EXPRESSION AND STRUCTURAL CHARACTERIZATION OF THE P8 MAJOR COAT PROTEIN FROM B5 FILAMENTOUS BACTERIOPHAGE

by

Htet Ei Bo

B.Sc., The University of British Columbia, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2011

Abstract

This thesis is based on work done on the expression, purification and structural characterization of the major coat protein of B5 bacteriophage. The major coat proteins of bacteriophages have long been utilized to understand membrane proteins and membrane-associated assembly. It is the special feature of the major coat protein to exist in different environments that holds the key to its involvement in phage assembly. The structure of p8 in the different environments, especially in the host membrane, has to be fully understood before the mystery of phage assembly can be solved. The major coat protein of B5, p8, has been chosen in this study because B5 infects Gram positive bacteria and the structure of p8 in an appropriate model membrane can better represent its native structure in the host membrane.

In Chapter 1, I introduce background information on filamentous phage, and the debate of major coat protein structure. The different structures that already exist for the major coat protein in virion, in host membranes, and during phage assembly are discussed.

In the next chapter, I present the steps required to obtain pure p8 using a heterologous bacterial expression system. The optimizations and considerations needed to express and purify p8 are discussed thoroughly. The considerations taken for p8 expression can essentially be applied to other membrane protein expression. In the same chapter, an I32C mutant of p8 is also designed, expressed and successfully purified. The technique used to introduce the single substitution mutation to p8 can be applied to other protein mutation experiments.

In the subsequent chapter, p8 structure is studied using circular dichroism (CD), nuclear magnetic resonance (NMR) and site directed labeling with a 6-bromoacetyl-2-dimethylaminonaphthalene (BADAN) fluorescence probe. The results from CD show that p8 has high alpha helicity when reconstituted into lipid compositions that represent the Gram positive membrane. Preliminary NMR experiments have been performed and conditions to obtain optimal NMR spectra have been explored. BADAN fluorescence labeling experiments have been trialed and have been shown to successfully indicate the local environment of residue 32 to which BADAN is attached. Finally, possible future work is discussed.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADEAS, N-(iodoacetylaminoethyl)-5-naphthylamine-1-sulfonic acid
BADAN, 6-bromoacetyl-2-dimethylaminonaphthalene
Cardio, cardiolipin
CD, circular dichroism
COSY, correlation spectroscopy
Da, Daltons
DNA, deoxyribonucleic acid
DPC, dodecylphosphocholine
HSQC, heteronuclear single quantum coherence
kDa, kilo daltons
LB, Luria-Bertani
MALDI-TOF, matrix assisted laser desorption ionization time of flight mass spectrometry
MBP, maltose binding protein
mRNA, messenger ribonucleic acid
NAPS, Nucleic Acid Protein Service Unit
Nickel-NTA, nickel-nitrilotriacetic acid
NMR, nuclear magnetic resonance
NOESY, nuclear overhauser effect spectroscopy
nt/s, nucleotide/protein
PCR, polymerase chain reaction
PI, pre induced
POPC, palmitoyl oleoyl phosphatidyl choline
POPG, palmitoyl oleoyl phosphatidylglycerol
RF, replicative form
r.p.m., revolutions per minute
SCAM, substituted cysteine accessibility method
SDS, sodium dodecyl sulphate
SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis
TOCSY, total correlation spectroscopy
Tris, tris(hydroxymethyl) aminomethane
UV, ultraviolet

Acknowledgements

I would like to give special thanks to Dr. S. K. Straus for giving me the opportunity to benefit from her vast knowledge and for going above and beyond her call of duty as a supervisor. Also, many thanks to Andrew Tait for giving me answers to my many questions and for giving me the confidence to think critically. I could not have done as much as I have without him. Guidance from Andrew and Suzana has definitely molded me into the scientist I am today. I need to acknowledge the past and present members of Straus Lab, all of whom contributed to and helped steer my path.

I would like to acknowledge the faculty and staff at UBC for their help. Jessie and Elena at Bioservices deserve acknowledgment for providing me with a nurturing environment to work in. I would also need to thank Mark Okon for his NMR expertise and his valuable time. The staff at Mass spectrometry, especially Marshall, deserves thanks for their input.

I would also like to thank my amazing parents for their unreserved emotional support and for wanting only the best for me. My loving fiancé deserves special acknowledgment for his support and for always looking out for me. I really could not have done it without his love and kindness.

I also need to thank my close friends for the laughs and tears we shared.

Dedication

For my mother and father,

My dad, for being the most intelligent and most driven man I know.

My mom, for being the most loving, most selfless woman I know.

Without them, I would not be where I am today.

Chapter 1: Introduction

1.1 Bacteriophages

Bacteriophages were discovered independently by Twort and Lond, and d'Herelle over 90 years ago [1]. Felix d'Herelle first discovered *Bacteriophagum intestinale* in 1918 [2]. Bacteriophages are viruses that infect bacteria and so were considered in the control of bacterial infections. The development of antibiotics such as penicillin during the 1940s, however, overshadowed the use of bacteriophages [1]. Even so, the benefits of bacteriophage therapy should not be discounted. Since bacteriophages are active against antibiotic resistant bacteria, they offer an alternative to antibiotics and could also be used in people with allergies to antibiotics [1]. Also, bacteriophages are self limiting: they infect and lyse target bacteria and once there are no more hosts to infect, the phage life cycle ends [1]. It was demonstrated that injection of 10^4 phages into mice that were also injected with 10^6 colony forming units of Escherichia Coli (*E. coli*) was enough to protect the mice against *E. coli* infection [1]. This proved that there was *in vivo* replication of the phages. Bacteriophages could be used against diarrhea by injecting a mixture of two phages specific against *E. coli* strains O9 and K30, 99 [1]. Soothill *et al* [3] showed that bacteriophages active against *Pseudomonas aeruginosa* and *Acinetobacter baumanii* could be used to control skin infections.

Although bacteriophage therapy produced promising results in early studies, there were downfalls to its use. For one, there was difficulty in selecting viable phages and the fact that phages have specificity to a narrow range of host strains became a limiting factor in protecting against general infections. A mixture of seven bacteriophages injected into beef to

counteract *Pseudomonas* strains was unsuccessful in controlling bacterial growth. It was found that only 57.2% of 1023 *Pseudomonas* strains were susceptible to the phage mixture [1]. Secondly, phage resistant strains were also developing. Bacteria might modify their receptor molecules which are needed for phage recognition and infection [1]. Also, it is difficult to prevent phage elimination by host defense mechanisms when the phages are administered systemically [1]. Upon the lysis of bacterial cells, there is the release of endotoxins which are released into the blood stream and could have implications for the host [1].

The use of phages in other applications has been explored. The narrow host specificity of a phage has been exploited to identify bacterial strains in bacteriophage typing. Bacteriophages are used as biological tracers of fecal contaminants because a high concentration of phages can be added to the sample and the phages can be detected in low concentrations [1]. Bacteriophages, especially the M13 filamentous phage, have been used in phage display. Here, the deoxyribonucleic acid (DNA) sequence of a ligand of interest is fused to the gene of phage coat proteins p3, p6 or p8 and the ligand is subsequently displayed on the mature phage [1]. A wide array of phage libraries has been constructed. This is partially due to the ability to quickly clone millions of variants of a ligand in a batch [1]. Phage libraries can be used to optimize and identify interactions between ligands and proteins.

According to the International Committee on Taxonomy of Viruses, viruses consist of 6 orders with 87 families, classified according to their morphological properties such as the presence of a tail and the type of genetic material they contain (double stranded DNA, single stranded DNA, double stranded RNA or single stranded RNA). There are also 65 independent families of viruses that are not classified into an order due to their highly varied

morphology [4]. Bacteriophages consist of 13 families and 30 genera. The different types of bacteriophages are illustrated in Figure 1.1. The tailed viruses make up the most diverse order known as the Caudovirale [4]. Of particular interest is the family of filamentous phages or Inoviridae. Filamentous phage B5, the subject of this thesis, is a member of this family.

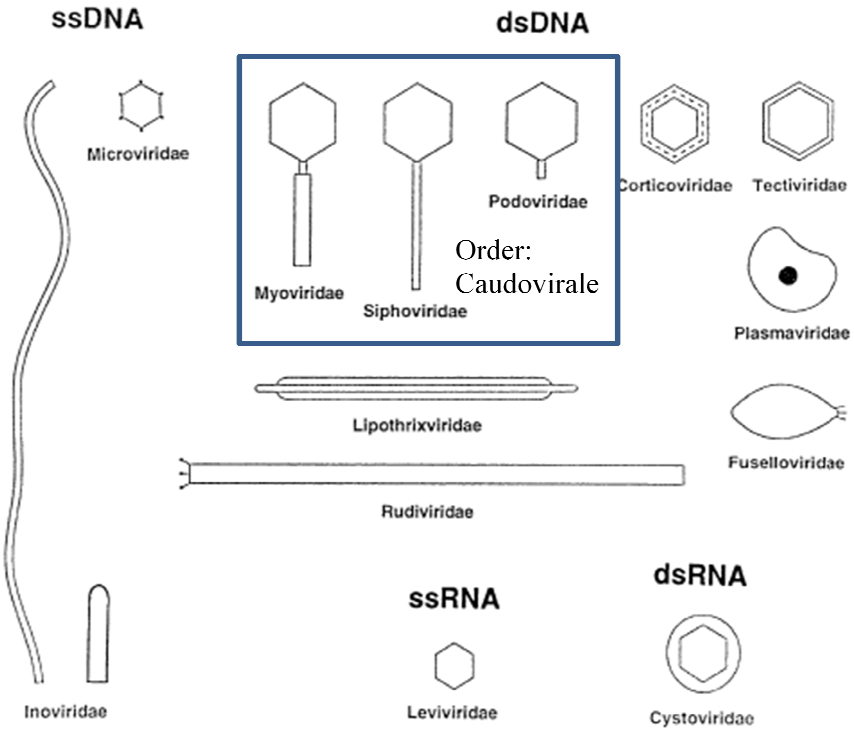

Figure 1.1- Representation of the morphology of common phage families. The family Inoviridae, which are the filamentous phages, is the family of interest in this project. The three families in the order Caudovirale are also identified. Figure is adapted from [5].

1.2 Filamentous phage

Filamentous phage M13, amongst other filamentous phages, was discovered in 1963 by Hofschneider while analyzing different *E. coli* bacterial strains present in waste water samples [6]. Filamentous phages make up the family Inoviridae. This family does not belong to an order but it has two genus classifications: Inovirus and Plectrovirus. The genus Inovirus

contains 36 species, among them the Enterobacteria phages M13, Ike, Pf1 and fd. The genus Plectovirus contains 6 species of Acholeplasma and Spiroplasma phages.

Filamentous phages infect hosts that express sex pili because the sex pili act as receptor sites for phage recognition and infection [7]. Filamentous phages are classified according to their pilus specificity. The Ff group of filamentous phages infects *E. coli* cells which express the fertility factor and so contain a sex pili. M13, fd and f1 bacteriophages are examples of Ff bacteriophages. These Ff bacteriophages have interchangeable gene products [7]. The phage Ike infects *E. Coli* that express N pili. Ike phages are 50% identical genetically to the Ff phages [7]. There are also phages that infect *Pseudomonas aeruginosa* via the P pili, such as the phages Pf1 and Pf3 [7].

Filamentous phages differ from other bacteriophages because they are extruded from the host bacteria continuously and the host is not lysed upon progeny extrusion. The infected host grows at a slower rate but appears to remain unaffected otherwise [6]. Filamentous phages are also the smallest of the bacteriophages. Filamentous phages are visible under an electron microscope and can be described as long flexible filaments varying in length from 720nm (Pf3) to 1.94µm (Pf1). X ray diffraction studies have shown that the fd phage has an outer diameter of 60Å, an inner diameter of 20Å and is about 0.87µm long.

1.3 B5 filamentous phage

Most filamentous phages are isolated from Gram negative bacteria. In 2002, Chopin *et al* [8] reported the discovery of the B5 filamentous phage. This phage infects the Gram positive bacteria *Propionibacterium freudenreichii* and has dimensions of 620nm in length

and 12nm in width. Figure 1.2 shows the electron micrograph of B5 phage from Chopin *et al* [8].

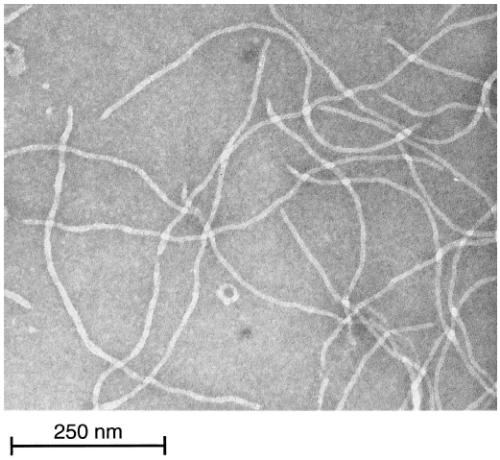

Figure 1.2-Electron micrograph of B5 phage. Figure is adapted from [8].

The B5 phage genome consists of 5806 bases and is organized in a similar way as other filamentous phage genomes. Its genome consists of genes II and V which are needed for DNA synthesis, genes VIII, III, VI, VII, IX which code for phage structural proteins and genes I, XI which code for proteins that aid in phage assembly and extrusion [8]. See the following section for proteins associated with the genes. The phage DNA genome has a total G+C content of 64% and has 10 open reading frames [8]. The gene for the major coat protein of B5 shows homology to the major coat protein gene in PH75 and Pf3 [8]. Figure 1.3 shows the sequence alignment between the three associated coat proteins.

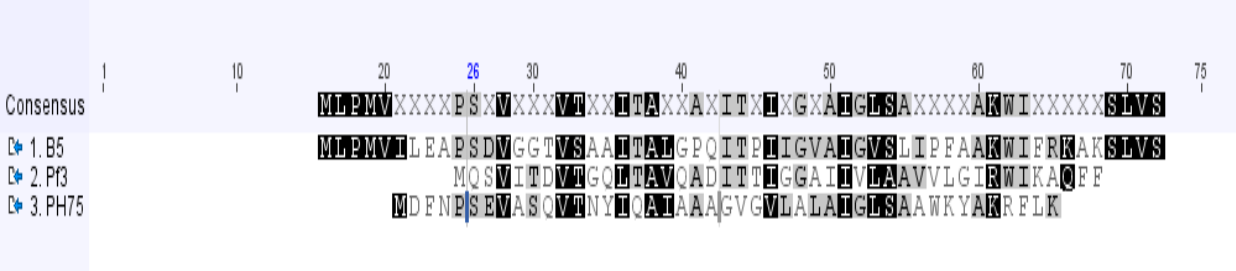

Figure 1.3- Sequence alignment of major coat protein from B5, PH75 and Pf3 phage. Conserved amino acids are in bold. The top sequence is the consensus sequence. Sequence similarity is analyzed using Genious Basic 5.3.4 [9].

The B5 phage differs from the more common Ff phages in a few ways. The major coat protein of Ff phages is synthesized with a signal sequence consisting of 23 amino acids which is then cleaved off before the insertion of the coat protein into the host membrane. The major coat protein of B5 phage, on the other hand, is not synthesized with a signal sequence. Because the B5 phage infects Gram positive bacteria, it lacks p4 which is needed for the formation of a pore in the outer host cell membrane.

1.4 Constituent proteins of filamentous bacteriophages

As alluded to in the previous section, filamentous bacteriophages have a simple genome. Consequently, phages consist of a very simple make up of structural proteins forming a protective arrangement around the phage DNA genome. Figure 1.4 gives a representation of the phage structure and illustrates the handful of proteins that make up a phage particle.

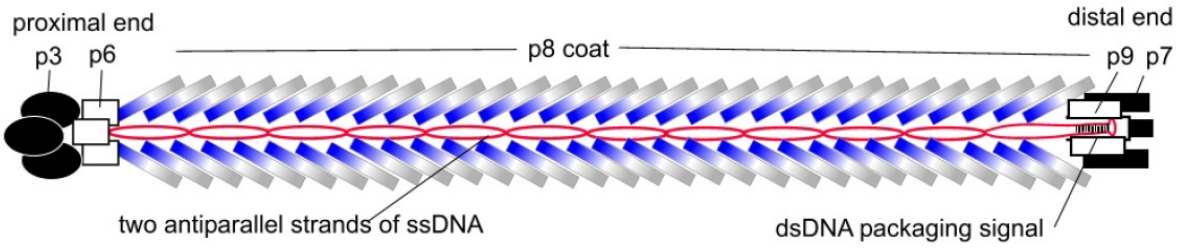

Figure 1.4- A representation of the final assembled structure of a filamentous phage. The p8 major coat protein forms a shell around the phage genome while minor coat proteins cap the ends of the virus. The proximal end of the phage extrudes last from the host and is responsible for host recognition while the diastral end extrudes first from the host. This figure is adapted from D. A Marvin, M. Symmons, S. K. Straus, manuscript in preparation.

The filamentous phage is arranged with major coat protein p8 forming a fish-scale like arrangement around the viral genome which consists of a single stranded, circular DNA [7]. In Ff phages, there are 2.3 nucleotides per major coat protein, while Pf1 has 1 and Pf3 has 2.4 [7]. For B5 phage, this is unknown. It is the structure of the major coat protein of the B5 phage that is of interest in this project. Its structure when reconstituted into lipid bilayers will be studied.

Phages can be classified according to the symmetry class, as defined by the packing of their major coat proteins, as identified by their diffraction patterns. Class I phage strains include Ff and Ike phages. Class II phage strains include Pf1 and Pf3 phages. Class I symmetry phages have $C_5S_{2.0}$ symmetry [10]. Class II symmetry phages have $C_1S_{5.4}$ symmetry [11]. p3, p6, p7 and p9 are minor coat proteins. The distal end of the phage emerges first from the host cell and contains 3-5 copies of p7 and p9. The proximal end of the phage is involved in phage entry and exit and contains 5 copies of each of p3 and p6. Phages lacking minor coat proteins appear as polyphages which are phages 10-20 times the length of a normal phage, are still tethered to the host and contain more than one complete circular phage genome. Phages lacking p3 and p6 were non infective, while phages lacking p7 and p9 were

still infective. Phages lacking p6 were also unstable [12]. p3 plays a role in not only phage-host cell recognition but also oligomerises to form a pore large enough for the phage to enter the host cell [13]. It is found that the N terminal two thirds of the p3 protein is essential for infectivity [14]. The whole phage structure is essentially held together by hydrophobic interactions between the apolar domains in the middle section of p8 and the apolar domains in the minor coat proteins. The remaining proteins encoded in the phage genome do not form part of the virus particle but are needed for phage assembly. p1, a 348 amino acid long protein, provides the energy needed to drive phage assembly and interacts with both thioredoxin and phage DNA during phage assembly [8]. Its C terminus is directed into the periplasm while the N terminus is directed into the cytoplasm. The protein p1 is seen to promote the formation of adhesion zones between the inner and outer membrane of Gram negative host membranes. These adhesion zones are the sites of phage extrusion [10]. p11 is a 108 residue protein whose genetic code is within gene I. It has basic residues at the N terminus, similar to p1, needed to interact with DNA. p4 exists in the outer membrane with its N terminus oriented towards the periplasm. It forms an oligomer of 10-12 subunits with an internal diameter of 80Å which acts as a pore for viral extrusion [15]. p4 functions as more than just an exit pore, it aids in phage assembly initiation because in the absence of p4, there is no visible build up of phage progeny in the host cell [7]. p2 controls the rate of replicative form (RF) DNA production [6]. p2 expression is negatively regulated by p5. p5 is 87 amino acids in length and forms a dimer, with its hydrophobic face buried away from the solvent. It binds tightly and cooperatively to phage DNA and protects phage DNA while it is in the host cytoplasm.

Each of the phage proteins plays an important role in assembly, a process which will be reviewed in the following section.

1.5 Filamentous phage assembly

Although many details of phage assembly are unknown at this juncture, a number of studies relating in particular to Gram negative phages have led to a number of models being proposed. By far the most studied phage in this regard is M13, which infects and replicates in *E. coli*

Entry of M13 into *E. coli* occurs when the phage protein p3 recognizes the F pilus of *E. coli*. The phage is then brought close to the host cell via retraction of the F pilus tip [6]. This hypothesis is known as the 'pilus retraction model' [6]. Once the phage is close enough to the host cell, infection occurs. p3 oligomerizes and forms a pore large enough for adsorption of the phage particle [13]. Glaser-Wuttke *et al* [13] have shown that p3 oligomerizes and forms large pores which stay open for seconds.

In addition to p3, adsorption of phage through the host *E. coli* outer membrane, periplasmic space and inner membrane requires the host proteins Tolerant Q (TolQ), TolR and TolA which are part of the Tol transport system [16]. It is thought that p3 interacts with domain III of TolA via its amino terminus [17]. This activates the TolQRA complex and leads to phage adsorption.

When phage is first adsorbed into the host, p6 is lost from the phage particle and the phage particle becomes unstable. This instability results in the full disassembly of the phage particle [6]. The phage coat protein p8 and other minor coat proteins are deposited into the

host inner membrane while the phage DNA enters the host cytoplasm where it is converted into a double stranded super coiled (RF) DNA via host enzymes [6].

The RF DNA is the template for not only phage DNA replication but also phage protein synthesis (transcription of RF DNA and translation). All 11 phage proteins are synthesized by transcription of RF DNA via the rolling circle mechanism and translation of the messenger ribonucleic acid (mRNA) produced [18]. The phage structural coat proteins p8, p3, p6, p7 and p9 are deposited into the host inner membrane upon translation. p1, p11 and p4 are proteins required for phage assembly and are deposited in the inner and outer (for p4) membranes of *E. coli* (Figure 1.5). The phage protein p5 homodimers bind to the newly synthesized phage DNA strand, thus separating the single stranded phage DNA from the rolling circle mechanism [6]. The DNA strand loops back on itself and a double stranded packaging signal is left exposed at the end of the DNA-p5 homodimer complex.

Interaction of the double stranded packaging DNA signal with the p1-thioredoxin complex at the host inner membrane triggers the formation of a pore. The continuous extrusion of the phage particle while it is being assembled requires the replacement of one p5 monomer in the DNA-p5 homodimer complex with two p8 coat proteins. This results in the production of the DNA-p8 complex. The formation of DNA-p8 complex is the pivotal step in phage assembly. The DNA is now protected by the phage coat protein p8. The protein p4 oligomerizes in the outer host cell membrane to form the pore for phage extrusion. The assembly process requires aggregation of p8 monomers and their assembly around the DNA complex, as well as changes in the structural properties of p8 in going from the non polar environment of the lipid bilayer to an aqueous environment upon extrusion [19]. Addition of p3 and p6 minor coat proteins to the extruding phage closes the phage particle and stops

phage assembly. The phage is then fully released into the medium. Figure 1.5 shows an illustration of the phage assembly process.

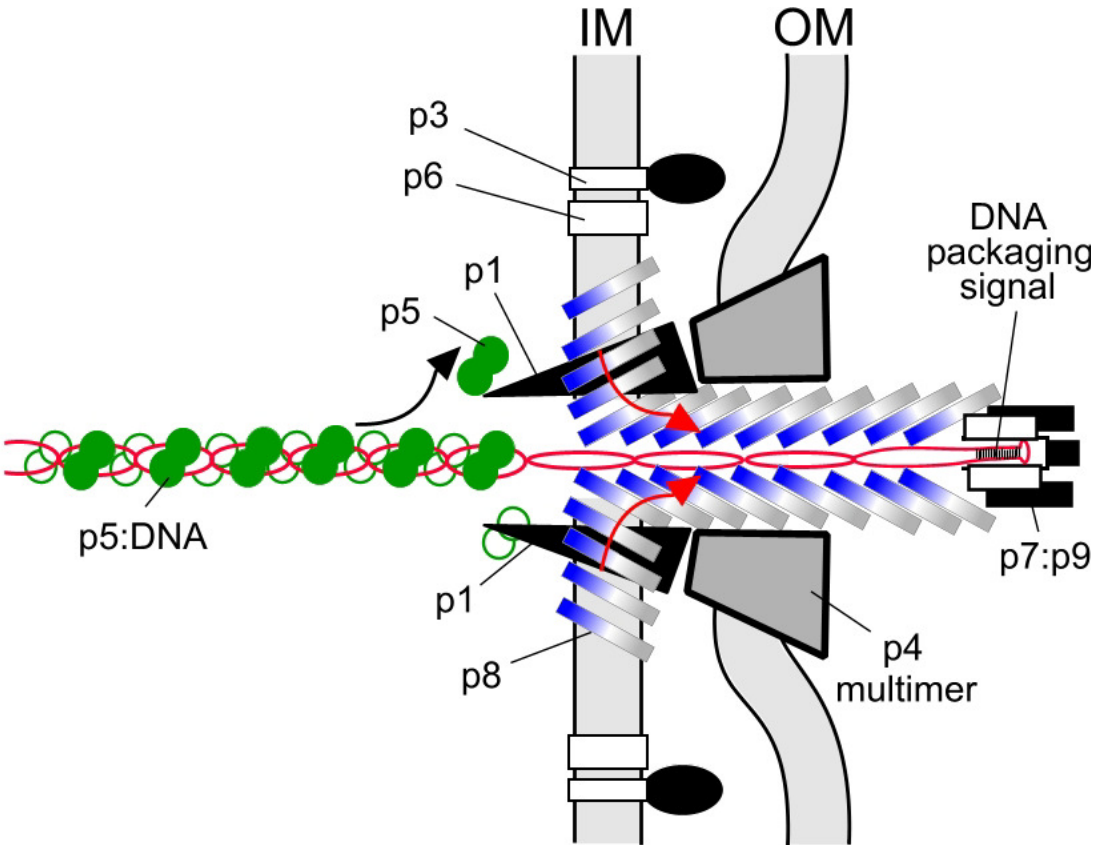

Figure 1.5- An illustration of the phage assembly process. Assembly is initiated when p1 interacts with the phage DNA packing signal. The major and minor coat proteins assemble around p1. Phage extrudes out of a pore created by p4 while p5 is replaced by p8 simultaneously. This figure is adapted from D. A Marvin, M. Symmons, S. K. Straus, manuscript in preparation.

The major coat protein p8 experiences environments of different polarity as it is being assembled. It goes from the non polar environment of the host lipid membrane where it is not associated with DNA to being associated with the phage DNA as it is shifted upwards within the host membrane and eventually extruded out into a polar environment. In the polar environment, the p8 protects the phage DNA and interacts with it. The p8 protein is very important in the phage assembly process. The structure of p8 in the different environments

has been a great point of debate. The reasons for the debate will be discussed in the following sections. It is the goal of this thesis to understand the structure of p8 during the assembly process.

1.6 Major coat protein p8

The p8 coat protein is the major component of the filamentous phage virion. Therefore understanding how its structure may change during assembly is of great importance. In the following section, I will therefore focus on a brief review of what is known about p8.

The major coat proteins of the different phage families are very similar. They are about 50 amino acids in length and consist of an acidic N terminal end, a hydrophobic central region and a basic C terminal end [14]. The major coat proteins of the Ff phages differ only by one amino acid: residue 12 in f1 and fd is an aspartic acid but in M13, it is an asparagine [14]. The major coat protein in B5 phage is not synthesized with a signal sequence while the major coat protein of Ff phages are synthesized with a 23 amino acid leader signal sequence which is cleaved off upon membrane insertion.

