Structural Studies on Nucleocapsid Proteins of Shrimp White Spot Syndrome Virus

ZHUANG YING

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2004

Acknowledgment

I would like to express my heartfelt appreciation and gratitude to my supervisor Prof. Hew Choy Leong, for his patience, encouragement and guidance during the course of the project.

I would like to thank Dr. Song Jian Xing for the technical support for my NMR experiments and for stimulating discussion on my project.

Special thanks go out to Dr. Li Zheng Jun, Dr. Lin Qing Song and Ms. Tang Xu Hua for their useful scripts. Many thanks to my friends in Department of Biological Sciences and other departments or institutes, who made me feel so much at home and made my stay in NUS a pleasant learning experience.

Finally, I wish to thank The National University of Singapore for granting me a Research Scholarship.

Content

SUMMARY

Shrimp aquaculture, an important industry worldwide, has been unfortunately devastated by the spread of viral pathogens during the last few decades. Shrimp white spot syndrome virus (WSSV) is the major pathogen out of the more than 20 reported viral agents causing huge economic losses to the industry worldwide.

This project aimed to characterize and elucidate structures of representative proteins covering the nucleocapsid of the shrimp white spot syndrome virus. VP15, which is a major necleocapsid protein, was found to be a non-structural protein after *in-vitro* purification procedure. However, from the DNA-Protein interaction assay, the protein showed the ability to bind dsDNA. VP24, another supposed major nucleocapsid protein, was highly hydrophobic. A protocol was developed to solubilize and purify this protein. Furthermore, Anti-VP24 was also successfully produced and showed a high specificity.

Future study of nucleocapsid proteins of WSSV can be focused on identification more nucleocapsid protein to achieve the capsid priteins *in-vitro* assembly. It will provide a wealth of information for structural based drug design.

CHAPTER 1. INTRODUCTION

1. Literature Review

1.1 Research Background

Penaeid shrimp is one of the main seafood diets for human. Shrimp aquaculture has become an important industry worldwide and large-scale shrimp farming has arisen sharply due to huge consumer demands during the last few decades (Pa'ez-osuna *et al*., 2001). In the year 2000 alone, the world's shrimp production totalled 4,168,400 tons and more than thirty species were cultured. However, intensive cultivation, inadequate sanitation of the shrimp industry and worldwide trade of life stock have unfortunately promoted the spread of vial pathogens. Shrimp diseases, especially those caused by viruses, present the biggest threat to the shrimp farming industry.

Since baculovirus penaei was first discovered in 1974, approximately 20 shrimp viruses have been reported in cultured penaeid shrimp. Among them, white spot syndrome virus (WSSV), considered to be a new virus (Chen *et al*., 1997) and first appeared in the early 1990s in Taiwan, is the causative agent of a major pandemic that has led to severe mortalities of cultured shrimp worldwide (Lightner and Redman *et al*., 1998). The economic loss is phenomenal in the range of multi-billion U.S. dollars and the disease has devastated the shrimp industry in many coastal countries in Asia and more recently in Central and South America. Clinical signs of the disease include a red color to the entire body and appendages along with small subcutaneous white spots. Histological examination reveals prominent intranuclear inclusion bodies in cuticular epithelium, subcutis and connective tissues. Except for the use of nonspecific immunostimulants such

as the bacterial glycans (with minimal protection) and low-density culture conditions, there is at present no effective treatment protocol for this disease.

Most organs and tissues of shrimp, except for hepatopancreatocytes and epithelial cells of the midgut, can be infected by WSSV showing an obvious symptom of white spots on shrimp's cuticle (Wang *et al*., 19999) (Fig.1.1). The virus has a broad host range, including other invertebrate aquatic organisms, such as crab and crayfish (Flegel *et al*., 1997) (Fig.1.2). Therefore the virus is not only a major threat to the shrimp industry but also to the marine environment.

1.2 WSSV (White Spot Syndrome Virus)

Based on the morphology, genomic structure and composition and phylogenetic analyses, WSSV is a member of the genus *Whispovirus* within a new virus family called *Nimaviridae*, referring to the thread-like polar extension on the virus particle (*Nima*: Latin for thread).

1.2.1 Biological Properties

WSSV, circulating ubiquitously in the haemolymph of infected shrimp, can infect most organs and tissues, except for hepatopancreatocytes and epithelial cells of the midgut, which are regarded as refractory tissues (Wang *et al*., 1999). Upon infection, the viruses are observed first in the stomach, gill and cuticular epidermis of the shrimp, and subsequently in other tissues of mesodermal and ectodermal origins (Chang *et al*., 1996). WSSV has an extremely broad host range. Almost all the species of peneaid shrimp are susceptible to its infection. Moreover the virus can infect a wide range of aquatic crustaceans ranging from salt and brackish water penaeids and crabs to fresh water crayfish.

Fig.1.1 WSSV infected shrimp
(The presence of white spots on shrimp cuticle)

Fig.1.2 WSSV infected crab
(The presence of white spots on shrimp cuticle)

WSSV replication initiates in the nucleus, where virions are assembled. The virions are released from the infected cell by cell rupture. Infected animals show gross signs of lethargy, such as lack of appetite and slow movement. Characteristic for infected shrimps are the white spots on the exoskeleton.

1.2.2 Microheterogeneity in WSSV Isolates

Only a single species within the genus *Whispovirus* has been identified today (WSSV). Various geographical isolates with genotypic variability (variants) have been identified within this species. Other names that are used in literature for WSSV include HHNBV (hypodermal and hematopoietic necrosis baculo-like virus of *P. chinensis)*, RV-PJ (rod-shaped virus of *Penaeus japonicus)*, SEMBV (systemic ectodermal and mesodermal baculovirus) and WSBV (white spot Bacilliform virus).

1.2.3 Morphology

Virions of the type species of *Whispovirus*, WSSV, are large, ovoidal particles of about 300-350 nm in length and 70-120 nm in width, with a tail-like appendage at one end (Fig1.3 and Fig1.4). The virion consists of a rod-shaped nucleocapsid with a tight-fitting capsid layer surrounded by a double envelope (Fig.1.5). Isolated nucleocapsids have a crosshatched appearance and a size of about 300 x 70 nm. A multifilament appendage is often seen attached at the narrow end in the purified virion (Huang C *et al*., 2001), which is excluded in the intact virion from the direct negative staining of WSSV-infected shrimp hemolymph (Wongteerasupaya *et al*., 1995). The nucleocapsid, which contains a DNA-protein core bounded by a distinctive capsid layer, is wrapped often singly into an envelope to shape the virion (Durand *et al*., 1997).

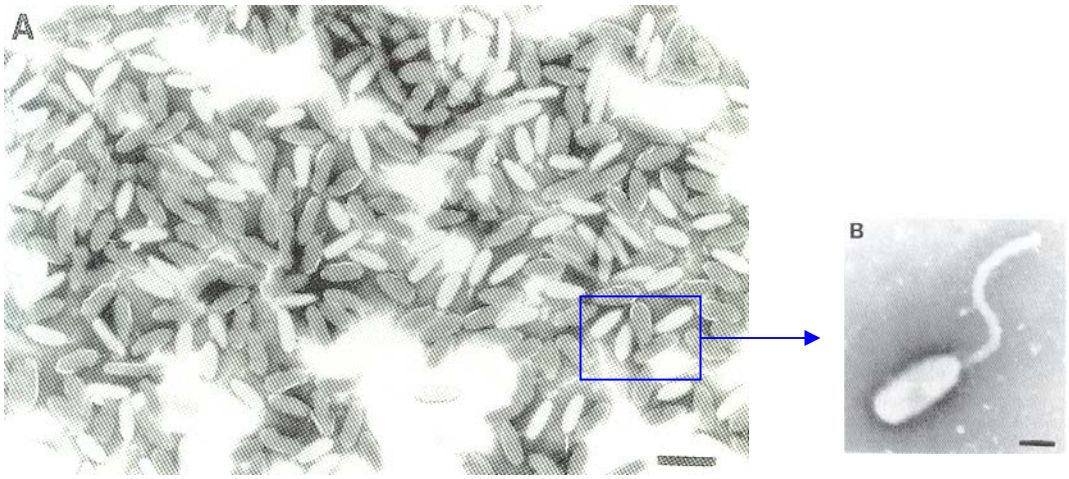


Fig.1.3 Electron micrographs of purified intact WSSV virions
(Fig. A Scale Bar=416nm; and Fig B, Scale Bar=104nm)

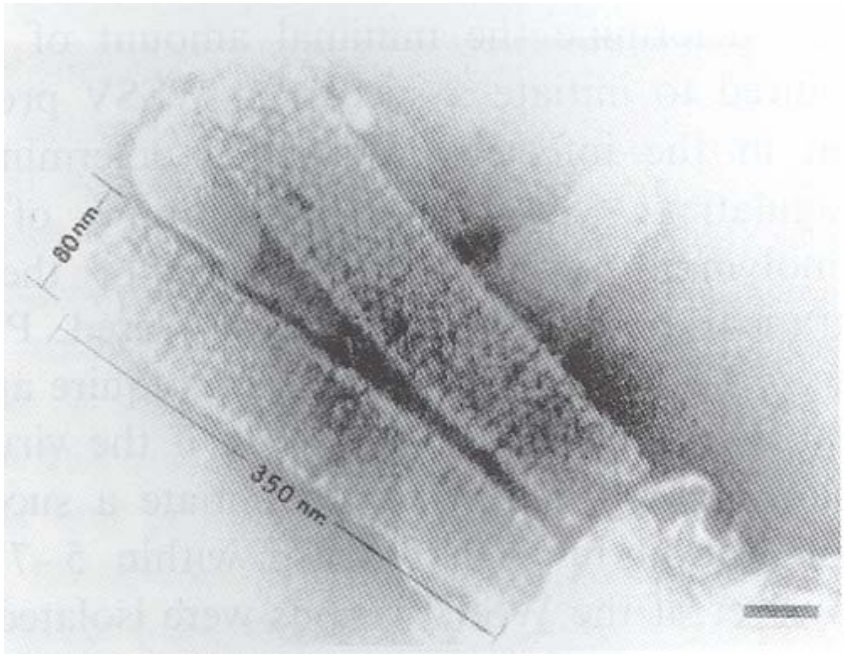
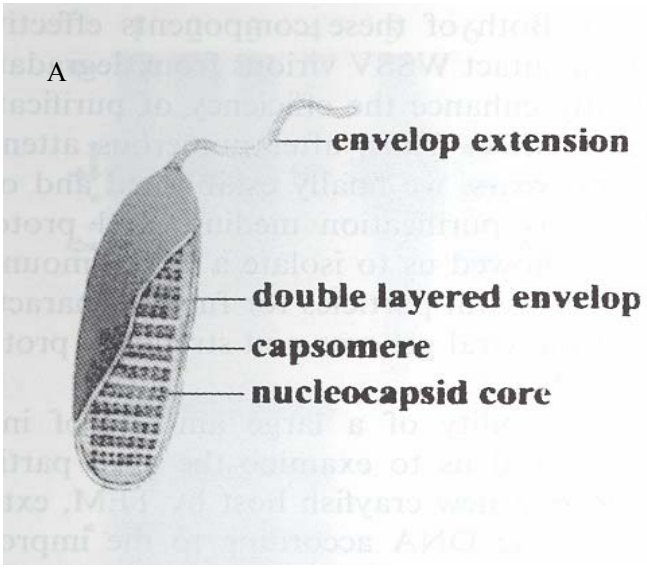
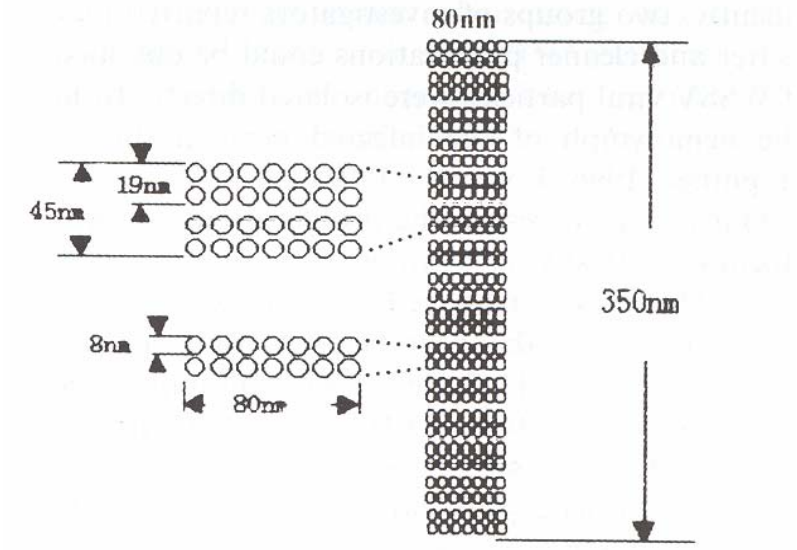


Fig.1.4 Nucleocapsids of WSSV
(Scale Bar=50nm)

A





B

Fig.1.5 Proposed WSSV particle structures (Fig A and B)

DNA replication and de novo envelop formation of WSSV take place in the nucleus. (Durand *et al*., 1997; Wang CH *et al*., 2000).

1.2.4 Virion Proteins

In general, the viral proteins are divided into three temporal classes: early proteins synthesized prior to the replication of the DNA, and the intermediate and late proteins synthesized after the onset of DNA replication (Jensen *et al*., 1996). The late proteins synthesized from 5 to 6 h post-infection are assumed to be viral structural proteins.

Based on the separation of the proteins from the purified WSSV virions by SDS-PAGE, more than 20 bands ranging from 5~ 190 kDa were visible with Coomassie blue staining (Fig 1.6) (Huang *et al.,* 2002a). In an attempt to identify the viral proteins from the 181 theoretical ORFs, proteins of the purified WSSV were separated by two-dimensional gel electrophoresis (2-DE). More than sixty protein spots were revealed as detected by silver staining.

Structural proteins including the envelope proteins and nucleocapsid proteins, are important due to their involvement in the interaction between virus and host, such as attachment to receptors and fusion with cell membranes. Up to date, at least 5 genes encoding major structural proteins of WSSV have been reported and been named according to their sizes in SDS-PAGE (Zhang *et al*., 2002a, 2002b; Huang *et al*., 2002a, 2002b). The envelope contains three major proteins, VP28, VP19 and VP26, whereas the nucleocapsid contains two, VP24 and VP15 (Fig.1.7). VP28 is most likely located on the surface of the virus particle and plays a key role in WSSV infection. A putative function of VP15, a highly basic protein with no hydrophobic regions, is that of a histone-like, DNA-binding protein.

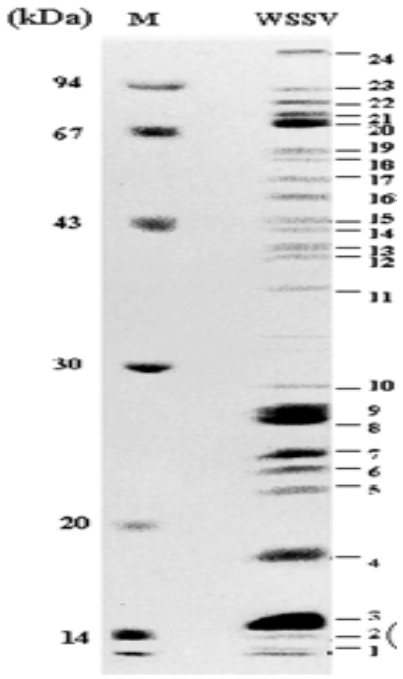


Fig.1.6 SDS-PAGE of proteins from purified WSSV virions (Coomassie staining)

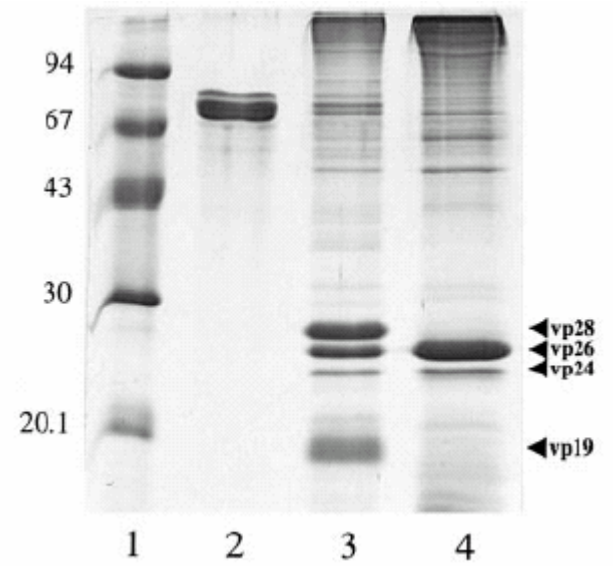


Fig.1.7 Coomassie brilliant blue-stained SDS–15% PAGE gel of purified WSSV.

Lane 1, low molecular mass protein marker. Lane 2, mock purification (uninfected). Lane 3, purified WSSV particles (including envelope proteins and nucleocapsid proteins). Lane 4, purified WSSV nucleocapsids.

1.2.5 Genome

WSSV is at present the largest animal genome that has been entirely sequenced. Sequencing of three different isolates revealed that the circular dsDNA genome contains around 305 kb (Fig.1.8). The genome has a G+C content of 41% uniformly distributed over the genome. Three percent of the WSSV genome consists of nine homologous regions (*hrs*) dispersed throughout the genome with different sizes, while the remaining 97% of the sequences are unique (Yang *et al*., 2001; van Hulten *et al*., 2001). The WSSV genome has the capacity to encode 181 presumptive open reading frames (ORFs) of 50 amino acids or more, ranging from 61 to 6077 amino acids (Yang *et al.,* 2001). Since the origin of replication is unknown, a guanine residue from the beginning of the largest BamHI fragment is designated as the starting point of the physical map of the WSSV genome (Yang *et al.,* 2001). The ORFs are present on both strands in almost equal proportions (54% forward, 46% reverse). A TATA box sequence is found in the promoter regions of 46% of the WSSV ORFs. Consensus poly (A) signal sequence can be found for 80% of the ORFs.

Several unique features of the WSSV genome are the presence of a very long ORF of 18,234 nt, with unknown function, a collagen-like ORF, and nine non-coding regions, dispersed along the genome, each containing a variable number of 250 bp tandem repeats.

1.2.6 Phylogeny

Based on the phylogenetic trees of the WSSV genes DNA polymerase (see figure 1.9), ribonucleotide reductase large (*rr1*) and small (*rr2*) subunits, protein kinase, thymidine-thymidylate kinase and endonuclease, no recent relationship was revealed for WSSV with members of other established large dsDNA virus families.

