Development and Use of Avian Pneumovirus Reverse Genetics Systems

Nicole Lynn Edworthy

A thesis submitted for the qualification of Ph.D. at the Department of Biological Sciences, University of Warwick, UK (November, 2008)
Table of Contents

TABLE OF CONTENTS 1
LIST OF FIGURES VI
LIST OF TABLES X
LIST OF ABBREVIATIONS XII
ACKNOWLEDGEMENTS XIV
DECLARATIONS XV
ABSTRACT XVI

CHAPTER 1: INTRODUCTION 1
1.1 Introduction 2
1.1.1 Pneumovirus taxonomy and virion structure 2
1.1.2 APV incidence and distribution 4
1.2 Virus pathogenesis 6
1.2.1 Disease: turkey rhinotracheitis, swollen head syndrome, and co-infection 6
1.2.2 Tissue tropism 7
1.2.3 Vaccines 8
1.3. *Pneumovirinae* genome organization and protein products 10
1.3.1 *Pneumovirinae* genomes 10
1.3.1 *Pneumovirinae* genes and proteins 12
1.3.2.1 N protein 12
1.3.2.2 P protein 15
1.3.2.3 L protein 17
1.3.2.4 M2-1 and M2-2 proteins 17
1.3.2.5 F protein 22
1.3.2.6 G protein 24
1.3.2.7 SH protein 26
1.3.2.8 M protein 28
1.3.2.9 NS1 and NS2 proteins 28
1.4 Virus transcription and replication 30
1.4.1 Virus transcription 30
1.4.2 Transcriptional regulatory sequences 32
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4.3 Virus replication</td>
<td>37</td>
</tr>
<tr>
<td>1.4.4 Replication regulatory sequences</td>
<td>37</td>
</tr>
<tr>
<td>1.5 Reverse genetics systems</td>
<td>40</td>
</tr>
<tr>
<td>1.6 Aims</td>
<td>46</td>
</tr>
<tr>
<td>CHAPTER 2: MATERIALS AND METHODS</td>
<td>47</td>
</tr>
<tr>
<td>2.1 Solutions and buffers</td>
<td>48</td>
</tr>
<tr>
<td>2.2 Media</td>
<td>49</td>
</tr>
<tr>
<td>2.3 Mammalian tissue culture- maintenance of cell lines</td>
<td>50</td>
</tr>
<tr>
<td>2.3.1 Vero, BSC-1, and HEp-2 cells</td>
<td>50</td>
</tr>
<tr>
<td>2.3.2 BSR-T7 cells</td>
<td>50</td>
</tr>
<tr>
<td>2.4 Bacterial cultures</td>
<td>50</td>
</tr>
<tr>
<td>2.4.1 Growth of E. coli</td>
<td>50</td>
</tr>
<tr>
<td>2.4.2 Frozen competent bacteria</td>
<td>51</td>
</tr>
<tr>
<td>2.5 Standard cloning procedures</td>
<td>51</td>
</tr>
<tr>
<td>2.5.1 Reverse transcription-PCR</td>
<td>51</td>
</tr>
<tr>
<td>2.5.2 PCR</td>
<td>51</td>
</tr>
<tr>
<td>2.5.3 DNA purification</td>
<td>53</td>
</tr>
<tr>
<td>2.5.3.1 Gel extraction</td>
<td>53</td>
</tr>
<tr>
<td>2.5.3.2 Phenol/chloroform extraction</td>
<td>53</td>
</tr>
<tr>
<td>2.5.3.3 Ethanol precipitation</td>
<td>53</td>
</tr>
<tr>
<td>2.5.3.4 PCR product fragment purification</td>
<td>53</td>
</tr>
<tr>
<td>2.5.4 Agarose gel electrophoresis</td>
<td>54</td>
</tr>
<tr>
<td>2.5.5 Urea acrylamide gel electrophoresis</td>
<td>54</td>
</tr>
<tr>
<td>2.5.6 Enzyme digestions</td>
<td>54</td>
</tr>
<tr>
<td>2.5.7 Ligations</td>
<td>55</td>
</tr>
<tr>
<td>2.5.8 Transformation of bacteria</td>
<td>55</td>
</tr>
<tr>
<td>2.5.9 Small scale DNA preparations</td>
<td>55</td>
</tr>
<tr>
<td>2.5.9.1 Mini-preparations of DNA</td>
<td>55</td>
</tr>
<tr>
<td>2.5.9.2 Alkaline lysis</td>
<td>55</td>
</tr>
<tr>
<td>2.5.10 Large scale DNA preparations</td>
<td>55</td>
</tr>
<tr>
<td>2.6 Introduction of site-specific mutations into plasmid DNA</td>
<td>56</td>
</tr>
<tr>
<td>2.6.1 Modified QuickChange PCR method</td>
<td>56</td>
</tr>
<tr>
<td>2.6.2 Quick screen method</td>
<td>56</td>
</tr>
</tbody>
</table>
2.6.3 Pre-sequencing purification of DNA from quick screened colonies 57
2.7 Sequencing reactions with Big-Dye 58
2.8 Rescue systems 58
2.8.1 Helper virus rescue system 58
2.8.2 Plasmid helper rescue system 59
2.8.3 Full length genome rescue system 59
2.9 Preparation of cell lysates for CAT and luciferase assay 60
2.10 CAT assay 60
2.11 Luciferase assay 61
2.12 RNA Quantification 61
2.12.1 Northern blot 61
2.13 Growth and harvesting of virus stocks 62
2.13.1 Growth of non-recombinant virus stocks 62
2.13.2 Growth of recombinant virus stocks 63
2.14 Determination of virus titre by GFP expression assay 63
2.15 Nuclotide and protein sequence alignments 64
2.16 mRNA and protein simultaneous harvest 64
2.17 RNA probes 65
2.18 RNA standards 65
2.19 Imaging methods 66

CHAPTER 3: DEVELOPMENT AND OPTIMIZATION OF THE APV PLASMID RESCUE SYSTEM 67
3.1 Introduction 68
3.2 Development of the plasmid rescue system for APV 70
3.2.1 Cloning of the APV N gene helper plasmid 71
3.2.2 Cloning of the APV P gene helper plasmid 72
3.2.3 Cloning of the APV M2-2 ORF helper plasmid 73
3.3 Establishment and optimisation of the APV plasmid rescue system 73
3.3.1 Optimization of the amount of N10, P9, and pCITE L plasmid 76
3.3.2 Effect of varying the amount of pM2-1 and pM2-2 plasmid on reporter gene expression 78
3.4 Discussion 81
3.4.1 The APV plasmid rescue system 81
3.4.2 M2-1 and M2-2 in the APV plasmid rescue system

CHAPTER 4: EFFECT OF MUTATION OF THE GS AND GE SIGNALS ON MRNA AND PROTEIN EXPRESSION IN THE APV PLASMID RESCUE SYSTEM

4.1 Introduction
4.2 Comparison of mRNA and protein expression levels from GS sequence mutants
4.3 Activity of transcriptional signal mutations in the APV plasmid rescue system
4.3.1 Gene junction mutations
4.3.1.1 Generation of IR mutant 1
4.3.1.2 Generation of IR mutant 2
4.3.1.3 Generation of IR mutant 3
4.3.1.4 Generation of IR mutant 4
4.3.2 Determination of effect of mutation of the pCATLUC minigenome gene junction
4.3.3 Generation of dual GS sequence point mutations
4.3.4 Effect of dual GS sequence point mutation
4.4 Conclusions

CHAPTER 5: HETEROLOGOUS AND CHIMERIC PROTEIN FUNCTION IN THE APV PLASMID RESCUE SYSTEM

5.1 Introduction
5.2 Effect of APV/PVM and APV/RSV heterologous helper plasmid sets on protein expression in the APV plasmid rescue system
5.3 Effect of APV and hMPV heterologous helper plasmid sets on protein expression in the hMPV plasmid rescue system and APV plasmid rescue system
5.4 Functionality of M2-1 APV/RSV chimeras in the APV plasmid rescue system
5.5 Discussion
CHAPTER 6: RESCUE OF APV FROM CDNA 140
6.1 Introduction 141
6.2 Production of GFP-expressing APV 143
6.2.1 Cloning of the eGFP full-length genome 144
6.2.3 Growth and eGFP expression in cell culture of virus CF2ESH 11 152
6.2.4 CF2ESH 11 virus expression of SH, G and eGFP mRNAs 157
6.2.5 CF2ESH 11 virus replication in turkeys 157
6.3 Production of GFP-expressing APV SH gene deleted viruses 162
6.3.1 Further characterization of virus CF2EdSH(dG) 164
6.4 Discussion 173
6.4.1. Detection of eGFP in vitro 173
6.4.2. Detection of eGFP in vivo 173
6.4.3. Viruses with SH deleted genomes 174

CHAPTER 7: CONCLUSIONS 176
7.1 Conclusions 177
7.2 Future work 179

BIBLIOGRAPHY 182

APPENDIX A: PRIMER SEQUENCES 201

APPENDIX B: CONSTUCT SEQUENCES 203
pCAT LUC dicistronic minigene construct 204
pCF2 cassetted APV genome construct 208

APPENDIX C: CONFOCAL IMAGES
Time-lapse movies are included as DVD presentation and as flick books.

APPENDIX D: PUBLISHED DATA 221
**List of Figures**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Genome Organization of the <em>Pneumovirinae</em>.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Model of the <em>Pneumovirinae</em> virion.</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Paramyxovirus ribonucleoprotein capsid structure.</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>Model of Paramyxovirus transcription.</td>
<td>31</td>
</tr>
<tr>
<td>1.5</td>
<td>Model of Paramyxovirus replication.</td>
<td>38</td>
</tr>
<tr>
<td>1.6</td>
<td>Model of the negative sense RNA virus full length rescue system.</td>
<td>41</td>
</tr>
<tr>
<td>1.7</td>
<td>Schematic of the APV minigenome and dicistronic minigenome constructs.</td>
<td>43</td>
</tr>
<tr>
<td>1.8</td>
<td>Model of the negative sense RNA virus plasmid rescue system.</td>
<td>44</td>
</tr>
<tr>
<td>1.9</td>
<td>Model of the negative sense RNA helper virus rescue system.</td>
<td>45</td>
</tr>
<tr>
<td>3.1</td>
<td>Sequence of the APV subgroup A (isolate CVL-14/1) M2-1 and M2-2 ORFs.</td>
<td>69</td>
</tr>
<tr>
<td>3.2</td>
<td>Cloning of the N10 plasmid.</td>
<td>71</td>
</tr>
<tr>
<td>3.3</td>
<td>Cloning of P9 plasmid.</td>
<td>72</td>
</tr>
<tr>
<td>3.4</td>
<td>Cloning of pM2-2: PCR for M2-2 fragment and restriction digest of M2-2 fragment and vector.</td>
<td>74</td>
</tr>
<tr>
<td>3.5</td>
<td>Screen of M2-2 plasmid clones for correct inserts.</td>
<td>75</td>
</tr>
<tr>
<td>3.6</td>
<td>Effect of APV N plasmid concentration on CAT protein expression in the APV A PRS.</td>
<td>77</td>
</tr>
<tr>
<td>3.7</td>
<td>Effect of P plasmid variation on CAT protein expression in the APV PRS.</td>
<td>77</td>
</tr>
<tr>
<td>3.8</td>
<td>Effect of APV L plasmid concentration on CAT protein expression in the APV A PRS.</td>
<td>78</td>
</tr>
<tr>
<td>3.9</td>
<td>Effect of APV M2-1 plasmid concentration on CAT and LUC protein expression in the APV A PRS.</td>
<td>80</td>
</tr>
</tbody>
</table>
Figure 3.10 Effect of APV M2-2 plasmid concentration with and without M2-1 plasmid present on CAT protein expression in the APV A PRS.

Figure 4.1 Effect of mutation of the dicistronic minigenome plasmid LUC GS sequence on percentage activity.

Figure 4.2 CAT protein and LUC protein levels of selected GS sequence mutants.

Figure 4.3 CAT protein and LUC protein levels of selected GS sequence mutants.

Figure 4.4 Dicistronic minigenome (mRNA sense) gene junction mutants.

Figure 4.5 Generation of fragments A and B.

Figure 4.6 Generation of fragment C.

Figure 4.7 Generation of IR mutant 1.

Figure 4.8 Agarose gels examining of restriction digestion products for IR mutant 1 cloning.

Figure 4.9 Generation of IR mutant 2.

Figure 4.10 Generation of fragment D.

Figure 4.11 Generation of IR mutant 3.

Figure 4.12 Generation of fragments E and F.

Figure 4.13 Generation of fragment G.

Figure 4.14 Generation of IR mutant 4*.

Figure 4.15 Generation of IR mutant 4.

Figure 4.16 Determination of effect of mutation of the pCATLUC minigenome gene junction on CAT and LUC protein expression.

Figure 4.17 P values determined using the Paired T Test for significance between the IR mutants 1 to 4 for CAT and LUC protein values.

Figure 4.18 Agarose gel examining PCR products for cloning of L GS sequence dual mutants.
Figure 4.19 Effect of duel GS point mutation on reporter protein expression.

Figure 4.20 Percentage activity of gene junction mutation \textit{in vitro} and \textit{in vivo}.

Figure 5.1 Effect of heterologous substitution of RSV plasmids.

Figure 5.2 Effect of heterologous substitution of PVM plasmids.

Figure 5.3 Effect of heterologous substitution of APV and hMPV N, P or L plasmids.

Figure 5.4 Effect of heterologous substitution of APV and hMPV minigenome or M2-1 plasmids.

Figure 5.5 RSV M2-1 protein deletion mutants.

Figure 5.6 Composition of APV/RSV chimeric M2-1 proteins.

Figure 5.7 The effect of APV/RSV M2-1 chimeric protein on CAT reporter protein expression in the APV PRS.

Figure 5.8 The effect of APV/RSV M2-1 chimeric protein on CAT and LUC reporter protein expression in the RSV PRS.

Figure 6.1 Full length APV cloned genomes.

Figure 6.2 Generation of pCF2 mutants- overview of cloning strategy.

Figure 6.3 Generation of eGFP fragment for insertion into the CF2 genome.

Figure 6.4 Generation of genome vector for insertion of the eGFP gene into the CF2 genome.

Figure 6.5 Screening of clones by restriction digestion for substitution of the eGFP gene for the SH gene to create the CF2EdSH genome.

Figure 6.6 Generation of SH fragment for insertion into the CF2ESH genome.

Figure 6.7 Generation of genome vector for insertion of the SH gene into the CF2EdSH genome.

Figure 6.8 Screening of clones by restriction digestion for insertion of the SH gene to create the CF2ESH genome.

Figure 6.9 eGFP expression of virus CF2ESH 11 in BSR-T7 cells.
Figure 6.10 part I: Time lapse photography of infection of cells with virus CF2ESH 11, 30 to 80 min.

Figure 6.10 part II: Time lapse photography of infection of cells with virus CF2ESH 11, 90 to 140 min.

Figure 6.10 part III: Time lapse photography of infection of cells with virus CF2ESH 11, 150 to 200 min.

Figure 6.11 Northern blot of CF2ESH 11 virus RNA with G, SH or eGFP specific probes.

Figure 6.12 Confocal microscopy of CF2ESH infected turkey sinus 3 days p.i.

Figure 6.13 Confocal microscopy of CF2ESH infected turkey sinus 5 days p.i.

Figure 6.14 Confocal microscopy of mock infected turkey sinus 3 or 5 days p.i.

Figure 6.15 real-time PCR measurements of number of copies of the N gene (shown as log 10 copies per 10 mg of tissue).

Figure 6.16 eGFP expression of viruses CF2EdSH 34 and CF2EdSH(dG) 56 in BSC-1 cells.

Figure 6.17 part I: Time lapse photography of infection of BSC-1 cells with virus CF2EdSH 56, at 0, 10 and 20 min.

Figure 6.17 part II: Time lapse photography of infection of BSC-1 cells with virus CF2EdSH 56, at 30, 40 and 50 min.

Figure 6.17 part III: Time lapse photography of infection of BSC-1 cells with virus CF2EdSH 56, at 60, 70 and 80 min.

Figure 6.18 Northern blot of virus RNA with G, SH or eGFP specific probes.

Figure 6.19 real-time PCR measurements of number of copies of the N gene (shown as log 10 copies per 10 mg of tissue).
List of Tables

Table 1.1 Nucleotide homologies of the pneumovirus genomes. 11

Table 1.2 Amino acid and nucleotide lengths of the pneumovirus ORF and proteins. 13

Table 1.3 Percentage nucleotide homology and amino acid identity of the pneumovirus N ORF and proteins. 14

Table 1.4 Percentage nucleotide homology and amino acid identity of the pneumovirus P ORF and proteins. 16

Table 1.5 Percentage nucleotide homology and amino acid identity of the pneumovirus L genes and proteins. 18

Table 1.6 Nucleotide homologies of the pneumovirus M2 mRNA. 18

Table 1.7 Percentage nucleotide homology and amino acid identity of the pneumovirus M2-1 genes and proteins. 20

Table 1.8 Percentage nucleotide homology and amino acid identity of the pneumovirus M2-2 genes and proteins. 21

Table 1.9 Percentage nucleotide homology and amino acid identity of the pneumovirus F genes and proteins. 23

Table 1.10 Percentage nucleotide homology and amino acid identity of the pneumovirus G genes and proteins. 25

Table 1.11 Percentage nucleotide homology and amino acid identity of the pneumovirus SH genes and proteins. 27

Table 1.12 Percentage nucleotide homology and amino acid identity of the pneumovirus M genes and proteins. 29

Table 1.13 Percentage nucleotide homology and amino acid identity of the pneumovirus NS1 and NS2 genes and proteins. 29

Table 1.14 Pneumovirus gene start (GS) sequences. 33

Table 1.15 Pneumovirus gene end (GE) sequences. 35

Table 1.16 Pneumovirus intergenic region lengths in nucleotides. 36

Table 1.17 Nucleotide homologies of the pneumovirus Le and Tr sequences. 39

Table 4.1 Sequences of the mutant GS signals. 87
Table 4.2 Primers and template pairs used for generation of duel GS sequence mutants.

Table 5.1 Amino acid identities of the APV N, P, L and M2-1 protein sequences with the homologous proteins of hMPV, PVM and RSV.

Table 5.2 Pneumovirus consensus gene start and gene end sequences.

Table 6.1 S/P ratios of turkeys prior to and three weeks post inoculation with mutant virus CF2ESH 11.

Table 6.2 S/P ratios of turkeys prior to and three weeks post inoculation with mutant virus CF2EdSH 56.
Abbreviations

A  adenine
APV  avian pneumovirus
bp  base pairs
C  cytosine
CAT  chloramphenicol acetyl transferase
cDNA  complementary DNA
CIU  cell infecting units
CMV  cytomegalovirus
CPE  cytopathic effect
C-terminus  carboxyl terminus
d.p.i.  days post infection
dATP  2’ deoxyadenosine 5’ triphosphate
dCTP  2’ deoxycytidine 5’ triphosphate
dGTP  2’ deoxyguanosine 5’ triphosphate
DIG  digoxigenin
DNA  deoxyribonucleic acid
dNTPs  equimolar mix of dATP, dCTP, dGTP and dTTP
dTTP  2’deoxythymidine 5’ triphosphate
EtOH  ethanol
FCS  fetal calf serum
G  guanine
hMPV  human metapneumovirus
hr  hour/hours
IRES  internal ribosome entry site
kb  kilobase
LUC  luciferase
min  minute/minutes
MOI  multiplicity of infection
mRNA  messenger RNA
N-terminus  amino terminus
o/n  over night
ORF  open reading frame
p.i.  post infection
PRS  plasmid rescue system
PVM  pneumonia virus of mice
RLU  relative light units
RNA  ribonucleic acid
RSV  respiratory syncytial virus
rt  room temperature
SDS  sodium dodecyl sulphate
SDW  sterile distilled water
sec  second/seconds
SeV  Sendai virus
ssDNA  single-strand DNA
T  thymine
U  uracil
VSV  vesicular stomatitis virus
Acknowledgements

I would like to thank my supervisor Prof. A.J. Easton for his help and encouragement. I would also like to thank Drs. A.C. Marriott, R. Ling, A. Nenninger and the members of the pneumovirus laboratory for help and advice. Many thanks to my husband, my mother, my grandmother, and all my family and friends for immeasurable support. This PhD would not have got done without you.
Declarations

I hereby declare that all the results presented in this thesis were obtained by myself under the supervision of Prof. A.J. Easton, unless otherwise stated. This thesis has not been submitted for a degree in any other institution.
Abstract

Avian pneumovirus (APV) has remained an important pathogen of domestic fowl since its isolation in the 1970s. A reverse genetics system for APV was developed that affords direct manipulation and analysis of the molecular biology, pathogenicity, and tropism of APV. Using a synthetic minigenome system, the M2-1 protein was found to enhance transcription but not be essential for replication and the APV M2-2 protein was shown to inhibit transcription of a reporter gene. The viral cis-acting sequences were mutated to determine their role in transcription. Initially, a series of mutations originating from vaccine candidates were introduced into the gene end sequence of the LUC gene. The levels of LUC reporter protein expression in the mutants was 40-70% of normal, thus demonstrating a mechanism for reduction of virus immunogenicity as the result of a single point mutation. Heterologous rescue of the APV minigenome was carried out using plasmids expressing the RSV, PVM and hMPV proteins and showed that homologous protein: protein interactions were necessary for minigenome transcription. An APV cloned virus rescue system (Naylor et al., 2004) was used to create APV viruses which contained the gene encoding enhanced green florescent protein (eGFP) either within intact APV, or in mutants lacking the SH and G genes or lacking the SH gene alone. It was demonstrated that the SH and G genes are not essential for APV replication in vitro and in vivo and that the APV genome is capable of accepting insertions of foreign material. Expression of eGFP from the recombinant viruses was investigated in vivo in turkeys at 3 and 5 days post infection. eGFP was found in the sinus tissue of the birds infected with the virus containing the full complement of virus genes in addition to that encoding eGFP.
Chapter 1:
Introduction to Avian Pneumovirus and Reverse Genetics Systems
1.1 Introduction

1.1.1 Pneumovirus taxonomy and virion structure

Avian pneumovirus (APV), also known as avian metapneumovirus (Lamb et al., 2000) and previously known as turkey rhinotracheitis virus, is a member of the virus family *Paramyxoviridae*, the subfamily *Pneumovirinae* and the genus *Metapneumovirus* (Bermingham *et al.*, 2007; Cavanagh and Barrett, 1988; Collins and Gough, 1988; Njenga *et al.*, 2003; Pringle, 1995; Pringle, 1996; Pringle, 1998; Yu *et al.*, 1991). Within the *Pneumovirinae* subfamily, the viruses of the *Pneumovirus* genus are the best studied, especially the human pathogen respiratory syncytial virus (RSV). The viruses of the *Pneumovirinae* have an enveloped, non-segmented, negative-sense RNA genomes that are 13 to 15kb in length. The APV genome encodes eight mRNAs and directs the expression of nine proteins (figure 1.1 and section 1.3) (Collins and Gough, 1988; Ling and Pringle, 1988). APV is classified into the subfamily *Pneumovirinae* and genus *Metapneumovirus* based on electron microscopy studies, mRNA profiles, genome sequences, and the molecular weights of its proteins (Bachi and Howe, 1973; Bhella *et al.*, 2002; Cavanagh and Barrett, 1988; Compans *et al.*, 1967; Joncas *et al.*, 1969).

Sub-Family: *Pneumovirinae*

Genus *Pneumovirus*:

Human, bovine and ovine respiratory syncytial viruses (RSV, BRSV and ORSV), Pneumonia virus of mice (PVM). PVM lacks the overlap between the M2-1 and L genes.

Genus *Metapneumovirus*:

Avian pneumovirus (APV), Human metapneumovirus (hMPV)

![Figure 1.1 Genome organization of the Pneumovirinae.](image)
The virions of the pneumoviruses and metapneumoviruses are spherical but pleomorphic (the appearance of two or more distinctly different forms is common) and approximately 200nm diameter (Collins and Gough, 1988; Gough et al., 1988b; Wyeth et al., 1986). The virion consists of the ribonucleoprotein (RNP) complex surrounded by the matrix (M) protein and a cell-derived lipid envelope with the attachment glyco- (G) and fusion (F) proteins protruding 13-14nm from the surface (Collins and Gough, 1988) (figure 1.2). A small hydrophobic (SH) glycoprotein has also been isolated, although the exact location of the SH protein in the virion is not known. The RNP complex consists of the negative sense RNA genome tightly bound with nucleocapsid (N) protein and associated with phospho- (P) and large polymerase (L) proteins (Huang et al., 1985; Peeples and Levine, 1979). This complex has been visualized using negative staining and electron microscopy for the Paramyxoviridae and shows a distinctive herring-bone structure (figure 1.3). The APV nucleocapsid (the N protein associated with the RNA genome) is smaller (12-15nm in diameter) than the 18nm helical nucleocapsid of the Paramyxovirinae (Alexander, 1991; Collins and Gough, 1988).

**Figure 1.2 Model of the Pneumovirinae virion.** The pneumovirus virus virion is made up of the RNP complex surrounded by the matrix (M) protein and a host-derived lipid membrane with the viral glycoproteins (the fusion and attachment proteins) embedded (reproduced from (Dimmock et al., 2006)).
1.1.2 APV incidence and distribution

APV was first observed as an agent causing respiratory disease in turkeys and chickens in South Africa in the 1970s (Buys and Du Preez, 1980; Buys et al., 1989a). The virus was isolated as the causative agent of respiratory disease in turkeys and as an agent contributing to swollen head syndrome in chickens (section 1.2) during outbreaks that occurred in France and England (Giraud et al., 1986; McDougall and Cook, 1986; Wilding et al., 1986). Viral proteins have been detected in turkeys using immunofluorescence (IF) (Jones et al., 1986; Jones et al., 1988) and immunoperoxidase (IP) assays (O’Loan and Allan, 1990). APV isolated from turkeys has also been shown to be able to infect chickens, verified by IP, virus isolation, and histology (Jones et al., 1987; Majo et al., 1995).

APV appears to be present in flocks of domestic birds in most areas of the world, with serological evidence of APV infection reported in most of Europe, South Africa, Asia (Lu et al., 1994), Israel (Weisman et al., 1988), Japan (Otsuki et al., 1996), south and central America (Dani et al., 1999), China, Nigeria (Owoade et al., 2006), and the USA.
(Bennett et al., 2004; Cook et al., 1999; Goyal et al., 2000). Virus has been isolated from birds in France, Great Britain, Italy (Alexander, 1991), South Africa (Buys and Du Preez, 1980) and Israel (Weisman et al., 1988) (reviewed by Jones (1996)).

APV primarily infects domestic flocks of turkeys and chicken, but may also persist in populations of wild birds. APV causes major economic losses in infected flocks and turkeys as young as one day old are susceptible to infection (Gough et al., 1988a). Once introduced into an area, the spread of APV is rapid with morbidity reaching 100% (Alexander, 1991; Anon., 1985; Cook et al., 1991; Gough et al., 1988a), though how the virus spreads from farm to farm is unknown. Stuart (1986) suggested APV could be spread via migratory birds as it has been shown to infect wild fowl, although the role of wild birds in the spread of APV to and between domestic populations remains unclear (Bennett et al., 2002; Bennett et al., 2004; Gough et al., 1988a; Gough et al., 2001; Gough et al., 1994; Lwamba et al., 2002; Turpin et al., 2008). Evidence for APV in wild birds comes from its ability to cause disease following experimental infection in pheasants, and guinea fowl (Gough et al., 1988a), the detection of antibodies to APV in guinea fowl (Picault et al., 1987) and the isolation of virus from wild flocks of pheasants (Gough et al., 2001) and wild Canada geese (Bennett et al., 2002). APV RNA has also been found in blue-winged teals (Bennett et al., 2002). There have been no reports of natural or experimental infection outside of avian species (Naylor and Jones, 1993).

Four subgroups of APV have been established based initially on the relatedness of the sequence of the G protein and serological cross reactivity. APV subgroups A and B have been found in Europe, Africa, Asia, and the Middle East (Banet-Noach et al., 2005), (Collins et al., 1993; Cook et al., 1993a; Hafez et al., 2000; Juhasz and Easton, 1994; Li et al., 1996; Mase et al., 2003). Subgroup C was initially isolated in Colorado, but has now been found in the states of Minnesota, North Dakota, South Dakota, Iowa and Wisconsin in the USA as well as in France and Korea (Bennett et al., 2004; Cook et al., 1999; Dar et al., 2002; Goyal et al., 2000; Seal, 1998). Subgroup D was found in France in farmed ducks (Bayon-Auboyer et al., 2000), although this subgroup has not been seen outside of these samples isolated in 1985.
1.2 Virus pathogenesis

1.2.1 Disease: turkey rhinotracheitis, swollen head syndrome, and co-infection.

APV causes respiratory illness in domestic fowl and was initially characterised as the pathogen causing outbreaks of turkey rhinotracheitis (TRT). APV primarily infects turkeys and chickens, although seroconversion of wild birds with asymptomatic infection has been reported, as described above. In experimentally infected turkeys, APV causes coughing, and ocular and sinus discharge. Associated clinical signs are sneezing, head shaking, egg loss and production of eggs with thinned shells (Anon., 1985; Cook et al., 1996; Cook et al., 2000; Jones et al., 1988; Lister and Alexander, 1986; O'Brien, 1985; Sugiyama et al., 2006; Wyeth et al., 1987). In broiler breeders, extreme cases of 98% drop in egg production have been observed (Naylor and Jones, 1993).

In chickens, APV infection has been linked to swollen head syndrome (SHS) (Aung et al., 2008; Gough et al., 1994; Jones et al., 1987; Jones et al., 1991; Lister and Alexander, 1986; Lu et al., 1994; Morley and Thomson, 1984; Nakamura et al., 1997; O'Brien, 1985; Pattison et al., 1989; Picault et al., 1987; Tanaka et al., 1995; Wyeth et al., 1987), although secondary infection clearly plays a role in the disease. An agent that is now known to be APV was first implicated when the virus was isolated from chickens with SHS (Buys et al., 1989b; Jones et al., 1991; Nakamura et al., 1997; Picault et al., 1987). While APV antibodies have been found in commercial chicken flocks, antibodies were found in birds with and without SHS or respiratory disease (Cook et al., 1988) and SHS has also been shown to occur in the absence of APV infection (Droual and Woolcock, 1994; Shirai et al., 1993). It seems clear that APV infection alone is not sufficient to induce, or may not be the sole cause of, SHS. However, as attempts to reproduce SHS experimentally in boiler chicks by 32 days with co-infection of APV and the most commonly found co-isolate in SHS (E. coli) failed, the role of APV in this syndrome remains unclear (Al-Ankari et al., 2001).

In the field, APV infection is seldom the only pathogen infecting commercial flocks and co-infection can result in up to 90% mortality (Alexander, 1991; Gough et al., 1988b), although mortality without co-infection is more normally 2-5% (Jing et al., 1993). APV
infection appears to have an immunosuppressive effect in turkeys (Chary et al., 2002a; Jones et al., 1992; Timms, 1986) such that prior APV infection rendered hemorrhagic enteritis virus vaccine approximately 50% less effective than vaccination alone (Chary et al., 2002b). This immunosuppression may leave birds open to bacterial infection (Jirjis et al., 2004; Van de Zande et al., 2001), possibly increasing the susceptibility of turkeys to secondary infection by damaging the ciliated and non-ciliated epithelia cells of the turbinates (Majo et al., 1996). A mechanism of ‘viral priming’ has been proposed, based on the finding that APV infection prior to infection with Ornithobacterium rhinotracheale was show to cause respiratory disease where bacterial infection alone did not (Marien et al., 2005). Experimental co-infection of APV and E. coli O78:K80 resulted in increased respiratory symptoms and lesions in chickens (Majo et al., 1997). Co-infection of APV with Bordetella avium and Pasteurella-like organisms showed an increased range of virus infection to the heart, liver, spleen, kidneys and caecal tonsils for a short time post infection, and increased the number of sites of infection of Pasteurella-like organisms (Cook et al., 1991). Mycoplasma gallisepticum infection following APV infection was more invasive and increased levels of mortality were reported, although the severity of illness in dually infected birds was not affected (Naylor et al., 1992). Al-Ankari et al. (2001) found in broiler chicks that co-infection of APV and a mix of pathogenic E. coli strains exacerbated the disease presentation in the birds.

1.2.2 Tissue tropism
As with all pneumoviruses, APV is thought to be primarily a respiratory pathogen though the exact tropism of the virus has been shown to be strain and host (chickens or turkeys) dependent (Cook et al., 1993b). APV subgroup A virus was found in the turbinates, sinuses, upper and lower trachea, lungs and air sacs of experimentally infected turkeys periodically up to 10 d.p.i. whereas APV subgroup B was found mainly in the upper respiratory tract (Buys et al., 1989a; Cook et al., 1991; Jones et al., 1988; McDougall and Cook, 1986; Van de Zande et al., 1999). Indirect immunofluorescence experiments confirmed the presence of APV antigen in the turbinates and trachea but not lungs or air sacs (Jones et al., 1988). Turkey poults and chickens experimentally infected with APV showed that the virus associated with the epithelial ciliated cells of the turbinates and
trachea (Majo et al., 1995).

In chickens, \textit{in vitro} experiments have shown ciliostasis in tracheal organ cultures and syncytia formation in chicken embryo cell cultures following of APV infection (Gough et al., 1988a). In experimentally infected chickens, virus was isolated from the respiratory tract (nasal tissue, sinuses, trachea, and lung), although it was found most abundantly in the upper respiratory tract tissues. In these experiments, no virus was isolated from the kidney, liver, duodenum, bursa of Fabricius or caecal tonsils. Using IP, virus was found most consistently in nasal turbinates up to 5 days post infection (d.p.i.) and was found occasionally in sinuses and trachea on days 4 and 5 d.p.i. Histology of the upper respiratory tract tissues showed changes most strikingly in the nasal turbinates and less severely in the sinuses and trachea (Catelli et al., 1998).

The association of APV infection with egg drop syndrome suggested that the virus was capable of infecting the reproductive tract. However, direct experimental evidence is mixed, with Giraud et al. (1986) finding no evidence of transmission of the virus to eggs in turkeys, but several other groups finding APV in the reproductive tract of broiler breeders and laying turkeys (Anon., 1985; Lister and Alexander, 1986; O'Brien, 1985; Wyeth \textit{et al}., 1987). In chickens, Jones et al. (1988) found APV in the reproductive tract of laying hens, as well as in the upper respiratory tract, but not in blood or ovaries and APV antigen has been detected by IP in oviduct in laying hens (Cook \textit{et al}., 2000). Shedding of small amounts of APV from turkeys up to 14 d.p.i. has been reported (Cook \textit{et al}., 1991), again suggesting that APV is able to infect tissues outside of the respiratory tract.

\subsection*{1.2.3 Vaccines}

The production of live-attenuated APV vaccines that cannot revert to virulence in the field, or killed vaccines that effectively protect the birds from further infection is an ongoing challenge for APV research. A method of priming with live-attenuated virus followed by administration of inactivated virus was shown to be required to protect turkeys and chickens against both respiratory symptoms and loss of egg production.
resulting from APV challenge (Cook et al., 1996; Cook et al., 2000). Sugiyama et al. (2006) showed that the choice of challenge virus was important in establishing efficacy of a vaccine against egg drop and respiratory symptoms. Problems of reversion to virulence were found with a live attenuated APV subgroup A vaccine which was shown to persist in turkeys (Catelli et al., 2006). Additionally, Kapczynski et al. (2008) showed that a simple intranasal vaccination with an inactivated APV did not protect turkeys from APV challenge, indicating that the method of vaccine delivery also needs to be taken into consideration. Encouragingly, initial trials of a cold-adapted strain of APV showed that it protects turkeys for 14 weeks without causing disease (Patnayak and Goyal, 2006; Patnayak et al., 2002). Additionally, Ganapathy and Jones (2007) showed that chicks could be protected with an APV subgroup B live attenuated vaccine for up to 49 days after vaccination. Due to the possibility of APV vaccination affecting the outcome of infection with other avian pathogens (section 1.2.1), concerns about the use of APV vaccines remain.

The mechanisms of protection of APV vaccines have been investigated and it is clear that more research is required in this area. Interestingly, it has been shown that pouls can be vaccinated even when they have circulating maternal antibodies to APV without decrease in protection (Catelli et al., 1998; Cook et al., 1989). Jones et al. (1992) showed that even with immunosuppression of antibody production using cyclophosphamide, turkey pouls vaccinated with a live attenuated APV strain were protected from challenge 21 days later. For an APV subgroup C vaccine, the absence of antibodies in turkey pouls following vaccination still resulted in protection from challenge with APV (Patnayak and Goyal, 2004; Patnayak et al., 2002).

The immunopathogenesis of the virus is not well studied, but it has been shown that APV may mediate the inhibition of T cells (Chary et al., 2002a), trigger the accumulation of CD4+ T cells, and the release of cytokines (Liman and Rautenschlein, 2007). Specifically, APV induces increased CD4+ T cell proliferation in the spleen and Harderian gland of turkeys. APV infection also induced humoral and cell-mediated responses as well as protective antibodies in turkeys (Liman and Rautenschlein, 2007).
1.3. *Pneumovirinae* genome organization and protein products

1.3.1 *Pneumovirinae* genomes

The APV subgroup A isolate CVL14/1 genome is 13370 base in length, which comparable to the hMPV isolate 001 genome (13387 bases) but shorter than either the RSV strain S2 (15190 bases) or PVM strain 15 (14887 bases) genomes (for Genbank accession numbers, see section 2.15). As can be seen in figure 1.1, all pneumoviruses encode three proteins that form the RNP complex (the N, P, and L proteins), three glycoproteins (the G, F and SH proteins), a structural matrix protein (M) and an M2 gene that has two overlapping ORFs. The M2 mRNA produces an anti-termination factor (M2-1) and a protein of unknown function (M2-2).

The APV genome differs from the RSV and PVM genomes in several respects. APV lacks the NS1 and NS2 genes (Randhawa *et al.*, 1997) and the genes of APV differ in order from RSV and PVM: the F and M2 genes of APV are found between M and SH instead of between G and L as is seen for RSV and PVM. Additionally, PVM has been shown to encode a second P protein (P-2) from the P gene (Barr *et al.*, 1994). For APV, only the M2 gene and possibly the N gene express more than one protein (Alvarez and Seal, 2005).

A comparison of representative genomes of APV subgroups A, C (full genome sequences for subgroups B and D are not currently available), hMPV, RSV and PVM was carried out (table 1.1, section 2.15). Overall, the representative viruses of the subfamily *Pneumovirinae* have between 40 and 66% homology, with the APV subgroup A genome having the highest levels of homology with the APV subgroup C genome (59%), followed closely by the hMPV genome (57%) and more distantly by the RSV and PVM genomes (39 and 43%, respectively). Interestingly, the APV subgroup C genome has the highest level of homology with hMPV (66%). Lwamba *et al.* (2005) speculated that the high levels of homology seen between the APV subgroup C and hMPV SH, G, and L sequences indicated an evolutionary relationship.
Table 1.1 Nucleotide homologies of the pneumovirus genomes. The percentage homologies (in green boxes) of the genomes of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15. Genomes were not rearranged to account for the differences in gene order of the viruses, possibly leading to artificially low levels of homology between the pneumovirus and the metapneumovirus genomes. The levels of nucleotide homology and amino acid identity for the individual virus genes and proteins are detailed in tables 1.3 through 1.13.

<table>
<thead>
<tr>
<th>Percentage homology (%)</th>
<th>APV subgroup A</th>
<th>APV subgroup C</th>
<th>hMPV</th>
<th>PVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV subgroup C</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hMPV</td>
<td>57</td>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PVM</td>
<td>43</td>
<td>45</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>RSV</td>
<td>39</td>
<td>41</td>
<td>40</td>
<td>52</td>
</tr>
</tbody>
</table>

The percentage homologies of the pneumovirus genomes were compared as described in section 2.15. Genomes were not rearranged to account for the differences in gene order of the viruses, possibly leading to artificially low levels of homology between the pneumovirus and the metapneumovirus genomes. The levels of nucleotide homology and amino acid identity for the individual virus genes and proteins are detailed in tables 1.3 through 1.13.
1.3.2 Pneumovirinae genes and proteins

1.3.2.1 N protein

The subfamily pneumovirus N proteins are well conserved in length, with APV subgroup A, B and RSV having 391 amino acids, APV subgroup C and hMPV having 394 amino acids, and PVM having 393 amino acids (table 1.2). The levels of homology between the pneumovirus N genes and of amino acid identity between the N proteins were also determined (table 1.3). The levels of nucleotide homology between the APV subgroups A, B and C ranged from 65% to 76%, with the highest levels seen between subgroups A and B (76%). Within the genus Metapneumovirus, the highest levels of homology were seen between the APV subgroup C and the hMPV N genes (65%). The nucleotide homology levels within the pneumovirus genus and between the pneumovirus and metapneumovirus genera were lower, with the highest level being between the PVM and RSV N proteins (62%). The protein identity levels showed a similar pattern, although here the highest level was seen between APV subgroups A and B (91%). The amino acid identity levels within the Metapneumovirus genus were higher than their nucleotide homology levels and ranged from 69 to 91%. As before, the level of homology between APV subgroup C and hMPV (88%) was higher than that seen between APV subgroup C or hMPV and any other virus.

The model for the functional processes undertaken by the pneumovirus RNP complex suggests the N protein is an essential structural pneumovirus protein that encapsidates the viral RNA and binds to the P and M2-1 proteins. Experimental evidence for this model is extensive. The bRSV N protein can spontaneously form nucleocapsid structures with RNA even when expressed in the absence of any other viral proteins (Maclellan et al., 2007). Khatter et al. (2000) showed that the bRSV N protein has the ability to encapsidate RNA, is integral to RNA synthesis, and binds to the P protein. Evidence for N:P protein binding was found when the RSV N and P protein complex was shown to be required for the formation of inclusion bodies in cells. The RSV M2-1 protein was also bound to the N protein in these bodies (Garcia et al., 1993). The RSV N, P and M2 proteins have been shown to co-localize using immunofluorescence staining (Ghildyal et al., 2002). The bRSV N protein has been shown to have specific, separate regions to
<table>
<thead>
<tr>
<th>Length</th>
<th>APV subgroup A</th>
<th>APV subgroup B</th>
<th>APV subgroup C</th>
<th>APV subgroup D</th>
<th>hMPV</th>
<th>PVM</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1176 (391)</td>
<td>1176 (391)</td>
<td>1185 (394)</td>
<td>Not available</td>
<td>1185 (394)</td>
<td>1183 (393)</td>
<td>1176 (391)</td>
</tr>
<tr>
<td>P</td>
<td>837 (278)</td>
<td>840 (279)</td>
<td>885 (294)</td>
<td>Not available</td>
<td>885 (294)</td>
<td>888 (295)</td>
<td>726 (241)</td>
</tr>
<tr>
<td>M2*</td>
<td>Not available</td>
<td>Not available</td>
<td>Not available</td>
<td>Not available</td>
<td>746</td>
<td>875</td>
<td>961</td>
</tr>
<tr>
<td>M2-1</td>
<td>762 (561)</td>
<td>186 (222)</td>
<td>748 (184)</td>
<td>Not available</td>
<td>555</td>
<td>187</td>
<td>176</td>
</tr>
<tr>
<td>M2-2</td>
<td>561 (186)</td>
<td>222 (73)</td>
<td>555 (216)</td>
<td>Not available</td>
<td>184</td>
<td>216</td>
<td>297</td>
</tr>
<tr>
<td>G</td>
<td>1176 (391)</td>
<td>Not available</td>
<td>1758 (585)</td>
<td>1170 (389)</td>
<td>1170 (389)</td>
<td>711</td>
<td>1092</td>
</tr>
<tr>
<td>F</td>
<td>1617 (538)</td>
<td>Not available</td>
<td>1614 (537)</td>
<td>Not available</td>
<td>1614 (537)</td>
<td>1614 (537)</td>
<td>1725 (574)</td>
</tr>
<tr>
<td>M</td>
<td>765 (254)</td>
<td>805 (266)</td>
<td>765 (254)</td>
<td>Not available</td>
<td>765</td>
<td>774</td>
<td>771</td>
</tr>
<tr>
<td>SH</td>
<td>525 (174)</td>
<td>Not available</td>
<td>528 (175)</td>
<td>Not available</td>
<td>528</td>
<td>279</td>
<td>195</td>
</tr>
<tr>
<td>NS-1</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>342</td>
<td>420</td>
<td></td>
</tr>
<tr>
<td>NS-2</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>113</td>
<td>139</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.2 Amino acid and nucleotide lengths of the pneumovirus ORF and proteins.

The lengths of the ORFs (nucleotides, in bold) and proteins (amino acids, in italics) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup B, APV subgroup C, APV subgroup D, hMPV isolate 001 and PVM strain 15 (section 2.15). Shown are the lengths of the N (nucleocapsid), P (phospho-), L (large polymerase), M2 (second matrix), M2-1 (second matrix ORF 1), M2-2 (second matrix ORF 2), G (glyco-), F (fusion), M (matrix), SH (small hydrophobic), NS-1 (first non-structural), NS-2 (second non-structural) open reading frames (ORFs) and proteins. *For M2, the length indicated is the number of nucleotides in the M2 mRNA.
Table 1.3 Percentage nucleotide homology and amino acid identity of the pneumovirus N ORF and proteins. The percentage homologies of the nucleotides of the N ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup B isolate 98103, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.

<table>
<thead>
<tr>
<th></th>
<th>APV subgroup A</th>
<th>APV subgroup B</th>
<th>APV subgroup C</th>
<th>hMPV</th>
<th>PVM</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV subgroup A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APV subgroup B</td>
<td>76</td>
<td>91</td>
<td>70</td>
<td>68</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>APV subgroup C</td>
<td>65</td>
<td>66</td>
<td>71</td>
<td>69</td>
<td>43</td>
<td>39</td>
</tr>
<tr>
<td>hMPV</td>
<td>65</td>
<td>66</td>
<td>75</td>
<td>88</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td>PVM</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>51</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>RSV</td>
<td>48</td>
<td>50</td>
<td>49</td>
<td>50</td>
<td></td>
<td>62</td>
</tr>
</tbody>
</table>

Table 1.3 Percentage nucleotide homology and amino acid identity of the pneumovirus N ORF and proteins. The percentage homologies of the nucleotides of the N ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup B isolate 98103, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.
bind to itself, as well as the P and M2 proteins (Samal et al., 1993). Little of the N protein sequence appears to be superfluous, as two-hybrid system experiments showed that deletion of any region of the RSV N protein obliterated N protein binding (Garcia-Barreno et al., 1996; Slack and Easton, 1998).

1.3.2.2 P protein

The subfamily pneumovirus P proteins vary in length from the RSV protein of 241 residues to the 295 residue PVM P protein (table 1.2). The APV subgroup C and hMPV P proteins are identical in length (294 residues) whereas the APV subgroup A and B P proteins are shorter, at 278 and 279 residues, respectively. The levels of homology or identity of the P genes and proteins were determined (table 1.4). The highest levels of nucleotide homology were seen between APV subgroups A and B (69%) and between APV subgroup C and hMPV (67 1%). The P protein identity levels were also highest between APV subgroups A and B at 72% and between APV subgroup C and hMPV P proteins (68%). RSV and PVM had no more than 42% nucleotide homology or 27% amino acid identity with any of the metapneumoviruses and only 49% nucleotide homology or 33% amino acid identity between themselves.

The pneumovirus P protein is essential, making up part of the viral polymerase complex. The bRSV P protein has been shown to bind the N and L proteins in co-immunoprecipitation studies (Khattar et al., 2001). The regions of the RSV P protein that are involved in binding have been found to be two regions in carboxyl-terminus of the P protein for N protein binding by two hybrid studies, co-immunoprecipitation and the formation of inclusion bodies (Garcia-Barreno et al., 1996; Slack and Easton, 1998). Further RSV two-hybrid system studies showed that the RSV P protein can bind to itself, N protein, and NS1 protein (Hengst and Kiefer, 2000). The P protein also appears to be responsible for host specificity for human and bovine RSV (Buchholz et al., 2000; Schmidt et al., 2004).
Table 1.4 Percentage nucleotide homology and amino acid identity of the pneumovirus P ORF and proteins. The percentage homologies of the nucleotides of the P ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup B isolate 98103, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.
1.3.2.3 L protein
The L proteins of the pneumovirus subfamily vary in length from 2004 amino acids for APV subgroups A and C to 2164 amino acids for RSV (table 1.2). The L gene sequences for APV subgroups B and D are not currently available. The levels of nucleotide homology for the L gene are over 50% for all members of the pneumovirus subfamily (table 1.5), with the highest levels found between hMPV and APV subgroup C at 71%. The APV subgroup A L gene has a 63% nucleotide homology with those of the APV subgroup C and hMPV, but only 55 and 50% homology with those of PVM and RSV, respectively. The levels of identity between the L proteins are similar, with APV subgroup C and hMPV having 80% amino acid identity and APV subgroup A having identical amino acid identity percentages as nucleotide homology percentages to APV subgroup C and hMPV.