When in the host membrane, p8 is oriented with its N terminal end directed towards the outer membrane and its C terminal end directed towards the host cell cytoplasm [14]. When in the phage, 2700 copies of the major coat protein form a layer around the viral DNA molecule. And because the p8 proteins are held together by hydrophobic interactions in the mid section of p8, the phage is resistant to proteases, salts, detergents and extreme pH. Phages can be lysed mechanically or by chloroform [6].

In the virion, the acidic N terminus of p8 is oriented towards the exterior of the phage. The negative charge on the N terminus is thought to be responsible for dissolving the phage in solution. The basic C terminus of p8 points towards the interior of the phage and interacts with the negatively charged DNA via its lysine residues (4 in total) [20]. There are non-specific electrostatic interactions independent of DNA sequence between the positively charged lysine residues and the negatively charged DNA [20].

1.6.1 The structure of the major coat protein in the virion

At the end of assembly, p8 is in the virion. The structures of the major coat proteins in M13 and fd phage have been studied extensively. The protein itself measures 1nm by 7nm and has been shown by X ray diffraction studies to be at a 20^{o} angle from the long axis of the filament [10]. The p8 molecules overlap with neighboring p8 molecules and have interlocked side chains with about 2.7nm space between the N terminal ends of adjacent subunits [10].

Welsh *et al* [21] have determined the structure of the Pf3 coat protein to 3.1Å resolution. The Pf3 coat protein appears as a single gently curved helix at an angle of 20.4^{o} (±2.3) with respect to the virion axis [21]. They found that Pf3 and Pf1 coat proteins have a few differences in structure even though they belong to the same helix symmetry (class II). Pf3 is almost completely alpha helical, while Pf1 is helical except for the first five amino acids [21]. Also, the nucleotide/protein ratio (nt/s) for Pf3 is 2.4 while the nt/s for Pf1 is 1.0 [21]. The Pf1 coat protein is thought to interact with the enclosed phage DNA via positively charged side chains at the C terminal end, but the Pf3 coat protein interacts via both positively charged side chains and aromatic side chains [21].

Zeri *et al* [22] have shown that the fd major coat protein is a straight, slightly curved helix except for a kink near residue 39. The residues 1-5 are unstructured and the rest of the protein is alpha helical [22]. Residues 8 to 38 have a 23^{o} tilt while residues 40 to the C terminus have a smaller tilt of 18^{o} [22]. This kink in the helical structure gives the protein a similar shape as its membrane bound form, as solved in micelles [22]. The kink in the protein is thought to be important in maintaining a low energy structure when the coat protein is in the host membrane and to optimize side chain interactions at the C terminus of the protein where the protein would interact with the phage DNA inside the phage [22]. Marvin *et al* [23] have however, recently shown that p8 does not adopt a kinked helical structure, but rather forms a continuous helix. This model agrees with all experimental data such as fiber diffraction, solid state NMR and mutational studies.

1.6.2 The structure of major coat protein during extrusion

Although much remains to be resolved about the structure of p8 during assembly, it is known that the p8 of M13 not only experiences environment changes in going from the nonpolar host membrane to the polar aqueous environment, but that it also experiences a shift in position within the membrane. Papavoine *et al* [15] describe the interactions between p8 monomers in phage assembly by combining the data from solution state NMR performed by both Papavoine *et al* [15] and Marvin *et al* [24]. In Figure 1.6, p8 proteins -1, 0 and 1 make up the original layer 0 in the phage coat that will be discussed. First, there is an interaction of the C terminal lysine residues in proteins -1, 0 and 1 with the phage DNA. This interaction causes the initial protein orientation changes needed in phage assembly [15]. The lysine residues are present on one side of the amphipathic helix and can interact with DNA while

phenylalanine residues at positions 65 and 68 are present on the other side of the amphipathic helix and play a role in anchoring the protein to the membrane. Upon interaction with DNA, proteins in layer 0 are in the correct orientation to interact with the proteins in the layer preceding them, i.e. layer 11 which comprises proteins 5, 6 and 11 [15]. Coat protein layer 0 then shifts up by 16.15Å [15]. This shift allows the proteins in the layer -11 that follows, to enter phage assembly and interact with layer 0 [15]. The interaction between p8 proteins in successive layers of the phage coat allows stabilization of the phage particle as it extrudes. Every layer of coat protein that is incorporated shifts the proteins -1, 0 and 1 in the original layer 0 by 16.15Å but it will take the incorporation of three layers to fully extrude the layer 0 out of the host inner membrane because it is 48.45 Å thick [15].

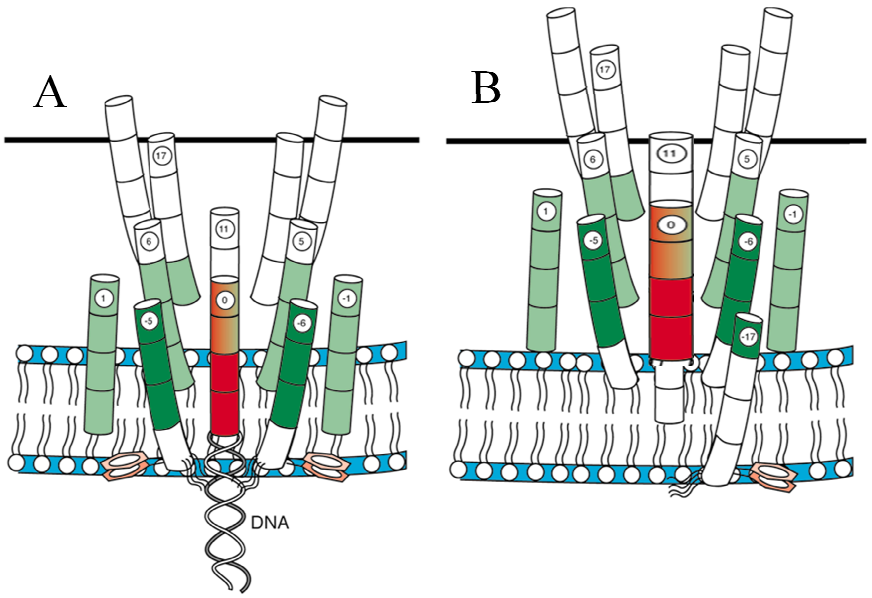

Figure 1.6- Cartoon showing the interaction of successive layers of p8 protein as the phage is assembled. Layer 0 represents the original layer and layer -11 represents the layer that follows. Layer 0 is seen in A to be extruding out from the inner membrane and in B, it has shifted up by16.15A upon the incorporation layer -11 (presented in red). Adapted from [15].

1.6.3 The structure of the major coat protein in membrane

The shape of the p8 protein from M13 phage when present in the host membrane is of much debate. The shape is slightly different, depending on what biophysical technique has been applied in the study. This stems from the fact that in the biophysical techniques used to study its structure, p8 is reconstituted into different lipid environments and the structure of the internally unstable p8 changes depending on the environment it is reconstituted in.

Solution nuclear magnetic resonance (NMR) is a technique used in the study of the major coat protein structure. The coat protein is reconstituted into micelles with lipid content representative of the bacteria host membrane. Papavoine *et al* [15] reconstituted the M13 major coat protein in sodium dodecyl sulphate (SDS) and dodecylphosphocholine (DPC). They found that the M13 protein is composed of three main regions: the amphiphathic helix (residues 8 to 16), a hinge region (residues 17-24) and a transmembrane hydrophobic helix (residues 25-45). The amphipathic helix is found on the surface of the micelle almost at a right angle to the transmembrane helix [25]. In the amphipathic helix, the residues are arranged around the helix based on their side chains. Amino acids with large side chains such as lysine 8 and phenylalanine 11 are found on the side of helix that is directed away from the solvent while amino acids with smaller side chains such as alanine 9 and serine 13 are found on the side of the helix that is directed towards the solvent [15]. The hinge region of the M13 coat protein acts as a flexible linker between the amphiphathic helix and the hydrophobic helix. The hydrophobic helical region of M13 is located in the micelle and amino acid residues are equally distributed across the helix independent of their side chains, although aromatic residues tryptophan 26, phenylalanine 42 and phenylalanine 45 are found at the

ends of the hydrophobic helix indicating that these aromatic residues are important in anchoring the coat protein [15].

Solid state NMR is another important technique in the study of the M13 major coat protein. Solid state NMR involves reconstituting the M13 protein into lipid bilayers. Solid state NMR studies have shown that the fd coat protein consists of a 16Å amphipathic helical segment which lies parallel to the lipid bilayer surface and a 35Å transmembrane helical segment which lies at a tilt of 16^{o} with respect to the lipid bilayer normal [19]. The two helices are connected by a short loop made of residues threonine 19 and glutamic acid 20 [19]. The shape of M13 protein solved by solid state NMR is that of an L shape.

Site Directed Labeling (SDL) is yet another method used to solve the structure of the major coat protein by reconstituting p8 in lipid vesicles. The shape of the major coat protein obtained from SDL experiments appears to adopt an I shape. It consists of a single gently curved helix tilted by 18^{o} with respect to the normal of a unilamellar vesicle membrane and with maximum curvature at residue 20 [26]. Even though SDL is thought to provide the closest representation of the host membrane environment, the helix shows changes in structure when placed in slightly different membrane environments. Stopar *et al* [27] have shown that the first 7 residues in the N terminus are unstructured when the protein is placed in 22:1 PC and if placed in 14:1 PC, the first 14 residues are unstructured. M13 protein shows a tilt of 33 ^{o} in 14:1 PC, while it has a tilt of 19^{o} in a 20:1 PC [28].

Spruijt *et al* [29] have performed SDL experiments on M13 major coat protein. They engineered M13 mutants with single cysteine substitutions, attached an N-(iodoacetylaminoethyl)-5-naphthylamine-1-sulfonic acid (ADEAS) probe and performed single cysteine scanning fluorescence microscopy. It was found that M13 had a banana-like

structure; it was a single slightly curved and tilted molecule with a flexible hinge [29]. The flexible hinge loses its alpha helicity when M13 is placed in the thinner 14:1 PC membrane as compared to a 18:1 PC membrane [29].

Stopar *et al* [30] investigated the anchoring of M13 by 6-bromoacetyl-2-dimethylaminonaphthalene (BADAN) labeling. The change in fluorescence of an intrinsic tryptophan 26 residue is analyzed when the M13 protein is placed in 22:1 PC versus 14:1 PC. A blue shift in fluorescence of about 700cm^{-1} is observed [30]. This shift corresponds to the tryptophan going into a more non polar environment, suggesting that the tryptophan molecule is 'sinking' into the lipid bilayer as the bilayer thickness increases [30]. The residue 46 is also tested for its anchoring ability. This residue is mutated into a cysteine residue and a BADAN probe is attached onto the cysteine. BADAN shows a blue shift in fluorescence when moved into a non polar environment. When the p8 protein from M13 with the BADAN labeled cysteine residue 46 is placed in a 22:1 PC versus 14:1 PC, there is a blue shift of only 70cm^{-1} [30]. From these results, it was concluded that the anchoring strength of the C terminus is five times stronger than that of the N terminus. The C terminus consists of two phenylalanine resides and four lysine residues which provide strong anchoring of the protein. The N terminus consists of aromatic and charged residues glutamate, tyrosine and tryptophan [30]. Meijer *et al* [31] showed that the C terminus of the coat protein is important in anchoring the protein. They increased the length of the transmembrane domain of the protein and found that the N terminus of the transmembrane domain moved into a more polar environment while the C terminus was shown to be unaffected [31]. This result adds to the hypothesis that the C terminus of the protein has a stronger anchoring ability than the N terminus of the M13 protein [30]. Tryptophan, lysine and phenylalanine are important in

anchoring a transmembrane protein in the membrane. These positively charged and aromatic residues are present in the interfacial region of the lipid bilayer which is the region where the phosphate head groups exist.

1.7 I Vs L- The debate

As described above, the discrepancy in the shape of the major coat proteins of M13 and Pf1 comes from the fact that the protein is sensitive to its environment and has lead to a debate in the phage literature as to whether the major coat protein is an I- or an L- shaped helix. In solid state NMR, the protein is reconstituted into oriented bilayers which could restrict the N terminus of the protein and 'press' it between the successive bilayers [25]. This could force the protein into an L shape, formed by two alpha helices connected by a flexible hinge. When solution state NMR is performed, the protein is reconstituted into micelles. Micelles have strong curvature that could force the protein to curve [25]. In SDL experiments, however, the proteins are reconstituted into fully hydrated vesicles which closely mimic the host cell membrane environment [25]. However, one should note that attaching the large probes needed in fluorescent labeling experiments could potentially also affect the protein overall structure. Figure 1.7 shows p8 structures in micelles as studied using solution state NMR and bilayers as studied using solid state NMR. Structures are obtained from Protein Data Bank.

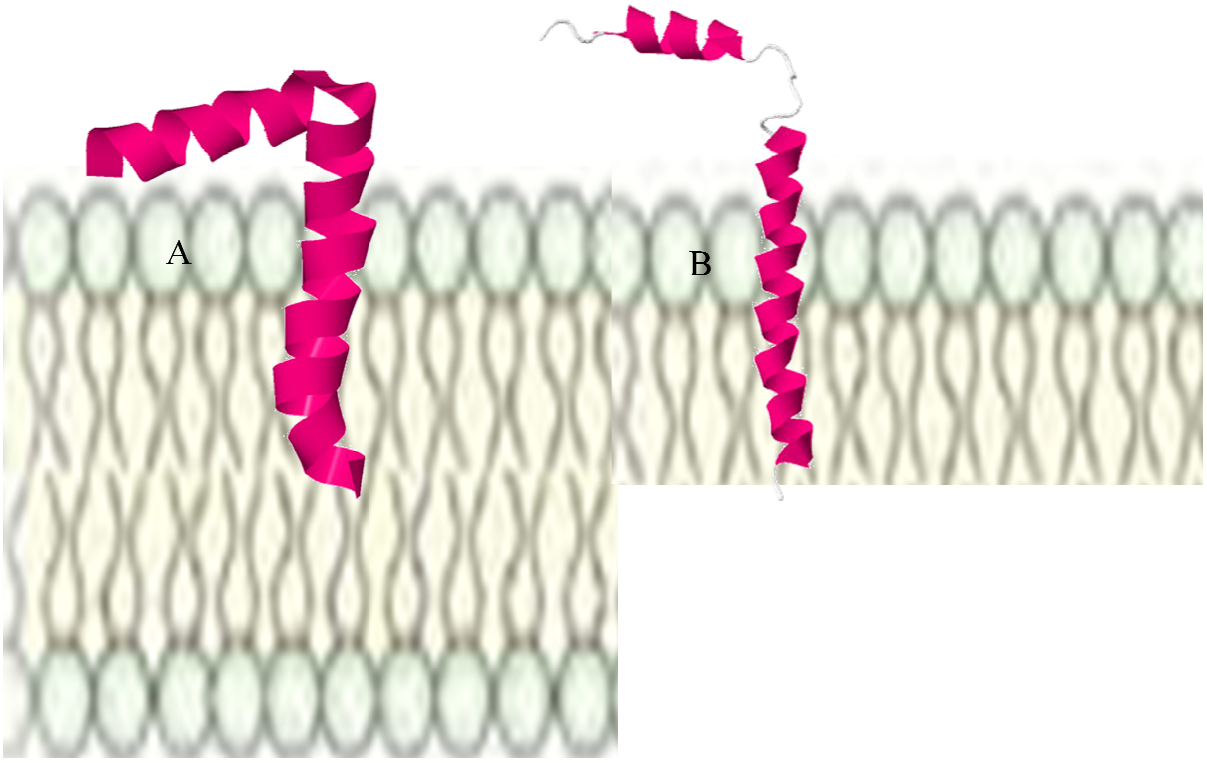

Figure 1.7- Structures of major coat protein as obtained from PDB. Figure A shows the structure of the major coat protein of fd phage in lipid bilayers studied via solid state NMR (PDB ID: 1MZT). Major coat protein is in an L conformation with a short hinge region. Figure B shows the structure of the major coat protein of Pf1 phage in lipid micelles studied via solution state NMR (PDB ID:2KSJ). Major coat protein is more curved and has a longer flexible hinge region.

In order to solve the structure of p8 in its native environment in the host membrane, p8 should be reconstituted in an environment which accurately presents that native environment.

Since p8 structure and, as described in the previous section, orientation depend so strongly on the nature of the membrane, and since it is very difficult to construct an accurate model of Gram negative membrane (due to their complexity), the I vs L debate is likely to rage on. One possible solution, however, is to study p8 structure in Gram positive membranes which are easier to mimic than Gram negative membranes. Gram negative membranes contain a periplasmic region between the inner and outer membrane. The periplasmic space will affect the structure of the coat protein and contribute to the native

conformation of p8. Its effect should therefore not be ignored when studying p8 native structure. Gram positive membranes are composed of an inner lipid bilayer membrane and a peptidoglycan cell wall. The inner lipid bilayer can be mimicked with the use of appropriate lipids which commonly occur in bacterial membranes such as palmitoyl oleoyl phosphatidyl choline (POPC), palmitoyl oleoyl phosphatidylglycerol (POPG) and cardiolipin (cardio). A number of studies from our own laboratory have demonstrated the validity of these membrane compositions as models for Gram positive bacteria [32] [33]. The major coat protein of B5 filamentous phage is a perfect candidate since B5 infects Gram positive bacteria. The aim of this thesis is to express the p8 of B5 phage using a bacterial expression system, to reconstitute p8 into a model membrane and to perform structural studies using techniques such as CD, NMR and fluorescence.

Chapter 2: Expression of P8

2.1 Introduction

In order to study p8 structure, the major coat protein has to be either synthesized or expressed. Peptide synthesizers have been used to synthesize many peptides and short proteins. They offer a way to produce peptides chemically. Solid phase peptide synthesis is suitable for proteins or peptides which are shorter than 30 amino acids. Synthesis of longer peptides is possible but yields will be affected and peptides with single mutations or deletions could be synthesized in much higher quantities than the full length, desired peptide. Since p8 of B5 phage is a 57 amino acid long protein, if peptide synthesis is used, the yield of the p8 protein would be low and many fragments would be produced.

The expression of p8 in *E. coli* can also prove to be a difficult task which involves screening numerous conditions due to the fact that it is a heterologous membrane protein. Most membrane proteins are insoluble by nature and insoluble proteins get sequestered into inclusion bodies once expressed in the cytoplasm. These inclusion bodies can kill the *E. coli* or slow down its growth tremendously. Once it is fully optimized, however, heterologous expression in *E. coli* can be a cost effective way to isolate relatively large yields of proteins. Many considerations have to be taken into account when designing a system to express a membrane protein.

The vector used has to be first considered. p8 has to be rendered soluble by choosing an appropriate vector. The vectors chosen in this case are the pMAL-c2X and pMAL-p2X vectors from New England Biolabs. The same vectors have previously been used by Andrew Tait in our lab to successfully express the membrane protein U24 from Human Herpes Virus-

6. The pMAL vector series has a coding region for maltose binding protein (MBP) which is known to solubilize insoluble proteins, thus increasing protein expression yields by reducing the formation of insoluble aggregates. The pMAL-p2X vector codes for MBP with its signal sequence which causes it to be sequestered in the periplasm of *E. coli* cells, whereas in the pMAL-c2X vector, the signal sequence is deleted and MBP exists in the cytoplasm. This provides the option of expressing the passenger protein in either the periplasm or the cytoplasm. Expression in the cytoplasm generally results in high yields of protein expression but expression in the periplasm has its advantages. The periplasmic space of *E. coli* provides an oxidizing environment in which proteins that require the formation of disulphide bonds can escape degradation which would otherwise occur if these stabilizing bonds are not allowed to form. It has also been shown that periplasmic expression yields more protein so in the case of p8, both types of expression are trialed.

The strains of *E. coli* used for expression are also an important consideration. Origami 2 strain is a mutant strain where the reductases which keep the thiols of cysteine residues in a reduced form are inactivated. This not only allows for the formation of cysteine bonds in the *E. coli* cytoplasm, it might also provide an environment where protein degradation by host enzymes is reduced. The downfall of this strain is that it grows slowly, although Tait *et al* [34] have shown that for a lower cell pellet mass, a higher yield of the target protein is produced for Origami cells compared to C41 cells.

Once expression is optimized, p8 has to be purified from the cell lysate since proteins native to *E. coli* will also be expressed. The purification system used for p8 purification is the commonly used nickel-nitrilotriacetic acid (NTA) column which has an affinity for the flexible hexa-histidine tag added to the N terminus of the p8 sequence. The tag can be

displaced from the beads by buffer with high imidazole concentration. The simplicity of this purification system and the fact that it is widely used for a variety of proteins makes this a good method for p8 purification. In the expression of p8, the use of MBP as a solubility enhancer and hexa-histidine tag as an affinity tag for purification proved to be effective in producing good p8 yields.

The aim is to produce a high yield of p8 using the above mentioned considerations. Although expressing a non native membrane protein is difficult and is subject to many variables, careful consideration of solubility, vector and bacteria strains is a good starting point.

Once p8 is successfully expressed, the structure and lipid-dependent membrane interaction of p8 structure will be studied (Chapter 3). In this thesis, p8 structure is studied using circular dichorism (CD), ^{1}H and ^{15}N NMR. The study of p8 using ^{15}N NMR requires that p8 be ^{15}N labeled. Also, p8 structure is studied using 6-bromoacetyl-2-dimethylaminonaphthalene (BADAN) fluorescence. This requires the expression and purification of a cysteine mutant, to which the BADAN label will be attached.

2.2 Methods and materials

2.2.1 Insert sequence preparation

In order to express a protein in *E. coli*, its gene sequence has to be transformed. The *E. coli* can then be induced to express the heterologous protein. However, the route to protein expression is not that simple. The gene sequence of the protein of interest has to be first modified and optimized for expression in *E. coli*.

As a starting point, we obtained the p8 amino acid sequence from the National Center for Biotechnology Information (NCBI), accession number: AF428260.1. The nucleotide sequence is derived by back translation of the amino acid sequence using Expasy Translate. The next step is to optimize the nucleotide sequence to account for codon bias. *E. coli* has codon bias: this happens because some of the tRNAs needed to translate a heterologous protein are rare in *E. coli*. Therefore, if codon bias is not taken into account, heterologous protein expression levels will be low because of the depletion of these rare codons. The rare codons in the protein sequence have to be changed into codons that are more common in *E. coli*. Bacterial strains such as ArcticExpress (RP) and ArcticExpress (DE3) RP can be used as well to overcome codon bias. These strains have extra copies of arginine and proline tRNA genes that recognize the arginine codons AGA and AGG, and the proline codon CCC. In our case, we optimized the sequence by using an online tool from UC Riverside[1] which identifies rare codons (lower than 10% threshold).

At the 5' and 3' ends of the sequence, GAATTC codons are added. Guanine forms 3 hydrogen bonds with a cytosine on the complementary DNA strand. This stabilizes the double stranded DNA at its ends. The fact that the adenine and thymine form two hydrogen bonds with nucleotides on the complementary strand eases strand separation during the polymerase chain reaction (PCR) step (section 2.2.2). These 6 extra nucleotides at the ends also allow for efficient binding of the restriction endonuclease to the DNA sequence.

The sequence is then modified to include restriction enzyme sites, a hexahistidine tag and a thrombin cut site. The restriction enzymes allow the insert to be 'cut', forming sticky ends on both ends. The vector plasmid into which the p8 sequence has to be inserted has

[1] http://faculty.ucr.edu/~mmaduro/codonusage/usage.htm

complementary sticky ends. This will allow the p8 sequence to be inserted into the vector and to be ligated. The plasmid that forms after ligation can then be transformed into *E. coli* for expression of the desired protein. The restriction enzymes chosen in this project are BamHI and HindIII because they are commonly used and are readily available. The hexa-histidine tag forms a flexible linker between p8 and the fusion protein. MBP is used to increase the solubility of the passenger protein. The thrombin cut site allows the fusion protein to be cleaved off from p8 upon purification.

The sequence from each optimization step is shown in Table 2.1 . The full sequence is represented as a double stranded DNA sequence which is then divided into 4 shorter fragments, referred to as oligos, with overlapping nucleotides. These fragments are analyzed for hairpin loops. Oligos are ordered from Nucleic Acid Protein Service Unit (NAPS) at 100 nmole scale.

Table 2.1- Table showing the p8 sequence from each step of optimization and modification.

	Sequence
Original p8 amino acid sequence	MLPMVILEAPSDVGGTVSAAITALGPQITPIIGVAIGVSLIPFAAKWIFR KAKSLVS
Corresponding p8 nucleotide sequence (as reversed translated using Expasy)	atgcttcccatggtgattttggaggctccgagcgatgtcggtggcactgtttctgctgcgatcacggc tctcgggccgcagatcacgccgattatcggcgtggctatcggtgtttctctgattccgttcgcagccaag tggattttccgcaaggcaaagagtctggtttcctga
p8 sequence optimized to remove rate codons	atg **ctg** **ccg** atg gtg att **ctg** gag gct ccg agc gat gtc ggt ggc act gtt tct gct gcg atc acg gct **ctg** **ggt** ccg cag atc acg cca att atc ggc gtg gct atc ggt gtt tct ctg att ccg ttc gca gcc aag tgg att ttc cgc aag gca aag **agc** ctg gtt tcc tga • Codons that are changed to account for rare codons are in bold and underlined
Addition of	GAA TTC **GGA TCC** CAT CAC CAT CAT CAT CAC AGC AGC GGC

	Sequence
BamHI, **HindIII** and *thrombin cut sites*, and hexahistidine tag	*CTG GTG CCG CGC GGC AGC* atg **ctg ccg** atg gtg att **ctg** gag gct ccg agc gat gtc ggt ggc act gtt tct gct gcg atc acg gct **ctg ggt** ccg cag atc acg **cca** att atc ggc gtg gct atc ggt gtt tct ctg att ccg ttc gca gcc aag tgg att ttc cgc aag gca aag **agc** ctg gtt tcc tga **AAG CTT** GAA CAC • The sequence coding for additional cut sites and tags are shown in capital letters and the original p8 sequence is identified by small letters with the optimized codons in bold and underlined as before. • Note that codons not identified by special font are added to increase the overall bulk of the sequence, to optimize access to either the restriction enzyme digest site or the thrombin cut site.
Double stranded full p8 gene sequence	**5′GAATTCGGATCCCATCACCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCAT** 3′CTTAAGCCTAGGGTAGTAGTAGTAGTAGTGTCGTCGCCGGACCACGGCGCGCCGTCGTA **GCTGCCGATGGTGATTCTGGAGGCTCCGAGCGATGTCGGTGGCACTGTTTCTGCTGCGA** CGACGGCTACCACTAAGACCTCCGAGGCTCGCTACAGCCACCGTGACAAAGACGACGCT **TCACGGCTCTGGGTCCGCAGATCACGCCAATTATCGGCGTGGCTATCGGTGTTTCTCTG** AGTGCCGAGACCCAGGCGTCTAGTGCGGTTAATAGCCGCACCGATAGCCACAAAGAGAC **ATTCCGTTCGCAGCCAAGTGGATTTTCCGCAAGGCAAAGAGCCTGGTTTCCTGA** TAAGGCAAGCGTCGGTTCACCTAAAAGGCGTTCCGTTTCTCGGACCAAAGGACT **AAGCTTGAACAC3′** TTCGAACTTGTG5′
Fragments ordered	Fragment 1: **GAATTCGGATCCCATCACCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCATGC** **TGCCGATGGTGATTCTGGAGGCTCCG** Fragment 2: ATTGGCGTGATCTGCGGACCCAGAGCCGTGATCGCAGCAGAAACAGTGCCACCGACATCGC TCGGAGCCTCCAGAATCACCATCG Fragment 3: **GTCCGCAGATCACGCCAATTATCGGCGTGGCTATCGGTGTTTCTCTGATTCCGTTCGC**

	Sequence
	Fragment 4: GTGTTCAAGCTTTCAGGAAACCAGGCTCTTTGCCTTGCGGAAAATCCACTTGGCTGCGAAC GGAATCAGAGAAACACCGA

2.2.2 Polymerase chain reaction (PCR)

Since the insert is designed in four separate sequences with overlapping nucleotides, overlap PCR is performed to fuse the four DNA fragments together. PCR reamplification is then performed to create more copies of a DNA fragment which is essential in order to increase transformation efficiency. In order to perform PCR and to generate the DNA which encodes p8, the following components were mixed:

1) 2 µl of 10 mM deoxyribonucleotide triphosphate (dNTP) (Fermentas)
2) 10 µl of X10 pfu buffer (Fermentas)
3) Primer and oligo concentrations of 2.25 µM
4) 1.25 units of pfu in a total volume of 44 µl.

