12-15-2007

Purification and structural analysis of Newcastle disease virus V protein and flowering locus T (FT) protein

Swapna Jayapalan

PURIFICATION AND STRUCTURAL ANALYSIS OF NEWCASTLE DISEASE VIRUS V PROTEIN AND FLOWERING LOCUS T (FT) PROTEIN

By

Swapna Jayapalan

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Chemistry
in the Department of Chemistry

Mississippi State, Mississippi

Dec 2007

PURIFICATION AND STRUCTURAL ANALYSIS OF NEWCASTLE DISEASE VIRUS V PROTEIN AND FLOWERING LOCUS T (FT) PROTEIN

By

Swapna Jayapalan

Approved:

John K. Young Assistant Professor of Analytical Chemistry, Department of Chemistry (Major Professor)	David O. Wipf Professor of Analytical Chemistry Department of Chemistry (Committee Member)
Svein Saebo Professor of Physical Chemistry Department of Chemistry (Committee Member)	Stephen Foster Associate Professor of Physical Chemistry Department of Chemistry (Graduate coordinator)

Gary L. Myers
Interim Dean of the College of
Arts and Sciences

Name: Swapna Jayapalan

Date of Degree: Dec 2007

Institution: Mississippi State University

Major Field: Chemistry

Major Professor: Dr. John K. Young

Title of Study: PURIFICATION AND STRUCTURAL ANALYSIS OF NEWCASTLE DISEASE VIRUS V PROTEIN AND FLOWERING LOCUS T (FT) PROTEIN

Pages in Study: 81

Candidate for Degree of Master of Science

Newcastle disease virus (NDV) is one of the paramyxovirus that has been studied at length since this virus infects the birds of all species. NDV is highly virulent in chickens and results in a high mortality rate because the V protein of NDV is found to inhibit the avian immune response system. No drugs are available for treating NDV therefore, determining the structure of V protein would help in developing a drug that can inactivate the V protein, thereby increasing the host immune response against viral infection. The research here is focused on purification and initial structural analysis of the V protein of NDV. The V protein was purified by gel filtration chromatography and the structure was studied using fluorescence and CD spectroscopy, and NMR. The results suggested that the V protein is unstructured.

The research was also aimed at purification and structural analysis of the flowering locus T (FT) protein, which is found to play a major role in the initiation of flowering in plants. Structural analysis of the FT protein may help in finding the possible domains of the FT protein that interacts with other plant proteins, leading to flowering. The FT protein was purified by ion exchange chromatography and the structure was studied by fluorescence and CD spectroscopy. The fluorescence data suggested that the FT protein might be folded, where as the CD data were inconclusive. More accurate secondary structure information about the protein could be obtained using NMR, however; since the concentration of the FT protein was too low (0.007 mM), NMR study was not possible.

DEDICATION

I would like to dedicate this research to my loved ones, Krupa and Venu Jayapalan (mom & dad), Sudheer Jayapalan (brother), Sandhya (sister) and to my loving husband, Dr. Kumaraguru Raja, who has always encouraged and supported me.

ACKNOWLEDGMENTS

The pleasure of completing this work lies in the opportunity given to acknowledge those who have made it possible. I gratefully express my deep sense of gratitude and acknowledgements to the Department of Chemistry, Mississippi State University in general and Dr. Stephen Foster, graduate co-ordinator in particular for allowing me to pursue my degree in Master of Science, in Chemistry.

I express my profuse thanks to my major research advisor, Dr. J. K. Young, without whom this research would not have been possible. He has always guided and encouraged me through out my research. I would like to thank my committee members, Dr. Wipf and Dr. Saebo, for their support. I am greatly indebted to my colleagues in the lab, Arezue, Karthik, and Bulent, for their valuable assistance and co-operation during my research. I finally thank my fellow friends for their kind support and timely help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS, AND NOMENCLATURE

RNA	ribonucleic acid
NDV	Newcastle Disease virus
HN	hemagglutinin-neuraminidase
F	fusion
M	matrix
P	phospho
NP	nucleocapsid
L	large–polymerase
mRNA	messenger RNA
Neu5Ac	sialic acid N-acetylneuraminic
HR	heptad repeat
HR1	heptad repeat one
HR2	heptad repeat two
TM	transmembrane
SV	Simian virus
ICAP-AES	inductively coupled argon plasma-atomic emission spectrometry
SDS-PAGE	Sodium dodecyl sulfate-polyacryl amide gel electrophoresis
DNA	deoxy ribonucleic acid
DDB	damage specific DNA-binding protein
BP	beta-propeller

BPA	beta-propeller A
BPB	beta-propeller B
BPC	beta-propeller C
IFN	interferon
STAT1	signal transducers and activator of transcription factor 1
NiV	Nipha virus
RT-PCR	reverse transcriptase-polycyclic reactions
E. coli	Escherichia coli
kDa	kilodalton
UV	ultraviolet
kb	kilobase
LB	Luria-Bertani
IPTG	Isopropyl beta-D-1-thiogalactopyranoside
M9	minimal
Ni-NTA	Nickel-nitrilotriacetic acid
mM	millimolar
CD	circular dichroism
NMR	nuclear magnetic resonance
λ	wavelength (nm)
λ_{em}	wavelength of emission (nm)
θ	Ellipticity in degrees
2D HSQC	two-dimensional heteronuclear single quantum correlation

INEPT	insensitive nuclei enhanced by polarization and transfer
J	coupling constant
FT	flowering locus T protein
RKIP	RAF-kinase inhibitor protein
CO	CONSTANS protein
GFP	green fluorescent protein
$(NH_4)_2SO_4$	ammonium sulphate
DTT	dithiothreitol
KCl	potassium chloride
$(-N(CH_3)_3^+)$	quaternary amine

Chapter I

Purification and Structural Analysis of Newcastle Disease Virus V Protein

1.1 Paramyxovirus

Paramyxoviruses are lipid bilayer enveloped viruses that contain a single-stranded, left-handed helical coil ribonucleic acid (RNA) genome of negative sense. Negative sense RNA means that the RNA itself does not act as a messenger RNA.[1] They occur worldwide, and infection is transmitted via ingestion or aerosol.[2] These viruses infect a wide variety of hosts causing severe clinical symptoms.[3] Viruses such as measles, mumps, and parainfluenza virus cause chronic pneumonia, fever, and respiratory tract diseases in humans.[4] These viruses also infect animals, such as Bovine Respiratory Syncytial virus, which causes severe bronchitis and pneumonia in cows.[5] Some of these viruses are highly fatal resulting in a high mortality rate. For example, Nipha virus, which is named after the village from which it emerged in Malaysia in 1998, has spread rapidly and caused severe encephalitis and respiratory diseases in pigs with a high mortality rate. This led to high economic loss to the pig industry in Malaysia. Nipha virus was also transmitted to humans, especially to those working in pig farming, causing severe febrile encephalitis with a high mortality rate.[6] Paramyxoviruses not only infect humans and animals, but also a wide variety of avian species. For example, Newcastle Disease virus

(NDV) affects chickens and many other birds causing chronic respiratory and nervous disorders, leading to paralysis and death. To prevent this, all chickens in the world are vaccinated against this virus. Newcastle Disease virus is still one of the important pathogens affecting the economy of the poultry industry.[2] These viruses have the potential to cross over to the human population through infected animals.[6]

In the past few years, the proteins encoded by the Paramyxovirus have been extensively studied because no drugs are available for treating viral infections caused by these viruses. Vaccination is the only way to control the epidemic of the viral infection.[7] Therefore, the study of viral proteins helps to understand the mechanism by which the virus functions which in turn help to design new potential drug targets.

1.1.1 Structural Proteins of Paramyxovirus

Paramyxoviruses are commonly spherical in shape having a diameter of 150-350 nm and are enveloped by a lipid bilayer membrane, which is derived from the infected host cell membrane.[1] Figure 1.1 shows the structure of a paramyxovirus virion. The negative stranded RNA genome contains about 15,000 nucleotides that codes for six genes and seven different structural proteins in the virus. They are the hemagglutinin-neuraminidase (HN) gene that encodes for HN protein, Fusion (F) gene that encodes F protein, Matrix (M) gene that encodes for M protein, phospho (P) gene that encodes for P and V proteins, nucleocapsid (NP) gene that encodes for NP protein and, the large–polymerase (L) gene that encodes for L protein.[1] An additional protein named V protein is obtained from the P gene when the messenger RNA (mRNA) skips a nucleotide base.[8] The negative stranded helical RNA genome in Paramyxovirus serves as the template for

the synthesis of mRNA and replication of the virus in the infected host cell.[9] The schematic diagram of paramyxovirus genome is shown in Figure 1.2. The outer surface

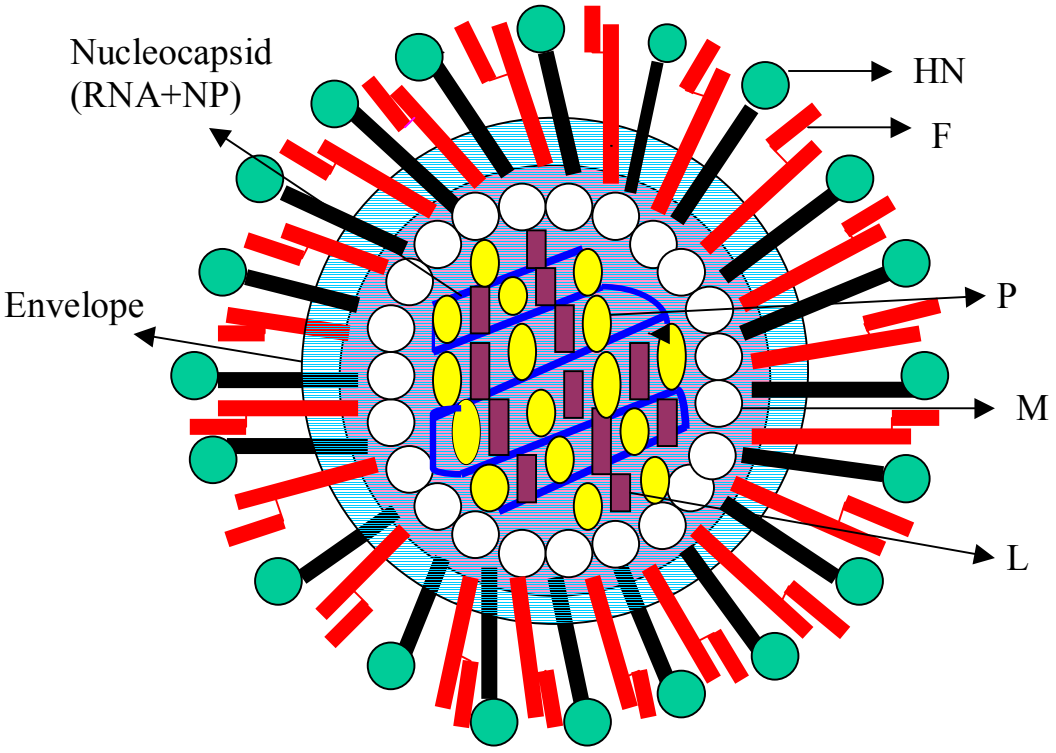


Figure 1.1: Paramyxovirus virion structure.[9]

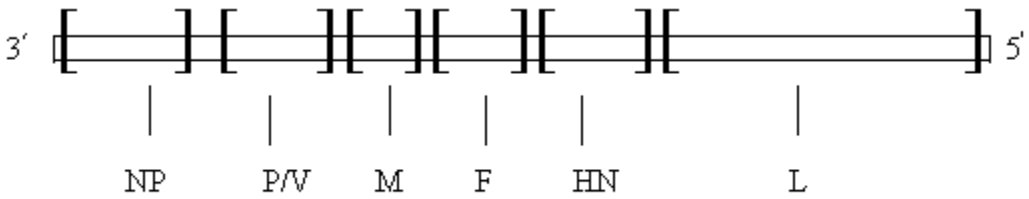


Figure 1.2: Schematic diagram of paramyxovirus genome.[9]

of the virus is covered by two glycoproteins, the HN glycoprotein and the F glycoprotein. These glycoproteins are inserted through the lipid bilayer membrane by a polypeptide-anchoring segment, while most of their structure protrudes forming viral spikes.[10] The HN glycoproteins contain about 565 to 582 amino acid residues and they play a key role in the attachment of the virus to the host cell membrane.[1] This protein is a multifunctional glycoprotein having both hemagglutination activity (agglutination of erythrocytes) and neuraminidase activity (binding to sialic acid receptors on the surface of the host cell), hence the name HN glycoprotein. The fusion of the viral membrane with the host cell membrane is promoted by the fusion protein.[10] The F protein consists of two components, F_1 and F_2, which are linked together by disulfide bonds. The F proteins contain about 540 to 580 amino acid residues.[1] Below the lipid bilayer of the virus is the M protein, which constitutes the most abundant protein in the virus. M proteins contain about 341 to 375 amino acid residues and play a very important role in the assembly of the viral proteins and also in the budding of the virus.[1, 11] Inside the viral membrane is present the negative-stranded helical RNA, which forms a stable complex with the nucleoprotein known as nucleocapsid.[11] The nucleoproteins contain about 489 to 553 amino acid residues and help in virus replication. Associated with the nucleocapsid is a polymerase complex, which is composed of phosphoprotein and large protein. The length of P protein is highly variable with in the virus family. For example, the P protein of parainfluenza viruses contain about 507 to 603 amino acid residues, where as the P protein of rubulaviruses contain about 245 to 397 amino acid residues.[1] In some viruses such as the NDV, an additional protein called the V protein is produced from P gene during transcription,

when the mRNA transcriptase skips a nucleotide base while reading the P gene. As a result the C-terminal half of the V protein results in cysteine rich regions. This V protein is usually found in virus-infected cells and the function of this V protein is still uncertain.[8] The L protein is the largest protein in length containing about 2,200 amino acid residues, however, it is the least abundant structural protein of the virus. The L protein helps in the replication of the virus in the host cell.[1]

The functions of these structural viral proteins will be discussed in detail in the NDV section. Newcastle disease virus is one of the paramyxovirus that is studied at length because the outbreak of this virus infects the birds of all species. In chickens, the mortality rate is high, which leads to economic losses in the poultry industry.[8] No drugs are available for treating the infection caused by NDV.[2] Therefore, studying the structure of the proteins encoded by the NDV helps in the development of drugs against viral infection.

1.2 Newcastle Disease Virus

NDV is named after the village it was first identified in England in 1927.[12] It occurs naturally and infects all species of birds causing severe respiratory and nervous disorders.[3] The disease is found to be highly communicable and the infection is transmitted via ingestion or aerosol.[2] NDV is highly virulent in chickens and results in a high mortality rate. In 2002, in California, an NDV outbreak in a commercial poultry farm infected 3.5 million birds, which lead to complete depopulation of the whole farm in order to prevent the spread of the disease.[13] The virus is so virulent that some times birds die without any symptoms, therefore the Newcastle disease is on list 1 contagious disease

of poultry.[13] The virulence of this virus is not clearly understood, but it is found to depend on the strain of the virus.[14] Currently, the disease is present worldwide and no drugs are available for the disease. Vaccination is the only preventive measure to control the spread of the disease.[3] NDV is also on the list of viruses that could be used for bioterrorism because this virus is found to inhibit the host interferon system.[15] The interferon system increases the host immune response against viral infection by producing alpha/beta interferones, which are proteins that inhibit the replication of the virus particles in the host cells. Thus, the interferones act as antiviral agents. The NDV is found to inhibit the production of interferones, thereby playing an interferon-antagonist role.[14]

1.3 Structural Proteins of NDV

The NDV is an enveloped virus, which is roughly spherical in shape having a diameter around 100-500 nm.[1] The single stranded, left handed helical coil negative RNA genome of NDV is 15,186 nucleotides long, which encodes seven structural proteins: the hemagglutinin-neuraminidase and fusion protein (which form the external surface proteins), the matrix protein, phosphoprotein, nucleocapsid protein, V protein, and the large–polymerase protein (which form the internal proteins).[16] Figure 1.3 shows the schematic diagram of NDV genomic organization.

1.3.1 HN and F protein

The HN and F proteins are two viral glycoproteins that play a combined role in the infection process.[17] The HN protein of NDV contains 577 amino acids and helps in the attachment of the virus to the host cell membrane. The amino acid sequence of HN

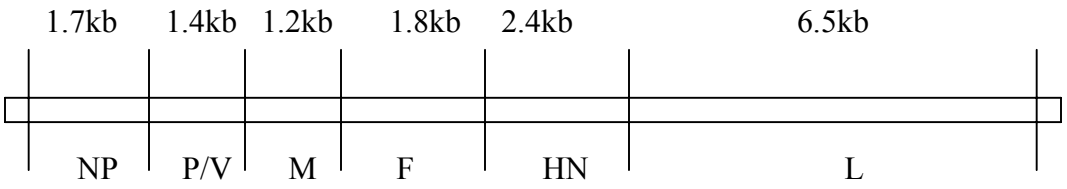


Figure 1.3: Schematic diagram of NDV genomic organization.[2]

protein reveals that hydrophobic amino acids are present at the N-terminal (amino terminal) end and hydrophilic amino acids are present at the C-terminal (carboxyl terminal) end.[17] The N-terminal end of HN glycoprotein has a stack like structure that is inserted into the cytoplasm of the host cell and a short helical domain that extends into the transmembrane region followed by a long helical structure and a globular head that is projected to the outside. The HN glycoprotein is found as a tetramer on the membrane surface, consisting of a pair of dimers, which are linked together by disulfide bonds.[18] The HN protein is a multifunctional protein having a significant role in both hemagglutination and neuraminidase activity. The residues E401, R174, R416, and Y526 present on the N-terminal end interact with the acetamido methyl group of sialic acid N-acetylneuraminic (Neu5Ac) receptor molecules present on the terminal gangliosides of the host cell and initiate agglutination of red blood cells (hemagglutination).[19] Figure 1.4 shows the chemical structure of Neu5Ac. In addition to binding to sialic acid receptors the HN protein undergoes a conformational change and initiates enzymatic cleavage of the sialic acid receptor linkage in the host cell by its neuraminidase activity. The HN

neuraminidase activity prevents the self-aggregation of viral particles in the infected cells.[11] However, the exact mechanism by which HN protein acts as a binding and catalytic site is still a controversial issue. The HN protein not only plays an important role in the adsorption of the virus to the host cell but also promotes fusion between the viral and the infected cell surface by activating the fusion protein.[10]

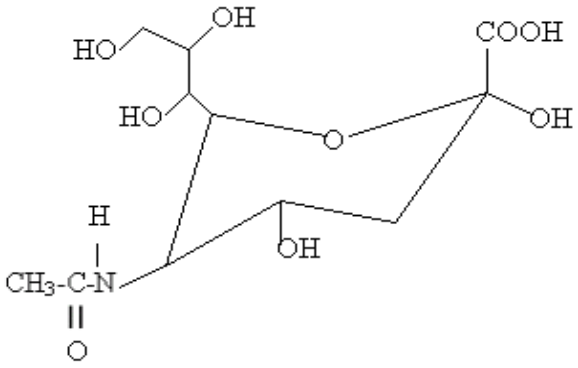


Figure 1.4: Chemical structure of Neu5Ac.[10]

The F protein of NDV contains 553 amino acids and is found to mediate the fusion between the viral and the host cell membranes. The F protein exists as a homotrimer on the membrane surface where the three monomers are inter-coiled.[20] Sequence analysis of F protein suggests that the hydrophobic amino acids are present at the C-terminal end and the hydrophilic amino acids are present at the N- terminal end. This feature indicates that the C-terminal end of the F protein is inserted into the viral lipid membrane followed by a small helical domain that extends into the transmembrane region, which is followed by the N-terminal end that is protruded out side.[21] All the F proteins of NDV are synthesized as an inactive precursor F_0, which is cleaved into

biologically active polypeptides F_1 and F_2 by the proteolytic enzymes of the host cell. Both the F_1 and F_2 polypeptides are linked together by disulfide bonds (-S-S-).[3] Due to this cleavage, the N-terminal end of F_1 polypeptide contains a highly conserved hydrophobic fusion peptide sequence (10-15 amino acids long), which inserts into the target membrane. The F_1 polypeptide has two heptad repeat (HR) regions: one heptad repeat (HR1) region is located close to the N-terminal end of the F_1 polypeptide and the second heptad repeat (HR2) is located adjacent to the transmembrane domain.[22] Figure 1.5 shows the schematic diagram of F protein of paramyxovirus. The HR2 domain has a leucine zipper motif, which continues 4-5 helical turns. Morrison and her research group showed that substituting two leucine residues with alanine inhibited the fusion activity

