H5

THE HUMAN PARAINFLUENZA VIRUS 3 FUSION PROTEIN: CLONING, MAPPING, SEQUENCE ANALYSIS AND EXPRESSION

A Thesis Submitted to the

School of Graduate Studies

University of Ottawa

In Partial fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Microbiology and Immunology

School of Medicine

By

Marie-José Côté

ABSTRACT

To map the genes of human parainfluenza virus 3, two cDNA libraries were made, one from polyadenylated RNA isolated from HPIV3-infected cells, and a second from genomic RNA. By cross-hybridization analysis, 5 nonhomologous groups of clones were characterized in the mRNA-derived library. The coding assignment for each group of clones was determined by translation of individual HPIV3 mRNAs following hybridization-selection (Storey, D.G., 1987, Ph. D. thesis, University of Ottawa, Ottawa, Ontario). cDNA clones representing each of 5 HPIV3 genes, NP, P/C, M, F and HN, were used as hybridization probes to identify genomic RNA-derived cDNA clones spanning more than one gene. A sixth group of clones representing the HPIV3 L gene was identified and the gene order of the HPIV3 genome was determined to be: 3'-NP-P/C-M-F-HN-L-5'. The sequences of the 5'-ends of the HPIV3 NP, P/C, M, F and HN mRNAs were determined by primer extension. Sequence analysis of HPIV3 genomic cDNA clones spanning more than one gene demonstrated consensus sequences at the 3'-terminus, 3'-UCCUNUUUUC-5', and at the 5'-terminus, 3'-UUNAUNNUUUUU-5', of each gene. The consensus sequences are believed to be polymerase start and polyadenylation signals, respectively, and similar sequences are found at the beginning and the end of genes in other members of the family *Paramyxoviridae*. Sequence analysis of the HPIV3 gene boundaries also identified an intergenic sequence, 3'-GAA-5', lying between the two consensus sequences.

The second part of the project focused on the HPIV3 fusion protein. Sequence analysis was carried out with 4 overlapping cDNA clones representing all of the F gene and the complete nucleotide sequence of

the HPIV3 F gene was determined. The HPIV3 F gene is 1851 nucleotides long and contains a single long open reading frame coding for 539 amino acids. The predicted molecular weight of the unglycosylated precursor protein, F_0, is 60,031. Four potential carbohydrate acceptor sites were identified. Comparison of the HPIV3 F_0 amino acid sequence with the F_0 amino acid sequences of other members of the family *Paramyxoviridae* showed conservation of hydrophobic regions and cysteine residues and, to a lesser extent, proline residues. The HPIV3 F_0 amino acid sequence demonstrated a high degree of similarity with the amino acid sequence of the bovine parainfluenza virus 3 F_0 protein and less, but still considerable similarity, with the amino acid sequences of the F_0 protein of Sendai virus and human parainfluenza virus 1.

The third part of the project involved expression of the HPIV3 fusion protein using vaccinia virus recombinants. The expressed proteins had molecular weights similar to the HPIV3 F_0 and F_1 proteins and a similar F_0/F_1 ratio. Site-directed mutagenesis was used to convert individual cysteine codons of the HPIV3 F gene into serine codons. Expression and analysis of cysteine mutants of the HPIV3 fusion protein demonstrated that 10 of the 11 cysteines of the fusion protein were essential for correct processing. The F_0 precursor of these 10 mutants were expressed at low levels and not cleaved. These experiments also strongly suggested that cysteine residues 63, on F_2, and 192, on F_1, are responsible for the disulfide bond between F_1 and F_2.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ken Dimock for giving me the opportunity to work with him in his laboratory. His help, support, patience and his way to make time for us when needed made this work possible.

I wish to thank Drs. C. Yong Kang and Douglas G. Storey for their collaboration and interaction on the HPIV3 project. Without them this project would not have gone as far.

I am also thankful for the support, ideas, friendship and laughs brought by the 4110 members, Ken Garson and Donald G. Murphy and the 4111 members, Laura L. Anweiler and Donelda Lahey.

An Ontario Graduate Studentship was received for three years of this work and the University of Ottawa also provided financial support.

DEDICATION

This thesis is dedicated to Benoit

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

143B	Human osteosarcoma thymidine kinase negative cells
AMV	avian myeloblastosis virus
ATP	adenosine triphosphate
BES	N,N-bis[2-hydroxyethyl]-2-aminoethanesulfonic acid
bp	base pair
BSA	bovine serum albumin
BUdR	5-bromodeoxyuridine
cDNA	complementary deoxyribonucleic acid
CV-1	African green monkey kidney cells
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytidine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddNTP	dideoxynucleotide
ddTTP	dideoxythymidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
EDTA	ethylene diamine tetraacetic acid
EtBr	ethidium bromide
FBS	fetal bovine serum
HPIV3	human parainfluenza virus type 3

Ig	immunoglobulin
IgG	immunoglobulin G
IPTG	isopropylthiogalactoside
K	kilo
LB	Luria broth
LCDC	Laboratory Centre for Disease Control
LLC-MK2	Rhesus monkey kidney cells
MEM	minimal essential medium
MOI	multiplicity of infection
mRNA	messenger ribonucleic acid
NENS	100 mM NaAc pH 5.1, 20 mM EDTA, 200 mM NaCl and 1.0% SDS
nm	nanometer
NP40	Nonidet P40
PBS	phosphate-buffered saline
PEG	polyethylene glycol
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
SSC	3M NaCl, 0.3 M trisodium citrate
STE	100 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM EDTA
TAE	Tris-acetate buffer
TBE	Tris-borate buffer
TBS	Tris-buffered saline
TE	10 mM Tris-HCl pH 7.2 and 1 mM EDTA
TMNS	100 mM Tris-HCl pH 7.1, 10 mM $MgCl_2$, 50 mM NaCl and 500 mM sucrose
XGal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1 GENERAL INTRODUCTION

1. HUMAN PARAINFLUENZA TYPE 3 VIRUS: HISTORICAL BACKGROUND

Pathogenesis and epidemiology

Human parainfluenza virus type 3 (HPIV3) belongs to the *Paramyxovirus* genus in the family *Paramyxoviridae* along with human parainfluenza viruses types 1, 2 and 4 (Kingsbury *et al.*, 1978; Matthews, 1982). This group of viruses (HPIV1, 2, 3 and 4) is second only to respiratory syncytial virus as an important cause of respiratory tract illness in infants and children (Chanock and McIntosh, 1985). The symptoms resulting from human parainfluenza virus infections range from mild upper respiratory tract involvement to severe lower respiratory tract illnesses such as croup, pneumonia and bronchiolitis (Glezen and Denny, 1973; Chanock and McIntosh, 1985). Parainfluenza virus infections occur worldwide, do not seem to follow a seasonal pattern and happen early in life. Reinfection, especially with HPIV3, is known to commonly occur later in life, usually causing mild upper respiratory tract symptoms, but can occur in children less than 2 years of age (Glezen *et al.*, 1984; Chanock and McIntosh, 1985).

Human parainfluenza virus types 1 to 4 have been shown to typically cause mild upper respiratory tract infections. However, severe human parainfluenza virus 1 and 2 infections are usually located in the larynx and upper trachea and cause the croup syndrome. Severe HPIV3 infections involve the lower trachea and bronchioles which may result in pneumonia and bronchiolitis (Chanock and McIntosh, 1985).

Serological surveys have provided epidemiological information about the age at which HPIV3 infections occur. At least 60% of children under the age of 2 and up to 80% of children by the age of 4 have been infected with HPIV3 (Parrott *et al.*, 1962; Chanock and McIntosh, 1985). A study of 121 children followed from birth to 1 year of age, 90 of which were followed for up to 2 years, showed that the incidence of HPIV3 infections was approximately 70% per year during the first 2 years of life. From the same study, of 63 children followed from birth to age 3, only one escaped infection (Glezen *et al.*, 1984). The importance of HPIV3 as a human pathogen and the high transmissibility of the virus give us good reasons to investigate HPIV3 at the molecular level.

Taxonomy of *Paramyxoviridae*

The family *Paramyxoviridae* is divided into three genera based on the properties of the viral attachment protein. Human (and also bovine) parainfluenza virus 3 belongs to the genus *Paramyxovirus* and possesses hemagglutinin and neuramidase activity as do the other members of the genus: Newcastle disease virus (avian), mumps virus (human), parainfluenza virus types 1 (human, murine: Sendai virus), 2 (human), 4 (human) and 5 (canine: simian virus 5). The two other genera of the family *Paramyxoviridae* are: *Morbillivirus* which includes measles virus (human), canine distemper virus, peste-des-petits ruminants (ovine), rinderpest virus (bovine); and *Pneumovirus*, the main members of which are human and bovine respiratory syncytial viruses. Morbilliviruses lack neuraminidase activity and pneumoviruses typically lack both neuraminidase and hemagglutinin activity (Matthews, 1982; Morrison, 1988). All through the thesis, the term paramyxovirus will be used as a general term to

designate members of the family *Paramyxoviridae*.

Information about the genome organization of members of the family *Paramyxoviridae* should help to verify if this classification scheme reflects their evolutionary relationships.

General characteristics of paramyxoviruses

Despite the clinical importance of HPIV3, most of the characterization of paramyxoviruses has been done with Newcastle disease virus, simian virus 5 and Sendai virus, because for some time propagation of HPIV3 *in vitro* was not simple. Paramyxoviruses are pleomorphic but usually roughly spherical. The 120-300 nm particle possesses surface projections or spikes measuring approximately 8 nm in length. These spikes are embedded in a lipid bilayer, derived from the host cell membrane. A helical, elongated nucleocapsid is found inside the envelope. Paramyxoviruses have single-stranded RNA genomes of negative polarity (- sense) which, by definition, is the opposite sense to mRNA (+ sense). This implies that the linear genomic RNA, with a molecular weight of approximately 5×10^6, is not infectious. Negative sense polarity also signifies that the virus must carry its own RNA-dependent RNA polymerase in order to synthesize mRNA and replicate the genome (Kingsbury *et al.*, 1978).

Replication of paramyxoviruses

The first two steps in infection by paramyxoviruses are mediated by the two different types of glycoprotein forming the spikes on the viral envelope. The hemagglutinin-neuraminidase (HN) is responsible for the attachment of the virus to a cell receptor. This process facilitates fusion

of viral and cellular membranes and permits nucleocapsid entry into the host cell (Chen *et al.*, 1971). This last step is mediated at physiological pH by the other glycoprotein: the fusion protein (F) (Scheid and Choppin, 1974; Kingsbury, 1985).

Once the nucleocapsid is in the host cell cytoplasm, transcription of the negative stranded genome is initiated. It is believed that the RNA polymerase complex starts transcription at the 3'-end of the genome and continues to the 5'end of the genome, capping and polyadenylating the mRNA corresponding to individual genes as it goes. The mechanism for stopping and starting transcription is believed to be mediated by consensus signal sequences at each gene boundary (Gupta and Kingsbury, 1984; Kingsbury, 1985). Viral mRNAs are then translated into proteins. After translation, the nucleoprotein, phosphoprotein and the polymerase protein associate with the newly synthesized RNA genome to form the nucleocapsid (Portner and Kingsbury, 1976). The matrix protein is believed to be incorporated into the plasma membrane immediately after translation. The two glycoproteins are translated on membrane-bound ribosomes and then transported and processed via the same pathway used for cellular membrane bound proteins (Nagai *et al.*, 1976a; Kingsbury, 1985).

The mechanism by which the RNA polymerase complex switches to genome replication from transcription is not understood. This mechanism must prevent the RNA polymerase complex from responding to the stop and start signals at the gene boundaries so that a full length, plus-strand antigenome, to be used as a template for genome synthesis, can be synthesized (Kingsbury, 1985). The translation/replication switch is believed to be mediated by the availability of newly synthesized viral

proteins (Carlsen *et al.*, 1985).

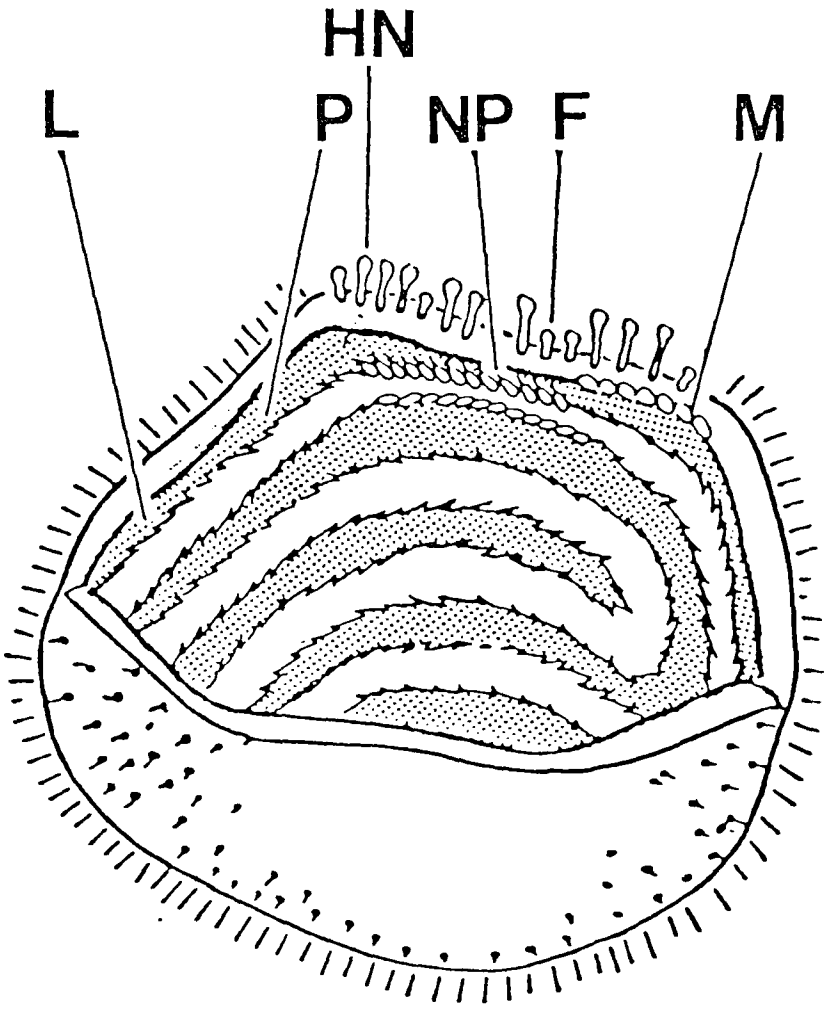
After insertion of the viral glycoproteins in the infected cell membrane, the matrix protein associates with the membrane area containing the viral glycoproteins. The nucleocapsid then recognizes these regions, possibly by interacting with the matrix protein, and assembly of the virion is completed when the new virus buds from the cell (Nagai *et al.*, 1976a).

Structural characterization of HPIV3

When this project was initiated, little was known about the molecular biology of HPIV3. It had been reported, however, that the molecular weight of HPIV3 genomic RNA was approximately 4.6 x 10^6 (Storey *et al.*, 1984) corresponding to approximately 15,000 nucleotides. The structural characterization of HPIV3 virions had also been described (Guskey and Bergtrom, 1981; Storey *et al.*, 1984; Jambou *et al.*, 1985; Sanchez and Banerjee, 1985; Wechsler *et al.*, 1985). Basically, 7 major viral structural proteins were identified by SDS-polyacrylamide gel electrophoresis and by comparisons of their properties with those of the corresponding Sendai virus structural proteins. A typical HPIV3 virus particle is depicted in Figure 1.

Associated with the nucleocapsid are the following proteins: the L (large) protein, the putative polymerase; the P (for polymerase associated or phosphorylated) protein which is believed to be associated with the polymerase complex; and the NP (nucleocapsid) protein, the most abundant viral protein, which holds together the nucleocapsid and protects the viral genome (Chanock and McIntosh, 1985). Their approximate molecular weights are 195,000 (195K) for L, 87K for P and 67K for NP (Storey *et*

Figure 1. Schematic diagram of a typical paramyxovirus. A segment of the envelope has been removed. L indicates the large protein corresponding to the putative polymerase; P, the phosphoprotein or polymerase-associated protein; HN, the hemagglutinin-neuraminidase glycoprotein; NP, the nucleoprotein; F, the fusion glycoprotein and M, the matrix or membrane protein. Taken from Kingsbury (1977).


HN
L
P
NP
F
M

al., 1984; Jambou *et al.*, 1985; Sanchez and Banerjee, 1985; Wechsler *et al.*, 1985). The M (matrix) protein, with an approximate molecular weight of 35K, is associated with the envelope components of the virion and is probably responsible for the interactions between the viral nucleocapsid and envelope components during assembly (Chanock and McIntosh, 1985). It has been shown by phosphorus labelling that NP, P and M are phosphorylated (Jambou *et al.*, 1985; Wechsler *et al.*, 1985). By glucosamine labelling, three glycoproteins have been identified: HN (69K), the hemagglutinin/neuraminidase protein which mediates the attachment of the virus to host cell receptors and F_0 (60K), the uncleaved precusor of the third glycoprotein, F_1 (46K), the fusion protein. The cleavage of F_0 yields an active disulfide-linked protein $F_{1,2}$, which is responsible for virus entry into the host cell (Chanock and McIntosh, 1985; Storey *et al.*, 1984; Jambou *et al.*, 1985; Sanchez and Banerjee, 1985; Wechsler *et al.*, 1985). Additional low molecular weight proteins have been described by individual groups: a 12K polypeptide was claimed to be the F_2 protein by Jambou *et al.* (1985); a 21K protein named VP8 was identified by Sanchez and Banerjee (1985), but the origin of this protein could not be determined.

2. GENOME ORGANIZATION OF PARAMYXOVIRUSES

Once the structural proteins of a virus have been identified and characterized, typically, the next step in understanding viral structure and function would be to examine the genome organization of the virus. When this project began, knowledge of paramyxovirus genome organization was accumulating through the use of ultraviolet inactivation and molecular cloning techniques.

Gene order of paramyxoviruses

Ultraviolet irradiation has been shown to cause premature termination of transcription and has been used in the past to determine the distance of a particular gene from its promotor (Hackett and Sauerbier, 1975). The principle is that the larger a target, the more likely it will be hit and therefore inactivated by ultraviolet light. If two genes A and B have the same promotor and gene A is proximal to the 5' end of the cistron, gene A is transcribed before gene B. Each time ultraviolet radiation hits gene B, gene A can continue to be transcribed and translated, however, when radiation hits gene A, transcription of both genes can no longer occur. Gene B, therefore, being distal from the promotor, would be more sensitive to the ultraviolet irradiation than gene A. Analysis of the results of ultraviolet inactivation are obtained by plotting the relative rates of gene-specific protein synthesis as a function of ultraviolet light dosage (Hackett and Sauerbier, 1975).

The application of this technique to negative stranded RNA viruses is slightly different. Abraham (1979) as well as Pons and Rochovansky (1979) used ultraviolet inactivation to determine if transcription of each segment

of influenza virus, an orthomyxovirus, was initiated independently. They showed that the rate of transcriptional inactivation was directly proportional to the target (segment) size. Their results indicated that each of the 8 segments of influenza virus possesses its own promotor.

Ball (1977) as well as Flamand and Delagneau (1978) used this technique with the rhabdoviruses, vesicular stomatitis virus and rabies virus which possess, like paramyxoviruses, a non-segmented genome. Their results showed that the rate of transcriptional inactivation was not directly proportional to the size of each gene. In fact, the entire genome behaved as a single transcriptional unit with a single promotor at the 3'-end. The rate of inactivation was proportional to the distance of a particular gene from the 3'-end of the genome. Ultraviolet inactivation of vesicular stomatitis virus transcription demonstrated that the order of the genes, from the 3'-end of the genome, is N (nucleocapsid), NS (non-structural), M (matrix), G (glycoprotein) and L (polymerase) (Ball, 1977). These results were confirmed by generating cDNA clones containing genomic sequences and by sequence analysis of the vesicular stomatitis virus genome (Rose, 1980).

Ultraviolet inactivation of transcription was done with two members of the genus *Paramyxovirus*, Newcastle disease virus and Sendai virus (Glazier *et al.*, 1977; Collins *et al.*, 1980). For both viruses, the data suggested that transcription originates from a single promotor at the 3'-end of the viral genome. For Newcastle disease virus, the rates of inactivation of transcription suggested the following gene order: 3'-NP-P-(F_0,M)-HN-L-5' (Collins *et al.*, 1980). For Sendai virus, which is believed to be closely related to HPIV3, the gene order was different: 3'-NP-F_0-M-P-HN-L-5' (Glazier *et al.*, 1977).

An alternate way to determine the gene order of negative stranded RNA viruses is by generating cDNA clones from the viral genomic RNA. Dowling *et al.* (1983) cloned Sendai virus genomic sequences following random priming of cDNA synthesis. A probe specific for the 3'-end of the viral genome was used to locate the 3' proximal clone. The order of the genomic clones was then determined relative to the 3' proximal clone by restriction analysis. The final order was further confirmed by dot blot hybridization. The authors then assigned a protein to each of the ordered clones by translation of hybridization-selected mRNA.

Paterson *et al.* (1984a) generated cDNA clones of simian virus 5 by using oligo(dT) to prime cDNA synthesis from infected cell polyadenylated RNA. After selecting simian virus 5 specific clones with a genomic RNA probe, clones corresponding to the different viral proteins were identified by analysing translation products of hybrid-selected mRNA. The authors also used another technique for identification of clones by translating mRNA that did not hybridize to a particular clone and identifying the corresponding protein product by default. Simian virus 5 seems to produce mRNA species that correspond to more than one gene as a result of readthrough of transcription termination signals. By combining the characterization of these polycistronic mRNAs and the sequence similarities with clones representing individual simian virus 5 genes, it was possible to determine the gene order of simian virus 5.

By cloning, Sendai virus gene order of the 3'-end region of the genome is 3'-NP-P/C-M-...-5' (Dowling *et al.*, 1983; Shioda *et al.*,1983). C is a non-structural protein seen in Sendai virus infected cells. Sequence analysis of cDNA clones corresponding to the P protein gene demonstrated that the C protein is coded by an overlapping open reading

frame using an alternate AUG in the P protein gene (Giorgi *et al.*, 1983; Shioda *et al.*, 1983). The gene order reported for simian virus 5 using a similar approach, is identical: 3'-NP-P/V-M-F-HN-L-5' (Paterson *et al.*, 1984a). V is also a non-structural protein coded by an open reading frame on the P protein gene. P mRNA differs from V mRNA by having two additional G residues, not templated by the genome, which cause a switch in reading frame and permit P translation (Thomas *et al.*, 1988). The results of molecular cloning experiments therefore, show that, in general, all paramyxoviruses have a similar gene order and highlight the similarities between paramyxovirus and rhabdovirus genomic organization. Also, the cloning results of Dowling *et al.* (1983), Shioda *et al.* (1983) and Paterson *et al.* (1984a) confirm the gene order of Newcastle disease virus predicted by ultraviolet mapping (Collins *et al.*, 1980). However, the gene order found for Sendai virus by ultraviolet mapping (Glazier *et al.*, 1977) differs from the others and appears to have been incorrect.

Ultraviolet inactivation cannot account for internal RNA polymerase initiation sites, the effects of secondary structure on transcription or post-translational regulation of viral protein synthesis. This method was used successfully for analysis of rhabdovirus genome organization. Regulation of paramyxovirus transcription or protein synthesis may be more complex than rhabdoviruses. This may explain the difficulty in applying ultraviolet mapping techniques to paramyxoviruses.

Gene boundary organization of paramyxoviruses

Negative stranded RNA viruses must have a virion associated RNA-dependent RNA polymerase in order to synthesize mRNAs, using the genome as a template, as a first step after infection. Once translation of

viral mRNAs has started, viral specific protein products direct the viral genome to become a template for the production of genome replicas (Kingsbury, 1977). The signal by which the switch from transcription to replication mode occurs is not known. Ultraviolet irradiation studies showed that the Sendai virus genome behaved as a single transcriptional unit. It was postulated that the virus possesses a single transcriptional promotor at the 3'-end of the genome (Glazier *et al.*, 1977). The two possible ways by which viral monocistronic transcripts can be synthesized from a single promotor are: 1) post-transcriptional endonucleolytic cleavage of a polycistronic transcript (cleavage model) or 2) termination and reinitiation of transcription (stop-start model).

In order to understand more about these events, Gupta and Kingsbury (1984) studied nucleotide sequences at the boundaries of Sendai virus genes. A conserved consensus sequence was found at the end of each gene: 3'-UNAUUCUUUUU-5' (genome sense), as well as at the beginning of each gene: 3'-UCCCANUUUC-5' (genome sense). The intergenic sequences found were GAA between NP and P, P and M, M and F, F and HN, and GGG between HN and L (Gupta and Kingsbury, 1984). Non-coding sequences of approximately 50 nucleotides at each end of the genome were considered to be leader sequences (Leppert *et al.*, 1979; Dowling *et al.*, 1983; Re *et al.*, 1983; Shioda *et al.*, 1983). A similar sequence was found at the 3' end of vesicular stomatitis virus genome (Colonno and Banerjee, 1976). The functions of leader sequences for paramyxoviruses are unknown. However, it is believed that they play a role in the regulation of vesicular stomatitis virus transcription/replication as preferred binding sites for the nucleocapsid protein. Binding of nucleocapsid protein would start encapsidation and prevent transcription

termination at the end of the leader, resulting in a full length genome replica (Blumberg *et al.*, 1983; Banerjee, 1987). Since transcription is initiated at the 3'-end of the genome, the leader sequence would be acting as a binding site for the polymerase and it as been shown that transcription of the leader occurs before transcription of other genes (Banerjee, 1987).

Gupta and Kingsbury (1984) believe that their data support the stop-start model. The consensus sequences found at the gene boundaries are thought to be signal sequences for the viral transcriptase: the sequence found at the end of each gene would act as a signal and a template for the addition of poly (A), the intergenic sequence surrounded by the consensus sequences would be a stop signal and the consensus sequence found at the beginning of each gene would be a start signal for the transcriptase.

In summary, the gene order of some paramyxoviruses found using molecular cloning techniques was: 3'-NP-P/C(orV)-M-F-HN-L-5'. Ultraviolet mapping was previously used to investigate paramyxovirus gene order but inconsistent results suggested that this method was unreliable. Sequence analysis of paramyxovirus genomes showed the presence of consensus sequences at the gene boundaries. These sequences are believed to be involved in transcriptional regulation. The first two objectives of this project will be to examine the genetic organization of HPIV3 by elucidating the gene order and analysing gene boundary sequences. Comparison of HPIV3 genomic organization with the organization of other paramyxovirus genomes should help us to test if the present classification scheme, based on hemagglutination and neuraminidase activity, reflects the evolutionary relationships of this family of viruses.

3. PARAMYXOVIRUS FUSION PROTEIN

Role in pathogenicity

As mentioned earlier, the virions of HPIV3 and other paramyxoviruses possess, on their surface, two glycoproteins: HN, which mediates the adsorption of the virus to the host cell and has both hemagglutination and neuraminidase activities (Chen *et al.*, 1971), and $F_{1,2}$, which exhibits fusion activity and is responsible for viral penetration of the host cell by fusion of the viral and cellular membranes. $F_{1,2}$ also mediates hemolysis and the cell-cell membrane fusion which leads to syncytium formation (Scheid and Choppin, 1974). The formation of multinucleated giant cells by cell-cell membrane fusion is one of the principal cytopathic effects of paramyxoviruses observed in tissue culture. Virus-induced syncytium formation provides an alternate way for paramyxovirus infections to spread, from cell to cell, in addition to release of the virus by budding. Merz *et al.* (1980) showed the importance of the fusion protein in pathogenesis. They prepared antibodies specific for each glycoprotein of simian virus 5 and observed the effect, in culture, of inhibition of the individual biological activities. They demonstrated that antibodies directed against HN, which blocked virus adsorption, were not effective against cell to cell spread of the virus. Antibodies specific for the fusion protein, however, completely prevented progression of infection. This stresses the important role of the fusion protein in the spread of paramyxovirus infections from cell to cell.

To become activated, the fusion glycoprotein precursor, F_0, must be cleaved by a host protease with trypsin-like activity (Homma and Ohuchi, 1973; Scheid and Choppin, 1974; Nagai *et al.*, 1976b). Nagai *et al.* (1976b)

compared avirulent and virulent strains of Newcastle disease virus in different cell lines and showed that the F_0 precursors of avirulent strains exhibited cleavage and production of infectious particles in only few cell lines. In other cell lines, uncleaved F_0 protein and low infectious titers were observed. Virulent strains however, demonstrated cleavage of the glycoprotein and high infectious virus titres in all cell lines. The fact that some cell types are unable to cleave the fusion glycoprotein precursor, a necessary event for viral penetration, demonstrated that host range, virulence and probably cell tropism depend on the presence of the appropriate activating protease.

Structure of the fusion protein

The Sendai virus F_0 precursor glycoprotein has a molecular weight of 69K and yields, after cleavage, a disulfide-linked complex $F_{1,2}$ which can be detected by SDS-polyacrylamide gel electrophoresis under non-reducing conditions (Scheid and Choppin, 1977). F_1 was determined to have a molecular weight of approximately 53K, while the molecular weight of F_2 was found to be approximately 13K. Similar results were obtained when simian virus 5 and Newcastle disease virus $F_{1,2}$ complexes were analyzed (Scheid and Choppin, 1977). Terminal analysis of F_0, F_1 and F_2 failed to identify a free N-terminus for F_0 and F_2 while an N-terminal phenylalanine was identified for F_1. This led to the conclusion that F_0 and F_2 possess the same N-terminus and that cleavage of F_0 yields a phenylalanine residue at the N-terminus of F_1. The following peptidic order for F_0 was suggested: $X\text{-}NH_2\text{-}F_2\text{-}...\text{-}Phe\text{-}F_1\text{-}...\text{-}COOH$ (Sheid and Choppin, 1977).

After fusing the membranes of Sendai virions with erythrocyte membranes, Lyles (1979) used these fused membranes to prepare inside-out vesicles containing the Sendai virus glycoproteins. Portions of viral glycoproteins normally on the surface of the virion were inside the vesicles. Protease treatment of these inside-out vesicles digested a small portion of F_1. It was concluded that the Sendai virus fusion glycoprotein was a transmembrane protein with a hydrophobic transmembrane region near the C-terminus of F_1 and that F_2 was distal to the membrane.

Processing of the fusion protein

The paramyxovirus fusion protein is a membrane glycoprotein which is believed to be translated, processed and transported to the infected cell surface via the normal cell pathway (Nagai *et al.*, 1976a). The first step of this pathway involves insertion of the nascent peptide into the rough endoplasmic reticulum. This is mediated by the attachment of a signal recognition particle to a hydrophobic region on the protein, the signal peptide, typically located at the N-terminus of the nascent polypeptide. The entire ribosome-mRNA-nascent polypeptide complex docks with the rough endoplasmic reticulum via signal recognition particle interaction with its receptor on the endoplasmic reticulum membrane, the docking protein. After docking, cotranslational translocation of the polypeptide across the membrane is accomplished (Blobel, 1980). There have been some questions about whether membrane proteins are translocated cotranslationally or posttranslationally (Verner and Schatz, 1988). However, Chao *et al.* (1987) showed that posttranslational translocation of the influenza virus hemagglutinin protein was possible only when the protein was truncated. Otherwise, the influenza virus hemagglutinin

protein was cotranslationally translocated. Using a similar approach, it was demonstrated that the deletion of the signal peptide led to the synthesis of unglycosylated protein products that were not detected at the cell surface but accumulated in the cytoplasm (Gething and Sambrook, 1982, Sekikawa and Lai, 1983). This last experiment demonstrated the importance of the signal peptide for translocation of viral membrane proteins.

After insertion into the rough endoplasmic reticulum, a protein is transported from the endoplasmic reticulum through the Golgi apparatus membranes to the plasma membrane. During transport, co- and post-translational modification occurs including glycosylation, disulfide bond formation, proteolytic cleavage and fatty acid acylation (Rothman, 1981; Freedman, 1984).

Initially, studies of intracellular processing of paramyxovirus fusion proteins used Newcastle disease virus as the model (Nagai *et al.*, 1976b; Seto *et al.*, 1981). More recent studies have used measles virus. Glycosylation of measles virus fusion protein appears to begin as soon as the protein is inserted into the endoplasmic reticulum (Sato *et al.*, 1988). The formation of disulfide bonds is also assumed to be an early event in membrane protein post-translational modification (Freedman, 1984). Such events would precede proteolytic cleavage of F_0 and its transport to the cell surface (Sato *et al.*, 1988). Even though it was demonstrated that cleavage of Sendai virus F_0 into F_1 and F_2 can occur outside the cell (Lamb *et al.*, 1976; Seto *et al.*, 1981), cleavage of Newcastle disease virus fusion protein is an intracellular event (Nagai *et al.*, 1976b; Seto *et al.*, 1981). Cleavage of F_0 appears to occur either in the *trans* Golgi compartment (which is the last compartment of the Golgi apparatus) or in

a cell compartment just after transport of the protein through the *trans* Golgi membranes (Morrison *et al.*, 1985). Measles virus fusion protein cleavage is believed to occur intracellularly as well, but can occur at the cell surface (Sato *et al.*, 1988). McGinnes *et al.* (1985) as well as Morrison *et al.* (1987) reported that Newcastle disease virus fusion protein undergoes a conformational change prior to proteolytic cleavage. By using monoclonal antibodies, they showed that the change in conformation was due to disulfide bond alterations which occurred during transport between the rough endoplasmic reticulum and the *trans* Golgi.

Fatty acid acylation was shown to occur late in the processing pathway of Newcastle disease virus fusion protein. It was suggested that fatty acid acylation may be essential for transport of certain glycoproteins to the plasma membrane (Schmidt, 1982; Chatis and Morrison, 1982).

Involvement of the F_1 N-terminus in fusion

The fusion protein is a disulfide-linked protein which requires proteolytic cleavage and, as a result, a conformational change leading to its activation. The conformational change resulting from cleavage activation, was shown to occur by ultraviolet circular dichroism and by detergent binding studies: the inactive F_0 bound approximately 27 molecules of triton X-100 per peptide and the active $F_{1,2}$ bound approximately 67 molecules of detergent (Hsu *et al.*, 1981). The detergent binding studies showed an increased hydrophobicity after cleavage of the fusion protein precursor. The importance of fusion protein cleavage led Gething *et al.* (1978) to purify the fusion protein in order to analyse the N-terminal sequence of F_1, since the N-terminus of F_2 was blocked. The

amino acid sequence of the N-terminus of Sendai virus F_1 was shown to be hydrophobic, with six amino acid residues at positions identical to those in the N-terminus of HA_2, the smaller glycoprotein generated by cleavage of the influenza virus hemagglutinin glycoprotein precursor (HA). The hydrophobicity of F_1 suggested an involvement in viral and cellular membrane interactions (Gething *et al.*, 1978). The amino acid sequence of the N-terminus of the F_1 proteins of Newcastle disease virus, simian virus 5 and measles virus were shown to be highly conserved (Scheid *et al.*, 1978; Varsanyi *et al.*, 1985). Such data added further emphasis to the possibility of direct involvement of the F_1 N-terminus in fusion and the necessity for cleavage of the precursor, F_0, to liberate the hydrophobic F_1 N-terminus.

Oligopeptides mimicking the N-terminal region of the F_1 protein of several paramyxoviruses inhibited infectivity, cell fusion and hemolysis (Richardson *et al.*, 1980). The inhibiting oligopeptides were found to bind to the cell and not to the virus and are believed to be competing with the F_1 N-terminus for a receptor site on the cell surface (Richardson and Choppin, 1983).

Interaction between the fusion protein and other viral proteins

The literature contains conflicting reports about possible involvement of the HN protein in fusion. Liposomes reconstituted with Sendai virus glycoproteins were shown to have hemolysis and cell fusion activity. When reconstituted membranes contained only the fusion protein, fusion could be demonstrated only when wheat germ agglutinin was added to agglutinate liposomes (Hsu *et al.*, 1979). These results suggest that agglutination was necessary to bring the fusion protein in close proximity

to the membrane and effect fusion. An HN deficient Sendai mutant with less than 5% of the amount of HN protein as wild type, however, was shown to retain fusion properties (Gibson *et al.*, 1988). These results are conflicting with those of Hsu *et al.* (1979).

The other HN enzymatic activity, neuraminidase, has also been implicated in cell fusion. Merz and Wolinsky (1981) showed that fusing strains of mumps virus had less active neuraminidase than nonfusing strains. This was further shown when a fusing variant of a nonfusing mumps virus parent was selected after growth in the presence of a neuraminidase inhibitor (Waxham and Wolinsky, 1986).