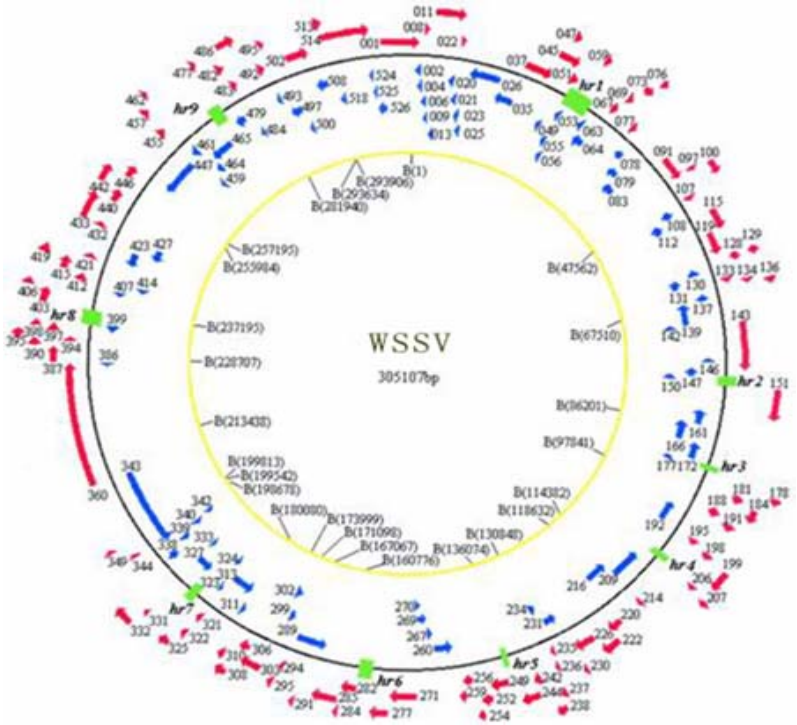


Fig.1.8 Genetic map of WSSV

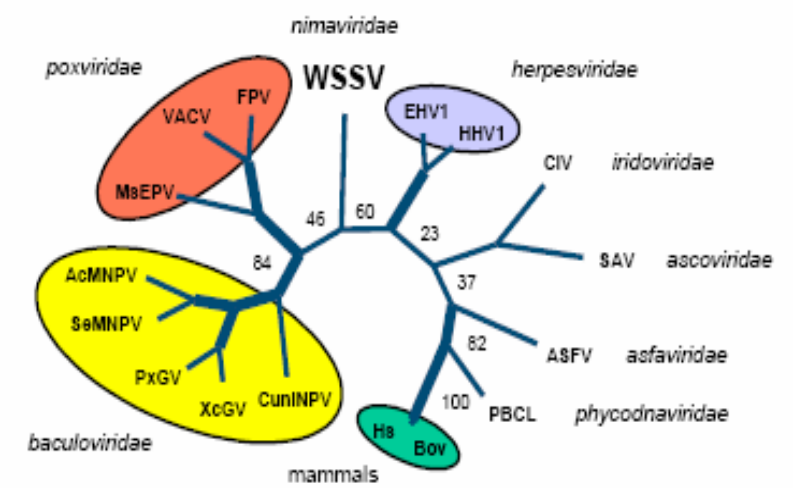


Fig.1.9 Bootstrap analysis (100 replicates) of an uprooted phylogenetic tree of the WSSV protein DNA polymerase.
The numbers at the branches indicate frequency of the clusters and frequencies over 70% are indicated by thick lines.

2. Structural Genomics

The explosive growth of genetic sequence information has offered us comprehensive collections of the protein sequences found in many living organisms. Most of these proteins are not experimentally characterized. Although half of the proteins that are encoded in sequenced eukaryotic genomes have computationally recognized homology to at least one well-characterized domain (Lander, E. S. *et al.*, 2001), functional interpretation of these matches is fraught with difficulty. Functional changes over evolutionary time (Devos, D. and Valencia *et al.*, 2000) and database errors (Brenner *et al.,* 1999) confound reliable computational prediction of the precise roles of newly discovered genes. Even proteins with recognized domains are often scattered with regions of unmatched sequence. Therefore, most of the residues in putative gene products lack any computational annotation, and there exists no general experimental approach to directly ascertain their molecular role. The challenge of understanding these gene products has led to the development of functional genomics methods, which collectively aim to improve the genetic sequence with biological understanding. Structural genomics is one such approach, with unique promise to reveal the molecular function (Ashbumer, *et al.,* 2000) of protein domains.

Structural genomics projects aim to provide an experimental or computational three dimensional model structure for all of the tractable macromolecules that are encoded by complete genomes. Their experimental structures and computational models are expected to yield new insight into the molecular function and mechanism of thousands of proteins. (Steven E. Brenner *et al.,* 2001)

2.1 Technology

Structural biologists have embraced high-throughput biology by developing and implementing technologies that will enable the structures of hundreds of protein domains to be resolved in a relatively short period of time. There are now many new developments in X-ray crystallography, NMR spectroscopy, cryoelectron microscopy, robotic crystallization, high performance computing, PCR recombinant technology and high-level protein expression system. These advancements have made it feasible to mount a system/organism-specific structural biology program to determine the three dimensional structures of the majority if not all proteins in a particular organism (Burley *et al.,* 1999; Chance *et al.,* 2002). Structural Genomics, this new concept is now being developed in many laboratories. This large scale, high throughput concerted effort will attempt to determine the structures of representative proteins (in terms of protein domain and folding) and should provide a better understanding of the physiology of the organism under investigation. Besides elucidating the functions of the presumptive proteins based on the determined structures (Sharpiro *et al.,* 2000; Norin and Sundstrom *et al.,* 2002), the study will also provide the structural information necessary for drug targeting and design.

2.2 Process

The principles of experimental structural genomics are largely the same as those for traditional structural biology, but differ in motivation, automation and scale (Fig.1.10). The key to the success of this scientific venture is the ability to optimize the structure-determination process, so as to reap economies of scale as centres increase their throughput. The flow of information and material between projects of one structural genomics program is shown in the figure (Fig.1.11).

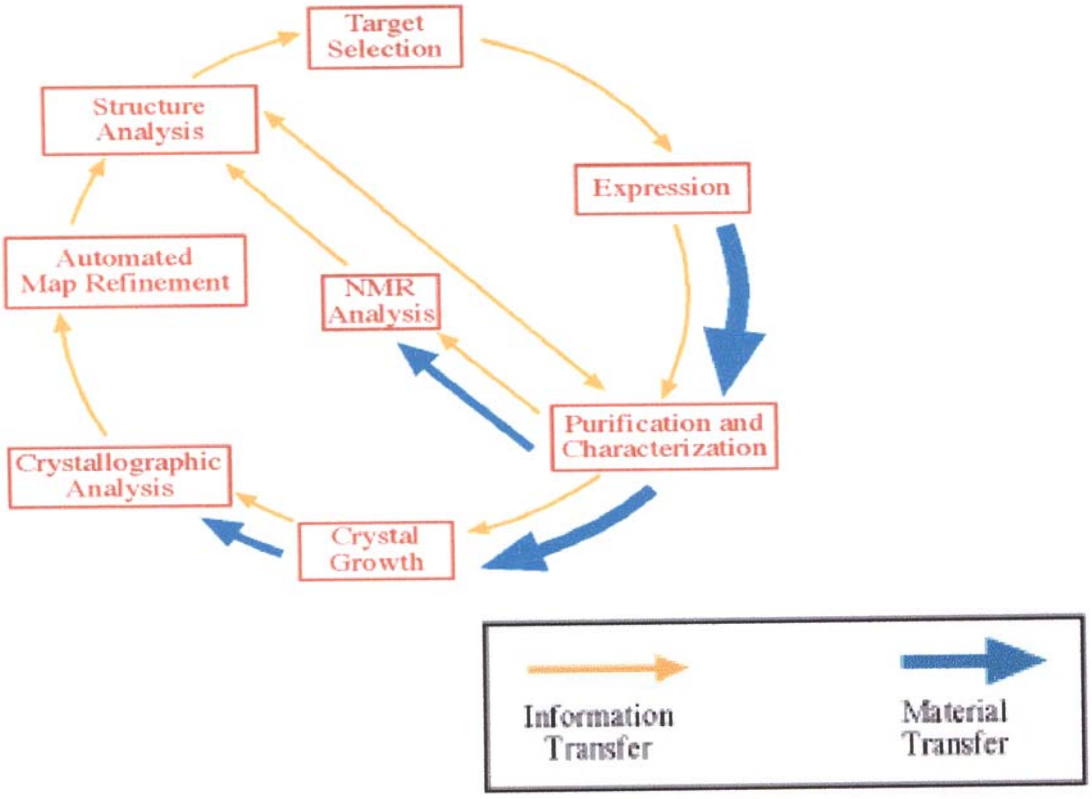


Fig.1.10 Process of the structure genomic

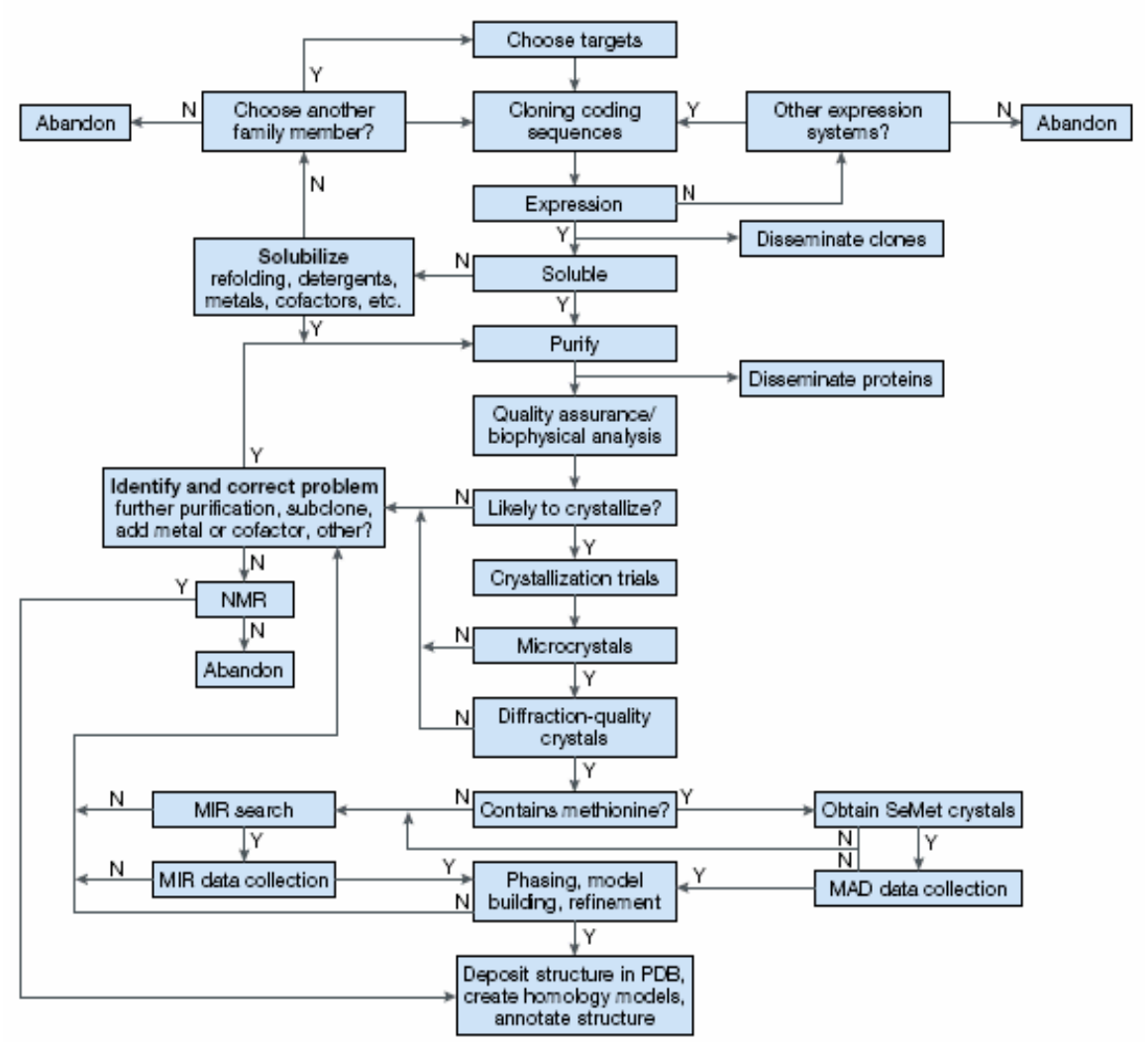


Fig.1.11 Processes involved in high-throughput structural genomics using X-ray crystallography. N indicates that a process has failed and Y that it has succeeded. (MIR, multiple isomorphous replacement; an alternative to multiple anomalous dispersion (MAD) phasing for structure determination; NMR, nuclear magnetic resonance; SeMet, selenomethionine.)

Structural genomics provides a solid foundation for the future of structural biology research. It will revolutionize biochemistry and molecular biology, making use of the three-dimensional structure information. Just as one can expect to find sequences for most genes of interest in public databases, structural genomics promises to offer a comparably comprehensive library of experimental and computational models. These will reveal new functions, indicate molecular mechanisms and explicate mutations.

2.3 WSSV as A Model for Structural Genomics Study

WSSV provides an important and interesting model for structural genomics study. It has a relatively small genomes with only approximately 181 opening reading frames, much smaller than many of the organisms such as Mycoplasma genitalium, Mycobacterium tuberculosis, Saccaromyces cerevisiae, E.*coli* (*Escherichia coli*) and C. *elegans* selected for similar structural genomic studies (Chance *et al.,* 2002). This would make the study more manageable for a small research team in Singapore.

Many of WSSV presumptive ORFs do not have any match with known proteins and functions. Structural genomics therefore offers a unique opportunity to elucidate the function of these proteins through structural determination and prediction. Structural elucidation of WSSV will provide many insightful mechanisms of host-pathogen interaction, the maturation and assembly of the virus particles and many other important questions. Most importantly, it will provide a wealth of information for structural based drug design.

3. The Objective of This Research

The characterization of the structural proteins and their genomic sequence is of major importance to determine the taxonomic position of the virus. The project aims to elucidate the three dimensional structure of the virus capsid protein and its assembly. The structures of representative proteins (in terms of protein domain and folding) should provide a better understanding of the physiology of the organism under investigation. Furthermore, the structure and interaction of the capsid proteins may explain the unique morphological features of this virus. Besides elucidating the functions of the presumptive proteins based on the determined structures, this study will also provide the structural information necessary for drug targeting and design.

CHAPTER 2. NUCLEOCAPSID PROTEIN VP15 OF WSSV

1. Summary

DNA fragment encoding the proteins VP15 was PCRed out from WSSV genomic DNA and cloned into pGEX-6P-1 vector (N-terminal GST-tagged) with restriction sites BamHI / Xho I. The sequence confirmed vector was transformed into BL21 cell for expression. Over-expressed and soluble proteins was further ^{15}N isotope-labeled by growing BL21 cells in the M9 medium with ^{15}N -ammonium sulfate as the sole nitrogen source. The GST fusion protein was purified by affinity chromatography on the glutathione Sepharose 4B beads. The N-terminal GST tag was cleaved by PreScission Protease on-column at 4℃ degree with overnight incubation. Pre-equilibrium SP Sepharose column was used for further purification of this extremely basic protein. Collected protein was confirmed as target VP15 by N-terminal sequence. The purified VP15 was further used to perform NMR experiments. The two-dimensional NMR ^{1}H-^{15}N Heteronuclear Single-Quantum Correlation (HSQC) Spectra was collected for ^{15}N isotope-labeled samples. Gel mobility shift assay were further carried out to determine whether there were any specific DNA binding site for VP15. CD spectroscope was used to determine whether any structure change happened during DNA-protein binding process.

2. VP 15

2.1 Introduction

Previous studies showed the virion nucleocapsid consists of three major proteins of VP26, VP24 and VP15 (Van Hulten *et al*., 2001a). However, based on the experimental result of our lab, VP26 is not a capsid protein, but rather an envelope protein. So nucleocapsid of WSSV virions presumably consists of two major proteins VP15 and VP19.

2.2 Sequence and Location of the 15 kDa Protein Gene on the WSSV Genome

The complete sequence of WSSV has recently been determined (Van Hulten *et al*., 2001a; Yang *et al.,* 2001) and the N-terminal amino acid sequence determined for VP15 was found to be encoded by ORF109. This ORF is located as a single-copy gene from nucleotide position 163996 to 164238 on the WSSV genome. It partially overlaps with a small putative ORF of 11 kDa, ORF110, which is in the reverse orientation on the WSSV genome (Fig. 2.1). Both ORFs are flanked by two large ORFs (ORF108 and ORF111), encoding putative proteins of 174 and 132 kDa, respectively. The experimentally determined N terminus of VP15 is not located at the putative start of ORF109, but at the valine at position 21, which is the amino acid following the methionine encoded by the second ATG (Fig. 2.2). The first ATG of this ORF is apparently not used (Van Hulten *et al*.2002), This ATG is in an unfavourable Kozak (1989) context (TTC**ATG**A) for efficient translation initiation, whereas the second ATG is in a favourable Kozak context (AAA**ATG**G).

