UNIVERSITI PUTRA MALAYSIA

SEQUENCE ANALYSIS OF THE L GENE OF NEWCASTLE DISEASE VIRUS STRAIN AF2240

ENI KUSUMANINGTYAS

FSAS 2003 26

SEQUENCE ANALYSIS OF THE L GENE OF NEWCASTLE DISEASE VIRUS STRAIN AF2240

ENI KUSUMANINGTYAS

MASTER OF SCIENCE
UNIVERSITI PUTRA MALAYSIA
2003

SEQUENCE ANALYSIS OF THE L GENE OF NEWCASTLE DISEASE VIRUS STRAIN AF2240

By

ENI KUSUMANINGTYAS

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

July 2003

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

SEQUENCE ANALYSIS OF THE L GENE OF NEWCASTLE DISEASE VIRUS STRAIN AF2240

By

ENI KUSUMANINGTYAS

July 2003

Chairperson: Professor Datin Khatijah Mohd. Yusoff, Ph.D.

Faculty: Science and Environmental Studies

Newcastle disease is an avian disease which causes a devastating effect in commercial poultry production. The causative agent for this fatal disease is Newcastle disease virus (NDV) which contains six structural proteins; nucleocapsid (NP), phospho (P), matrix (M), fusion (F), haemagglutinin-neuraminidase (HN) and large (L) proteins. The L protein of NDV is important not only as a multifunctional enzyme but also as an RNA-dependent RNA polymerase. This protein may carry out all of the enzymatic steps in transcription such as initiation, elongation and cotranscriptional modification of RNAs including capping, methylation and polyadenylation. Therefore, this study was carried out to determine the L gene sequence and its predicted translated product, and to analyse the primary and secondary structures of the L protein of the local NDV strain AF2240. In order to identify the conserved functional domains and motifs within the L protein, the amino acid composition of the L protein of strain AF2240 was compared with those of other NDV strains.

The L gene was divided into 10 fragments which were then amplified by RT-PCR, cloned into pGEM T-Easy vector and transformed into *Escherichia coli* strain TOP 10. Sequencing was done in both directions (forward and reverse) in order to confirm the correct sequence which was then analysed using the Expasy and Workbench tools analysis.

The coding sequence of the L gene of NDV strain AF2240 contains 6615 nucleotides (from the start codon ATG to stop codon TAA) with a single large open reading frame (ORF) that encodes 2204 amino acids with estimated molecular weight of 249 kDa. The L protein contains six conserved domains which were proposed to play an important role in the transcription and replication processes. Comparison with other NDV strains showed that they could be divided into two groups based on the deletion and insertion located at amino acids 1287 to 1316. The region containing this compensatory frameshift mutation in strain AF2240 shares the same amino acid sequence with strains B1 Takaaki, Clone 30 and F48E9 (group A). Strains B1, LaSota, Beaudette C and ZJ1 contain a different set of amino acid sequence within this particular region (group B).

The above compensatory mutation has changed the predicted hydrophobicity and charge of the protein. Hydropathy profile between amino acids 1290 to 1300 showed that group A contains hydrophobic amino acids while group B contains hydrophilic amino acids. This frameshift region does not correlate with viral pathogenicity and it takes place in Domain V of L protein which is proposed to play a role in replication. Since Domain V is also involved in protein folding, it is suggested that this mutation may affect the structure and function of the L protein.

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

ANALISA JUJUKAN GEN L VIRUS PENYAKIT NEWCASTLE STRAIN AF2240

Oleh

ENI KUSUMANINGTYAS

Julai 2003

Pengerusi: Profesor Datin Khatijah Mohd. Yusoff, Ph.D.

