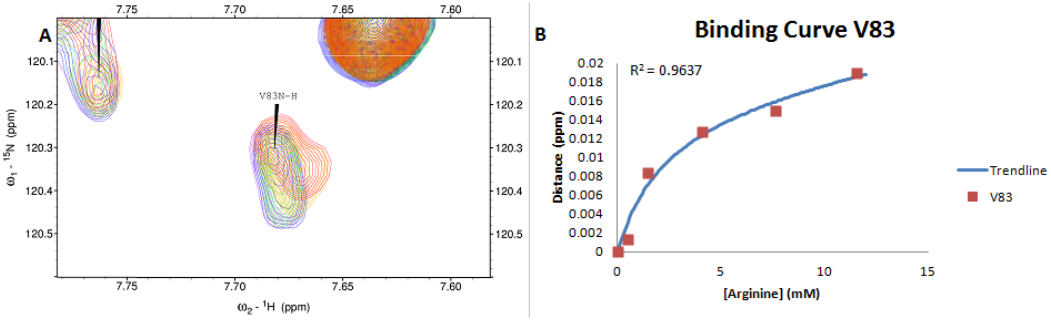


Figure 4.27. Arginine binding to residue V83. (A) The chemical shift perturbation distances measured upon arginine titration. (B) Collected data was used to estimate $B_{max}$ and $K_d$ values of 0.017ppm and 2.4mM respectively using the ligand binding function. These numbers strongly suggest the presence of specific binding of arginine to OEP16.

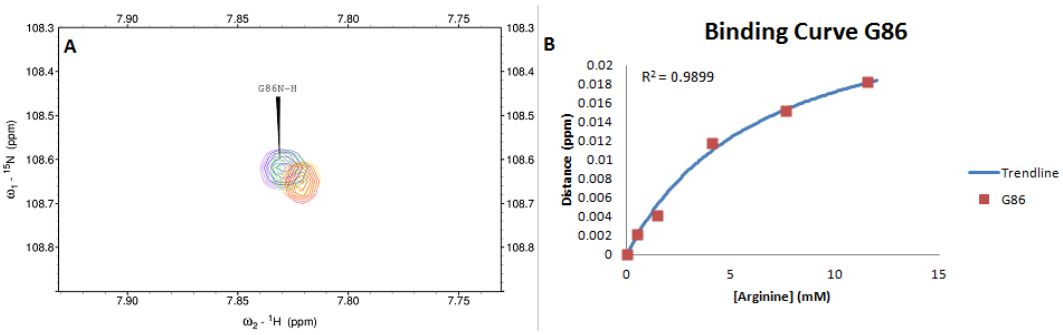


Figure 4.28. Arginine binding to residue G86. (A) The chemical shift perturbation distances measured upon arginine titration. (B) Collected data was used to estimate $B_{max}$ and $K_d$ values of 0.029ppm and 6.5mM respectively using the ligand binding function. These numbers strongly suggest the presence of specific binding of arginine to OEP16.

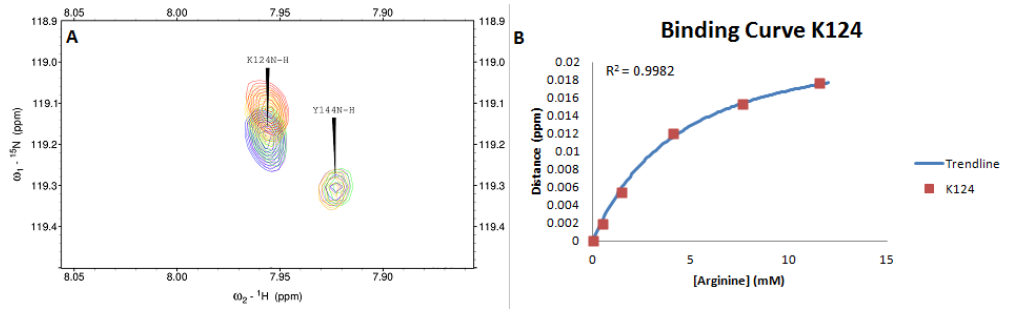


Figure 4.29. Arginine binding to residue K124. (A) The chemical shift perturbation distances measured upon arginine titration. (B) Collected data was used to estimate $B_{max}$ and $K_d$ values of 0.027ppm and 4.9mM respectively using the ligand binding function. These numbers strongly suggest the presence of specific binding of arginine to OEP16.

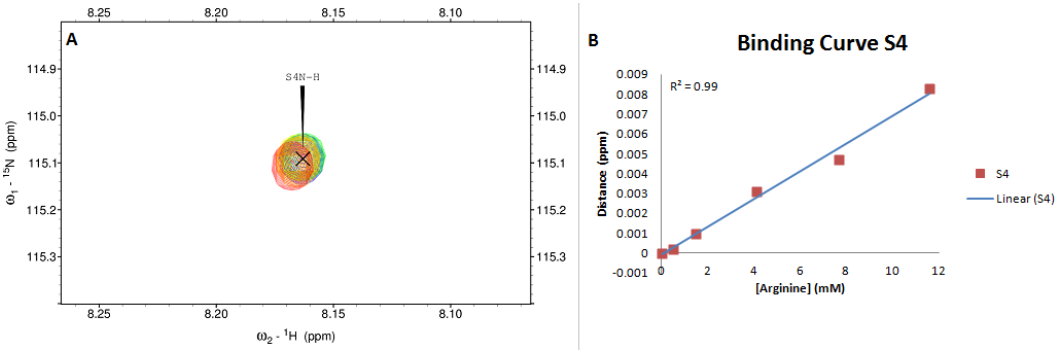


Figure 4.30. Arginine binding to residue S4. (A) The chemical shift perturbation distances measured upon arginine titration. (B) Collected data fits very well to a linear curve which strongly suggests any binding present is non-specific.

Due to the small chemical shift perturbation distances reported by the arginine titration experiment, the estimated $B_{max}$ provides a clearer comparison of ligand binding, therefore a $B_{max}$ has been estimated for each resolved residue in the $^{15}$N-HSQC spectra and compared to the secondary structure of OEP16 reported by TALOS, and is shown in figure 4.31. The six residues described above display a much larger $B_{max}$ than the other residues (summarized in table 4.4), which may be indicative of their importance in arginine binding.

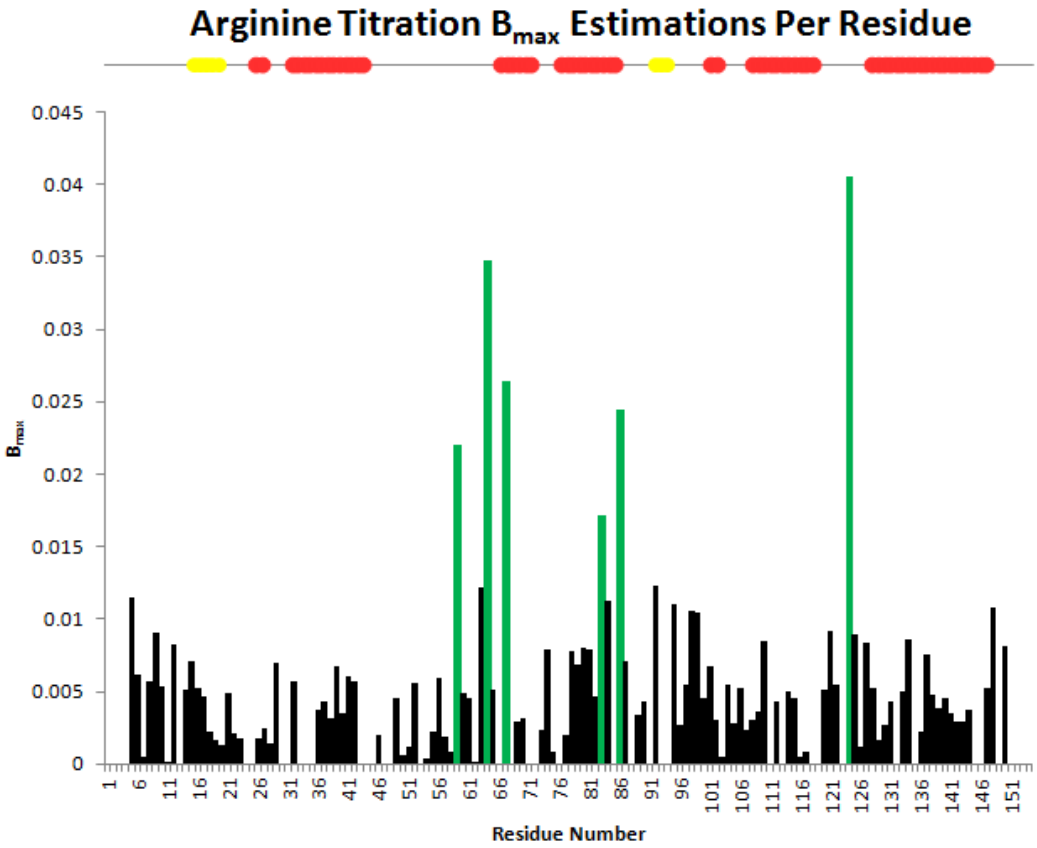


Figure 4.31. Estimated $B_{max}$ as a function of residue number. The chemical shift perturbation values are minute, but six residues display a nonlinear relationship upon arginine titration and can easily be seen in the plot. Secondary structure is plotted above with lines denoting loop regions and circles indicating α-helical content calculated by TALOS as well as identification of the transmembrane (TM) helices.

Using TALOS+ and the arginine titration data, a model of OEP16 can be made as it traverses the membrane, shown in figure 4.32. The red residues indicate helices throughout the protein and green represent residues that are responsible for ligand binding. Further analysis is addressed in the discussion section.

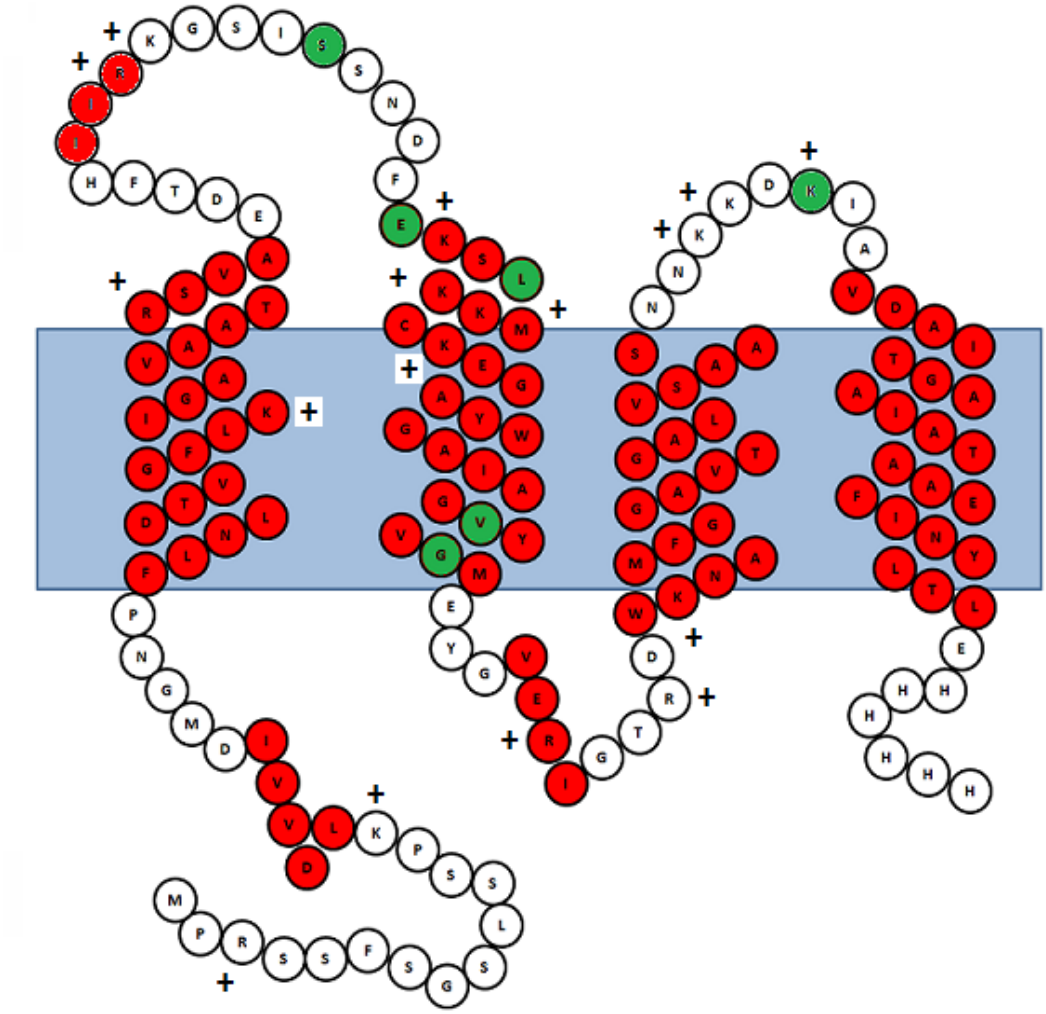

Figure 4.32. Model of OEP16 using TALOS calculations and primary structure. Red residues indicate likely α-helices formation. Arginine and Lysine residues are labeled with a + to show that there are nearly twice as many positively charged species on the side of the membrane opposite of the protein termini which suggests that the loops between helix one and two and between three and four are on the cytoplasmic side of the outer chloroplast membrane by the 'positive inside rule' [106]. The amino acids suggested to be involved with ligand binding are colored green. The location of the green-labeled residues help confirm that the protein is oriented with the two loop regions on the cytosolic side as they are needed to bind amino acids from the cytosol and transport them into the chloroplast.