Furthermore, the TATA-box, 87 nucleotides upstream of the second ATG, has a

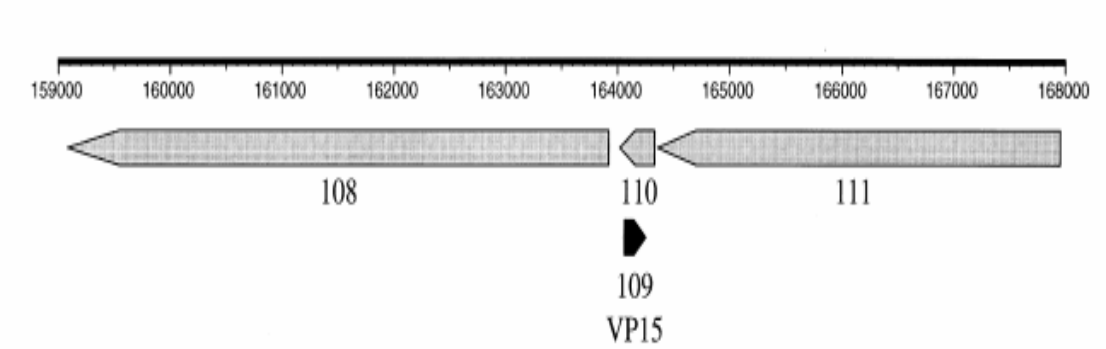
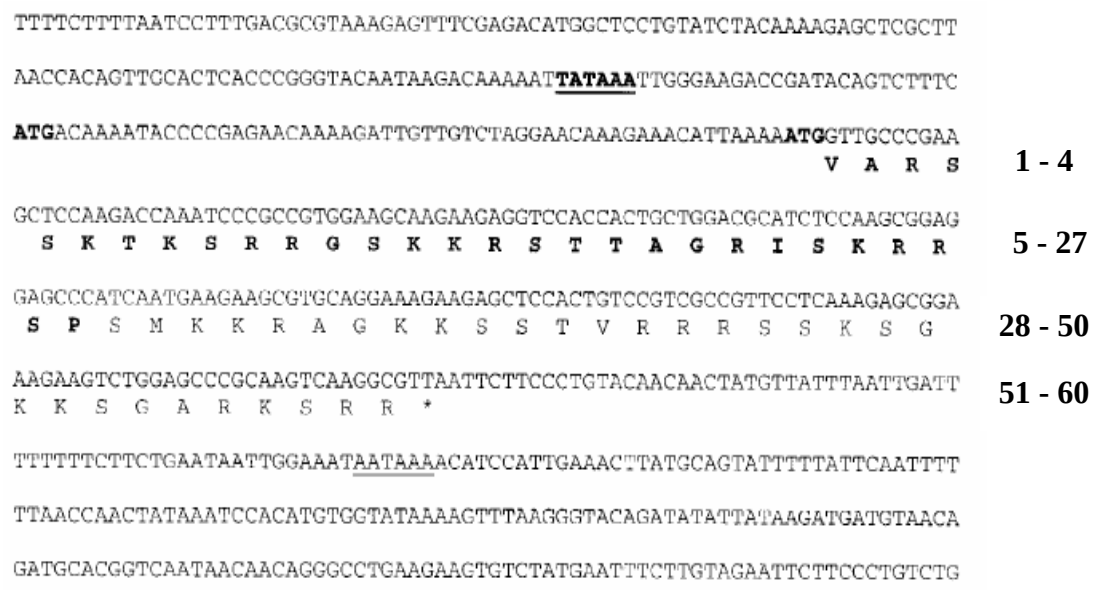


Fig.2.1 Location of WSSV VP15 on the WSSV genome

5'



3'

Fig.2.2 Nucleotide and protein sequence of VP15. (The N-terminal sequenced amino acids are in bold type, the location of the putative TATA-box is underlined and in bold and the putative poly(A) signal is underlined)

position relative to a putative transcriptional start, which also favours this second ATG. As no signal peptide was predicted for the first nor for the second ATG, the most likely explanation for the presence of the valine at the N terminus is the removal of the N-terminal methionine by N-terminal processing enzyme(s) as is observed for more than 60% of eukaryotic and prokaryotic proteins (Giglione *et al*., 2000). When the second residue is a valine, removal of the N-terminal methionine is favoured (Flinta *et al*., 1986). A poly(A) signal is present 62 nucleotides downstream of the translation stop codon (Fig.2.2). ORF109, starting at the preferred (second) ATG, encodes 61 amino acids.

2.3 Characteristic of Protein VP15

The theoretical molecular mass of the 60 amino acid protein after removal of the N-terminal methionine is only 7102.26Da, but it migrates as a 15 kDa protein in SDS PAGE. Its aberrant migration may be caused by the fact that this viral protein has an extreme basic pI value of 13.2. The protein has an amino acid composition rich in lysine (22 %), arginine (23%) and serine (25 %). No potential sites for *N*-linked glycosylation (N-²P′-[ST]-²P′) or for *O*-glycosylation (Hansen *et al*., 1998) could be identifed. Eight possible phosphorylation sites ([ST]-X-X-[DE] or [ST]-X-[RK]) were found within VP15. No cysteine residues are present, and therefore VP15 is unable to form any disulfide bridges. No hydrophobic regions are present in this protein (Fig. 2.3) and it has a strong positive charge along the whole length of the polypeptide.

A putative function for this highly basic protein is that of a histone-like, DNA-associated protein. Its small size, positive charge at physiological pH (26.9 at pH 7.0) and its presence within the nucleocapsid fraction are in agreement with this hypothesis. Homology searches using blast programs (Altschul *et al*., 1997) showed that

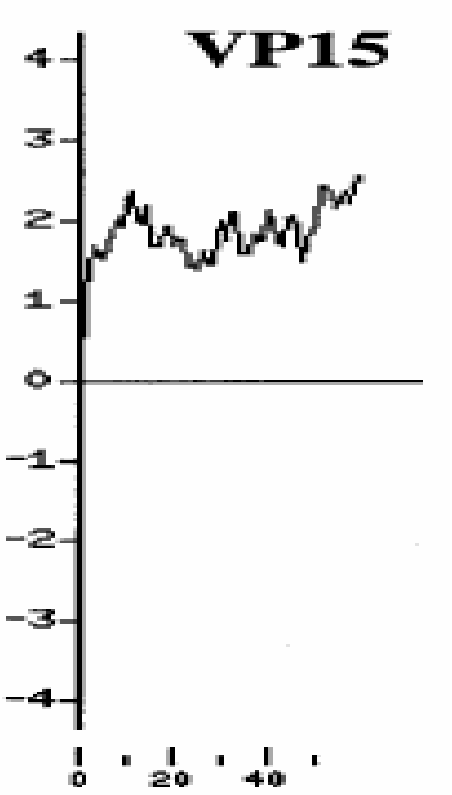


Fig.2.3 Hydrophilicity plots of VP15

(The amino acid number is on the abscissa, the hydrophilicity value on the ordinate.)

VP15 has homology with several DNA-binding proteins of eukaryotic origin. Comparison of VP15 with DNA-binding proteins of baculoviruses, e.g. with AcMNPV DNA-binding protein p6.9 (Wilson and Price *et al.,* 1988) and *Lymantria dispar* MNPV (LdMNPV) DNA-binding protein p6.9 (Kuzio *et al*., 1999), revealed an overall pair wise amino acid sequence similarity & identity of 47% & 26% with AcMNPV and 45%& 27% with LdMNPV. This homology with other DNA-binding proteins is the result of the high content of basic amino acids (e.g. arginine) and serine, characteristic of these proteins. This further substantiates the putative function of VP15 as a DNA-binding protein.

2.4 Studies on VP15

This study aims to select proper expression system to express VP15 and to use chromatography for its purification. Furthermore, gel mobility shift was also studied for DNA-protein interaction, to identify whether there is any specific DNA binding site. Some structural biology technologies, such as, NMR (nuclear magnetic resonance), CD (circular dichroism) were used for protein, protein-DNA structural elucidation.

3. Introduction to Biomolecular NMR

When a sample is placed in a magnetic field and is subjected to radiofrequency at an appropriate frequency, nuclei in the sample will absorb the energy. Absorption of energy by the nuclear spins causes transitions between different energy levels. The energy absorbed by the nuclear spins induces a voltage that can be detected by a suitably tuned coil of wire, amplified and then the signal displayed as free induction decay.

The NMR phenomenon was first detected by Bloch and Purcell independently (Bloch *et al.,* 1946; Purcell *et al.,* 1946). In order to induce magnetic resonance, an oscillatory magnetic field has to be applied at the frequency which corresponds to the difference between two energy levels of the nuclear spins. This was originally achieved by scanning various different frequencies (continuous wave or CWNMR). A major breakthrough in NMR technology is the concept of Fourier Transform NMR spectroscopy (Ernest and Anderson *et al.,* 1966). In FTNMR, all spins of a particular isotope can be excited in one experiment and the resulting resonance signals can be converted to frequency spectrum via Fourier Transform. FT NMR is therefore more time efficient and sensitive.

NMR spectroscopy was firstly used for the structural determination of small molecules in organic chemistry. The study of protein by NMR was hampered until the late 1970s. This is mainly because a large number of protons that a protein contains will make the 1D proton spectrum extremely overlapped and largely uninterpretable. The application of 2D spectra makes it possible to study small proteins by NMR. In two-dimensional NMR spectroscopy the second dimension is another time domain which can be established by introducing another pulse and systemically increase the time

between the two pulses.

The strategy for protein structural determination by 2D experiments was first developed by Wuthrich and his colleagues in the early 1980s (Wuthrich *et al.,* 1986). The two dimensional NMR spectra are useful for determining structure of proteins with less than 80 amino acid residues. However, with the increasing molecular weight of protein, the 2D spectra become severely overlapped and crowded. This render it is impossible to identify adjacent spin systems by only utilizing a single common resonance frequency of the connecting cross peaks.

Since 1990s, significant progress has been made for determining protein structures using multidimensional NMR experiments. Compared with the NMR structures of 10 years ago, the structures being determined today have much higher molecular weight, precision and are completed much faster. A lot of technological advances are required for the improvements. Some of the most important aspects are certainly the availability of high field spectrometers (500-900MHz) and multidimensional techniques. Equally important are the development of multi-channel spectrometer hardware, indirect detection and the application of cryoprobes. Other advances include improvements in computers hardware and software, plus advances in techniques for calculating protein structures using NMR data.

4. Introduction to Circular Dichroism

Inherently asymmetric chromophores (uncommon) or symmetric chromophores in asymmetric environments will interact differently with right- and left-circularly polarized light resulting in two related phenomena. Circularly-polarized light rays will travel through an optically active medium with different velocities due to the different indices of refraction for right- and left-circularly polarized light called optical rotation or circular birefringence. The variation of optical rotation as a function of wavelength is called optical rotary dispersion (ORD). Right- and left-circularly polarized light will also be absorbed to different extents at some wavelengths due to differences in extinction coefficients for the two polarized rays called circular dichroism (CD). Optical rotary dispersion enables a chiral molecule to rotate the plane of polarized light. ORD spectra are dispersive (called a cotton effect for a single band) whereas circular dichroism spectra are absorptive. The two phenomena are related by the so-called König-Kramers transforms.

The phenomenon of circular dichroism is very sensitive to the secondary structure of polypeptides and proteins. Circular dichroism (CD) spectroscopy is a form of light absorption spectroscopy that measures the difference in absorbance of right- and left-circularly polarized light (rather than the commonly used absorbance of isotropic light) by a substance. It has been shown that CD spectra between 260 and approximately 180 nm can be analyzed for the different secondary structural types: alpha helix, parallel and antiparallel beta sheet, turn, and other (Fig.2.4). Modern secondary structure determinations by CD are reported to achieve accuracies of 0.97 for helices, 0.75 for beta

sheet, 0.50 for turns, and 0.89 for other structure types (Manavalan and Johnson *et al.,* 1987).

Besides defining the secondary structure characteristics of proteins and peptides, CD can also provide rapid structural comparisons (e.g. between mutants), is ideally suited to exploring the extent and rate of structural changes, and can probe interactions (protein-ligand; protein-protein; protein-nucleic acid etc.). The full potential of CD is realised when it is used in conjunction with other biophysical techniques.

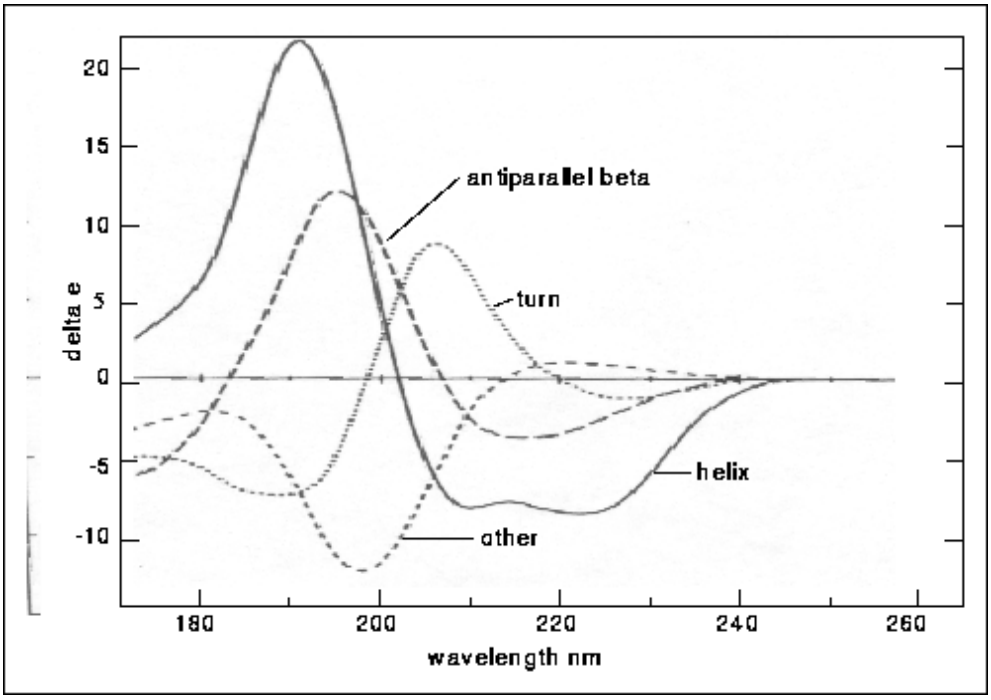


Fig.2.4 Standard curve of CD spectrum

5. Material and Methods

Construction Plasmid of 6P-1-VP15

183 bp DNA fragment encoding the proteins was amplified out from WSSV genomic DNA using PCR and cloned into pGEX-6P-1 vector (N-terminal GST-tagged) with restriction sites *Xho*I and *Bam*HI.

5.1 PCR Amplification of Gene VP15

VP15 DNA Sequence:

5' atggttgccc gaagctccaa gaccaaatcc cgccgtggaa gcaagaagag gtccaccact gctggacgca tctccaagcg gaggagccca tcaatgaaga agcgtgcagg aagaagagc tccactgtcc gtcgccgttc ctcaaagagc ggaaagaagt ctggagcccg caagtcaagg cgttaa ***3'***

The 183 bp gene was PCR amplified by two primers (5'-:***CGC GGA TCC*** **GTT GCC CGA AGC TCC AAG** -3' and 5'- ***CGC CCG CTC GAT*** **TTA ACG CCT TGA CTT GCG GGC** -3') from the pGEX-6P-1vector. The primers used in the PCR were specially designed to incorporate *Xho*I and *Bam*HI restriction sites to the flanks of the VP15 fragment for subsequent cloning into the expression vectors. (Table2.1& Table 2.2)

5.2 Construction of 186bp PGEX 6P-1-VP15 Gene

The PCR fragment (5 µL) was run on a 1.5 % agarose gel. 45µL PCR product was purified and concentrated by ethanol precipitation. Both the amplified VP15 fragment and the pGEX-6P-1 plasmid vector were doubly digested (Table 2.3) with *Xho*I and *Bam*HI restriction enzymes (New England Biolabs). The cut DNA fragments were analyzed by gel electrophoresis and purified by gel extraction.

Components	Volume (μL)
DNA template(WSSV genome)	1
Pfu DNA Polymerase 10X Buffer with $MgSO_4$	5
Forward Primer (10 μM)	1
Reverse Primer (10 μM)	1
dNTP (10 μM each)	1
Pfu DNA Polymerase (3 U/μL)	1
Deionized water	40
Total	**50**

Table2.1: Components in PCR cycle

Step	Cycle	Temperature (ºC)	Time
1	1	95	1min
2	30	94	30 s
		50	30 s
		72	30 s
3	1	72	15 min
4	1	15	Hold

Table2.2: Cycling parameters of PCR cycle

Components	Volume (μL)
*Xho*I	1
*Bam*HI	1
Buffer 2 (10X)	2
BSA (10X)	2
DNA (vector / insert)	6/14
Deionized water	8/0
Total	**20**

Table2.3: Double Enzyme Digestion

Vector was dephosphorylated. A 1:3 molar ratio of vector: insert was used for the cloning of the VP15 into vectors. The ligation was performed at 16℃ for 18 h.

Ethanol Precipitation

45μL PCR product was mixed with 55 μL water. Same volume CH_3Cl was added. 10 μL NaAC was applied to the sample. 2.5 Volume 100% ethanol was added to the mixture and left it at -80℃ for 10 mins. After discarding the supernatant, 2.5 volume 70% ethanol was further added to the tube and centrifuge at 13,000 rpm for 10 mins. Supernatant was discarded and the tube was dried on 50-60℃ by heat block. 20 μL M.Q water was added to get the concentrated DNA.

Vector Dephosphorylation

The mixture of all components (Table 2.4) was incubated at 37℃ for 30 mins. Another 2μL Apase stock was added and incubated at 37℃ for 30 mins again. The sample was kept at in 75-85℃ for 10 mins and later at room temperature.

Gel Extraction

DNA was run on 1.5% agarose gel. The DNA fragment was excised from the gel and weighed. The 300 μL buffer QG was added to 100 μL gel, and incubated at 50℃ for 10 mins (mix every 2-3 mins by votex). 100 μL isopropanol was added to the sample and loaded the mixture to the QIA quick column, centrifuged for 1 min. 0.7 mL of buffer PE was added and centrifuged for 1 min. The flow through was discarded. Finally, QIA quick column was placed into a clean 1.5 mL microcentrifuge tube. The DNA was eluted with 20 μL water.

Ligation

Following heat inactivation, the DNA insert was ligated to dephosphorylated

Components	Volume (µL)
Digested Vector DNA	20
10X Apase stock (Alkaline Phosphatase)	2
(Buffer: 0.5 µL, Apase: 0.5µL, Water: 4µL)	
Buffer (10X)	3
Deionized water	5
Total	**30**

Table2.4: Component in Vector Dephosphorylation

pGEX-6P-1 vector using 1:2 molar ratio of vector: insert. The ligation reaction (Table 2.5) was performed at 16℃ for 18 h.

Preparation of Competent E. *coli* Cells

A tube of glycerol stock of DH5-α *E. coli* cells was thawed and 50 µL was used to innoculate 2 mL of Lauria Bertiani (LB) medium. This was shaken in a 37℃ incubator set at 230 rpm for 18 h. Of this culture, 100 µL was transferred to 250 mL of LB medium and left to shake at 230 rpm at 37 ℃ till the optical density at 600 nm (OD_{600}) reached a value of 0.3 to 0.6. The bacterial culture was centrifuged at 4000 x g for 15 min at 4 ℃. The pellet was resuspended on ice in 1/20 volume TSS [LB broth (pH 6.1), 10 % polyethylene glycerol, 5 % DMSO (dimethylsulfoxide, Sigma) 10 mM $MgCl_2$, 10 mM $MgSO_4$; final pH 6.5]. The competent cells were kept on ice for 10 min, then aliquoted in 100 µL and snap-frozen in liquid nitrogen before storage at -80℃.

Transformation of *E.coli* Competent Cells

100 µL competent cell was thawed on ice before the ligation mixture was added. The cells were kept on ice for 30 minutes. After heat shock at 42℃ for exactly 60 seconds, the cells were kept on ice for an additional 2 minutes. 800µL LB medium was added to the cells and mixed by inverting up and down. The cells need to be incubated at 37℃ for 1 hour before plated onto LB agar plate. The plate was incubated overnight at 37℃.

Small Scale Plasmid DNA Isolation and Purification

Selected colonies were grown in 5 mL LB medium supplemented with 100 µg/µL ampicillin. The culture was shaken in a 37℃ incubator set at 250 rpm for 18 h. This was

Components	Volume (µL)
T4 DNA ligase	1
10X T4 DNA Ligase buffer	2
Dephosphorylated vector	6
Insert	11
Total	**20**

Table2.5: Ligation reaction

followed by centrifuging 5 mL of the cultures at 10,000 x g for 10 min. The plasmid DNA was extracted using the Minipreps DNA Purification System (QIAgene).