Another finding adds to the confusion: Portner *et al.* (1987a), using monoclonal antibodies, identified four antigenic sites on the Sendai virus HN protein, one of which did not appear to be associated with either hemagglutination or neuraminidase activity. However, antibodies specific for this site inhibited hemolysis, suggesting that antibodies binding to HN interfere with F glycoprotein activity. Again, this supports the hypothesis that HN may play a role in fusion. However, the monoclonal antibody binding to HN may itself interfere directly with the fusion process.

In summary, the paramyxovirus fusion protein is an important factor in pathogenesis: fusion of viral and cellular membranes permits the entry of the nucleocapsid into the cell cytoplasm at an early step of infection and cell-cell fusion is an alternative way for virus to spread from cell to cell. To become active, the F_0 precursor must be cleaved by a host trypsin-like enzyme which yields a disulfide-linked $F_{1,2}$ complex. Cleavage liberates the hydrophobic N-terminus of F_1 which is believed to interact with membranes during fusion.

Because of the importance of the fusion protein during the replicative cycle of HPIV3, the third and fourth objectives of this project will be to analyse the entire sequence of the HPIV3 F gene and to compare the HPIV3 fusion protein amino acid sequence with sequences of other paramyxovirus fusion proteins. This will provide information on the evolutionary and taxonomic relationships among the different members of this family and will also allow us to identify conserved sequences among the paramyxovirus fusion protein genes which may be important in the structure and the function of the fusion protein.

The fifth objective of the project will be to investigate the importance of cysteine residues for the maintenance of structure and/or function of the HPIV3 fusion protein, using site-directed mutagenesis. This will allow us to identify cysteine residues involved in the disulfide bond between F_1 and F_2.

4. STATEMENT OF THE OBJECTIVES OF THE STUDY

The objectives of the study are the following:

1) To elucidate the gene order and gene boundary organization of the HPIV3 genome.

2) to compare HPIV3 genome organization with the genomes of other paramyxoviruses to test the present classification scheme of this family and to have a better understanding of their evolutionary relationships.

3) To determine the nucleotide sequence of the HPIV3 fusion protein gene.

4) To compare the amino acid sequence of the HPIV3 fusion protein with other paramyxovirus fusion protein sequences.

5) To investigate the importance of cysteine residues for structure and function of the HPIV3 fusion protein.

CHAPTER 2 MATERIALS AND METHODS

1. TISSUE CULTURE AND VIROLOGICAL METHODS

Cells

The Rhesus monkey kidney cell line, LLC-MK2, and the African green monkey kidney cell line, CV-1, were obtained from Flow laboratories (Rockville, MD) and from the American Type Culture Collection (Rockville, MD) respectively. The human osteosarcoma thymidine kinase negative cell line, 143B, was kindly provided by Dr. J. Campione-Piccardo from the Laboratory Centre for Disease Control (LCDC), Ottawa. All cell lines were grown in autoclavable minimal essential medium (MEM, Gibco Canada, Montreal, Quebec) and 5% or 10% heat inactivated fetal bovine serum (FBS, Gibco). The cell monolayers were grown in 60 or 100 mm diameter polystyrene dishes (Nunc, Kamstrup, Denmark or Corning, Corning, NY) and incubated at 37°C in a 5% CO_2 atmosphere in a Shellab incubator (Johns Scientific, Toronto, Ontario).

Viruses

Human parainfluenza virus type 3 (HPIV3), strain C-243 was obtained from D.A. McLeod from the Laboratory Centre for Disease Control, Ottawa. LCDC originally obtained strain C-243 from the American Type Culture Collection (strain 47885). Wild type vaccinia virus (strain WR) and the recombinant vaccinia virus containing the *lacZ* gene (Vsc8) were kindly provided by Dr. D. Harnish (McMaster University, Hamilton), with the permission of Dr. B. Moss (National Institute of Allergy and Infectious Diseases, Bethesda, MD).

Infections

Confluent plates of LLC-MK2 or CV-1 cells were infected with human parainfluenza virus 3 at a multiplicity of infection (MOI) of 5 plaque forming units per cell. After addition of the HPIV3 inoculum, the cells were incubated for 1 hour at 37°C to allow the virus to adsorb. The inoculum was then removed and fresh MEM containing 5% FBS was added. Subsequently, the infected cell plates were incubated for 24-48 hours.

Vaccinia virus infected CV-1 cells, scraped from tissue culture plates were frozen and thawed three times and treated 15 minutes at 37°C with 250 μg/ml of trypsin-EDTA (Gibco) (Ozols, D. personal communication, Mackett *et al.*, 1985). The released virus was allowed to adsorb for one hour on either CV-1 or 143B cells at a MOI of 1 or 10 plaque forming units per cell. After removal of the inoculum, fresh MEM with 5-10% FBS was added to the plate and incubation was continued for 6-36 hours.

Vaccinia virus plaque assay and plaque purification (Ozols, D. personal communication, Mackett *et al.*, 1985).

After trypsin treatment, aliquots of serial 10 fold dilutions of vaccinia virus were incubated for one hour on either CV-1 or 143B cells in 60 mm diameter plates. After removal of the inoculum, CV-1 cells were overlayed with 1 x MEM, 1% low melting temperature agarose (IBI, New Haven, CT) and 5% FBS. For 143B cells, 10% FBS was used in addition to 25 μg/ml of 5-bromodeoxyuridine (BUdR, Boehringer Mannheim Biochemicals, Dorval, Quebec). After 48 hours of infection, a second overlay containing 1 x MEM, 1% low melting temperature agarose and 300 μg/ml of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (XGal, Boehringer Mannheim Biochemicals) was applied.

To plaque purify recombinant vaccinia virus, individual blue plaques were resuspended in 250 μl of MEM and trypsinized for 15 minutes at 37°C as described above. Between 10 to 50 μl of the plaque suspension was diluted to 500 μl, and the aliquot was incubated for one hour on a 60 mm diameter culture of CV-1 cells. The culture was then overlayed as described earlier.

Transfection of DNA in eukaryotic cells

Confluent cultures of LLC-MK2 cells were split 1:4 6 to 18 hours before transfection. Following the method of Chen and Okayama (1987), 20 μg of DNA were added to 50 μl of 2.5 M $CaCl_2$ and diluted to 500 μl with H_2O. The DNA-$CaCl_2$ solution was then added to an equal volume of 2 x BBS (50 mM BES buffer, 280 mM NaCl, 1.5 mM Na_2HPO_4 adjusted to pH 6.96). The transfection solution was incubated at room temperature for 10 to 20 minutes and added dropwise to the cell medium with gentle swirling. Cultures were incubated at 37°C in a 5% CO_2 atmosphere for approximately 12 hours, washed twice with MEM and incubated at 37°C for an additional 36 hours before harvesting.

2. RNA ISOLATION AND ANALYSIS

Isolation of cellular mRNA

Uninfected and HPIV3-infected LLC-MK2 cells were scraped from plates and pelleted at 200 x g for 5-10 minutes (Sorvall GLC-2B). After a wash with Tris-buffered saline (TBS: 25 mM Tris-HCl, pH 7.2, 137 mM NaCl, 5.6 mM glucose, 5.0 KCl, 0.7 mM Na_2HPO_4), the cells were resuspended in 1 x TMNS (100 mM Tris-HCl, pH 7.1, 10 mM $MgCl_2$, 50

mM NaCl, 500 mM sucrose) to a cell concentration of approximately 2.5 x 10^7 cells/ml. 25 μg/ml of polyvinyl sulphate, 30 μg/ml of spermine, triton X-100 and sodium deoxycholate to final concentrations of 1% (V:V) each were added to the cells. The cells were disrupted with 5 passes of the pestle in a Dounce homogenizer and debris were pelleted at 12,100 x g for 5 minutes at 4°C (Beckman J2-21M). An equal volume of 2 x NENS (100 mM NaAc, pH 5.1, 20 mM ethylene diamine tetraacetic acid (EDTA), 200 mM NaCl, 1.0% sodium dodecyl sulfate (SDS)) containing 500 μg/ml of proteinase K (Boehringer Mannheim) was added to the supernatant and the mixture was incubated at 37°C for 30 minutes. The supernatant was then extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1, V:V), once with chloroform/isoamyl alcohol (24:1, V:V) and total cellular RNA was precipitated with 2.5 volumes of ethanol. All RNA precipitations were done at -20°C, overnight.

Isolation of polyadenylated RNA

Precipitated total cellular RNA was dissolved in 200 μl of low salt buffer (10 mM Tris-HCl, pH 7.2, 1 mM EDTA, 0.05% SDS), incubated 2 minutes at 80°C and immediately chilled on ice. NaCl was added to a final concentration of 0.5 M and the RNA sample was loaded on an oligo(dT) cellulose (Boehringer Mannheim) column equilibrated with high salt buffer (10 mM Tris-HCl, pH 7.2, 1 mM EDTA, 0.4 M NaCl, 0.2% SDS). Non-polyadenylated RNA was then eluted from the column with high salt buffer (approximately 8 ml), and polyadenylated RNA was eluted with low salt buffer (approximately 6 ml). After adding NaCl to a final concentration of 0.3 M, the polyadenylated RNA was precipitated with 2.5 volumes of ethanol.

Isolation of HPIV3 genomic RNA (Storey *et al.*, 1984).

Medium removed from infected cells was clarified for 5 minutes at 4,500 x g and 4°C (Heraeus-Christ minifuge 2). Virus was then pelleted for 1.5 hours, at 112,400 x g and 4°C in a SW 28 (Beckman ultracentrifuge L8-55M) or SW 27 rotor (Beckman ultracentrifuge L3-50). After resuspension of the viral pellet in NENS buffer containing 200 μg/ml of proteinase K (200 μl for every 30 ml of medium), the samples were incubated at 37°C for 30 minutes. The viral RNA was then extracted twice with phenol/chloroform/isoamyl alcohol, once with chloroform/ isoamyl alcohol and precipitated with 2.5 volumes of ethanol.

Isolation of RNA from vaccinia virus infected cells

To 200 μl of infected cell lysate (1/5 of a 100mm diameter plate, see section 9), 20 μl of 10% SDS, 20 μl of 0.5 M EDTA and 32 μl of proteinase K (5 mg/ml) were added and the sample was incubated for 30 minutes at 37°C. The sample was then extracted twice with phenol/ chloroform/isoamyl alcohol, once with chloroform/isoamyl alcohol and nucleic acids were precipitated with 0.3 M NaCl and 2.5 volumes of ethanol.

Glyoxal agarose gel electrophoresis of RNA

RNA was dissolved in 5 μl of diethyl pyrocarbonate-treated water (DEP-H_2O) and added to 15 μl of sample buffer containing 10 mM Na phosphate, pH 7.0 (6.1 mM Na_2HPO_4, 3.9 mM NaH_2PO_4), 0.9 M (6%) deionized glyoxal (BDH ville St-Laurent, Quebec) and 50% dimethylsulfoxide (DMSO), final concentration. After heat denaturation at 60°C for 15 minutes, the RNA sample was loaded on a 1.2% agarose gel

and electrophoresed either at 50 volts in a HE 33 Minnie submarine gel unit (Hoefer Scientific, San Francisco, CA) for 1.5 hours or at 100 volts in a HE 99 Max submarine gel unit (Hoefer Scientific) for approximately 4 hours. The electrophoresis buffer was 10 mM Na phosphate, pH 7.0.

Northern blots

Glyoxylated RNA was transferred from agarose gels to Biodyne transfer membrane (PALL membrane, ICN Biochemicals, Montreal, Quebec) immediately after electrophoresis. The gel was placed in contact with the membrane as described by Maniatis *et al.* (1982) and blotted passively for 24 hours using 20 x SSC (3M NaCl, 0.3 M trisodium citrate) as blotting buffer . The nylon membrane was dried and then baked under vacuum at 80°C for 1 hour.

3. DNA ISOLATION AND ANALYSIS

DNA isolation from vaccinia virus-infected cells

DNA was isolated from vaccinia virus-infected cells using essentially the same procedure as for isolation of RNA from vaccinia virus infected cells described in section 2.

Small scale isolation of plasmid DNA from *E. coli* (mini-prep)

The isolation of plasmid DNA was essentially a modification of the method of Birnboim and Doly (1979). Overnight cultures (2 ml) of *E. coli* were grown in Luria broth (LB, 1.0% tryptone, 0.5% yeast extract, 0.5% NaCl) containing 50 μg/ml of ampicillin. Cells were pelleted and resuspended in 150 μl of solution I (50 mM glucose, 10 mM EDTA, 25 Mm

Tris-HCl, pH 8.0 and 8 mg/ml of lysozyme). After incubation at room temperature for 10 minutes, the bacteria were lysed by adding an equal volume of solution II (2% SDS, 0.2 M NaOH) and incubated at room temperature for an additional 10 minutes. 0.5 volumes of solution III (60 ml 8 M KAc, 19 ml glacial acetic acid, 21 ml H_20) was then added and the mixture was incubated on ice for 10 minutes. After centrifugation for 15 minutes in a microcentrifuge (Fisher), 300 μl of 2-propanol was added to the supernatants and the mixtures were incubated on ice for 2-10 minutes. The samples were then centrifuged for 15 minutes, pellets were resuspended in 100 μl H_2O and extracted once with phenol/chloroform/ isoamyl alcohol and once with chloroform/isoamyl alcohol. 0.5 volumes of 8M NH_4Ac and 2.5 volumes of ethanol were added for precipitation. All DNA precipitations were done by incubating the sample on ice for 10 minutes or overnight at -20°C. DNA was then pelleted in a microcentrifuge for 15 minutes at 4°C and mini-preps were resuspended in 50 μl of H_2O containing 1 μg of ribonuclease A (RNAse A).

Large scale isolation of plasmid DNA (maxi-prep)

This procedure was essentially the same as the small scale procedure with the following modifications. A 500 ml bacterial culture containing 50 μg/ml ampicillin was grown to an optical density (600nm, Beckman DU-8B spectrophotometer) of 0.4-0.6 and then plasmid was amplified overnight with 170 μg/ml of chloramphenicol. Bacterial cells were also grown overnight without amplification. Cells were pelleted at 5,860 x g for 10 minutes (Sorvall RC2-B) and resuspended in 10 ml of solution I. After incubation with solutions I, II and III, the samples were centrifuged for 20 minutes at 71,900 x g and 4°C using a SW 27 or SW 28 rotor. To the

supernatant, 0.66 volumes of 2-propanol were added and the DNA was pelleted at 12,100 x g for 30 minutes at 4°C (Beckman, J2-21 or J2-21M). Pellets were resuspended in 4.5 ml of TE (10 mM Tris-HCl, pH 7.2 and 1 mM EDTA) per 500 ml of culture, 4.95 g of CsCl was added, the volume was determined and ethidium bromide (EtBr) was added to a final concentration of 80 μg/ml. The samples were then centrifuged at 12,100 x g for 10 minutes at room temperature (Beckman, J2-21 or J2-21M) to remove debris and then the supernatant was centrifuged at 289,000 x g using a VTi65 rotor for 10 to 16 hours at 20°C. After the plasmid band was collected, 3 to 5 extractions were done with an equal volume of either 2-propanol saturated with NaCl or 2-butanol saturated with H_2O. Four volumes of H_2O were added to the sample and the DNA was precipitated by adding 2 volumes of ethanol. After a 10 minute incubation on ice, the plasmid DNA was pelleted at 12,100 x g for 30 minutes at 4°C. The pellet was then dissolved in TE to a concentration of 1 mg/ml.

Agarose gel electrophoresis

DNA electrophoresis was done in 0.7 to 1.5% agarose gels (IBI or Biorad, Mississauga, Ontario) using Tris-acetate buffer (TAE, 40 mM Tris-acetate, 2 mM EDTA) or Tris-borate buffer (TBE, 89 mM Tris-borate, 89 mM boric acid, 2.5 mM EDTA). Sample were prepared and electrophoresed using either a HE 33 Minnie submarine gel apparatus (50 volts for 1.5 to 2 hours) or a HE 99 Max submarine gel apparatus (100 volts for 3.5 to 4 hours).

DNA fragment isolation from agarose gels

DNA was stained in the gel with 0.5 μg/ml EtBr and the appropriate fragment was located using ultraviolet light. DEAE-cellulose membrane (Schleicher and Schuell, Keene, NH) was inserted in the gel just beyond the DNA band. Electrophoresis was continued for about 15 minutes in the dark to transfer the DNA fragment to the membrane. The fragment was then eluted from the membrane in 200 μl of 10 mM Tris-HCl, pH 7.2, 1 mM EDTA and 1.5 M NaCl at 37°C for 12-15 hours. Fresh buffer was added to the membrane and elution of the remaining DNA was done at 56°C for 1 hour. The elution buffer containing the DNA was then pooled and extracted once with phenol/ chloroform/isoamyl alcohol and once with chloroform/isoamyl alcohol. The DNA fragment was then precipitated with ethanol. Alternatively, electroelution into high salt was used. The DNA fragment of choice was cut from the gel and electroeluted in 0.5 x TBE using the IBI unidirectional electroeluter. Fragments were electroeluted for 30-50 minutes at 100 volts. The eluted DNA was then extracted twice with 2-butanol and precipitated with ethanol using 20 μg of glycogen (Boehringer Mannheim) as carrier.

Purification of oligonucleotides

Synthetic oligonucleotides (University of Ottawa, Biotechnology Research Institute) were purified by electrophoresis through a 20% polyacrylamide gel in 1 x TBE. 1/10 volume of 50% glycerol was added to the sample prior to electrophoresis at 150 volts for approximately 90 minutes using the Biorad Mini-Protean II Electrophoresis Cell. Bromophenol blue marker dye was run in a separate well. The oligonucleotide was located by ultraviolet shadow casting, cut from the

gel and eluted in 100 μl of 0.5 M NH_4Ac and 1 mM EDTA for 1 hour at 37°C. This last step was repeated once with fresh buffer. The samples were pooled and extracted twice with phenol/chloroform/isoamyl alcohol, once with chloroform/isoamyl alcohol and the DNA was precipitated with 0.5 volume of 8 M NH_4Ac, 2 volumes of ethanol and 20 μg of glycogen.

4. RESTRICTION AND MODIFICATION OF DNA

Enzymes

Restriction endonucleases and DNA modifying enzymes were supplied by Bethesda Research Laboratories (Gaithersburg, MD), Boehringer-Mannheim, New England Biolabs (Beverly, MA), Pharmacia Canada (Dorval, Quebec) and Promega Biotech (Madison, WI) unless another supplier is indicated.

Restriction digests

All DNA restriction digests were done in buffers and at temperatures recommended by the suppliers. Between 0.5 and 10 μg of DNA was incubated for 1 hour at 37°C with approximately 5 units of enzyme per μg of DNA in a final volume of 20 μl. Restriction digests were stopped by the addition of 0.1 volume of Endo-R stop solution (50% glycerol, 1% SDS, 100 mM EDTA, 0.1% bromophenol blue).

Synthesis of cDNA (Rutledge *et al.*, 1988)

HPIV3 genomic RNA (5 μg) was heated 1 minute at 90°C with 10 μg of random primers (hexadeoxyribonucleotides, Pharmacia) in 4 μl of H_2O and then chilled on ice. The RNA/primer sample was incubated at 42°C

for 1 hour in 100 mM Tris-HCl, pH 8.4, 130 mM KCl, 10 mM $MgCl_2$, 2.5 mM dithiothreitol (DTT) with 1.3 mM of each of the four deoxyribonucleotides (dNTPs), 100 μCi [α-^{32}P] dATP (800 Ci/mmol, Amersham, Oakville, Ontario) and 800 units/ml of avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences, St. Petersburg, FL) in a volume of 43 μl. After adjusting the reaction mixture to approximately 100 mM Tris-HCl, pH 7.8, 14 mM $MgCl_2$, 1 mM DTT, 50 μg/ml of bovine serum albumin (BSA), 1-2 units of ribonuclease H (RNAse H) and 20-25 units of *E. coli* DNA polymerase I were added. The sample, with a final volume of 86 μl, was incubated at 15°C for 2 hours. The reaction products were then size-fractionated by chromatography on a Sepharose CL-4B column equilibrated with 10 mM Tris-HCl, pH 8.0, 1 mM EDTA and 300 mM NaCl.

3' recessed end filling (Maniatis *et al.*, 1982)

0.1-1.0 μg of DNA in 50 mM Tris-HCl, pH 7.2, 10 mM $MgCl_2$, 0.1 mM DTT and 50 μg/ml BSA were incubated with 80 μM of dNTPs and 1-5 units of the large fragment of DNA polymerase I (Klenow fragment) at room temperature for 15-30 minutes in a final volume of 20 μl.

Ligation

50 to 500 ng of cDNA or a restriction fragment were mixed with 50 ng of linearized plasmid or 100 ng of linearized M13 replicative form and incubated in 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP with or without polyethylene glycol (PEG 8000) at 15°C overnight with 5 units of T4 DNA ligase in a final volume of 10-20 μl.

5. PREPARATION OF PROBES AND HYBRIDIZATION

Southern blots (Maniatis *et al.*, 1982)

The conditions used were essentially the same as described for the Northern blot in section 2, except that before transfer, DNA was denatured by soaking the gel in 1.5 M NaCl, 0.5 M NaOH for 30-60 minutes, followed by neutralization in 1 M Tris-HCl, pH 8.0, 1.5 M NaCl for 30-60 minutes.

Colony and plaque lifts

Bacterial colonies and M13 plaques were lifted onto Colony/Plaque Screen hybridization transfer membrane discs (Dupont, Mississauga, Ontario) and then processed as recommended by the manufacturer.

Nick-translation (Maniatis *et al.*, 1982)

0.5 to 1.0 μg of DNA in 50 mM Tris-HCl, pH 7.2, 10 mM $MgCl_2$, 0.1 mM DTT and 50 μg/ml BSA, were incubated with 20 μM dCTP, dGTP, dTTP, 100 μCi of [α-^{32}P] dATP (800 Ci/mmol, Amersham), 6.67 x 10^{-4} units of deoxyribonuclease I (DNAse I) and 5 units of DNA polymerase I at 15°C for 1.5 hours. Unincorporated dNTPs were removed using a Sephadex G-50 spin column as described in Maniatis *et al.* (1982). The nick-translated probe was then denatured for 15 minutes at room temperature in 0.33 M NaOH and neutralized with HCl and Tris-HCl, pH 7.0.

Preparation of single-stranded specific probes

Approximately 100 ng of single-stranded DNA were incubated with 2-12 ng of primer (M13 sequencing primer (17-mer) or probe primer (16-mer), New England Biolabs) in 0.2 M Hepes, pH 6.9, 20 mM $MgCl_2$, 5 mM DTT and 140 mM KCl at 55°C for 5 minutes and allowed to cool in a heating block until the temperature reached 35°C. The reaction volume was doubled by adding 25 μCi of [α-^{32}P]dATP (800 Ci/mmol, Amersham), dCTP, dGTP and dTTP to final concentrations of 50 μM each and 1 unit of Klenow fragment. The sample was incubated for 30 minutes at room temperature and then for 30 minutes at 30°C. The removal of unincorporated nucleotides and the denaturation of single-stranded probes was done as described for nick-translated probes. Probes synthesized using probe primer were not denatured.

Hybridization

Filters with bound DNA or RNA were prehybridized in hybridization buffer (6 x SSC, 0.1% SDS, 0.1% polyvinyl pyrrolidone, 0.1% BSA, 0.1% Ficoll 400, 100 μg/ml salmon sperm DNA) for 1 hour at 65°C. The probe was then added to fresh hybridization buffer and incubated with the membrane overnight at 65°C. The filters were then rinsed with 2 x SSC, 0.2% SDS, followed by 3 washes of 15 minutes each with the same solution and 3 washes of 15 minutes each with 0.2 x SSC, 0.2% SDS. After a final wash of 5 minutes with 3 mM Tris base, the membrane was placed in a cassette (Picker, Ottawa) and exposed to X-ray film (Dupont Cronex, E.I. Dupont De Nemours, Wilmington, DE) at -80°C.

6. DNA AND RNA SEQUENCING

5' end labelling of primer

Approximately 100 ng of oligonucleotide in 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA were incubated with 250 μCi [τ-^{32}P]ATP (3,000 Ci/mmol, Amersham) and 20 units of T4 polynucleotide kinase at 37°C for 30 minutes. The labelled primer was then purified by electrophoresis in 20% polyacrylamide gels as described earlier in section 4.

Sequencing from single-stranded or plasmid DNA

Plasmid DNA was denatured with 200 mM NaOH, 0.2 mM EDTA at room temperature for 5 minutes. The DNA mixture was neutralized with 300 mM NaAc, pH 5.0 and DNA was precipitated with 2 volumes of ethanol. Alkaline denaturation was not necessary for single-stranded DNA obtained following the technique of Dale *et al.* (1985). Chain-termination sequencing was done using either Klenow enzyme and a modification of the method of Sanger *et al.* (1977) (G. Bellemare and C. Lemieux), or modified T7 DNA polymerase (Sequenase, United States Biochemical Corporation, Cleveland, OH). Sequencing using the Sequenase kit was done according to the manufacturer's instructions. Most of the sequencing was done using Klenow enzyme, as follows. 2 to 3 μg of plasmid DNA or 1 μg of M13 single-stranded DNA was annealed with 4 to 50 ng of primer (M13 17-mer sequencing primer or specific primers synthesized with an Applied Biosystems 380B DNA synthesizer, University of Ottawa Biotechnology Research Institute) in 10 mM Tris-HCl, pH 8.0 and 10 mM $MgCl_2$ in a volume of 9 μl, by incubating 2 minutes at 90°C, 10 minutes at 55°C and

slowly cooling the sample to 35°C. The DNA/primer hybrid preparation was aliquoted into 4 tubes for the different A, C, G and T reactions. The A reaction contained 12 μM ddATP and 25 μM dCTP, dGTP and dTTP; the C reaction contained 40 μM ddCTP, 3.2 μM dCTP and 35 μM dGTP and dTTP; the G reaction contained 200 μM ddGTP, 3.2 μM dGTP and 35 μM dCTP and dTTP; and the T reaction contained 600 μM ddTTP, 3.2 μM dTTP and 35 μM dCTP and dGTP. Each reaction contained 6.5 μCi (3.2 μM, final concentration) [α-^{32}S]dATP (>400 Ci/mmol., Amersham) and 0.5 units of Klenow enzyme. Reaction mixtures with a final volume of 5 μl were incubated 30 minutes at 42°C and chased 30 minutes at 42°C with approximately 300 μM dNTPs in a final volume of 7 μl. The reactions were stopped with 1 volume of 90% formamide buffer containing 10 mM EDTA, 0.3% xylene cyanol, bromophenol blue and orange G and were incubated at 95°C for 3 minutes prior to loading.

Sequencing from mRNA (primer extension) (Collins *et al.*, 1986; Storey *et al.*, 1987)

^{32}P-labelled primer (1.5 x 10^6 cpm) was incubated with 100 μg of total RNA extracted from HPIV3-infected cells in a volume of 20 μl at 95°C for 1 minute and then chilled on ice. The sample was aliquoted into 4 tubes for the A, C, G and T reactions. The final concentration of dNTPs was 100 μM in each reaction and the appropriate ddNTP was at 75 μM. Each reaction, with a final volume of 10 μl, was incubated for 30 minutes at 42°C in 100 mM Tris-HCl, pH 8.4, 130 mM KCl, 10 mM $MgCl_2$ and 2.5 mM DTT, containing 80 μg/ml of actinomycin D and 15 units of AMV reverse transcriptase. After precipitation of the nucleic acids with NH_4Ac and ethanol, the pellet was resuspended in formamide buffer as

described above and incubated 5 minutes at 95-100°C just before loading.

Sequencing gel

Sequencing samples were electrophoresed on the IBI sequencing apparatus Model STS-45. Two types of gel were used: 8% acrylamide, 0.4% bis-acrylamide and 8 M urea or 6% acrylamide, 0.3% bis-acrylamide and 8 M urea. The running buffer was 1 x TBE and electrophoresis was carried out at a constant power of 55 watts. Sequencing gels were either bound to a glass plate treated with silane A-174 (BDH), fixed 20-30 minutes with 10% methanol, 10% acetic acid and dried 30 minutes in an oven at 60°C, or lifted from the glass plate onto Whatman (3MM) paper and dried under vacuum at 80°C for 30 minutes. Gels were then exposed to X-ray film (Dupont) at -80°C.

7. M13 PHAGE

All M13 procedures were done according to Messing *et al.* (1983) with some modifications.

M13 plaque assay (Messing, 1983)

Serial 10 fold dilutions of M13 phage (culture supernatants or phage resuspended in low salt TE) were added to a mixture of 200 µl of fresh *E. coli*, 10 µl of 100 mM isopropylthiogalactoside (IPTG), 50 µl of 2% XGal and 3 to 5 ml of melted soft agar (1.0% tryptone, 0.8% NaCl, 0.6% agar, 0.01% vitamin B_1). The mixture was then poured on a B plate (1.5% agar, 1.0% tryptone, 0.8% NaCl, 0.01% vitamin B_1) and incubated at 37°C overnight.

Single-stranded phage DNA analysis directly from culture supernatants

Each M13 plaque was propagated at 37°C in 5-10 ml of 2YT (1.6% tryptone, 1.0% yeast extract, 0.5% NaCl) for 7-9 hours in *E. coli* JM101, or 10-12 hours in *E. coli* MV1190 cells. The cells were then pelleted at 4,500 x g for 10 minutes (Heraeus-Christ minifuge 2) and 1 µl of 10% SDS as well as 1 µl of Endo·R stop solution (described in section 3) were added to 10-20 µl of each supernatant. The samples were then electrophoresed on a 0.7% agarose gel in 1 x TBE, at 50 volts for 12-16 hours (Max submarine gel apparatus).

C-test

5 to 10 µl of each M13 supernatants were incubated with 1 µl of 10% SDS and 1 µl of Endo·R stop solution for 1 hour at 65°C. The samples were then electrophoresed as described above.

Isolation of single-stranded M13 phage DNA

To supernatants of M13 infected cultures, 0.2 volumes of 20% PEG, 3.5 M NH_4Ac were added and the mixture was incubated 30-60 minutes on ice. The samples were centrifuged at 12,100-17,400 x g for 15 minutes (Beckman, J2-21 or J2-21M), the viral pellet was resuspended in 300 µl of 10 mM Tris-HCl, pH 7.2, 1 mM EDTA, 300 mM NaCl and incubated 30 minutes on ice. The samples were then centrifuged 2 minutes in a microfuge and the supernatants were extracted once with phenol, twice with phenol/chloroform/isoamyl alcohol and once with chloroform/isoamyl alcohol. Single-stranded M13 phage DNA was then precipitated with 0.5 volumes of 8 M NH_4Ac and 2 volumes of ethanol.

Isolation of M13 replicative form

This procedure is essentially the same as the small scale isolation of plasmid DNA described in section 3, with the following modifications. M13-infected *E. coli* from 5-10 ml overnight cultures were resuspended in 500 μl of solution I. Equal volumes of solutions II and III were added and after centrifugation, the supernatant was added to 1 ml of 2-propanol. The pellet, resuspended in 300 μl of TE was extracted twice with phenol/ chloroform/isoamyl alcohol and once with chloroform/isoamyl alcohol. M13 replicative form DNA was then precipitated with NH_4Ac and ethanol.

8. GROWTH AND TRANSFORMATION OF BACTERIAL CELLS

Transformation of competent *E. coli* with plasmid DNA

10 ml of 2YT was inoculated with *E. coli* and incubated overnight at 37°C without shaking. 10 ml of the overnight culture was used to inoculate 100 ml of P medium (20 mM KPO_4, pH 7.0, 15 mM $(NH_4)_2SO_4$, 10 mM $MgSO_4$, 1.8 μM $FeSO_4$, 1% casamino acids and 0.25% glucose) and the culture was incubated with shaking at 37°C until an optical density (600 nm) of 0.3-0.4 was reached. The cells were then centrifuged at 2,600 x g for 10 minutes (Sorvall RC2-B) at 4°C and washed with 10 mM NaCl. The pellets were resuspended in 0.5 volumes of 50 mM $CaCl_2$, incubated on ice for 15 minutes and then pelleted again. The pellets were resuspended in 0.1 volumes of 50 mM $CaCl_2$, 16% glycerol, quick frozen in an ethanol/dry ice bath, and stored at -80°C. For transformation, competent cells were thawed on ice and 200 μl were added to 10-20 μl of a ligation reaction. The mixture was incubated on ice for 30 minutes and

then heated 90 seconds at 42°C. 800 μl of LB was added and the sample was shaken vigourously for 1 hour at 37°C. The culture was then plated on LB + ampicillin (50 μg/ml) agar plates and incubated at 37°C overnight.

Transformation of competent *E. coli* with M13 single-stranded or replicative form DNA (Messing, 1983)

A 5 ml overnight culture of JM101 or MV1190 in 2YT was diluted 1:1000 and grown at 37°C until an optical density (600nm) of 0.6-0.7 was reached. The cells were then centrifuged at 4,500 x g for 5 minutes at 4°C (Heraeus-Christ minifuge 2) and resuspended in 0.5 volumes of 50 mM $CaCl_2$. After an incubation on ice of 20 minutes, the cells were pelleted again and resuspended in 0.1 volumes of 50 mM $CaCl_2$. A 300 μl aliquot of competent cells was added to M13 DNA and the mixture was incubated on ice for 40 minutes. After a heat shock at 42°C for 2 minutes, 10 μl of 100 mM IPTG, 50 μl 2% of XGal, 200 μl of fresh *E. coli* and 3 to 5 ml of melted soft agar were added to each tube. Mixtures were then poured on B agar plates and incubated at 37°C overnight.

9. PROTEIN ANALYSIS

Radiolabelling of HPIV3 and vaccinia virus proteins

24 hours following infection with HPIV3 or 1 hour following infection with vaccinia virus, cells were starved for 2 hours with leucine-free MEM (Flow Laboratories) and then labelled with 20 μCi/ml of L-[4,5-^{3}H] leucine (139 Ci/mmol, Amersham) for 4 hours. Uninfected cells were labelled in an identical manner.

Cell lysates

Uninfected or HPIV3 or vaccinia virus-infected cells from each 100mm diameter culture dish were washed once with TBS, scraped into 4 ml of TBS and centrifuged at 200 x g for 5 minutes (Sorvall GLC-2B). Cells were resuspended in 500 μl of RIPA buffer without detergents (50 mM Tris-HCl, pH 7.2 and 150 mM NaCl) and transferred into a microfuge tube. 500 μl of RIPA buffer containing 2 x detergents (50 mM Tris-HCl, pH 7.2, 150 mM NaCl, 0.2% SDS, 2% sodium deoxycholate, 2% triton X-100, 2 mM phenylmethyl-sulfonyl fluoride (PMSF) and 2 mM benzamidine·HCl (BHCl)) were added and the samples were incubated 1-2 minutes on ice. Cell debris and nuclei were removed by centrifugation in a microcentrifuge for 30 seconds. Cell lysates were then stored at -80°C.

Immunoprecipitation

Usually 1/3 of a plate (approximately 300 μl of cell lysate) was immunoprecipitated with 20 mg/ml of protein A Sepharose CL-4B beads (Pharmacia or Sigma Chemical Company, St. Louis, MO) and either 20-30 μl of rabbit anti-HPIV3 serum (kindly provided by D. McLeod from LCDC) or 5 μl of rabbit polyclonal monospecific antiserum (kindly provided by R. Ray from U. of Alabama, Birmingham, AL)(Ray and Compans, 1987) or monoclonal antibodies (kindly provided by R. Rydbeck from Karolinska Institute, Stockholm Sweden)(Rydbeck *et al.*, 1986). The samples were allowed to mix either for 2 hours or overnight at 4°C. After 5 washes with RIPA buffer containing 1 x detergents or with 1% NP40 in STE (100 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM EDTA), 15 μl of 2 x sample buffer (62.5 mM Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 0.02%

bromophenol blue and 10% β-mercaptoethanol) were added to the beads. Samples were then heated at 95°C for 5 minutes and either loaded on a SDS-polyacrylamide gel or stored at -20°C.

Polyacrylamide gel electrophoresis (Laemmli, 1970)

Analysis of cell proteins was done by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The resolving gel consisted of 10% acrylamide, 0.27% bis-acrylamide, 0.1% SDS and 750 mM Tris-HCl, pH 8.8. The stacking gel consisted of 5% acrylamide, 0.13% bis-acrylamide, 0.1% SDS and 66 mM Tris-HCl, pH 6.8. Electrophoresis was carried out at 150 volts in 25 mM Tris base, 192 mM glycine, 0.1% SDS buffer using the Bio-Rad Mini-Protean II Electrophoresis Cell. After electrophoresis, gels were fixed in 30% methanol, 10% acetic acid for 20-30 minutes and prepared for fluorography with Amplify (Amersham) for 25-30 minutes. Gels were dried under vacuum for 1 hour at 60°C, placed in a cassette (Picker) and exposed to X-ray film (Dupont) at -80°C.