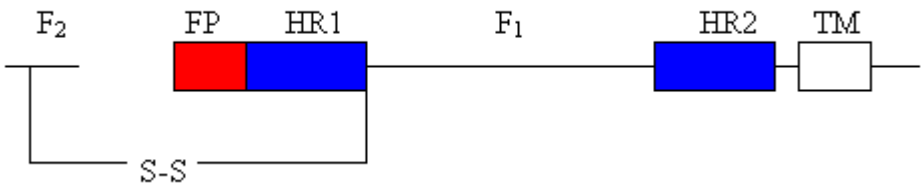


Figure 1.5: Schematic diagram of F protein of paramyxovirus. Where F_2 and F_1 refer to the polypeptides and FP, HR1, HR2, and TM refers to the fusion peptide, heptad repeat ne, heptad repeat two, and transmembrane. S-S indicates the disulfide link between F_2 and F_1 polypeptides.[23]

due to the loss in the structure of the helix. This suggests that leucines interact with the peptide fusion sequence during the fusion process.[24] Many models have been proposed to

explain the fusion process, but there are many controversial issues that remain unsolved.[25] The current model proposes that when the HN protein binds to the sialic receptors on the surface of the host cell, it undergoes a conformational change, which in turn triggers a conformational change in the fusion protein. As a result, the fusion peptide sequence is activated and is inserted into the target membrane. The HR1 and HR2 domains form a stable complex, which is thought to pull the viral and the target membranes together resulting in fusion events.[25] The fusion of this peptide sequence with the host membrane is pH independent, therefore, viral infection occurs at the host cell plasma membrane.[21] However, what conformational structure of the F protein inserts into the target membrane and the mechanism involved in the fusion process is still unclear.[25] Young *et al.* have studied the peptide sequences of HR1 and HR2 regions and determined the structure of these peptides by Nuclear magnetic resonance (NMR) spectroscopy and molecular modeling. The results suggested that both HR1 and HR2 polypeptides form trimers having helical structure. Binding studies between HR1 and HR2 peptides revealed that these peptides could interact with each other forming a rosette-like structure (arrangement of helices in a circular form like the pettles of a rose).[26]

1.3.2 M protein

Below the lipid bilayer of the virus is present the M protein, which forms a shell like core arround the inner surface of the viral membrane and constitutes the most abundant protein in the virus. The M gene of NDV is 1092 nucleotide base pairs long that encodes the M protein of 364 amino acids.[27] This protein is highly basic (14% basic and 8% acidic amino acids) and is hydrophobic in nature. Therefore, it is close to the

peripheral region of the viral lipid layer.[28] These proteins contain short motifs called late domains that interact with the host vacuolar protein-sorting pathway and use that pathway for their viral replication process.[29] To date, the structure and function of M protein is not known. However, some studies suggested that M interacts with other M molecules and forms a two dimensional crystalline array.[30] M proteins are known to maintain the assembly of the viral proteins in the infected cells. In the presence of M protein, the cells co-expressed releasing NP, F, or HN proteins. Whereas, in the absence of M protein the cells could not release NP, F, or HN proteins. This suggest that the M protein is required for the assembly and release of other viral proteins.[29] There are many contradictory conclusions drawn about the M protein. Some studies suggested that M protein interacts with the F and HN proteins, which is important for maintaining the virus assembly. It is also known that the M protein interacts with the nucleocapsid protein enhancing the virus budding process, which involves pinching off viral particles (buds) from the plasma membrane of the infected cell. However, the mechanism by which the M protein interacts with other viral proteins and its role in budding process is not clearly understood.[11]

1.3.3 P protein

The P protein of NDV contains 484 amino acids with a molecular weight of 42 kDa.[31] The C-terminal end of P protein contains high concentrations of serine and threonine amino acids, which are known to undergo in-vivo phosphorylation. Hence theses proteins are named phosphoprotein.[31] The P interacts with other P molecules and forms oligomers. By deletion analysis it was established that the C-terminal region of P protein plays a major role in P-P interaction. However, what causes the oligomerization

of P protein in infected cells is still unknown.[32] The P protein is found to form a complex with the L protein. By using deletion analysis it was confirmed that the amino acid residues 412 to 479 present on the C-terminal half of P protein binds to amino acid residues 21 to 350 present on the N-terminal end of L protein.[33] The P protein also forms a stable complex with the NP protein, which plays a major role in viral replication. However, it was found that different domains of P protein interact with the C-terminal end of NP, therefore, more methodical studies need to be carried out to establish P-NP complex formation.[34] Studies on the P and NP proteins have suggested that nucleocapsid like particles are formed only when the P protein interacts with the NP protein. Therefore, it is believed that the P protein acts as a channel of communication for the NP protein and prevents the unrestrained nucleocapsid like particle formation.[34] The P protein plays a main role in the biological replication and transcription of the virus in the infected cells of the host. However, the molecular nature of the P protein and its role in the assembly of the virus is still unclear.[32]

1.3.4 NP protein

The NP protein of NDV contains 489 amino acids and has a calculated molecular weight of 53 kDa. The N-terminal end of NP forms a complex with the RNA genome, which serves as the helical core nucleocapsid structure of NDV.[35] The NP interacts with other NP molecules forming herringbone-like structures (long flexible structures). This was supported by electron microscopy studies, which suggested that when NP was expressed in yeast cells in the absence of other viral proteins, herringbone-like particles were formed.[36] Deletion analysis suggested that the N-terminal 294 amino acids play a

major role in NP-NP interactions.[37] The sequence analysis suggested that N-terminal amino acid residues 362 to 371 of NP are highly conserved among other NDV strains, while the C-terminal amino acid residues are highly variable.[35] The NP forms a complex with the viral RNA resulting in nucleocapsid like particles. This was established from an experiment, which indicated that when Sf9 cells containing RNA were infected with NP, they showed the formation of nucleocapsid like particles.[38] The NP also forms a complex with P protein. The deletion analysis studies revealed that the C-terminal amino acid residues from 1 to 26 of NP form a complex with the P protein, which is believed to play a significant role in virus replication and budding.[34]

1.3.5 L protein

The L protein contains 2223 amino acids and has a molecular weight of 250 kDa. The L protein is the largest and the least abundant protein of the virus, suggesting that the L protein might have an enzymatic role rather than structural.[39] To date, the structure of the L protein is not known. It was recently found that L interacts with other L molecules forming oligomers. Deletion analysis studies suggested that the amino acid residues from 1-174 present at the N-terminal end of the L protein are involved in L-L interactions.[33] Sequence analysis of the L protein revealed that the N-terminal end of L protein is highly conserved among paramyxoviruses and is thought to be involved in enzymatic activities such as binding to P and NP proteins to form a complex that plays a major role in virus budding process.[40] It is believed that the L protein may be multifunctional in nature playing a significant role in viral transcription and replication.[39]

1.3.6 V protein

The V protein is formed from the P gene when the mRNA transcriptase skips a nucleotide base, G (guanidine), at position 484. Therefore, the N-terminal half of both the V and P proteins share the same amino acid sequences, but the C-terminal amino acids are highly variable.[31] In the past the V protein was categorized as a "non-structural protein" because it was found only in the virus infected avian host cells but not in the virus particles. Later in 1991, Samson and his group concluded that the V protein is a structural protein found in the virions; however, the concentration of the protein was found to be too low therefore, they were not able to establish where in the virion the V protein resides. The exact function and role of the V protein in NDV life cycle is still not completely understood. [41]

So far, the structure of the V, M, P, NP, and L proteins encoded by the Newcastle disease virus have not been determined because the study of these proteins is difficult due to their large size.[34, 35, 37, 38] Therefore, the research here is focused on studying the structure of the V protein since it is the smallest protein encoded by NDV. Also, the V protein is found to inhibit the avian immune response system against viral infection.[14] Therefore, determining the structure of V protein would help in synthesizing a drug that can inactivate the V protein, thereby increasing the host immune response against viral infection.

1.4 V Protein

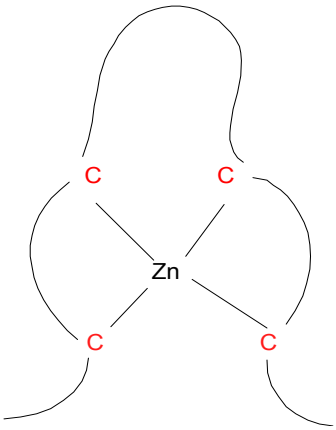
The V protein of NDV is found only in infected avian host cells and has a molecular weight of 25 kDa. Sequence analysis revealed that the N-terminal 164 amino acid residues of the V protein are identical to the P protein where as the C-terminal amino

acids of both the V and P proteins are highly variable.[42] To date, the biological function of the V protein is not known, however, it is believed that it might play a multifunctional role. Since the N-terminal end of the V protein is identical to the P protein it is speculated that the V protein may also bind to the L protein analogous to that of the P protein.[43] Sequence analysis studies on the V protein revealed that the C-terminus is well preserved among the paramyxoviruses and contains cysteine (C) rich regions. The C-terminal region of NDV V protein contains seven cysteine residues, and are arrangement in a manner similar to a zinc finger motif.[31] Figure 1.6 shows the cysteine rich C-terminus amino acid sequence encoded by the NDV V protein and the schematic representation of the zinc finger motif found in the NDV V protein. These cysteine rich motifs show high affinity for zinc, which was proved by metal binding assays carried out on Simian virus (SV) V protein. The assay confirmed that the unique C-terminal domain of the V protein binds to zinc more strongly than other divalent metal ions such as magnesium, calcium, and cadmium. However, the role of this zinc binding is still not established. The main drawback of this metal binding assay is that it does not give any quantitative information about how many zinc atoms are binding to one molecule of V protein.[43] Therefore, using inductively coupled argon plasma-atomic emission spectrometry (ICAP-AES) it was confirmed that two zinc atoms bind to one molecule of SV V protein. To understand, the role of cysteines present in the unique C-terminus domain of V protein in binding with zinc, Paterson[8] and his research group carried out mutations to alter the cysteine residues to alanine. However, the SDS-PAGE (sodium dodecyl sulfate-polyacryl amide gel electrophoresis) results suggested that the cysteine altered V proteins did not express after

purification, indicating that changing the cysteine residues to alanine may have distorted the structure of the V protein.[8]

GHRREHSISYTMGGVTTISVCNPSCSPIRAZPRQYSCTCGSCPATCRLCASASDDV

(a)



(b)

Figure 1.6: (a) cysteine rich C-terminus amino acid sequence encoded by the NDV V protein and (b) schematic representation of zinc finger motif found in the NDV V protein.[44]

Studies on the C-terminal domain of the SV V protein suggested that it interacts with damage specific DNA-binding protein (DDB), which is known to repair the damaged DNA (deoxy ribonucleic acid). DDB is a heterodimer consisting of DDB1 and DDB2 subunits. It was proposed that when a cell contains damaged DNA, DDB binds to it and repairs it and during this process the cell cycle is slowed down. Lamb *et al.*[45] have concluded that when HeLa T4 cells were expressed with V protein changes in the cell

cycle were observed; however, when the HeLa T4 cells were expressed with the V protein lacking in conserved C-terminal region no changes were observed in the cell cycle.[45] Further, the X-ray crystallography studies suggested that SV V protein forms a complex with a subunit of DDB1. DDB1 has a three beta-propeller (BP) structure and the propellers are named (A to C) as BPA, BPB, and BPC. The SV V is found to interact with the BPC. The BPC has a pocket opening, which accommodates the N-terminal helical region of the V protein, indicating that the V protein has specific binding to DDB1. It is speculated that the binding of V protein to DDB1 alters the DNA repair function of DDB. These studies suggest that V protein interacts with DDB, although to date, the biological function of V protein is not fully understood.[44]

Until now, the known anticipated function of the V protein is the inhibition of the host immune response system to viral infection.[42] The interferon (IFN) system is a major component of the host defense mechanism against viral infection, which activates antiviral IFN regulatory factors, such as IFN-alpha/beta that inhibit virus replication and thereby control the spread of viral infection in the host.[14] Huang[14] and his research group have demonstrated that the C-terminal portion of the V protein of the NDV inhibits the IFN signaling pathway necessary for synthesizing the antiviral IFN regulatory factors, by lowering the concentration of signal transducers and activator of transcription (STAT1) protein.[14] To study this possibility, parental NDV and recombinant NDV lacking the V protein were induced in human 2fTGH cells. The separated cell lysates were transferred from SDS-PAGE to a nitrocellulose membrane and tested against STAT1 specific antiserum. The results suggested that the parental NDV infection lowered the

concentration of STAT1 protein compared to the recombinant virus lacking the V protein. Further, the role of C-terminus domain of the V protein in the STAT1 protein degradation was studied by inducing the human 2fTGH cells with the NDV V protein and the V protein lacking in C-terminus domain. The protein expression tests suggested that the C-terminus of the NDV V protein lowered the concentration of the STAT1 protein compared to the V protein lacking in C-terminus domain. These studies indicate that the carboxyl terminus of the NDV V protein acts as an IFN antagonist (blocking the IFN signals) by lowering the concentration of the STAT1 protein, resulting in viral infection in the host. However, the mechanism by which the V protein degrades the STAT1 protein still remains unknown.[14]

Mutation analysis studies suggested that by altering a single amino acid residue in the V protein sequence of Nipha virus (NiV) altered the ability of the V protein to function as an IFN antagonist. The mutation of E125 to G in the C-terminus region of the NiV V protein resulted in a protein that still facilitated IFN signaling blockage, indicating that glycine (G) plays an important role in the capability of the V protein to function as an IFN antagonist. However, the mutation of P206 to K in the C-terminus region of the V protein resulted in the protein that did not lower the concentration of STAT1 protein. This suggests that altering a single amino acid residue in the V protein sequence completely distorts the IFN antagonist function of the V protein.[46] However, this IFN antagonist function of V protein is observed only in avian host cells but not in human cells.[16]

Many biologists have proposed that the V protein might play a role in RNA synthesis because it shares a common N-terminal domain with the P protein, which is known to be involved in RNA synthesis and replication. To investigate this hypothesis, mini genome RNA plasmids with and without SV V protein were constructed. The plasmids were then transfected into T7 cells and polyadenylated to form mRNA. The mRNAs were then tested for reverse transcriptase-polycyclic reactions (RT-PCR). The results suggested that the cells transfected with plasmids without V protein produced a large number of PCR products that were confirmed from the intensity of the SDS-agarose band, whereas the cells transfected with plasmids with V protein showed less intensity bands. This possibly suggests that V protein plays an inhibitory role in RNA synthesis and replication. However, what causes the V protein to inhibit the RNA synthesis is still not obvious.[47]

The V protein is also found to play an important role in the virulence of the virus.[48] This was established by carrying out V protein virulence studies in animal models. When SV was transnasally inoculated into mice it caused deadly pneumonia compared to recombinant viruses without V protein. Similar studies established that NDV replicated faster in embryonated chicken eggs compared to the mutated virus, which was incapable of expressing the V protein. Studies carried out on the mutated measles virus without the V protein indicated that it reproduced at a slower rate in the rat lungs compared to the wild type measles virus, which reproduced at a faster rate and caused high mortality rate. These studies indicate that V protein plays a major role in the

virulence of the virus. However, what function of V protein causes this virulence is still undefined.[48]

The research here is focused on recombinant expression of the V protein of NDV in the bacteria *Escherichia coli* (*E. coli*) and purification of the protein for structural studies. Compared to the other NDV viral proteins, the V protein is the smallest protein having a molecular weight of 25 kDa,[42] therefore, it is a good target for structural studies. Moreover, not much research has been done on the V protein, therefore, the availability of the structure of the V protein would help in designing drugs that can inactivate the V protein thereby increasing the host immune response system against viral infection.

1.5 Recombinant expression and purification of V protein

The deoxyribonucleic acid (DNA) sequence encoding the V protein was cloned into a Novagens pET-24b expression vector that was obtained from Ms. Trudy Morrison, department of Molecular Genetics and Microbiology, University of Massachusetts Medical School. The DNA encoding the V protein is obtained from an Australian NDV strain, which contains 717 nucleotides and encodes a 239 amino acid polypeptide, having a molecular weight of 27 kDa. The vector is a double stranded circular DNA capable of autonomous replication, which can be cut using restriction enzymes for insertion and expression of the target DNA.[49] Initially, the purification of the V protein cloned into the Novagens pET-24b vector was attempted by Dr. John K. Young (Department of Chemistry, Mississippi State University), however, the results suggested that the protein was insoluble and formed inclusion bodies. To solve this problem, the DNA sequence coding for the V protein was cut from the pET-24b vector using restriction enzymes and

inserted into a Novagens pET-14b vector. The pET-14b vector is designed in such a way that when the DNA of interest is cloned into the specific restriction sites it produces a protein having an N-terminal His-tag sequence.[49] The His-tag contains six consecutive histidine residues that are tagged to the protein of interest in order to bind to the specific column material. Adding a His-tag to the protein may increase the solubility of the protein because the His-tag sequence contains polar amino acids such as serine and positively charged amino acids such as arginine. The His-tag also helps in the purification of the protein by binding the target protein to the nickel resin column.[50] The pET-24b vector was restriction digested (RD) with the restriction enzymes Nde I and BamH I to cut out the DNA that encodes the protein. Figure 1.17 shows the schematic representation of pET-24b plasmid with inserted V DNA sequence.

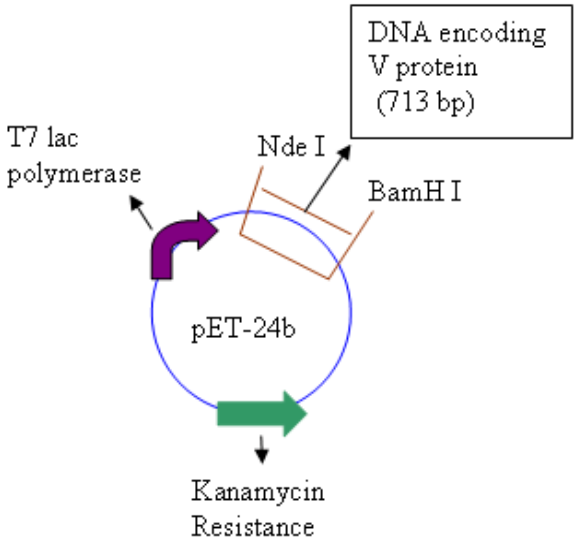


Figure 1.7: Schematic representation of pET-24b vector with inserted V DNA sequence.[49]

The Nde I and BamH I restriction enzymes cut the pET-24b plasmid at a specific site such that the DNA fragment for V protein is obtained. The cut sites for Nde I and BamH I restriction enzymes are shown in the Figure 1.8. The digested DNA sample was then loaded on an agarose gel that contained ethidium bromide. After the gel electrophoresis, the gel was then observed under a UV (ultraviolet) illuminator. The ability to take the gel picture was not available, so the observations were hand drawn. The agarose gel of pET-24b restricted digested DNA sample is described Figure 1.9. The gel results suggested that the RD samples 2-4 contained correct DNA fragments. The gel showed three bands for these samples and the bands were identified based on the number of nucleotide bases. The first band corresponds to the vector with DNA that encodes the V protein, which appeared around 5 kb (kilo base), the second band corresponds to vector without the V DNA that appeared around 4 kb, and the third band corresponds to V DNA appeared around 1kb. The band representing the V DNA was then cut from the gel using a razor and the V DNA was isolated using a DNA isolation kit..

Figure 1.8: The cut sites for Nde I and BamH I restriction enzymes. Where G, A, T, and C stand for nucleotides gaunidine, adenosine, thymidine, and cytosine. The arrows indicate the cut site.[49]

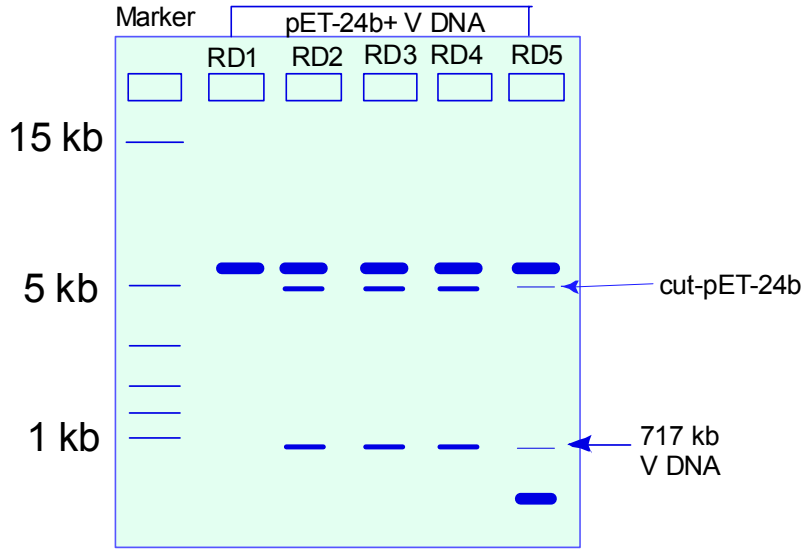


Figure 1.9: Agarose gel of restriction digested pET-24b DNA sample. Where RD indicates restriction digestion samples (1-5).

