THE UNIVERSITY
of EDINBURGH

THE UNIVERSITY of EDINBURGH

Investigating alternative AUG codon usage in avian influenza A virus segment 2

Elizabeth Billington

College of Medicine and Veterinary Medicine

The Royal (Dick) School of Veterinary Studies

The Roslin Institute

Dissertation submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2021

Declaration of authentication

Pirbright, November 2021

I hereby declare that this thesis entitled “Investigating alternative AUG codon usage in avian influenza A virus segment 2” was produced by myself, that the work and illustrations herein are my own except where otherwise explicitly acknowledged. This work has not been submitted, either in whole or in part, for any other degree or professional qualification of any other institution.

Elizabeth Billington

Acknowledgements

I would like to thank my supervisors, Professor Paul Digard and Dr Holly Shelton for their advice, patience, and support throughout my PhD. These four and a bit years would not have been possible without them.

I would also like to thank all the members of the labs I have worked in. A research project is a team effort, and they have all helped in multiple ways throughout my project. I would particularly like to thank Rute for being my pathfinder and taking me under her wing in the first few months of the project, and Khalid for his endless patience when I moved to Pirbright and couldn't find anything. As always, huge thanks to Dr Sam Lycett for her help with the bioinformatics that began as a small side project and grew into a thesis chapter.

I would like to thank the Bake Off Chat in Edinburgh and the PhD Girl Gang in Pirbright for the friendship, love, and support they've given me. Brunch dates, days out, and many rants about the science and supervisors all helped to get me through. An extra thank you to Charlotte, who had to live with me during the writing of this thesis, which I'm sure was a struggle. I would also like to thank my non-scientific friends, and my partner, for trying their best to understand and empathise with the struggles, and their unflinching confidence that I would work it out.

Another big thank you to my parents, Jon and Mary, who have always supported me and believed in my ability to do what I wanted to do. I do appreciate the attempts to remember anything beyond "bird flu" and the least useful of questions; "Is your experiment working now?"

Table of Contents

Lay Summary

PB1-F2 is a non-essential protein produced from segment 2 of influenza A viruses (IAVs) to block a host immune response to viral infection. In IAV infected birds, PB1-F2 can be produced as a full-length protein or potentially as various fragments from internal start codons within the PB1-F2 sequence. In this project, we determined whether PB1-F2 fragments were expressed from internal start codons and whether these C-terminal PB1-F2 fragments could function equivalently to the full length PB1-F2 protein. In addition, we looked to determine if abrogation of the internal start codons had any impact on virus replication in vitro.

Bioinformatic analysis of IAV segment 2 sequences showed conservation of the C-terminus of PB1-F2 even in viruses with an N-terminal stop codon. This indicated the internal AUG codons can be used independently to produce C-terminal fragments. Using tagged proteins, protein products were produced that matched to specific AUG codon initiation of expression. We show producing C-terminal fragments has limited impact on virus growth in cells compared to full length protein, however when measuring genome replication alone, N-terminal stop codons had significant differences compared to the wild type. The fragments are equally as effective as the full length protein in antagonising the interferon (IFN) response. This is important, as it indicates that attempting to remove PB1-F2 expression by introducing a stop codon in the N-terminus will not prevent functional fragments being expressed. If the virus only needs the C-terminus of PB1-F2 to be functional, changes may need to be made to how the presence of PB1-F2 in sequences is reported.

Thesis Abstract

Influenza A viruses (IAVs) have a segmented, negative sense RNA genome. PB1-F2 is an IAV accessory protein encoded by segment 2, in the +1 reading frame. Avian IAVs predominantly encode full length PB1-F2s, whereas human IAVs often have stop codons resulting in C-terminus truncations or ablation of PB1-F2 expression. One reported function of PB1-F2 is innate immune antagonism, which requires C-terminal motifs. Full length PB1-F2 is translated from AUG 4 of segment two, however there are often one or more in frame downstream AUGs (AUGs 7, 8, or 9). Although there have been previous reports of expression from AUGs 7-9, no one has matched a protein product to a specific AUG codon. There have also been reports of an increase in N-terminal truncations of PB1-F2 in recent years. The aim of this PhD was to assess AUG 7-9 usage and possible phenotypes of the products, and assess whether the increase in N-terminal truncations allowed continued expression of possibly functional C-terminal fragments.

Bioinformatic analysis of avian IAV segment 2 sequences indicates conservation of open reading frames encoding the PB1-F2 C-terminus, particularly in domestic birds, despite some subtypes acquiring N-terminal stop codons that persist through several years. Conservation of the C-terminus from AUGs 7 and 9 leads to the hypothesis that these serve as independent initiation codons for the C-terminus. C-terminal fragment expression from specific AUG codons has been shown in 293T cells using tagged proteins. Persistence of N-terminal truncations suggests that only producing C-terminal fragments has no detrimental effects on the virus. This could be a method of host-specific adaptation of the virus.

Segment 2 mutants were generated in an avian H5N1 IAV background, which differed in the presence or absence of the AUG start codons or stop codon positions in PB1-F2. Significant differences in viral polymerase activity, measured using mini-replicon assays in avian cells, were observed for some stop codons, but none of the AUG mutants. It was also found that any C-terminal expression of PB1-F2 is sufficient to antagonise the poly(I:C) induced IFN response in avian cells, raising questions on the true minimal requirements for PB1-F2 function. Annotation of segment 2 sequences may need to be adapted to account for continued expression of C-terminal fragments from sequences with an N-terminal truncation.

List of Abbreviations

Adenine Nucleotide Translocator 3 (ANT3)

Amino Acid (aa)

Avian Influenza Viruses (AIVs)

Chicken Embryonic Fibroblasts (DF1)

c-Jun Terminal Kinase (JNK)

ClusterMatcher (CM)

ClusterPicker (CP)

Complementary RNA (cRNA)

Cycloheximide (CHX)

Cytopathologic Effects (CPE)

Defective Ribosomal Product (DRiP)

DEFRA (Department of Environment, Food, and Rural Affairs)

Dimethyl Sulphoxide (DMSO)

Dulbecco's Modified Eagle Medium (DMEM)

Endoplasmic Reticulum (ER)

Ethylenediaminetetraacetic Acid (EDTA)

Eukaryotic Initiation Factors (eIFs)

Foetal Bovine Serum (FBS)

Global Influenza Surveillance and Response System (GISRS)

Highly Pathogenic Avian Influenza (HPAI)

Hours Post Transfection (hpt)

Human Embryonic Kidney 293T Cells (293T)

Influenza A Viruses (IAVs)

Interferon (IFN)

Interferon Stimulated Genes (ISGs)

Intranasal (IN)

Intratracheal (IT)

Intravenous (IV)

Japanese Quail Fibrosarcoma Cells (QT-35)

Leica Application Suite (LAS)

Low Pathogenicity Avian Influenza (LPAI)

Luria-Bertani broth (LB)

Madin-Darby Canine Kidney Cells (MDCK)

Major Histocompatibility Complex (MHC)

Metric Tonnes (MT)

Mitochondrial Targeting Sequence (MTS)

National Centre for Biotechnology Information, Influenza Virus Resource (NCBI IVR)

NOD-like Receptor (NLR)

Nuclear Dot Protein 52 (NDP52)

Nuclear Localisation Signal (NLS)

Nucleotide-binding Oligomerisation Domain-like Receptor X1 (NLRX1)

Parkin RBR E3 Ubiquitin Protein Ligase (PRKN)

Pattern Recognition Receptor (PRR)

Permeability Transition Pore Complex (PTPC)

Phosphatidylinositol-3 Kinase (PI3K)

Plaque Forming Unit (PFU)

Polyinosinic:polycytidylic Acid (poly I:C)

Polymerase Chain Reaction (PCR)

Porcine Alveolar Macrophages (PAM)

Protein Kinase C (PKC)

Protein Kinase R (PKR)

Reactive Oxygen Species (ROS)

Receptor Binding Site (RBS)

Ribonucleoprotein (RNP)

RIG-I-like Receptors (RLRs)

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Tetramethylethylenediamine (TEMED)

Toll-like Receptors (TLRs)

Tracheal Organ Cultures (TOCs)

Translational Initiation Sites (TIS)

Tris Glycine Methanol (TGM)

Tu Translation Elongation Factor, Mitochondrial (TUFM)

Ultraviolet (UV)

Upstream ORF (uORF)

Voltage-dependent Anion Channel 1 (VDAC1)

World Health Organisation (WHO)

World Organisation for Animal Health (OIE)

List of Figures

List of Tables

Chapter 1: Introduction

General Introduction

Influenza A virus (IAV) is an orthomyxovirus with a segmented, single stranded, negative sense RNA genome. The genome segments are able to reassort, which can facilitate host jumps. Although the reservoir host is aquatic birds, numerous spillover events have resulted in the establishment of circulating lineages in mammalian hosts including canines, equines, and humans. IAVs are frequently transmitted to domestic poultry through the faecal-oral route (Webster & Govorkova, 2014).

Host barriers to IAV include HA binding specificity, polymerase activity, and viral control of the host immune system. Avian influenza viruses (AIVs) are of major concern as both a zoonotic threat to human health and an economic and food security issue. In 2019, the EU produced 13,471 tonnes of poultry meat, of which 82% was broiler chicken, and consumed 11,834 tonnes of poultry meat (European Commission, 2022). There have been numerous incidences of AIVs infecting poultry workers (European Centre for Disease Prevention and Control, 2021; Ma et al., 2019), and the virus is considered to be a novel pandemic risk. Zoonotic IAV infections from avian or swine hosts have previously stably adapted to a human host, most notably with pandemic strains from 1918 and 2009. Host switching is usually polygenic, and previous pandemic strains have used a variety of methods to successfully jump the species barrier (Taubenberger & Kash, 2010). The World Health Organisation (WHO) reports a total of 861 confirmed cases of human infection with H5N1 avian influenza virus (AIV) and 455 deaths from 2003-2020 (World Health Organisation, 2020).

All eight segments of the genome must be packaged together to make an infectious virus, allowing reassortment if two viruses enter the same cell. Reassortment can be the source of dramatic shifts in virus phenotype. New viruses most frequently arise in Asia (Russell et al.,

2008), and can be spread by migrating birds or livestock transport. Chicken is a major protein source globally, and the global poultry industry was valued at $310 billion in 2020 (The Business Research Company, 2021). Freight transport was found to have a significant impact on AIV diffusion throughout China (Lu et al., 2017; Webster & Govorkova, 2014). Economic losses due to AIV outbreaks can be devastating to small farmers.

In 2014-15, a highly pathogenic avian influenza (HPAI) outbreak in the US resulted in >$1 billion dollars lost in reduced broiler chicken exports alone, and federal expenditures of $879 million to cover costs of depopulation, cleaning and disinfection, and indemnities for lost birds (Ramos et al., 2017). HPAI outbreaks have increased in frequency in recent decades. HPAI H5 viruses descended from Goose/Guangdong/1996 have become endemic in poultry populations in much of Asia (V. Martin et al., 2006). The burden of AIV on the poultry industry is heaviest in regions where the virus is endemic and therefore more difficult and expensive to control (Alexander, 2007).

IAV Genome Organisation

Avian influenza A virus (IAV) is a single stranded negative sense RNA virus with a segmented genome (Figure 1.1). The eight segments vary in size from 890-2,341 nucleotides (Noda et al., 2006), and total ~13.6kb. Each of the eight IAV segments encodes at least one structural gene, and almost all produce one or more non-structural accessory proteins (Table 1) (Firth & Brierley, 2012; Pinto et al., 2020).

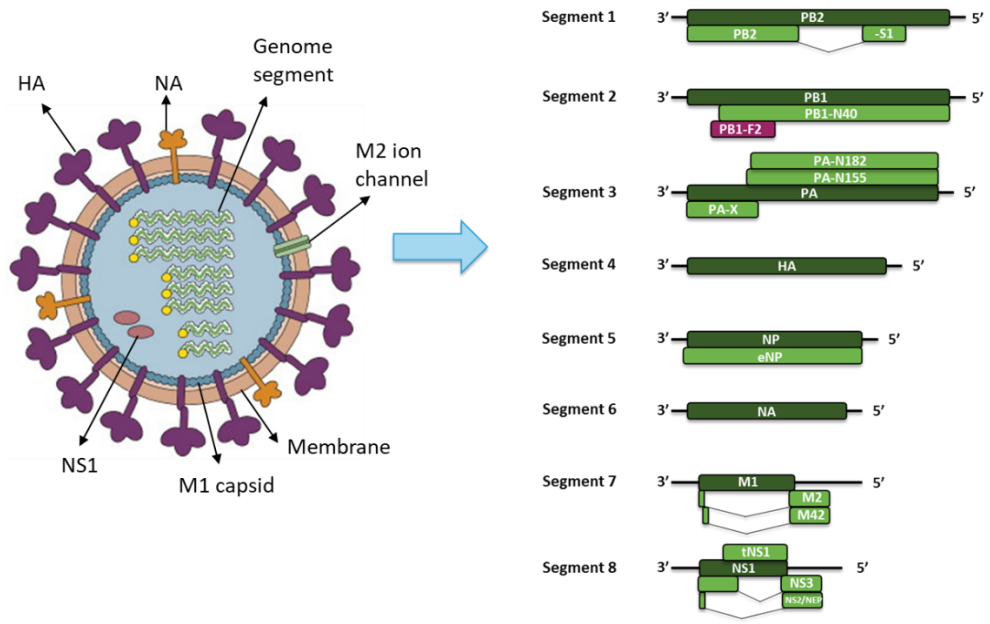

Figure 1.1 IAV Genome

Influenza A viruses have eight negative sense RNA segments, encased in a phospholipid membrane studded with viral glycoproteins HA and NA. Each of these segments produces as essential protein (dark green), and the majority also code accessory proteins (light green), such as PB1-F2 (purple). Adapted from (Pinto et al., 2020; Racaniello, 2014).