PCR was performed on the Minicycler from MJ Research. The PCR programs used are "hhv6touch 1", "PCR re-amplification" and "colony PCR". "Hhv6touch 1" is the program used to perform overlap PCR. The details of these programs are given in appendix A.1.

Overlap PCR condenses oligo 1 with oligo 2 and oligo 3 with oligo 4 using appropriate primers. The oligo that is formed is half the length of the full p8 sequence. The half-length oligos are amplified via PCR re-amplification PCR. Overlap PCR again condenses the two half length oligos to form the full length p8 sequence. The sequence is

then re-amplified via PCR reamplification. Figure 2.1 shows a simple cartoon representation of the basic steps involved in overlap PCR.

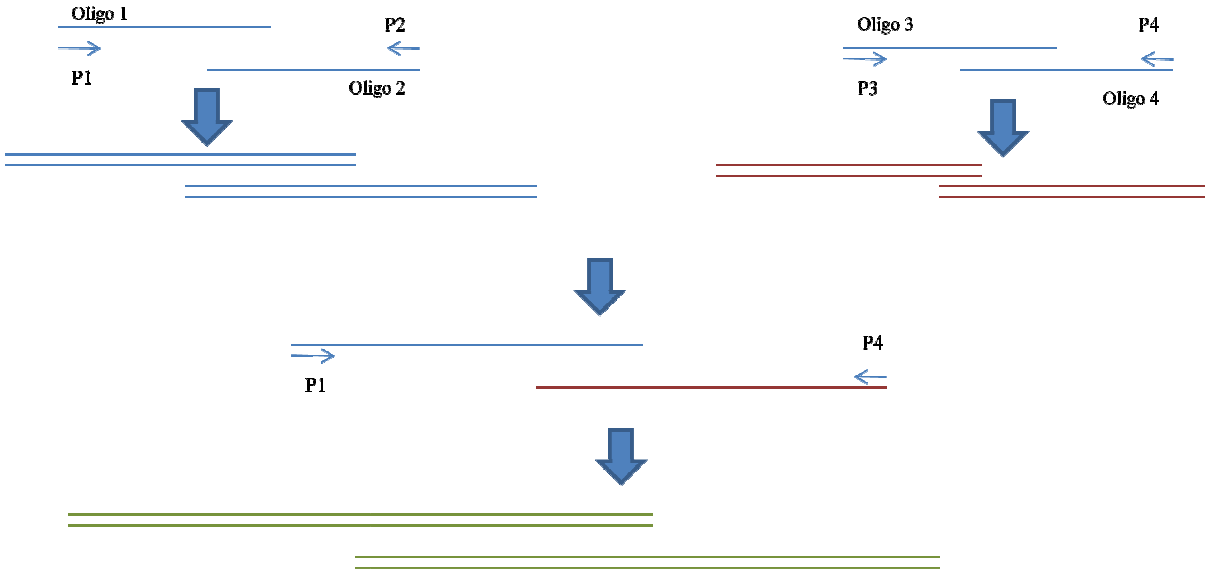

Figure 2.1- Cartoon representation of overlap PCR. 'P' here stands for primer. Primer and oligo sequences for overlap PCR of p8 sequence can be found in Table 2.2.

PCR involves a few basic steps and cycles, as illustrated in Figure 2.1. First, the DNA has to be denatured ie. the double stranded DNA is denatured into its 2 complementary single stranded DNA strands. The denaturation step normally occurs at 95^{o}C for 30s. The strands then undergo an annealing cycle. In this cycle, the primers anneal to the complementary base pairs on the DNA strands. The next cycle is the extension cycle where a complementary strand is synthesized from the 3' end of the primer by DNA polymerase. There are many things to take into account when optimizing PCR. The common factors that affect the purity and yield of PCR are annealing temperature, primer length, elongation time and temperature, and cycle number. They were all optimized here, according to findings in literature [35-37].

The physical properties of engineered primers and oligos can be predicted on Net Primer available through Primerbiosoft[2]. The properties of the primers are summarized in Table 2.2. Primers are ordered from NAPS at 25 nmole scale.

Table 2.2- Sequences and properties of the optimal 4 primers and oligo fragments used in PCR for p8 of B5 phage. Note that sequence 1-4 are the same as Fragment 1-4 in Table 2.1.

	Sequence	Tm (^{o}C)	G+C content (%)
Primer 1	GAATTCGGATCCCATCACC	53.0	52.63
Primer 2	ATTGGCGTGATCTGCGG	55.9	58.82
Primer 3	GTCCGCAGATCACGCCAAT	58.2	57.89
Primer 4	GTGTTCAAGCTTTCAGGAAACCAG	56.6	45.83
Sequence 1	GAATTCGGATCCCATCACCATCATCATCACAGCAG CGGCCTGGTGCCGCGCGGCAGCATGCTGCCGATG GTGATTCTGGAGGCTCCG	75.5	62.0
Sequence 2	ATTGGCGTGATCTGCGGACCCAGAGCCGTGATCG CAGCAGAAACAGTGCCACCGACATCGCTCGGAGC CTCCAGAATCACCATCG	74.9	60.0
Sequence 3	GTCCGCAGATCACGCCAATTATCGGCGTGGCTATC GGTGTTTCTCTGATTCCGTTCGC	71.8	55.1
Sequence 4	GTGTTCAAGCTTTCAGGAAACCAGGCTCTTTGCCT	72.7	50.0

[2] http://www.premierbiosoft.com/netprimer/index.html

	Sequence	Tm (^{o}C)	G+C content (%)
	TGCGGAAAATCCACTTGGCTGCGAACGGAATCAG AGAAACACCGA		

After PCR is completed, it was necessary to visualize the weight of the DNA product to ensure that the correct product was formed. This was achieved using agarose gels. The protocol for visualizing the DNA product is in appendix A.2.

2.2.3 Expression vector plasmids

Once the fully optimized p8 gene sequence is obtained, it is inserted into an expression vector. In the case of p8, the sequence will be inserted into pMAL expression vectors. The pMAL expression vectors code for maltose binding protein (MBP) and p8 protein fused to MBP as a fusion protein, for reasons explained below. pMAL p2X and pMAL c2X vectors from New England Biolabs were used.

The vectors used enable p8 expression with a MBP fusion protein. MBP will increase the solubility of its passenger protein, p8 which is an insoluble membrane protein. MBP is thought to have intrinsic molecular chaperone properties [38]. Kapust *et al* [38] have shown that MBP increases the solubility of 6 proteins which are difficult to express due to their insolubility. Fusing an insoluble protein to any protein which is soluble in nature will not necessarily increase the solubility of the passenger protein. MBP has been shown to be a much better fusion protein than other common fusion proteins such as GST and TRX [38]. The way in which maltose binding protein increases the solubility of a protein is illustrated in

Figure 2.2. Once expressed, the passenger protein is thought to be in an intermediate state known as the folding intermediate. From this state, it could fold into its native structure or it could form insoluble aggregates [38]. In order for the intermediate to escape the latter fate, MBP sequesters its passenger protein [38]. This reduces the number of protein in the intermediate state and will eventually allow the folding of the passenger protein into its native sequence. Also, with the passenger protein sequestered, the chances of interaction between the passenger proteins will decrease and the probability of the formation of insoluble aggregates will decrease [38]. Thus, MBP aids in increasing protein solubility by reducing aggregation. It should be noted that MBP acts as a solubility enhancer, it does not infer spontaneous folding abilities onto its passenger protein [39]. Louis *et al* [39] have shown that there are cases where the protein is deemed soluble by MBP but the protein is still unable to attain native structure [39]. Sachdev *et al* [40] have suggested that the solubility enhancement of MBP is only seen when the passenger protein is fused to the C terminus of MBP.

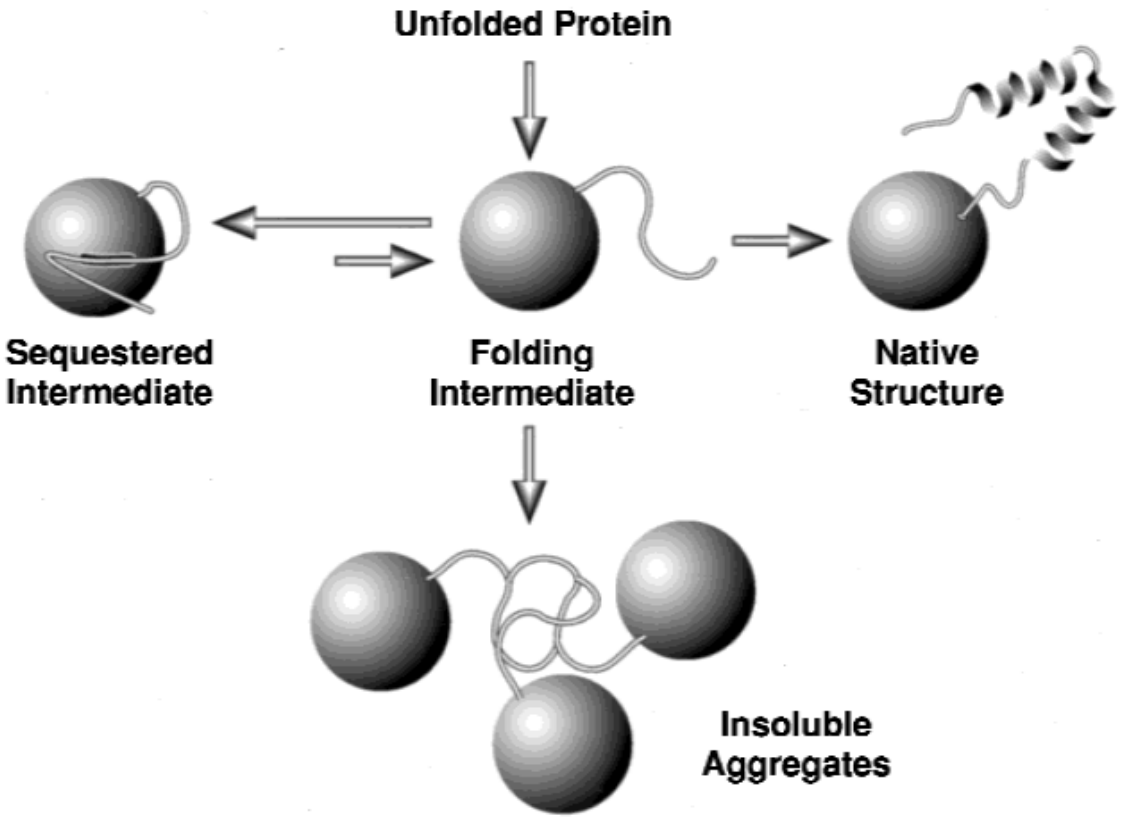

Figure 2.2- An illustration showing how MBP may promote solubility of passenger protein. MBP sequesters the passenger protein and allows the passenger protein to follow the path to native structure by reducing the amount of passenger proteins in the folding intermediate state. Adapted from [38].

As stated before in Section 2.2.1, a hexahistidine tag is placed between the MBP and p8 as a flexible linker. Its affinity to nickel beads is exploited by using nickel-NTA column to purify p8. Adding a hexa-histidine tag at the N terminus of the passenger protein does not interrupt the ability of MBP to infer solubility onto the passenger protein [41].

2.2.4 Restriction enzyme endonuclease digestion, ligation and transformation

In order for the p8 sequence to be inserted into the vector plasmid for expression, the p8 gene sequence has to be first cut with restriction enzymes to create sticky ends. The p8 with the sticky ends can then be ligated into a vector plasmid, also cut with the same

restriction enzymes. The protocols for restriction enzyme digestion and ligation are in appendix A.3.

Once successfully ligated, the plasmid containing the p8 sequence can then be transformed into *E. coli* strains for propagation of the plasmid. The strains used in this project for DNA propagation are SURE and XL1-BLUE (from Stratagene). The transformation protocol can be found in appendix A.3. The colonies have to be checked to ensure that they contain the plasmid with the p8 sequence rather than just the vector plasmid that re-ligated. Colony PCR is used to screen for colonies that contain the desired plasmid. In colony PCR, primers specific to the pMAL plasmid (pMAl seq and pMAL seq rev) are used. These primers attach to the plasmid at sites on either end of the p8 sequence cloning site. If the p8 sequence does not insert successfully into the plasmid, colony PCR would result in DNA around 200bp (the number of base pairs covered by the forward and reverse pMAL primers), but if the p8 sequence is successfully inserted into the plasmid, colony PCR would result in DNA at 200bp plus the number of base pairs in the p8 sequence (245bp). The pMal seq and pMal seq rev primer sequences are as follows:

pMAL seq primer: GGTCGTCAGACTGTCGATGAAGCC

pMAL seq rev primer: CGCCAGGGTTTTCCCAGTCACGAC

The plasmids can then be extracted from the correct colony using QIAGEN plasmid miniprep (from QIAGEN). The protocol for colony PCR can be found in appendix A.4 and the QIAGEN plasmid miniprep is performed by following the protocol in appendix A.5.

2.2.5 Protein expression

Once the correct plasmid is obtained, it is transformed into *E. coli* and protein expression is induced. High yield protein expression requires optimization by changing many different conditions including different media composition, different bacteria for expression and different growth conditions. The bacterial systems trialed in this project are C41 (DE3) and Origami 2 strains from Lucigen and EMD Biosciences, respectively. These cell strains have to be first made competent in order to be used for expression. The protocol for making competent cells is in appendix A.6.

In this project, two types of expressions are carried out: expression in 2% Luria-Bertani (LB) broth and expression in minimal M9 media. Expression in LB provides the bacteria with all the nutrients it needs for growth, hence, maximum yield can be expected from expression in LB. LB is notably the most commonly used expression media for that reason. The p8 obtained from LB expression can be used for ^{1}H NMR. However, the aim of this project is to solve the p8 structure in a lipid environment using NMR and in order to effectively use NMR, well resolved spectra are required. Therefore, N^{15} NMR has to be performed in conjunction with ^{1}H NMR because this type of 2D experiment allows for a better spread of peaks. Expression of protein in minimal M9 media is essential if the protein is to be labeled with ^{15}N or ^{13}C or if it is to be selectively labeled for ^{15}N or ^{13}C NMR experiments. M9 contains only the minimum nutrients needed by *E. coli* to grow and the LB component is removed. The sources of carbon and nitrogen come from dextrose and ammonium chloride respectively. The carbon and nitrogen sources can be replaced with isotopically labeled sources.

The cell strains used in this project (Origami and C41) were also given great consideration. Origami cells are an oxidizing strain of *E. coli* cells. They are *trxB gor* deletion mutants and so contain thioreodoxin and gluthathione in their oxidized form which results in the accumulation of active oxidized alkaline phosphatase. Xiong *et al* [42] have shown that the heterologous proteins expressed in Origami cells appear in soluble fractions instead of in inclusion bodies. Expression of membrane proteins in soluble form, rather than inclusion bodies facilitates purification.

Upon expression, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is used to analyze the protein expression. A stacking gel layer and a separating gel layer are made using recipes found in appendix A.7. The running Buffer for the SDS gels is Tris/Glycine: 12.14 g Tris, 57.6 g Glycine and 4 g SDS in 4 L deionised water. To visualize proteins of a lower molecular weight, a Tris/Tricine gel can be run. The running buffer is composed of a cathode buffer and an anode buffer. The cathode buffer is 0.1 M tris, 0.1 M tricine and 0.1 % SDS. X5 Anode buffer is 0.2 M tris-Cl, pH 8.9. Gels are stained and visualized by following the protocol detailed in appendix A.8.

2.2.5.1 Protein expression- in 2% LB

The temperature conditions tested for C41 and Origami strains are 37 ^{o}C, 18 ^{o}C, 30 ^{o}C and room temperature. 37 ^{o}C is the recommended temperature for *E. coli* growth. 18 ^{o}C has been shown by Tait *et al* [34] to result in a slightly reduced protein yield but greatly decreased protein degradation.

The protocol for expression in 2% LB is as listed below:

1) A single colony is inoculated into 5 ml 2 % LB Broth with 0.05 mg/ml carbenicillin (CBC).
2) The 5 ml culture is incubated at 37 ^{o}C, shaken at 225 revolutions per minute (r.p.m.) for 5 hours.
3) A 100 ml 2 % LB Broth with 0.1 mg/ml ampicillin is inoculated with 1ml of the 5 ml culture.
4) The 100 ml culture is incubated at 37 ^{o}C, shaken at 225 r.p.m. for 15 hours.
5) 25 ml of the 100 ml overnight culture is centrifuged at 3750 r.p.m, 4 ^{o}C for 10 min.
6) The supernatant is discarded and the cell pellet is resuspended in 10 ml LB which is used to inoculate 1 L 2 % LB Broth with 0.1 mg/ml ampicillin.
7) The 1 L culture is incubated at 37 ^{o}C and shaken at 225 r.p.m. until *E. coli* reaches the log phase (OD_{600} between 0.5 and 0.7)
8) The cells are induced with 1 mM isopropyl-beta-D-thiogalactopyranosid (IPTG).
9) The cells are grown for an additional 3-4 hrs at 37 ^{o}C.
10) The culture is harvested by centrifuging the cells at 5000 r.p.m. at 4 ^{o}C for 15 min.
11) The supernatant is removed and the cell pellet is stored at -80 ^{o}C.

Cells to be grown at 18 ^{o}C are put in water bath at 4 ^{o}C for 30 min after the OD_{600} reaches between 0.5 and 0.7. The cells are then induced with 1 mM IPTG and allowed to grow overnight at 18 ^{o}C and harvested.

2.2.5.2 Protein expression- in M9

E. coli grow slower and protein production is reduced when grown in M9 media. As a result, the protein production in *E. coli* grown in M9 media might be dramatically different

than protein production in *E. coli* grown in LB. Optimizations have to be explored. Marley *et al* [43] used Basal Vitamins Eagle Media to increase the yield of *E. coli* growth in M9 media. Ferric citrate was shown by Hartmann *et al* to increase the growth rate of *E. coli* [44]. The effects of X1000 MEM vitamin solution (similar to Basal Vitamins Eagle Media), thiamine and ferric citrate on p8 expression were investigated.

The recipe for M9 is as follows:

1) 200 ml autoclaved X5 M9 salts (14 g dibasic sodium phosphate, 3 g monobasic potassium phosphate, 0.5 g sodium chloride, 1 g ammonium chloride)
2) 750 ml autoclaved deionised water
3) 10 g dextrose, 0.241 g magnesium sulphate, 0.088 g ferric citrate and 0.1 g thiamine filter sterilized in 40 ml deionised water
4) 1 ml calcium chloride
5) 10 ml X1000 MEM Vitamin solution (Invitrogen)

Upon induction, an extra 0.05 g/L thiamine is added to the cell culture. For ^{15}N protein labeling, ammonium chloride is replaced with isotopically labeled ammonium chloride (Cambridge Isotopes).

2.2.6 Protein purification

Protein over-expressed from a bacterial system has to be purified due to the extensive array of host proteins expressed along with the desired protein. Specific to this project, the p8 protein with its fusion protein has to be purified from this protein mixture. In order to purify p8, however, the *E. coli* cells containing the p8 protein have to be lysed using a french press.

The detailed protocol for cell lysis is given in appendix A.9. The cell lysate can then undergo column purification.

The fusion protein, maltose binding protein, is used as solubility enhancer in p8 expression but its affinity for amylose can be exploited by using amylose beads to purify the p8 protein. The use of amylose beads for purification, however, is not efficient because amylose beads are unstable, expensive and can only be used a few times [41]. Also, interaction between MBP and amylose might not be strong enough in some cases to be used as the basis for purification [41]. When an amylose bead column was used to purify MBP-p8 complex from the cell protein mixture in this project, it was not successful (results not shown).

The hexa-histidine linker between MBP and p8 can also be used for purification. The hexa-histidine linker has an affinity for nickel and so, a nickel-NTA column, also known as a his-trap column, can used for purification of the MBP-p8. The cell lysate (identified as the supernatant) from lysis of *E. coli* cells expressing MBP-p8 was loaded onto a his-trap column and the MBP-p8 complex attached to the nickel beads due to the hexa-histidine tag. The column was washed three times with twice the column volume of binding buffer (0.5M sodium chloride, 20mM potassium phosphate dibasic, 10mM imidazole, 0.5% triton X, pH 7.4) which washes the column of proteins that do not bind to the beads (forming the binding fractions, B_1, B_2 and B_3). The MBP-p8 bound to the nickel beads was then be competitively displaced by washing the column with a solution of high imidazole concentration (forming the elution fraction). The imidazole will compete with and replace the hexa-histidine tag from the beads. There may be disadvantages in using the hexa-histidine tag as a means of purifying the protein. In some instances, the hexa-histidine tag might not be exposed and so

will not interact with the nickel beads. A complete protocol for nickel column purification is detailed in appendix A.10.

2.2.7 Thrombin digest

It has to be clear that at this point, the protein that is expressed and purified is the MBP-p8 complex. MBP, as already stated earlier, is encoded in the pMAL plasmid. MBP is linked to p8 via a flexible hexa-histidine tag. To separate the passenger p8 protein from the hexa-histidine tag and MBP, a thrombin cut site (amino sequence LVPRGS) has been designed to precede the p8 DNA sequence. The thrombin cleaves after the proline-arginine amino acids if they are exposed to thrombin in the three dimensional structure. Figure 2.3 highlights the different parts that make up the MBP-p8 that is expressed, as well as the site of thrombin digest.

Thrombin digest is performed by adding 100U of thrombin to the purified protein mixture from the his-trap and the digestion reaction is allowed to continue for 18 hours at room temperature while shaking. The digestion mixture is stopped by adding p-aminobenzamidine-agarose to the digestion mixture. Appendix A.11 has a more detailed protocol.

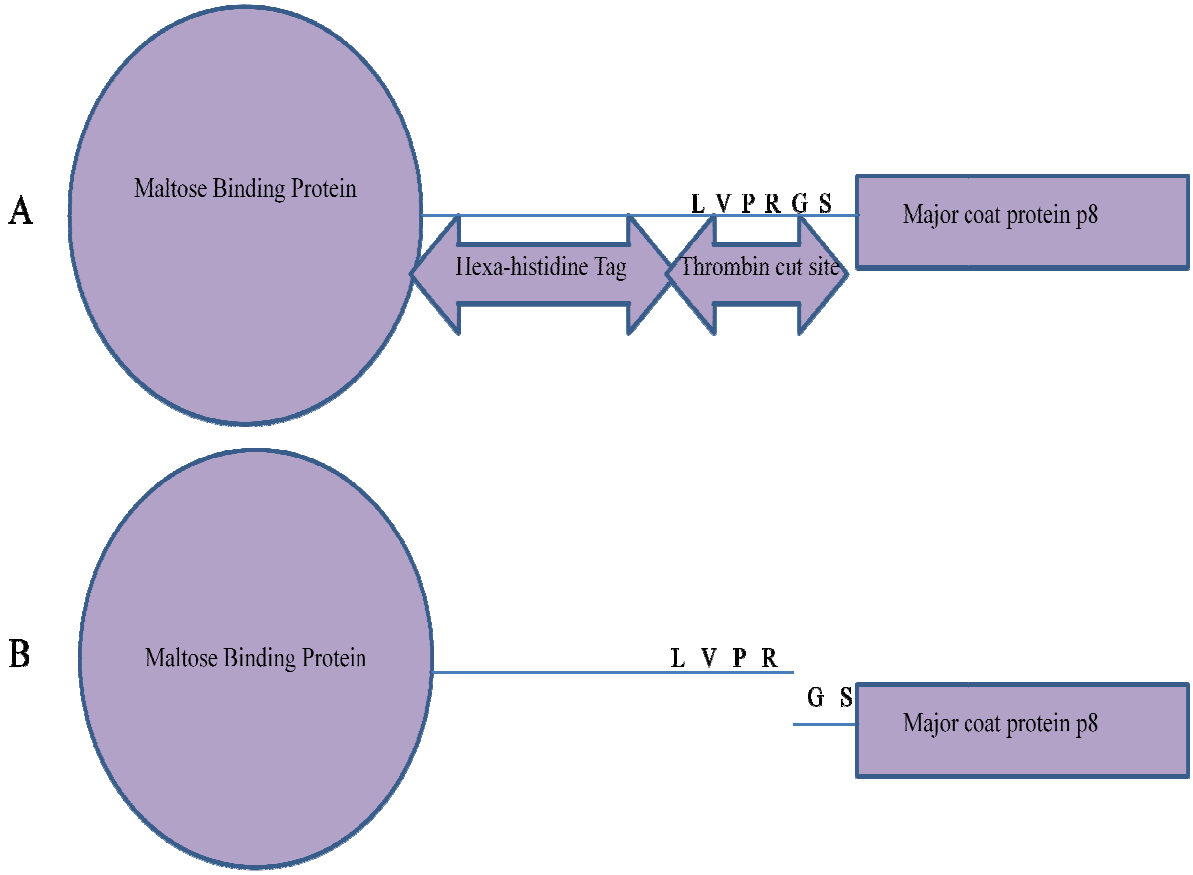

Figure 2.3- Cartoon representation of the MBP-p8 complex and the effect of thrombin digestion. The MBP-p8 complex (A) has a molecular weight of 50.66 kDa. Note that the p8 protein after thrombin digestion contains the glycine-serine amino acids from the LVPRGS thrombin cut site because thrombin cuts after the arginine. The extra G-S residues adds to the predicted molecular weight of the isolated p8 which becomes 5.96kDa. The MBP containing a fragment of the thrombin cut site in B has a molecular weight of 44.7 kDa.