The L protein is the large polymerase subunit and has been shown to be the main catalytic unit of the polymerase complex for negative-strand RNA viruses (Grdzelishvili et al., 2005; Hercyk et al., 1988; Ogino et al., 2005). As discussed above, the L protein for RSV binds the P protein to make up the viral RNA dependent RNA polymerase. Analysis of the polymerase proteins of five negative sense RNA viruses (Sendai virus, Newcastle disease virus, measles virus, vesicular stomatitis virus and rabies virus) defined six conserved domains (Poch et al., 1990). No direct study of the APV L protein functionality has been carried out.

1.3.2.4 M2-1 and M2-2 proteins
The M2 mRNA of the pneumovirus subfamily is unique, with no homologue in any other system. The organization of the M2 mRNA is unusual in that it contain two overlapping ORF, both of which produce viral proteins. Translation of the M2-2 protein from the M2 mRNA is coupled to translation of the M2-1 protein as the overlapping region of the M2-1 and M2-2 ORFs has been shown to direct the ribosome back to the start codon of the M2-2 ORF from the M2-1 stop codon (Ahmadian et al., 2000; Gould and Easton, 2005; Gould and Easton, 2007). The M2 mRNAs of the pneumoviruses vary in length from 727 nucleotides for hMPV to 952 nucleotides for RSV (table 1.2). The levels of
Table 1.5 Percentage nucleotide homology and amino acid identity of the pneumovirus L genes and proteins. The percentage homologies of the nucleotides of the L ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup B isolate 98103, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.

<table>
<thead>
<tr>
<th></th>
<th>APV subgroup A</th>
<th>APV subgroup C</th>
<th>hMPV</th>
<th>PVM</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV subgroup A</td>
<td>63</td>
<td>63</td>
<td>50</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>APV subgroup C</td>
<td>63</td>
<td>80</td>
<td>48</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>hMPV</td>
<td>63</td>
<td>71</td>
<td>49</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>PVM</td>
<td>55</td>
<td>54</td>
<td>55</td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>RSV</td>
<td>50</td>
<td>50</td>
<td>52</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.6 Nucleotide homologies of the pneumovirus M2 mRNA. The percentage homology of the mRNAs of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.

<table>
<thead>
<tr>
<th>Percentage homology (%)</th>
<th>APV subgroup A</th>
<th>APV subgroup C</th>
<th>hMPV</th>
<th>PVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV subgroup C</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hMPV</td>
<td>63</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PVM</td>
<td>35</td>
<td>37</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>RSV</td>
<td>39</td>
<td>39</td>
<td>41</td>
<td>50</td>
</tr>
</tbody>
</table>
homology for the M2 mRNAs of the pneumovirus subfamily also vary widely, with APV subgroup C and hMPV having the highest levels of homology at 71% and PVM having no more than 38% homology with any of the pneumoviruses (table 1.6). The RSV M2-1 is the largest of the pneumovirus M2-1 proteins at 194 amino acids (table 1.2). The M2-1 proteins of APV subgroups A and C, hMPV and PVM are smaller, but similar in length at 186, 184, 187, and 176 amino acids, respectively. The level of amino acid identity (table 1.7) between the pneumoviruses is low (below 40%) apart from between APV subgroup C and hMPV (84%) and between APV subgroup A and APV subgroup C (70%) or hMPV (72%).

The M2-1 protein is produced from the first ORF of the M2 gene. The RSV M2-1 protein has been shown to be hydrophilic and basic (Collins and Wertz, 1985) and was characterized as the second RSV envelope-associated matrix protein (Huang et al., 1985). The RSV M2-1, L, N, P proteins and genome RNA have been shown to co-localize in cells (Garcia-Barreno et al., 1996; Li et al., 2008), suggesting a potential role for M2-1 protein. The expression of RSV M2-1 protein along with the N, P, L proteins and the genome RNA significantly enhanced viral full-length transcript production, suggesting that the M2-1 protein acts as a transcriptional elongation factor (Collins et al., 1995; Collins et al., 1996; Fears and Collins, 1999; Grosfeld et al., 1995; Yu et al., 1995).

The RSV M2-2 protein is 90 amino acid in length, similar to the 93 residue PVM M2-2 protein. In contrast, the M2-2 proteins of APV and hMPV are considerably shorter (71 or 73 amino acids, table 1.2). However, the M2-2 ORFs and proteins of the pneumovirus subfamily lack any significant sequence homology or amino acid identity, apart from hMPV and APV subgroup C that have a 64% nucleotide homology (table 1.8).

A protein expressed from the second ORF of the M2 mRNA was detected for RSV and PVM, but not for APV, by western blot. These M2-2 proteins for RSV and PVM were visualized in cells by immunofluorescence and shown to localize in the cytoplasm to inclusion bodies (Ahmadian et al., 1999). The M2-2 protein is not required for virus
Table 1.7 Percentage nucleotide homology and amino acid identity of the pneumovirus M2-1 genes and proteins. The percentage homology of the M2-1 ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.
<table>
<thead>
<tr>
<th></th>
<th>APV subgroup A</th>
<th>APV subgroup C</th>
<th>hMPV</th>
<th>PVM</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV subgroup A</td>
<td></td>
<td>20</td>
<td>26</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>APV subgroup C</td>
<td>49</td>
<td></td>
<td>56</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>hMPV</td>
<td>52</td>
<td>64</td>
<td></td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>PVM</td>
<td>36</td>
<td>32</td>
<td>35</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>RSV</td>
<td>31</td>
<td>29</td>
<td>30</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.8 Percentage nucleotide homology and amino acid identity of the pneumovirus M2-2 genes and proteins. The percentage homology of the M2-2 ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.
growth and can be deleted from RSV to give a virus which is an attenuated *in vitro* and *in vivo* (Bermingham and Collins, 1999; Jin *et al.*, 2000b). For hMPV, a deletion of the M2-2 ORF in a recombinant virus resulted in an increased number of mutations to the genome, up-regulation of transcription, and attenuated growth in hamsters. As in RSV, the hMPV M2-2 protein appears to play a role in regulation of transcription (Schickli *et al.*, 2008).

### 1.3.2.5 F protein

The F genes of the pneumoviruses are 1614 to 1725 nucleotides in length and have levels of nucleotide homology ranging from 45 to 70% (tables 1.2 and 1.9). As with the other APV genes, the highest levels of homology are seen between the APV subgroup C and the hMPV F genes. The *Pneumovirinae* F protein is a glycoprotein which is expressed as a precursor F0 protein that is subsequently cleaved into two protein (F1 and F2). Disulfide bonds covalently link the F1 and F2 proteins. The RSV F1 component is involved in fusion of the virus envelope with the host cell membrane (González-Reyes *et al.*, 2001; Zimmer *et al.*, 2002; Zimmer *et al.*, 2001). Expression of the RSV F protein in recombinant vesicular stomatitis virus has been shown to be sufficient for cell infection and fusion (Kahn *et al.*, 1999).

The F protein is essential for virus growth and has several functions beyond cell fusion. The F protein may also be responsible for the host specificity human and bovine RSV. Recombinant bRSV viruses lacking the SH and/or G genes were shown to specifically infect bovine cells and similar recombinant RSV mutants to specifically infect human cells (Karron *et al.*, 1997). Also the RSV F protein has been shown to have 15 heparin binding domains (six in the F2 subunit and nine in the F1 subunit) that may be involved in binding or infectivity of the virus (Feldman *et al.*, 2000; Schmidt *et al.*, 2004). The RSV F protein is responsible for inhibition of T cell proliferation (Schlender *et al.*, 2002; Schlender *et al.*, 2003). In human and bovine RSV, a 27 amino acid peptide that is released upon cleavage of the F0 protein may play a role in immune evasion (Begona Ruiz-Arguello *et al.*, 2002) though no analogous protein is seen in APV, PVM or hMPV.
Table 1.9 Percentage nucleotide homology and amino acid identity of the pneumovirus F genes and proteins. The percentage homology of the F ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.
The active domains of the RSV F protein have been probed using deletion mutagenesis. In this way, it has been shown that the F protein can bind to the G protein independently of any SH protein being present and the domain responsible for binding has been determined to be between amino acids 335 and 574 in the F₁ subunit region of the F protein (Schmidt et al., 2004). No information on the capacity of the APV F protein to interact with the G protein has been reported.

1.3.2.6 G protein

The G protein of the pneumoviruses varies greatly in length, from 236 and 298 residues for hMPV and RSV, respectively, to 585 residues for APV subgroup C. The APV subgroup C G protein is 194 residues and 196 residues longer than the APV subgroup A and subgroup B G proteins and has no more than 12% amino acid identity with any of the pneumoviruses (Govindarajan and Samal, 2004). Generally, for the virus sequences examined, there is a lack of significant nucleotide or amino acid homology between any of the G genes or proteins of the viruses of the pneumovirus subfamily (tables 1.2 and 1.10).

For RSV, the G protein is the viral attachment protein and has been shown to be a type 2 glycoprotein (Levine et al., 1987; Wertz et al., 1985). The search for a cellular receptor has shown that the RSV G protein can bind to heparin and glycosaminoglycans (GAGs) at a heparin-binding domain (Feldman et al., 2000; Schmidt et al., 2004) and that the hMPV G protein also binds cellular GAGs (Thammawat et al., 2008). The G gene has been deleted from recombinant RSV, resulting in a significantly attenuation virus (Bermingham and Collins, 1999; Teng et al., 2001). In a primary human epithelial airway cell culture system, while the RSV F protein is the only required glycoprotein for infection, the G protein was found to be necessary for efficient spread of the virus (Mellow et al., 2004; Zhang et al., 2002).

The G protein may also have a role in immune evasion, as a second form of G protein that is shorter that the full-length G protein and soluble has been shown to be produced by RSV (Roberts et al., 1994). It may be that this form of G protein is used as an
Table 1.10 Percentage nucleotide homology and amino acid identity of the pneumovirus G genes and proteins. The percentage homology of the G ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup C, APV subgroup D isolate Fr/85/2, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.
immunomodulator or as a decoy for the immune system (Johnson and Graham, 1999; Johnson et al., 1998). Johnson et al. (2004) showed that the removal G protein did not reduce vaccine-associated disease, but the removal did decrease vaccine protectivity.

For APV, recombinant viruses have been created which lacks the G genes. It was found that APV lacking the G gene grew well in cell culture but was attenuated in turkeys (Ling et al., 2008; Naylor et al., 2004). Similar results were seen with the hMPV recombinant virus lacking the G gene (Biacchesi et al., 2005; Biacchesi et al., 2004b). The effects of SH and G gene deletions on virus growth is further discussed in chapter 6.

1.3.2.7 SH protein

The SH protein of the pneumovirus subfamily, like the G proteins, vary greatly in length from 64 residues for RSV to 183 residues for hMPV (table 1.2) and have no significant nucleotide homology or amino acid identity (table 1.11). The function of the SH protein for the pneumoviruses is not well defined.

The RSV SH protein has been shown to be a type 2 integral membrane protein (Collins and Mottet, 1993; Feldman et al., 2001; Olmsted and Collins, 1989). Recombinant viruses carrying a deletion in the SH gene have been shown to replicate in vitro and in vivo, though a recombinant RSV lacking the SH gene was attenuated (Bukreyev et al., 1997). The deletion of the SH and/or G genes from an hMPV cloned virus was also carried out and gave the following results: the G deleted virus was not attenuated in vitro, but was attenuated in hamsters, the SH deleted virus was not attenuated in vitro or in vivo, and the G/SH deleted virus was slightly attenuated in vitro and attenuated in hamsters (Biacchesi et al., 2004a; Biacchesi et al., 2004b).

For APV, recombinant viruses has been created which lacks the SH gene or the SH and the G gene. It was found that APV lacking the SH gene showed a syneytial phenotype in cell culture. This virus also had attenuated growth in cell culture and in turkeys (Ling et al., 2008; Naylor et al., 2004). APV recombinant viruses are further discussed in chapter 6.
Table 1.11 Percentage nucleotide homology and amino acid identity of the pneumovirus SH genes and proteins. The percentage homology of the SH ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.
1.3.2.8 M protein
The M proteins of the pneumovirus subfamily do not vary greatly in length (254 to 266 residues, table 1.2) and the levels of nucleotide homology and amino acid identity for the metapneumoviruses are high (67 to 76% and 72 to 84%, respectively, table 1.12). Little is known about the APV M protein, but for RSV the M protein was shown to be the matrix protein and to associate with the viral envelope (Peeples and Levine, 1979). The RSV M protein was found to associate with the RNP complex and may inhibit transcription via association with the N-terminal 110 amino acids of the M2-1 protein (Ghildyal et al., 2002; Li et al., 2008). The M protein is also required for production of infectious virus (Teng and Collins, 1998).

1.3.2.9 NS1 and NS2 proteins
The NS1 and NS2 genes are found only in PVM and RSV and no homologue has been found in the metapneumoviruses. The NS1 genes are 343 and 420 nucleotides for PVM and RSV, respectively, encoding proteins of 113 and 139 amino acids. The NS2 genes are 471 and 375 nucleotides (156 and 124 amino acids) respectively for PVM and RSV (table 1.2) and there is no detectable nucleotide homology or amino acid identity when comparing the NS1 or NS2 genes of RSV and PVM (table 1.13). For RSV, the NS1 protein has been shown to bind the M and P proteins in a two-hybrid system, but no interaction of NS2 with any other viral protein was found (Evans et al., 1996; Hengst and Kiefer, 2000). The RSV NS1 and NS2 genes can be deleted from cloned virus and result in a virus that is attenuated in chimpanzees (Bermingham and Collins, 1999; Jin et al., 2000b; Teng and Collins, 1999; Whitehead et al., 1999). Most interestingly, the RSV NS1 and NS2 proteins have been shown to act together as mediators of resistance to the anti-viral interferon response and as inhibitors of interferon activation (Bossert and Conzelmann, 2002; Bossert et al., 2003; Schlender et al., 2000). As the metapneumoviruses have no analogous proteins, it has been speculated that one of the other viral proteins might act in a homologous manner; however there has been no report of such a function for any of the APV proteins to date.
Table 1.12 Percentage nucleotide homology and amino acid identity of the pneumovirus M genes and proteins. The percentage homology of the M ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, APV subgroup B, APV subgroup C, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.

<table>
<thead>
<tr>
<th></th>
<th>APV subgroup A</th>
<th>APV subgroup B</th>
<th>APV subgroup C</th>
<th>hMPV</th>
<th>PVM</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV subgroup A</td>
<td>84</td>
<td>76</td>
<td>76</td>
<td>37</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>APV subgroup B</td>
<td>71</td>
<td>73</td>
<td>72</td>
<td>37</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>APV subgroup C</td>
<td>70</td>
<td>67</td>
<td>87</td>
<td>37</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>hMPV</td>
<td>68</td>
<td>69</td>
<td>76</td>
<td>37</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>PVM</td>
<td>48</td>
<td>49</td>
<td>48</td>
<td>47</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>RSV</td>
<td>48</td>
<td>50</td>
<td>47</td>
<td>51</td>
<td>53</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.13 Percentage nucleotide homology and amino acid identity of the pneumovirus NS1 and NS2 genes and proteins. The percentage homology of the NS1 or NS2 ORFs (green boxes) and percentage identity of the proteins (grey boxes) of RSV strain RSS-2 and PVM strain 15 were compared as described in section 2.15.

<table>
<thead>
<tr>
<th>NS1 homology or identity</th>
<th>RSV</th>
<th>PVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSV</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>PVM</td>
<td>47</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NS2 homology or identity</th>
<th>RSV</th>
<th>PVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSV</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>PVM</td>
<td>44</td>
<td></td>
</tr>
</tbody>
</table>
1.4 Virus transcription and replication

As with other viruses with RNA genomes, the same viral polymerase complex replicates the viral genome and transcribes the viral mRNAs. There is a lack of clarity as to how the polymerase complex switches from transcription to replication mode. For VSV it was shown that the extent of terminal complementarity, rather than the terminal sequences, determined the levels of replication and transcription. Replication of DI viruses with high complementarity was increased when compared to wildtype VSV (Wertz et al., 1994). For RSV the switch may involve several factors, including genome encapsidation, speed of processivity of the polymerase, and levels of the M2-2 accessory protein (Banerjee et al., 1977; Cowton et al., 2006; Lippincott et al., 2001).

Purified RSV genome RNA is non-infectious and the minimal infectious unit has been shown to be the genome RNA together with the N, P and L proteins (Marriott and Easton, 1999). The key components in replication and transcription are the signalling sequences in the viral genome (the leader, trailer, gene start and gene end sequences), the N protein encapsidated negative sense RNA genome, the N protein encapsidated positive sense RNA antigenome, and the polymerase complex (consisting of the P and L proteins). The M2-1 and M2-2 proteins are non-essential accessory proteins that may affect transcription and replication for some viruses.

1.4.1 Virus transcription

A model for APV transcription has been developed based on the well studied RSV, VSV and SeV models (reviewed in (Cowton et al., 2006))(figure 1.4). It has been established that the negative sense RNA viruses carry out sequential transcription of genes with a single point of initiation of transcription, proceeding from the 3’ to 5’ end of the genome (Abraham and Banerjee, 1976; Ball and White, 1976). Transcription by the viral polymerase complex of the encapsidated RNA genome produces capped, poly-adenylated mRNA transcripts. These mRNAs are produced in a stop-start process, in which the viral polymerase complex starts transcription at the gene start (GS) sequence and transcribes the mRNA until it reaches the gene end (GE) sequence. The mRNA is then polyadenylated, transcription is terminated, and the transcript is released (Barik, 1993).
Figure 1.4 Model of Paramyxovirus transcription. In the model of Paramyxovirus transcription the viral polymerase complex (blue) binds to the 3’ end of the genome, begins transcription of the first gene at the gene start sequence and produces a capped (Cap) mRNA with a polyA tail ((A)_n) added at the gene end sequence. The polymerase complex then either scans through the intergenic region and transcribes the next gene, or disassociates from the genome and starts again at the 3’ end of the genome. In this way, a gradient of mRNA transcripts is produced.
The polyadenylation of the mRNA is carried out by the viral polymerase, probably via a mechanism of ‘reiterative transcription’ at a series of U residues (Barr and Wertz, 2001). How and why the polymerase stops the iteration and releases the transcript is unknown. However, termination of the upstream gene must occur before re-initiation of transcription of the downstream gene.

The re-initiation of transcription of the polymerase complex at downstream gene start sequence is relatively inefficient. Once the polymerase complex fails to re-initiate at the next available gene it disengages from the template and can only re-initiate by binding to the 3’ end of the genome RNA where the process begins again. The result is the generation of a gradient of transcription where the steepness of the gradient is determined by the efficiency of the re-initiation event. Hence, the genes at the 3’ end of the genome are more frequently transcribed than those at the 5’ end of the genome.

1.4.2 Transcriptional regulatory sequences

When first sequenced, the discrete gene start (GS) and gene end (GE) sequences of parainfluenza virus type 3, a member of the Paramyxoviridae family, were said to “almost invite foreign gene insertions.” (Brian Murphy, quoted by Schmidt et al. (2004)). Indeed, any gene flanked by the GS and GE signals can be transcribed (when present as negative sense RNA that is encapsidated with the viral nucleocapsid protein) by the viral polymerase.

The GS sequences of the pneumovirus subfamily are highly conserved sequences of nine to thirteen bases (table 1.14). Each gene of APV subtype A has a GS sequence of 5’ GGGACAAGU 3’ (mRNA sense), apart from the L gene that has a GS sequence of 5’ AGGACCAAU 3’ (mRNA sense). The genes of RSV strain RSS-2 have a consensus GS sequence of 5’ GGGGCAAAUA 3’ (mRNA sense) that is found upstream of each gene apart from the M2 and L genes. PVM stain 15 and hMPV isolate 001 have more variation in their GS sequences, but a nine or thirteen base (respectively) sequence is found before each gene with no more than three deviations from the consensus sequence.
Table 1.14 Pneumovirus gene start (GS) sequences. The source of the sequences by Genbank number is detailed in section 2.15. The non-consensus bases are indicated in bold and in blue. ‘r’ indicates a purine at this position and ‘y’ indicates a pyrimidine. All sequences are shown 5’ to 3’ mRNA sense.

<table>
<thead>
<tr>
<th>GS sequence</th>
<th>APV</th>
<th>hMPV</th>
<th>RSV</th>
<th>PVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS1</td>
<td>N/A</td>
<td>N/A</td>
<td>GGGGCAAAUA</td>
<td>AGGACAAAGU</td>
</tr>
<tr>
<td>NS2</td>
<td>N/A</td>
<td>N/A</td>
<td>GGGGCAAAUA</td>
<td>AGGACAAAGU</td>
</tr>
<tr>
<td>N</td>
<td>GGGACAAGU</td>
<td>GGGACAAGUGAAA</td>
<td>GGGGCAAAUA</td>
<td>AGGAUAUU</td>
</tr>
<tr>
<td>P</td>
<td>GGGACAAGU</td>
<td>GGGACAAGUCAAA</td>
<td>GGGGCAAAUA</td>
<td>AGGAUAUU</td>
</tr>
<tr>
<td>M</td>
<td>GGGACAAGU</td>
<td>GGGACAAGUAAAAA</td>
<td>GGGGCAAAUA</td>
<td>AGGACAAAU</td>
</tr>
<tr>
<td>F</td>
<td>GGGACAAGU</td>
<td>GGGACAAGUAAAAA</td>
<td>GGGGCAAAUA</td>
<td>AGGAUAUU</td>
</tr>
<tr>
<td>M2</td>
<td>GGGACAAGU</td>
<td>GGGACAAUCAUA</td>
<td>GGGGCAAUUG</td>
<td>AGGAUAAGU</td>
</tr>
<tr>
<td>SH</td>
<td>GGGACAAGU</td>
<td>GGGAUUAAUGCA</td>
<td>GGGGCAAAUA</td>
<td>AGGACAAAU</td>
</tr>
<tr>
<td>G</td>
<td>GGGACAAGU</td>
<td>GGGACAAGUAGUU</td>
<td>GGGGCAAAUA</td>
<td>AGGAUGAGU</td>
</tr>
<tr>
<td>L</td>
<td>AGGACCAAU</td>
<td>GAGACAAUAAACA</td>
<td>GGGACAAUAUG</td>
<td>AGGACAAU</td>
</tr>
<tr>
<td>consensus</td>
<td>GGGACAAGU</td>
<td>GGGACAAAUAAGUA</td>
<td>GGGGCAAAUA</td>
<td>AGGAYAARU</td>
</tr>
</tbody>
</table>

Sequence Table 1.14 Pneumovirus gene start (GS) sequences. The source of the sequences by Genbank number is detailed in section 2.15. The non-consensus bases are indicated in bold and in blue. ‘r’ indicates a purine at this position and ‘y’ indicates a pyrimidine. All sequences are shown 5’ to 3’ mRNA sense.
For APV, hMPV and RSV, the GS sequences have a series of three G residues at the 5’ end (mRNA sense) and are A/G rich. The first seven bases of the metapneumoviruses APV and hMPV GS consensus sequences match exactly, although the pneumoviruses PVM and RSV GS sequences have no such degree of conservation. Mutational analysis of the APV GS sequence was carried out and it was shown that any mutation was detrimental to transcription, apart from a change of the C at position 5 to a U (Edworthy and Easton, 2005). Kuo et al. (1997) carried out mutational analysis of the RSV GS sequence, showing that mutations at positions 1, 3, 6, 7 and 9 resulted in loss of expression from the associated gene, whereas mutations to position 5 were well tolerated. Interestingly, the RSV vaccine strain candidate cpts248/404 has been shown to have a mutation at position 9 (U to C, mRNA sense) of the M2 GS sequence and this mutation alone causes virus attenuation in mice (Whitehead et al., 1998). Taken together, this suggests that mutations in GS sequence mutations may attenuate virus \textit{in vivo} via down regulation of viral genes.

The GE sequences of the pneumovirus subfamily are less well conserved than the GS sequences (table 1.15), but are functionally important. For example, for RSV, the GS and GE signals of clinical isolates of RSV were compared and, while the GS sequences were conserved, the GE sequences varied. It was speculated changes in the GE sequence of the G gene down-regulated F protein expression (Moudy \textit{et al.}, 2003; Wertz and Moudy, 2004). Indeed, it has been shown that the GE sequence is necessary to terminate mRNA transcription and allow re-initiation at the downstream GS sequence. Loss of the GE sequence results in a high level of read-through mRNAs being produced and these read-through mRNAs do not produce protein efficiently from the second ORF (Kuo \textit{et al.}, 1996b).

The nucleotides between the GE sequence of one gene and the GS sequence of the next gene are termed the intergenic region nucleotides. For RSV, there is little variation in the levels of polymerase that scan through the region although the length of this region varies from 1 to 53 nucleotides, (table 1.16)(Kuo \textit{et al.}, 1996a). The substitution of foreign sequence into the intergenic region of RSV had no effect on transcription,
<table>
<thead>
<tr>
<th>GE sequence</th>
<th>APV</th>
<th>hMPV</th>
<th>RSV</th>
<th>PVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS1</td>
<td>N/A</td>
<td>N/A</td>
<td>UAGUUA AUAU AAAA</td>
<td>UAGUUA AAU AAAA</td>
</tr>
<tr>
<td>NS2</td>
<td>N/A</td>
<td>N/A</td>
<td>UAGUAA UUU AAAA</td>
<td>UAGUUA U AG AAAA</td>
</tr>
<tr>
<td>N</td>
<td>AGU AAUU AAAAAA</td>
<td>AGUA AUU AAAAAA</td>
<td>GAGUUA AU AAAAAAA</td>
<td>UAUUUA AAU AAAA</td>
</tr>
<tr>
<td>P</td>
<td>AGU UAUG AAAAAA</td>
<td>AGUU UAAU AAAAAA</td>
<td>GAGUUA AU AAAAAAA</td>
<td>UAGUUA AAU AAAA</td>
</tr>
<tr>
<td>M</td>
<td>AGU CAUU AAAAAA</td>
<td>AGUU AUU AAAAAA</td>
<td>AAGUUA AU AAAAAA</td>
<td>UAGUUA AAU AAAA</td>
</tr>
<tr>
<td>F</td>
<td>AGU UAUU AA AAAA</td>
<td>AGUU AAU AAAAAA</td>
<td>UAGUUA AAU AAAA</td>
<td>UAGUUA C AAAAAA</td>
</tr>
<tr>
<td>M2</td>
<td>AGU UAUU AA AAAA</td>
<td>AGUU AAU AAAAAA</td>
<td>UAGU CAU AA AAAA</td>
<td>UAGUUA AAU AAAA</td>
</tr>
<tr>
<td>SH</td>
<td>AUU UAAUU AAAAAA</td>
<td>AGUU AAUU AAAAAA</td>
<td>UAGUUA UAU AAAA</td>
<td>UAGUUA AAU AAAA</td>
</tr>
<tr>
<td>G</td>
<td>AUA AU AAAAAAAA</td>
<td>AGUU AAC AAAAAA</td>
<td>UAGUUA UUU AAAA</td>
<td>UAGUUAU AU AAAAAA</td>
</tr>
<tr>
<td>L</td>
<td>AGU UAUU AAAAAA</td>
<td>AGUU AAUU AAAAAA</td>
<td>UAGUUA UUU AAAA</td>
<td>UAGUUA C AAAAAA</td>
</tr>
<tr>
<td>consensus</td>
<td>AGU (n)$<em>{2-5}$ (A)$</em>{n}$</td>
<td>AGUU (n)$<em>{3-4}$ (A)$</em>{n}$</td>
<td>UAGUUA (n)$<em>{2-4}$ (A)$</em>{n}$</td>
<td>UAGUUA (n)$<em>{1-2}$ (A)$</em>{n}$</td>
</tr>
</tbody>
</table>

**Table 1.15 Pneumovirus gene end (GE) sequences.** The source of the sequences by Genbank number is detailed in section 2.15. The non-consensus bases are indicated in bold and in blue. ‘r’ indicates a purine at this position and ‘y’ indicates a pyrimidine. All sequences are shown 5’ to 3’ mRNA sense.
<table>
<thead>
<tr>
<th>IR length</th>
<th>APV</th>
<th>hMPV</th>
<th>IR length</th>
<th>RSV</th>
<th>PVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-P</td>
<td>3</td>
<td>2</td>
<td>NS1-NS2</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>P-M</td>
<td>1</td>
<td>8</td>
<td>NS1-N</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>M-F</td>
<td>2</td>
<td>34</td>
<td>N-P</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>F-M2</td>
<td>2</td>
<td>41</td>
<td>P-M</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>M2-SH</td>
<td>29</td>
<td>11</td>
<td>M-SH</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>SH-G</td>
<td>8</td>
<td>126</td>
<td>SH-G</td>
<td>44</td>
<td>2</td>
</tr>
<tr>
<td>G-L</td>
<td>60</td>
<td>612</td>
<td>G-F</td>
<td>53</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F-M2</td>
<td>45</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 1.16 Pneumovirus intergenic region lengths in nucleotides. The number of nucleotides found between the GE and GS sequences of the genes of the pneumoviruses is shown. The source of the sequences by Genbank number is detailed in section 2.15.
suggesting that there is a great degree of tolerance of sequence variation in these regions of the genome (Kuo et al., 1996a). The intergenic regions of APV, hMPV and PVM are similarly divergent, ranging from 1 to 60 nucleotides for APV, 2 to 189 nucleotides for hMPV, and 2 to 56 nucleotides for PVM (table 1.17).

Looking specifically at the Le sequence and initiation of transcription at the 3’ end of the genome, two models have been proposed. The first, supported by work done on SeV and other paramyxoviruses, suggests that the viral polymerase complex initiates transcription at the 3’ end of the genome, transcribing until the polymerase complex recognizes the first GS sequence and then releasing a short leader RNA before beginning transcription of the first gene. The second model, supported by work done on VSV and other rhabdoviruses, proposes that the polymerase complex initiates transcription directly at the GS sequence, without producing an mRNA from the leader region. A consistent feature of both models is that there are regions of leader sequence that are required for transcription and that the polymerase complex binding sequence is contained within the first 11 nucleotides of the leader region (Cowton et al., 2006).

1.4.3 Virus replication
The model for replication of the negative sense RNA virus genome (figure 1.5) utilizes the same viral polymerase complex as is used for viral transcription. However, during replication, when the polymerase binds to the 3’ terminus leader region of the genome, the polymerase commences transcription of the genome immediately and fails to recognize any of the transcription signals. In this way, an ‘antigenome’ copy of the genome is created and immediately encapsidated with N protein (Cowton et al., 2006). This antigenome has been shown to be a necessary intermediate to genome replication, as the template from which new genomes can be copied (reviewed in (Banerjee et al., 1977)).

1.4.4 Replication regulatory sequences
For replication, the extreme 5’ and 3’ terminal sequences contain the signals to direct the polymerase complex to replicate the genome. At the 3’ end of the genome, the leader
Figure 1.5 Model of Paramyxovirus replication. In the model of Paramyxovirus genome replication, the viral polymerase complex bind at the 3’ end of the genomic RNA and transcribes to the 5’ end, producing an antigenome RNA. The polymerase complex then bind to the 3’ end of the antigenome RNA and transcribes a copy of the genomic RNA.
Table 1.17 Nucleotide homologies of the pneumovirus Le and Tr sequences. The percentage nucleotide homologies of the Le sequences (blue boxes) and Tr sequences (green boxes) of RSV strain RSS-2, APV subgroup A isolate CVL141, hMPV isolate 001 and PVM strain 15 were compared as described in section 2.15.

<table>
<thead>
<tr>
<th>Percentage homology (%)</th>
<th>APV</th>
<th>hMPV</th>
<th>PVM</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV</td>
<td>100</td>
<td>41</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>hMPV</td>
<td>22</td>
<td>100</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>PVM</td>
<td>13</td>
<td>35</td>
<td>100</td>
<td>69</td>
</tr>
<tr>
<td>RSV</td>
<td>16</td>
<td>50</td>
<td>32</td>
<td>100</td>
</tr>
</tbody>
</table>
(Le) sequence extends from base 1 to the GS sequence of the first gene. At the 5' end of the genome, the trailer (Tr) sequence extends from the end of the L gene GE sequence to the end of the genome. As with the GS and GE sequences, these regulatory regions are able to direct replication of any sequence they flank. For the pneumovirus subfamily, the Le sequences are most highly conserved between PVM and RSV (69% nucleotide identity, table 1.17), but APV, PVM and RSV all have regions of high conservation at the ends of the Le region. The first 11 bases for all three viruses are identical apart from the fourth base and a second region of five A nucleotides is conserved for all four viruses at the 5' end of the Le region. For the Tr regions, there is no more than 50% nucleotide identity between any of the pneumoviruses (table 1.17), however the last 12 nucleotides are conserved for all four viruses apart from one base in PVM. This suggests a very well conserved process of replication for the various pneumoviruses.

1.5 Reverse genetics systems

The term ‘reverse genetics systems’ for negative sense RNA viruses refers to systems of introducing mutations into the viral genome and determining the resulting phenotype. In order to be able to manipulate the RNA genome of RNA viruses, it is first necessary to ‘convert’ it into a DNA copy that produces the RNA genome in the presence of a DNA dependent-RNA polymerase. There are three reverse genetics systems- the helper virus system, the plasmid rescue system and the full-length rescue system (Theriault et al., 2004).

The full-length rescue system aims to produce virus entirely from cDNA in vitro. The requirements for the system are a DNA genome plasmid, a DNA dependent RNA polymerase, the viral nucleocapsid and polymerase proteins and any additional helper proteins (figure 1.6). This system has been developed for RSV and hMPV (Bukreyev et al., 1997; Herfst et al., 2004). Firstly a method for creating the RNA genome from the cDNA clone was required. By cloning the genome under the control of a T7 DNA dependent RNA polymerase promoter, transcription resulted in a copy of the antigenome RNA being produced in the cytoplasm of the cell. The T7 DNA dependent RNA polymerase can be provided either by a recombinant vaccinia virus expressing T7 DNA
Figure 1.6 Model of the negative sense RNA virus full length rescue system. The components required to create a reverse genetics system for a negative sense RNA viruses are show here. VVT7 (A) is an engineered vaccinia virus used to constitutively express T7 DNA dependent RNA polymerase (T7 RNA pol) in infected HEp2 cells (B). These cells are then transfected with a genome plasmid (C) under the control of a T7 polymerase promoter (T7 Pro) that produces an RNA copy of the antigenome upon transcription by the T7 RNA polymerase, which is then terminated using a hepatitis ribozyme. Simultaneously, the cells are transfected with expression plasmids for the helper proteins (D). For RSV, these are the N, P, L and M2-1 proteins. These cells then have all the components necessary to form RNP complexes, transcribe and replicate the genome (E). Transcription and replication of the cloned RNA genome result in the production of infectious virus which can be assayed (F).
dependent RNA polymerase (Fuerst et al., 1986) or by using BSR-T7 cells that have been engineered to express T7 DNA dependent RNA polymerase (Buchholz et al., 1999). With the T7 polymerase present in the cells, expression plasmids for the N, P and L proteins (under the control of a T7 promoter) were transfected into cells to provide the helper proteins in the cytoplasm of the cell. In this way, the anti-genome RNA, and the N, P, L and M2-1 proteins are able to form a functional unit and the synthetic genome is replicated and transcribed. By using the T7 polymerase, these RNA transcribed in the cytoplasm as is seen in natural pneumovirus infection, rather than in the nucleus. When the entire virus genome is used, infectious virus is produced and can be assayed.

This system has the advantage of producing infectious virus that can be assayed in vitro and in vivo. The effect of mutations to the viral genome can be assessed in terms of viral fitness, RNA production, and tissue tropism. However, pneumovirus genomes are 13 to 16kb and cDNAs are technically difficult to create and manipulate. Hence historically before the creation of a full-length rescue system, a minigenome was used to determine the minimal requirements for viral replication and transcription. This minigenome was used to establish helper virus and plasmid rescue systems, as described below.

In the plasmid rescue system, a simpler synthetic ‘minigenome’ (figure 1.7) is used in place of the full-length genome clone. A minigenome has a reporter gene inserted between the viral replication and transcription signals and the reporter protein expression can be used as a measure of RNP complex function. This is termed a plasmid rescue system (figure 1.8) and has advantages and disadvantages compared to the full-length rescue system. In the plasmid rescue system no virus is produced, whereas full-length system produces infectious virus. However, the plasmid rescue system allows manipulation of not only the cis-acting non-translated genome sequences but also of the helper proteins.

The helper virus rescue system uses infectious virus to provide the trans-acting proteins to replicate and transcribe the minigenome in vitro (figure 1.9). The difference between the helper virus system and the plasmid rescue system is that in the former the helper
Figure 1.7 Schematic of the APV minigenome and dicistronic minigenome constructs. For the minigenome (A), the gene end (GE) signal (5’ UAGUUAUU 3’, mRNA sense), gene start (GS) signal (5’ GGGACAAGU 3’, mRNA sense), leader (Le) sequence and trailer (Tr) sequence were cloned flanking the CAT (chloramphenical acetyl transferase) reporter gene under the control of the T7 RNA polymerase promoter (T7) in the pt7.2 plasmid. The RNA fragment resulting from transcription by the T7 RNA polymerase and the hepatitis delta virus ribozyme sequence (δ) creates the exact viral Le sequence by self cleaving (indicated with the blue arrow). For the dicistronic minigenome (B), an addition reporter gene was added, a luciferase gene. The APV minigenomes were provided by Dr. J. Smith (Marriott et al., 2001).
Figure 1.8 Model of the negative sense RNA virus plasmid rescue system. The components required to create a reverse genetics system for a negative sense RNA viruses are show here. VVT7 (A) is an engineered vaccinia virus used to constitutively express T7 DNA dependent RNA polymerase (T7 RNA pol) in infected HEp2 (B) cells. These cells are then transfected with a minigenome plasmid (C) under the control of a T7 polymerase promoter (T7 Pro) that produces an RNA copy of the anti-minigenome upon transcription by the T7 RNA polymerase, which is then terminated using a hepatitis ribozyme. Simultaneously, the cells are transfected with expression plasmids for the helper proteins (D). For RSV, these are the N, P, L and M2-1 proteins. These cells then have all the components necessary to form RNP complexes, transcribe and replicate the genome (E). Transcription of the minigenome produces reporter mRNA which, after translated by the cell, produces reporter protein (F).
The components required to create a helper virus reverse genetics system for a negative sense RNA viruses are shown here. Recombinant BSR-T7 cells (A) are engineered to constitutively express T7 DNA dependent RNA polymerase (T7 RNA pol). These cells are then transfected with a minigenome plasmid (B) under the control of a T7 polymerase promoter (T7 Pro) that produces an RNA copy of the anti-minigenome upon transcription by the T7 RNA polymerase, which is then terminated using a hepatitis ribozyme. Simultaneously, the cells are infected with wildtype virus which provides the N, P, L and M2-1 helper proteins (C). These cells then have all the components necessary to form RNP complexes, transcribe and replicate the genome (D). Transcription of the minigenome produces reporter mRNA which is then translated by the cell to produce reporter protein (E).
proteins are provided by infection with ‘helper’ virus and in the latter the helper proteins are provided from transfected expression plasmids. Helper virus rescue and plasmid rescue have been used to establish the minimal cis-acting requirements in the viral genome to direct transcription and replication. Helper virus rescue uses live virus and can differ from the virus-free systems in its sensitivity to mutations (Fearns et al., 2000; Marriott et al., 2001). From a technical stand point, this technique requires the least amount of cloning before experiments can be attempted and hence historically has been the first system established for a virus.

The drive for the development of a reverse genetics system for negative sense RNA viruses is two-fold. The creation a virus entirely from cDNA is an excellent tool for primary research on virus function, as it allows direct manipulation of the genome using standard DNA cloning techniques and hence point mutations, insertions, deletions and rearrangements can be introduced and the effects assessed. The results of this process are 'rationally designed' viruses that are ideal for vaccination. They can be created to be multivalent, safe, effective, and unlikely to revert to virulence.

1.6 Aims

The aims of this project were to firstly develop the APV plasmid based reverse genetics system and then to use the APV reverse genetics systems to:

- investigate viral protein functions
- investigate intergenic region mutations
- quantify RNA levels of gene start sequence mutants
- create a eGFP tagged APV virus cloned entirely from cDNA in order to investigate viral tropism in vitro and in vivo.
Chapter 2: Materials and Methods
2.1 Solutions and buffers
All chemicals described were of AnalaR grade and supplied by Sigma-Aldrich Co. Ltd. or BDH unless otherwise stated.

Glyoxal running buffer pH 7: 10mM NaPO₄ buffer, made by combining 5mM NaH₂PO₄ and 5mM Na₂HPO₄ at a ratio (v:v) of 39:61 and adjusted to pH 7 using NaOH or H₃PO₄ as necessary.

Hybridization buffer: 50% formamide, 5X SSC, 5X Denhardt’s solution (Sambrook et al., 1989), 0.5% SDS, 50µg/ml torula yeast RNA, 0.1mg/ml boiled salmon sperm DNA.

Loading dye (6X): 0.25% bromophenol blue, 0.25% xylene cyanol FF and 30% (w/v) glycerol.

MOPS buffer (X10) pH 7: 0.2M MOPS (3-(N-Morpholino)propanesulfonic acid), 50mM anhydrous sodium acetate (NaOAc), 10mM ethylenediamine tetraacetic acid (EDTA).

Phenol/chloroform: A mixture of equal volumes of TE saturated phenol and chloroform.

Phosphate buffered saline (PBS), pH 7.4: 136mM NaCl, 2.7mM KCl, 1.3mM KH₂PO₄, 1.4mM Na₂HPO₄.

SSC (X20) pH 7: 3M NaCl, 0.3M sodium citrate (Na₃C₆H₅O₇).

SSPE buffer (X20) pH 7.4: 3M NaCl, 0.2M sodium dihydrogen phosphate (NaH₂PO₄), 20mM EDTA.

TAE (50X): 2M Tris, 0.05M EDTA (pH 8.0), 1M acetic acid.

TBE (10X) pH 8.1-8.5: 0.89M Tris, 0.89M Boric acid, 25mM EDTA.
TfbI ph 5.8: 30mM Potassium acetate, 100mM rubidium chloride, 10mM calcium chloride, 50mM manganese chloride, 15% v/v glycerol, pH adjusted with 0.2M filter sterilized acetic acid.

TfbII ph 6.5: 10mM MOPS, 75mM calcium chloride, 10mM rubidium chloride, 15% glycerol, pH adjusted with 1.0M filter sterilized potassium hydroxide.

Tris-EDTA (TE) solution: 10mM Tris (pH 7.5), 1mM EDTA.

Tris-HCL buffer: Tris (tris(hydroxymethyl)aminomethane), pH adjusted with 10M HCl.

Trypsin: 0.25% (w/v) in PBS pH 7.4 and stored at -20 C.

Versene: 0.002% (w/v) EDTA (disodium salt) and 0.002% (w/v) phenol red.

2.2. Media

GMEM-NEAA: Glasgow's minimal essential medium supplemented with non-essential amino acids and 0.2mM glutamine.

LB (Luria Bertani) medium: 10g/L Bacto tryptone, 5g/L yeast extract and 10g/L NaCl. For LB agar, 15g/L of bacto-agar was added. The medium was sterilized by autoclaving at 121°C for 20min.

MEMH: Eagle's minimal essential medium supplemented with 20 mM Hepes/NaOH, pH 7.2, lacking bicarbonate.

SOB (Super Optimal Broth) medium: 20g/L Bacto tryptone, 5g/L yeast extract, 0.5g/L NaCL, 2.5mM KCL (pH 7.0). The medium was sterilized by autoclaving at 121°C for 20min. The solution was cooled before the addition of 5ml/L of sterile 2M MgCl₂.

SOC medium: SOB medium supplemented with glucose (final concentration 20mM).
2.3 Mammalian tissue culture- maintenance of cell lines

2.3.1 Vero, BSC-1, and HEp-2 cells

BSC-1, Vero, and HEp2 cells were grown at 37°C in GMEM-NEAA supplemented with either 5% (BSC-1 and Vero cells) or 10% (HEp-2 cells) (v/v) foetal calf serum (FCS). Cells were grown at 37°C in an atmosphere of 5% (v/v) CO₂ and passaged when the monolayers reached confluence. The cell monolayer was separated from the surface of the tissue culture flask by washing briefly with versene solution, adding a mixture of versene and trypsin (1:1), and incubation at 37°C until the cells began to detach. The cell suspension was pipetted briefly and a proportion of the cells were resuspended in fresh medium, placed in a new flask and returned to an incubator at 37°C.

2.3.2 BSR-T7 cells

BSR-T7 cells are BSC-1 cells permanently transformed with a plasmid that constitutively expresses bacteriophage T7 DNA dependent RNA polymerase (Buchholz et al., 1999). The plasmid was maintained by antibiotic selection with antibiotic G-418 sulphate (Promega). The cell line was kindly provided by Professor K.-K. Conzelmann. Cells were maintained in GMEM-NEAA with10% FCS, 0.2mM glutamine, and 100 units/ml penicillin and 100µg/ml of streptomycin added. 1µg/ml of antibiotic G-418 sulphate was added every other cell passage. Cells were passed in the same way as described above.

2.4 Bacterial cultures

2.4.1 Growth of *E. coli*

Transformed *E. coli* (XL1-Blue or DH5α strain) was plated out onto LB agar plates containing 100µg/ml ampicillin and grown at 37°C to obtain single colonies. One colony was then picked and grown up at 37°C or 30°C for 8 to 16hr in 5 ml LB medium supplemented with 100µg/ml ampicillin. For larger cultures, 100 ml of LB with 100µg/ml ampicillin was inoculated with 100μl of a culture grown overnight.

2.4.2 Frozen competent bacteria

Frozen competent *E. coli* stocks were generated by adding 1ml of a 5ml culture grown from a single colony for 16 hr, to 100ml of warmed LB medium. Bacteria were grown to
an OD$_{600}$ of 0.39 in a 250ml conical flask at 37°C with shaking. Bacteria were then chilled on ice for 5min. Once chilled, the bacteria were pelleted at 2000rpm for 15min at 4°C and resuspended in 40ml TfbI buffer. After chilling on ice for 1hr, the bacteria were pelleted as before and resuspended in 4ml TfbII buffer. After chilling on ice for 2 to 5hr, 100µl aliquots were snap frozen in liquid nitrogen and stored at -70°C. All solutions were kept cold and chilled pipettes were used (Dr. P. Gould, personal communication).

2.5 Standard cloning procedures
2.5.1 Reverse transcription-PCR
Reverse transcription-PCR (RT-PCR) reactions were carried out using RevertAid™ H Minus M-MuLV Reverse Transcriptase (MBI Fermentas) following the manufacturer’s instructions.

2.5.2 PCR
Polymerase chain reactions (PCR) were carried following the individual manufacturer’s instructions. The three most commonly used polymerases were Pfu Turbo DNA Polymerase (Stratagene), Elongase Polymerase Mix (Life Technologies) and Pwo DNA Polymerase (Roche Applied Science).