In order to insert the V DNA into the cloning site of pET-14b vector, the pET-14b vector need to be cut open at the specific cloning site. Therefore, the vector was first restriction digested to be cut open at specific cloning sites using the restriction enzymes Nde I and BamH I. The digested pET-14b vector DNA sample was then loaded on an agarose gel to check whether the pET-14b vector was cut open. After the gel electrophoresis, the results suggested that RD samples 2 and 3 contained the correct size bands. The first band corresponded to the pET-14b vector DNA that appeared around 5 kb and the second band corresponded to the cut pET-14b vector, which appeared close to the first band because the cut DNA sequence was very short (about 9 bases). Therefore,

the cut DNA fragment was hardly noticeable on the gel. Figure 1.10 describes the agarose gel of pET-14b sample after restriction digestion. The bands representing the cut pET-14b DNA was then cut and the DNA was isolated.

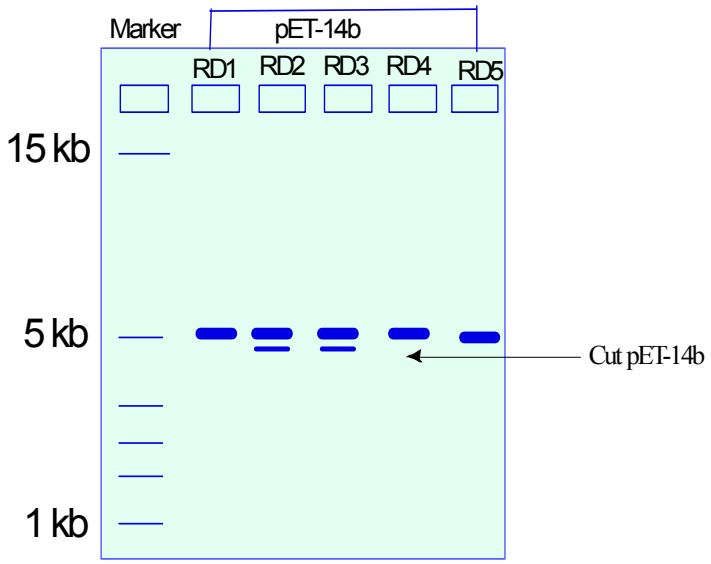


Figure 1.10: Agarose gel of pET-14b vector after restriction digestion. Where RD indicates restriction digestion samples (1-5).

Next, the DNA that encodes the V protein and the cut pET-14b vector have been joined to form a plasmid and this process of joining the two DNA fragment together is known as ligation.[49] The DNA that encodes the V protein was ligated into the cut pET-14b vector by using Novagens T4 DNA ligase. The T4 DNA ligase is a bacterial enzyme that catalyzes the ligation of two DNA fragments involving covalent bond formation

between 3'-hydroxyl groups and 5'- phosphate groups.[49] The schematic representation of isolation of V DNA from pET-24b vector is shown in Figure 1.11. The schematic representation of restriction digestion of pET-14b vector is shown in Figure 1.12. The schematic representation of ligation of V DNA into cut pET-14b vector is shown in Figure 1.13. The ligation reaction mixture was then tested on an agarose gel for the right plasmid with V DNA. Figure 1.14 describes the agarose gel of ligated samples of pET-14b vector with V DNA. The gel results suggested that the ligated (L) samples 2 and 3 have correct pET-14b plasmid with V DNA. Therefore, those ligated samples were then transformed into the *E. coli* cells for protein overexpression.

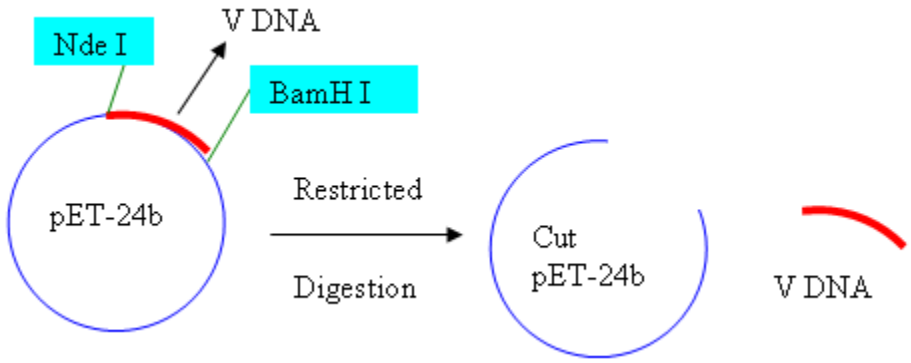


Figure 1.11: Schematic representation of isolation of V DNA from pET-24b vector.

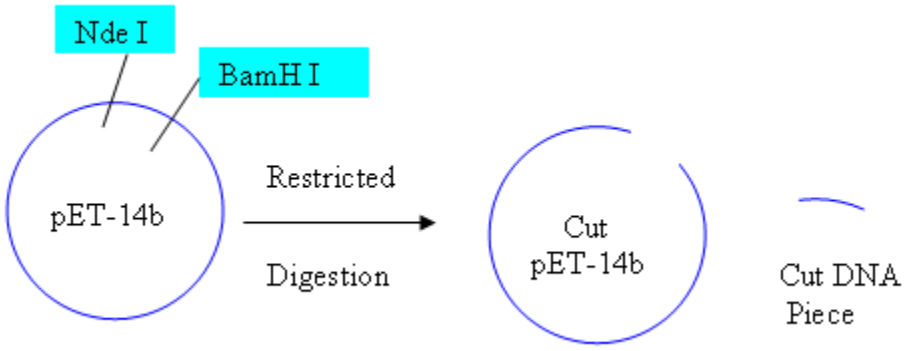


Figure 1.12: Schematic representation of restriction digestion of pET-14b vector.

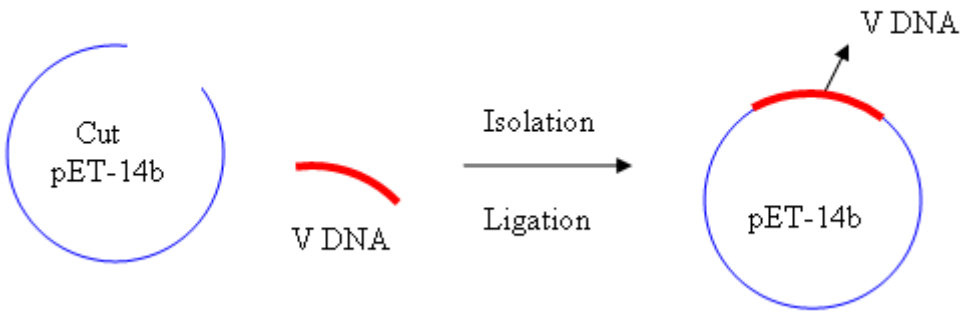


Figure 1.13: Schematic representation of ligation of V DNA into cut pET-14b vector.

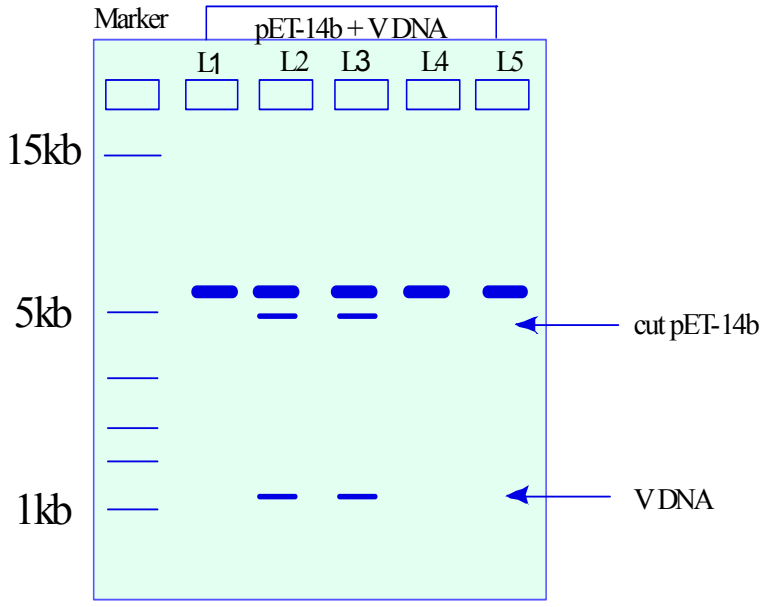


Figure 1.14: Agarose gel of digested samples of pET-14b vector with V DNA. Where L denotes the ligated samples (1-5).

Transformation is a process by which foreign DNA is inserted into a bacterial cell.[48] The ligated DNA is transformed into the XL1 blue cells (*E. coli* strain).[50] The transformation was performed according to protocol [51] and the cells were spread on Luria-Bertani (LB) agar plate with ampicillin antibiotic. The pET-14b vector is designed in such a way that it gives the cells harboring the plasmid resistance to ampicillin antibiotic. After overnight incubation, the cells that had resistance to ampicillin antibiotic grew on the LB agar plate. To isolate the plasmid DNA from the cells, single colonies of the transformant cells were extracted from the LB agar plate and inoculated into LB broth

media with ampicillin antibiotic. The LB media is a nutrient rich media, therefore the cells grow at a faster rate.[51] The plasmid DNA was then isolated from the cells and some amount of isolated plasmid DNA was again restriction digested with enzymes Nde I and BamH I to check whether the cells have the correct plasmid with V DNA. The gel electrophoresis results suggested that the cells have the right plasmid with V DNA. The correct plasmid with V DNA was then transformed into Rosetta cells and the colonies were tested for protein overexpression. The Rosetta cells were selected for transformation because they have the ability to express rare codons like GGA into amino acid that it codes.[49]

15.1 V Protein Overexpression

Protein overexpression is a process in which the DNA sequence encoded by a protein is transcribed into mRNA, which is then translated into the amino acid sequence.[49] To test for protein expression, the culture with plasmid DNA is induced with IPTG (isopropyl beta-D-1-thiogalactopyranoside). IPTG induces the activity of the T7 RNA polymerase gene present in the host cells (rosetta cells), which then binds to the lac promoter, encoded by the vector. This binding then triggers the lac promoter to translate the target DNA into the protein, which is then analyzed by SDS-PAGE.[49] Figure 1.15 shows the schematic diagram of the V protein overexpression. Both the before and after induction samples of the cultures were loaded on the SDS-PAGE. Figure 1.16 shows SDS-PAGE gel for the V protein overexpression in LB. Since adding His-tag would increase the molecular weight of the V protein; therefore, the protein is expected to have

a molecular weight higher than 27 kDa. After confirming that the cells were able to express the V protein, a glycerol stock of rosette cells expressing the V protein was made. All further growths were made from the glycerol stock. To get large amounts of protein, culture stock was grown at large scale in M9 (minimal) media and tested for protein overexpression. M9 is a minimal nutrient media; therefore, the cells grow at a slow rate, synthesizing their own amino acids, thereby giving more time for the protein to become soluble in the media. And also, this media can be prepared using labeled carbon or nitrogen sources, which are required for protein structural studies by NMR (nuclear magnetic resonance) spectroscopy.[51] The SDS-PAGE results suggested that the cells were able to express the target protein in M9 media, therefore, the overnight culture growth was centrifuged and the cell pellet was collected and stored at –80 °C. Figure 1.17 shows SDS-PAGE gel of V protein overexpression grown in M9 media.

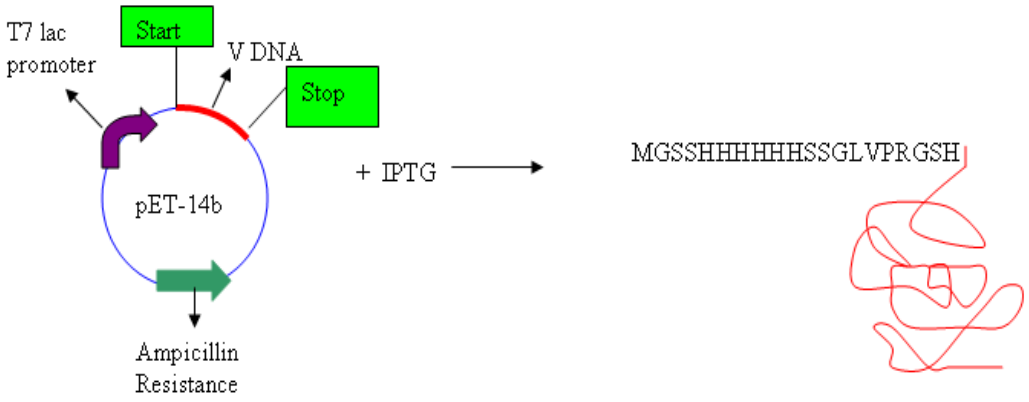


Figure 1.15: Schematic diagram of the V protein overexpression showing the His-tag.

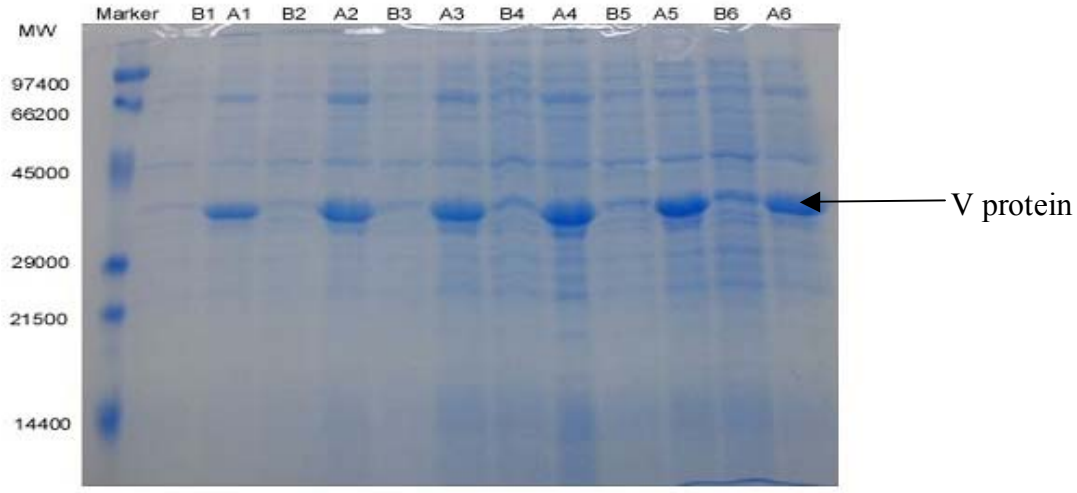


Figure 1.16: SDS-PAGE gel of the V protein overexpression in LB. Where B and A indicate the before and after IPTG induction samples of colonies 1-6. The numbers on the left side of the gel indicate the size of the maker.

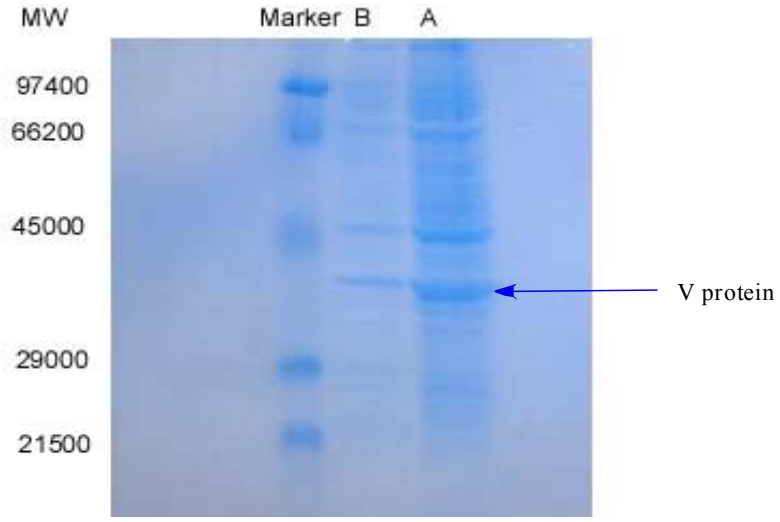


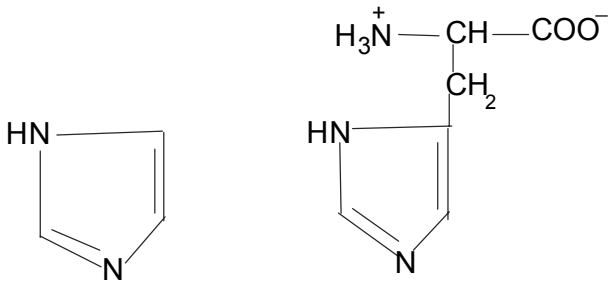
Figure 1.17: SDS-PAGE gel of the V protein overexpression grown in M9 media. Where B and A indicate the before and after IPTG induction samples.

In order to study the structure of the protein, the target protein must be obtained in pure form, free from other protein molecules in the complex mixture. Therefore, the purification of the protein is an important step for the structural studies of the target protein. The protein purification process involves different steps, which separate the target protein from other protein molecules in the mixture based on their size, chemical and binding properties.[52] The chromatographic techniques used in this research are size-exclusion and affinity chromatography.

1.5.2 Purification of unlabeled V protein

The cell pellet expressing the V protein was purified by first dissolving the cell pellet in 10 mM Tris, pH 7 buffer, then sonicated for cell lyses, and the solution was centrifuged. To the supernatant sample solution, ammonium sulfate salt is added and stirred. Ammonium sulfate salt is used in the protein purification process to remove any cell debris through precipitation.[53] The sample solution was again centrifuged and the supernatant solution was purified by using a Ni-NTA (nickel-nitrilotriacetic acid) agarose column chromatography. The His-tagged proteins are generally purified using Ni-NTA column chromatography because the matrix is composed of Ni-NTA attached to Sepharose® CL-6B. The His-tag is expected to bind tightly to the Ni in the matrix because histidines contain the imidazole ring as a part of their structure, which shows high affinity for Ni due to the presence of lone pair of electrons on nitrogen. Therefore, the unwanted proteins can be eluted out.[54] The chemical structure of histidine and imidazole is shown Figure 1.18. The Ni-NTA resin shows significant selectivity to proteins containing His-tag.[543] The column material was prepared by suspending Ni-NTA

agarose beads in wash buffer, which contained 10 mM Tris, pH 7 buffer with low concentration (10 mM) of imidazole and small amount of zinc. Adding imidazole to the buffer improves the purification process because it binds to the Ni in the column matrix and minimizes the binding of untagged proteins.[54] After the column is equilibrated with wash buffer, the sample was then loaded onto the column and then the column was washed with 20 column volumes to remove unbound proteins. The flow through was collected and tested for protein by SDS-PAGE. After this step, the outlet was connected to the fraction collector and the column was run using an elution buffer, which contained 10 mM Tris, pH 7 buffer with high concentration of imidazole (50mM). At a high concentration of imidazole, the affinity between the imidazole in the buffer and the Ni is increased, which helps to elute the His-tagged proteins from the column matrix.[54] The column effluent was then collected in the fraction collector. To test for the target protein, a sample of each fraction collected was analyzed by SDS-PAGE. The SDS-PAGE gel of



Imidazole Histidine

Figure 1.18: Chemical structure of histidine and imidazole.[54]

V protein expression after Ni-NTA column purification is shown in Figure 1.19. The SDS-PAGE results suggested that most of the protein did not bind to the column, which is concluded from the sample flow through (SFT) and from the flow through (FT) collected after the wash. The fractions collected after the sample elution suggested that sample amount of protein did bind to the column (seen in fraction 2); however, the protein was not obtained in a pure form. These results suggest that there may be a possible interaction between the histidines present in the His-tag and the zinc present in the buffer. The SDS-PAGE results also suggested that some amount of protein did bind to the column (shown in column for fraction 2); however, it was not pure. Therefore, the purification of V protein was attempted using a size-exclusion chromatography technique.

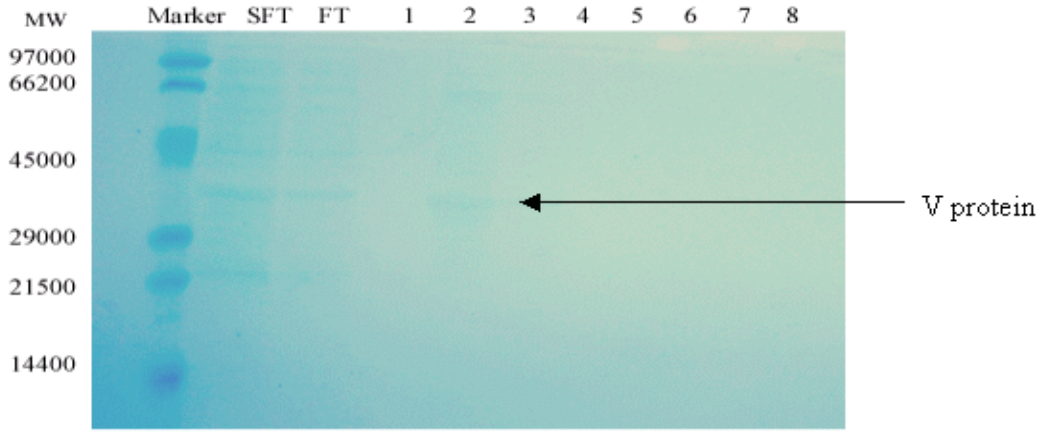


Figure 1.19: SDS-PAGE V protein expression gel for fractions collected from the Ni-NTA column. Where SFT and FT indicate the columns for sample flow through and wash flow through, and the numbers 1-8 indicate the fraction of the samples collected from the Ni-NTA column.