10. IMMUNOFLUORESCENCE

Approximately 10,000 CV-1 cells were plated on each well of a Multitest slide (Flow Laboratories). Cells in some wells were infected with HPIV3 or vaccinia virus at a MOI of 1 plaque forming units per cell for 6-24 hours. The slides were rinsed in PBS (8.1 mM Na_2HPO_4, 1.9 mM $NaH_2PO_4 \cdot H_2O$, 137 mM NaCl and 2.7 mM KCl) for 5 minutes, cells were fixed 3-5 minutes with acetone and then the slides were dried. A dilution (1:500) of anti-HPIV3 serum in PBS was applied to each well and slides were incubated 30-60 minutes at 37°C. Slides were then washed 3 x in

PBS for 5 minutes and a dilution (1:20 or 1:30) of fluorescein-conjugated goat anti-rabbit IgG (Jackson Immunoresearch Laboratories, West Grove, PA) or sheep anti-mouse Ig (Amersham) was applied. Slides were incubated at 37°C for 30 minutes, washed 3 x for 5 minutes with PBS and dried. A drop of glycerol in carbonate buffer (50 mM Na_2CO_3, 450 mM $NaHCO_3$, pH 8.0) was added to each well and a cover slip was placed on the wells. Slides were examined using an ultraviolet microscope (Laborlux K, Wild-Leitz Canada Ltd) .

11. HEMOLYSIS

CV-1 cells from each 100mm diameter culture dish were rinsed once with TBS, scraped into 4 ml of TBS and centrifuged at 200 x g for 5 minutes (Sorvall GLC-2B). Cells were resuspended in 900 μl of TBS and frozen and thawed 3 times. 100 μl of TBS-washed (twice) human red blood cells were added to the lysates and the mixtures were incubated overnight at 37°C. Samples were then centrifuged in a microcentrifuge for 1 minute to remove debris and red blood cells and the optical density (550 nm) of the supernatant was determined and recorded.

CHAPTER 3 GENE MAPPING OF HUMAN PARAINFLUENZA VIRUS 3

INTRODUCTION

The first objective of this project was to elucidate the genome organization of HPIV3. A comparison of the genome organization of several paramyxoviruses will provide a better understanding of the evolutionary relationships within the family *Paramyxoviridae*. The gene order of some paramyxoviruses has been elucidated using ultraviolet mapping or molecular cloning techniques. Since cloning techniques appear to be more reliable and certainly more direct, HPIV3 genome organization was investigated by molecular cloning.

To determine the gene order of HPIV3 the strategy was to use a bank of cDNA clones synthesized from polyadenylated RNA isolated from HPIV3 infected LLC-MK2 cells and to group these clones by cross-hybridization. The identification of the viral protein product represented by each group was made possible by hybrid-selection and translation of individual HPIV3 mRNAs (Storey, D.G., 1987; Dimock *et al.*, 1987). After this first step, a second HPIV3 clone bank was generated using cDNA synthesized from HPIV3 genomic RNA. The genomic bank was predicted to contain cDNA clones spanning two or more genes. A representative member of each group of clones was used to probe the genomic clone library to identify adjacent genes and to deduce the gene order.

To investigate the organization of the HPIV3 gene boundary regions, the strategy was to determine the nucleotide sequence of cDNA clones spanning more than one gene. In addition, information concerning the 5' terminus (messenger sense) of the viral genes was obtained by direct

sequence analysis of RNA isolated from infected cells using specific primers that hybridized near the 5'-end of each viral mRNA.

A complete gene map of the HPIV3 genome will be presented and discussed. HPIV3 genome organization will be compared with that of other paramyxoviruses.

RESULTS

1.Characterization of the HPIV3 mRNA cDNA library

The HPIV3 mRNA-derived cDNA library was synthesized by Dr. K. Dimock (Dimock *et al.*, 1987). cDNA synthesis was primed with oligo(dT) and polyadenylated RNA isolated from HPIV3-infected LLC-MK2 cells was used as template (Maniatis *et al.*, 1982). After fractionation by chromatography on Sepharose 4B, the cDNA was ligated into pBR322 using *Bam*HI linkers, and the ligation mix was used to transform *E.coli* strain RR1. Hybridization of the library with ^{32}P-labeled fragments of HPIV3 genomic RNA revealed that approximately 5% of the ampicillin-resistant colonies contained HPIV3-specific sequences.

Restriction endonuclease digestion with *Bam*HI was initially done on approximately 50 viral specific clones in order to identify clones possessing the largest inserts. These inserts were then isolated from agarose gels, labeled by nick-translation and used to probe a membrane disc containing DNA from clones of the library. The preliminary results of the colony hybridization permitted the identification of clones that did not hybridize to any of the inserts. Inserts of these clones were used as probes in a second round of colony hybridization. Using the cross-hybridization method, 5 distinct groups of HPIV3 mRNA clones were

identified which did not cross-hybridize with each other. Each group was presumed to represent a distinct HPIV3 gene.

An HPIV3 protein was assigned to each group by Dr. D.G. Storey (Dimock *et al.*, 1987; Storey, 1987) by analysis of *in vitro* translation products of hybridization-selected mRNA using clones representative of each group. DNA from each representative clone was fixed to nitrocellulose discs and after hybridization with RNA isolated from HPIV3-infected or uninfected cells, the discs were washed and the bound RNA was eluted and translated *in vitro* using wheat germ extracts or rabbit reticulocyte lysates in the presence of [^{35}S]methionine. The ^{35}S-labeled products were then analyzed by SDS-polyacrylamide gel electrophoresis.

The cross-hybridization and hybrid-select translation results are shown in Table 1. pPI3 (approximately 1,000 bp) was representative of a group of 6 clones assigned to the matrix (M) protein gene. A group of 3 clones was represented by clone pPI28 (approximately 900 bp) and assigned to the phosphoprotein (P/C) gene. Clone pPI47 (approximately 1,800 bp), was representative of the most abundant group (31 members) of cDNA clones, assigned to the nucleocapsid (NP) protein. The last two groups of clones were assumed to represent the two glycoproteins. The group of clones with 4 members, was represented by pPI10 (approximately 900 bp) and was assigned to the hemagglutinin-neuraminidase (HN) gene. The *in vitro* translation product of HN mRNA was not glycosylated in the *in vitro* translation systems, however, a polypeptide of the same size was detected in HPIV3-infected LLC-MK2 cells treated with the glycosylation inhibitor, tunicamycin, confirming the assignment. Clone pPI14 (approximately 1,600 bp) was the only clone that could be assigned to the

TABLE 1

GENE ASSIGNMENT OF HPIV3 mRNA-DERIVED cDNA CLONES

GROUP[1]	REPRESENTATIVE CLONE	NUMBER OF CLONES[2] IN THE GROUP	GENE ASSIGNMENT[3]
A	pPI3	6	M
B	pPI10	4	HN
C	pPI14	1	F
D	pPI28	3	P
E	pPI47	31	NP

[1] Distinct groups of clones were identified by cross hybridization analysis

[2] 50 clones were analyzed

[3] Gene assignments were based on hybrid-selection and *in vitro* translation

fusion protein (F) gene. This assignment was not as definitive as the others because of the very low efficiency of translation of the F mRNA in both translation systems, but a product was identified.

A HPIV3 mRNA was assigned to each gene by Dr. D.G. Storey (Dimock *et al.*, 1987; Storey, 1987). Northern blots containing polyadenylated RNA from HPIV3-infected cells were probed with ^{32}P-labeled inserts representing each HPIV3 gene to make the assignments.

The second step in the characterization of the cDNA clones was to determine the orientation of each viral-specific insert with respect to the HPIV3 genome, for sequencing purposes. To determine the orientation of each cloned sequence, a representative clone of each of the 5 genes was digested with *Bam*HI and ligated into the *Bam*HI site of the replicative form of phage M13mp10. M13 is a single-stranded DNA bacteriophage infecting *E. coli*. When the phage replicates, the single-stranded DNA is converted into a double-stranded replicative form. Both forms can be isolated: double-stranded replicative form from infected *E. coli* cells and the single-stranded form from the phage released into the growth medium. The principal feature of M13 is that the phage always packages the + strand of the replicative form (Messing, 1983). After transformation of the M13 recombinant replicative form into *E. coli* JM101, 10 colorless phage plaques were collected for each group and amplified (Chapter 2). Since the subcloning was done at a single restriction site, the subclones should be a mixture of clones having either orientation of the insert. The C-test technique was used to identify clones with inserts in opposite orientations. Because the + strand is always present in the phage, clones with inserts in the opposite orientation should hybridize to each other. After incubation, the samples were electrophoresed on agarose gels and if

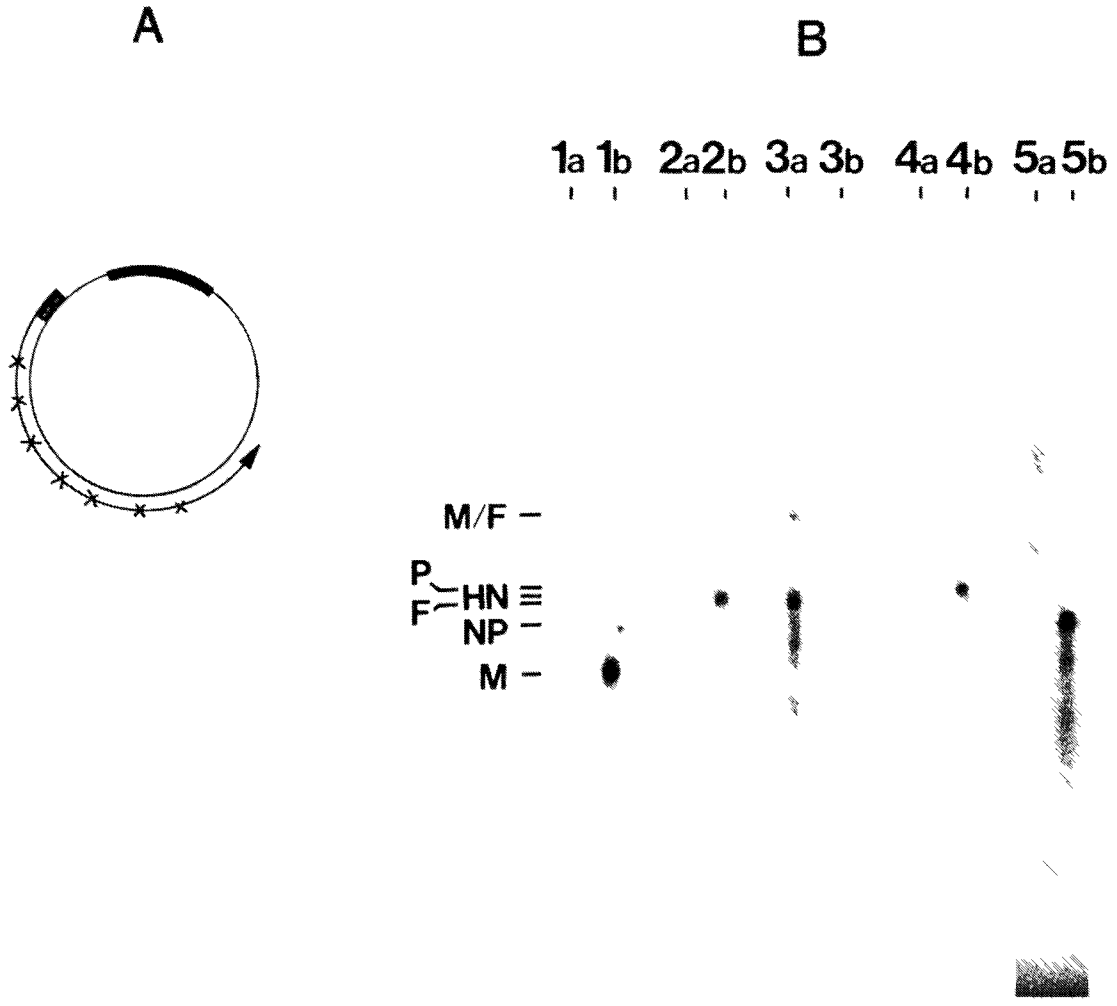
the two clones have inserts of the opposite orientation, they will hybridize to each other and the resulting figure eight-like form will be retarded in agarose gels. For each of the 5 genes, two M13 clones with inserts of opposite orientation were identified and selected.

After isolation of the M13 clones with inserts of opposite orientations, strand-specific probes were made by priming cDNA synthesis with the M13 (16-mer) probe primer. As shown in panel A of Figure 2, this primer directs DNA synthesis to proceed away from the insert, around the M13 genome. Limited by deoxyribonucleotide concentration, DNA synthesis should not proceed far enough to copy the insert itself. These ^{32}P-radiolabeled cDNA hybrids were used, without denaturation, to probe Northern blots containing polyadenylated RNA isolated from HPIV3-infected cells. The results are shown in Figure 2, panel B. Clones pPI3:6, pPI10:18, pPI14:e, pPI28:29 and pPI47:46, corresponding to the M, HN, F, P and NP genes, all hybridized to their specific HPIV3 mRNAs. The conclusion is that these inserts are genomic (-) strand sense: the 3'-end of each of these inserts corresponds to the 5'-end of the mRNA. The HPIV3 specificity of the set of clones that did not hybridize was confirmed by priming cDNA synthesis with M13 (17-mer) sequencing primer. This time, cDNA synthesis proceeded through the insert and, after denaturation, the ^{32}P-labeled cDNA was used for Northern blot hybridization. As expected, all probes hybridized to their respective mRNAs (data not shown).

2. Characterization of the HPIV3 genomic cDNA library

A viral genomic clone bank was constructed in order to determine the HPIV3 gene order. HPIV3 genomic RNA was isolated from virus

Figure 2. Orientation of representative clones of 5 HPIV3 genes. Panel A shows an example of the M13 strand-specific probes used for this experiment. The heavy bar represents the single-stranded insert, the black block represents the annealed probe primer, the arrow indicates the direction of second strand synthesis and the crosses indicate [^{32}P] label. In panel B, polyadenylated RNA was isolated from HPIV3-infected cells, denatured with glyoxal, electrophoresed on a 1.2% agarose gel and transferred to a nylon membrane. Strips were hybridized with strand specific probes. Lane 1a, pPI3:3 (M); 1b, pPI3:6 (M); 2a, pPI10:12 (HN); 2b, pPI10:18 (HN); 3a, pPI14:e (F); 3b, pPI14:f (F); 4a, pPI28:23 (P); 4b, pPI28:29 (P); 5a, pPI47:44 (NP); 5b, pPI47:46 (NP). The positions of HPIV3 specific mRNAs are indicated to the left of panel B.


A
B
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b
M/F
P
HN
F
NP
M

collected from the culture fluid of infected LLC-MK2 cells and cDNA synthesis was primed using random hexanucleotides. Double-stranded cDNA was size-fractionated by chromatography on a Sepharose CL-4B column. The fractions containing the largest double-stranded cDNA were blunt-end ligated into the *Sma*I site of the replicative form of M13 mp10.

The genomic library was first characterized by hybridization with nick-translated plasmids representing each of the HPIV3 genes. The M13 plaques were lifted onto a membrane and, after the DNA was fixed, membranes were probed with ^{32}P-labeled plasmids pPI3 (M), pPI10 (HN), pPI14 (F), pPI28 (P) and pPI47 (NP). We were particularly interested in clones that hybridized with more than one probe, and therefore, should contain intergenic sequences. The results of this first screening are summarized in Table 2.

The first attempt at genomic cloning yielded 94 colorless plaques, 24 of which hybridized to the HPIV3-specific probes. The others were assumed to contain either sequences not covered by the probes or sequences from the L gene, which was assumed to cover almost half the viral genome and did not appear to be represented in the mRNA-derived cDNA bank. It was possible that some of the colorless plaques were simply M13 mutants that no longer expressed β-galactosidase. There were 19 clones that hybridized to only one probe: 7 to NP, 1 to P, 2 to M, 3 to F and 6 to HN. 4 clones were shown to hybridize to two probes: 3 to both P and M and 1 to both F and HN. Clones spanning the HN/L region could not be identified since we did not have an L specific probe. 1 clone showed hybridization to probes specific for 4 genes: NP, P, M and F. The clone covering the NP/P/M/F regions, mpPIgX (or for close friends, Mr. X), had an insert of approximately 4.5 Kbp.

TABLE 2

CHARACTERIZATION OF THE HPIV3 GENOMIC RNA-DERIVED cDNA LIBRARY

POSITIVE HYBRIDIZATION	NUMBER OF POSITIVE CLONES[1]
NP	7
NP-P/C	-
P/C	1
P/C-M	3
M	2
M-F	-
F	3
F-HN	1
HN	6
NP-P/C-M-F	1

1 Probes specific for the NP, P/C, M, F and HN genes of HPIV3 were used to screen 94 recombinant M13 plaques

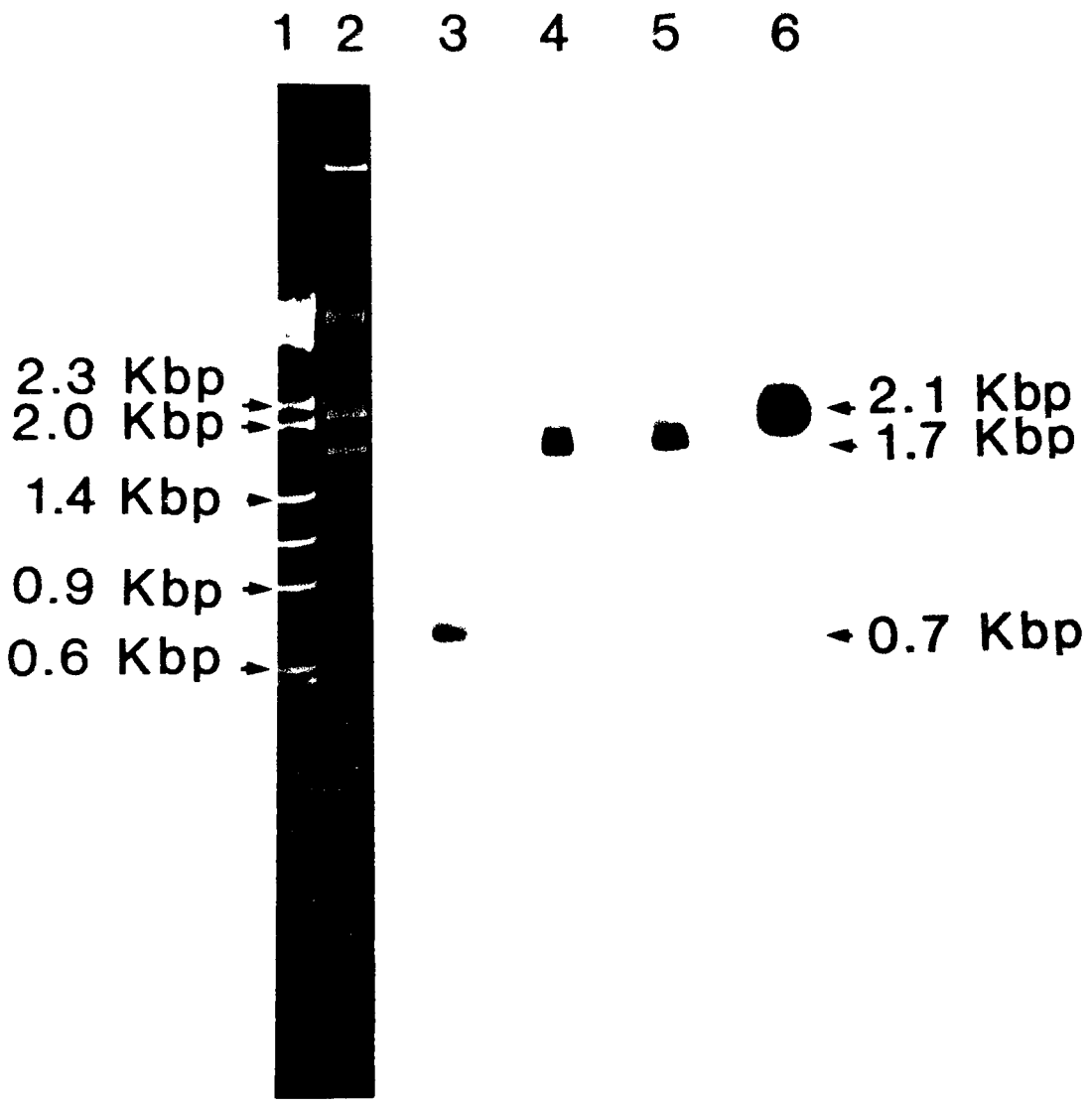
To characterize mpPIgX further, the replicative form DNA was digested with *Eco*RI and *Bam*HI and electrophoresed in an agarose gel. The resulting fragments were transferred from the agarose gel to a nitrocellulose membrane. Strips containing all the resulting DNA fragments were probed with nick-translated plasmids representing NP (pPI47), M (pPI3), F (pPI14) and P (pPI28). The results are shown in Figure 3. Hybridization with a probe representing HN (pPI10) was done, but no hybridization was detected.

*Eco*RI and *Bam*HI digestion yielded 3 fragments of 0.7, 1.7 and 2.1 Kbp. The sum of these 3 fragments suggests an insert of 4.5 Kbp in length. The small fragment of 0.7 Kbp corresponded to the NP gene portion. The medium sized fragment of 1.7 Kbp hybridized to probes representing M and F. The large fragment of 2.1 Kbp. corresponded to P gene sequences.

3. Characterization of HPIV3 gene boundaries

Primer extension directly off mRNA was used as a method to locate the 5'-ends (mRNA sense) of the HPIV3 genes. Oligonucleotides complementary to sequences approximately 50 nucleotides away from the 5'-end of each mRNA were synthesized by the University of Ottawa Biotechnology Research Institute. Selection of oligonucleotide primers was done using sequence information from Dimock *et al.* (1986), Luk *et al.* (1986), Cote *et al.* (1987), Prinoski *et al.* (1987) and Storey *et al.* (1987). The sequences of these oligonucleotides (20-mers) are presented in the Appendix. Each oligonucleotide was labeled at the 5'-end with [γ-^{32}P] ATP and T_4 polynucleotide kinase and purified, after the labeling reaction, by electrophoresis on a polyacrylamide gel. Each labeled

Figure 3. Characterization of clone mpPIgX by Southern blot analysis. The replicative form of clone mpPIgX was digested with restriction enzymes *Bam*HI and *Eco*RI and electrophoresed in several lanes of a 1.5% agarose gel. The DNA in one track was stained with ethidium bromide (lane 2) and the remainder was transferred to nitrocellulose. Strips were hybridized with nick-translated plasmids representing NP, M, F or P/C genes (lanes 3-6). An HN-specific probe (pPI10) showed no detectable hybridization. Ethidium bromide staining: lane 1, X174 replicative form DNA (*Hae*III digest) and lambda DNA (*Hind*III digest); 2, mpPIgX replicative form. Hybridization probes were: lane 3, pPI47(NP); 4, pPI3(M); 5, pPI14(F); 6, pPI28(P).


1 2 3 4 5 6
2.3 Kbp
2.0 Kbp
1.4 Kbp
0.9 Kbp
0.6 Kbp
2.1 Kbp
1.7 Kbp
0.7 Kbp

oligonucleotide was used, in the presence of reverse transcriptase and mixtures of deoxy- and dideoxyribonucleotides, to prime DNA synthesis using total RNA isolated from HPIV3-infected cells as a template. The primers, which were genomic sense, should anneal to the corresponding viral mRNA and prime cDNA synthesis which should continue to the 5'-end of the mRNA template. Each set of primer extension reactions were electrophoresed on 8M urea, 8% polyacrylamide gels. The primer extension results for NP, P, M, F and HN mRNAs are shown in Figure 4. Primer extension with 2 different L specific oligonucleotide primer mRNA was attempted several times without success, which may be attributed to the low levels of the L mRNA in HPIV3-infected cells.

Lanes corresponding to reactions containing ddATP, ddCTP, ddGTP and ddTTP are indicated for each HPIV3 gene in Figure 4, as well as a lane for a reaction containing no inhibitor. The last two nucleotides were difficult to identify because of the intensity of the signal corresponding to termination of DNA synthesis at the 5'-end of each mRNA. However, DNA sequencing of genomic clones identified the missing two nucleotides: 3'-UC-5' in each case. The corresponding sequences found at the 3'-end (genome sense) of each gene are shown in Figure 5. Two additional observations were made in this experiment: readthrough of cDNA synthesis past the strong stop signal was observed for genes NP and F. An example is shown in Figure 5. Elango *et al.* (1988) made the same observation for the NP gene of mumps virus.

The genomic cDNA library provided clones spanning more than one gene which should include intergenic sequences. After subcloning fragments of these clones into M13, the sequences contained in these gene boundary regions were determined and are displayed in Figure 6. The

Figure 4. Sequence of the 5' termini of 5 HPIV3 mRNAs. Primer extension using specific primers homologous to 5 of the HPIV3 genes as indicated in each panel was done in the presence or absence of dideoxynucleotides. The dideoxynucleotide used in each of the reactions is indicated above the corresponding lane. N indicates no dideoxynucleotide was included. The primer extension reactions were electrophoresed on a 8M urea-8% polyacrylamide gel.

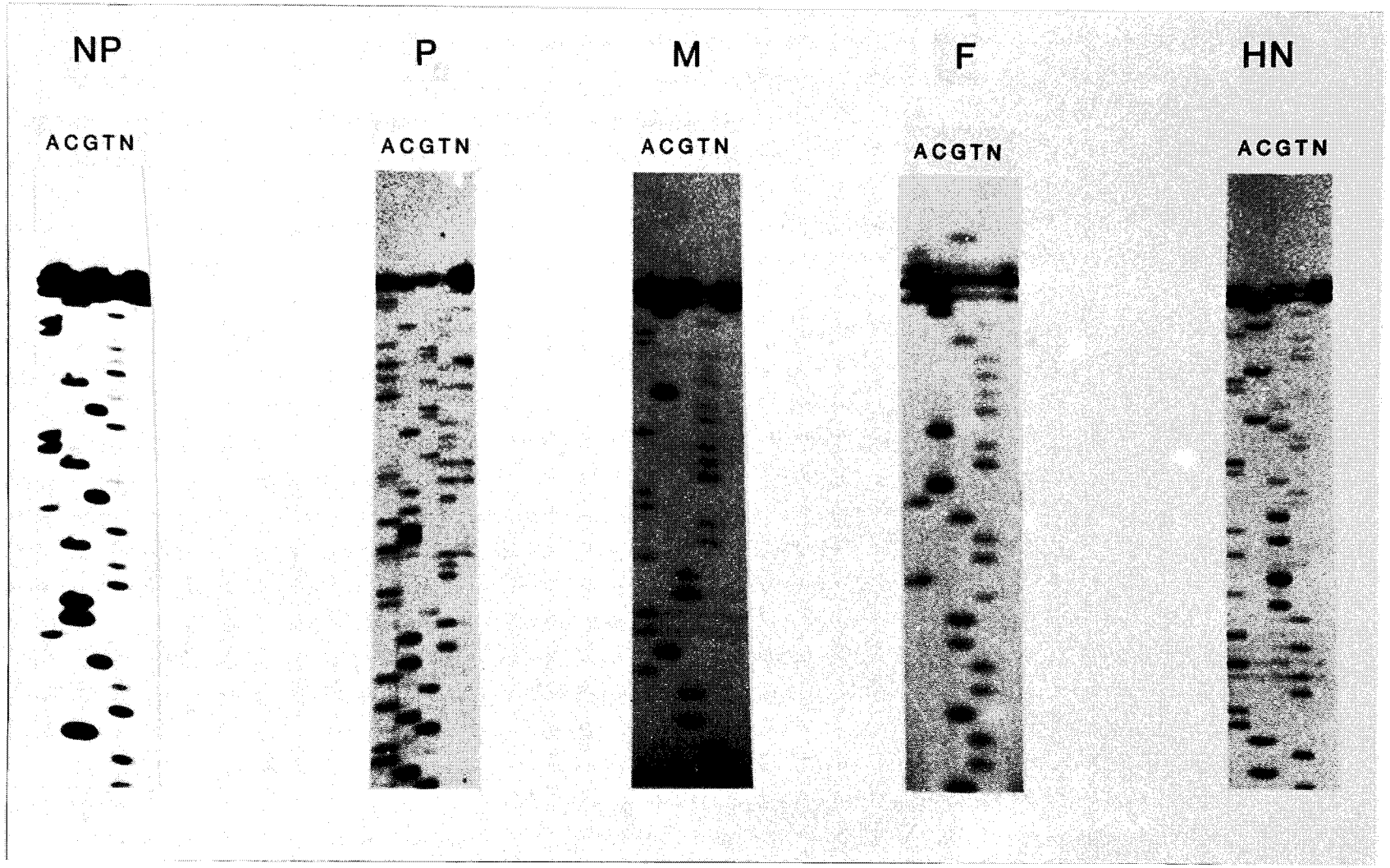

NP
ACGTN
P
ACGTN
M
ACGTN
F
ACGTN
HN
ACGTN

Figure 5. Identification of a putative leader-NP mRNA species. Primer extension using primers specific for the HPIV3 NP and M mRNAs was done in the presence or absence of dideoxynucleotides as described in the legend of Figure 4 except the exposure time which was longer in this case.

ACGTN ACGTN

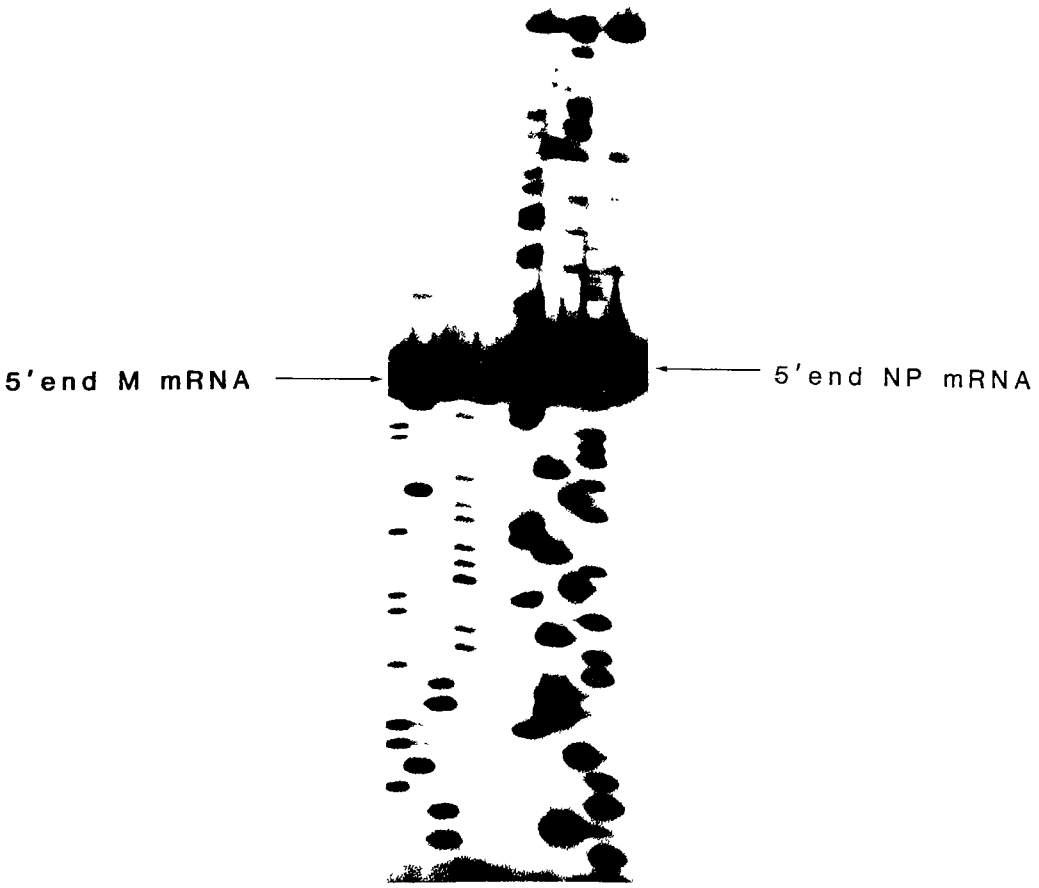
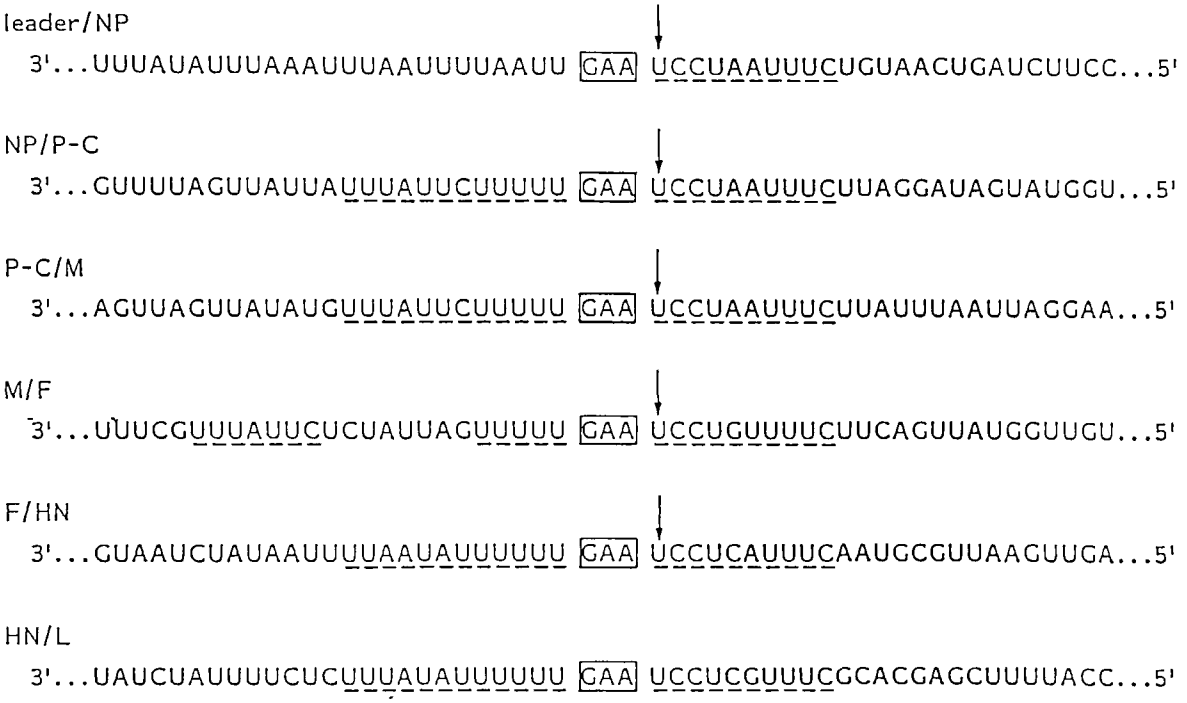


Figure 6. Nucleotide sequence of each of the HPIV3 gene boundaries. Nucleotide sequence is shown in genomic RNA sense (-). The putative polymerase stop and start sequences are indicated by dotted lines, the intergenic sequences are boxed and arrows indicate mRNA starts found by primer extension.


leader/NP
3'...UUUAUAUUUAAAUUUAAUUUUAAUU GAA UCCUAAUUUCUGUAACUGAUCUUCC...5'
NP/P-C
3'...GUUUUAGUUAUUAUUUAUUCUUUUU GAA UCCUAAUUUCUUAGGAUAGUAUGGU...5'
P-C/M
3'...AGUUAGUUAUAUGUUUAUUCUUUUU GAA UCCUAAUUUCUUAUUUAAUUAGGAA...5'
M/F
3'...UUUCGUUUAUUCUCUAUUAGUUUUU GAA UCCUGUUUUCUUCAGUUAUGGUUGU...5'
F/HN
3'...GUAAUCUAUAAUUUUAAUAUUUUUU GAA UCCUCAUUUCAAUGCGUUAAGUUGA...5'
HN/L
3'...UAUCUAUUUUCUCUUUAUAUUUUUU GAA UCCUCGUUUCGCACGAGCUUUUACC...5'

NP-P/C sequence was analyzed from a fragment of mpPIgX (Dimock, K., unpublished observation); the P/C-M region was sequenced by using a subclone of genomic clone mpPIgMP (Prinoski, K.A., unpublished observation); the M-F region was determined by sequencing a subclone of mpPIgX, mpPIgX3 (Cote *et al.*, 1987); the genomic clone $mpPIg14_1$ contained the F-HN region (Cote *et al.*, 1987); clones $mpPIg10_6$ and $mpPIg10_7$ provided the HN-L sequences (Cote *et al.*, 1987; Storey *et al.*, 1987). The leader-NP sequence was obtained by chemically sequencing the 3'-end of HPIV3 genomic RNA labeled with cytidine 3',5'-[5'-^{32}p] bis phosphate and T_4 RNA ligase (Dimock *et al.*, 1986). Dimock *et al.* (1986) also obtained the leader-NP sequence by analysing sequences cloned into M13mp10 following polyadenylation of HPIV3 genomic RNA and oligo(dT)-primed cDNA synthesis.