A typical Pfu Turbo DNA Polymerase reaction was carried out as follows:

5µl 10X Pfu buffer
0.4µl dNTP mix (100mM)
DNA (0.01µg to 1ug)
10µl forward primer (10pmol/µl)
10µl reverse primer (10pmol/µl)
1µl Pfu Turbo DNA Polymerase
Sterile distilled water (SDW) to 50µl total volume

The solution was then heated for one cycle to 95°C for 2min; followed by 30 cycles of 30sec at 95°C, 30sec at 45°C and 1min at 72°C; and one cycle of 10min at 72°C.
A typical Elongase Polymerase Mix reaction was carried out as follows:

<table>
<thead>
<tr>
<th>Mixture A</th>
<th>Mixture B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1µl dNTP (10mM)</td>
<td>2µl buffer A</td>
</tr>
<tr>
<td>1µl forward primer (10pmol/µl)</td>
<td>8µl buffer B</td>
</tr>
<tr>
<td>1µl reverse primer (10pmol/µl)</td>
<td>1µl Elongase DNA Polymerase</td>
</tr>
<tr>
<td>DNA (0.01µg to 1µg)</td>
<td>19µl SDW</td>
</tr>
<tr>
<td>SDW to 20µl total volume</td>
<td></td>
</tr>
</tbody>
</table>

Mixtures A and B were combined on ice in pre-chilled PCR eppendorf tubes. The mix was then subjected to PCR consisting of one cycle of 30sec at 94°C; followed by 30 cycles of 30sec at 94°C, 30sec at 50 to 60°C and 1min at 68°C; and one cycle of 10min at 68°C.

A typical *Pwo* DNA Polymerase reaction was carried out as follows:

DNA (0.01µg to 1µg)
0.5µl *Pwo* DNA polymerase
2µl forward primer (10pmol/µl)
2µl reverse primer (10pmol/µl)
1µl 10X buffer + Mg
1µl dNTPs (10mM)
SDW to 50µl total volume

The mix was then subjected to PCR consisting of one cycle of 2min at 94°C; 30 cycles of 30sec at 94°C, 30sec at 60°C and 1min at 72°C; and one cycle of 7min at 72°C.

**2.5.3 DNA purification**

**2.5.3.1 Gel extraction**

Fragments of DNA were separated on agarose gels and excised using a clean razor blade. The DNA was extracted from these gel slices by the use of a QIAquick Gel Extraction.
Kit (Qiagen) according to the manufacturer’s instructions and the DNA was eluted into 30 to 50μl of SDW. Care was taken to expose the DNA to as little UV as possible, especially plasmids that were larger than 10kb.

2.5.3.2 Phenol/chloroform extraction
To purify DNA without gel extraction or the use of a kit, an equal volume of buffer-saturated phenol: chloroform (1:1) was added to the DNA solution. After mixing well (vortexed for 10sec, except for high molecular weight DNA which were gently rocked), the solution was centrifuged for 2 to 3min at 13,000rpm. The aqueous layer was then carefully removed to a new tube, being careful to avoid the interface. The above process was repeated until an interface is no longer visible. Where necessary, an equal volume of EB was added to the phenol/chloroform to remove any remaining DNA, followed by mixing, centrifuging and removal of the aqueous layer as before.

2.5.3.3 Ethanol precipitation
To purify DNA out of a solution, the volume of the DNA sample was measured and the salt concentration adjusted by adding 1/10 volume of 3M sodium acetate, pH 5.2. The solution was mixed well and 2 to 2.5 volumes of cold 100% ethanol (calculated after salt addition) added. The solution was mix well and placed on ice or at -20°C for ≥20 min. The solution was then centrifuged at 13,000rpm for 10 to 15min and the supernatant was carefully decanted. 1ml of 70% ethanol was added, the solution mixed, and centrifuged briefly. The supernatant was carefully decanted and air dried at 37°C for 10min. The pellet was then resuspended in an appropriate volume of SDW.

2.5.3.4 PCR product fragment purification
Purification of DNA which did not require gel extraction was also carried out using the QIAquick gel extraction kit PCR purification protocol and buffer PB (Qiagen).

2.5.4 Agarose gel electrophoresis
Purified plasmid DNA, the results of restriction enzyme digests, and PCR products were all routinely analyzed by agarose gel electrophoresis. The percentage agarose was
determined by the size of the DNA fragments to be differentiated, but usually ranged between 0.5 % (w/v) to 1.5 % (w/v) agarose. Agarose was melted in 1X TBE buffer to give the appropriate concentration and allowed to cool prior to the addition of a 10mg/ml solution of ethidium bromide to give a final concentration of 1µg/ml. Samples containing loading dye were subjected to electrophoresis at 180 to 100V until the DNA fragments being analyzed had migrated sufficiently to be separated. The DNA fragments were visualized on a UV transilluminator (UV products Inc.).

2.5.5 Urea acrylamide gel electrophoresis
RNA fragments were purified by urea gel electrophoresis. The gel was made from 2.5ml of 20% acrylamide in 7M urea/1X TBE solution, 9.6ml of 7M urea in 1X TBE, 26.2µl of 25% APS (Ammonium Persulfate, Sigma) and 11.8µl of TEMED (N,N,N,N - Tetramethyl-Ethylenediamine, Sigma). TBE buffer (X1) was used as the running buffer. Prior to loading on the gel, the RNA samples were prepared by adding an equal volume of loading buffer (RPA kit III, Ambion). The samples were incubated at 95ºC for 3min and loaded onto the gel. The gel was electrophoresed using a Bio-Rad tank for approximately 3hr at 50V. The RNA in the gel was visualized under UV light

2.5.6 Enzyme digestions
DNA was digested following the manufacturer’s instructions with restriction endonucleases (NEB, Fermentas, or Invitrogen). The typical volumes of DNA used when the concentration was unknown were 10µl of product from alkaline lysis small scale DNA preparations and 1µl for DNA from Qiagen Plasmid Mini Kits (Qiagen) and GenElute™ Plasmid Mini Kits (Sigma) unless otherwise stated.

2.5.7 Ligations
DNA was ligated using T4 DNA Ligase (Fermentas) using as close to a ratio (w:w) of 1:3 vector to insertion fragment DNA as possible, using the manufacturer’s instructions. Ligations were carried out for 12 to 16hr at 16ºC or for 48hr at 4ºC.
2.5.8 Transformation of bacteria
Transformation of competent *E. coli* with DNA was carried out as follows. 200µl aliquots of competent bacteria were thawed on ice and 100µl of the bacteria was incubated with 10µl of DNA mix for 30 min, on ice. The samples were transferred to 42°C for 30sec and then chilled on ice for 2 min. 900µl of LB medium was added and the samples were shaken for at least 45min at 37º C. 50 to 200µl of the suspension was then spread on LB agar plates supplemented with 100µg/ml ampicillin and incubated at 37ºC overnight.

2.5.9 Small scale DNA preparations
2.5.9.1 Mini-preparations of DNA
For plasmid DNA pure enough for sequencing reactions to be carried out, either QIAGEN Plasmid Mini Kits (Qiagen) or GenElute™ Plasmid Miniprep Kits (Sigma) were used. Aliquots (1 to 5ml) of transformed bacteria in LB medium containing the appropriate antibiotics were allowed to grow for 12 to 16hr at 37ºC or 30˚C and the plasmid DNA was then purified from it following the manufacturer’s instructions.

2.5.9.2 Alkaline lysis
To screen large numbers of transformed colonies, an alkaline lysis method was utilized as detailed in Sambrook et al. (1989) which is a modification of the methods of Birnboim (1979) and Ish-Horowicz (1981). 1.5 ml of culture was used, and a phenol chloroform purification step was included prior to ethanol precipitation of the DNA.

2.5.10 Large scale DNA preparations
For large scale preparations, 100 to 400ml of LB medium containing 100µg/ml ampicillin was inoculated with transformed *E. coli* and grown overnight at 37ºC. The DNA was purified using either QIAGEN Plasmid Maxi Kits (Qiagen) or GenElute™ HP Plasmid Kits (Sigma) following the manufacturer’s instructions. The DNA was then dissolved in 500ul to 3ml sterile distilled water. The concentration of these DNA preparations was then determined using a NanoDrop™ Spectrophotometer (Thermo Scientific). Measurements were taken in duplicate.
For low copy number vectors or large plasmids, one colony was picked from a plate and added to 2ml LB containing 100µg/ml ampicillin. The solution was shaken at 30°C for 24hr and then all of the mixture added to 400ml LB with containing 100µg/ml ampicillin. This was again left to shake at 30°C, for at least 24hr. The Sigma maxi prep was then used as per manufacturer’s instructions, except that the DNA was eluted with 3ml 70°C EB by centrifuging at 4000rpm for 5 min. The DNA was then further concentrated by EtOH precipitation (section 2.5.3.3) and dissolved in 52µl SDW

2.6 Introduction of site-specific mutations into plasmid DNA

2.6.1 Modified QuickChange PCR method

The standard method for introducing mutations into plasmid DNA was by mutagenic PCR. This was used to introduce insertions, deletions or point mutations into plasmid DNA and is a modification of the QuikChange® Site-Directed Mutagenesis Kit (Stratagen) protocol. PCR was carried out on 5µl template DNA (0.1µg/µl) combined with 1µl forward primer (5.5pmol/µl), 1µl reverse primer (5.5pmol/µl), 1µl dNTP (10mM), 5µl 10X PCR buffer, 1µl Pfu turbo (Stratagene) and 37µl SDW. The solution was heated in a PCR block for one cycle at 80°C for 2min, one cycle at 94°C for 30sec, 18 cycles of 94°C for 30sec followed by 55°C for 60sec and 68°C for 30 min. The PCR product was digested with either 2µl Dnp 1 (Stratagene) at 37°C for 8hr or 1µl at 37°C for 16 hr. The enzyme then was inactivated by heating the solution to 75°C for 15min. The resulting DNA was transformed into BL21-AI One Shot® Chemically Competent E. coli (Invitrogen) by adding 3µl of product to one vial of bacteria, incubating for 30min on ice, heating to 42°C for 30sec, and then cooling on ice for 2min. 250µl SOC media (Invitrogen) was added to each vial and the vials were incubated at 37°C for approximately 1hr while being shaken at 225rpm. An aliquot (150µl) of this media was spread onto an LB plate with the appropriate antibiotics and incubated at 30°C for 2 days (Clive Naylor, personal communication).

2.6.2 Quick screen method

Screening of the DNA from the resulting colonies was carried out using the following modified DNA mini preparation method (Dr. R. Ling, personal communication). Here,
4.5ml of a 5ml over night culture was centrifuged for 10min at approximately 3000rpm. The supernatant was removed and the pellet resuspended with vortexing and pipetting in 147µl of P1 solution (Qiagen) and transferred to a large eppendorf tube. To this, 147µl P2 solution (Qiagen) was added and the tube inverted 6 times to mix. 206µl N3 solution (Qiagen) was then added and the tube inverted 6 times to mix. After centrifuging the mixture for 10min at 13,000rpm, the supernatant was removed, added to a tube containing 1ml EtOH, and vortexed. This mixture was centrifuged for 12min at 13,000rpm and the supernatant discarded. To the remaining pellet of DNA, 0.5ml of 70% EtOH was added and mixed by vortexing. The tube was then centrifuged for 5min at 13,000rpm. The supernatant was discarded, the tube centrifuged briefly again and any remaining liquid removed. The pellet was then air dried at 37°C for 10min and resuspended in 20µl EB (Qiagen). The resulting DNA could be screened by restriction digestion in the presence of RNase. A typical restriction digest for screening of the full length APV subgroup A clones contains:

2µl DNA solution
0.25µl RNase (0.5mg/µl)
0.5µl MfeI restriction enzyme
0.5µl 10X REact buffer 4
1.75µl SDW

This solution was incubated, as described in the manufacturer’s instructions, at 37°C for 1 to 4 hr.

2.6.3 Pre-sequencing purification of DNA from quick screened colonies
Colonies which gave a correct digestion pattern were then prepared for sequencing as follows. 0.25ml of PB (Qiagen) was added to mini prep for screening preparation and was then passed through a QIAprep® Spin Miniprep Kit (Qiagen) column into a large eppendorf by centrifuging for one min at 13,000rpm. The column was then washed with 0.75ml PE solution (Qiagen). The wash was discarded and the column centrifuged again for one min at 13,000rpm. The column was then removed to a clean eppendorf, and the
DNA eluted from it with 50µl EB (Qiagen) preheated to 70°C. The DNA from this procedure is of good enough quality to be used in sequencing reactions (section 2.7)

2.7 Sequencing reactions with Big-Dye

Sequencing reactions for the Departmental Sequencing Service prepared as follows: 5.5pmol of primer was added to the recommended amount of DNA (see below) in a total volume of 10µl. The table below shows the recommended amounts of DNA, based on template type and size.

<table>
<thead>
<tr>
<th>Template</th>
<th>Amount DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR Products</td>
<td></td>
</tr>
<tr>
<td>100 to 200bp</td>
<td>2ng to 5ng</td>
</tr>
<tr>
<td>200 to 500bp</td>
<td>5ng to 17ng</td>
</tr>
<tr>
<td>500 to 1000bp</td>
<td>8ng to 33ng</td>
</tr>
<tr>
<td>1000 to 2000bp</td>
<td>17ng to 67ng</td>
</tr>
<tr>
<td>&gt;2000bp</td>
<td>33ng to 83ng</td>
</tr>
<tr>
<td>Plasmids</td>
<td></td>
</tr>
<tr>
<td>Up to 15kb</td>
<td>250 to 500ng</td>
</tr>
</tbody>
</table>

For large templates (over 15kb), at least 1µg of DNA was combined with 1µl of primer (5.5pmol/µl) and SDW was added to 6µl.

The mixture was then submitted to the Departmental Sequencing Service, where PCR was performed on a Hybaid omnigene PCR machine.

2.8 Rescue systems

2.8.1 Helper virus rescue system

The APV helper virus rescue system consisted of the transfection of DNA into BSR-T7 cells infected with helper virus. BSR-T7 cells were transfected when 50 to 60% confluent. The BSR-T7 cells were infected with APV at an MOI of 1 pfu/cell for 1hr at 37°C. Plasmid DNA was prepared by adding 1µg DNA to 90µl of Opti-MEM
(GibcoBRL, Life technologies) to which 3µl of Fugene (GibcoBRL, Life technologies, UK) had been added drop-wise, and the mixture incubated at room temperature for 5 to 15 min. A further 400µl of Opti-MEM was added to the lipid/DNA mix, and this was used to overlay BSR-T7 cells in a 12 well plate. 0.5ml of GMEM-NEAA with 5% FCS was added to each well. Cells were harvested 48hr later for CAT or luciferase assays.

2.8.2 Plasmid helper rescue system
The APV plasmid rescue system consisted of the transfection of DNA into HEp2 cells infected with recombinant T7 vaccinia virus. HEp2 cells in a 12 well plate were transfected when 50 to 60% confluent. The HEp2 cells were infected with recombinant T7 vaccinia virus (VFT7.3) at an MOI of 1 pfu/cell for 1hr at 37°C. Plasmid DNA was prepared by adding 1µg of DNA to 100µl of Opti-MEM. A second mixture of 8µl of Lipofectin (Invitrogen) added to 100µl of Opti-MEM was incubated at room temperature for 30 to 45min. The DNA was then mixed with the lipid and left at room temperature for 15min. A further 400µl of Opti-MEM was added to the lipid/DNA mix, and this was used to overlay the HEp2 cells after the VFT7.3 virus was removed and the cells washed with Opti-MEM. The plate was incubated at 37°C. 0.5 ml of GMEM-NEAA with 20% FCS was added to each well after approximately 24hr and cells were harvested 48hr later for CAT or luciferase protein.

2.8.3 Full length genome rescue system
BSR-T7 cells in 6 well (30mm diameter) plates were transfected when 90% confluent. The DNA mixture to be transfected consisted of: 0.3µg pCITE N gene plasmid, 0.3µg pCITE P gene plasmid, 0.24µg pCITE M2-1 gene plasmid, 0.15µg pCITE L gene plasmid and 3.0µg full length genome plasmid. All APV genes in pCITE vectors were provided by Dr. Roger Ling. Opti-MEM (Invitrogen) was added to the DNA mixture to a volume of 250µl/well. For each well, 10µl of Lipofectamine 2000 was added to 240µl of Opti-MEM and left at room temperature for 5min. The transfection reagent and Opti-MEM mixture was then added to the DNA and Opti-MEM mixture, vortexed to mix, and left at room temperature for 30min. Meanwhile, the media on the cells was removed and replaced with 1.5ml of GMEM-NEAA supplemented with 1% FCS per well. The DNA,
transfection reagent, Opti-MEM mixture was then added to the well in a drop wise manner and the transfection returned to 33°C. Each well was supplemented with 2ml/well of GMEM-NEAA with 1% FCS after approximately 24 hr. Cells were then incubated for 5 days or until 90% CPE is visible. Virus was then harvested as before, aliquoted, and stored at -70°C.

### 2.9 Preparation of cell lysates for CAT and luciferase assay

Cells from each well were scraped into the medium with a sterile scraper and the media containing the cells transferred to a sterile eppendorf tube. The cells were pelleted by centrifugation in a microcentrifuge for 1min at 13,000rpm. The supernatant was discarded and the cells were washed with 0.5ml of PBS, resuspended in 150µl of CAT lysis buffer (Roche) and incubated at room temperature for 30min. The nuclei were removed by centrifugation for 10min at 13,000 rpm and the supernatant either stored at -70°C or analyzed immediately.

### 2.10 CAT assay

The reporter CAT protein was assayed with an ELISA technique (Roche). An aliquot of the cell lysate (50ul) was mixed with 150µl of dilution buffer made from 10mg/ml of blocking agent (Roche) in PBS and added to a CAT microtitre plate (Roche) coated with anti-CAT antibody and incubated at 37°C for 1hr. The wells were washed five times with PBS containing 0.1% (v/v) Tween-20 before the addition of 200µl of anti-CAT-digoxigen (DIG) conjugated antibody diluted to 2µg/ml in dilution buffer, to each well. The reaction was and incubated at 37°C for 1hr. The wells were washed as above and 200µl of anti-DIG-POD (digioxigenin-peroxidase) antibody, diluted to 150units/ml in dilution buffer, was added to each well followed by incubation at 37°C for 1hr. The wells were washed again and 200µl of ABTS (2,2'-azino-di-[3-ethylbenzthiazoline sulfonate (6)] diammonium salt, Roche) substrate added. The plate was left to develop color for approximately 15min on an ELISA plate reader (Labsystems MultiscanRC, Lifesciences international Ltd.) at a wavelength of 405nm.
2.11 Luciferase assay

Cell lysates generated by the method described in section 2.9, were assayed for the presence of luciferase by an enzyme reaction that resulted in the production of light. 10µl of cell lysate was mixed with 50µl of luciferase reagent which was provided in a kit to assay for luciferase enzyme activity (Promega) and measured immediately in a Labsystems luminoskan luminometer and expressed as relative light units (RLU).

2.12 RNA Quantification

2.12.1 Northern blot

Gel Electrophoresis:
RNA was separated on either a formaldehyde or glyoxal gel. Formaldehyde gels were made of 1X MOPS (v/v) buffer, 1% (w/v) agarose, 1/56 (v/v) formaldehyde and SDW to 100ml. Agarose was dissolved in 1X MOPS buffer, cooled slightly and the formaldehyde added immediately prior to pouring. The running buffer consisted of 1X MOPS buffer, 1/56 formaldehyde and SDW to 800ml. Prior to loading on the formaldehyde gel, samples were heated to 65°C for 15min as: 25% of the mRNA from a well from a 6 well plate (25µl), 7.5µl formaldehyde, 22.5µl formamide, 6µl 10X MOPS buffer, and 0.3µl ethidium bromide (EtBr, 10mg/ml). The gel was subjected to 80 to 100V for approximately 1.5hr. A ladder lane was loaded with 2µl of size ladder (Millennium ladder, Ambion), 5µl loading buffer (Promega) and 18µl SDW. After electrophoresis, the gel was washed for 10min in SDW and then for 10min in 10X SSC.

Glyoxal gels were made of 1% (w/v) agarose and SDW and were submersed in glyoxal running buffer for electrophoresis at 3 to 4 volts per cm between the electrodes, with the buffer circulated using a pump. Samples were made up of 5µl of 200ng/µl RNA stock plus 5µl glyoxal gel load dye (Ambion). A size ladder lane was loaded with 0.1µl of size ladder (Millennium ladder, Ambion), 5µl loading buffer (Promega) and 4.9µl SDW. Prior to loading on the gel, samples were heated to 50°C for 30 to 60min and then chilled on ice. The gel was subjected to 80 to 100V for approximately 1.5hr.
**Turbo Blot**

A TurboBlotter™ system (Schleicher & Schuell, Inc., Keene, NH) apparatus was used following the manufacturer’s instruction for capillary transfer of the RNA from the gel to a neutral nylon membrane (Hybond-N membrane, Amersham Biosciences), using 20X SSC as transfer buffer for 2 to 3 hr.

**Pre-hybridization and hybridization**

The blot was gently dried and baked at 80°C for 30 to 120 min. The blot was then rinsed in 2X SSC or washed in 20mM TrisHCl pH 8 at 68°C for 15 min and transferred carefully to a small hybridization roller bottle. After at least 30 min of pre-hybridization in 5ml of hybridization buffer at 68°C, the denatured (by boiling) probe was added. Hybridization was allowed to proceed overnight.

**Blot processing and development**

The blot was then rinsed in 2X SSPE with 0.1% SDS at 68°C for 15 min twice, followed by two washes in 0.1X SSPE with 0.1% SDS at 68°C for 15 min each. The detection protocol was then carried out using CSPD (disodium 3-(4-methoxyspiro \{1,2-dioxetane-3,2’-(5’-chloro)tricyclo[3.3.1.13,7]decan\}4-yl) phenyl phosphate, Roche) and the Wash and Block Buffer Set (Roche) following the manufacturer’s instructions, except the membrane was blocked in blocking buffer for at least 60 min and incubated with the CSPD for 15 min at 37°C. Exposures were carried out using professional paper, black-and-white (Kodak) and developed using either a Curix 60 (AGFA) machine or a Universal Hood (Bio-Rad). Using the Bio-Rad machine and QuantityOne software (version 4.6.3, Bio-Rad), digital exposures were taken every 10 min for 2 hr.

**2.13 Growth and harvesting of virus stocks**

**2.13.1 Growth of non-recombinant virus stocks**

APV subgroup A virus stocks were generated by infecting sub-confluent monolayers of Vero cells in roller bottles at a low multiplicity of infection (0.5 pfu/cell) in 10ml GMEM-NEAA with 2% (v/v) FCS, and incubation at 33°C. Virus was allowed to absorb for 1 hr before 10ml of GMEM-NEAA with 1% FCS and CO₂ were added. Bottles were
incubated at 33°C until 90% CPE was observed (usually 11 to 14 days post infection). Virus was harvested using glass beads, aliquoted, and stored at -70°C.

2.13.2 Growth of recombinant virus stocks
APV recombinant virus stocks were generated by transfection of BSR-T7 cells in 6-well plates using Lipofectamine 2000 (Invitrogen, following manufacturer’s instructions) with the helper plasmids pCITE N (0.3µg/well), pCITE P (0.3µg/well), pCITE M2-1 (0.24µg/well), and pCITE L (0.15µg/well), and a cDNA genome plasmid (3µg/well). After 4 to 5 days, the virus was harvested and the entire product was overlaid onto a confluent large flask of BSC-1 cells. The cells and virus were incubated at 33°C for 1hr before 10ml of GMEM-NEAA with 1% FCS was added. The cells and virus were then left at 33°C for approximately 4 days if the virus was expected to form syncytia and 6 to 12 days if it was not expected to form syncytia. Once extensive CPE was present, the virus was harvested using glass beads and used to infect confluent roller bottles of BSC-1 cells (about 9.5X10^8 cells) at approximately 30% of a large cell culture flask (about 5.4X10^7 cells per flask) per roller bottle. The virus was allowed to absorb at 33°C for 1hr before an additional 10ml of GMEM-NEAA supplemented with 1% FCS was added to the bottles. Four and six days p.i. glass beads were used to harvest virus CF2EdSH 56 and virus CF2SH 11. The supernatant was cleared of cellular debris by centrifuging the media/virus/cell debris mixture for 30min at 4,000g, removing all but 250µl of the supernatant to a fresh universal, adding 5ml of PBS, resuspending the cell pellet by vortexing, centrifuging the mixture for 30min at 4,000g, and adding the supernatant to the supernatant from the first centrifugation step. Aliquots were the stored at -70°C.

2.14 Determination of virus titre by GFP expression assay
For recombinant viruses which expressed GFP, a modified protocol for titre determination using fluorescent cell-counting was carried out (Philipson, 1961). BSC-1 cells were grown to 90% confluency in a 96-well plate. Virus stocks were serially diluted 1:10 GMEM-NEAA medium supplemented with 1% FCS with and 100µl of diluted solution was overlaid onto cells for 1hr at 33°C. The virus solution was then removed and replaced with 200µl/well of GMEM-NEAA medium supplemented with 1% FCS.
Plates were incubated at 33°C for 48 hr. Cells expressing GFP were then counted and the number of cell infecting units (CIF) per ml determined.

2.15 Nucleotide and protein sequence alignments
Alignments were carried out using the MegAlign program (DNASTAR Ltd.) using the alignment function ClustalV, derived from the program ClustalX (Thompson et al., 1997). Using GeneDoc (Nicholas et al., 1997), the Blosum 62 scoring table was used to determine the percentage homology or identity. All gene homology calculations were carried out using the sequence of the ORF except as indicated. Unless otherwise stated, the nucleotide and protein sequences used were hMPV isolate 00-1 (van den Hoogen et al., 2001)(Genbank accession number AF371337), PVM strain 15 (Thorpe and Easton, 2005)(Genbank accession number AY743910), RSV strain RSS-2 (Tolley et al., 1996)(Genbank accession number U39662), APV subgroup A, strain CVL14/1 (Li et al., 1996; Ling et al., 1995; Ling and Pringle, 1988; Randhawa et al., 1996b; Yu et al., 1992)(Genbank accession numbers U22110, X63408, U65312 and U39295), APV subgroup B (Randhawa et al., 1996a)(Genbank accession numbers U37586, AM293283, and AM411375), APV subgroup C (Govindarajan and Samal, 2005)(Genbank accession number AY590688), APV subgroup D (Bayon-Auboyer et al., 2000)(Genbank accession number AJ288946). It should be noted that when entire genomes were compared, no change in the order of the genes was carried out. This meant that analogous genes were not compared, as the gene order of the pneumoviruses differs from that of the metapneumoviruses (figure 1.1).

2.16 mRNA and protein simultaneous harvest
To extract protein and mRNA, tissue culture cells were rinsed once with PBS and harvested. The cells were centrifuged for 1min 13,000g, resuspended in PBS, and then divided into two equal aliquots- one to be used for protein extraction and the other for mRNA extraction. The mRNA extraction was carried using the Sigma direct mRNA isolation kit (using the protocol for pelleted cells) following the manufacturer’s instructions. Protein extraction was carried out as described in section 2.9.
2.17 RNA probes

RNA probes for northern blots were created as described in ‘The DIG System User's Guide for Filter Hybridization” (Boehringer Mannheim/Roche). Alternatively, RNA probes were created using 0.2 to 1µg of linearized plasmid DNA as follows. The DNA was combined with 2µl of DIG-NTP mix (Roche), 4µl of 5X polymerase buffer (Fermentas), and 10 to 50units T7 or T3 RNA polymerase (Fermentas). This mixture was incubated at 37°C for 2 hr. Then 1µl of DNase (Ambion, mixing gently as per protocol) was added and the mixture incubated for approximately 15min at room temperature. 50µl SDW was added and the RNA purified using a Chromaspin 200 Kit (Clontech), including a pre-spin of 700g for 5min. The RNA was eluted from the column, divided into 3 aliquots and stored at -70°C until required.

2.18 RNA standards

RNA standards were created for northern blot as follows (Dr. R. Ling, personal communication). DNA with the sequence of interest in the correct orientation was linearized in RNase-free conditions as follows. Proteinase K digestion was then carried out by combining 50µl of the digested template with 2.9µl of 10X SDS, 0.58µl of 0.5M EDTA, and 5µl of Proteinase K (1mg/ml, Sigma). The solution was incubated at 55°C for approximately 2 hr. The DNA was then purified by phenol/chloroform extraction and EtOH precipitation using RNase-free conditions. The DNA was then examined by electrophoresis and the concentration determined using the NanoDrop™ Spectrophotometer.

RNA fragments were generated from the DNA using a MEGAScript Kit (Ambion) following the manufacturer’s instructions. The RNA was then separated by electrophoresis on a urea acrylamide gel (section 2.5.5). Once separated on the gel, the RNA fragments were visualized under UV light using UV shadowing sheets (Ambion). The RNA was extracted from the gel by wrapping the gel in cling film, exposing the gel to UV light and marking the position of the RNA fragments on the cling film. The RNA fragments were then cut out using a scalpel and put in 1.5ml eppendorf tubes. 350µl of pre-warmed Elution solution (RPA III kit, Ambion) was added to the slices and incubated.
at 37°C overnight. The RNA was isolated by removing the liquid to a new tube and subjecting the solution to phenol/chloroform extraction followed by EtOH precipitation of the RNA. The RNA pellet was then resuspended in 11μl RNase-free H₂O. The RNA was measured as above with the NanoDrop™ Spectrophotometer.

2.19 Imaging methods
GFP expression in cells was captured using a Leica SP2 confocal microscope at 40X magnification. The microscope was connected to a computer that could process the images using Leica software. In this way, it was possible to set the excitation wavelength specifically for GFP (488nm) and to capture the resulting emitted light in the appropriate range and produce an image digitally. For sliced tissues, a dye-separation protocol was carried out due to the high levels of background florescence seen. By setting the negative control tissues as the background and removing this light from the captured image of GFP expressing tissues, the cells expressing GFP could be clearly seen. Protocols for image capture and background substitution were provided by Leica.
Chapter 3:
Development and optimization of the APV plasmid rescue system
3.1 Introduction
The sub-family *Pneumovirinae* consists of the pneumoviruses avian pneumovirus (APV), respiratory syncytial virus (RSV), pneumonia virus of mice (PVM), and human metapneumovirus (hMPV). By analogy to RSV, the three proteins likely to be required for replication and transcription of the APV genome are the nucleocapsid (N) protein, the phosphoprotein (P) and the large polymerase subunit (L) protein. In addition, for RSV, the presence of the M2-1 protein has been found to increase viral transcription (Fears and Collins, 1999). APV, like all viruses with a non-segmented negative-sense RNA genome, cannot produce infectious virus from the genome alone. In pneumoviruses, the minimal unit of infectivity is the RNP complex made up of the negative sense RNA genome encapsidated by N protein and associated with the L and P proteins. Reverse genetics systems that reconstitute this RNP complex using cloned components have been used to determine the role of the viral proteins, the functional regions of the proteins, and the function of conserved non-translated regions of the genome (section 1.3 and 1.4).

An APV reverse genetics system was developed by creating an APV sub-genomic mini-replicon plasmid expressing either CAT or CAT and LUC protein under the control of authentic APV replication and transcription signals (Marriott et al., 2001; Smith, 2001) (section 1.4). These ‘minigenomes’ were used to investigate the sequence requirements for transcription and replication of the minigenome in the presence of APV. This rescue system was useful for investigating the role of specific non-translated regions of the genome, but it could not be used to look at the effect of mutations in APV genes. The development of a plasmid-based rescue system (PRS) (section 1.5) would allow investigation of the role of the N, P and L proteins. The pneumoviruses also express a unique M2 mRNA, which contains two ORFs (M2-1 and M2-2, figure 3.1) for which there are no homologous or analogous genes in other Paramyxoviruses.

In the case of RSV, the M2-1 protein expressed in a PRS has been shown to greatly enhance mRNA production by acting as an elongation and anti-termination factor (Collins et al., 1999a; Cuesta et al., 2000; Fears and Collins, 1999; Hardy and Wertz, 1998; Teng and Collins, 1998), while the M2-2 protein inhibited viral RNA
replication and transcription (Collins et al., 1999a; Collins et al., 1996; Hardy and Wertz, 1998; Teng and Collins, 1998). In the case of PVM, the M2-1 protein has been shown to greatly (more than 10-fold) to enhance reporter protein expression when present at low levels and to inhibit RNA synthesis when present at higher levels in the PRS (Dibben,
HMPV recombinant virus lacking the M2-1 or M2-2 ORFs grew to good levels \textit{in vitro}, but could not replicate \textit{in vivo} without the M2-1 ORF present. Virus lacking the M2-2 ORF grew to very low levels \textit{in vivo} but protected hamsters from challenge with hMPV, suggesting it would make a good vaccine candidate (Buchholz et al., 2005). APV lacking the M2-2 ORF replicated normally \textit{in vitro} but showed impaired replication and a reduced antibody response in turkeys (Ling et al., 2008).

The aim of this section of the project was to develop a functional APV subgroup A (isolate CVL14/1) PRS, to optimize the PRS, and to use the PRS to investigate the role of the M2-1 and M2-2 proteins on protein production from minigenome constructs.

3.2 Development of the plasmid rescue system for APV

It was hypothesized, based on the established RSV PRS (Yu \textit{et al.}, 1995), that the APV RNA-dependent RNA polymerase (consisting of the L, N and P proteins) could be reconstituted in cell culture from cDNA plasmids and would recognize a synthetic APV genome or minigenome (also referred to in the literature as a genome analog). Activity of the polymerase could be measured by assaying for a reporter gene expressed in place of the APV genes from the minigenome. Here, two APV minigenomes were used: a CAT protein expressing minigenome (pCAT minigenome) and a CAT and LUC protein expressing dicistronic minigenome (pCATLUC). The levels of protein expressed from these minigenomes were used to measure the functionality of the N, P, and L expressing helper plasmids.

The pCAT minigenome and pCATLUC have been described (Marriott \textit{et al.}, 2001; Randhawa \textit{et al.}, 1997; Smith, 2001). The L helper plasmid (pCITE L) was provided by Dr. R. Ling and the M2-1 and M2 plasmids were provided by Dr. H. Stokes. It was therefore necessary to clone the APV N, P and M2-2 ORFs into a pTM1 or pT7.2 vectors in order to have a full set of the proposed plasmid rescue system components. These protein expression vectors have a multi-cloning site flanked by a T7 polymerase promoter/IRES sequence and a T7 terminator sequence.
3.2.1 Cloning of the APV N gene helper plasmid

RT-PCR was used to clone the APV N gene from RNA into the pT7.2 vector (Novagen) as follows. Cellular RNA was extracted from APV subgroup A (isolate CVL-14/1) infected cells (courtesy of Prof. A. J. Easton), and a reverse transcriptase reaction was carried out using approximately 0.2µg of RNA, primers APVNTOPOF and APVNTOPOR (appendix A), and RevertAid H minus M-MuLV reverse transcriptase (Section 2.5.1). This was followed by performing PCRs using Pfu polymerase with primers APVNTOPOF and APVNTOPOR (figure 3.2) and confirmed to produce a DNA fragment of the expected size by gel electrophoresis (sections 2.5.2 and 2.5.4).

![Figure 3.2 Cloning of the N10 plasmid](image)

**Figure 3.2 Cloning of the N10 plasmid.** Following RT-PCR, an approximately 1170bp N gene DNA fragment was visualized on an agarose gel (A). Ligation of this fragment into the pT7.2 vector resulted in three clones: N10 and N11 gave the expected 3592bp and 1178bp fragment restriction digestion pattern whereas N12 lacked the 1178bp fragment that denotes N gene insertion (B).

The product of this reaction was then purified (section 2.5.3) and digested with restriction enzymes SalI and NcoI. The produce was ligated into a pT7.2 vector that had been similarly digested and then treated with shrimp alkaline phosphatase (Fermentas Life Sciences) to prevent re-ligation of the ends of the DNA as per the manufactures.
instructions. Following transformation (section 2.5.8) of the ligation product into DH5α E.coli competent cells and small scale plasmid purification (section 2.5.9.1), the DNA from three colonies (N10-N12) was digested with the restriction enzymes NcoI and SalI and separated by gel electrophoresis as before to check for ligation of the N gene into the pT7.2 vector (figure 3.2). Samples N10 and N11 were found to produce fragments of the expected sizes (3592 and 1178bp) and were sequenced (section 2.7). Plasmid pN10 had the expected sequence and the DNA subjected to large scale plasmid purification.

### 3.2.2 Cloning of the APV P gene helper plasmid

The APV P gene was cloned into the pPTM 1 expression vector (Moss et al., 1990). PCR was carried out on an APV P gene clone (pBSII SK+ P plasmid, provided by Dr. Joanne Smith) using Pwo DNA polymerase (example in section 2.5.2) with primers TRTP4 and TRTP2 (appendix A). A fragment of the desired length (approximately 850bp) was detected following gel electrophoresis (figure 3.3).

**Figure 3.3 Cloning of P9 plasmid.** Visualization of P gene PCR product of approximately 850bp (A) and results of restriction digestion of plasmid P1, P3, P6, and 9 with NcoI and BamHI by agarose gel electrophoresis. Here, the P1, P3 and P9 clones give the expected digestion pattern of 836bp and 5.4kb fragments (B).

The PCR product was purified by phenol/chloroform extraction followed by ethanol precipitation and then digested with the restriction enzymes NcoI and BamHI followed by
fragment separation in an agarose gel. The approximately 850bp fragment was excised from the agarose gel and ligated into a PTM1 vector which had been digested previously with NcoI and BamHI enzymes. The ligation reaction was transformed into XL-1 Blue competent bacteria and nine colonies were subjected to small-scale plasmid purification as above. The resulting samples were digested with the restriction enzymes NcoI and BamHI and examined by gel electrophoresis. The clones with the correct insert produced two fragments of 5.4kb and 836bp. As can be seen in figure 3.2, samples 1, 3, 6 and 9 had the correct insertion. E. coli transformed with samples 1 and 9 was then subjected to large-scale plasmid purification as above. The clones were sequenced as above and clone P9 was shown to contain the correct sequence.

3.2.3 Cloning of the APV M2-2 ORF helper plasmid
The APV M2-2 ORF was cloned from the existing APV subgroup A (isolate CVL-14/1) M2 plasmid. A PCR reaction was carried out on this plasmid with Elongase DNA polymerase (section 2.5.2) and primers APVAM2-2FOR and APVAM2-2REV (appendix A). The M2-2 fragment was visualized on an agarose gel (figure 3.4), extracted, and was digested with restriction enzymes NcoI and EcoRI. The fragment was then purified and ligated into the pT7.2 vector which had been subjected to restriction digestion as above. Colonies were screened for correct insertion (section 2.6.2) using restriction enzymes NcoI and EcoRI and colonies 16, 17, 22, 24 and 33 were found to have the correct digestion pattern of fragments of 231bp and 3651bp (figure 3.5) on an agarose gel. The DNA from these colonies was then sequenced. Preparations 16-18, 22, 24 and 33 were found to have gained a base in the non-coding region of the vector at base 3424 and preparation 16 to have an additional change of a G→A at base 3557. Preparation M2-2 24 was arbitrarily chosen out of the preparations with a base insertion (which should have no effect on protein production) to be use in all future experiments.

3.3 Establishment and optimization of the APV plasmid rescue system
Initial attempts to establish an APV PRS were carried out using the APV pCAT or pCATLUC minigenomes and APV helper and M2-1 plasmids described above based on the existing RSV PRS (Marriott et al., 1999). Plasmid rescue system attempts were
carried out by transfecting sub-confluent monolayers of HEp2 cells (approximately 10^6 cells/well) which had been infected with vTF7-3 recombinant vaccinia virus expressing T7 polymerase (Fuerst et al., 1986) at a multiplicity of infection (MOI) of 1 pfu/cell. The DNA for all replicate samples was combined with an appropriate amount of Lipofectin (Invitrogen) and Opti-MEM medium, measured into aliquots and transfected into HEp2 cells as described in section 2.8.2. After 48hrs, the cells were harvested (section 2.9), pelleted, and lysed using 150ul of CAT lysis buffer to isolate the reporter proteins. The cell lysates were then assayed (sections 2.10 and 2.11) and the total amount of CAT protein per well calculated. Each sample was repeated 2-3 times within each assay and the whole procedure was repeated on at least two separate occasions. Means were calculated for occasion the whole experiment was performed and these were used as the independent replicates for calculation of the standard error.

Figure 3.4 Cloning of pM2-2: PCR for M2-2 fragment and restriction digest of M2-2 fragment and vector. PCR was carried out on the pM2 using primers M2-2 FOR and APV A M2-2 REV pT7.2 and the approximately 250bp fragment was visualized on an agarose gel (A). The M2-2 PCR fragment and Pt7.2 vector were subjected to restriction digestion and visualized: a 3650 and 6bp fragments for Pt7.2 and an approximately 220bp fragment for M2-2.
Initial testing resulted in a PRS able to produce levels of CAT protein at least 100-fold above background using 0.2\(\mu\)g per well N10 plasmid, 0.1\(\mu\)g per well P9 plasmid, 0.1\(\mu\)g per well pCITE L plasmid, 0.01\(\mu\)g per well M2-1 plasmid, 0.2\(\mu\)g per well pCAT

Figure 3.5 Screen of M2-2 plasmid clones for correct inserts. DNA isolated from M2-2 colonies 7-40 (1-6 data not shown, all without inserts) was subjected to restriction digestion with EcoRI and Ncol. Colonies with the M2-2 ORF inserted would be expected to show fragments of 3650 and 231bp (circled) whereas the vector alone would be 3650bp. Colonies 16 (A), 17, 22, 24 (B), and 33 (C) gave the expected pattern and were fully sequenced.
minigenome or pCATLUC plasmid and enough pT7.2 vector plasmid to make the total DNA amount up to 1µg per well. Optimization of the plasmid rescue system was carried out by varying each of the helper plasmids in turn, while keeping the rest of the plasmids at the concentrations listed above.

3.3.1 Optimization of the amount of N10, P9, and pCITE L plasmid
Varying the amounts of N10 helper plasmid from 0.01 to 0.4µg per well resulted in a range of activity from 572 to 4392pg per well at 0.4 and 0.05µg per well of N10 plasmid, respectively. The negative controls of 0µg per well of N10 plasmid and APV (-) (no L plasmid included) gave 39 and 4pg per well, respectively, of CAT protein. Cat protein expression at 0.05 to 0.2µg per well of plasmid was over 1000-fold above the 0µg N10 plasmid control. Further increase in the amount of N10 plasmid, 0.3 and 0.4µg per well, resulted in a decrease in CAT protein production by 62 and 86%, respectively (figure 3.6). This indicated that the use of 0.05 to 0.2µg of plasmid per well should be used for maximum reporter protein expression.

The P9 helper plasmid was added from 0.05 to 0.4µg per well, resulting in a range of activity from 1444 to 4969pg per well at 0.1 and 0.2µg per well of P9 plasmid, respectively. The negative controls of 0µg per well of P9 plasmid and APV (-) gave 7 and 13pg per well, respectively, of CAT protein. A higher level of plasmid, 0.4µg per well, resulted in a decrease in CAT protein production by 41% (figure 3.7). Hence, the use of 0.1 or 0.2µg of plasmid per well should be used for maximum reporter protein expression.

The effect of variation of the pCITE L helper plasmid from 0.0001µg per well to 0.4µg per well was similar to that seen for the N and P plasmids. Variation resulted in a range of activity from 1908 to 7441pg per well at 0.01 and 0.2µg per well of pCITE L plasmid, respectively. The negative control of 0µg per well of pCITE L plasmid gave 26pg per well of CAT protein. Cat protein expression at 0.05 to 0.4µg per well of plasmid was over 2000-fold above the 0µg pCITE L plasmid control (figure 3.8). This indicated
Figure 3.6 Effect of APV N plasmid concentration on CAT protein expression in the APV A PRS. The amount of N plasmid added to the transfection was varied from 0 to 0.4 ug/well. The results are the average of at least three independent experiments (except for 0.5 ug/well which was determined on two days) and the bars show the standard error of the means.

Figure 3.7 Effect of P plasmid variation on CAT protein expression in the APV PRS. The amount of P plasmid added to the transfection was varied from 0 to 0.4 µg/well. The results are the average of at least three independent experiments and the bars show the standard error of the means.
that the use of 0.05 to 0.4µg of plasmid per well should be used for maximum reporter protein expression.

The development of a PRS produced a set of plasmids that could be used in either the PRS with a minigenome or in full-length genome rescue system with a full-length cloned genome. All subsequent transfections were performed with 0.1 to 0.2µg per well of N10 plasmid, 0.1µg per well of P9 plasmid and 0.1µg per well of pCITE L plasmid.

3.3.2 Effect of varying the amount of pM2-1 and pM2-2 plasmid on reporter gene expression

As discussed in section 3.1, the APV M2 gene contains two ORFs (M2-1 and M2-2) that are coupled via an overlapping region. Here, the M2-1 and M2-2 ORF were investigated separately and together for their effect on production of reporter protein from the pCATLUC dicistronic minigenome (section 1.5). Having a second reporter gene in the

Figure 3.8 Effect of APV L plasmid concentration on CAT protein expression in the APV A PRS. The amount of L plasmid added to the transfection was varied from 0 to 0.4µg/well. The results are the average of at least three independent experiments and the bars show the standard error of the means.
minigenome provides the opportunity to investigate the effect of the M2 proteins on sequential genes expression.

Variation of the M2-1 plasmid from 0µg per well to 0.4µg per well increased the amount of CAT protein production from 3224pg per well to a maximum of 15222pg per well at 0.025µg per well of M2-1 plasmid. There was a plateau of CAT production from 0.005µg per well to 0.1µg per well of M2-1 plasmid. M2-1 plasmid amounts greater than 0.1µg per well decreased CAT protein production by approximately 26% (figure 3.9).

For APV, the effect of the addition of 0 to 100ng of M2-2 plasmid to the PRS, with and without 10ng of M2-1 plasmid, was a gradual increase in inhibition of CAT and LUC protein production. At levels of 1ng/well M2-2 plasmid, CAT protein was still detected at more than 1000pg per well. However, the addition of M2-1 plasmid to the transfection mixture increased the CAT protein expression 2-3 fold. Interestingly, when more than 1ng of M2-2 plasmid was added, CAT expression decreased to almost zero, even with an equal amount of M2-1 plasmid present. For LUC protein, the protein expression pattern was identical to that seen for CAT protein in the presence of M2-2 plasmid with and without M2-1 plasmid present (figure 3.10).
Figure 3.9 Effect of APV M2-1 plasmid concentration on CAT and LUC protein expression in the APV A PRS. The amount of M2-1 plasmid added to the transfection was varied from 0 to 0.4µg/well. The results are the average of at least three independent experiments (except for APV (-) which was determined on one day) and the bars show the standard error of the means.
3.4 Discussion

3.4.1 The APV plasmid rescue system

This work describes the development and optimization of a robust APV PRS. The effect of varying the N, P, L and M2-1 plasmids was measured and it was shown that there was a range of acceptable concentrations for all four plasmids. This suggested that the N, P, L and M2-1 proteins were not required in exact ratios, but that there was a minimum and a maximum amount of each protein (relative to a fixed amount of minigenome plasmid) required for the PRS to function. For the N, P and L helper plasmids, a minimum of 0.1µg of plasmid out of a total reaction containing 1µg input DNA was ideal for

Figure 3.10 Effect of APV M2-2 plasmid concentration with and without M2-1 plasmid present on CAT protein expression in the APV A PRS. The amount of M2-2 plasmid added to the transfection was varied from 0 to 100ng/well with no M2-1 plasmid added (blue) and with 0.01µg/well of M2-1 plasmid added (red). The results are the average of at least two independent experiments and the bars show the standard error of the means.
maximum efficiency and more than 0.4µg per well of any of the helper plasmids appeared to be detrimental to reporter protein expression. For the M2-1 plasmid, any concentration above 5ng/well enhanced reporter protein expression. Similarly, the maximum PVM PRS reporter protein production was achieved using a range of ratio of reporter plasmids (Dibben, 2006; Dibben and Easton, 2007). For RSV, it was found that maximum production of genomic RNA was achieved with N, P and L genes transfected at a 12:5:1 molar ratio (Yu et al., 1995).

3.4.2 M2-1 and M2-2 in the APV plasmid rescue system

With the establishment of the APV PRS, it was possible to investigate the role of the M2-1 and M2-2 proteins in viral transcription. The PRS was the appropriate rescue system with which to investigating the minimal requirements for production of a fully functional RNP complex as Collins et al. (1999a) showed that M2-1 protein could be produced from the anti-genomic plasmid or helper virus in the full-length or helper virus rescue systems, respectively, by promiscuous transcription by the vaccinia virus polymerase. It was shown that M2-1 protein could be produced in sufficient amounts to have an enhancement effect on transcription in the absence of any M2-1 expression plasmid being present (Collins et al., 1999a).