Pelleted bacterial cells were resuspended in 250 μL P1 buffer. 250 μL P2 Buffer were added and the tube was inverted 4-6 times in less 5 mins. 350 μL N3 Buffer was added and inverted 4-6 times. The content was centrifuged for 15 mins at 4℃ at 13,000 rpm. Supernatant was applied to a QIA prep column, centrifuged for 30- 60 s and the flow through was discarded. QIA prep spin column was washed by 0.5 mL Buffer PB and centrifuged for 30-60 s and flow-through was discarded. To remove residual wash buffer another centrifuge was done to the spin column. To elute the DNA, 50 or less M.Q water was added to the center of QIA prep column and centrifuged for 1 min.

Analysis of DNA Purity and Concentration

To analyze the amount and purity of DNA, spectrophotometric readings at $OD_{260\ nm}$ and $OD_{280\ nm}$ were taken after DNA purification. The ratios of $OD_{260\ nm}/OD_{280\ nm}$ were taken as the basis of DNA purity.

Verification of 6P-1-VP15 Construct

The plasmid DNAs and control vector were digested with *Bam*HI and *Xho*I restriction enzyme (New England Biolabs), according to the instructions (same as above). Each reaction mixture was incubated at 37℃ for 2 h. The reaction mixtures were stopped by heating at 75℃ for 15 min. 5 μL of each reaction mixture was analyzed by gel electrophoresis to verify for the correct size.

DNA Sequencing of the PCR Fragment

The PCR fragment clone was then sequenced using BigDye® Terminator v 3.0 Cycle Sequencing kit from both sense and antisense orientations. The DNAMAN sequence analysis software (Lynnon BioSoft, version 4.15) was then used to assemble and analyze the DNA sequences obtained. The components and cycling parameters of the sequencing reaction are shown in Table 2.6 and 2.7 respectively.

5' pGEX Sequencing Primer 5'-d[GGGCTGGCAAGCCACGTTTGGTG]-3'
3' pGEX Sequencing Primer 5'-d[CCGGGAGCTGCATGTGTCAGAGG]-3'

The precipitation of the PCR fragment was carried out according to the following protocol.

80 μL mixture (3 μL 3M NaOAC pH: 4.6; 62.5 μL Non-denatured 95% ethanol; 14.5 μL deonized water) was added to the PCR products and stood for 15 mins. The tube was spinned at Max speed for 20 mins and the supernatant was discarded. 500 μL 70% Ethanol was added and kept still for another 15 mins. The tube was spined for 5 mins at 13,000 rpm and the supernatant was aspirated carefully. The tube was left to dry thoroughly overnight.

Automated sequencing was then performed using the ABI Prism® 3100 DNA Automated Sequencer (Perkin Elmer).

5.3 Expression System of 6P-1-VP15

Transformation Plasmid 6P-1-VP15 into Expression Competent Cell BL21 (DE3)

The DNA plasmid (1 μL) was added to 100 μL KCM (100 mM KCl, 30 mM $CaCl_2$, 50 mM $MgCl_2$). Frozen BL 21 competent cells (100 μL) were added to the DNA/KCM mixture. The mixture was gently mixed, kept on ice for 30 mins and plated onto a LB agar plate pretreated with ampicillin. The plate was incubated overnight at 37℃.

Components	Volume (μL)
Double-stranded DNA template (0.4 μg)	1
Terminator Ready Reaction Mix	4
Primer (10 pmol)	1
Deionized water	4
Total	10

Table2.6: The components in DNA sequencing PCR

Step	Cycle	Temperature (^{o}C)	Time
1	25	95	30 s
		50	5 s
		60	4 min
2		4	Hold

Table2.7: Cycling parameters for DNA sequencing PCR

Determination of Target Protein Solubility

A single colony was picked up from plate to inoculate 5 mL LB medium containing 100ug/mL ampicilin (LBA) and grown overnight at 37℃ with shaking. 2.5 mL overnight culture was further inoculated into 50 mL LBA and grown with vigorous shaking until the OD_{600} was about 0.5. 1 mL sample (non-induced control) was taken out, centrifuged and resuspended in 50 μL SDS-PAGE sample buffer. IPTG was added to the rest of culture to a final concentration of 0.5 mM to induce protein expresssion for further 3-3.5 hours. 1 ml sample was collected, centrifuged and resuspended in 100 μL SDS-PAGE sample buffer. The rest of culture was harvested and centrifuged at 5000 rpm for 10 minutes at 4℃. The cell pellet can be resuspended in 5 mL lysis buffer and further sonicated thoroughly. The lysate was centrifuged at 16,000 rpm for 20 minutes. 100 μL supernatant (soluble) was taken out and added to 100 μL 2 × SDS-PAGE sample buffers. Alquots of pellet (insoluble protein) were resuspended in 100 μL 2 × SDS-PAGE sample buffer.

All the 4 samples were boiled for 5 minutes and microcentrifuged. 10 μL of supernatant was loaded into each lane of the SDS-PAGE gel.

Unlabeled Protein Expression

For unlabeled protein, a single colony of host cell harboring the construct 6P-1-VP15 was picked up from the agar plate and inoculated into 10 mL LB medium containing 100μg/mL ampicillin. The overnight culture was further inoculated into 1 liter fresh LB medium with the same concentration of ampicillin. The cells were grown at 37^0C with vigorous shaking until the $O.D._{600}$ reached about 0.5. IPTG (isopropyl -D-thiogalactoside) was then added to a final concentration of 0.5 mM to induce the expression for an additional 3 hours. The cells were harvested by centrifugation at 5000

rpm for 20 minutes. The cell pellet was kept frozen at -20℃ or directly resuspended in lysis buffer for further purification.

^{15}N Uniformly Labeled Sample

The production of ^{15}N uniformly labeled sample is similar to that of unlabeled sample, except that M9 minimal media containing ^{15}N NH_4Cl and ^{12}C glucose as the sole nitrogen source and the carbon source was used to grow the cell. Protein experiment was induced at a cell OD_{600} around 0.4-0.5. Long induction time (4 hours) is used to increase the yield of expressed protein.

1 liter M9 minimal medium contain: 1 g NH_4Cl, 1 g Glucose, 0.0147g $CaCl•2H_2O$, 0.493 $MgSO_4•7H_2O$, 7.52g Na_2HPO_4, 3g KH_2PO_4, 0.5g NaCl.

5.4 Protein Purification (Fig. 2.5)

5.4.1 Pre-Column Treatment

Resuspended pellet in 20mL of lysis buffer [2mM DTT, 500mM NaCl, 1mM EDTA, pH 8.0 (adjusted by Tris)] was sonicated thoroughly and then inclusion bodies and cell debris were spun down at 18,000 rpm for 30 minutes. The supernatant was discarded and the pellet was resuspended in 20mL GST column binding buffer (the same as lysis buffer). The insoluble cell debris was removed by centrifugation.

5.4.2 GST (glutathione S-Transferase) Affinity Chromatography

The soluble denatured protein was loaded into a 2mL GST resin column which had been equilibrated with binding buffer. After the GST-tagged protein was bound onto the resin, the resin was further washed thoroughly with 30 mL washing buffer [2mM DTT, 500mM NaCl, 1mM EDTA, pH 8.0 (adjusted by Tris)] until the absorption at 280 nm of the washing was very low.

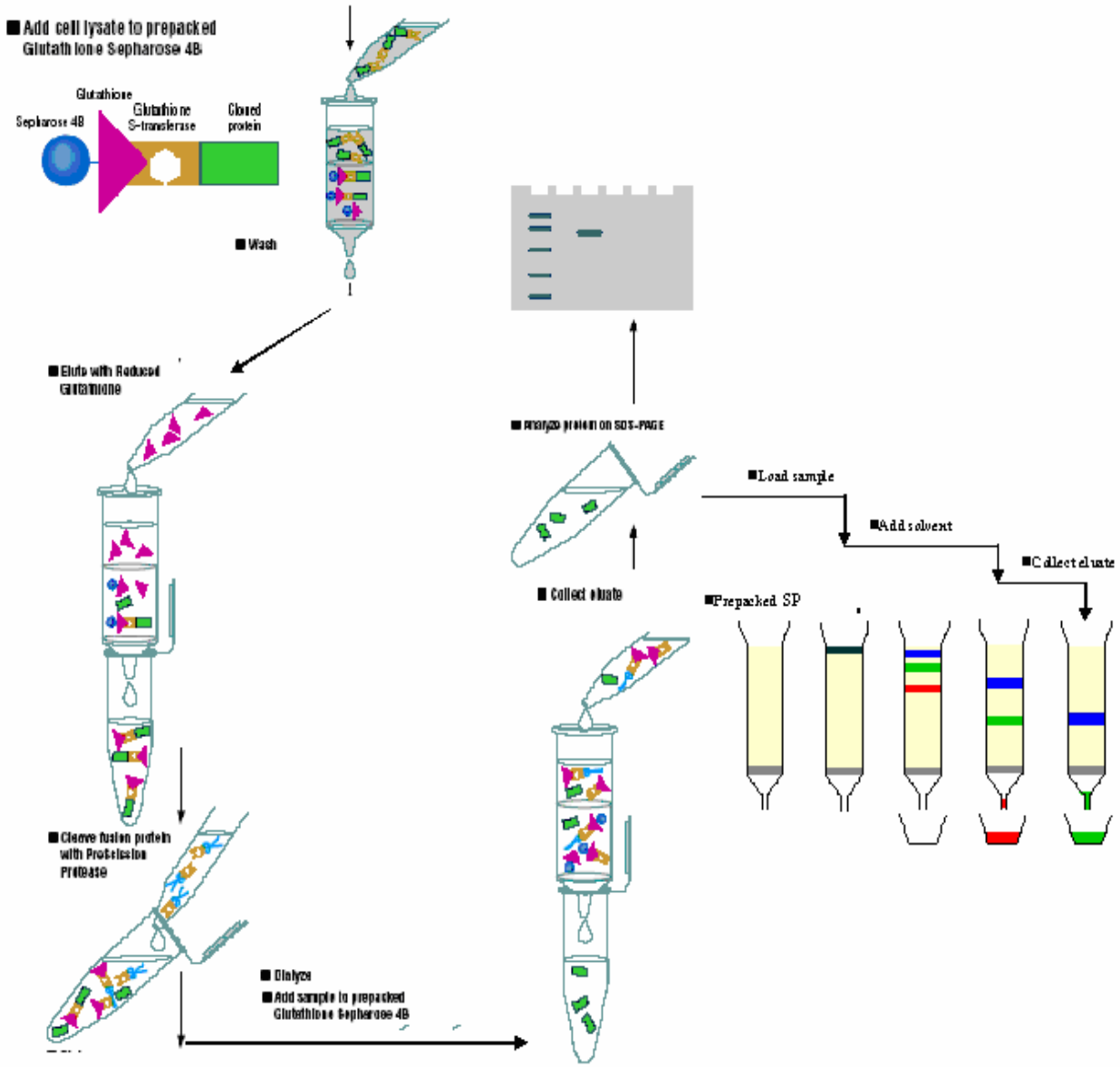


Fig.2.5 Protein Purification Procedure

GST-tagged protein can be eluted out by a small amount of elution buffer [2mM DTT, 500mM NaCl, 1mM EDTA, 20mM reduced glutathione pH 8.0 (adjusted by Tris)]. For every step, 0.5 mL sample was collected and boiled with SDS loading buffer to be further checked by SDS-PAGE.

5.4.3 PreScission™ Protease Cleavage

The yield and concentration of protein was estimated by measuring absorbance at 230nm. GST-tag can be removed by cleavage with PreScission™ Protease at 4℃ overnight. The completion of cleavage was checked by SDS-PAGE. Cleaved protein wish GST and PreScission™ Protease was dialyzed and loaded to column with GST resin again to remove the GST and protease.

5.4.4 Ion Exchange Column

The flow through which contained protein VP15 was then loaded onto a SP Sepharose column. The mobile phase was equilibrated by washing buffer. A slow flow rate (0.5mL/min) of elution buffer [2mM DTT, 2M NaCl, 1mM EDTA, pH 8.0 (adjusted by Tris)] was used to run the column. The fractions with the highest absorbance at 280 nm were pooled. The purity of the fractions was checked by SDS-PAGE and the most purified fractions were pooled and dialyzed in buffer to remove NaCl. The purified protein was buffer exchanged to NMR buffer (50mM phosphate buffer with 10% D_20, pH 6.0 and 0.5 M Na_2SO_4) by ultra-filtration using Centriprep3 (Milipore) and concentrated to 0.5 mL for NMR experiments.

The protein was sequence identified by N-terminal Sequence and MALDI-MS.

5.5 Characterization of 6P-1-VP15

5.5.1 2-D NMR Experiments

^{1}H-^{15}N-HSQC experiments were carried out on a Bruker AVANCE 500 MHz spectrometer equipped with pulse field gradients and cryo-probe. ^{15}N uniformly labeled sample samples were contained in 5mm Shigemi NMR tube and the experimental temperature was kept at 4℃ .

5.5.2 CD Experiments

CD experiments were carried out on Jasco1-800 spectropolarimeter. In parameters settings, the scanned wavelength was set between 190nm to 260 nm and scanned 3 times for every sample. 200μL unlabelled sample of 6P-1-VP15 was loaded to the CD tube and the buffer was used as the reference spectrum.

5.5.3 Gel Mobility Shift Assay

The gel mobility shift assay was performed primarily as described by Ausubel 1995. An appropriate amount of different DNA fragments amplified by PCR from WSSV was added into the purified 6P-1-VP15. Some short oligonucleotides including single strand and double strand DNA (annealed by two single oligos) were also designed to react with the proteins. The reaction mixtures were mixed and incubated at 4℃ for 2 hours, analyzed by 3% agarose gel electrophoresis and photographed under ultraviolet light.

Annealing Oligo

(a) Resuspending the Oligonucleotides: both complementary oligonucleotides were resuspended at the same molar concentration, using Annealing Buffer.

(b) Annealing the Oligonucleotides: Equal volumes of both complementary oligos (at

equimolar concentration) were mixed in a 1.5mL microfuge tube and placed in a standard heatblock at 90 – 95℃ for 5 mins. After slow cooling to room temperature for 45-60 minutes, the tubes were stored on ice or at 4℃ ready to use.

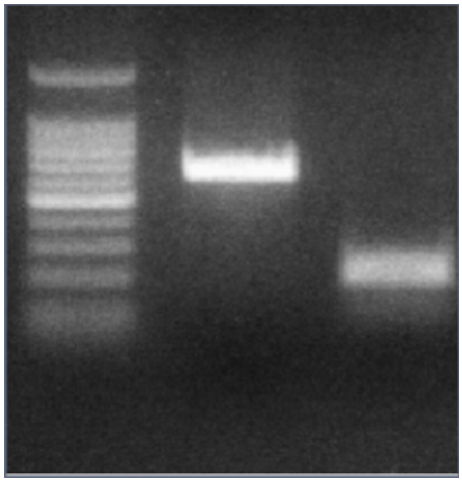
(Annealing Buffer: 10mM Tris, pH 7.5 - 8.0, 50mM NaCl, 1mM EDTA)

6. Result and Discussion

6.1 PGEX-6P-1 Expression System

183 bp DNA fragment encoding the proteins was PCRed out from WSSV genomic DNA (Fig2.6) and cloned into pGEX-6P-1 vector (N-terminal GST-tagged) (Fig.2.7) with restriction sites *Bam*HI/*Xho*I.

pGEX system is the most powerful system yet developed for the cloning and expression of recombinant proteins in *E. coli*. GST fusion proteins are produced in *E. coli* cells containing a recombinant pGEX-6P-1 plasmid. Protein expression from a pGEX plasmid is under the control of the *tac* promoter, which is induced using IPTG. Target genes are cloned in pGEX plasmids under control of strong bacteriophage T7 promoter; expression is induced by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so selective and active that, when fully induced, almost all of the cell resources are converted to target gene expression; the desired product can reach more than 50% of the total cell protein a few hours after induction. Although this system is extremely powerful, it is also possible to attenuate expression levels simply by lowering the concentration of IPTG. Decreasing the expression level may enhance the yield of the soluble fraction of some target proteins. Once established in a non-expression host e.g. DH5α, target protein expression may be initiated by transferring the plasmid into an expression host containing a chromosomal copy of the T7 RNA polymerase gene under *lac* operator control. In this case, expression is induced by the addition of IPTG to the bacterial culture. Two types of T7 promoter and several hosts that differ in their stringency of suppressing basal expression levels are available, providing great flexibility and the ability to optimize the expression of a wide variety of target genes.



Marker Postive control VP15

Fig.2.6 PCR product of VP15 from WSSV genome

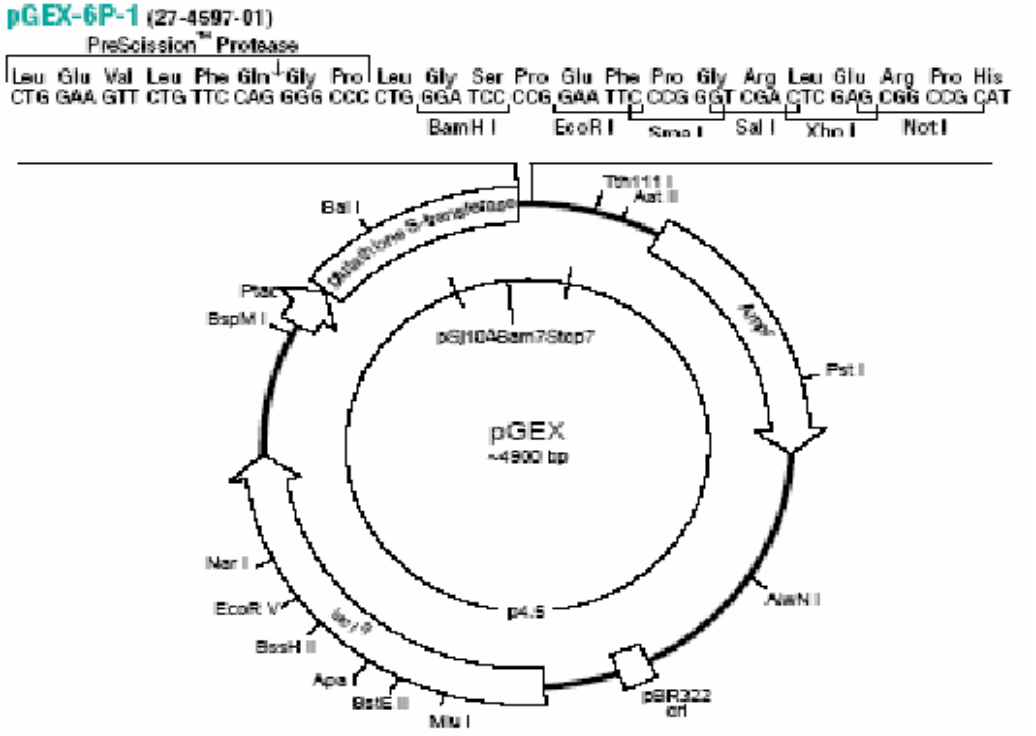


Fig.2.7 Vector information of pGEX-6P-1

Some other different expression systems were also been tried to express this VP15 protein such as pET and secreted system. However, the protein was either not expressed or expressed in minimal quantiting to warrant any further study.