DISCUSSION

In order to better understand HPIV3 genome organization, the first two objectives of this project were to elucidate the gene order of HPIV3 as well as to determine the gene boundary sequences. A comparison of paramyxoviruses genome organization should increase our knowledge concerning the taxonomic relationships within this family of viruses.

1. HPIV3 gene order

Non-segmented negative stranded RNA viruses have the feature of transcribing discrete mRNAs for each gene instead of a full length polycistronic + sense transcript. We exploited this feature, in the sense that it allowed construction of a cDNA library containing clones

corresponding to individual HPIV3 genes. The cross-hybridization experiments provided a way to group the clones of this mRNA-derived cDNA library according to the corresponding HPIV3 gene. 5 different groups were obtained and a representative clone was identified for each group. A viral protein product for each of the groups was assigned by *in vitro* translation of hybrid-selected mRNAs (Storey, 1987; Dimock *et al.*, 1987). Table 1 shows the number of clones obtained for each gene group. If these numbers reflect the relative amount of each mRNA in infected-cells, it can be concluded that NP message is more abundant than the other viral mRNAs. In fact, Northern blots containing polyadenylated RNA isolated from HPIV3 infected cells and probed with sequences representing each of the genes, showed that the signal corresponding to the NP message is the strongest. The absence of L clones in this clone library may reflect low levels of L mRNA.

To map the genes on the HPIV3 genome, viral specific cDNA clones spanning more than one gene were needed. The idea was to screen a genomic library with probes corresponding to each gene and observe which genes shared boundaries, that is, were adjacent. Table 2 shows the results of grouping the genomic cDNA clones by hybridization. The number of clones containing sequences from more than one gene was not very high. This could be explained by the fact that probes used for this experiment were nick-translated HPIV3 cDNA clones derived from mRNA following oligo(dT) priming. These clones contain an average of approximately 1 Kbp of sequence from the 5'-end (genome sense) of each HPIV3 gene. Therefore, identification of any genomic clone that did not have several hundred bp of sequence from the 3'-end (genome sense) of the adjacent upstream gene would not be possible with these probes.

However, 3 clones hybridized to both P/C and M probes and show that these 2 genes are next to each other. These results can be explained by the fact that the M gene is 1150 bp in length and the M cDNA probe was 1020 bp in length, therefore detection of clones covering the P-M junction was quite efficient. Clone mpPIgX, which was 4.5 Kbp in length and included NP-P/C-M-F sequences, was analyzed by hybridization following restriction endonuclease digestion to elucidate which genes were next to each other. Figure 3 shows that the restriction fragment of 1.7 Kbp hybridized to both M and F probes. Figure 2 also shows that M and F probes identified a common mRNA species which is believed to be an M-F polycistronic readthrough transcript and is observed frequently in HPIV3-infected cells (Storey, 1987; Dimock *et al.*, 1987). Together, these observations indicate that the M gene lies between the P and F genes and the NP gene has to be next to the P/C gene or the F gene. Sequence analysis showed that genomic clone mpPIg14$_1$ spanned the F and HN boundary (Cote *et al.* 1987; Storey *et al.* 1987), which placed the HN gene next to the F gene, therefore the NP gene must be adjacent to the P gene. 3' terminal clones synthesized using polyadenylated HPIV3 genomic RNA as a template and oligo(dT) as a primer showed hybridization to NP clones (Dimock, K., unpublished observation). These results demonstrated that the NP gene is the 3' proximal gene on the HPIV3 genome. Because we did not have a probe representing the L protein gene, screening for a HN-L clone was not possible, however sequence analysis of two genomic clones which hybridized with the HN probe, showed sequences which extended beyond the boundary of the HN gene into another gene, assumed to be the L gene (Storey *et al.*, 1987). The conclusion can be drawn from these data that the HPIV3 gene order (genome sense) is 3'-NP-P/C-M-F-

HN-L-5'. This result is identical to the gene order published by Spriggs and Collins (1986a) for HPIV3. The HPIV3 gene order is also similar to the gene order determined by other laboratories for other members of the family *Paramyxoviridae*: bovine parainfluenza virus 3 (Sakai *et al.*, 1987; Suzu *et al.*, 1987), Sendai virus (Gupta and Kingsbury, 1984), measles virus (Richardson *et al.*, 1985; Hasel *et al.*, 1987), canine distemper virus (Russell *et al.*, 1985), Newcastle disease virus (Collins *et al.*, 1980; Ishida *et al.*, 1986; Wilde *et al.*, 1986; Sato *et al.*, 1987a,b; Yusoff *et al.*, 1987), simian virus 5 (Paterson *et al.*, 1984a) and mumps virus (Elango *et al.*, 1988).

2. HPIV3 gene boundary organization

Specific priming of cDNA synthesis near the 5'-end of HPIV3 mRNA allowed us to locate the 5' termini of the HPIV3 NP, P, M, F and HN genes. The strong termination signals seen for each gene in Figure 4 are labeled in Figure 6 with arrows and are believed to represent the 5' terminus of each gene. The F gene primer extension reactions exhibited sequence extending beyond the strong termination signal. With a longer exposure, it was possible to recognize nucleic acid sequence of the M gene . This M-F primer extension readthrough together with the quantity of M-F transcript detected by Northern blot hybridization, suggest that the polymerase termination signal at the end of M is only about 50% efficient at terminating transcription. What is more surprising though, was the observation of a substantial amount of readthrough during the NP primer extension. After a prolonged exposure to X-ray film, the NP primer extension reactions showed a sequence of approximately 40-50 nucleotides beyond the 5' terminus of the NP gene. The possibility that

this sequence represents synthesis from an antigenome template was ruled out since the primer extension reactions done using the same amount of primers specific for genes other than NP (with the exception of F gene) did not exibit a signal beyond the termination site. However, the sequence which could be read suggested that this represents the leader. Sequencing this region, which is presumed to be the + strand leader, was attempted without much success. Failure to sequence this area may be due to a high degree of secondary structure. The polymerase failure to recognize a stop signal at the end of the leader sequence during mRNA transcription is one explanation for the presence of molecules with both NP and leader sequences. Once the polymerase eventually recognizes the stop signal at the end of NP, termination of the mRNA would occur resulting in a leader-NP mRNA species. Another possibility could be a malfunction of the mechanism that prevents the polymerase from recognizing the stop signal at the end of the NP gene during replication, resulting in a premature termination of antigenome synthesis. An RNA species corresponding in size to the NP message and hybridizing with an NP probe has been detected in HPIV3 virions (Murphy, D.G., personal communication). The leader-NP RNA molecules containing signals for packaging may be encapsidated and packaged into virions.

The gene boundary sequences in Figure 6 are the same as those described by Spriggs and Collins (1986a). The only differences are not at the level of consensus or intergenic sequences, but in the non-translated region of the L gene. This region was not directly sequenced from mRNA which may mean that our cDNA clone covering this region is not representative of the viral RNA pool. Alternatively, differences emerged between the two viruses after passage in each laboratory. The consensus

sequence found at the beginning of each viral gene is conserved, 3'-UCCUNUUUUC-5'. This sequence is believed to be the polymerase transcriptional start signal (Gupta and Kingsbury, 1984). The consensus sequence found at the beginning of each of the HPIV3 genes is very similar to the Sendai virus consensus sequence, 3'-UCCCANUUUC-5' (Shioda *et al.*, 1983; Gupta and Kingsbury, 1984) as well as to that of bovine parainfluenza virus 3, 3'-UCCUNNUUNC-5' (Sakai *et al.*, 1987). Conserved consensus sequences at the beginning of genes are seen for other paramyxoviruses: 3'-UGCCCAUC-5' for Newcastle disease virus (Jorgensen *et al.*, 1987; McGinnes *et al.*, 1987; Sato *et al.*, 1987a,b; Yusoff *et al.*, 1987); and 3'-CCCCGUUUA-5' for respiratory syncytial virus (Collins *et al.*, 1986); but sequences at the beginning of genes are not as well conserved within the genomes of several other members of the family: 3'-U(U/C)C(G/U)(G/U)N(C/U)(U/C)U-5' for mumps virus (Elango *et al.*, 1988) and 3'-UCC(U/C)NN(G/U)A(U/C)C(U/A)-5' for measles virus (Cattaneo *et al.*, 1987). Paramyxoviruses can be classified in three different groups based on the degree of conservation of consensus sequences within the genome: Newcastle disease virus and respiratory syncytial virus possess consensus sequences which do not vary; HPIV3, bovine parainfluenza virus 3 and Sendai virus have consensus sequences that show some variation from one gene to another; mumps virus and measles virus consensus sequences show a high degree of variation from gene to gene.

A putative polymerase stop signal is found at the end of Sendai virus genes (Gupta and Kingsbury, 1984) and we found a similar consensus sequence, 3'-UUNAUNNUUUUU-5', at the 5'-end of HPIV3 genes, identical to the sequence of Spriggs and Collins (1986a). The U rich

region is believed to act as a signal for polyadenylation. Our laboratory, (Cote *et al.*, 1987; Prinoski *et al.*, 1987) as well as others, made the observation that the consensus sequence at the end of the M gene is disrupted by 8 nucleotides (Spriggs and Collins, 1986a; Galinski *et al.*, 1987a; Luk *et al.*, 1987; Spriggs *et al.*, 1987a). It is believed that this interruption in the putative stop signal would promote the production of the M-F readthrough transcripts seen on Northern blots. Once more the 5'-end consensus sequence of HPIV3 is most similar to the corresponding sequences in Sendai virus, 3'-UNAUUCUUUUU-5' (Gupta and Kingsbury, 1984; Shioda *et al.*, 1983) and bovine parainfluenza virus 3, 3'-NUNNUUUUU-5' (Sakai *et al.*, 1987). Consensus sequences are also found in other paramyxoviruses: 3'-ANUCU$_{5-7}$-5', for Newcastle disease virus (Jorgensen *et al.*, 1987; McGinnes *et al.*, 1987; Sato *et al.*, 1987a,b; Yusoff *et al.*, 1987); for respiratory syncytial virus, there are two types of consensus sequences, 3'-UCANUANAUUUU-5' or 3'-UCAAUNNNNUUUU-5' (Collins *et al.*, 1986); for mumps, 3'-ANNUNU$_{6-7}$-5' (Elango *et al.*, 1988). The putative polymerase stop signal does not seem to be very well conserved among members of the *Paramyxoviridae* family, the only common features of these consensus sequences are that the sequence is A and U rich and that there is a poly U tract.

HPIV3 intergenic sequences, GAA, are well conserved within the genome. Spriggs and Collins (1986a) confirmed our results with one exception, they determined the HN-L intergenic sequence to be GAN. However, Galinski *et al.* (1988) found a GAA for HN-L. Galinski *et al.* (1988) observed the sequence GAT at the end of the L gene just before the leader sequence. HPIV3, bovine parainfluenza virus 3 (GAA, Sakai *et al.* 1987), Sendai virus (GAA or GGG, Gupta and Kingsbury, 1984) and

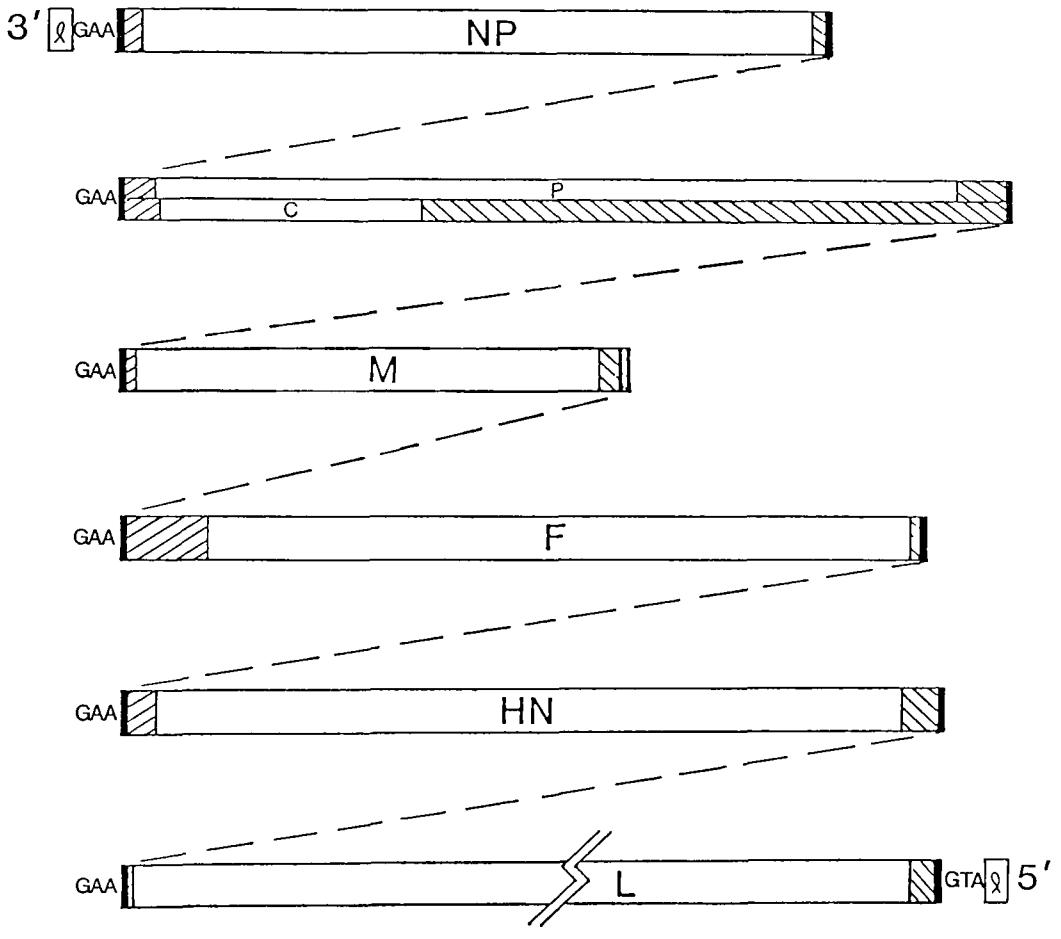
measles virus (GAA or GCA, Cattaneo *et al.*, 1987) possess trinucleotide intergenic sequences. However, other paramyxoviruses have intergenic sequences which exibit no consensus sequence and differ in length: from 1 to 7 nucleotides for mumps virus (Elango *et al.*, 1988); from 1 to 52 nucleotides for respiratory syncytial virus (Collins *et al.*, 1986); for simian virus 5, either 2 or 4 nucleotides (Hiebert *et al.*, 1985); and from 1 to 48 nucleotides for Newcastle disease virus (Ishida *et al.*, 1986; Sato *et al.*, 1987a, b). From these comparisons, the *Paramyxoviridae* family can be divided into two groups; those with conserved trinucleotide intergenic sequences, for example HPIV3, bovine parainfluenza virus 3, Sendai virus and measles virus, and those with variable intergenic sequences, for example, mumps virus, simian virus 5, Newcastle disease virus and respiratory syncytial virus.

3. HPIV3 genome organization

In the last few years, interest in the molecular biology of HPIV3 has led to complete sequence determination of the genome. Each gene, with the exception of the L gene, has been sequenced by 3 or 4 different groups. The gene order and the gene boundary data in combination with sequence analysis of HPIV3 genes, obtained by our laboratory and others, has resulted in elucidation of the HPIV3 genome organization as depicted in Figure 7.

At the 3' terminus (genome sense) of the HPIV3 genome, there is a sequence which is similar to leader sequences normally found at the 3'-end of linear negative sense single-stranded RNA virus genomes (Colonno and Banerjee, 1976; Leppert *et al.*, 1979; Dowling *et al.*, 1983; Re *et al.*, 1983; Shioda *et al.* 1983; Billeter *et al.*, 1984; Kurilla *et al.*, 1985; Dimock

Figure 7. Organization of the HPIV3 genome. The HPIV3 coding sequence for each gene is represented by the open boxes. Hatched boxes indicate the untranslated sequences, the black boxes correspond to the putative transcriptional stop and start consensus sequences and the open boxes at each end of the genome, with the letter *l* inside, represent the leader sequences. The intergenic sequences are represented by letters corresponding to their sequence.


3′
GAA
NP
GAA
C
P
GAA
M
GAA
F
GAA
HN
GAA
L
GTA
5′

et al., 1986; Ishida *et al.*, 1986; Sakai *et al.*, 1987; Elango *et al.*, 1988). So far, the leader sequence of paramyxoviruses (Sendai virus, measles virus, Newcastle disease virus, human and bovine parainfluenza virus 3, mumps virus) are all 55 nucleotides in length, with the exception of measles virus, which has a 56 nucleotide leader.

The NP gene is the 3' proximal gene. The open reading frame encodes a nucleocapsid (NP) protein of 515 amino acids with a predicted molecular weight of 57,800 (Galinski *et al.*, 1986a; Jambou *et al.*, 1986; Sanchez *et al.*, 1986). The NP gene is followed by the gene coding for the polymerase-associated or phosphorylated (P) protein. Sequence analysis of mRNA-derived cDNA clones of the P gene of HPIV3, revealed the presence of two overlapping reading frames (Galinski *et al.*, 1986b; Luk *et al.*, 1986; Spriggs and Collins, 1986b). The largest open reading frame encodes the P protein of 602 amino acids with a predicted molecular weight of 67,500. There is a second open reading frame which appears to encode a protein of 199 amino acids with a predicted molecular weight of 23,300. The translation initiation codon of this second open reading frame is 7 nucleotides downstream from the first AUG and would start translation in a +1 reading frame. This protein is believed to be the counterpart of the non-structural C protein found in some paramyxoviruses. The matrix or membrane (M) protein gene is next on the HPIV3 genome. The open reading frame encodes a protein of 353 amino acids with a predicted molecular weight of 39,500 (Galinski *et al.*, 1987a; Luk *et al.*, 1987; Prinoski *et al.*, 1987; Spriggs *et al.*, 1987a). As shown in Figure 6, the consensus gene end sequence in the M gene is disrupted by 8 nucleotides which are believed to be responsible for frequent transcription through the M/F junction. The next gene on the HPIV3

genome is the fusion (F) protein gene which encodes a protein of 539 amino acids with a predicted molecular weight of 60,031 (Spriggs *et al.*, 1986; Cote *et al.*, 1987; Galinski *et al.*, 1987b). Unlike the mRNAs of other HPIV3 genes, the mRNA transcribed from the F gene possesses a long 5' non-translated region. This region may be a regulatory sequence responsible for the low levels of fusion protein synthesis (Storey *et al.* 1984; Storey, 1987). The hemagglutinin-neuraminidase (HN) glycoprotein gene is immediately downstream of the F gene and has an open reading frame which encodes a protein of 572 amino acids with a predicted molecular weight of 64,200 (Elango *et al.*, 1986; Galinski *et al.*, 1987b; Storey *et al.*, 1987). The last, but not the least, gene on the HPIV3 genome, is the large or polymerase (L) protein gene. The open reading frame is predicted to encode a protein of 2,233 amino acids with a predicted molecular weight of 256,000 (Galinski *et al.*, 1988). Downstream of the L gene, at the extreme 5'-end of the HPIV3 genome, there is a sequence of 44 nucleotides in length which is very similar to the complement of the leader sequence found at the 3'-end of the HPIV3 genome. The leader sequence is believed to be involved in initiation of transcription and replication of the HPIV3 genome (Galinski *et al.* 1988).

4. Comparison of the organization of the HPIV3 genome with other paramyxoviruses

The HPIV3 gene order deduced from the cloning results is very similar to the order found for many of the other members of the family, ***Paramyxoviridae***, 3'-NP-P/C(orV)-M-F-HN-L-5'. However, there are differences for some members of the family. Respiratory syncytial virus, for instance, has 4 additional genes , although the overall gene order is

basically the same, 3'-1C-1B-N-P-M-1A-G-F-22K-L-5' (Collins *et al.*, 1986). Two other members possess an extra gene, named SH, which codes for a small hydrophobic protein. This extra gene is situated between the F and HN genes on the mumps virus and simian virus 5 genomes (Hiebert *et al.*, 1985; Elango *et al.*, 1988). The simian virus 5 SH gene contains an open reading frame which encodes a polypeptide of 44 amino acids with a molecular weight of 5,000. The protein is found in simian virus 5 infected-cells but its function is unknown (Hiebert *et al.*, 1985).

Other specific differences are seen in the genome organization of paramyxoviruses. Polycistronic mRNAs are unusual in eucaryotic cells. However, several paramyxovirus use different strategies to synthesize 2 different proteins from the P/C (or V) gene. HPIV3, bovine parainfluenza virus 3, Sendai virus, measles virus and canine distemper virus appear to share the same strategy: translation of the C protein initiates independently, several nucleotides downstream of the initiation site for the P protein in a + 1 reading frame (Giorgi *et al.*, 1983; Barrett *et al.*, 1985; Bellini *et al.*, 1985; Galinski *et al.*, 1986b; Luk *et al.*, 1986; Spriggs and Collins, 1986b; Sakai *et al.*, 1987). The respiratory syncytial virus P gene does not code for any protein other than the P protein (Morrison, 1988). However, in the case of Newcastle disease virus, simian virus 5 and mumps virus, the issue seems to be more complicated. On the basis of peptide maps, it was concluded that the protein C (or V) was translated from the same open reading frame as P (Collins *et al.*, 1982; Herrler and Compans, 1982; Paterson *et al.*, 1984a). Sequence analysis of an avirulent strain of Newcastle disease virus shows only one other open reading frame downstream from the first AUG in a -1 frame (Sato *et al.*, 1987a). Morrison (1988) described monoclonal antibodies specific for the Newcastle

disease virus P protein that would precipitate, *in vitro,* translated products corresponding to C or V. However, Thomas *et al.* (1988) demonstrated that simian virus 5 transcribes two P/C (or V) specific mRNA species which differ only by 2 non-templated G's. These extra G's allow for translation of the P protein from one mRNA species while the mRNA transcribed, without modification, from the P gene codes for the C (or V) protein. The two proteins P and C (or V) of simian virus 5 share N-terminal sequences. It is possible to speculate that the same strategy may be used by mumps virus and Newcastle disease virus.

This information as well as the sequence similarity among different paramyxovirus genes leads us to reassess the actual taxonomic classification of this family. As described before the family *Paramyxoviridae* is divided into 3 genera, *Paramyxoviruses, Morbilliviruses* and *Pneumoviruses*. This classification was based on structural features and the properties of the attachment proteins of these viruses (Matthews, 1982). This was not a bad choice since the *Pneumovirus,* respiratory syncytial virus is very different from the others and should be in a genus by itself. However, respiratory syncytial virus with its 10 genes, seems to be as different from the other paramyxoviruses as the rhabdovirus, vesicular stomatitis virus. Perhaps respiratory syncytial virus should be in a different family. The *Morbilliviruses*, measles and canine distemper virus, have been shown to have a high degree of similarity and maybe the lack of the neuraminidase activity would keep them in a separate genus. Where the classification needs to be reconsidered is within the genus *Paramyxovirus*. HPIV3, bovine parainfluenza virus 3, Sendai virus share conserved trinucleotides intergenic sequences, the absence of an SH gene, a similar strategy for P/C (or V) expression. This would put simian virus

5, mumps virus and Newcastle disease virus in a different category, perhaps a new genus. A summary of these different particularities is shown in Table 3. There are still too many unknowns to classify these 3 viruses together.

TABLE 3

GENOME ORGANIZATION OF THE FAMILY *PARAMYXOVIRIDAE*

genus	virus	number of genes	length of intergenic sequences	P/C(orV) gene organization
Morbillivirus	measles	6	3	2 ORF[7]
Morbillivirus	CDV[1]	6	3	2 ORF
Paramyxovirus	HPIV3[2]	6	3	2 ORF
Paramyxovirus	BPIV3[3]	6	3	2 ORF
Paramyxovirus	Sendai	6	3	2 ORF
Paramyxovirus	NDV[4]	6	variable	?
Paramyxovirus	Mumps	7	variable	?
Paramyxovirus	SV5[5]	7	variable	editing[8]
Pneumovirus	RSV[6]	10	variable	1 ORF[9]

1 Canine distemper virus
2 Human parainfluenza virus 3
3 Bovine parainfluenza virus 3
4 Newcastle disease virus
5 Simian virus 5
6 Respiratory Syncytial virus
7 Open reading frame
8 The gene transcribes 2 specific mRNA species which differ by 2 non-templated G's
9 Absence of C (orV) protein

CHAPTER 4 SEQUENCE ANALYSIS OF THE HPIV3 FUSION PROTEIN GENE.

INTRODUCTION

HPIV3 possesses on its envelope a glycoprotein, the fusion protein, which mediates fusion between the viral and cellular membranes. This protein is also responsible for cell to cell spread of the virus. The fusion protein is synthesized as an inactive precursor, F_0, which must be cleaved by a host trypsin-like enzyme to yield an active disulfide-linked molecule, $F_{1,2}$ (Chapter 1).

Because of the importance of the HPIV3 fusion protein in pathogenesis (Chapter 1), the third objective of this project was to determine the complete nucleotide sequence of the F gene and the fourth objective was to compare the deduced amino acid sequence of the fusion protein of HPIV3 with the fusion protein sequences of other paramyxoviruses. This will help to better understand the evolutionary relationships between different species of paramyxoviruses and identify conserved sequences among paramyxovirus fusion proteins that are likely to be important for its structure and/or function.

Four overlapping clones covering the entire HPIV3 fusion protein gene and its flanking regions were used for sequence analysis. A sequential series of overlapping clones was generated (Dale *et al.*, 1985) and the resulting subclones were sequenced using the chain termination method (Sanger *et al.*, 1977).

In this chapter, the complete nucleotide sequence of the HPIV3 F gene will be presented along with the predicted amino acid sequence. The hydropathy profile and the location of cysteine residues, proline residues and potential glycosylation sites in the HPIV3 fusion protein will be compared with the same features of other paramyxovirus fusion proteins. The significance of fusion protein sequence similarities will also be discussed.

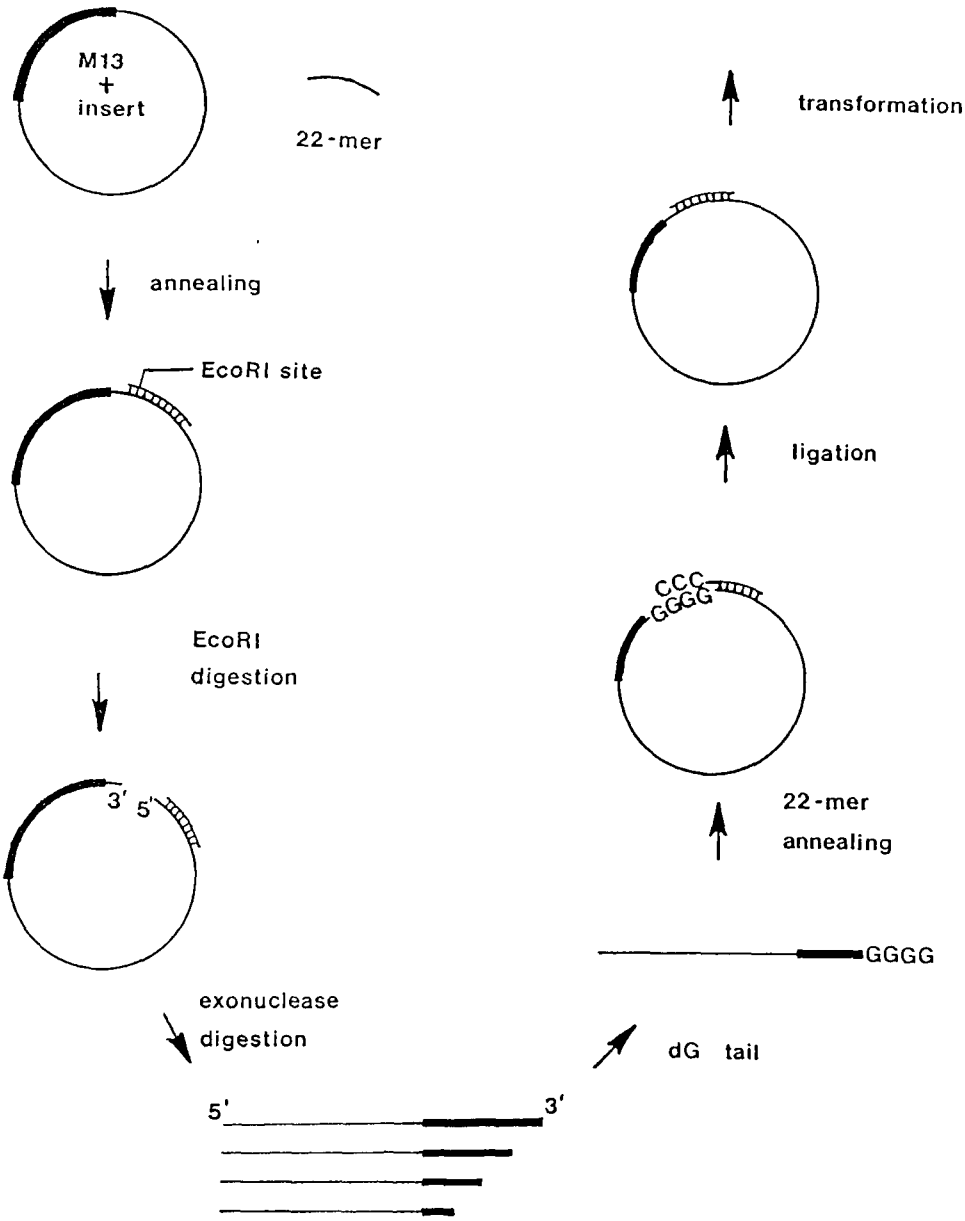
RESULTS

1. Production of a sequential series of overlapping clones for sequencing

A particular advantage of the M13 bacteriophage system for recombinant work and sequence analysis is its property of encapsidating single-stranded DNA into phage particles which can be easily recovered in the growth medium of infected *E. coli*. The M13 single-stranded genome carrying the insert of choice can then be used as a template for cDNA synthesis with the Klenow enzyme and chain termination inhibitors in order to analyse the sequence of an insert. cDNA synthesis is initiated with a 17-mer primer which hybridizes to a M13 region just upstream of the insert. Under these conditions, nucleotide sequence of foreign DNA can be analysed for up to 300-400 nucleotides. Subcloning fragments of large inserts is usually necessary for complete sequence analysis. Dale *et al.* (1985) described a rapid technique to produce a sequential series of overlapping clones using M13 single-stranded DNA. The principle of producing sequential overlapping clones as applied in this thesis is shown in Figure 8.

A 22-mer oligonucleotide was annealed to M13 single-stranded DNA.

Figure 8. Strategy for the production of a sequential series of overlapping clones for DNA sequencing. The solid bar represents the insert. Adapted from Dale *et al.* (1985).

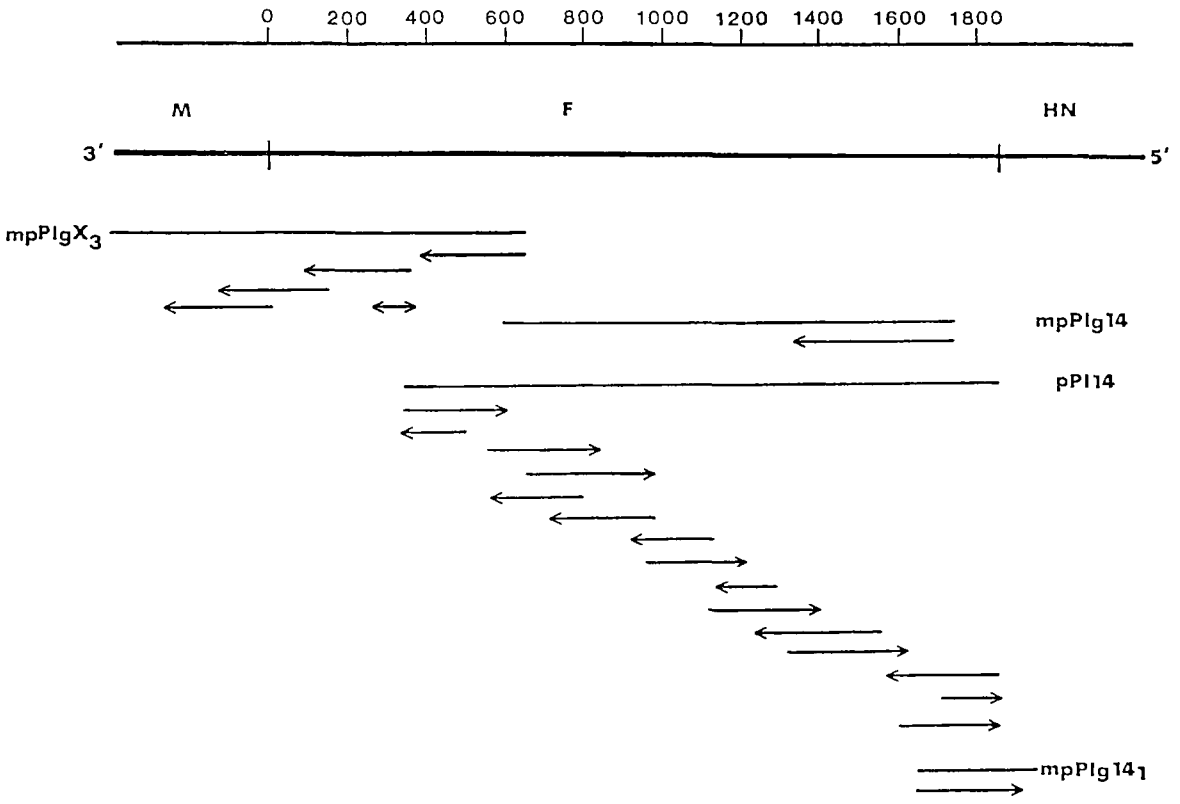

M13
+
insert
22-mer
annealing
EcoRI site
EcoRI
digestion
3'
5'
exonuclease
digestion
5'
3'
dG tail
GGGG
22-mer
annealing
CCC
GGGG
ligation
transformation

The annealing of the primer generates an *Eco*RI endonuclease site in the polylinker region near the 3'-end of the cloned insert. After digestion of the DNA with *Eco*RI, within the annealed region, the 3' to 5' exonuclease activity of T4 DNA polymerase was used to remove nucleotides from the 3'-end of the insert. Stopping the exonuclease activity at different time points generated deletions of different lengths. An oligo (dG) tail was then added to the 3'-end of the insert and a second annealing was done with the 22-mer oligonucleotide (which possesses an oligo (dC) tail). The 3'-end of the insert anneals with the intact 5'-end of the M13 polylinker site and brings 5' and 3'-ends of the M13 DNA together. After ligation, the M13 single-stranded DNA was transformed into *E. coli* (JM101) and the DNA isolated from resulting plaques was screened by size and then sequenced.

2. Sequencing strategy of the HPIV3 fusion protein gene

Four overlapping clones covering the HPIV3 F gene and flanking regions were used for sequencing. As shown in Figure 9, mpPIgX$_3$, a subclone of mpPIgX, covers the M-F junction as well as the 3'-end (genome sense) of the gene. pPI14 and mpPIg14 correspond to most of the F gene up to the 5'-terminus. mpPIg14$_1$ covers the F-HN junction. mpPIgX$_3$, pPI14 and mpPIg14$_1$ were cloned into M13mp19 in both orientations and sequential series of overlapping clones were produced for both orientations in order to sequence as much of the F gene as possible in both directions. This strategy is shown in Figure 9. The chain-terminating inhibition method of Sanger *et al.* (1977) adapted to the M13 system by Messing *et al.* (1983) was used for sequencing the HPIV3 fusion protein gene.

Figure 9. Strategy for sequencing the HPIV3 F gene (Cote *et al.*, 1987). The positions of cDNA clones synthesized from HPIV3 F mRNA (pPI14) and genomic RNA (mpPIgX3, mpPIg141 and mpPIg142) are shown. The numbers are base pairs. Arrows indicate the direction of sequencing and the approximate number of nucleotides sequenced in each subclone constructed by the sequential deletion method of Dale *et al.* (1985).


0
200
400
600
800
1000
1200
1400
1600
1800
M
F
HN
3'
5'
mpPIgX3
mpPIg14
pPI14
mpPIg141

3. HPIV3 fusion protein gene: nucleotide sequence and predicted amino acid sequence (Cote *et al.*, 1987)

All the sequencing data obtained were analysed using the International Biotechnologies Inc. software. The complete nucleotide sequence of the fusion protein gene of HPIV3 is shown in Figure 10, in the mRNA sense. The F gene is 1851 nucleotides long including the A residue immediately following the M-F intergenic sequence and the 6 A residues preceding the F-HN intergenic sequence. The dotted lines represent the consensus putative polymerase stop and start sequences at each end of HPIV3 genes, as discussed in Chapter 3. The 8 nucleotides which interrupt the putative polymerase stop sequence of the M gene are also shown. The intergenic trinucleotides (CTT in mRNA sense) are located between the putative polymerase stop and start sequences at the M-F and F-HN boundaries.