Segment	Segment length (bp)*	Protein products	Expression mechanism	Protein length (aa)*	Molecular mass (kDa)*	Main function	Reference
1	2341	PB2	Canonical translation initiation	759	85.7	Polymerase subunit	(Fields & Winter, 1982)
		PB2-S1	mRNA splicing	508	55	RIG-I inhibitor	(Yamayoshi et al., 2016)
2	2341	PB1	Canonical translation initiation	757	86.6	Polymerase subunit	(Winter & Fields, 1982)
		PB1-N40	Alternative AUG initiation (translation reinitiation)	718	82.4	Unknown, interacts with polymerase	(Wise et al., 2009)
		PB1-F2	Alternative AUG initiation (leaky scanning)	87	10.5	Immune antagonist/virulence factor	(Chen et al., 2001)
3	2233	PA	Canonical translation initiation	716	84.2	Polymerase subunit	(Fields & Winter, 1982)
		PA-X	Frameshift	252	29	Endonuclease activity, host shut-off	(Jagger et al., 2012)
		PA-N155	Alternative AUG initiation	562	62	Unknown	(Akkina, 1990; Muramoto et al., 2013)
		PA-N182	Alternative AUG initiation	535	60		
4	1778	HA	Canonical translation initiation	566	61.5	Receptor binding	(Winter et al., 1981)
5	1565	NP	Canonical translation initiation	498	56.1	Polymerase subunit, vRNP formation	(Winter & Fields, 1981)
		eNP	Translation initiation in UTR	504	56.8	(virulence factor, unpublished)	(Wise et al., 2019)
6	1413	NA	Canonical translation initiation	454	50.1	Virus release, sialidase	(Fields et al., 1981)
		NA43	Alternative AUG initiation (leaky scanning)	439	50	Virus release, sialidase	(Machkovech et al., 2019)
7	1027	M1	Canonical translation initiation	252	27.8	Matrix protein	(Lamb et al., 1981)
		M2	Canonical translation initiation + mRNA splicing	97	11	Ion channel	(Lamb et al., 1981; Wise et al., 2012)
		M42	Splicing + leaky scanning	99	13	M2 variant	(Wise et al., 2012)
8	890	NS1	Canonical translation initiation	230	26.8	IFN antagonist	(Inglis et al., 1979)
		NS2/NEP	Canonical translation initiation + mRNA splicing	121	14.2	vRNP export	(Inglis et al., 1979; O'Neill et al., 1998)
		NS3	Canonical translation initiation + mRNA splicing	187	21	Unknown	(Selman et al., 2012)
		NEG8	Unknown	216	23.8	Unknown, bioinformatically theorised	(Clifford et al., 2009)

Table 1.1 Viral Proteins

Table summarising protein products of each segment, including expression mechanism and function. Protein length and molecular mass given for lab strain A/Puerto Rico/8//1934 (H1N1).

Each segment forms a ribonucleoprotein (RNP) complex, by associating with NP along the length of vRNA and binding a panhandle structure formed of complementary 5' and 3' vRNA termini to the viral polymerase complex (Dadonaite et al., 2019; Urbaniak & Markowska-

Daniel, 2014). While bound to NP, only a fraction of vRNA is exposed along the RNP, limiting the availability of segment interactions (Arranz et al., 2012).

The segmented genome organisation allows for reassortment events to encourage genetic diversity, and overlapping ORFs within the segments provide a range of accessory proteins to finetune infection.

There is a significant correlation between genome length and the proportion of overlapping ORFs (Chirico et al., 2010; Pavesi et al., 2018). Products of overlapping genes are usually accessory proteins, as in IAV, although some short overlaps can be used for gene regulation (Chirico et al., 2010). Accessory proteins are often under evolutionary constraints due to the need to maintain essential protein sequences (Zell et al., 2007). Overlapping genes have been found to be enriched for high-degeneracy amino acids, which may help alleviate these constraints (Pavesi et al., 2018).

New ORFs produced by overlap can have useful functions, and introduce genomic novelty (Belshaw et al., 2007; Chirico et al., 2010). The evolution of overlapping genes is thought be assisted by the high mutation rate of RNA viruses, allowing introduction of start and stop codons in unused ORFs. The overlap would, however, increase deleterious effects of some mutations by affecting multiple genes with a single change (Belshaw et al., 2007).

IAV Virion Structure

IAV virions are pleiomorphic, however lab cultures are commonly spheres, 80-100nm in diameter (Mosley & Wyckoff, 1946). Some strains such as A/Udorn/72 (H3N2), and many clinical isolates, have a filamentous morphology of up to 30 μm in length (Dadonaite et al., 2016).

Virions consist of a membrane wrapped collection of vRNPs within an M1 capsid shell, decorated with both viral and host cellular proteins.

When viewed by electron microscopy, vRNPs are organised in a 7+1 configuration in both spherical and filamentous virions (Noda et al., 2006). M1 matrix protein forms an ordered helix layer adjacent to the membrane that can attract viral glycoproteins to the budding site, and affect morphology of the virion (Calder et al., 2010; Chlanda et al., 2015). The M1 layer is present for the entire length and tip of filamentous particles (Chlanda et al., 2015). Glycoproteins HA, NA, and the M2 ion channel are inserted into membranes at the endoplasmic reticulum (ER), and then trafficked to the plasma membrane at the budding site (Copeland et al., 1986; Dou et al., 2018). Host proteins can be collected via specific recruitment or taking advantage of cytoplasmic abundance at the budding site (Shaw et al., 2008). There is a 'conserved architecture' of virions that includes a consistent ratio of viral proteins, and a variable cellular protein component that corresponds to host species (Hutchinson et al., 2014).

The pleiomorphic morphology of IAVs is thought to be related to multiple passage in vitro, (Hirst & Hutchinson, 2019) although M1 sequence also impacts morphology (Digard lab, unpublished data).

IAV Replication

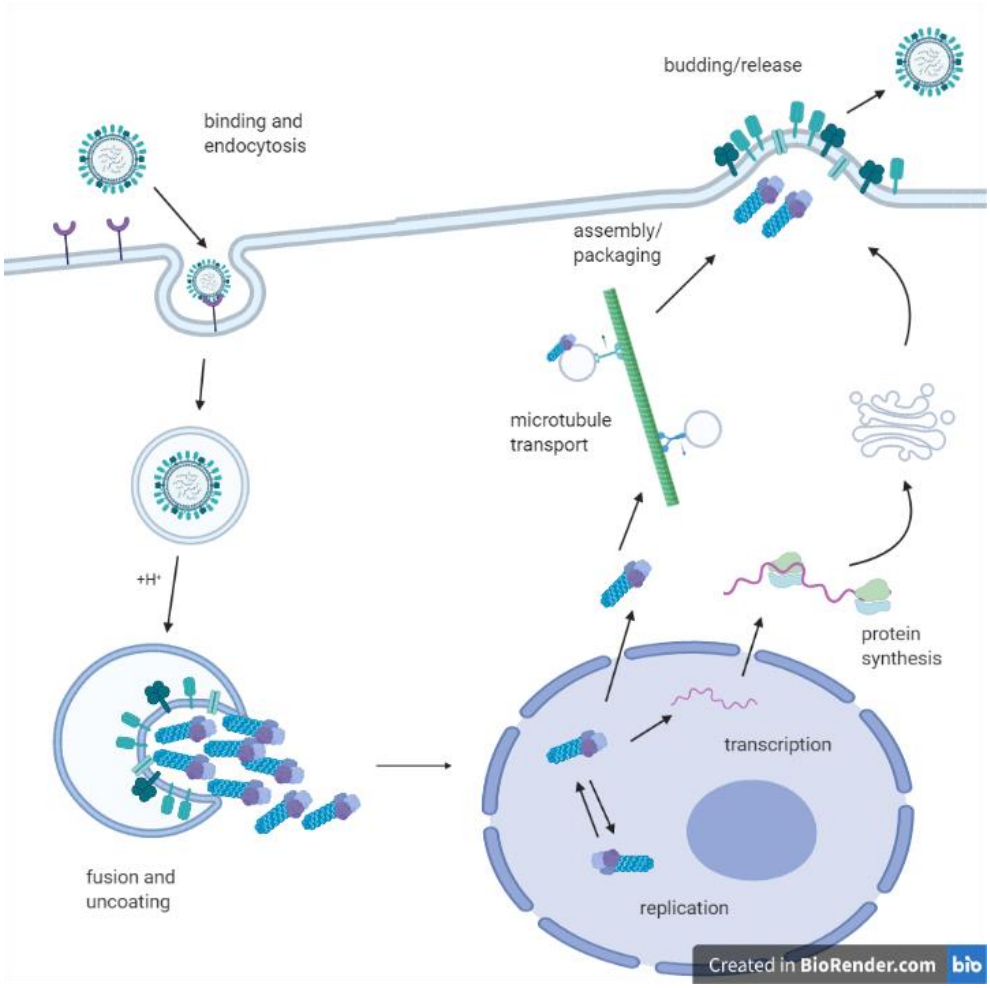

Figure 1.2 IAV Replication Cycle

HA binds to sialic acids and facilitates virions entering the cell through receptor-mediated endocytosis. Endosome acidification triggers membrane fusion and release of the vRNPs into the cytoplasm. vRNPs are transported into the nucleus for replication and transcription. Viral mRNAs are translated to produce new viral proteins. Progeny vRNPs are exported to the cytoplasm and trafficked to the plasma membrane. New virions are assembled and released by budding. Cellular and viral structures not to scale.

Virus Entry

Replication begins with the HA protein binding sialic acid receptors on the outside of the host cell. HA is a trimer linked by disulphide bonds, with each subunit having structurally distinct domains. A triple strand coiled-coil of helices inserts in the membrane, and a globular antiparallel β-sheet head domain contains the receptor binding site (Wilson et al., 1981).

Receptor binding is one of the primary host restrictions of IAV (Connor et al., 1994). The HA receptor binding site (RBS) consists of a series of conserved residues in a shallow depression in the HA head that are able to form hydrogen bonds to the sialic acid receptor (Skehel & Wiley, 2000). However, preferences can be overcome via point mutations of the HA RBS (H9 T180; H3 Y98F, H183F, L194A) (Sealy et al., 2020; Skehel & Wiley, 2000) or through a high enough inoculum (Szewczyk et al., 2014). Glycosylation of HA is also able to modulate antigenicity (Sealy et al., 2020). HA from avian viruses preferentially bind α 2,3 linked sialic acids found in abundance in bird respiratory and enteric tracts, and the human gut epithelium. HA from human viruses binds α 2,6 linked sialic acids found in the trachea and nasal epithelium. Deeper within the structure of the human lungs, α 2,3 linked sialic acids can also be found. Several ‘mixing pot’ species possess both linkages, while ducks only possess α 2,3 linked sialic acids (Nicholls et al., 2008). HA binding sialic acids typically has a relatively low affinity, which coupled with NA activity, allows for the virus to move through mucus to the epithelium until several molecules have bound to give a firm attachment of the virus to the host cell (Shaw & Stertz, 2018). NA is a tetramer with four-fold symmetry stabilised by bound metal ions. It functions as a sialidase to allow release of progeny virus from infected cells (Varghese et al., 1983).

Once sufficiently bound to sialic acid receptors, the virus is internalised via clathrin-mediated endocytosis (Lakadamyali et al., 2004). Vesicles

are scissioned by dynamin (Shaw & Stertz, 2018). However, there is also evidence that IAVs can exploit clathrin-independent endocytosis to enter the host cell with equal efficiency (Lakadamyali et al., 2004). Filamentous IAVs in particular enter by micropinocytosis, as they are usually too large to enter in clathrin-coated pits. Once trafficked to a low-pH compartment, alterations in M2 cause fragmentation of the filament leading to membrane fusion (Rossman et al., 2012).

vRNP Release

HA is cleaved by trypsin into HA1 and HA2. Introduction of a low pH environment after cleavage activates fusion (Maeda & Ohnishi, 1980). Fusion is a rapid process mediated only by HA2. Pre-treatment of fowl plaque virus with pH 5.2 buffer caused up to 85% inactivation of fusion activity, indicating fusion requires irreversible changes in HA2 (Whitel et al., 1982). The acid-induced conformational change of HA2 charges a pocket near the cleavage site and exposes a 22 residue N-terminal fusion peptide. The fusion peptide inserts itself into the endosomal membrane and draws viral and cellular membranes towards each other (Tsurudome et al., 1992). A C-terminal transmembrane domain anchors HA2 in the viral membrane. This causes a hairpin structure to bring the membranes in proximity. As the hairpins collapse, the membranes become close enough to fuse (Bullough et al., 1994; Dou et al., 2018). HA causes the outer leaflet of the membranes to curve the opposite way to the inner leaflet. This stress between the leaflets forms an unstable pore that develops into a stable channel and then to membrane fusion (Skehel & Wiley, 2000).