| Residue | $K_d$ (mM) | $B_{max}$ (ppm) |
|---|---|---|
| S59 | 5.8 | 0.023 |
| E64 | 6.9 | 0.036 |
| L67 | 8.4 | 0.025 |
| V83 | 2.4 | 0.017 |
| G86 | 6.5 | 0.029 |
| K124 | 4.9 | 0.027 |

Table 4.4. Estimated $K_d$ and $B_{max}$ values. Six residues displayed specific binding from the arginine titration experiments.

### 4.7.6 $D_2O$ Titration Results

Another method that was used to look at the structure of OEP16 in SDS micelles was to perform a titration experiment that revealed information regarding the solvent accessible areas of the protein. This was done by titrating increasing amounts of $D_2O$ into a solution of $^{15}$N-labled OEP16 in SDS micelles and then performing a 2D $^{15}$N-HSQC experiment and monitor the changes in signal intensity for each resonance. A more detailed explanation of this experiment is discussed in the materials and methods section 3.3.6.2.

Amides that contain a deuterium atom rather than a proton will not provide an NMR signal when an $^{15}$N-HSQC is performed. Therefore as the concentration of $D_2O$ is increased, the signal intensity from the $^{15}$N-HSQC will decrease. If the amide backbone of the protein is equally accessible to solvent, the rate of intensity decrease will be the same for all residues. However, if parts of the protein are protected, either by a detergent micelle or by the tertiary fold of the protein, the rates of intensity decay will vary based on solvent accessibility. Differences in exchange rate can provide evidence for the presence of a tertiarily folded protein. Additionally, rate of proton exchange performed by the amide backbone is further retarded due to the presence of hydrogen bonds. The

results of the $D_2O$ titration are shown in figure 4.33 as a plot of the rate of intensity decrease versus residue number. The larger the rate decrease, the more solvent exposed residues. Figure 4.33 shows a trend of rising and falling intensities that correlate well with the helical residues calculated by TALOS. The fastest rate of decay is located at either termini of the protein and between residues D48 and F63. Other peaks in the plot are shown between residues E92 and M104, and between N120 and D128. These results are in good agreement with the TALOS and data for where the transmembrane regions of the protein are, and therefore provide additional evidence that OEP16 in SDS micelles form a tertiarily folded structure.

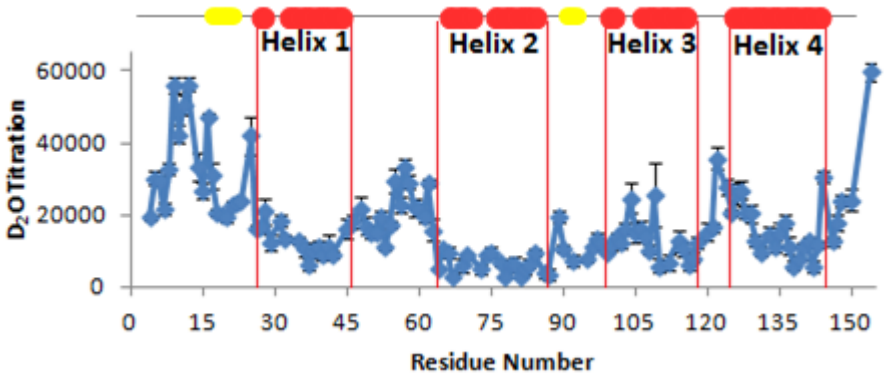


Figure 4.33. Rates of intensity decay upon titration of $D_2O$. High rates of decay indicate areas of the protein where the amide hydrogen is easily exchanged. Lower rates of decay suggest regions of the protein that are protected, either by the detergent micelle or protein intra-contact points. Also indicated by the red lines are the transmembrane helices from TALOS data. The data here agrees well with TALOS and relaxation experiments regarding the location of the transmembrane helices.

# 5. Discussion of Results

## 5.1 Purification of OEP16

### 5.1.1 Purification and Reconstitution using Method 1

Despite the short comings of the original purification method described by Ni and Zook *et al* [25], called Method 1, it provides the best method to date for how to purify and refold OEP16 in SDS micelles. This method provides higher yields of properly folded OEP16-SDS complexes compared to Method 2 described in this dissertation, which was optimized for isolation of OEP16 in mild detergents. More importantly it has the ability to obtain large amounts of monodispersed OEP16 in SDS micelles without aggregation. This contrast to the new method where increased aggregation is observed seen by SEC studies (section 4.4). Even though the Method 1 is still used for SDS, additional modifications where needed to obtain higher yields of $OEP16^{His}$. The improvements of this method were designed to better control the purification and refolding process.

The first improvement to Method 1 was to use an HPLC to purify OEP16 from solubilized inclusion bodies. This was previously done using a gravity column sometimes aided by a peristaltic pump to drive the mobile phase. The improvement provided here is a real time observation of protein elution that could be autonomously collected using a fraction collector in order to get the highly purified product. The second improvement is the masking. Masking refers to partially refolding OEP16 through dilution to 3M urea while introducing 1% SDS to protect (or mask) the strongly hydrophobic regions. Previously this was done by hand, slowly diluting the protein solution in urea drop-wise over an hour with the detergent buffer. It was hypothesized early in my studies that this was a major point where protein starts to aggregate and was

lost due to improper folding. Instead the HPLC was used to control the flow of solution. The HPLC flow rate (in mL/min) was calculated using the simple expression:

$$Flow\ Rate = \frac{\frac{8V_1}{3} - V_1}{60}$$

$V_1$ (units of mL) is the volume of purified OEP16 in 8M Urea. The HPLC was then programmed to flow the dilution buffer at the calculated flow rate for one hour, dripping into the OEP16 protein solution stirring at 200rpm. This allowed an extremely precise level of control for slowly diluting the solution which resulted in minimal protein loss as observed in the final HPLC runs to refold the protein seen in figure 4.1b.

Significant yields of purified and refolded OEP16 in SDS micelles were obtained for all isotropically expressed protein. The 80% perdeuterated $^{13}C$, $^{15}N$ sample preparation resulted in an average of 600µL solutions that contained 1mM (16.6mg/mL) OEP16. The $^{15}N$ labeled sample used for the NOESY-HSQC, TOCSY-HSQC, relaxation, and titration experiments had a yield of 1.5mM (24.8mg/mL) for 1mL of solution. Each sample was additionally purified via gel filtration to remove any aggregates and ensure a monodispersed sample.

5.1.2 Purification and Reconstitution of OEP16 using Method 2

The modified purification and reconstitution of OEP16 provided several benefits over the initial purification scheme developed in Method 1. While the initial method was optimized for SDS, only very low yields could be achieved with milder detergents. The optimized purification Method 2 allows purification and refolding of the recombinant $OEP16^{His}$ in a broad variety of milder detergents. Furthermore, Method 2 removed the necessity to partially mask the strongly hydrophobic areas of OEP16 by dilution. The strongly hydrophobic areas of the protein are masked directly on the column with a 1% SDS, 8M Urea buffer prior to removal of urea. This allowed for a stepwise gradient to be

performed to slowly remove the SDS and urea while introducing a milder detergent. Method 2 therefore allowed folding of OEP16 in a single HPLC run directly on the column in a large range of mild detergents (listed in figure 4.2).

In Method 1the only detergents suitable for purification and reconstitution were SDS and SLS with low yields achieved in $C_{12}E_8$ and β-DDM, while SB12 and Triton X-100 could not elute the protein from the column at all. OEP16 can now be reconstituted in ionic, zwitterionic, and non-ionic detergents with good yields. However this new method does not work with all detergents, such as CHAPS, β-OG, and SB12.

Cymal-4, β-OG, DPC, SB12, and CHAPS were first selected for their ability to form relatively small micelles (<25kDa). OEP16 is a relatively small protein compared to other membrane proteins and therefore concentration of protein-micelle complexes without co-concentration of empty micelles becomes challenging, especially with larger micelle sizes. It is also important to keep the protein-micelle size small for NMR experiments in order to maintain fast correlation times. Rotational diffusion becomes much slower as the particle becomes larger, and therefore $T_2$ values become longer, which is problematic for sufficient resolution of NMR spectra. Yet other detergents, such as Triton X-100 and β-DDM were also tested so that comparisons could be drawn between this study and the results reported by Ni and Zook *et al* [25].

The problem with the reconstitution of OEP16 in mild detergents is that the protein obtained in these a few of these detergents is stable only at relatively low concentrations (<2mg/mL or <12μM) and is far too low for NMR studies. Concentration attempts usually result in the formation of precipitate in the form of two dimensional lamellar sheet-like crystals. This may prove useful if solid state NMR is used to explore the structure of OEP16, but becomes impossible to perform solution NMR experiments. If in depth NMR solution spectroscopy is to be used to study OEP16 in these conditions,

it will be necessary to find conditions where the protein remains soluble for extended periods of time (days to weeks). Future studies may include the micro dialysis method for screening conditions developed by Ni and Zook *et al* [25]. Several conditions have already been attempted, including various ionic strengths, pH, glycerol concentrations, and temperatures.

## 5.2 Circular Dichroism Investigations

### 5.2.1 CD Spectra and the Secondary Structure Analysis

All CD spectra measured for OEP16 display the characteristics of an α-helical protein with a double minima at 220nm and 208nm with a strong maximum at 192nm. The intensities of these bands differ slightly between different OEP16-detergent complexes, suggesting a range of α-helical content depending on the detergent used. The CD deconvolution shows a range of α-helical content (33% to 71%) for different detergents. However, in all detergents used, α-helices are the major secondary structure element of OEP16 with almost no evidence of β-sheet formation, which is also in agreement with the results published by Ni and Zook *et al* [25].

The α-helical content for all stable OEP16 detergent complexes consistent with a protein that can traverse a membrane four times and thereby agrees well with the four transmembrane helical model suggested by Linke *et al* [24] that estimates that the transmembrane α-helical portion of the protein constitutes 50% of the residues. DPC, SB12, β-DDM and Cymal-4 show an increased α-helical content and likely provide an environment that is favorable for formation of α-helices in the loop regions.

Although Ni and Zook *et al* describes a purification / reconstitution method that works well with SDS and SLS samples, smaller amounts of $OEP16^{His}$ in β-DDM micelles for CD spectroscopy was obtained using Method 1. Using Method 1 reported by Ni and Zook, a secondary structure containing only 36% α-helices is reported [25] for OEP16-β-

DDM complexes. In contrast, using the modified purification Method 2, the OEP16-β-DDM complexes displayed 54% α-helical content, as determined from CD experiments. This adds evidence that the modified Method 2 purification is superior to the Method 1 when milder detergent conditions are used for refolding OEP16.

β-OG is considered insufficient to purify and stabilize OEP16, but very small amounts can be eluted from the column and precipitates several hours later. CD data shows that there is only 33% α-helical content in the OEP16-β-OG micelle which is significantly less than other conditions studied. This also does not provide enough α-helical secondary structure to traverse the membrane four times. This, combined with a low $T_m$ (compared to other OEP16 samples in 2x cmc detergent), suggests that the protein may be partially folded or misfolded in β-OG.

