UNIVERSITI PUTRA MALAYSIA

PRODUCTION OF THE NUCLEOCAPSID PROTEIN OF A SWINE NIPAH VIRUS ISOLATE IN ESCHERICHIA COLI

ONG SWEE TIN

FSAS 2002 57

PRODUCTION OF THE NUCLEOCAPSID PROTEIN OF A SWINE NIPAH VIRUS ISOLATE IN *ESCHERICHIA COLI*

By

ONG SWEE TIN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Master of Science

December 2002

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Master of Science

PRODUCTION OF THE NUCLEOCAPSID PROTEIN OF A SWINE NIPAH VIRUS ISOLATE IN *ESCHERICHIA COLI*

By

ONG SWEE TIN

December 2002

Chairman: **Associate Professor Dr. Tan Wen Siang**

Faculty: **Science and Environmental Studies**

Nipah virus (NiV) possesses a nonsegmented, single-stranded negative sense RNA genome that contains six structural genes arranged in the order of 3' N-P-M-F-G-L 5'. The nucleocapsid (N) gene of Nipah virus isolated from swine was amplified from the viral RNA by reverse transcription polymerase chain reaction (RT-PCR). The nucleocapsid (N) gene of Nipah virus was cloned into the bacterial expression vector, pTrcHis2, for intracellular expression in three *Escherichia coli* strains: TOP 10, BL 21 and SG 935. The N protein was expressed as a 63 kDa fusion protein containing the *myc* epitope and His-tag at its C-terminal end. The amount of the fusion protein expressed in strain SG 935 was significantly higher than the other two strains, and was detected by the anti-*myc* antibody, anti-His and swine anti-NiV serum. The N gene sequence shared 99% homology with that of Nipah virus isolated from human. The coding region of N protein of NiV was cloned into different vectors and subsequently introduced into different *E. coli* strains. The yield of the N protein produced in different vectors and different hosts was compared. It was found that the amount of N protein expressed by the pTrcHis2 vector containing the *trc* promoter in *E. coli* SG 935 was four-fold higher than that of vector pRSETB and pGEX-4T-1 in *E. coli* strain BL 21 series. Deletion of the N- or/and C-terminal

region of the N protein revealed that the N-terminal region plays a role in N protein solubility, but a mutant ($MN50_{fus}$) containing the second half of the N protein showed the highest expression level in all the three *E. coli* strains. Lowering the growth temperature of *E. coli* cell cultures to 25°C improved the solubility of the full-length and truncated N_{fus} protein from 50% to 80%. This study addresses the fundamental problems encountered in production of Nipah viral N protein in *E. coli* which may be useful as an alternative antigen for the detection of anti-NiV in swine.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Master Sains

PENGHASILAN PROTEIN NUKLEOKAPSID DARIPADA VIRUS NIPAH ISOLAT KHINZIR DALAM *ESCHERICHIA COLI*

Oleh

ONG SWEE TIN

Disember 2002

Pengerusi: **Prof. Madya Dr. Tan Wen Siang**

Fakulti: **Sains dan Pengajian Alam Sekitar**

Virus Nipah (NiV) mempunyai seutas genom RNA negatif yang tidak berbahagi. Genom ini mengandungi enam gen struktur yang tersusun dalam turutan 3' N-P-M-F-G-L 5'. Gen nukleokapsid (N) daripada virus Nipah yang dipencilkan daripada khinzir telah digandakan daripada RNA virus dengan teknik transkripsi berbalik tindak balas rantaian polymerase (reverse transcription polymerase chain reaction). Gen nukleokapsid (N) telah diklonkan ke dalam vektor bakteria, pTrcHis2, untuk pengekpresan intrasel dalam tiga strain *Escherichia coli*: TOP 10, BL 21 dan SG 935. Protein N telah diekspres sebagai protein gabungan yang bersaiz 63 kDa dan mengandungi epitope *myc* dan His-tag di hujung C. Jumlah protein gabungan yang diekspres dalam strain SG 935 adalah lebih tinggi daripada dua strain yang lain, dan boleh dikesan dengan antibodi anti-*myc* antibody, anti-His dan serum khinzir anti-NiV. Turutan gen N berkongsi 99% kesamaan dengan virus Nipah yang dipencilkan daripada manusia. Bahagian yang mengekodkan protein N telah diklon ke dalam vektor berlainan and dimasukkan ke dalam strain *E. coli* yang berlainan. Perbandingan jumlah protein N yang terhasil dengan vector dan host yang berlainan telah dijalankan. Jumlah protein N yang diekspres oleh vektor pTrcHis2 yang mengandungi promoter *trc* dalam *E. coli* SG 935 didapati empat kali ganda lebih