6.2 Expression and Purification of the Protein

DNA sequence analysis confirmed that plasmid 6P-1-VP15 was transformed into BL21 cell for expression. At 37℃, when the cells grow to the exponential phase, the production of GST-tagged VP15 could be induced by adding IPTG to a final concentration of 0.5 mM. After the cells were lysed by sonication and centrifugation, GST-VP15 was released into the supernatant due to its high hydrophilicity.

GST Affinity Chromatography

Fusion proteins were purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B. This affinity chromatography step was used to remove most of the contaminating proteins to facilitate the later purification steps. Unbound protein was removed by washing the column extensively with binding buffer. Elution buffer containing 20 mM GSH was used to elute the bounded GST-tagged protein from beads. This single step using GST affinity chromatography can improve the purity of the proteins to above 80%.(Fig.2.8)

SDS-PAGE technique is used to assess the purity and to estimate the molecular weight of the protein.

The pGEX-6P-1 plasmids were for inducible, high level intracellular expression of genes or gene fragments as fusions with *Schistosoma japonicum* GST. GST occurred naturally as a 26 kDa protein that can be expressed in *E. coli* with full enzymatic activity.

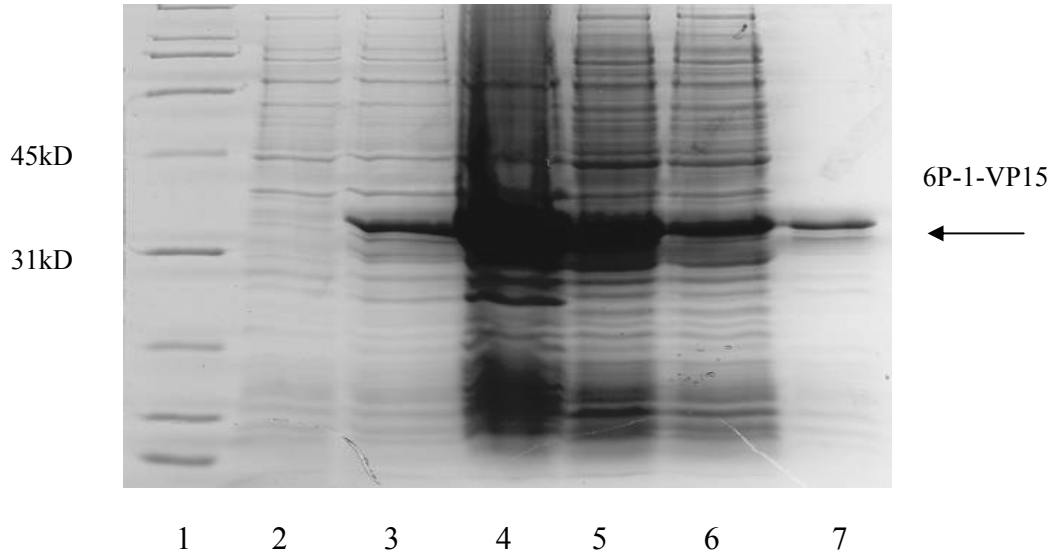


Fig.2.8 Protein Expression and GST Affinity Column Purification

(1) Low molecular weight marker; (2)Total protein extract of non-induced culture; (3)Total protein extract of induced culture containing GST-VP15 fusion protein; (4)Inclusive body after dialysis and centrifugation step; (5) Supernatant after 18,000 rpm centrifugation step; (6) (Flow-through fraction from GST affinity column; (7) 6P-1-VP15 eluted by 20 mM GSH from GST affinity column

Cleavage recognition sequence was located immediately upstream from the multiple cloning site on the pGEX-6P-1 plasmids.

Precession Cleavage

Following the elution of the GST fusion protein, eluate was dialyzed the extensively against PreScission Cleavage Buffer in order to remove reduced glutathione from the sample. 1 μL (2 U) of PreScission Protease was added for each 100 μg of fusion protein in the eluate and incubated at 5℃ for overnight. (Fig.2.9)

PreScission Protease is a fusion protein of GST and human rhinovirus 3C protease. The protease specifically recognizes the amino acid sequence Leu-Glu-Val-Leu- Phe-Gln ¯ Gly-Pro cleaving between the Gln and Gly residues. Since the protease is fused to GST, it is easily removed from cleavage reactions using Glutathione Sepharose.

Once digestion was complete, the sample was loaded to a washed and equilibrated Glutathione Sepharose column to remove the GST portion of the fusion protein and the PreScission Protease from the protein of interest. (Fig.2.10)

Ion Exchange Chromatography

The last SP Sepharose cation chromatography was used to remove the other remaining contaminating proteins (Fig.2.11). SDS-PAGE showed that single peak composed entirely of the VP15, with the purity above 95% was obtained. At this stage, the protein was normally pure enough for NMR analysis.

From the data, VP15 migrated as a 15KDa protein on SDS-PAGE instead of its expected 7 KDa Molecular weight. That is probably because the protein has an extremely basic pI value of 13.2 due to its amino acid composition rich in lysine (22 %), arginine (23%) and serine (25 %). In SDS-PAGE, for proteins with highly basic residues tend to

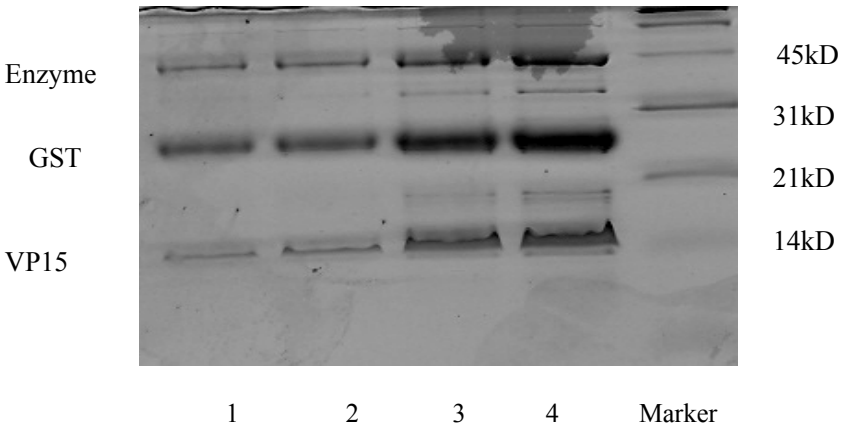


Fig.2.9 Protein Digestion of PreScission
(1—4 lanes: different fractions after PreScission Cleavage)

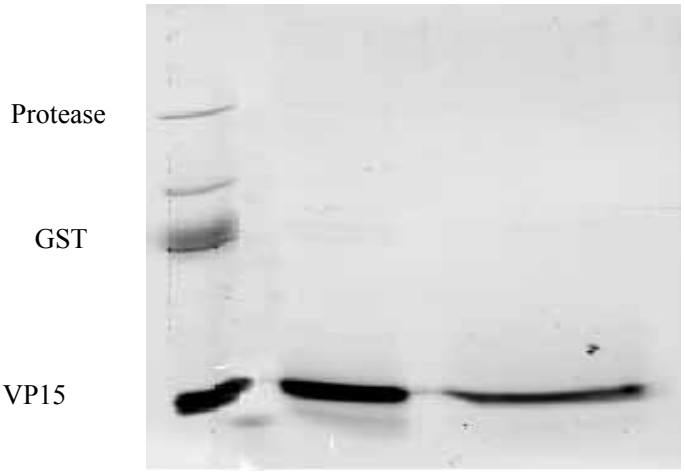
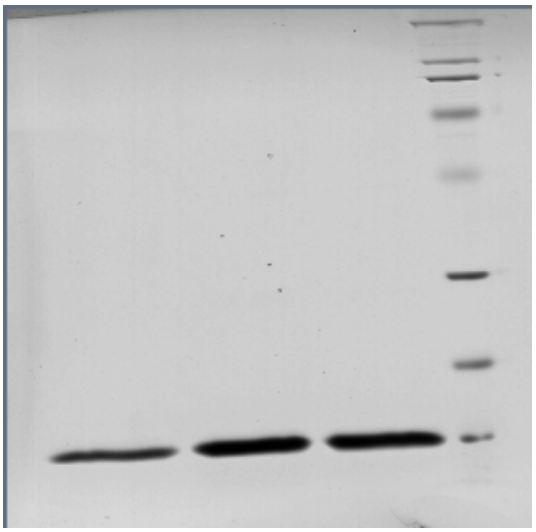


Fig.2.10 Remove GST and Enzyme
(1) Sample after protease digestion; (2) and (3) different eluted fraction after GST beads



VP15

1 2 3 marker

Fig.2.11 Final purified by SP Sepharose
1, 2, 3 different eluted fractions after SP Sepharose column purification

have an apparent molecular weight because basic amino acids are highly attracted to the negative charge on SDS, resulting in the increase of the mass:charge ratio of the SDS-protein complex, thus lowering electrophoretic mobility.

On-Column Purifcation

Based on the characteristics of PreScission Protease which is fused with GST, target proteins can be purified using either on-column purification or digestion following elution from the matrix (as stated above). As the latter method is more flexible, this method was chosen to purify the target proteins in the initial stage of experiment. However, on column digestion purification provides a more convenient way. The fusion protein can be digested with the appropriate site-specific protease while the fusion protein is bound to Glutathione Sepharose 4B. Cleavage of the bound fusion protein eliminates the extra step of separating the released protein from GST since the GST moiety remains bound to the matrix while the cloned protein is eluted using wash buffer. So this method was used mostly used to get the target protein. (Fig.2.12)

Protein Identification

The purified protein was identified by N-terminal Sequence Determination and MALDI-MS Molecular Weight Analysis.

Expected: Sequence of Expression Protein:
(cleavage site)**G P L G S** +(protein) V A R S S K T K S R R G S K K R S T T A G R I S K R R S P S M K K R A G K K S S T V R R R S S K S G K K S G A R K S R R *
Theoretical Mw: 7102.26 D

Result of MALDI WM: 7112.2 D, which was around 10 Dalton different with expected MW.

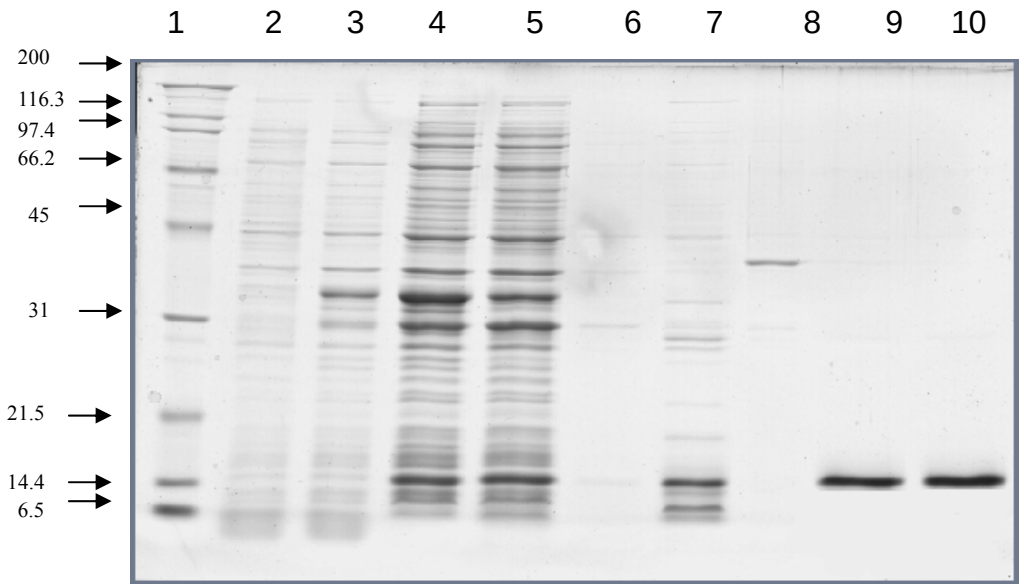


Fig.2.12 On-column purification procedure

(1)Low molecular weight marker; (2)Total protein extract of non-induced culture (3)Total protein extract of induced culture containing GST-VP15 fusion protein; (4) Supernatant after 18,000 rpm centrifugation step; (5) Flow-through fraction from Glutathione Sepharose 4B column (6) Last wash fraction; (7) Cleavage product eluted after on-column digestion with PreScission Protease; (8) Flow-through fraction from SP Sepharose column; (9) and (10) VP15 eluted from SP Sepharose column

Result of N-terminal sequence: Altogether 11 cycles was carried out from G to K (GPLGS+VARSSK), which matched the N-terminal sequence of target protein

^{15}N Labeled Protein Purification for NMR Study

The advances of molecular biology techniques have facilitated the development of biomolecular NMR greatly. The various over-expression systems not only make the production of ^{13}C, ^{15}N labeled protein feasible, but also simplify the purification steps as well. The most commonly used method for isotopic enrichment of proteins is overexpression in *E. coli* using minimum medium containing ^{13}C and ^{15}N as the sole carbon and/or nitrogen source. *E. coli* can efficiently take up and incorporate into proteins exogenously supplied isotopically-labeled compounds into target proteins. The amino acid metabolic pathways of *E. coli* are also well understood, which allow them to be genetically blocked or altered if required to obtain the desired enrichment pattern.

The ^{15}N labeled VP15 migrated a little bit slower than unlabeled VP15 in SDS-PAGE (Fig.2.13) because its molecular weight was greater than the unlabeled one due to the incorporation of N^{15} label.

Although the prokaryotic bacterial expression system can be widely used for protein expression, many eukaryotic proteins have been found to fold improperly in bacteria and therefore cannot be used for structural studies. Lack of post-translational modification, such as glycosylation, phosphorylation by the host *E.coli* is responsible for this phenomenon. In that situation, expression in yeast could be an alternative way to obtain the folded eukaryotic proteins.

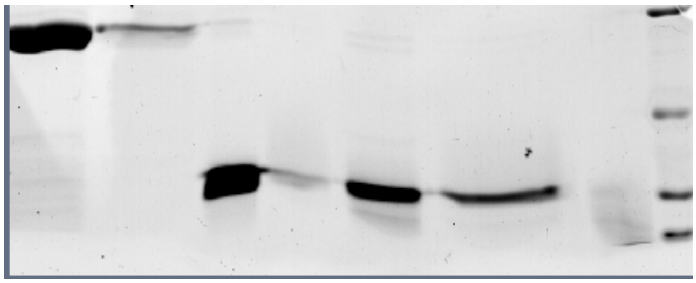


Fig 2.13 Purification of ^{15}N labeled VP15

Lane (1) (2): labeled 6P-1-VP15 fusion protein; (3) ^{15}N labeled VP15 protein; (4) and (5) unlabeled VP15 protein as control

6.3 Characterization of VP15

6.3.1 Prediction of VP 15 Secondary Structures

Amino Acid:	**10|**	**20|**	**30|**
	MVARSSKTKS	**RRGSKKRSTT**	**AGRISKRRSP**
Secondary Structure:	CCCCCCCCCC	CCCCCCCCCC	CCCECCCCCC
Amino Acid:	**40|**	**50|**	**60|**
	SMKKRAGKKS	**STVRRRSSKS**	**GKKSGARKSR**
Secondary Structure:	CCCHCCCCCC	EEEEECCCCC	CCCCCCCCCC

From prediction, VP15 was mostly composed of random coil and it probably was an unstructured protein.

6.3.2 CD spectra of VP15

CD (Circular Dichroism) is a quick, low resolution technique for determining protein secondary structure. CD experiment was carried out to analyze the secondary structure of VP15. The buffer was used a control. (Fig.2.14)

CD spectra between 260 and approximately 180 nm (Far UV-CD) can be analyzed for the different secondary structural types of proteins: alpha helix, parallel and antiparallel beta sheet, turn, and other. For a folded protein, its secondary structure can be detected from different peaks:

- Far UV-CD of Random Coil (RC):

  Positive at 212 nm Negative at 195nm

- Far UV-CD of β-sheet:

  Positive at 197 nm Negative at 218nm

- Far UV-CD α-helix:

Positive $_{\text{perpendicular}}$ at 192 nm Negative at 222 nm and Negative $_{\text{parallel}}$ at 209nm

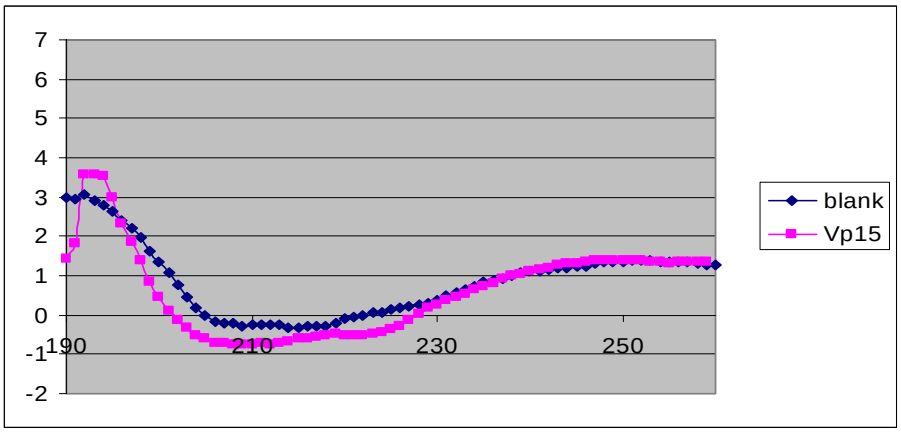


Fig.2.14 CD spectra of VP15

However, from the spectra (Fig.2.14), no significant difference can be traced between the protein and the control. Even in different buffers (Fig.2.15), no structural change was observed. The difference among all curves was mostly before 205nm, which was effected by different salt conditions. The CD data here suggested that VP15 was a poorly-folded protein and was consistent with the predicted result.

6.3.3 2D ^{1}H-^{15}N HSQC Spectrum

Two-dimensional NMR specta can explore and exploit two types of interactions ("through-bond coupling" and "through-space coupling") between same (homo) or different (hetero) nuclei in a molecule to get very useful conformational information.

The ^{1}H-^{15}N HSQC (Heteronuclear Single-Quantum Correlation) is a 2D experiment with one ^{1}H frequency and one ^{15}N frequency. Each amino acid residue (except proline) gives one signal that corresponds to the N-H amide group. Depending on pH and solvent exposure side chain NHx groups can be visible, too. Folded Proteins or protein domains display a broad distribution of NMR frequencies resulting in a good spread-out of signals in the 15N-HSQC spectra. Unfolded amino acid residues share similar NMR frequencies resulting in heavy overlap of signals.

The HSQC experiment is in fact a double INEPT experiment. This experiment correlates protons with their directly attached heteronuclei. Proton magnetization is detected while the low-gamma nuclei evolved during the evolution time. Because of the detection of the high frequency nuclei, this sequence is very sensitive, so ^{1}H-^{15}N HSQC experiment was carried to further analyze and identify the secondary structure of VP15.