Size-exclusion or Gel filtration chromatography is one of the widely used techniques for protein purification. The Gel filtration separates molecules based on their size. The large molecules are separated first, followed by the small molecules.[53] The column is usually packed with chemically inert porous spherical particles (a few micrometers (μm) in size). To perform this separation, the column is first equilibrated with the mobile phase, so that the porous of the matrix particles and the space between particles is filled with the mobile phase. After equilibrating the column with mobile phase, the sample was then loaded on the column. During the flow through, the sample molecules diffuse in and out of the column matrix. The small sample molecules diffuse further into the matrix because of their small size and therefore they are retained longer in the column matrix. The large molecules on the other hand are unable to diffuse into the column matrix because of their large size and therefore they are passed through the column at a faster rate along with the mobile phase and are eluted first.[55] The schematic diagram of gel filtration column separating different sized molecules is shown in Figure 1.20.

The gel filtration column used in the purification of V protein is a Hiload ™ 16/60 Superdex ™ 75 (Amersham Biosciences). It is made of glass and has length of 60 cm and internal diameter of 16 mm. The Superdex column matrix contains dextran, which is covalently bonded to 34 μm cross-linked agarose beads. The column has high efficiency and good resolution for proteins having a molecular weight range of 3000-70000.[56] The calculated molecular weight of NDV V protein (239 amino acids long) is 27516 g/mol, therefore, Hiload™ 16/60 Superdex™ 75 column was chosen. The column was first

equilibrated with the mobile phase, which contained 50 mM Tris, pH 7 buffer and then the sample was injected manually into the sample injector loop. The sample components were then separated based on their size. To find the fractions that contain the protein of right size, the samples were analyzed by SDS-PAGE. The gel filtration chromatogram of V protein is shown in Figure 1.21. After the SDS-PAGE electrophoresis, the fractions that contain the protein were identified. The SDS-PAGE gel of the fractions that contained the protein after gel filtration purification is shown Figure 1.22. The SDS-PAGE results suggested that the fractions 29-35 contain some amount of V protein and also the V protein could be obtained in a pure form, as shown by fractions 28 and 29. Therefore, the purification of the V protein was restricted only to the gel filtration chromatography.

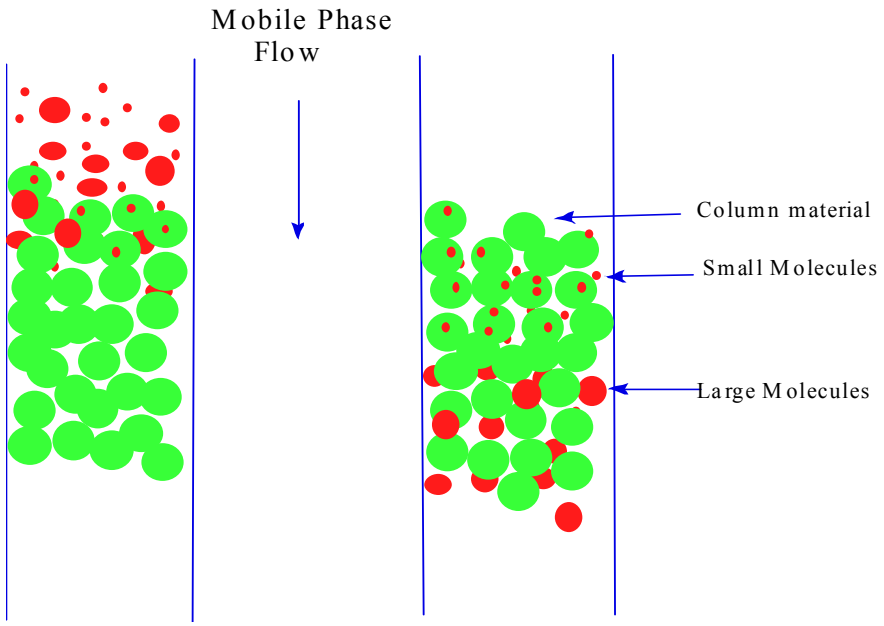


Figure 1.20: Schematic diagram of gel filtration column separating different sized molecules.[53]

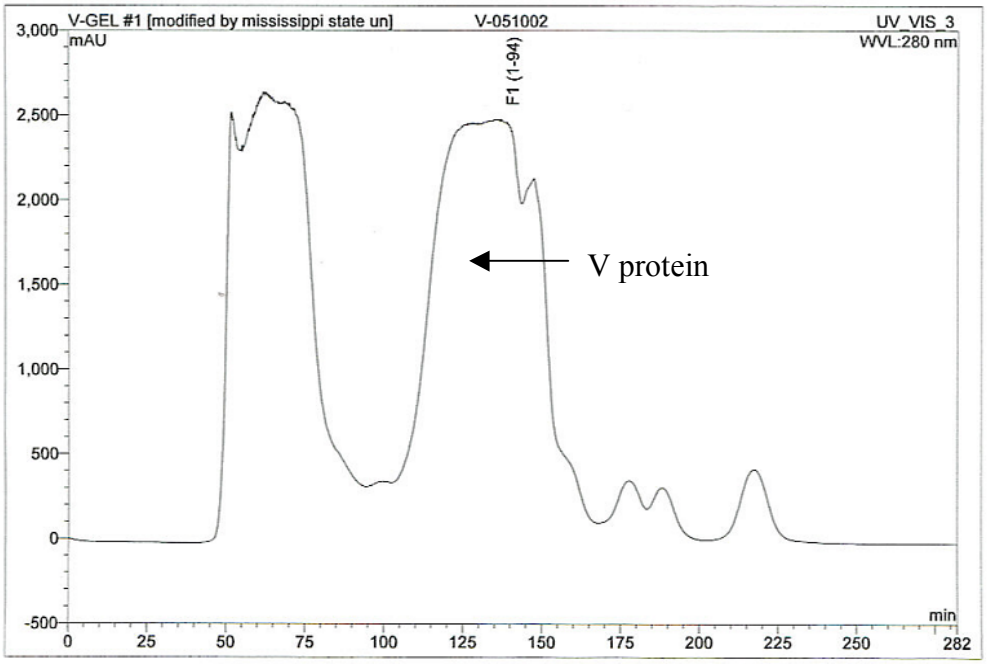


Figure 1.21: Gel filtration chromatogram of the V protein

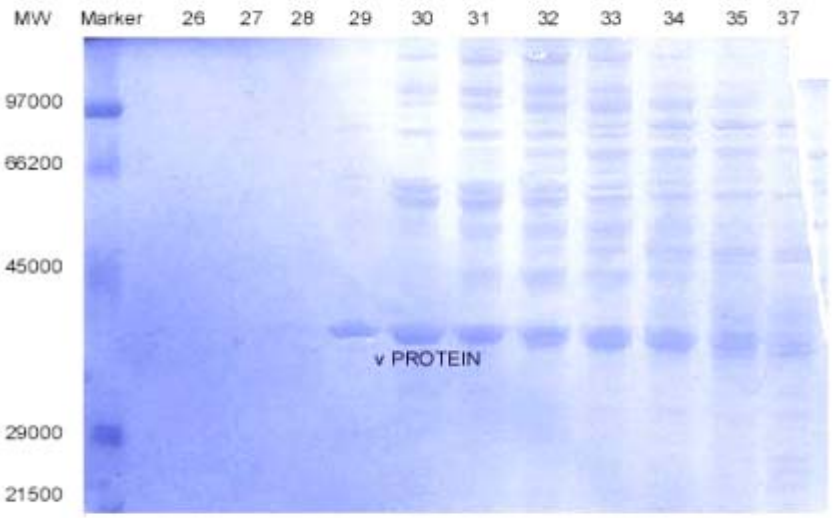


Figure 1.22: SDS-PAGE gel of the fractions that contained the V protein after gel filtration purification.

1.5.3 Purification of ^{15}N labeled V protein

After developing a gel filtration purification method for V protein, the culture stock expressing the V protein was then grown in labeled M9 media, which contained ^{15}N labeled ammonium chloride. The labeled nitrogen is required for structural studies of protein by NMR spectroscopy.[51] The culture was then tested for protein overexpression. Both the before and the after induction samples of the cultures were loaded analyzed by SDS-PAGE. The SDS-PAGE gel of V protein overexpression grown in ^{15}N labeled M9 media is shown in Figure 1.23. The SDS-PAGE results suggested that the cells were able to express the target protein therefore, the overnight culture growth was centrifuged and the cell pellet was collected and stored at –80 °C.

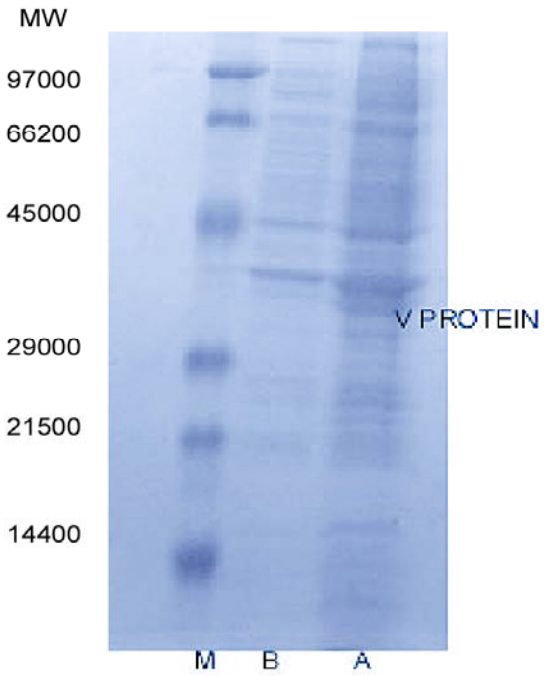


Figure 1.23: SDS-PAGE gel of V protein overexpression grown in labeled M9 media.

The cell pellet expressing the V protein was then purified using gel filtration chromatography. The experiment was repeated in a similar manner as discussed in the previous sections. SDS-PAGE gel of labeled V protein expression after the purification by gel filtration technique is shown in Figure 1.24. The SDS-PAGE results suggested that

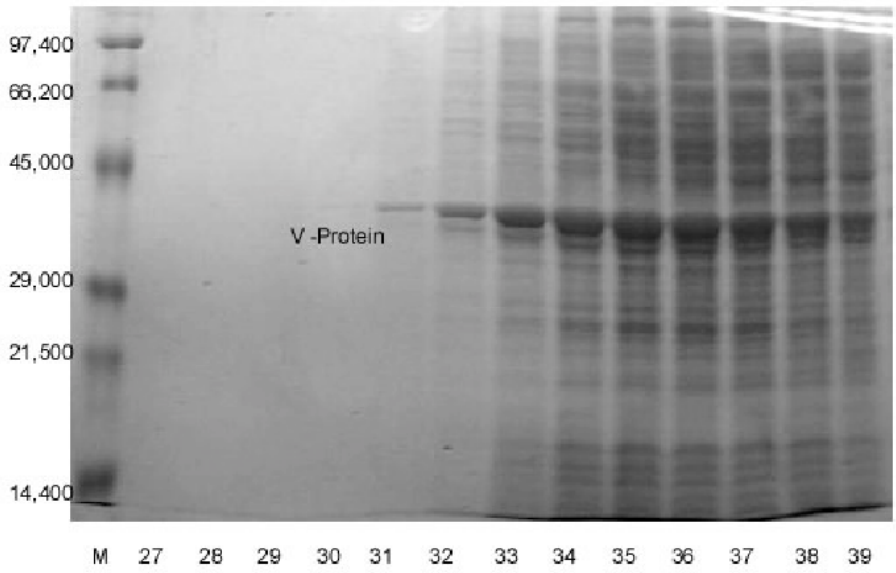


Figure 1.24: SDS-PAGE gel of labeled V protein expression after the protein purification by gel filtration technique.

the fractions 30-32 contain the V protein without any other non-target proteins. Therefore, the fractions 30-32 were mixed together and there concentration (0.5 mM) was measured using UV-Visible spectrophotometer. The sample solution (30 ml) was further concentrated by using a sample concentrator till the sample volume was reduced to 0.5 ml. The 0.5 ml concentrated sample was then transferred into a vial, frozen in liquid

nitrogen and stored at -80° C. This purified V protein sample was then used for NMR studies. The remaining solutions from lanes 33-39 were mixed and purified again using the gel filtration technique. The samples collected from the column were analyzed again for the right size protein using SDS-PAGE. Figure 1.25 shows the SDS-PAGE gel of

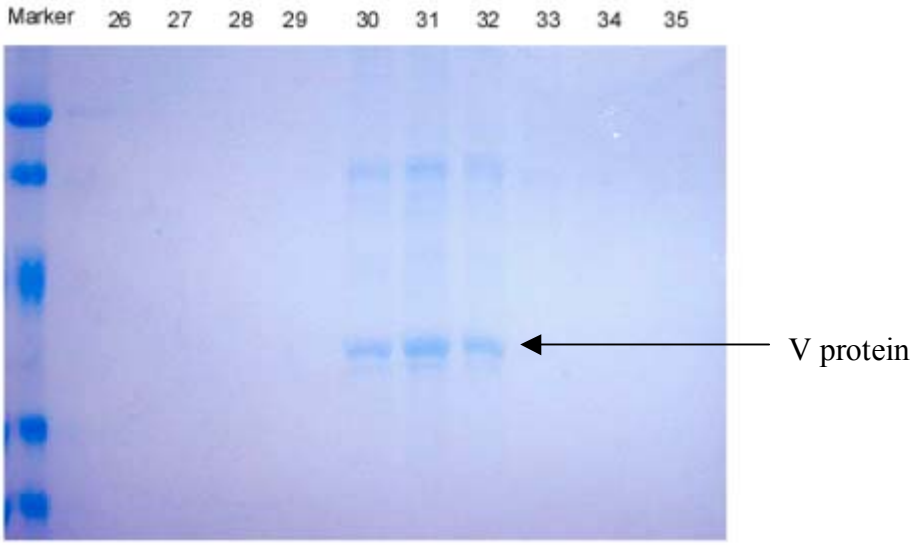


Figure 1.25: SDS-PAGE gel of labeled V protein expression after the concentrated protein is purified by gel filtration technique.

labeled V protein expression after the sample mixture (33-38) was purified by gel filtration technique. The SDS-PAGE results (Figure 1.24) suggested that the samples 30-32 contain the V protein. Therefore, the samples from 30-32 were mixed and their concentration (0.2 mM) was measure using UV-Visible spectrophotometer. The sample solution was then transferred into vials, frozen in liquid nitrogen, and stored at -80° C.

This V protein sample was then used for CD (circular dichroism) and fluorescence analysis, to study the secondary structure of the V protein.

1.6 Structural analysis of V protein

To study the structure of the protein various techniques are now available.[57] Florescence spectroscopy and Circular Dichroism (CD) were used in this research to obtain the basic information about the secondary structure of the V protein. After obtaining a basic idea about the secondary structure of the V protein, more detailed structural information about the protein was obtained using NMR.

1.6.1 Fluorescence Spectroscopy

Fluorescence spectroscopy is widely used to study the secondary structure of proteins because it is highly sensitive and data acquisition is easy.[58] Fluorescence spectroscopy involves excitation of a molecule to a higher energy level by using a light of certain wavelength and this electronically excited molecule rapidly relaxes to the ground state by emitting the absorbed radiation. Therefore, the emission spectrum is independent of the absorption spectrum.[58] Molecules that fluoresce are known as flourophores and the most intense fluorescence is shown by flourophores that contain aromatic hydrocarbons. Aromatic hydrocarbons have π-π^* transitions that requires lower energy and therefore, absorption occurs at longer wavelength. Hence π^*-π transitions have greater quantum efficiency.[57, 59] This technique is applied to study the secondary structure of proteins because proteins contain aromatic amino acids such as tryptophan, tyrosine, and phenylalanine. When compared to tyrosine and phenylalanine, tryptophan shows absorption maximum due to the presence of indole ring system in the structure that

involves π-π* transitions. Tryptophan is found to have maximum absorption at wavelength (λ) of 280-290 nm and therefore, the fluorescence emission of tryptophan is typically observed, even if the protein contains all three aromatic amino acid residues.[60] If tryptophan is absent then the fluorescence emission of tyrosine is studied because it shows absorption maximum than phenylalanine.[58, 59] The fluorescence emission of tryptophan is sensitive to its environment in the protein and this is used as a probe of the protein secondary structure analysis.[60] Normally, tryptophan by itself shows a maximum fluorescence emission at a wavelength of 350 nm in polar solvents. If the protein is unfolded, tryptophan is more exposed to free water molecules therefore it shows maximum fluorescence emission near 350 nm. However, if the protein is folded, the tryptophan residues may be subjected to rigid surroundings, which causes a decrease in the wavelength of fluorescence emission of tryptophan.[58] The maximum fluorescence emission observed for tryptophan residues present in folded proteins is near 320-330 nm. In proteins, which are random coiled the maximum fluorescence emission for tryptophan is observed near 340 nm, which suggests that tryptophan residue may be exposed to free water molecules.[60]

Figure 1.26 shows the fluorescence spectrum of V protein. The fluorescence emission spectrum of the V protein shows a maximum emission at 340 nm, which suggests that the V protein is random coiled. The V protein contains 4 tryptophans, which may be present at the surface of the V protein and in contact with the polar groups present inside the protein molecule. Therefore, it is concluded that V protein may be random coiled in structure.

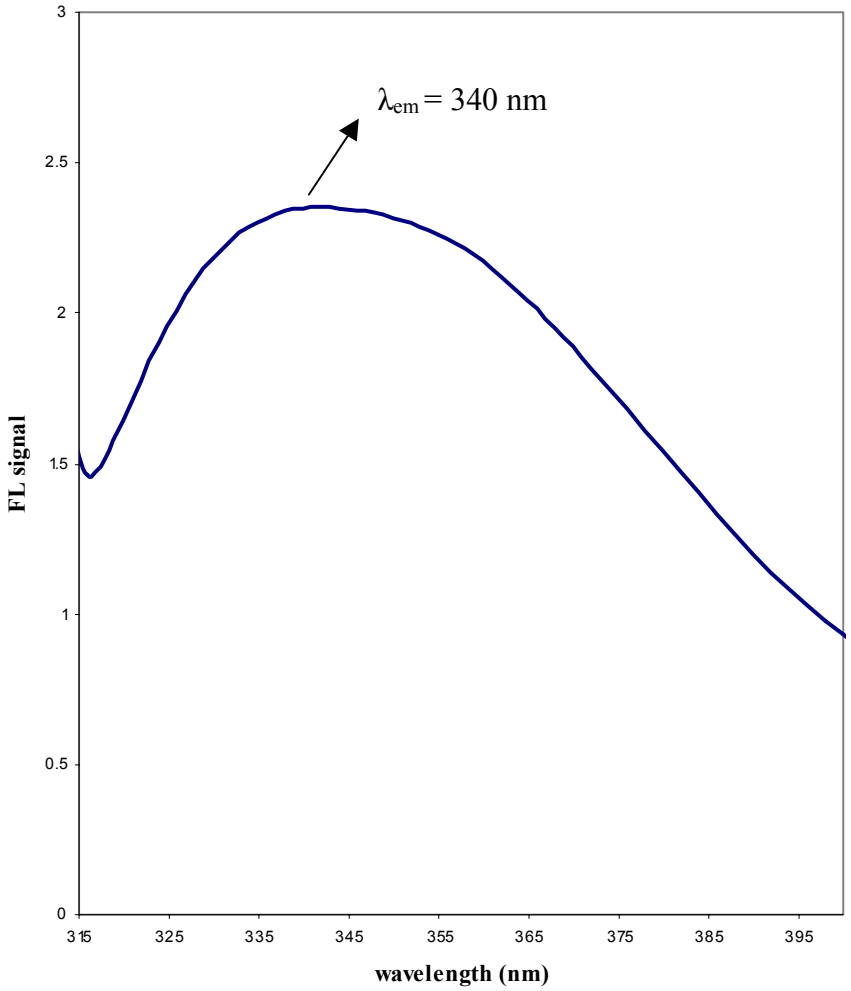


Figure 1.26: Fluorescence spectrum of the V protein.