SB12 is considered unable to refold OEP16 because detergent concentrations of 10x cmc were required to maintain protein solubility. Without the high detergent concentration the protein precipitated onto the Ni-NTA column. All structural investigations were performed on OEP16 in 10x cmc SB12. A detailed discussion on why 10x cmc SB12 is needed to maintain solubility is discussed in secion 5.4.

5.2.2 Melting Profiles Observed by Circular Dichroism

With the exception of SDS, the melting profiles show that there is a fairly narrow $T_m$ range (less than 10°C) between the different OEP16-detergent complexes. The sigmoidal shape of the melting curves also suggests a cooperatively unfolding mechanism, although the amount of cooperativity observed is considerably different among the different detergents used. OEP16 in SDS micelles are extremely stable, and does not show evidence of unfolding (through CD) at temperatures greater than 100°C *and* with 5.8M urea. However, this is not a unique phenomenon; the c-rings of ATP-synthase are "resistant to boiling in SDS" stated in the 2002 publication by Arechaga *et*

*al* [107]. Therefore it is very probable that OEP16 maintains tertiary structure even under strong denaturing conditions.

5.3 The Multimeric State of OEP16 by DLS and SEC

The quality of DLS data appeared to be partially dependent on the amount of detergent needed to maintain a 2x cmc concentration. The OEP16-detergent complexes produced from lower cmc detergent systems, such as DPC and β-DDM with cmc values of 0.5% and 0.01% respectively, appeared to have a narrow distribution of radii in the histograms calculated by DLS. This suggests a well behaved, monodispersed sample at 3.8nm and 4.3nm for OEP16 in DPC and β-DDM respectively. In contrast, the OEP16-detergent complexes produced with high cmc detergents such as Cymal-4 and SDS (0.37% and 0.3% respectively, displayed more polydispersed histograms; the polydispersed histograms are likely a result of the overlap of similarly sized OEP16-detergent complexes and empty detergent micelles which scatter more light due to the increased concentration, or possibly differing oligomeric states. DLS results of OEP16 in SB12 (data not shown) displayed an extremely polydispersed sample which is likely due to the significant amount of detergent needed (10x cmc) to maintain the folded state. From DLS the size of the OE16-detergent particles to be about 4nm which suggests a multimeric state close to a hexameric value shown by values in table 4.2. Although the DLS results are unable to confirm an exact multimeric state, they agree with SEC studies that OEP16 may form multimers.

The oligomeric state determined by SEC and DLS results differ significantly compared to the monomeric state determined from NMR relaxation results. Obtaining absolute sizes on the OEP16-detergent complex proves to be more difficult than soluble proteins. Each method utilized in these studies has both benefits and short comings regarding their ability to estimate a molecular weight and multimeric state.

DLS data for protein-detergent complexes of this size (~2-4nm) is difficult to analyze and approaches the lower limit (~1nm) for practical scattering experiments. This is due to several problems. This first is that scattering intensity increases as the *sixth* power of the particles diameter [108]. That means that scattering is going to be much less intense for a smaller particle. Therefore if an equilibrium exists between oligomeric states, the largest multimeric state is going to present a signal that is much larger than the monomer; this means that even if the equilibrium of the multimeric state favors the monomer, the majority of the scattered will be from the multimer.

One of the major pitfalls of SEC of membrane proteins is that standard curves are generally prepared using water soluble proteins. Therefore I must assume that a protein-micelle complex will diffuse in and out of the porous gel matrix the same way aqueous proteins do and that the protein, nor the detergent interacts with the column material. Also, because detergent is required for the membrane protein buffer, but may unfold soluble proteins, the standard protein buffer is not identical to the membrane protein buffer. However, despite these shortcomings of both SEC and DLS, it appears that OEP16 most likely forms a multimer in SDS micelles.

## 5.4 Stability of OEP16 in Various Detergent Complexes

Ni and Zook *et al* suggested that the chain length, net charge, cmc, and micelle size had minimal-to-no influence on the potential of the detergent to successfully fold and stabilize OEP16 [25]. However, combined with the data reported in this manuscript, it is evident that the presence of a negative charge is extremely beneficial to the ability to fold $OEP16^{His}$. However, the location of the negative charge within the head group of the detergent molecule may also be important. SDS and sodium lauryl sarcosinate (SLS) are both anionic and were particularly successful at purifying OEP16 with high yields using Method 1[25]. In the new Method 2, DPC was particularly successful in purifying

refolded OEP16 that displayed a high $T_m$ in melt profiles as well as maintained a high level of monodispersity in DLS experiments.

OEP16 in DPC micelles could not be concentrated, however. Concentration was attempted through ultrafiltration using a 30kDa MWCO filter using a wide range of spinning speeds and times. Regardless of how the concentration was attempted or what buffer condition was tried, OEP16 in DPC precipitated at concentrations greater than 5mg/mL. This does not seem to be an isolated incident; DPC is widely used in low concentration protein characterization techniques, but only one crystal structure (ompF) has ever been solved using DPC as the detergent [98]. This may suggest that as a general rule, DPC cannot maintain protein solubility at high protein concentrations; or that DPC may promote the formation of β-sheet structures. This hypothesis is supported by observation of the 2D 15N-HSQC spectra of OEP16 in DPC micelles that quickly precipitated out of solution and that ompF is a β-sheet protein [98].

On the other hand, OEP16 in SB12 micelles remained soluble only in ~10x cmc detergent and displayed a significantly lower $T_m$ value, even though DPC and SB12 are both zwitterionic C12 detergents. It can be argued that this difference is caused by the differences in the position of the positive and negative charges: the positively charged quaternary amine group is directly bound to the alkyl chain in SB12 (and CHAPS) while the negatively charged phosphate group is directly bound to the alkyl chain in DPC (see figure 5.1). This could have an influence on the stability of OEP16 as the 4-helix model of OEP16 predicts that a number of positively charged amino acid residues are located in the extrinsic loop regions of the protein seen in figure 5.2. OEP16 contains six arginine residues and eleven lysine residues accounting for over 10% of the amino acid content of wild type OEP16. The electromagnetic repulsion of these positive charges can be resolved through the use of anionic detergents as seen in SDS and SLS, or through a

zwitterionic compound where the positively charged residue can interact with the negative charge as observed in DPC. However, due to the positive quaternary amide on SB12, the repulsion of the positively charged amino acids with the positively charged head group could lead to the instability of the protein- micelles. This argument can also explain why OEP16 precipitates onto the Ni-NTA column while attempting to refold the protein in CHAPS. Although CHAPS has a very different molecular structure, it has the same charge layout as SB12 which could cause protein instability by repulsion of positively charged amino acids with the positively charged portion of the head group. Experiments including the use of 10x cmc CHAPS have not been attempted as done for SB12.

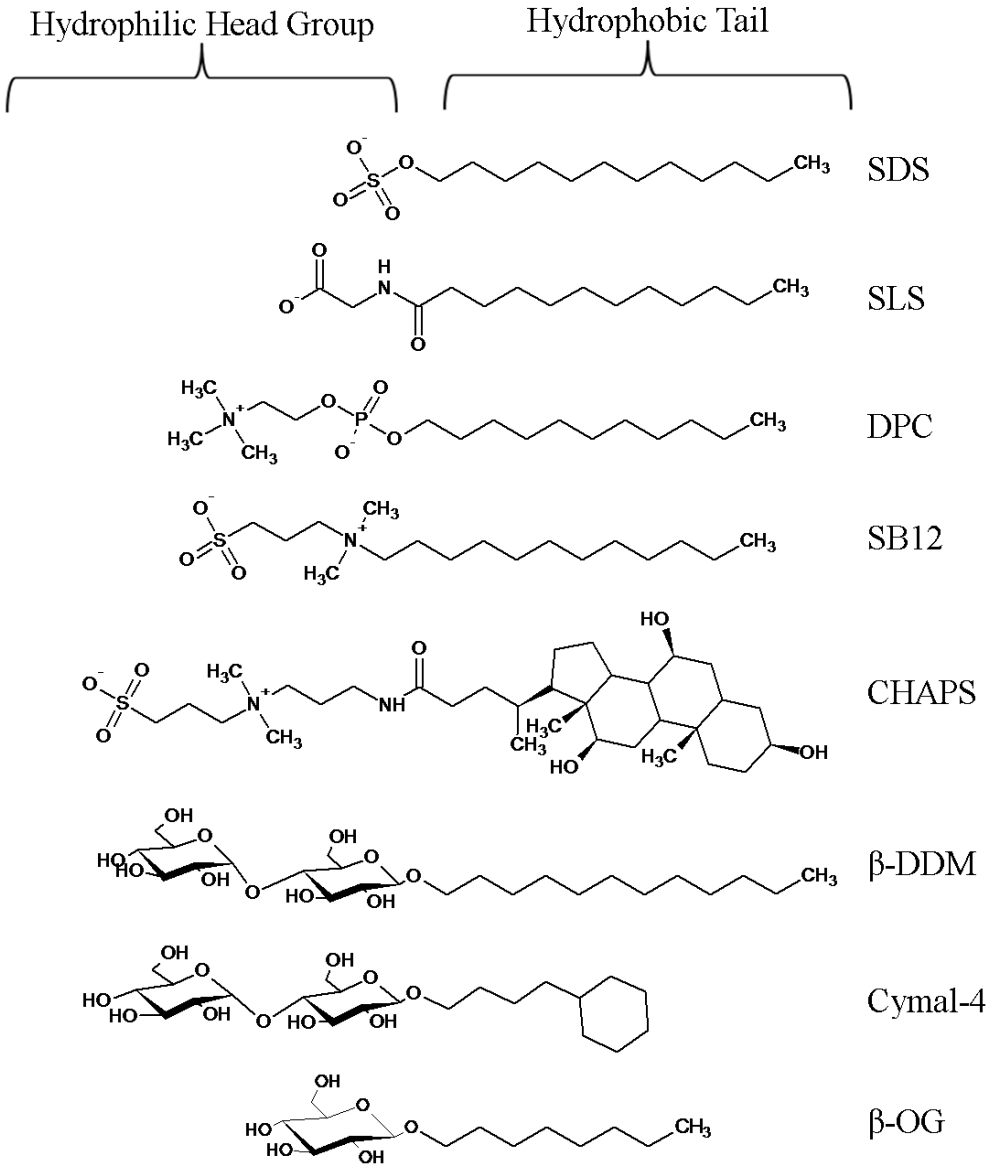


Figure 5.1. Molecular structures of different detergent molecules used to purify and refold OEP16. In general, detergents that have a large polar head group and an alkyl chain >8 carbons work well. It also seems significant that if it is a charged species, that a negative charge should be in close proximity to the alkyl chain.

There are several possibilities why OEP16 in β-OG reveals a poor α-helical content by CD studies and does not remain soluble. The most obvious possibility is that the short chain length of β-OG (C8) is not enough to protect the hydrophobic regions of OEP16. However I stipulate that the contributing factor is the size of the non-ionic polar

head group. Even though the alkyl chain is significantly smaller than β-DDM, it remains about the same physical length as Cymal-4 as seen in figure 5.1. Therefore it is likely that the additional stability provided by Cymal-4 and β-DDM is through the addition of a glucopyranoside group to the hydrophilic side of the surfactant to form a maltopyranoside moiety. This can also be seen by the ability for Triton X-100 to properly fold the protein. The longest carbon chain of Triton X-100 in the hydrophobic tail is only eight carbons long, and therefore it is possible that the increased stability compared to β-OG is a result of the very large polar non-ionic head group.

## 5.5 Nuclear Magnetic Resonance Spectroscopy

TALOS calculations predict that OEP16 forms an exclusively α-helical protein with flexible loop regions (primarily at the protein termini and the loop between helices one and two). The overall alpha helix content calculated by TALOS is about 55% which fits very well with the deconvolution results of the CD spectra and the hydropathy plot analysis performed by Linke *et al* [24] of about 50%. For the first time there is experimental evidence that describes where in the protein these α-helices are likely to reside. The TALOS estimated α-helical content and order parameters combined with the $D_2O$ titration experiment and relaxation studies provide detailed information about where the protein exists within the membrane and can be summarized by figure 5.2 and are used to construct the models illustrated by figures 4.32 (in methods section) and 5.4 The TALOS calculations from chemical shift data agree very well with the data obtained from relaxation and $D_2O$ titration experiments which provides increased confidence in the backbone assignment. Additionally, where there is evidence of secondary structure from TALOS calculations (5.2a), there is an increase in the $T_1/T_2$ ratio (5.2c) as well as a decrease in the rate of intensity decay (5.2d) which are all indicators of protein structure.