For RSV, the components of the RNP complex were found to be the N, P, L proteins and viral (-) RNA, with the M2-1 protein acting as an anti-termination and elongation factor necessary for transcription of more than one gene from a minigenome or full length cloned genome (Collins et al., 1999a; Collins et al., 1996; Fears and Collins, 1999; Hardy and Wertz, 1998). The RSV M2-1 protein was shown to enhance synthesis of full-length mRNA of long (1,212 or 1,780nt) genes (up to a 615-fold increase) more than a short (274nt) gene (less than a 2-fold increase) (Fears and Collins, 1999). The N, P, and L proteins along with RNA genome were also found to be the minimal requirements for protein production from a minigenome in the PVM PRS (Dibben and Easton, 2007). A PRS for APV was developed here whereby the N, P and L proteins were able to produce reporter protein from an APV minigenome in the absence of any source of M2-1 protein. The APV M2-1 protein was found to enhance reporter protein production 2 to 5-fold.
only. Hence, unlike RSV but similar to PVM, the M2-1 protein was not necessary for multiple gene transcription.

The addition of M2-2 plasmid to the APV PRS had a marked effect on reporter protein production. This suggests that the M2-2 protein is likely to be produced by the virus, although it has not been detected as a protein expressed during APV infection by western blot (Ahmadian et al., 1999). The effect of the M2-2 plasmid on protein expression was modulated by the presence of the M2-1 plasmid. Gould and Easton (2007) showed that the overlapping region of the M2-1 and M2-2 ORF can direct coupled translation of the second ORF following translation of the first ORF. However, they found that the APV M2-1/M2-2 overlapping region was 64-times less efficient at directing coupled translation than the M2-1/M2-2 overlapping regions of PVM or RSV. This suggests that if the M2-2 protein is produced, it is a very low level. Hence, the work here supports the theory, based on RSV and PVM, that APV does express an M2-2 protein that it is a down regulatory protein that balances the enhancing effect the M2-1 protein.
Chapter 4:
Effect of mutation of the GS and GE signals on mRNA and protein expression in the APV plasmid rescue system
4.1 Introduction

With the establishment of a robust plasmid rescue system (PRS) for APV (chapter 3), the minimal requirements for protein production from the APV minigenome were found to be the N, P and L proteins though this was enhanced by the presence of the M2-1 gene. The functionality of the minigenome demonstrated that the non-translated cis-acting leader (Le), trailer (TR), consensus gene start (GS), and consensus gene end (GE) sequences are sufficient to direct transcription by the viral polymerase complex. These findings are consistent with the current model of pneumovirus transcription (section 1.4), whereby the polymerase complex transcribes mRNAs by travelling in a progressive stop–start manner from the 3’ to the 5’ end of the genome directed by the GS and GE sequences which flank each gene (Dickens et al., 1984; Randhawa et al., 1997). Dicistronic minigenomes have been used for mutational analysis of the gene junction regions in hRSV (Kuo et al., 1997; Kuo et al., 1996b). By mutating the gene start sequence between the reporter genes (5’ GS-CAT-GE-IR-GS-Luc-GE 3’, antigenome sense, appendix B), the effect on transcription of the LUC gene can be compared to the transcription of the CAT gene, which should be unaffected by the mutation. Similarly, the effect of mutation to the GE sequence of the CAT gene on termination or re-initiation should be reflected in changes in the amounts of CAT and LUC mRNAs produced.

For APV, the conserved GS sequence is identical for all genes except for that preceding the L gene. The APV consensus GS sequence is GGGACAAGU (antigenome sense) (Li et al., 1996; Ling et al., 1995; Ling et al., 1992; Randhawa et al., 1996b; Yu et al., 1991; Yu et al., 1992), whereas the L GS sequence has three differences (underlined), giving a sequence of \( \text{AGGACC\underline{AA}U} \) (Randhawa et al., 1996b) (section 1.4). The gene end sequences are more varied, with a consensus sequence of uAGUtAnnn(A\(_n\)), and the intergenic regions vary in length from one to 60nt. The APV pCATLUC minigenome contains the conserved GS sequence before both the CAT and the LUC genes. It has the P gene end sequence (UAGUUAUGAAAAAA, antigenome sense) following the CAT gene and the L gene end sequence (UAGUUAUAAAAA, antigenome sense) following the LUC gene. A five nucleotide intergenic region, introducing a C\(_{la}\)I site for cloning.
purposes (A/UCGAU, antigenome sense), separates the CAT GE sequence and the LUC GS sequence.

It has been previously shown that mutations introduced into specific nucleotide positions of the GS sequence have different effects on LUC reporter gene expression (Edworthy and Easton, 2005). The aim of this section was to use the pCATLUC minigenome to determine the effect of a set of naturally occurring mutations to the gene junction (GJ) sequence. The GJ consists of nucleotides found between the genes, including the upstream gene GE sequence, the intergenic region (IR) sequence and the downstream gene GS sequence. The level of correlation between mRNA production and protein production following transcriptional signal mutation to the pCATLUC minigenome was first examined by comparing reporter gene mRNA levels and reporter protein levels.

4.2 Comparison of mRNA and protein expression levels from GS sequence mutants

Based on previous studies in the RSV PRS (Marriott et al., 1999), it was hypothesized for the APV PRS that reporter proteins were produced in the same ratios as reporter gene mRNAs though this was not directly confirmed. In order to test this, GS sequence mutations that were known to decrease LUC protein production from the pCATLUC minigenome in the APV PRS (Edworthy and Easton, 2005) were examined to determine if this decrease was due to a loss of mRNA production. CAT protein or mRNA production was used as the control for each sample, as CAT production should be unaffected by mutations to the LUC GS sequence. Mutant plasmids were named to indicate the base of original nucleotide and the mutant base. For example, mutant plasmid G1A has the first base of the LUC GS sequence changed from a G to an A.

The first sets of mutants investigated altered the first base and fifth base of the GS sequence (table 4.1). These four mutants represented those with the least (C5U), greatest (G1C and G1U), and a modest (G1A), effect on reporter gene expression (figure 4.1)(Edworthy and Easton, 2005). The second set of mutations used to investigate mRNA and protein production included mutant A7U, G2U, G3A, A4C, and G8A. The GS mutant plasmid A7U has a change of an A at base seven of the consensus GS sequence to a U. The mutation introduced an AUG into the GS sequence (5’
GGGACAUGU 3’, mRNA sense, start codon in bold), creating an alternate start codon and alternate ORF out of frame with the LUC ORF. Mutants G2U and A7U were selected due to their low level of LUC protein production, mutants G3A and A4C for their middle level LUC protein production and mutant G8A for its high level of LUC protein production (figure 4.1, taken from Edworthy and Easton, 2005).

<table>
<thead>
<tr>
<th>GS mutant</th>
<th>GS sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wildtype</td>
<td>GGGACAAGU</td>
</tr>
<tr>
<td>G1A</td>
<td>AGGACAAGU</td>
</tr>
<tr>
<td>G1C</td>
<td>CGGACAAGU</td>
</tr>
<tr>
<td>G1U</td>
<td>UGGACAAGU</td>
</tr>
<tr>
<td>G2U</td>
<td>UGGACAAGU</td>
</tr>
<tr>
<td>G3A</td>
<td>GGAACAGU</td>
</tr>
<tr>
<td>A4C</td>
<td>GGGCCAAGU</td>
</tr>
<tr>
<td>C5U</td>
<td>GGGUAAGU</td>
</tr>
<tr>
<td>A7U</td>
<td>GGGACAAGU</td>
</tr>
<tr>
<td>G8A</td>
<td>GGGACAAAU</td>
</tr>
</tbody>
</table>

**Table 4.1 Sequences of the mutant GS signals.** ‘Wildtype’ indicates the consensus APV gene start sequence. All sequences are shown 5’ to 3’, mRNA sense.

By carrying out a standard PRS transfection experiment (section 2.8.2), reporter protein and mRNA were collected in sufficient quantities to visualize on northern blots and to measure in protein assays using the same transfection sample. This was done by using the mRNA and protein simultaneous harvest protocol as described in section 2.16. The northern blot membranes were incubated with a RNA probe specific for CAT and LUC mRNA (section 2.17) and all blots were repeated to confirm initial findings.

All constructs showed similar levels of CAT protein expression, apart from mutant A4C that produced approximately 2.8-fold more CAT protein seen from transfection of the pCATLUC minigenome (figures 4.2 and 4.3). Levels of CAT mRNA expression were similarly consistent, with mutant A4C having a darker band that the other samples.

The level of LUC protein produced from these constructs ranged from 408 RLU/10^6 cells for the unmuted pCATLUC plasmid (labelled ‘Wildtype’, figure 4.2) to no signal
for mutants G1C (figure 4.2) and G2U (figure 4.3). The plasmids that directed synthesis only of low levels of reporter protein showed no LUC mRNA production (mutants G1C, G1U, G2U, and A7U). Plasmids G1A, G3A and A4C gave moderate levels of LUC protein and had moderate levels of LUC mRNA production. The highest levels of LUC protein production were produced by the plasmids C5U and G8A (figures 4.2 and 4.3).

For the GS sequence mutants investigated, the levels of reporter protein expressed from pCATLUC minigenome constructs correlated well to levels of mRNA expressed from the
Figure 4.2 CAT protein and LUC protein levels of selected GS sequence mutants. Protein production from GS sequence mutants (A) was compared to mRNA production as measured by northern blot using a RNA probe specific for the LUC and CAT genes (B). The CAT gene is 66obp and the LUC gene is 1650bp. ‘Wildtype’ is the unmutated pCATLUC dicistronic minigenome and ‘no MG’ is the negative control of no minigenome included in the PRS experiment.
Figure 4.3 CAT protein and LUC protein levels of selected GS sequence mutants. Protein production from GS sequence mutants (A) was compared to mRNA production as measured by northern blot using a RNA probe specific for the LUC and CAT genes (B). The CAT gene is 66obp and the LUC gene is 1650bp. ‘Wildtype’ is the unmutated pCATLUC dicistronic minigenome. Both lanes of the left part of panel B derive from the same exposure of a single blot; background levels across the blot were uneven. ‘No MG’ is the negative control of no minigenome included in the PRS experiment.
constructs. The faint upper band seen in both northern blots is consistent with the mRNA products of read-through transcription. This does not appear to change significantly in intensity, suggesting that the GS sequence mutations did not affect mRNA termination at the gene junction. This would agree with the discontinuous model of transcription where an intact downstream gene start signal is not required for termination and polyadenylation of the upstream mRNA.

4.3 Activity of transcriptional signal mutations in the APV plasmid rescue system
4.3.1 Gene junction mutations
Reverse genetics systems have played a major role in the analysis and development of protective, non-pathogenic, APV vaccine strains. APV A field strain #8544, an actual vaccine strain isolated from an outbreak in northern England in 1985, was found to protect turkeys from challenge with APV. However, it was also shown to produce post-vaccination disease in some cases, even following passage in Vero cells (Williams et al., 1991a; Williams et al., 1991b). Naylor & Jones (1994) and Naylor et al. (2007) determined that this was due to the presence of a sub-population of virulent virus. Plaque purification produced 12 viruses that were free of the disease-causing sub-populations. These isolates had varying degrees of sequence conservation and differing abilities to induce protective immunity. Complete sequencing of the original isolate (field strain #8544), of the attenuated (virus P20), and of selected plaque-purified isolates (viruses C, K, F, H, and L) revealed a set of point mutations which produced four phenotypes for the virus: protective against APV challenge but severely pathogenic (isolate #8544), protective but some pathogenicity (virus P20), protective and non-pathogenic (viruses C and K), or non-protective and non-pathogenic (viruses F, H and L). Briefly, compared to the pathogenic progenitor strain isolate #8544, the P20 virus had nine base changes: one change in the F gene start sequence, two in the F gene itself (coding), one in the M2-1 gene (coding), one in the SH-G gene junction, one in the G gene (non-coding), and two in the L gene (non-coding). The protective viruses C and K were found to have the same sequence as the P20 virus. Virus L differed in sequence from P20 in having two bases that matched isolate #8544 (F gene start and L gene) and extensive mutation to the G gene (55 A to G mutations in the coding region). Viruses F and H differed from P20 by
one mutation each to the SH-G gene end sequence. The four SH-G gene junctions found upon sequencing the progenitor strain 8544 (IR mutant 1), the protective virus C (IR mutant 2), and the two non-protective viruses H and F (IR mutant 3 and IR mutant 4, respectively) are shown in figure 4.4. The IR mutant 1 represents the naturally occurring SH-G gene junction for virus strain 8544 and does not represent a sequence that is a mutation of the ‘wt’ virus. All other viruses were sequenced only across the 11 regions of mutation found between isolate #8544 and P20 and were found to have identical sequence to virus P20 (Naylor et al., 2007).

**Gene Junction Mutations**

\[
\text{pCATLUC minigenome gene junction (GJ):}\n\text{UAAUAGUUAUGAAAAAUUCGAUGGGGACCAAGUAACCAUG} \\
\text{IR mutant 1 (pathogenic, protective strain #8544 GJ):}\n\text{UAAUAGAUAUUUGACAACACUAGUGCCAAAUAUAAUAGGCAACAG} \\
\text{UAAUUAUUUAUUAAAGAAAGGUCGGGACAAGUAUCUCUAUG} \\
\text{IR mutant 2 (non-pathogenic, protective viruses C, K and P20 GJ):}\n\text{UAAUAGAUAUUUGACAACACUAGUGCCAAAUGAUAGGCAACAG} \\
\text{UAAUUAUUUAUUAAAGAAAGGUCGGGACAAGUAUCUCUAUG} \\
\text{IR mutant 3 (non-pathogenic, non-protective virus H GJ):}\n\text{UAAUAGAUAUUUGACAACACUAGUGCCAAAUGAUAGGCAACAG} \\
\text{UAAUUAUUUAUUAAAGAAAGGUCGGGACAAGUAUCUCUAUG} \\
\text{IR mutant 4 (non-pathogenic, non-protective virus F GJ):}\n\text{UAAUAGAUAUUUGACAACACUAGUGCCAAAUGAUAGGCAACAG} \\
\text{UAAUUAUUUAUUAAAGAAAGGUCGGGACAAGUAUCUCUAUG} \\
\]

**Figure 4.4** Dicistronic minigenome (mRNA sense) gene junction mutants. Shown here are the IR mutant 1-4 gene junction sequences (antigenome sense), with stop and start codons shown in blue, point mutation position in red, and the GE and GS sequences underlined.
As the only mutations in common in the non-protective (viruses F and H) and the protective (viruses C, K and P20) genomes were single base changes in the SH-G gene junction, it was hypothesized that the differences in the ability of these viruses to elicit protection to APV challenge may be due to changes in the amounts of SH and G proteins expressed by the virus. The effect of the mutations on gene expression was investigated by replacing the existing dicistronic pCATLUC minigenome gene junction with the SH-G gene junction and carrying out point mutations to create the four desired gene junction sequences.

4.3.1.1 Generation of IR mutant 1
The mutation of the pCATLUC minigenome gene junction to the SH-G gene junction was carried out using a series of overlapping primers. As described in section 2.6.1, it is possible to insert large fragment into plasmid using a modified QuickChange site directed mutagenesis protocol. The first step was to create a large fragment that contained the desired SH-G gene junction of 150-200 nucleotides that matched the start of the LUC gene and the end of the CAT gene. To create the insertion fragment, a series of PCRs were carried out (shown diagrammatically in figures 4.5 to 4.7) using primers ranging from 19 to 84 bases (appendix A). PCRs 1 and 2 were carried out using the pCATLUC minigenome as the template DNA. A 20 base forward primer (primer DSMG IR F) which bound from 200bp upstream of the end of the CAT gene stop codon was used with a reverse primer (primer IntMut/DSMG R) containing the final 15 bases of the CAT gene sequence followed by a non-binding region of 69 bases. This PCR gave a fragment of 269bp which was named fragment A. For fragment B, a PCR was carried out using a 40 base forward primer (primer IntMut/DSMG F) which contained the final 31 bases of the replacement gene junction followed by the first 19 bases of the LUC gene, and a 19 base reverse primer (primer DSMG IR R) that complimented the LUC gene 186 bases downstream of the LUC gene start codon (figure 4.5). PCR for fragment C was carried out using fragments A and B and the full length cloned APV CASA genome) (Naylor et al., 2004) as the template (figure 4.6). Fragment C contained the desired progenitor gene junction flanked by sequence complementary to the CAT and LUC genes.
A

PCRs 1 and 2 template:

**pCATLUC minigenome**

![Diagram of pCATLUC minigenome]

5’ TAATAGTTATGAAAAATCGATGGGACAAGTAACC 3’

**Primer:** DSMG IR F
**Product:** Fragment A
**Primer:** IntMut/DSMG R

**Primer:** DSMG IR R
**Primer:** IntMut/DSMG F
**Product:** Fragment B

B

**DNA ladder (bp)**

<table>
<thead>
<tr>
<th>Ladder</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.5 Generation of fragments A and B.** A) Using the pCATLUC minigenome as the template for PCR, primers DSMG IR F and IntMut/DSMG R were used to introduce 69bp (in green) of SH-G gene junction sequence to the 3’ end of a 200bp fragment (in light green) of the pCATLUC plasmid (fragment A). For fragment B, the pCATLUC plasmid was again used as the template for PCR and primers IntMut/DSMG F and Primer DSMG IR R were used to introduce 31bp (in green) of SH-G gene junction region sequence to the 5’ end of an 179bp fragment (in light blue) of the pCATLUC minigenome. B) PCR product fragments A (269bp) and B (210bp) were visualized on an agarose gel.
Figure 4.6 Generation of fragment C. A) Fragments A and B were used in PCR with the APV full length cloned genome (APV CASA) to create fragment C. Fragment C has the SH-G intergenic flanked by pCATLUC minigenome sequence. B) PCR product fragment C (458bp) was visualized on an agarose gel.
Fragment C was then used in the modified QuickChange site directed mutagenesis protocol (section 2.6.1), with the pCATLUC minigenome construct as template (figure 4.7). Twenty-four colonies were selected and used for small-scale DNA purification. The resulting plasmid DNA was screened for the presence of the insertion by digestion with \textit{Cla}1 and \textit{Sph}1 enzymes. The presence of the desired insertion into the plasmid resulted in a single band of 5383bp, whereas the plasmid without insertion gave bands of 675bp and 4661bp. As seen in figure 4.8, samples 7, 9, 11, 16-20, 22, 24, 26, 28-30 had the desired restriction digest fragment pattern and samples 7, 9, 11, 16-20, 22, 24 were

---

**Figure 4.7 Generation of IR mutant 1.** Fragment C was used in the modified quick change protocol PCR to introduce the modified IR mutant 1 intergenic region into the pCATLUC minigenome, resulting in the desired IR mutant 1 plasmid.
sequenced using primers DSMG check 5 and DSMG check 6. All samples were found to have a G→A change at base 4480 in the CAT gene that was non-coding. Sequencing of samples 7, 11, 16-18, and 19 confirmed insertion of the desired gene junction. Sample 18 was chosen to be used for IR mutant 1 plasmid in all further experiments.

4.3.1.2 Generation of IR mutant 2

In order to generate the second set of mutations to the gene junction to create IR mutant 2 (figure 4.4, A to G change at base 28 of the IR mutant 1 gene junction), QuickChange site directed mutagenesis was utilized as above. It had been observed that, even for point mutations, primers or DNA fragments in excess of 40 bases were necessary when using

![Agarose gels examining of restriction digestion products for IR mutant 1 cloning. Thirty samples were digested with ClaI and SphI enzymes for loss of the ClaI site found in the unmutated pCATLUC minigenome intergenic region. Introduction of the desired intergenic region should result in a single band of 5383bp, whereas digestion of the pCATLUC minigenome gives two bands (4661 and 675bp). Samples 7, 9, 11, 16-20, 22, 24, 26 and 28-30 gave the desired bands and samples 7, 11, 16 to 19 were sequenced with primers check 5 and 6. Based on the results of this sequencing, sample 18 was fully sequenced and used in all further experiments.](image)
this protocol (Dr. O. Dibbin, personal communication). Two 80 base primers, IRmut2 FOR long and IRmut2 REV long (appendix A), were prepared (figure 4.9). These primer pairs were used in the method for full-length clone modification protocol (section 2.6) with IR mutant 1 plasmid as the template in the PCR. This protocol resulted in one sample, sample 7, with the desired gene junction mutations for IR mutant 2. This plasmid was fully sequenced and used in all further experiments.

**Pt7.2 backbone**

**Figure 4.9 Generation of IR mutant 2.** IRmut2 REV long and IRmut2 FOR long are complementary primers which match the gene junction region of the IR mutant 2 plasmid apart from the A->G mutation at base 28 of the gene junction sequence (antigenome sense). Primers IRmut2 REV long and IRmut2 FOR long were used in the modified QuickChange site-directed mutagenesis PCR to introduce the IR mutant 2 intergenic region into the IR mutant 1 plasmid (in yellow), resulting in the desired IR mutant 2 plasmid with an A to G mutation at base 28 of the gene junction (antigenome sense).
4.3.1.3 Generation of IR mutant 3

The strategy used to generate the IR mutant 3 plasmid was to create a large fragment containing the desired gene junction and to use this fragment in QuickChange site directed mutagenesis PCR. PCR was carried out using DSMG IRF with IR3 REV LONG as primer pairs and IR mutant 2 DNA as the template and resulted in the generation of fragment D (figure 4.10). Using fragment D with the IR mutant 2 plasmid as the template in QuickChange site directed mutagenesis PCR (as above), the IR mutant 3 plasmid was generated (figure 4.11). The resulting product was transformed, the DNA isolated, and sixteen colonies were screened by sequencing with the primer DSMG check 5. IR mutant 3 sample 66 was fully sequenced and used as the IR mutant 3 plasmid in all further experiments.

**Figure 4.10 Generation of fragment D.** A) Using the IR mutant 2 plasmid as the template for PCR, primers DSMG IR F and IR mut 3 REV LONG were used to introduce 67bp (in orange) of IR mutant 3 intergenic region sequence to the 3’ end of a 200bp fragment (in light green) of the CAT gene (Fragment D). B) PCR product fragment D (267bp) was visualized on an agarose gel.
**Primer:**

Fragment D

**Product:**

IR mutant 3

5' TAA

TGAATTTTGACAACACTAGTGCCAAAT

G

ATAGGCAACAG

TATTATTTAATTTAAAAAGAAAGGTCGGGACAAGTATCTCTCTATG 3'

**FL quick-change PCR template:**

**IR mutant 2 plasmid**

GS  CAT gene  GE  IR  GS  LUC gene  GE

5' TAATGAATTTTGACAACACTAGTGCCAAATGATAGGCAACAG

TATTATTTAATTTAAAAAGAAAGGTCGGGACAAGTATCTCTCTATG 3'

**Figure 4.11 Generation of IR mutant 3.** Fragment F was used in the modified QuickChange site-directed mutagenesis PCR to introduce the IR mutant 3 gene junction sequence into the IR mutant 2 plasmid, resulting in the desired IR mutant 3 plasmid.
4.3.1.4 Generation of IR mutant 4

The strategy attempted to create IR mutant 4 was to PCR very long fragments which spanned the IR and could be used full-length quick-change mutagenesis. PCR to generate the desired fragments was carried out using Elongase polymerase (as above) with primers DSMG IR F and IR mut 4 REV LONG and the IR mutant 1 plasmid as the template (figure 4.12). PCR resulted in the desired fragments (fragments E and F) which were then used to generate a very long fragment (fragment G) using PCR (figures 4.13). Fragment G was purified as above and used in FL quick-change mutagenesis PCR (figure 4.14) to create the IR mutant 4 plasmid.

Transfection as above resulted in approximately 200 colonies per plate, of which 12 were selected for small scale DNA preparation. The preparations were then sequenced using the primer DSMG check 5. Sequencing showed IR mutant 4 sample 22 to have the desired mutations, although this sample also contained a coding change to base 4055 in the LUC gene. By subjecting both sample 22 and the pCATLUC plasmid to restriction digestion and ligation it was possible to replace the coding mutation at base 4055 and produce a correct plasmid (figure 4.15).

Both plasmids were subjected to sequential restriction digest with BstB1 and Sph1 and the desired fragments purified (figure 4.15). Ligation resulted in 24 clones that were screened by sequencing using primer DSMG check 5. IR mutant 4 sample 23 was chosen to be used in future experiments as the IR mutant 4 plasmid.
**Figure 4.12 Generation of fragments E and F.** Using the IR mutant 1 plasmid as the template for PCR, primers DSMG IR F and IR4 REV LONG were used to introduce 67bp (in dark blue) of IR mutant 4 gene junction sequence to the 3’ end of a 200bp fragment (in light green) of the CAT gene (fragment E). For fragment F, the IR mutant 1 plasmid was again used as the template for PCR and primers IR mut 4 FOR LONG and primer DSMG IR R were used to introduce 79bp (in dark blue) of IR mutant 4 gene junction sequence between an 179bp fragment (in light blue) of the LUC gene and a 13bp fragment of the CAT gene (in light green). B) PCR products fragment E (left, 267bp) and fragment F (right, 271bp) were visualized on an agarose gels.
Figure 4.13 Generation of fragment G. Fragments E and F were used in PCR with the IR mutant 1 plasmid to create fragment G. Fragment G has the IR mutant 4 gene junction sequence flanked by IR mutant 1 sequence, with changes of an A to G change at base 28 and an A to a T at base 48 (antigenome sense).
Figure 4.14 Generation of IR mutant 4*. Fragment I was used in the modified QuickChange site-directed mutagenesis PCR to introduce the IR mutant 4 gene junction into the IR mutant 2 plasmid, resulting in the IR mutant 4 plasmid. Sequencing of plasmid IR mutant 4 showed a point mutation of a C to a T at base 4055 (antigenome sense) in the LUC gene, which would change the amino acid for that codon from a proline to a serine.
Figure 4.15 Generation of IR mutant 4. A) Plasmid IR mutant 4* was found to contain a mutation in the LUC gene at base 4055 (antigenome sense). In order to correct the mutation, the IR mutant 4 plasmid and the pCATLUC minigenome were digested with enzymes BstBI and SphI to give 488 and 4895bp fragments. The 488bp fragment from pCATLUC was ligated with the 4895bp fragment from the IR mutant 4* plasmid, producing an IR mutant 4 plasmid with an unmutated LUC gene. B) Digestion products were visualized on an agarose gels, with the IR mutant 4* as the vector (left) and the pCATLUC minigenome as the insert. The insert fragment was visible only when larger volumes were added to the gel (right, indicated with an arrow).
4.3.2 Determination of effect of mutation of the pCATLUC minigenome gene junction

Plasmids IR mutant 1, IR mutant 2, IR mutant 3 and IR mutant 4 were transfected into HEp2 cells with the APV PRS helper plasmids and cell lysates were assayed for reporter protein (sections 2.8.2, 2.9, 2.10 and 2.11). The mean of the three independent experiments was determined and standard error calculated and is shown in figure 4.16.

When introduced into the PRS, all minigenomes produced levels of CAT protein similar to that of minigenome IR mutant 1 (p-values of 0.4319 to 0.7032, figure 4.17), although the mean values for IR mutants 3 and 4 are lower than those seen for IR mutants 1 and 2 (30.0 and 31.5ng verses 27.5 and 26.6ng, respectively). Luc protein production from the set of IR mutants ranged from 40.4 to 103.2 RLU per well, with IR mutant 1 producing the highest levels of protein. IR mutant 2 and IR mutant 3 produced similar levels of LUC protein (98.2 and 74.0 RLU per well on average). The lowest levels of LUC protein production were seen for IR mutant 4, at an average of 40.6 RLU per well. IR mutants 1 had significantly higher levels of LUC protein production than either IR mutant 3 (p=0.0197) or IR mutant 4 (p=0.0067). IR mutant 4 produced significantly less LUC protein than all other mutants (p-values from 0.0067 to 0.0140, figure 4.17).
The levels of CAT protein production from IR mutants 1 to 4 were also compared to CAT protein production from the pCATLUC minigenome that was included as the positive control. Modification of the CAT-LUC gene junction in this plasmid to be the longer SH-G gene junction found in IR mutant 1 (figure 4.4) resulted in consistently higher levels of CAT production (data not shown, 23.7ng for the pCATLUC minigenome versus 30.0ng for IR mutant 1 plasmid), indicating that production of protein from the first ORF of the dicistronic minigenome was affected by mutation of the gene junction.

Levels of LUC protein production from the IR mutant 1 were, conversely, consistently lower than the levels produced from the pCATLUC minigenome (95.5 verses 171.3 RLU per well, respectively). The control experiments (no L plasmid added to the PRS) gave an average of 0.0ng of CAT protein per well and an average 0.6 RLU of LUC protein per well.

The levels of CAT protein production from IR mutants 1 to 4 were also compared to CAT protein production from the pCATLUC minigenome that was included as the positive control. Modification of the CAT-LUC gene junction in this plasmid to be the longer SH-G gene junction found in IR mutant 1 (figure 4.4) resulted in consistently higher levels of CAT production (data not shown, 23.7ng for the pCATLUC minigenome versus 30.0ng for IR mutant 1 plasmid), indicating that production of protein from the first ORF of the dicistronic minigenome was affected by mutation of the gene junction. Levels of LUC protein production from the IR mutant 1 were, conversely, consistently lower than the levels produced from the pCATLUC minigenome (95.5 verses 171.3 RLU per well, respectively). The control experiments (no L plasmid added to the PRS) gave an average of 0.0ng of CAT protein per well and an average 0.6 RLU of LUC protein per well.

Figure 4.17 P values determined using the Paired T Test for significance between the IR mutants 1 to 4 for CAT and LUC protein values. Using the Paired T Test of significance, the CAT protein (top) and LUC protein (bottom) p-values for IR mutants 1 to 4 were compared (labelled 1 to 4). The pairs that are significantly different (based on a 0.05 p-value significance level) are in bold.
4.3.3 Generation of dual GS sequence point mutations

The L GS sequence is the only non-consensus APV GS sequence and differs by three nucleotides from the consensus GS sequence (AGGACCAAU, differences underlined). Edworthy and Easton (2005) showed that the introduction of the L GS sequence in place of the consensus GS sequence before the LUC ORF in a dicistronic minigenome resulted in 47% of the activity seen for the consensus GS sequence. Interestingly, when each mutation was introduced individually into the consensus GS sequence (mutants G1A, A6C, and G8A), the percentage activities of the constructs were similar (mutant G1A was 46%, A6C was 52%, G8A was 59% activity), as was the combination of the first and last mutations together (mutant G1A/G8A, 40% activity). These experiments showed that the effect of the mutations was not cumulative (Edworthy and Easton, 2005) and also indicated that the L gene expression is determined by the GS sequence as well as the 5’ proximal location of the L gene in the genome RNA.

The potential role of the GS in pathogenicity was raised following analysis of a naturally occurring mutation in a pathogenic isolate of the APVA CVL-14 virus (Dr. R. Ling, personal communication). This pathogenic isolate was generated by passaging the non-pathogenic APVA CVL-14 virus through turkeys until infection caused disease. This pathogenic isolate was fully sequenced and three regions of mutation were found, including a mutation to the L GS sequence. This mutation resulted in the L gene having the GS sequence 5’ GGGACC AAU 3’, which has only two changes compared to the standard consensus GS sequence (changes in bold). No differences in transcription levels were seen in northern blot experiments of the pathogenic strain mRNA compared to the wildtype virus mRNA, suggesting that the mutation to the L GS sequence did not affect L gene transcription. The L GS mutation was introduced into the CASA APV full-length genome and the virus rescued from this was used to infect turkeys. However, none of the viruses produced with the CASA backbone were viable (Prof. A. Easton, personal communication).

In order to determine the potential effect of these mutations on gene expression, two mutants were created containing two out of the three possible L GS sequence changes.
They were named G1A/A6C (GS sequence $\text{AGGAC}_{\text{AGU}}$) and A6C/G8A (GS sequence $\text{GGGAC}_{\text{AAU}}$), indicating the original nucleotide, its position in the GS sequence and the mutant nucleotide.

The construction of the A6C/G8A and G1C/A6C dual GS sequence mutants was carried out as follows. PCR was performed using Pfu DNA polymerase (Stratagene) following the manufacturer’s instructions. The mutagenic forward primers for each reaction and different dicistronic minigenome templates are listed in table 4.2. The reverse primer for all these PCRs was the T7prom primer (appendix A). The PCR products were used to create mutated fragments for ligation into the APV pCATLUC minigenome. The products of PCR were sized on an agarose gel, as shown in figure 4.18. PCR products were then purified by gel extraction to remove template DNA, as described in section 2.5.3.1. The purified PCR product was then digested with ClaI and SphI enzymes, the products separated on an agarose gel, and the 675bp band extracted. The vector was obtained by digesting the pCATLUC minigenome with ClaI and SphI enzymes and purified as for the insertion fragments. Ligation of the mutated insertion fragments into the vector fragment was then carried out and the ligation mixture was transformed into competent $E.\ coli$ DH5α. Plasmid DNA was prepared and screened by sequencing of the insert region. Samples with the correct insertion were subjected to large-scale preparation and fully sequenced.

<table>
<thead>
<tr>
<th>PCR name:</th>
<th>DNA template:</th>
<th>Forward primer:</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1A/A6C PCR 1</td>
<td>G1A</td>
<td>ADM G1A/A6C</td>
</tr>
<tr>
<td>G1A/A6C PCR 2</td>
<td>A6C</td>
<td>ADM G1A/A6C</td>
</tr>
<tr>
<td>A6C/G8A PCR 3</td>
<td>A6C</td>
<td>ADM A6C/G8A</td>
</tr>
<tr>
<td>A6C/G8A PCR 4</td>
<td>G8A</td>
<td>ADM A6C/G8A</td>
</tr>
<tr>
<td>G1A/A6C PCR 5</td>
<td>DSMG</td>
<td>ADM G1A/A6C</td>
</tr>
<tr>
<td>A6C/G8A PCR 6</td>
<td>DSMG</td>
<td>ADM A6C/G8A</td>
</tr>
</tbody>
</table>

Table 4.2 Primers and template pairs used for generation of dual GS sequence mutants. For each PCR, a DNA template was selected that was either unmutated (the DSMG plasmid) or had been mutated at the LUC GS sequence. Primers were designed to introduce one or two point mutations into the LUC GS sequence. Full primer sequence can be found in appendix A.
4.3.4 Effect of dual GS sequence point mutation

The effect of the introduction of two of the three non-consensus GS sequence bases to the LUC gene GS sequence on LUC protein production was calculated as ‘percentage activity’ (i.e. the LUC:CAT protein ratio for the mutant plasmid divided by the LUC:CAT protein ratio for the wildtype plasmid, multiplied by 100) as described previously (Edworthy and Easton, 2005) (figure 4.19). The effect of dual mutation for mutant G1A/A6C was to reduce the activity of the dicistronic minigenome to 25% of the wildtype pCATLUC minigenome. Mutant A6C/G8A gave 40% the activity of the unmutated wildtype pCATLUC minigenome. The low level of activity seen for mutant G1A/A6C, even lower than that for the L GS sequence, suggests that the third change (G8A) may compensate for the effect of the first two. The activity of the mutant G8A alone was the highest of the L GS single point mutations with 59% activity. For all of the dual and triple mutants, the resultant effect on gene expression was not an average or an addition of the effects seen with the single mutations. However, the two point mutations with lowest levels of activity (G1A with 46% and A6C with 52%) when present in combination did result in the dual mutant with the lowest activity levels (G1A/A6C with 25%).

Figure 4.18 Agarose gel examining PCR products for cloning of L GS sequence dual mutants. G1A/A6C PCR 1 (PCR 1), G1A/A6C PCR 2 (PCR 2), A6C/G8A PCR 3 (PCR 3), A6C/G8A PCR 4 (PCR 4), G1A/A6C PCR 5 (PCR 5), and A6C/G8A PCR 6 (PCR 6) all produced the desired approximately 1960bp band (as indicated by the arrow) using the template and forward primer combination listed in table 4.3. The reverse primer for all reactions was the T7prom primer.

![Agarose gel examining PCR products for cloning of L GS sequence dual mutants](image)
Conclusions

The value of the minigenome system as a fast and reliable way to measure the effect of mutation of the cis- and trans- acting signals of the APV genome was investigated by examining the effect of GS sequence mutations on mRNA and protein production simultaneously. Northern blot analysis to detect LUC mRNA showed a strong correlation between mutation of the GS sequence, a change in its ability to act as a transcriptional promoter, and a change in the levels of protein detected from the reporter genes. Of particular interest was mutant A7U, whereby mutation of the A at position seven of the GS sequence resulted in an alternative start codon being introduced upstream of the LUC gene start codon. It was possible that some of the loss of functional protein

Figure 4.19 Effect of dual GS point mutation on reporter protein expression. Mutants G1A/A6C and A6C/G8A were cloned (chequered bars). The average and standard error of the percentage activity was taken from the means of two or three independent experiments of PRS experiments. Error bars indicate standard error. Data for mutants G1A, A6C, G8A, G1A/G8A, and ‘L GS’ (having the L non-consensus GS sequence before the LUC gene) is reproduced here from Edworthy and Easton (2005) for comparison, all mutations being either single or dual point mutations found in the non-consensus L GS sequence. ‘Wildtype’ is the unmutated pCATLUC minigenome.

4.4 Conclusions

The value of the minigenome system as a fast and reliable way to measure the effect of mutation of the cis- and trans- acting signals of the APV genome was investigated by examining the effect of GS sequence mutations on mRNA and protein production simultaneously. Northern blot analysis to detect LUC mRNA showed a strong correlation between mutation of the GS sequence, a change in its ability to act as a transcriptional promoter, and a change in the levels of protein detected from the reporter genes. Of particular interest was mutant A7U, whereby mutation of the A at position seven of the GS sequence resulted in an alternative start codon being introduced upstream of the LUC gene start codon. It was possible that some of the loss of functional protein
observed for this mutant was due to use of the introduced codon for translation. However, Northern blot analysis showed that the mutation resulted in a loss of LUC mRNA production. For all the GS sequence mutants examined, the loss of LUC protein correlated directly with a loss of LUC mRNA production. The GS sequence mutations appeared to directly affect transcription of mRNAs rather than producing read-through or truncated transcripts. This result agrees with similar analysis done for the PVM and RSV where the gene start sequences mutations that affected levels of reporter protein also affected mRNA production (Dibben and Easton, 2007; Kuo et al., 1996b).

The L GS non-consensus sequence has been shown to down-regulate mRNA production (Edworthy and Easton, 2005). It was hypothesized that the APV L gene start sequence contained changes which in groups of two out of the three would also reduce mRNA, and hence protein, production in an additive manner. Here, L GS sequence dual mutants showed a non-additive effect.

The plasmid rescue system was used to investigate the effect of naturally occurring gene junction mutations on upstream and downstream protein expression. The mutations were observed when an APV strain (field stain #8544) was found to be both protective and have some post vaccine virulence. From this virus strain, a less virulent virus (virus P20) was purified, as well as non-protective/non-virulent virus stains (viruses H and F) and protective/non-virulent virus stains (viruses A-E, G, J, K and M). Sequence changes in the SH-G gene junction were found to be the sole differences between some of the protective and non-protective viruses. By cloning these gene junction mutations into the dicistronic minigenome used in the APV PRS, a number of conclusions can be drawn. IR mutants 2-4 all have an A to G change at base 4377. This is the only difference between the pathogenic strain #8544 virus SH-G gene junction and the non-pathogenic virus SH-G gene junctions. This suggested that the base change may contribute to the non-pathogenic phenotype. However, no effect on protein production due solely to the mutation was seen. Looking at the SH-G gene junction of the non-protective, non-pathogenic viruses (IR mutant 3 and 4), the gene junction sequence differs by one base in the GE sequence from the non-pathogenic but protective viruses. However, only in the
case of IR mutant 4 does the GE sequence mutation affect mRNA levels of the
downstream gene. The down-regulation of the downstream gene via mutation of the GE
sequence in the case of IR mutant 4 may contribute to the non-protective phenotype.
However, this effect on downstream gene expression was found only with the two IR
mutant 4 mutations. Thus, down-regulation of G gene expression may be one mechanism
by which the virus becomes non-protective.

The dicistronic minigenome gene junction mutation results presented here were found to
agree well with analysis of identical gene junction mutations introduced into the SH-G
gene junction of the cloned APV CASA genome (Naylor et al., 2007). Naylor et al.
(2007) mutated the APV CASA genome SH-G gene junction to be identical to the gene
junction of IR mutants 2 to 4 (cloned viral genomes rC, rH, and rF, respectively). The
recombinant genomes rC, rH and rF were used in the full-length rescue system to create
recombinant viruses. Unfortunately, it was not possible to directly determine the amounts
of SH and G protein produced from these viruses, as no antibodies were available.
However, the SH and G gene mRNA levels were determined for these viruses by
quantitative northern blot. The SH and G gene mRNA levels from the original viruses
#8544, C, and F (virus H was unavailable for analysis) were also determined and the
effect of the mutations compared to the effect of the IR mutant gene junctions on CAT
and luc protein expression. The ‘activity’ of each of the gene junction mutants was
determined comparing the expression of the upstream protein (CAT) or mRNA (SH RNA
and SH-G read-through mRNA) to the downstream protein (LUC) or mRNA (G gene
RNA)(Naylor et al., 2007). The ‘percentage activity’ was determined as described for the
GS mutants previously (Edworthy and Easton, 2005) by dividing the mutant activity level
by the relevant wild-type activity level (virus #8544 for all viruses and minigenome IR
mutant 1 for IR mutants 1 to 4) and multiplying by 100 (Edworthy and Easton,
2005)(figure 4.20). The average CAT and LUC protein level ratios from the IR mutants
3 and 4 were all within the range of the two recombinant viruses examined. The activity
level of IR mutant 2 was closer to that of the virus C than the recombinant virus rC, but
for all measures of activity, the GE found in IR mutant 4 gave the lowest activity.
For RSV, deletion of the GE sequence from a CAT minigenome in the RSV PRS reduced CAT protein expression to 1.6% of wildtype production. Northern blots of the mutant showed a loss of the CAT mRNA. Kuo et al. (1996) hypothesized that this loss of expression from the GE sequence deletion mutant was likely to be due to a disruption of the polyadenylation of RNA transcripts. Using a RSV dicistronic minigenome, deletion of the CAT GE sequence resulted in 16% CAT expression and 0.9% LUC protein expression. Northern blotting showed that CAT-LUC read-through mRNA was being produced primarily, with only a small amount of LUC mRNA. They concluded that the viral polymerase could not recognize the downstream GS sequence if the upstream GE sequence was deleted (Kuo et al., 1996b).

Determining the effect of a mutation on viral gene expression and the effect of the change in gene expression on phenotype is not always straightforward. The PRS can be used for rapid screening of gene junction mutants for their effect on upstream and downstream

**Figure 4.20 Percentage activity of gene junction mutation in vitro and in vivo.**

The percentage activity was calculated as described by Edworthy and Easton (2005) for the IR mutant genomes 1-4. The percentage activity for the same gene junction mutations in naturally occurring APV virus mutants (virus #8544, virus C and virus F) and recombinant APV viruses (virus rC, rH and rF)(Naylor et al., 2007). For the in vivo experiments with the IR mutant genomes, the mean of three independent experiments was calculated and the standard error determined.
gene expression. These mutations can then be used in the production rationally designed vaccine candidates that are protective and attenuated.
Chapter 5:  
Heterologous and chimeric protein function in the 
APV plasmid rescue system
5.1 Introduction

The development of a plasmid-based rescue system (PRS) for APV, described in chapter 3, provided a direct way to investigate the function of the trans-acting N, P, L and M2-1 proteins as well as the cis-acting leader, trailer, and gene junction non-translated genomic sequences. The substitution of functionally, or genetically, equivalent proteins from related viruses in place of one or more of the helper plasmids (heterologous rescue) has been used to investigate the role of conserved residues in protein function. The result of transfection of a set of homologous helper plasmids together with the genome or minigenome of a related virus can also provide information about the importance of conservation of cis-acting non-translated sequences. Heterologous rescue attempts using reverse genetics have been used to show the close functional relationship of the cis-acting and trans-acting signals of the metapneumoviruses APV subgroup C and human metapneumovirus (Govindarajan et al., 2006) and to generate candidate vaccine viruses (Pham et al., 2005). For RSV, rescue of a bovine RSV (bRSV) minigenome was achieved using ovine and human RS helper virus, showing that the bRSV cis-acting signals can be recognized by ovine RSV and human RSV polymerase complex proteins (Yunus et al., 1999). A similar approach has been described for parainfluenza virus types 1 and 3, which were shown to be able to rescue a Sendai virus DI genome when the RNA template was encapsidated by parainfluenza virus NP protein. The inability of either parainfluenza virus to rescue the Sendai virus genome when encapsidated with Sendai virus NP protein suggested that the NP:polymerase interaction was virus specific (Curran and Kolakofsky, 1991). Using the established APV plasmid-based rescue system, heterologous rescue of the RSV, PVM and hMPV minigenome by APV N, P, L and M2 proteins was attempted.

Heterologous rescue attempts to determine the level of functional homology between viruses. The other basic way of predicting functional homology is by comparing protein identity levels. As described in chapter 1, but using the clone sequences from chapter 3, the percentage of amino acid identity of the pneumovirus N, P, and L and M2-1 proteins was carried out as described in section 1.15 and is presented in table 5.1. Amino acid
identity here refers to the number of residues that match exactly between the two sequences. It was also useful to consider conserved substitutions in a percentage identity calculation (here referred to as percentage identity with conserved substitutions). This calculation includes residues whose comparison yields a greater than zero score in the Blosum 62 scoring table using GeneDoc software. For the purposes of this discussion, levels of amino acid identity that are 60% or above will be termed ‘high’, levels between 60 and 35% ‘low’, and levels of 34% and below ‘very low’.

Comparing the N, P, L and M2-1 proteins of these pneumoviruses, overall APV has the highest levels of homology with hMPV, followed by PVM and RSV. Previously, it was shown that the APV helper plasmids were unable to functionally replace the RSV helper plasmids in the RSV PRS (Stokes, 2003), as might be expected from the low levels of homology between these proteins. The APV subgroup A N protein was 391 residues in length (Li et al., 1996) and has a high (68%) level of amino acid identity with the 394 residue hMPV N protein (van den Hoogen et al., 2001), rising to 86% when conservative substitutions are included. A region of 98% identity was found in the carboxyl-terminal half of the APV N protein, from residues 238-290. The APV N protein was found to have low levels of homology to the 393 residue PVM N protein and 391 residue RSV N protein (Thorpe and Easton, 2005; Tolley et al., 1996) with 42 and 39% amino acid identity, respectively.

<table>
<thead>
<tr>
<th></th>
<th>hMPV</th>
<th>PVM</th>
<th>RSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV N</td>
<td>68% (86%)</td>
<td>42%</td>
<td>39%</td>
</tr>
<tr>
<td>APV P</td>
<td>55% (69%)</td>
<td>25%</td>
<td>24%</td>
</tr>
<tr>
<td>APV L</td>
<td>63% (79%)</td>
<td>50%</td>
<td>43%</td>
</tr>
<tr>
<td>APV M2-1</td>
<td>71% (84%)</td>
<td>35%</td>
<td>39%</td>
</tr>
</tbody>
</table>

Table 5.1 Amino acid identities of the APV N, P, L and M2-1 protein sequences with the homologous proteins of hMPV, PVM and RSV. The percentage amino acid identity, calculated as described in section 2.15, is shown here in pair-wise comparisons of the APV N, P, L or M2-1 protein with the equivalent hMPV, PVM or RSV protein. For comparison of the APV and hMPV proteins, included in parenthesis is a percentage identity calculation including similar amino acid substitutions based on a Blosum 65 amino acid similarity matrix.
The APV P protein, at 287 residues (Ling et al., 1995; Ling and Pringle, 1988), was shorter than the hMPV and PVM P proteins which are 294 and 295 residues in length, respectively (Thorpe and Easton, 2005; van den Hoogen et al., 2001), but longer than the RSV 241 residue protein (Tolley et al., 1996). Overall, the APV P protein has lower homology with the hMPV, PVM, and RSV P proteins than was seen for the N proteins. APV P has 55% identity with hMPV (69% with similar residue substitution), but the first 28 residues are 75% identical and the carboxyl-terminus half of the protein (residues 139 to 278) are 72% identical. The carboxyl-terminus has two regions of 94% and 83% identity (residues 168 to 203 and 227 to 256, respectively). The APV P protein has 25% and 24% amino acid identity with PVM and RSV, respectively, but the region spanning residues 168 to 203 shows a 61% amino acid identity across the pneumovirus sub-family.