2.2.8 Complete isolation of p8

Once the MBP-p8 fusion complex is digested, both p8 and MBP exist in the protein mixture. Separation of p8 from MBP is crucial in obtaining pure p8. Complete purification of p8 can be done via either high pressure liquid chromatography (HPLC) or Q-sepharose column purification.

2.2.8.1 High pressure liquid chromatography (HPLC)

One technique commonly used to isolate a specific protein in a mixture of proteins is HPLC. p8 is a membrane protein and so can be expected to be more non-polar than MBP which is a soluble protein. So separation of p8 from MBP can be achieved by reverse phase HPLC where solvent of decreasing polarity is used to elute the proteins in a protein mixture. The protein solution that is loaded on the HPLC column consists mainly of MBP, p8 and incomplete MBP-p8 fusion complexes.

In order to perform HPLC, the protein mixture, post thrombin digest, has to undergo acetone precipitation, extensive dialysis and eventually be lyophilized to obtain a powder protein sample. The details of this protocol are in appendix A.12. 50 mg lyophilized protein sample is dissolved in 2.5 ml trifluoroethanol, and then sonicated for 30 minutes. 2.5 ml of Buffer A (90 % water, 10 % acetonitrile, 0.1 % Trifluroacetic acid) is then added and sonicated for 25 minutes. 2 drops of hexafluoroisopropanol is added and sonicated for another 10 minutes. The sample is then filtered for HPLC.

HPLC was done using a Waters 600 with Waters 2996 Photodiode Array Detector for 229 nm UV detection. The HPLC column is the Jupiter 10u C4 300A column from Phenomenex with C4 beads. The gradient is listed in Table 2.3 with Buffer and Buffer B (10 % water, 90 % acetonitrile, 0.1 % Trifluroacetic acid). HPLC peaks are collected and lyophilized for 2 days and sent o mass spectrometry facility in the Department of Chemistry to confirm that p8 is obtained in pure form (purity about 95.4%).

Table 2.3- The gradient flow of buffer A and B over the HPLC run time of 80minutes. Note that the gradient curve 6 refers to a curve for the gradient change in solvent composition running through the column, as provided by Waters.

	Time(min)	Flow(ml/min)	% A	% B	Curve
1		10.00	100.0	0.0	
2	5.00	10.00	100.0	0.0	6
3	10.00	10.00	25.0	75.0	6
4	75.00	10.00	0.0	100.0	6
5	80.00	10.00	100.0	0.0	6

2.2.8.2 Q sepharose column purification

An alternative to HPLC purification is the use of a double column purification protocol where the protein mixture is run through a nickel-NTA column and a Q sepharose column sequentially. When purified using this method, the protein mixture might not be as pure as when purified by HPLC.

When the protein mixture, post thrombin digest, is loaded onto a his-trap column, MBP will separate from p8 due to the fact that the MBP protein contains the hexa-histidine tag after thrombin digest (see Figure 2.3). The hexa-histidine tag has a high affinity for the nickel beads in the his-trap and so MBP will attach to the nickel beads while p8 would flow through the column (forming the sample fraction, H_s). MBP would eventually get displaced from the beads when a solution with high imidazole concentration is washed through the column (forming the elution fraction, H_e).

Q sepharose is as an anion exchange column. MBP has an isoelectric point of about 5.0 and p8 has an isoelectric point of 9.5. This makes MBP negatively charged when in the dialysis buffer of pH 7.4 while p8 is positively charged. Hence, when the protein mixture

containing both MBP and p8 is loaded onto the Q sepharose column, p8 would not attach to the Q sepharose beads and would flow through the column (forming the sample fraction, Q_s) while the MBP and other impurities would attach to the column and eventually get eluted by washing the column with a solution containing a high salt content (forming the elution fraction, Q_e). A complete protocol for the double column purification of p8 is in appendix A.13.Double column purification

2.2.9 Cysteine mutant p8

Stopar *et al* [30] have conducted experiments testing the membrane anchoring ability of the major coat protein from M13 phage. In the N terminal membrane interface, aromatic charged residues Glu20, Tyr21, Tyr24 and Trp 26 anchor the protein; in the C terminus, interfacial residues Phe42, Phe45, Lys40, Lys43, Lys44 anchor the protein. Stopar *et al* set out to investigate which terminal end contributed more to M13 anchoring. The threonine 46 residue (representative of the C terminus) is mutated to a cysteine residue and a BADAN label is attached to the cysteine residue via its thiol group. Figure 2.4 shows the chemical structure of BADAN.

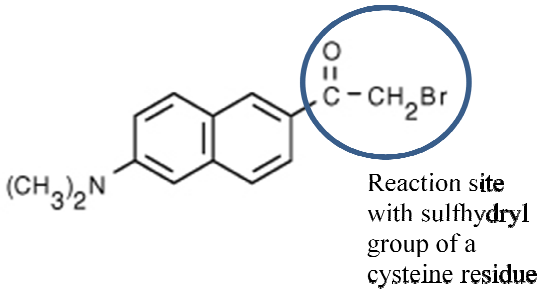

Figure 2.4- Chemical structure of BADAN fluorescence probe. The reaction site is highlighted.

BADAN is sensitive to its environment and will display a shift in fluorescence maximum depending on the environment it is in. In going from a polar environment to a non-polar environment such as that of the lipid bilayer, the maximum fluorescence is blue shifted [30]. For M13, the fluorescence of the intrinsic tryptophan residue 26 and of a BADAN labeled cysteine mutant at position 22 (representative of the N terminus) was also measured [30]. When this protein is reconstituted into lipid bilayers of different thickness, Stopar *et al* showed that the fluorescence of tryptophan 26 showed a blue shift of 700cm^{-1} in going from 14:1phosphatidylcholine (PC) to a thicker 22:1 PC bilayer [30]. The BADAN label at position 22 showed a blue shift of 350cm^{-1}. The BADAN label on residue 46 however, showed only about a 70cm^{-1} shift [30]. This means that residues 26 and 22 had moved into the non polar environment of the lipid bilayer with increasing the membrane thickness whereas residue 46 remained in a polar environment and did not shift into the lipid bilayer as much as residue 26 did [30]. The ability of a quencher to access the tryptophan and the BADAN label was also measured. The quencher showed decreasing ability to access and quench residue 26 while the quenching of residue 46 remained unchanged as the lipid bilayer thickness was increased [30].

A similar approach is followed here to measure the anchoring of B5 major coat protein in membrane. Analysis of the sequence similarities between the p8 major coat protein

(AF428260.5), M13 major coat protein (CAA23861) and Pf3 major coat protein (AAA88387) is done using Geneious Pro 5.3.4 (see Figure 2.5). Sequence alignment of the three sequences shows that the residues lines up well in terms of hydrophobicity and sequences have similar transmembrane segments. Site directed mutagenesis is used to create the cysteine mutant at amino acid residue 32 which line up well with the isoleucine 22 which resides in the N terminus anchoring region of M13.

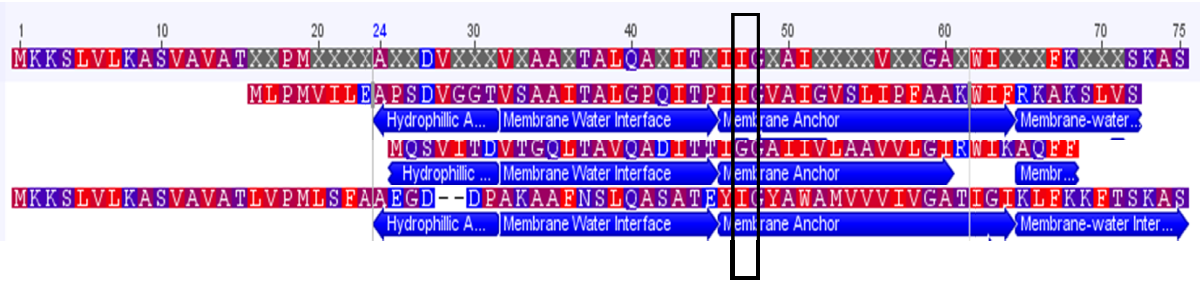

Figure 2.5- Sequence alignment of B5, Pf3 and M13 (from top to bottom). Note the top sequence is the consensus sequence of all three sequences. The domains labeled are with reference to M13 whose domains have been identified by Vos *et al* [25]. It should be noted that M13 coat protein is synthesized with a leader sequence, the first 23 amino acids of its sequence, which is cleaved off before insertion into membrane. The box highlights the M13 residue mutated for BADAN labeling by Stopar *et al* [30], and corresponding residues for B5 and Pf3, identified through alignment. The alignment is done using Geneious Basic 5.3.4 [9].

2.2.9.1 Primer construct

To construct the I32C p8 mutant, site directed mutagenesis is performed on the p8 plasmid using a primer which incorporates the mutation. QuickChange Site Directed Mutagenesis (SDM) System from Stratagene is the commonly used method for introducing a mutation. In QuickChange SDM, the forward and reverse primers are exactly complementary. The primers anneal to the template DNA and replicate the plasmid DNA with the mutation. The newly synthesized plasmid, however, has a nick in the strand [45]. Because of this nick, the newly synthesized plasmid cannot be used as a template for amplification [45]. As a result, the

concentration of plasmid with the mutation that can be obtained is low. Increasing the amount of parental DNA plasmid does not help this problem because there will be a higher amount of hemimethylated DNA which is the methylated parental DNA (which does not contain the mutation) annealed with non methylated newly synthesized DNA (which will contain the mutation) [45]. Hemimethylated DNA is resistant to DpnI digestion and so will be present in the PCR mixture after DpnI digestion.

For SDM of p8, however, a protocol in Liu *et al* was followed [45]. In this protocol, forward and reverse primers are not completely complementary. The primers designed are:

Forward Primer→5': gcc ***gca*** aat cgg cgt gat ctg cgg:3'

Reverse Primer→5': att ***tgc*** ggc gtg gct atc ggt gtt tct ctg: 3'

The 5' end of the primers overlap (about 3-6 nucleotides) and contain the mutation. The primers have an extended non overlapping 3' end. The synthesized plasmid DNA does not contain a nick because the 3' non overlapping region of the primers is made long enough to bridge the nick [45]. The newly synthesized strand can now be used as a template for amplification thus increasing the PCR efficiency. The melting temperature of the non-overlapping sequence of the primers ($T_{m,no}$) is 5-10^{o}C higher than the melting temperature of the overlapping sequence $T_{m,pp}$ (Table 2.4): this allows the primers to bind to the newly synthesized DNA and participate in amplification of the mutated DNA [45]. This type of primer design allows for the use of less parental DNA, thus increasing DpnI digestion efficiency [45]. Also, the problem of the forward primers annealing with the reverse primers, and vice versa, is removed since the primers only have a few complementary nucleotides at the 5' end. Primers are ordered from NAPS at 25 nmole scale.

Table 2.4- Physical properties of the forward and reverse primers designed to introduce a cysteine mutation at residue 32 in the p8 sequence obtained using Net Primer Software available on Premier Biosoft International[3]

	Forward Primer	Reverse Primer
Tm of Primer (^{o}C)	79.23	78.19
Tm overlapping region (^{o}C)	40.60	40.60
Tm non overlapping region (^{o}C)	57.08	57.32
$\Delta G_{formation}$ of hairpin loops (kcal/mol)	-5.01	-1
G+C %	66.6	53.3

2.2.9.2 PCR

The PCR protocol followed for SDM is different from the protocols listed in appendix A.1. A new program, as given by Liu *et al* is used for SDM PCR. Details of the protocol and the PCR reaction mixture compositions are given in appendix B.1.

1 ul of the PCR mixture is transformed into 50 ul of Xl-1 Blue cells by heat shock.

The protocol as detailed in appendix A.5 is performed to extract the plasmid with the I32C p8 sequence. The purified plasmid was sent for sequencing to NAPS.

2.2.9.3 Expression, purification and isolation of I32C p8

Expression of the cysteine mutant of p8 is done in Origami cells using the same expression protocol as the non labeled p8 (section 2.2.5.2). Purification of I32C p8 is done following the same purification protocol as the wild type p8 with the exception of the

[3] http://www.premierbiosoft.com/netprimer/index.html

addition of 1mM dithiothreitol (DTT) to all the purification buffers. DTT is a reducing agent and aids in keeping the thiol group of the cysteine residue in its reduced state. This prevents the formation of intermolecular disulphide bridges which would result in aggregation of p8 and the formation of dimers.

2.2.9.4 Labeling the cysteine residue of p8 with BADAN

The mutant p8 is then BADAN labeled by following the protocol provided by Invitrogen (see section 3.4.1). The labeling is performed in an oxygen free environment using buffers purged with argon. The reaction is also kept in the dark as much as possible.

2.3 Discussion

2.3.1 p8 DNA sequence and incorporation in pMAL plasmid

The p8 insert sequence was successfully designed and PCR was used to prepare the full length p8. The next step which required the ligation of the p8 insert into the vector plasmid through restriction enzyme digestion and ligation proved to be a difficult one.

At first, the protocol followed was such that the p8 sequence was cut with the restriction enzymes from Fermentas sequentially. After each restriction enzyme cutting step, the p8 DNA sequence was run on an agarose gel, it was visualized using UV light and the fragment at the correct molecular weight was cut. The sequence was then extracted using the QIAGEN gel extraction kit. The p8 sequence with the sticky ends was then ligated with the vector plasmid which also has complementary sticky ends. The ligation mixture was then transformed directly into an appropriate *E. coli* strain. A few colonies would grow but upon extraction of the plasmid by QIAGEN plasmid miniprep and sequencing of the plasmid, the

chromatograph would show an extensive amount of mixed signals and the sequence would contain a lot of 'N' nucleotides which meant the sequence was not fully resolved. Resolution of a chromatograph is expected to decrease near the end of the sequence but poor resolution was obtained for the plasmid throughout the chromatograph. Baselines also dominated the chromatograph. One reason could be the low concentration of the plasmid. The concentration of the plasmids sent for sequencing (as measured at OD_{260}) are between 75 ng/µl and 100 ng/µl which is below the desired concentration of 200 ng/µl. Also, the DNA/protein, as measured by OD_{260}/OD_{280} of the samples were between 1.6 and 1.8. The ratio should be 1.8 for a pure DNA but the samples were still in an acceptable range. The presence of contamination in the plasmid sample could also result in bad signal. Common contaminants include salts and nucleases. Contamination could also result from overloading the QIAGEN miniprep column.

A different protocol was implemented and this included the use of Fast Digest Enzymes from Fermentas, ligase from New England Biolabs and a new ligation protocol. The new ligation protocol involved adding both restriction enzymes directly to a mixture of insert and vector. The DNA ligase was then added to the digestion mixture. Multiple inserts could ligate in the vector and the vector could also re-ligate after being cut since all the fragments produced from the restriction enzyme digestion were still present in the ligation mixture and the unwanted fragments were not removed by first running an agarose gel. This new protocol forgoes the agarose gel step and also the subsequent exposure to UV light when cutting the fragments out of the gel. The frequent exposure of the sequence to UV light could cause damage to the DNA which could be an explanation for the failed transformation attempts from the previous protocol. The new protocol resulted in many colonies. The

colonies were tested using colony PCR to ensure that they contained the vector and the correct insert.

Looking at the gel in Figure 2.6A, we can see that 10 colonies were picked for the pMAL-p2X plasmid transformation. Colonies 6 and 7 look like they contain the p8 sequence since the plasmids in these colonies contain about 200 base pairs more than the other colonies. The extra 200 base pairs would correspond to the p8 sequence which has been successfully inserted into the vector. Sequencing results confirmed that colonies 6 and 7 indeed contained the correct insert. The pMAL-c2X plasmids (Figure 2.6B) also produced good colonies upon transformation and 1 colony was obtained with the correct insert.

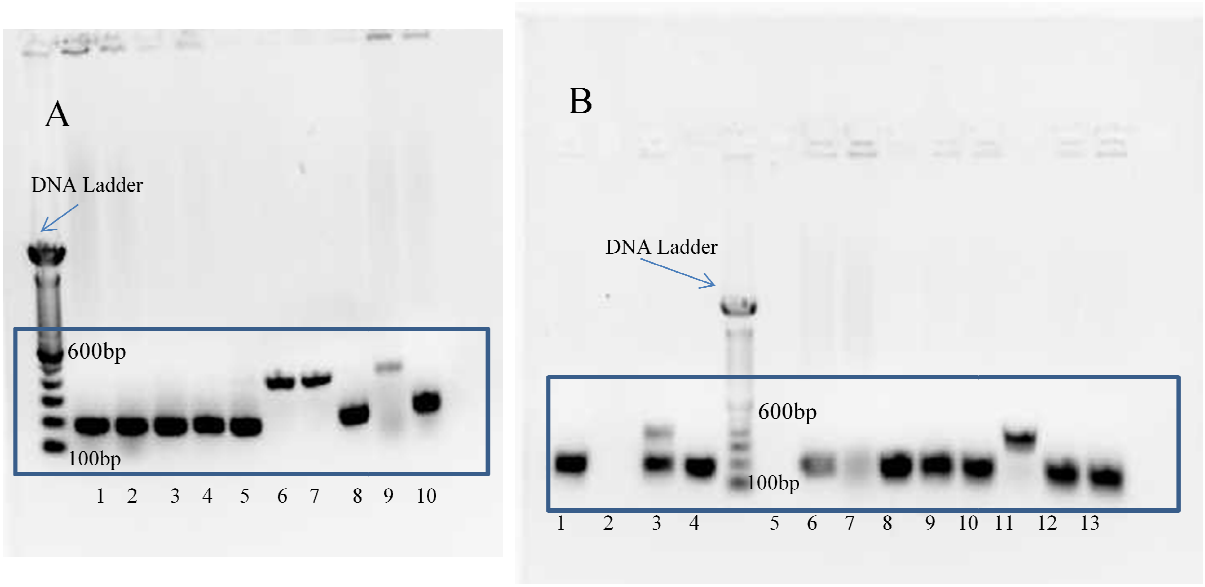

Figure 2.6- Agarose gel showing colony PCR results of SURE cells transformed with p2X plasmid (A) and c2X plasmid (B). Note that the lanes marked 'ladder' represent the100bp DNA ladder. The molecular weight corresponding to the bands of the ladder have been identified as much as possible. The numbers on the other lanes correspond to the colony which was picked for colony PCR.
In A, lanes 6 and 7 (which represent the DNA material extracted from colonies 6 and 7) contain the p2X plasmid with the desired insert. Lanes 6 and 7 can be seen to be about 200 base pairs above the rest of the lanes. The 200 base pairs represent the p8 sequence successfully ligated into the p2X vector plasmid. In B, lane 11 (which represents the DNA material extracted from colony 11) contains the c2X plasmid with the desired insert. Lane 11 can be seen to be about 200base pairs above the rest of the lanes. The 200 base pairs represent the p8 sequence successfully ligated into the c2X vector plasmid.

2.3.2 Protein expression- in 2% LB

The pMAL plasmids containing the p8 sequence insert were transformed into Origami and C41 *E. coli* cells for expression of p8. A few conditions are trialed, as listed in Section 2.2.5.1. These include varying both the plasmid used and the temperature of expression. The two plasmids to be trialed are the pMAL-c2X and pMAL-p2X plasmids.

Expression at 30^{o}C and room temperature did not result in appreciable yield (results not shown). Expression at 18^{o}C and 37^{o}C, however, produced comparable yield. Figure 2.7Figure 2.7 shows two bands visible in the SDS-PAGE: the upper band represents the MBP-p8 complex (50.6kDa) and the lower band represents the Maltose Binding Protein (44.7kDa) without the p8 or with incomplete p8 (refer to Figure 2.3 for a diagram explaining

the MBP-p8 complex). Expression of p8 in pMAL-c2X at 37°C and pMAL-p2X at 18°C in Origami cells yielded thicker bands at 50kDa. The C41 strain-based expressions showed that expressing p8 in pMAL-c2X at 37°C resulted in the optimal expression.

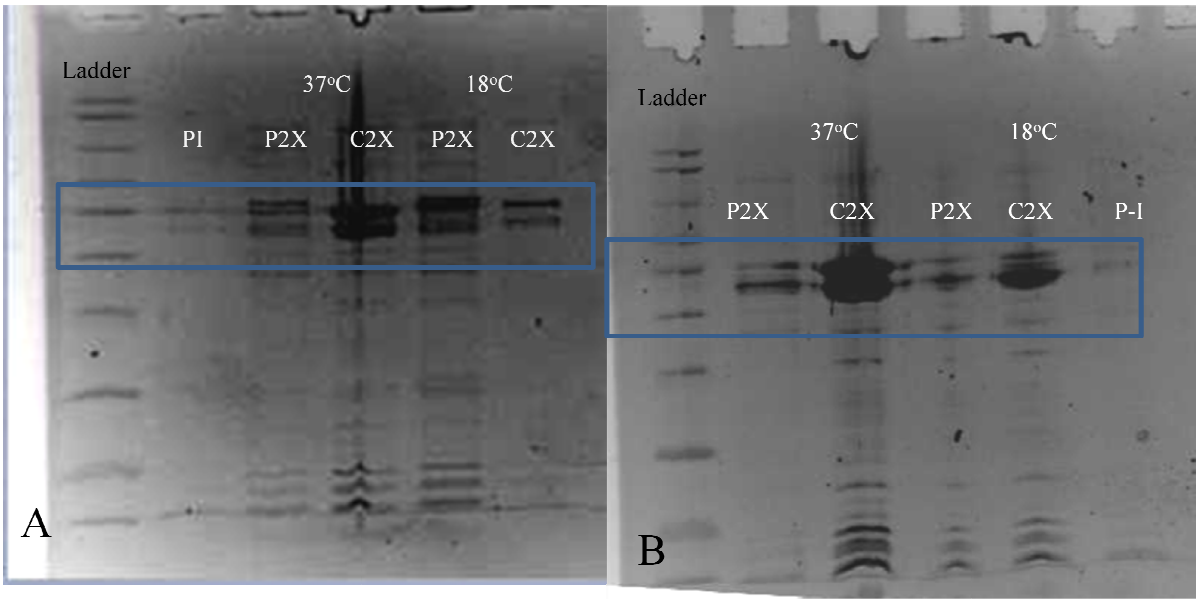

Figure 2.7- Analysis of large scale expression of Origami (A) and C41 (B) E. coli strains using p2X and c2X plasmids, at 37°C and 18°C. Gel is an invision stained SDS gel. The lanes (from left to right) in figure A correspond to Bench Mark His-tagged Ladder, *E. coli* expression pre induction, pMAL-p2X plasmid expression in 37°C, pMAL-c2X plasmid expression in 37°C, pMAL-p2X plasmid expression in 18°C and pMAL-c2X plasmid expression in 18°C. The lanes (from left to right) in figure B correspond to Ladder, pMAL-p2X plasmid expression in 37°C, pMAL-c2X plasmid expression in 37°C, pMAL-p2X plasmid expression in 18°C, pMAL-c2X plasmid expression in 18°C and the pre induced (PI) sample. The blue box highlights the region of the gel which contains the molecular weight range between 40kDaltons(Da) and 50kDa. Note the double bands observed in this region correspond to MBP-p8 complex (50.6 kDa) and MBP (44.7 kDa).

The length of incubation time of *E. coli* cells post induction is also tested for optimal protein expression. The results are shown in the Gel Diagram in Figure 2.8 below. Incubating *E. coli* for 3 hours post induction results in a good protein expression level, although 3-4 hour incubation times show roughly comparable yields.

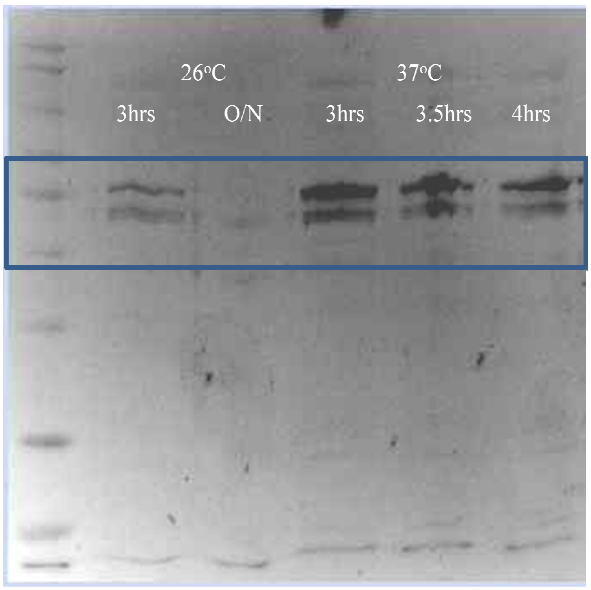

Figure 2.8- Invision stained SDS gel showing the analysis of the effect of post induction growth time on the expression levels of pMAL-c2X Origami cells. The lanes (from left to right) are Mark 12 ladder, expression at 26°C with 3hrs incubation post-induction, expression at 26°C with overnight incubation post-induction, expression at 37°C with 3hrs incubation post-induction, expression at 37°C with 3.5hrs incubation post-induction and expression at 37°C with 4hrs incubation post-induction. Note that 26°C expression is trialed as a possible temperature for *E. coli* expression but as evident in this gel, expression yields are not comparable to 37°C expression and so is not pursued. The blue box highlights the region of the gel which contains the molecular weight range between 55.4 kDa and 36.5 kDa. Note the double bands observed in this region correspond to MBP-p8 complex (50.6 kDa) and MBP (44.7 kDa).

2.3.3 Protein expression- in minimal media M9 for ^{15}N labeling

Expression in LB is useful in producing the optimal protein expression yield because it provides the *E. coli* with all the nutrients it needs to grow and divide. However, the second part of my project involves solving the structure of p8 using a series of ^{1}H and ^{15}N NMR experiments. ^{15}N labeled p8 can be obtained by expressing *E. coli* in M9 media with an ^{15}N source. However, as already mentioned, *E. coli* growth in minimal media is slow and expression yields can be reduced tremendously. A few optimizations relating to M9 expression have been investigated including the addition of X1000 MEM vitamin solution, thiamine and ferric citrate.

MEM added to the M9 solution for C41-C2X expression at 37°C had a positive effect on the growth of *E. coli.* Roughly 2.52±1.42 g of cells per 1L cell culture was produced from

the expression of C41-C2X in M9 media with the addition of MEM vitamin solution. Without MEM vitamin solution, the yield is 1.91±0.40 g. The doubling time of C41 cells grown in M9 without MEM and with MEM are 103 min and 85 min respectively. A summary of expression with and without MEM can be found in Figure 2.9 below.

The addition of ferric citrate also reduced the doubling time of C41 expression from 103 min (C41 cells grown in M9) to 88 min (C41 cells grown in M9 with ferric citrate).