^{1}H-^{15}N HSQC Spectra was collected for ^{15}N isotope-labeled samples and the broad resonance peaks were obtained, which resulted in poorly dispersed HSQC spectra

(Fig.2.16). This result suggested that VP15 was almost unstructured.

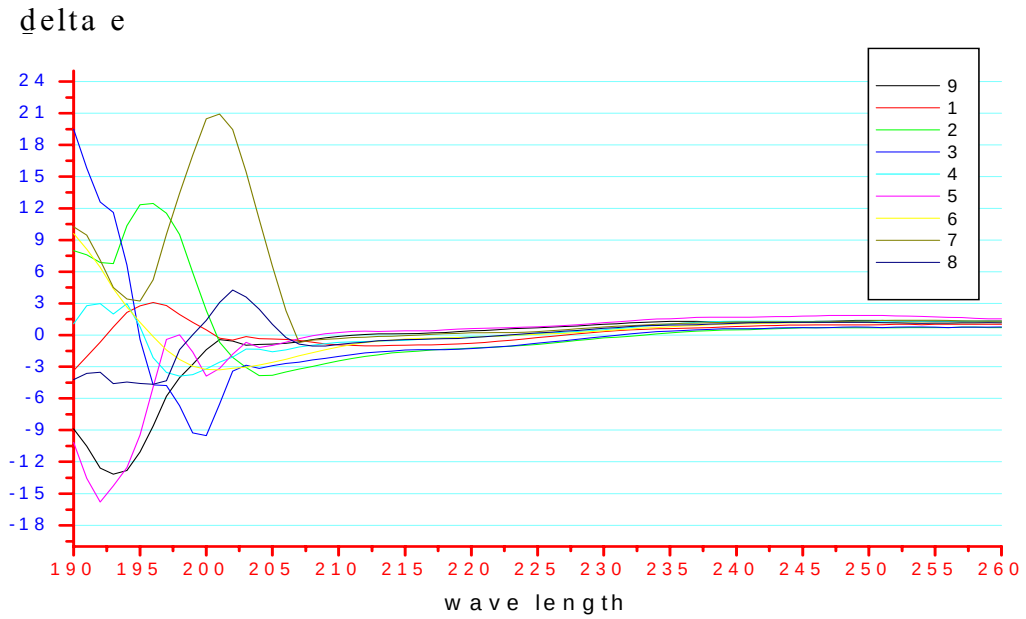


Fig.2.15 CD spectra of VP15 in different buffers

1. 500mM NaCl,20mM Tris, pH. 9.5
2. 250mM NaCl,10mM Tris, pH. 9.5
3. 50mM NaCl,10mM Tris, pH. 9.5

4.500mM NaCl,20mM Tris, pH. 7.0
5.250mM NaCl,10mM Tris, pH. 7.0
6. 50mM NaCl, 10mM Tris, pH. 7.0

7.500mM NaCl,20mM Tris, pH. 4.0
8.250mM NaCl,10mM Tris, pH. 4.0
9. 50mM NaCl, 10mM Tris, pH. 4.0

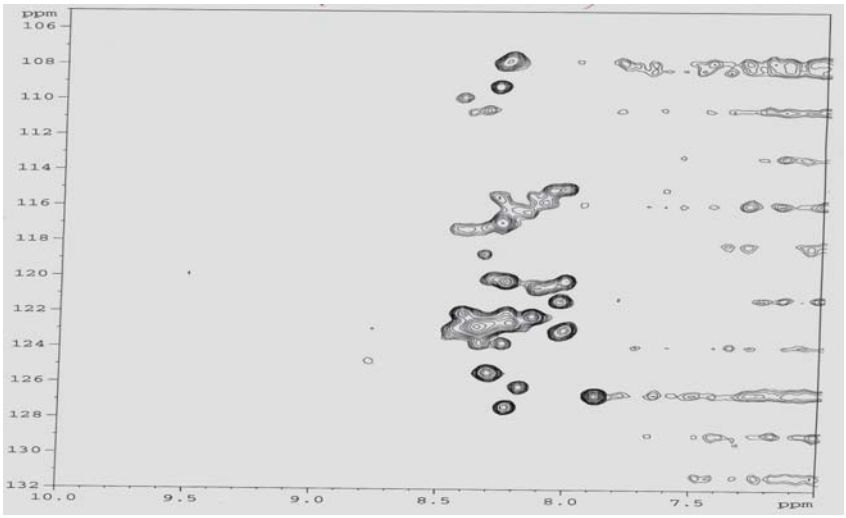


Fig.2.16 The ^{1}H-^{15}N HSQC spectrum of ^{15}N labeled VP15

6.3.4 DNA-Protein Interaction and Gel Mobility Shift

VP15 was a major structural protein of WSSV. However, its secondary structure was mostly composed of random coil based on above result.

VP15 had a strong positive charge along the whole length of the polypeptide. A putative function for this highly basic protein is that of a histone-like, DNA-associated protein. Its small size, positive charge and its presence within the nucleocapsid fraction are in agreement with this hypothesis. The DNA-Protein interaction capability might be important in DNA packing and understanding how it gives rise to the precise assembly pathway from this interaction. The capability of VP15 to bind DNA was tested and gel mobility shift was carried out to select the binding site if there was any. Also, it was important to investigate whether there are any major structural conformational changes associated with the binding of VP15 to DNA.

The gel shift, or electrophoretic mobility shift assay provides a simple and rapid method for detecting DNA-binding proteins. It is based on the differential mobility of free DNA and DNA-protein complexes in native (non-denaturing) polyacrylamide or agarose gels. It could be observed that complexes of protein VP15 and DNA migrated more slowly than free DNA fragments through agarose gel (Fig.2.17), which suggested that VP15 was a DNA binding protein and didn't select specific DNA sequence to bind. As control, there was no migration observed in protein A with DNA fragments (Fig.2.18).

Furthermore, the assay was also performed by incubating a purified protein with ssDNA(single strand DNA) and dsDNA (double strand DNA). From the Fig.2.19, no migration could be observed in protein with ssDNA and very weak band of protein with dsDNA.

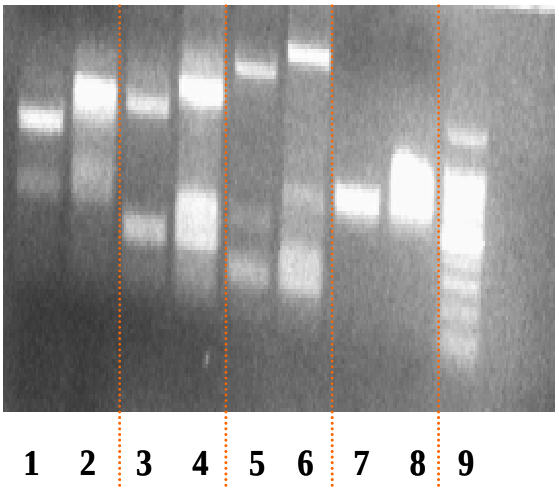


Fig.2.17 Different DNA fragments and DNA fragments/VP15 mixture

(1, 3, 5, 7 DNA alone; 2, 4, 6, 8 DNA+ VP15)

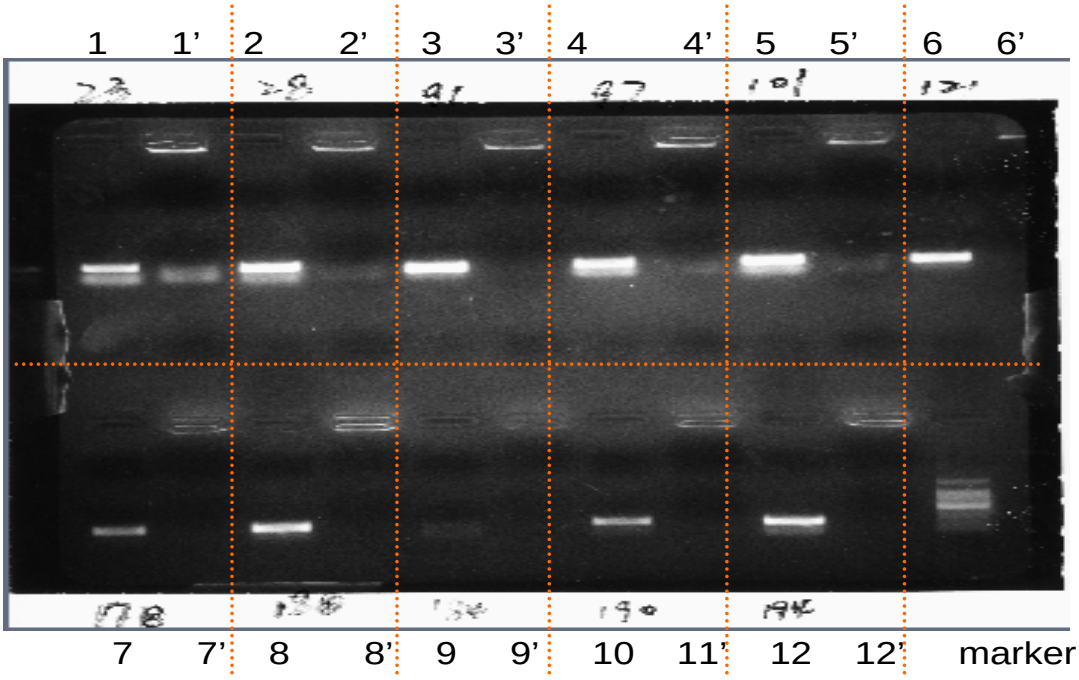


Fig.2.18 VP15/DNA mixture and Protein A/DNA mixture as control

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, control protein A with different ORF of WSSV;
1', 2', 3', 4', 5', 6', 7', 8', 9', 10', 11', 12', VP15 with different ORF of WSSV

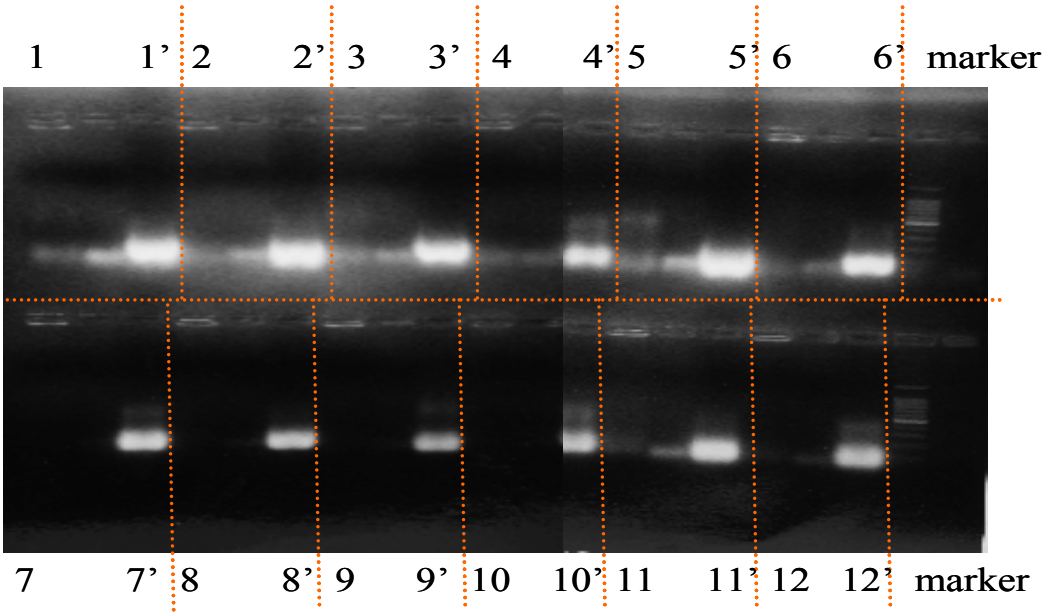


Fig.2.19 VP15/ssDNA oligos and VP15/dsDNA oligos

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, VP15 with dsDNA
1', 2', 3', 4', 5', 6', 7', 8', 9', 10', 11', 12', VP15 with ssDNA

From the above experiments, we could see that VP 15 was a non-specific dsDNA binding protein. In above films, some of DNA bands with VP15 were missing. It was probably because the DNA-protein complex was too big to enter the gel matrix. In additional, it also may because of the incorrect protein: DNA ratio. The amount of protein should be carefully titrated to get a satisfactory result.

The structural status of VP15 with DNA was checked by CD. However, from Fig.2.20, no significant difference was observed in wavelength 200nm—260 nm between protein alone and protein with DNA. The difference before 200nm may be interfered by DNA. VP15 still appeared like an unstructured protein even with the DNA.

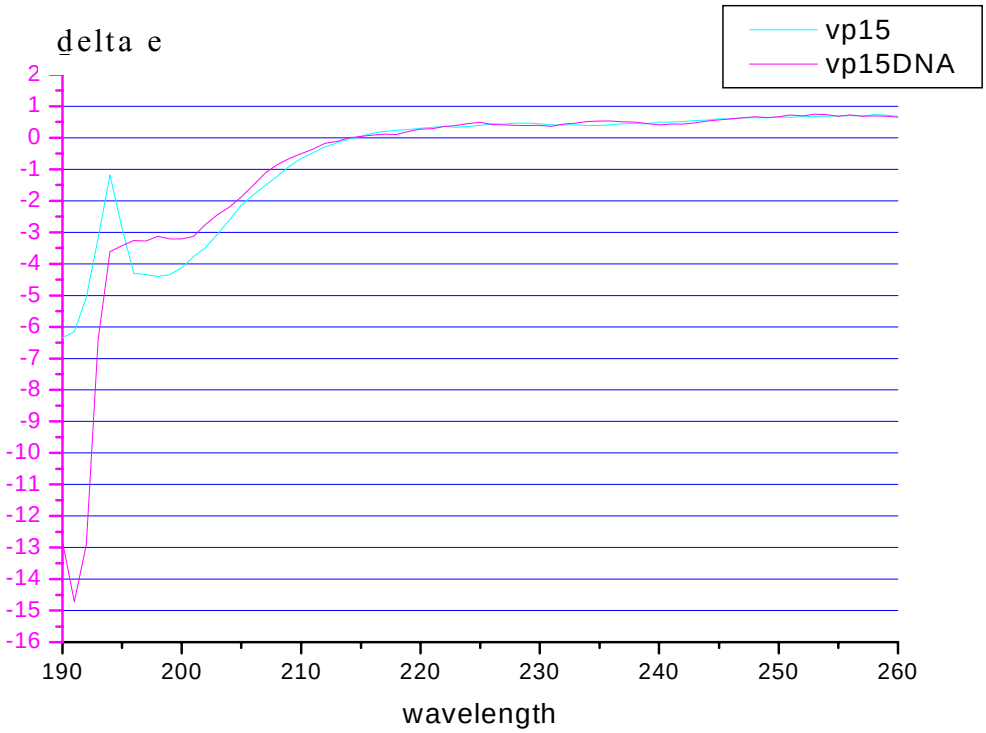


Fig.2.20 CD spectra of VP15 and VP15+DNA

7. Conclusion and Future Work

7.1 Conclusion

This study aimed to characterize the nucleocapsid proteins of white spot syndrome virus. VP15 was the major capsid protein of WSSV. The full length protein VP15 was cloned and expressed in *E. coli* and first purified by GST-affinity chromatography and further purified using ion exchange chromatography. The same expression system and purification steps were employed to produce heavy isotope labeled sample for the NMR studies.

The Secondary structure analysis based on CD spectra and 2-D NMR spectra indicated that this protein VP15 was a random coil. This protein was further characterized by gel mobility shift assay and the result showed that it was a non-specific DNA binding protein. However, it only interacted with double stand DNA instead of single strand DNA. Further more, there was no distinct difference between 2-D structures of the protein and protein with DNA from CD spectra.

7.2 Future work

1. To study the protein VP15 *in vivo* and elucidation the structure of the native protein. It is important to characterize the native structure of VP15, whether there is any post-translative modification of the protein.
2. Assembly of the nucleocapsid by protein-protein, protein-DNA interaction *in vitro*. It is needed to study these interactions under conditions closely approaching those *in vivo* inside the cell (N.J.Dimmock *et al,* 1946). Since pH and ionic strength play an important role on the formation of aggregates of capsid proteins, they need to be carefully adjusted in order to achieve the capsid proteins *in vitro* assembly. With the

complexes formed in these interactions, the dynamic nature, processivity, and high precision of these localized interactions can be captured.

CHAPTER 3. NUCLEOCAPSID PROTEIN VP24 OF WSSV

1. Summary

cDNA of VP24 was subcloned into pGEX-6P-1 vector and overexpressed as a GST fusion protein. Due to the strong hydrophobic domain in N-terminal, the expressed protein occurred as insoluble inclusion body. Its solubility in different detergents with different concentrations was compared. Lysis buffer containing 0.3% N-laurylsarcosine (sarkosyl) was chosen in which the protein exhibited a high solubility. After obtaining the soluble protein, affinity chromatography was used for its purification. Antibody of this protein was generated by purified protein. It was used as antigen to immunize mice by intra dermal injection. Antibody was generated by this method and can be used to further identify the location of this protein in WSSV virions in order to distinguish the difference of envelope proteins VP26 and VP28.