The APV L protein was 2004 residues long (Randhawa et al., 1996b), which was similar to the hMPV L protein (2005 residues; (van den Hoogen et al., 2001), but shorter than the PVM and RSV L proteins (2040 and 2165 residues, respectively)(Thorpe and Easton, 2005; Tolley et al., 1996). The APV L protein has 63% amino acid identity with the hMPV L protein, rising to 79% when including similar residue substitutions. When compared to PVM or RSV, the APV L protein has 50% and 43% amino acid identity, respectively. However, regions of 100% conservation exist of up to 11 consecutive residues (for example, APV residues 1185 to 1195) that are conserved in all four viruses.

The APV M2-1 protein was 186 residues in length (Ling et al., 1992) and has high level (71%) of amino acid identity with the 187 residue hMPV M2-1 protein (van den Hoogen et al., 2001). The level of homology rises to 84% when including similar residue substitutions. The APV and hMPV M2-1 have eight regions of 6 to 17 consecutive residues that are conserved. The levels of amino acid identity between the APV M2-1 protein and the RSV or PVM M2-1 proteins are lower, at 39% or 35% identity, respectively.

The minigenomes used in the PRS contain only non-translated nucleotides from the viral genomes (section 1.5). Considering the non-translated regions of the pneumovirus...
genomes, comparisons of the leader sequences (section 1.4.4) showed that APV has low levels of nucleotide identity with either hMPV (41%), PVM (53%) or RSV (53%), however 10 of the first 11 nucleotides from the viral 3’ end are identical for APV, RSV and PVM. The trailer region of APV (section 1.4) has less nucleotide identity with the other pneumoviruses- hMPV (22%), RSV (16%), and PVM (13%)- but all four viruses have identical sequence for 11 out of the 12 nucleotides at the viral 5’ end. All four virus genomes have a similar gene start (GS) sequence for their first gene, with hMPV and APV containing identical motifs of 5’GGGACAAUG 3’(mRNA sense). The PVM and the RSV first (NS1) gene start sequences contain one and two base differences each, with sequences of 5’AGGACAAGU 3’ and 5’GGGGCAAAAU 3,’ respectively (table 1.14). From mutational analysis of the APV consensus gene start sequence which was found before the first (N) gene, it has been shown that mutation of the first position from a G to an A residue decreases activity of the GS sequence by approximately 50%, as does mutation of position 4 from an A to a C residue or mutation of position 8 from an G to an A residue (figure 4.1) (Edworthy and Easton, 2005). However, as shown in figure 4.19, the effect of multiple mutations to the APV gene start sequence was not an additive effect (section 4.3.4). The gene end (GE) sequences of pneumoviruses are poorly conserved but have the consensus sequence of uAGuUa(n)2;3(A)4-6 (table 5.2). Kuo et al. (1997) showed that the diverse RSV-A2 GE sequences were equally able to facilitate mRNA production except for the NS1 and NS2 GE sequences (60% as efficient). The RSV-S2 GE sequences are identical to those found for RSV-A2 (Kuo et al., 1997) except for the N and P GE sequences which have an additional A at the 3’ end, the M GE sequence which has one fewer A at the 3’ end, and the G GS sequence which has sequence AGUCAUUAAAAA instead of the A2 G GE sequence of AGUUACUUAAAAA (5’ to 3’, (+) sense, differences in bold). The intergenic regions of pneumoviruses vary greatly in length and content and for RSV-A2 it was shown that the intergenic region did not play a role in transcription (Kuo et al., 1996a).
Previously, Stokes (2003) used the RSV PRS to investigate the functioning of chimeric N and M2-1 proteins with the aim to isolate functional regions of the RSV N and M2-1 proteins. It was shown that the PVM M2-1 protein, although 40% identical to the RSV M2-1 and retaining a conserved Cys3-His1 motif, could not functionally replace the RSV M2-1 protein in a RSV PRS using lacZ as the reporter protein. Chimeras were created with RSV and APV M2-1 protein and used to identify two functional regions that separately affected production of the first (CAT) gene (the N-terminal 44 residues) and both the first and second (LUC) genes (residues 93-152) from the RSV pCATLUC minigenome (Stokes, 2003) (Dr. A. Marriott, personal communication). Similarly, it was shown that the APV N protein could not replace the RSV N protein in the RSV PRS. Truncated versions of the RSV N proteins were non-viable in the RSV PRS and hence chimeric RSV/PVM N proteins were created to examine the RSV N protein functional regions. Using these chimeras, it was showed that residues 352-369 are essential for function of the N protein and binding to the P protein (Stokes, 2003; Stokes et al., 2003).

The aim of this section of the project was to determine the level of functionality of each of the RSV, PVM and hMPV N, P, L and M2-1 proteins in the APV PRS. For hMPV, the APV N, P, L and M2-1 proteins were substituted individually into the hMPV PRS to determine their ability to functionally replace the equivalent hMPV protein and the hMPV N, P, L and M2-1 proteins were tested collectively for their ability to rescue the APV minigenome. The ability of the APV N, P, L and M2-1 proteins to collectively

<table>
<thead>
<tr>
<th></th>
<th>Consensus GS sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Consensus GE sequence&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV</td>
<td>GGGACAAGU</td>
<td>nAGUtAnnn(A&lt;sub&gt;n&lt;/sub&gt;)</td>
</tr>
<tr>
<td>hMPV</td>
<td>GGGACAArU</td>
<td>UAGUUA(n)&lt;sub&gt;2-3&lt;/sub&gt;(A)&lt;sub&gt;4-6&lt;/sub&gt;</td>
</tr>
<tr>
<td>PVM</td>
<td>AGGAyAArU</td>
<td>UAGUUAnu(A&lt;sub&gt;n&lt;/sub&gt;)</td>
</tr>
<tr>
<td>RSV</td>
<td>GGGGCAAUa</td>
<td>uAGUuAnnn(A&lt;sub&gt;n&lt;/sub&gt;)</td>
</tr>
</tbody>
</table>

Table 5.2 Pneumovirus consensus gene start and gene end sequences. Positions that have only a purine (r) or pyrimidine (y) are marked.

<sup>a</sup>shown here 5’ to 3’, (+) sense
rescue the RSV, PVM and hMPV mini-genomes was also investigated. The APV M2-1 protein hypothesised functional domains were also investigated utilizing a set of RSV/APV M2-1 protein chimeras.

5.2 Effect of APV/PVM and APV/RSV heterologous helper plasmid sets on protein expression in the APV plasmid rescue system

Plasmids containing the RSV or PVM N, P, L or M2-1 genes (provided by Dr. H. Stokes, Dr. A. Marriott and Dr. O. Dibben) were individually substituted into the APV PRS and the effect on protein production from the pCAT APV minigenome assessed. The RSV helper plasmids have been successfully used to rescue RSV minigenome (Marriott and Easton, 1999) and were provided by Dr. A. Marriott for use in these experiments. The PVM helper plasmids have also been fully assessed and found to be functional in the PVM plasmid rescue system (Dibben, 2006; Dibben and Easton, 2007) and were provided by Dr. O. Dibben for use in these experiments. Experiments were carried out as described in section 3.3, with substitutions and variations in amount of heterologous plasmid transfected as indicated.

The RSV helper plasmids were investigated for their ability to functionally replace the equivalent APV helper plasmids. As can be seen in figure 5.1, the RSV N plasmid at concentrations of 0.05, 0.1, or 0.4 µg/well failed to support minigenome rescue at levels significantly above the control containing no RSV N plasmid. Substitution of RSV P and L plasmids at 0.2 and 0.4 or 0.1 and 0.3 µg/well, respectively, gave levels of CAT protein production that were not significantly higher than the respective negative controls (i.e. when of no P or no L plasmid was added). The RSV minigenome was found to have some activity in the APV PRS, however the level was only 0.27% of that measured using the APV minigenome. As the M2-1 protein was not absolutely required for APV transcription (section 3.3.2), the high levels of protein production in the absence of M2-1 protein were expected and substitution of the RSV M2-1 plasmid at either 0.01 or 0.025 µg/well gave levels of CAT protein production that were not significantly above those seen in the absence of any M2-1 protein.
Figure 5.1 Effect of heterologous substitution of RSV plasmids. The RSV N, P, L, RSV pCAT minigenome plasmid (A) or the M2-1 plasmid (B) was substituted for the APV analogue in the APV plasmid rescue system and compared to the levels of CAT protein achieved using a homologous set of helper plasmids and the APV pCAT minigenome plasmid. The CAT protein levels for the positive control (‘All APV plasmids’) that has no RSV plasmid substitution and negative controls of no plasmid for each of the helper plasmids were measured. The mean and standard error of two to four independent experiments was taken and plus or minus one standard error utilized for error bars.
The PVM helper plasmids were also investigated for their ability to functionally replace the equivalent APV helper plasmids. As can be seen in figure 5.2, the addition of the PVM N plasmid at concentrations of 0.05, 0.1, or 0.4 µg/well failed to support minigenome rescue. Substitution of PVM P and L plasmids at 0.2 and 0.4 or 0.1 and 0.3 µg/well, respectively, gave levels of CAT protein production that were not significantly higher than the negative controls. The PVM minigenome was found to have some activity in the APV PRS, however the level was only 1.7% of that measured using the APV minigenome. As the M2-1 protein was not essential for APV transcription, the high levels of protein production in the absence of M2-1 protein were expected and substitution of the PVM M2-1 plasmid at either 0.01 or 0.025 µg/well gave levels of CAT protein production which were not significantly above those seen in the absence of any M2-1 protein.

Thus, none of the PVM or RSV helper proteins were found to be similar enough in protein or RNA binding ability to be able to functionally replace the equivalent APV proteins in the APV PRS. Interestingly, both PVM and RSV minigenomes showed very low levels of activity, supporting previous findings that nucleocapsid complexes of related viruses can support replication and transcription of another virus in the same genus (Buchholz et al., 1999; Curran and Kolakofsky, 1991; Govindarajan et al., 2006; Halpin et al., 2004; Pelet et al., 1996; Theriault et al., 2004; Yunus et al., 1999).

5.3 Effect of APV and hMPV heterologous helper plasmid sets on protein expression in the hMPV plasmid rescue system and APV plasmid rescue system

The ability of the closely related APV and hMPV plasmid rescue system components to functionally replace each other was assessed by substituting the heterologous N, P, L, M2-1 or pCAT minigenome plasmid into the APV or hMPV plasmid rescue system. The hMPV plasmid rescue has previously been established (Herfst et al., 2004) and the hMPV pN, pP, pL and pCAT minigenome were the generous gift of Prof. A. Osterhaus. Plasmid rescue experiments were carried out as described in section 3.3, using 0.2 µg/well
Figure 5.2 Effect of heterologous substitution of PVM plasmids. The PVM N, P, L (A), M2-1 plasmid or the PVM pCAT minigenome plasmid (B) was substituted into the APV plasmid rescue system and compared to the levels of CAT protein achieved using a homologous set of helper plasmids and the APV pCAT minigenome plasmid. The mean and standard error of two to four independent experiments was taken.
pN and pCAT minigenome plasmids, 0.1µg/well pP and pL plasmids and 0.01µg/well pM2-1 plasmids with substitutions of heterologous plasmid as indicated.

The result of exchanging the N, P or L plasmids is shown in figure 5.3. As can be seen, heterologous substitution of the N or L plasmids into either the APV or hMPV PRS did not result in CAT protein production significantly above the negative control samples, labelled as either ‘Plasmid absent from hMPV PRS’ or ‘Plasmid absent from APV PRS.’ Interestingly, the hMPV P protein did show activity in the APV PRS, although at a level 0.6% of that seen using the APV P plasmid (‘All APV… plasmids’ column, figure 5.4).

![Figure 5.3 Effect of heterologous substitution of APV and hMPV N, P or L plasmids. APV and hMPV heterologous substitution of N, P or L gene plasmids was carried out. The average and standard error three independent experiments for each sample were taken. A negative control was included for each sample set, labelled as ‘Plasmid absent from hMPV PRS’ or ‘Plasmid absent from APV PRS’.

This level of activity was also unexpected as the P proteins of APV and hMPV have the lowest levels of amino acid identity of all the helper proteins, at 55%. Also, experiments with hMPV and APV subgroup C showed that an hMPV full-length cloned genome could
be rescued with one substitution in the helper plasmid set at 23%, 5%, and 6% with APV subgroup C L, N, and P plasmids, respectively. Conversely, a cloned APV subgroup C full-length genome could be rescued to 28%, 10%, and 12% using hMPV L, N, and P plasmids, respectively (Govindarajan et al., 2006). In both cases the heterologous P plasmid was least able to function in the full-length rescue system, unlike the APV subgroup A P plasmid used here in the PRS. This may be due to the high level of amino acid identity between the APV subgroup C P protein and the hMPV P protein (67%) (Govindarajan et al., 2006).

Rescue of the hMPV pCAT minigenome plasmid using APV helper plasmids and rescue the APV pCAT minigenome plasmid using hMPV helper plasmids was also carried out. As can be seen in figure 5.4, CAT protein production was obtained. Although not able to fully replace the equivalent minigenome, rescue using the hMPV and APV heterologous minigenome plasmids generated 29% (18000ng of CAT for heterologous rescue versus 62000pg for homologous rescue) and 26% of the CAT protein of the homologous plasmid rescue experiments (4000ng of CAT for heterologous rescue versus 15000pg for homologous rescue). This was not surprising given the high level of homology of the cis-acting sequences in the APV subgroup A and hMPV genomes and the finding that an hMPV full-length cloned genome was rescued by a set of APV subgroup C helper plasmids to 52% virus recovery. Conversely, rescue of an APV subgroup C full-length cloned genome was recovered with hMPV helper plasmids at 47% of homologous virus recovery (Govindarajan et al., 2006).

The action of the heterologous M2-1 plasmid was tested in the hMPV and APV plasmid rescue systems. As shown in figure 5.4, the M2-1 plasmid was not able to functionally replace the homologous plasmids as not significantly more CAT protein was produced from these samples than in the absence of any M2-1 protein in the system. This result was consistent with the results of experiments using the full-length rescue system where virus was rescued using a heterologous M2-1 plasmid (Govindarajan et al., 2006).
5.4 Functionality of M2-1 APV/RSV chimeras in the APV plasmid rescue system

It was not expected that the RSV helper plasmids could replace the APV plasmids, as the percentage identity between their N, P, L and M2-1 genes was low (24 to 43%, section 5.1). This property was utilized to create chimeric RSV/APV N protein and chimeric

![Graph](image)

**Figure 5.4 Effect of heterologous substitution of APV and hMPV minigenome or M2-1 plasmids.** APV and hMPV heterologous substitution of M2-1 gene and minigenome plasmids was carried out. The average and standard error were taken from three independent experiments. A positive (no substitution) and a no M2-1 plasmid control were included in each plate.

RSV/APV M2-1 proteins which were used to assess the functionality of a sub-section of the RSV proteins (Stokes, 2003; Stokes et al., 2003; Zhou et al., 2003). This strategy was found to be useful as truncated protein can lack the structure of a complete protein and be mis-folded even if they have retained essential residues.
Investigation of the function of regions of the RSV M2-1 protein was initially carried out by excluding the M2-1 protein from the RSV PRS or by deleting the M2-1 gene start signal from full-length cloned genomes (Collins et al., 1995; Fearns and Collins, 1999; Mason et al., 2003). These studies established the M2-1 protein as a transcriptional elongation and anti-termination factor that was necessary for efficient transcription of mRNAs downstream of the first ORF of either the RSV mini-genome or full-length genome. The RSV M2-1 protein has also been shown to bind P protein, RNA and to activate Rel A (of the NFκβ family) (Cartee and Wertz, 2001; Mason et al., 2003; Reimers et al., 2005). Stokes (2003) carried out an initial investigation of the functional regions of the RSV M2-1 gene by creating three deletion mutants of the M2-1 expression plasmid which were truncated at amino acid 44, 92 or 153 (figure 5.5). In summary, the resulting truncated M2-1 proteins were tested for their ability to enhance mRNA production in the RSV PRS and it was found that these mutants had lost all ability to enhance mRNA production in the PRS. Additionally, it was shown that the APV M2-1 protein was not able to functionally replace the RSV M2-1 protein in the RSV PRS.

![Figure 5.5 RSV M2-1 protein deletion mutants.](image)

Figure 5.5 RSV M2-1 protein deletion mutants. Shown here is a diagrammatic representation of RSV M2-1 deletion mutant proteins. Amino acids were deleted from the carboxyl terminus and numbers represent amino at the start and end of the protein.

A chimeric protein strategy was also attempted (Stokes, 2003; Stokes et al., 2003), whereby sections of the RSV gene sequence were replaced with the equivalent APV M2-1 sequence, as shown in figure 5.6. Regions of conserved sequence (three consecutive
residues) between the APV and RSV M2-1 protein sequences were utilized to divide the protein into four sections. Six chimeric APV/RSV M2-1 protein expression plasmids were created, as shown in figure 5.6, and shown to produce proteins of the expected size (Stokes, 2003)(Dr. A. Marriott, personal communication).

![Diagram of APV/RSV chimeric M2-1 proteins]  

**Figure 5.6 Composition of APV/RSV chimeric M2-1 proteins.** A: Diagrammatic representation of APV/RSV M2-1 chimeric proteins. Blue sections represent amino acids derived from the RSV M2-1 protein and gray sections represent amino acids derived from the APV M2-1 protein. B: The protein residue composition of each chimera is shown. The residues from either the APV (in bold) or RSV M2-1 proteins which comprise Sections I-IV are presented under ‘Section composition’.

In this study, the APV/RSV M2-1 chimeras were tested in the APV PRS to determine if sub-sections of the APV M2-1 sequence could enhance protein production from the APV pCATLUC minigenome. The RSV M2-1 protein was shown to be unable to functionally
replace the APV M2-1 protein in the APV PRS (section 5.2, figure 5.1). Introduction of chimeras 1 to 6 into the PRS at 0.01µg/well as described in section 3.3 was carried out and the levels of CAT and LUC protein measured as described in section 2.9 to 2.11. Controls of APV M2-1 plasmid, RSV M2-1 plasmid and no M2-1 plasmid were included, as the APV PRS produces high levels of reporter protein even in the absence of the M2-1 plasmid. For each independent experiment, the protein values were calibrated against the control of the levels of CAT and LUC activity seen in the absence of the M2-1 plasmid and this value was set at 100%. The standard error of the means of five independent experiments was utilized for the error bars and a negative control of transfection lacking the L plasmid was included for each experiment (figure 5.7).

As can be seen in figure 5.1 and 5.8, the control experiment of addition of the RSV M2-1 plasmid in place of the APV M2-1 plasmid in the APV PRS was unable to enhance reporter protein expression. The APV M2-1 protein was also confirmed to be non-essential in the APV PRS. The addition of 0.01µg/well of APV M2-1 plasmid increased CAT protein production by more than four-fold and LUC relative light activity levels by more than three-fold above no M2-1 levels, confirming the results of the experiments in section 3.3.2. Substitution of chimeric M2-1 plasmids 2, 3, 4 failed to produce levels of CAT protein significantly above the levels seen in the absence of any M2-1 plasmid. Substitution of chimeric M2-1 plasmids 5 and 6 produced low (6-20% over no M2-1 control) levels of CAT protein. Chimera 1, however, showed enhancement of both CAT and LUC protein production, with two-fold and 3.5 fold increases, respectively. Substitution of chimeric M2-1 plasmids 2, 3 and 6 produced low (6-20% over no M2-1 control) levels of LUC protein and substitution of M2-1 chimeras 4 and 5 failed to produce levels of LUC protein significantly above the levels seen in the absence of any M2-1 plasmid. Hence, while none of the M2-1 chimeric plasmids was able to fully replace the APV M2-1 plasmid in terms of the 3’ genome proximal gene (the CAT gene) transcription enhancement, the addition of 0.01µg/well of a M2-1 gene that has the first
Figure 5.7 The effect of APV/RSV M2-1 chimeric protein on CAT reporter protein expression in the APV PRS. Each well was transfected with 0.2µg of pN10 and pCATLUC, 0.1µg of pP9 and pCITE L. For chimeric substitutions, 0.01µg/well of chimeric M2-1 plasmid was added. ‘RSV M2-1’ indicates that the RSV M2-1 expression plasmid was added to the transfection mixture in place of the APV M2-1 plasmid. The APV (+), a positive control, had 0.01µg of APV pM2-1 added to the transfection mixture. APV (-), a negative control, included the pN10, pP9 and pM2-1 in the transfection mixture but lacked the pCITE L plasmid. All transfection mixtures were made up to 1µg/well with pT7.2 vector plasmid. CAT protein expression (A) and LUC protein expression (B) were measured 48hr post-transfection. The standard error was determined from three to five independent replicates.
44 residues substituted with RSV M2-1 sequence was able to enhance the second (LUC) gene transcription to wildtype levels.

The high level of activity of chimera 1 but not the other chimeras suggests that section II along with either section III and/or section IV (figure 5.6) of the M2-1 protein are necessary for transcriptional enhancement. This was supported by the data obtained using chimera 4, which contains section I but not sections II, III or IV, and which was the only chimera to give no enhancement of transcription of either the CAT or the LUC gene. No chimera was constructed which contained sections II and IV, however chimera 6 contained both sections II and III; this chimera gave low levels of both CAT and LUC protein expression.

Chimera 5, with sections I and II, stimulated a very low level of CAT production (110% +/-9%), but not significantly more than chimera 4 (100% +/-4) and neither gave significant LUC production. This suggests that sections I or sections I and II are not sufficient to enhance protein production in the APV PRS. Chimeras 2 and 3 did not produce enhancement of CAT production, but did (at low levels) enhance LUC production, suggesting that section IV may have a small role in second gene transcription enhancement.

5.5 Discussion
Heterologous substitution, the substitution of genes from other viruses, into reverse genetics systems has been used extensively to determine the level of functional relatedness between viruses, to create vaccine candidate viruses and to study the functional regions of the proteins essential to virus genome transcription and replication (Bailly et al., 2000; Biacchesi et al., 2005; Buchholz et al., 2000; Clements-Mann et al., 2001; Curran and Kolakofsky, 1991; Dimock and Collins, 1993; Durbin et al., 2000; Govindarajan et al., 2006; Halpin et al., 2004; Muhlberger et al., 1999; Newman et al., 2002; Pelet et al., 1996; Pham et al., 2005; Schmidt et al., 2000; Schmidt et al., 2001; Skiadopoulos et al., 2003; Stokes, 2003; Stokes et al., 2003; Theriault et al., 2004; Yunus et al., 1999). The ability of proteins to functionally substitute between viruses would be
most expected between closely related viruses. For example, full-length rescue of hMPV was achieved using a set of APV subgroup C helper plasmids and an APV subgroup C synthetic genome was rescued by a set of hMPV helper plasmids (Govindarajan et al., 2006). The authors showed that the polymerase complex of either virus could recognize the cis-acting signals in the heterologous genome, indicating a close functional relationship between these viruses. However, while hMPV and APV subgroup C have a high level of identity between their leader regions (65% nucleotide identity), trailer regions (72% nucleotide identity) and N, P, L and M2-1 proteins (N proteins 89% amino acid identity, P proteins 68%, M2-1 proteins 85%, L proteins 80%) (Govindarajan et al., 2006; Govindarajan and Samal, 2005), APV subgroup A has less identity with hMPV both in the amino acid identity of its protein or in the nucleotide identity of its non-translated regions (section 5.1).

Based on a ‘Jennerian’ vaccination approach, heterologous substitution of closely related virus proteins could introduce a level of attenuation of replication to the virus without loss of protection (Bailly et al., 2000). A similar approach was used to create a live-attenuated vaccine candidate for hMPV by substituting the APV P gene into a hMPV genome. This resulted in a virus with the unique and desirable phenotype of having increased growth in cell culture but being attenuated in vivo (Pham et al., 2005).

The results presented here show that substitution of the hMPV pP or pN plasmids into an APV subgroup A background retained only very low levels of ability to support minigenome protein expression (figure 5.3), as did the hMPV minigenome plasmid (figure 5.4). Interestingly, while the APV minigenome plasmid could also function in the hMPV PRS, none of the other APV helper plasmids gave levels of activity in the hMPV PRS significantly above the negative controls. This could be due to greater fidelity of the APV RNP complex proteins, such that a change to the N, P or L proteins interferes with the protein:protein binding required to form the RNP complex. Compared to experiments using the rescue systems where a hMPV full-length genome was shown to be rescued to varying degrees by sets of hMPV and APV subgroup C RNP complex proteins (Govindarajan et al., 2006) and to produce viable virus with substitutions in the genome
of the APV subgroup C N or P genes (Pham et al., 2005), the APV subgroup A was less able to functionally replace sequence from hMPV. Generally however, it follows that if heterologous substitution in PRS experiments results in a lower level of functionality, it may be a useful substitution to consider when creating a live attenuated vaccine strain of a virus. However, as this may result in over-attenuation, it was useful to have a selection of mutations of known effect to choose from.

Heterologous substitutions in rescue systems have been used extensively to investigate virus relatedness and protein function. Heterologous rescue of the human parainfluenza virus type 1 strain Washington/20993/1964 (HPIV1 WASH/64) cloned genome was successfully carried out using HPIV3 helper plasmids or a mixture of HPIV1 and HPIV3 helper plasmids. The amino acid identity between these helper proteins varies from 62% and 61% for the N and L proteins, respectively, to only 29% for the P proteins. These viruses also have extensive nucleotide homology between their cis-acting sequences. This indicates a high level of functional activity between these proteins- the HPIV3 N can recognize the HPIV1 genome and the HPIV3 L:P polymerase complex can act on this template to produce functional RNP complex and anti-genome RNA. More interestingly, the data shows that the HPIV3 L protein can recognize a template bound with HPIV1 N and P protein. Newman et al. (2002) used heterologous rescue to confirm that the well conserved sequences in P proposed to be responsible for L:P binding can act in the context of an otherwise poorly conserved protein. Rescue of a Sendai virus (SeV) minigenome cDNA by PIV3 N, P and L proteins was also shown to be possible, confirming that with an N protein which was from the same virus as the polymerase complex (L:P) recovery was possible of a heterologous RNA genome. This suggests that protein: protein binding (ie P:N) was virus specific whereas protein:RNA binding (N:RNA) can be heterologous for PIV3 and SeV (Curran and Kolakofsky, 1991; Pelet et al., 1996). Similar results were achieved using the Zaire Ebola virus full-length rescue system with heterologous helper plasmids from the closely related Reston Ebola virus (71-89% overall amino acid identity) or the more distantly related filovirus Marburg virus (46-58% amino acid identity). This showed again the importance of protein:protein interactions over protein:RNA interactions in the RNP complex (Theriault et al., 2004).
Chimeric proteins have been used to evaluate the functional regions of the RSV helper proteins in the RSV PRS. Stokes (2003) and Marriot (personal communication) substituted the APV/RSV chimeras into the RSV PRS at 0.025, 0.05, 0.1, 0.2, 0.4 and 0.8µg/well. As shown in figure 5.7, none of the chimeras restored reporter protein production to the level seen when RSV M2-1 was included at 0.05µg/well. APV/RSV chimera 3 gave the highest levels of both reporter proteins, with a high (>60%) amount of CAT protein production when 0.4 and 0.8µg/well of chimeric plasmid were used. APV/RSV chimeras 1 and 2 showed moderate levels of CAT expression (40-59% of wildtype) at 0.8µg/well of chimeric plasmid. Chimeras 4 and 5 gave low levels of CAT production (20-39% of wildtype) at 0.2, 0.4 and 0.8µg/well. Chimera 6, with only the final 154-194 bases from the RSV M2-1 protein, showed no activity above the negative control in which no M2-1 plasmid was added to the transfection. Luciferase activity in the presence of these chimeras was even lower than that seen with CAT assay, with the greatest enhancement being 20% of that seen with RSV M2-1 protein for chimera 3 at 0.8µg/well plasmid; however, enhancement was seen across the range of concentrations added. Chimera 5 gave moderate (3-19%) levels of LUC protein production, reaching 7% at 0.8µg/well added. Chimeras 2 and 4 gave low (<5%) levels of LUC expression, even at 0.8µg/well of plasmid added. Chimeras 1 and 6 gave no enhancement of LUC expression above the negative control of no M2-1 plasmid added. From this data, it was proposed that the RSV M2-1 protein had two functional domains, one that enhanced production of the leader proximal gene (here the CAT gene) between residues 1-44 and a second transcriptional enhancer for both genes between residues 93-152 (Stokes, 2003)(Dr. A. Marriott, personal communication). It would be expected that the chimera proteins might show sections that were important for both the APV M2-1 and the RSV M2-1 protein: protein interactions. However, as the proteins are different lengths and have low levels of amino acid identity (39%), it is not surprising that different regions will be important for different virus protein interactions.

Protein functionality studies on the bRSV N protein have shown any amino acid deletion to be very detrimental to N protein activity and deletion of certain residues in the C-
N-termini also obliterated N protein function (Khattar et al., 2000). Further studies utilizing RSV/PVM N chimera proteins showed that only six residues at the C-term end of the protein were essential for viral replication and N-P protein binding (Stokes et al., 2003). With the APV M2-1 protein, initial studies of heterologous substitution into the RSV PRS found that the APV M2-1 protein could not functionally replace the RSV M2-1 protein. It was also shown that truncated versions of the RSV M2-1 protein did not enhance reporter protein production in the RSV PRS. Subsequently, the APV M2-1 protein sequence was utilized to create a chimeric protein which could be used to investigate the functionality of regions of the RSV M2-1 protein. The functionality of APV/RSV M2-1 chimeras 1 to 6 (figure 5.6 and 5.7) in the RSV PRS was determined by Dr. H. Stokes and Dr. A. Marriott. They found that the APV/RSV M2-1 chimera 3

![Percentage (%) CAT or LUC protein production](chart.png)

**APV/RSV M2-1 chimeric plasmids (0.8µg/well)**

**Figure 5.8 The effect of APV/RSV M2-1 chimeric protein on CAT and LUC reporter protein expression in the RSV PRS.** The ability of the chimeric proteins to direct transcription was measured by calculating the level of CAT or LUC protein production in the presence of 0.025 to 0.8µg/well of chimeric M2-1 protein as a percentage of the levels of protein produced with 0.05µg/well of the RSV M2-1 plasmid. Shown here is the maximum value achieved when adding chimeric plasmids at 0.025 to 0.8µg/well (adapted from Stokes, 2003)(Dr. A. Marriott, personal communication).
retained the ability to produce both CAT and LUC reporter proteins from a RSV pCATLUC mini-genome. Chimeras 1 and 2 enhanced CAT protein production but failed to enhance LUC production. Chimeras 4 and 5 gave increased levels of both reporter proteins, but not to the extent seen for chimera 3. Chimera 5, which has only the final 154-194 bases from RSV, failed to enhance expression of either reporter protein. Hence two functional regions were identified, the N-terminal 44 residues for production of the leader-proximal gene (the CAT gene) and residues 93-152 for expression of both genes in the RSV pCATLUC minigenome (Stokes, 2003).

Using the same chimeras in the APV PRS, only chimera 1 gave any level of functionality. This was surprising, as chimera 1 lacks the first 44 residues that were shown to be important function in the RSV PRS. Chimera 1 contained the first 1-44 residues derived from the RSV M2-1 protein and the rest from the APV M2-1 protein. If the first 1-44 residues were essential for APV transcription enhancement, chimeras 4-6 should have some function, but not chimeras 1-3. The functionality of chimera 1 but not chimeras 2 or 3 implies that section II was essential for M2-1 protein transcriptional enhancement activity; however chimeras 5 and 6 that retain this section are non-functional. Non-functionality of chimeras 4-6 suggests that section IV was essential for efficient transcription from the minigenome. However as chimeras 2 and 3 (which retain this section) are non-functional, the data suggest that both sections II and IV are needed for M2-1 protein transcriptional enhancement activity. The limited functionality of any of the chimeras apart from chimera 1 suggests that APV M2-1 has different functional regions than the RSV M2-1 protein and that these regions may be non-consecutive. Also, although the amount of M2-1 chimera was appropriate for the APV PRS, it may be that over-expression of the protein would result in higher levels of activity and this would be interesting to investigate.

This work lays the foundation for future experiments investigating the functionality of the N, P, L and M2-1 by creating chimeric proteins with hMPV, PVM, or RSV. By using sequence comparison, possible functional regions can be identified and chimeras created to test mutations which might result in complete loss of function if just deleted from the
gene sequence. Also, as has been show for hMPV and APV subgroup C, the substitution of APV subgroup A genes into the hMPV full-length genome might result in viable vaccine candidates and increase the options available for varying the levels of attenuation of a vaccine candidate.
Chapter 6:
Rescue of APV from cDNA clones
6.1 Introduction

The ability to create a virus entirely from cDNA is a powerful tool and is the ultimate goal of reverse genetics studies. Using a cloned copy of the viral genome and helper plasmids, it is possible to produce a virus entirely from cDNA (section 1.5). For the pneumovirus subfamily, viruses created from cloned DNA have been created for APV (Govindarajan et al., 2006; Naylor et al., 2004; Naylor et al., 2007), RSV (Buchholz et al., 1999; Collins et al., 1995), PVM (Krempl et al., 2007) and hMPV (Biacchesi et al., 2004a; Biacchesi et al., 2004b; Herfst et al., 2004). At least one virus from most of the families of non-segmented negative sense RNA viruses have been created from cloned DNA. Viruses have been created for many of the members of the Paramyxoviridae family (in addition to the pneumoviruses above) including Sendai virus (Garcin et al., 1995), measles virus (Radecke et al., 1995), rinderpest virus (Baron and Barrett, 1997), human parainfluenza virus type 3 (Durbin et al., 1997; Hoffman and Banerjee, 1997), simian virus 5 (Karron et al., 1997), Nipah virus (Yoneda et al., 2006), canine distemper virus (Gassen et al., 2000), and Newcastle disease virus (Krishnamurthy et al., 2000).

Amongst the Rhabdoviridae family viruses, rescue systems have been described for rabies virus (Schnell et al., 1994) and vesicular stomatitis virus (Lawson et al., 1995; Whelan et al., 1995). Full length rescue systems (FLRS) in which infectious virus is recovered have been developed for Marburg virus (Enterlein et al., 2006) and Zaire ebolavirus of the Filoviridae family (Theriault et al., 2004), as well as for the Bornaviridae family virus borna disease virus (Perez et al., 2003). The ability to create a virus entirely from cloned cDNA has become an essential instrument for the study of negative sense RNA viruses.

FLRS are being used for a wide range of studies of viruses, including creation of rationally designed vaccine candidates (Biacchesi et al., 2005; Biacchesi et al., 2004b; Buchholz et al., 2005; Cheng et al., 2005; Cheng et al., 2001; Collins et al., 1995; Hoffman and Banerjee, 1997; Jin et al., 1998; Skiadopoulos et al., 2004), determination of the function of naturally occurring mutations isolated from field strains (Naylor et al., 2007), investigation of protein function (Bermingham and Collins, 1999; Buchholz et al., 2005; Buchholz et al., 1999; Bukreyev et al., 1997; Cheng et al., 2005; Cheng et al., 2005; Cheng et al., 2005; Cheng et al., 2005; Cheng et al., 2005).
Collins et al., 1995; Jin et al., 2000a; Jin et al., 2000b; Naylor et al., 2007; Skiadopoulos et al., 2004; Techaarpornkul et al., 2001; Teng et al., 2000; Theriault et al., 2004), and determination of the ability of the virus to be used as a vector (Govindarajan et al., 2006; Karron et al., 1997; Skiadopoulos et al., 2004). The goal of rational vaccine design is to create vaccines by introducing a small number of attenuating mutations in order to fine-tune the balance of attenuation and protection of a candidate-vaccine strain (Buchholz et al., 2006; Collins and Murphy, 2005; Collins et al., 1999b). The individual point mutations found to produce cold-passaged, temperature sensitive, or naturally occurring mutations have been analysed for their ability to attenuate the virus (Firestone et al., 1996; Karron et al., 1997; Naylor et al., 2007; Patnayak and Goyal, 2004; Patnayak and Goyal, 2006; Patnayak et al., 2003). Deletion mutations have also been generated which would not be likely to occur naturally—such as whole gene deletions—and hence would be unlikely to revert to virulence (Biacchesi et al., 2005; Biacchesi et al., 2004b; Bukreyev et al., 1997; Karron et al., 1997; Ling et al., 2008; Naylor et al., 2004; Whitehead et al., 1999). The addition of foreign genes to a virus genome has also allowed multivalent vaccine-candidate strains to be produced (Buchholz et al., 2006; Collins and Murphy, 2005; Durbin et al., 2000; Schmidt et al., 2001). Insertion of a reporter gene, such as green florescent protein (GFP), into the viral genome can serve as a marker of recombinant virus production and expression of GFP by a recombinant virus also allows visualization of virus in vitro and in vivo, thereby increasing the speed of analysis and production of slow growing viruses (Govindarajan et al., 2006; Techaarpornkul et al., 2001).

By analogy to RSV, APV has three glycoproteins expressed on its surface; the F (fusion) protein, the G (glycoprotein) protein and the SH (small hydrophobic) protein. While the roles of the F and G proteins have been well defined (sections 1.3.2.5 and 1.3.2.6), the function of SH protein is unknown. Full-length cloned RSV, hMPV and APV viruses have been created which lack one or more of the viral glycoproteins (Biacchesi et al., 2005; Biacchesi et al., 2004b; Bukreyev et al., 1997; Karron et al., 1997; Ling et al., 2008; Naylor et al., 2004; Techaarpornkul et al., 2001; Whitehead et al., 1999). For RSV, deletion of the SH gene from the genome resulted in no change in phenotype and
showed that the SH protein was not required for viral infectivity (Techaarpornkul et al., 2001). Teng and Collins (1998) showed that the F gene was required for the production of infectious particles and production of infections virus from an F-deleted genome has only been possible by complementing the genome with a F protein producing cell line (Oomens and Wertz, 2004b) or substituting an analogous glycoprotein gene in place of the F gene (Buchholz et al., 1999; Cartee et al., 2003; Oomens and Wertz, 2004a; Stope et al., 2001). The RSV G gene was not required for the production of infectious virus (Teng and Collins, 1998; Wertz and Moudy, 2004). The deletion of the hMPV SH gene from the genome also showed the SH protein to be non-essential in vitro and in vitro (Biacchesi et al., 2004a; Biacchesi et al., 2004b). However, for APV, an SH gene deleted virus produced a syncytial plaque phenotype, suggesting a role in regulation of the cell-cell fusion process. Recombinant APV lacking an SH or G gene were also shown to replicate poorly in turkeys compared to wildtype APV (Ling et al., 2008).

Green florescent protein (GFP), derived from the *Aequorea Victoria* jellyfish, is a frequently used reporter gene in reverse genetics studies. Enhanced GFP (eGFP) has been developed to give greater fluorescence and increased expression in mammalian cells (Chalfie et al., 1994; Inouye and Tsuji, 1994; Prasher et al., 1992). The initial aim of the work described in this chapter was to insert an eGFP ORF either in place of the SH gene or after the SH gene in the CF2 APV cloned genome (Naylor et al., 2007). These recombinant viruses would then be used to determine sites of replication in vivo and to follow virus replication in vitro.

### 6.2 Production of GFP-expressing APV

The APV clones described here were based on those constructed by Dr. R. Ling (Ling et al., 2008). A modified version of this genome, the CF2 genome, was later constructed as a semi-cassotted APV subgroup A cloned genome. The CF2 APV genome construct differs from the APV subgroup A isolate CVL14/1 virus by the insertion of restriction enzyme sites into the intergenic regions between the genes (all gene junctions except the M-F and F-M2 junctions) in order to facilitate cloning (appendix B) as this modification was found to increase the recombinant virus growth rate compared to the fully cassotted
The CF2 APV genome was utilized to create a recombinant APV genome containing an eGFP gene inserted before the SH gene. In this way, the virus could be visualised in cell culture and in turkey organs (figure 6.1).

**Figure 6.1 Full length APV cloned genomes.** The APV full-length cloned genome pCF2 (first, top to bottom) was modified such that a gene encoding eGFP was additional gene inserted between the M2 and SH genes (second) or substituted for the SH gene (third and fourth). G* indicated that this gene may not be expressed, as the eGFP GE sequence was deleted during cloning (third).

### 6.2.1 Cloning of the eGFP full-length genome

In order to create a cloned APV virus that expressed GFP, the GFP ORF was cloned into the CF2 genome. The cloning strategy used for inserting the GFP gene into the M2-SH gene junction was carried out as described below. The CF2 genome containing the inserted eGFP gene was named CF2ESH. Due to the placement of restriction sites in the genome, the cloning of the CF2ESH genome was done via an intermediate genome that
lacked the SH ORF (named genome CF2EdSH) (figure 6.2). Also shown in figure 6.2, and discussed later in section 6.3, is a genome created whilst attempting to create the

CF2EdSH genome which lacks a functional G gene, called genome CF2EdSH(dG). The mutant genome CF2EdSH 34 was created first by removing the SH gene and replacing it

Figure 6.2 Generation of pCF2 mutants- overview of cloning strategy. The CF2 genome (A) was digested with restriction enzymes AgeI and EagI (as indicated with yellow and red stars) and ligated with a PCR generated eGFP fragment (B) that had been similarly digested. The ligation produced the CF2EdSH genome (C), which was then digested with restriction enzymes EagI and SmaI (as indicated with red and green stars) and ligated with a PCR generated SH fragment (D) that had been similarly digested. This ligation produced the CF2ESH genome (F).
with the eGFP gene fragment. To insert a copy of the eGFP gene in place of the SH gene whilst retaining the SH transcription signals, primers containing AgeI and EagI enzyme restriction sites along with the SH gene start and gene end sequences were used to create an eGFP fragment. The plasmid pEGFP-C1 (BD Biosciences Clontech, GenBank Accession #: U55763) was used in a PCR (section 2.5.2) with primers FPSHP and FPHMES (appendix A). The product (figure 6.3) was purified (section 2.5.3.1) and the DNA digested with AgeI enzyme. Following purification, the fragment was digested with the enzyme EagI and the fragment purified (figure 6.3).

This fragment was then ligated into the similarly digested and purified CF2 genome DNA (figure 6.4). The ligation product was transformed into E. coli, producing 70 colonies. The recombinant bacteria were grown at 30°C over night, plasmid DNA was purified from all 70 colonies (section 2.6.2), and digested with EcoRV and EagI enzymes to provide an initial screen for the presence of the eGFP insert. Preparations 34-36, 47-49, 56 gave the expected fragments of 1409 and 15118bp, whereas the vector plasmid alone produced a single fragment of approximately 15kbp (figure 6.5). These preparations were further purified, sequenced with primer TRT22K2 (appendix A), and preparations 56 and 34 were found to contain the eGFP insert. The insertion regions of preparations CF2EdSH 56 and CF2EdSH 34 were fully sequenced and found to have the following unexpected mutations:

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Bases</th>
<th>Mutation</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF2EdSH 34</td>
<td>8104</td>
<td>deletion</td>
<td>eGFP-G intergenic region</td>
</tr>
<tr>
<td>CF2EdSH(dG) 56</td>
<td>8073 to 8080</td>
<td>deletion</td>
<td>eGFP GE sequence</td>
</tr>
<tr>
<td>CF2EdSH(dG) 56</td>
<td>8085 to 8240</td>
<td>deletion</td>
<td>eGFP GE sequence</td>
</tr>
<tr>
<td>CF2EdSH(dG) 56</td>
<td>8870</td>
<td>A→C</td>
<td>M2-eGFP intergenic region</td>
</tr>
<tr>
<td>CF2EdSH(dG) 56</td>
<td>8894</td>
<td>A→C</td>
<td>M2-eGFP intergenic region</td>
</tr>
</tbody>
</table>

The CF2ESH genome was created by inserting the SH gene into the eGFP-G intergenic region of the CF2EdSH 34 genome. The SH gene was amplified by PCR from the cassetted APV genome using the primers SHSmaIU and GEagIL, as above. The SH PCR
Figure 6.3 Generation of eGFP fragment for insertion into the CF2 genome. PCR was carried out on the plasmid pEGFP C1 (Clontech) using primers FPSHP and FPSHMES (A). PCR produced an eGFP fragment flanked by *Age* I and *Eag* I restriction digest sites, as well as the SH GE and GS sequences. The product was examined on an agarose gel (B) and the 846bp eGFP gene fragment (lane 1, arrow) was purified. The eGFP fragment was then digested with enzymes *Age* I and *Eag* I to produce an 832bp fragment (C, lane 1, arrow).
Figure 6.4 Generation of genome vector for insertion of the eGFP gene into the CF2 genome. The genome CF2 (A) was digested with AgeI and EagI enzymes. The digestion product (B) was examined on an agarose gel (C, lane 1). As indicated with an arrow in lane 1, sequential restriction digestion and purification produced a single fragment larger than 10kbp.
product was purified and digested with restriction enzymes Smal and Eagl (figure 6.6). The CF2EdSH 34 vector DNA was similarly digested and purified (figure 6.7), ligated with the SH insertion fragment and transformed into E. coli. Small-scale preparation of plasmid DNA was carried out from 32 colonies and the DNA was digested with EcoRV and EagI, as above (figure 6.8). Preparation CF2ESH 11 was shown to have the correct digestion pattern and the insertion region of this plasmid was fully sequenced. This region matched the desired sequence, except for the following deletions in the eGFP-SH intergenic region (genome sense of CF2ESH 11 genome): a T was deleted at base 8724 and an A was deleted from between bases 8714 and 8715. All mutant virus genome plasmids were amplified poorly in E. coli, so DNA from small scale preparations was

**Figure 6.5 Screening of clones by restriction digestion for substitution of the eGFP gene for the SH gene to create the CF2EdSH genome.** Small scale plasmid preparation was carried out on clones 34 to 59 and these samples were subjected to restriction digestion with EcoRV and EagI enzymes. Preparations with the desired substitution (eGFP gene in place of the SH gene) should have a 1409bp DNA fragment. Preparations 34, 35, 47-49, 56, and 59 (in red) were selected as possible candidates.
amplified using the Illustra™ TempliPhi 100 Amplification Kit (GE Healthcare) following the manufacturer’s protocol.

**Figure 6.6 Generation of SH fragment for insertion into the CF2ESH genome.** PCR was carried out on the CASA genome plasmid using primers SHSmaIU and GEagIL (provided by Dr. R. Ling) (A). PCR produced a SH ORF fragment flanked by EagI and SmaI restriction digest sites, as well as the SH GE and GS sequences. The product was examined on an agarose gel (B) and the 681bp SH gene fragment (lane 1, arrow) was purified. This fragment was then digested with enzymes Agel and EagI to produce a 629bp fragment (C, lane 1, arrow).
Figure 6.7 Generation of genome vector for insertion of the SH gene into the CF2EdSH genome. The genome CF2EdSH (A) was digested with SmaI and EagI enzymes. The digestion product (B) was examined on an agarose gel (C, lane 1). This fragment was then purified to produce a single DNA fragment larger than 10kbp (C, lane 3).
6.2.3 Growth and eGFP expression in cell culture of virus CF2ESH 11

Full-length rescue was carried out as described in section 2.8.3 by transfection of the required plasmids into 6-well plates of BSR-T7 cells (Ling et al., 2008). Transfected cells were examined for eGFP expression after 2 days to check for the production of recombinant virus (figure 6.9). Virus infected cells were harvested (section 2.13.2) at 4 to 6 days post inoculation (p.i.). A 2ml aliquot of this supernatant was then used to infect a small flask of confluent BSC-1 cells at 33°C and the virus was then harvested by freeze/thawing when 80% CPE was present (6 to 12 days p.i.). The virus was ultimately amplified to prepare stocks as described in section 2.13.2. The titre of the virus was determined by examination of GFP expression in infected cells from 10-fold serial
dilution of virus stock in 96 well plates of BSC-1 cells (section 2.14). The titre for virus CF2ESH 11 was determined to be $6.5 \times 10^6$ CIU/ml.

Examination of the eGFP expression of virus CF2ESH 11 was carried out by photographing infected BSC-1 cells using confocal microscopy. BSC-1 cells were grown to confluency in 35mm glass-bottom dishes (IWAKI brand, Asahi techno glass) and were infected with 1ml of virus (MOI of 2.17 cell infecting units (CIU) per cell) for 1hr at 33°C, the medium removed, and the cells overlaid with Leibovitz’s L-15 medium (Invitrogen). Photographs were taken between 5hr and 33hr 20min p.i. every 10min in a heated chamber (33°C) with the 489nm Argon Laser line (figure 6.10). Infection was monitored over 2300min, with the majority of eGFP expression occurring between 300 and 2000min (5hr and 33hr 20min). The expression of eGFP increased in intensity and number of cells infected increased until cell death started to occur (1900 and 2000min, figure 6.10 part III, appendix C).