M1 protein forms a bridge between vRNPs and the viral membrane, likely mediated by a positive charge on the protein and a negatively charged vRNP. M1 is a key transport regulator for vRNPs, whose activity can be inhibited by amantadine (Martin & Helenius, 1991). The M2 ion channel is a homotetramer that is vulnerable to amantadine class antivirals. The channel is primarily proton selective and pH

selective. As the pH inside the vesicle decreases, proton channel activity is inhibited and M2 activates a cation (K^+) transport activity. This switch in activity allows accumulation and retention of protons in a low pH vesicle (Leiding et al., 2010). The N-terminus is exposed to the lumen of the endosome during acidification, encouraging proton permeability between pH 5-6.5 (Chizhmakov et al., 1996). When the M2 ion channel allows proton entry into the endosome M1 dissociates from the vRNPs. The pH dependence on dissociation was confirmed by transient acidification of the cytoplasm. A NH_4Cl-prepulse protocol showed dissociation but no reassociation as pH approached neutral, suggesting that low pH has an irreversible effect on M1. Dissociation allows nuclear import of vRNPs (Bui et al., 1996).

vRNA Replication and Transcription

Upon release into the cytoplasm, vRNPs separate from M1 and enter the nucleus. Transport through nuclear pores is mediated by NP (Martin & Helenius, 1991). NP contains multiple nuclear localisation signals (NLSs), and is able to mediate facilitated translocation through interacting with importins (Wu et al., 2007). NLS1 (residues 3-13) of NP can regulate nuclear-cytoplasmic shuttling of NP by binding either importin-α or CRM1. The choice of binding partner is partly determined by the phosphorylation status of S9, Y10, and Y296 residues of NP. Phosphorylation of S9 and Y10 reduces interactions between NP and importin-α, so they are dephosphorylated in early stages of infection to promote nuclear import (Zheng et al., 2015). PB2 is able to accumulate in the nucleus in the absence of other viral proteins (Jones et al., 1986). PB2 D701N enhances nuclear localisation. D701N disrupts a salt bridge in the C-terminus of PB2, which exposes an NLS. The NLS can interact with importin-α while PB2 is bound to a vRNP, and has been found to increase import efficiency, possibly through affecting vRNP orientation at the nuclear pore (Sediri et al., 2015).

Once the vRNPs are in the nucleus, genome replication and transcription to mRNA can begin. Within the vRNP, vRNA is able to associate with the polymerase through conserved sequences at each termini, while the rest of the segment is bound to NP (Arranz et al., 2012; Fodor, 2013). NP is essential for elongation of RNA chains and prevents transcription termination by binding both template and newly synthesised RNA (Honda et al., 1988). A positive stranded complementary RNA (cRNA) is produced first, and is then used as a template for vRNA synthesis (Taylor et al., 1977). cRNA becomes bound to NP and the polymerase to form a cRNP. A second polymerase is recruited to the cRNP to allow cRNA replication to new vRNA. The cRNA intermediate prevents the need for host primers in genome replication by acting as a template to produce the new vRNA (Te Velthuis & Fodor, 2016). *In vivo,* both cRNA and mRNA are synthesised during the transcription phase. However, the requirement for a host-derived cap and polyadenylated tail on mRNAs compared to the uncapped full length copy of cRNA has resulted in different initiation and termination mechanisms that are not fully understood (Te Velthuis & Fodor, 2016; Vreede & Brownlee, 2007). Initiation of transcription requires interaction of the viral polymerase with host Pol II to facilitate cap-snatching and the addition of a G residue to the 3' end of the template. Polyadenylation is achieved through a series of U residues near the 5' end and transcription is terminated by steric hindrance as the vRNA 5' end remains bound to the polymerase (Poon et al., 1999; Vreede & Brownlee, 2007).

Viral mRNA transcription is primed by cleavage of host capped mRNAs by the polymerase through its endonuclease activity. PB2 binds the cap and PA acts as the endonuclease, with a preference for host small nuclear RNAs, small nucleolar RNAs, and promoter-associated capped small RNAs (Gu et al., 2015; Sikora et al., 2014). Capped mRNAs are cleaved 10-14 nucleotides from the 5' cap, and act as primers to initiate transcription (Plotch et al., 1981). Once

cleaved, the 3'-terminus of the capped primer moves from the endonuclease active site to the polymerase active site, where it is brought into proximity with the 3' vRNA template (Te Velthuis & Fodor, 2016). Once viral mRNAs have been produced, host splicing machinery is recruited to segments 7 and 8 (Shaw & Stertz, 2018). A small fraction of additional mRNAs are processed to produce M2 (Lamb et al., 1981).

Replication involves interaction between two viral polymerases. Under the *trans*-acting model, replication is initiated by the *trans*-polymerase binding to GTP at the 3' end of the vRNA. Unlike during transcription, at the termination of replication the 5' end of the vRNA must be released from the vRNA-associated polymerase to allow replication by the *trans*-acting polymerase. The newly synthesised cRNA 3' end then binds to the trans-acting polymerase to form the cRNP complex (Fodor, 2013; Te Velthuis & Fodor, 2016).

Once replicated, progeny vRNPs are exported via the CRM1 pathway (Neumann et al., 2000). Interactions between CRM1, NEP, and M1 mediate vRNP export, in a 'daisy chain' model where M1 interacts with vRNP and NEP, which interacts with CRM1 to facilitate export (Akarsu et al., 2003; Neumann et al., 2000). It has been reported that vRNP and cRNP complexes are regulated differently, with vRNPs being selectively exported while cRNPs are restricted to the nucleus (Tchatalbachev et al., 2001). However, more recent work using highly sensitive strand-specific qRT-PCR suggests that both vRNPs and cRNPs can be exported from the nucleus, and the selection of vRNPs for packaging happens due to preferential M1 binding at the assembly site (Chaimayo et al., 2017). It has been suggested that the polymerase conformation varies between cRNP and vRNP complexes due to differing promoter structures (Tchatalbachev et al., 2001). The compactness of polymerase domains appears to change depending on the type of RNA bound, favouring transcriptase or replicase activity as required (Thierry et al., 2016). Chaimayo et al. also suggests

structural differences as a mechanism of preferential binding of M1-vRNP. cRNPs may expose nuclear export signals of NP that are not on the surface of vRNP complexes. Differences in the exposed surfaces of NP may also determine preferential M1-vRNP interactions at the plasma membrane (Chaimayo et al., 2017). Outside of the nucleus, vRNPs are transported to the budding site via the microtubule network (Shaw & Stertz, 2018). Meanwhile, HA, NA, and M proteins are all trafficked through the ER-Golgi secretory network toward the plasma membrane (Shaw & Stertz, 2018). The proteins are folded and glycosylated in the ER before being assembled into HA trimers and NA tetramers. The Golgi apparatus modifies the glycan chains and mediates esterification with fatty acids (Szewczyk et al., 2014).

Viral Budding

HA and NA cluster in areas of the membrane rich in cholesterol and sphingolipids known as lipid rafts (Shaw & Stertz, 2018). Cholesterol can increase curvature of plasma membranes to encourage fusion (Chen & Rand, 1997; Yesylevskyy et al., 2013). HA can cause coalescence of raft domains to form a viral 'budozone' (Rossman & Lamb, 2011). NP and vRNPs are recruited to M1 at the plasma membrane. M2 and cholesterol temporarily stabilise the budding membrane to allow assembly, and recruitment of a full set of vRNPs (Rossman & Lamb, 2011; Szewczyk et al., 2014). Parts of vRNA are exposed to allow RNA-RNA interactions between segments. A redundant, plastic network of segment interactions aids in genome packaging, while allowing room for reassortment or mutation (Dadonaite et al., 2019). IAV segments package into a 7+1 configuration, with the central segment likely to be a polymerase subunit (White & Lowen, 2018).

M1 is essential for virus assembly. It attaches to areas of the plasma membrane containing virus glycoproteins and actively promotes

budding (Lohmeyer et al., 1979). M1 crosslinks the cytoplasmic tails of HA and NA and forms a helical net under the membrane to provide structure. M1 polymerisation at the budding site may be an elongation mechanism for filamentous particles (Hay, 1974; Rossman & Lamb, 2011). M2 then becomes involved in membrane scission and virion release, by restructuring the plasma membrane to have the curvature for cleavage of the new virion from the host cell (Schmidt et al., 2013).

Release of New Virions

NA mediates release of the new virion, but there must be a functional balance with HA. HA makes up ~80% of the virion membrane protein complement; its sialic acid binding activity keeps virions on the cell surface until NA removes the sialic acid and releases the virion (Rossman & Lamb, 2011; Szewczyk et al., 2014). If NA activity is low, virions may aggregate on the cell surface instead of seeking new host cells. A lower HA affinity for sialic acid can evolve to compensate for weak NA activity (White & Lowen, 2018).

During replication, HA binds sialic acid receptors and triggers endocytosis. The acidic environment of the endosome triggers membrane fusion and release of the vRNPs. The vRNPs are then trafficked to the nucleus for replication. Progeny vRNPs and viral proteins are trafficked to the plasma membrane, where M2 and NA mediate budding and virion release. The replication cycle may be finetuned by accessory proteins and immune antagonists, such as PB1-F2, in response to immune pressure.

IAV Segment Translation Methods

Canonical Initiation

During canonical eukaryotic translation initiation, a pre-initiation complex is formed of the ribosomal 40S subunit, various eukaryotic initiation factors (eIFs) and a $tRNA^{Met}$. The complex binds to secondary structures within the 5’ UTR of the mRNA and begins scanning

downstream for a start codon. The $tRNA^{Met}$ binds to an AUG codon, triggering dissociation of the eIFs and binding of the 60S ribosomal subunit. Following peptide chain elongation and termination of translation, the ribosomal subunits are recycled (Jackson et al., 2010). In theory, binding of $tRNA^{Met}$ to the AUG codon relies on strong Kozak consensus (CC**A**CCAUG**G**) to prevent binding to non-AUG codons, or AUG codons with a weak Kozak consensus (Kozak, 1984). However, this is not always effective.

Exceptions to canonical translation initiation can be either programmed or incidental. Incidental exceptions can be regarded as translational 'noise', and are likely to be lost over time due to lack of purifying selection. Heterogeneity of translation and the introduction of noise is associated with increasing proteome diversity, but can also be involved in regulation of expression (Sonneveld et al., 2020; Thattai & Oudenaarden, 2001). Purifying selection reduces genetic diversity by selecting against deleterious mutations, in order to preserve biological function (Cvijović et al., 2018). In contrast, programmed exceptions are usually strongly conserved and involve a significant proportion of ribosomes (Firth et al., 2012).

Canonical initiation relies on recognition of an AUG codon with a Kozak consensus, but is still susceptible to varying pressures on different ORFs.

Non-AUG Initiation

Translation initiation at non-AUG codons has been reported in multiple contexts, including during viral infection and as a form of translational regulation of the human proteome (Dasso & Jackson, 1989; Fritsch et al., 2012; Prasad et al., 2016; Schwab et al., 2004).

Non-AUG codons can be efficiently utilised as translational initiation sites (TIS) with the proper Kozak consensus sequence. Some non-AUG codons can reach up to half the protein expression as the

consensus Kozak sequence CACCAUGG, with CACCCUGG being the most efficient (De Arce et al., 2018).

Within the Kozak consensus, the -3 and +4 positions are most important. The AUG-anticodon interaction is usually strong enough to trigger initiation regardless of the surrounding nucleotides (De Arce et al., 2018). However, non-AUG codons are more susceptible to their surrounding Kozak sequences than AUG codons due to mismatches between non-AUG codon and the anticodon. The interaction between the preinitiation complex and the surrounding nucleotides provides an 'energy boost' to allow initiation (De Arce et al., 2018).

During viral infection, a subset of stress-response mRNAs are upregulated for translation initiation by protein kinase R (PKR). This subset was found to be enriched for initiation at non-canonical start codons. PKR is a serine-threonine kinase that is activated by dsRNA. Upon activation, PKR phosphorylates the eIF2 α subunit to inactivate the transcription factor and reduce available pre-initiation complexes within the cell (Gale & Katze, 1998; Meurs et al., 1990). When there is a lack of eIF2 available, a series of stress response mRNAs, such as the ATF transcription factor family are expressed (García et al., 2006; Guerra et al., 2006).