The first ATG is located between positions 194-196 and begins a large open reading frame of 1617 nucleotides. A polypeptide of 539 amino acids is predicted to be encoded by this large open reading frame. No other open reading frame of more than 114 nucleotides starting with an ATG was found. The amino acid sequence predicted from the large open reading frame is also shown in Figure 10. The estimated molecular weight of the unglycosylated product encoded by the HPIV3 F gene is 60,031.

A hydropathy plot of the predicted amino acid sequence of the fusion protein shows 3 principal regions of hydrophobicity. These regions are underlined in Figure 10. The first major hydrophobic domain is believed to be the signal peptide (amino acids 1 to 18) and the site for cleavage of this peptide is thought to occur between Cys18 and Glu19 by analogy with the Sendai virus fusion protein, where the cleavage site is between

Figure 10. The complete nucleotide sequence of the HPIV3 F gene and flanking regions (Cote *et al.*, 1987). The cDNA sequence is shown in the mRNA strand sense (+) in the 5' to 3' direction. The sequence includes the intergenic sequences (CTT) between the M and F genes and also the F and HN genes. The putative transcriptional initiation sites of the F and HN and the putative transcriptional termination site and polyadenylation signals of the M and F genes are underlined (---). The predicted amino acid sequence of the unglycosylated polypeptide, F_0, is also shown. The arrow indicates the cleavage site between F_1 and F_2 (Spriggs *et al.*, 1986). The asterisk indicates the translational initiation site and the three major hydrophobic sequences are underlined(___): A, the signal peptide; B, the F_1 amino terminus; C, the transmembrane region. Potential glycosylation sites are boxed.

```
AGCAAATAAGAGATAATCAAAAACTTAGGACAAAAGAAGTCAATACCAACAACTATTAGCAGCCACACTCGCTGGAACAAGAAAGAAGGGATAAAAAAAGT   75

TTAACAGAAGAAACAAAAACAGAAAGCACAGAACACCAGAACAACAAGATCAAAACACCCAACCCATTCAAAACGAAAATCTCAAAAGAGATTGGCAACACAACAAA  182

CACTGAACATC ATG CCA ACC TCA ATA CTG CTA ATT ATT ACA ACC ATG ATT ATG GCA TCT TTC TGC CAA ATA GAT ATC ACA AAA  265
            Met Pro Thr Ser Ile Leu Leu Ile Ile Thr Thr Met Ile Met Ala Ser Phe Cys Gln Ile Asp Ile Thr Lys   24

CTA CAG CAT GTA GGT GTA TTG GTT AAC AGT CCC AAA GGG ATG AAG ATA TCA CAA AAC TTT GAA ACA AGA TAT CTA ATT TTG  346
Leu Gln His Val Gly Val Leu Val Asn Ser Pro Lys Gly Met Lys Ile Ser Gln Asn Phe Glu Thr Arg Tyr Leu Ile Leu   51

AGC CTC ATA CCA AAA ATA GAA GAT TCT AAC TCT TGT GGT GAC CAA CAG ATC AAG CAA TAC AAG AGG TTA TTG GAT AGA CTG  427
Ser Leu Ile Pro Lys Ile Glu Asp Ser Asn Ser Cys Gly Asp Gln Gln Ile Lys Gln Tyr Lys Arg Leu Leu Asp Arg Leu   78

ATC ATT CCT TTA TAT GAT GGA TTA AGA TTA CAG AAG GAT GTG ATA GTG TCC AAT CAA GAA TCC AAT GAA AAC ACT GAC CCC  508
Ile Ile Pro Leu Tyr Asp Gly Leu Arg Leu Gln Lys Asp Val Ile Val Ser Asn Gln Glu Ser Asn Glu Asn Thr Asp Pro  105

AGA ACA AAA CGA TTC TTT GGA GGG GTA ATT GGA ACT ATT GCT CTG GGA GTG GCA ACC TCA GCA CAA ATT ACA GCG GCA GTT  589
Arg Thr Lys Arg Phe Phe Gly Gly Val Ile Gly Thr Ile Ala Leu Gly Val Ala Thr Ser Ala Gln Ile Thr Ala Ala Val  132

GCT CTG GTT GAA GCC AAG CAG GCA AGA TCA GAC ATT GAA AAA CTC AAG GAA GCA ATC AGG GAC ACA AAC AAA GCA GTG CAG  670
Ala Leu Val Glu Ala Lys Gln Ala Arg Ser Asp Ile Glu Lys Leu Lys Glu Ala Ile Arg Asp Thr Asn Lys Ala Val Gln  159

TCA GTC CAG AGC TCC ATA GGA AAT TTG ATA GTA GCA ATT AAA TCG GTC CAG GAT TAT GTC AAC AAA GAA ATC GTG CCA TCA  751
Ser Val Gln Ser Ser Ile Gly Asn Leu Ile Val Ala Ile Lys Ser Val Gln Asp Tyr Val Asn Lys Glu Ile Val Pro Ser  186

ATT GCG AGA TTA GGT TGT GAA GCA GCA GGA CTT CAG TTA GGA ATT GCA TTA ACA CAG CAT TAC TCA GAA TTA ACA AAC ATA  832
Ile Ala Arg Leu Gly Cys Glu Ala Ala Gly Leu Gln Leu Gly Ile Ala Leu Thr Gln His Tyr Ser Glu Leu Thr Asn Ile  213

TTC GGT GAT AAC ATA GGA TCG TTA CAA GAA AAA GGG ATA AAA TTA CAA GGT ATA GCA TCA TTA TAC CGC ACA AAT ATC ACA  913
Phe Gly Asp Asn Ile Gly Ser Leu Gln Glu Lys Gly Ile Lys Leu Gln Gly Ile Ala Ser Leu Tyr Arg Thr Asn Ile Thr  240

GAG ATA TTC ACA ACA TCA ACA GTT GAT AAA TAT GAT ATT TAT GAT CTA TTA TTT ACA GAA TCA ATA AAG GTG AGA GTT ATA  994
Glu Ile Phe Thr Thr Ser Thr Val Asp Lys Tyr Asp Ile Tyr Asp Leu Leu Phe Thr Glu Ser Ile Lys Val Arg Val Ile  267

GAT GTT GAC TTG AAT GAT TAC TCA ATC ACC CTC CAA GTC AGA CTC CCT TTA TTA ACT AGA CTG CTG AAC ACC CAG ATT TAC 1075
Asp Val Asp Leu Asn Asp Tyr Ser Ile Thr Leu Gln Val Arg Leu Pro Leu Leu Thr Arg Leu Leu Asn Thr Gln Ile Tyr  294

AAA GTA GAT TCC ATA TCA TAC AAC ATC CAA AAC AGA GAA TGG TAT ATC CCT CTT CCC AGC CAC ATC ATG ACA AAA GGG GCA 1156
Lys Val Asp Ser Ile Ser Tyr Asn Ile Gln Asn Arg Glu Trp Tyr Ile Pro Leu Pro Ser His Ile Met Thr Lys Gly Ala  321

TTT CTA GGT GGA GCA GAT ATC AAA GAA TGT ATA GAA GCA TTC AGC AGT TAT ATA TGC CCT TCT GAT CCA GGA TTT GTA CTA 1237
Phe Leu Gly Gly Ala Asp Ile Lys Glu Cys Ile Glu Ala Phe Ser Ser Tyr Ile Cys Pro Ser Asp Pro Gly Phe Val Leu  348

AAC CAT GAA ATG GAG AGC TGT TTA TCA GGA AAC ATA TCC CAA TGT CCA AGA ACC GTG GTC ACA TCA GAC ATT GTT CCA AGA 1318
Asn His Glu Met Glu Ser Cys Leu Ser Gly Asn Ile Ser Gln Cys Pro Arg Thr Val Val Thr Ser Asp Ile Val Pro Arg  375

TAT GCA TTT GTC AAT GGA GGA GTG GTT GCA AAT TGT ATA ACA ACC ACA TGT ACA TGC AAC GGT ATT GGC AAT AGA ATC AAT 1399
Tyr Ala Phe Val Asn Gly Gly Val Val Ala Asn Cys Ile Thr Thr Thr Cys Thr Cys Asn Gly Ile Gly Asn Arg Ile Asn  402

CAA CCA CCT GAT CAA GGA GTA AAA ATT ATA ACA CAT AAA GAA TGT AAT ACA ATA GGT ATC AAC GGA ATG CTG TTC AAT ACA 1480
Gln Pro Pro Asp Gln Gly Val Lys Ile Ile Thr His Lys Glu Cys Asn Thr Ile Gly Ile Asn Gly Met Leu Phe Asn Thr  429

AAT AAA GAA GGA ACT CTT GCA TTT TAC ACA CCA AAT GAT ATA ACA TTA AAC AAT TCT GTT TCA CTT GAT CCA ATT GAC ATA 1561
Asn Lys Glu Gly Thr Leu Ala Phe Tyr Thr Pro Asn Asp Ile Thr Leu Asn Asn Ser Val Ser Leu Asp Pro Ile Asp Ile  456

TCA ATC GAG CTC AAT AAG GCC AAA TCA GAT CTA GAA GAG TCA AAA GAA TGG ATA AGA AGG TCA AAT CAA AAA CTA GAT TCC 1642
Ser Ile Glu Leu Asn Lys Ala Lys Ser Asp Leu Glu Glu Ser Lys Glu Trp Ile Arg Arg Ser Asn Glu Lys Leu Asp Ser  483

ATT GGA AAT TGG CAT CAA TCT AGC ACC ACA ATC ATA ATT GTT TTG ATA ATG ATA ATT ATA TTG TTT ATA ATT AAT GTA ACG 1723
Ile Gly Asn Trp His Gln Ser Ser Thr Thr Ile Ile Ile Val Leu Ile Met Ile Ile Ile Leu Phe Ile Ile Asn Val Thr  510

ATA ATT ATA ATT GCA GTT AAG TAT TAC AGA ATT CAA AAG AGA AAT CGA GTG GAT CAA AAT GAT AAA CCA TAT GTA TTA ACA 1804
Ile Ile Ile Ile Ala Val Lys Tyr Tyr Arg Ile Gln Lys Arg Asn Arg Val Asp Gln Asn Asp Lys Pro Tyr Val Leu Thr  537

AAC AAA TGA CAGATCTATAGATCATTAGATATTAAAATTATAAAAAACTTAGGAGTAAAGTTACGC                                      1851
Asn Lys ---                                                                                                539
```

Cys25 and Glu26 (Blumberg *et al.*, 1985). The second hydrophobic domain is the N-terminus of F_1 (amino acids 110 to 135). It has been demonstrated that F_0 is cleaved after Arg109 to generate the new F_1 N-terminus (Spriggs *et al.*, 1986). This second hydrophobic domain is preceded by a highly hydrophilic region. The last hydrophobic domain is probably the transmembrane region of the fusion protein (amino acids 494 to 516) preceding a highly charged cytoplasmic tail (amino acids 517 to 539). Four potential N-linked glycosylation sites (Asn-X-Ser/Thr) were identified in the HPIV3 F_0 protein; all of these were found in the F_1 coding region and are boxed in Figure 10. The fourth potential acceptor site (amino acids 508 to 510) is found within the transmembrane region of the HPIV3 F_0 protein and, therefore, may not be used.

DISCUSSION

The third objective of this project was to determine the nucleotide sequence of the HPIV3 F gene and predict the amino acid sequence of the HPIV3 fusion protein. The fourth objective was to compare the HPIV3 fusion protein with those of other paramyxoviruses to have a better understanding of the evolutionary relationships within the family *Paramyxoviridae* and to identify conserved sequences among paramyxovirus F genes that may be crucial for structure and/or function.

1. General comparison of the HPIV3 fusion protein sequence with other published HPIV3 and paramyxovirus fusion protein sequences

The HPIV3 F gene nucleotide sequence and the predicted amino acid sequence are shown in Figure 10 and were described above. Except for

the L gene of HPIV3, the sequences of all of the HPIV3 genes have been published by at least 3 independant groups. Spriggs *et al.* (1986) and Galinski *et al.* (1987b) have also published the nucleotide sequence and the predicted amino acid sequence of the HPIV3 fusion protein gene. Our results (Cote *et al.*, 1987) as well as those of Galinski *et al.* (1987b) confirm the findings of Spriggs *et al.* (1986). There are, overall, a total of 13 nucleotide differences between the 3 published sequences. All the differences are between two sequences only and the third sequence confirms one of the other two. There are 9 nucleotide differences between the sequence shown in Figure 10 and the results of Spriggs *et al.* (1986) and Galinski *et al.* (1987b). Two of the differences (nucleotides 97 and 142) are in the untranslated region preceding the initiation codon. In the coding region, four differences (nucleotides 1096, 1297, 1384 and 1387) do not change the amino acid sequence. The remaining three differences (nucleotides 1077, 1299, 1541) would result in a change of amino acid residue. From these three changes of amino acid residue (Arg/Lys295, Lys/Thr369 and Ala/Ser449) only one would result in a net charge change (Lys/Thr369). One explanation for these differences could be the different passage histories of HPIV3 in the three laboratories. It is also possible that the F gene sequence determined from cDNA clones may not be representative of the pool of viral genomes. However, direct reverse transcriptase dideoxynucleotide sequencing of the HPIV3 F gene using genomic RNA as a template and F gene-specific oligonucleotide primers (Prinoski *et al.*, manuscript in preparation) showed, at the nucleotide level, only 1 difference between the sequence obtained by direct RNA sequencing and the sequence determined from cloned cDNA. This suggests that different passage histories might best explain the

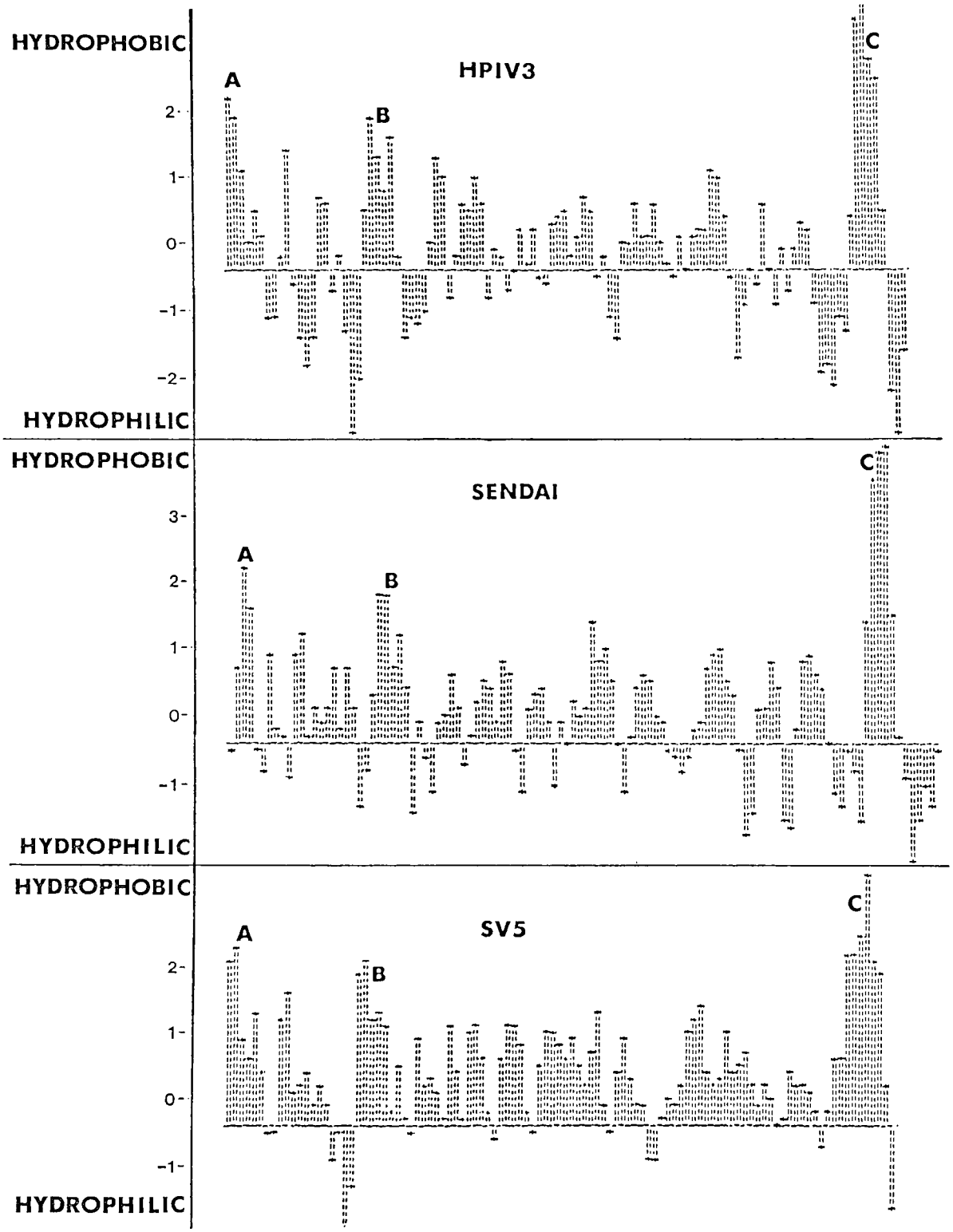
differences seen between the 3 published nucleotide sequences of the HPIV3 F gene.

Recently, the sequences of several paramyxovirus F genes have been published: bovine parainfluenza virus 3 (Suzu *et al.*, 1987), human parainfluenza virus 1 (Merson *et al.*, 1988), different strains of Sendai virus (Blumberg *et al.*, 1985; Hsu and Choppin, 1984; Miura *et al.*, 1985; Shioda *et al.*, 1986), measles virus (Richardson *et al.*, 1986; Buckland *et al.*, 1987), canine distemper virus (Barrett *et al.*, 1987), rinderpest virus (Hsu *et al.*, 1988), simian virus 5 (Paterson *et al.*, 1984b), Newcastle disease virus (Chambers *et al.*, 1986; McGinnes and Morrison, 1986; Sato *et al.*, 1987a), mumps virus (Waxham *et al.*, 1987) and respiratory syncytial virus (Collins *et al.*, 1984; Elango *et al.*, 1985; Baybutt and Pringle, 1987). All 11 fusion protein sequences show common features such as size and location of hydrophobic regions, potential glycosylation sites and cysteine and proline residues. These features will be discussed in more detail in the following sections.

2. Hydrophobicity of the HPIV3 fusion protein and comparison with other paramyxovirus fusion proteins

The hydropathy plot (Kyte and Doolittle, 1982) of the HPIV3 F_0 protein depicts a protein that is generally hydrophobic in nature with 3 highly hydrophobic regions as shown in Figure 11: (A) the signal peptide, (B) the F_1 N-terminus and (C) the transmembrane region. However, there are also 2 regions that are particularly hydrophilic: the C-terminus of F_2 and the cytoplasmic C-terminal tail. Figure 11 also shows a comparison of the hydropathy plots of the F_0 proteins of HPIV3, Sendai virus and simian virus 5. It is clear that these hydrophobic regions have a similar

Figure 11. Hydropathy plots of the F gene of HPIV3, Sendai virus and simian virus 5. The plots represent an average of the hydropathic indices (relative hydrophobic/hydrophilic nature) of 9 consecutive amino acids. The plots of the F_0 proteins of Sendai virus and simian virus 5 were generated from sequence data of Blumberg *et al.* (1985) and Paterson *et al.* (1984b). The letters represent highly hydrophobic regions: A, signal sequence; B, F_1 N-terminus; C, transmembrane region.

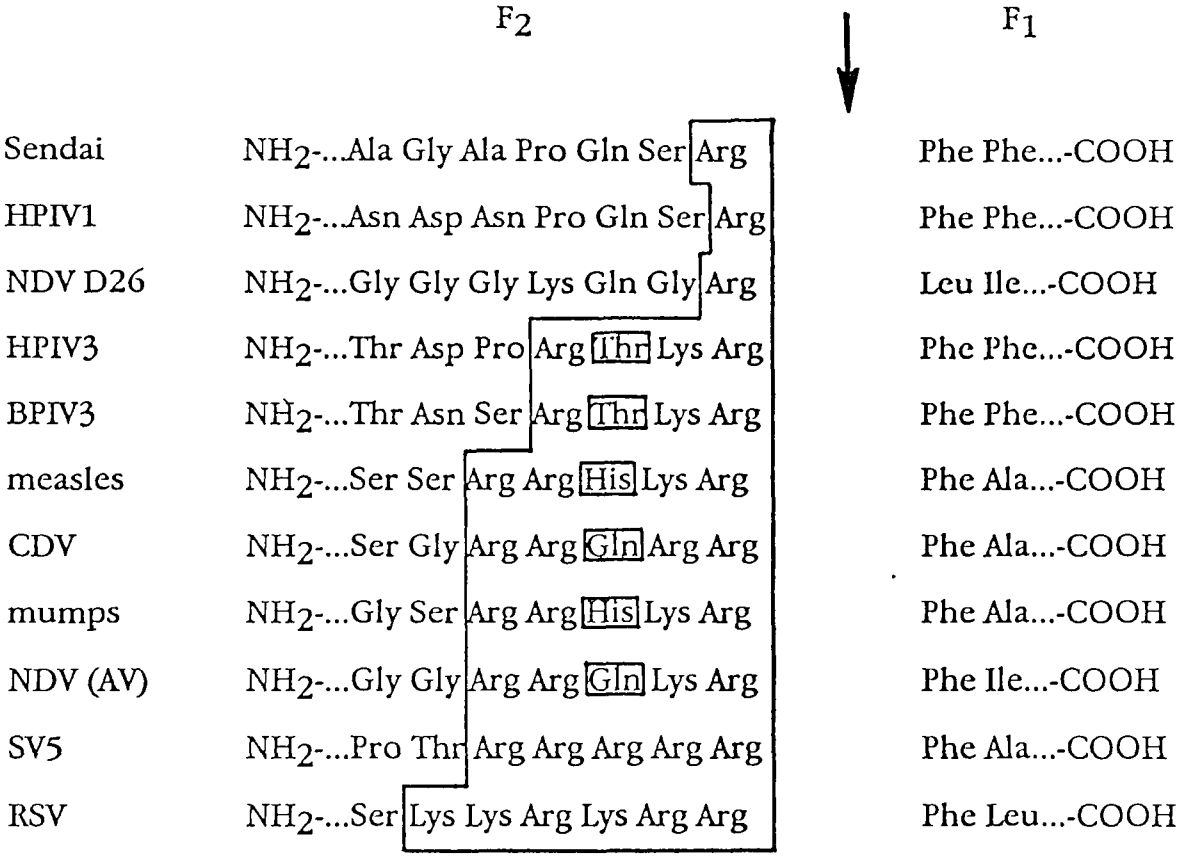

HYDROPHOBIC
HPIV3
A
B
C
2
1
0
-1
-2
HYDROPHILIC
HYDROPHOBIC
SENDAI
A
B
C
3
2
1
0
-1
HYDROPHILIC
HYDROPHOBIC
SV5
A
B
C
2
1
0
-1
HYDROPHILIC

location in each of the fusion proteins. Comparisons with the F_0 proteins of other paramyxoviruses yielded similar results (data not shown).

The highly hydrophobic region recognized at the N-terminus of the F_0 protein is believed to be the signal peptide which mediates membrane translocation of the protein (Blobel, 1980 and reviewed by Verner and Schatz, 1988). A typical signal peptide is 13-30 amino acids long and contains a stretch of 7 or 8 apolar residues. Sekikawa and Lai (1983) showed that a deletion of 11 amino acids from the 16 amino acid signal peptide of the influenza A virus hemagglutinin prevented expression of the hemagglutinin at the surface of the cell and resulted in its accumulation in the cytoplasm. The mutant hemagglutinin was not glycosylated, which suggested that the mutant protein had been synthesized on free ribosomes instead of on membrane associated ribosomes. Gething and Sambrook (1982) published similar results. These experiments indicate that the signal peptide appears to be essential for glycosylation and transport of the membrane protein to the cell surface.

Located just before the cleavage site of F_0, at the C-terminus of F_2, is a hydrophilic region, composed of basic residues such as arginine and lysine which has been termed the connecting peptide or cleavage-activation site. As shown in Figure 11, the hydrophilicity of the cleavage-activation site varies from one paramyxovirus to another. The number of basic residues varies from 1 for Sendai virus (Blumberg *et al.*, 1985; Hsu and Choppin, 1984; Miura *et al.*, 1985; Shioda *et al.*, 1986) to 6 for respiratory syncytial virus (Collins *et al.*, 1984; Elango *et al.*, 1985; Baybutt and Pringle, 1987) as shown in Figure 12. The HPIV3 F_0 protein cleavage-activation site contains 3 basic amino acid residues with an intervening threonine. Sendai virus mutants resistant to trypsin treatment

Figure 12. Amino acid sequences of paramyxovirus cleavage-activation sites. Basic amino acid residues are boxed and the arrow indicates the cleavage site between F_1 and F_2. Sendai virus (Miura *et al.*, 1985; Shioda *et al.*, 1986); HPIV1, human parainfluenza virus 1 (Merson *et al.*, 1988); NDV D26, avirulent strain of Newcastle disease virus (Sato *et al.*, 1987a); HPIV3 (Spriggs *et al.*, 1986; Cote *et al.*, 1987; Galinski *et al.*, 1987b); BPIV3, bovine parainfluenza virus 3 (Suzu *et al.*, 1987); measles virus (Richardson *et al.*, 1986; Buckland *et al.*, 1987); CDV, canine distemper virus (Barrett *et al.*, 1987); mumps virus (Waxham *et al.*, 1987); NDV (AV), Newcastle disease virus, Australian Victoria strain (Chambers *et al.*, 1986; McGinnes and Morrison, 1986; SV5, simian virus 5 (Paterson *et al.*, 1984b); RSV, respiratory syncytial virus (Collins *et al.*, 1984; Elango *et al.*, 1985; Baybutt and Pringle, 1987).


F2
F1
Sendai
NH2-...Ala Gly Ala Pro Gln Ser Arg
Phe Phe...-COOH
HPIV1
NH2-...Asn Asp Asn Pro Gln Ser Arg
Phe Phe...-COOH
NDV D26
NH2-...Gly Gly Gly Lys Gln Gly Arg
Leu Ile...-COOH
HPIV3
NH2-...Thr Asp Pro Arg Thr Lys Arg
Phe Phe...-COOH
BPIV3
NH2-...Thr Asn Ser Arg Thr Lys Arg
Phe Phe...-COOH
measles
NH2-...Ser Ser Arg Arg His Lys Arg
Phe Ala...-COOH
CDV
NH2-...Ser Gly Arg Arg Gln Arg Arg
Phe Ala...-COOH
mumps
NH2-...Gly Ser Arg Arg His Lys Arg
Phe Ala...-COOH
NDV (AV)
NH2-...Gly Gly Arg Arg Gln Lys Arg
Phe Ile...-COOH
SV5
NH2-...Pro Thr Arg Arg Arg Arg Arg
Phe Ala...-COOH
RSV
NH2-...Ser Lys Lys Arg Lys Arg Arg
Phe Leu...-COOH

were shown to have a single amino acid change at position 116, which is the position of the single basic residue of the Sendai virus F_0 protein cleavage-activation site (Hsu *et al.*, 1987; Itoh and Homma, 1987). Furthermore, by comparing the cleavage-activation site of the fusion protein of avirulent and virulent Newcastle disease virus strains, Toyoda *et al.* (1987) as well as Glickman *et al.* (1988) demonstrated that virulence increases with the number of paired basic amino acid residues. There appears to be a relationship between pathogenicity and cleavability of the fusion protein of Newcastle disease virus. An increased number of basic residues would be expected to increase the efficiency of cleavage of F_0 proteins, which is necessary for activation of the fusogenic properties of the fusion protein. Recently, Paterson *et al.* (1989) tested the relationship between the number of basic residues at the cleavage-activation site and the cleavability of the fusion protein of simian virus 5. Normally, the cleavage-activation site of simian virus 5 consists of 5 arginine residues. They observed that a minimum of 4 arginines was necessary for cleavage, and therefore for activation. 2 or 3 arginine residues were not sufficient to result in cleavage by host cell proteases. Cleavage could be accomplished by addition of trypsin, however, the protein was not active. The mutant with 1 arginine residue could not be cleaved with trypsin. This experiment demonstrated not only that the number of basic residues is important in order for cleavage of the fusion protein to occur, but also that cleavage alone does not necessarily activate the protein. There may also be a correlation between the intracellular site of cleavage and the number of basic residues in the cleavage-activation site of fusion proteins. Sendai virus with 1 arginine residue at the F_0 cleavage-activation site is cleaved at the surface of infected-cells (Lamb *et al.*,

1976; Seto *et al.*, 1981) while Newcastle disease virus with two paired basic residues (total of 4 basic residues) is cleaved inside the infected-cell (Nagai *et al.*, 1976b; Seto *et al.*, 1981). It is not known where in the cell the HPIV3 fusion protein precursor is cleaved. It is possible that with the sequence, Arg-Thr-Lys-Arg, the HPIV3 F_0 protein cleavage occurs inside the host cell, however, the data obtained by Paterson *et al.* (1989) would argue the contrary. This matter needs to be investigated.

Adjacent to the cleavage-activation site on the amino acid sequence of the F_0 protein, is the hydrophobic F_1 N-terminus. This region is very conserved within the *Paramyxoviridae* family and is believed to play an important role by interacting with membranes during fusion (Gething *et al.*, 1978; Richardson *et al.*, 1980; Richardson and Choppin, 1983). Recently, the hydrophobicity of the N-terminus of F_1 and its relationship with membranes was investigated. Constructs fusing regions of membrane proteins such as influenza hemagglutinin (HA) or the gene III protein (III) of coliphage f1 and the hydrophobic F_1 N-terminus of simian virus 5 or Sendai virus were made to test if the N-terminus of paramyxovirus F_1 can act as a transmembrane region. The results of these experiments showed that the F_1 N-terminal region is close to the threshold of hydrophobicity necessary for membrane interaction. The N-terminus did not act like a membrane anchor when located internally within a protein, with or without an adjacent hydrophilic cleavage-activation site (Davis and Hsu, 1986; Paterson and Lamb, 1987). However, if placed at the end of a protein, in the absence of a neighboring cleavage-activation site, the F_1 N-terminus was sufficiently hydrophobic to interact with membranes (Paterson and Lamb, 1987). These elegant experiments demonstrated that after cleavage, the free hydrophobic N-terminus of F_1, now located at the

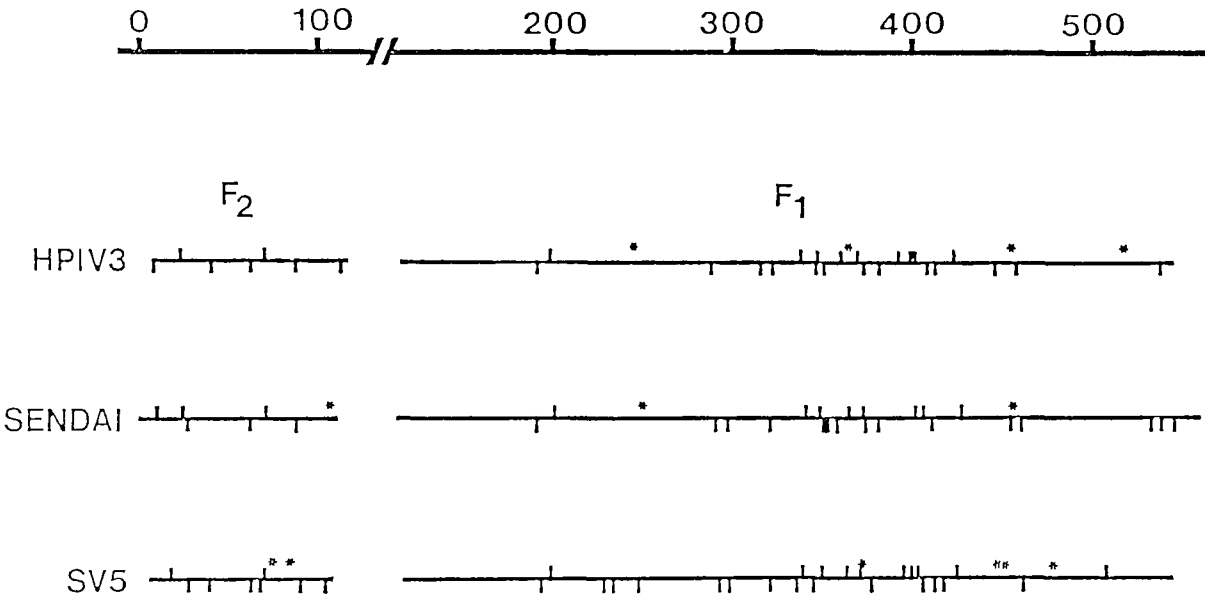
end of a polypeptide and free of the hydrophilic cleavage-activation site, which reduces the overall hydrophobic nature of that region of the F_0 protein, reaches a hydrophobic threshold that allows interaction with lipid membranes.

The most hydrophobic region of the fusion protein is the transmembrane region, near the C-terminus of F_1, preceding the hydrophilic C-terminal cytoplasmic tail. Constructs of the influenza virus hemagglutinin lacking the transmembrane region were expressed in simian cells (Gething and Sambrook, 1982). Their results showed that the mutant hemagglutinin was glycosylated but secreted in the medium, thus confirming the role of this hydrophobic region as the transmembrane domain which prevents complete translocation and secretion of the protein (Gething and Sambrook, 1982). Mutations made in the cytoplasmic tail of the influenza hemagglutinin showed that this hydrophilic region, usually inside the cell, influences the rate and the efficiency of transport of the protein to the surface of the cell (Doyle *et al.*, 1985). It is believed that the cytoplasmic domain may be recognized by the viral matrix (M) protein as a late step during assembly (Lyles, 1979).

3. Location of HPIV3 fusion protein cysteine and proline residues and glycosylation sites and comparison with other paramyxovirus fusion proteins

Another feature of the paramyxovirus fusion protein sequence is the conservation of the location of cysteine residues, proline residues and some potential glycosylation sites. Figure 13 demonstrates the alignment of HPIV3, Sendai virus and simian virus 5 F_0 proteins. Eight of 11 cysteines and 11 of 18 prolines of the HPIV3 F_0 protein are clustered in

Figure 13. Comparison of cysteine and proline residues and potential glycosylation sites in the predicted amino acid sequences of the HPIV3, Sendai virus and simian virus 5 F_1 and F_2 proteins (Cote *et al.*, 1987). The amino termini are on the left. Vertical lines represent the positions of cysteine (above) and proline (below) residues and the asterisks represent the potential N-glycosylation sites (Asn-X-Ser/Thr). Sequencing data for Sendai virus and simian virus 5 F_1 and F_2 proteins are from Blumberg *et al.* (1985) and Paterson *et al.* (1984b).


0
100
200
300
400
500
F2
F1
HPIV3
SENDAI
SV5

F_1 between amino acids 282 and 453. In the fusion protein of Sendai virus, cysteines and prolines are also concentrated in the same part of F_1. In addition, many of the cysteines, and to a lesser extent the prolines, in the F_1 protein of simian virus 5 can also be aligned with those of HPIV3 and Sendai virus F_1 proteins. Morrison (1988) observed that all the paramyxovirus F_0 proteins sequenced to date demonstrate the same remarkable conservation of cysteine residues: 8 in a cluster in F_1, 1 in the N-terminus of F_1 and another in F_2. This last cysteine residue must be involved in disulfide bonding of F_1 and F_2. Additional cysteines are found in some paramyxovirus F_2 polypeptides and in respiratory syncytial virus F_1 (Morrison, 1988). The number of conserved cysteine and proline residues in F_1 suggests a high degree of folding and possibly an important role in stabilizing the tertiary structure of the fusion protein.

To a much lesser extent, glycosylation sites are conserved among paramyxovirus fusion proteins. In Figure 13, 4 potential glycosylation sites are shown for the HPIV3 F_0 protein, all in the F_1 coding region. The relative positions of the HPIV3 F_0 protein glycosylation sites were compared with those of other paramyxovirus F_0 proteins and in particular, in Figure 13, with the Sendai virus and simian virus 5 F_0 proteins. The first glycosylation site (amino acids 238 to 240) corresponds to one in an identical position in the Sendai virus F_0 protein. The second (amino acids 359 to 361) corresponds to one at an equivalent position in the simian virus 5 F_0 protein. There is a glycosylation site at a corresponding position also in mumps virus and Newcastle disease virus F_0 proteins (Morrison, 1988). The third (amino acids 446 to 448) glycosylation site is in a similar position in all three of the F_0 proteins shown in Figure 13. The fourth potential acceptor site (amino acids 508 to 510), found within

the membrane anchor region, does not correspond to a glycosylation site of either the Sendai virus or simian virus 5 F_0 protein, and is probably not used.