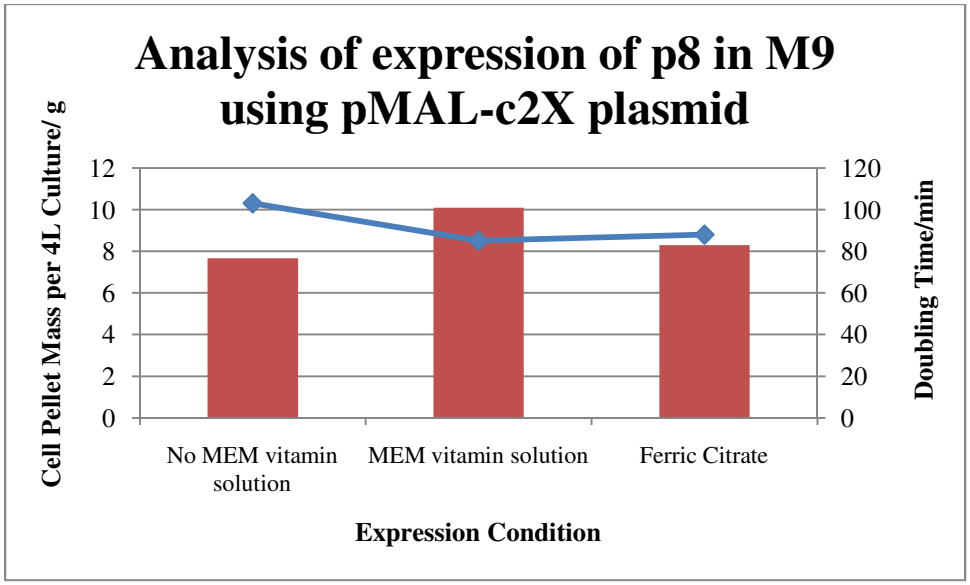

Figure 2.9- Graph comparing the effect of MEM and ferric citrate on C41 expression yields (as represented by cell pellet mass per 4L culture in red) and doubling time (in blue).

When expressing ^{15}N labeled p8, the expression system with the best yield is C41 *E. coli* expression using the pMAL-c2X plasmid at 37°C. 10g of C41 cells is produced from a 4L expression culture. It should be noted that the pMAL-c2X Origami expression system, which works well when the *E. coli* is grown in 2% LB, produces three bands instead of the usual double bands when expressed in M9, as visible by SDS gel analysis in Figure 2.10. The middle band, the thinnest of the bands, is the band representative MBP-p8 but there is a band

at 55.4 kDaltons (kDa) when expressed in M9. One reason for the additional band 55.4 kDa could be because Origami 2 cells are deficient in the gene responsible for coding for the amino acid leucine. They are unable to synthesize the amino acid leucine and so, have to take it up from the media. The Origami cells will not have a source of leucine when grown in M9 minimal media. The upper band of higher molecular weight might come from the leucine residues in p8 sequence being replaced by other available amino acids. An expression was performed with the addition of 0.2 g/l of leucine. This resulted in the band at 55kDa being less prominent but still present.

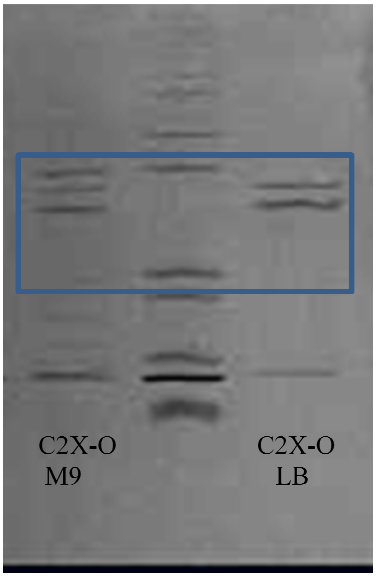

Figure 2.10- SDS gel analysis of p8 purified when expressing pMAL-c2X plasmid in Origami cells at 37°C. The lanes (from left to right) are p8 purified from M9 expression, Mark 12 ladder, p8 purified from LB expression. The upper band (at 55.4kDa) obtained from Origami expression in M9 corresponds to a protein larger than the MBP-p8 protein complex (which corresponds to only 44.7kDa). This protein is not seen in LB expression. The blue box highlights the region of the gel which contains the molecular weight range between 55.4 kDa and 36.5 kDa. Note the double bands observed in this region correspond to MBP-p8 complex (50.6 kDa) and MBP (44.7 kDa).

2.3.4 Protein purification

When purification was first attempted, the pMAL-c2X Origami was expressed at 18^{o}C. This did not result in any visible protein bands when the eluant from the his-trap was analyzed on a SDS gel. pMAL-c2X Origami 18^{o}C expression does produce an MBP-p8 protein band (band at ~50kDa in the supernatant well), but upon purification, the expected double bands were not observed in the eluant fraction.

Changing the expression system to pMAL-c2X Origami at 37^{o}C produces a good yield of p8-MBP protein (22 ±4 mg per 1 L expression culture). There are two thick bands visible for the elution fraction in

Figure 2.11.

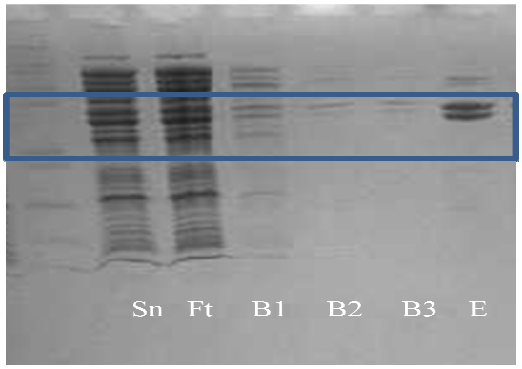

Figure 2.11- SDS-PAGE analysis of the purification of p8 expressed in pMal-c2X Origami *E. coli* cells at 37^{o}C. The lanes (from left to right) are: Mark 12 ladder, supernatant from cell lysis that is loaded onto the his-trap column (Sn), flowthrough as the supernatant is loaded onto the his-trap column (Ft), the first column wash with binding buffer (B_1), the second column wash with binding buffer bind 2 (B_2), the third column wash with binding buffer (B_3), and the elution fraction as the MBP-p8 complex which bound to the beads are eluted using an elution buffer with high imidazole concentration (E). The highlighted box shows the region of the gel which corresponds to the molecular weight range between 55.4 kDa and 36.5 kDa. Note the double bands observed in this region correspond to MBP-p8 complex (50.6 kDa) and MBP (44.7 kDa).

As expected, expression of p8 in M9 had an effect on p8 yields. Looking at Figure 2.12, there is still a band at 50kDa (which corresponds to the MBP-p8 complex) but only 3.2 mg of MBP-p8 is produced from a 1 L of M9 expression culture.

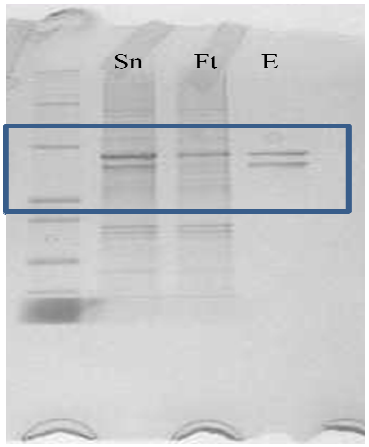

Figure 2.12- SDS-PAGE analysis of the purification of N15 labeled p8 expression in C41 E. coli cells using pMAL-c2X plasmid. The lanes (from left to right) are: Mark 12 ladder, supernatant from cell lysis that is loaded onto the his-trap column (sn), flowthrough as the supernatant is loaded onto the his-trap column (ft), and the elution fraction as the MBP-p8 complex which attached to the beads are eluted using an elution buffer with high imidazole concentration (E). The box highlighted shows the region of the gel which corresponds to the molecular weight range between 55.4 kDa and 36.5 kDa. Note the double bands observed in the elution lane correspond to MBP-p8 complex (50.6 kDa) and MBP (44.7 kDa).

Many conditions were trialed and many optimizations were made during the course of this project. A summary of yields is given in Figure 2.13.

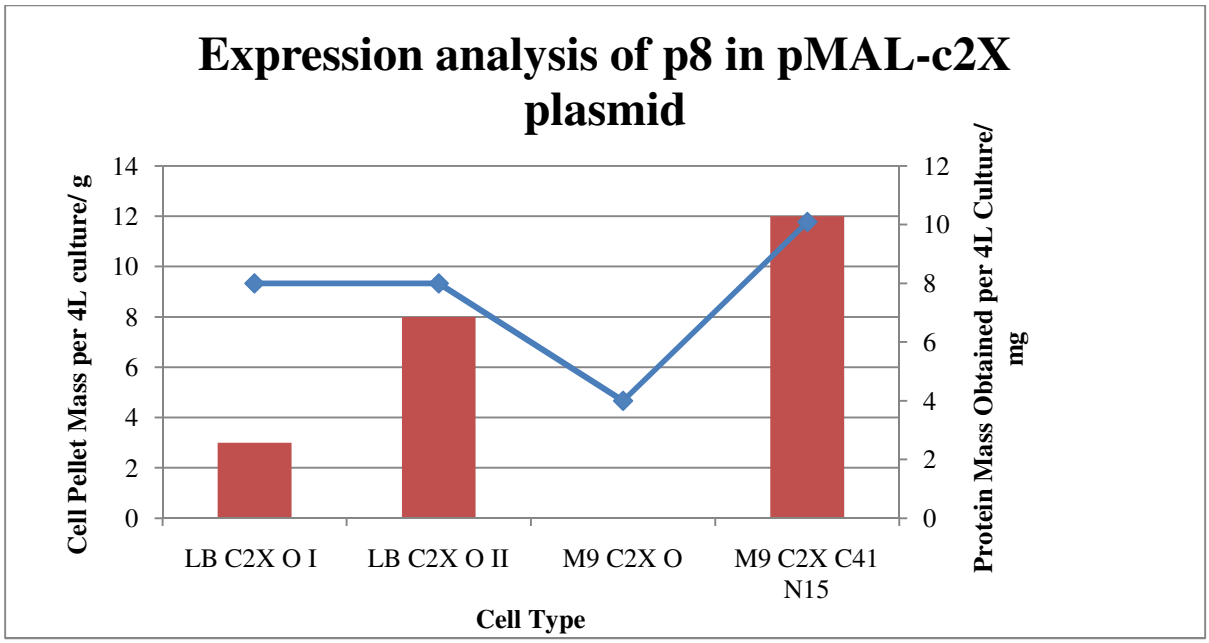

Figure 2.13- Summary of the yields of cell pellet mass (in blue) and protein mass (in red) the different expression systems. Note that M9 C2X O yield for protein mass per 4L culture was 0.05mg. LB C2X O I represents the yield from HPLC purification of the protein expressed in Origami cells, while LB C2X O II represents the yield from Q sepharose purification of the protein expressed in Origami cells. Expressions carried out in M9 are purified using Q Sepharose.

2.3.5 Thrombin digest

Upon thrombin digest, the upper band which represents MBP-p8 will disappear because the p8 is cleaved from the fusion protein (see Figure 2.14). The lower band (maltose binding protein alone) will remain since it is unaffected by thrombin. A band at 5.8 kDa (p8) should be visible upon SDS gel analysis but the band is not visible on the SDS gel at this stage. A mass spectrometry trace for purified p8 post-digestion shown in appendix C.1 confirms the molecular weight and the presence of p8 protein in the protein solution. Figure 2.14 shows the proteins obtained pre- and post-digestion.

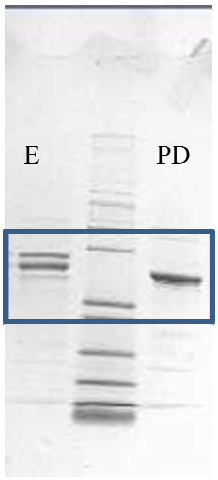

Figure 2.14- SDS-PAGE analysis of the effect of thrombin digestion on the MBP-p8 protein complex. The lanes (from left to right) are the elution fraction from his-trap purification of cell lysate (E), Mark 12 ladder, the sample obtained post digestion (PD). The highlighted box shows the region of the gel which corresponds to the molecular weight range between 55.4 kDa and 36.5 kDa. Note the double bands observed in the elution lane correspond to MBP-p8 complex (50.6 kDa) and MBP (44.7 kDa). The PD sample lane contains a dark band which corresponds to the MBP band.

2.3.6 High pressure liquid chromatography

Under the HPLC conditions outlined in section 2.2.8.1, p8 is eluted between 68-70 minutes. There are a few peaks visible on the chromatograph (Figure 2.15). The most dominant peak is the MBP (~44.7 kDa), followed by a few peaks which represent the MBP with incomplete p8 (~44.6 kDa) and finally the p8 protein is the least dominant peak (5.96 kDa). The yield of p8 protein upon HPLC purification is 3 mg per 4 L expression batch of Origami in 2% LB using pMAL-c2X plasmid. A typical HPLC trace is seen in Figure 2.15.

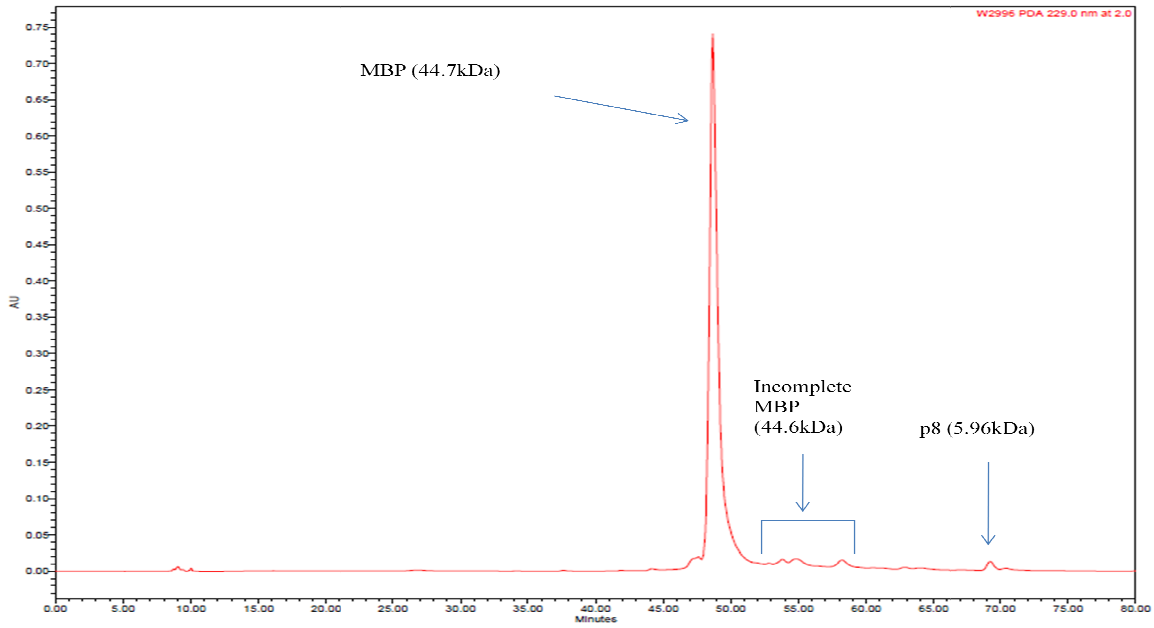

Figure 2.15- A HPLC trace from the purification of the protein solution after thrombin digest. As expected, the solution contains free MBP and p8, along with other partially cleaved products and incomplete MBP which could be a result of degradation. The major coat protein p8 is eluted between 68 and 70 min due to its hydrophobicity compared to the other proteins.

2.3.7 Q-sepharose column purification

When HPLC was used to purify the protein mixture from the M9 expression of C41 cells, the yield was low. 2.5 g of cells from a 1L M9 expression produced 32 mg MBP-p8, resulting in 0.05 mg of pure p8 after HPLC purification. The yield was too low for any further structural studies. The purification of p8 using his-trap and Q-sepharose sequentially was therefore trialed to improve yields.

The protein flow through from nickel-NTA column purification and the subsequent Q sepharose column purification, which corresponds to p8, is shown to be 6 kDa (see appendix C.2 for a mass spectrometry data). This corresponds to the p8 protein with a potassium adduct. Figure 2.16 shows the SDS-PAGE analysis of the nickel-NTA and Q-sepharose column purification of the ^{15}N labeled p8 expressed in M9 using C41 *E. coli* cells and

pMAL-c2X plasmid. The band at 6kDa is not easily visible on the SDS gel but mass spectrometry analysis confirms the identity of the protein in the elution fraction. The yield of p8 protein from 1 L culture of C41 cells purified using Q sepharose column is 3 mg, which compares favorably to the 0.05 mg obtained previously.

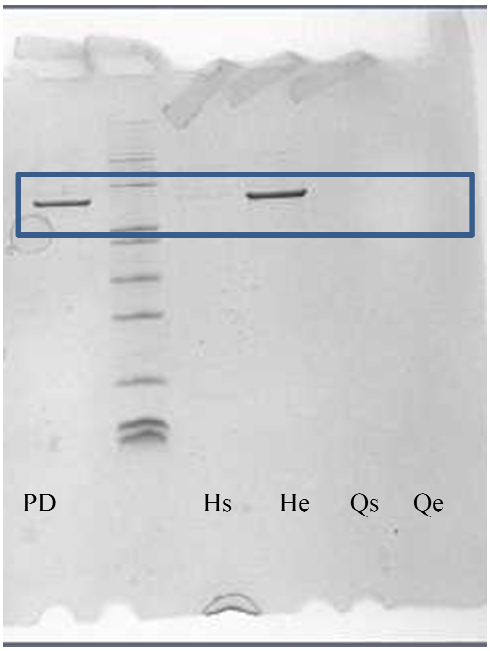

Figure 2.16- SDS-PAGE analysis of nickel-NTA and Q sepharose column purification of ^{15}N labeled p8 expressed in C41 cells using the pMAL-c2X plasmid. The lanes (from left to right) are protein post digestion (PD), Mark 12 ladder, the flowthrough from the his-trap as the post digestion sample is loaded onto the column (H_s), elution from the his-trap column as the MBP bound to the column via its hexa-histidine tag is eluted (H_e), the flowthrough from the Q-sepharose column as the H_s from the histrap is loaded onto the Q-sepharose column (Q_s), elution from the Q-sepharose column as the contaminants bound to the column is eluted (H_e). Note that the band observed in PD and H_e lanes correspond to MBP (45kDa). The Qs sample contains the p8 (not visible but confirmed by mass spectrometry) while Qe contains protein contaminants such as partially degraded p8. The highlighted box shows the region of the gel which corresponds to the molecular weight range between 55.4 kDa and 36.5 kDa.

2.3.8 Cysteine mutant p8

The protocol by Liu *et al* [45] has been successfully followed to produce the I32C p8 mutant. The I32C mutant is expressed in Origami cells using the pMAL-c2X plasmid. The purification of the mutant is done by following the Q-sepharose column purification protocol as wild type p8 except with the addition of 1 mM DTT to all purification solutions. However,

the addition of DTT when purifying the protein mixture obtained from Origami cells resulted in a complication: the distinct double bands obtained upon purification of wild type p8 were not seen. Looking at Figure 2.17, we can see a dominant band at ~45 kDa which represents the MBP but the band above it, the 50 kDa MBP-p8 band is very faint. Substitution of the isoleucine residue with a cysteine should not affect the yield of protein obtained significantly especially since the conditions used were the same. A hypothesis is that the addition of DTT caused refolding of I32C p8 into a conformation where the His-tag was no longer exposed. In this case, the p8-MBP would not bind to the nickel beads. 2% Triton X was added to solve this problem. The binding buffer already contains 0.5% Triton X but bringing the concentration to 2% allowed unfolding of the MBP-p8 complex and exposed the His-tag once again. Figure 2.18 shows that the use of binding buffer with 2% Triton X enabled the I32 MBP-p8 to bind to the beads and be eluted as the characteristic double band once again.

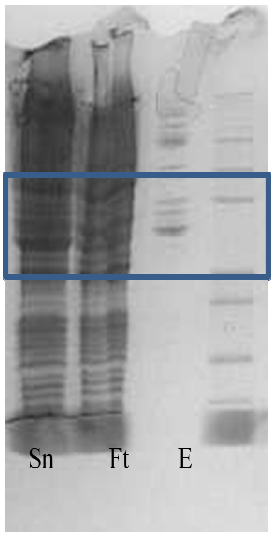

Figure 2.17- SDS-PAGE analysis of the purification of I32C-p8 p8 with the addition of 1 mM DTT to all purification buffers. The lanes (from left to right) are supernatant from cell lysis that is loaded onto the his-trap column (Sn), flowthrough as the supernatant is loaded onto the his-trap column (Ft), and the elution fraction as the MBP-p8 complex which attached to the beads are eluted using an elution buffer with high imidazole concentration (E). It can be seen that in the elution fraction (E), the MBP-p8 band (at 50 kDa) is only a very faint band, compared to purification of the native p8 (see Figure 2.11). The highlighted box shows the region of the gel which corresponds to the molecular weight range between 55.4 kDa and 36.5 kDa.

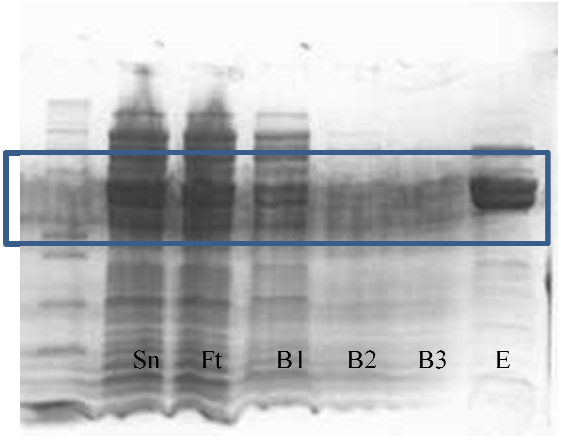

Figure 2.18- SDS-PAGE analysis of purification of I32C p8 using binding buffer with 1mM DTT and 2% triton X. The lanes (from left to right) are: Mark 12 ladder, supernatant from cell lysis that is loaded onto the his-trap column (Sn), flowthrough as the supernatant is loaded onto the his-trap column (Ft), the first column wash with binding buffer (B_1), the second column wash with binding buffer bind 2 (B_2), the third column wash with binding buffer bind 3 (B_3), and the elution fraction as the MBP- I32C p8 complex which bound to the beads are eluted using an elution buffer with high imidazole concentration (E). The highlighted box shows the region of the gel which corresponds to the molecular weight range between 55.4 kDa and 36.5 kDa. Note the double bands observed in this region correspond to MBP-I32C p8 complex (~50.6 kDa) and MBP (44.7 kDa).

The I32C mutant is further purified and isolated using double column purification. 6 mg of I32C p8 was obtained from a 4 L Origami expression using pMAL-c2X plasmid. Appendix C.3 shows the mass spectrometry spectrum obtained for the I32C p8 mutant. The purified I32C p8 was then successfully BADAN labeled.

2.4 Conclusion

The major coat protein of B5 phage is successfully expressed in E. coli. Various optimizations and considerations have been performed to obtain a good yield of p8. The cysteine mutant of p8 (I32C p8) is also successfully engineered, expressed and labeled with BADAN fluorescence probe.

Chapter 3: Structure Characterization of P8

3.1 Introduction

In order to understand how a filamentous phage is assembled, it is important to understand the structures of the different components of phage assembly at each step of the assembly process. As outlined in Chapter 1, assembly involves minor coat proteins p9, p7, p3 and p6 which reside in the host membrane and come together during phage assembly. First, p9 and p7 come together to form the distal end of the virion, which extrudes first from the membrane. Then p8 proteins assemble so that the phage DNA is coated as it extrudes from the host. Finally, p3 and p6 interact to cap the proximal end of the virion.

To dissect this process, it is important to build models of each step. The aim of this thesis, as already discussed in Chapter 1, is to determine the structure of p8 in membranes in order to gain insight into how p8 might interact in the membrane right before virion assembly. Specifically, we are trying to assess whether: 1) p8 adopts an alpha helical structure in a membrane environment; 2) whether it adopts an I or L shape to maximize interactions between the helix and the membrane bilayer; and 3) how membrane composition affects structure.

When building a model, it is best to start with a simplified system. Although B5 filamentous phage assembly obviously involves many players, it is simplest to start with a membrane which contains only the major coat protein. As discussed in Chapter 1, the advantage of studying B5 is that it infects and replicates in the Gram positive bacteria *Propionibacterium freudenreichii*, which is much simpler to model as compared to their Gram negative counterparts.

Gram positive bacteria membranes contain mostly anionic phospholipids such as POPG and cardiolipin [46]. Cheng *et al* [32] have shown that 1:1 POPC/POPG and 1:1 cardio/POPG can be used as models to represent *Staphylococcus aureus* and *S. epidermidis* while 1:1 palmitoyl oleoyl phosphatidylethanolamine (POPE)/POPG is an accurate model for representing *Bacillus cereus*. They showed the importance of using an appropriate membrane model to represent bacterial membranes and that even within the Gram positive bacteria, different model membranes have to be applied. 1,2-di-dodecyl-glycero-3-phosphocholine (DPC) is yet another commonly used lipid for studying proteins using solution state NMR, mainly due to the stability and small size of the micelles formed [47]. The lipid compositions used for structural determination of p8 are 1:1 POPC/POPG, 1:1 POPG/cardio, and DPC micelles. These are the representatives of a Gram positive bacteria membrane which will be used here.

In this Chapter, we will therefore explore the secondary structure of p8 in a variety of membrane environments to see what the effect of composition is, as well as present the results from preliminary experiments using NMR. These experiments will form the basis for future work on the structure of p8 in membrane bilayers. Finally, the results from fluorescence experiments shed light into how p8 is inserted into the membrane.

3.2 Circular dichroism spectroscopy (CD)

CD is done to identify the secondary structure composition of p8. As already discussed, the structure of p8 is very dependent on the environment that it is in and since we want to identify the secondary structure of p8 in its native host membrane environment, lipid micelles which closely represent its native environment are necessary.

CD uses circularly polarized light to measure the differences in absorption between left-handed polarized light and right-handed polarized light which arises due to the optically active, chiral nature of a protein. CD works within a range called the far-UV range (wavelengths 190 nm- 250 nm). In this range, peptide bonds act as chromophores. CD reflects a general secondary structure composition of the protein and cannot be used to determine the contribution of a specific residue to structure.

3.2.1 Sample preparation

The following samples were prepared:

1) 0.25 mg p8 dissolved in 470 µM POPC/POPG in 10 mM tris, pH 7.5. Final protein to lipid ratio being 1:9.44.
2) 0.25 mg p8 dissolved in 7.5 mM DPC in 750 µM phosphate buffer (monosodium phosphate and disodium phosphate), pH 7.5. Final protein to lipid ratio being 1:75.
3) 0.25 mg p8 dissolved in 470 µM POPG/cardiolipin in 10mM tris, pH 7.5. Final protein to lipid ratio being 1:9.44.

In addition to these samples, blanks used for CD were made by resuspending 0.25 mg p8 in 500 µl of 10 mM tris buffer (blank for POPC/POPG) or in 500 µl of 750 µM phosphate buffer (blank for DPC). Likewise, the lipid blanks for CD were made using lipid stocks:

1) 3760 µM 1:1 POPC: POPG in 10 mM tris buffer
2) 15 mM DPC in 750 µM phosphate buffer
3) 3760µM 1:1 POPG: cardio in 10 mM tris buffer

Each of the above samples was prepared three times separately and a CD spectrum was recorded for each of the three.

3.2.2 Experimental

CD was performed on Jasco J-710 spectropolarimeter with a sensitivity of 100 millidegrees, bandwidth of 1.5 nm, response time of 4 s, continuous scanning speed of 50 nm/min and accumulation of 3 scans to increase signal-to-noise ratio. Readings are taken at 30^{o}C for the wavelength range 185 nm to 225 nm.