Stress can alter the initiating start codon used, including expression of immune epitopes. Host CUG initiation has been identified during IAV infection (Machkovech et al., 2019).

A subset of ribosomes actively scans 5' to 3' for CUG initiation codons. CUG initiation is eIF2 independent. eIF2 is often targeted during translation inhibition, so an independent mechanism allows stressed cells to express antigenic peptides for immunosurveillance (Schwab et al., 2004).

Leaky Scanning

Leaky ribosomal scanning is an important mechanism of proteome expansion. Ribosomes recognise the Kozak consensus sequence around AUG codons to facilitate translation initiation (Kozak, 1984). The most efficient translation initiation motif is CACCAUGG, where the -3/+4 positions have the greatest impact. AUG codons upstream of the initiation site have a preference for a pyrimidine at -3, whereas initiating AUGs usually have a purine in this position (Kozak, 1981, 1984). In IAV segment 2, leaky scanning occurs when the preinitiation complex skips over the weaker Kozak sequencing of AUGs 1-3 in favour of binding AUG 4 with a stronger consensus (Wise et al., 2011).

Frameshift

IAVs utilise frameshifting to express PA-X, an endonuclease involved in host gene expression shut-off. There are several methods of frameshifting, resulting from either a pause in translation, or slippage of the mRNA within the ribosome. Some RNA structures such as stem-loops or pseudoknots encourage frameshifting. It is thought that the time taken for the ribosome to unwind the secondary structure allows frameshifting to occur (Firth et al., 2012). In the case of PA-X, a slow to translate CGU codon in the A site within a UCC_UUU_CGU sequence facilitates tRNA realignment. The rare CGU codon causes a pause that allows some ribosomes to shift frame by re-pairing tRNAs at the P site from UUU to UUC. Both these codons are translated by an AAG anticodon, which has higher affinity for UUC, allowing the frameshift (Jagger et al., 2012).

Reinitiation

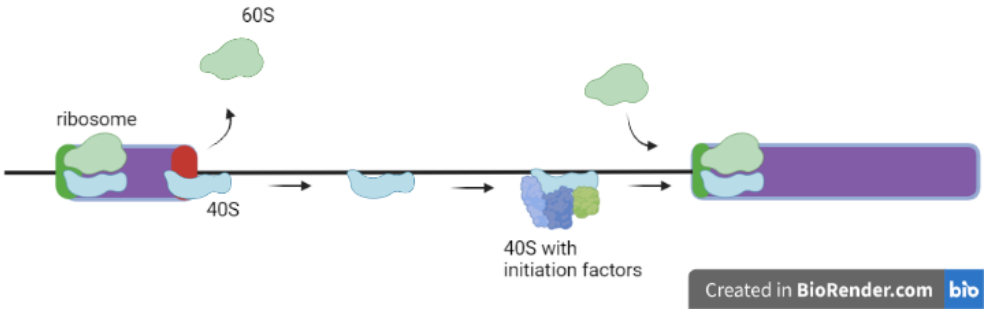

Figure 1.3 Translation Reinitiation

Black line indicates mRNA, with purple boxes for ORFs. Green and red boxes for AUG and STOP codons, respectively. The presence of a STOP codon induces dissociation of the ribosome. The 40S subunit can continue scanning, and reacquire subunits of the preinitiation complex to resume translation at a downstream start codon.

Translation reinitiation following termination of an upstream ORF (uORF) can be used to express a downstream ORF that would normally not attract ribosomes to be efficiently expressed. The ribosome halts translation at the stop codon of the uORF, but the 40S subunit continues scanning rather than being recycled (Figure 1.3). The preinitiation complex can reform, and translation begin again at the downstream AUG codon (Figure 1.3) (Kozak, 2001). Reinitiation is most efficient with a short uORF of less than 13 codons, due to the need to retain some eukaryotic initiation factors (eIFs) on the 40S ribosome (Hronová et al., 2017; Kozak, 2001). In viral systems, the ORFs typically overlap in different reading frames (Powell, 2010).

In IAV segment 2, reinitiation is used to balance the expression of PB1-F2 and PB1-N40. Ribosomes ignore the weak Kozak consensus around AUGs 1 and 2 by leaky scanning to initiate at AUG 3 in frame 2. The AUG 3 ORF is only two codons long, so initiation is immediately followed by termination of translation and continued scanning of the 40S subunit. Due to proximity, the 40S subunit scans past AUG 4 before re-acquiring the necessary eIFs to form the preinitiation

complex, leading to expression of PB1-N40 from AUG 5 in frame 1 (Wise et al., 2011).

Alternative mRNA Splicing

Segments 7 and 8 of IAV undergo alternative splicing to produce accessory proteins (Lamb et al., 1980, 1981). IAV remodels the cellular spliceosome to encourage viral mRNA splicing and suppress host gene expression. The bulk of the polymerase can block the assembly of the spliceosome components at a splice site in viral mRNA (S. R. Shih et al., 1995). NS1 can inhibit complete spliceosome recruitment and interact with several cellular proteins involved in splicing to achieve viral control of the process (Dubois et al., 2014). Segment 7 $mRNA_3$ has been proposed to negatively regulate gene expression. Proximity of the $mRNA_3$ 5' splice site vs the $mRNA_2$ 5' splice site to the viral polymerase binding site is used to control alternative splicing of M1 mRNA (Shih et al., 1995). NS1 also appears to have a functional role in M mRNA splicing, through either direct interaction with the mRNA or through interactions with host factors (Mor et al., 2016).

Splicing of segment 7 has been shown to affect host range of IAVs (Bogdanow et al., 2019). Through multiple studies on various host proteins, it is clear that altering the splicing machinery will negatively affect viral replication (Artarini et al., 2019; Ren et al., 2019).

Start-Snatching

It has previously been reported that an upstream AUG (uAUG) codon in the 5' UTR of segment 5 can be used to extend the ORF and produce an extended eNP protein (Wise et al., 2019). There have also been reports that expression can initiate in the 5' UTR of eukaryotic mRNAs (Haimov et al., 2017). Cap-snatching from host mRNAs is a key factor in viral protein expression (Plotch et al., 1981; Te Velthuis

& Fodor, 2016). Cap snatching is accepted to produce a fusion of host and viral mRNA by using the host cap as a primer (Plotch et al., 1981).

Although IAVs prefer non-coding host RNAs as their cap-snatching substrate (Gu et al., 2015), if the host 5' UTR contains an AUG codon upstream of the IAV mRNA, a hybrid host-viral protein can be expressed. This production of hybrid proteins has been termed 'start-snatching' (Ho et al., 2020).

The hybrid proteins can be produced either as a host-derived 5' extension, or as an out of frame novel protein (Upstream Frankenstein ORF (UFO)). Among IAV segments, 10-15% of cap-snatched sequences were found to contain an uAUG, although many segments have STOP codons before the canonical viral AUG codon, limiting expression of hybrid proteins (Ho et al., 2020). In segment 2, start-snatching results in a frame 3 product of 77aa named PB1-UFO. PB1-UFO was found to be conserved in human H1N1, H3N2, and H5N1 sequences, and had an impact on viral virulence in mice at high infectious doses (Ho et al., 2020).

IAV utilises several non-canonical expression mechanisms to expand its proteome (Table 1.1). Several other RNA viruses also utilise non-canonical mechanisms. For example, *Caliciviridae* use leaky scanning and reinitiation mechanisms while members of the *Respirovirus* genus can initiate translation at ACG or GUG codons (Firth & Brierley, 2012). This thesis focuses on PB1-F2, which is expressed by leaky scanning (Chen et al., 2001).

Accessory proteins of IAV

IAV has a small genome of only 13kb, and has evolved multiple methods to expand the proteome. Each of the eight IAV segments codes at least one structural gene, and almost all produce one or more non-structural accessory proteins (Firth & Brierley, 2012). Several accessory proteins are expressed by leaky ribosomal scanning,

resulting in N-terminally truncated versions of the essential product. This method may be associated with re-initiation events, as in segment 2 (Wise et al., 2009). Ribosomal frameshifting can also produce accessory proteins. Other segments utilise the host splicing machinery to produce novel ORFs. Segment 8 may have a negative sense ORF that is indicative of an ambisense strategy, but expression of this predicted protein has not yet been confirmed (Sabath et al., 2011). There are also possible microRNA coding regions in the virus genome that could have a regulatory role in virus replication (Vasin et al., 2014). Accessory proteins can have multiple roles designed to adapt the virus to the host and maintain optimal virulence and efficient replication. Many accessory proteins finetune the virus control over the host immune response, while others allow a functional redundancy in case an essential protein ORF is damaged due to lack of proof-reading. For example, NA43 can maintain low viral titres if the canonical NA AUG codon is removed, although it does not appear to be important for viral replication in the presence of full-length NA *in vitro* (Machkovech et al., 2019).

Segment 1

PB2-S1 was discovered during an RT-PCR screen of viral mRNAs to look for novel proteins encoded by spliced mRNAs. It is produced as a splice variant that deletes nucleotides 1513-1894 of the PB2 reading frame, with a premature stop at position 1936. PB2-S1 was commonly expressed in human H1N1 isolates pre-2009, but was not present in 2009 pandemic strain (Yamayoshi et al., 2016). PB2-S1 preserves the PB2 mitochondrial targeting sequence, and functions as an IFN antagonist via RIG-I and MAVS (Carr et al., 2006; Yamayoshi et al., 2016). PB2-S1 is able to decrease viral polymerase activity in minireplicon assays through interacting with PB1, but appears to have no effect on viral fitness *in vitro* or in mouse models (Yamayoshi et al., 2016).

Segment 2

PB1-N40 is an N-terminally truncated accessory protein of PB1 produced from AUG 5 by leaky scanning, with an unknown function. PB1-F2 is an accessory protein expressed by leaky scanning from AUG 4 in a +1 reading frame, that was discovered during a search for antigenic peptides recognised by CD8+ T cells (Chen et al., 2001). PB1-UFO is expressed from frame 3 as a host-viral hybrid protein by start-snatching (Ho et al., 2020). Segment 2 translation will be discussed in detail in Chapter 2.

Segment 3

Accessory proteins of segment 3 were first discovered during a screen of antigenic cross-reactivity between polymerase subunits. Additional smaller proteins were visualised during immunoprecipitation of PA proteins with monoclonal antisera (Akkina, 1990). The precipitated proteins were later identified as N-terminally truncated forms of PA, designated PA-N155 and PA-N182, through mutational analysis. Loss of PA-N155 resulted in delayed replication in MDCKs, and reduced pathogenicity in mouse models. PA-N155 has a supportive, but unclear role in viral replication, while PA-N182 did not cause any significant change in viral fitness (Muramoto et al., 2013).

PA-X is the product of an overlapping reading frame that produces a fusion domain of the N-terminal endonuclease domain and C-terminal X domain via frameshift. It was discovered following an assessment of segment 3 conservation that did not match known RNA structures. PA-X has host shut-off function similar to PA, and is able to modulate host gene expression during infection. Differential gene expression appears to be focused on pathways affecting lymphocyte activation and inflammatory responses (Jagger et al., 2012).

Segment 5

An upstream AUG codon in the 5’ untranslated region of swine-origin segment five has recently been identified as producing an extended version of NP, named eNP, in a preprint (Wise et al., 2019). The addition of six amino acids at the N-terminal of NP has an unclear function, but expression of eNP has been seen to affect pathogenicity of IAV in both mice and pigs (Wise et al., 2019).

Segment 6

A recent attempt to profile translation initiation in IAV infected cells identified an N-terminally truncated form of NA, dubbed NA43 after its AUG codon. NA43 expression is not dependent on the presence or absence of the canonical NA start site expressing full length NA. It is enzymatically active and expressed on the cell surface. Some low levels of viral growth are supported by NA43 in the absence of full length NA (Machkovech et al., 2019).

Segment 7

Segment 7 produces multiple differently spliced mRNAs. The fourth mRNA of segment 7 encodes M42, a functional variant of M2 with an alternate ectodomain. The spliced mRNAs share a 3’ splice acceptor site, but are differentiated by 5’ splice donor sites (Wise et al., 2012). M2 is an integral membrane protein that forms a homotetramer and has protein channel activity. The altered ectodomain of M42 causes a change in localisation from perinuclear to plasma membrane localised. However, this altered localisation does not appear to prevent viral replication (Wise et al., 2012).

Segment 8

A mutation in NS1 during serial passage in mice resulted in a secondary splice product, named NS3. D125G created a splice site for NS3 to be expressed. The additional product caused an increase in viral replication in the mouse host. NS3 is thought to be expressed in

a minority of strains and is associated with adaptation to a new host (Selman et al., 2012).

The NS1 mRNA can also use downstream AUG codons to initiate translation. The second and third AUGs of the mRNA, at codon positions 79 and 81, are well conserved in circulating human strains. In PR8, truncated NS1 products differ from the full length protein in cytoplasmic localisation and IRF3 inhibition. In Udorn however, there is no role for truncated NS1 proteins in the RIG-I pathway, although they may have other functions (Kuo et al., 2016).