4. Similarities between HPIV3 and other paramyxovirus fusion protein

There are many sequence similarities among paramyxovirus fusion proteins. HPIV3 F_0 protein amino acid sequence shows the highest degree of similarity with bovine parainfluenza virus 3 F_0 protein (80%)(Suzu *et al.*, 1987). Sendai virus and human parainfluenza virus 1 come next with fairly high per cent similarities, 43 and 41% respectively (Cote *et al.*, 1987; Merson *et al.*, 1988). The group of Newcastle disease virus, simian virus 5 and mumps virus F_0 proteins show 30, 30 and 28% similarity respectively, with HPIV3 F_0 protein. The F_0 proteins of measles virus and respiratory syncytial virus are identical to HPIV3 at 24% of their amino acid positions.

As discussed in Chapter 3, HPIV3, bovine parainfluenza virus 3, Sendai virus and human parainfluenza virus 1 appear to be closely related to each other. The similarity of their F_0 protein sequences support this statement. Newcastle disease virus, Simian virus 5 and mumps virus do not appear to be as closely related to HPIV3 as Sendai virus and since their genome organization differs from that of HPIV3, they may form at least a different subset within the family. More information is needed. Measles virus and respiratory syncytial virus F_0 proteins show the lowest degree of similarity with the HPIV3 F_0 protein which suggest that they belong to different genera.

CHAPTER 5 EXPRESSION AND ANALYSIS OF CYSTEINE MUTANTS OF THE HPIV3 FUSION PROTEIN

INTRODUCTION

Comparisons of the predicted amino acid sequences of the fusion proteins of paramyxoviruses show that cysteine residues are in similar positions. The most striking observation is the conserved cluster of 8 cysteine residues located at the beginning of the second half of F_1. This begs the question of the importance of this region to the function of fusion proteins. Nucleotide sequencing of the F gene of antigenic variants of both Sendai virus and Newcastle disease virus, selected with monoclonal antibodies inhibiting fusion and virus neutralization, showed that amino acid substitutions were occuring in the cysteine cluster (Portner *et al.*, 1987b; Toyoda *et al.*, 1988). Portner *et al.* (1987b) suggested that the region of the cysteine cluster may be located near the F_1 N-terminus when the protein is folded and antibody binding at this site could be inhibiting fusion. Toyoda *et al.* (1988) suggested that the domain located within the cysteine cluster may be directly involved in structure and function of the fusion protein. This statement was supported by the results of Hull *et al.* (1987); they isolated a measles virus mutant which was able to cause cell-cell fusion in the presence of an oligopeptide that inhibits fusion. The nucleotide sequence of this mutant showed 3 amino acids changes, two of which were located in the cysteine cluster domain.

These data strongly suggest that the cysteine residues in paramyxovirus fusion proteins, in particular, the cysteine cluster, are of major importance in maintaining the tertiary structure and, possibly, the function of fusion proteins. The fifth objective of this study, therefore, was to examine the importance of cysteine residues for the structure and/or function of the HPIV3 fusion protein and to identify the residues responsible for the linkage between F_1 and F_2.

The HPIV3 F gene was expressed using vaccinia virus as a vector. This orthopoxvirus has a double-stranded DNA genome of approximately 187,000 bp which codes for approximately 100 early and 100 late proteins. Vaccinia virus replicates in the cytoplasm of cells, with a broad host range, and uses its own transcriptional machinery (Mackett and Smith, 1986; Moss and Flexner, 1987).

Our strategy was to generate a recombinant vaccinia virus containing a full length construct of the HPIV3 F gene and to follow expression of the fusion protein by immunofluorescence and polyacrylamide gel electrophoresis. Then, using the M13 bacteriophage system and site-specific mutagenesis, individual cysteine codons of the HPIV3 F gene were converted into serine codons. The expression of altered HPIV3 F genes in vaccinia virus was analysed by reducing and non-reducing polyacrylamide gel electrophoresis. The expression of the HPIV3 F gene using vaccinia virus and the importance of cysteine residues for HPIV3 fusion protein structure and/or function will be discussed.

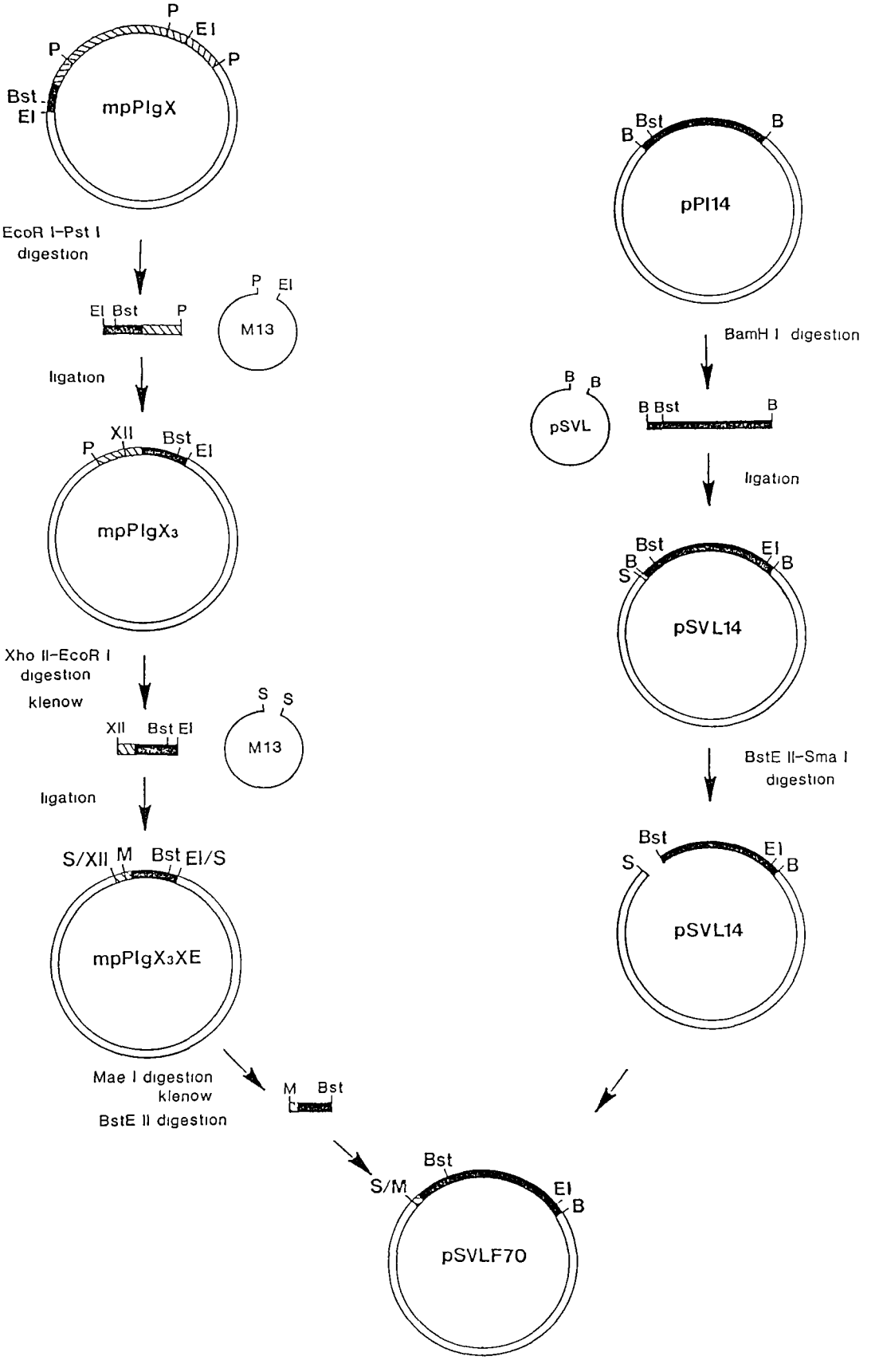
RESULTS

1. Construction of a full length HPIV3 F gene

To construct a full length cDNA clone covering the entire HPIV3 F gene sequence, 2 independent clones were used; pPI14, a viral mRNA-derived cDNA clone covering approximately 1,500 nucleotides of the F gene, and mpPIgX, a 4,500 bp genomic RNA-derived cDNA clone 650 bp of which corresponds to the 5'-end (mRNA sense) of the F gene, including the translation initiation site. These 2 clones are shown in Figure 14 along with a description of the construction of the full length copy of the HPIV3 F gene.

A fragment which includes 650 nucleotides of the HPIV3 M gene and 650 nucleotides of the HPIV3 F gene was cut from clone mpPIgX using *Eco*RI and *Pst*I. The M-F fragment was cloned into the *Eco*RI-*Pst*I site of M13 replicative form DNA and the resulting phage was named mpPIgX$_3$. A *Xho*II-*Eco*RI fragment covering approximately 220 nucleotides of the HPIV3 M gene and 650 nucleotides of the F gene was ligated into the *Sma*I site of M13 replicative form DNA after the ends had been filled using Klenow enzyme. The resulting clone was named mpPIgX$_3$XE. In the meantime, a *Bam*HI fragment covering approximately 1,500 nucleotides of the HPIV3 F gene, was cloned into the *Bam*HI site of plasmid pSVL (Pharmacia). The resulting clone, pSVL14, and mpPIgX$_3$XE share a *Bst*EII site located 383 nucleotides from the 5'-end of the HPIV3 F gene. Therefore, the next and last step of the construction was to ligate fragments of the 2 clones together to yield a full length copy of the HPIV3 F gene. To accomplish this last step, mpPIgX$_3$XE was digested with *Mae*I endonuclease, the 3' recessed ends were filled in and the DNA was

Figure 14. Strategy for the construction of a full length DNA copy of the HPIV3 F gene. The solid bars correspond to the HPIV3 F sequences and sequences of other HPIV3 genes are represented by the hatched areas. The letters represent the following restriction sites: B, *Bam*HI; Bst, *Bst*EII; EI, *Eco*RI; M, *Mae*I; P, *Pst*I; S, *Sma*I; XII, *Xho*II.


mpPIgX
EcoR I–Pst I
digestion
M13
ligation
mpPIgX3
Xho II–EcoR I
digestion
klenow
M13
ligation
mpPIgX3XE
Mae I digestion
klenow
BstE II digestion
pPI14
BamH I digestion
pSVL
ligation
pSVL14
BstE II–Sma I
digestion
pSVL14
pSVLF70

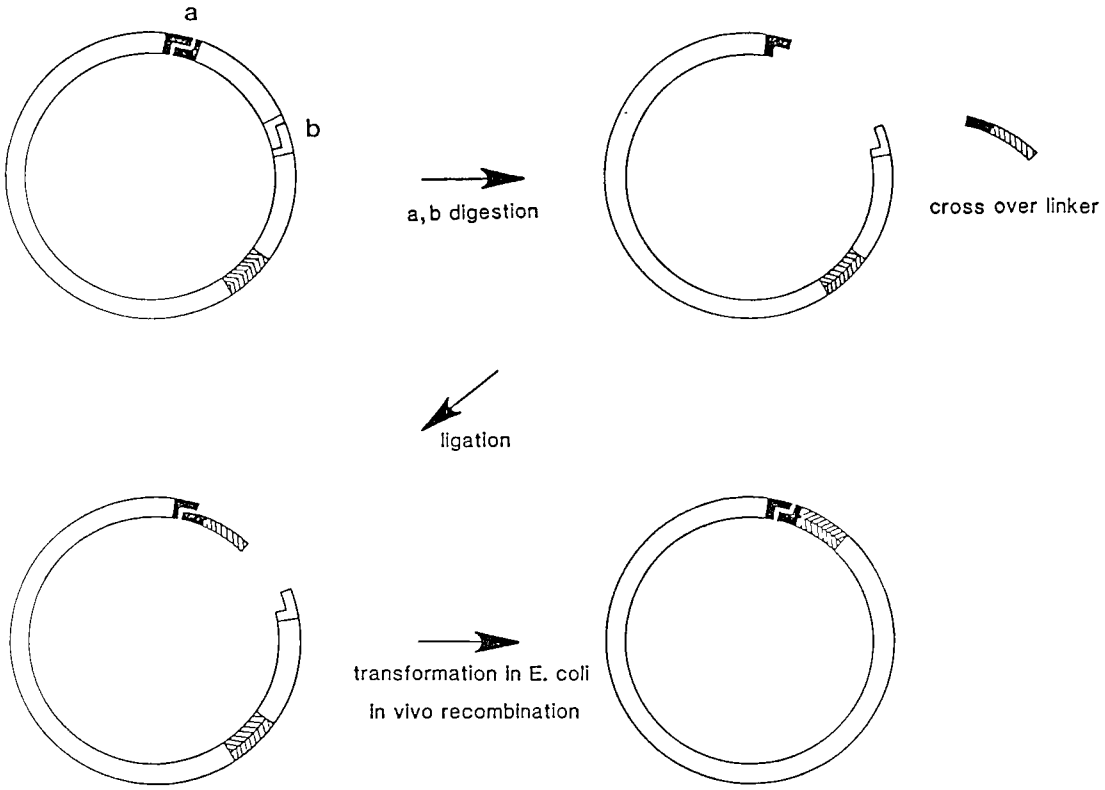
then digested with *Bst*EII. The resulting fragment, containing 70 nucleotides of the HPIV3 M gene and 383 nucleotides of the 5'-end of the HPIV3 F gene, was cloned into the *Sma*I-*Bst*EII site of pSVL14. The new clone, containing a full length copy of the HPIV3 F gene plus 70 bp corresponding to the 3'-end of the M gene was named pSVLF70.

2. Crossover linker mutagenesis (Sung *et al.*, 1986)

Clone pSVLF70 contained 70 nucleotides from the HPIV3 M gene, the intergenic sequence between the M and F genes, 1,851 nucleotides corresponding to the entire HPIV3 F gene and approximately 20 residues from the 3' poly (A) tail of the F mRNA. Since the HPIV3 F gene open reading frame is 1617 nucleotides in length, the clone in question contains 264 bp upstream and approximately 60 bp downstream that, in all probability, would not be used for expression of the protein in the vaccinia virus system. Crossover linker mutagenesis is a rapid and elegant method for deletion of specific regions of DNA such as, in this case, the elimination of non-translated regions.

The principle of the crossover linker mutagenesis method is shown in Figure 15. Sites a and b are restriction sites. The region contained between site a and the hatched area is the non-translated region to be deleted. After digestion with restriction endonuclease at sites a and b, the resulting plasmid is purified and ligated to an oligonucleotide complementary to the 3' overhang of site a and containing several (in this case 12) nucleotides corresponding to the begining of the region to be conserved (the hatched area). After ligation, linear plasmid DNA is selected by agarose gel electrophoresis and used to transform *E. coli.* The region not included in the oligonucleotide is eliminated by *in vivo*

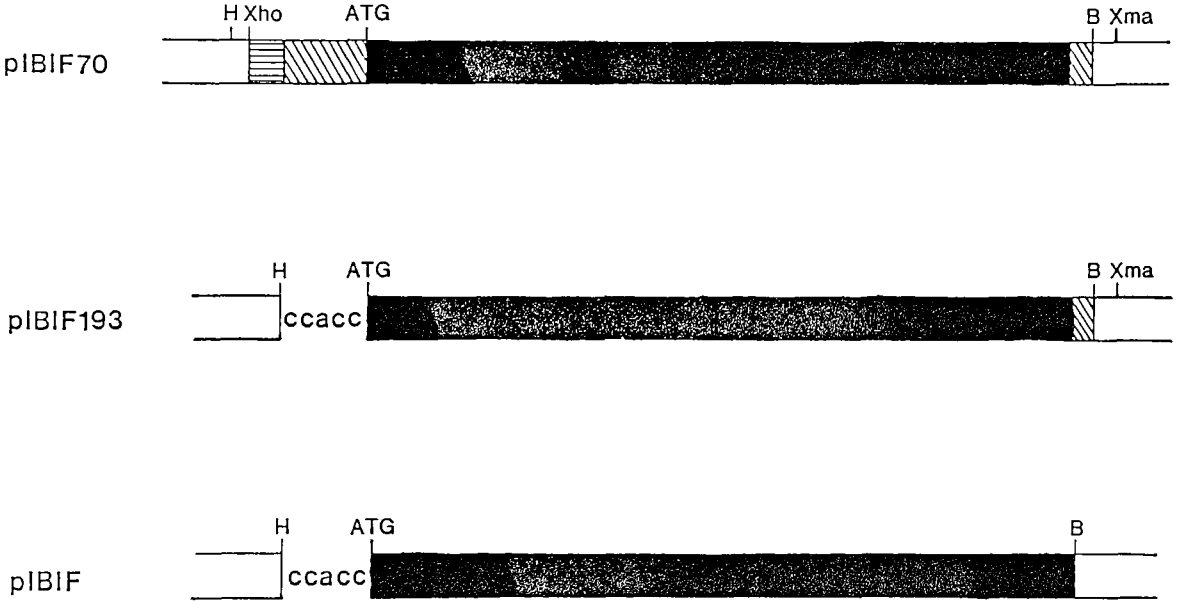
Figure 15. Strategy for crossover linker mutagenesis. The black box corresponds to restriction site a and restriction site b is represented by an open box. The hatched areas are homologous sequences.


a
b
a, b digestion
cross over linker
ligation
transformation in E. coli
in vivo recombination

recombination. The synthetic oligonucleotide may also contain extra sequences such as restriction sites or regulatory sequences.

Two crossover linker constructions were performed on the HPIV3 F gene in order to remove the non-translated regions. Before mutagenesis, for technical purposes, pSVLF70 was digested with *Xho*I and *Bam*HI and the resulting fragment, corresponding to the entire insert, was cloned into the *Xho*I-*Bam*HI site of plasmid pIBI (IBI). The resulting clone, pIBIF70 is shown in Figure 16. The first crossover linker experiment was done to remove sequences upstream of the coding region of the F gene. The sequence of the oligonucleotide used in the experiment is shown in the Appendix. For the 5'-end (mRNA sense), the oligonucleotide was designed to: 1) regenerate site a, in this case a *Hind*III site, 2) add 5 nucleotides corresponding to the preferred context for translational initiation by eukaryotic ribosomes (Kozak, 1986) and 3) include 12 nucleotides corresponding to the start of the HPIV3 F gene open reading frame. The resulting clone, pIBIF193, is shown in Figure 16. The second crossover linker experiment was done with pIBIF193, to eliminate the 3' non-translated region of the HPIV3 F gene. The oligonucleotide used in this experiment was designed to delete the non-translated region, including the poly (A) tail and to generate a *Bam*HI site at the 3'-end of the coding region so that the construct could be easily recloned into other vectors. The resulting clone, pIBIF, is shown in Figure 16 and contains, between *Hind*III and *Bam*HI restriction sites, the entire coding sequence of the HPIV3 F gene. The nucleotide sequences at both ends of the HPIV3 F gene coding region were confirmed by sequence analysis.

Figure 16. Three different constructs of the HPIV3 F gene. The black boxes represent the coding sequence of HPIV3 F. The diagonally hatched boxes correspond to non-translated sequences of HPIV3 F and the horizontally hatched box represents HPIV3 M sequences. The letters represent the following restriction sites: B, *Bam*HI; H, *Hind*III; Xho, *Xho*I; Xma, *Xma*I.

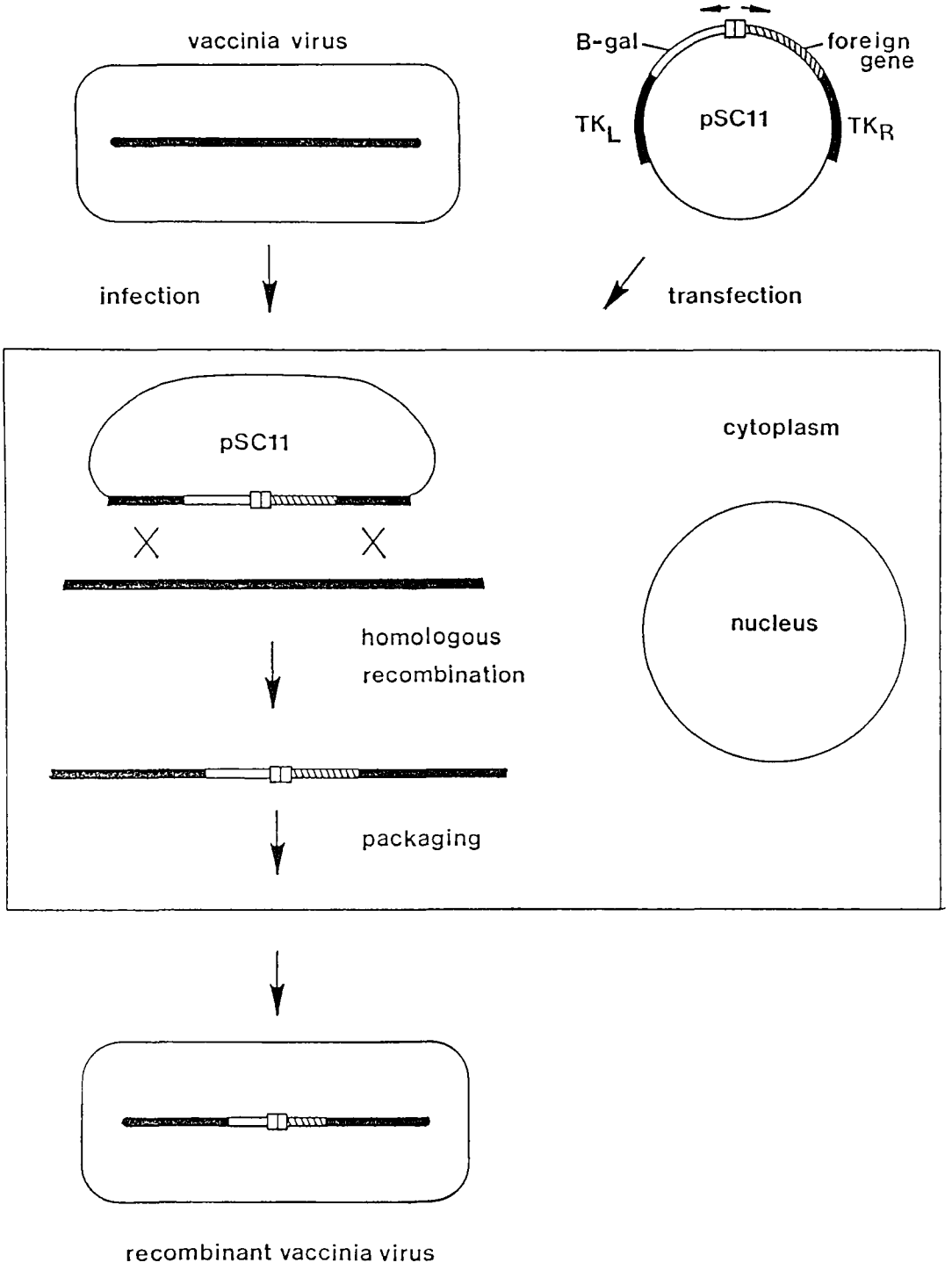

pIBIF70
H Xho
ATG
B Xma
pIBIF193
H
ccacc
ATG
B Xma
pIBIF
H
ccacc
ATG
B

3. Construction of vaccinia virus recombinants expressing the HPIV3 fusion protein

In order to study the effects of different mutations on the structure and function of the HPIV3 fusion protein, a system expressing the fusion protein as the only HPIV3 specific protein, that would allow manipulation of the F gene, was mandatory. Vaccinia virus was chosen as the system for expression of wild type and mutated fusion proteins of HPIV3. The first step in generating recombinant vaccinia virus was to subclone the F gene into the *Sma*I site of plasmid pSC11 (Chakrabarti *et al.*, 1985). pSC11 is shown in Figure 17 and contains 2 vaccinia virus promotors flanked by sequences from the vaccinia virus thymidine kinase gene; the P11 promotor is derived from a late gene encoding an 11K polypeptide and the P7.5 promotor is derived from a viral gene coding for a 7.5 polypeptide expressed throughout the replicative cycle (Mackett and Smith, 1986). pSC11 also contains the *E. coli* β-galactosidase gene under the control of the P11 promotor. The F gene was placed under the control of the P7.5 promotor.

To produce recombinant vaccinia virus, infection of eukaryotic cells with wild type vaccinia virus was done at a low MOI (0.1 plaque forming units/cell for LLC-MK2 cells). An hour after the beginning of infection, the inoculum was aspirated, 10ml of fresh MEM was added to the infected cells and transfection was performed (Chapter 2). During the 48 hours following the infection-transfection, the rare event of homologous recombination between the thymidine kinase gene sequences on the plasmid and the vaccinia genome is expected to occur. Homologous recombination allows the β-galactosidase gene as well as the F gene to become part of recombinant vaccinia virus genomes which will be

Figure 17. Strategy for the production of recombinant vaccinia virus. Black areas represent vaccinia virus sequences. The two white boxes correspond to vaccinia virus promotors and arrows indicate the direction of transcription. The open regions represent the β-galactosidase gene and the hatched areas represent the foreign gene to be expressed. Adapted from Mackett and Smith (1986).


vaccinia virus
B-gal
foreign gene
pSC11
TK_L
TK_R
infection
transfection
pSC11
cytoplasm
nucleus
homologous recombination
packaging
recombinant vaccinia virus

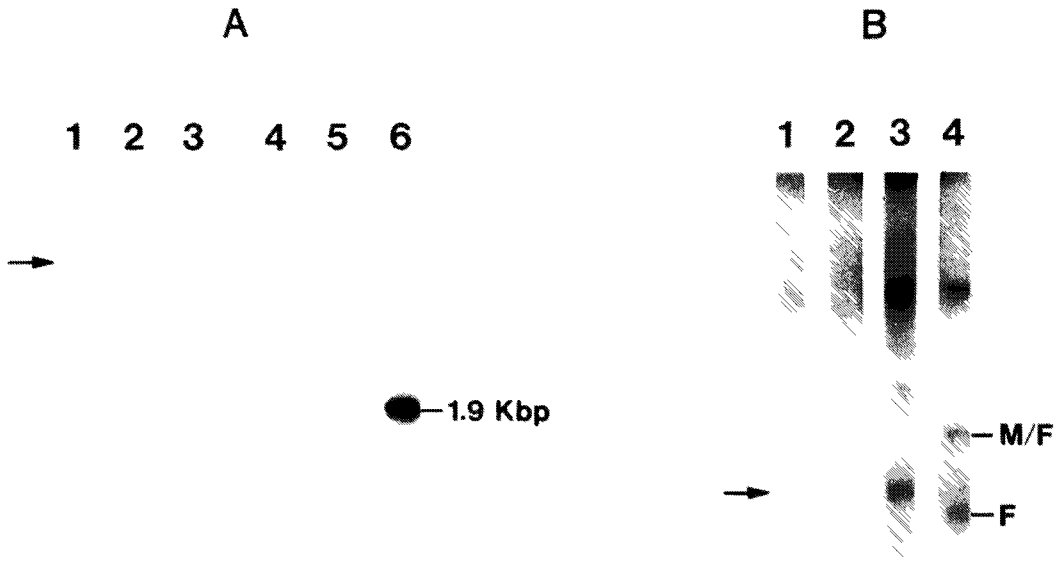
packaged yielding virions able to express both β-galactosidase and the fusion protein.

Virus harvested from infected-transfected cells is a mixture of wild type vaccinia virus, spontaneous thymidine kinase mutants and recombinant virus. The next 2 steps in the procedure permit the selection of recombinant vaccinia virus. 143B cells were infected with virus harvested from the infected-transfected cells and incubated in MEM containing 5-bromodeoxyuridine (BUdR) to select thymidine kinase negative virus. Phosphorylation of BUdR, a nucleoside analogue, by thymidine kinase, promotes its incorporation into viral DNA, a lethal event. Recombinant vaccinia virus (as well as spontaneous thymidine kinase mutants) no longer possesses a functional thymidine kinase gene since, during recombination, it was disrupted by the β-galactosidase and F genes. Recombinant virus, therefore, can grow in the presence of BUdR (Moss and Flexner, 1987). However, most of the plaques which appeared on 143B cells were spontaneous thymidine kinase mutants rather than recombinants. A second method of selection was necessary in order to efficiently identify recombinant vaccinia virus. Expression of the β-galactosidase gene permits identification of recombinant virus. A second agarose overlay, containing XGal, allowed identification of the recombinant vaccinia virus within hours. When XGal is processed by β-galactosidase, the blue product diffuses into the agar surrounding recombinant virus plaques. Blue plaques were picked and recombinant viruses were plaque purified 3 times and amplified for production of recombinant vaccinia virus stocks. Recombinant viruses were then checked for the presence of the foreign gene.

Recombinant vaccinia virus containing the entire HPIV3 F gene including the non-translated regions as well as the 70 nucleotides from the M gene was named VF70; virus containing the F gene with the 5' non-translated region (mRNA sense) deleted was named VF193; and virus made from the F gene with all the non-translated regions deleted was named VF.

Southern and Northern blots of DNA and RNA isolated from non-infected cells and from cells infected with Vsc8 (a recombinant vaccinia virus expressing only the β-galactosidase gene) and recombinant vaccinia virus VF193 were probed with a nick-translated insert isolated from pIBIF193. Panel A in Figure 18 shows a high-molecular weight band in lane 3 corresponding to the DNA genome of VF193. The band of 1.9 Kbp seen in lane 6 corresponds to the band expected after an *Eco*RI digest of the VF193 genome. *Eco*RI cuts the F gene 61 bases from the 3'-end (mRNA sense) and also approximately 400 bp upstream of the initiation codon in vaccinia virus sequences. These results demonstrate that VF193 was carrying the HPIV3 F gene and that no major rearrangements were detected. DNA from both non-infected cells and Vsc8-infected cells showed no hybridization with the probe. Panel B of Figure 18, lane 3, shows that a species of RNA, in VF193-infected cells, slightly larger than the HPIV3 F mRNA (lane 4) hybridized with the F gene specific probe. The size of the VF193 F mRNA is larger because the vaccinia virus promotor P7.5 is located 400 bp upstream of the HPIV3 F gene sequence. The high molecular weight band in lane 3 on the Northern blot is assumed to be the result of contamination of the RNA with VF193 DNA. The high molecular weight band in lane 4 on the Northern blot corresponds to HPIV3 genomic RNA.

Figure 18. Characterization of recombinant vaccinia virus VF193 by Southern and Northern blot analysis. In panel A, DNA was isolated from uninfected, Vsc8-infected or VF193-infected CV-1 cells, digested (or not) with the restriction enzyme *Eco*RI, electrophoresed on a 1.0% agarose gel and transferred onto a nylon membrane. Hybridization was done with a F gene specific nick-translated *Hind*III-*Bam*HI fragment of pIBIF193. Lanes 1 and 4, uninfected cells; 2 and 5, Vsc8-infected cells; 3 and 6, VF193-infected cells. DNA of lanes 4, 5, 6 was digested with *Eco*RI prior to electrophoresis. The arrow shows the position of a band in lane 3. In panel B, total RNA was isolated from uninfected CV-1 cells, Vsc8-infected cells, VF193-infected cells and HPIV3-infected cells, denatured with glyoxal, electrophoresed on a 1.2% agarose gel and transferred onto a nylon membrane. Lane 1, uninfected CV-1 cells; 2, Vsc8-infected cells; 3, VF193-infected cells; 4, HPIV3-infected cells.


A
1 2 3 4 5 6
1.9 Kbp
B
1 2 3 4
M/F
F

To characterize the product expressed by the recombinant vaccinia virus, ^{3}H-labeled proteins immunoprecipitated from lysates of non-infected, Vsc8-infected, VF193-infected and VF70-infected cells were electrophoresed on a SDS-polyacrylamide gel as shown in Figure 19. Rabbit fusion protein-specific antiserum was used for the immunoprecipitations shown in lanes 1-4. Two proteins corresponding to F_0 and F_1 were immunoprecipitated from lysates of HPIV3-infected cells (lane 4) and are also found in VF193-infected cells (lane 3). The same proteins, F_0 and F_1 were immunoprecipitated from VF70-infected cell lysates using rabbit anti-HPIV3 serum (lane 5), fusion protein-specific monoclonal antibodies (lane 6) and fusion protein-specific antiserum (lane 7).

CV-1 cells infected with recombinant vaccinia virus VF showed strong immunofluorescence using rabbit anti-HPIV3 serum and fluorescein-conjugated goat anti-rabbit IgG as shown in Figure 20. Uninfected and Vsc8-infected cells exhibited no specific fluorescence.

4. Site-specific mutagenesis of cysteine codons to serine codons

After setting up the expression system for the HPIV3 fusion protein, the next step was to mutate individually each cysteine codon, TGC or TGT, into a serine codon, TCC or TCT. Serine was chosen since, having an oxygen atom instead of a sulfur atom, it is sterically very similar to a cysteine residue. The mutations involved only one nucleotide change: G to C. Figure 21 shows the position of each cysteine residue in the HPIV3 fusion protein and the substitution that was used to convert each cysteine codon to a serine codon. Mutations was carried out using M13 bacteriophage carring the HPIV3 F gene and the Muta-Gene kit (Bio-

Figure 19. Characterization of the expressed protein products of VF193 and VF70 by SDS-polyacylamide gel electrophoresis. [^{3}H]leucine-labeled proteins were isolated from uninfected, Vsc8-infected, HPIV3-infected, VF193-infected and VF70-infected CV-1 cells, immunoprecipitated with different sera or monoclonal antibodies and electrophoresed on a SDS-10% polyacrylamide gel. Lanes 1 to 4 were immunoprecipitated using rabbit fusion protein specific antiserum. Lane 1, ininfected cells; 2, Vsc8-infected cells; 3, VF193-infected cells; 4, HPIV3-infected cells. Lanes 5, 6, 7 represent immunoprecitated proteins from VF70-infected cells. Lane 5, rabbit anti-HPIV3 serum; 6, mouse monoclonal antibodies specific for the fusion protein; 7, rabbit fusion protein specific antiserum.

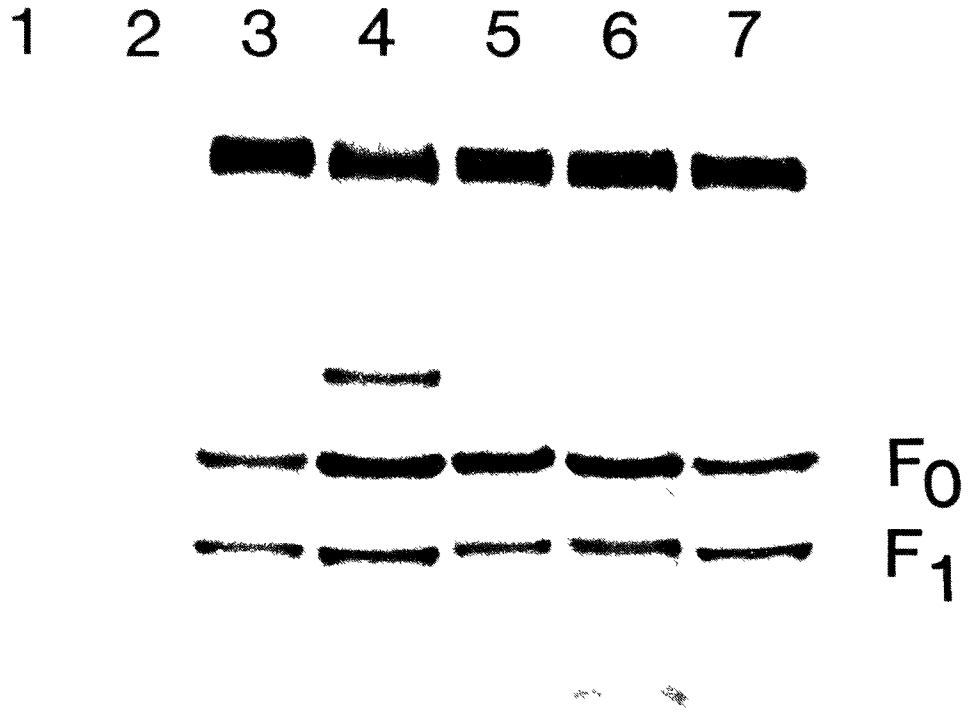

1 2 3 4 5 6 7
F_0
F_1

Figure 20. Immunofluorescence of recombinant vaccinia virus-infected cells. CV-1 cells were grown on a Multitest slide and infected (MOI of 1 plaque forming units/cell) with Vsc8 (B), VF (C), HPIV3 (D) or not infected (A). After 24 hours, Cells were fixed in acetone at room temperature for 3 minutes, incubated with rabbit anti-HPIV3 serum at 37°C for 30-60 minutes, washed and incubated with fluorescein-conjugated goat anti-rabbit IgG at 37°C for 30 minutes.