Fakulti: Sains dan Pengajian Alam Sekitar

Penyakit sampar ayam telah mengakibatkan kerugian dalam sektor peternakan. Agen yang mengakibatkan jangkitan ini ialah virus penyakit Newcastle (NDV) yang terdiri daripada 6 protein struktur iaitu: nukleokapsid (NP), fosfo (P) matrik (M), fusion (F), hemaglutinin-neuraminidase (HN) dan besar (L). Protein L bukan sahaja berperanan sebagai enzim yang mempunyai pelbagai fungsi malah ia juga adalah "*RNA-dependent RNA polymerase*". Protein ini menjalankan semua tindakan enzimatik dalam transkripsi seperti permulaan, pemanjangan dan modifikasi RNA seperti penudungan, pemetilan dan poliadenilasi. Oleh itu, projek ini adalah untuk menentukan jujukan gen L dan produk yang dijangka serta untuk menganalisis struktur primer dan sekunder protein L NDV tempatan strain AF2240. Untuk mengenalpasti fungsi domain dan motif terpelihara dalam protein L, perbandingan asid amino protein L strain AF2240 dengan strain NDV yang lain telah dijalankan.

Gen L telah dibahagikan kepada 10 serpihan dan diamplifikasikan dengan RT-PCR, diklonkan dalam vector pGEM T-Easy dan ditransformasi ke dalam *Escherichia coli* strain TOP 10. Penjujukan dibuat dari kedua-dua arah untuk

memastikan jujukan yang betul dan jujukan ini dianalisa dengan menggunakan program Expasy dan Biology Workbench.

Jujukan pengkodan gen L strain NDV AF2240 mengandungi 6615 nukleotida (dari kodon pemula ATG hingga kodon penamat TAA) yang mengkodkan 2204 asid amino dengan anggaran berat molekul 249 kDa. Protein L mengandungi 6 domain terpelihara yang berperanan penting dalam proses transkripsi dan replikasi. Perbandingan dengan strain yang lain menunjukkan protein L ini boleh dibahagikan kepada dua kumpulan berdasarkan kepada pemotongan dan penyelitan di kawasan asid amino 1287 hingga 1316. Kawasan yang mengandungi mutasi rangka bacaan pada strain AF2240 mempunyai jujukan asid amino yang sama dengan strain B1 Takaaki, Klon 30 dan F48E9 (kumpulan A). Strain B1, LaSota, Beaudette C dan ZJ1 mengandungi jujukan asid amino yang berbeza pada kawasan khas ini.

Mutasi kompensatori tersebut telah menukar kehidrofobian dan cas protein. Profil hidropati di antara asid amino 1290 hingga 1300 menunjukkan bahawa kumpulan A mengandungi asid amino hidrofobik manakala kumpulan B mengandungi asid amino hidrofilik. Mutasi rangka bacaan tidak berhubung kait dengan kepatogenan virus dan ini berlaku pada Domain V protein L yang memainkan peranan penting dalam replikasi. Oleh sebab Domain V juga terlibat dalam lipatan protein, maka adalah bahawa dicadangkan mutasi ini boleh mempengaruhi struktur dan fungsi protein L.

ACKNOWLEDGEMENTS

In the name of Allah, the most Gracious and the most Merciful.

First and foremost, my sincere gratitude and appreciation are to my supervisor Prof. Datin Dr. Khatijah Mohd Yusoff for her generous supervision, guidance and encouragement throughout this work, her valuable advice and constructive criticism from the beginning of this research till the final review of the thesis. My sincere gratitude also goes to Assoc. Prof. Dr. Tan Wen Siang for his guidance, discussion, support and advice and Assoc. Prof. Dr. Zulkeflie Zamrod for his assistance, discussion, suggestion and allowing me to use the facility in the Centre for Gene Analysis and Technology (CGAT), Universiti Kebangsaan Malaysia (UKM).

I would also like to thank the Agency for Agricultural Research and Development, (AARD) Jakarta for awarding me the scholarship which made this study possible, and the Research Institute for Veterinary Science (RIVS), Bogor, Indonesia for allowing me to continue my study at the Universiti Putra Malaysia.