Host Differences in AIV Infection

The original reservoir host of IAVs is aquatic waterfowl (Halvorson et al., 1985; Webster et al., 1992). The long relationship between IAV and its original reservoir is thought to be the reason for waterfowls high tolerance, and the relatively low rate of viral evolution within waterfowl compared to other hosts (Webster et al., 1992; Wille & Holmes, 2020; Yoon et al., 2014).

Through multiple spillover events, poultry and mammalian lineages have developed (Vandegrift et al., 2010; Wasik et al., 2019). Over time, interspecies transmission events have developed host-specific gene pools. Further evolutionary divergence has resulted in independently evolving host-specific strains (Webster et al., 1992). Development of a stable virus lineage from a zoonotic infection requires both virus adaptation and a suitable host ecology (Wille & Holmes, 2020). The genetic drift necessary to develop distinct genetics of each segment depends on the host species barriers and geographic isolation; for example, American and Eurasian avian lineages are defined by isolation, not host or subtype (Wille & Holmes, 2020; Zell et al., 2012). Phylogenetic analyses suggests long term co-evolution, particularly in avian hosts, with indications that Anseriformes and Charadriiformes have been infected with influenza viruses for millennia (Wille & Holmes, 2020). There are likely to have been several introductions of

the virus from waterfowl to poultry before a stable lineage developed (Carnaccini & Perez, 2020).

Variation in Avian hosts

Influenza evolution varies between wild birds and poultry. Waterfowl usually circulate low pathogenicity avian influenza (LPAI), due to the equilibrium of being the original reservoir host. However, HPAI can be transmitted from poultry and cause severe disease with high mortality (Chen et al., 2005). The faecal-oral transmission between wild birds involves a 'latent' phase within water sources (Fourment & Holmes, 2015). Contaminated water in wetland habitats has previously been postulated as a method of long-term persistence of IAVs. Infectivity is dependent on several factors, including temperature, salinity and conductance, but multiple avian IAV subtypes are persistent over several months (Brown et al., 2007; Ramey et al., 2020; Stallknecht et al., 1990). During this period, the virus is not replicating, which slows the nucleotide substitution rate compared to poultry (Fourment & Holmes, 2015).

AIVs can infect >105 bird species (Nuñez & Ross, 2019). In addition to binding preferences between mammals and birds, viruses circulating in different avian species differ in their fine receptor-binding specificity and must adapt when switching, for example, from ducks to chickens (Petersen et al., 2012). It is thought that most AIV transmission from wild to domestic birds is through indirect contact, i.e., contaminated water. Local wild birds can be considered a 'bridge' between migratory birds and domestic poultry (Elbers & Gonzales, 2020; Lu et al., 2017).

Wild birds are more resistant than poultry and can maintain infection without severe disease, allowing reassortment between strains (Nuñez & Ross, 2019). However, there is still variation in susceptibility within wild migratory birds. Multiple studies of migratory birds that stop over at Delaware Bay, USA found that Ruddy Turnstones are highly

susceptible to IAV (Hanson et al., 2008; Kawaoka et al., 1988). The high rate of infection may be due to the preference of Ruddy Turnstones for wetlands where AIV may survive in shallow pools of water, over sandbars that lack stagnant water and are washed by the tide twice a day (Hanson et al., 2008). Pigeons are considered resistant to flu, and in primary cell culture have been seen to have less α 2, 3 linked sialic acid than turkey cells. In tracheal organ cultures (TOCs), turkey cells were shown to be susceptible to a wide range of species-specialised viruses. Serial passage showed adaptive AIV mutations, particularly in the receptor binding pocket and cleavage site of HA (Petersen et al., 2012).

Domestic ducks have had a major role in the generation and maintenance of H5N1 HPAI in southern China. Since 2002 there has been an increase in wild waterfowl suffering clinical signs and severe disease form HPAI infection (Nguyen et al., 2005). Domestic ducks and geese are closely related to their wild counterparts, but differences in their environments and level of contact with other poultry have changed their role in transmission and maintenance of IAV (Cardona et al., 2009). HPAI develops from LPAI during circulation in poultry. A study of the evolutionary dynamics of a H7N1 outbreak in northern Italy March 1999-February 2001 showed the original LPAI virus evolved into LPAI-1 and HPAI lineages. All the HPAI isolates sampled had a common ancestor that dated from the period of LPAI circulation. Minority variants of LPAI H7N1 had some key HPAI mutations, further suggesting the LPAI virus as the precursor of the HPAI lineage (Monne et al., 2014).

Ducks and chickens are known to have differing tolerances to AIV infection. Ducks have a minimal response to LPAI, whereas chickens infected with LPAI may suffer high morbidity/mortality. During a HPAI H5N1 outbreak in Korea affecting two broiler breeding farms, two layer chicken farms, two duck breeding farms, and one quail breeding farm in 2006/7, the chickens and quails suffered high mortality, while the

ducks had no mortality and only a drop in egg production. When an isolate from this outbreak was used for experimental infection, both chickens and quails died earlier and shed to higher titres than ducks (Jeong et al., 2009). When infected with A/chicken/Shanghai/441/2009 (H9N2) ducks responded faster, but with a lower level of inflammatory cytokines and a persistent cellular response, whereas chickens had an excessive but delayed inflammatory cytokine response and inadequate cellular response. Intranasal infection showed ducks seroconverted two days later than chickens. Humoral immunity appears to be an important factor in duck resistance to AIV (Yang et al., 2019). When intranasally infected with A/Chicken/Korea/IS/06 (H5N1), ducks began to seroconvert from 4 days post infection (dpi), and the antibody titres were maintained until 28dpi, when the experiment ended (Jeong et al., 2009).

Genetic Basis of Variation

The family of Mx proteins are interferon stimulated genes (ISGs), originally discovered in inbred mice (Lindenmann et al., 1963). Mx proteins are GTPases with varying antiviral functions in different mammalian hosts (Verhelst et al., 2013). Mx is not thought to be antiviral against influenza in chickens (Benfield et al., 2008; Bernasconi et al., 1995) or wild ducks (*Anas platyrhynchos*) (Bazzigher et al., 1993). A comparative study of Mx diversity in teal (*Anas crecca carolinensis*), wigeon (*Anas americana*), mallard (*Anas platyrhynchos*), pintail (*Anas acuta*), and northern shoveler (*Anas clypeata*) showed nucleotide diversity in the *Mx* gene is being maintained at different levels in different duck species. Mallards had the lowest total, silent, and nonsynonymous site diversity of the five species. Low sequence diversity at an immune-related loci could affect how frequently mallards are found to be AIV positive (Dillon & Runstadler, 2010).

There are several reports that PB1-F2 activity, or lack thereof, is host-dependent (Chen et al., 2001; Deventhiran et al., 2016). The different phenotypes induced by PB1-F2 may be caused by differences in host immune systems, i.e. chickens lack the dsRNA sensor RIG-I, and are more susceptible to PB1-F2 pathogenesis than ducks which express RIG-I (Leymarie et al., 2014). In ducks, RIG-I is rapidly upregulated in the lung as part of the early innate immune response against HPAI A/Vietnam/1203/04 (H5N1), but is not significantly induced by LPAI A/mallard/British Columbia/500/2005 (H5N2) (BC500) (Barber et al., 2010). RIG-I is also absent in Red Jungle Fowl, which resembles ancestral chickens, in addition to modern chicken lines. This suggests that chickens may have lost RIG-I prior to domestication. They have retained MDA5, which detects long dsRNA in mammals, and converges with the RIG-I signalling cascade at MAVS, triggering IFNβ induction and ISG expression (Barber et al., 2010). There is evidence that in chickens, MDA5 is able to detect shorter dsRNA strands, and help compensate for the loss RIG-I (Hayashi et al., 2014). MDA5 siRNA knockdown in DF1s can be rescued by transfection of pigeon, duck, or goose RIG-I to restore IFN-β production, but MAVS siRNA knockdown is not rescued in this manner. There is evidence that RIG-I is a more efficient inducer of the antiviral response than MDA5, though this may be an artefact of the methods used (Hayashi et al., 2014; Shao et al., 2015; Xu et al., 2015). Transfection of chicken DF1 cells with duck RIG-I rescues 5'ppp RNA detection and IFNβ promoter activity, leading to lower viral titres of BC500 than WT DF1 cells (Barber et al., 2010). Pigeon RIG-I is a weaker IFN inducer than either of the waterfowl duck or goose, although it shares 78.87% amino acid identity with duck RIG-I. Goose RIG-I was the strongest inducer of IFN-β, Mx and PKR during IAV infection (Shao et al., 2015; Xu et al., 2015). There is a RIG-I independent pathway through TLR7 that contributes to IAV detection and IFNα expression in both chickens and ducks (Barber et al., 2010).

Variation in Mammalian Hosts

Although interspecies transmission is not uncommon, only a limited range of subtypes have developed stable mammalian lineages (Yoon et al., 2014). Distinct lineages have developed in humans, pigs, and horses, that are adapted to the host (Zell et al., 2012). Phylogenetic distance between hosts is thought to be a predictor of the evolutionary success of a cross-species transmission, as this is most common within a host class, i.e. Aves (Wille & Holmes, 2020). IAVs are highly adaptable, due to their error prone polymerase and segmented genome (Vandegrift et al., 2010). Despite this inherent adaptability, successful human-to-human transmission of a zoonotic virus remains rare. A second level of adaptation is required to allow a series of shifts in the influenza replication cycle; from faecal-oral to respiratory transmission; the lower temperature of the human respiratory tracts compared to the avian gastrointestinal tract; changes in HA binding and NA cleavage; differences in host proteins involved in genome replication; and adaptation to new methods of immune antagonism (Wasik et al., 2019).

One of the primary host restrictions of IAV is sialic acid linkage. Mammalian adapted viruses prefer α 2, 6 linked sialic acid whereas avian viruses prefer α 2, 3 linked sialic acids (Long et al., 2019). Pigs are considered high risk for novel pathogenic virus generation because they have both linkages at the site of IAV infection. H5Nx human-human transmission has been reported within households, but is generally limited (Nuñez & Ross, 2019).

The majority of sporadic avian spillover to humans is due to H5N1 or H7N9 viruses. Both subtypes have high case fatality rates (approx. 52%, 39%, respectively), but human-human transmission is rare (Ciminski et al., 2021). H5 Goose/Guangdong and Asian H7N9 viruses have had several zoonotic events with limited transmission. Acquiring a preference for airborne transmission is polygenic, and currently is

not fully understood (Sutton, 2018). An analysis of several seroprevalence studies showed swine workers have a higher likelihood of having antibodies for swine-derived H1N1 (~10%), than poultry workers for avian viruses (<4%), which suggests that transmission from swine to humans may be more common than avian zoonosis (Ciminski et al., 2021).

We know there have been multiple mammalian infections with avian H9 viruses, that have never reached stable circulation due to poor transmission. Asian H9 AIVs have been shown to be replication competent in ferrets, but have poor transmission. In mink, some isolates have the mammalian adaptation E627K in PB2, and a canine H3N2 circulating strain had a H9N2 PA segment (Carnaccini & Perez, 2020). In addition to lineages that have not yet developed, there are lineages that have died out. Horses were once considered a 'dead end' IAV host, before cross-species transmission to dogs (Landolt, 2014). However, despite the cross-species transmission, several equine lineages (H7N7, Eurasian H3N8), and a canine lineage (H3N8) have not been isolated for >10 years, so are considered extinct. This is likely to be due to a lack of susceptible hosts to maintain transmission chains (Landolt, 2014; Wille & Holmes, 2020).

Genetic Basis of Variation

Cellular kinases can have a large impact on success on viral replication. Phosphorylation of a protein can promote scaffolding activities, and affect how efficiently it can interact with binding partners or modulate interactions. Tyrosine, serine, threonine, and lipid kinases have all been identified as having effects on IAV, from entry to vRNP export, to modulation of the host immune system (Meineke et al., 2019). Phosphatidylinositol-3 kinase (PI3K) and c-Jun terminal kinase (JNK) are differently regulated by NS1 proteins of different viral strains. A/mallard/Huadong/S/2005 (H5N1) can induce JNK activation to trigger autophagy and inhibits the PI3K pathway. PR8 is unable to

activate JNK pathway, and A/chicken/Shanghai/F/98 (H9N2) activates the PI3K in DF1 and CEF cells (Zhang et al., 2019). Phosphorylation of NS1 is a host adaptation marker. T215 is a phosphorylation target that is present in all circulating human IAV strains 1920-2008. In avian strains, residue 215 is a proline (Hsiang et al., 2012).

The polymerase is an important factor in determining host range of IAVs. While avian-sourced PB1 segments are associated with pandemic viruses, PB2 has long been recognised as a vital host determinant. Two major amino acid changes associated with host adaptation are E627K and D701N. E627K was first identified in the early nineties by Subbarao et al. Although PB2 was already known to affect host range, this study isolated the specific mutation that allowed escape from a host restricted phenotype (Subbarao et al., 1993). They discovered that lysine vs glutamine at position 627 was separated along host species between human and avian influenza strains.