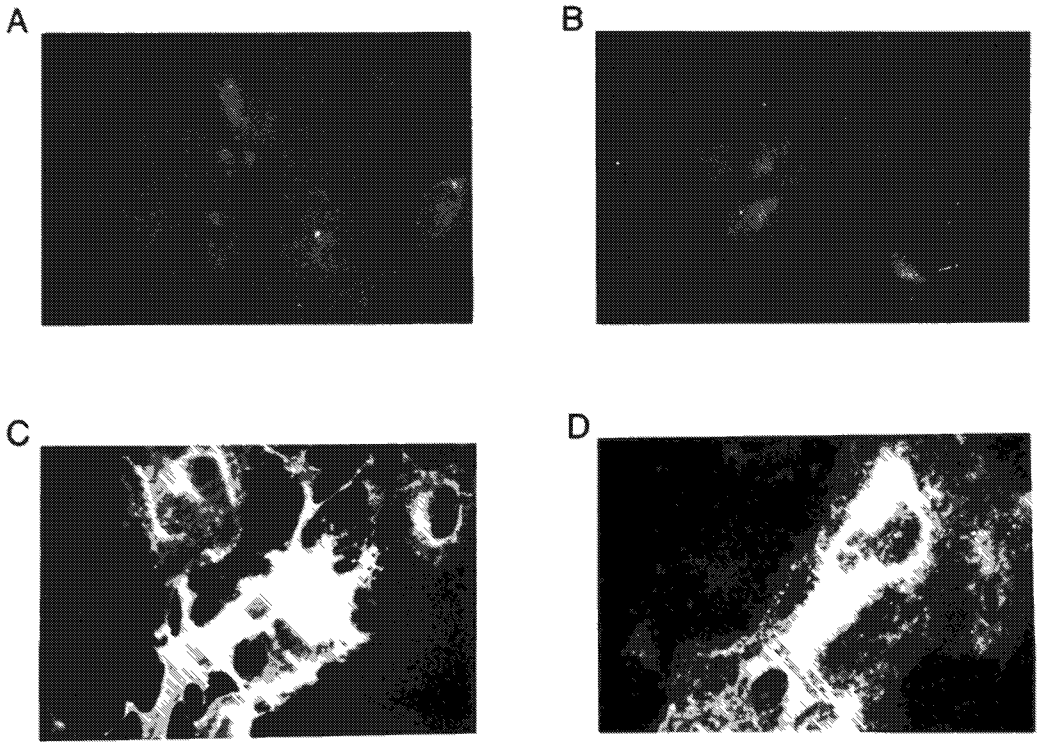
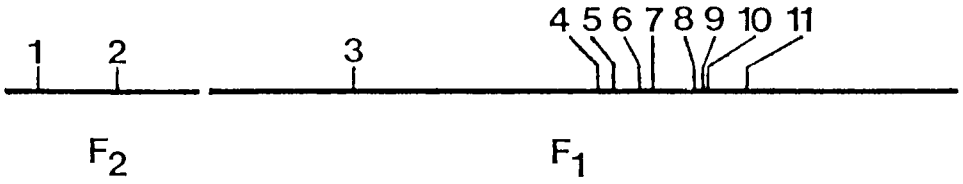

A
B
C
D

Figure 21. Positions and mutation of cysteine residues on the HPIV3 F_1 and F_2 proteins. The N-terminus is on the left and each bar represents a cysteine residue as numbered. For each cysteine residue, the amino acid position in the F_0 protein, the original codon and the mutation are shown.



CYSTEINE NUMBER	AMINO ACID NUMBER	ORIGINAL CODON	MUTATED CODON
1	CYS 18	TGC	TCC
2	CYS 63	TGT	TCT
3	CYS 192	TGT	TCT
4	CYS 331	TGT	TCT
5	CYS 340	TGC	TCC
6	CYS 355	TGT	TCT
7	CYS 363	TGT	TCT
8	CYS 387	TGT	TCT
9	CYS 392	TGT	TCT
10	CYS 394	TGC	TCC
11	CYS 417	TGT	TCT

Rad). The principle of the method is shown in Figure 22 (Kunkel, 1985). M13 grown in an *E. coli* mutant deficient in dUTPase (*dut*-) and uracyl N-glycosylase (*ung*-) carries a number of uracils in its genome (U-DNA). After isolation of the single-stranded U-DNA genome, an oligonucleotide containing a single mismatch is annealed to the U-DNA and used to prime complementary strand synthesis. After ligation, the double-stranded DNA is transformed into wild type *E. coli* where the U-DNA strand is preferentially inactivated leaving the complementary strand, with the single mismatch, to replicate.

Using specific oligonucleotides having strategically located C residues instead of G's, specific mutations were introduced into the HPIV3 F gene in each of the 11 cysteine codons. The oligonucleotides used for mutagenesis are listed in the Appendix. After isolation of M13 replicative form DNA for each mutant, the insert was isolated following *Bam*HI and *Hind*III restriction endonuclease digestion, the 3' recessed ends were filled in and the fragments were cloned into the *Sma*I site of pSC11. The nucleotide sequence in the region of each mutation was determined, for each of the resulting pSCF clones, using specific oligonucleotide primers. The sequencing results are shown in Figure 23.

5. Expression and characterization of cysteine mutants

Vaccinia virus recombinants of the HPIV3 fusion protein gene containing single altered cysteine codons were isolated and named VFCys18, VFCys63, VFCys192, VFCys331, VFCys340, VFCys355, VFCys363, VFCys387, VFCys392, VFCys394 and VFCys417. CV-1 cultures were infected with Vsc8, VF and with the 11 cysteine mutants. ^{3}H-labeled cells were lysed with RIPA buffer, proteins were immunoprecipitated with

Figure 22. Strategy for site-specific mutagenesis of M13 single-stranded DNA. *Dut* and *ung* correspond to dUTPase and uracyl N-glycosylase enzymes.

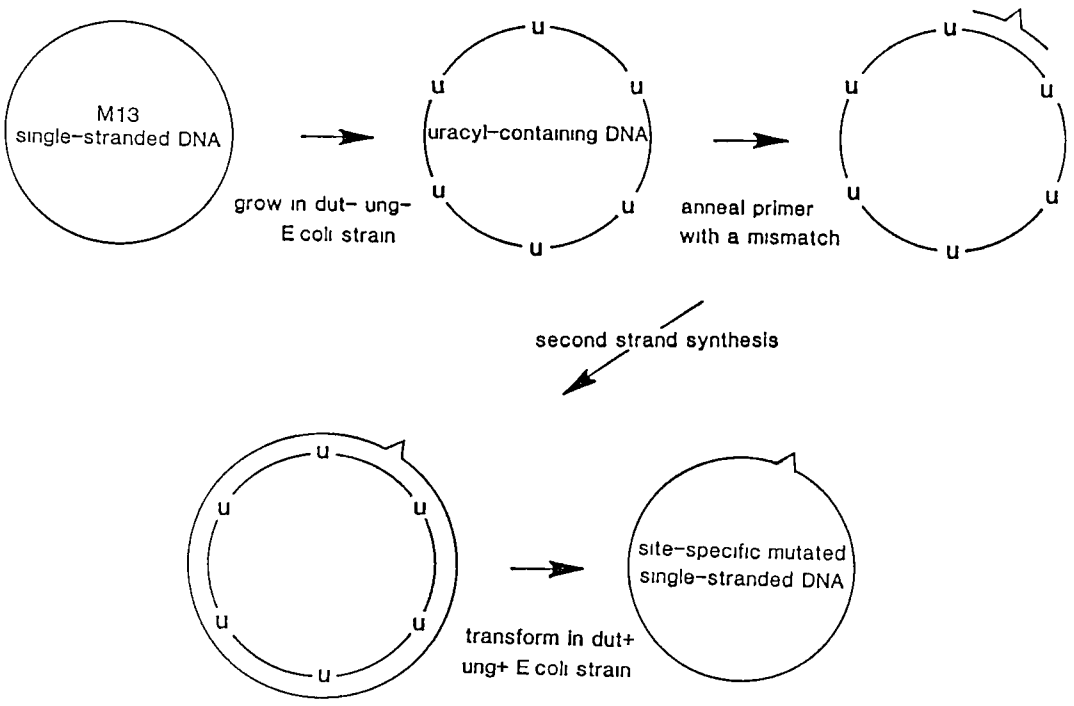
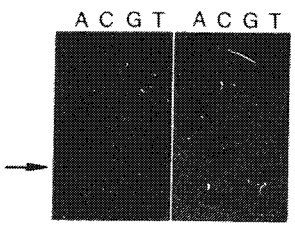

M13
single-stranded DNA
grow in dut- ung-
E coli strain
uracyl-containing DNA
anneal primer
with a mismatch
second strand synthesis
transform in dut+
ung+ E coli strain
site-specific mutated
single-stranded DNA

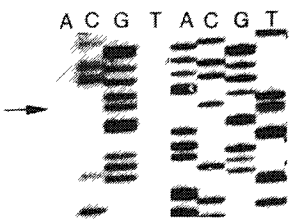
Figure 23. Sequence analysis of mutant of the HPIV3 fusion protein gene. Sequencing reactions using specific primers for the HPIV3 F gene were done in the presence dideoxynucleotides. The dideoxynucleotide used in each of the 4 reactions is indicated above the corresponding lane. The sequencing reactions were electrophoresed on a 8M urea-8% polyacrylamide gel. For each cysteine residue, the first set of 4 reactions represents the wild type sequence and the second set represents the mutated sequence. The arrows indicate the positions of the changes from a G to a C.

CYS 18

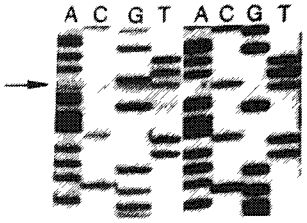


CYS 63

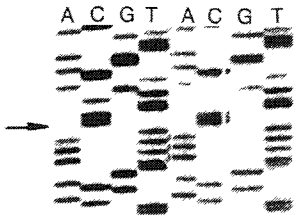
CYS 192



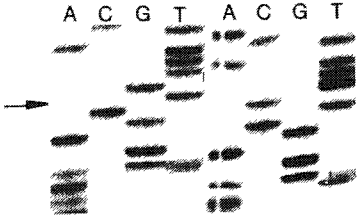
CYS 331



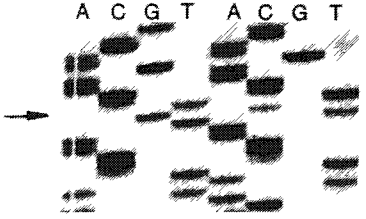
CYS 340



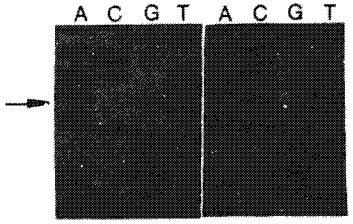
CYS 355



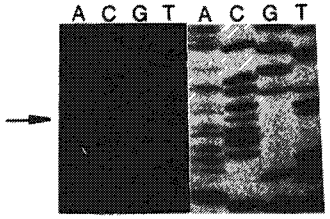
CYS 363



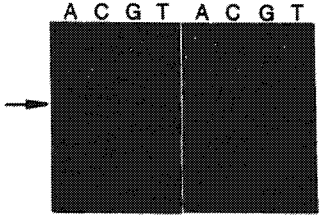
CYS 387



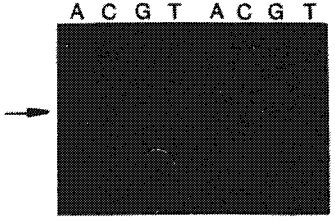
CYS 392



CYS 394



CYS 417



rabbit anti-HPIV3 serum and electrophoresed on a SDS-polyacrylamide gel following reduction with β-mercaptoethanol. The results are shown in Figure 24. The HPIV3 fusion protein with a serine instead of a cysteine at position 18 (VFCys18) had an electrophoretic profile similar to the wild type fusion protein, that is, F_0 and F_1 bands. However, all of the other mutant fusion proteins migrated as single bands at the position of F_0. Extra bands under F_0 seen for cysteine mutant 394 and 417 are seen faintly in the Vsc8 lane and probably represent degradation products; no bands corresponding to F_1 were detected.

Another set of Vsc8, VF and VFCys18 to VFCys417 infections were processed the same way except that the immunoprecipitates were not reduced prior to SDS-polyacrylamide gel electrophoresis. The results of this experiment are shown in Figure 25. Wild type and VFCys18 showed a doublet corresponding to F_0 and $F_{1,2}$. Although the positions of $F_{1,2}$ and F_0 are indicated, it is not known which species corresponds to either the upper or the lower band. VFCys63 and VFCys192 both showed a band running slightly slower than the doublet and a second band even more slowly was seen for VFCys63. All the other mutant proteins (VFCys 331 to 417) had bands running between F_0 and $F_{1,2}$.

DISCUSSION

The fifth objective of this project was to investigate the importance of cysteine residues for structure and/or function of the HPIV3 fusion protein and to identify the cysteine residues believed to be responsible for the disulfide bond between F_1 and F_2.

Figure 24. Characterization of cysteine mutants of the HPIV3 fusion protein by reducing SDS-polyacrylamide gel electrophoresis. [^{3}H]leucine-labeled recombinant vaccinia virus infected-cells were harvested using RIPA buffer, proteins were immunoprecipitated with rabbit anti-HPIV3 serum and run on a SDS-10% polyacrylamide gel. Lane 1, Vsc8-infected cells; 2, VF-infected cells; 3, VFCys18; 4, VFCys63; 5, VFCys192; 6, VFCys331; 7, VFCys340; 8, VFCys355; 9, VFCys363; 10, VFCys387; 11, VFCys392; 12, VFCys394; 13, VFCys417.

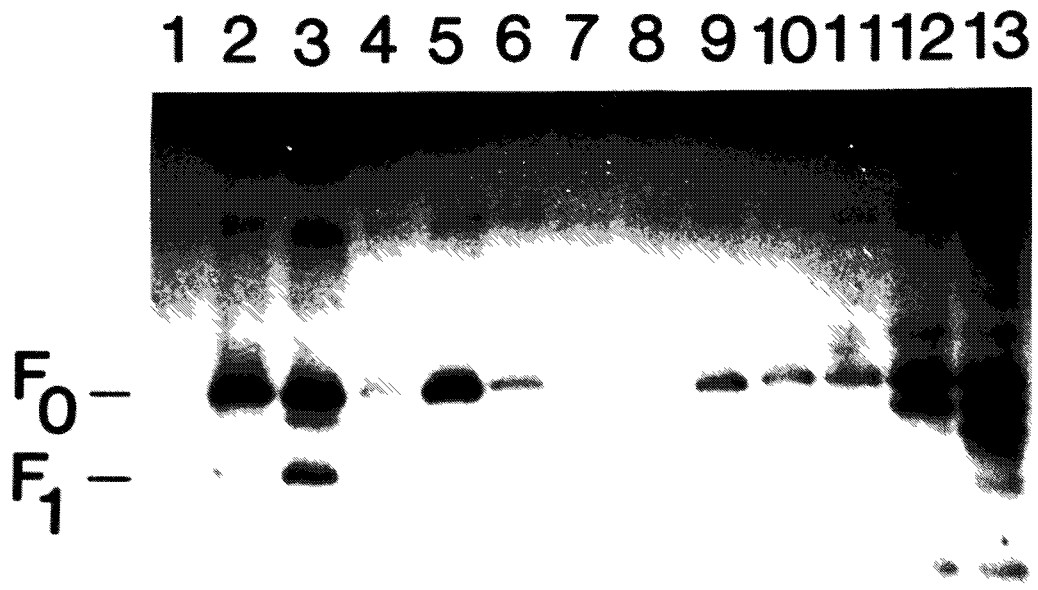
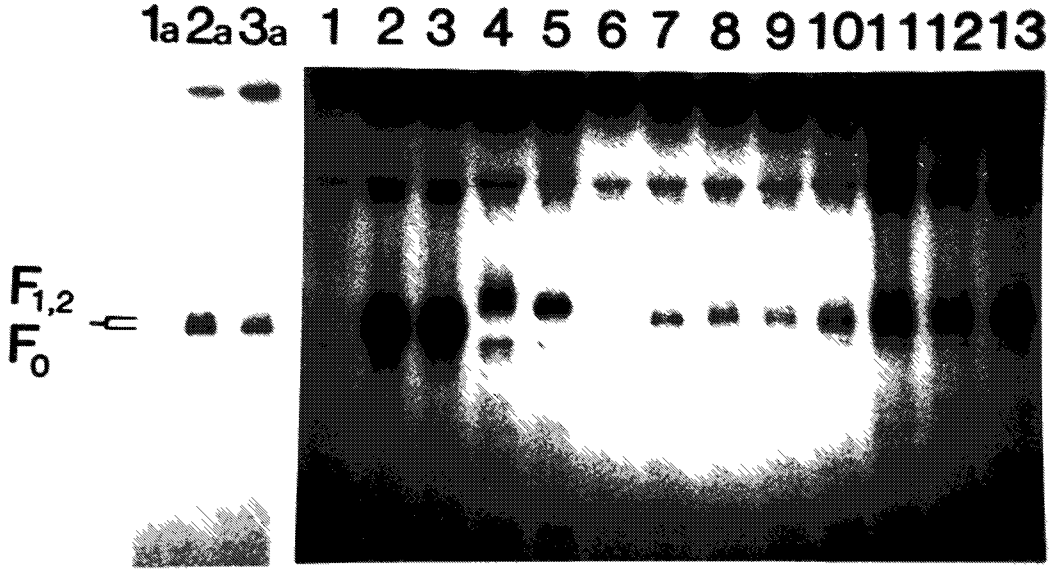

1 2 3 4 5 6 7 8 9 10 11 12 13
F_0
F_1

Figure 25. Characterization of cysteine mutants of the HPIV3 fusion protein by non-reducing SDS-polyacrylamide gel electrophoresis. [^{3}H] leucine-labeled recombinant vaccinia virus infected-cells were harvested using RIPA buffer, proteins were immunoprecipitated with rabbit anti-HPIV3 serum and run on a SDS-10% polyacrylamide gel under non-reducing condition. Lane 1, Vsc8-infected cells; 2, VF-infected cells; 3, VFCys18; 4, VFCys63; 5, VFCys192; 6, VFCys331; 7, VFCys340; 8, VFCys355; 9, VFCys363; 10, VFCys387; 11, VFCys392; 12, VFCys394; 13, VFCys417. Lanes 1a, 2a, 3a, shorter exposure of lanes 1, 2, 3.


1a 2a 3a 1 2 3 4 5 6 7 8 9 10 11 12 13
F1,2
F0

1. Characterization of the HPIV3 fusion protein expressed using a vaccinia virus vector

A first attempt at expressing the HPIV3 fusion protein was done using a transient expression system. The F gene was cloned into plasmid pSVL under the control of a SV40 late promotor. The recombinant plasmid was then transfected into COS 1 cells which are CV-1 cells expressing the SV40 large T antigen, required for SV40 replication. This system had been successfully used in this department (Garson and Kang, 1986), however, expression of the HPIV3 fusion protein could not be detected using this system. Southern blot analysis demonstrated presence of the plasmid, but no F mRNA was detected by Northern blot analysis. A possible explanation for this result is the presence of cryptic splicing sites in the HPIV3 F gene. Since HPIV3 replicates in the cytoplasm of infected cells, HPIV3 mRNAs are not spliced. However, the F gene contains consensus sequences which may be recognized by cellular splicing enzymes. This could prevent functional F mRNA from reaching the cytoplasm and, accordingly, expression of the viral glycoprotein. As a result, vaccinia virus was chosen as an expression system because of its cytoplasmic site of replication. Vaccinia virus has been used to express a variety of viral glycoproteins, including the fusion and HN proteins of HPIV3 (Spriggs *et al.*, 1987b), the fusion protein of Newcastle disease virus (Espion *et al.*, 1987), the fusion and HN proteins of simian virus 5 (Paterson *et al.*, 1987), the fusion and the hemagglutinin proteins of measles virus (Drillien *et al.*, 1988) and the fusion protein of human respiratory syncytial virus (Olmsted *et al.*, 1986; Wertz *et al.*, 1987).

Both vaccinia virus recombinants VF70 and VF193 expressed proteins corresponding to the F_0 precursor and the F_1 protein (Figure 19). The pattern of fusion protein glycosylation is assumed to be normal since the electrophoretic mobility of F_0 and F_1, expressed from VF70 and VF193, is similar to that of the corresponding HPIV3 glycoprotein. F_0 appears to be cleaved properly and the F_0/F_1 ratio is similar to the ratio in HPIV3-infected cells. VF193 lacks the 193 nucleotide 5' non-translated region of the HPIV3 F gene. VF70 contains this region and also possesses 70 nucleotides from the 3'-end of the HPIV3 M gene and the intergenic sequence, GAA. No increase or decrease of F_0 or F_1 expression was seen as a result of eliminating the 5' non-translated region of the HPIV3 F gene. One might expect a change in the level of expression of the fusion protein if the sequence was involved in translational regulation, however, since the F gene is under transcriptional and maybe translational regulation of vaccinia virus, it may not be surprising that elimination of the HPIV3 F gene 5' non-translated region did not lead to a change in its level of expression.

Usually, the biological activity of the fusion protein in HPIV3-infected cells can be detected either by observing syncytia formation resulting from cell-cell fusion or by hemolysis. Vaccinia virus cytopathic effects appeared very early during infection and were so extensive that syncytia formation could not be detected in VF-infected cells. Therefore, the biological activity of the fusion protein expressed by vaccinia virus, was assessed in a hemolysis assay. Membrane fragments of infected-cells with fusion protein at the surface, were added to a suspension of erythrocytes. Fusion between the erythrocyte and the infected-cell membranes occurs and since erythrocytes do not possess a membrane

repair mechanism, the erythrocytes are lysed (Homma *et al.*, 1976). The hemolysis assay worked very well with HPIV3-infected cells, however, hemolysis could not be demonstrated with vaccinia virus-infected cells expressing the HPIV3 fusion protein. Hsu *et al.* (1979) demonstrated that reconstituted membranes containing only the Sendai virus fusion protein required the presence of an agglutinin to elicit hemolytic activity. These results suggested that the attachment protein (HN) or some other agglutinin is necessary to bring the cell membranes together so that fusion can take place. Wheat germ agglutinin, therefore, was added to the hemolysis assay but did not result in increased hemolytic activity in VF-infected cells. A recombinant vaccinia virus containing the HPIV3 HN gene (VHN) was isolated, amplified and used to coinfect CV-1 cells along with VF. This combination also did not show any hemolysis. There are a few possible explanations for the lack of fusion activity for the HPIV3 fusion protein expressed by VF. It is possible that vaccinia virus gene products may interfere with fusion activity. To test this possibility, HPIV3-infected cells were superinfected 18 hours later with vaccinia virus and processed for hemolysis. The results showed hemolytic activity equal to HPIV3-infected CV-1 cells (data not shown). This demonstrated that once the fusion protein is expressed on the cell surface, vaccinia virus gene products do not interfere with hemolytic activity. There is also a possibility that the HPIV3 fusion and HN proteins expressed by vaccinia virus are not presented properly at the cell surface because of the absence of another viral product, such as the HPIV3 M protein, which is believed to interact with membrane area containing the viral glycoproteins. Simian virus 5 fusion protein was shown to be biologically active on its own (Paterson *et al.*, 1985), but maybe HPIV3 fusion protein

requires the involvement of another viral gene product(s). The ratio of the fusion protein to HN or M proteins may be another requirement for the fusogenic activity of the HPIV3 fusion protein. However, more investigation needs to be done in order to solve this problem.

2. Characterization of cysteine mutants

To investigate the importance of cysteine residues for the maintenance of structure and/or function of the HPIV3 fusion protein, site-specific mutants of the fusion gene were generated by changing each cysteine codon into a serine codon. After confirmation of each mutation by sequence analysis, a recombinant vaccinia virus was constructed for each cysteine mutant. Following analysis by SDS-polyacrylamide gel electrophoresis under reducing conditions, none of the cysteine mutants, except for the one with a mutation at cysteine 18, showed a protein product with the electrophoretic mobility of F_1. This suggests that fusion proteins with altered cysteines, other than at position 18, are not cleaved into $F_{1,2}$, and therefore are not functional (Homma and Ohuchi, 1973; Scheid and Choppin, 1974; Nagai *et al.*, 1976b). Cleavage may not occur for two possible reasons: 1) the loss of a disulfide bond may produce a change in the conformation of the protein making the cleavage-activation site inaccessible to the enzyme responsible for cleavage of F_0 into $F_{1,2}$ or, 2) a change in conformation may prevent proper transport of the fusion protein to the intracellular compartment where the host enzyme with trypsin-like activity may reside, or to the surface of the cell, where cleavage may also occur. The fusion protein mutant with a serine at position 18 is the only one which appeared to be cleaved properly. This cysteine is probably located at the end of the signal

peptide, and may get cleaved from the F_0 protein along with the signal peptide during translocation and, therefore, would not be expected to play any major role in maintaining the tertiary structure of the fusion protein. Cysteine 18 is also the only cysteine residue in the HPIV3 fusion protein that is not conserved in fusion proteins of other members of the family *Paramyxoviridae*.

Cysteine mutants 63 to 417 expressed their protein products at low levels. Since the substitution of serine for cysteine residues probably results in conformational changes in the fusion protein, it is possible that the altered fusion proteins may no longer be recognized by antibodies directed against the fusion protein. Therefore, we attempted to used anti-HPIV3 rabbit serum to detect fusion proteins on Western blots. This experiment failed to reveal any fusion protein, even with cells infected with wild type. Another attempt was made using non-neutralizing HPIV3 fusion protein specific monoclonal antibodies, kindly provided by K.L. van wyke Coelingh (National Institute of Allergy and Infectious Diseases, Bethesda, MD)(Coelingh *et al.*, 1989). It was hoped that non-neutralizing monoclonal antibodies might detect linear epitopes and not conformational epitopes and, therefore, might identify the denatured F_0 and F_1 proteins. This second attempt also failed. The other possible explanation for the low levels of expression of altered fusion proteins is that the incorrectly folded fusion proteins may be blocked in transport or unstable and as a result are degraded more rapidly than the normal F gene product. Mutation of two cysteines located in extracellular domains of bovine rhodopsin yielded rhodopsins that were expressed at very low levels. These two cysteine residues are conserved in all vertebrate rhodopsins and appear to be essential for biological activity (Karnik *et al.*, 1988).

Since cysteines 63 to 417 are remarkably conserved among the family *Paramyxoviridae*, one may expect them to be essential for the maintenance of the tertiary structure and since proper folding is believed to be necessary for efficient intracellular transport (Gething *et al.*, 1986), it would not be surprising to see dramatic changes in the level of expression, processing and turnover of the fusion protein.

The results obtained by electrophoresis of the fusion proteins under non-reducing conditions demonstrated that the protein product of the cysteine 18 mutant was behaving like the wild-type fusion protein; the doublet of F_0 and $F_{1,2}$ was easily recognized. However, cysteine mutants 331 to 417 yielded a broad protein band between F_0 and $F_{1,2}$. Cysteine residues 331 to 417 are part of the cysteine cluster (Figure 21), therefore conversion of anyone of them into a serine residue appears to prevent cleavage of F_0. A change in the migration of the mutant fusion protein in non-reducing conditions is expected because of the disruption of a disulfide bond, but it was not possible to pair the cysteine residues since there were no characteristic patterns in their electrophoretic mobilities. The fusion proteins altered at cysteine residues 63, located on F_2, and 192, located on F_1 near the hydrophobic N-terminus, had the same electrophoretic mobility, slightly slower than the $F_0/F_{1,2}$ doublet. This strongly suggests that cysteines 63 and 192 normally form the disulfide bond that links F_2 to F_1. Because these two cysteines are quite a distance from each other as compared to the ones in the cysteine cluster, rupture of the disulfide bond between cysteine 63 and 192 might be expected to result in a F_0 protein with a more relaxed tertiary structure which would migrate more slowly under non-reducing conditions. The disruption of a disufide bond in the cluster might not affect mobility as

drastically since this region contains not only the remaining cysteines, but also a large number of proline residues and likely has a high degree of folding.

It was noticed, under both reducing and non-reducing conditions, that an additional protein was present in the lanes of the cysteine 63 mutant. Since this band was seen following both types of electrophoresis conditions, it was assumed that the slow migration was due to increased size of the protein. A close look at the amino acid sequence of this altered F gene showed that a new potential carbohydrate acceptor site was generated; Asn-Ser-Cys was converted to Asn-Ser-Ser. If this new potential acceptor site is glycosylated some of the time, this may explain the presence of two species of F_0 for this particular mutant.

CHAPTER 6 CONCLUSIONS

The first objective of this project was to elucidate the genome organization of HPIV3. A comparison of the genome organization of HPIV3 with several members of the family *Paramyxoviridae* will provide a better understanding of the evolutionary relationships within this family of viruses.

The strategy for determining the gene order of HPIV3 was to generate cDNA libraries from HPIV3 mRNA and genomic RNA, respectively. By cross-hybridization analysis, 5 nonhomologous groups of clones were identified. Viral proteins NP, P, M, F and HN were assigned to each group by hybridization-selection and translation of individual HPIV3 mRNAs (Storey, 1987). cDNA clones representing each of the HPIV3 genes were used to identify, in the genomic cDNA library, clones spanning more than one gene. Using this approach, a sixth group of clones, representing the HPIV3 L gene, was identified and the gene order of HPIV3 was determined to be: 3'-NP-P/C-M-F-HN-L-5'.

The organization of the HPIV3 gene boundary regions was investigated by determining the nucleotide sequence of cDNA clones spanning more than one gene. The nucleotide sequences of the 5' termini of HPIV3 mRNAs were obtained by primer extension analysis. Consensus sequences found at the beginning and the end of each gene are believed to be polymerase initiation and termination sites and a trinucleotide intergenic sequence, GAA, was also identified.

A comparison of the genome organization of members of the family *Paramyxoviridae* including HPIV3 shows a conserved gene order. However, some differences are apparent: some paramyxoviruses have extra genes,

such as simian virus 5, mumps virus and respiratory syncytial virus; there are different strategies for the use of a second open reading frame in the P/C (or V) gene; there are differences in the degree of conservation of the number of nucleotides in the intergenic sequences. These observation suggest, that the genus *Paramyxovirus* should be subdivided into at least two groups: the parainfluenza viruses and Sendai virus in one group and Newcastle disease virus, mumps virus and simian virus 5 in another group.

Because of the importance of HPIV3 fusion protein in pathogenesis, the second part of the project was to determine the complete nucleotide sequence of the F gene. The predicted amino acid sequence of the HPIV3 fusion protein was compared with the sequences of the fusion proteins of other members of the family *Paramyxoviridae* to identify conserved sequences among paramyxovirus fusion proteins that are likely to be important for structure and/or function of the fusion protein.

The complete nucleotide sequence of the HPIV3 F gene was determined by sequencing 4 overlapping cDNA clones. The HPIV3 F gene was 1851 nucleotides long and contained a long open reading frame with a coding capacity for a polypeptide of 539 amino acids with a predicted molecular weight of 60,031. Four potential glycosylation sites were identified. A hydropathy plot showed three major hydrophobic regions corresponding to a signal peptide, the N-terminus of F_1 and a transmembrane region.

Comparison of the amino acid sequence of the HPIV3 fusion protein with the sequences of the fusion proteins of other members of the family *Paramyxoviridae* showed common features, such as overall hydrophobicity and the conserved location of cysteine residues and, to a lesser extent, proline residues. Ten cysteine residues were remarkably conserved in all

the fusion proteins examined. The HPIV3 F_0 protein showed a very high degree of similarity to the F_0 proteins of bovine parainfluenza virus 3 (80%), Sendai virus (43%) and human parainfluenza virus 1 (41%) (Cote *et al.*, 1987; Suzu *et al.*, 1987; Merson *et al.*, 1988). The F_0 protein of Newcastle disease virus, mumps virus and simian virus 5 showed a lower degree of similarity with HPIV3 F_0 protein supporting the idea that the classification of the genus *Paramyxovirus* should be reassessed.

The final objective of this project was to investigate the importance of cysteine residues for the structure and/or function of the HPIV3 fusion protein and to identify the residues responsible for the disulfide linkage between F_1 and F_2.

A vaccinia virus expression system was chosen for expression of the HPIV3 fusion protein. Recombinant vaccinia virus expressed products that had the same characteristics as the HPIV3 F_0 and F_1 proteins. Site-specific mutagenesis was used to change individual cysteine codons into serine codons. Each cysteine mutant was expressed using vaccinia virus and the gene products were analyzed by reducing and non-reducing SDS-polyacrylamide gel electrophoresis.

10 of the 11 cysteines of HPIV3 fusion protein appeared to be essential for structure and/or function of the fusion protein since none of the 10 mutant proteins were cleaved. It is not clear if the lack of cleavage is due to a structural conformation that inhibits cleavage or to a problem in intracellular transport of the proteins, which prevents access to the host enzyme required for cleavage-activation. The same mutant proteins appeared to be expressed at very low levels, which is believed to be due to degradation of improperly folded proteins. The possibility that a conformational change prevents recognition by the antibodies used for

immunopreciptation cannot be ruled out, however. Analysis of the expressed proteins under non-reducing conditions suggests that cysteine residues 63 and 192 form the disulfide linkage between F_1 and F_2, since non-denatured products of these two mutants had the same electrophoretic mobility and different from the others.

These results stress the importance of cysteine residues for structure and function of the fusion protein and explain their remarkable conservation within the family *Paramyxoviridae*.

During the last set of experiments, several questions were raised and future experiments may help to answer them; 1) It is not known if the fusion protein of HPIV3 is cleaved intracellularly or extracellularly. Experiments such as pulse-chase labeling and protease digestion of extracellular protein would allow us to determine which form of the fusion protein, (F_0 or $F_{1,2}$), reaches the cell surface; 2) It is not known if the low levels of expression of the fusion protein cysteine mutants are due to rapid turnover of the mutant proteins or if conformational changes in the mutant fusion proteins prevent recognition by antibodies. Production of antibodies against fusion protein oligopeptides may generate antibodies that recognize linear epitopes and may be used for Western blot analysis; 3) The biological assay used to test any future mutations of the HPIV3 fusion protein has to be developed. Co-expression of the HPIV3 M gene along with the HN and F gene may be tested for hemolysis activity. Infection of cells that do not demonstrate early cytopathic effects caused by vaccinia virus may allow syncytium formation to be visualized.

Other regions of the fusion protein sequence could be mutated, for example basic residues could be added to the cleavage-activation site to see if the ratio of F_0/F_1 decreases with the number of basic residues added. Mutations can also be introduced at other amino acid residues such as prolines, or residues of the hydrophobic N-terminus of F_1. There is no need to say that determination of the structure of the fusion protein by X-ray crystallography would be priceless for the planning of future mutations.

LIST OF REFERENCES

Abraham, G. (1979) The effect of ultraviolet radiation on the primary transcription of Influenza virus messenger RNAs. *Virology* **97**: 177-182.

Ball, L.A. (1977) Transcriptional mapping of vesicular stomatitis virus *in vivo*. *J. Virol.* **21**: 411-414.

Banerjee, A.K. (1987) Transcription and replication of rhabdoviruses. *Microbiol. Rev.* **51**: 66-87.

Barrett, T., Shrimpton, S.B. and Russell, S.E.H. (1985) Nucleotide sequence of the entire protein coding region of canine distemper virus polymerase-associated (P) protein mRNA. *Virus Res.* **3**: 367-372.

Barrett, T., Clarke, D.K., Evans, S.A. and Rima, B.K. (1987) The nucleotide sequence of the gene encoding the F protein of canine distemper virus: a comparison of the deduced amino acid sequence with other paramyxoviruses. *Virus Res.* **8**: 373-386.

Baybutt, H.N. and Pringle, C.R. (1987) Molecular cloning and sequencing of the F and 22K membrane protein genes of the RSS-2 strain of respiratory syncytial virus. *J. gen. Virol.* **68**: 2789-2796.

Bellini, W.J., Englund, G., Rozenblatt, S., Arnheiter, H. and Richardson, C.D. (1985) Measles virus P gene codes for two proteins. *J. Virol.* **53**: 908-919.

Billeter, M.A., Baczko, K., Schmid, A. and ter Meulen V. (1984) Cloning of DNA corresponding to four different measles virus genomic regions. *Virology* **132**: 147-159.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7**: 1513-1523.

Blobel, G. (1980) Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. USA* **77**: 1496-1500.