tinggi daripada vektor pRSETB dan pGEX-4T-1 dalam siri strain *E. coli* BL 21. Pemotongan bahagian termina N- atau/dan C- protein N menunjukkan bahawa bahagian hujung N memainkan peranan dalam keterlarutan protein N, tetapi mutan ($MN50_{fus}$) yang mengandungi bahagian kedua protein N menunjukkan tahap ekspresi tertinggi di antara tiga strain *E. coli*. Apabila suhu pertumbuhan sel *E. coli* diturunkan kepada 25°C, keterlarutan protein N_{fus} berpanjangan penuh atau terpotong bertambah 50% hingga 80%. Kajian in membincangkan masalah asas yang ditemui dalam penghasilan protein N virus Nipah dalam *E. coli* yang mungkin berguna sebagai antigen gantian untuk mengesan anti-NiV dalam khinzir.

ACKNOWLEDGEMENTS

I would like to sincerely acknowledge the chairman of the supervisory committee, **Assoc. Prof. Dr. Tan Wen Siang** for his guidance, suggestions, never-ending patience, his kindness and his willingness to help throughout the course of this project. Without his valuable advice and support, it would be impossible for me to complete my Master of Science project. I have indeed gained so much from him throughout the study.

My sincere appreciation also goes to the members of the supervisory committee, **Prof. Datin Dr. Khatijah Yusoff** and **Dr. Sharifah Syed Hassan** for providing prompt and helpful advice when required.

I also deeply appreciated for all the invaluable help and guidance from all my beloved friends, especially Tam Yew Joon, Kho Chiew Ling, Dr. Majid Eshagi, Wong Sing King, Ho Kok Lian and Kok Wai Ling for making my time in the laboratory joyable and pleasant. I thank them for their kind help, thoughtful comments and friendship. Many thanks are also conveyed to all the members in the Virology Laboratory.

Of course, my deepest appreciation will always go to my family for their love, patience, support and sacrifices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A_{540}	absorbance at wavelength 540 nm
A_{600}	absorbance at wavelength 600 nm
ATP	adenosine triphosphate
α	alpha
aa	amino acid
Amp	ampicillin
ATG	start codon
β	beta
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pair
BSA	bovine serum albumin
°C	degree Celsius
cDNA	complimentary DNA
C-terminus	carboxy terminus
DEPC	diethyl pyrocarbonate
dH_2O	distilled water
DNA	deoxyribonucleic acid
DTT	1,4-dithiothreitol
dNTPs	deoxynucleoside triphosphate
EDTA	ethylenediamine tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
h	hour
HCl	hydrochloric acid

CHAPTER 1

INTRODUCTION

In March 1999, an outbreak of fatal viral encephalitis was reported in Malaysia and Singapore (Harcourt *et al.*, 2000). The causative agent for this disease, Nipah virus (NiV) was isolated and named after the place where the outbreak first occurred. The virus causes respiratory and neurological syndromes and deaths in swine. The outbreak which claimed 105 human lives were controlled only after more than 1 million swine in the affected areas were culled (Chua *et al.*, 2000) which nearly abolished the swine industry in Malaysia.

Electron microscopic analysis of the NiV isolated from a human tissue revealed that it has structural features that similar to those of the family *Paramyxoviridae* (Chua *et al.*, 2000). In terms of nucleotide sequence, NiV and Hendra virus (HV) are closely related to each other but they could not be grouped within any of the established genera of the subfamily *Paramyxovirinae* and thus these viruses were suggested to be grouped as members of a new genus within the subfamily (Harcourt *et al.*, 2000).