1.6.2 Circular Dichroism (CD)

To get more information about the secondary structure of the V protein, CD was used. CD is a powerful tool that gives an insight into the secondary structure of proteins.[61] CD technique involves passage of circularly polarized light through the sample. The direction of the circularly polarized light is split using a modulator into two components: the right handed and left handed circularly polarized light. The modulator is usually a piezoelectric quartz crystal. After passing the two components of circularly polarized through the sample, the difference in the absorption of right handed and left handed circularly polarized light by the sample is measured by the CD instrument. In other words, if the sample absorbs any one of the circularly polarized components more than the other, then the subsequent combination of the two components gives rise to an elliptically polarized light. The CD spectrum then displays the ellipticity of the transmitted light in degrees (θ) at a given wavelength.[62]

For a molecule to be CD active, it should be optically active. Only chiral molecules are optically active and therefore they are CD active. Since all the amino acids except glycine are chiral in nature, proteins are CD active.[61] Proteins contain amide, aromatic and disulfide chromophores that are CD active. Among these chromophores, the amide group is the most abundant since it is a part of the backbone structure of the polypeptide chain in proteins. The π-π* and *n*-π* transitions shown by the amide chromophore in the proteins absorb strongly in the far UV CD region (190-250 nm); therefore, this region forms the finger print region for the secondary structure analysis of proteins. The aromatic chromophores dominate the near-UV CD region (250-300 nm)

and the sulphide chromophores dominate the region above 300 nm and therefore these regions in the CD spectrum give information about the tertiary structure of the proteins.[61] From the CD analysis of proteins of known secondary structure, it is established that if the protein shows two negative bands at approximately 222 nm and 210 nm, and a positive band at approximately 192 nm, that a helical protein structure is suggested. If the protein shows one negative band at approximately 215 nm and a positive band at approximately 196 nm, that a beta strand protein in structure is suggested.[62] However, if the protein shows a positive peak near 218 nm and a negative peak near 200 nm, that a unstructured protein in structure is suggested.[61] Figure 1.27 shows the reproduced CD spectrum of alpha, beta, and random coiled structural proteins. The CD spectrum of the V protein is shown in Figure 1.28.

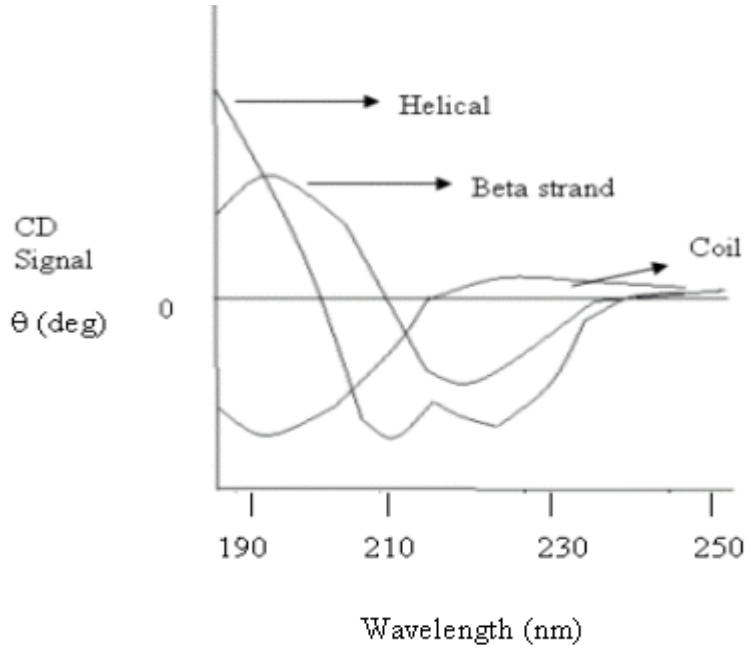


Figure 1.27: CD spectrum shown by alpha, beta, and random coiled structural proteins.[61]

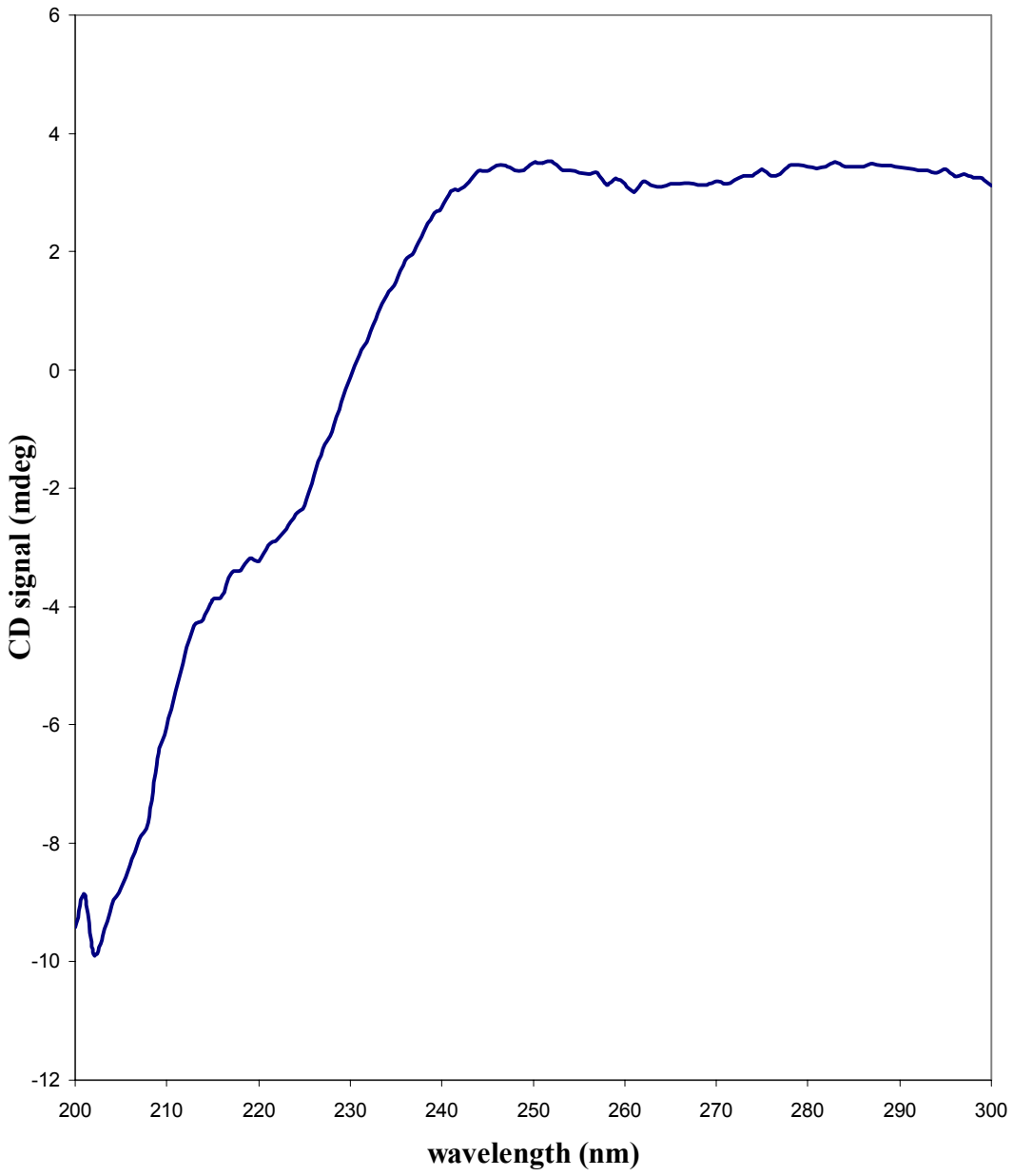


Figure 1.28: CD spectrum of the V protein

The CD spectrum of the V protein shows a negative peak around 202 nm and a positive peak around 247 nm. Even though, the positive CD signal is observed at a higher wavelength for the V protein compared to 218 nm, which is usually observed for random coiled proteins; however, the pattern of the CD signal looks similar to the proteins having random coiled structure. Therefore, it is concluded that V protein may be largely unstructured. Both, the fluorescence and the CD analysis of the V protein gave similar results and suggest that the V protein is unstructured or random coiled.

1.6.3 NMR

To find whether the V protein is folded or unfolded, a high-resolution technique such as NMR was used. NMR involves the interaction between electromagnetic radiation and nuclei placed in an external magnetic field. Nuclei having spin ½, such as ^{1}H (proton), ^{13}C (carbon), ^{15}N (nitrogen), and ^{31}P (phosphorus) are generally used in NMR studies because they have only two energy levels; therefore, the energy transfer is more efficient than the nuclei having spin greater than ½.[57] When radio frequency pulses are applied to these nuclei placed in an external electromagnetic field, they absorb the radio frequency and start resonating. Since the resonating frequency of every nuclei is different in a given magnetic field strength, studying the resonance frequencies of the nuclei in a given molecule can give information about the structure of the molecule.[63] The NMR experiment normally considered with an isotope-labeled protein is a two-dimensional heteronuclear single quantum correlation (2D HSQC) experiment because it is quick and the data obtained can be quickly analyzed.[64] The experiment involves measurement of the chemical shift correlations between the proton and a heteronuclei (^{15}N or ^{13}C).[57] Since the

V protein is ^{15}N labeled, the chemical shift correlations between the amide ^{1}H and ^{15}N is measured using 2D ^{1}H-^{15}N HSQC. Every amide proton directly attached to the amide nitrogen appears as a cross peak in the 2D -^{1}H-^{15}N HSQC spectrum. Except proline, all the backbone amino acid residues have directly bonded amide proton to amide nitrogen bond. Therefore, the number of cross peaks in the 2D ^{1}H-^{15}N HSQC spectrum should be equal to the number of amino acid residues present in the protein.[57, 64]

The pulse sequence for the 2D ^{1}H-^{15}N HSQC experiment, is shown in Figure 1.29. The pulse program is "the insensitive nuclei enhanced by polarization and transfer" (INEPT) pulse. After the first 90° pulse on the ^{1}H, the magnetization that originates on the NH proton is transferred from amide ^{1}H to ^{15}N through one bond N-H coupling constant ($^{1}J_{(NH)}$). The ^{15}N magnetization that is evolved during this time period (t_1) is refocused to the amide ^{1}H by applying an 180° pulse on ^{1}H in the middle of this period.[57] After this step, both the ^{1}H proton and ^{15}N chemical shifts are decoupled using a decoupling pulse sequence. At this point, a gradient pulse is applied to eliminate water, which can be effectively dephased. During this period, which is a constant time t_1 period, ^{15}N nuclei chemical shifts are evolved. Finally, the magnetization is transferred from ^{15}N back to the amide protons using a reverse-INEPT pulse sequence and during this time period that is t_2 period, proton chemical shifts are evolved and are recorded. This is followed by a group of preservation of equivalent pathways (PEP) pulses to increase the amount of data. The 2D ^{1}H-^{15}N HSQC spectra are plotted taking ^{1}H chemical shifts on the X-axis and ^{15}N chemical shifts on theY-axis.[57, 63] If the protein is unfolded, the peaks are all grouped in one region of the spectrum due to the overlapping of the ^{1}H and ^{15}N chemical shifts;

however, if the protein is well folded, the ^{1}H-^{15}N cross peaks are dispersed in the spectrum and each peak is evidently visible. Therefore, 2D ^{1}H-^{15}N HSQC experiment gives information about the basic secondary structure of the protein.[57]

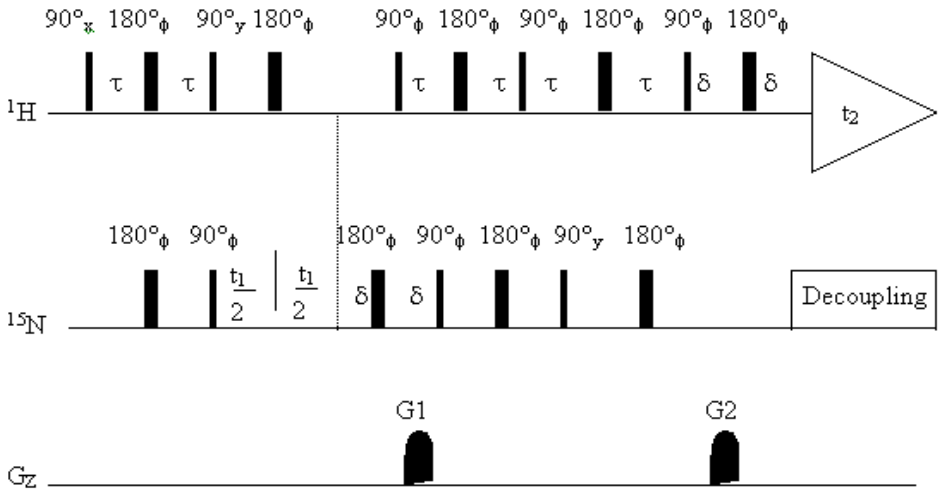


Figure 1.29: Pulse sequence involved in 2D ^{1}H-^{15}N HSQC. The thin and thick lines indicate the 90° and 180° pulses and t_1 and t_2 indicate the evolution period for nitrogen and protein chemical shifts. The G1, and G2 indicate the gradient pulse applied on water. The τ and δ are delay times.[57]

The 2D ^{1}H-^{15}N HSQC spectra of the V protein suggests that most of the peaks are grouped in one region having a proton chemical shift of 8 ppm (parts per million) and nitrogen chemical shift of 124 ppm, which is a characteristic feature of random coiled protein.[64] This suggests that the V protein may be random coil in structure. Figure 1.30 shows the 2D ^{1}H-^{15}N HSQC spectrum of the V protein. The two peaks aligned horizontally in a straight line having nitrogen chemical shift 112 ppm indicate the

presence of NH_2 groups because for each N-H group a single peak is observed in the 2D ^{1}H-^{15}N HSQC spectrum. Although there are few peaks that are well spread in the spectrum; however, a conclusion can be drawn that the V protein may be random coiled in structure.

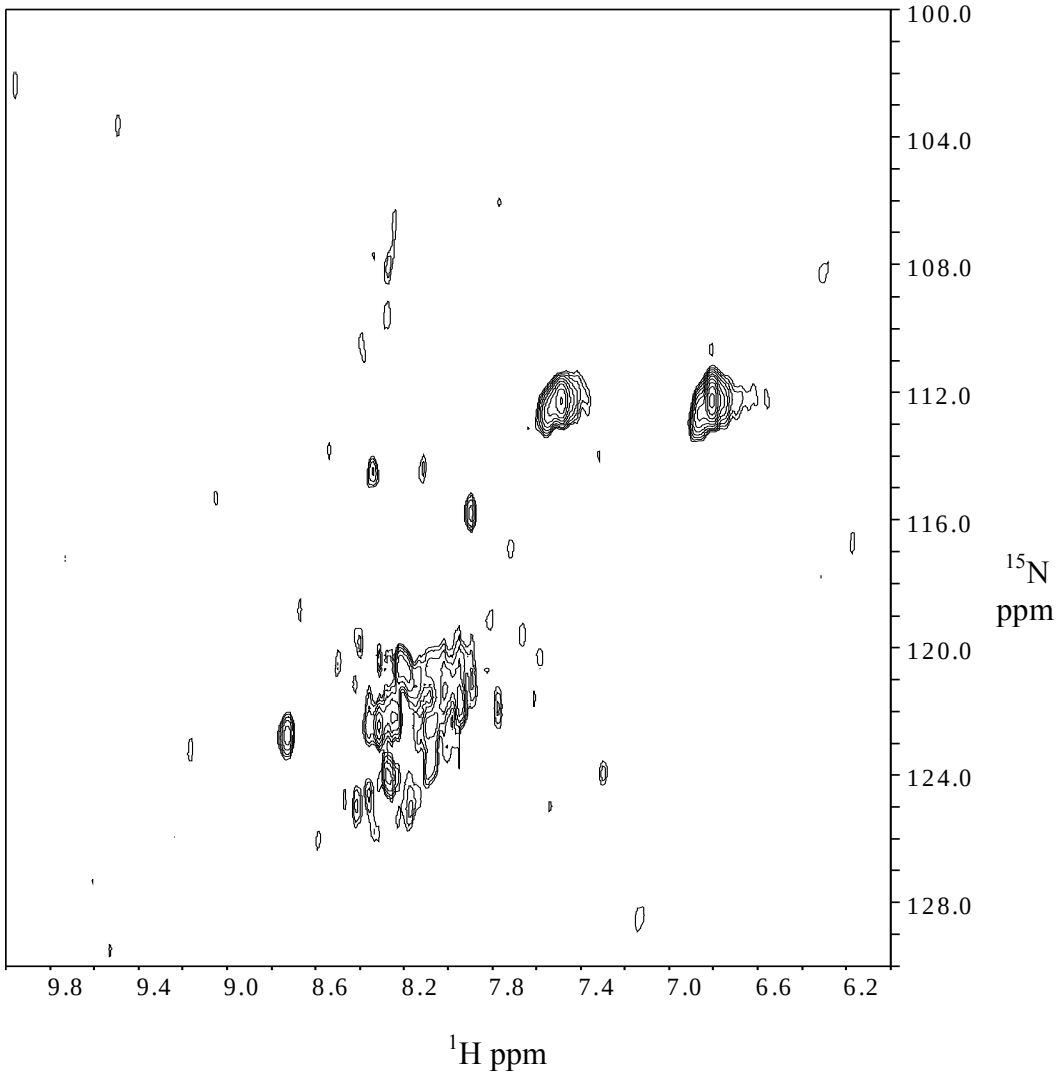


Figure 1.30: 2D ^{1}H-^{15}N HSQC spectrum of the V protein. Where ^{1}H chemical shifts are plotted on the X-axis and ^{15}N chemical shifts on the Y-axis.

1.7 Conclusion

The results obtained from fluorescence, CD, and 2D ^{1}H-^{15}N HSQC spectroscopy suggests that the V protein may be unstructured. Since the V protein is found only in the avian infected embryo cells, there is a possibility that the V protein may be unstructured or the V protein may be structured in its natural environment. Many native factors may possibly influence the structure of the V protein. There is a possibility that the protein may be structured in presence of other proteins such as DDB, STAT1 etc. due to protein-protein interactions.

There is another method, which can be used to purify the target protein. In some studies, [65, 66] NDV proteins were purified using a glutathione S-transferase (GST) tag method. GST is a small domain that is tagged to the protein of interest similar to the His-tag method. Once the GST-tagged protein is isolated, it then purified by using GST•Mag^{TM} purification kit. In order to do this, the recombinant DNA needs to be redesigned. This process involves several steps such as polycyclic chain reaction (PCR) in order to find the correct restriction site, transformation into BL-21 cells, isolation of the target DNA and the purification of the target protein. This whole process may take 3-6 months and also, to cut the target protein from the GST-tag, thrombin enzyme is used, which results in the protein that starts with ArgGly (arginine and glycine) instead of M (methionine).[49] Adding ArgGly to the protein may possibly disrupt the structure of the protein and also if the target protein has ArgGly in the sequence, then the thrombin may cut those sites resulting in small proteins. Therefore, the GST-tag method should be investigated.

REFERENCES

1. Aacheson, N. H; Richardson, C; D. Kolakofsky, *Fundamental Virology*. Third Edition ed.; Lippincott-Raven Publishers: Philadelphia, **1996,** 577-604.

2. B. S. Seal; D. J. King; H. S. Sellers, *Dev Comp Immunol* **2000,** 24, 257-268.

3. T. G. Morrison, *Virus Res* **1988,** 10, 113-135.

4. http://en.wikipedia.org/wiki/Paramyxoviridae (accessed May 2007).

5. N. Bennett; J. Ellis; C. Bonville; H. Rosenberg; J. Domachowske, *Immunization Strategies for the Prevention of Pneumovirus Infections*. 2007, 6, 169-82.

6. K. B. Chua, *J Clin Virol* **2003,** 26, 265-275.

7. http://www.wingswest.net/pigeons/health/paramyxo.html (accessed May 2007).

8. R. G. Paterson; G. P. Leser; M. A. Shaughnessy; R. A. Lamb, *Virology* **1995,** 208, 121-131.

9. M. J. H. Edward K. Wagner *Basic Virology*. 2 ed.; Blackwell: MA, USA, 2004.

10. E. Villar; I. M. Barroso, *Glycoconj J* **2006,** 23, 5-17.

11. T. Takimoto; A. Portner, *Virus Res* **2004,** 106, 33-45.

12. http://jnci.oxfordjournals.org/cgi/content/full/91/20/1708 (accessed May 2007).

13. C. J. Pedersen, Senne, D. A., Woolcock, P. R., Kinde, H., King, D. J., Wise, M. G., Panigrahy, B., Seal, B. S., *J Clin Microbiol* **2004,** 42, 2329-2334.

14. Z. Huang; S. Krishnamurthy; A. Panda; S. K. Samal, *J Virol* **2003,** 77, 8676-8685.

15. M. Park, Shaw, M. L., Munoz-Jordan, J., Cros, J. F., Nakaya, T., Bouvier, N., Palese, P., Garcia saastre, A., Basler, C. F., *J virol* **2003,** 77, 1501-1511.