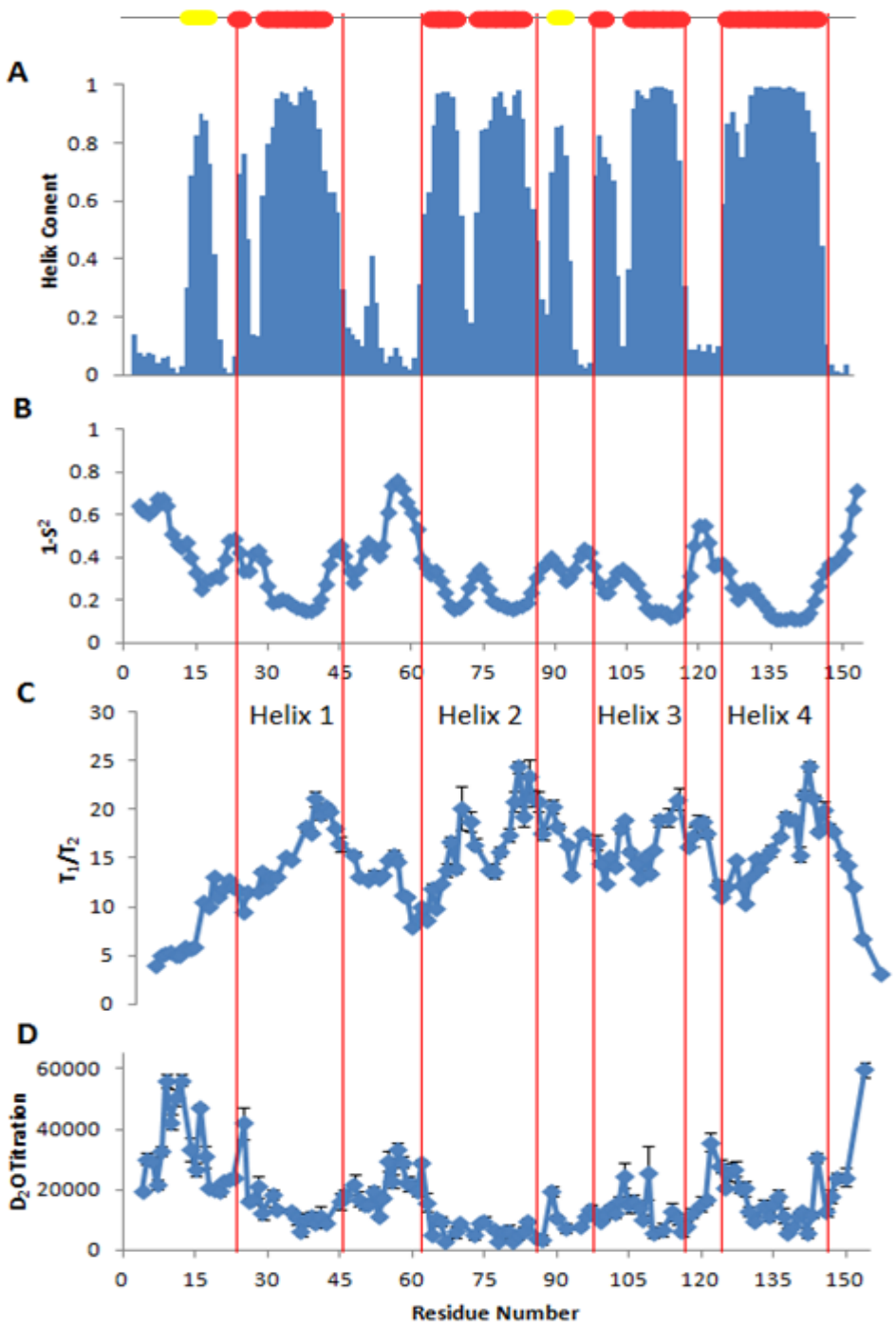


Figure 5.2. Comparison of TALOS calculations to relaxation data and $D_2O$ titration data. (A) TALOS calculated secondary structure and (B) calculated order parameters (displayed as $1\text{-}S^2$) fit very well with the (C) relaxation data and (D) $D_2O$ titration data. The agreement between all experiments provides confidence in residue assignment and secondary structure. With the helices indicated in red it the data shows clearly the transmembrane regions of the protein.

The secondary structure prediction by TALOS suggests that there is a small roughly 3 residue break within each of the four TM helical regions: one centers around N27, the second centers around G74, the third at about G106, and the fourth at T131 as illustrated in figure 5.3. However, the NOESY data still suggests the presence of helical content due to the presence of an NOE between the alpha hydrogen ($H_\alpha$) of N27 and the $H_\alpha$ of D31 in the NOESY-$^{15}$N-HSQC experiment. It is possible that the helices may be more distorted than completely broken; therefore the data my imply that there is a kink or gap in the helices. The location of each of these distortions is interesting as well: they all are at or near the N-terminus of the transmembrane helices of OEP16. Therefore when the N-terminus of the protein is oriented 'down' in the membrane, the breaks for helix one and three are closer to the bottom, while the breaks for helix two and four are located at the top as illustrated in figure 5.3. It is possible that these flexible regions function either for specificity of substrate diffusion, or to open and close as a way to diffuse the amino acids into the chloroplast.

This presents an interesting point in what orientation the protein is inserted into the membrane. A rule dubbed "the positive-inside rule" by von Heijne and Gavel refers to the phenomenon that an integral membrane protein will insert into the membrane in an orientation where the positively charged lysine and arginine residues are located on the inside of the cell membrane [106]. The idea is that the most positive extrinsic region of an integral membrane protein will orient on the same side of the membrane that it was expressed. For OEP16 which is encoded by the nuclear genome, this means that it will orient with the most positively charged residues are on the outside of the chloroplast. Therefore, according to figure 4.32, loops two and three are located on the cytosolic side of the membrane.

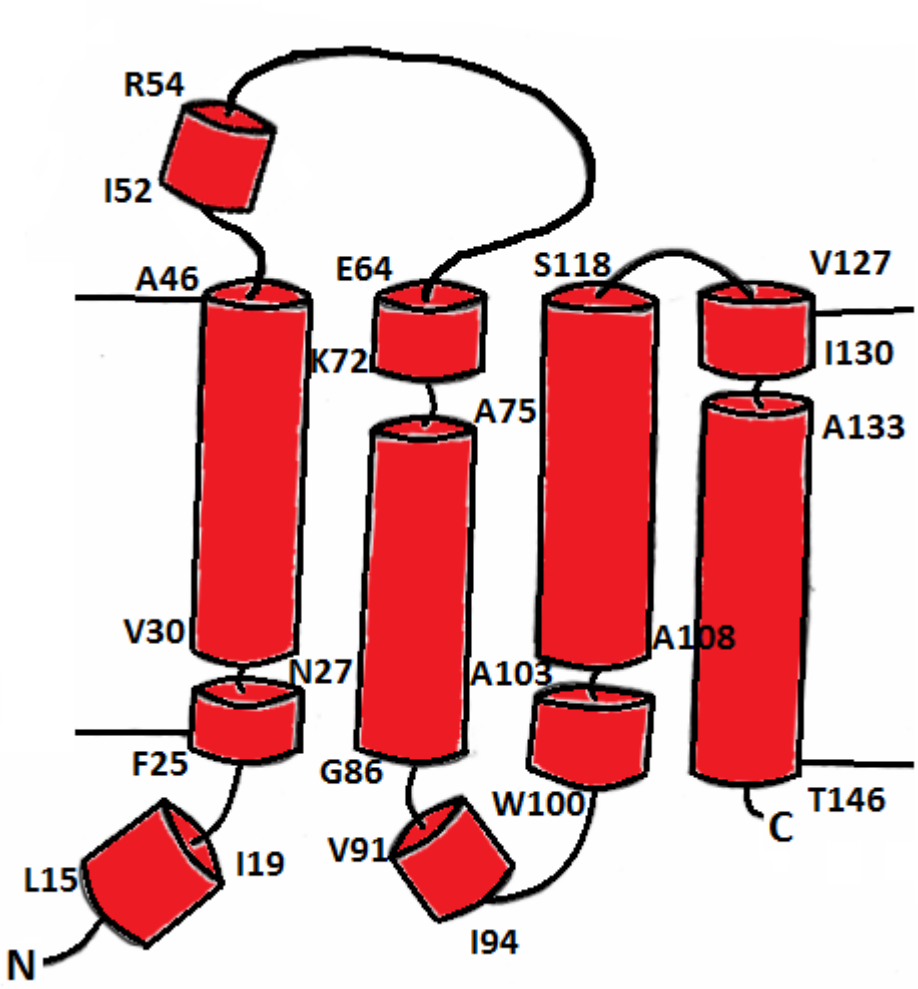


Figure 5.3. Model of OEP16 estimated from TALOS calculations, relaxation data, and the $D_2O$ titration experiment. This model draws attention to the characteristics of the helices, specifically the transmembrane helices where there is a gap or kink located at the C-terminus of each helix as well as the possible solvent exposed extrinsic helices.

All the NMR data and calculations collected regarding the structure of OEP16 agreed with each other. In addition, it also agrees with observations seen previously in published literature. The location of the transmembrane helices was suggested by Linke *et al* using hydropathy plot data. Experimental data from the NMR experiments and calculations performed in this study agree very well with Linke's model (figure 5.4) [24]. Additional CD results obtained in this study as well as those published by Ni and Zook *et*

*al* suggest that the alpha helical content of OEP16 in SDS micelles is about 50%, TALOS reports a slightly higher helical content at about 55% [25].


Figure 5.4. Location of transmembrane helices determined from different experimental data. TALOS secondary structure (green) and order parameter calculations (purple) agree very well with the location of the four transmembrane helices estimated from relaxation (red) and $D_2O$ titration (blue). The location of the helices by the experimental NMR data are also in agreement with the location previously suggested by Linke *et al* (yellow) from hydropathy plot analysis [24]. Each experiment individually confirms the location of the transmembrane helices.

Relaxation data on OEP16 taken at two different temperatures, 37°C and 25°C strongly suggest a hydrodynamic radius of OEP16-SDS particles to be smaller than the radius calculated by DLS and gel filtration chromatography. However, NMR is the only experimental method used that looks specifically at the protein, rather than the complete protein-micelle; therefore certain considerations may be needed to be made regarding how membrane proteins interact with the detergent micelle.

Figure 5.5 shows the four helix bundle of OEP16 in the anionic SDS detergent. The anionic nature of the SDS molecule creates a strongly negatively charged

environment along the outside of the transmembrane region of the proteins. In contrast, the extrinsic lysine and arginine residues create a positively charged environment at the top and bottom of the four helix bundle. Therefore the charge distribution of the protein micelle complex is not uniform throughout the OEP16-SDS micelle complex. It may be possible that anisotropic tumbling is a result. In order for the OEP16-SDS complex to tumble along the red arrow in figure 5.5, a large rearrangement of the water-shell is necessary: the partially negative oxygen atom of the water molecules that were associating with the positive lysine and arginine charges now require to arrange so that the partially positive water protons can associate with the negatively charged SDS micelle. However, no rearrangement is needed if the protein rotates in the direction of the green arrow in figure 5.5, therefore it is possible that OEP16 in SDS micelles tumble anisotropically.