NiV has a nonsegmented, single-stranded negative sense RNA genome of approximately 18.2 kb (Wang *et al.*, 2001). The virion is enveloped by a lipid bilayer containing two transmembrane glycoproteins: a cell receptor binding protein known as G, and a separate fusion (F) protein (Chua *et al.*, 2000). Lining the interior surface of the viral envelope is the matrix (M) protein, which is believed to link the

glycoproteins to the viral nucleocapsid or ribonucleoprotein (RNP) complex. The latter is made up of many copies of nucleoprotein (N) that encapsidates the RNA genome. Associated with the RNP are the phosphoprotein (P) and the large protein or RNA polymerase (L). These three components are believed to be involved in replication as well as transcription of viral RNA (vRNA) (Lamb & Kolakofsky, 1996). The N protein is the most abundant structural protein and the essential component of the helical nucleocapsid (Yu *et al.*, 1998).

NiV is classified internationally as a biosecurity level 4 (BSL4) agent (Daniels *et al.*, 2001) which is defined as a biological agent that causes severe human disease and is a serious hazard to employees; it is likely to spread to the community and there is usually no effective prophylaxis or treatment available. Hence propagation of the NiV can only be conducted under physical containment level 4 (PC4) conditions. The strict application of this guideline has made the propagation of the NiV difficult in Malaysia. Therefore, it is of important to produce the viral antigens via recombinant DNA technology.

The *Escherichia coli* system was chosen for expressing the recombinant N protein because *E. coli* has rapid doubling time and the ability to grow in inexpensive media. Furthermore, *E. coli* is one of the best understood organisms that has been applied in the development of many molecular cloning techniques. Thus, the objectives of this study were as follows:

(I) to determine the nucleotide sequence of the nucleocapsid gene (N) of the Nipah virus (NiV) swine isolate VRI 2794/99;

(II) to clone the N gene into an *E. coli* vector;

(III) to produce the nucleocapsid protein in *E. coli* and;

(IV) to study the solubility of the N protein in *E. coli.*

CHAPTER 2

LITERATURE REVIEW

2.1 General introduction of Nipah virus

Nipah virus (NiV) is a newly recognized zoonotic virus. The virus was first isolated in 1999 during an outbreak of viral encephalitis in several parts of Malaysia. It has caused disease in animals and humans, through direct contact with infectious animals. The NiV was named after a village, Sungai Nipah, where it was first isolated (Chua *et al.*, 2000). NiV is closely related to another newly emergent virus, the Hendra virus (HeV; Yu *et al.*, 1998). Both of these viruses are members of the *Paramyxoviridae* family (Wang *et al.*, 2001).

2.2 Occurrence

The outbreaks of encaphelitis in pigs and humans in several states of Peninsular Malaysia (Perak, Negeri Sembilan, and Selangor), which were reported between September 1998 and April 1999, resulted in the death of 105 humans (Nor *et al.*, 1999; Chua *et al.*, 2000) and in culling of about 1.1 million pigs (Eaton, 2001; Field *et al.*, 2001). During March 1999, 11 abattoir workers in Singapore developed a febrile illness (1 fatality) caused by NiV, following close contact with pigs imported from the outbreak areas in Malaysia (Paton *et al.*, 1999).

2.3 Nature of the virus

2.3.1 Virus classification

NiV is a novel, unclassified member of the family *Paramyxoviridae*. Members of *Paramyxoviridae* possess intermediate genome sizes from 15.1 to 15.9 kb, and have traditionally been described as having a 'uniform genome size' (Pringle, 1991; Rima *et al.*, 1995). The universality of this family feature is now challenged with the discovery of the larger genomes found in NiV and HeV (Harcourt *et al.*, 2000; Wang *et al.*, 2001). HeV and NiV have a genome size of 18.2 kb, which is much larger than the relatively uniform genome sizes of all other known members of the *Paramyxovirinae*. When more sequence data became available, detailed phylogenetic analyses of the HeV and NiV genes indicated that the overall sequence homology of these two viruses with other members of the family was limited (Wang *et al.*, 2001). Indeed the homology detected was lower than that observed among most existing members of any single genus (Gould, 1996; Wang *et al.*, 1998; Yu *et al.*, 1998; Harcourt *et al.*, 2000; Wang *et al.*, 2000;).