16. M. S. Park; A. Garcia-Sastre; J. F. Cros; C. F. Basler; P. Palese, *J Virol* **2003,** 77, 9522-9532.

17. N. S. Millar; P. Chambers; P. T. Emmerson, *J Gen Virol* **1986,** 67, 1917-1927.

18. J. P. M. Langedijk, Daus, F. J., Oirschot, J.T.V., *J Virology* **1997,** 71, 6155-

6167.

19. H. Connaris; T. Takimoto; R. Russell; S. Crennell; I. Moustafa; A. Portner; G. Taylor, *J Virol* **2002,** 76, 1816-1824.

20. C. J. Russell; T. S. Jardetzky; R. A. Lamb, *J Virol* **2004,** 78, (24), 13727-13742.

21. T. G. Morrison, *Biochim Biophys Acta* **2003,** 1614, 73-84.

22. K. A. Gravel; T. G. Morrison, *J Virol* **2003,** 77, 11040-11049.

23. K. A. Baker; R. E. Dutch; R. A. Lamb; T. S. Jardetzky, *Mol Cell* **1999,** 3, 309-319.

24. J. K. Young; R. P. Hicks; G. E. Wright; T. G. Morrison, *Virology* **1998,** 243, 21-31.

25. L. W. McGinnes; T. G. Morrison, *J Virol* **2006,** 80, 2894-2903.

26. J. K. Young; Abramowitz, M. C.; Morrison, T.G., *J virology* **1999,** 73, 5945-5956.

27. L. W. McGinnes; T. G. Morrison, *Virology* **1987,** 156, 221-228.

28. P. Chambers; N. S. Millar; S. G. Platt; P. T. Emmerson, *Nucleic Acids Res* **1986,** 14, 9051-9061.

29. H. D. Pantua; L. W. McGinnes; M. E. Peeples; T. G. Morrison, *J Virol* **2006,** 80, 11062-11073.

30. G. Mottet; V. Muller; L. Roux, *J Gen Virol* **1999,** 80, 2977-2986.

31. M. Steward; I. B. Vipond; N. S. Millar; P. T. Emmerson, *J Gen Virol* **1993,** 74, 2539-2547.

32. F. Jahanshiri; M. Eshaghi; K. Yusoff, *Arch Virol* **2005,** 150, 611-618.

33. B. Cevik; S. Smallwood; S. A. Moyer, *Virology* **2007,** 363, 189-97.

34. C. L. Kho; W. S. Tan; B. T. Tey; K. Yusoff, *Arch Virol* **2004,** 149, 997-1005.

35. B. S. Seal; J. M. Crawford; H. S. Sellers; D. P. Locke; D. J. King, *Virus Res* **2002,** 83, 19-29.

36. T. M. Myers; A. Pieters; S. A. Moyer, *Virology* **1997,** 229, 322-335.

37. C. L. Kho; W. S. Tan; B. T. Tey; K. Yusoff, *J Gen Virol* **2003,** 84, 2163-2168.

38. W. Errington; P. T. Emmerson, *J Gen Virol* **1997,** 78, 2335-2339.

39. T. Takimoto; T. Bousse; A. Portner, *Virus Res* **2000,** 70, 45-53.

40. K. Yusoff; N. S. Millar; P. Chambers; P. T. Emmerson, *Nucleic Acids Res* **1987,** 15, 3961-3976.

41. A. C. R. Samson; I. Levesley; P. H. Russell, *J Gen Virol* **1991,** 72, 1709-1713.

42. M. Sun; T. A. Rothermel; L. Shuman; J. A. Aligo; S. Xu; Y. Lin; R. A. Lamb; B. He, *J Virol* **2004,** 78, 5068-5078.

43. P. Liston; D. J. Briedis, *Virology* **1994,** 198, 399-404.

44. T. C. Li, X.; Garbutt, K. C.; Zhou, P.; Zheng, N., *Cell* **2006,** 124, 105-117.

45. G. Y. Lin; R. A. Lamb, *J Virol* **2000,** 74, 9152-9166.

46. K. Hagmaier; N. Stock; S. Goodbourn; L. F. Wang; R. Randall, *J Gen Virol* **2006,** 87, 3649-3653.

47. Y. Lin; F. Horvath; J. A. Aligo; R. Wilson; B. He, *Virology* **2005,** 338, 270-280.

48. S. E. Witko; C. Kotash; M. S. Sidhu; S. A. Udem; C. L. Parks, *Virology* **2006,** 348, 107-119.

49. David L. Nelson, *Leninger Principles of Biochemistry*. fourth ed.; W. H. Freeman: 2004.

50. *Novagen Catalog*. EMD Biosciences: USA, 2006/2007.

51. J. Sambrook, E. F. Fritsch, T. Maniatis., *Molecular Cloning: A Laboratory Manual*. second ed.; Cold Spring Harbor Laboratory Press: Plainview, N.Y., 1989.

52. http://en.wikipedia.org/wiki/Protein_purification (accessed May 2007).

53. R. K. Scopes, *Protein Purification, Principles and Practice*. second ed.; Springer-Verlag: New York Inc., 1987.

54. Q. Inc., *The QIAexpressionist™, A hand book for high-level expression and purification of 6X Histagged proteins*. fifth ed.; QIAGEN: 2003.

55. P. Biotech, *Gel Filtration, Principles and Methods*. 6 ed.; Rahms i Lund: Sweden.

56. http://www.amersham.com/ (accessed May 2007).

57. J. C. W. J. F. A. G. P. N. J. Skelton, *Protein NMR Spectroscopy, Principles and Practice*. Academic press: San Diego, California, 1996.

58. A. S. Ladokhin, in: *Encyclopedia of Analytical Chemistry*, R. A. Meyers, John Wiley & Sons Ltd: Chichester, 2000, 5762-5779.

59. H. Skoog, Neiman, *Principles of Instrumental Analysis*. Fifth ed.; Brooks /Cole: Singapore, 2005, 355-379.

60. R. P. A. Brockhinke, P. Dittrich, K. Kohse-Hoinghaus, *Appl Phys* **2000,** 71, (B), 755-763.

61. K. N. Nina Berova, Robert W. Woody, *Circular Dichroism, Principles and Applications*. second ed.; Wiley & Sons: New York, 2000.

62. N. C. P. Sharon M. Kelly, *Biochim Biophys Acta* **1997,** 1338, 161-185.

63. http://en.wikipedia.org/wiki/Nuclear_magnetic_resonance(accessed May 2007).

64 David, G. Reid, *Protein NMR Techniques,* Humana Press, New Jersey,1997.

65. P. Lie; J. Zhu; F. Gao; P. Tien; Z. Rao; G. F. Gao, *Acta Cryst* **2003**, D5a, 1296-1298.

66. S. H. Hassan; C. Wirblich; M. Forzan; P. Roy, *J Virol* **2001**, 75, 8356-8367.

Chapter II

Purification and Structural Analysis of Flowering Locus T (FT) Protein

2.1 Flowering

Human life is greatly affected by the flowering of plants, since most food (vegetables, fruits, grains, etc.) results from flowering. Flowers, also used in the textile industry, especially the fibers that results from the flowers are used in making clothing, such as cotton and linen. Some flowers offer great medicinal value because drugs, such as morphine, are obtained from the flowering of plants.[1] Therefore, flowering is an important development in the life cycle of plants. The development of flowers by plants is controlled by time and environmental factors.[2] Classical studies have shown that the photoperiod (short and long day length) plays an important role in floral induction. However, modern studies revealed that in response to the photoperiod a signal transmission substance called FT is produced in the leaf that triggers the pathway for floral development.[2, 3] Scientists believe that the activation of the FT gene in the leaves is dependant on the photoperiod. The FT protein, which is formed from the transcription of the FT gene is found at the apex of the plants shoot and is found to initiate the flowering process. Yet, so far, the pathway followed by the FT protein and how the FT protein initiates the floral development in plants is still unknown.[3] Fascinatingly, it is found that the amino acid sequence of FT protein is similar to the mammalian RAF-kinase inhibitor protein (RKIP). Mammalian RKIP is found to play a central role in the regulation of

several protein synthesis pathways. This suggests that the FT protein may also play a similar role as RKIP in the pathway for flowering.[3]

So far, the structure of the FT protein of poplar tree has not been determined.[3] Therefore, finding the structure of FT protein will help in studying the biological functions of the protein. Structural analysis may also help in determining the possible domains of the FT protein that interact with other proteins, leading to flowering in plants. In order to study the structure of the protein, the protein must be in a pure form. Therefore, the research here is focused on overexpression, purification and initial structural studies of the FT protein to determine if more complex structural studies by NMR are possible.

2.2 FT protein

Most of the research on FT protein was carried out using *Arabidopsis thaliana* (commonly called Arabidopsis) as a model plant because of its small genome size and short life cycle. The life cycle of Arabidopsis takes about 5 weeks starting from germination of the seed to the production of seed.[4] The photoperiod plays a very important role in the reproductive phase of Arabidopsis plants. The leaves of the plants detect the changes in the photoperiod, due to the presence of photoreceptors. The flowering in Arabidopsis takes place in response to long day lengths.[2] Studies suggested that Arabidopsis plants grown under short day lengths when exposed to a single long day length showed early floral induction than plants grown only under short day lengths. These results suggested that the primordial, which is initially supposed to develop into a leaf at the plant shoot apex, has transformed into a floral meristem on exposure to long

day lengths.[2, 3] In Arabidopsis, CONSTANS (CO) and FT genes that are found in the leaves play a central role in floral induction in response to long day lengths. This was confirmed by mutation analysis, which suggested that when CO and FT genes were mutated, the development of floral meristems was delayed even though the plants were exposed to long day lengths.[5] Model studies suggested that on exposure to long day lengths (about 12 hours) the gene CO present in the leaves is expressed, which indicates that the CO protein is acting as a photoperiod sensor. Once the CO is activated it then triggers the expression of the FT gene. The activation of the FT gene then initiates the flowering process. These results suggest that the flowering process is sensitive to photoperiod.[6]

The protein encoded by the FT gene is known as the FT protein, which is approximately 20 kDa in molecular weight.[3] Although the FT gene is expressed in the leaves, the FT protein is found in the floral meristem of the plants shoot apex. This was confirmed from a gene fusion experiment, which involved fusion of genes that encode FT protein and the green flourescent protein (GFP) and expressing them in *ft-7* plant mutants. The results suggested that the *ft-7* plant mutants developed flowers early when compared to *ft-7* plants. Also the fusion protein was detected in the seeds of the *ft-7* plant mutants.[7] The grafting of meristem showed the presence of the FT protein, which indicates that the FT protein moves from leaves to the meristem of the plants. However, the pathway followed by the FT protein is not clearly understood. These studies suggest that the FT protein plays a role in initiating transition from the vegetative to the reproductive state (flowering). [3, 7] Studies on FT also suggested that when the FT mRNA

is overexpressed it caused early floral development and vice versa caused late floral development.[3]

To understand the activity of the FT protein, interactions between FT with other transcription factors have been studied.[2] The hybrid analysis proved that FT interacts with a basic leucine zipper (bZIP) flowering date (FD) protein. The FD protein is expressed in the nucleus of the cells present at the plant shoot apex of seedlings. Interaction of FD and FT protein carried out in yeast cells showed that both the proteins co-exist in the nucleus of the yeast cells.[8] Studies also suggested that the FD protein expresses well in the presence of FT indicating that FT activates the FD protein; however, how it activates is still unclear.[2, 8] There are some speculations that the FD and the FT protein form a transcriptional complex, which in turn induces floral transition at the plants shoot apex. Deletion analysis studies suggest that the C-terminus of the FD protein interacts with the FT protein. Further studies suggested that mutant plants with both FD and FT showed early flowering than plants with only FD indicating that the activity of the FD is dependant on FT.[8] The transition from vegetative leaf to flower is a slow gradual stepwise process. Experiments on plant mutants revealed that the absence of FD protein in wild type Arabidopsis plants showed a delay in the transition from leaf primordial to floral meristem. Where as the introduction of high levels of FT in the meristem, in the absence of FD, showed a direct transition from leaf primordial to floral meristem. The significance of FT in floral initiation was established by mutation analysis. Plants having mutated FT and FD proteins showed complete lack of floral meristem.[3] However, the site of action is different for both the FD and FT protein. The FT protein is

expressed in the cotyledons and the leaves of seedlings, where as the FD protein is expressed only in the shoot apex. This suggests that the FT protein expressed in the leaves is carried to the plant shoot apex, which then interacts with the FD protein and initiates flowering in the plants. However, how the FT protein is carried to the plant shoot apex is still not clearly understood.[8]

The research here is focused on recombinant expression of the FT protein from poplar trees in the bacteria *E. coli* and purification of the protein for initial structural studies. Poplar trees belong to the same family as Arabidopsis, but produce flowers more than once (perennial tree) in their lifetime whereas Arabidopsis produces flowers only once (annual plant) in their lifetime. Poplar trees have a very long life span ranging from 100 to 200 years and produce flowers only after 7 or 10 years. This means that they have a long vegetative phase before the reproductive phase. Once they start flowering, they produce flowers every season. Therefore, studying the structure of FT protein of poplar trees would help to understand how flowering is induced in poplar trees every season in response to the photoperiod.[9]

2.3 Recombinant FT protein expression

The DNA sequence encoding the FT protein was cloned into a pET-14b vector by Ms. Nursen Binbuga from Dr. Cetin Yuceers lab (Department of Forestry, Mississippi State University). The DNA sequence encoding the FT protein under study is isolated from the poplar tree, which encodes 174 amino acids and has a molar mass of 19602.1 g/mol. The schematic diagram of pET-14b vector with inserted FT DNA is shown in Figure 2.1. The DNA sequence analysis carried out by Dr. Yuceers lab suggested that the

DNA sequence encoding the FT was inserted correctly into the vector. The plasmid with FT DNA was then transformed into Novagens BL21(DE3)pLysS cells and tested for protein overexpression by SDS-PAGE. After confirming that the cells were able to express the FT protein, a glycerol stock was made and stored at –80 °C. All further growths were made from this stock. To obtain large amounts of protein, the culture stock was grown at large scale in LB media and tested for protein overexpression (see section *15.1* of Chapter I). Figure 2.2 shows the SDS- PAGE gel of FT protein expression, grown in 1-liter LB media. The SDS-PAGE results suggested that the cells were able to express the target protein; therefore, the overnight culture growth was centrifuged and the

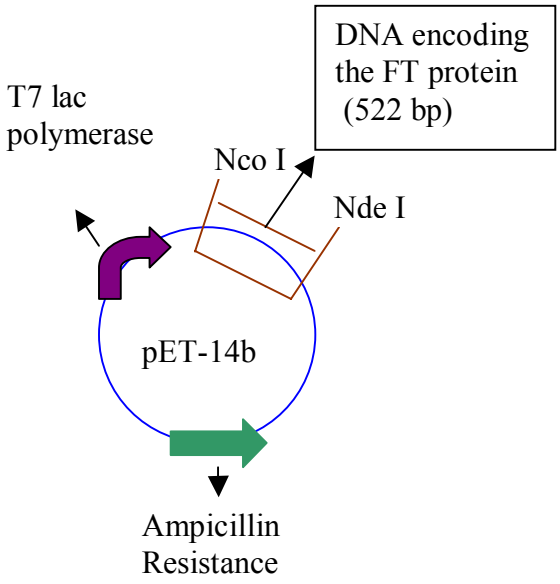


Figure 2.1: Schematic diagram of pET-14b vector with inserted FT DNA.

cell pellet was collected and stored at –80 °C. The cell pellet expressing the FT protein was then subjected to solubility test in 50 mM Tris, pH 7 buffer, which involved stirring and sonication steps. The sample solution was then centrifuged and a sample of both the pellet and the supernatant solution were tested for protein by SDS-PAGE. To the remaining supernatant solution ammonium sulfate ($(NH_4)_2SO_4$) salt at a concentration

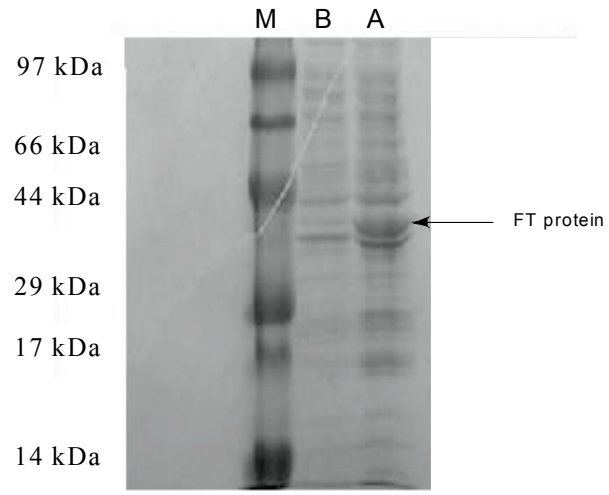


Figure 2.2: SDS-PAGE gel of FT protein overexpression grown in LB media. Where M, B, and A indicate the marker, and before and after IPTG induced samples.

of 10% was added and stirred. $(NH_4)_2SO_4$ is generally used to increase the ionic strength of the solution, which initially helps in precipitating the cell debris. However, higher concentrations of $(NH_4)_2SO_4$ helps in precipitating the soluble proteins.[10] The solution was then centrifuged and a sample of the supernatant solution was tested for protein by SDS-PAGE. Figure 2.3 shows the SDS-PAGE gel of FT protein (cells grown at 37 °C),

after the solubility test. The SDS-PAGE results suggested that most of the protein is present in the pellet. Therefore, to solve this solubility problem, the culture stock expressing the FT protein was grown in LB media at lower temperatures. Growing the cells that express the FT protein at a lower temperature helps in the solubility of the proteins because the cells grow at a slow rate synthesizing their own amino acids, thereby giving more time for the protein to become soluble in the media. From the solubility test it was found that the FT protein was fairly soluble when the cells that express the target

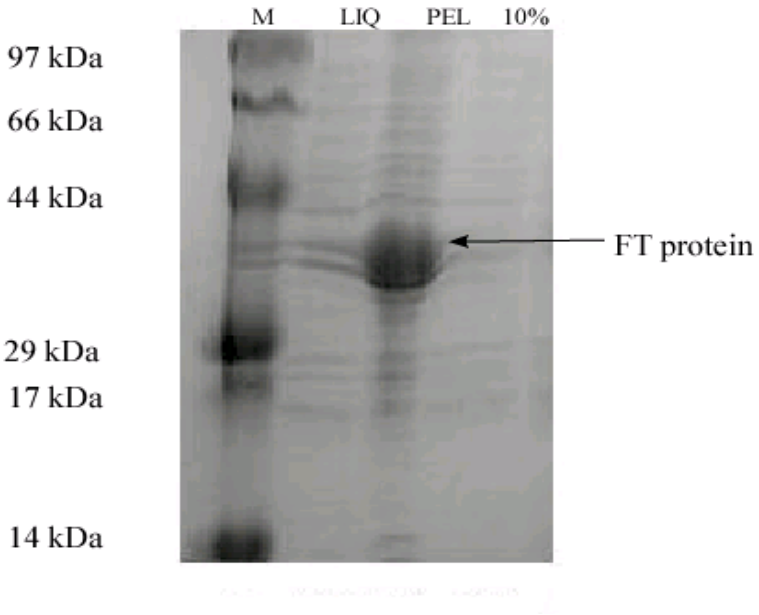


Figure 2.3: SDS-PAGE gel of FT protein (cells grown at 37 °C), after the solubility test. Where M, LIQ, and PEL indicate the columns for the marker, liquid sample, and the pellet sample. The 10% indicates the column for the supernatant sample solution after adding 10 % $(NH_4)_2SO_4$.

protein were grown at 30 °C. The SDS-PAGE gel of FT protein (cells grown at 30 °C in LB), after the solubility test is shown in Figure 2.4. Therefore, the cultural stock containing the cells expressing the FT protein was grown at large scale in M9 media at 30 °C and tested for protein overexpression. The SDS-PAGE results suggested that the cells were able to express the target protein in M9 media; therefore, the overnight culture growth was centrifuged and the cell pellet was collected and stored at –80 °C. Figure 2.5 shows the SDS-PAGE gel of FT protein expression grown in M9 media at 30 °C.