There are indications that the importance of PB2 627 varies among mammals, however 627K is the preferred amino acid in humans and mice (Shinya et al., 2007). Influenza viruses from different species do not all acquire 627K during mouse adaptation. Some viruses will acquire D701N, but it is rare for a virus to acquire both mutations (Liang et al., 2019; Ping et al., 2010; Shinya et al., 2007). D701N has previously been noted in adaptation of avian H5N1 and seal H7N7 viruses to mice, and aided H5N1 transmission to guinea pigs (Suttie et al., 2019). The D701N mutation removes a salt bridge that normally sequesters a nuclear localisation sequence (NLS) of PB2, allowing increased binding to importin 1α. Exposure of the PB2 NLS increases the efficiency of nuclear localisation of vRNPs (Sediri et al., 2015). Acquisition of E627K rather than D701N appears to be linked to the host protein ANP32. Avian H7N9 in *Anp32*$^{-/-}$ mice developed D701N rather than E627K. Acquisition of E627K is driven by low polymerase activity and interaction with ANP32A (Liang et al., 2019).

Interactions with ANP32 proteins form a major stumbling block for zoonotic viruses. ANP32A and B will not support replication of an avian virus in mammalian cells and vice versa. The avian ANP32A (excluding ratites) has a 33 amino acid insertion compared to the human orthologue that is required to support the avian polymerase. Removal of the insertion is sufficient to prevent support of avian viral polymerase, possibly due to removing a SUMOylation site (Long et al., 2016, 2019). The only mammalian orthologue to support avian polymerase activity without inserting the 33 amino acids is swine ANP32A. Swine ANP32A will support a lower level of polymerase activity through pro-viral positions 106V and 156S increasing binding affinity between ANP32A and the polymerase (Peacock et al., 2020). Some flightless birds such as ostriches do not have the 33aa insertion, so result in mammalian signatures of PB2 during avian influenza virus replication (Long et al., 2016). Binding affinity between the viral polymerase and ANP32A can also be affected by viral genomic RNA (Long, Mistry, et al., 2019).

AIVs are zoonotic and epizootic, leading to development of distinct lineages in different hosts (Webster et al., 1992). As these different hosts exert different immune pressures on the virus, variation in their immune antagonists is required. There are also changes in sialic acid receptor and genome replication machinery between hosts (Long et al., 2019). Adaptation to different host mechanisms has resulted in host differences in AIV infection.

Pathogenesis of IAV Infection

IAV pathogenesis can vary for a variety of reasons, including species, route of exposure, and co-infections. It is well reported that gallinaceous poultry is more likely to suffer severe clinical disease than wild waterfowl (Nuñez & Ross, 2019; Spackman et al., 2010; Yang et al., 2019).

A study of LPAI H7 viruses found that turkeys had more severe clinical disease than chickens or ducks, including significantly higher shedding and mortality (Spackman et al., 2010).
In aquatic birds, IAV primarily replicates in the intestinal tract and is mainly spread by the faecal-oral route (Yoon et al., 2014). Swallowing infected water sources both exposes birds to AIVs and dilutes the acidic gizzard fluid to prevent virion degradation. Seawater can act as a neutral buffer. A/duck/Bavaria/1/77 (H1N1) was found to be able to replicate to titres of 10^5 PFU/ml from gizzard fluid dilutions of 1:3 upwards, and to be able to replicate in neat gut fluid (Han et al., 2019). Although ducks infected with AIVs do shed orally, oral shedding is generally to lower titre or for shorter duration than cloacal shedding (Hénaux & Samuel, 2011) . During spread in domestic birds, several subtypes have acquired a preference for replicating in the respiratory tract, which can aid transmission in densely populated poultry sheds (Yoon et al., 2014).

A/Turkey/Ontario/7732/66 (H5N1) (Ty/Ont) infection of avian lymphocytes caused severe lymphoid depletion through induction of DNA fragmentation leading to apoptosis. Apoptosis was also induced during MDCK infection (Hinshaw et al., 1994). Apoptosis is thought to be necessary for efficient IAV replication (Ampomah & Lim, 2020; Mehrbod et al., 2019). IAVs can induce apoptosis to prevent an immune response to infection (Chen et al., 2001), but hosts can use rapid apoptosis as a mechanism of resistance, by limiting viral replication (Chang et al., 2015; Kuchipudi et al., 2012). IAV can modulate host apoptotic signalling during infection to promote efficient replication and virion release (Zhirnov & Klenk, 2007). The relationship between IAV and apoptosis appears to change over the course of infection (Mehrbod et al., 2019). It has been suggested that IAV modulation of the host IFN response could be an indirect apoptosis control mechanism (Ampomah & Lim, 2020).

Overexpression of *bcl2*, or blocking caspase 3 activation have been seen to reduce virus production (Kuchipudi et al., 2012). Proto-oncogene *bcl2* is an apoptosis inhibitor that was able to protect MDCK from DNA fragmentation during Ty/Ont and PR8 infection. Although the authors note that DNA fragmentation and cytopathologic effects (CPE) appear to be temporally linked, and co-expression of *bcl2* altered localisation of NS1 and NP, they were unable to conclusively identify the viral proteins responsible (Hinshaw et al., 1994). H5N1 NS1 has now been linked to induction of apoptosis in A549 and NCI-H292 human lung epithelial cell lines. NCI-H292 cells transfected with NS1 had significantly higher levels of apoptosis than those transfected with the empty vector control. It is theorised that NS1 may induce apoptosis though an intrinsic mitochondrial pathway that is related to cytochrome c release (Bian et al., 2017). In addition to NS1, accessory protein PB1-F2 is also associated with inducing apoptosis, particularly in immune cells. PB1-F2 primarily activates mitochondrial apoptosis pathways, through interaction with host proteins to form non-specific ion channels (Yoshizumi et al., 2014; Zamarin et al., 2005).

Secondary bacterial infections following influenza can result in hospitalisation or death of both patients with pre-existing lung disease and those who were previously healthy (Rynda-Apple et al., 2015). In a mouse model, a low dose of PR8 IAV was survivable, but a low dose of PR8 IAV followed by a high dose of *S. aureus* was not (Lee et al., 2010). PB1-F2 has been reported to affect susceptibility of mammalian hosts to secondary bacterial infections in numerous studies (Cheung et al., 2020; Pinar et al., 2017; Weeks-Gorospe et al., 2012).

The mechanisms behind the severity of secondary bacterial infections are complex and multifaceted, relying on the interface of viral, bacterial and host immune interactions. Common pathogens identified include opportunists *S. pneumoniae*, *S. aureus*, *H. influenzae*, and *P. aeruginosa.* Damage to the epithelium and reduced ciliary action result in a loss of barrier function which allows increased adherence of

bacteria (Martin-Loeches et al., 2017). IAV infection can also reduce the numbers of alveolar macrophages, in some cases through PB1-F2 induced apoptosis. Immune exhaustion and IAV induced changes to macrophage and neutrophil recruitment can contribute to secondary bacterial infections (Rynda-Apple et al., 2015).

IAV pathogenesis can be highly variable. PB1-F2 has been found to impact multiple pathogenesis factors, such as shedding, apoptosis, and secondary bacterial infections (Alymova et al., 2014; James et al., 2016; Zamarin et al., 2005).

PB1-F2 Structure and Function in IAV Infection

PB1-F2 is a short protein of 87-101 amino acids that is well conserved in avian influenza viruses (Leymarie et al., 2014). Perhaps due to its small size and variability, little structural analysis has been performed on PB1-F2, with only a single partial crystal structure from PR8 available (Bruns et al., 2007) (Figure 1.4). Predicted structures have been suggested, while many groups rely on NMR data or spectroscopy to determine PB1-F2 structure (Pasricha et al., 2018; Solbak et al., 2013).

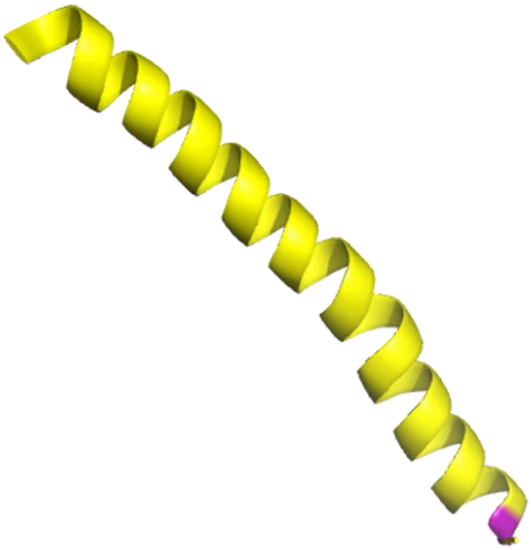

Figure 1.4 Partial PB1-F2 Structure
PR8 PB1-F2 partial structure (aa50-87) from Bruns 2007, visualised in Pymol. (10.2210/pdb2HN8/pdb). AUG 9 highlighted in purple.

PB1-F2 length depends on the strain, but is most commonly 87-90 amino acids long. PB1-F2 is a short-lived protein maximally expressed at ~5 hrs post infection with a reported half-life of one hour when expressed in MDCKs during PR8 infection (Zamarin et al., 2006). However, stability is highly variable and dependent on strain and cell type. There is a basic amphipathic helix in the C-terminal that mediates mitochondrial localisation (Zamarin et al., 2005). It has been postulated that differences in hydrophobicity patterns may affect function, and contribute to the strain specific nature of the protein (Mahardika et al., 2019).

NMR data suggests there are two short α-helices in the N-terminus with an extended helix in the C-terminal, connected by a flexible hinge (Krumbholz et al., 2011). There is evidence that some strains, particularly highly pathogenic ones, may have two helices in the C-terminus (Alymova et al., 2014; Gibbs et al., 2003; Kamal et al., 2018; Yamada et al., 2004). While PR8 is generally reported to have a single continuous α helix in the C-terminal, multiple H5N1 avian strains are predicted to have a helix-loop-helix in the C-terminal (Mahardika et al.,

2019; Solbak et al., 2013). Most of the structural data for PB1-F2 comes from tagged or chemically synthesised peptides, rather than virus expressed protein. The addition of tags may affect protein folding (Gibbs et al., 2003; Solbak et al., 2013; Yamada et al., 2004).

Most PB1-F2 functions are localised to the C-terminus, which contains an α-helix with a mitochondrial targeting sequence. It has been postulated that is some strains, there is also a helix in the N-terminus, however many PB1-F2 proteins have a non-structured N-terminal (Krumbholz et al., 2011). Three alternative start codons (AUGs 7-9) allow for expression of C-terminal fragments of unknown significance (Zamarin et al., 2005). Although the full length ORF is well conserved in avian viruses, there are few suggestions of function for the N-terminus. Functional motifs such as the mitochondrial targeting sequence, virulence and stability markers are commonly identified in the C-terminus. PB1-F2 is highly variable, so there are many reports of ‘minimal requirements’ for a mitochondrial targeting sequence (MTS), and virulence markers or motifs may also vary between proteins (Alymova et al., 2014; Cheng et al., 2017; Košík et al., 2015; Yamada et al., 2004).

In chickens full length PB1-F2 suppresses pathogenicity and lengthens the transmission window by prolonging virus shedding. The enhanced transmission may provide a positive pressure to maintain a full length PB1-F2, when in mammals there is a bias towards truncated proteins (James et al., 2016).

The amino acid sequence 68-72 appears to be important in C-terminus structure and pathogenicity. The lab strain PR8 has 68-ILVF-71, which is associated with a continuous extensive C-terminal helix and low pathogenicity in mice. The 1918 pandemic strain has 68-TPVS-71 while a recent H5N1 has 68-TQGS-71. Threonine, glycine and serine are known helix breakers, and are proposed to enhance pathogenicity in mice and form a helix-loop-helix structure within the C-terminus

(Alymova et al., 2014). Instability of 1918 PB1-F2 has been further isolated to residues 68 and 69, which affect cytoplasmic vs mitochondrial localisation (Park et al., 2019).

A sequence swap of 68-ILVF-71 and 68-TQGS-71 between PR8 and A/Hong Kong/156/1997 H5N1 swapped the proteins' stability. Half-life may also be affected by ubiquitination of lysines at positions 73, 78, and 85 (Cheng et al., 2017). These lysines form a major ubiquitin cluster. The cluster is extremely well conserved in human and avian PB1 sequences present in GenBank – every virus that codes a full length PB1-F2 has the C-terminus lysine cluster (Košík et al., 2015). Ubiquitin is directly involved in protein degradation, MHC (major histocompatibility complex) peptide presentation, and viral recognition. Both 1918 and PR8 PB1-F2 were found to be heavily ubiquitinated in infected cells, and degraded through the ubiquitin-proteasome pathway (Park et al., 2019). Mutating the lysine residues to arginines increases H5N1 PB1-F2 stability, inducing protein accumulation and enhancing both the antibody response to the protein and levels of IFN antagonism (Košík et al., 2015). In contrast, for the 1918 PB1-F2, stabilisation by addition of MG132 resulted in lower IFN inhibition, possibly due to inhibition of a codegradation mechanism (Park et al., 2019).