Figure 6.9 eGFP expression of virus CF2ESH 11 in BSR-T7 cells. BSR-T7 cells were grown to a confluent monolayer and transfected with genome CF2ESH 11, pCite L, pCite M2-1, pCite N and pCite P plasmids. After 48hrs, the cells were examined by microscopy under white light (A) and UV (B) microscope for eGFP expression.
Figure 6.10 part I: Time lapse photography of infection of cells with virus CF2ESH 11, 300 to 800 min. Confluent monolayers of BSC-1 cells were infected with virus CF2ESH 11 at 2.17 CIU per cell. eGFP expression was visualized using confocal microscopy.
Figure 6.10 part II: Time lapse photography of infection of cells with virus CF2ESH 11, 900 to 1400min.
Figure 6.10 part III: Time lapse photography of infection of cells with virus CF2ESH 11, 1500 to 2000min.
6.2.4 CF2ESH 11 virus expression of SH, G and eGFP mRNAs

In order to determine the levels of SH, G and eGFP mRNA being produced from mutant virus CF2ESH 11, northern blots were carried out (section 2.12). Total RNA was extracted from infected BSC-1 cells using a RNeasy® mini kit (Qiagen) following the manufacturer’s instructions 4-6 day p.i. and 1µg of RNA was added to each of the sample lanes. The SH and G gene probes and standards were provided by Dr. R. Ling. The eGFP RNA probe and standards were created (section 2.17 and 2.18) from p(-)GFP and p(+)GFP expression plasmids, respectively, provided by Dr. A. Marriot. The RNA size markers (RNA Millennium Markers, Ambion) were probed with a complementary RNA probe provided by Dr. R. Ling.

As can be seen in figure 6.11 (A), G gene mRNA was expressed from mutant virus CF2ESH 11 and non-mutated virus CF2. Using the G gene standards, virus CF2ESH 11 produced approximately 1x10^8.5 molecules of G mRNA from 1µg of total RNA and virus CF2 produced between 1x10^8.5 and 1x10^8 molecules of G mRNA from 1µg of total RNA. Expression SH RNA was examined (figure 6.11, B) and mutant virus CF2ESH 11 and non-mutated virus CF2 showed detectable levels of production. Using the SH gene standards, both virus produced between 1x10^8.5 and 1x10^8 molecules of SH mRNA from 1µg of total. Examining eGFP RNA (figure 6.11, C), as expected virus CF2ESH 11 alone showed expression. The eGFP standards were larger than the viral eGFP transcripts as they were produced from a DNA template that did not terminate at the end of the eGFP gene. On the basis of the quantity of the standards, the amount of eGFP mRNA produced in 1µg of total RNA from virus CF2ESH 11 was between 1x10^9.5 and 1x10^9 molecules of eGFP mRNA.

6.2.5 CF2ESH 11 virus replication in turkeys

Mutant virus CF2ESH 11 was used to infect turkeys. These experiments were conducted by Dr. N. Eterradossi and Dr. D. Toquin at the Agence Française de Securite Sanitaire des Aliments (AFSSA) in France. Specific pathogen free (SPF) turkeys were hatched and kept in isolators, blood samples taken, and one group of nine birds with one control group of six birds were established. The experimental group was then infected via intra
Figure 6.11 Northern blot of CF2ESH 11 virus RNA with G, SH or eGFP specific probes. Total RNA was extracted from BSC-1 cells infected with virus CF2ESH 11 (lanes labelled 11), virus CF2 (lane labelled C), and mock infected cells (lane labelled (-)). Lanes labelled $10^{9.5}$ to $10^7$ are standards (number of molecules of relevant gene RNA). A) The membrane was probed with G gene specific RNA. The mRNA transcript of the G gene is predicted to be 1.2kb. The blot was exposed for 1hr and 12min. B) The membrane was probed with SH gene specific RNA. The mRNA transcript of the SH gene is predicted to be 0.5kb. The blot was exposed for 20min. C) The membrane was probed with eGFP gene specific RNA. The mRNA transcript of the eGFP gene is predicted to be 0.7kb. The blot was exposed for 1hr and 53min. The ladder for all blots is an RNA ladder with markers from 0.5 to 9.0kb.
nasal inoculation with 0.1ml CF2ESH 11 virus per bird, and the SPF birds with 0.1ml of MEMH medium as the mock-inoculated controls. At three and five days post inoculation two birds from each group were culled and organs taken for histology and quantitative real-time PCR experiments. These tissues were stored in 10% paraformladehyde. The organs taken were: sinuses, upper trachea, lower trachea (with bronchi), lung, oviduct (in females), spleen, liver, thymus, bursa of Fabricus, and M. pectoralis. The remaining birds were culled at 21 days p.i., with blood samples taken for serological tests (see below). Samples were passed back to Warwick for analysis.

The respiratory tissue samples (sinuses, lung, upper trachea, and lower trachea) were examined for eGFP expression. One fourth of each organ sample was removed from paraformladehyde and soaked for at least 24 hours in PBS before being snap frozen in Tissue-Tek O.C.T.™ compound (Sakura Finetek). The embedded tissues were then sliced using a freezing microtome (OTF 5000, Bright) and 5µm sections were mounted with Gel Mount™ Aqueous Mounting Medium (Sigma-Aldrich) on slides. The slide cover slips were immediately sealed and examined for eGFP fluorescence using a confocal microscope. Four or more slices were examined for each bird for all respiratory organs. The fixed tissue had a high level of auto-fluorescence, but it was possible to separate this signal from the eGFP signal using the dye separation protocol in the Leica confocal software, which labelled auto-fluorescence red and eGFP fluorescence green. As can be seen in figures 6.12 and 6.13, sinus tissues from birds infected with virus CF2ESH 11 showed clear eGFP expressing cells from both birds on day 3 and from one of the birds on day 5. Sinus from birds mock infected showed no GFP expression and no other organ showed any eGFP expression (figure 6.14).
Figure 6.12 Confocal microscopy of CF2ESH infected turkey sinus 3 days p.i. Sinus tissues were removed from turkeys 3 days p.i. with virus CF2ESH 11 and paraformaldehyde fixed. Confocal microscopy was used to separate the auto fluorescence signal from the eGFP signals (A), with samples no. 5 (top) and no. 6 (bottom) representing two independently treated birds.
Figure 6.13 Confocal microscopy of CF2ESH infected turkey sinus 5 days p.i. Sinus tissues were removed from turkeys 5 days p.i. with virus CF2ESH 11 and paraformaldehyde fixed. Confocal microscopy was used to separate the auto fluorescence signal from the eGFP signals (sample no. 7).

Figure 6.14 Confocal microscopy of mock infected turkey sinus 3 or 5 days p.i. Sinus tissues were removed from turkeys 3 days (top) or 5 days (bottom) following mock infection. Confocal microscopy was used to separate the auto fluorescence signal, using the settings determined from positive samples. None of the samples showed any eGFP signal.
Dr. N. Eterradossi and Dr. D. Toquin also examined each tissue sample for N gene mRNA expression using quantitative real-time PCR and tested the blood of the birds culled at 21 days p.i. for seroconversion using an ELISA. Figure 6.15 shows the results of the quantitative real-time PCR experiments (Dr. N. Eterradossi and Dr. D. Toquin, personal communication). This was a more sensitive measurement of virus RNA than GFP expression in tissue sections, however the general trend was in agreement with the GFP found in tissue section. The highest signals were seen in the sinuses for birds infected with virus CF2ESH 11 at both three and five days p.i. However, detectable signal was also seen in all of the respiratory tissues on both days. RNA was also seen in all other tissues studies at 3 days p.i., though at lower levels, and in all tissues except for the liver and bursa of fabricius 5 days p.i. No RNA was detected in any of the mock infected controls. Overall, virus CF2ESH 11 appears to reach all deeper organs and muscle tissues (Dr. N. Eterradossi and Dr. D. Toquin, personal communication).

Examination of the blood samples from the birds by Dr. N. Eterradossi and Dr. D. Toquin showed positive ratios of sample (S) to reference positive (P) readings in ELISA experiments using an APV subgroup A antigen for both viruses (Table 6.1). Virus CF2ESH 11 had an S/P mean ratio of 0.798, whereas virus CF2EdSH(dG) 56 had an S/P mean ratio of 0.153. None of the control bird showed an S/P mean ratio. This indicated that the virus is capable of stimulating immunity in turkeys.

6.3 Production of GFP-expressing APV SH gene deleted viruses
Mutant genomes CF2EdSH 34 and CF2EdSH(dG) 56 were created during the cloning of genome CF2ESH. These deletion mutants were interesting as whole gene deletion has been considered as a method of attenuating APV for vaccine use (Naylor et al., 2007). The substitution of the eGFP gene in place of the SH gene preserves the number of genes in the genome and allows for visualization of attenuated virus growth.

Virus was produced from genomes CF2EdSH 34 and CF2EdSH(dG) 56 and the viruses were used to infect BSC-1 cells after which the cell culture phenotype was determined.
Figure 6.15 real-time PCR measurements of number of copies of the N gene (shown as log 10 copies per 10mg of tissue). Drs. D. Toquin and N. Eterradossi analysed the combined mRNA and vRNA of tissue samples from CF2ESH 11 virus infected turkeys (‘ESH,’ green boxes) at 3 d.p.i. (A) and 5 d.p.i (B) for expression of copies of the N gene using real-time PCR. Values in parenthesis are outside of the linear range of the real-time PCR control and undetectable reactions (Ct>40) are marked as ‘0.’ The samples with duplicate positive values in the linear range of detection are shaded dark gray, those with only one such value are shaded medium gray and those with positive values which were outside of the linear range are shaded light gray. Samples from mock infected SPF (specific-pathogen-free) turkeys were also tested for each organ at 3 and 5 d.p.i and all had undetectable levels of expression. The oviduct tissue samples were not examined. Data provided by Dr. N. Eterradossi and Dr. D. Toquin (personal communication).
The deletion of the SH gene in both these genomes was expected to result in viruses with syncytial phenotype in cell culture (Ling et al., 2008). Confluent monolayers of BSC-1 cells infected with either virus CF2EdSH 34 or CF2EdSHdG 56 were photographed using confocal microscopy under UV light as above (figure 6.16).

### 6.3.1 Further characterization of virus CF2EdSH(dG)

The titre for virus CFEdSH 56 was determined to be $1.1 \times 10^3$ CIU/ml. As before, 1ml of virus was used to infect confluent BSC-1 cells in a 6-well single dish, giving an CIU per cell of $3.7 \times 10^{-4}$. Timelapse photography of virus infection was also carried out as for virus CF2ESH. The infection was monitored over 800min (13hr 20min), and showed the formation of a syncytial plaque. Images of cells under ultraviolet and transmitted light are shown in figure 6.17 parts I-III and appendix C. EGFP expression is first seen in a single cell, which then fuses with neighbouring cells to form a multi-nucleated syncytial cell.

### Table 6.1 S/P ratios of turkeys prior to and three weeks post inoculation with mutant virus CF2ESH 11

<table>
<thead>
<tr>
<th>Time of sampling</th>
<th>Virus inoculated</th>
<th>Number of birds tested</th>
<th>S/P mean ratio</th>
<th>S/P min-max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to inoculation</td>
<td>None</td>
<td>5</td>
<td>-0.024</td>
<td>-0.028 to -0.016</td>
</tr>
<tr>
<td>Three weeks Post inoculation</td>
<td>None (control birds)</td>
<td>3</td>
<td>-0.024</td>
<td>-0.026 to -0.023</td>
</tr>
<tr>
<td>CF2ESH 11</td>
<td>5</td>
<td>0.798</td>
<td>0.542 to 1.045</td>
<td></td>
</tr>
</tbody>
</table>

Virus immune system stimulation was measured by comparing the sample (S) to reference positive (P) ELISA readings using a subgroup A ELISA antigen. Five birds were tested prior to inoculations and 8 birds were tested three weeks post inoculation (three control birds which had not been infected, five birds infected by the intranasal route with virus CF2ESH 11). The mean S/P ratio was calculated for each group of turkeys and the range of ratios is presented (Data provided by Dr. N. Eterradossi and Dr. D. Toquin, personal communication).

The deletion of the SH gene in both these genomes was expected to result in viruses with syncytial phenotype in cell culture (Ling et al., 2008). Confluent monolayers of BSC-1 cells infected with either virus CF2EdSH 34 or CF2EdSHdG 56 were photographed using confocal microscopy under UV light as above (figure 6.16).
Figure 6.16 eGFP expression of viruses CF2EdSH 34 and CF2EdSH(dG) 56 in BSC-1 cells. Recombinant virus was produced as before in BSR-T7 cells and virus was then used to infect BSC-1 cells. The cells infected with virus CF2EdSH 34 were examined by microscopy (A). The cells infected with virus CF2EdSH(dG) 56 were examined by confocal microscopy under 489nm laser line (B) and with transmitted light (C).
Figure 6.17 part I: Time lapse photography of infection of BSC-1 cells with virus CF2EdSH 56, at 0, 100 and 200min. EGFP expression (left) and bright field (right) was images were taken using confocal microscopy.
Figure 6.17 part II: Time lapse photography of infection of BSC-1 cells with virus CF2EdSH 56, at 300, 400 and 500 min. EGFP expression (left) and bright field (right) was images were taken using confocal microscopy.
Figure 6.17 part III: Time lapse photography of infection of BSC-1 cells with virus CF2EdSH 56, at 600, 700 and 800min. EGFP expression (left) and bright field (right) was imaged using confocal microscopy.
In order to determine the levels of SH, G and eGFP RNA being produced from mutant genomes CF2EdSH(dG) 56, northern blots were carried out as above using the probes and standards as described for virus CF2ESH 11. As can be seen in figure 6.18 (A), G gene RNA was expressed only from non-mutated virus CF2. A larger (approximately 2.0kb) mRNA can be seen produced from the mutant virus CF2EdSH(dG) 56. This may be a read through transcript containing the eGFP and G genes resulting from the loss of the eGFP GE signal from this genome, as a read through transcript would be approximately 2.0kb. Expression of SH mRNA was examined (figure 6.18, B) and, as expected, only virus CF2 showed detectable levels of production. For eGFP expression (figure 6.18, C), it was expected that for all eGFP gene mutant viruses an eGFP specific mRNA would be detected, however no eGFP expression was observed for virus CF2EdSH(dG) 56.

Mutant virus CF2EdSH(dG) 56, having been shown to express eGFP in cell culture, was examined in turkeys. The design of the experiment was as described above. The experimental group was infected by intra nasal inoculation with 0.5ml of CF2EdSH(dG) 56 virus per bird. Organs were taken for histology and quantitative real-time PCR experiments at days three and five post inoculation. The respiratory tissue samples (sinuses, lung, upper trachea, and lower trachea) were examined for eGFP fluorescence as above, however none showed eGFP expression.

Dr. N. Eterradossi and Dr. D. Toquin examined the tissue samples for N RNA expression using quantitative real-time PCR (real time RT PCR) and tested the blood of the 21 days p.i. birds for sero-conversion using ELISA as for virus CF2ESH 11. RNA was detected for virus CF2EdSH(dG) 56, mainly in the sinuses, but also at low levels in the bursa of fabricius and pectoralis muscle 3 days p.i. and in the thymus and bursa of fabricius at 5 days p.i. (figure 6.19). Virus CF2EdSH(dG) 56 replicated less well and in fewer organs than virus CF2ESH 11 (Dr. N. Eterradossi and Dr. D. Toquin, personal communication). Examination of the blood samples from the birds by Dr. N. Eterradossi and Dr. D. Toquin showed positive ratios of sample (S) to reference positive (P) readings in ELISA experiments using an APV subgroup A antigen for virus CF2EdSH(dG) (table 6.2), with.
6.18 Northern blot of virus RNA with G, SH or eGFP specific probes. Total RNA was extracted from BSC-1 cells infected with virus CF2EdSH 56 (lane labelled 56), virus CF2 (lane labelled C), and mock infected cells (lane labelled (-)). A) The membrane was probed with G gene specific RNA. Lanes labelled 109.5 to 1077.5 are standards in molecules of gene RNA. The mRNA transcript of the G gene is predicted to be 1.2kb. The blot was exposed to UV for 1hr and 12min before photographing. B) The membrane was probed with SH gene specific RNA. Lanes labelled 109.5 to 1077.5 are standards in molecules of SH gene RNA (right). The mRNA transcript of the SH gene is predicted to be 0.5kb. The blot was exposed to UV for 20min before photographing. C) The membrane was probed with eGFP gene specific RNA. The mRNA transcript of the eGFP gene is predicted to be 0.7kb. The blot was exposed to UV for 1hr and 53min before photographing.
Drs. D. Toquin and N. Eterradossi analysed the combined mRNA and vRNA of tissue samples from CF2EdSH 56 virus infected turkeys (‘EdSH,’ pink boxes) at 3 d.p.i. (A) and 5 d.p.i (B) for expression of copies of the N gene using real-time PCR. Values in parenthesis are outside of the linear range of the real-time PCR control and undetectable reactions (Ct>40) are marked as ‘0.’ The samples with duplicate positive values in the linear range of detection are shaded dark gray, those with only one such value are shaded medium gray and those with positive values which were outside of the linear range are shaded light gray. The oviduct tissues were not examined. Samples from mock infected SPF (specific-pathogen-free) turkeys were also tested for each organ at 3 and 5 d.p.i and all had undetectable levels of expression (Dr. N. Eterradossi and Dr. D. Toquin, personal communication).

**Figure 6.19 real-time PCR measurements of number of copies of the N gene (shown as log 10 copies per 10mg of tissue).**

<table>
<thead>
<tr>
<th></th>
<th>Sinus</th>
<th>Upper Trachea</th>
<th>Lower Trachea</th>
<th>Lungs</th>
<th>Spleen</th>
<th>Liver</th>
<th>Thymus</th>
<th>Bursa of Fabricius</th>
<th>Pectoralis Muscle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 d.p.i.</td>
<td>E dSH</td>
<td>(3.88/4.45)</td>
<td>0/ (4.30)</td>
<td>0/0</td>
<td>0/0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(3.85)</td>
</tr>
<tr>
<td></td>
<td>E dSH</td>
<td>6.31/6.31</td>
<td>(4.64)/3.95</td>
<td>0/0</td>
<td>0/ (3.73)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.53/0</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 d.p.i.</td>
<td>E dSH</td>
<td>4.63/ (4.33)</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0</td>
<td>0</td>
<td>(4.18)/5.25</td>
<td>5.39/0</td>
</tr>
<tr>
<td></td>
<td>E dSH</td>
<td>4.81/ (4.19)</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The table below shows the S/P ratios of turkeys prior to and three weeks post inoculation with mutant virus CF2EdSH 56. Virus immune system stimulation was measured by comparing the sample (S) to reference positive (P) ELISA readings using a subgroup A ELISA antigen. Five birds were tested prior to inoculations and 8 birds were tested three weeks post inoculation (three control birds which had not been infected, five birds infected by the intranasal route with virus CF2EdSH 56). The mean S/P ratio was calculated for each group of turkeys and the range of ratios is presented (Dr. N. Eterradossi and Dr. D. Toquin, personal communication).

<table>
<thead>
<tr>
<th>Time of sampling</th>
<th>Virus inoculated</th>
<th>Number of birds tested</th>
<th>S/P mean ratio</th>
<th>S/P min-max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to inoculation</td>
<td>None</td>
<td>5</td>
<td>-0.024</td>
<td>-0.028 to -0.016</td>
</tr>
<tr>
<td>Three weeks post inoculation</td>
<td>None (control birds)</td>
<td>3</td>
<td>-0.024</td>
<td>-0.026 to -0.023</td>
</tr>
<tr>
<td></td>
<td>CF2EdSH 56</td>
<td>5</td>
<td>0.153</td>
<td>0.022 to 0.445</td>
</tr>
</tbody>
</table>

Table 6.2 S/P ratios of turkeys prior to and three weeks post inoculation with mutant virus CF2EdSH 56.
an S/P mean ratio of 0.153. This S/P mean ratio indicates that the virus is capable of stimulating immunity in turkeys.

6.4 Discussion
6.4.1. Detection of eGFP in vitro

The introduction of eGFP into the APV viral genome allowed visualization of living virus-infected cells. The time-lapse experiments showed the spread of virus in cell culture and the development of syncytia formation. Interestingly, there was some suggestion of the possible production of structures reminiscent of ‘membrane nano tubes’ as have been described for HIV-infected T cells (Sowinski et al., 2008). Higher resolution time-lapse microscopy would be necessary to discern such structures, but these nano tubes may represent an alternative means by which the virus infects nearby cells.

Introduction of the eGFP gene into the APV genome also showed that the genome is capable of accepting large foreign gene insertions, suggesting that recombinant APV genomes may be used in the creation of multivalent vaccine. Since inserting a gene could potentially alter expression levels of the downstream genes, the pathogenesis of the virus may be affected (Krempl et al., 2007; Naylor et al., 2007). It could be important to move the eGFP tag to the end of the genome to prevent the insertion of an extra gene affecting mRNA production. As seen with the RSV vaccine strain candidate cpts248/404, the down regulation of the M2 gene due to a mutation in its GS sequence resulted in virus that was attenuation in mice (Whitehead et al., 1998). However, the addition of eGFP as an extra gene in genome CF2ESH 11 did not decrease the production of the downstream gene (the G gene) RNA (figure 6.19) compared to the CF2 genome. Further experiments with multiple deliberate gene deletions could use other reporter genes as placeholders, allowing for redundancy in virus tracking proteins.

6.4.2. Detection of eGFP in vivo

The eGFP expression of the CF2ESH 11 virus also allowed visual identification of virus-infected cells within turkey tissue samples. Visible levels of eGFP were found in the sinus tissues of the CF2ESH 11 virus in both infected birds at 3 days p.i. and one bird at 5
days p.i., demonstrating virus replication in the respiratory tract. Quantitative real-time PCR confirmed the presence of the virus at the highest levels in the respiratory tract, but also showed low levels of virus in all tissues sampled in at least one bird on day 3 or day 5 p.i. In contrast, quantitative real-time PCR on the CF2EdSH(dG) virus infected bird organs showed a trend for the virus to be restricted to the upper respiratory tract. eGFP expression was not detected in these tissues by confocal microscopy, suggesting very low levels of virus replication. However, as it was not possible to grow the CF2EdSH(dG) virus to the same titre as the CF2ESH virus, it may be that the differences in ability to infect tissues was due to the difference in titre of virus used to inoculate the turkeys.

Previously (section 1.2) APV was described as being found primarily in the upper respiratory tract of birds, with regular reports of virus being isolated from the reproductive tract (Cook et al., 1991; Cook et al., 2000; Jones et al., 1988). The effect of APV infection on egg production would suggest some virus infection beyond the respiratory tract tissues. Additionally, for the virus to spread as rapidly (section 1.2) as is seen, would suggest that virus may be being shed by non-oral routes. The spread of APV is an area of concern and APV tissue tropism in domestic fowl and sentinel birds requires further investigation.

6.4.3. Viruses with SH deleted genomes

With the creation of mutant viruses lacking the SH and G genes, both of these genes have been shown to be unnecessary for APV viral replication in vitro and in vivo. Recombinant virus could grow with either the SH or both the SH and G genes deleted and showed a syncytial phenotype subsequently demonstrated to be due to the deletion of the SH gene (Ling et al., 2008).

The cloning and resulting sequence of the mutant APV genomes CF2ESH 11, CF2EdSH 34, and CF2EdSH(dG) 56 showed the difficulties inherent in manipulating large expression plasmids. Point mutations were seen in both genomes and the deletion of the GE signal in CF2EdSH(dG) 56 suggests that the APV genome has regions that are prone to mutation (areas with a high number of AT repeats, for example). However, the ability
of these viruses to replicate in turkey tissues suggests that the APV genome is tolerant of mutations and these mutations could be used to attenuate the virus in vivo. Future work will include the full in vivo characterization of virus CF2EdSH 34.

The use of the SH gene-deleted or SH and G gene-deleted viruses as vaccine candidates will be an avenue of future work. Both of these viruses are attenuated in vitro and the virus lacking both the SH and G genes is also attenuated and more limited in its tropism in vivo. The fine-tuning of the tropism of any future vaccine strains will be important, especially as farms seek to use herd protection to vaccinate flocks more quickly. A virus with increased tropism without increased disease would give a better chance of full protection of the flock.
Chapter 7:
Conclusions
7.1 Conclusions

Establishment of the APV plasmid rescue system

The aims of this thesis revolved around creating and using reverse genetics tools to expand our understanding of the molecular biology of avian pneumovirus (APV). The development of a plasmid based rescue system for APV as a RNA virus has allowed targeted manipulation of the virus. With this robust system established and producing levels of reporter protein equivalent to that seen for the RSV plasmid rescue system (Marriott et al., 1999)(Dr. A. Marriott, personal communication), further characterization of the limits of the system was possible. Analysis of mRNA and protein levels produced from the GS sequence mutants demonstrated that they were equivalent, confirming that the GS sequences affect only mRNA production and do not result the production of aberrant mRNA transcripts. These experiments also showed that protein quantification from the plasmid rescue system was a reliable method of assessing the effects of mutation of the cis-acting transcription signals.

Role of N, P, L, M2, SH and G proteins

Using the plasmid rescue system, the APV M2-1 protein was shown to be a transcriptional regulator, although to a lesser extent than the M2-1 proteins of RSV and PVM. The maximum enhancement in the presence of the APV M2-1 protein was only three-fold, unlike the ten-fold enhancement seen for RSV and PVM (Collins et al., 1999a; Dibben and Easton, 2007; Fears and Collins, 1999). It was also found that the APV M2-2 protein does have an effect on gene expression from the minigenome plasmid, suggesting that it may be a factor in the fine-tuning transcription for the pneumoviruses.

The optimal levels of the N, P and L helper proteins for efficient rescue were determined and it was shown that the amount of N plasmid was critical, with the addition of too much plasmid being as detrimental as too little plasmid. The P plasmid similarly proved to have an optimal amount, after which reporter protein expression from the minigenome was inhibited. However, the L plasmid proved to have a minimal required amount, after which addition of more plasmid failed to enhance or inhibit reporter gene expression.
Using the system for the generation of infectious recombinant viruses, the roles of the G and SH glycoproteins were also investigated. The SH and G genes were shown to be non-essential but enhancing for APV grown in vitro or in vivo, as has been seen for RSV (Bukreyev et al., 1997; Karron et al., 1997; Techaarpornkul et al., 2001). However, deletion of the SH gene from APV resulted in a syncytial virus phenotype in cell culture and the deletion of both the G and SH genes attenuated the virus in cell culture compared to the eGFP insertion virus. Hence, the APV SH protein appears to have a role in regulation of host cell fusion, although this may be via interaction with other APV proteins. This contrasts significantly with observations of the equivalent SH gene-deleted RSV where no change in cytopathic effect was observed. The APV G and SH genes are also required for efficient virus growth.

Transcriptional regulation in vaccine design
APV vaccines have been difficult to produce, as an attenuated live vaccine must not revert to virulence and a killed vaccine must be immunogenic enough to protect the birds. It was shown that single point mutations can be responsible for vaccine candidate virus phenotypes, here protective and non-protective strains. Indeed, the only differences found between two of the strains were point mutations to the GE sequence of the SH gene. The roles of these gene end sequence mutations in the vaccine strain pathogenicity and ability to induce protection from challenge were determined. Mutations to the GE sequence were found to affect production of the downstream gene, as predicted from the stop-start model of paramyxovirus transcription. This was confirmed by carrying out northern blots of the RNA from these vaccine candidate viruses where it was found that the levels of read-through transcripts were increased and the levels of monocistronic mRNA decreased (Naylor et al., 2007). The G gene, as a surface glycoprotein, would be expected to be one of the major antigens for APV antibody production. Since antibody production is not the sole indicator of protection (section 1.2), the G gene may also have an interaction with the innate immune system. For future APV vaccines, it will be important to have an SH/G balance that both produces protectivity and is non-pathogenic in birds (Naylor et al., 2007).
For Paramyxoviruses, general factors in determining the levels of gene expression are the position in the genome of the gene and the sequence of gene junctions. The APV L gene was shown to have both positional and GS sequence regulation, as its GS sequence is less active than the consensus GS sequence. This down regulation appears to be important, as even with only one or two of the three mutations of the L GS, sequence transcription levels are decreased to approximately 50% of genes with the consensus sequence.

**APV tissue tropism**

The APV full-length rescue system (Naylor et al., 2004) was used to create an eGFP-expressing cloned avian pneumovirus. Three genomes were cloned in which the eGFP gene was introduced in addition to the SH gene alone or in place of both the SH and G genes and finally where the eGFP gene was introduced as an additional transcription unit. These eGFP expressing viruses showed that the APV genome tolerates deletion of the SH and G genes and replacement with a heterologous gene. These deletions affected the phenotype of viral infection in vitro, confirming that the deletion of the SH gene from the APV genome results in a syncytial phenotype (Ling et al., 2008). The deletion of the SH and G genes affected the tropism of the virus in vivo, which may be due to either virus attenuation, to a change in the surface glycoproteins of the virus, or to the lower dose of virus administered. Analysis of mRNA levels by real-time PCR is a more sensitive method of detection of virus compared to tissue examination as determined by detection of virus RNA in tissues that were negative for eGFP expression. However, real-time PCR cannot show the area (the epithelial cells, for example) of the tissue that is infected. Both methods found the highest levels of virus were seen in the sinuses. The finding that the virus (either with or without SH and G genes) can spread to the liver and spleen suggests that a low level of viraemia is occurring during infection. This has not been the case with other pneumoviruses (Dr. A. Easton, personal communication).

**7.2 Future work**

The tools and preliminary research are now complete to create a panel of mutations with known effects on virus pathogenicity and protectivity. A set of mutations that should be tried is the deletion of entire non-essential genes in turn and the introduction of mutated
genes in their place. The determinants of APV tropism are likely to be the surface
glycoproteins so the effect of SH, F and/or G deletions should be determined. The APV
eGFP SH delete virus that has been produced should be put into turkeys to be compared
to the non-SH delete and SH/G delete viruses.

The failure of any of the N, P, L or M2-1 proteins from RSV, PVM or hMPV to
functionally replace the equivalent protein in the APV plasmid based rescue system
suggests that it should be possible to create chimeras with these viruses. Chimeric
proteins can be used to broadly define the functional regions of the proteins. Once the
active regions of the protein are defined, point mutations can be introduced to pin point
key residues.

The question of how the virus switches from transcription to replication is still unsolved.
Further investigation of the effect of M2-1 and M2-2 proteins on replication might
provide new avenues of research. The RNA levels for the genome in the presence of M2-
1 and M2-2 proteins should be measured to distinguish between effects on replication and
on transcription.

The non-transcribed regions of the APV genome also remain to be fully investigated.
The dicistronic minigenome could be used to investigate the efficacy of the different GE
sequences and saturation mutagenesis carried out. This would provide additional means
of down regulating key genes and changing the virus’s tropism.

The results of the vaccine candidate gene end mutation studies show that PRS can be
used as a model for transcription that is comparable to whole virus infection. This is in
contrast to Marriott et al., (2001) who found that the PRS differed from helper virus
rescue in its ability to recognize and replicate the genome or minigenome of related
viruses. One interesting aspect of having three possible reverse genetics systems are the
experiments which show correlation between systems, highlighting regions that function
consistently when manipulated in the presence or absence of whole virus. Here,
comparison of mutation in the IR investigated in both systems shows that similar values
can be obtained across systems. However this may not be true for the leader and trailer replication signals (Fearns et al., 2000; Smith, 2001) and needs to be further defined.
Bibliography


Buys, S. B., and Du Preez, J. H. (1980). A preliminary report on the isolation of a virus causing sinusitis in turkeys in South Africa and attempts to attenuate the virus. *Turkeys* **28**, 36 (but cont. on 46- which I don't have- also oddly the extra page isn't listed in citations of this paper).


Giraud, P., and others, a. (1986). A possible viral candidate for the aetiology of turkey rhinotracheitis. The Veterinary Record 118, 81.


Taxonomy" (M. H. V. Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B.
vesicular stomatitis viruses from DNA. Proc Natl Acad Sci U S A 92(10), 4477-81.
Levine, S., Klaiber-Franco, R., and Paradiso, P. R. (1987). Demonstration that
glycoprotein G is the attachment protein of respiratory syncytial virus. J Gen
Virol 68 ( Pt 9), 2521-4.
Association of respiratory syncytial virus m protein with viral nucleocapsids is
mediated by the m2-1 protein. J Virol 82(17), 8863-70.
Li, J., Ling, R., Randhawa, J. S., Shaw, K., Davis, P. J., Juhasz, K., Pringle, C. R.,
reactions following infection of turkeys with avian Metapneumovirus (aMPV)
G genes of turkey rhinotracheitis virus and their intergenic regions reveals a gene
order different from that of other pneumoviruses. J Gen Virol 73 ( Pt 7), 1709-15.
Deletion of the SH gene from avian metapneumovirus has a greater impact on
virus production and immunogenicity in turkeys than deletion of the G gene or
Lippincott, Williams, and Wilkins (2001). Chapter 45: Respiratory Syncytial Virus,
Infectious Agent. 4th Edition ed. In "Fields Virology" (D. M. Knipe, and P. M.
Howley, Eds.). Courier Westford, USA.
Bulletin 56(8), 637-663.
Swollen head syndrome in Taiwan - isolation of an avian pneumovirus and
Lwamba, H. C., Alvarez, R., Wise, M. G., Yu, Q., Halvorson, D., Njenga, M. K., and
Seal, B. S. (2005). Comparison of the full-length genome sequence of avian


Appendix A: Primers Sequences
All sequences are listed 5’ to 3’

APVNTOPOF
CGGAACAAGGGATGTCTCTTTGAAAGTATTAGACT

APVNTOPOR
TGAGTCAAGGGTTACTCAAATTTGGATCATCTCT

TRTP4
CTTCCCTGAAGGC

TRTP2
ATAACTACAGATCAAGATTG

APVAM2-2FOR
GAGAGTGACATGGCAATGGT

APVAM2-2REV
GGGATCCCTAATTATATGAGG

IntMut/DSMGR
GTCCCGACCTTTTTTTAATTAATAAATACTGTTGCCTATTATTTGGCACTAG
TGTTGTCAAAAATTCAATTACGCCCGCCCTG

IntMut/DSMGF
ATTAAAAAGAAAGGTCGGGACAAGTATCTCTATATGAAGAC

DSMGIR R
GACATTTCGAAGTACTCAG

DSMG check 1
ATGAGTGGCAGGGCGGGCG

DSMG check 2
GCTGGCGATTCAGGTTCATC

DSMG check 3
TTATTTATCGGAGTTGCAGT

DSMG check 4
TACTTCGAAATGTCCGTTCG

DSMG check 5
GGCCTATTTCCCTAAAGG
Appendix A: Primers Sequences
All sequences are listed 5' to 3'

DSMG check 6
CCGGCATAAAGAATTGAAG

DSMG check 7
CCTGCGTTATCCCTGATTTC

DSMG check 8
CCTT TAGGAAATAG

DSMG check 9
GAGCAGACAAGCCCGTCA

DSMG check 10
CATT CATTTAAGGTCAGTTG

IRmut2 FOR long
GGGCGGGGGGTGTAATGAAATTTGACAAACACTAGTGCGAAAATGATAGGCAACAG
TATTATTTAATTAAAAAGAAAGGTGCG

IRmut2 REV long
CCCGACCTTTCTTTTTAATTTAAACACTTGCGCTATCATATTGGCAGTGTGAGCAAGGGCGA

DSMGIR F
CCAGTTTTGATTTAAACGTG

IR3 REV LONG
CCCGACCTTTCTTTTTAATTTAAACACTTGCGCTATCATATTGGCAGTGTGAGCAAGGGCGA

T7prom primer
TAATACGACTCAGATATAG

FPSHP
GGGCGGGGACAAGTATATACACTGTGATTTTGACAAACACTAGTGCGAAAATGATAGGCAACAG
TATTATTTAATTAAAAAGAAAGGTGCG

FPSHMES
CCCGACCTTTCTTTTTAATTTAAACACTTGCGCTATCATATTGGCAGTGTGAGCAAGGGCGA

TRT22K2
AAAGATTTTGAGTTTAGCTAT
Appendix B: Construct Sequences
1 ctgcattaat gaatcggcga aagcgccggg agagcgggtt tcgcgtattgg gcgcgtcttcc
61 gcttcctcgc tcaactgactc gctgctgcctg tcgcgttcgggc tcgcgcgac gcgtctcagct
121 cactcaaggg cggtataacg gtattccaca gaatcaggcgg ataacgcaggg aaagaacatg
181 tgagccaaag ggcagcagaa ggcagacag ggctaaaggg ccgcgtttcgt ggcttttttcc
241 cataggtcgc gccccctctga cgagcatac aaatacgcag gctcaagtcg gaggctgcgca
301 aacccgacag gactataaag ataccagcgct tttccccccgt gaagctcctct cctgctgtcct
361 cctccctccga ccttcgccgct tacccgatac ctctcgcggct tcctccccctt cgggaagcggtg
421 gcgcgttcctt atagctcagc ctgtgaggt ctcagttcgg ttaggtctgt tgcgtccaaag
481 ctgggctgtg tgcagcaacc cccgcgtccag cccgcacgct gcgccttcctc ccgtaactat
541 ctgtcttgagct ccaacccccgg aagacacgc a ttagtgccag cgcgtcattc ggcgcttacgc
601 aggattagca gcgcaggtga tgtaggcggg gctacagagt tcttgaagtg gtgcgctacac
661 tacggctca ctagagac ctaatttggt atcttgccgct tcgcgtgacgc agttaccccc
721 gggaaaaaggt tgtgtagctc tggtcgcgcg ccacaaaaac ccgcgtgtag cgggtgctttt
781 ttcggtgttcg cagcagcagt taccgcgacga aaaaaaggt ctcagaagag tctttttgtc
841 ttgttttggc agcagcagat taccgcgaag aaaaaaggt tcttaagaga ttttaatcct
901 agattatccaa aagagatcct ctcctagatc cttttaaat ctaaatagga ttttaaatca
961 atctaaagta tatatagta aactttgcctt gacagttacc aatgcttaat ctagtgacga
1021 cctatctcag cgcacgtgtct atttcggtca tccatatggg cctgacctcc cgtctgcttag
1081 ataactacga tacgggaaggg cttaccatct gcgcctagtgct tgcagatgt acgcggacagc
1141 ccacgctcag cgcgccgaga tttatcagca ataacccagc cagcggcggc gcgcggacgc
1201 agaagtggtgc tgcgaacttc taccgcctcc anaccgtcta ttaatgttg ccgggaagct
1261 agagtaagta gttccgcagt taattgtttg cgcacggttg ttgctctgctc tacaggccatc
1321 gtgtgtgctc gctgctggtt ggtattggtct ctaattgcgt cgcgtccttc ccgcgttgtcctga
1381 cagaggtcat gtcccccat gttgtgaca aaaaaaggt cgcgtcctcg ccctccgctc
1441 gtgtgcagaa gtaatgggca gcgcagtttcttcagtt ctcctcagtt cgggaagcggtg
1501 tctcttactg tcatgcctac gcgaatagtc ctttatgtgat gctgtgagta ctcacaacag
1561 tcattctggag aataaggtgt gcggcgccag gtggctgtctt ggcgcgcgtc aatacgcggat
1621 aataccgcgc cacatagcag aacaaaaata ggctctcataa tcgggagaag ccctctcgagg
1681 cggaaaaactc caagttcagc aggctgtggg gcagtgaacc cacctctgtcc
1741 cccactgtat ccctcagcatc ttctacttctt accacgggttc gttggtgagc aaaaaacagga
1801 aggcaaaatgc cgcacaaaaa gggaataagg gcgcagcggc aatgctggaat actcaactgc
2G2term-2

1861 ttcttttctt aatattatgt aagcattttat cagggttatt gtctctagag cggatatcata
1921 ttgtgaatgta ttttagaaa taaacaataa ggggttccgc gcacaattttg cccgaagagtg
1981 ccaacctacgcc tctaaagaaac cattattaccc atgcattaaa cctataaa taggccgtatc
2041 acgaggcccc ttcgctctgcg gcgtttcggt gatgactggtt aaaacctctcg acacatgcag
2101 ttcctcggagc cgctcacacgc ttgtctgttaa cgccagctcg ccggagcagca agcccgtcag
2161 gcgcgctgacgc cgggtgttgg gcgggtcctgg gcgtgctcata actatgcccgc atccagacag
2221 attgtactgca gatggcacca tattcggtgtg gaaatacgc cagatgctgt aaggagaaaa
2281 taccgcaatcg gcgcccatttc gcattaactggtatcgtgagc ggtaagaggg gcgtcgcgtg
2341 cgggctcctct gcgtattcag cccagaacctg tttatttcag ttataaatag tttacaataaa

2401 agcataagcacaatatttt cacaataaaa gcattttttt cactgcatta tagtgtgtgtt
2461 tttcctcaac cactataatcg atctttatcct gcctgctcga aagcggcgggc cgcggcgcctt
2521 ctagagcttt ttagagacca tcattagccg ctagcttttag

2581 ttttaagttt ttatcaactt acatcatttt aaggtcaggc tgttttcaggg gtagaatatat
2641 gtaggagtt gattttgctt tattacctag aattacagcg catcttttcgc cccctctttg

2701 gctttatatga ggtctctct tattttttctt gcgtgctgatg ttccgcagtt accttttccgt
2761 acctctgccca aacaaacaac tccccgcggc aaccttttccc cgggtgttacc ttgcagctggc
2821 acgtaatacca cgatctccttt ttcgctcattc gtctttcctgt gcctaaaaac aacaacggcgc
2881 gcggagaagtt accggctgcgc atcgctgagtt gcagctggcga cacctgctgc gaagatgttg
2941 gggtgctggga caagatgga ttccaattca cgggagccct cctgatacgct ttggtagacttta
3001 atccagacacta cccgcttcgc aacgactagaag atctgctgctc gcctgctcaca gtaagctatg
3061 tctccagact gcagctctactc atctttgctca atcaaggcgct tgtgctgcttc gcgatttgttt
3121 acataacggc acataactctc tagaactctct gcgcctcttttga attaaqccc
3181 acggttttccc gcgtcataaggccttcacaacgc gtcgctcattc gaaatggacac aacgattgcc
3241 accgcggcagg gtttactact cccctcgggt gtagctgatg aagtctcagtgt
2G2term-2

3301 agcccatact ctgtcctgat acctggcaga tggaaacctct tggcaacgcc ttcccccgact <Luc'
3361 tcttagaga gggagagcgc accagaagca atttctgtga aattagataa atctgtatttg <Luc'
3421 tcaatcgag tgcgttttggc gaagaagggag aatagggttg gcaccagcag cgacacttga <Luc'
3481 atcttgtaat cctgaaggtct cctcagaac agctcctctct caaatctata cattaagacg <Luc'
3541 actcgaatact cacatataca atatccgagtt gtagtaaaca ttccaaaccg gtagtggaat <Luc'
3601 ggaacacac tttaaatgc agtatccgga atgtattgtg tgccaaaat aggatctcttg <Luc'
3661 gcagcgcga atctcagcga ggcatcttctc tgaagcagag cgacaccttt aggcagacca <Luc'
3721 gtagatccag agagttgata gtagctgca attgctctgt cctctagcga ggactcctgc <Luc'
3781 aaaaaatcgat attcatataaa accgggaggt agatgaagatg tgacgaactg gtacatcggac <Luc'
3841 tgaatcctct gtgaatcctg attagaatcc atgataaataa ttttttttgt gattggagac <Luc'
3901 tttttttgca ctgtccaaat tttttgtgca ccccttttgg aaacgaaacac caggtgaggc <Luc'
3961 tgccaaatgc cctactcgtt gagcaatctca gttccattat aatgtcgttt cgccgggca <Luc'
4021 actgcaactc cgataaataa acgcgccccac accggctataa aagatggaag agaagtttaca <<DSGCHECK3...<<
     >>.DSMG Check 6...<< <Luc'
4081 ctgcacatercgagatgctcttgct atttcttttgc atgccttctgc atgtgctgcat <Luc'
4141 cgaacggaca ttccgagact ctcagctctga atcttctgtc cctcagatag tgcgcatctgta <Luc'
4201 aaagcaattg ttccaggaac cagggctat ctttctcatag ccttatgcaag ttgctcttcca <Luc'
4261 gcggctcctc tctccagcgg ctataagtagc gcccgggttc ttctttgtat ttggcttgctt <Luc'
4321 tccatggtttc ctctctccctg ctgcttttctca ataactatag cagccccccc tgccactcat <Luc'
     gene start <<.......<<
     ClaI site >>...>>
     gene end- P <<............<
     <<...OMCHECKS...<<
     <<.......CAT'.......<<
4381 cgcagtactg ttgtaattca ttaaggattc tgccgacagtg gaagccatca cagacggcat <CAT'
4441 gatgaacctcg aatcgcccag gcagtcagcga ccttgctgcc ttcgtataaa tattttgcca <DSGCHECK2...<<
     <<.......CAT'.......<<
06 Apr 2009

Sequence Data

File Name: F2_pUCRL4 annotated.cm5, dated 31 Mar 2009

Description: CASAF2

Printed: 1 to 16309 bps (Full)

F2_pUCRL4, 16309 bps DNA Circular

TTG to CTC mutation >>>

<<L ORF.<<

File Name: F2_pUCRL4 annotated.cm5, dated 31 Mar 2009

Description: CASAF2

Printed: 1 to 16309 bps (Full)
1381  gttacaagat agtacatgtt tccctcacttg tatcactaaa tgcaggacgt cctccatgtt
<........................................L ORF........................................>
1441  tttagttcga gcacacacag tggtagcag aagagatct ttcacactct tgtgaatataa
<........................................L ORF........................................>
1501  atccccatgt tctctcttag ttgcatcagt tccctcatt gccacccctt cacaaggtc
<........................................L ORF........................................>
1561  gattattctg ctgatactcc ccattacag cacatactct gtgggtagat gatgatacatt
<........................................L ORF........................................>
1621  ggcacttttcc aacatccgtg atacaaccct tagattaggc tatcactag cagtctggaa
<........................................L ORF........................................>
1681  ctaagatttc ccagctcctc caccagtgaa gtagataatt cttagttcag gtttgtttaat
<........................................L ORF........................................>
1741  caacttgccct atacatccct ttacatctac cttacacctc gtggagcat atgcacaaagt
<........................................L ORF........................................>
1801  ataaagggtg atgtgtagcc aaggcattaa agtcatgtag tttttgtcag aatgtgtgcc
<........................................L ORF........................................>
1861  agactgtcta gctgtactgc atgtgcattt gaaagttggtc aacactaacta agatgacttt
<........................................L ORF........................................>
1921  cgggaacagc agtagggcgc gagtggtgctc tccagaggtt gtcacatccag tggagatac
<........................................L ORF........................................>
1981  tgttttcttt atggtaggtgc cattgtcttg catgttcttg caccctgcgca aacgtgtgcat
<........................................L ORF........................................>
2041  aaccatgcgc atatccgagtt aacacgcatc ttaaactttt aacattgcac ccgtcagccat
<........................................L ORF........................................>
2101  tatatttcata cctcagttga caggttttttg ctgcaactag tccacctcag cattgcaat
<........................................L ORF........................................>
2161  ccattggtctc tccatctgtg ctcactgtttct taccgacct acaaaccagt ttttgaacct
<........................................L ORF........................................>
2221  gataactccc agaaccattca atacagagtct atcgtaaacg atagttcttc tttttacttt
<........................................L ORF........................................>
2281  gggctccagg aacacctgctc tcatcctcgc ccaaaagcta ttgtactcctg tataagcct
<........................................L ORF........................................>
2341  gctctatggga tctctctacta tcccaatgtc cttgtttttct catgttggcc aatggcattag
<........................................L ORF........................................>
2401  gagataaggtt tggagctca taagaaacca tgtgaagttt atgaagcct ggtctgttatt
<........................................L ORF........................................>
2461  gaattcatct ccccactcctt ctggaaactt actattcctga tatgtagata gaatcaaac
<........................................L ORF........................................>
2521  agtgcacag cggcagacaa gagcagcctgt tagcgcctct actatgtgtg taccattaaca
<........................................L ORF........................................>
2581  gaagttttccg gttgtgagtt tgggtttatatt tattctgcta cggagctttg gtaaataaaac
<........................................L ORF........................................>
F2_pUCRL4

ctccttcata taccaagt cgtctgtt cggattaag atatgtgat cagtcacaat
..............L ORF......................

ccttcctaca cctttgaaat tcaggtgccc ctggaaattt gggcggaca ttatgtcaaat
..............L ORF......................

acctctcag atagttcata gtagttcttg tttaggctga cgcctagtcg atgttcaccc
..............L ORF......................

cacacacagt acactaatcc cacaacttat agcatattgg aacacaagtt taatgtcctc
..............L ORF......................

atccgaaat ctctcactca accgttctgtt tataggaatag tgccttaaat gaaagttgtgt
..............L ORF......................

agttctatac gcggctactg atgatggaa atcctagggc atgctggctca ctgcaagctc
..............L ORF......................

gttagagaa ttgacgctca taaaccgagg tagaagtggt tttaacagct tatatggag
..............L ORF......................

acccagtttt ccaagcata ttaagttag agagcgtcgg agggcaagta cggctttgta
..............L ORF......................

gaccacctg attttgccata tggttttaa tttgtctttt ttttttcttg aataaggctcg
..............L ORF......................

tctgtgttaa acagctacag attttttttct tttggtcatt gaaactaccc atggtgcctt
..............L ORF......................

cggacctctt tttctctttg tttttttttc agctgagaat ttttctacaag ttataacctgg
..............L ORF......................

cttctgtgct attgttaact aataatatac taaagctagac aacatgcttg gagatgatac
..............L ORF......................

acctgagaac tccaccccc ccatgacat ttctctatatt tttttggaag gctgcacaca
..............L ORF......................

aagaatccga ttgtgtcctc tcagatatct ttgtgtcagtt atagggtccag actcaagtat
..............L ORF......................

tctacccaac aggcccaagt tgagcaacat cagttggaact gctctgtaaa tatctacacc
..............L ORF......................

acttatagca gaagttcttt ggaggataatt tggattggaat ttggatcggg caatcacgat
..............L ORF......................

tacaoctttt tcagctttggt ggaatggga ttcttcatata caccctcctca gcctcagttg
..............L ORF......................

gtacacaggt gatatatctt ccaatactgt gcctattttc tcctctattct gcctgaaatg
..............L ORF......................

cacagcacta tcaactgaaca aggggttggg tgcacaaggt agcacaactag ttaagtgcgtt
..............L ORF......................

ttaataatg tcgtttgtaa ccttagctcg ctctttctgca ccaataagttt gagggtctctt
..............L ORF......................

catgtagttg gtaagagttg ctcctatcatt ttggtctgtg ttcaacagtg ttaagaagaa
..............L ORF......................