Phylogenetic analysis of the nucleotide sequence of the N, P, M, F, G and L open reading frames (ORFs) from representatives of the subfamily *Paramyxovirinae* showed that NiV consistently formed a unique cluster which was more closely related to the morbiliviruses and the respiroviruses than to the rubulaviruses (Wang *et al.*, 2001). These observations suggested that NiV and HeV should be considered as members of a new genus Henipahvirus (Hendra + Nipah) within the *Paramyxoviridae* (Wang *et al.*, 2000; Chan *et al.*, 2001).

2.3.2 Virus morphology

NiV is pleomorphic and enveloped virus with a long helical nucleocapsid. The virus particles vary in size from 120 to 500 nm (Chua *et al.*, 2000). A schematic diagram of the NiV virion is shown in Figure 2.1. There are glycoprotein spikes with the length of approximately 8-12 nm protruding out from the surface of the enveloped virus and these spikes can be readily visualised by electron microscopy (Lamb & Kolakofsky, 1996). The envelope contains two transmembrane glycoproteins, a cell receptor binding protein (G) and a separate fusion (F) protein. Typical "herringbone" nucleocapsid structures, approximately 1.67 ± 0.07 μm in length and with an average diameter of 21 nm were observed in infected cells by means of negative stain preparations (Chua *et al.*, 2000).

2.3.3 Virion and genome structure

The complete genome sequence of NiV has been determined (Harcourt *et al.*, 2000). It comprises a negative single stranded, nonsegmented RNA of approximately 18.2 kb in length. As with other *Paramyxovirinae*, there are six major transcription units (Figure 2.2) in the order 3'-N-P-M-F-G-L-5' (Wang *et al.*, 2001) encoding six major structural proteins; nucleocapsid protein (N), phosphoprotein (P), matrix protein (M), fusion protein (F) and glycoprotein (G) or attachment protein, and large protein (L) or RNA polymerase.