CDProRun software[4] was used to analyze CD data and predict the structural composition of p8 in lipid micelles. CDProRun software which has the programs CDSSTR (variable selection), CONTIN (ridge regression), and SELCON3 (self consistent method) are used to calculate the secondary structure composition of p8 [48]. CDSSTR, CONTIN and SELCON3 are estimation programs which use a variety of protein databases with known conformation. CONTIN for example uses different datasets and chooses the dataset that gives the best fit.

CDPRoRun program accepts ellipticity values in Delta epsilon (De) units. Ellipticity values are converted to De units using:

$$\text{Ellipticity in De units} = \text{Ellipticity}*((0.1*MW)/ (P*C*3298))$$

where MW is the molecular weight of the protein in Daltons, P is the path length in cm and C is the concentration in mg/ml.

The molar ellipticity per residue is calculated for the readings using the formula:

$$[\theta]_r = ([\theta]_{obs} (MW/10c*d*\text{number of residues})$$

where $[\theta]_{obs}$ is the ellipticity measured experimentally, c is the concentration of the sample in mg/ml and d is the optical path length of the cell in cm.

---

[4] http://lamar.colostate.edu/~sreeram/CDPro/main.html

3.2.3 Results

The CD results show that p8 is mainly alpha helical in all the tested lipid micelles. The major coat proteins of the extensively studied M13, Pf3 and Pf1 phages are also alpha helical, though the helical content varies slightly depending on the lipid environment the coat protein is reconstituted in [15, 28].

The graphs in Figure 3.1 show the mean ellipticity per residue for each of the three lipid compositions in which p8 was reconstituted. All CD spectra show a characteristic graph for an alpha helical protein with minima at around 222 nm and 208 nm, and a maximum at 190 nm.

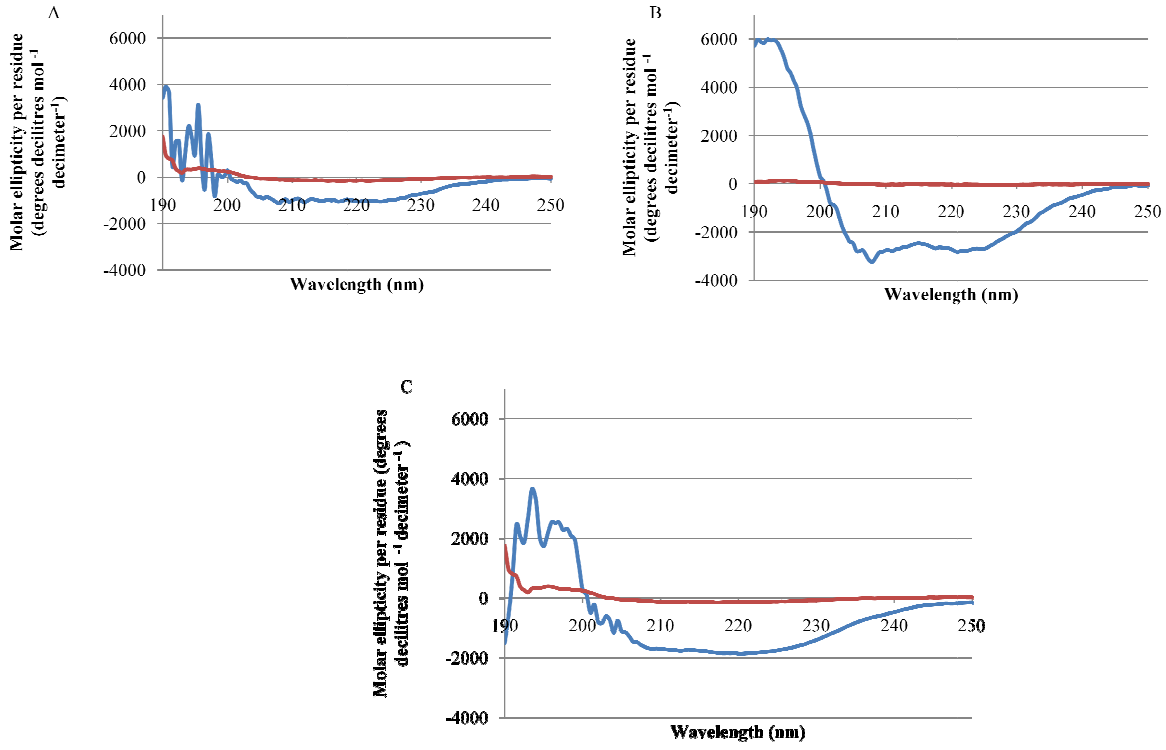

Figure 3.1- Far-UV CD spectra of p8 reconstituted in A) 10mM tris, 470 µM POPC/POPG, B) 10mM phosphate buffer, 7.5 mM DPC in 750 µM phosphate buffer and C) 10mM tris, 470 µM POPG/cardio. CD is performed at 30°C. Note the red plot represents p8 reconstituted in the appropriate buffer and is characteristic of a random coil.

Although all CD traces in Figure 3.1 show that p8 is alpha helical, the degree of alpha helicity is clearly different (from qualitative analysis of the plots). This degree of alpha helicity has to be quantitatively analyzed to obtain the relative secondary composition of p8 in the three lipid compositions. This is done using the programs SELCON3, CDSTRR, and CONTIN which are available through CDProRun software [48]. The relative secondary composition of p8 in the 3 lipid environments are shown in Table 3.1. The errors are calculated using standard deviation of the three CD samples for each lipid environment. It can be seen that p8 is mostly alpha helical. Its alpha helical content is greatest in DPC micelles, followed by POPG/ cardio micelles.

Table 3.1- Secondary structure of p8 in different lipid compositions. Secondary structure composition is calculated using SELCON3, CDSTRR, and CONTIN. Errors are calculated using standard deviations of the three samples ran for each lipid composition. Note that the distorted helix and distorted strand compositions are estimated by considering specific residues in the terminal ends of helices or strands as contributors to distortion.

Lipid	% Regular Helix	% Distorted Helix	% Regular Strand	% Distorted Strand	% Turn	% Random Coil
DPC	60.8±6.0	27.2±2.4	1.0±0.4	2.7±1.6	3.9±0.6	8.3±0.3
POPC/POPG	26.1±1.0	13.1±0.1	9.8±5.5	13.5±4.2	10.2±3.9	27.5±12.1
POPG/Cardio	42.5±3.4	21.4±0.4	5.0±2.1	7.5±0.5	9.1±1.3	14.9±0.3

3.3 H NMR and ^{15}N NMR

When determining the structure of the major coat protein, it is simplest to start by examining its structure in micelles using solution state NMR methods. Although using micelles may force p8 to adopt an unnaturally curved structure (as discussed in section 1.7) and result in a model that is not necessarily biologically relevant, it is nevertheless an important starting point to establish whether p8 is completely helical and to determine its three-dimensional arrangement.

There are two possible approaches when undertaking structure elucidation of a 57 amino acid protein by solution state NMR. The first is to rely solely on ^{1}H NMR to assign spectra and extract structural constraints. In this case, assignments are based on the use of homonuclear NMR experiments such as correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), and nuclear overhauser effect spectroscopy (NOESY) [49-51]. COSY experiments allow ^{1}H nuclei which are three bonds away to be coupled to each other through J^{3} coupling. TOCSY experiments extend the correlation to every ^{1}H nuclei in an amino acid residue. Finally, with the knowledge of the protein primary sequence, NOESY can be used to establish through-space connectivity between residues. If a protein is alpha helical, connectivity between the ^{1}H nuclei of *i* and *i+4* residues can be established because alpha helices typically have geometry of 3.6 residues per turn. Once resonances have been identified in a NOESY spectrum, the intensities or volumes of the peaks are used to extract distance information (because of r^{-6} dependence of the NOE effect, where r is the distance between two ^{1}H nuclei). From a host of distance constraints, the "optimal" three dimensional arrangement of all atoms can be calculated using programs such as xplor, CNS, or a host of other alternatives.

Proteins which are mainly alpha helical typically do not display well resolved homonuclear spectra, even if they are small proteins (e.g. 50 residues or less) because of poor dispersion of resonances. It is therefore necessary to try and improve the resolution of the spectrum by using heteronuclear experiments, such as ^{15}N-^{1}H or ^{13}C-^{1}H experiments. Adding a ^{15}N dimension will often improve the resolution and drastically reduce resonance overlap. In this case, peak assignment is based on a ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) experiment, 3D ^{15}N-edited NOESY-HSQC, 3D HNCA, and/or ^{1}H-^{15}N IPAP-HSQC [52]. Here again, structural constraints are obtained from NOE distances.

In order to perform the NMR experiments listed above, the optimal sample conditions which yield the best resolved spectra have to be found. Temperature, pH, buffer and lipid compositions can all be varied to find the optimal sample conditions which will ultimately be used to get the complete NMR data set required to solve the structure of p8 in micelles.

3.3.1 Sample preparation

When reconstituting p8 in DPC, the ratio of protein to DPC has to be given some thought. Protein to lipid ratios should be kept low in order to reduce the overall size of the micelles formed and also careful consideration of ratios should be made to ensure one protein molecule is incorporated per micelle. Kallick *et al* [47] concluded that if DPC concentrations are too low in the study of the major coat protein structure, there may be some coat proteins in the micelle and some out of the micelle. This would give rise to the slow exchange of proteins between the two environments and cause doubling of peaks in NMR. Kallick *et al* [47] found that a much higher concentration of DPC is needed to obtain good NMR spectra for

fd and Pf1 coat proteins. They suggested that this is because major coat proteins are larger and have some of its residues in the micelle while smaller would exist on the micelle surface.

For ^{1}H NMR, the sample was prepared by combining 3 mg p8 and 21.1 mg deuterated DPC resuspended in 2 ml chloroform and drying this solution into a thin film using nitrogen gas and dried in vacuum overnight. The thin film is resuspended in 550 µl 10 mM phosphate buffer (monosodium phosphate and disodium phosphate) and sonicated for half an hour. 8-10 % v/v of D_2O is added to the sample and the sample is transferred to a 5mm NMR tube. Overall, the sample is composed of 1.0 mM p8 reconstituted in 100 mM DPC in 10mM phosphate buffer with 8-10 % v/v D_2O.

The same protocol is followed for the preparation of ^{15}N p8 sample for HSQC except a few different sample conditions and sample preparation techniques are trialed. The conditions trialed are as listed in Table 3.2. Note that p8 was also reconstituted in 100mM SDS micelles. Solubility under this condition was low and so was not pursued further.

Table 3.2- List of the sample conditions trialed for optimizing ^{15}N NMR of labeled p8 in 100 mM DPC micelles. Note that all solutions appeared clear except for p8 dissolved in MES which was cloudy with large suspensions. Sample 3 was not subjected to NMR.

Sample	1	2	3	4	5	6	7
Buffer	100 mM phosphate	100 mM phosphate	20 mM MES	100 mM phosphate	100 mM phosphate	100 mM phosphate	100 mM phosphate
Protein Concentration	1.0 mM	0.5 mM	0.5 mM	0.5 mM	0.5 mM	0.5 mM	0.5 mM
pH	7.5	7.5	7.5	7.5	7.5	4.5	9.5
Temperature/ ^{o}C	25	45	-	25	45	25	25
Sample Preparation	Protein dissolved in chloroform	Protein dissolved in chloroform	Protein dissolved in chloroform	Protein dissolved directly in buffer	Protein dissolved directly in buffer	Protein dissolved directly in buffer	Protein dissolved directly in buffer

3.3.2 Experimental

NMR experiments were performed on a 500 MHz and 600 MHz Burker spectrometer (McIntosh lab) equipped with a cryoProbe with the help of Mark Okon from McIntosh lab at the chemistry department. The exact experimental details and parameters used are stated in the individual NMR spectrum captions. Data processing was done using NMRViewJ.

3.3.3 Results

For sample 1 (Table 3.2), a TOCSY and NOESY spectrum were recorded. The fingerprint region for each of these spectra is shown in Figure 3.3 and Figure 3.2.

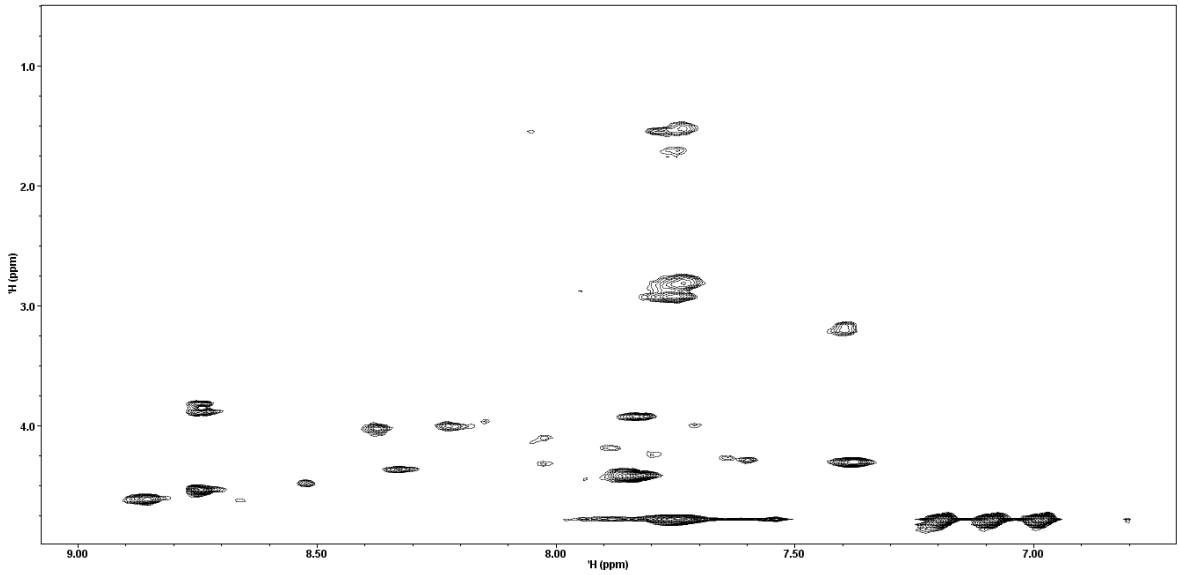

Figure 3.2- TOCSY spectrum of 1 mM p8 (in 100 mM DPC, 10mM phosphate buffer and 10% D_2O, pH 7.4). Experiment was performed at 25°C.

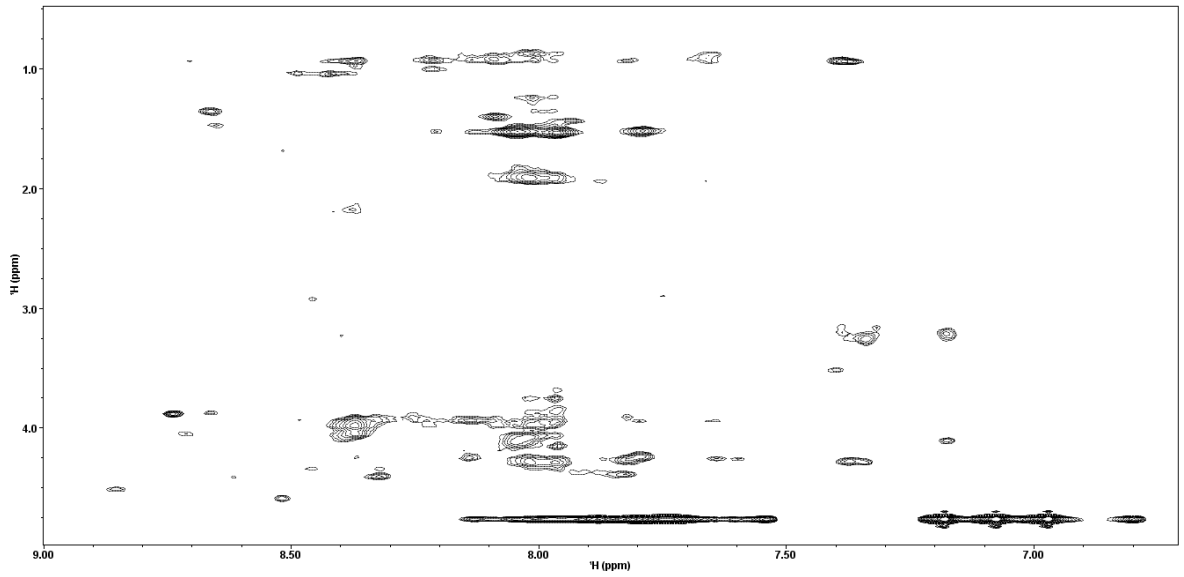

Figure 3.3- NOESY spectrum of 1 mM p8 (in 100 mM DPC, 10mM phosphate buffer and 10% D_2O, pH 7.4). Experiment was performed at 25°C.

Many peaks overlap in the 2D ^1H NMR spectrums of p8 and it is difficult to assign peaks so heteronuclear experiments were tried. However, labeling p8 with ^{15}N was not an easy task, as already explained in section 2.3.3. The yield was tremendously reduced. Because of the small yield of labeled p8 obtained, a few ^1H test experiments were performed to identify the conditions that give the best spread of peaks. pH, protein concentration and temperature are varied.

In addition to changing the sample conditions, different sample preparation techniques were also explored. NMR samples were prepared by dissolving p8 in chloroform and drying with nitrogen to form a thin film of p8 which is then resuspended in an appropriate buffer and lipid. It is possible however that when p8 is directly dissolved in chloroform during sample preparation, it could form something reminiscent of S-form micelles observed when phage particles are exposed to chloroform and form spheroids [53]. These spheroids are known as the S-form and are about 0.4µm in diameter. The phage DNA

sticks out of the sphere and is exposed to the solvent. These S form spheres contain major coat proteins which have reduced alpha helicity (50% instead of close to 100% alpha helicity). If p8 does form these S form spheres during the sample preparation step or simply a smaller aggregate, it would be difficult to break the spheres and reconstitute p8 into lipid micelles for NMR. This would not only mean that the p8 structure obtained from solution state NMR is not of the desired form, it would result in problems with solubility, aggregation and problems associated with the formation of large particles with low tumbling rate. It turns out that forgoing the chloroform step and resuspending p8 directly in buffer and dried lipid resulted in a much more resolved ^{1}H NMR spectra, as shown in the figure that follows.

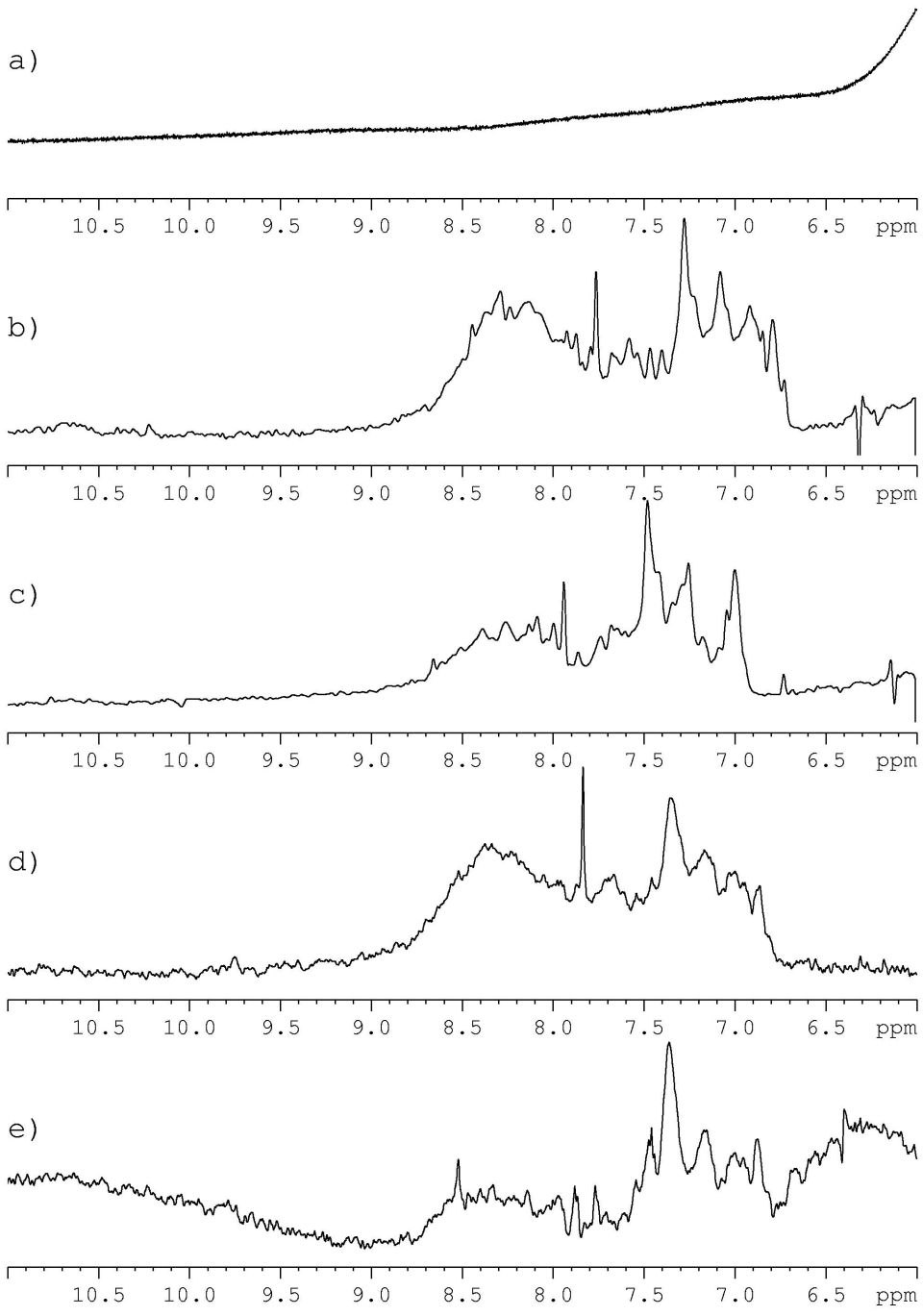

Figure 3.4- ^{1}H NMR spectra obtained using a 500 MHz Burker spectrometer. The protein was dissolved in chloroform during the sample preparation step for a) but dissolved directly in DPC-phosphate solution for the others. Experiment was performed with 0.5 mM p8 in 100 mM DPC, 10mM phosphate buffer and 10% D2O, a) pH 7.5 at 25^{o}C, b) pH 7.5 at 25 ^{o}C, c) pH 7.5 at 45 ^{o}C, d) pH 4.5 at 25 ^{o}C, e) pH 9.5 at 25 ^{o}C.

The spectra a,b,c,d,e correspond to samples 1, 4, 5, 6 and 7 of Table 3.2. The number of scans and spectral sweep was a) 16 scans, 10302 Hz, b) 250 scans, 9615 Hz, c) 1000 scans, 10302 Hz, d) 512 scans, 7978 Hz, e) 512 scans, 7978 Hz.

From the results of ^{1}H NMR on p8 dissolved directly in the DPC-phosphate solution (Figure 3.4b) and, p8 dissolved first in chloroform (Figure 3.4a), it can be concluded that forgoing the chloroform does indeed result in a spectrum with a better resolution of peaks at the H_N region (8.0-11.0 ppm). The sample of temperature 25^{o}C (Figure 3.4b) was seen to produce a slightly better signal than at 45^{o}C (Figure 3.4c). Also the sample at pH 7.5 (Figure 3.4b) produces a better H_N signal than pH 4.5 and 9.5 that were trialed (Figure 3.4d and e).

The next step is to perform HSQC for ^{15}N labeled p8 in DPC micelles using the optimized conditions. 2 HSQC experiments are performed for both types of sample preparation techniques that are discussed. The spectrum in Figure 3.5 is obtained from p8 sample dissolved in chloroform during the preparation step. There is a broad peak between 7.75-8.75 ppm ^{1}H which represents the H_N protons. There are also peaks which represent the side chains. The two peaks on the upper part of the spectrum (^{1}H = 7.0 ppm-8.0 ppm, ^{15}N = 112 ppm) could represent glutamine 27. While the peak at ^{1}H = 8.2-8.8 ppm, ^{15}N = 108 ppm could represent glycine residues. Negative peaks (not shown in Figure 3.5) observed around ^{1}H = 7.6 ppm and ^{15}N = 115 ppm could be from arginine 50. Not too much information can be extracted from the HSQC spectrum in Figure 3.5 due to presence of broad peaks. Broad peaks are characteristic of aggregation. Reducing aggregation will result in narrower, well defined peaks.

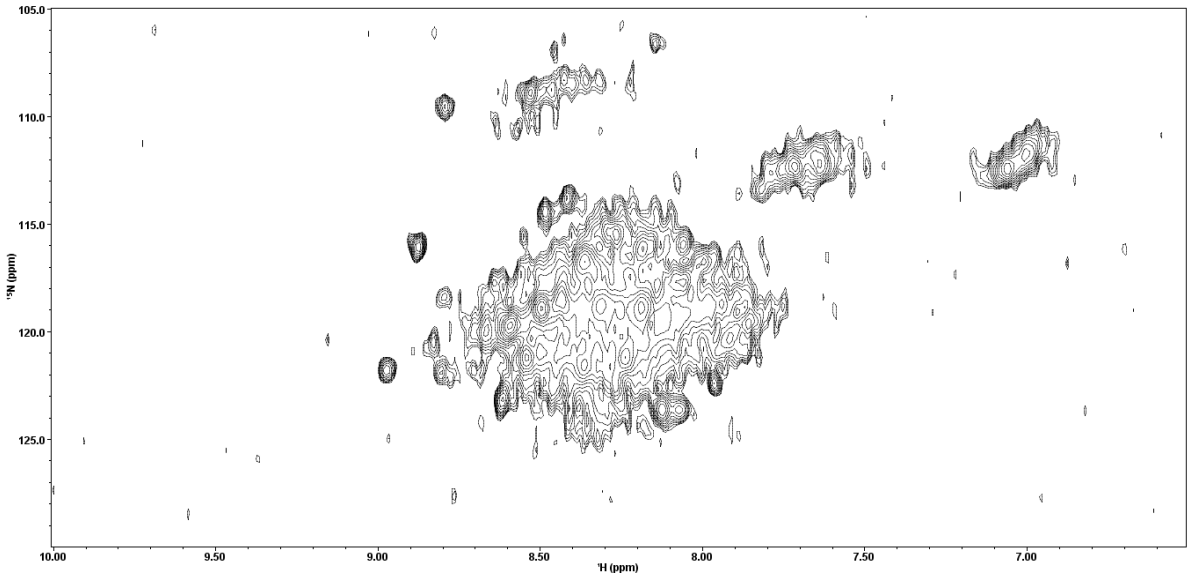

Figure 3.5- HSQC spectrum of 0.5 mM p8 (in 100 mM DPC, 10mM phosphate buffer and10 % D_2O, pH 7.4) obtained using a 500MHz Burker spectrometer with 640 scans at 40^{0}C. The spectral sweep in the direct (^{1}H) dimension is 8012 Hz and 1519 Hz in the indirect (^{15}N) dimension. Note that this sample corresponds to sample 1 in Table 3.2 which is where p8 is dissolved in chloroform during sample preparation.