2. Introduction to VP24

2.1 Characterization of VP24 and Its Similarity with VP26 and VP28

VP24 ORF, encompassing 627nucleotides, encodes 208 amino acids. Computer analysis of the 208 amino acids showed that a strong hydrophobic region was present at the N-terminus of VP24 (Fig.3.1), including a putative transmembrane *a*-helix formed by amino acids 6 through 25. The algorithm of Garnier *et al*., (1978) predicted several other α-helices and β-sheets along the protein. It is remarkable that VP28, VP26 and VP24 roughly have the same size (206 amino acids) but have distinct electrophoretic mobilities. This may be due to differences in isoelectric points, conformational differences or post-translational modifications. Homology searches with WSSV VP24 were performed against GenBank, EMBL, SWISS-PROT and PIR databases using FASTA, TFASTA and BLAST, but no significant homology with structural proteins from other large DNA viruses could be found. Surprisingly, statistically significant similarity was found with the sequence of two other WSSV virion structural envelope proteins, VP26 and VP28 (Van Hulten *et al*., 2000), with 41 and 46% amino acid similarity, respectively. An alignment of the three WSSV proteins was made using ClustalW (Thompson *et al*., 1994), and revealed several conserved regions (in the hydrophilicity plots (Fig. 3.1). These residues might represent a transmembrane region Fig3.2). In the N-terminal region a well conserved stretch of amino acids is observed at positions 15±30. A strong hydrophobic region with a *a*-helix is observed for all three proteins, or be involved in the interaction of the structural proteins to form homo- or hetero- multimers. In two other conserved regions, around positions 88±102 and positions 138±148, the algorithm of Garnier *et al* (1978) predicted a α-helices and β-sheets for all three proteins. (Fig3.1 and 3.2)

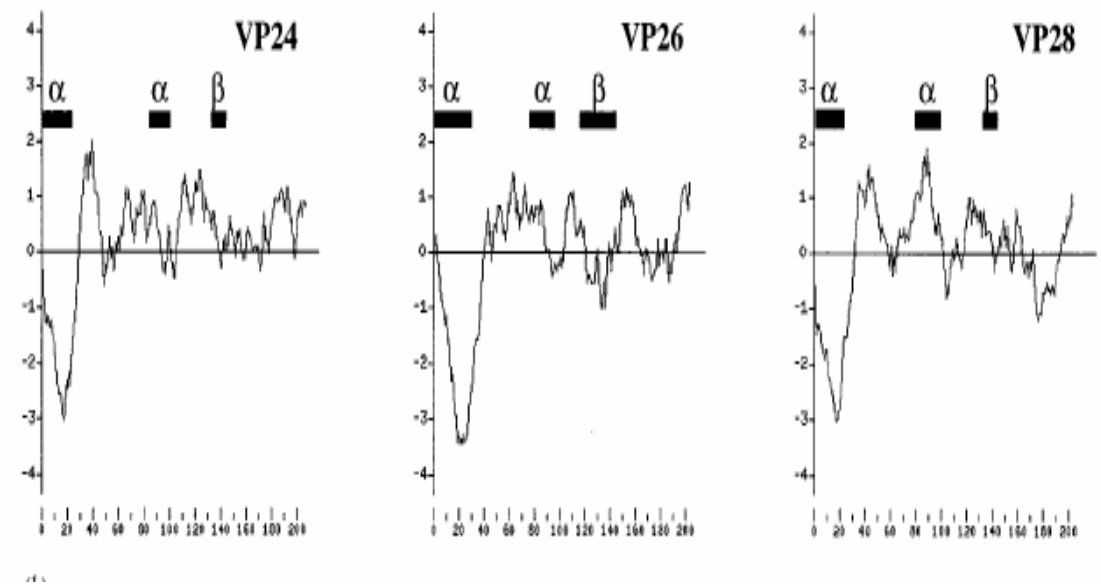


Fig.3.1 Hydrophilicity plots of WSSV VP24, VP26 and VP28
(The amino acid number is on the abscissa, and the hydrophilicity value on the ordinate. *a*-Helices and *β*-sheets are indicated)

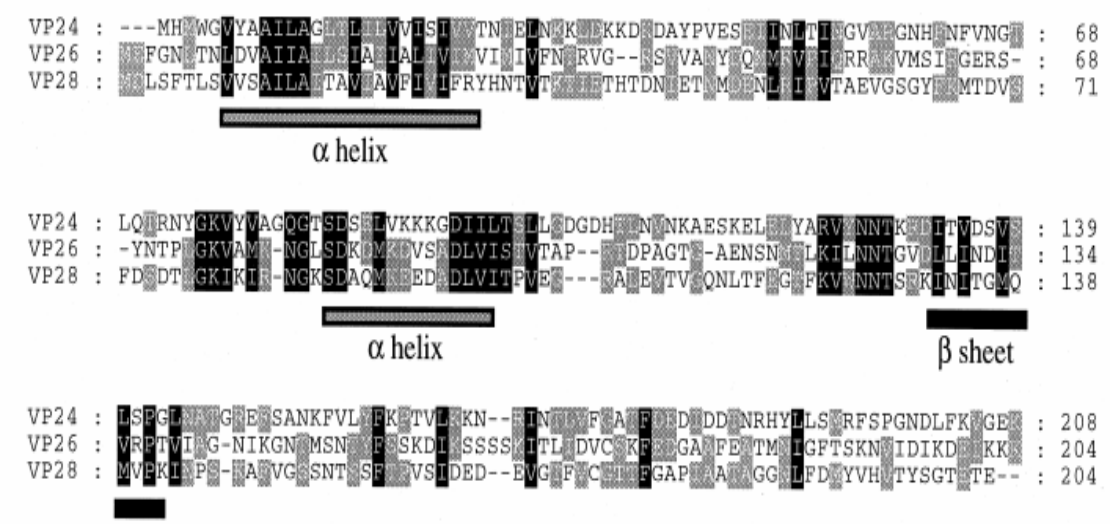


Fig.3.2 Amino acid sequence alignment of VP24, VP26and VP28 (shading is used to indicate the occurrence (black 100%, grey 67%) of similar amino acids. Conserved α-helices and β-sheets are indicated)

Two-thirds of the residues conserved in the three proteins were hydrophobic and might be involved in the folding of the proteins, giving them a similar structure. The nucleocapsid protein VP24 had a basic character with isoelectric points 8.7, whereas the envelope protein (VP28 and VP26) were more acidic. As there is a high homology at the amino acid level among the three structural WSSV proteins, and conserved domains are present, there is reason to believe that their structures are similar. The presence of the hydrophobic domain indicates that these proteins most probably are capable of forming homo and hetero multimers. Studies on the interaction of these proteins and their location in the virion are required to substantiate this hypothesis. A way to explain the high degree of amino acid similarity of the three structural WSSV proteins is to assume that these genes have evolved by gene duplication. Nucleotide comparisons supported this hypothesis, as significant homology was found. Alignment of VP24, VP28 and VP26, revealed that VP24 has 40% nucleotide identity with VP26 and 43% with VP28, whereas VP26 has 48% nucleotide identity with VP28. The data presented here strongly suggest that these three WSSV structural protein genes share a common ancestor. The most surprising observation might be that these proteins have evolved to give proteins with different functions in the WSSV virion, i.e. in the nucleocapsid and the envelope. Such a situation is unusual in animal DNA viruses, although a parallel may exist for the virion glycoproteins of alpha herpes viruses as their genes might have evolved by duplication and subsequent divergence (McGeoch *et al.,* 1990).

2.2 Studies on VP24

This study aims to select proper cloning and expression system to express VP24 and develop the proper protocol to solubilize this protein and identify the location of this protein in WSSV virions, which is very important to further elucidate the structure of this major nuceocapsid protein.

3. Material and Method

3.1 PCR Amplification of Gene VP24

VP24 DNA Sequence

atgcacatgt gggggttta cgccgctata ctggcgggtt tgacattgat actcgtggtt
atatctatag ttgtaaccaa catagaactt aacaagaaat tggacaagaa ggataaagac
gcctaccctg ttgaatctga aataataaac ttgaccatta acggtgttgc tagaggaaac
cactttaact ttgtaaacgg cacattacaa accaggaact atggaaaggt atatgtagct
ggccaaggaa cgtccgattc tgaactggta aaaaagaaag gagacataat cctcacatct
ttacttggag acggagacca cacactaaat gtaaacaaag ccgaatctaa agaattagaa
ttgtatgcaa gagtatacaa taatacaaag agggatataa cagtggactc tgtttcactg
tctccaggtc taaatgctac aggaagggaa ttttcagcta acaaatttgt attatatttc
aaaccaacag ttttgaagaa aaataggatc aacacacttg tgtttggagc aacgtttgac
gaagacatcg atgatacaaa taggcattat ctgttaagta tgcgattttc tcctggcaat
gatctgttta aggttgggga aaaataa

The 627 bp gene was PCR amplified by two primers (5'-: ***CGC GGA TCC CAC ATG TGG GGG GTT TAC*** -3' and 5'- ***CGC CCG CTC GAG TTA TTT TTC CCC AAC CTT AAA C*** -3') from the pGEX-6P-1vector. The primers used in the PCR were specially designed to incorporate *Xho*I and *Bam*HI restriction sites to the flanks of the VP24 fragment for subsequent cloning into the expression vectors.

3.2 Construction of 6P-1-VP24

627 bp DNA fragment encoding the proteins was PCRed out from WSSV genomic DNA and cloned into pGEX-6P-1 vector (N-terminal GST-tagged) with restriction sites BamH1/XhoI.

3.3 Expression System of 6P-1-VP24

3.3.1 Transformation Plasmid 6P-1-VP15 into Expression Competent Cell BL21 (DE3)

3.3.2 Determination of Target Protein Solubility

A single colony was picked up from plate to inoculate 5 ml LB medium containing 100ug/mL ampicilin (LBA) and grown overnight at both 37℃ and 30℃ with shaking. 2.5 mL overnight culture was further inoculated into 50 mL LBA and grown with vigorous shaking until the OD_{600} was about 0.5. 1 mL sample (non-induced control) was taken out, centrifuged and resuspended in 50 μL SDS-PAGE sample buffer. IPTG was added to the rest of culture to a final concentration of 0.5 mM to induce protein expresssion for further 3-3.5 hours. 1 mL sample was collected, centrifuged and resuspended in 100 μL SDS-PAGE sample buffer. The rest of culture was harvested and centrifuged at 5000 rpm for 10 minutes at 4℃. The cell pellet was resuspended in 5 mL lysis buffer and further sonicated thoroughly. The lysate was centrifuged at 16000 rpm for 20 minutes. 100 μL supernatant (soluble) was taken out and added to 100 μL 2 × SDS-PAGE sample buffers. Aliquot (insoluble protein) was resuspended in 200 μL 1× SDS-PAGE sample buffer.

All the 8 samples were boiled for 5 minutes and microcentrifuged. 10 μL of supernatant was loaded into each lane of the SDS-PAGE gel.

3.3.3 Preparation of Inclusion Bodies

Inclusion bodies were isolated from the crude cell lysate by centrifugation and washed twice with 1X IB Wash Buffer (20 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% Triton X-100) to remove loosely associated contaminants.

The induced culture was harvested by centrifugation at 5,000 rpm for 15 min at 4℃. The supernatant was removed and discarded. The cell pellet then was thoroughly resuspended in 0.1 culture volume of 1X IB Wash Buffer and mixed for full resuspension. Subsequently the suspension was cooled on ice to 4℃ to prevent heating during cell

breakage. Lysozyme was added to a final concentration of 100 mg/mL from a freshly prepared 10 mg/mL stock in water. It was then incubated at 30℃ for 15 min followed by swirling and sonicating on ice with an appropriate tip until cells were lysed and the solution was no longer viscous. Samples were kept cold to prevent proteolytic degradation of the target protein. The inclusion bodies were collected by centrifugation at 16,000 rpm for 30 min. The supernatant was removed and the pellet was thoroughly resuspended in 0.1 culture volume of 1X IB Wash Buffer, centrifuged and thoroughly resuspended in 0.1 culture volume of 1X IB Wash Buffer again. The suspension was transferred to a clean centrifuge tube with known tare weight. The inclusion bodies were collected by centrifugation at 16,000 rpm for 10 min. The supernatant was decanted and by tapping the inverted tube on a paper towel the last traces of liquid was removed. The tube was weighed and the wet weight of the inclusion bodies was obtained by subtracting the tare weight.

3.3.4 Solubilization and Refolding

From the wet weight of the inclusion bodies to be processed, the amount of 1X IB Solubilization Buffer (50 mM CAPS, pH 11.0) necessary was calculated to resuspend the inclusion bodies at a concentration of 10–20 mg/ml. At room temperature, the calculated volume of 1X IB Solubilization Buffer supplemented with 0.3% N-lauroylsarcosine was prepared. The calculated amount of 1X Solubilization Buffer/N-lauroylsarcosine from above was added to the inclusion bodies and gently mixed. Large debris can be broken up by repeated pipetting and incubated at room temperature for 15 min. The sample was clarified by centrifugation at 16,000 rpm for 30 min at room temperature. The supernatant containing the solubilized protein was transferred into a clean tube, avoiding pelleted

debris.

3.3.5 Dialysis Protocol for Protein Refolding

Dialysis was performed at 4℃ for 4 hours each with at least two buffer changes of greater than 50 times the volume of the sample (20 mM Tris-HCl, pH 8.5). The dialysis was continued through two additional changes (4 h each) with the dialysis buffer lacking DTT.

3.4 Protein Purification by Glutathione S-Transferase Affinity Chromatography

The soluble denatured protein was loaded into a 2ml GST resin column which had been equilibrated with binding buffer. After the GST-tagged protein was bound onto the resin, the resin was further washed thoroughly with 30 ml washing buffer until the absorption at 280 nm of the washing was very low. GST-tagged protein can be eluted out by a small amount of elution buffer [20mM Tris-HCl, 20mM reduced glutathione pH 8.5 (adjusted by Tris)]. For every step, 0.5 ml sample was collected and boiled with SDS loading buffer to be further checked by SDS-PAGE.

3.5 Identification by MALDI-MS (matrix assisted laser desorption ionization)

The band of the fusion protein was cut off from the gel and was cut into fine piece. In-gel digestion was performed by trypsin and the sample was analyzed by MALDI-MS and searched the result by MS-FIT.

3.6 Preparation of Antibody

1. The purified 6P-1-VP24 fusion protein was used as antigen to immunize mice by intradermal injection once every 2 weeks over an 8-week period. Antigen (5 μg) was

mixed with an equal volume of Freund's complete adjuvant (Sigma) for the first injection. Subsequent injections were conducted using 5 μg of antigen mixed with an equal volume of Freund's incomplete adjuvant (Sigma).

2. Four days after the last injection, mice were exsanguinated, and antisera collected. The titres of the antisera were 1:20000 as determined by ELISA. ELISA was performed essentially as described by Harlow & Lane (1998). The immunoglobulin (IgG) fraction was purified by protein A–Sepharose (Bio-Rad) (Sambrook *et al.*, 1989) and stored at –70 °C. The optimal dilution of purified IgG, after serial dilutions, was 1:1000, as determined by ELISA.

3.6.1 Western Blot Protocol

The sample was performed electrophoresis and proteins were transferred to Hybond ECL (Transfer Buffer: 12mM Tris-HCl, pH 8.3, 96mM glycine, 20% (v/V methanol).

At this step, non-specific binding sites were blocked by immersing the membrane in Blocking solution (Blocking solution: 2% (w/V) fatty aicd –free bovine serum albumin (BSA) in Wash solution) for 1 hour at room temperature on an orbital shaker followed by rinsing membrane for 2x2 minutes of wash buffer (Wash solution: PBS-Tween 0.01% (v/v) Tween 20 in PBS).

The primary antibody was diluted 1:2000 in blocking solution and the membrane was incubated in diluted primary antibody for 3 hours at room temperature on an orbital shaker. After rinsing the membrane with three changes of wash buffer (5 minutes for each change), the membrane was further incubated with diluted HRP labeled secondary antibody for 1 hour at room temperature on the orbital shaker. Another 3 times (5mins each) changes of fresh wash solution to membrane were carried out and finally the signal

was detected by ‘Western blotting detection reagents RPN2132 RPN 2133 kit’ (Pharmacia)

4. Result and Discussion

4.1 Construction Plasmid of 6P-1-VP15

627 bp DNA fragment (Fig.3.3) encoding the proteins was PCRed out from WSSV genomic DNA (Fig2.6) and cloned into pGEX-6P-1 vector with restriction sites *Bam*H1/*Xho*I.

4.2 6P-1-VP24 Was Expressed As Inclusion Bodies

In some cases, fusion protein solubility can be dramatically increased by lowering the growth temperature. Experiment with growth temperatures in the range of 20 to 30℃ was tried combined with altering level of induction by decreasing IPTG concentration to 0.1 mM.

At both 37℃ and 25℃, when the cells grew to the componential phase, the production of GST-tagged VP24 could be induced by adding IPTG to the concentration in the arrange of 0.1 mM to 1 mM. After the cells were lysed by sonication and centrifugation, 6P-1-VP24 was expressed as the inclusion bodies (Fig3.4).

Bacterial expression has greatly expanded the biochemical analysis of many proteins. However, the great drawback of the bacterial system has been the accumulation of many expressed proteins in a highly insoluble form (Stewart, 1990). In some case, other expression systems such as yeast and baculovirus expression systems can be alternatively used to yield soluble protein. However, for VP24 no cysteine is present in its amino acid sequence so it should not make much difference even using different system to express this high hydrophobic protein.

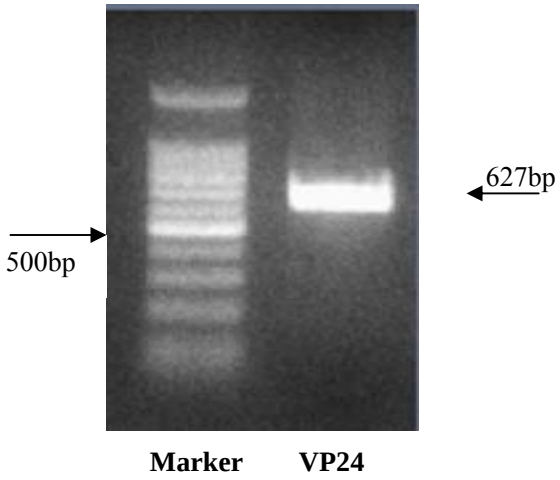


Fig 3.3 PCR product of VP24 from WSSV genome

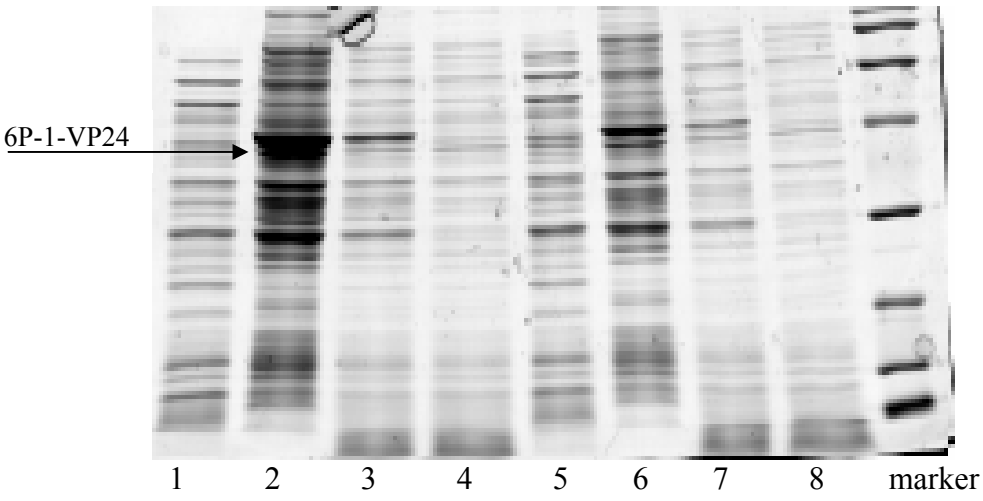


Fig.3.4 6P-1-VP24 was expressed as inclusion bodies.
Lane (1) Whole cell before induction at 37℃ (2) Whole cell after 3 hours IPTG induction at 37℃ (3) The pellet of cell lysate after induction at 37℃ (4) The supernant of the cell lysate after induction at 37℃ (5) Whole cell before induction at 25℃ (6)Whole cell after 3 hours induction at 25℃ (7) The pellet of cell lysate after induction at 25℃ (8) The supernant of the cell lysate after induction at 25℃

Over-expression of target protein in *E coli* as inclusion bodies has some advantages over cytoplasmic or periplasmic expression in both fields of research and manufacturing. Inactivation of over-expressed heterologous protein as inclusion bodies enables host cells to accumulate large amount of protein products without any cytotoxicity to the cell. The highly aggregated form of protein in inclusion bodies also protects it from proteolytical degradation. Inclusion bodies also simplify the downstream purification process since most contaminating proteins in the soluble fraction can be discarded during the pre-column steps.