F2_pUCRL4

3901 gttcatataa actctctcat ccatctcat tgctgttact agtgaacctca tgtgtgttaa
<................................L ORF....................................>

3961 cacctctgtg aaaaaatctg gtgtgcttct aataaaaactc ctataaaca caacaggttc
<................................L ORF....................................>

4021 cccctcccccc aactgcaatg gtatgcttcat gaacaagttt gacagtctgcc tgtcatctgc
<................................L ORF....................................>

4081 taacctgaag aatccttataa ctttggataa agttctatta agctcatgta atatgtgtgc
<................................L ORF....................................>

4141 acctgcaagt ctagatctt ttggcttgac aagacacttg atgtacaacc agaaattcgg
<................................L ORF....................................>

4201 aactataaga ctagtcacaa atgactcccc tggaaattct cacttggac agagacctcc
<................................L ORF....................................>

4261 aatagctttc atactagtct tgtatgtcat cagtatgttg tttatccaggg gacccacgct
<................................L ORF....................................>

4321 cagaaacctc ttaattgcag cagagatact tatgcctttct gatttggtaggg tcttgctcat
<................................L ORF....................................>

4381 aacctgagat cttctttgaaa ctaaagtctc cccctccttc aacttgtgctc ctatgttata
<................................L ORF....................................>

4441 gtaggcatcc ttcaacagcag ttaacatctt gatggcgaga ctataatcag cctgatttac
<................................L ORF....................................>

4501 agtttgagca cctgtgagtc gcacaggttt actcacatct atagactgttt tgtcctccatt
<................................L ORF....................................>

4561 cagcagtgat gtgagctgca cccgattctct cagacacacc acatacagga gtgatattgc
<................................L ORF....................................>

4621 ttccatggct cacatctttt gcaccacgcc tcctatgcca cccatgtgga acctatacag
<................................L ORF....................................>

4681 tcacctttgtc tcgggtatctg ggtctatgtg ctagatccca cctgtgtagc ggggagcatg
<................................L ORF....................................>

4741 tctataagtg cagatcatgg tgtgagagca tacagtcaga tgtaaccagc agaaaggtct
<................................L ORF....................................>

4801 ttgactccca tgcagtcggt ctagaacact gcagacacact gaagtagctct ctgatctaaa
<................................L ORF....................................>

4861 ggcttggtta aacctgctca aatcagcact tattgagca ctgtctatat aattttgta
<................................L ORF....................................>

4921 gctatcactt ttccgtttttt ttgtagacga tagttggagac ttgagcttaca tgtttcttgg
<................................L ORF....................................>

4981 taattctaga ttctcatacc tagtaatgtt cttgaaagaa aatsgcacta tggctacgta
<................................L ORF....................................>

5041 taaccaatttc tcagcaagga tctggacttg cttttcttttt ccaggttgca ttgcaaaccat
<................................L ORF....................................>

5101 acacacctaca ctagtccecc ttcccccccc tggtaaggag accacgctgt cttgctcatt
T to C mutation >
<................................L ORF....................................>

5161 gttcaatataa actctctcat ccatctcat tgctgttact agtgaacctca tgtgtgttaa
<................................L ORF....................................>
5161 caagtcactct gcctactacc catatatattt gagatcacttt tgtttaaat tagcatcttt
<..................................................<
5221 taagtagaag tctagcaccg tctctgttag cgtcatctct agtctgtcct gagactccaat
L ORF......................................
5281 cctggaatcct ctctgtatgt tttgcggaaat gtaggtaagt ggcataacgg accatagtgaag
<..................................................<
5341 aagatttggaa ggagatattg ctttatcatt cagtactatc tccaaactgg tctgttaggg
L ORF......................................
5401 tctataagaat ctttggccaa aactgacccc aggggtcctcc aagatccttt ccacacacaca
<..................................................<
5461 ttccagtggt tttgtgtatg ttttgcatt cagcattctc ttcccccttc gactccatgt
L ORF......................................
5521 tggatggatc tttatcttttg gcctactttt gtaggctcct acgaaaacctt tgcctactct
L ORF......................................
5581 cagaatggaag ggccccctca tctccgaag ggccaccaga ctttacattt ttgcagtccctc
L ORF......................................
5641 actgttatcc ccctactcatc ccatactcctt ccctcttcttct acacagcagat gctccgaaat
L ORF......................................
5701 gcggatataa aaatactct ctcgacgttc aagcaatcctt ttattccaca gaagcttaat
L ORF......................................
5761 tattttcgac ctatctcccct cttgcttccag ctattcatgc tggtcctggt atctatgattg
L ORF......................................
5821 ctccacgat gttttctgttg tccgccttcct gtagcgagct ctcctctcc ctatatttact
L ORF......................................
5881 agctatgcca ttcaaatcgct tattccggaat ctttacactg aattgtgcat gttcagtcctt
L ORF......................................
5941 ccgcacattt tgcgtcctata tggatcctct ccaatctcttt acaatcccat accccctcatt
L ORF......................................
6001 gcggatgtgcc ctcaataact catcagtcctc aagatcactt cgggtgagta actcaccaaaca
L ORF......................................
6061 taatttgcgc tctatactccat ggcggcttttg ggcaactattt aagcaattgc tcacccacac
L ORF......................................
6121 acaaaagtttt gcattaatcct tgttaaggtc aaggtcctttc caagtcagga gttgattgta
L ORF......................................
6181 ggtcacaataa catattcttc tacctttctgg acttatgatt atacacccga aagaacaaac
L ORF......................................
6241 aagaaaaaga aagttcagtc tattttctgg taatataagga atcctccttcgctattg
L ORF......................................
6301 aacacattaa gacattgatga cttcattcta ctatcaacaa gatcaaacc agttgccctagg
L ORF......................................
6361 ccagtcaggt aagttcttca cttgaatcag atctatttt tttcccccttc ctgggtgtccgt
L ORF......................................
7681  gtacagtatac actgcggcta tacatatcac cataagaaat ttgcacatgc aaacagcttt
<.............................G......................................<
7741  cagacttttct tgttcaggcc gttgttaatc cttccataatt tcttgtctga gtgcacacctg
<.............................G......................................<
7801  gtgcagtagt tggtgtgctcc tgggtggttg accaccaatc tctatctctc ccatgtagtt
<.............................G......................................<
7861  ttctgacactc ctctaaactc gactgttccaa ctatgacgct aacagtgaat gcaaatgtga
<.............................G......................................<
7921  ctgtgcaggt cagccccaaa gctgatagga ctatagccaa tatgtacctc ctccccgatgt
<.............................G......................................<
7981  ccagccgagaa ccccaactgca gtgtgatagtt cactgtgtgcc ctgaaccata tatagtttgg
<.............................G......................................<
8041  accccataga gatacttgtc cccgccgttc ttttaatta aataatactg ttgcctatta
A to G mutation >
T to G mutation >
8101  ttggcacta gtgttgccaat aatctattag ttgtagcagc acttagttgc tgtcccaattg
<<..................SH..........................................<<
8161  ttgggtcaga tgcattcctt gtaatcctcg ttgtgaccac aggtttttcc aacacactcc
<.............................SH......................................<
8221  accactaaat ctataaaatt caaatccccca cccctccccgct ccctcaagca gcttttagtt
<.............................SH......................................<
8281  ttgagatacc tgaacctcacc ttgcacaaat tttttctctta ccaggcagtc agggggagta
<.............................SH......................................<
8341  ctacgtcttt cattggttgt cagttaaagt tgattttctt gtagtgcctc tatcatgtta
<.............................SH......................................<
8401  ctgggggtag tatagattgc agctgtagtt gttgctgcttc ctgataaatt ctgcttgcag
<.............................SH......................................<
8461  ttgtggacct cctgcacctgt gaatgccagt ttccacagaca atccgatcgt agcaaaagg
<.............................SH......................................<
8521  aagatcaag ctaagattgc acaaatcaca gcaacacaac tcaactagat cctgcagcaa
<.............................SH......................................<
8581  gaattgcacc ttgacctgat cactgtccgt ttgaagcag tgtctgaccc aaggttgacg
<.............................SH......................................<
8641  gtgagatgcta ttatgacttg tccacacctt ttggacttt atagcttaac cggtttaaat
TT to CC mutation >>>
8701  taactaatta tataggttat atataagctaa aactcaaaa ttctttgcac actccacaca
<<<<.................................M2-2..................................<
8761  ttgcttcctt tacttagtgac catatttgct gctttcctta ttaaatcact
<<<<.................................M2-2..................................<
8821  ccagctcgca atactataat ccatatatcct acgaaccata cacaatccaa gttcattgca
<<<<.................................M2-2................................<
8881 cttaatttt gctgtcacc c ttctgcaggg tattaccatt gcattgtca ctctctggcc
<.....................................................M2-2......................................<<<
8891 ctgcagatag ctgtagatg atcttgccaa gtttttgag tttttctctg ggaagcggct
<.....................................................M2-1......................................<<
9001 tgaggcctgt gatcaaacct gcgggtatct ttcctcatatc aaacataagat agtatcagat
<.....................................................M2-1......................................<<
9061 tgtgtagcgc cacatgttta gggcttccaa gcattgaact tctggtagtt ttcacatcat
<.....................................................M2-1......................................<<
9121 tttctggag ctgttttatata atgttgatac ggcgtgatga ggctgctgat tgggtgattg
<.....................................................M2-1......................................<<
9181 tttgtttccc ttctatatata tttgtaaaca ctttgaacaa aaatcctgag
<.....................................................M2-1......................................<<
9241 ttctatatct ctgcttcgcc cctgatatca aggataggcc atctgtctctg tctgttgttc
<.....................................................M2-1......................................<<
9301 tcaataatgg gttaagcatg tagtttgctc tgaataagag cacatgtcgt ggcaggagcc
<.....................................................M2-1......................................<<
9361 aataatattg gttaaatgta caggatgac cccggttgca tttgcccggt atttcatatc
<.....................................................M2-1......................................<<
9421 tgccaggtatt ctcgctagttt attcctactt cttccaattt taataactta aactgacata
<.....................................................M2-1......................................<<
<..<F ORF...<<
9481 agccatgctg ctatgacctg tcgtgccccgg gtaatcatgc tttggttgg gtttccggac
<.....................................................F ORF......................................<<
9541 tctgggaacag taatatatca ctgcaagtta aaagaagatt cctagaatga ctagcactac
<.....................................................F ORF......................................<<
9601 tattgctata gcacatcccag cctgctcttt ggcctatbtgg cctagcagtt cgttagactt
<.....................................................F ORF......................................<<
9661 atctattaag tctctggatc tatctatact cttcaaatct tggtaaagtg caacattgaa
<.....................................................F ORF......................................<<
9721 ctgatcctca gggattataa attggtttaattgattcaca aactggagttct ttttatggt
<.....................................................F ORF......................................<<
9781 cctctgtccg cctacaacct tgctcaattgc ttaacaagttg ttataatgg tttatgttc
<.....................................................F ORF......................................<<
9841 agctttcgttg tgggaatgt gggtgcaccc tttgctagc gtttttatta tccctacttt
<.....................................................F ORF......................................<<
9901 atgtctacct atggagcaac ttacactctc taacaagac actagaccccc ctaggggggt
<.....................................................F ORF......................................<<
9961 taaggtcatc ataactgacag ggtgtctacc tgtgtgactt tgtcatgggt atttgatatg
<.....................................................F ORF......................................<<
10021 cgatatgtta tagttgcact ctctcaacttc tagggccaca ttaatgccag tgtctgtgct
<.....................................................F ORF......................................<<
10081 acaaaaataca taatcatccc ttacctacaa atcatctca ttaggataat aagctgtaga
<.....................................................F ORF......................................<<
 GCCAGCATTT GTACAGTACC ACCCTTGATC TTCTCTCAGT ATAACAAGCAT AATTGCCTTT
\[\text{ORF}\]

 CTTTTCTCTA CAGAGTGTTG CTGCCAACC CCTCCAAACAA GGTGTCTCAA TCACGCGAA
\[\text{ORF}\]

 TATGGGCACT TGTAACATAT AAACGACGTT TCCATCAA ACACAAATCA ATATACAAA
\[\text{ORF}\]

 CCCCTTCCTT CTAACCATGG CACAGTGTT CCACATCAA CTAATCGTC TGATGAAGT
\[\text{ORF}\]

 TGCCATTCTG TTAATGGTC ACAAAGTTC ATCACTGGTC ATTAATCAA GAGACACAGC
\[\text{ORF}\]

 TGATGTGATA CTCGCAATG CAGAAGTTG CTCACCAAC TACGGAAACC TTCTGATTTT
\[\text{ORF}\]

 TTGGCCAAAA CTAATTGAGC TCTTTATATC TGATATTG CATTGGTCT GTGATAAGC
\[\text{ORF}\]

 AGGAGTCAAT TTTTTACTTA TAAATTCTT TGGGTCAATG ACTGCAGTTG CTAGTACCT
 T TO C MUTATION
\[\text{ORF}\]

 CACACATTC CCTAATGTGG ATACTGCCTC ATTTGTGTC CCGAGGCACT TCTTAATGAC
\[\text{ORF}\]

 CTTCACCTCT CCTCTCAATC TAATTGCTC TGCAAGTCTG ACACAGCTG TTACGGCAAGC
 T TO C MUTATION
\[\text{ORF}\]

 AGCAATCGCA AACAACAGTGT CAATGGCAAC CGCACAAGAT CTAACGTTCC TTGGTGAGGA
 > T TO A MUTATION
 > T TO C MUTATION
\[\text{ORF}\]

 TAGTCTGCTT TCTTAGCCA CTGGACAGC TGACACTGTT TTGAGCTCC TCAAAGCATT
\[\text{ORF}\]

 CCTTGAGAT ACTAACTCAG TGCACATTAG GCTGGTGCCA TCAATGCAAG TGTGTTCTC
 C TO T MUTATION
\[\text{ORF}\]

 AACATCCCT ATTTCAAGGT TAAATACATT CGTATAACAC CCTGCTCTTA ACACACTTT
\[\text{ORF}\]

 ATAACCCTCA GTTACAGTC TGACAGGATT TTCAATGAT GTTTCTGATA TGCACTACT
\[\text{ORF}\]

 AGGATTAGAT TAAAGGAAC AATAGAGACA GATTCTTACA TCCATACATT GGCTCCCAAT
 C TO T MUTATION
\[\text{ORF}\]

 TTTTTATTGA CTATAACATT AACAGATTTA TGCTTTAGGG TACCCTCAAC GTGACCTCAG
<< M. <<

 TACATACTG GTTCTTTGAT GTATCCATGT TTCCCGAGAC TTCTGCTTG
<< M. <<

 TACATACTG GCCAGTTCTA CTATAACTTG CATACCTGCA CCCCATTCT TAAAGATCCC
<< M. <<
F2_pUCRL4

11281  tttagggtg ttcatgttca tatcaagat cagccctgca tatggagcta tccttgcttg
      <...............................M...............................<

11341  agttatagcc tggctgttct ccccaacttat tcgcatcttc acagttgctg atctctagtc
      <...............................M...............................<

11401  cttgatagag acagcttttaa tatatgcagg tatggttatg gcacccccct tctctatgtc
      <...............................M...............................<

11461  tataaaatca cacaatgcca tcaagtcatg agtctttcct ccaacagacc tcacatctgt
      <...............................M...............................<

11521  cattatgtc gagaccatgc catatggttt caaggtagtc agataaatgg acctcacaatc
      <...............................M...............................<

11581  acacacagtc aatgtccccg acatccagct actgtactca tcaagtgcta ctgcaagcact
      <...............................M...............................<

11641  gacggcaaat ttcctaggta gtcagacat ggcagcacct tgggctgtcg cattgacctg
      <...............................M...............................<

11701  gacgcacaggg cttgtccgggg atgtctgtgt cttgggtgttg attgataag ttttcagtgtg
      <...............................M...............................<

11761  atctaaaac actgsgtcag gcgtacttga tgggaatag ggaacaccata cagtcagtgt
      <...............................M...............................<

11821  tgcaggtta gagtcttttt caatggagtc aacctgaacg gcttgacgtgt agggacccc
      <...............................M...............................<

11881  ctgataagtg tcaattatat aagactccat tttgacctgt cccgctagca ttttttcata
      ..........M.................<<  addition >>..>>

11941  actacagatc aagattgtat atgtccttct cttccctatc agattcagac tctccactct
      <...............................P...............................<

12001  cttcttcact ctctgacctt ccaacttgac tcctgtgcttc taggaaccttg ttcaactctt
      <...............................P...............................<

12061  tgcttttct gcgaagcttc acgtacctat ctctcatatt tgcctctctg gttctctctt
      <...............................P...............................<

12121  cttcctcatct cttcagactt tttcttttag cttctgtcct gatgtcgttg atcagctctt
      <...............................P...............................<

12181  cttctcatccc tatcattgca tcaagtattc cattctgcgc ggcggttgcc cagcagcattg
      <...............................P...............................<

12241  caatattcga ttgctttgac atgccccata tcatgtcata cttctctctt atgctttctta
      <...............................P...............................<

12301  atcttgcttc atatagaggt gatgtgtat ctttttcctc aatgtcaat attgatgact
      <...............................P...............................<

12361  tttctcatct atctcttctc gagatctgaca acagctctag cgccctctct tccagtttag
      <...............................P...............................<

12421  tatatttttc tacatctgggg tttgtaaaact tgaaccttatt agaccttttt gcaagacctcag
      <...............................P...............................<

12481  tactggtaga aagtctctgtg tcaagtggcgt cggagttttt gatatatag
      <...............................P...............................<
F2_pUCRL4

12541 attctatcc tttgaccttg gggacaggag cttgttagg tttatacag gatccctttg
<............................P...............................<

12601 gttggttggg gttgcacaca ggggttaatg gcaccttttc acaatttgt ttaatagct
<............................P...............................<

12661 ccccaacta aagaagcact ccagcagagg tggctctctt tgaatctctg taggcatcgg
<............................P...............................<

12721 ccatttttgc agcctgcttg cccatctta gataatctct gcttccaggg aaagacattg
<............................P...............................<<

12781 ttacctgtcc ccggccgcca tatttttaa ttacctaat tggatgatc tctcatcactc
addition >>>..............
<...............................N...............................<<

12841 tcccccctagg taggatgttg ctgctctccct ctctcatctca tgtgaccccta aggcgcctag
<............................N...............................<

12901 gttgatcttg tgttctcttc tcaatgtcgg tgcataacct tcagcagag caaacagctc
N.............................N..............................

12961 aaggttgagg gctcgacctt tgtcatgcc tataatacca agggggcggg cgttccccag
<............................N...............................<

13021 gacaacacgc gcacaaatttg gcacgttgtg caatgatataa agacccgctt tggaactctg
<............................N...............................<

13081 gcgtgactgt atgaacccgt acacgctcct catttccctc actaggtcat agacctcaga
<............................N...............................<

13141 tacctgcctc aacctcagctt ggacagatat atggcccaac attatatgtg tggagattcg
<............................N...............................<

13201 tgcataagca cccgcgcgca gcattgttgg cccctgcccct ccagctggca taaaatatt
<............................N...............................<

13261 cacaacagg ctccctctcc tgcctccgggt ggttgtacttt cgcagtgctt tacagttatcc
<............................N...............................<

13321 aataaataga ttctctgaat acetctcttt ctcaaaacatt ctaaagaatg attggcaat
<............................N...............................<

13381 ccttggtatg tccatctccg ggtaccggtga tttggcatcg cttaaatccc ttgacgacag
<............................N...............................<

13441 ccggatagca gtgtcagggc ctacctcaac agttgagggc aacttggtga aatgagggc
<............................N...............................<

13501 accaatgcat aatagaaata ccggagcact caaggcagac gccttttggt tctttgtgag
<............................N...............................<

13561 ttgccatcttt tccctacacc tttgctcccct tggctctctt tgaatgctct gcagracatc
<............................N...............................<

13621 agcccaacca aacacactag ctgtaatgtt tagcaacctct gcctttctga cagaacactc
<............................N...............................<

13681 tgccgaaggt tgtacaatct tgtgcataac accctgtact tcctacccgg aatcttccag
A to T mutation >
<............................N...............................<

13741 tatgttggag gttctacacg ctcttagagc agtactaatg tactgctagc ctacccctg
F2_pUCRL4

13801 tgcaggctca tagtcgggtg gttttgcaaa caacaccata ccgcataatg tggatacttg
<.................................N...................................<

13861 cggctgcaagt tcggaggttg tgatgcagct ggtacaccga acatcccttc taattgtata
<.................................................................<

13921 ctgagagttct tcaagaattg catggtggta ctccaagctca ctgagtctaa tactttcaag
<.................................................................<

13981 agacatatcgg atttggacttg tcccatatttt tttctagaaac ctggttggaat gcgttttttt
<.N.<<

addition >>..>>

14041 ctctgtcttat agtgagtcgt attacgcgta aagtggcgct aatcatgtgc atagctgttt
14101 cctgtgtgaa atttgttatcc gtcacaattt ccacacaaca tacgagccgg aagcataaag
14161 tgtaaagcct ggggtgcccta atgagtgagc taactcacaat taattgctgtt ggcgetcactg
14221 cccgctttccc agtccggaaa cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcgc
14281 gggagagggcg gtttgctgtat tgggcgcctct tccgcctcct cgcgcactga ctgcgtgcgc
14341 tcgtgctgttc ggtgctggcgcc agcggtatca gtcactcaa aagcggttaat acggttatcc
14401 aacagatcag gggataacgc agggaaagac atgtgagcga aaggcagaca aagggccagg
14461 aaccgtaaaa aggcgcggctt gttggctttt ttccataggc tccgcctcctc tcagagcat
14521 cacacaaatcc gaccgcctag tcagagttgg cgaacccgca caggactata aagatacag
14581 gcttttccc ctcggagatcc cctcgtgcgc tctctgttgc gcacccgcgc gttacccgga
14641 taccctgctgc ccttttctcc ttgaggagcc gttggctttt ttctatagtc gcgctgtagg
14701 tatctcagtt cgggtgtcgg ctctgtgcct ccagctggct cgggtcagca agcccctgctt
14761 cagcgcagcc gcgcggccctt atccgcttaac tattcgctttt gtagtccacc ggtagagacac
14821 gacttacgc cacgcggccgc agccactcgt gacagagtaga gcagagcgag gtatgtaggc
14881 gggtgcacac agttctttga gtttgtgacct aactacgcgt aacactagga aacagtatttt
14941 ggtatctgcg ctctgtgcga gcacgccctt tccggaaaaa gatgggttgtg ctcttgccct
15001 ggcaaaaca ccacgcctgg tagcgtgcttg ttttttttt cgaagccgcgc gattacgcgc
15061 agaaaaaag gatctcaaga agatcttttg ttctttttct cgggtttcgt cgtctcaagtgg
15121 aacgaaact cacgctttag aaggtttttgct aatgaggattt caaagaggt cttcacttag
15181 atctttttta attaaaaatg aagtttttaa tcactcttaa gtaatatagga gtaaacttgg
15241 tctgacagtt acaaagtctt atacagctag gcaccaactct cagctgtctt ctctattctgt
15301 tcacccattt ttgcctgact ccgcgtcgtg tagataacta cgatacggga gggcttacca
15361 tctggccccc gttgctgctat gataccgcga gacccacgct ccagtcgtcc agatttatca
15421 gcataaaacc agccggccgg aaggggcgcag cgcaagactg gtcctgcaac ttatatcgc
15481 tcctaccaag ctacatattg ttgccgggaa gctagagtaa gtagctcgc gcgttaatgt
15541 ttgcgcacagc tttttgcctg tacgtcagcc atcggtggtg gcgcctgcgt gcgttgatag
F2_pUCRL4

gcttcattca gttcgggttc ccaacgatca aggcgaattc catgatcccc catgtagtgc
aaaaaagcgg ttatgtccttt cggtcctccg atcgttgcataa gaaatagttt ggccgcaagtgaattggg
TTATCACTCA TGGTTATGGCATGCCTGCATATAATCTCTTTA CTGTCTGCG AATCATGAAGAAGGCTCTGCTGC
GGTTGCGACCC GATGTATA TCCGCGAAAC TCTCAAGGAT CTACCCTGCTG
AAAGTGCTCA TCCATGGAAA ACCTCTCTGG GGCGAACA AATCAAGGAT ATTACCCA AAGTGCACCT CTACCCTGC
TTGAGATCCA GTTCGATGTA ACCCAGCGT GCACCCAACT CATCTCCAC ATCTTTACT
TTACCAGCG TTTCTGGGTG AGCAAAAAA AAGGCGAAA ATGCCGAAA AAAGGAAATA
AGGGCGACAC GGAATGTTG AATATCTCATA CTCTTCCTTT TCCAATATTA TTGAGCATT
TATCACGGTT ATATGGTCAT GAGCGGATAC ATATTTGAA GATATTAGAA AAAATAAACA
ATAGGGGTTC CGCGCAATT TCCCCGAAAA GTGCGAACCTG AGTCTAAGA AACCATTATT
ATCATGACAT TACATCATAA AATAGGCGT ATCACCGGTC CTTTCGTC
Appendix D: Published Data
Development of a reverse-genetics system for *Avian pneumovirus* demonstrates that the small hydrophobic (SH) and attachment (G) genes are not essential for virus viability

Clive J. Naylor,1 Paul A. Brown,1 Nicole Edworthy,2 Roger Ling,2 Richard C. Jones,1 Carol E. Savage1 and Andrew J. Easton2

Avian pneumovirus (APV) is a member of the genus *Metapneumovirus* of the subfamily *Pneumovirinae*. This study describes the development of a reverse-genetics system for APV. A minigenome system was used to optimize the expression of the nucleoprotein, phosphoprotein, M2 and large polymerase proteins when transfected into Vero cells under the control of the bacteriophage T7 promoter. Subsequently, cDNA was transcribed from the virion RNA to make a full-length antigenome, which was also cloned under the control of the T7 promoter. Transfection of the full-length genome plasmid, together with the plasmids expressing the functional proteins in the transcription and replication complex, generated APV in the transfected cells. The recombinant virus was passaged and was identified by cytopathic effect (CPE) that was typical of APV, the presence of a unique restriction-endonuclease site in the cDNA copy of the genome and immunofluorescence staining with anti-APV antibodies. Replacement of the full-length wild-type antigenome with one lacking the small hydrophobic (SH) protein and the attachment (G) genes generated a virus that grew more slowly and produced atypical CPE with syncytia much larger than those seen with wild-type virus.

INTRODUCTION

Avian rhinotracheitis is an important disease that is responsible for major economic losses in domestic poultry throughout most of the world (Jones, 1996). The most severe clinical effects are seen in turkeys, but there are significant economic losses in commercial chickens and it has been implicated in swollen head syndrome. In the 1980s, the combined efforts of several groups showed the cause to be a pneumovirus, named *Avian pneumovirus* (APV) (Jones et al., 1986; Wilding et al., 1986; Cavanagh & Barrett, 1988; Collins & Gough, 1988; Ling & Pringle, 1988), which is now classified as a member of the family *Paramyxoviridae*, subfamily *Pneumovirinae*, genus *Metapneumovirus*. It is distinguished from members of the genus *Pneumovirus* by its different gene order (Ling et al., 1992; Yu et al., 1992) and the absence of non-structural protein genes NS1 and NS2 (Randhawa et al., 1997). Recently, human metapneumovirus, a second member of the metapneumoviruses, was described, which is responsible for serious respiratory disease in children (van den Hoogen et al., 2001, 2002, 2003).

Manipulation of the genome and study of genome function in APV have been limited, due to the absence of a methodology whereby precise changes can be made. Reverse-genetics systems have been developed for other viruses of the family *Paramyxoviridae*, which have enabled specific mutations to be introduced into the virus genome and the subsequent phenotypic consequences determined. These have included *Human parainfluenza virus 3*, *Human respiratory syncytial virus* (HRSV), *Bovine respiratory syncytial virus*, *Newcastle disease virus*, *Sendai virus*, *Rinderpest virus* and *Simian virus 5* (Marriott & Easton, 1999). These analyses have identified many important common features of the molecular biology of these viruses. However, analysis of HRSV by using reverse genetics has identified several aspects that differentiate it from other paramyxoviruses. Rescue studies with HRSV have shown that the minimal replicative unit is comprised of the nucleocapsid protein (N), phosphoprotein (P), M2 protein, RNA polymerase (L) and full-length viral genome in the antigenome sense (Collins et al., 1995). This paper describes the development of a reverse-genetics system for...
APV. The system was used to produce APV that entirely lacked the SH and G genes, indicating that these genes are not essential for viability in tissue culture.

**METHODS**

**Preparation of a full-length DNA copy of the APV genome.**

A full-length DNA copy of the APV genome was constructed in a series of PCR and ligation steps. The sequence of the APV genome is available from GenBank under accession no. AY640317. The cloning strategy is illustrated in Fig. 1 and all oligonucleotide primers used are shown in Table 1. RNA was extracted from Vero cells infected with APV (strain LAH A; Lohmann) by using an RNAeasy kit (Qiagen). Two overlapping cDNAs were generated by reverse transcription (RT) for 1.5 h at 42 °C by using Superscript II (Invitrogen) (Fig. 1a). These RT reactions used the antigenomic sense primers APVLead and M2 start, corresponding to the leader (Invitrogen) (Fig. 1a). These reactions were used as templates for eight further PCRs using Pfu polymerase (Stratagene) (Fig. 1c). These products, APV1, APV2, APV3, APV4+5, APV6, APV7, APV8 and APV9+10, were generated with primers listed in Table 1 that created SalI or XhoI sites at each end of the products. Digestion and ligation of these sites regenerates the viral sequence whilst eliminating the cloning sites, enabling insertion of the next fragment by using the same method.

A T7 promoter sequence was added to the leader end to generate APV T7 and a Hepatitis delta virus ribozyme (HDVR) sequence was added to the trailer end to generate APV9+10HDVR. The HDVR was amplified from pOLT5 (Peters et al., 1999) by using the HDVR start and HDVR back primers (Fig. 1b). This product was ligated onto APV9+10 and the correct construct was amplified with APV11.9 XhoI+ and HDVR back.

APV T7 and APV9+10HDVR were cloned into the low-copy plasmid pCTPE (a modified version of pOLT5 with HDVR, T7 terminator and lacZ sequences removed to improve cloning efficiency; Fig. 1d). The remaining PCR products were cloned into a modified form of pUC18 with the SalI site changed to an EcoRI site by site-directed mutagenesis (QuickChange; Stratagene) (Fig. 1d). Site-directed mutagenesis was also used to correct coding changes that were identified by sequencing (Imperial College School of Medicine, London, UK) and to introduce an SfiI restriction site into APV3. The latter introduced conservative (lysine to arginine) changes at aa 299 and 300 of the predicted F protein.

Sequential ligation of SalI/XhoI-cut fragments generated clones containing APV T7+3 or APV6–10HDVR (Fig. 1e). Attempts to clone APV4+5 into these constructs were not successful. PCRs on APV T7+3/APV4+5 and APV4+5/APV6–10 ligation carried out in the presence of SalI and XhoI were performed by using primers CTPE 240− and G11− or M2 mid and CTPE 190−, respectively (Fig. 1f). Exonuclease I (USB) was used to remove remaining oligonucleotides and an overlap PCR was used to join the two products, with T7 APVLead2 and CTPE 110− being added after two cycles (Fig. 1g). No primers were present in the first two cycles, in order to allow the fragments to anneal and copy each other from the overlap prior to the addition of the PCR primers. A 15-5 kb product was obtained, which was circularized by ligation and transformed into Escherichia coli (Fig. 1h). No colonies were identified by sequencing (Imperial College London Medical School service, London, UK) and to introduce an SfiI restriction site into APV3. The latter introduced conservative (lysine to arginine) changes at aa 299 and 300 of the predicted F protein.

After cloning, the full-length cDNA copy of the genome was modified between positions 3828 and 3831 (antigenome sense) by site-directed mutagenesis (QuickChange; Stratagene) to introduce a convenient SfiI restriction site. This also resulted in a conservative substitution at aa 299 and 300 of the F gene, with lysines being converted to arginine residues. Confirmation of the presence of the mutation after passage of the recombinant virus was obtained by RT-PCR amplification of the region of the genome between nt 3625 and 4063 using primers F6+ and F7-B (Table 1).

A defined deletion mutant of APV was generated by high-fidelity PCR (10 cycles) amplification of the full-length clone to introduce a deletion in the region of the SH gene. The amplicon was designed to terminate at genome position 5363, 1 nt prior to the SH gene transcription start signal, and to restart at position 5965, 2 nt prior to the G gene transcription start signal. Long-distance PCR (Thiel et al., 2003) was employed to amplify the full-length genome clone so as to remove the entire SH gene by using primers SH omit+ and SH omit−. The PCR mixture was run on an agarose gel to reveal solely the expected band (approx. 15.0 kb). This was self-ligated and used to transform competent cells. Resulting colonies were screened by multiplex PCR and restriction-enzyme analysis and a single clone from almost 1000 was identified as intact. This was sequenced across the M2-G gene junction region and it was found that several bases had been lost from the start of the G gene. The amplicon terminated at position 5363, as expected. However, the sequence unexpectedly continued from a point 12 nt into the ORF and hence the G gene lacked its transcription start, translation start and a small section of coding region. The entire sequence of the mutant genome was sequenced to confirm that this deletion was the only mutation present.

**Preparation of plasmids expressing the virus replication proteins.** N, P and M2 gene sequences from strain APV-A were amplified by PCR and cloned. RNA was extracted (RNasy; Qiagen) from infected cells, reverse-transcribed (Superscript II; Invitrogen) and amplified by PCR (Pfu polymerase; Stratagene) using primers that amplified the gene to include the start codon (N and P) or introduced the T7 promoter sequence (M2 and L) immediately prior to the start codon of each gene, as shown in Table 1. The downstream primer in all cases terminated beyond the gene’s stop codon. N and P genes were cloned into the Smal site of pCI (Promega) downstream of a bacteriophage T7 promoter, allowing gene expression under the control of the cytomegalovirus promoter. T7M2 was cloned into the Smal site of the modified pUC18 mentioned above.

The I gene was cloned in sections into the EcoRV site of pCTPE by using the sequential approach used for the complete viral genome. In order, APV9+10, APV8, APV7 and APV2 were ligated into pCTPE. It proved impossible to add the final section of the gene (APV7LStart), representing approximately the first 400 bp of the gene, in a similar manner. However, the full I gene, together with the pCTPE plasmid, was prepared as a single PCR product (Pfu polymerase; Stratagene) following their ligation. The blunt-ended product was circularized by ligation and the mixture was treated with the restriction endonuclease DpnI to remove the original methylated plasmid and leave only the desired product for transformation into E. coli. Three resultant clones were sequenced fully and were found to be free of coding errors.

**Testing of support proteins in a minigenome.** The four support genes were tested functionally by using a cloned APV minigenome in which the virus genome had been modified so that all genes were replaced by a single copy of the chloramphenicol acetyltransferase (CAT) reporter gene, flanked by the virus leader and trailer regions. This adopted the method of Randhawa et al.
(1997) except that, in this instance, bacteriophage T7 polymerase was generated by a Fowlpox virus recombinant (FPT7; Britton et al., 1996) at an m.o.i. of 1 p.f.u. per cell and lipofectamine 2000 (Invitrogen) was used to transfect DNA into Vero cells. CAT reporter-gene expression was measured 48 h after transfection by ELISA, as described previously (Ahmadian et al., 2000).

**Fig. 1.** Schematic representation of cloning strategy for generation of a cDNA clone representing the APV genome. The primers used in specific steps are given in parentheses. The position of *SalI* (s) and *XhoI* sites (x) at the termini of fragments are shown. Details of the procedures are given in the text.
Virus rescue. The method adopted followed that used for the minigenome, except that full-length antigenome (wild-type or deletion) replaced the minigenome and incubation times were increased to allow time for any virus to produce detectable cytopathic effect (CPE). Fowlpox virus will destroy most avian cells before this time, but it is unable to package in mammalian Vero cells; hence, additional

<table>
<thead>
<tr>
<th>Fragment generated/primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV71</td>
<td></td>
</tr>
<tr>
<td>APVLead</td>
<td>CGAGAGACCGATTCAAGGCAGG</td>
</tr>
<tr>
<td>APVLead ext</td>
<td>ACGAGACCGATTCAAGCCAGCTTCTTCT</td>
</tr>
<tr>
<td>T7 APVLead1</td>
<td>TAAATACGACTCTACATAGGACGAGAAAAACGATTCAAGGCAGG</td>
</tr>
<tr>
<td>T7 APVLead2</td>
<td>TAAATACGACTCTACATAGGACGAGAAAAACGACATCAGG</td>
</tr>
<tr>
<td>APV 1.1 Sal−</td>
<td>CTTCAAGTGTTCGGGTTGCGTGCACC</td>
</tr>
<tr>
<td>APV2</td>
<td></td>
</tr>
<tr>
<td>APV 1.1 Xho+</td>
<td>GCCATGTTACAAAGGCTCAGGCC</td>
</tr>
<tr>
<td>APV 3.1 Sal−</td>
<td>CATTGCAATGTTGTTGTCGACATTTC</td>
</tr>
<tr>
<td>APV3</td>
<td></td>
</tr>
<tr>
<td>APV 3.1 Xho+</td>
<td>GCCAGGTATTTCGGAGGACACTTACGGC</td>
</tr>
<tr>
<td>APV 4.6 Sal−</td>
<td>GCAGAGATTTCGCGTGCAGTACCTTC</td>
</tr>
<tr>
<td>APV4 + 5</td>
<td></td>
</tr>
<tr>
<td>APV 4.6 Xho+</td>
<td>CAAGTGAGATCTCGAGGCGGAAATTC</td>
</tr>
<tr>
<td>APV 7.6 Sal−</td>
<td>GATCGTATTTCAACTCGAGAATCTAACCACGACC</td>
</tr>
<tr>
<td>APV6</td>
<td></td>
</tr>
<tr>
<td>APV 7.6 Xho+</td>
<td>GCTCGATATTCAACTCGAGAATCTAACCACGACC</td>
</tr>
<tr>
<td>APV 8.2 Sal−</td>
<td>GGGTATCTAGATGTGTCGACAGATGTG</td>
</tr>
<tr>
<td>APV7</td>
<td></td>
</tr>
<tr>
<td>APV 8.2 Xho+</td>
<td>GCCGGGACACAGAACAACATCTCTCAGGCAC</td>
</tr>
<tr>
<td>APV 9.9 Sal−</td>
<td>GCTATGATTTTTCCCGATCGACAAAGCAC</td>
</tr>
<tr>
<td>APV8</td>
<td></td>
</tr>
<tr>
<td>APV 9.9 Xho+</td>
<td>GTACATCCAGTGTCTTTTCGTAGGG</td>
</tr>
<tr>
<td>APV11.9 Sal−</td>
<td>CAGTGAAGTGTTCGTTCGACATG</td>
</tr>
<tr>
<td>APV9 + 10</td>
<td></td>
</tr>
<tr>
<td>APV11.9 Xho+</td>
<td>GCTGAGGTTGACATAGTCAGGC</td>
</tr>
<tr>
<td>APV Trail ext</td>
<td>ACGAGAAAACACCGTATATCATCAAAATTTTACC</td>
</tr>
<tr>
<td>HDVR start</td>
<td>GGGTCCGCACTGACATCCTC</td>
</tr>
<tr>
<td>HDVR back</td>
<td>CTTCTCGCTAGGACCTACG</td>
</tr>
</tbody>
</table>

N gene

| N start +                 | GTCAAAATGTCTCCTGGAAG                     |
| P start −                 | CAGGGAAGAGACATGTTCAC                     |
| P start +                 | GTAACAAATGTCTTTCCGG                      |
| P stop − ext              | GACCTTGCCATTATATTTTTACTACAGATCAAG        |

M2 gene

| T7-M2                     | TTAATACGACTCTACATAGGACGAGAAAATACCGAGGAG  |
| M2-1st −                  | GCATTGCAATTTATTTTATGTCGACCCCC           |
| M2 start                  | GATGCTCATGGCGAATTC                      |
| M2 mid                    | CCAGAGATTCAATGCTTTTAAAAGACCC            |

Additional primers

| T7L                       | TTAATACGACTCTACATAGGACGAGAAAATACCGAGGAG |
| CTP 240−                  | GGCAGCACTGTTCTTTCTCGCC                 |
| CTP 190−                  | CGAGATCCTTAGCGCCGACCTTC                |
| CTP 110−                  | CGCCGACATACCGGTGGGGAAG                 |
| G1−                       | CGAGATCCTAGCCTTAAACATGAGGAATGG         |
| SH omit +                 | TCGGGGACAAGTATCTCAATGGGG               |
| SH omit −                 | CTCCTCTTTGGACTATGCTCAATTGTGTTTT        |
| F6+                      | GTTAGAGCAATTAACAGAATGCC                |
| F7-B                     | CTACCATACTGACGGTGTC                    |

Table 1. Sequence of primers used in the generation of PCR fragments
means of limiting *Fowlpox virus* replication were unnecessary (Britton *et al.*, 1996).

Control transfections of two types were also prepared whereby some cell sheets, containing 10^6 cells, were inoculated with *Fowlpox virus* (at an m.o.i. of 1 p.f.u. per cell) and transfection reagent, whereas others received solely the latter. The transected cell sheet was viewed, freeze–thawed and filtered (0-2 μm) to remove cellular debris and to eliminate any possibility of effect from residual *Fowlpox virus*. The clarified material was used to infect new cells. Resultant cell sheets were examined daily for signs of CPE typical of APV.

If CPE was not evident, material was passaged further. When CPE was detected, RNA was extracted for RT-PCR by using an RNase kit (Qiagen) with the optional DNase digestion step. RNA was isolated 4 days post-infection. The region of the genome encoding the F gene containing the novel SstII mutation was amplified in the case of both full-length and deleted recombinant genome-derived material and a further PCR was used to amplify between the M2 and G genes in the case of the latter. All RT-PCR products were generated by using primers that were designed to anneal initially to genome RNA and were sequenced. Standard PCR conditions used were 10 s at 94°C, 20 s at 50°C and 60 s at 72°C for 30 cycles using Taq DNA polymerase (Promega). The F gene-derived products were incubated with the restriction endonuclease SstII to assess their cleavability.

Confluent Vero cells were inoculated with the passaged material. After 4 days, these were incubated with APV-specific polyclonal antiserum that was raised in turkeys. After washing, this was further incubated with fluorescein isothiocyanate (FITC)-linked anti-turkey antiserum that was raised in turkeys. After washing, this was further incubated with APV-specific polyclonal antiserum that was raised in turkeys. After washing, this was further incubated with APV-specific polyclonal antiserum that was raised in turkeys. After washing, this was further incubated with APV-specific polyclonal antiserum that was raised in turkeys.

All RT-PCR products were generated by using primers that were designed to anneal initially to genome RNA and were sequenced. Standard PCR conditions used were 10 s at 94°C, 20 s at 50°C and 60 s at 72°C for 30 cycles using Taq DNA polymerase (Promega). The F gene-derived products were incubated with the restriction endonuclease SstII to assess their cleavability.

Confluent Vero cells were inoculated with the passaged material. After 4 days, these were incubated with APV-specific polyclonal antiserum that was raised in turkeys. After washing, this was further incubated with fluorescein isothiocyanate (FITC)-linked anti-turkey conjugate to detect infected cells (Baxter-Jones *et al.*, 1986) and then viewed by using fluorescence microscopy and photographed.

**RESULTS AND DISCUSSION**

The minigenome construct was used to identify functional copies of the genes encoding components of the virus ribonucleoprotein complex that are responsible for RNA synthesis. Plasmids expressing the APV N, P, M2-1 and L genes and the minigenome plasmid were transfected into cells that had previously been infected with FPT7. After 48 h, expression of the CAT reporter gene in the minigenome was assessed by ELISA. Fig. 2 shows that cotransfection of all four support-gene plasmids directed the synthesis of CAT protein. In the absence of the L gene plasmid, no reporter-gene expression was detected, as shown in Fig. 2. Optimum levels of the various plasmids to achieve maximal expression of the CAT gene were established by using this system and were determined to be 400 ng plasmid encoding the N protein, 100 ng plasmid encoding the P protein, 200 ng plasmid encoding the L protein and 10 ng plasmid encoding the M2-1 protein per well of a 12-well plate. Whilst expression of the CAT reporter gene was observed in the absence of the M2-1 protein, levels were low and expression was enhanced at least 100-fold in its presence. This is similar to the situation seen with HRSV (Collins *et al.*, 1995).

A full-length copy of the APV genome was constructed in such a way that a unique restriction-enzyme site, not present in the sequence of the wild-type APV, was inserted into the genome. This generated an SstII site within the F gene coding region and altered the lysines at aa 299 and 300 to arginines. The plasmid carrying the modified full-length genome was under the control of the T7 promoter in such a way as to generate a positive-sense antigenome RNA. The plasmid was transfected into FPT7-infected cells together with plasmids expressing the APV N, P, M2-1 and L genes, as before. After the initial transfection, it was not possible to clearly identify areas of CPE, so at 6 days, cells were harvested and freeze–thawed and clarified supernatant was used to infect a fresh cell monolayer. Cellular changes typical of APV infection of Vero cells became apparent after 3 days, with patches of syncytial CPE visible across the entire cell sheet. After a further 2 days, this had led to generalized CPE that, 1 day later, led to multifocal destruction and detachment of the cell sheet. After harvesting this second passage in Vero cells, the titre was approximately 10^6 TCID_{50} ml^{-1}, which is typical of the yield obtained with Vero cell-adapted APV field isolates (Naylor & Jones, 1994).