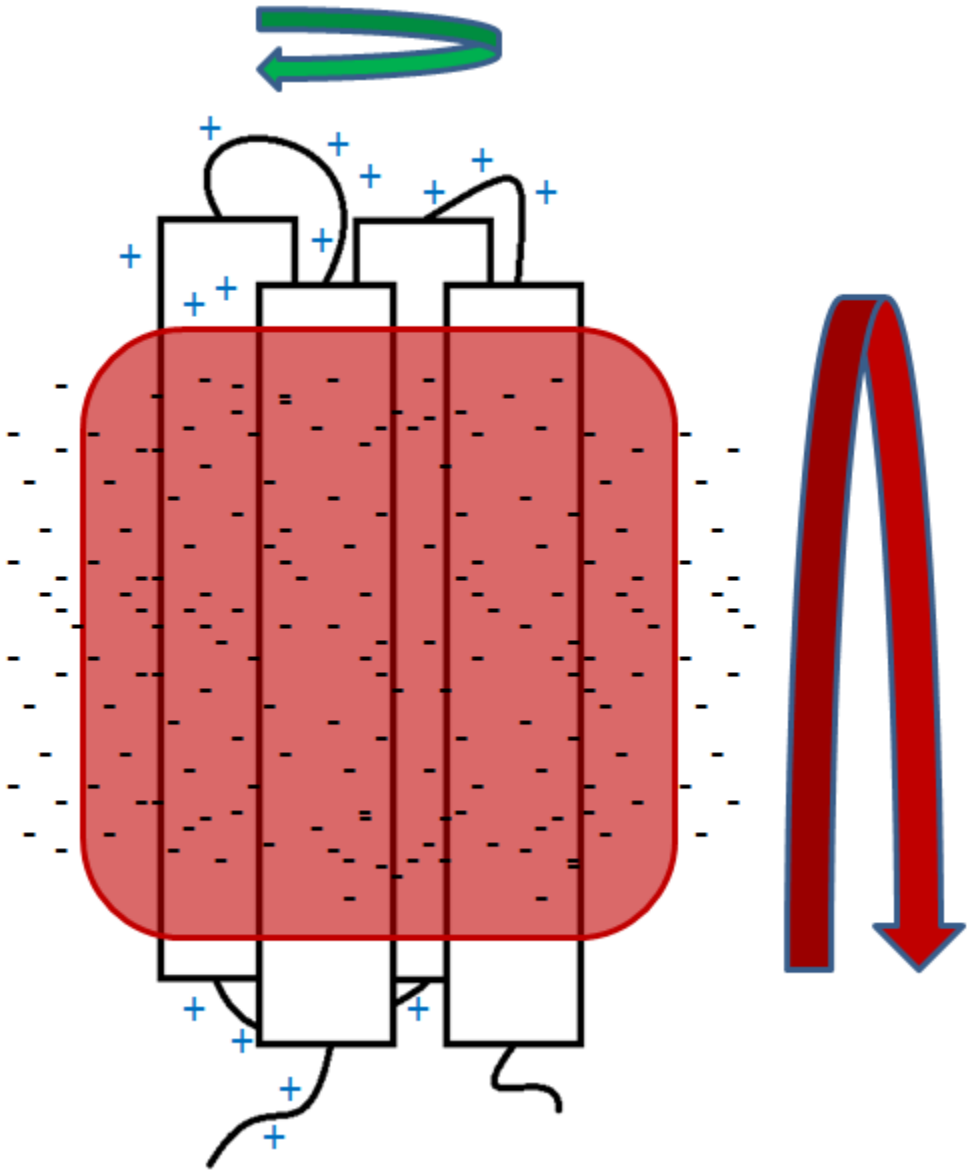

Figure 5.5. A schematic diagram of the four helix bundle OEP16 in the anionic SDS micelle. The positive charges at the top and bottom of OEP16 and the negative charges of the SDS micelle provide a charge distribution that is not uniform and result in anisotropic tumbling due to large watershell arrangement neeeded to tumble in the direction of the red arrow. Additionally, because OEP16 is not covalently bound to the micelle, it may be probable that the rotation of OEP16 is different along the green arrow compared to the micelle rotation. These considerations must be made when attempting to obtain information regarding the size of OEP16 from relaxation data.

Another consideration that needs to be made when considering the motion of only the protein is how the hydrophobic OEP16 interacts with the SDS micelle. It may

be possible that OEP16 rotates along the green arrow at a different speed compared to the SDS micelle. There are no bonds formed between OEP16 and the SDS micelle that would require the micelle to rotate at the same speed as the protein. This will then also contribute to anisotropic motion of the OEP16-SDS complex. It is possible that these considerations could be resolved with additional relaxation experiments discussed in future outlook section 6.2.

If only the rotation of OEP16 is being measured and the resulting molecular weight is ~32kDa, this is close to an OEP16 dimer. This would be useful because it is highly unlikely that a monomeric, four helix bundle OEP16 can provide a 10Å pore to diffuse amino acids through. A multimer is much more likely, and would agree with the DLS and SEC results that also suggest OEP16 in SDS is a multimer. A useful NMR experiment is a pulsed gradient spin-echo experiment developed by Nesmelova *et al* as a way to obtain a translational diffusion coefficient and compare the resulting hydrodynamic radius to the values obtained from rotational diffusion NMR experiments and DLS measurements [109].

To test if the helices packed together in a tertiary protein fold, an α-helical wheel was constructed for all of the transmembrane regions of the protein. The $D_2O$ titration experiment was then used to elucidate which residues within the transmembrane region are more solvent exposed, shown in figure 5.6. For all helices, it appears that roughly half of one side of the transmembrane region is solvent exposed while the other side is less accessible to proton exchange and therefore less solvent exposed. It is likely due to the hydrophobic effect that the transmembrane helices face each other in the tertiary folded state. The least solvent exchanged residues tend to be the glycine and alanine residues, which are capable of forming very tight packing within a helix. Additionally helix one and two have multiple charged residues on the solvent exposed side of the

helical wheel which may be involved with forming the channel pore. Helix three appears to have three contact points that may provide insight into how the four helix bundle is packed depicted as a general schematic in figure 5.7 should OEP16 form a trimer.

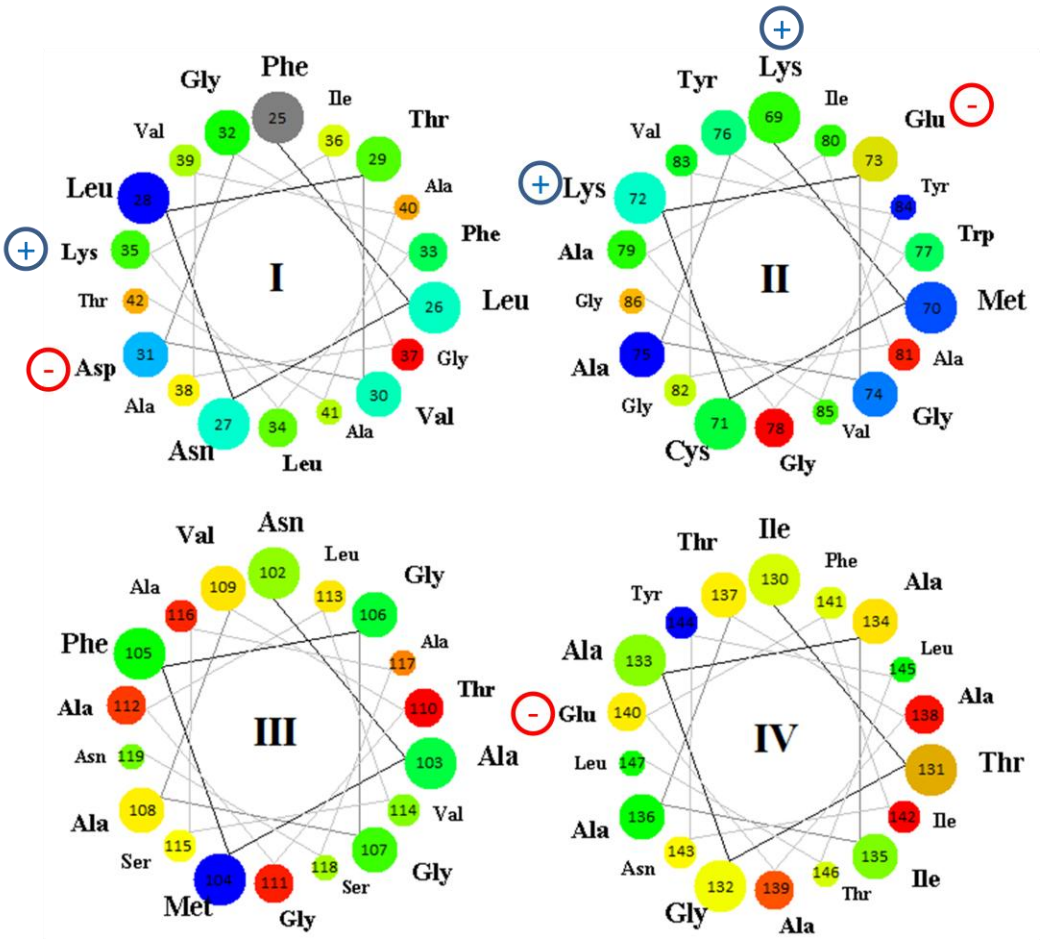


Figure 5.6. Helical wheel showing the four α-helical TM regions of the protein. Included is a color coded representation on the solvent exposed region of the helix based on NMR $D_2O$ titration results: Red is the most solvent protected, with blue being the most solvent accessible. The residue becomes more solvent exposed the more blue it becomes.

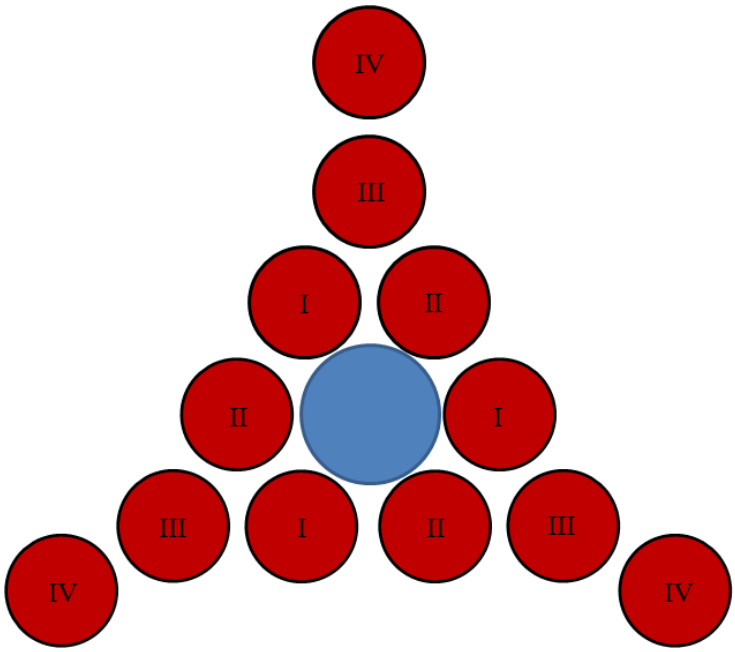


Figure 5.7. A model of OEP16 based on information obtained from the helical wheel analysis of the $D_2O$ titration experiment. Helix one and two form the channel pore (blue) to allow diffusion of amino acids, helix three stabilizes helix one and two while helix four may help stabilize helix three. The model is shown as a trimer, but other multimeric states are allowed using this model: helices one and two form the pore, stabilized by helix three, and helix four sticks out from the center like a bicycle spoke.

The presence of chemical shift perturbation upon titration of arginine, an amino acid that is known to be translocated into the chloroplast by OEP16, gives strong evidence that OEP16 is in a conformation that binds arginine, especially considering that the binding curve presents in a non-linear fashion with estimated $K_d$ values between 3mM and 8mM consistent for selective pore diffusion seen by the sugar selective porin ScrY [105]. Therefore it may be possible that OEP16 exists in its functional state in SDS micelles.

## 5.6 Protein Crystallography

OEP16 crystals greater than 1mm in length were grown during this study, but did not display a morphology indicative of a well ordered, three dimensional crystal lattice. The two dimensional shape of the crystal that could easily bend in solution upon

inversion suggest more of a lamellar sheet orientation rather than a 3D crystal. These crystals are very similar in morphology to those that were grown by Linke *et al* that diffracted to a resolution less than 20Å [16]. Home source X-Ray diffraction could not reveal any diffraction peaks. Due to the thin layers formed I did not attept further x-ray crystallographic studies. However, because the crystals may be able to some degree in a medium, solid state NMR may prove a useful tool for future investigations.