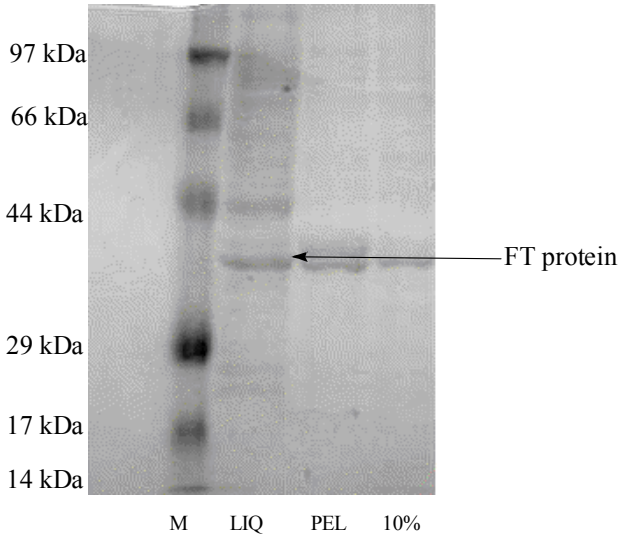


Figure 2.4: SDS-PAGE gel of the FT protein (cells grown at 30 °C in LB), tested for solubility. Where M, LIQ, and PEL indicate the columns for the marker, liquid sample, and the pellet sample. The 10% indicates the column for the supernatant sample solution after adding 10 % $(NH_4)_2SO_4$.

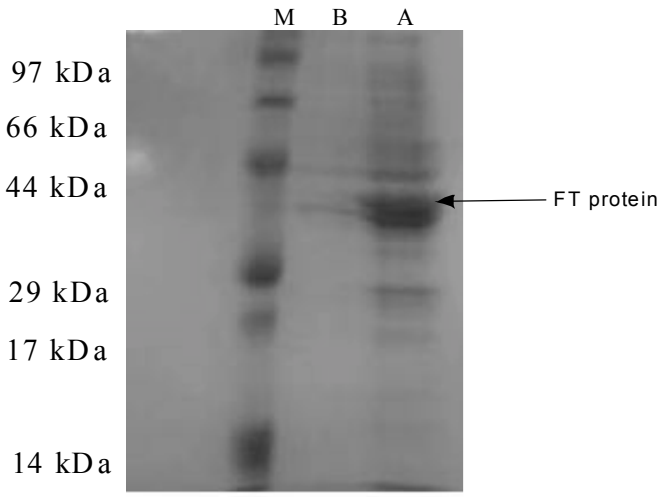


Figure 2.5: SDS-PAGE gel of FT protein overexpression (cells grown at 37 °C in M9 media). Where M indicates the column for the marker, and B and A indicate the before and the after IPTG induced samples.

2.4 Purification of FT protein

In order to study the structure of the protein, the protein must be obtained in a pure form. Therefore, the cell pellet expressing the FT protein grown in M9 media was purified and the purification process involved dissolving the cell pellet in 50 mM Tris, pH 7 buffer containing dithiothreitol (DTT). DTT is a strong disulfide bond reducing agent therefore, adding DTT would help in preventing the disulfide bond formation in the proteins.[11]After the cell pellet expressing the FT protein was tested for solubility, a sample of both the supernatant solution and the pellet was tested for the FT protein by SDS-PAGE. Figure 2.6 shows the SDS-PAGE gel of FT protein expression, after the

solubility test. The SDS-PAGE results suggested that some amount of the FT protein was soluble. Therefore, the remaining supernatant solution was purified by gel filtration

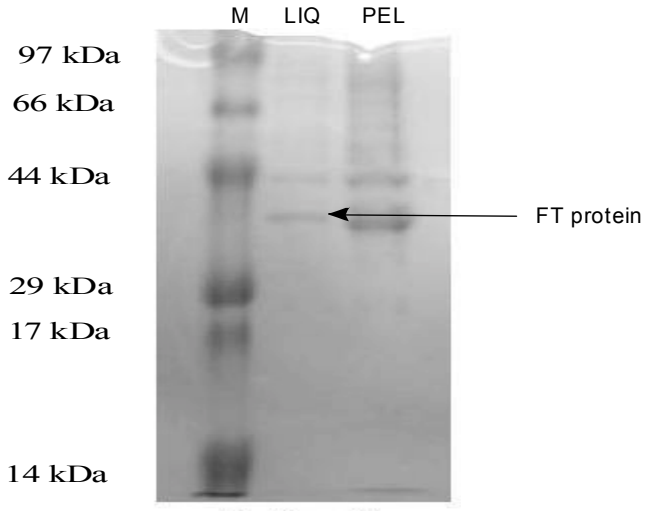


Figure 2.6: SDS-PAGE gel of the FT protein expression (cells grown at 30 °C in M9 media), after the solubility test. Where M, LIQ, and PEL indicate the marker, liquid and pellet sample.

chromatography (see section *1.5.2* of Chapter I). The gel filtration chromatogram of the FT protein is shown in Figure 2.7. The samples collected after the gel filtration purification was tested for the FT protein by SDS-PAGE. The SDS-PAGE results suggested that the fractions 34-43 contain the FT protein. The fractions containing the FT protein were then mixed and the solution was further purified using the ion exchange chromatography. Figure 2.8 shows the SDS-PAGE gel of the fractions that contained the FT protein after the purification by gel filtration technique.

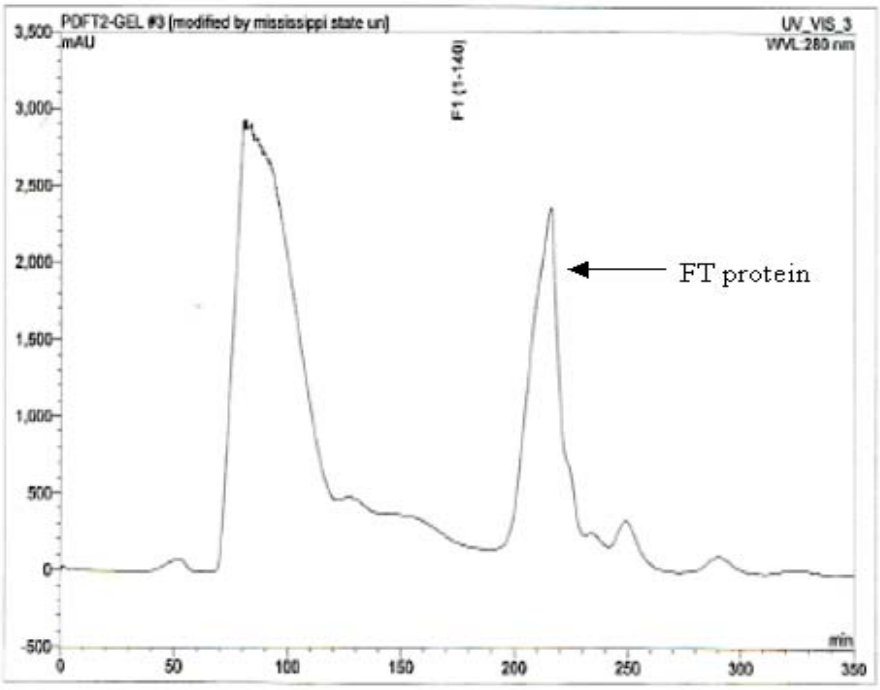


Figure 2.7: Gel filtration chromatogram of the FT protein.

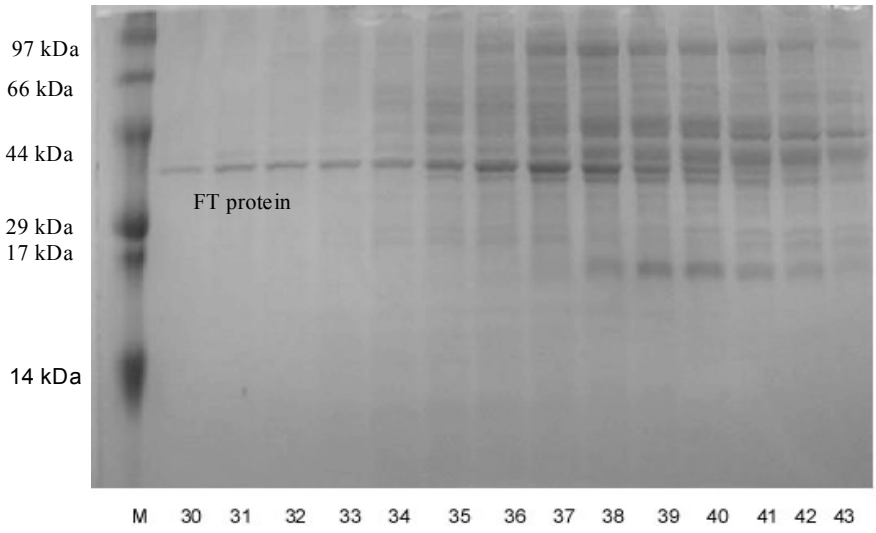


Figure 2.8: SDS-PAGE gel of the fractions that contained the FT protein after purification by gel filtration technique. The numbers 30-43 indicate the number of the fraction collected.

2.4.1 Ion Exchange Chromatography

Ion exchange chromatography is a technique in which molecules are separated based on their net charge.[12] From the amino acid sequence analysis the number of negatively charged groups (aspartic and glutamic acid) were found to be 18 and the number of positively charges groups (lysine and arginine) were found to be 19. Thus, the net charge on the FT protein was found to be +1 positive charge. Therefore, the purification of the FT protein was attempted using cation exchange chromatography. The cation exchange column used was UNO^{TM} S-12 from Bio-Rad. The column is packed with support material (resin) that contains negatively charged sulfonyl groups ($-SO_3^-$) to which positively charged cations are bound. These bound cations are exchangeable, therefore they are known as "counter-ions." [13] Figure 2.9 shows the sulfonyl functional group used in cation exchanger resin. Generally, ion exchange columns are packed with polystyrene resin. The principle behind the cation exchange column chromatography is that when the protein sample containing positively charged groups is run through the column, they bind tightly to the negatively charged column matrix displacing the counter cation, hence the name cation-exchange chromatography.[12] Finally, to elute the protein

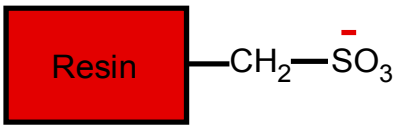


Figure 2.9: Sulfonyl functional group used on cation exchanger resin.[14]

of interest, the concentration of the salt in the buffer is increased, which decreases the interaction between the charged groups of the bound proteins and the column matrix by increasing the interaction between the ions of the salt and the column matrix. The proteins are then eluted in the order in which they bind. The weakly bound proteins are eluted first and the tightly bound proteins are eluted last.[12]

The experiment was carried out first by equilibrating the column with starting buffer (A), which contained 50 mM Tris at pH 7 and DTT. After the column is equilibrated, the sample is then loaded into a sample loop. The unbound substances are then washed out using the starting buffer for 10 minutes. The sample was then eluted out using 50 mM Tris, pH 7 elution buffer (B), containing high concentration (250 mM) of KCl salt. This increases the competition between the potassium ions (K^+) and the positively charged groups present in the proteins that interact with the resin matrix, causing the proteins to elute.[13] This method of eluting the protein with a mobile phase of different composition is known as gradient elution method.[12] The gradient used in this research is a continuous gradient. To elute the bound protein, the gradient was shifted from 100% A to 100% B and then the gradient was holded for 10 minutes and then the column was again re-equilibrated with wash buffer. The cation exchange chromatogram of FT protein is shown in Figure 2.10. The sample fractions collected were tested for protein by SDS-PGE. The SDS-PAGE results suggested that the FT protein did not bind to the cation exchange resin, which was evident from the chromatogram that the protein was eluted out in the wash buffer. This suggests that the positive charge might be buried inside the protein molecule and the surface charge of the protein may be negative.

Therefore, there was no exchange of counter ion. Thus, the protein was not retained on the column and therefore, it was eluted out. Therefore, a purification attempt was made using an anion exchange column.

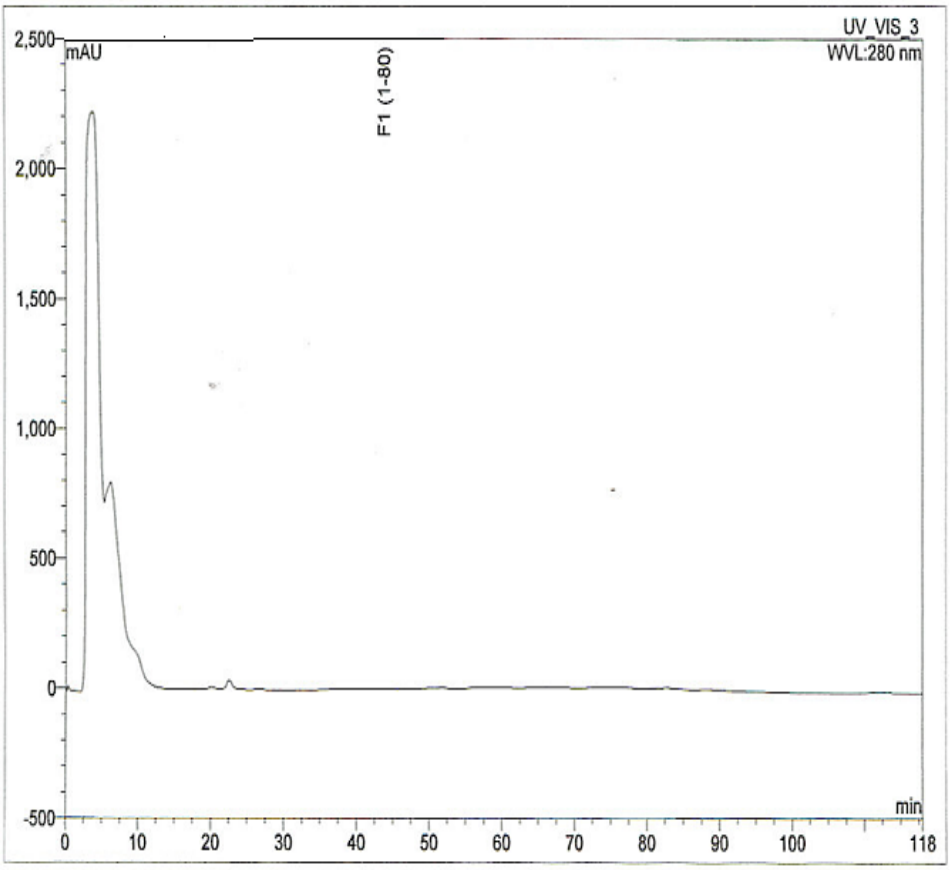


Figure 2.10: Cation exchange chromatogram of the FT protein

In the anion exchange chromatography the resin matrix is positively charged (usually quaternary amines are used) to which negatively charged proteins bind. The anion exchange column used was UNO^{TM} Q-12 from Bio-Rad. The column is packed

with support material containing positively charged quaternary amines ($-N(CH_3)_3^+$) to which negatively charged anions are bound.[13] Figure 2.11 shows the quaternary amine functional group used on anion exchange resin. After equilibrating the column with

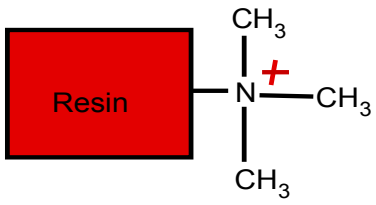


Figure 2.11: Quaternary amine functional group used on anion exchange resin.[14]

starting buffer, the sample was then loaded into the sample loop. The unbound substances are then washed out using the starting buffer for 10 minutes. The sample was then eluted out using 50 mM Tris, pH 7 buffer, containing high concentration (250 mM) of KCl salt. This increases the competition between the chloride ions (Cl^-) and the negatively charged groups present in the proteins to bind to the resin matrix, causing the proteins to elute.[14] The gradient method was used to elute the bound protein, which was discussed previously. The anion exchange chromatogram of the FT protein is shown in Figure 2.12. The sample fractions collected were tested for the FT protein by SDS-PGE. Figure 2.13 shows the SDS-PAGE gel of fractions collected after purification of the FT protein by anion exchange column. The SDS-PAGE results suggested the presence of the FT protein in the fraction number 16. This suggests that the target protein was bound to

the anion exchange resin and it was eluted out during the gradient elution along with the non-target proteins, indicating that the surface charge of the FT protein may be negative.

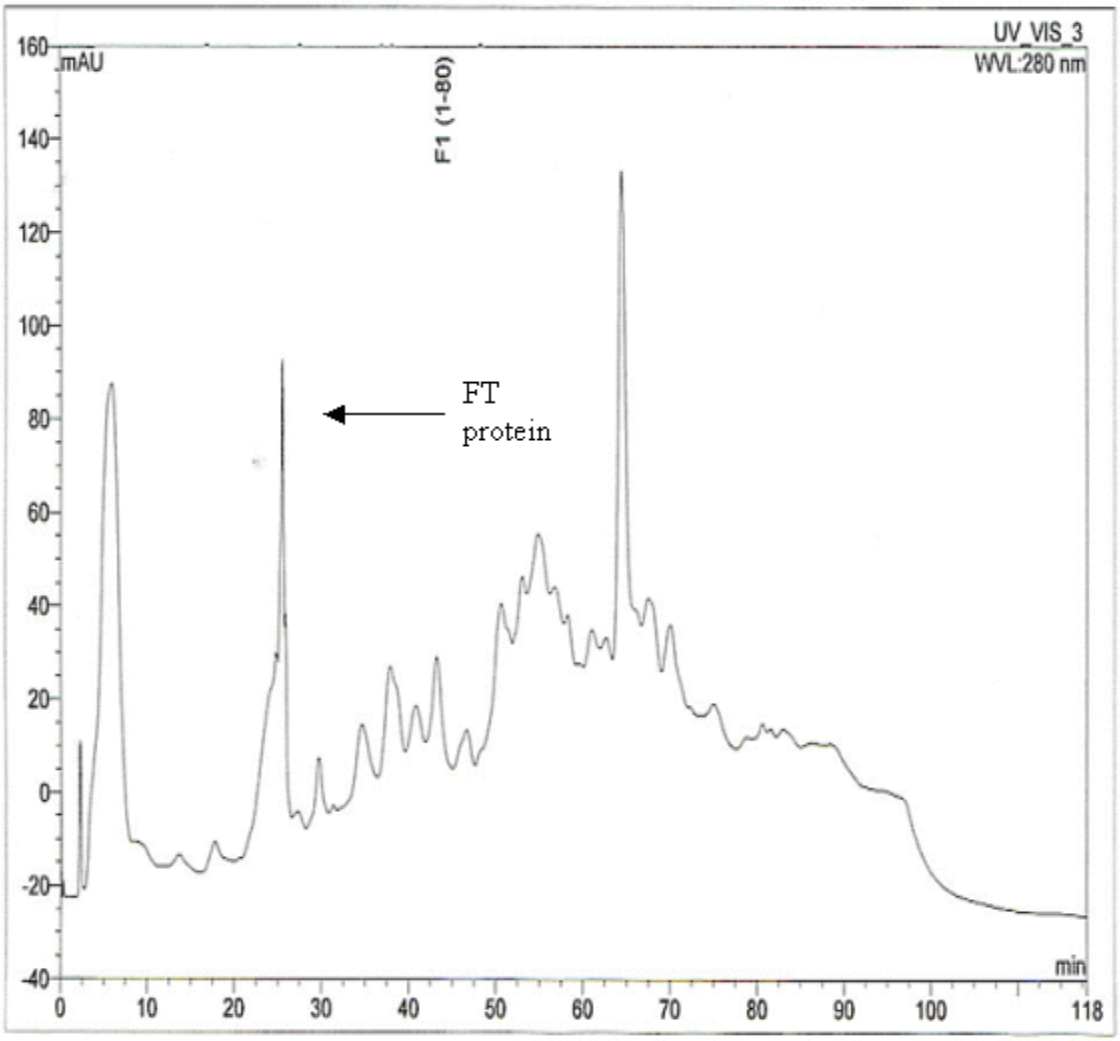


Figure 2.12: Anion exchange chromatogram of the FT protein.

The concentration of the fraction was measured using UV-Visible spectrophotometer. This sample was then used for protein structural studies by fluorescence and CD spectroscopy.

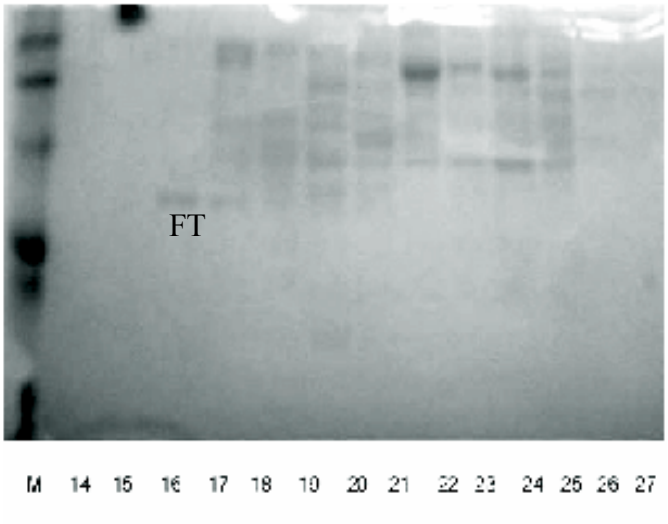


Figure 2.13: SDS-PAGE gel of fractions collected after purification of the FT protein by anion exchange column.