RNA was extracted from the passaged virus in the presence of DNase I and a 438 bp region of the F gene containing the altered sequence was amplified by RT-PCR and sequenced to confirm the presence of the altered bases between positions 3828 and 3831. In addition, the product was digested with SstII and the two expected cleavage products were clear on an agarose gel, as shown in Fig. 3. A control PCR in which the reverse transcriptase was not included did not produce any DNA fragments (data not shown). These data indicate clearly that the passaged virus was derived from the cloned cDNA and that the amplified
When unfixed cell monolayers that had been infected 3 days previously with the fully competent recombinant virus were incubated with APV hyperimmune serum and stained with anti-turkey–FITC conjugate, intense fluorescence that was absent from the uninfected controls was observed (Fig. 4).

In order to extend these observations and to investigate the requirement of specific genes for growth in vitro, a second APV genome construct was generated in which the entire transcribed region encoding the SH gene was deleted. The sequence of the entire insert in the plasmid bearing the recombinant virus genome was determined and showed that the cloning procedure had also generated an extension of the deletion to include the conserved G gene start, together with the first 11 nt of the G gene coding region. The deletion in the G gene included the translation initiation codon. No other unexpected sequence alterations were present. The plasmid was transfected into cells together with the support plasmids, as before. Obvious CPE was not detectable initially in the transfected cells but, following passage of clarified cell extracts, very small syncytial areas were seen after 6 days. When this was passaged again, the number of syncytial patches increased, but still required 6 days to be readily visible. In contrast to the wild-type rescue, neither generalized CPE nor monolayer detachment was observed and this was considered to be a clear indication of altered behaviour in cell culture. When passaged a third time, giant syncytia could be seen (Fig. 5). This CPE is strikingly different from that seen with conventional strains of APV. In addition, the deletion mutant was found consistently to generate yields of $10^{5.3}$ TCID$_{50}$ per standard well of cells. This contrasts with the value of $10^{6.8}$ TCID$_{50}$ per well that was seen with the

---

**Fig. 3.** RT-PCR amplification of cDNA from APV. A region of the F gene of APV was amplified between nt 3620 and 4063 of the APV genome RNA for both a wild-type non-recombinant and the recombinant virus derived from a full-length cDNA clone. Virus had been passaged twice in tissue culture following the original transfection and RNA was prepared 4 days after infection. Lane 1, 443 bp fragment amplified from non-recombinant wild-type virus; lane 2, DNA fragment from lane 1 digested with SstII; lane 3, size markers; lane 4, 443 bp fragment amplified from the recombinant virus recovered from a full-length cDNA clone; lane 5, DNA fragment from lane 4 digested with SstII, showing the generation of two novel fragments.

**Fig. 4.** Immunofluorescence staining of cells infected with rescued wild-type virus. Cells were stained 3 days post-infection. (a) Cells were infected with virus recovered from transfections with a full-length genome and virus antigen was detected by immunofluorescence. (b) Cells were infected with FPT7 alone and anti-APV antibodies were used to show lack of cross-reactivity with the control cells.
wild-type virus and suggests that the deletion mutant may be impaired slightly in its growth in vitro.

To confirm the presence of the deletion in the recombinant \(\Delta\text{SH}/\text{G}-\text{APV}\) by RT-PCR, genomic RNA of the virus, isolated after three passages in tissue culture, was amplified between the M2 gene (antigenome position 5273) and the G gene (antigenome position 6605). Genomic RNA was prepared and treated with DNase I before amplification to ensure that no DNA from the original transfection had been carried forward during virus passage. This produced a product of 690 bp in size, in contrast to the 1330 bp product that would be expected from a virus with a full complement of genes (Fig. 6). The sequence of the shortened fragment was identical to that of the original plasmid in the junction region. No fragment was obtained in the absence of reverse transcriptase. This confirmed that the recombinant virus lacked the regions of the SH and G genes that had been deleted in the cDNA.

The function of the SH protein in APV and HRSV is unknown at present. Its removal without apparent loss of viability is a finding similar to that of Bukreyev et al. (1997) for HRSV and shows that functional similarity extends across the subfamily Pneumovirinae. The timing and character of the CPE produced by this virus was indistinguishable from that seen with the rescued wild-type virus. The \(\Delta\text{SH}-\text{RSV}\) recombinant was also attenuated when administered to chimpanzees (Whitehead et al., 1999) and it be will be interesting to explore whether \(\Delta\text{SH}/\text{G}-\text{APV}\) is also attenuated. The similarity between the members of the genera Pneumovirus and Metapneumovirus may extend to being able to remove other genes that have been shown to be non-essential in HRSV, such as the second ORF of the M2 gene (Schmidt et al., 2002; Jin et al., 2003). Similarly, additional genes may be inserted into the APV genome.

By analogy with HRSV, the G gene of APV is anticipated to encode the attachment protein that mediates the initial interaction between the virus and host cell (Levine et al., 1987). Despite its central role in the virus life cycle, the HRSV G protein has been shown to be dispensable for virus growth in cell culture. Karron et al. (1997) reported the characterization of an HRSV vaccine candidate that was shown to contain a deletion that removed both the SH and G genes. Reverse-genetics approaches subsequently confirmed that loss of the G gene in HRSV, either singly or together with removal of the SH gene, did not result in loss of infectivity in vitro (Techaarpornkul et al., 2001; Teng et al., 2001). This is consistent with the observations.
reported here for ΔSH/G-APV and further strengthens the similarity of the two systems. However, deletion of the HRSV G gene was shown to affect the efficiency of replication in certain cell types (Teng et al., 2001). It was suggested that this effect was primarily at the level of attachment and entry, although the precise reasons for the host cell-dependent differences are not yet clear. As with the SH- and G-deleted vaccine candidate, recombinant virus analysis showed that the G protein is important for infection in vivo. It will be of interest to study the replication characteristics of ΔSH/G-APV in various cell types and in vivo.

The new ability to manipulate the APV genome will lead to fundamental questions being addressed concerning gene function. Several recombinant HRSV viruses have been shown to be attenuated (Marriott & Easton, 1999). Commercially derived, live APV vaccines for administration to poultry have been available for over 10 years in Europe and these have led to major improvements in disease control (Cook, 2000). However, doubts remain about their performance under certain conditions, especially where vaccine administration may not be optimal and significant numbers remain unvaccinated (Jones, 1996; Cook, 2000). APV vaccines are known to revert to a virulent state under experimental conditions (Naylor & Jones, 1994) and there is evidence to suggest that this also occurs in the field, especially where vaccine administration is poor (Naylor et al., 2003). It is widely believed that the current vaccines may represent the maximum stability that is achievable by empirical cell passage (Naylor et al., 2003) and attempts to protect by using recombinant and DNA vaccines have to date conferred protection markedly inferior to that conferred by the current products (Yu et al., 1994; Kapczynski & Sellers, 2003). Killed vaccines have been shown to be poorly effective unless preceded by an initial live vaccination (Cook, 2000). Current work in our laboratory, involving the study of previously derived empirical vaccines, together with associated progenitor strains and virulent revertants (Naylor & Jones, 1994), is starting to identify genome regions where specific mutations may significantly alter pathogenesis. It is anticipated that the reverse-genetics system described here will open new avenues for the development of live vaccines to improve on the empirical types that are currently in use. It is also probable that such developments may have directly useful applications in studies of the recently discovered human metapneumovirus (van den Hoogen et al., 2001).

ACKNOWLEDGEMENTS

This work was supported in part by EC contract no. QLK2-CT-2002-01699 and Lohmann Animal Health.

REFERENCES


Mutational analysis of the avian pneumovirus conserved transcriptional gene start sequence identifying critical residues

Nicole L. Edworthy and Andrew J. Easton

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Seven of the eight genes in the avian pneumovirus (APV) genome contain a conserved 9 nt transcriptional start sequence with the virus large (L) polymerase gene differing from the consensus at three positions. The sequence requirements of the APV transcriptional gene start sequence were investigated by generating a series of mutations in which each of the nine conserved bases was mutated to each of the other three possible nucleotides in a minigenome containing two reporter genes. The effect of each mutation was assessed by measuring the relative levels of expression from the altered and unaltered gene start sequences. Mutations at positions 2, 7 and 9 significantly reduced transcription levels while alterations to position 5 had little effect. The L gene start sequence directed transcription at levels approximately 50% below that of the consensus gene start sequence. These data suggest that there are common features in pneumovirus transcriptional control sequences.

Avian pneumovirus (APV) causes acute respiratory infection in domestic poultry throughout most of the world (Jones, 1996) and it has been detected in several wild bird species, although the disease impact in these species is not yet clear (Bennett et al., 2002, 2004; Lwamba et al., 2002; Shin et al., 2000, 2002). APV is classified as a member of the family Paramyxoviridae, the subfamily Pneumovirinae and the genus Metapneumovirus (Cavanagh & Barrett, 1988; Collins & Gough, 1988; Ling & Pringle, 1988). It is distinguished from members of the genus Pneumovirus by the order of its genes – 3’-N-P-M-F-M2-SH-G-L-5’ (Ling et al., 1992; Randhawa et al., 1996, 1997; Yu et al., 1992a, b) and the absence of non-structural protein genes (Randhawa et al., 1997).

RNA synthesis in pneumoviruses requires a ribonucleoprotein complex comprising the nucleocapsid (N) protein bound to the genomic RNA together with the phosphoprotein (P) and large (L) polymerase protein. Efficient, progressive, transcription is enhanced by the M2-1 protein, a feature that is unique to the members of the subfamily Pneumovirinae (Collins et al., 1995; Naylor et al., 2004). Transcription of virus mRNA is carried out by the virion-associated polymerase complex in a progressive stop–start manner from the 3’ to the 5’ end of the genome (Dickens et al., 1984; Pringle & Easton, 1997), and is directed by gene start and gene end sequences that flank the transcription units. The gene order in non-segmented negative-strand RNA viruses is significant, as their transcription strategy results in more mRNA transcripts from genes at the 3’ end of the genome, with levels of mRNA progressively decreasing in a step-wise manner to the 5’ end. In Human respiratory syncytial virus (HRSV) and APV, the gene start sequence is conserved in all of the genes except for the L gene (Li et al., 1996; Ling et al., 1992, 1995; Randhawa et al., 1996, 1999, 1992a), although the HRSV gene start sequence is 10 nt in length (Collins et al., 1986; Kuo et al., 1996). The APV consensus is GGGCAAGU in mRNA sense (Li et al., 1996; Ling et al., 1992, 1995; Randhawa et al., 1996; Yu et al., 1991, 1992a) and the L gene start sequence has three differences (underlined), giving a sequence of AGGACAAU (Randhawa et al., 1996). A consensus gene end sequence at the 3’ end of the transcription unit is thought to be involved in termination and polyadenylation (Bukreyev et al., 1996; Jacobs et al., 2003; Ling et al., 1992; Randhawa et al., 1996). Termination and polyadenylation of the mRNA occur at the gene junction, followed either by progression of the polymerase across the non-transcribed intergenic region to the next gene or relocation of the polymerase complex to the genome 3’ terminus to begin the process again. The reinitiation of transcription of the next gene then occurs (Collins et al., 1986).

Here, we report the effect of mutation of the APV gene start sequence on gene expression, and identify the key residues in the gene start sequence that control transcription. Each of the nine bases of the conserved APV gene start sequence was mutated to the three other possible nucleotides. A dicistronic minigenome for APV, similar to the dicistronic HRSV minigenome (Kuo et al., 1996), was kindly provided by Dr J. Smith, University of Warwick, UK. The minigenome was constructed using the APV leader and
were determined. For each mutant, three to eight individual
unaltered consensus gene start sequence for the CAT gene
genes expressed from the dicistronic minigenome with an
reporter gene expression. The values for the two reporter
level of Luc gene expression was normalized to the CAT
mined as described previously (Marriott
protein production from the reporter genes were deter-
the gene start mutations on gene expression, the levels of
minigenome plasmid used in the study contained CAT
nome. The virus genome sense RNA transcribed from the dicis-
Fig. 1.
Diagram of negative-strand RNA dicistronic minige-
the P gene end sequence generated a ClaI site that was used
 sequence in the vector. In addition, a non-consensus APV L
gene start sequence and an ‘intermediate’ mutant with two
of the three bases unique to the non-consensus L gene start
sequence were also created using mutagenic primers. All
mutations were confirmed by sequencing. Details of the
primer sequences are available upon request from the
authors.

The dicistronic minigenome, shown diagrammatically in
Fig. 1, was transfected into BSR-T7 cells (Buchholz et al.,
1999) that had previously been infected with APV, as
described by Marriott et al. (2001). To quantify the effect of
the gene start mutations on gene expression, the levels of
protein production from the reporter genes were deter-
mined as described previously (Marriott et al., 2001) and the
level of Luc gene expression was normalized to the CAT
reporter gene expression. The values for the two reporter
genes expressed from the dicistronic minigenome with an
unaltered consensus gene start sequence for the CAT gene
were determined. For each mutant, three to eight individual
transfection experiments were averaged and a standard
development determined.

The levels of expression of the Luc protein, relative to the
level of the CAT protein, compared with the wild type for
each of the mutations were calculated and are presented in
Fig. 2. Although the levels of reporter gene expression from
the mutants covered a wide range, all but one mutation
(C5U) decreased activity to below 60 % of the control.
Consideration of all of the mutations showed that changes
from a G to an A, in which the purine nature of the base is
conserved, were better tolerated than alteration of a G
residue to either a U or C. Mutations were more detrimental
(any change resulting in a value below 25 % of control) when
present at positions 2, 7 and 9. At position 1 the change from
a G to an A was the best tolerated mutant with the other
alternatives giving very low (G to U) or no (G to C) gene
expression. At positions 2 and 3, the change of the conserved
G residue to any other nucleotide reduced expression
significantly, with little difference between the three mutants
at each position, although the introduction of a U residue at
position 2 reduced expression more than the A or C
substitutions. The nucleotide that tolerated mutation with
the least effect on transcription was at position 5, where the
introduction of a U residue gave little reduction in reporter
gene expression. Overall, levels of gene expression of the
mutants ranged from 0 % for G1C to almost wild-type levels
for C5U.

Of particular interest for APV was the set of mutations that,
individually and collectively, form the L gene start sequence.
It was found that these three mutations caused a similarly
decreased level of expression (40–50 %) when present
individually or when combined in the L gene start sequence,
and also that a change of two of the three bases (G1A/G8A)
that make up the L gene start sequence similarly reduced
gene expression (Fig. 2). It is interesting to note that there
was no additive effect of the multiple mutations. The
conserved APV gene start sequence appears to be very
important for efficient transcription and the differences in
the L gene start sequence from the consensus maintained in
the other genes may, coupled with the effect of the position
of the L gene in the APV genome, be an important factor in
transcriptional regulation of the L gene.

It should be noted that the A7U mutant introduces a
translational start codon, AUG, which may alter the protein
reading frame from the naturally occurring one. This will
affect the observed activity of the Luc enzyme produced
from these transcripts, and hence the level of expression for
this mutant cannot be considered to be truly representative
of transcriptional capacity. Results of CAT ELISA and Luc
enzyme assays have previously been shown to closely match
levels of mRNA transcription from minigenomes measured
by Northern blot analysis (Marriott et al., 1999). To clarify
the effect of the introduction of this start codon, and to
confirm that the mutations were affecting mRNA transcrip-
tion rather than having an indirect effect on another aspect
of gene expression, Northern blot analysis was carried out.
using mRNA isolated from HEp2 cells previously infected with recombinant vaccinia virus expressing T7 RNA polymerase in a plasmid-based rescue system as described by Marriott et al. (1999), with plasmid amounts optimized for APV (0.4 μg of the N plasmid and dicistronic minigenome, 0.2 μg of the P and L plasmids and 0.02 μg of the M2-1 plasmid) transfected into approximately 2 × 10⁶ cells using Lipofectin (Invitrogen). A range of mutants were chosen to represent the different levels of Luc activity seen in Fig. 2 and this included the A7U mutant (Fig. 3). A riboprobe containing sequences from both CAT and Luc genes was generated by transcription from an APV dicistronic minigenome and the blots were processed as described by Marriott et al. (1999). It can be seen that the levels of Luc mRNA relative to those of CAT mRNA are consistent with the protein levels determined for the mutants. The mRNA of the mutant found to have the least effect, C5U, had both the CAT and Luc band at a similar relative intensity to that of the wild-type minigenome. The mutations with the most detrimental effect in the protein assay were G1C where no Luc protein was detected and mutant G1U which expressed Luc protein at levels of approximately 5% of the wild-type control. From Fig. 3 it can be seen that no Luc mRNA can be detected for either of these mutants. In contrast, mutant G1A had an intermediate effect on Luc protein production and it can be seen that the Luc mRNA band is of moderate intensity compared with the wild-type Luc mRNA and their CAT mRNA levels are of similar intensity. It can be seen that mutant A7U did not direct the synthesis of detectable levels of Luc mRNA, indicating that this mutation had a significant effect on transcription.

The APV gene start sequence is similar to that of HRV, the representative virus for the subfamily Pneumovirinae. Mutational analysis of the gene start sequence has been done for HRV (Kuo et al., 1997) and features similar to the data obtained here were found. At position 1, the APV and the HRV sequences are intolerant of change, with no mutation achieving above 40% of wild-type sequence activity and a change from a G to an A was the best tolerated mutation. The favouring of the A at position 1 was unexpected because this nucleotide is not found naturally in any HRV gene start sequence, though it is seen in the APV L gene start sequence. At positions 2 and 3 of the HRV consensus sequence, any change significantly reduced gene expression, but a change from a G to an A had the least effect.
on gene expression. Position 5 for both HRSV and APV was more tolerant of mutation and mutation to a U gave almost 100% activity in both systems. A change at position 6 for HRSV significantly reduced gene expression, but the effect of mutation at this position was less dramatic for APV. For both APV and HRSV a mutation at position 5 to a C residue gave the highest level of expression. At position 7, for both viruses, any change decreased the expression levels to below 20% of controls. Position 9, for both viruses, was least tolerant of mutation of any kind, with any change decreasing expression levels to below 10% of the wild-type level. Overall, both viruses were relatively sensitive to change at positions 7 and 9, and position 5 was relatively insensitive to change.

The APV gene start sequence mutation data presented here differ from those of HRSV in that the APV gene start sequence showed a trend to more readily accept changes from a G to an A. However, the APV gene start sequence was also more sensitive to mutation overall than that of HRSV and none of the mutations resulted in enhanced activity of the gene start sequence. The APV gene start sequence also tolerated best the changes at the positions found to differ from the consensus in the L gene start sequence over any other possible nucleotide. This trend was also found at positions 4 and 10 of the HRSV gene start sequence (mRNA sense) in the gene start sequence for the L gene if the A at position 9 of the HRSV gene start sequence is considered to be an insertion rather than a substitution (Kuo et al., 1997). The data presented here, together with those of Kuo et al. (1997) for HRSV, indicate that there is a consistent pattern for the nucleotides important for initiating efficient transcription by pneumoviruses. In addition, the expression of the L gene is regulated not only by genome location but also by its unique gene start sequence, which is less efficient than the consensus present in the other genes.

Acknowledgements

We would like to thank Dr Joanne Smith for providing the APV dicistronic minigenome used in this work and Dr Tony Marriott for constructive comments. This work was supported by EC contract no. QLK2-CT-2002-01699.

References


Avian metapneumovirus SH gene end and G protein mutations influence the level of protection of live-vaccine candidates

Clive J. Naylor,1 Roger Ling,2 Nicole Edworthy,2 Carol E. Savage1 and Andrew J. Easton2

1Department of Veterinary Pathology, University of Liverpool, Leahurst, Neston, Cheshire, UK
2Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

A prototype avian metapneumovirus (AMPV) vaccine (P20) was previously shown to give variable outcomes in experimental trials. Following plaque purification, three of 12 viruses obtained from P20 failed to induce protection against virulent challenge, whilst the remainder retained their protective capacity. The genome sequences of two protective viruses were identical to the P20 consensus, whereas two non-protective viruses differed only in the SH gene transcription termination signal. Northern blotting showed that the alterations in the SH gene-end region of the non-protective viruses led to enhanced levels of dicistronic mRNA produced by transcriptional readthrough. A synthetic minigenome was used to demonstrate that the altered SH gene-end region reduced the level of protein expression from a downstream gene. The genomes of the remaining eight plaque-purified viruses were sequenced in the region where the P20 consensus sequence differed from the virulent progenitor. The seven protective clones were identical, whereas the non-protective virus retained the virulent progenitor sequence at two positions and contained extensive alterations in its attachment (G) protein sequence associated with a reduced or altered expression pattern of G protein on Western blots. The data indicate that the efficacy of a putative protective vaccine strain is affected by mutations altering the balance of G protein expression.

INTRODUCTION

Vaccines must demonstrate two key characteristics to be successful: they must be attenuated for disease and their administration must be able to induce an appropriate level of protection in the vaccinee. Protection from infection generally involves the induction of a suitable immune response from both the humoral and cellular arms. Much attention has focused on the molecular basis of attenuation in vaccine candidates, but there is little understanding of the factors that relate to the establishment of protective immunity other than the inclusion of appropriate epitopes in antigenic proteins. Avian metapneumovirus (AMPV) is responsible for major economic losses in domestic poultry throughout most of the world (Jones, 1996). The most severe clinical effects of infection are seen in turkeys, but there are significant economic losses in commercial chickens where it has also been implicated in swollen head syndrome. AMPV is a member of the family Paramyxoviridae, subfamily Pneumovirinae and genus Metapneumovirus (Cavanagh & Barrett, 1988; Collins & Gough, 1988; Ling & Pringle, 1988), and is distinguished from members of the genus Pneumovirus by the order of its genes (3’-N-P-M-F-M2-SH-G-L-5’) (Ling et al., 1996, 1997; Yu et al., 1992a, b) and the absence of non-structural protein genes (Randhawa et al., 1997). The only other member of the genus Metapneumovirus is the species Human metapneumovirus (van den Hoogen et al., 2002). Several AMPV vaccine candidates have been described, but most suffer from problems of either incomplete protection and/or reversion to pathogenicity (Catelli et al., 2006). The genetic basis for field strain attenuation and vaccine instability is not clear and an understanding of this is important for the rational design of vaccine candidates.

AMPV type A field strain #8544 was isolated in northern England in 1985 during a severe outbreak of turkey rhinotracheitis that rapidly spread across the whole country. When tested as a candidate vaccine, it proved to be highly protective against virulent challenge, but also caused severe post-vaccine disease in a number of cases (Williams et al., 1991a). Subsequently, strain #8544 was passaged extensively in Vero cells in our laboratory to yield an uncloned virus (named P20) that protected turkeys against virulent AMPV challenge (Williams et al., 1991b). However, on occasions, the putative P20 vaccine produced respiratory disease, typical of AMPV infection, during serial passage in turkeys. This was ascribed to a small subpopulation
of virulent virus in less than 1 in 10^5 infectious doses (Naylor & Jones, 1994). To remove the virulent subpopulation, stock P20 was plaque purified and 12 derivative viruses were prepared as candidate vaccines and, as anticipated, all were shown to be free of the virulent subpopulation.

Here, we describe the protection conferred by these 12 plaque-purified viruses and demonstrate that the loss of ability to confer protection from virulent challenge is associated with an alteration in the balance of protein expression, particularly the attachment (G) glycoprotein.

**METHODS**

**Viruses.** The vaccine candidate P20 was used to infect monolayers of Vero cells, and 12 individual plaques were picked and purified. These purified viruses, designated A–H and J–M, were grown in Vero cells and used to inoculate 1-day-old turkey pouls.

**Clinical assessment of plaque-purified viruses.** An established challenge model was used to assess the ability of the plaque-purified viruses to confer protection from pathogenic virus challenge (Naylor et al., 1997a; Williams et al., 1991a). One-day-old turkey pouls were divided into 13 groups, each comprising ten birds. Twelve groups received intraocular inoculations, each with a different virus (A–M), at a dose of 10^7.0 TCID50 per poult. The final group was not inoculated. Clinical disease was assessed on a daily basis. At 3 weeks of age, half of the uninoculated control group was combined in one room, together with all of the birds from the inoculated groups and all were challenged by intraocular inoculation with virulent field virus strain #8544 at a dose of 10^7.0 ID50 per poult. This ensured that all birds receiving a virulent challenge were in identical conditions and received the same challenge. For each bird, the clinical signs were assessed prior to its identification by wing band number and then scored on a daily basis until signs ceased, using a four-point scoring system, with a score of 0 representing no signs, 1 representing clear nasal exudates, 2 representing turbid nasal exudates and 3 representing swollen infraorbital sinus and/or frothy eye (Catelli et al., 2006). The disease severity for each group was expressed by summing the daily mean scores for the entire period of clinical disease to give a single cumulative score for each. It has been shown previously that turkeys with no protection typically achieve a score of between 8 and 10 on this scale, whereas fully protected birds score between 0 and 1. These scores have been shown to correlate closely with the level of virus replication in the trachea, which, together with the nasal turbinates, constitute the key tissues for replication of the virus (Naylor et al., 1997a; Williams et al., 1991b). At 3 weeks of age, prior to combining the groups, serum was collected from all birds prior to challenge to provide a baseline for AMPV serology by ELISA to confirm that virus had replicated and induced an antibody response.

**Sequencing of viruses.** The complete genomic sequences of field virus strain #8544, vaccine P20, and two protective (C and K) and two non-protective (F and H) plaque-purified viruses were determined. RNA was extracted using an RNeasy kit (Qiagen) and cDNA was prepared as overlapping leader, central and trailer regions, using Superscript II reverse transcriptase (Invitrogen). The cDNAs were amplified with PfuTurbo (Stratagene) for 12 cycles and then amplified further to provide three overlapping PCR products for sequencing. The PCR products were treated with exonuclease I and shrimp alkaline phosphatase (GE Healthcare) to remove primers and dNTPs, respectively, and sequenced. If any region of sequence showed a difference to the other viruses, the process was repeated and the sequence difference confirmed using a new RNA extraction from the relevant virus followed by RT-PCR spanning the base(s) in question.

The remaining eight cloned vaccine viruses were sequenced in all regions where any differences were found between strain #8544, vaccine P20 and purified viruses C, F and H. This included the putative mutation together with a flanking sequence of about 250 bases in both directions. Subsequently, the G gene of purified virus L was fully sequenced after detection of additional mutations at the start of the G gene.

**Western blotting of cloned viruses.** Protein expression of selected viruses was studied by Western blotting. Viruses F, K, G and L were grown in Vero cells and 10^5.5 TCID50 of each was purified by ultracentrifugation through a 25% sucrose cushion followed by protein separation on 10% SDS-PAGE gels. The proteins were electroblotted onto nitrocellulose and incubated with serum raised during natural infection of turkeys with strain #8544. As a control, a separate gel was silver stained to confirm that similar amounts of material were present in each sample (not shown). The bound antibodies were detected using horseradish peroxidase-conjugated secondary antibodies and diaminobenzidine as the substrate.

**Generation of recombinant viruses.** Recombinant viruses, based on the subgroup A AMPV strain CVL-14/1, were created by reverse genetics. Full-length cDNA copies of the virus genome were prepared with modifications to the sequence to insert restriction enzyme recognition sites before the N gene initiation codon, in the L gene untranslated region and before the gene-start signals of the P, M, SH, G and L genes. The parental clone, designated F2, was modified by introducing the sequence of the SH gene-end and SH–G intergenic region from protective C and unprotective F and H viruses. This was achieved by PCR amplification of the region from the intergenic M2–SH to the intergenic region of the SH–G genes including restriction sites enabling it to be inserted into F2 digested with the same enzymes. Virus was rescued from these plasmids in the T7 RNA polymerase-expressing cell line BSR-T7 (Buchholz et al., 1999) following transfection with plasmids containing the full-length anti-genome and the N, P, M2-1 and L genes in pCITE4 (Novagen). Virus was propagated in BS-C-1 cells and cells infected at the third passage were used for RNA extraction. RNA was extracted from BS-C-1 cells infected with viruses C and F or recombinant viruses using an RNeasy Midi kit (Qiagen). RNA (1 µg per lane) was analysed by Northern blotting using glyoxal gels along with serial dilutions of in vitro transcribed [generated using Megascript T7 or T3 kits (Ambion) and purified on 5% acrylamide/urea gels] corresponding to the SH or G genes to allow quantification of RNA levels. Blots were probed with antisense RNA probes generated with a Maxi-script kit (Ambion) directed against the SH or G genes and containing 5% of a probe generated against the Millenium markers (Ambion), which were included on each gel. Hybridized and washed blots were wrapped in cling film and exposed to an imaging plate for reading in an FLA5000 imager (Fuji). Analysis of 16-bit TIFF files was carried out with TOTALLAB 2003 (nonlinear dynamics) and Microsoft EXCEL to give the number of copies of each specific RNA (µg total RNA)^−1.

**Use of a dicistronic minigenome to measure the effects of different SH gene-end regions on a downstream reporter.** Overlap-extension PCR was used to alter the gene-end region of the first gene in a dicistronic minigenome, as described previously (Edworthy & Easton, 2005), to that of strain #8544 or of virus C, F or H sequence. The sequences in the minigenome constructs contained the desired SH gene end and the subsequent intergenic region sequence of the appropriate virus (Fig. 1). The minigenomes were rescued in a recombinant vaccinia T7 virus-based transfection system,
and chloramphenicol acetyl transferase (CAT) and luciferase (Luc) reporter expression were measured as described previously (Edworthy & Easton, 2005).

RESULTS

Clinical assessment of plaque-purified virus from P20 vaccine stock

All of the viruses were attenuated for disease and no clinical signs were seen after inoculation with any of the purified viruses. After challenge with virulent virus, the progression of clinical signs was followed and details are given in Table 1. Only seven of the purified viruses generated high levels of protection, two were partially protective and three (viruses F, H and L) had similar levels of disease to the unvaccinated and challenged control group. Sera from all vaccinated groups were weakly positive prior to challenge, showing that the viruses had replicated to levels sufficient to induce an immune response (Table 1). There was no significant difference in antibody titres among groups receiving different viruses.

Table 1. Cumulative clinical scores of poults inoculated at 1 day of age with plaque-purified viruses and challenged at 3 weeks of age

Animals were bled to prepare sera for ELISA 1 day before challenge.

<table>
<thead>
<tr>
<th>Plaque-purified virus</th>
<th>Disease post-challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Description of clinical signs</td>
</tr>
<tr>
<td>A</td>
<td>None</td>
</tr>
<tr>
<td>B</td>
<td>Mainly clear nasal exudate</td>
</tr>
<tr>
<td>C</td>
<td>None</td>
</tr>
<tr>
<td>D</td>
<td>Mainly clear nasal exudate</td>
</tr>
<tr>
<td>E</td>
<td>None</td>
</tr>
<tr>
<td>F</td>
<td>Severe signs</td>
</tr>
<tr>
<td>G</td>
<td>None</td>
</tr>
<tr>
<td>H</td>
<td>Severe signs</td>
</tr>
<tr>
<td>J</td>
<td>None</td>
</tr>
<tr>
<td>K</td>
<td>None</td>
</tr>
<tr>
<td>L</td>
<td>Severe signs</td>
</tr>
<tr>
<td>M</td>
<td>None</td>
</tr>
<tr>
<td>None</td>
<td>Severe signs</td>
</tr>
</tbody>
</table>

*Log₂ ELISA titres pre-challenge. Titres >6.0 indicate significant APMV antibody titres. All groups showed low but significant antibody titres. Differences between protected and unprotected groups were not significant.
Nucleotide sequences of protective and non-protective viruses

The complete sequences of the genomes of the pathogenic field strain #8544 and the protective virus P20 derived from it were determined. As shown in Table 2, comparison of the consensus sequences of the #8544 and P20 genomes identified a total of nine base substitutions that had occurred during the attenuation process. Only three of these differences resulted in changes in amino acid sequence: two in the F protein and one in the M2-1 protein. When the two complete genomic sequences of the fully protective viruses C and K were determined, they were found to be identical to the P20 consensus sequence, indicating this to be the dominant sequence found in the population of viruses in the P20 stock. In contrast, the genomes of two (F and H) of the three viruses conferring negligible protection differed from those of the protective viruses only in the SH intergenic region where there were single point substitutions in the SH transcription termination (gene-end) sequence. A summary of the sequences in the SH gene-end, SH–G intergenic and G gene-start sequences is shown in Fig. 1.

Sequencing of the remaining eight plaque-purified viruses in regions spanning the 11 mutations that arose during the conversion of strain #8544 to P20 found them all to be identical to the P20 consensus sequence, indicating this to be the dominant sequence found in the population of viruses in the P20 stock. In contrast, the genomes of two (F and H) of the three viruses conferring negligible protection differed from those of the protective viruses only in the SH intergenic region where there were single point substitutions in the SH transcription termination (gene-end) sequence. A summary of the sequences in the SH gene-end, SH–G intergenic and G gene-start sequences is shown in Fig. 1.

Western blots of protective and non-protective viruses

Proteins prepared from Vero cells infected with selected representative plaque-purified viruses (protective G and K; non-protective F and L) were analysed by Western blotting using anti-AMPV serum prepared in birds infected twice by the respiratory route (Fig. 3). For all viruses, a strong band was seen at 55 kDa, which has been identified previously as the F protein (Cavanagh & Barrett, 1988; Collins & Gough, 1988; Ling & Pringle, 1988). Protective viruses K and G both produced a band of approximately 85 kDa corresponding to the G protein (Cavanagh & Barrett, 1988; Collins & Gough, 1988; Ling & Pringle, 1988). As shown previously, the heavily glycosylated protein was seen as a faint, diffuse band. Larger bands were also seen, which have been described as G protein multimers (Cavanagh & Barrett, 1988; Collins & Gough, 1988; Ling & Pringle, 1988). For non-protective viruses, the intensity of the G protein bands was much lower relative to the G protein bands of protective viruses. This suggested that expression of the G protein was reduced in the viruses that did not confer protection from infection compared with those that protected against challenge.

Table 2. Sequence differences among strain #8544, P20 vaccine and plaque-purified viruses A–M

The nature, location and resulting amino acid alteration, if any, of each of the differences are shown. The amino acid alteration is given as that occurring between the pathogenic progenitor strain #8544 and the P20 vaccine derived from it.

<table>
<thead>
<tr>
<th>Nucleotide position*</th>
<th>Genome location</th>
<th>Strain #8544</th>
<th>Vaccine P20</th>
<th>Amino acid alteration on attenuation</th>
<th>All protective clones</th>
<th>Virus F</th>
<th>Virus H</th>
<th>Virus L†</th>
</tr>
</thead>
<tbody>
<tr>
<td>2941</td>
<td>F gene start</td>
<td>C</td>
<td>U</td>
<td>None</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>C</td>
</tr>
<tr>
<td>3553</td>
<td>F</td>
<td>A</td>
<td>G</td>
<td>V→A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>3825</td>
<td>F</td>
<td>C</td>
<td>U</td>
<td>E→K</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>5055</td>
<td>M2-1</td>
<td>U</td>
<td>C</td>
<td>K→R</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>5140</td>
<td>M2-1</td>
<td>A</td>
<td>G</td>
<td>None</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>5929</td>
<td>SH untranslated</td>
<td>U</td>
<td>C</td>
<td>None</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>5946</td>
<td>SH gene end</td>
<td>A</td>
<td>A</td>
<td>None</td>
<td>A</td>
<td>U</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>5949</td>
<td>SH gene end</td>
<td>U</td>
<td>U</td>
<td>None</td>
<td>U</td>
<td>U</td>
<td>A</td>
<td>U</td>
</tr>
<tr>
<td>6358</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>None</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>10022</td>
<td>L</td>
<td>A</td>
<td>C</td>
<td>None</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>11624</td>
<td>L</td>
<td>A</td>
<td>G</td>
<td>None</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
</tr>
</tbody>
</table>

*Sequences are shown in the 3′→5′ genomic sense.
†Sequence changes in the G gene are not shown.
Effect of SH gene-end mutations on levels of SH–G dicistronic RNAs and G gene expression

The effects of the SH gene-end mutations on transcription of the downstream G gene were determined using RNA extracted from cells infected with the plaque-purified viruses. This was studied further using recombinant viruses containing the same mutations in the SH–G intergenic region.

Total cellular RNA was prepared from virus-infected cells, separated by electrophoresis, blotted and probed. The results of samples probed with plus-strand-specific SH and G gene probes are shown in Fig. 4(a, b). The SH gene probe showed that the wild-type (F2, equivalent to #8544) and protective virus SH gene ends (Fig. 4b, tracks C and rC) allowed only negligible readthrough into the G gene, as shown in Table 3. In contrast, a significant level of SH–G gene readthrough mRNA was seen for all non-protective vaccine SH gene ends (Fig. 4a, b, tracks F, rF and rH). The ratios of G gene transcripts to SH gene transcripts (SH+ dicistronic SH–G), as calculated from G- and SH-probed blots, are shown in Table 3. In general, the relative amount of G gene transcribed was lower in the non-protective viruses, but only to a marked degree for rF.

To determine the effect of the SH gene-end mutations directly on the level of downstream gene protein expression, a reverse-genetics approach involving a dicistronic synthetic minigenome was used. A similar approach has been used previously to determine the effect of mutations in AMPV gene-start sequences (Edworthy & Easton, 2005). A minigenome was constructed containing two reporter genes, CAT and luc, with the CAT gene positioned to be transcribed first from the synthetic genome. The wild-type gene-end, intergenic and second gene-start sequences were those found in strain #8544, and constructs containing the SH gene-end sequences from plaque-purified viruses C, F and H were also generated. The plasmids containing the

Fig. 2. Comparison of the sequences of the G genes of the P20 and plaque-purified L viruses. The nucleotide sequence of the G gene of P20 virus is shown with the amino acid sequence of the predicted G protein below. The nucleotide sequence of the G gene of virus L is shown above the P20 virus sequence and differences in its G protein sequence are shown above.

Fig. 3. Western blot analysis of AMPV proteins. Proteins from Vero cells infected with vaccine stock P20 and plaque-purified viruses F, G, K and L were purified by ultracentrifugation and separated by SDS-PAGE, blotted and probed with polyclonal antisera collected from turkeys infected with APV under experimental conditions. A similarly purified but uninfected cell preparation was also used. Molecular masses are indicated.
minigenomes under the control of the bacteriophage T7 promoter, together with plasmids expressing the AMPV N, P, L and M2-1 genes, also under the control of a T7 promoter (Naylor et al., 2004), were transfected into cells infected with a recombinant vaccinia virus expressing the T7 RNA polymerase.

The levels of expression of CAT protein (measured by ELISA) and Luc (in arbitrary units) were expressed as a ratio (Luc/CAT) for each minigenome. In Table 3 and Fig. 4(c), each ratio is given as a percentage of the ratio of the wild-type #8544 virus (which gave the highest ratio). This approach provided a relative measure of expression of the second gene compared with the first gene for the various sequences. The level of expression of the luc gene in protective virus C was almost identical to that of the parental virus. However, for the sequences from the two non-protective viruses, the level of expression of the luc gene was significantly reduced compared with the protective viruses, with the sequence from virus F causing the most reduction (43% of the control). This demonstrated that the alteration in the SH gene-end sequence that altered the level of readthrough mRNA at the SH–G gene junction also reduced the level of expression of the downstream gene. This was in agreement with the reduction in the level of G protein seen in the Western blots of the virus proteins (Fig. 3).

**Fig. 4.** Analysis of the effect on transcription and gene expression of alterations in the SH gene-end sequence. (a) Northern blot of mRNA extracted from cells infected with the plaque-purified viruses C and F, or recombinant viruses with the SH gene end corresponding to #8544 (F2) or to clones C, F or H (rC, rF or rH, respectively). The blot was probed with a G gene-specific probe. The positions of molecular size markers are indicated on the right. For quantification purposes, various amounts of AMPV-specific RNA synthesized in vitro were included on the gel and are indicated as the log_{10} of the number of molecules. The positions of the SH, G and readthrough mRNAs are indicated. Also indicated is the position of the antigenome RNA, which was also detected by the negative-sense probe. (b) Northern blot similar to that described in (a) but probed with an SH gene-specific probe. (c) Graph showing the relative levels of expression of the luc gene compared with the CAT gene in a dicistronic AMPV minigenome assay. The minigenomes contained the SH gene-end and SH–G gene intergenic regions from the viruses indicated (see also Fig. 2). The levels of expression were determined as described previously (Edworthy & Easton, 2005).
Table 3. Effect of different SH gene-end sequences on readthrough and downstream gene expression

The quantification of mRNA was derived from Northern blots of RNA from virus-infected cells (Fig. 4a,b). The relative levels of expression of the luc gene compared with the CAT gene in a dicistronic minigenome construct are presented as the mean of data derived from Luc and CAT assays on plasmid rescue of dicistronic minigenomes, as shown in Fig. 4(c). Ratios are expressed as a percentage of that obtained with the wild-type strain #8544 SH gene-end sequence.

<table>
<thead>
<tr>
<th>Virus/SH–G sequence in minigenome</th>
<th>Fractional readthrough at SH gene end*</th>
<th>Ratio of G RNA to SH+SH–G RNA†</th>
<th>Luc/CAT ratio as a percentage of #8544‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>#8544</td>
<td>0.009</td>
<td>0.423</td>
<td>100</td>
</tr>
<tr>
<td>Virus C</td>
<td>0.012</td>
<td>0.347</td>
<td>99</td>
</tr>
<tr>
<td>rC§</td>
<td>0.007</td>
<td>0.466</td>
<td></td>
</tr>
<tr>
<td>Virus F</td>
<td>0.314</td>
<td>0.294</td>
<td>43</td>
</tr>
<tr>
<td>rF§</td>
<td>0.683</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td>rH§</td>
<td>0.231</td>
<td>0.380</td>
<td>79</td>
</tr>
</tbody>
</table>

†Comparison of the fraction of G RNA relative to transcripts initiated at the upstream gene (SH + SH–G dicistronic RNA) determined from SH- and G-probed Northern blots.
‡Comparison of Luc (measured as relative light units) to CAT (measured as pg CAT) ratios, expressed as a percentage of the highest value.
§Virus on which the recombinant virus was based.

DISCUSSION

It is generally accepted that a small number of alterations in genome sequence can lead to dramatic differences in pathogenicity. This has led to the proposal that it will be possible to develop rationally designed vaccine candidates lacking pathogenicity determinants. However, another important aspect of a vaccine is its ability to establish protection from challenge. The parameters that affect the ability to confer protection have been little studied, but it will be necessary to clarify what they are for successful vaccine design. In our current study, analysis of 12 plaque-purified viruses derived from the original attenuated and protective P20 stock showed that all were still attenuated, but three failed to confer significant protection when tested as candidate vaccines (Table 1). The remaining plaque-purified viruses gave either complete or near-complete protection from challenge and presumably represented the majority of the protective virus in the original vaccine stock.

The complete genomic sequences of the parental pathogenic strain #8544 and the P20 vaccine derived from it were determined to locate the mutations associated with the loss of pathogenicity and the inability to protect from infection. This indicated that there were only nine nucleotide changes between the two viruses, only three of which resulted in amino acid changes, two in the F protein and one in the M2-1 protein, whilst of the remainder, four were silent mutations and two occurred in a non-coding region.

The complete genome sequences of two protective viruses, C and K, and two non-protective viruses, F and H, were determined. The protective viruses were identical in sequence to the P20 consensus sequence. In contrast, the non-protective viruses, F and H, both contained a single mutation, not present in the protective viruses, in the transcription termination signal of the SH gene (Fig. 1). The mRNA profile of the parental and selected plaque-purified viruses showed that the level of transcriptional readthrough at the SH–G gene junction in the non-protective viruses was considerably greater than that seen for protective viruses. In addition, a similar alteration in the frequency of SH–G readthrough was seen in recombinant viruses engineered to contain the altered SH gene-end sequence (Fig. 4a; Table 3). When the levels of gene expression for a series of dicistronic minigenomes containing SH gene-end sequences from either the protective or non-protective viruses was assessed, the level of protein expression for protective viruses was seen in recombinant viruses engineered to contain the altered SH gene-end sequence (Fig. 4a; Table 3).

A Western blot of the P20 vaccine and the four plaque-purified viruses using polyclonal antiserum raised during infection of turkeys by a natural route under experimental conditions showed antibody responses primarily to the F protein, irrespective of whether the viruses were protective or not (Fig. 3). Assuming that there were similar responses in the inoculated birds, this indicates that, for AMPV, an antibody response to the F protein alone is not sufficient to protect birds from challenge. The detection of high levels of antibodies directed against the F protein after infection by a natural route is in contrast to the antibody profile seen for other similar viruses in mammals. However, the immune systems of mammals and birds are significantly different. It should be borne in mind that a previous study showed that virus stock P20 fully protected bursectomized turkeys where antibody was undetectable (Jones et al., 1992). Equally, turkeys with high levels of circulatory antibody can be unprotected (Naylor et al., 1997b). This implies that viruses that do not protect are likely to have lost T-cell rather than B-cell epitopes. It is possible that the non-protective viruses may not be expressing sufficient levels of key T-cell epitopes on the G protein gene.

Whilst Western blotting of the proteins from selected protective and non-protective viruses only showed a weak diffuse band in the region corresponding to the reported size for the G protein, this was noticeably reduced in blots of non-protective viruses (Fig. 3). Taken together, these
data strongly suggest that an alteration in the balance of level of expression of virus proteins, in particular a reduction in the relative level of the G protein, can significantly reduce the potential protective efficacy of a vaccine candidate.

Alterations of transcription termination signals have been shown to have an effect on transcriptional readthrough, but did not result in any significant difference in the ability of the altered viruses to replicate in vitro or in vivo (Tran et al., 2004). An A→G substitution (genomic sense) at position 5937 (strain #8544 numbering) in the SH gene end inadvertently introduced into a recombinant AMPV has been shown to produce increased readthrough at the SH–G gene junction and reduced expression of G mRNA (R. Ling and A. J. Easton, unpublished data).

The conservation of strain #8544 sequences in plaque-purified virus L at positions 2941 and 11624 might lead to the expectation that it would be more virulent and more protective than the other clones. However, the G protein gene in virus L also contained extensive mutations, which may have resulted in the reduced level of protection to challenge as well as the absence of virulence (Fig. 2). The predicted protein sequence of the G protein of strain L contained more potential O-linked glycosylation sites than was the case for the G protein of the other plaque-purified viruses. The additional potential glycosylation sites were at positions 109 (I→T), 130 (I→S), 170 (I→T) and 217 (M→T), whilst two were lost at positions 45 (T→I) and 160 (S→P; Fig. 2). The origin of the G gene mutations in virus L is unclear, but it may have been present as a subpopulation in the original strain #8544 isolate or have been generated some time later during passage in tissue culture. A possible explanation for the A→G transitions in the genomes of negative-strand RNA viruses involving cellular adenine deaminase enzymes has been described (Bass et al., 1989).

Reduction of G gene expression or sequence variations in it therefore appear to reduce protection afforded by potential vaccine viruses. The importance of the G protein in generating a fully protective response has been demonstrated in respiratory syncytial virus using a recombinant virus in which the G gene was deleted (Johnson et al., 2004). However, this contrasts with the data from human metapneumovirus (HMPV) where a virus lacking the G gene was shown to confer effective protection to African Green monkeys in challenge trials (Biacchesi et al., 2005). The reasons for the differences seen with HMPV are not clear, and it is possible that this reflects differences in the host species or in the challenge protocols in the different systems. In the current study, it was possible that the reduction in protection observed in protective viruses was due in part to a reduction in replication rates of non-protective viruses. Serology indicated that replication occurred, but this may not be a measure of the relative levels of in vivo replication when comparing protective and non-protective viruses.

The data presented here clearly show that the features required in an effective vaccine – the ability to replicate in the host without causing disease and the ability to confer protection from subsequent infection – are quite different. Also, the factors that determine whether a vaccine candidate can confer protection from disease can be very subtle and can be affected by as little as a single point mutation. The data also show that a critical factor for AMPV in determining the ability to confer protection is the balance in levels of G protein expression compared with other virus proteins, even when the other major antigen, the F protein, is expressed at effective levels. It will be of interest to explore further the potential effects of altering the balance of other virus proteins to determine whether this is a G protein-specific effect or whether it applies more generally to the overall balance of virus protein expression.

ACKNOWLEDGEMENTS

This work was supported by funding from the UK Biotechnology and Biological Sciences Research Council, grant no. BB/D012171/1.

REFERENCES