PB1-F2 has been postulated to have a variety of functions broadly grouped into

- PB1-F2 aggregation/cytotoxic effects
- mitochondrial apoptosis
- modulation of the host innate immune response in a strain-dependent manner
- PB1 interactions

The functions may vary according to virus background and the host the virus is infecting.

PB1-F2 Aggregation and Cytotoxic Effects

PB1-F2 can adopt a β-sheet conformation and oligomerise into amyloid-like structures in membrane-mimicking environments (Ajjaji et al., 2016; Chevalier et al., 2010). Membrane charge and pH may help to regulate conformational changes of PB1-F2 (Vidic et al., 2016). Full length PB1-F2 has been seen to form long filaments, whereas C-terminal fragments form smaller aggregates. Aggregated PB1-F2 is particularly damaging to membranes containing cholesterol or cardiolipin (Ajjaji et al., 2016). Cytotoxicity of PB1-F2 is highly reliant on the protein's structure (Vidic et al., 2016). A recent study found that A/Zhejiang/DTID-ZJU01/2013 H7N9 had a greater likelihood of aggregating than A/Vietnam/1203/2004 H5N1 or lab strain WSN, confirming the strain specific nature of this process (Cheung, Lee, et al., 2020).

Aggregated PB1-F2 can be incorporated into the phagolysosome and activate the NLRP3 inflammasome, or promote apoptosis by inducing membrane damage and cytotoxicity (Bruns et al., 2007; Cheng et al., 2017; Vidic et al., 2016). Aggregated H7N9 PB1-F2 can alter mitochondrial morphology and hinder aggregation of MAVS through steric hindrance. Unaggregated MAVS is destabilised and can be targeted for lysosomal or proteasomal degradation (Cheung, Lee, et al., 2020). PB1-F2 mediated NLRP3 inflammasome activation is related to viral-induced mitochondrial dysfunction. PB1-F2 can induce mitochondrial dysregulation which results in the release of reactive oxygen species (ROS) within the host cell. Inflammasome-driven pyroptosis (an inflammatory form of programmed cell death) causes increased sensitivity to secondary bacterial infection (Pinar et al., 2017).

Mitochondrial Apoptosis

There is a considerable body of evidence that PB1-F2 can interact with and alter function of mitochondria through a variety of mechanisms, as summarised in Figure 1.5. PB1-F2 attacks the mitochondria through protein-protein interactions (Zamarin et al., 2005) and by directly damaging membranes (Vidic et al., 2016). PB1-F2 targets the mitochondria to disrupt host processes and promote apoptosis, particularly in immune cells. PB1-F2 has been found to consistently activate Bax/Bak and cause membrane disruption (McAuley et al., 2010; Yeganeh et al., 2018). Restriction by NLXR1 and the possibility of PB1-F2 forming its own ion channel, similar to the anti-bacterial membrane attack complex are less well reported (Jaworska et al., 2014). There are also conflicting reports over localisation.

PB1-F2 of PR8 interacts with two components of the mitochondrial permeability transition pore complex (PTPC) when overexpressed in 293T cells; adenine nucleotide translocator 3 (ANT3) and voltage-dependent anion channel 1 (VDAC1) (Zamarin et al., 2005). ANT3 is located on the inner mitochondrial membrane, and VDAC1 on the outer. Pull-down experiments indicate PB1-F2 C-terminus interacts with ANT3 and the N-terminus with VDAC1, suggesting PB1-F2 may be present in the intermembrane space of the mitochondria, acting as a bridge to promote PTPC conformation. PB1-F2 mediated mitochondrial permeabilisation is ANT3-dependent (Zamarin et al., 2005). The difference in PB1-F2 mediated apoptosis between immune and epithelial cells may be affected by the different expression levels of ANT3 and VDAC1 between cell lines (Krumbholz et al., 2011). In response to pro-apoptotic stimuli, ANT3 and VDAC1 undergo a conformational change to form nonspecific pores in their respective mitochondrial membranes. The two interact to form the PTPC and dissipate membrane potential while releasing apoptotic mediators (Zamarin et al., 2005). MAVS has also previously been associated with regulation of apoptosis through interactions with VDAC1 (Guan et al.,

2013). PB1-F2's variable function may be partially dependent on its interactions with MAVS as either a pro-apoptotic or IFN stimulating protein (Varga & Palese, 2011).

PB1-F2 associates with the negative inner membrane of mitochondria through electrostatic interactions and forms a stable highly ordered complex containing >3 molecules (Yoshizumi et al., 2014). The protein has a high affinity for membranes and can alter membrane integrity, triggering mitochondrial dysfunction in immune cells (Leymarie et al., 2014). Mitochondrial membrane potential is lost through depolarisation of transmembrane potential (Yamada et al., 2004). PB1-F2 of PR8 can induce mitochondrial fragmentation, and abnormal structures that inhibit mitochondria-mediated innate immune responses (Yoshizumi et al., 2014). PB1-F2 may detain the cell cycle during S-phase to promote mitochondrial fragmentation (Yamada et al., 2004). PB1-F2 in the presence of vRNA can trigger mitochondrial DNA release into the cytoplasm via TOM40 (Moriyama et al., 2020). Mitochondrial DNA triggers an innate immune response via TLR9, NLRP3, and cGAS pathways (Piantadosi, 2020). Mitochondrial DNA can also be released by Bax/Bak induced apoptosis, which can be triggered by PB1-F2 (Mcauley et al., 2010; Piantadosi, 2020).

PB1-F2 can be vulnerable to host restriction factors at the mitochondria. The nucleotide-binding oligomerisation domain-like receptor X1 (NLRX1) binds PB1-F2 to prevent the induction of apoptosis (Jaworska et al., 2014). NLRX1 is a pattern recognition receptor (PRR) in the NOD-like receptor (NLR) family that is thought to localise to the mitochondrial matrix, through a localisation signal in the N-terminus effector domain. NLXR1's interaction with PB1-F2 plays a central role in the survival of alveolar macrophages, reducing inflammation and restricting viral spread in the early stages of infection (Jaworska et al., 2014). In the early stages of infection, NLXR1 is able to bind PB1-F2 at mitochondria and may compete for interactions with components of the PTPC. In later stages, PB1-F2 is thought to

overwhelm host restriction and induce apoptosis (Jaworska et al., 2014).

PB1-F2's apoptotic activities can be enhanced by acting as an 'immune cell trap'. Increased expression of chemokines Csf3 and Cxcl2 increases infiltration of leukocytes such as neutrophils and monocytes into the lung, where they are more vulnerable to infection and PB1-F2 induced apoptosis (Cheung et al., 2020).

Possibly as a consequence of the lack of a mitochondrial targeting sequence (MTS), PB1-F2 from A/Hong Kong/156/1997 (HK156) H5N1 does not induce apoptosis in human monocytic U937 cells (Chen et al., 2010). In porcine alveolar macrophages (PAM), avian H5N1 PB1-F2 does induce apoptosis, and shows high levels of caspase 3/7 activation (Chang et al., 2015). Phosphorylation of threonine 27 and serine 35 of lab adapted H1N1 PB1-F2 is critical for inducing apoptosis in monocytes but this activity is not seen in H5N1 protein, despite the residues being conserved (Chen et al., 2010). PB1-F2 expression induces expression of host genes associated with cell death, inflammatory responses and neutrophil chemotaxis (Hao et al., 2020).

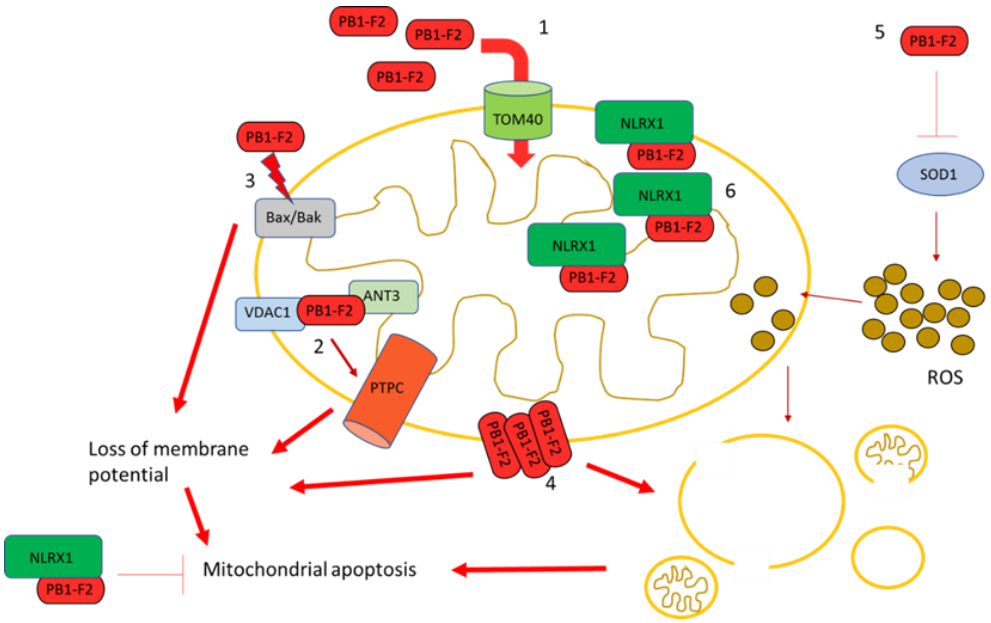

Figure 1.5 Mitochondrial Interactions of PB1-F2 PB1-F2 causes loss of membrane potential or damages membranes to induce mitochondrial apoptosis.

1) PB1-F2 enters mitochondria through the TOM40 channel, 2) It interacts with VDAC1/ANT3 to form the PTPC, 3) PB1-F2 activates proapoptotic proteins Bax/Bak, 4) PB1-F2 oligomers disrupt mitochondrial membranes, 5) Repression of SOD1 causes increase in ROS, which damage mitochondria, 6) Host protein NLRX1 restricts PB1-F2 activity

Inspired by (Jaworska et al., 2014; Kamal et al., 2018; Klemm et al., 2018; Krumbholz et al., 2011; Yoshizumi et al., 2014; Zamarin et al., 2005)

Immune Modulation by PB1-F2

Several virulence markers have been described in various PB1-F2 proteins, mostly at C-terminal positions. One of the best studied is N66S, which was present in the 1918 pandemic strain PB1-F2. In mice, this marker is associated with increased weight loss, higher viral titres, and cytokine dysregulation in the lungs (Conenello et al., 2007). When a pandemic 2009 strain A/California/04/2009 (Cal/09) PB1-F2

was restored to a full length ORF, 66S was found to increase expression of pro-inflammatory cytokines IL1β and RANTES but had no effect on the susceptibility of mice to secondary bacterial infections (Hai et al., 2010). In a mouse model, HK156 (H5N1) N66S was found to suppress early IFN responses, allowing uncontrolled replication that caused a surge of cytokine and chemokine production from 3dpi. This delayed signal resulted in increased leukocyte infiltration and more severe immunopathology in the lungs (Conenello et al., 2011).

In an analysis of PB1-F2 sequences N66S was found to be enriched in avian H2N2 and H3N2 subtypes compared to human isolates. All H5N1 isolates from human hosts containing the mutation were HP viruses. Unexpectedly for a virulence marker, all bar one avian isolates containing N66S were considered to be LPAI (Pasricha et al., 2013). Following this unexpected result, in turkey poults, a triple reassortant virus with a human source PB1 segment coding PB1-F2 N66S failed to cause clinical signs of infection and was not transmissible to contacts. Turkey poults infected with WT 66N virus developed diarrhoea from 5dpi and the virus was able to be reisolated from contacts. Both groups seroconverted, though N66S had lower HAI titres at 10dpi. PB1-F2 activity may also have been affected by the presence of two noninflammatory motifs, H75 and Q79 (Deventhiran et al., 2016). 1918 PB1-F2 with N66S has been found to dysregulate innate immune pathways in macaques, promoting fatal infection (Conenello et al., 2011). N66S may be a host-specific virulence factor that promotes inflammation only in mammalian hosts.

Over 80% of H5N1 PB1-F2 sequences contain four or more virulence-associated residues; L62, S66, R75, R79, and L82. The majority of virus strains coding these residues are of avian origin (Smith & McCullers, 2013). These inflammatory markers have been found in pandemic strains from 1918 and 1968, but have been lost in descendant viruses. Extracellular PB1-F2 mediated immunopathology (such as NLRP3 inflammasome activation) may have a large effect on

virulence in pandemic IAVs, which is then lost as the virus adapts to the human host (Cheung et al., 2020). L62, R75, R79, and L82 inflammatory markers are associated with increased leukocyte infiltration into the lungs. L82 encourages aggregation, which may cause activation of pattern recognition receptors (PRRs), leading to an increased inflammatory response (McAuley et al., 2017). PB1-F2 proteins containing L62, S66, R75, R79, and L82 are associated with enhanced pathogenicity of secondary bacterial infections in mammals (Weeks-Gorospe et al., 2012). Extracellular, possibly aggregated, PB1-F2 containing these markers can be necrotic rather than apoptotic. Necrosis of lung epithelial cells may repress antibacterial immunity and increase bacterial adhesion, promoting secondary bacterial infection (Cheung et al., 2020).