4.3 Identity of the Fusion-Protein

The expressed protein was cut off, performed in-gel digestion and identified by in MALDI-MS (Fig.3.5).

Based on the search result of MS-FIT (Tab.3.1), the Mouse score is quite high for the top three searches in protein bank (all are protein VP24 in different name in different database). The result confirmed that the expressed fusion GST-fusion protein was 6P-1-VP24.

4.4 Solubility of 6P-1-VP24 in Different Denaturants

In order to obtain the best conditions to solubilize this protein, different denaturants with different concentrations were tried (Fig3.6).

Denaturants' concentration and different culture temperature (at alkaline pH) can covalently modify primary amines on the target protein. N-lauroylsarcosine (Sarkosyl) is a sodium salt of alkyl anionic detergent(Fig3.7). It can be used to solubilize most, if not all, GST fusion proteins and still maintain enzymatic activity (John V, 1992).

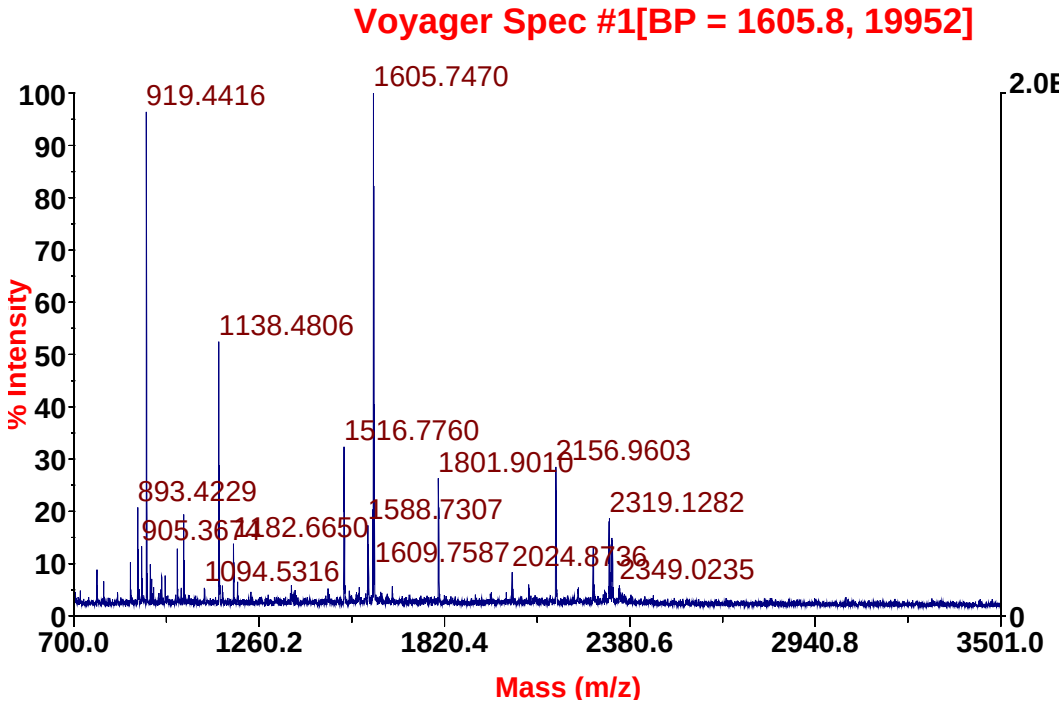


Fig.3.5 MALDI-MS spectra of 6P-1-VP24

Result Summary

MOWSE Score	# (%) Masses matched	Protein MW (Da) /pI	Species	NCBInr.030428 Accession #	Protein Name
1.57E+05	8/28 (28%)	23147.6 / 8.62	SHRIMP WHITE SPOT SYNDROME VIRUS	17158106	wsv002
1.57E+05	8/28 (28%)	23148.5 / 6.92	SHRIMP WHITE SPOT SYNDROME VIRUS	29291839	VP24
1.57E+05	8/28 (28%)	23175.6 / 8.63	SHRIMP WHITE SPOT SYNDROME VIRUS	19481650	WSSV058
1.02E+05	11/28 (39%)	29488.3 / 6.33	UNREADABLE	4389298	

Table.3.1 MS-FIT search result

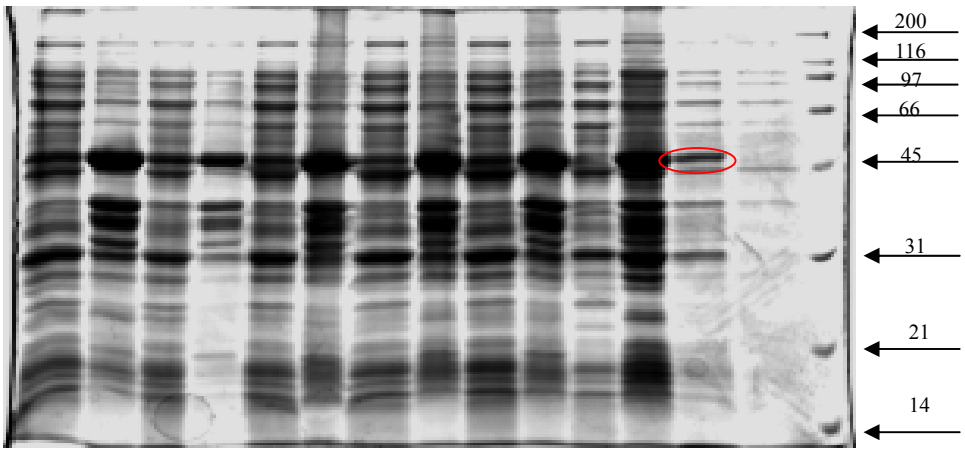


Fig 3.6a Solubility in different detergents

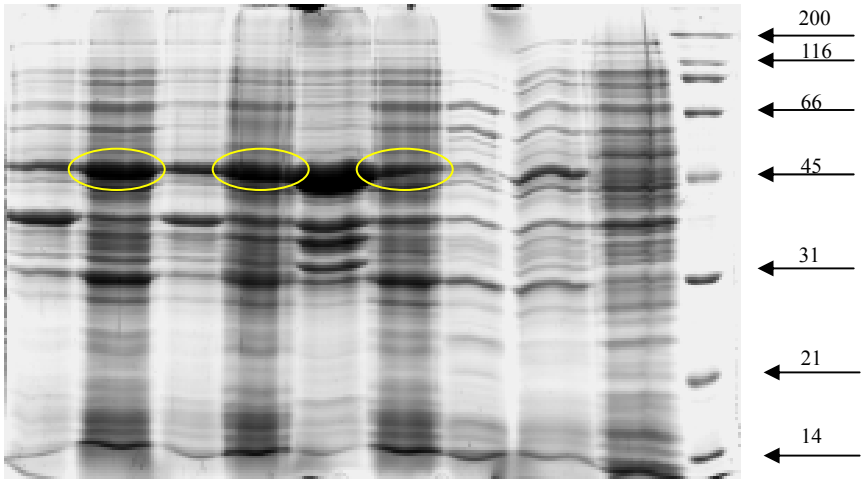


Fig3.6b Solubility ain different detergents

(P: pellet S: supernatant -: before IPTG induction; +: after induction)

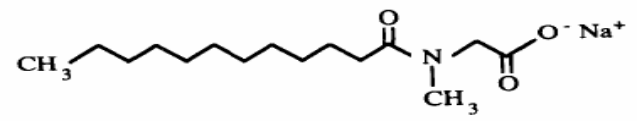


Fig3.7 Chemical structure of sakosyl

4.5 Protein Refolding

The intermolecular association of hydrophobic domain during folding is believed to play a role in the formation of inclusion bodies. For proteins with cysteine/cysteine residues, improper formation of disulphide bonds in the reducing environment of the *E. coli* cytoplasm may also contribute to incorrect folding and formation of inclusion bodies. Fortunately, there is no cysteine reside in this protein.

In order to restore the protein to its unique, functional, three-dimensional conformation from inclusion bodies, the denaturing and refolding steps would need to be optimized for each protein on a case-by-case basis. Two denaturants are usually employed to solubilize the inclusion bodies, namely guanidine hydrochloride and urea. Guanidine hydrochloride is a more powerful denaturant than urea, and its solubilizing effect on different chemical groups in proteins is different from that of urea. However, it's very complicated to check the protein status with this denaturants. N-lauroylsarcosine does not impair refolding and purification, the residual detergent level following dialysis is often 100-fold lower than the concentration in the solubilization step, typically less than 0.003%.

Many factors will affect the recovery rate of refolding such as protein concentration, temperature, cosolvents, thiol reagent, and refolding methods. Refolding at a low protein concentration will generally increase the yield of soluble protein but subsequent protein concentration will be tedious. As to the refolding methods, rapid dilution or slow dialysis can be used and they have their own advantages over each other.

4.6 Purification of the Protein by GST Affinity Column

The dialyzed protein was loaded to GST affinity column to remove the most

contaminant proteins. The result was showed in Fig.3.8.

In the process of expression, the GST alone was expressed. So the GST expressed was bound to the affinity column too and was eluted down with the GST-fusion protein. Further purification steps are needed if more purity is needed. However, since no GST existed in WSSV virus, the fusion protein with GST as the antigen to immunize mouse had few effect to the polyclonal antibody.

4.7 Production of Antibody of 6P-1-VP24

The produced 6P-1-VP15 was used as antigen to obtain the polyclonal antibody from mouse. The specificity of this antibody could be investigated from the result of western blot assay below (Fig.3.9)

The result proved that VP24 is one the of major WSSV structure proteins. The western blot assay showed that this polyclonal antibody of VP24 was quite specific and the specificity will be enough to carry out the immuno-electron Microscope studies to further confirm the location of VP 24 in WSSV virions.

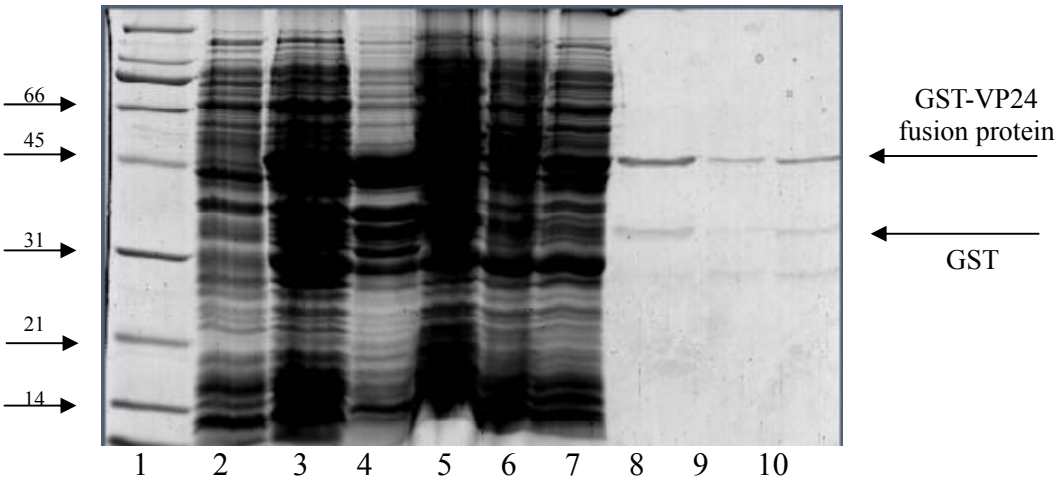


Fig.3.8 Purification of the protein by GST affinity column

Lane (1) Low molecular weight marker;(2)Total protein extract of non-induced culture (3)Total protein extract of induced culture containing GST-VP24 fusion protein; (4) Supernatant containing detergent; (5)Pellet after centrifuge (6) Flow-through fraction from Glutathione Sepharose 4B column (7)Wash fraction (8)(9)(10) elution

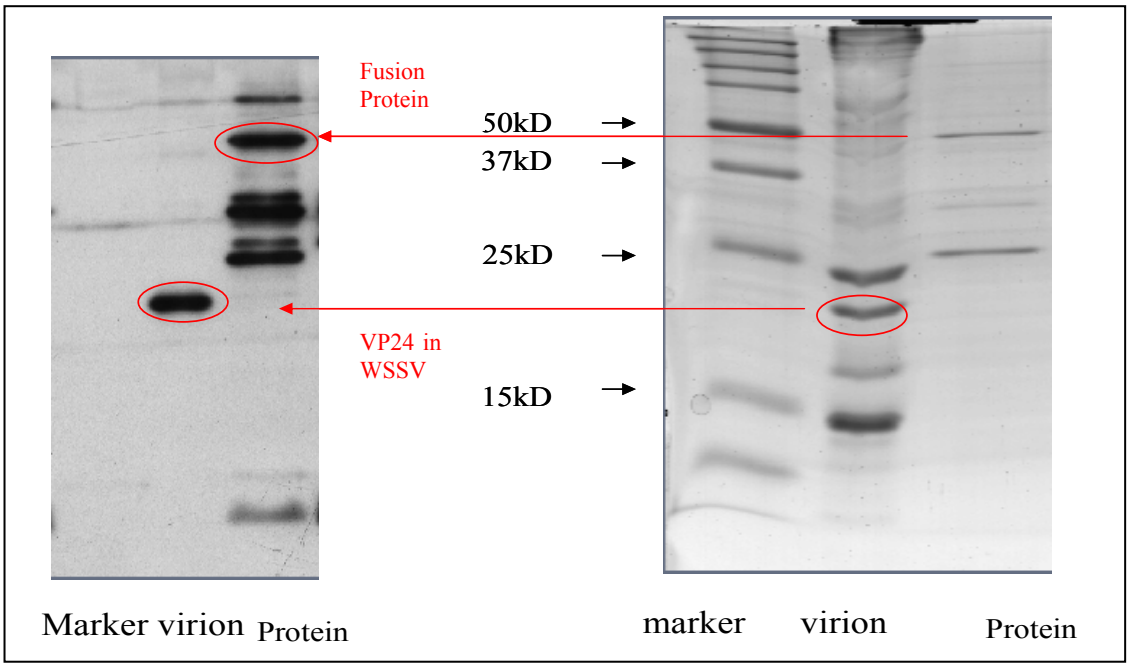


Fig.3.9 Identify the specificity of Anti-VP24 by western blots
Virion: WSSV virion ; protein: 6P-1-VP24 fusion protein

5. Future Work

1. Due to the statistically significant similarity with the sequence of two identified envelope proteins VP26 and VP28, immuno-electron microscopy is needed to carry out to identify the localization of VP24 in WSSV.

2. The GST fusion protein need to be further purified by suitable chromatography after cleavage the GST tag for structure studies.

CHAPTER 4 REFERENCES

Ashburner, M. *et al. Nature Genet.* 25, 25–29 (2000)

Bloch, F. *Phys. Rev*. 70, 460–474 (1946)

Brenner, S. E. *Trends Genet.*15, 132–133 (1999)

Burley.S.k. *et al. Nature Genetics.* 23, 151-157 (1999)

Chance M.R. *et al. Protein Science*. 11, 723-738 (2002)

Chang P.S. *et al. Identification of white spot syndrome.* (1996)

Chou H.Y. *et al. Dis.Aquaat.org.* 23, 165-173 (1995)

Devos, D. & Valencia, *Proteins.* 41, 98–107 (2000)

Durand S. *et al. Dis Aquat Org* 29, 205–211 (1997)

Feng Y. *et al. J. Virol.* 75, 11811–11821 (2001)

Garnier, J., Osguthorpe, D. J. & Robson, B. *J. Mol. Biol.* 120, 97–120 (1978)

Green E.D *Nat. Rev. Genet. 2* , 573–583 (2001)

Harlow, E. Lane, D. *Antibodies: A Laboratory Manual*. 553–612 (1988)

Hendrik, M. *et al. J. Gen. Virol.* 84, 1517–1523 (2003)

Huang, C. *et al. Virus Res.* 76,115-125 (2001)

Huang C. *et al. J. Gen. Virol*. 83, 2385-2392 (2002b)

Huang C. *et al. Mol. Cell. Proteomics* 1, 223-231 (2002)

John V, F. *et al. Analytical biochemistry* 210, 179-187 (1993)

Jensen O. N. *et al. J Virol*, 70, 7485-7497 (1996)

Joachim Frank. *et al. Annu. Rev.Biophys.Biophmol.Struct.* 31, 303-19 (2002)

Lander, E. S. *et al. Nature* 409, 860–921 (2001)

Lo C.F. *et al*. *Dis. Aquat.Org.* 27, 215-225 (1996)

Lo. C.F. *et al. Dis.Aquat.Org*. 35, 175-185(1999)

Marielle C. W. van Hulten. *et al. J. Gen. Virol.* 81, 2525–2529 (2000)

McGeoch, D. J. *et al. J. Gen. Virol.* 71, 2361–2368 (1990)

N.J.Dimmock. *et al. Introduction to modern virology,* Fifth edition (2001)

Purchell, E. M. *et al. Phys. Rev*. 69,37-38 (1946)

Pa'ez-osuna F. *et al. Environmental Management* 28, 131–140 (2001)

Sambrook, J. *et al. Molecular Cloning: A Laboratory Manual*. 2nd edition (1989)

Sattler, M. *et al. Prog. NMR Spectrosc.* 34, 93-201 (1999)

Song J. *et al. Biochem Cell Biol.* 76, 177-88 (1998)

Steven E. Brenner, *Nature Review Genetics*. 2, 801-809 (2001)

Stwart Frankel. *et al. Natl.Acad.Sci.USA*. 88, 1192-1196 (1991)

Thompson,D. *et al. Nucleic Acids Research* 22, 4673-4680 (1994)

Todd, A. E. *et al. J. Mol. Biol.* 307, 1113–1143 (2001)

Van Hulten. *et al. Virology* 266, 227–236 (2000)

Venter, J. C. *et al. Science* 291, 1304–1351 (2001)

Wang Y.G.. *et al. Dis Aquat Organ* 39,1-11 (1999)

Wongteerasupaya C. *et al. Dis Aquat Org* 21, 69–77(1995)

Wuthrich, K. *NMR of Proteins and Nucleic Acids Wiley*, New York (1986)

Yang F. *et al. J.Virol.* 75, 11811-11820 (2001)

Yorgo Modies. *et al. the EMBO Journal.* 21, 4754-4762 (2002)

Zhang X. *et al. Virus Research*. 79, 137-144 (2001)

Zhang X *et al. J. Gen.Virol.* 83, 471-477 (2002)

Zhang X., Xu X., Hew C.L. *Virus Res* 79,137-144 (2001b)

Zhang X. *et al. J Gen Virol* 83:1069-1074 (2002b)

Zhiyuan Gong and Choy L. Hew. *Biochemistry.* 33, 15149-15157 (1994)