A second HSQC spectrum is obtained from the second type of p8 preparation where p8 is not dissolved in chloroform. Two peaks are present on the upper part of the spectrum (^{1}H = 7.0 ppm-8.0 ppm, ^{15}N = 112 ppm), similar to the corresponding peaks in Figure 3.5, and could represent glutamine 27. The amide peaks in the second HSQC spectrum, also appear in the range from 7.75-8.75ppm. Overall, the spectrum obtained for the condition which is found to be optimal through ^{1}H NMR trials gave a similar HSQC spectrum to the sample condition initially obtained for the HSQC spectrum in Figure 3.5. This thesis contains preliminary optimizations on HSQC sample conditions for p8 in DPC micelles. Dissolving p8 directly in DPC-phosphate solution and forgoing the chloroform step does improve the sample quality but to improve the resolution of the HSQC spectrum, other optimizations have to be tested (such as those discussed in Chapter 4).

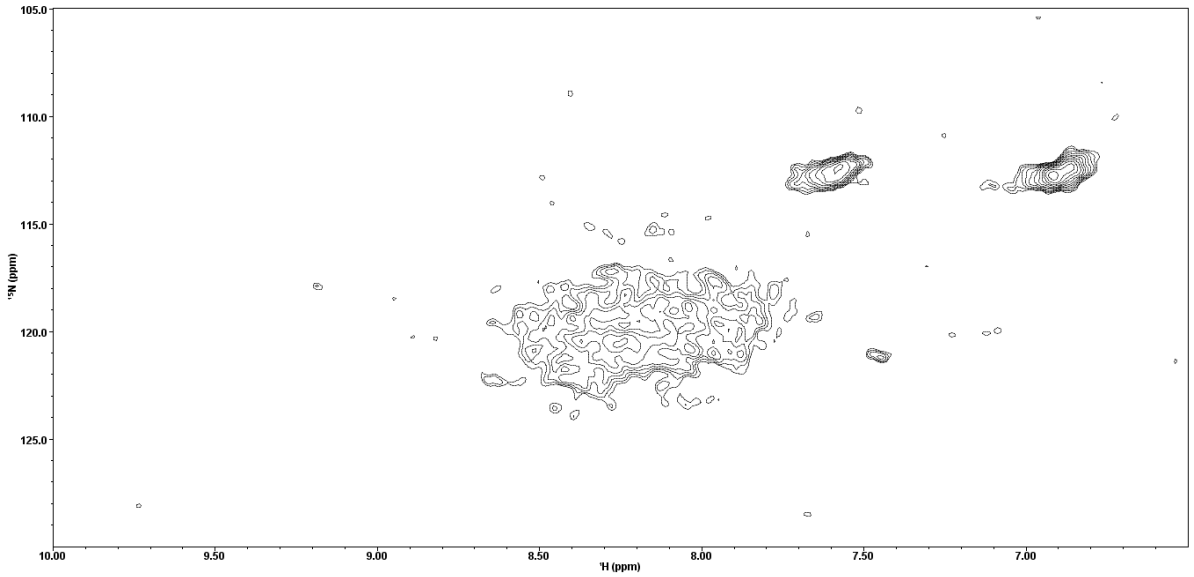

Figure 3.6- HSQC spectrum of 0.5 mM p8 (in 100 mM DPC, 10mM phosphate buffer and10 % D_2O, pH 7.4) obtained using a 600MHz Burker spectrometer with 640 scans at 25^{0}C. The spectral sweep in the direct (^{1}H) dimension is 9615 Hz and 1945 Hz in the indirect (^{15}N) dimension. Note that this sample corresponds to sample 4 in Table 3.2 which is where p8 is dissolved directly in the DPC-phosphate solution during sample preparation.

3.4 BADAN fluorescence

The attachment of the BADAN fluorescence, as already explained in Section 2.2.9, allows the estimation of the location in a membrane bilayer of the residue to which BADAN is attached. The BADAN label is sensitive to its environment and its fluorescence will go through a blue shift upon entering a less polar environment such as that of a membrane.

Once a BADAN label has been attached to the cysteine residue of the I32C p8 mutant, the labeled protein sample is resuspended in POPC/POPG, DPC, POPG/cardio lipid compositions. These lipids are chosen because p8 has already been shown to exist in an alpha helical conformation in these lipid compositions and so will be a relevant model for p8 structural studies.

3.4.1 Sample preparation

BADAN probe is attached to the I32C p8 via its cysteine residue. The steps below are followed for BADAN probe attachment:

1) Dissolve 0.5 mg (100 μM) p8 directly in 10mM phosphate buffer (pH 7.4) or 10mM tris buffer (pH 7.4)
2) Prepare 1 mM BADAN probe stock solution in acetonitrile
3) Add BADAN probe stock solution to p8 in a 1:10 (mole to mole) p8 to BADAN ratio while stirring
4) Let the reaction proceed for 2 hours in the dark at room temperature
5) Dialyze out excess BADAN by dialyzing the mixture for 48 hours against phosphate buffer or tris buffer (as appropriate), changing the buffers every 12 hours.

All steps are performed in the dark and all the buffers used are purged with argon gas for at least 15 minutes. Steps 1-4 are performed in a glove bag charged with argon gas.

The BADAN labeled p8 samples which were resuspended in lipid are prepared as follows:

1) For p8-DPC sample, BADAN-labeled p8 in 10mM phosphate buffer was reconstituted in 20 mM DPC
2) For p8-POPC/POPG sample, BADAN-labeled p8 in 10mM tris buffer was reconstituted in 1.88 mM POPC/POPG.
3) For p8-POPG/cardiolipin sample, BADAN-labeled p8 in 10 mM tris buffer was reconstituted in1.880 mM POPG/cardiolipin.

A buffer blank (the appropriate buffer) and a reagent blank (p8 resuspended in the appropriate buffer and lipid with similar protein to lipid ratios as the labeled sample) were prepared as well. All solutions are sonicated for half an hour.

3.4.2 Experimental

Fluorescence measurements were performed using a Cary Eclipse fluorescence spectrophotometer from Varian. The excitation wavelength was 387 nm and the emission wavelength was scanned in the range of 450 nm and 600 nm with a 5 nm band pass in both the excitation and detection light paths. The maximum emission wavelength was 550 nm for BADAN in an aqueous environment.

The wavelength of maximum absorption can be converted to wavenumbers for comparison with wavenumber values given in Stopar *et al* [30]. The maximum wavelength of absorption (in nm) can be converted into wavenumbers (in cm) by:

Wavenumbers (ν_{max}) = 1/ [wavelength in nm * 10^{-7}]

3.4.3 Results

The resulting emission spectra for I32C BADAN labeled p8 and the emission spectra for BADAN in different solvent environments are illustrated in the figures below.

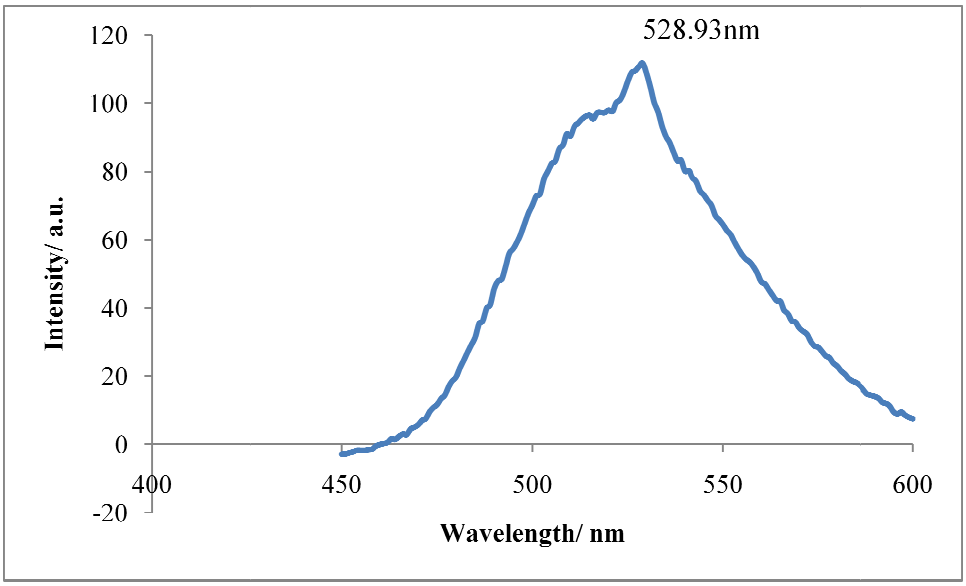

Figure 3.7- Fluorescence emission spectrum for BADAN in I32C p8 reconstituted into DPC at a lipid to protein ratio of 200. The fluorescence maximum is at 528.93 nm and v_{max} is 18.9*10^3 cm^{-1}.

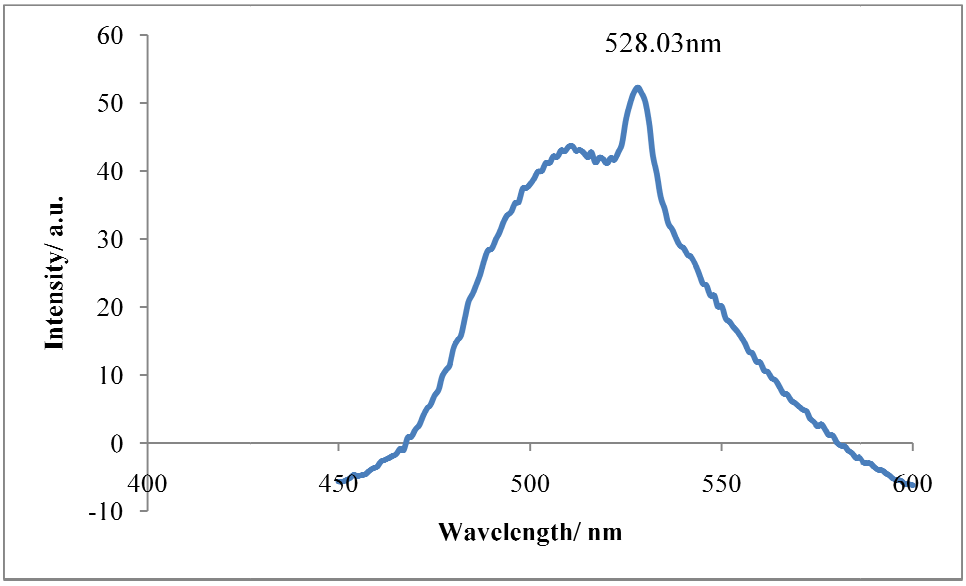

Figure 3.8- Fluorescence emission spectrum for BADAN in I32C p8 reconstituted into POPC/POPG at a lipid to protein ratio of 200. The fluorescence maximum is at 528.03 nm and v_{max} is 18.9 *10^3 cm^{-1}.

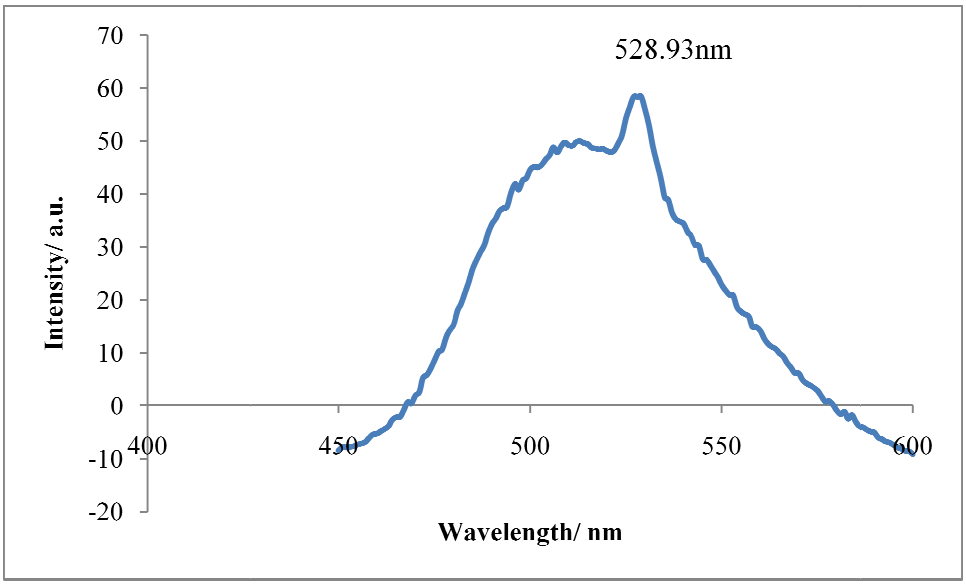

Figure 3.9- Fluorescence emission spectrum for BADAN in I32C p8 reconstituted into POPG/Cardiolipin at a lipid to protein ratio of 200. The fluorescence maximum is at 528.93 nm and v_{max} is 18.9 $*10^{3}$ cm^{-1}.

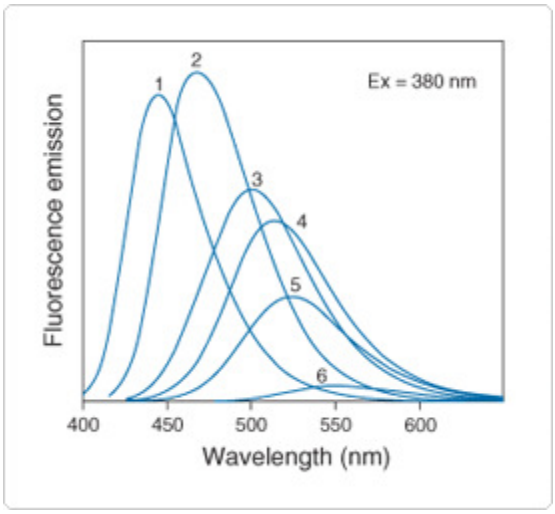

Figure 3.10- Fluorescence spectrum for BADAN in 1) toluene, 2) chloroform, 3) acetonitrile, 4) ethanol, 5) methanol and 6) water taken from Invitrogen Thiol Reactive Probe Handbook[5]. It can be seen that BADAN emission is typically around 550 nm in water. This is taken to be the reference emission wavelength for BADAN in water for this thesis.

[5] http://www.invitrogen.com/site/us/en/home/References/Molecular-Probes-The-Handbook/Thiol-Reactive-Probes/Thiol-Reactive-Probes-Excited-with-Ultraviolet-Light.html

The maximum emission occurs at 528.93 nm for DPC and POPG/ Cardio lipid compositions and at 528.03 nm for POPC/ POPG lipid compositions. These represent a blue shift from the emission maximum of 550 nm for BADAN in an aqueous environment (figure 3.10). This means that BADAN is in a less polar environment. It is therefore plausible to hypothesize that the BADAN label attached to p8 at residue 32 is close to the lipid-water interface (a less polar environment than the aqueous environment).

3.5 Discussion

The secondary and tertiary structure of p8 were investigated in this thesis and although some work is preliminary, the following information has been obtained: 1) secondary structure composition of p8 in micelles; 2) optimal sample conditions (found thus far) for heteronuclear solution state NMR; 3) the relative position of residue 32 of p8 as compared to the equivalent residue 22 in M13.

From CD data, p8 is found to exist in a mostly alpha helical state in all three lipid micelles (DPC, POPC/POPG, POPG/cardiolipin), although the extent of alpha helicity is greatest in DPC (88 % regular alpha helicity). Since the major coat proteins of M13, Pf3 and Pf1are known to be mostly alpha helical, all three lipid compositions used seem to provide the environment necessary for p8 to exist in its native alpha helical structure.

The data from TOCSY and NOESY cannot be utilized alone to solve the structure of p8. There is much resonance overlap. Initial ^{15}N HSQC performed for labeled p8 showed signs of aggregation and different sample conditions were tested in order find an optimal condition which would yield a well resolved HSQC spectrum. Preparing the sample by dissolving p8 directly in the buffer (Figure 3.4b), rather than first in chloroform (Figure

3.4a), was seen to give the best spectrum which had good signal intensity and dispersion in the H_N region of 8.0-11.0 ppm. The following sample condition (Figure 3.4b), was found to be optimal: 25^{o}C, pH 7.5, 0.5 mM p8 in 10 mM phosphate buffer and 100 mM DPC. This condition lead to the best spread of H_N peaks in the region between 8.0 ppm and 11.0 ppm. This condition would most likely yield the best HSQC spectrum for further structural elucidation.

In the optimal condition found for solution state NMR, DPC is the lipid of choice. Although DPC is not an anionic lipid, it is still very commonly used in solution state NMR. One reason is that DPC micelles rotate freely in solution with tumbling time in the range required for NMR. Vinogradova *et al* [54] have shown that detergents with medium acyl chain lengths, such as DPC, provide an environment which allows most proteins to maintain their native conformation. Once an acceptable solution state NMR spectrum is obtained for proteins in DPC, other lipid compositions can be tested. In our case, p8 can be reconstituted in lipids that represent Gram positive membranes such as POPC/POPG and POPG/cardiolipin. These lipids have to be studied using solid-state NMR.

Preliminary fluorescence data was obtained from BADAN labeling of I32C p8 at residue 32. The data showed that residue 32 of p8 is in an environment which is less polar than water (because of the blue shift in fluorescence maxima at around 528 nm instead of 550 nm). The equivalent residue 22 in the major coat protein of M13 is in the lipid-water interface of a membrane and has a blue shift in its fluorescence maxima as well, although it has a greater blue shift with fluorescence maximum of around 508 nm [30]. Thus it can be concluded from the preliminary data that residue 32 of p8 might exist in the lipid-water interface, although more work is needed to calculate the exact depth within a membrane.

It is possible to create a standard set of BADAN fluorescence maxima values when the probe is in different depths of a membrane. These standard values can then be used to predict the depth within a membrane at which a BADAN labeled residue is located. Koehorst *et al* [55] have created such standard values by reconstituting BADAN labeled M13 mutants in 18:1 PC/ 18:1 PG micelles. They found the wavenumber of a completely water exposed BADAN to be 19100 cm^{-1} and the wavenumber of BADAN at the hydrophobic core of a membrane bilayer to be 20500 cm^{-1}. The wavenumbers obtained for BADAN labeled I32C p8 in the three lipid compositions tested are around 18910 cm^{-1}. This number is smaller than the v_{max} for water exposed BADAN probe, as found by Koehorst *et al*. An explanation of this might be that since BADAN is very sensitive to its environment, a standard set of fluorescence values that is specific to the particular lipid composition is needed. In this case, we need to establish a standard set of values for BADAN in different depths of each of DPC, POPC/POPG, and POPG/cardio lipid compositions.

However, the fact that there is a blue shift in maximum fluorescence for BADAN labeled I32C p8 when compared to BADAN maximum fluorescence in water means that the BADAN on I32C p8 is in a more non polar environment such as the lipid-water interface, though this cannot directly be related back to a specific position in the membrane as yet.

Although much work done presented in this thesis is preliminary, the findings reported here set the stage for further work on p8. The secondary structure of p8 in three lipid compositions have been determined, optimal conditions for further solution state NMR studies of p8 have been explored and finally, preliminary BADAN labeling experiments have been used to conclude that residue 32 of p8 might indeed exist in the lipid-water interface of a membrane. The question of I vs L for the structure of the major coat protein still exists and

more work has to be done to solve the structure of p8, the major coat protein of B5 filamentous phage.

Chapter 4: Future Work

Phage assembly is a good system for studying membrane mediated assembly because it takes only a handful of proteins to initiate and complete the assembly process. The dynamic nature of p8, along with the ability of phage structural proteins to not only accumulate around p1 during phage assembly, but also to interact synchronically with each other, contributes to the highly efficient phage assembly process which is capable of producing 4-6 progeny every 5 minutes [56].

The interaction and assembly of p8 is the pivotal step in phage assembly. Studying the structure of p8 in host membrane is fundamental to understanding the phage assembly process better. Much work is still needed in order to explore the structure of p8 in a membrane environment. Even though there is an abundance of biophysical techniques for the study of protein structure, membrane proteins represent less than 1% of protein structures in the Protein Data Bank. This is because numerous optimizations are required to find conditions that agree with a membrane protein's particular physical characteristic, whether it be appropriate buffers for solubilising the protein or appropriate pH conditions for maintaining protein structure.

In this project, p8 has been successfully expressed and purified. Circular dichroism experiments show that the p8 adopts a mainly alpha helical structure in different membrane compositions. Preliminary ^{1}H and ^{15}N NMR experiments including NOESY, TOCSY and HSQC have been performed. We have also identified a good sample preparation technique for p8. To improve the HSQC spectrum further, different buffers, pH and salt conditions should be investigated. Selecting an appropriate buffer for protein NMR is the basis for obtaining a good NMR spectrum. Only phosphate and MES buffer were used in this project

but with the vast amount of knowledge on protein NMR available, a suitable buffer with appropriate salt concentrations which would improve the protein NMR spectrum must exist. The pH of the sample can be adjusted to coincide with the minimum H_N exchange with the solvent which would reduce the line broadening effects of solvent exchange. Once optimal conditions for ^{15}N HSQC have been found, NOESY-HSQC can be done on p8. NOESY-HSQC would give a 3-dimensional spectrum and could potentially be used to solve for p8 structure.

Other structural characterization methods, besides NMR, can be used to solve for p8 structure when in host membrane. Neutron refraction can be performed to give information about p8 structure in membrane which can be used in conjunction with NMR structural data. Neutron refraction gives structures with resolutions of 1.5-2.5Å [57]. Neutrons are refracted by the nuclei, not the electron cloud and so can provide information about hydrogen atom position. Neutron refraction has been used to highlight structural changes in peptide upon going from one environment to another [58]. Also, the structure of p8 in virion has to be solved. The structure of major coat proteins M13, Pf3 and Pf1 in virion have been solved by X-ray crystallography and a similar investigation can be done on the major coat protein of B5.

Preliminary site directed labeling (SDL) experiments have been done on p8 to investigate the local environment of residue 32 of I32C p8 (representative of the N terminal end of p8 transmembrane domain) in DPC, POPC/POPG, POPG/cardio. The data obtained for I32C p8 in DPC adds to the solution state NMR of wild type p8 in DPC micelles. Solid state NMR can be performed to provide structural information about wild type p8 in POPC/POPG and POPG/cardio vesicles or aligned bilayers. In this way, we will have both

fluorescence and NMR data for p8 in the three lipid environments. The immediate environment of the C terminus of the transmembrane domain should also be investigated. A suitable mutation of a residue in the C terminus of p8 should be made and a BADAN labeling should be attached. The anchoring strengths of the C and N terminal membrane-water interface of p8 can then be compared, similar to what is done using M13 by Stopar *et al* [30].

When studying membrane assembly of structures which require multiple copies of a protein to associate amongst themselves, such as in phage assembly where 2700 copies of p8 have to assemble, the question of oligomerization during membrane assembly has to be investigated. p8 is thought to oligomerize even before the assembly process occurs. Nagler *et al* [59] have shown that M13 coat protein formed dimers in the membrane through specific helix-helix interactions. Through single cysteine mutations and analysis of disulphide bond formation, Nagler *et al* concluded that M13 formed close contacts with neighboring M13 monomers through residues 23, 27, 31 and 35 [59]. M13 monomers were thought to assemble into a sheet-like structure within the membrane where a monomer interacts with two neighboring monomers [59]. The sheets of M13 then bind to phage DNA as it is extruded (see Figure 4.1). This allows for minimal change in M13 topology upon extrusion [59]. Oligomerization should be investigated for p8 of B5 phage. Cysteine mutations of p8 can be made to test for close associations between p8 monomers. This will provide some information on the general organization of p8 just before it is assembled into the phage particle.

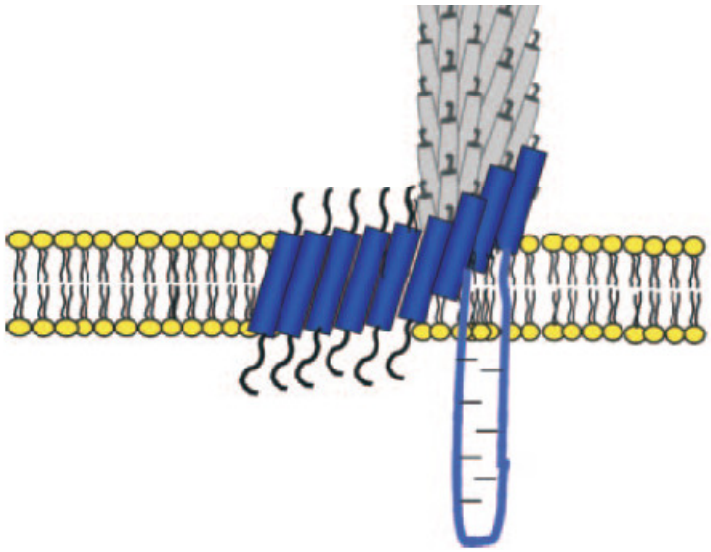

Figure 4.1- Diagram showing a new theory where sheets of mutimeric M13 proteins interact with phage DNA upon extrusion. Adapted from Nagler *et al* [59].

Looking at the bigger picture of phage assembly, there are other structural phage proteins that take part in the assembly process. The minor coat proteins p7, p9, p3 and p6 interact with each other and with p8 during the phage assembly process. The structures of the minor coat proteins in membranes have been well studied and their roles are well defined but their structure in host membrane should be investigated further. The dynamic structure of the phage proteins, when in host membrane, not only drive phage assembly, but also allow stabilizing interactions which enable the phage proteins to remain stable when going from one environment to another during the phage assembly process. These interactions should be investigated further in order to fully understand phage assembly.

Thorough knowledge of all phage protein structures and interactions will allow us to chip away at the mystery of phage assembly.

References

[1] Marks T., Sharp R., *Journal of Chemical Technology and Biotechnology* **2000**, *75*, 6-17.

[2] F., D'hérelle, *C. R. Soc. Biol.* **1918**, *81*, 1160-1162.

[3] Soothill J. S., *Journal of Medical Microbiology* **1992**, *37*, 258-261.

[4] **2010**.

[5] Calendar R., in *Phage Ecology and Phage Evolutionary Biology* (Ed.: Abedon S. T.), **2006**.

[6] Spruijt R., Wolfs C., Hemminga M., in *Encyclopedia of Molecular Biology*, **2002**.

[7] Russel, M., *Molecular Microbiology* **1991**, *5*, 1607-1613.

[8] Chopin M. C., Rouault A., Ehrlich S. D., Gautier M., *JOURNAL OF BACTERIOLOGY* **2002**, *184*, 2030-2033.

[9] Drummond A. J., Ashton B., Buxton S., Cheung M., Cooper A., Duran C., Field M., Heled J., Kearse M., Markowitz S., Moir R., Stones-Havas S., Sturrock S., Thierer T., Wilson A., **2010**.

[10] Marvin D., *Current Option in Structural Biology* **1998**, *8*, 150-158.

[11] Makowski L., Caspar D. L. D, *J. Mol. Biology* **1981**, *145*, 611-617.

[12] Lopez J, Webster R. E., *Virology* **1983**, *127*, 177-193.

[13] Glaser-Wuttke G., Keppne J., Rasched I., *Biochimica et Biophysica Acta* **1989**, *985*, 239-247.

[14] Calendar, Richard, *The bacteriophages, Vol. 2*, Oxford University Press, Oxford, **2006**.

[15] Papavoine C. H. M., Christiaans B. E. C., Folmer, Konings R. N. H., Hilbers C. W., *Journal of Molecular Biology* **1998**, *282*, 401-419.