Blumberg, B.M., Giorgi, C. and Kolakofsky, D. (1983) N protein of vesicular stomatitis virus selectively encapsidates leader RNA *in vitro*. *Cell* **32**: 559-567.

Blumberg, B.M., Giorgi, C., Rose, K. and Kolakofsky, D. (1985) Sequence determination of the Sendai virus fusion protein gene. *J. gen. Virol.* **66**: 317-331.

Buckland, R., Gerald, C., Barker, R. and Wild, T.F. (1987) Fusion glycoprotein of measles virus: nucleotide sequence of the gene and comparison with other paramyxoviruses. *J. gen. Virol.* **68**: 1695-1703.

Carlsen, S.R., Peluso, R.W. and Moyer, S.A. (1985) *In vitro* replication of Sendai virus wild-type and defective interfering particle genome RNAs. *J. Virol.* **54**: 493-500.

Cattaneo, R., Rebmann, G., Schmid, A., Baczko, K., ter Meulen, V. and Billeter, M.A. (1987) Altered transcription of a defective measles virus genome derived from a diseased human brain. *EMBO journal* **6**: 681-688.

Chakrabarti, S., Brechling, K. and Moss, B. (1985) Vaccinia virus expression vector: coexpression of β-galactosidase provides visual screening of recombinant virus plaques. *Mol. Cell. Biol.* **5**: 3403-3409.

Chambers, P., Millar, N.S. and Emmerson, P.T. (1986) Nucleotide sequence of the gene encoding the fusion glycoprotein of Newcastle disease virus. *J. gen. Virol.* **67**: 2685-2694.

Chanock, R.M. and McIntosh, K. (1985) Parainfluenza viruses. in B.N. Fields *et al.* (eds.) *Virology*. Raven Press, New York. p. 1241-1253.

Chao, C.C.-K., Bird, P., Gething, M.-J. and Sambrook, J. (1987) Posttranslational translocation of influenza virus hemagglutinin across microsomal membranes. *Mol. Cell. Biol.* 7: 3842-3845.

Chatis, P.A. and Morrison, T.G. (1982) Fatty acid modification of Newcastle disease virus glycoproteins. *J. Virol.* **43**: 342-347.

Chen, C., Compans, R.W. and Choppin, P.W. (1971) Parainfluenza virus surface projections: glycoproteins with haemagglutinin and neuraminidase activities. *J. Gen. Virol.* **11**: 53-58.

Chen, C. and Okayama, H. (1987) High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* 7: 2745-2752.

Coelingh, K.L. Van Wyke and Tierney, E.L. (1989) Antigenic and fuctional organization of human parainfluenza virus type 3 fusion glycoprotein. *J. Virol.* **63**: 375-382.

Collins, P.L., Hightower, L.E. and Ball, L.A. (1980) Transcriptional map for Newcastle disease virus. *J. Virol.* **35**: 682-693.

Collins, P.L., Wertz, G.W., Ball, L.A. and Hightower, L.E. (1982) Coding assignments of the five smaller mRNAs of Newcastle disease virus. *J. Virol.* **43**: 1024-1031.

Collins, P.L., Huang, Y.T. and Wertz, G.W. (1984) Nucleotide sequence of the gene encoding the fusion (F) glycoprotein of human respiratory syncytial virus. *Proc. Natl. Acad. Sci. USA* **81**: 7683-7687.

Collins, P.L., Dickens, L.E., Buckler-White, A., Olmsted, R.A., Spriggs, M.K., Camargo, E. and Coelingh, K.L. Van Wyke (1986) Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic struture and gene order. *Proc. Natl. Acad. Sci. USA* **83**: 4594-4598.

Colonno, R.J. and Banerjee, A.K. (1976) A unique RNA species involved in initiation of vesicular stomatitis virus RNA transcription *in vitro*. *Cell* **8**: 197-204.

Cote, M.-J., Storey, D.G., Kang, C.Y. and Dimock, K. (1987) Nucleotide sequence of the coding and flanking regions of the human parainfluenza virus type 3 fusion glycoprotein gene. *J. gen. Virol.* **68**: 1003-1010.

Dale, R.M.K., McClure, B.A. and Houchins, J.P. (1985) A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: Application to sequencing the corn mitochondrial 18 S rDNA. *Plasmid* **13**: 31-40.

Davis, N.G. and Hsu, M.-C. (1986) The fusion-related hydrophobic domain of Sendai F protein can be moved through the cytoplasmic membrane of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **83**: 5091-5095.

Dimock, K, Rud, E.W. and Kang, C.Y. (1986) 3'-Terminal sequence of human parainfluenza virus 3 genomic RNA. *Nucl. Acids Res.* **14**: 4694.

Dimock, K., Storey, D.G., Cote, M.-J. and Kang, C.Y. (1987) Cloning, coding assignments and mapping of human parainfluenza virus 3 genes. In B.W. Mahy and D. Kolakofsky (eds.) *The biology of negative strand viruses*. Elsevier Biomedical Press, Amsterdam. p.213-220.

Dowling, P.C., Giorgi, C., Roux, L., Dethlefsen, L.A., Galantowicz, M.E., Blumberg, B.M. and Kolakofsky, D. (1983) Molecular cloning of the 3'-proximal third of Sendai virus genome. *Proc. Natl. Acad. Sci. USA* **80**: 5213-5216.

Doyle, C., Roth, M.G., Sambrook, J. and Gething, M.-J. (1985) Mutations in the cytoplasmic domain of the influenza virus hemagglutinin affect different stages of intracellular transport. *J. Cell. Bio.* **100**: 704-714.

Drillien, R., Spehner, D., Kirn, A., Giraudon, P., Buckland, R., Wild, F. and Lecocq, J.-P. (1988) Protection of mice from fatal measles encephalitis by vaccination with vaccinia virus recombinants encoding either the hemagglutinin or the fusion protein. *Proc. Natl. Acad. Sci. USA* **85**: 1252-1256.

Elango, N., Satake, M., Coligan, J.E., Norrby, E., Camargo, E. and Venkatesan, S. (1985) Respiratory syncytial virus fusion glycoprotein: nucleotide sequence of mRNA, identification of cleavage activation site and amino acid sequence of N-terminus of F_1 subunit. *Nucl. Acids Res.* **13**: 1559-1574.

Elango, N., Coligan, J.E., Jambou, R.C. and Venkatesan, S. (1986) Human parainfluenza type 3 virus hemagglutinin-neuraminidase glycoprotein: nucleotide sequence of mRNA and limited amino acid sequence of the purified protein. *J. Virol.* **57**: 481-489.

Elango, N., Varsanyi, T., Kovamees, J. and Norrby, E. (1988) Molecular cloning and characterization of six genes, determination of gene order and intergenic sequences and leader sequence of mumps virus. *J. gen. Virol.* **69**: 2893-2900.

Espion, D., De Henau, S., Letellier, C., Wemres, C.-D., Brasseur, R., Young, J.F., Gross, M., Rosenberg, M., Meulemans, G. and Burny, A. (1987) Expression at the cell surface of native fusion protein of the Newcastle disease virus (NDV) strain Italien from cloned cDNA. *Arch. Virol.* **95**: 79-95.

Flamand, A. and Delagneau, J.F. (1978) Transcriptional mapping of rabies virus *in vivo*. *J. Virol.* **28**: 518-523.

Freedman, R.B. (1984) Native disulphide bond formation in protein biosynthesis: evidence for the role of protein disulphide isomerase. *Trends biochem. Sci.* **9**: 438-441.

Galinski, M.S., Mink, M.A., Lambert, D.M., Wechsler, S.L. and Pons, M.W. (1986a) Molecular cloning and sequence analysis of the human parainfluenza 3 virus RNA encoding the nucleocapsid protein. *Virology* **149**: 139-151.

Galinski, M.S., Mink, M.A., Lambert, D.M., Wechsler, S.L. and Pons, M.W. (1986b) Molecular cloning and sequence analysis of the human parainfluenza 3 virus mRNA encoding the P and C proteins. *Virology* **155**: 46-60.

Galinski, M.S., Mink, M.A., Lambert, D.M., Wechsler, S.L. and Pons, M.W. (1987a) Molecular cloning and sequence analysis of the human parainfluenza 3 virus gene encoding the matrix protein. *Virology* **157**: 24-30.

Galinski, M.S., Mink, M.A. and Pons, M.W. (1987b) Molecular cloning and sequence analysis of the human parainfluenza 3 virus genes encoding the surface glycoproteins, F and HN. *Virus Res.* **8**: 205-215.

Galinski, M.S., Mink, M.A. and Pons, M.W. (1988) Molecular cloning and sequence analysis of the human parainfluenza 3 virus gene encoding the L protein. *Virology* **165**: 499-510.

Garson, K. and Kang, C.Y. (1986) Identification of the v-rel protein in REV-T transformed chicken bone marrow cells and expression in COS 1 cells. *Biochem. Biophys. Res. Comm.* **134**: 716-722.

Gething, M.-J., White, J.M. and Waterfield, M.D. (1978) Purification of the fusion protein of Sendai virus: analysis of the NH_2-terminal sequence generated during precursor activation. *Proc. Natl. Acad. Sci. USA.* **75**: 2737-2740.

Gething, M.-J. and Sambrook, J. (1982) Construction of influenza heamagglutinin genes that code for intracellular and secreted forms of the protein. *Nature* **300**: 598-603.

Gething, M.-J., McCammon, K. and Sambrook, J. (1986) Expression of wild-type and mutant forms of influenza hemagglutinin: the role of folding in intracellular transport. *Cell* **46**: 939-950.

Gibson, S., Bundo-Morita, K., Portner, A. and Lenard, J. (1988) Fusion of a Sendai mutant deficient in HN protein (ts271) with cardiolipin liposomes. *Virology* **163**: 226-229.

Giorgi, C., Blumberg, B.M. and Kolakofsky, D. (1983) Sendai virus contains overlapping genes expressed from a single mRNA. *Cell* **35**: 829-836.

Glazier, K., Raghow, R. and Kingsbury, D.W. (1977) Regulation of Sendai virus transcription: evidence for a single promotor *in vivo*. *J. Virol.* **21**: 863-871.

Glezen, W.P. and Denny, F.W. (1973) Epidemiology of acute lower respiratory disease in children. *N. Engl. J. Med.* **288**: 498-505.

Glezen, W.P., Frank, A.L., Taber, L.H. and Kasel, J.A. (1984) Parainfluenza virus type 3: seasonality and risk of infection and reinfection in young children. *J. Inf. Dis.* **150**: 851-857.

Glickman, R., Syddall, R.J., Iorio, R.M., Sheehan, J.P. and Bratt, M.A. (1988) Quantitative basic residue requirements in the cleavage-activation site of the fusion glycoprotein as a determinant of virulence for Newcastle disease virus. *J. Virol.* **62**: 354-356.

Gupta, K.C. and Kingsbury, D.W. (1984) Complete sequences of the intergenic and mRNA start signals in the Sendai virus genome: homologies with the genome of vesicular stomatitis virus. *Nucl. Acids Res.* **12**: 3829-3841.

Guskey, L.E. and Bergtrom, G. (1981) High yield growth and purification of human parainfluenza type 3 virus and initial analysis of viral structural proteins. *J. gen. Virol.* **54**: 115-123.

Hackett, P.B., Sauerbier, W. (1975) The transcriptional organization of the ribosomal RNA genes in mouse L cells. *J. Mol. Biol.* **91**: 235-256.

Hasel, K.W., Day, S., Millward, S., Richardson, C.D., Bellini, W.J. and Greer, P.A. (1987) Characterization of cloned measles virus mRNAs by *in vitro* transcription, translation, and immunoprecipitation. *Intervirology* **28**: 26-39.

Hiebert, S.W., Paterson, R.G. and Lamb, R.A. (1985) Identification and predicted sequence of a previously unrecognized small hydrophobic protein, SH, of the paramyxovirus simian virus 5. *J. Virol.* **55**: 744-751.

Herrler, G. and Compans, R.W. (1982) Synthesis of mumps virus polypeptides in infected vero cells. *Virology* **119**: 431-438.

Homma, M. and Ohuchi, M. (1973) Trypsin action on the growth of Sendai virus in tissue culture cells. *J. Virol.* **12**: 1457-1465.

Homma, M., Shimizu, K., Shimizu, Y.K. and Ishida, N. (1976) On the study of Sendai virus hemolysis. 1. Complete Sendai virus lacking in hemolytic activity. *Virol.* **71**: 41-47.

Hsu, D., Yamanaka, M., Miller, J., Dale, B., Grubman, M. and Yilma, T. (1988) Cloning of the fusion gene of rinderpest virus: comparative sequence analysis with other morbilliviruses. *Virology* **166**: 149-153.

Hsu, M.-C., Scheid, A. and Choppin, P.W. (1979) Reconstitution of membranes with individual paramyxovirus glycoproteins and phospholipid in cholate solution. *Virology* **95**: 476-491.

Hsu, M.-C., Scheid, A. and Choppin, P.W. (1981) Activation of the Sendai virus fusion protein (F) involves a conformational change with exposure of a new hydrophobic region. *J. Biol. Chem.* **256**: 3557-3563.

Hsu, M.-C. and Choppin, P.W. (1984) Analysis of Sendai virus mRNAs with cDNA clones of viral genes and sequences of biologically important regions of the fusion protein. *Proc. Natl. Acad. Sci. USA* **81**: 7732-7736.

Hsu, M.-C., Scheid, A. and Choppin, P.W. (1987) Protease activation mutants of Sendai virus: sequence analysis of the mRNA of the fusion protein (F) gene and direct identification of the cleavage-activation site. *Virology* **156**: 84-90.

Hull, J.D., Krah, D.L. and Choppin, P.W. (1987) Resistance of a measles virus mutant to fusion inhibitory oligopeptides is not associated with mutations in the fusion peptide. *Virology* **159**: 368-372.

Ishida, N., Taira, H., Omata, T., Mizumoto, K., Hattori, S., Iwasaki, K. and Kawakita, M. (1986) Sequence of 2,617 nucleotides from the 3' end of Newcastle disease virus genome RNA and the predicted amino acid sequence of viral NP protein. *Nucl. Acids Res.* **14**: 6551-6564.

Itoh, M. and Homma, M. (1988) Single amino acid change at the cleavage site of the fusion protein is responsible for both enhanced chymotrypsin sensitivity and trypsin resistance of a Sendai virus mutant, TR-5. *J. gen. Virol.* **69**: 2907-2911.

Jambou, R.C., Elango, N. and Venkatesan, S. (1985) Proteins associated with human parainfluenza virus type 3. *J. Virol.* **56**: 298-302.

Jambou, R.C., Elango, N., Venkatesan, S. and Collins, P.L. (1986) Complete sequence of the major nucleocapsid protein gene of human parainfluenza type 3 virus: comparison with other negative strand viruses. *J. gen. Virol.* **67**: 2543-2548.

Jorgensen, E.D., Collins, P.L. and Lomedico, P.T. (1987) Cloning and nucleotide sequence of Newcastle disease virus hemagglutinin-neuraminidase mRNA: identification of a putative sialic acid binding site. *Virology*: **156**: 12-24.

Karnik, S.S., Sakmar, T.P., Chen, H.-B. and Khorana, H.G. (1988) Cysteine residues 110 and 187 are essential for the formation of the correct structure in bovine rhodopsin. *Proc. Natl. Acad. Sci. USA* **85**: 8459-8463.

Kingsbury, D.W. (1977) Paramyxoviruses. In D.P. Nayak (ed.) *The molecular biology of animal viruses.* Marcel Dekker, New York. p.349-382.

Kingsbury, D.W., Bratt, M.A., Choppin, P.W., Hanson, R.P., Hosaka, Y., ter Meulen, V., Norrby, E., Plowright, W., Rott, R. and Wunner, W.H. (1978) *Paramyxoviridae. Intervirology* **10**: 137-152.

Kingsbury, D.W. (1985) Orthomyxo- and aramyxoviruses and their replication. In B.N. Fields *et al.*(eds.) *Virology*. Raven Press, New York. p. 1157-1178.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283-292.

Kunkel, T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**: 488-492.

Kurilla, M.G., Stone, H.O. and Keene, J.D. (1985) RNA sequence and transcriptional properties of the 3' end of the Newcastle disease virus genome. *Virology* **145**: 203-212.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105-132.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lamb, R.A., Mahy, B.W.J. and Choppin, P.W. (1976) The synthesis of Sendai virus polypeptides in infected cells. *Virology* **69**: 116-131.

Leppert, M., Rittenhouse, L., Perrault, J., Summers, D.F. and Kolakofsky, D. (1979) Plus and minus strand leader RNAs in negative strand virus-infected cells. *Cell* **18**: 735-747.

Luk, D., Sanchez, A. and Banerjee, A.K. (1986) Messenger RNA encoding the phosphoprotein (P) gene of human parainfluenza virus 3 is bicistronic. *Virology* **153**: 318-325.

Luk, D., Masters, P.S., Sanchez, A. and Banerjee, A.K. (1987) Complete nucleotide sequence of the matrix protein mRNA and three intergenic junction of human parainfluenza virus type 3. *Virology* **156**: 189-192.

Lyles, D.S. (1979) Glycoproteins of Sendai virus are transmembrane proteins. *Proc. Natl. Acad. Sci. USA* **76**: 5621-5625.

Mackett, M., Smith, G.L. and Moss, B. (1985) The construction and characterisation of vaccinia virus recombinants expressing foreign genes. In D.M. Glover (ed.) *DNA cloning, a practical approach*. IRL Press, Oxford. Volume II. p.191-211.

Mackett, M. and Smith, G.L. (1986) Vaccinia virus expression vectors. *J. gen. Virol.* **67**: 2067-2082.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning. A laboratory manual.* Cold Spring Harbor, NY.

Matthews, R. (1982) Classification and nomenclature of viruses. *Intervirology* **17**: 1-181.

McGinnes, L.W., Semerjian, A. and Morrison, T.G. (1985) Conformational changes in Newcastle disease virus fusion glycoprotein during intracellular transport. *J. Virol.* **56**: 341-348.

McGinnes, L.W. and Morrison, T.G. (1986) Nucleotide sequence of the gene encoding the Newcastle disease virus fusion protein and comparisons of aramyxovirus fusion protein sequences. *Virus Res.* **5**: 343-356.

McGinnes, L.W., Wilde, A. and Morrison, T.G. (1987) Nucleotide sequence of the gene encoding the Newcastle disease virus hemagglutinin-neuraminidase protein and comparisons of paramyxovirus hemagglutinin-neuraminidase protein sequences. *Virus Res.* **7**: 187-202.

Merson, J.R., Hull, R.A., Estes, M.K. and Kasel, J.A. (1988) Molecular cloning and sequence determination of the fusion protein gene of human parainfluenza virus type 1. *Virology* **167**: 97-105.

Merz, D.C., Scheid, A. and Choppin, P.W. (1980) Importance of antibodies to the fusion glycoprotein of paramyxoviruses in the prevention of spread of infection. *J. Exp. Med.* **151**: 275-288.

Merz, D.C. and Wolinsky, J.S. (1981) Biochemical features of mumps virus neuraminidases and their relationship with pathogenicity. *Virology* **114**: 218-227.

Messing, J. (1983) New M13 vectors for cloning. In R. Wu, L. Grossman and K. Moldave (eds.) *Methods in Enzymology.* Academic Press, New York. Volume **101**. p.20-78.

Miura, N., Ohtsuka, E., Yamaberi, N., Ikehara, M., Uchida, T. and Okada, Y. (1985) Use of the deoxyinosine-containing probe to isolate and sequence cDNA encoding the fusion (F) glycoprotein of Sendai virus (HVJ) *Gene* **38**: 271-274.

Morrison, T.G., Ward, L.J. and Semerjian, A. (1985) Intracellular processing of the Newcastle disease virus fusion glycoprotein. *J. Virol.* **53**: 851-857.

Morrison, T.G., Peeples, M.E. and McGinnes, L.W. (1987) Conformational change in a viral glycoprotein during maturation due to disulfide bond disruption. *Proc. Natl. Acad. Sci. USA* **84**: 1020-1024.

Morrison, T. (1988) Structure, function, and intracellular processing of paramyxovirus membrane proteins. *Virus Res.* **10**: 113-136.
Moss, B. and Flexner, C. (1987) Vaccinia virus expression vectors. *Ann. Rev. Immunol.* **5**: 305-324.

Nagai, Y., Ogura, H. and Klenk, H.-D. (1976a) Studies on the assembly of the envelope of Newcastle disease virus. *Virology* **69**: 523-538.

Nagai, Y., Klenk, H.-D., Rott, R. (1976b) Proteolytic cleavage of the viral glycoproteins and its significance for the virulence of Newcastle disease virus. *Virology* **72**: 494-508.

Olmsted, R.A., Elango, G.A., Prince, G.A., Murphy, B.R., Johnson, P.R., Moss, B., Chanock, R.M. and Collins, P.L. (1986) Expression of the F glycoprotein of respiratory syncytial virus by a recombinant vaccinia virus: comparison of the individual contributions of the F and G glycoproteins to host immunity. *Proc. Natl. Acad. Sci. USA* **83**: 7462-7466.

Parrott, R.H., Vargosko, A.J., Kim, H.W., Bell, J.A. and Chanock, R.M. (1962): III. Myxoviruses: Parainfluenza. *Am. J. Public Health* **52**: 907-917.

Paterson, R.G., Harris, T.J.R. and Lamb, R.A. (1984a) Analysis and gene assignment of mRNAs of a paramyxovirus, simian virus 5. *Virology* **138**: 310-323.

Paterson, R.G., Harris, T.J.R. and Lamb, R.A. (1984b) Fusion protein of the aramyxovirus simian virus 5: nucleotide sequence of mRNA predicts a highly hydrophobic glycoprotein. *Proc. Natl. Acad. Sci. USA* **81**: 6706-6710.

Paterson, R.G., Hiebert, S.W. and Lamb, R.A. (1985) Expression at the cell surface of biologically active fusion and hemagglutinin/neuraminidase proteins of the paramyxovirus simian virus 5 from cloned cDNA. *Proc. Natl. Acad. Sci. USA* 82: 7520-7524.

Paterson, R.G., Lamb, R.A., Moss, B. and Murphy, B.R. (1987) Comparison of the relative roles of the F and HN surface glycoproteins of the paramyxovirus simian virus 5 in inducing protective immunity. *J. Virol.* **61**: 1972-1977.

Paterson, R.G. and Lamb, R.A. (1987) Ability of the hydrophobic fusion-related external domain of a paramyxovirus F protein to act as a membrane anchor. *Cell* **48**: 441-452.

Paterson, R.G., Shaughnessy, M.A. and Lamb, R.A. (1989) Analysis of the relationship between cleavability of a paramyxovirus fusion protein and length of the connecting peptide. *J. Virol.* **63**: 1293-1301.

Pons, M.W. and Rochovansky, O.M. (1979) Ultraviolet inactivation of Influenza virus RNA *in vitro* and *in vivo*. *Virology* **97**: 183-189.

Portner, A and Kingsbury, D.W. (1976) Regulatory events in the synthesis of Sendai virus polypeptides and their assembly into virions. *Virology* **73**: 79-88.

Portner, A., Scroggs, R.A. and Metzger, D.W. (1987a) Distinct functions of antigenic sites of the HN glycoprotein of Sendai virus. *Virology* **158**: 61-68.

Portner, A., Scroggs, R.A. and Naeve, C.W. (1987b) The fusion glycoprotein of Sendai virus: sequence analysis of an epitope involved in fusion and virus neutralization. *Virology* **157**: 556-559.

Prinoski, K., Cote, M.-J., Kang, C.Y. and Dimock, K. (1987) Nucleotide sequence of the human parainfluenza virus 3 matrix protein gene. *Nucl. Acids Res.* **15**: 3182.

Ray, R. and Compans, R.W. (1987) Glycoproteins of human parainfluenza virus type 3: affinity purification, antigenic characterization and reconstitution into lipid vesicles. *J. gen. Virol.* **68**: 409-418.

Re, G.G., Gupta, K.C. and Kingsbury, D.W. (1983) Genomic and copy-back 3' termini in Sendai virus defective interfering RNA species. *J. Virol.* **45**: 659-664.

Richardson, C.D., Scheid, A. and Choppin, P.W. (1980) Specific inhibition of paramyxovirus and myxovirus replication by olipeptides with amino acid sequences similar to those at the N-termini of the F_1 or HA_2 viral polypeptides. *Virology* **105**: 205-222.

Richardson, C.D. and Choppin, P.W. (1983) Oligopeptides that specifically inhibit membrane fusion by paramyxoviruses: studies on the site of action. *Virology* **131**: 518-532.

Richardson, C.D., Berkovich, A., Rozenblatt, S. and Bellini, W.J. (1985) Use of antibodies directed against synthetic peptides for identifying cDNA clones, establishing reading frames, and deducing the gene order of measles virus. *J. Virol.* **54**: 186-193.

Richardson, C.D., Hull, D., Greer, P., Hasel, K., Berkovich, A., Englund, G., Bellini, W., Rima, B. and Lazzarini, R. (1986) The nucleotide sequence of the mRNA encoding the fusion protein of measles virus (Edmonston strain): a comparison of fusion proteins from several different paramyxoviruses. *Virology* **155**: 508-523.

Rose, J.K. (1980) Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. *Cell* **19**: 415-421.

Rothman, J.E. (1981) The Golgi apparatus: two organelles in tandem. *Science* **213**: 1212-1219.

Russell, S.E.H., Clarke, D.K., Hoey, E.M., Rima, B.K. and Martin, S.J. (1985) cDNA cloning of the messenger RNAs of five genes of canine distemper virus. *J. gen. Virol.* **66**: 433-441.

Rutledge, R.G., Seligy, V.L., Cote, M.-J., Dimock, K., Lewin, L.L. and Tenniswood, M.P. (1988) Rapid synthesis and cloning of complementary DNA from any RNA molecule into plasmid and phage M13 vectors. *Gene* **68**: 151-158.

Rydbeck, R., Orvell, C., Love, A. and Norrby, E. (1986) Characterization of four parainfluenza virus type 3 proteins by use of monoclonal antibodies. *J. gen. Virol.* **67**: 1531-1542.

Sakai, Y., Suzu, S., Shioda, T. and Shibuta, H. (1987) Nucleotide sequence of the bovine parainfluenza 3 virus genome: Its 3' end and the genes of NP, P, C and M proteins. *Nucl. Acids Res.* **15**: 2927-2944.

Sanchez, A. and Banerjee, A.K. (1985) Studies on human parainfluenza virus 3: characterization of the structural proteins and *in vitro* synthesized proteins coded by mRNAs isolated from infected cells. *Virology* **143**: 45-54.

Sanchez, A., Banerjee, A.K., Furuichi, Y. and Richardson, M.A. (1986) Conserved structure among the nucleocapsid proteins of the *Paramyxoviridae*: complete nucleotide sequence of human parainfluenza virus type 3 NP mRNA. *Virology* **152**: 171-180.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.

Sato, H., Oh-hira, M., Ishida, N., Imamura, Y., Hattori, S. and Kawakita, M. (1987a) Molecular cloning and nucleotide sequence of P, M and F genes of Newcastle disease virus avirulent strain D26. *Virus Res.* **7**: 241-255.

Sato, H., Hattori, S., Ishida, N., Imamura, Y. and Kawakita, M. (1987b) Nucleotide sequence of the hemagglutinin-neuraminidase gene of Newcastle disease virus avirulent strain D26: evidence of a longer coding region with a carboxyl terminal extension as compared to virulent strains. *Virus Res.* **8**: 217-232.

Sato, T.A., Kohama, T. and Sugiura, A. (1988) Intracellular processing of measles virus fusion protein. *Arch. Virol.* **98**: 39-50.

Scheid, A. and Choppin, P.W. (1974) Identification of biological activities of paramyxovirus glycoproteins. Activation of cell fusion, hemolysis, and infectivity by proteolytic cleavage of an inactive precursor protein of Sendai virus. *Virology* **57**: 475-490.

Scheid, A. and Choppin, P.W. (1977) Two disulfide-linked polypeptide chains constitute the active F protein of paramyxoviruses. *Virology* **80**: 54-66.

Scheid, A., Graves, M.C., Silver, S.M. and Choppin, P.W. (1978) Studies on the structure and functions of paramyxovirus glycoproteins. In B.W. Mahy and R.D. Barry (eds.) ***Negative strand viruses and the host cell.*** Academic Press. London. p 181-193.

Schmidt, M.F.G. (1982) Acylation of viral spike glycoproteins: a feature of enveloped RNA viruses. *Virology* **116**: 327-338.

Sekikawa, K. and Lai, C.-J. (1983) Defects in functional expression of an Influenza virus hemagglutinin lacking the signal peptide sequences. *Proc. Natl. Acad. Sci. USA.* **80**: 3563-3567.

Seto, J.T., Garten, W. and Rott, R. (1981) The site of cleavage in infected cells and polypeptides of representative paramyxoviruses grown in cultured cells of the chorioallantoic membrane. *Arch. Virol.* **67**: 19-30.

Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. and Iwasaki, K. (1983) Sequence of 3,687 nucleotides from the 3' end of Sendai virus genome RNA and the predicted amino acid sequences of viral NP, P and C proteins. *Nucl. Acids Res.* **11**: 7317-7330.

Shioda, T., Iwasaki, K. and Shibuta, H. (1986) Determination of the complete nucleotide sequence of the Sendai virus genome RNA and the predicted amino acid sequences of the F, HN and L proteins. *Nucl. Acids Res.* **14**: 1545-1563.

Spriggs, M.K. and Collins, P.L. (1986a) Human parainfluenza virus type 3: messenger RNAs, polypeptide coding assignments, intergenic sequences, and genetic map. *J. Virol.* **59**: 646-654.

Spriggs, M.K. and Collins, P.L. (1986b) Sequence analysis of the P and C protein genes of human parainfluenza virus type 3: patterns of amino acid sequence homology among paramyxovirus proteins. *J. gen. Virol.* **67**: 2705-2719.

Spriggs, M.K., Olmsted, R.A., Venkatesan, S., Coligan, J.E. and Collins, P.L. (1986) Fusion glycoprotein of human parainfluenza virus type 3: nucleotide sequence of the gene, direct identification of the cleavage-activation site, and comparison with other paramyxoviruses. *Virology* **152**: 241-251.

Spriggs, M.K., Johnson, P.R. and Collins, P.L. (1987a) Sequence analysis of the matrix protein gene of human parainfluenza virus type 3: extensive sequence homology among paramyxoviruses. *J. gen. Virol.* **68**: 1491-1497.

Spriggs M.K., Murphy, B.R., Prince, G.A., Olmsted, R.A. and Collins P.L. (1987b) Expression of the F and HN glycoproteins of human parainfluenza virus type 3 by recombinant vaccinia viruses: contributions of the individual proteins to host immunity. *J. Virol.* **61**: 3416-3423.

Storey, D.G., Dimock, K. and Kang, C.Y. (1984) Structural characterization of virion proteins and genomic RNA of human parainfluenza virus. *J. Virol.* **52**: 761-766.

Storey, D.G. (1987) Structural characterization of human parainfluenza virus 3 and cloning of viral specific genes. Ph. D. thesis, university of Ottawa, Ottawa, Ontario.

Storey, D.G., Cote, M.-J., Dimock, K., Kang, C.Y. (1987) Nucleotide sequence of the coding and flanking regions of the human parainfluenza virus 3 hemagglutinin-neuraminidase gene: comparison with other paramyxoviruses. *Intervirology* **27**: 69-80.

Sung, W.L., Zahab, D.M., MacDonald, C.A. and Tam, C.S. (1986) Synthesis of mutant parathyroid hormone genes via site-specific recombination directed by crossover linkers. *Gene* **47**: 261-267.

Suzu, S., Sakai, Y., Shioda, T. and Shibuta, H. (1987) Nucleotide sequence of the bovine parainfluenza 3 virus genome: the genes of the F and HN glycoproteins. *Nucl. Acids Res.* **7**: 2945-2958.

Thomas, S.M., Lamb, R.A. and Paterson, R.G. (1988) Two mRNAs that differ by to nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* **54**: 891-902.

Toyoda, T., Sakaguchi, T., Imai, K., Inocencio, N.M., Gotoh, B., Hamaguchi, M. and Nagai, Y. (1987) Structural comparison of the cleavage-activation site of the fusion glycoprotein between virulent and avirulent strains of Newcastle disease virus. *Virology* **158**: 242-247.

Toyoda, T., Gotoh, B., Sakaguchi, T., Kida, H. and Nagai, Y. (1988) Identification of amino acids relevant to three antigenic determinants on the fusion protein of Newcastle disease virus that are involved in fusion inhibition and neutralization. *J. Virol.* **62**: 4427-4430.

Varsanyi, T.M., Jornvall, H. and Norrby, E. (1985) Isolation and characterization of the measles virus F_1 polypeptide: comparison with other paramyxovirus fusion proteins. *Virology* **147**: 110-117.

Verner, K. and Schatz, G. (1988) Protein translocation across membranes. *Science* **241**: 1307-1313.

Waxham, M.N. and Wolinsky, J.S. (1986) A fusing mumps virus variant selected from a nonfusing parent with the neuraminidase inhibitor 2-deoxy-2,3-dehydro-*N*-acetylneuraminic acid. *Virology* **151**: 286-295.

Waxham, M.N., Server, A.C., Goodman, H.M. and Wolinsky, J.S. (1987) Cloning and sequencing of the mumps virus fusion protein gene. *Virology* **159**: 381-388.

Wechsler, S.L., Lambert, D.M., Galinski, M.S., Heineke, B.E. and Pons, M.W. (1985) Human parainfluenza virus 3: Purification and characterization of subviral components, viral proteins and viral RNA. *Virus Res.* **3**: 339-351.

Wertz, G.W., Stott, E.J., Young, K.K.Y., Anderson, K. and Ball, L.A. (1987) Expression of the fusion protein of human respiratory syncytial virus from recombinant vaccinia virus vectors and protection of vaccinated mice. *J. Virol.* **61**: 293-301.

Wilde, A., McQuain, C. and Morrison, T.G. (1986) Identification of the sequence content of four polycistronic transcripts synthesized in Newcastle disease virus infected cells. *Virus Res.* **5**: 77-95.

Yusoff, K., Millar, N.S., Chambers, P. and Emmerson, P.T. (1987) Nucleotide sequence analysis of the L gene of Newcastle disease virus: homologies with Sendai and vesicular stomatitis viruses. *Nucl. Acids Res.* **15**: 3961-3976.

APPENDIX OLIGONUCLEOTIDE SEQUENCES

Crossover linker

for the 5'-end	5'-AGC TTC CAC CAT GCC AAC CTC A-3'
for the 3'-end	5'-CCG GAT CCT CAT TTG TTT GT-3'

Primer used for mutagenesis

CYS 18	5'-ATC TTT CTC CCA AAT AG-3'
CYS 63	5'-TAA CTC TTC TGG TGA CC-3'
CYS 192	5'-ATT AGG TTC TGA AGC AG-3'
CYS 331	5'-GAT ATC AAA GAA TCT ATA GAA GCA TTC-3'
CYS 340	5'-TTA TAT ATC CCC TTC TG-3'
CYS 355	5'-GGA GAG CTC TTT ATC AG-3'
CYS 363	5'-ATC CCA ATC TCC AAG AA-3'
CYS 387	5'-TGC AAA TTC TAT AAC AA-3'
CYS 392	5'-AAC CAC ATC TAC ATG CA-3'
CYS 394	5'-ATG TAC ATC CAA CGG TA-3'
CYS 417	5'-CAC ATA AAG AAT CTA ATA CAA TAG G-3'

HPIV3 F gene sequencing primers

5'-end	5'-AGC TTC CAC CAT GCC AAC CTC A-3'
#2	5'-TTC TAA CTC TTG TGG TGA CC-3'
#3	5'-CTG GTT GAA GCC AAG CAG GC-3'
#4	5'-GAA TTA ACA AAC ATA TTC GG-3'
KD4	5'-GAA TCA ATA AAG GTG AGA GT-3'
#5	5'-CCC TTT ATT AAC TAG ACT GC-3'
KP4	5'-GGA TTT GTA CTA AAC CAT GA-3'
#7	5'-ACT CTT GCA TTT TAC ACA CC-3'
#8	5'-TGG CAT CAA TCT AGC ACC AC-3'

HPIV3 oligonucleotides used for primer extension

NP	5'-TTT GAA ATT ATA GAG TTC CC-3'
P	5'-GCA TCG CTT TCC ATC AAC TC-3'
M	5'-GCA GAG TTA GTT ATA CTC AT-3'
F	5'-CTT GTT CCA GCG AGT GTG GC-3'
HN	5'-CGG ATT TGG ATT TGT CTG TT-3'
L	5'-GGT TCT TTC TCA ATT ATA TG-3'
L_1	5'-AGT ATG TCA GAT ACA GTG CC-3'
Leader	5'-GTT AAT TTT AAT TTA AAT TT-3'