I am greatly indebted to Dr. Majid Eshagi for valuable discussion, suggestions and comments on my experiments. My akcnowledgement also goes to the students in the Virology Laboratories, Dr. Priadarishni Ramanujam and Tang Yik Kiong who helped me to get NDV culture and eggs, Kho Chiew Ling, Ong Swee Tin, Amir Rabu, Dr Azri Azahar, and Nazrien Kaman for their discussion. I would also like to thank Badrul Hisyam from the Universiti Kebangsaan Malaysia and Heryanti Tahir from the Centre for Gene Analysis and Technology (CGAT),

Universiti Kebangsaan Malaysia who taught me automated sequencing; Lau Shie Hie, Bimo Ario Tejo, Gama Galundra, Yadi Suryadi and Prihasto Setyanto for their valuable discussion and shared experience about research in Malaysia. For lab members 143 and 132; Wong Sing King, Ho Kok Lian, Kok Wai Ling, Rafidah Saadun, Noor Suhana, Shaherny Zaid, Tan Geok Hun, Lalita Ambigai, Lee Tong Chuan and Chia Suit Leen, Raha Raus and Ferozheh, your friendship and support made the two years stay in Malaysia an enjoyable one.

I also wish to express my gratitude to Dr. Djoko Budiyanto (Director of AARD, Jakarta), Dr. Muhamad Djoko Said Damardjati (Director of the Committee of Human Resources), Dr. Haryono (former Director of PAATP), Dr. Mahendro (present Director of PAATP), Hartiningsih, MSc and staff (from PAATP Training Division), Dr Syamsul Bahri (former Director of Research Institute for Veterinary Science), Dr. Darminto (Present Director of Research Institute for Veterinary Science), Istiana, MSc and Indraningsih, MSc for giving support to continue my studies at Universiti Putra Malaysia.

I wish to specially thank my parents, brother and sister for their constant love, encouragement and support as well as Didik, Daryani, Emi, Suci, Edi, Taufiq, Erna, Joko, and Jarot for their support. Finally, I would like to extend my deep gratitude to many people, whose names may not be mentioned above.

This study was supported by PAATP Project ADB Loan No 1526 - INO and IRPA grants 01 - 02 - 04 - 0107 and 01 - 02 - 04 - 003 BTK/ER/006.

I certify that an Examination Committee on **21 July 2003** to conduct the final examination of **Eni Kusumaningtyas** on her **Master of Science** thesis entitled **"Sequence Analysis of the L Gene of Newcastle Disease Virus Strain AF2240"** in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follow:

Raja Noor Zaliha Raja Abd. Rahman, Ph.D.
Associate Professor
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
(Chairman)

Datin Khatijah Mohd. Yusoff, Ph.D.
Professor
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
(Member)

Tan Wen Siang, Ph.D.
Associate Professor
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
(Member)

Zulkeflie Zamrod, Ph.D.
Associate Professor
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
(Member)

GULAM RUSUL RAHMAT ALI, Ph.D.
Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 30 SEP 2003

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

Datin Khatijah Mohd Yusoff, Ph.D.
Professor
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
(Chairperson)

Tan Wen Siang, Ph.D.
Associate Professor
Faculty of Science and Environmental Studies
Universiti Putra Malaysia
(Member)

Zulkeflie Zamrod, Ph.D.
Associate Professor
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
(Member)

AINI IDERIS, Ph.D.
Professor/ Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 14 NOV 2003

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

ENI KUSUMANINGTYAS

Date: 26 SEP 2003

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Ala (A)	-	alanine
Arg (R)	-	arginine
Asn (N)	-	asparagine
Asp (D)	-	aspartic acid
bp	-	basepair
Cys (C)	-	cysteine
DNA	-	deoxyribonucleic acid
dNTP	-	dideoxynucleotide triphosphat
Gln (Q)	-	glutamine
Glu (E)	-	glutamic acid
Gly(G)	-	glycine
h	-	hour
HA	-	haemagglutinin
His (H)	-	histidine
HN	-	haemagglutinin neuraminidase
Ile (I)	-	isoleucine
kb	-	kilobase
kDa	-	kilodalton
Leu (L)	-	leucine
Lys (K)	-	lysine
M	-	molar
Met (M)	-	methionine
min	-	minute

ml	-	millilitre
mM	-	millimolar
ND	-	Newcastle disease
NDV	-	Newcastle disease virus
ng	-	nanogram
ORF	-	open reading frame
PCR	-	Polymerase Chain Reaction
pH	-	*Puissance hydrogene*
Phe (F)	-	phenylalanine
pmol	-	picomol
Pro (P)	-	proline
RBCs	-	red blood cells
RNA	-	ribonucleic acid
RT-PCR	-	Reverse Transcription Polymerase Chain Reaction
s	-	second
Ser (S)	-	serine
Tm	-	melting temperature
Taq	-	*Thermus aquaticus*
Thr (T)	-	threonine
Trp (W)	-	tryptophan
Tyr (Y)	-	tyrosine
µl	-	microlitre
v	-	volume
V	-	volt
Val (V)	-	valine