# 6. Future Outlook

## 6.1 Structure Determination by Solution NMR

The structural investigations of OEP16 in SDS micelles by NMR can be further extended. Several important experiments are yet to be performed in order to obtain structural restraints needed for solving the OEP16 protein structure. The most important experiment will be long range NOESY experiments using amino acid specific labels (Isoleucine, Leucine, and Valine) in order to restrict the number of peaks observed in the spectra. The protein will need to be almost completely deuterated in order to reduce spin diffusion. For these experiments, *E. coli* needs to be grown in >95% $D_2O$ with $^{15}NH_4Cl$, and $d_6$-Glucose while supplementing with $^{1}H/^{13}C/^{15}N$ ILV residues [110]. This will then deuterate (nearly) all of the protein, with the exception of isoleucine, leucine, and valine. Additionally $d_{25}$-SDS and $d_8$-glycerol must be used to prevent spin diffusion as well as prevent unwanted NOE cross peaks between the protein and SDS or glycerol. Using multiple NOE experiments, structural constraints can be obtained for OEP16 which will lead to a 3D structural model of the protein.

Although this strategy seems fairly straight forward in theory, the practice remains difficult. The isotopes required to grow specifically labeled OEP16 is expensive. Additionally there is a significant price tag on $d_{25}$-SDS and $d_8$-glycerol as well; these isotopes cannot be used during the purification and refolding of $OEP16^{His}$ and therefore need to be exchanged via dialysis directly prior to the NMR experiments. This is not particularly a problem for SDS, which is only at a concentration of about 0.3% in solution, but to remove 10% non-enriched glycerol from the buffer would require several dialysis steps and a significant amount of isotope enriched 10% glycerol solution would be needed. Therefore a technique must be developed where less $d_8$-glycerol is needed.

Although NOE restraints are the largest contributers to the tertiary structure of the protein, it may not be sufficient to solve the structure of OEP16. Therefore it is suggested that the use of a weakly aligning media should be used in order to obtain residual dipolar coupling (RDC) restraints. In a weakly aligning media the protein is assumed to take on an elliptical form with a major and minor axis. NMR spectra peaks (specifically the doublet splitting due to the dipole-dipole coupling) are augmented by the weakly aligned molecule based on how X-H bond vectors orients relative to the static magnetic field. The doublet splitting of a weakly aligned molecule is described as [9]:

$$\omega_{IS} = d_{IS} + \pi J_{IS}$$

where $J_{IS}$ describes the J-coupling between the I spin from the H atom and the S spin from the X atom. Therefore $d_{IS}$ is called the residual dipolar coupling and proves its use through the following relationship [9]:

$$d_{IS} = b_{IS}\frac{1}{2}\overline{(3cos^2\Theta_{IS} - 1)}$$

where $b_{IS}$ is a constant based on the spins used (generally $^{13}C$ or $^{15}N$ and $^{1}H$). The significant parameter from this relationship is $\Theta_{IS}$ which is the angle between the X-H bond vector and the static magnetic field. Using a set of geometrical relationships, this angle measurement is related to the angle measurement between the X-H bond vector and the long axis of the aligned molecule,

This method provides a very powerful tool for structure determination, but inserting OEP16 into a weakly aligning media will be challenging. The use of bicelles is a common method for obtaining weak alignment for soluble proteins: the bicelles stack similar to stacks of coins and the soluble proteins become wedged in between the coins, thus orienting them [9]. If a membrane protein can be incorperated into a bicelle, it is possible that solid state NMR can be used as well (RDC require *weekly* aligned molecules). Therefore an alternative method includes using a low percent

polyacrylamide gel that is soaked with the NMR sample and then squeezed into the NMR tube [111]. The gel then orients along the NMR tube which in turn orients the protein sample, allowing for RDC measurements to be made.

The above two methods are likely to be the routes carried out for the ultimate structure determination for OEP16. However alternative methods may be necessary should the above methods fail. The protein crystallization attempts failed to produce crystals that are well ordered and diffracted by X-rays, however these crystals may be candidates for electron diffraction, if they are flat enough. Additionally the use of solid state NMR may also be a possibility for high resolution structure investigations.

## 6.2 Additional NMR Relaxation Experiments and Analysis

Probably the greatest advantage that NMR has over X-ray crystallography is the ability to look at the dynamics of a protein in solution. X-ray crystallography requires that the protein be packed into a rigid, insoluble lattice so that reflected photons can add constructively to form a diffraction pattern, and therefore are incapable of experimentally looking at the protein dynamics in the native soluble state. One of the major ways of visualizing protein dynamics this is through relaxation analysis. Although the relaxation results in this study provide significant information about the protein, more can be done in this area.

An important NMR relaxation experiment is the heteronuclear NOE experiment. The heteronuclear NOE is characterized by the cross-relaxation constant, $\sigma_{IS}^{NOE}$ which can be defined [10]:

$$\frac{\sigma_{IS}^{NOE}\gamma_S}{\rho_I\gamma_I} = \eta_{IS}$$

where $\gamma_I$ and $\gamma_S$ are the gyromagnetic ratios of $^1$H and the heteroatom (usually $^{13}$C or $^{15}$N) respectively, and $\rho_I$ is the spin-lattice relaxation rate constant ($R_1$). $\eta_{IS}$ is called the NOE

enhancement, and is related to the correlation time, $\tau_c$ by the following equation adapted from Cavanagh [10]:

$$\eta_{IS} = \frac{\frac{6}{1+(\omega_I+\omega_S)^2\tau_c^2} - \frac{1}{1+(\omega_I-\omega_S)^2\tau_c^2}}{\frac{1}{1+(\omega_I-\omega_S)^2\tau_c^2} + \frac{3}{1+\omega_I^2\tau_c^2} + \frac{6}{1+(\omega_I+\omega_S)^2\tau_c^2}}$$

where ω is the larmor frequency for the I and S spins.

An additional experiment includes obtaining relaxation data at different magnetic field strengths. This may help us understand how OEP16 moves within a SDS micelle as well as aid in identifying which residues are undergoing chemical exchange.

## 7. Concluding Remarks

The methods for expressing, purifying, characterizing, and performing NMR spectroscopy on the transmembrane protein OEP16 provides a foundation for the structure determination of membrane proteins by NMR. The data gathered in this study represents a significant bulk of the experimental methods required for structure determination by NMR spectroscopy with only spatial constraints still needed for full structure determination, as described in section 6.1. The confidence of the nearly complete backbone assignment is in large due to the corroborating TALOS calculations, relaxation data, and titration data. Additionally many of these results agree with structure predictions based on hydrophobicity plots [24] that suggests that OEP16 forms a four transmembrane helix bundle. This work has presented for the first time *experimental* evidence that indicates where these helices are located in the primary structure of OEP16. Additionally, arginine titration experiments provide insights into the residues of OEP16 that are involved in binding of the substrate. $D_2O$ titration provides additional convincing evidence that OEP16 is tertiarily folded in SDS micelles. The methods described in this study can also be used in future membrane protein structure investigations.

References

1. Williams, D. A. & Lemke, T. L. (2002) *Foye's principles of medicinal chemistry*, 5th edn, Lippincott Williams & Wilkins, Philadelphia.

2. Nugent, T. & Jones, D. T. Membrane protein structural bioinformatics, *Journal of Structural Biology*.

3. Berg, J. M., Tymoczko, J. L. & Stryer, L. (2002) *Biochemistry*, 5th edn, W.H. Freeman, New York.

4. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res.* **28**, 235-242.

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

6. Wiener, M. C. (2004) A pedestrian guide to membrane protein crystallization, *Methods.* **34**, 364-372.

7. Rhodes, G. (2006) *Crystallography made crystal clear : a guide for users of macromolecular models*, 3rd edn, Elsevier/Academic Press, Amsterdam ; Boston.

8. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000) Potential for biomolecular imaging with femtosecond X-ray pulses, *Nature.* **406**, 752-757.

9. Levitt, M. H. (2008) *Spin dynamics : basics of nuclear magnetic resonance*, 2nd edn, John Wiley & Sons, Chichester, England ; Hoboken, NJ.

10. Cavanagh, J. (2007) *Protein NMR spectroscopy : principles and practice*, 2nd edn, Academic Press, Amsterdam ; Boston.

11. Kovacs, F. A., Fowler, D. J., Gallagher, G. J. & Thompson, L. K. (2007) A practical guide for solid-state NMR distance measurements in proteins, *Concept Magn Reson A.* **30A**, 21-39.

12. Van Horn, W. D., Kim, H. J., Ellis, C. D., Hadziselimovic, A., Sulistijo, E. S., Karra, M. D., Tian, C. L., Sonnichsen, F. D. & Sanders, C. R. (2009) Solution Nuclear Magnetic Resonance Structure of Membrane-Integral Diacylglycerol Kinase, *Science.* **324**, 1726-1729.

13. Russell, D. J., Hadden, C. E., Martin, C. E., Gibson, A. A., Zens, A. P. & Carolan, J. L. (2000) A comparison of inverse-detected heteronuclear NMR performance: Conventional vs cryogenic microprobe performance, *J Nat Prod.* **63**, 1047-1049.

14. Barnes, A. B., De Paepe, G., van der Wel, P. C. A., Hu, K. N., Joo, C. G., Bajaj, V. S., Mak-Jurkauskas, M. L., Sirigiri, J. R., Herzfeld, J., Temkin, R. J. & Griffin, R. G. (2008) High-field dynamic nuclear polarization for solid and solution biological NMR, *Appl Magn Reson.* **34**, 237-263.

15. Pohlmeyer, K., Soll, J., Steinkamp, T., Hinnah, S. & Wagner, R. (1997) Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane, *P Natl Acad Sci USA.* **94**, 9504-9509.

16. Linke, D., Frank, J., Holzwarth, J. F., Soll, J., Boettcher, C. & Fromme, P. (2000) In vitro reconstitution and biophysical characterization of OEP16, an outer envelope pore protein of pea chloroplasts, *Biochemistry-Us.* **39**, 11050-11056.

17. Moreira, D., Le Guyader, H. & Philippe, H. (2000) The origin of red algae and the evolution of chloroplasts, *Nature.* **405**, 69-72.

18. Allen, J. F., de Paula, W. B. M., Puthiyaveetil, S. & Nield, J. (2011) A structural phylogenetic map for chloroplast photosynthesis, *Trends Plant Sci.* **16**, 645-655.

19. Chloroplast. (2012) *Encyclopaedia Britannica. Encyclopaedia Britannica Online Academic Edition.* Encyclopaedia Britannica Inc. <http://www.britannica.com.ezproxy1.lib.asu.edu/EBchecked/topic/113761/chloroplast>.

20. Samol, I., Buhr, F., Springer, A., Pollmann, S., Lahroussi, A., Rossig, C., von Wettstein, D., Reinbothe, C. & Reinbothe, S. (2011) Implication of the oep16-1 Mutation in a flu-Independent, Singlet Oxygen-Regulated Cell Death Pathway in Arabidopsis thaliana, *Plant Cell Physiol.* **52**, 84-95.

21. Samol, I., Rossig, C., Buhr, F., Springer, A., Pollmann, S., Lahroussi, A., von Wettstein, D., Reinbothe, C. & Reinbothe, S. (2011) The Outer Chloroplast Envelope Protein OEP16-1 for Plastid Import of NADPH:Protochlorophyllide Oxidoreductase A in Arabidopsis thaliana, *Plant Cell Physiol.* **52**, 96-111.

22. Reinbothe, S., Quigley, F., Gray, J., Schemenewitz, A. & Reinbothe, C. (2004) Identification of plastid envelope proteins required for import of protochlorophyllide oxidoreductase A into the chloroplast of barley, *P Natl Acad Sci USA.* **101**, 2197-2202.

23. Reinbothe, S., Quigley, F., Springer, A., Schemenewitz, A. & Reinbothe, C. (2004) The outer plastid envelope protein Oep16: Role as precursor translocase in import of protochlorophyllide oxidoreductase A, *P Natl Acad Sci USA.* **101**, 2203-2208.

24. Linke, D., Frank, J., Pope, M. S., Soll, J., Ilkavets, I., Fromme, P., Burstein, E. A., Reshetnyak, Y. K. & Emelyanenko, V. I. (2004) Folding kinetics and structure of OEP16, *Biophys J.* **86**, 1479-1487.

25. Ni, D., Zook, J., Klewer, D. A., Nieman, R. A., Soll, J. & Fromme, P. (2011) Isolation, folding and structural investigations of the amino acid transporter OEP16, *Protein Expres Purif.* **80**, 157-168.

26. Steinkamp, T., Hill, H., Hinnah, S. C., Wagner, R., Rohl, T., Pohlmeyer, K. & Soll, J. (2000) Identification of the pore-forming region of the outer chloroplast envelope protein OEP16, *J Biol Chem.* **275**, 11758-11764.