2.5 Structural analysis of FT protein

To obtain the basic information about the secondary structure, the purified FT protein was then studied by using CD and fluorescence spectroscopy. The NMR technique was not used because the concentration of the purified protein was too low (0.007 mM). For protein NMR studies the concentration of the protein should be at least 1mM to obtain good data.[15] Therefore, only CD and fluorescence studies were carried out on the FT protein.

2.5.1 Fluorescence Spectroscopy

The theory behind the structural analysis of proteins by fluorescence spectroscopy is discussed in Chapter 1 (section *1.6.1*). The fluorescence spectrum of the FT protein is shown in Figure 2.14. The FT protein has two tryptophans, and the fluorescence emission spectrum of the FT protein shows maximum emission at wavelength 330 nm. If the protein is folded, a decrease in the wavelength of fluorescence emission (from 350 to 330 nm) for tryptophan is observed.[16] Therefore, the fluorescence emission spectrum of the FT protein suggests that the FT protein may be folded. Moreover, fluorescence gives information about the environment in which the tryptophan is present therefore, to get more information about the secondary structure, the FT protein was further analyzed by CD spectroscopy.

2.5.2 CD

The CD spectrum of FT protein shows a positive peak around 243 nm and two negative peaks around 203 and 206 nm. Figure 2.15 shows the CD spectrum of the FT protein. The pattern of the CD spectrum of FT protein does not match with the CD spectrum of proteins of known secondary structure (see Figure 1.27). Therefore, the CD results obtained for the FT protein were inconclusive. More accurate initial structural information could be obtained using NMR, however; since sufficient amount of the FT protein was not available for NMR studies, therefore, proper conclusion about the secondary structure of the FT protein is difficult.

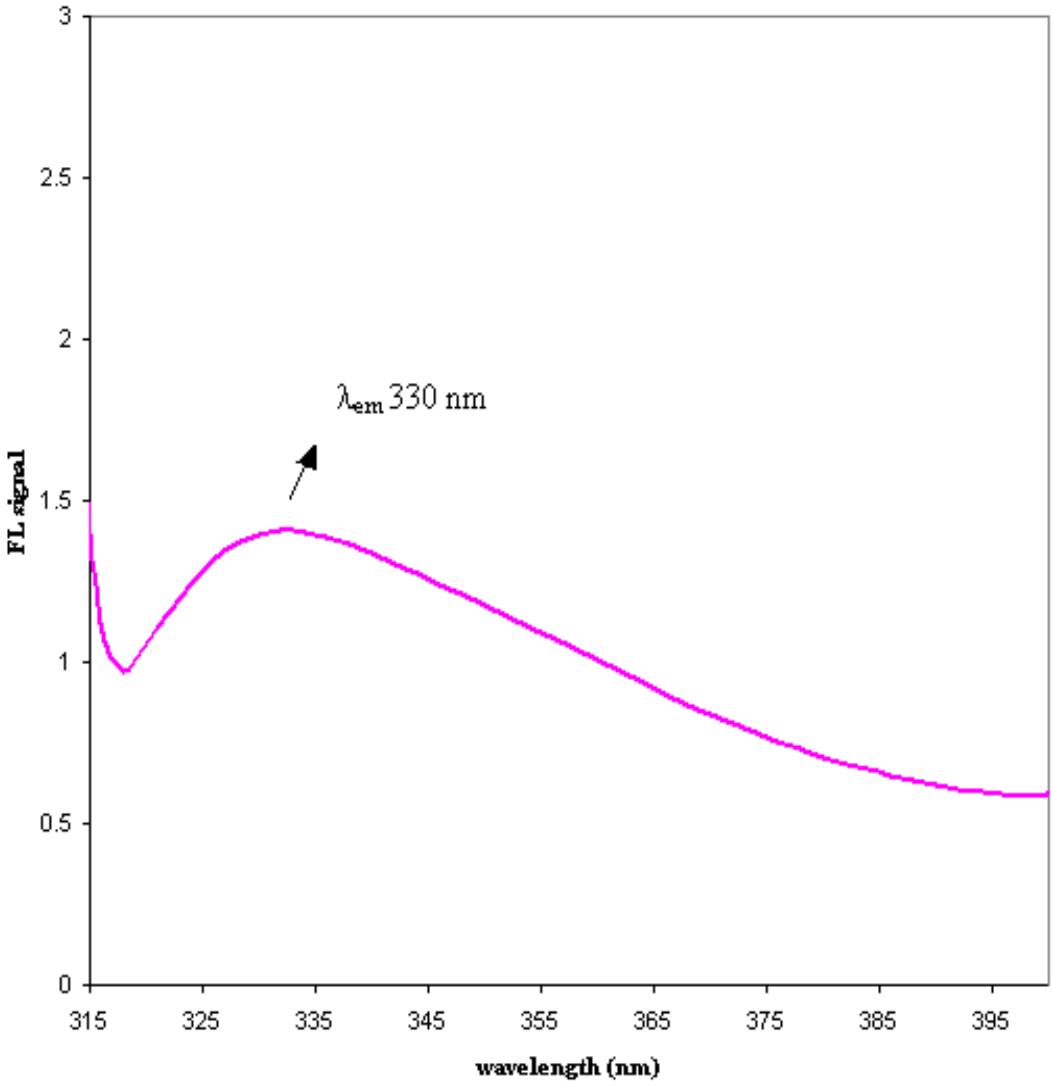


Figure 2.14: Fluorescence spectrum of purified FT protein.

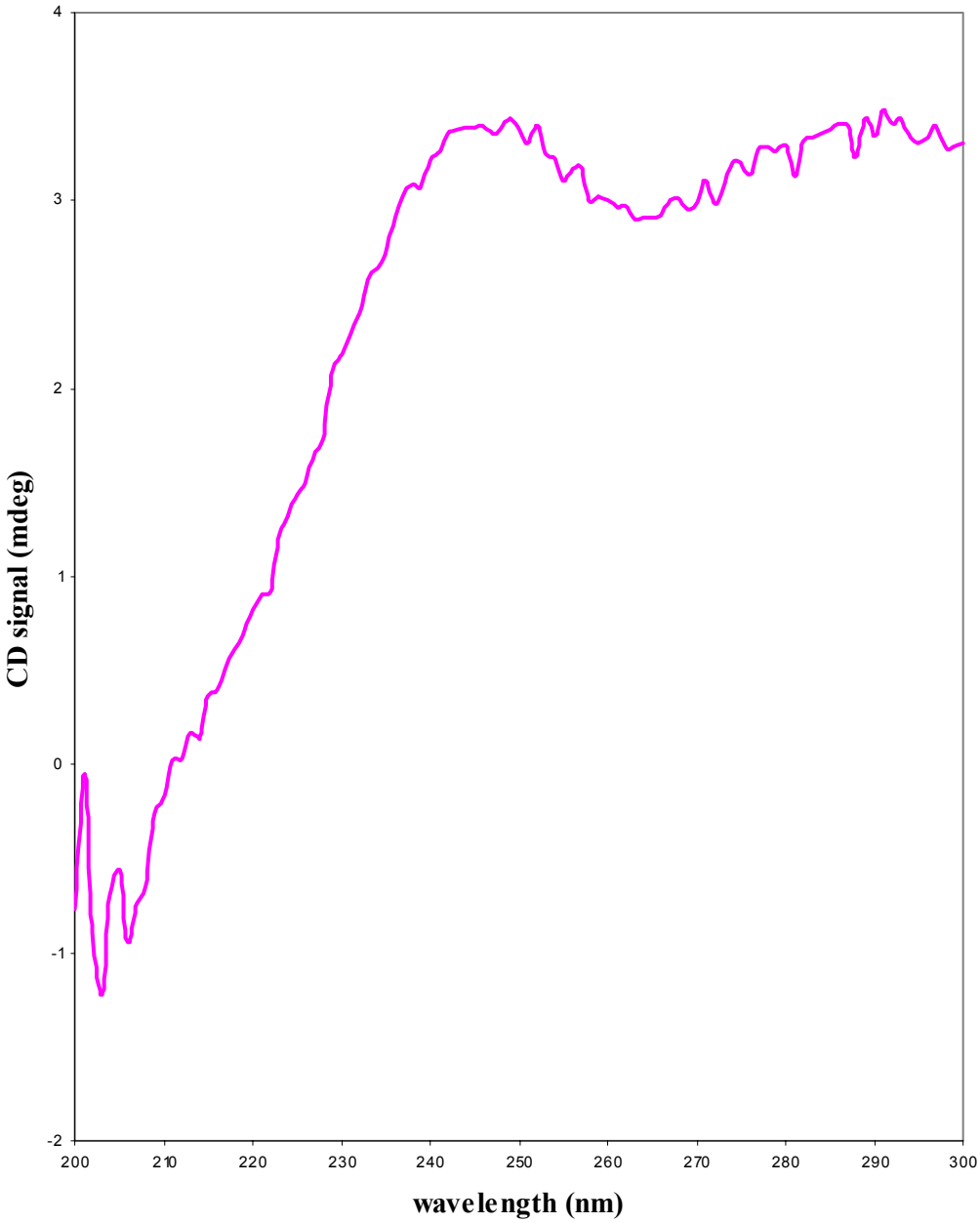


Figure 2.15: CD spectrum of purified FT protein. Where X-axis denotes the wavelength and the Y-axis denotes the ellipticity in millidegrees (mdeg).

2.6 Conclusion

Both the fluorescence and the CD spectroscopy gave contradicting results about the secondary structure of FT protein. However, fluorescence data is preferred over CD because the CD signal is weak when compared to the fluorescence signal of the FT protein. Also, when compared to the CD spectrum of V protein (see Figure 1.27 of Chapter I) it shows that the CD signal of FT protein is weak, which suggests that the concentration of FT protein may not have been sufficient to obtain a good CD signal. Therefore, a conclusion can be drawn that FT may be folded. More accurate secondary structure information about the protein could be obtained using NMR, but since the concentration of the FT protein was too low (0.007 mM), proper NMR study was not possible. Furthermore, only 1.5 mg of purified FT protein of concentration 0.007 mM was obtained when cells expressing the FT protein were grown in 1-liter M9 media. Therefore, to get 1.0 mM concentration sample at least 21 mg of FT protein is need, which indicates that the cells expressing the FT protein has to be grown in 20 liters of M9 media. In addition, for NMR studies the protein has to be labeled and the labeled isotopes are very expensive. Thus, further analysis of FT protein by NMR was not attempted. To get large amounts of the FT protein, the protein has to be more soluble. To fix this solubility problem, Dr. Yuceers lab is trying the restriction digestion on the pET-14b vector with FT DNA, such that the FT protein is His-tagged, which may help in solving this solubility problem.

REFERENCES

1. http://www.ucmp.berkeley.edu/anthophyta/anthophyta.html(accessed June 2007).

2. K. E. Jaeger; A. Graf; P. A. Wigge, *J Exp Bot* **2006,** 57, 3415-3418.

3. P. Teper-Bamnolker; A. Samach, *Plant Cell* **2005,** 17, 2661-2675.

4. http://csm.jmu.edu/biology/courses/bio480_580/mblab/thaliana.html (accessed June 2007).

5. T. Huang; H. Bohlenius; S. Eriksson; F. Parcy; O. Nilsson, *Science* **2005,** 309, 1694-1696.

6. P. Suarez-Lopez; K. Wheatley; F. Robson; H. Onouchi; F. Valverde; G. Coupland, *Nature* **2001,** 410, 1116-1120.

7. L. Corbesier; C. Vincent; S. Jang; F. Fornara; Q. Fan; I. Searle; A. Giakountis; S. Farrona; L. Gissot; C. Turnbull; G. Coupland, *Science* **2007,** 316, 1030-103.

8. M. Abe; Y. Kobayashi; S. Yamamoto; Y. Daimon; A. Yamaguchi; Y. Ikeda; H. Ichinoki; M. Notaguchi; K. Goto; T. Araki, *Science* **2005,** 309, 1052-1056.

9. C. Y. Hsu; Y. Liu; D. S. Luthe; C. Yuceer, *Plant Cell* **2006,** 18, 1846-1861.

10. R. K.Scopes, *Protein Purification, Principles and Practice* second ed.; Springer-verlag: New York, 1987.

11. T. E. Creighton, *Proteins, Structures and Molecular Properties*. second ed.; W. H. Freeman and Company: New York, 1997.

12. P. Biotech, *Ion Exchange Chromatography, Principles and Methods*. AA ed.; Vastra Aros tryckeri: Uppsala, Sweden.

13. http://www.bio-rad.com/LifeScience/pdf/Bulletin_9547.pdf (accessed June 2007).

14. http://www.proteinchemist.com/tutorial/iec.html(accessed June 2007).

15. W. J. F. John Cavanaugh, Arthur G. Palmer, Nicholar J. Skeleton, *Protein NMR Spectroscopy, Principles and Practice*. Academic Press: SanDiego, CA, 1996.

16. A. S. Ladokhin, in: *Encyclopedia of Analytical Chemistry*, R. A. Meyers, John Wiley & Sons Ltd: Chichester, **2000**, 5762-5779.

Chapter III

Materials and Methods

All the molecular biology supplies used for this research study were purchased from Fisher Scientific (Pittsburgh, PA), Novagen Inc. (Madison, WI), Sigma Chemicals (St. Louis, MO), Cambridge Isotope Laboratories, Inc. (Andover, MA), and Invitrogen Corporation (Carlsbad, CA), if not mentioned separately.

3.1 Growth Media, Bacterial Strains and Plasmids

The media used in the study were prepared using autoclaved water according to Sambrook *et al.*[1] *E. Coli* strains were grown overnight in an incubator (MAX^{Q} 4000, Barnstead/Lab-Line, A class) on either Luria-Bertani (LB) broth or M9 media in a 37 ^{o}C shaker at 250 rpm. Antibiotics were added to a concentration of 50 mg/ml and 34 mg/ml in case of ampicillin and chloramphenicol. All the antibiotics were stored at -20 ^{o}C.[1] The pET-24b and pET-14b vectors and the bacterial strains XL1-blue and Rosetta cells used in this study are from Novagen Incorporation. The DNA from the cells was isolated using the QIAprep® Miniprep kit (Qiagen).

3.2 Digestion of pET-14b and pET-24b vectors

The pET-14b and pET-24b vectors were restriction digested with the Nde1 and Bam HI (Invitrogens) restriction endonuclease.[1, 2] The restriction mix was prepared [2] as listed in Table 1. The reactions were mixed well and incubated at 37 ^{o}C for 3 hours. Then

the reaction mixture was loaded on an agarose gel and subjected to electrophoresis (Powerpac basic, Bio-Rad). The agarose gel was prepared by protocol [1] and the digested sample was then loaded on the agarose gel. After gel electrophoresis, DNA bands on the agarose gel were observed using the UV transilluminator (Slimline^{TM} Series) spectrometer and the V DNA band from the gel was cut out using a razor and then the V DNA was isolated using the QIAquick Gel Extraction Kit Protocol (Qiagen).

Table1 Restriction mixture for digestion of pET-14b and pET-24b vectors with enzymes BamH 1 and Nde1

pET-14b Restriction mixture	pET-24b Restriction mixture
pET-14b vector DNA	pET-24b vector DNA
10X Reaction 6 buffer	10X Reaction 6 buffer
Reaction enzyme BamH 1	Reaction enzyme BamH 1
Reaction enzyme Nde1	Reaction enzyme Nde1

3.3 Ligation of V DNA into pET-14b vector and Protein Overexpression

The ligation mixture of V DNA and pET-14b vector was prepared [1, 3] as listed in Table 2. The ligation mixture was mixed well and incubated overnight at 20 ^{o}C. The ligated plasmid was transformed into the *E. coli* cells the next day, following the transformation protocol as mentioned by Sambrook *et al.*[1] The spectrometer used for measuring optical density of cells is Spectronic 20 (Milton Roy Company). The

transformed bacterial colonies containing the V DNA + pET-14b plasmid were selected and tested for protein overexpression according to the protocol of Sambrook.[1] The concentration of IPTG used was 0.4 M.[2] After, confirming the transformed cells were able to express the V protein, the transformed cells were grown overnight in 1 liter LB. The overnight growth is then centrifuged (Sorvell® Legend RT) and the cell pellet was stored in -80 °C freezer (Copeland ® Dupoint Suva® Refrigerators).

Table 2 Ligation mixture of V DNA and pET-14b vector with T4 DNA ligase

Vector pET-14b DNA
V DNA
5X T4 DNA Ligase buffer
T4 DNA Ligase
Autoclaved water

3.4 V Protein Purification

The cell pellet was dissolved in 10 mM Tris, pH 7 buffer (Tris HCl + Tris Base), sonicated (Branson Sonifier 250) and centrifuged. The supernatant solution was then tested for the presence of protein by SDS-PAGE.[1] After confirming the presence of protein in the solution, the solution was then purified using the Ni-NTA column. The column was prepared using a 5 ml plastic syringe and filled with Ni-NTA agarose (Qiagen). The column inlet was then connected to the gradient maker (C. B. S. Scientific Co., GM 500) and the outlet was connected to the fraction collector (Bio-Rad, Model

2110). The wash buffer used contained 10 mM Tris, pH 7 buffer and 10 mM imidazole and the elution buffer contained 10 mM Tris, pH 7 buffer and 250 mM imidazole.

The protein sample was further purified using Hiload^{TM} 16/60 superdex^{TM} 75 gel filtration column (Amersham Biosciences). The mobile phase contained 50 mM Tris, pH 7 buffer and the flow rate was set to 3 ml/min. The column effluents were then collected by a fraction collector (Dionex). After testing the samples for protein using SDS-PAGE analysis, the samples that contained the protein were then mixed and their concentration was measured with a uv-visible spectrophotometer (Varian). After measuring the initial concentration, the sample solution was further concentrated using Amicon Ultra centrifugal device as per the manufacturers protocol (Millipore).

3.5 FT protein purification by Ion exchange chromatography

The cell pellet was dissolved in 10 mM Tris buffer and purified by using the same gel filtration column as mentioned for V protein. The effluents that contained the FT protein after gel filtration was further purified using the cation exchange column UNO^{TM} S-12 (Bio-Rad) and anion exchange column UNO^{TM} Q-12 (Bio-Rad). The wash buffer contained 50 mM Tris at pH 7 and 1 mM DTT (Cambridge Isotope Laboratories, Inc.) and the flow rate was set at 1 ml/min. The eluent buffer for both the cation exchange and for the anion exchange column contained 50 mM Tris at pH 7, 1 mM DTT, and 500 mM KCl (Sigma). [4, 5, 6, 7]

3.6 Protein Structural Studies

Fluorescence wavelength scans were collected on an AVIV model ATF-105/305 differential/ratio spectrofluorometer. The excitation wavelength used was 295 nm. The

CD spectra were recorded on an AVIV 202SF spectrophotometer, using 0.1-cm quartz cell. The sample was 0.4 mM V protein in 50 mM Tris, pH 7 buffer. The sample for FT protein was 0.007 mM protein in 50 mM Tris, pH 7 buffer. The wavelength range scanned was 200-300 nm. After subtracting the background and smoothing, all of the CD data were converted to degree of ellipticity [θ]. Further initial secondary structural information was obtained from 2D ^{15}N-HSQC NMR spectroscopy.[8] The NMR sample was 0.4 mM V protein in 50 mM Tris, pH 7 buffer. The 2D ^{15}N-HSQC spectra were collected on a Bruker 600 MHz spectrometer with an acquisition time of 2.65 seconds and a total of 256 scan acquisitions.

REFERENCES

1. J. Sambrook; E. F. Fritsch; T. Maniatis, *Molecular Cloning: A Laboratory Manual*. Second ed.; Cold Spring Harbor Laboratory Press: Plainview, N.Y., 1989.

2. H. Mosbah; A. Sayari; S. Bezzine; Y. Gargouri, *Protein Expression and Purification,* 47, **2006**, 516-523.

3. X. Liu; J. Liu, *Protein Expression and Purification,* 35, **2004**, 46-53.

4. J. Vasi; R. Morenwieser; K. Eriksson; R. Lemmens, S. Herzer., *Ion Exchange Chromatography Based Strategies for Purification of Viral Vectors*, ESACT Meeting, Granada, **2003**.

5. E. I. Trilisky; A. M. Lenhoff, *J Chromatogr*, A 1142, **2007**, 2-12.

6. R. K. Evans; D. K. Nawrocki; L. A. Isopi; D. M. Williams; D. R. Casimero; S. Chin; M. C. Chen, *J Pharm Sci*, 93, **2004**, 2458.

7. K. Eriksson; Y. Dasarathy; P. Moore, *Purification of Adenoviral Vectors Using Anion Exchange Chromatography*, Amersham Biosciences, Piscataway, NJ, 2000.

8. H. J. Kim; J. K. Young; G. L. Helms; J. N. S. Evans, *J Biomol NMR,* 24, **2002,** 269-270.