[16] Karlsson F., Borrebaeck C. A. K., Nilsson N., Malmborg-Hager A., *JOURNAL OF BACTERIOLOGY* **2003**, *185*, 2628–2634.

[17] Webster R.E., *Molecular Microbiology* **1991**.

[18] Ray, Dan S., *Journal of Molecular Biology* **1969**, *43*, 631-643.

[19] Opella S. J., Zeri J.C., Park S. H., *Annual Review Physical Chemistry* **2008**, *59*, 635-657.

[20] Hunter G. J., Rowitch D. H., Perham R. N., *Nature* **1987**, *327*.

[21] Welsh L. C., Symmons M. F., Sturtevant J. M., Marvin D. A., Perham R. N., *Journal of Molecular Biology* **1998**, *283*, 155-177.

[22] Zeri A. C., Mesleh M. F., Nevzorov A. A., Opella S. J., *PNAS* **2003**, 6458–6463.

[23] Marvin D. A., Welsh L. C., Symmons M. F., Scott W. R. P., Straus2, And S. K., *Journal of Molecular Biology* **2006**, *355*, 294-309.

[24] Marvin, D. A., Hale, R. D., Nave, C. & Helmer, Citterich M., *Journal of Molecular Biology* **1994**, *235*, 260-286.

[25] Vos W. L., Nazarov P. V., Koehorst R. B. M., Spruijt R. B., Hemminga M. A., *Cell Press* **2009**.

[26] Nazarovp. V., Koehorst R. B. M., Vos W. L.,Apanasovich V. V.,Hemminga M. A., *Biophysical Journal* **2007**.

[27] Stopar D., Strancar J., Spruijt R. B., Hemminga M. A., *Biophysical Journal* **2006**, *91*, 3341–3348.

[28] Stopar D., Spruijt R. B., Hemminga M. A., *Chemistry and physics of lipids* **2006**, *141*, 83-93.

[29] Spruijt R. B., Wolfs C. J. A. M., Hemminga M. A., *Biochemistry* **2004**, *43*, 13972-13980.

[30] Stopar D., Koehorst R. B. M., Spruijt R. B., Hemminga M. A., *Biochimica et Biophysica Acta* **2009**, 2217–2221.

[31] Meijer A. B., Spruijt R. B., Wolfs C. J. A. M., Hemminga M. A, *Biochemistry* **2001**, *40*, 8815-8820.

[32] Cheng J. T J., Hale J. S., Elliott M., Hancock R. E. W., Straus S. K., *Biochimica et Biophysica Acta* **2011**, *1808*, 622-633.

[33] Cheng J. T., Hale J.D. , Elliot M., Hancock R. E., Straus S. K., *Biophysical Journal* **2009**, *96*, 552-565.

[34] Tait A. R., Straus S. K., *Microbial cell factories* **2011**, *10*.

[35] Devor E. J., Tyra H., Barnes C., Clauer J., Frieberger J., Hermann H., Nutz N., Shafer L., in *Integrated DNA Technologies*.

[36] Rybicki, E., in *Molecular Biology Techniques Manual* (Ed.: James M. D. Coyne V. E., Reid S. J., Rybicki E. P), **1992**.

[37] Rychlik W., Spencer W. J., Rhoads R. E, *Oxford University Press* **1990**, *18*.

[38] Kapust R. B., Waugh D. S., *Protein Science* **1999**, *8*, 1668–1674.

[39] Louis J. M., Mcdonald R. A., Nashed N. T., Wondrak E. W., Jerina D. M., Oroszlan S., Mora P. T., *Eur. J. Biochem* **1991**, *199*, 361 -369.

[40] Sachdev D., Chirgwin J. M., *PROTEIN EXPRESSION AND PURIFICATION* **1998**, *12*, 122–132.

[41] Nallamsetty S.., Austin B. P., Penrose K. J., Waugh D. S., *Protein Science* **2005**, *14*, 2964–2971.

[42] Xiong S., Wang Y., Ren X., Li B., Zhang M., Luo Y., Zhang L., Xie Q., Su K., *World Journal of Gastroenterology* **2005**, *11*, 1077-1082.

[43] Marley J., Lu M., Bracken C., *Journal of Biomolecular NMR,* **2001**, *20*, 71-75.

[44] Hartmann A., Braun V., *Archives of Microbiology* **1981**, *130*, 353-356.

[45] Liu H., Naismith J. H, *BMC Biotechnology* **2008**, *8*.

[46] Epand R. M, Epand R. F., *Molecular BioSystems* **2009**, *5*, 580-587.

[47] Kallick D. A., Tessmer M. R., Watts C. R., Li C., *Journal of magnetic resonance* **1995**, 60-65.

[48] Sreerama N., Woody R. W., *Analytical Biochemistry* **2000**, *287*, 252-260.

[49] Jacobsen N. E., *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*, John Wiley & Sons, **2007**.

[50] Cavanagh J., Fairbrother W. J., Palmer A. G., Skelton N. J., *Protein NMR Spectroscopy: Principles and practice*, Academic Press, **1996**.

[51] Ernst R. R, Bodenhausen G., Wokaun A., *Principles of nuclear magnetic resonance in one and two dimensions, Vol. 14*, Clarendon Press, New York.

[52] Cooka G., Zhanga H., Parka S. H., Wanga Y., Opella S. J., *Biochimica et Biophysica Acta* **2010**, *1880*, 554-560.

[53] Griffith J., Manning M., Dunn K., *Cell* **1981**, *23*, 747-753.

[54] Vinogradova O., Sönnichsen F., Sanders C. R., *Journal of Biomolecular NMR,* **1998**, *4*, 381-386.

[55] Koehorst R. B. M., Spruijt R. B., Hemminga M. A., *Biophysical Journal* **2008**, *94*, 3945-3955.

[56] Kuhn A., Ploss M., *IOP Science* **2010**.

[57] Myles, D. A. A, *Current Opinion in Structural Biology* **2006**, *16*, 630–637.

[58] Zhao X. B., Pan F., Lu J. R., *J. R. Soc. Interface* **2009**, *6*, 659-670.

[59] Nagler C., Nagler G., Kuhn A., *JOURNAL OF BACTERIOLOGY,* **2007**, *189*, 2897-2905.

Appendices

Appendix A Additional Methods for p8 Protein Expression and Purification of Native p8

A.1 Polymerase chain reaction

The table below contain the detailed PCR cycles for Hhv6touch 1, re-amplification and colony PCR.

Table A.1- PCR cycles for "Hhv6touch 1", "Re-amplification" and "Colony PCR" protocols

	"Hhv6 touch 1"	"Re-amplification"	"Colony PCR"
Cycle 1 (X1)	95 ^{o}C 2 min	95 ^{o}C 2 min	100 ^{o}C 10 min
Cycle 2	95 ^{o}C 1 min 65 ^{o}C 1 min 72 ^{o}C 1 min X4	95 ^{o}C 1 min 55 ^{o}C 1 min 75 ^{o}C 1 min X30	95 ^{o}C 2 min X1
Cycle 3	95 ^{o}C hold 65 ^{o}C 1 min 72 ^{o}C 1 min X1	75 ^{o}C 7 min	94 ^{o}C 30 s 55 ^{o}C 30 s 72 ^{o}C 1 min X30
Cycle 4	95 ^{o}C 1 min 68 ^{o}C 1 min 72 ^{o}C 1 min X25		72 ^{o}C 10 min X1
Cycle 5 (1X)	72 ^{o}C 7 min		

A.2 Visualizing DNA on agarose gel

Prepare 30 ml 1.5% agarose gel in 1X TAE with 0.1 µl/ml $EtBr_2$. 1X TAE is used as the running buffer.

DNA samples are resuspended in X10 Blue Juice loading buffer from Invitrogen in 1:6 loading buffer to DNA ratio. 21 µl of samples is loaded per well. 8 µl of agarose gel ladder (Invitrogen) is diluted 1: 4 loading buffer to ladder ratio.

The gel is run at 100 V for about 20 minutes. The gel is visualized on a Molecular Imager-GelDOC XR (Bio-Rad).

A.3 Restriction enzyme endonuclease, ligation and transformation

A restriction digest reaction is set up with the following components:

1) 20.54 ng of full length insert and 200 ng of the vector
2) 1 µl Fast Restriction Enzyme BamHI and 1 µl Fast Restriction Enzyme HINDIII (fermentas)
3) 2 µl Fast Restriction Enzyme Buffer (fermentas) in a total reaction volume of 15 µl.

Once mixed, the digestion mixture is incubated for 5 minutes at 37 ^{o}C. The endonucleases are then deactivated by incubation at 80 ^{o}C for 20 minutes. Ligation is performed by mixing the following components:

1) 5 µl of the DNA digest reaction
2) 1 µl DNA ligase (New England Biolabs)
3) 2 µl ligase buffer supplied in a total reaction volume of 20 µl.

The ligation reaction is incubated for 20 minutes at room temperature and deactivated by incubating at 65 ^{o}C for 10 minutes. Transformation is performed by following the protocol below:

1) 2.5 µl of the ligation mixture is transformed into 50 µl of XL1 Blue *E. coli* cells.

2) Heat shock transformation is performed on the cells ie. 30 minutes on ice, 45 seconds at 42^{o}C and 3 minutes on ice.
3) Cells are allowed to recover by adding 800 µl SOC media and incubating in 37 ^{o}C , 225 r.p.m. for 1 hr.
4) Cells are centrifuged at 4.2 r.p.m. for 1 min and the cell pellet is re-suspended in 150 µl SOC.
5) 20 µl of cells is plated on 2 % LB with 0.05 mg/ml CBC and the plate is incubated in 37 ^{o}C incubator for 18 hours.

A.4 Colony PCR

A desired colony is inoculated into a PCR reaction mixture prepared as below:

1) 2 ul pfu buffer (Fermentas)
2) 0.4 µl pmal seq forward primer
3) 0.4 µl pmal seq reverse primer
4) 0.4 µl dNTP (Fermentas)
5) 15.8 µl deionised sterile water

Colony PCR protocol (Table A.1) is performed in the PCR mini cycler. The pfu DNA polymerase from Fermentas is added at the 95^{o}C step of the PCR protocol. 15 µl of the PCR product is loaded onto an agarose gel with 5:1 sample to Blue Juice Loading Buffer ratio for analysis.

Note that the same colony is used to inoculate both a PCR reaction mixture and a 5 ml 2% LB and 0.05 mg/ml CBC for overnight growth.

A.5 Plasmid miniPrep

The following protocol was followed for plasmid isolation:

1) 5 ml 2% LB with 0.05 mg/ml CBC is inoculated with a single colony.
2) The 5 ml culture is allowed to grow for 15 hours at 37 ^{o}C with 225 r.p.m. shaking.
3) QIAGEN Plasmid Extraction Kit (QIAGEN) is used to extract the plasmid.

A.6 Competent cells

Competent cells are made using $CaCl_2$; the protocol is adapted from Drew *et al*. The protocol is followed as below:

1) 5 ml of 2% LB Broth with 0.05 mg/ml Carbenicillin (CBC) is inoculated with the required colony of *E. coli* cells and incubated in a 37 ^{o}C shaker at 225 r.p.m. for 4 hours.
2) 2 ml of the culture is used to inoculate 100 ml 2% LB with 0.05 mg/ml CBC.
3) The 100 ml culture is incubated at 37 ^{o}C shaker at 225 r.p.m. until log phase (OD_{600} between 0.5-0.7).
4) The culture is transferred into two 50 ml Falcon tubes and cooled on ice for 10 min.
5) The culture is centrifuged at 2,600 g, 4 ^{o}C for 10 min.
6) The supernatant is discarded and the cell pellet is resuspended in 10 ml ice cold 0.1 M $CaCl_2$. The cells are incubated on ice for 10 min.
7) The cells are centrifuged at 2,600 g, 4 ^{o}C for 10 min. The supernatant is discarded and the cell pellet is re-suspended in 2 ml of ice-cold 0.1 M $CaCl_2$ plus 20% glycerol. The cells are incubated on ice for 10 min.

The cells are divided into 100 µl aliquots, snap frozen in liquid nitrogen and dry ice and stored at-80 ^{o}C.

A.7 SDS-PAGE

The stacking and separating gel layers are made by following the recipe below:

Table A.2- Recipe for separating and stacking gels for SDS-PAGE gel

	Separating gel	Stacking gel
Bis-acrylamide 40% (Biorad)	4.9 ml	0.81 ml
Tris/SDS pH 5.45	5 ml	1.55 ml
Water	1.93 ml	3.89 ml
50% glycerol	3.17 ml	----
10% ammonium persulfate (Biorad)	25 µl	50 µl
TEMED (Biorad)	5 µl	10 µl

A.8 Visualization of protein

Commassie staining

Gel is incubated in comassie staining solution (10% acetic acid, 50% methanol and 0.25 g commasie G25 in 1 L total volume) for 3hours with gentle shaking. Gel is destained using destaining solution (15% methanol, 10% acetic acid) and then visualized on the GelDoc XR (BioRad).

His-staining

For invision staining, the gel is incubated for 1 hr in 100 ml fixing solution, washed twice with 100 ml water for 10 min each wash. The gel is stained with 25 ml Invision-stain (Invitrogen) for 1 hour then washed twice with 100 ml 20 mM phosphate buffer for 10 min each wash.

A.9 French press and cell lysis

There are two separate protocols for the purification of origami and C41 cells.

For C41 cell, the protocol is,

1) C41 cell pellet from 1-2 L batch of cell grow-up is resuspended in 30 ml of lysis buffer (20mM potassium phosphate dibasic, 0.5M sodium chloride, pH 7.4).
2) Protease inhibitor cocktail for His-tagged proteins (1 ml per 20 g of cells) from Sigma Life Science and a spatula tip full of DNAse (Roche) is added to the suspension.
3) The suspension is loaded onto the French press where it is run through the french press three times, with the cells stored on ice between each cycle.
4) The cells are centrifuged in SA-300 rotor at 15500 r.p.m., 4 ^{o}C for 40 minutes.
5) The supernatant is discarded.
6) The cell pellet is re-suspended in solubilization buffer (20mM potassium phosphate dibasic, 0.5M sodium chloride, 10mM imidazole and 1% triton X, pH 7.4) and stirred for 2 hours at 4 ^{o}C.
7) The suspension is then centrifuged in SA-300 rotor at 15500 r.p.m., 4 ^{o}C for 40 minutes.

8) The supernatant is filter sterilized.

1) The supernatant is loaded onto a pre-equilibrated His column.

For Origami purification, the following protocol is followed:

2) Origami cell pellet from 1-2 L batch of grow up is resuspended in 30 ml solubilization buffer.
3) Protease inhibitor cocktail (1 ml per 20 g of cells) from Sigma Life Science and a spatula tip full of DNAse (Roche) is added to the suspension.
4) The suspension is run through the french press three times, with the cells stored on ice between each cycle.
5) The lysed cells are mixed at 4 ^{o}C for 2 hours.
6) The cell mixture is centrifuged in a SA-300 rotor at 15500 r.p.m. for 60-90 minutes.
7) The supernatant is filter sterilized.
8) The supernatant is loaded onto a pre-equilibrated His column.

A.10 Nickel-NTA beads column

Nickel-NTA beads (Qiagen) are packed into a re-useable column (Biorad). The column volume used for purification of p8 is 25 ml. The column purification protocol is as follows:

1) The column is first equilibrated by washing with 50 ml deionised water, 25 ml binding buffer (0.5M sodium chloride, 20mM potassium phosphate dibasic, 10mM imidazole, 0.5% triton X, pH 7.4),25 ml elution buffer (500mM sodium chloride, 20MM potassium phosphate dibasic, 500mM imidazole, 0.5% triton X, pH 7.4) then 50 ml Binding Buffer.
2) The filter sterilized sample is then loaded into the column at 2 ml/min.

3) The flow-through is collected and re-loaded onto the column at 2 ml/min.
4) The column is washed three times with 50 ml Binding Buffer at 5 ml/min (B1, B2, B3).
5) The bound protein is eluted by washing with 25 ml Elution Buffer at 5 ml/min (E).
6) Column is cleaned by washing with 50 ml deionised water and 25 ml 70% ethanol.
7) The protein eluant is dialyzed against a 1 L dialysis buffer (20mM potassium phosphate monobasic, 62.5mM sodium chloride, 0.5% triton X, pH7.4) overnight at 4 ^{o}C.

Dialysis is performed to remove the imidazole that is present in the eluant.

A.11 Thrombin digest

Protein sample from 4 litres of *E. coli* expression batch is digested with 100U of thrombin (GE Healthcare) for 18-22 hrs at room temperature. P-aminobenzamidine-agarose (Sigma Life Science) is added to the reaction to stop thrombin digestion. The protocol followed is as shown below:

1) 100ul of p-aminobenzamidine-agarose is centrifuged at 4200 r.p.m. on a table top centrifuged.
2) The layer of ethanol, which the agarose is stored in, is discarded.
3) The p-aminobenzamidine-agarose is equilibrated by resuspending in 200 µl of dialysis buffer and centrifuging at 4200 r.p.m.
4) The layer of dialysis buffer is then discarded
5) The p-aminobenzamidine-agarose pellet is added to the thrombin digestion reaction.
6) The mixture is vortexed and left to stand for 5 minutes.

7) The mixture is centrifuged at 4200 r.p.m. at 4 ^{o}C.

8) The pellet of p-aminobenzamidine-agarose is discarded.

A.12 Acetone precipitation

Protein undergoes three acetone precipitation cycles. Acetone is added to the protein solution, post thrombin digestion, in a 1:1 ratio and vortexed to mix properly. The first acetone precipitation is left in -20 ^{o}C freezer overnight. The second and third precipitation cycles are left for a minimum of 2 hours. After every precipitation cycle, the suspension is centrifuged at 4200 r.p.m., 4 ^{o}C for 10 minutes and the protein pellet is resuspended in 35 ml ice cold acetone (the next cycle). Acetone precipitation precipitates out any proteins in the solution and so gets rid of excess triton X.

The protein pellet after the third acetone precipitation cycle is dried, resuspended in nanopure water and dialyzed against nanopure water for 2 days while changing the dialysis water three times. This gets rid of excess salt. The protein solution is subsequently lyophilized.

A.13 Double column purification- nickel-NTA column and Q-sepharose

Nickel-NTA beads (Qiagen) are packed into a re-useable column (Biorad). Q sepharose beads (GE HealthCare) are packed into Biorad re-usable column. The column volume used for purification of p8 is 25 ml. The column purification protocol is as follows:

1) The his-trap is first equilibrated by washing with 50 ml deionised water.

2) The column is equilibrated by washing with 50 ml dialysis buffer (20mM potassium phosphate monobasic, 62.5mM sodium chloride, 0.5% triton X, pH7.4).

3) The filter sterilized sample is then loaded into the column at 2 ml/min. The flow-through from this is the H_s sample fraction.
4) The column is washed with 25 ml elution buffer (500mM sodium chloride, 20MM potassium phosphate dibasic, 500mM imidazole, 0.5% triton X, pH 7.4).
5) Column is cleaned by washing with 50 ml deionised water and 25 ml 70% ethanol.
6) The H_s sample is loaded onto a Q-sepahrose column also washed with 50 ml deionised water and equilibrated with 50 ml dialysis buffer. The flow-through from this is the Q_s sample fraction.
7) The column is washed with 25 ml buffer with high salt concentration (20mM potassium phosphate monobasic, 1M sodium chloride, 0.5% triton X, pH7.4).

Appendix B Additional Methods for the Protein Expression and Purification of I32C Mutant of p8

B.1 Site directed mutagenesis polymerase chain reaction

PCR mix was prepared as:

1) 143 ng of forward and reverse primer
2) 50 ng of template DNA
3) 1 µl of dNTP
4) 5 µl of pfu polymerase buffer
5) 2 µl of 25 mM $MgSO_4$
6) 40.2 µl of deionised water

The PCR protocol is followed as below:

Table B.1- Site directed mutagenesis PCR protocol followed for introducing a cysteine mutation at residue 32 of p8sequence

Cycle	Temperature/ ^{o}C	Time/ minutes
1- Denaturation of DNA	95	5
2- Amplification cycles	95	1
2- Amplification cycles	52	1
2- Amplification cycles	72	10
	*Repeat cycle 2 12 times	
3- Annealing step	41	1
4- Extension step	72	30

After the PCR reaction, the PCR mixture is cooled to 37 ^{o}C and 10 units of DpnI (Invitrogen) is added. The DpnI reaction is allowed to incubate at 37 ^{o}C for 1 hour.

Appendix C Additional Data for p8 Protein

C.1 MALDI-TOF-MS analysis of p8

The trace from mass spectrometry analysis of the p8 fragment post digestion is given in the figure below.

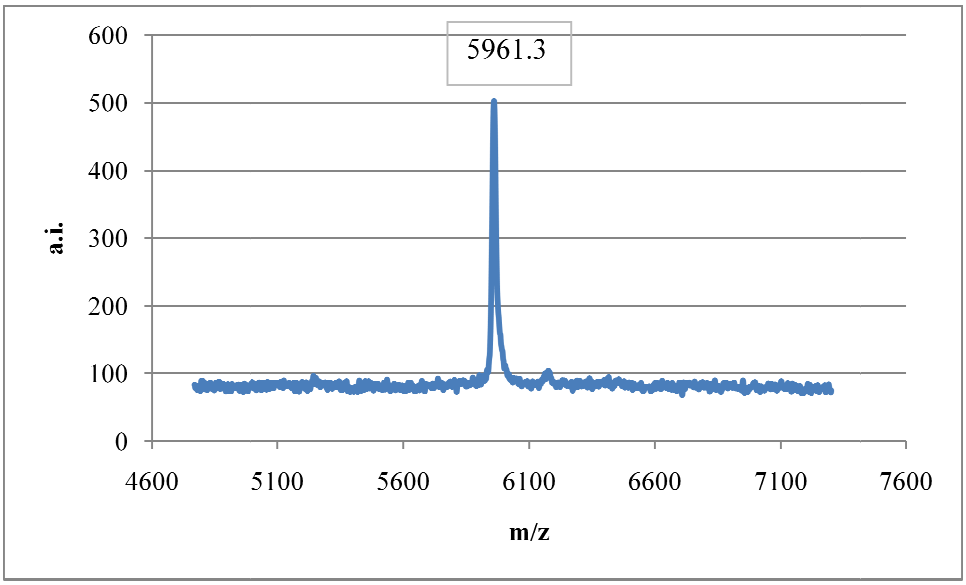

Figure C.1- Trace from MS using a MALDI-TOF MS with Bruker Biflex IV MALDI-TOF. The trace is of the MBP-p8 complex post-digestion. This post digestion sample, as already explained, contains both MBP and p8, although in this trace, only the peak from p8 is highlighted (peak at 5961.3 Da).

C.2 MALDI-TOF-MS analysis of ^{15}N labeled p8

The mass spectrometry trace of the Q_s sample from the double column purification of ^{15}N labeled p8 is shown in the figure below.

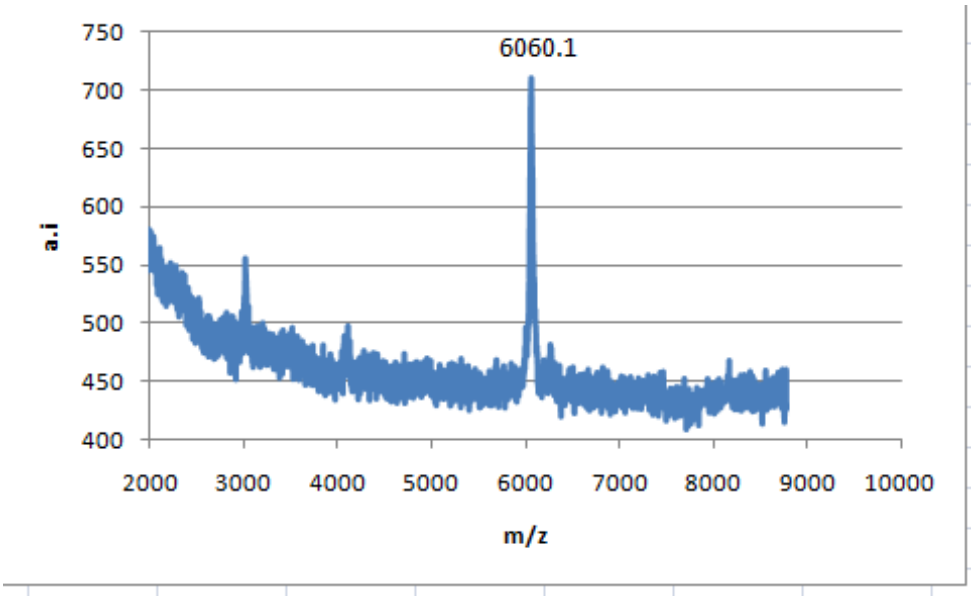

Figure C.2- Trace from MS using a MALDI-TOF MS with Bruker Biflex IV MALDI-TOF. The trace is of the Q_s sample from the double column purification of ^{15}N labeled p8. The molecular weight of the ^{15}N labeled p8 is 6060.1 Da.

C.3 MALDI-TOF-MS analysis of I32C p8

The trace of the Q_s sample from the double column purification of I32C p8 is shown in the figure below.

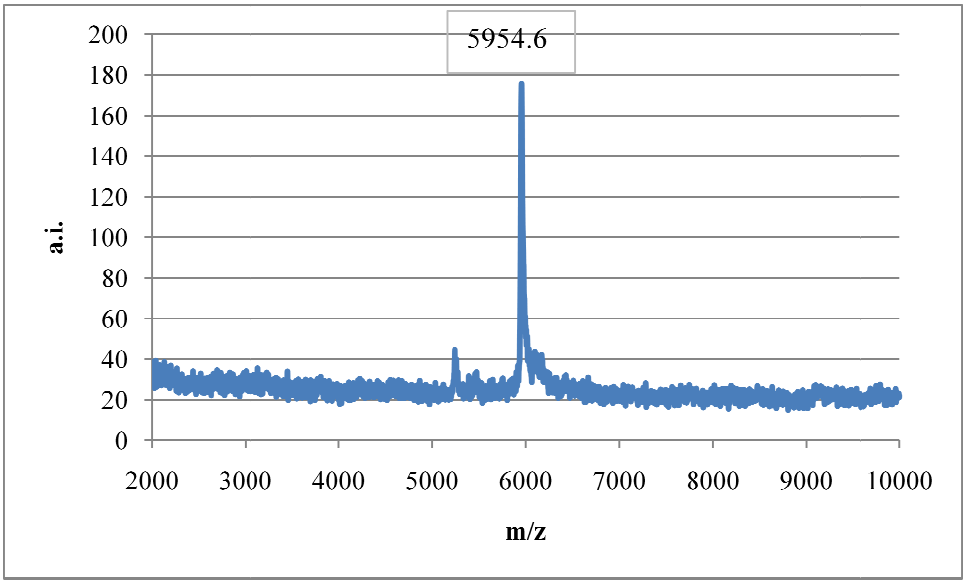

Figure C.3- Trace from MS using a MALDI-TOF MS with Bruker Biflex IV MALDI-TOF. The trace is of the Qs sample from the double column purification of I32C p8. The molecular weight of the I32C p8 is 5954.6 Da.