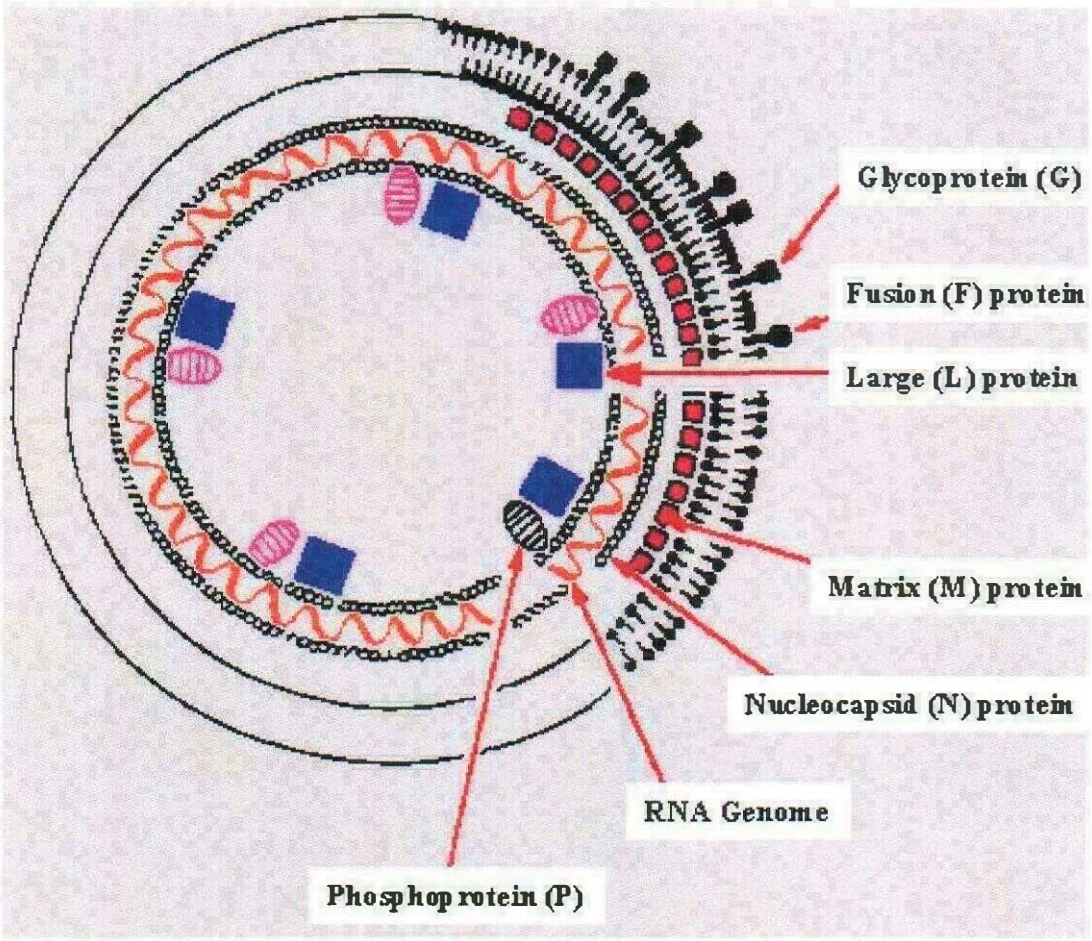

Figure 2.1: A schematic diagram of a Nipah virus (modified from Lamb and Kolakofsky, 1996).

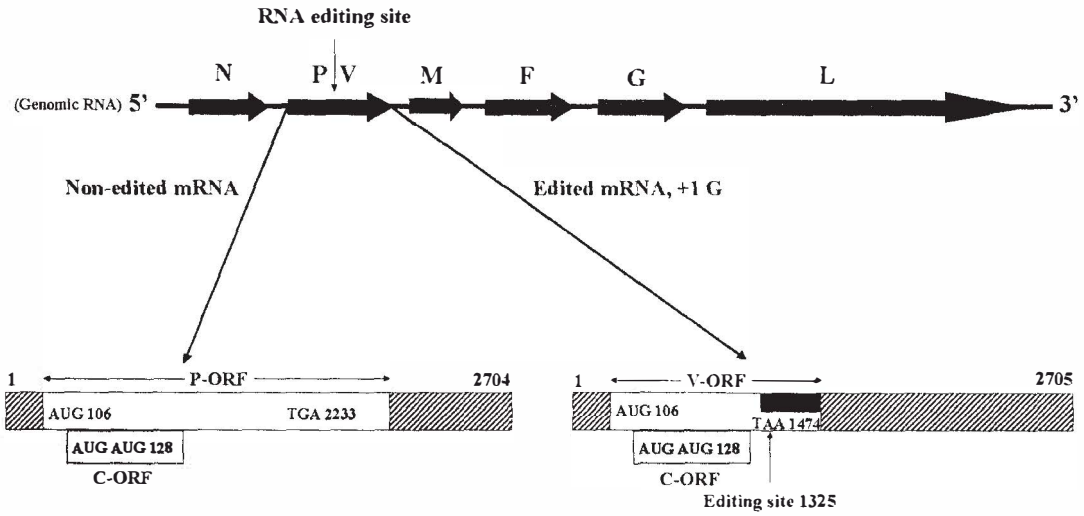

Figure 2.2: A schematic representative of the genome structure of NiV and the coding capacity of the P gene. The upper part represents the NiV genome (18.2 kb) and the location and order of the six major structural protein genes encoded by the anti-genome from 5' (left) to 3' end (right). The lower part contains the enlarged diagram showing the P gene coding strategy. The predicted P mRNA is 2704 nucleotides in length. The P protein is encoded by a faithful transcript of the viral genomic RNA from an ORF beginning at nucleotide 106 of the mRNA. A process known as pseudotemplated addition of G nucleotides or RNA editing generates the mRNA coding for the V protein. The RNA editing sites (vertical arrows) in the P genes of the paramyxoviruses are highly conserved. The addition of a G nucleotide at the RNA editing site (nucleotide 1325) allows access to a different reading frame (-1 relative to P). Therefore, the V protein contains the amino-terminal domain of the P protein joined to the cysteine-rich domain that is unique to the V protein (black box). The C protein (grey box) is expressed from an ORF that begins at nucleotide 128 (or 131) and overlaps P in the +1 frame. Hatched boxes indicate nontranslated regions of the mRNA (modified from Chua *et al.*, 2000; Wang *et al.*, 2001).