CHAPTER I

INTRODUCTION

Newcastle disease is one of the most important viral diseases of poultry in the world. It occurs in many countries and has a devastating effect on commercial poultry production (Spradbrow, 1999). The causative agent for this fatal disease is the Newcastle disease virus (NDV). This virus is a member of the genus *Rubulavirus* of the family *Paramyxoviridae* and it contains a single-stranded negative-sense non-segmented RNA genome (Murphy *et al.*, 1995). The genomic RNA encodes 6 proteins: nucleocapsid (NP), phospho (P), matrix (M), fusion (F), haemagglutinin-neuraminidase (HN) and large (L) proteins (Collins *et al.*, 1982; Steward *et al.*, 1993).

The L protein is important not only as a multifunctional enzyme but it also works as an RNA-dependent RNA polymerase. This protein might carry out all of the enzymatic steps in transcription such as initiation, elongation and cotranscriptional modification of RNAs including capping, methylation and polyadenylation (Schubert *et al.*, 1984; Banerjee, 1987; Yunus *et al.*, 1998). The L protein also plays an important role in polymerase complex formation with P protein (Horikami *et al.*, 1992) because it contains multiple domains that have been proposed to encode essential activities such as catalyzing RNA synthesis (Holmes and Moyer, 2002). Hamaguchi *et al.* (1983) demonstrated that when both L and P proteins were added to the genome template, the RNA synthetic activity was greatly stimulated while the P protein alone did not enhance but rather suppressed the activity.

Based on this information, it is important to understand the structure and functional role of the L protein. The local isolate velogenic viscerotropic NDV strain AF2240 was chosen since it is one of the most virulent NDV strains and often caused 100% mortality in susceptible chicken flocks (Lai, 1985).

In order to understand the structure and function of the L protein, the nucleotide sequence of the L gene of NDV strain AF2240 was sequenced and the primary and secondary structures of its protein were studied. Analysis of the primary structure allows prediction of secondary structure such as alpha helices, beta strands, beta turns and random coils, based on the characteristic of amino acids. On the other hand, comparison of the L protein of NDV AF2240 with other strains is also important to allow the prediction of the functional role of the L protein such as polymerase complex formation and polymerase activity as well as in transcription and replication of viral RNA.

However, the complete genome sequence of NDV strain AF2240 has not yet been determined. Only the HN (Tan *et al.*,1995), M (Jemain, 1999), F (Salih, 1999), NP (Kho *et al.*, 2001) and P (Kho *et al.*, 2002) genes of NDV strain AF2240 have been determined. Nevertheless, the nucleotide sequences of the L genes have been obtained from different strains of NDV; Beaudette C (Yusoff *et al.*, 1987), LaSota (de-Leew and Peeters, 1999), B1 (Sellers and Seal, 2000), B1 Takaaki (Nakaya *et al.*, 2001), clone 30 (Romer-Oberdorfer *et al.*, 1999), F48E9 (Chao *et al.*, 2001) and ZJ1 (Huang *et al.*, 2001).

This project was thus carried out with the following objectives:

1. to determine the nucleotide sequence of the L gene of NDV strain AF2240,
2. to predict and to analyse the protein translated from the gene and to analyse the primary and secondary structures of the predicted L protein; and
3. to compare the amino acid sequence of the L protein of NDV strain AF2240 with those of other strains such as Beaudette C, LaSota, B1, B1 Takaaki, clone 30, F48E9 and ZJ1.