27. Drea, S. C., Lao, N. T., Wolfe, K. H. & Kavanagh, T. A. (2006) Gene duplication, exon gain and neofunctionalization of OEP16-related genes in land plants, *Plant J.* **46**, 723-735.

28. Bruice, P. Y. (2001) *Organic chemistry*, 3rd edn, Prentice Hall, Upper Saddle River, N.J.

29. Kaur, H. (2009) *Spectroscopy*, Global Media, Meerut, IND.

30. Walker, J. M. (2005) *The proteomics protocols handbook*, Humana Press, Totowa, N.J.

31. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nat Protoc.* **1**, 2876-2890.

32. Edelhoch, H. (1967) Spectroscopic Determination of Tryptophan and Tyrosine in Proteins, *Biochemistry-Us.* **6**, 1948-&.

33. Mihalyi, E. (1968) Numerical Values of Absorbances of Aromatic Amino Acids in Acid Neutral and Alkaline Solutions, *J Chem Eng Data.* **13**, 179-&.

34. Wallace, B. A., Janes, R. W. & ebrary Inc. (2009) Modern techniques for circular dichroism and synchrotron radiation circular dichroism spectroscopy in *Advances in biomedical spectroscopy 1* pp. xii, 231 p., IOS Press,, Amsterdam.

35. Sreerama, N. & Woody, R. W. (2004) Computation and analysis of protein circular dichroism spectra, *Methods Enzymol.* **383**, 318-351.

36. Telfer, S. G., McLean, T. M. & Waterland, M. R. (2011) Exciton coupling in coordination compounds, *Dalton T.* **40**, 3097-3108.

37. Lees, J. G., Miles, A. J., Wien, F. & Wallace, B. A. (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space, *Bioinformatics.* **22**, 1955-1962.

38. Abdul-Gader, A., Miles, A. J. & Wallace, B. A. (2011) A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy, *Bioinformatics.* **27**, 1630-1636.

39. Podzimek, S. (2011) *Light scattering, size exclusion chromatography, and asymmetric flow field flow fractionation : powerful tools for the characterization of polymers, proteins and nanoparticles*, Wiley, Hoboken, N.J.

40. Mayer, M. G. & Jensen, J. H. D. (1955) *Elementary theory of nuclear shell structure*, Wiley, New York,.

41. Higman, V. A. (2012) Protein NMR - A Practical Guide in

42. Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E.,

Wenger, R. K., Yao, H. Y. & Markley, J. L. (2008) BioMagResBank, *Nucleic Acids Res.* **36**, D402-D408.

43. Cornilescu, G., Delaglio, F. & Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J Biomol Nmr.* **13**, 289-302.

44. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. (2009) TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J Biomol Nmr.* **44**, 213-223.

45. Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G. H., Eletsky, A., Wu, Y. B., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperski, T., Montelione, G. T., Baker, D. & Bax, A. (2008) Consistent blind protein structure generation from NMR chemical shift data, *P Natl Acad Sci USA.* **105**, 4685-4690.

46. Shen, Y., Vernon, R., Baker, D. & Bax, A. (2009) De novo protein structure generation from incomplete chemical shift assignments, *J Biomol Nmr.* **43**, 63-78.

47. Abragam, A. (1961) *The principles of nuclear magnetism*, Clarendon Press, Oxford,.

48. Lipari, G. & Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity, *J Am Chem Soc.* **104**, 4546-4559.

49. Kay, L. E., Torchia, D. A. & Bax, A. (1989) Backbone Dynamics of Proteins as Studied by N-15 Inverse Detected Heteronuclear Nmr-Spectroscopy - Application to Staphylococcal Nuclease, *Biochemistry-Us.* **28**, 8972-8979.

50. Cuatreca.P (1970) Protein Purification by Affinity Chromatography - Derivatizations Ofagarose and Polyacrylamide Beads, *J Biol Chem.* **245**, 3059-&.

51. Crowe, J., Dobeli, H., Gentz, R., Hochuli, E., Stiiber, D. & Henco, K. (1994) 6xffis-Ni-NTA Chromatography as a Superior Technique in Recombinant Protein Expression/Purification in pp. 371-387.

52. Walker, J. M. (1984) *Proteins*, Humana Press, Clifton, N.J.
53. Wu, Q., Merchant, F. A. & Castleman, K. R. (2008) *Microscope image processing*, Elsevier/Academic Press, Amsterdam ; Boston.

54. Devendittis, E., Palumbo, G., Parlato, G. & Bocchini, V. (1981) A FLUORIMETRIC METHOD FOR THE ESTIMATION OF THE CRITICAL MICELLE CONCENTRATION OF SURFACTANTS, *Analytical Biochemistry.* **115**, 278-286.

55. Walter, A., Suchy, S. E. & Vinson, P. K. (1990) SOLUBILITY PROPERTIES OF THE ALKYLMETHYLGLUCAMIDE SURFACTANTS, *Biochim Biophys Acta.* **1029**, 67-74.

56. Lemaire, M., Kwee, S., Andersen, J. P. & Moller, J. V. (1983) MODE OF INTERACTION OF POLYOXYETHYLENEGLYCOL DETERGENTS WITH MEMBRANE-PROTEINS, *Eur J Biochem.* **129**, 525-532.

57. Lemaire, M., Viel, A. & Moller, J. V. (1989) SIZE EXCLUSION CHROMATOGRAPHY AND UNIVERSAL CALIBRATION OF GEL COLUMNS, *Analytical Biochemistry.* **177**, 50-56.

58. Turro, N. J. & Yekta, A. (1978) Luminescent Probes for Detergent Solutions - Simple Procedure for Determination of Mean Aggregation Number of Micelles, *J Am Chem Soc.* **100**, 5951-5952.

59. Vanaken, T., Foxallvanaken, S., Castleman, S. & Fergusonmiller, S. (1986) ALKYL GLYCOSIDE DETERGENTS - SYNTHESIS AND APPLICATIONS TO THE STUDY OF MEMBRANE-PROTEINS, *Methods in Enzymology.* **125**, 27-35.

60. Vulliezlenormand, B. & Eisele, J. L. (1993) Determination of Detergent Critical Micellar Concentration by Solubilization of a Colored Dye, *Analytical Biochemistry.* **208**, 241-243.

61. le Maire, M., Champeil, P. & Moller, J. V. (2000) Interaction of membrane proteins and lipids with solubilizing detergents, *Bba-Biomembranes.* **1508**, 86-111.

62. Affymetrix (2012) Cymal-4, Anagrade <http://www.affymetrix.com/estore/browse/brand/anatrace/product.jsp?navMode=34000&productId=131598&navAction=jump&aId=productsNav#1_1>

63. Chattopadhyay, A. & London, E. (1984) FLUORIMETRIC DETERMINATION OF CRITICAL MICELLE CONCENTRATION AVOIDING INTERFERENCE FROM DETERGENT CHARGE, *Analytical Biochemistry.* **139**, 408-412.

64. Lorber, B., Bishop, J. B. & Delucas, L. J. (1990) PURIFICATION OF OCTYL BETA-D-GLUCOPYRANOSIDE AND REESTIMATION OF ITS MICELLAR SIZE, *Biochim Biophys Acta.* **1023**, 254-265.

65. Hjelmeland, L. M., Nebert, D. W. & Osborne, J. C. (1983) SULFOBETAINE DERIVATIVES OF BILE-ACIDS - NON-DENATURING SURFACTANTS FOR MEMBRANE BIOCHEMISTRY, *Analytical Biochemistry.* **130**, 72-82.

66. Womack, M. D., Kendall, D. A. & Macdonald, R. C. (1983) DETERGENT EFFECTS ON ENZYME-ACTIVITY AND SOLUBILIZATION OF LIPID BILAYER-MEMBRANES, *Biochim Biophys Acta.* **733**, 210-215.

67. Affymetrix (2012) Fos-Choline-12, Anagrade in

68. Vergis, J. M., Purdy, M. D. & Wiener, M. C. (2010) A high-throughput differential filtration assay to screen and select detergents for membrane proteins, *Analytical Biochemistry.* **407**, 1-11.

69. Schägger, H. (2006) Tricine-SDS-PAGE. **1**.

70. Compton, L. A. & Johnson, W. C. (1986) ANALYSIS OF PROTEIN CIRCULAR-DICHROISM SPECTRA FOR SECONDARY STRUCTURE USING A SIMPLE MATRIX MULTIPLICATION, *Analytical Biochemistry.* **155**, 155-167.

71. Hennessey, J. P. & Johnson, W. C. (1981) INFORMATION-CONTENT IN THE CIRCULAR-DICHROISM OF PROTEINS, *Biochemistry-Us.* **20**, 1085-1094.

72. Johnson, W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures, *Proteins-Structure Function and Genetics.* **35**, 307-312.

73. Manavalan, P. & Johnson, W. C. (1987) VARIABLE SELECTION METHOD IMPROVES THE PREDICTION OF PROTEIN SECONDARY STRUCTURE FROM CIRCULAR-DICHROISM SPECTRA, *Analytical Biochemistry.* **167**, 76-85.

74. Sreerama, N., Venyaminov, S. Y. & Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis, *Analytical Biochemistry.* **287**, 243-251.

75. Sreerama, N. & Woody, R. W. (2004) Computation and analysis of protein circular dichroism spectra in *Numerical Computer Methods, Pt D* (Brand, L. & Johnson, M. L., eds) pp. 318-351.

76. Pancoska, P. & Keiderling, T. A. (1991) SYSTEMATIC COMPARISON OF STATISTICAL-ANALYSES OF ELECTRONIC AND VIBRATIONAL CIRCULAR-DICHROISM FOR SECONDARY STRUCTURE PREDICTION OF SELECTED PROTEINS, *Biochemistry-Us.* **30**, 6885-6895.

77. Sreerama, N. & Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Analytical Biochemistry.* **287**, 252-260.

78. Chen, Y. H. & Yang, J. T. (1971) NEW APPROACH TO CALCULATION OF SECONDARY STRUCTURES OF GLOBULAR PROTEINS BY OPTICAL ROTATORY DISPERSION AND CIRCULAR DICHROISM, *Biochem Bioph Res Co.* **44**, 1285-&.

79. Chen, Y. H., Yang, J. T. & Chau, K. H. (1974) DETERMINATION OF HELIX AND BETA-FORM OF PROTEINS IN AQUEOUS-SOLUTION BY CIRCULAR-DICHROISM, *Biochemistry-Us.* **13**, 3350-3359.

80. Yang, J. T., Wu, C. S. C. & Martinez, H. M. (1986) CALCULATION OF PROTEIN CONFORMATION FROM CIRCULAR-DICHROISM, *Methods in Enzymology.* **130**, 208-269.

81. Sreerama, N., Venyaminov, S. Y. & Woody, R. W. (1999) Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy, *Protein Sci.* **8**, 370-380.

82. Sreerama, N., Venyaminov, S. Y. & Woody, R. W. (2001) Analysis of protein circular dichroism spectra based on the tertiary structure classification, *Analytical Biochemistry*. **299**, 271-274.

83. Sreerama, N. & Woody, R. W. (1993) A SELF-CONSISTENT METHOD FOR THE ANALYSIS OF PROTEIN SECONDARY STRUCTURE FROM CIRCULAR-DICHROISM, *Analytical Biochemistry.* **209**, 32-44.

84. Sreerama, N. & Woody, R. W. (1994) PROTEIN SECONDARY STRUCTURE FROM CIRCULAR-DICHROISM SPECTROSCOPY - COMBINING VARIABLE SELECTION PRINCIPLE AND CLUSTER-ANALYSIS WITH NEURAL-NETWORK, RIDGE-REGRESSION AND SELF-CONSISTENT METHODS, *J Mol Biol.* **242**, 497-507.

85. Sreerama, N. & Woody, R. W. (1994) POLY(PRO)II HELICES IN GLOBULAR-PROTEINS - IDENTIFICATION AND CIRCULAR DICHROIC ANALYSIS, *Biochemistry-Us.* **33**, 10022-10025.

86. Sreerama, N. & Woody, R. W. (2004) On the analysis of membrane protein circular dichroism spectra, *Protein Sci.* **13**, 100-112.

87. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989) Overcoming the Overlap Problem in the Assignment of H-1-Nmr Spectra of Larger Proteins by Use of 3-Dimensional Heteronuclear H-1-N-15 Hartmann-Hahn Multiple Quantum Coherence and Nuclear Overhauser Multiple Quantum Coherence Spectroscopy - Application to Interleukin-1-Beta, *Biochemistry-Us.* **28**, 6150-6156.

88. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989) 3-Dimensional Heteronuclear Nmr of N-15-Labeled Proteins, *J Am Chem Soc.* **111**, 1515-1517.

89. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990) 3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins, *J Magn Reson.* **89**, 496-514.

90. Clubb, R. T., Thanabal, V. & Wagner, G. (1992) A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue H-1(N), N-15, and C-13(') Chemical-Shifts in N-15-C-13-Labeled Proteins, *J Magn Reson.* **97**, 213-217.

91. Grzesiek, S. & Bax, A. (1992) Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr, *J Am Chem Soc.* **114**, 6291-6293.

92. Grzesiek, S. & Bax, A. (1992) Improved 3d Triple-Resonance Nmr Techniques Applied to a 31-Kda Protein, *J Magn Reson.* **96**, 432-440.

93. Wittekind, M. & Mueller, L. (1993) Hncacb, a High-Sensitivity 3d Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins, *J Magn Reson Ser B.* **101**, 201-205.

94. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995) Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes, *J Biomol Nmr.* **6**, 277-293.

95. Goddard, T. D. a. D. G. K. (2008) SPARKY 3 in, University of California, San Francisco,

96. Sreerama, N. & Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Anal Biochem.* **287**, 252-60.

97. Greenfield, N. J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, *Nat Protoc.* **1**, 2527-2535.

98. Kefala, G., Ahn, C., Krupa, M., Esquivies, L., Maslennikov, I., Kwiatkowski, W. & Choe, S. (2010) Structures of the OmpF porin crystallized in the presence of foscholine-12, *Protein Sci.* **19**, 1117-1125.

99. van Roon, H., van Breemen, J. F. L., de Weerd, F. L., Dekker, J. P. & Boekema, E. J. (2000) Solubilization of green plant thylakoid membranes with n-dodecyl-alpha,D-maltoside. Implications for the structural organization of the Photosystem II, Photosystem I, ATP synthase and cytochrome b(6)f complexes, *Photosynth Res.* **64**, 155-166.

100. Berjanskii, M. V. & Wishart, D. S. (2005) A simple method to predict protein flexibility using secondary chemical shifts, *J Am Chem Soc.* **127**, 14970-14971.

101. Cheng, N. S. (2008) Formula for the viscosity of a glycerol-water mixture, *Industrial & Engineering Chemistry Research.* **47**, 3285-3288.

102. Venable, R. M. & Pastor, R. W. (1988) FRICTIONAL MODELS FOR STOCHASTIC SIMULATIONS OF PROTEINS, *Biopolymers.* **27**, 1001-1014.

103. Squire, P. G. & Himmel, M. E. (1979) HYDRODYNAMICS AND PROTEIN HYDRATION, *Arch Biochem Biophys.* **196**, 165-177.

104. Gekko, K. & Noguchi, H. (1979) COMPRESSIBILITY OF GLOBULAR-PROTEINS IN WATER AT 25-DEGREES-C, *Journal of Physical Chemistry.* **83**, 2706-2714.

105. Schulein, K., Schmid, K. & Benzl, R. (1991) THE SUGAR-SPECIFIC OUTER-MEMBRANE CHANNEL SCRY CONTAINS FUNCTIONAL-CHARACTERISTICS OF GENERAL DIFFUSION PORES AND SUBSTRATE-SPECIFIC PORINS, *Mol Microbiol.* **5**, 2233-2241.

106. Vonheijne, G. & Gavel, Y. (1988) TOPOGENIC SIGNALS IN INTEGRAL MEMBRANE-PROTEINS, *Eur J Biochem.* **174**, 671-678.

107. Arechaga, I., Jonathan, P., Butler, G. & Walker, J. E. (2002) Self-assembly of ATP synthase subunit c rings, *Febs Lett.* **515**, 189-193.

108. Seinfeld, J. H. & Pandis, S. N. Atmospheric Chemistry and Physics - From Air Pollution to Climate Change (2nd Edition) in, John Wiley & Sons,

109. Nesmelova, I. V., Idiyatullin, D. & Mayo, K. H. (2004) Measuring protein self-diffusion in protein-protein mixtures using a pulsed gradient spin-echo technique with WATERGATE and isotope filtering, *J Magn Reson.* **166**, 129-133.

110. Metzler, W. J., Wittekind, M., Goldfarb, V., Mueller, L. & Farmer, B. T. (1996) Incorporation of H-1/C-13/N-15-{Ile, Leu, Val} into a perdeuterated, N-15-labeled protein: Potential in structure determination of large proteins by NMR, *J Am Chem Soc.* **118**, 6800-6801.

111. Chou, J. J., Gaemers, S., Howder, B., Louis, J. M. & Bax, A. (2001) A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles, *J Biomol Nmr.* **21**, 377-382.

# APPENDIX A

## HPLC METHODS FOR PURIFICATION OF OEP16$^{HIS}$

| Content | Urea+SDS | SDS | Injectant | Eluant | | |
|---|---|---|---|---|---|---|
| Time (min) | Buffer %A | Buffer %B | Buffer %C | Buffer %D | Flow Rate (mL/min) | Valve |
| -10 | 100 | 0 | 0 | 0 | 1 | To Column |
| 0 | 100 | 0 | 0 | 0 | 1 | To Column |
| 0 | 0 | 0 | 100 | 0 | 1 | To Column |
| 35 | 100 | 0 | 0 | 0 | 1 | To Column |
| 60 | 100 | 0 | 0 | 0 | 1 | To Column |
| 60.5 | 66 | 34 | 0 | 0 | 1 | To Column |
| 86.5 | 66 | 34 | 0 | 0 | 1 | To Column |
| 87 | 50 | 50 | 0 | 0 | 1 | To Column |
| 112 | 50 | 50 | 0 | 0 | 1 | To Column |
| 112.5 | 34 | 66 | 0 | 0 | 1 | To Column |
| 137.5 | 34 | 66 | 0 | 0 | 1 | To Column |
| 138 | 16 | 84 | 0 | 0 | 1 | To Column |
| 163 | 16 | 84 | 0 | 0 | 1 | To Column |
| 163.5 | 0 | 100 | 0 | 0 | 1 | To Column |
| 188.5 | 0 | 100 | 0 | 0 | 1 | To Column |
| 188.5 | 0 | 0 | 0 | 100 | 5 | To Waste |
| 189.8 | 0 | 0 | 0 | 100 | 5 | To Waste |
| 189.8 | 0 | 96.7 | 0 | 3.3 | 1 | To Column |
| 209.8 | 0 | 96.7 | 0 | 3.3 | 1 | To Column |
| 209.8 | 0 | 0 | 0 | 100 | 1 | To Column |
| 239.8 | 0 | 0 | 0 | 100 | 1 | To Column |
| 239.8 | 0 | 0 | 0 | 100 | 0 | To Column |

Table A.1. The original HPLC method for refolding OEP16 described in section 3.2.1. Buffers A, C, and D are controlled through a system of valves connected through pump A, Buffer B is connected to pump B and provides a system for establishing the step gradient needed.

| Content | Urea+SDS | Mild Det | Injectant | Elutant | GuHCl | Urea | | |
|---|---|---|---|---|---|---|---|---|
| Time (min) | Buffer %A | Buffer %B | Buffer %C | Buffer %D | Buffer %E | Buffer %F | Flow Rate (mL/min) | Valve |
| -10 | 0 | 0 | 0 | 0 | 100 | 0 | 1 | Column |
| 0 | 0 | 0 | 0 | 0 | 100 | 0 | 1 | Column |
| 0 | 0 | 0 | 100 | 0 | 0 | 0 | 1 | Column |
| 10 | 0 | 0 | 100 | 0 | 0 | 0 | 1 | Column |
| 10 | 0 | 0 | 0 | 0 | 100 | 0 | 1 | Column |
| 30 | 0 | 0 | 0 | 0 | 100 | 0 | 1 | Column |
| 30 | 0 | 0 | 0 | 0 | 0 | 100 | 1 | Column |
| 60 | 0 | 0 | 0 | 0 | 0 | 100 | 1 | Column |
| 60 | 100 | 0 | 0 | 0 | 0 | 0 | 1 | Column |
| 80 | 100 | 0 | 0 | 0 | 0 | 0 | 1 | Column |
| 80 | 90 | 10 | 0 | 0 | 0 | 0 | 1 | Column |
| 90 | 90 | 10 | 0 | 0 | 0 | 0 | 1 | Column |
| 90 | 80 | 20 | 0 | 0 | 0 | 0 | 1 | Column |
| 100 | 80 | 20 | 0 | 0 | 0 | 0 | 1 | Column |
| 100 | 70 | 30 | 0 | 0 | 0 | 0 | 1 | Column |
| 110 | 70 | 30 | 0 | 0 | 0 | 0 | 1 | Column |
| 110 | 60 | 40 | 0 | 0 | 0 | 0 | 1 | Column |
| 120 | 60 | 40 | 0 | 0 | 0 | 0 | 1 | Column |
| 120 | 50 | 50 | 0 | 0 | 0 | 0 | 1 | Column |
| 130 | 50 | 50 | 0 | 0 | 0 | 0 | 1 | Column |
| 130 | 40 | 60 | 0 | 0 | 0 | 0 | 1 | Column |
| 140 | 40 | 60 | 0 | 0 | 0 | 0 | 1 | Column |
| 140 | 30 | 70 | 0 | 0 | 0 | 0 | 1 | Column |
| 150 | 30 | 70 | 0 | 0 | 0 | 0 | 1 | Column |
| 150 | 20 | 80 | 0 | 0 | 0 | 0 | 1 | Column |
| 160 | 20 | 80 | 0 | 0 | 0 | 0 | 1 | Column |
| 160 | 10 | 90 | 0 | 0 | 0 | 0 | 1 | Column |
| 170 | 10 | 90 | 0 | 0 | 0 | 0 | 1 | Column |
| 170 | 0 | 100 | 0 | 0 | 0 | 0 | 1 | Column |
| 200 | 0 | 100 | 0 | 0 | 0 | 0 | 1 | Column |
| 200 | 0 | 0 | 0 | 100 | 0 | 0 | 5 | Waste |
| 200.4 | 0 | 0 | 0 | 100 | 0 | 0 | 5 | Waste |
| 200.4 | 0 | 0 | 0 | 100 | 0 | 0 | 1 | Column |
| 235.4 | 0 | 0 | 0 | 100 | 0 | 0 | 1 | Column |
| 235.4 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | Column |

Table A.2. The new HPLC method for purifying and refolding OEP16 described in section 3.2.2. Buffers A, C, D, E, and F are controlled through a system of valves connected through pump A, Buffer B is connected to pump B and provides a system for establishing the step gradient needed.

# APPENDIX B

## VISCOSITY OF WATER-GLYCEROL SOLUTIONS

Described below is the parameterization details for calculating the viscosity of a solution ($\eta_{gw}$) of glycerol and water as described by Chen [101]:

$$\eta_{gw} = \eta_g e^{A*\alpha}$$

$$\eta_g = 12100 e^{\frac{(-1233+T)T}{9900+70T}}$$

$$\eta_w = 1.790 e^{\frac{(-1230-T)T}{36100+360T}}$$

$$A = log\left(\frac{\eta_w}{\eta_g}\right)$$

$$\alpha = 1 - C_m + \frac{abC_m(1-C_m)}{aC_m + b(1-C_m)}$$

$$a = 0.705 - 0.0017T$$

$$b = (4.9 + 0.036T)a^{2.5}$$

$C_m$ is the glycerol concentration in mass, and T is the temperature in centigrade. This parameterization was designed to work for concentrations of glycerol ranging between 0% to 100% and temperatures between 0°C and 100°C [101].

# BIOGRAPHICAL SKETCH

James Duncan Zook was born in Champaign-Urbana, Illinois, on November 30, 1983. He attended the Stevens Point Area Catholic Schools and graduated from Pacelli High School in Stevens Point, Wisconsin, in May of 2002. James attended the University of Wisconsin-Eau Claire, majoring in Biochemistry and Molecular Biology and graduated Cum Laude in the spring semester of 2007. He entered the Graduate College at Arizona State University to pursue a doctorate in Biochemistry under the tutelage of Dr. Petra Fromme, studying membrane protein structure with emphasis on protein biophysical characterization and structure determination through solution state NMR.