CHAPTER II

LITERATURE REVIEW

2.1 Newcastle Disease

Newcastle disease (ND) is a respiratory disease in poultry with worldwide distribution causing a highly contagious and fatal disease with morbidity and mortality up to 100%. ND was first discovered by Kranevelt in Indonesia in 1926 and spread rapidly and widely after its first discovery. It was rapidly recognized in other parts of Asia (Korea, India and the Philipines) and in Newcastle-Upon-Tyne, England (Spradbrow, 1999).

The disease is indigenous to Asia, but endemic in parts of Africa, Europe and South America (Norton, 1994). The disease spread easily through contaminated food and water, direct contact and human as a vehicle for spreading. Based on clinical signs in chickens, Allan *et al.* (1978) and Hanson (1980) divided Newcastle disease into different forms:

1. Velogenic viscerotropic NDV (VVNDV): acute and lethal infection in all ages of chickens. Haemorrhagic lesions of gastrointestinal tract are often present.
2. Neurotropic velogenic (NVNDV): an acute, often lethal infection in all ages of chickens. There are respiratory and neurological lesions.
3. Mesogenic: less pathogenic, death usually occur in young birds.
4. Lentogenic: mild respiratory infection and asymptomatic enteric form.

2.2 Newcastle Disease Virus (NDV)

NDV is a member of the order *Mononegavirales* and family of *Paramyxoviridae*. *Paramyxoviridae* is divided into two subfamilies, the *Paramyxovirinae* and the *Pneumovirinae*. Prior to 1993, NDV belong to the *Paramyxovirus* genus. In 1993 the International Committee on the Taxonomy of Viruses (ICTV) rearranged the order of *Paramyxovirus* and placed NDV within the *Rubulavirus* genus (Rima *et al*., 1995).

2.2.1 Structural and Biological Activities of NDV

NDV virion is spherical or pleomorphic, often filamenteous and maybe polyploid (containing more than one genomic equivalent) in shape with a diameter of 100-300 nm comprising the envelope, capsid and genome (Allan *et al*., 1978; Alexander, 1991). The envelope is composed of protein, carbohydrate and lipid. Spikes on the envelope are approximately 8 nm in length. The spikes consist of an antigenic compound called haemagglutinin which trigger the host to produce antibodies which inhibit haemagglutination and neutralize the virus (Rott, 1964; Spradbrow, 1987).

The virus shows some biological activities. The haemagglutination activity allows NDV to agglutinate red blood cells (RBCs) due to the binding of the HN protein with surface receptor of RBCs. The neuraminidase activity is the ability of neuraminidase enzyme to release binding of NDV with the receptor of host cells that are agglutinated. Neuraminidase is apart of the HN protein involved in destroying

mucous and releasing new virus particles. The other activity is haemolysis of RBCs. Binding of viruses on the host receptor during virus propagation is followed by fusion between virus and host membrane resulting in the fusion of two or more cells. Fusion between viruses and the host causes haemolysis (Alexander, 1991).

2.2.2 NDV Genome

The genome of NDV is a nonsegmented, single stranded, negative sense RNA of approximately 15 kb in length. The RNA genome encodes six structural proteins in the order 3'-NP-P-M-F-HN-L-5'. Flanking the NP gene is the 3' extracistronic region known as the leader sequence and the 5' extracistronic sequence is known as the trailer region (Kolakofsky *et al.*, 1974; Chambers *et al.*, 1986; Khrishnamurty and Samal, 1998).

Characterization of strain Beaudette C genome leader, using end labeling of RNA and sequence analysis, showed that the 3' end of genome contains two leader RNA species of 47 and 53 nucleotides in length (Kurilla *et al.*, 1985). Extending the region beyond the leader sequence demonstrated an open reading frame (ORF) of amino acids which represent the amino terminus of the NP protein A schematic diagram of the NDV genome and virion is shown in Figure 1.

Moreover, Philips *et al.* (1998) reported that the 5' trailer showed high degree complementarities with the 3' end terminal leader. Some variations between the 5' terminal sequences of the different strains revealed the presence of alternative polyadenylation signals of the L gene that correspond to different trailer lengths.