PB1-F2 Inhibits Mitochondrial Innate Immune Functions

PB1-F2 can impair formation of the MAVS signalosome in a variety of ways, depending on virus strain. Common methods include dissipation of membrane potential, perturbation of mitochondrial dynamics to trigger mitophagy, or MAVS can be sequestered and blocked from binding other signalosome components (Cheung et al., 2020). Several viruses induce mitophagy as a way to attenuate the host immune response, as mitochondria can act as a scaffold for immune complexes such as the NLRP3 inflammasome. Damaged mitochondria are recognised by autophagic receptors that trigger PRKN-dependent (parkin RBR E3 ubiquitin protein ligase) mitophagy. The receptors are recruited to mitochondria by ubiquitin binding and autophagosome engulfment is mediated through associations with LC3B (R. Wang et al., 2020).

PB1-F2 proteins from PR8 and A/duck/Hubei/Hangmei01/2006 (H5N1) both induce mitophagy, through interactions with TUFM (Tu translation elongation factor, mitochondrial) and LC3B. PB1-F2 contains a LIR motif that can bridge between TUFM and LC3B to

trigger TUFM-dependent mitophagy and promote MAVS degradation to allow attenuation of the host innate immune response (Wang et al., 2020). Aggregated H7N9 PB1-F2 can destabilise MAVS, and block MAVS-TRIM31 interactions that are required for MAVS aggregation into an innate immune scaffold (Cheung, Lee, et al., 2020). PB1-F2 mediated mitochondrial fission will also inhibit signalosome formation as mitochondrial fusion is required. Interactions with nuclear dot protein 52 (NDP52) can target MAVS for autophagic degradation (Cheung et al., 2020).

NLRP3 is activated on sensing cell damage caused by infection, and is recruited to the mitochondria to form the NLRP3 inflammasome. The inflammasome activates caspase 1 and is involved in maturation and secretion of pro-inflammatory cytokines IL1β and IL18. The mitochondria play a central role in inflammasome activation (Moriyama et al., 2020). MAVS acts as a platform for NLRP3 to bind and oligomerise to activate the NLRP3 inflammasome. This activation can be inhibited by multiple PB1-F2 mechanisms depending on the virus strain. H7N9 PB1-F2 is able to block MAVS-NLRP3 interactions and prevent RNA-induced activation of the inflammasome, while WSN PB1-F2 prevents the activated inflammasomes' involvement in IL-1b maturation (Cheung, Ye, et al., 2020). Proteins from HPAI viruses are more potent intracellular suppressors than those from LPAI viruses, and only HPAI PB1-F2 proteins have been shown to activate the inflammasome as an extracellular aggregate (Cheung et al., 2020).

MAVS recruits IKK family members to activate NFκB and IRFs. PB1-F2 interferes at multiple steps along this pathway. PB1-F2 can bind to the transmembrane domain of MAVS, or target proteins further down the pathway, including IKKB, and CALCOCO2 to inhibit TRAF2/3 mediated IFN induction (Hao et al., 2020). PB1-F2 binding of MAVS or IKKB appears to be localisation dependent. In DF1 cells, mitochondrial PB1-F2 bound MAVS and inhibited the IFN pathway,

whereas cytoplasmic PB1-F2 had a stronger interaction with IKKB and inhibited translocation of NFkB to the nucleus (James et al., 2019).

PB1 – PB1-F2 Interactions

Analysing reports of PB1/PB1-F2 interactions is complicated by the interdependence of segment 2 protein expression (Firth & Brierley, 2012). For example, without PB1-N40, viruses with an intact PB1-F2 overexpress PB1 early in infection and replicate slowly (Wise et al, 2009). Not all papers have taken into account the overexpression of PB1-N40 induced by AUG 4 deletion. Introduction of an early stop codon within the PB1-F2 ORF is considered to be a less noisy method of removing PB1-F2 (Wise et al., 2009). However, this is further complicated by expression of C-terminal fragments of unknown functionality (Zamarin et al., 2005).

Multiple studies showing PB1-F2 effecting polymerase activity use PR8 polymerase segments with the PB1 gene of interest, which is a questionable strategy considering the extreme strain specificity of PB1-F2 (Chen et al., 2010; McAuley et al., 2010). Further muddying the waters are studies that use a ΔAUG 4 strategy, despite the effect this mutation can have on PB1-N40 (Mazur et al., 2008; McAuley et al., 2010). None of these studies have considered expression of C-terminal fragments. Confocal microscopy suggests that in some strains, a proportion of PB1-F2 isolates to the nucleus (Chen et al., 2010), where it could be involved in retaining PB1. Evidence for direct interactions between PB1 and PB1-F2 comes mostly from confocal microscopy (McAuley et al., 2010), but includes yeast two hybrid assays and immunoprecipitation from infected cells (Mazur et al., 2008). There is limited reproducibility of these results, suggesting PB1/PB1-F2 interactions may occur only in limited subtypes.

Thesis Aims

In this introduction, we identified several unanswered questions regarding PB1-F2. As an accessory protein, PB1-F2 is less well studied than some of the essential proteins of IAV. Much research has focused on the zoonotic risks of an avian PB1-F2 in a mammalian host (Kamal et al., 2015; Pasricha et al., 2018; Pinar et al., 2017; Schmolke et al., 2011), rather than expression from an avian virus in an avian host (James et al., 2016, 2019; Marjuki et al., 2010; Xiao et al., 2020).

One of the basic questions we cannot currently answer is 'What is the full range of proteins expressed from the PB1-F2 ORF?' While there have been reports of C-terminal fragment expression (Kamal et al., 2015; Zamarin et al., 2006), no investigation of the individual AUG codons has been performed. We also wanted to examine a report of increased prevalence of N-terminal truncations of PB1-F2 (Kamal et al., 2015), as this could also impact expression from the ORF. The majority of functional motifs in PB1-F2 have been localized to the C-terminus (Alymova et al., 2014; Cheng et al., 2017; James et al., 2016; Park et al., 2019; R. Wang et al., 2020), which raises questions regarding the minimal expression requirements for PB1-F2 to be functional during infection. A reinitiation mechanism has already been identified in segment 2 of IAV (Wise et al., 2011). Reinitiation within the PB1-F2 ORF could allow expression of functional fragments from an N-terminally truncated sequence.

We are currently unaware of any comparison of function between full length PB1-F2 and C-terminal fragments. In this thesis we explored how altering the start and stop codon composition of the PB1-F2 ORF affected polymerase activity, viral replication kinetics, and IFN antagonism *in vitro*. To our knowledge this is the first investigation of AUG 7 and 9 independent of one another rather than grouped as 'C-terminal AUGs' (Zamarin et al., 2006).

Previous work in the Shelton lab suggested a tight correlation between localization and stability of PB1-F2 proteins (James et al., 2019). We wanted to expand the analysis of the PB1-F2 panel to test this hypothesis, and begin to consider how protein structure could affect stability. There is currently little structural data available for PB1-F2 (Bruns et al., 2007), and what has been reported is mostly secondary structure helix positions (Chevalier et al., 2010; Solbak et al., 2013). Considering the known link between structure and function (Thornton et al., 1999; Uversky, 2019), this is a large gap in the literature.

In this thesis, we aimed to;

Investigate how the PB1-F2 ORF structure is conserved in different hosts, by a bioinformatic analysis of sequences. Conservation of full-length PB1-F2 sequences is known to vary between mammalian and avian viruses (James et al., 2016), so we extended this to consider maintenance of the C-terminus. The bioinformatics results were then used to inform our choice of mutants for experimental validation of expression.

Further explore the relationship between localization and stability, by expanding the stability analysis performed in James et al. (2019). We also investigated possible structural impacts on PB1-F2 stability and how varying functions between strains could impact protein structure.

Investigate how changing the level of C-terminal fragment expression could affect the PB1-F2 phenotype, and confirm whether C-terminal fragment expression alone is sufficient for PB1-F2 to functional *in vitro*.

The main conclusions reached include;

That the maintenance of the PB1-F2 ORF and likelihood of C-terminal fragment expression varies not only between avian and mammalian hosts, but also between different avian families.

That the relationship between localization and stability is not as linear as previously suggested in James et al. (2019).

Expression of C-terminal fragments in the absence of full length PB1-F2 is sufficient for function *in vitro*.

Chapter 2: Bioinformatic Analysis of the Open Reading Frame Structure of PB1-F2

Introduction

Reasons for and Methods of Avian Influenza A Virus Surveillance

Avian influenza is a known zoonotic risk and is endemic in poultry populations of several countries. Therefore, the virus is closely monitored on both national and international levels to attempt to prevent spread. WHO's Health Emergencies Programme monitors human cases of avian influenza and releases a weekly update on risk, confirmed infections, and deaths (World Health Organisation, 2021). However, WHO is a human public health agency so does not monitor avian infections.

In the UK, avian influenza (AIV) is a notifiable disease, which means that any suspected case or dead wild bird must immediately be reported to DEFRA (Department of Environment, Food, and Rural Affairs), or in Scotland, the Field Services Office. There is also a legal requirement to register any group of ≥50 poultry to allow tracking if AIV enters the country. If AIV is confirmed, a Restricted Zone may be enacted to prevent spread (Department for Environment, Food & Rural Affairs & Animal and Plant Health Agency, 2020). The UK is a member country of the World Organisation for Animal Health (OIE), which tracks international outbreaks of AIV.

The OIE requires member countries to report any HPAI AIV detected in domestic or wild birds and any LPAI of H5 or H7 subtype detected in poultry. The OIE also details biosecurity measures aimed at preventing AIV and control strategies should an outbreak occur (OIE - World Organisation for Animal Health, 2020).

Surveillance of AIV is often limited to subtype prevalence in bird populations based on sequences of HA and NA glycoproteins, but on occasion some studies are able to look at the other genes in more

detail to examine viral evolution and possible virulence factors (Kayed et al., 2019; Rimondi et al., 2018; Sarkar et al., 2012). Internal viral genes are sequenced to determine presence of zoonotic risk mutations or known virulence markers of specific segments (Sarkar et al., 2012), or to track the reassortment of internal genes (Kayed et al., 2019; Lycett et al., 2019; Rimondi et al., 2018). In many parts of the world, surveillance is done in response to disease outbreaks rather than being a long term project (Rimondi et al., 2018), which can make it hard to elucidate long term trends.

Segment 2 surveillance is important due to multiple previous pandemics having PB1 genes of avian origin. 1918, 1957, and 1968 pandemic viruses all sourced HA and PB1 from avian precursors (Pappas et al., 2008; Wendel et al., 2015). Several studies have shown that introducing an avian PB1 into a human influenza backbone can increase viral polymerase activity and enhance viral replication (Pappas et al., 2008; F. Wang et al., 2020; Wendel et al., 2015). Segment 2 also codes for the accessory protein and virulence factor PB1-F2, which has been reported to enhance viral virulence in mammalian hosts (Conenello et al., 2007; Kamal et al., 2015; Mettier et al., 2021). Unfortunately, many IAV segment 2 sequences in National Centre for Biotechnology Information, Influenza Virus Resource (NCBI IVR) are not annotated for PB1-F2, which is expressed from an alternative reading frame. Comprehensive annotation of segment 2 sequences could help assess pandemic risks.

IAV Lineage-Specific Differences in PB1-F2 ORF Maintenance

It is well reported that there are differences in PB1-F2 length between avian and mammalian viruses (Ajjaji et al., 2016; James et al., 2016; Schmolke et al., 2011; Zell et al., 2007). A previous analysis found that 96% of avian influenza A isolate segment 2 sequences in GenBank encoded a PB1-F2 protein of at least 79 residues, defined as an “intact” gene (Zell et al., 2007). After the H1N1 09pdm strain rose to

prominence, with a truncated PB1-F2 gene, only 7% of H1N1 human isolates coded a complete PB1-F2 ORF (defined as ≥87 amino acids in this paper) (Leymarie et al., 2014).

An analysis of H5N1 viruses found that 11.1% of PB1-F2 amino acids in avian hosts were conserved, compared to a far more substantial 45.5% of amino acids in human PB1-F2 sequences, with conserved residues spread throughout the ORF (Pasricha et al., 2013). A previous study determined that 40 nucleotide sequence positions in the PB1-F2 ORF were considered to be at risk of stop codon introduction with a single nucleotide change (Zell et al., 2007), which is a substantial proportion for a small protein. The amino acids conserved in avian hosts included residues previously identified as required for a functional mitochondrial targeting sequence (MTS) in the C-terminus of the protein (Pasricha et al., 2013).

Possible PB1-F2 Protein Products

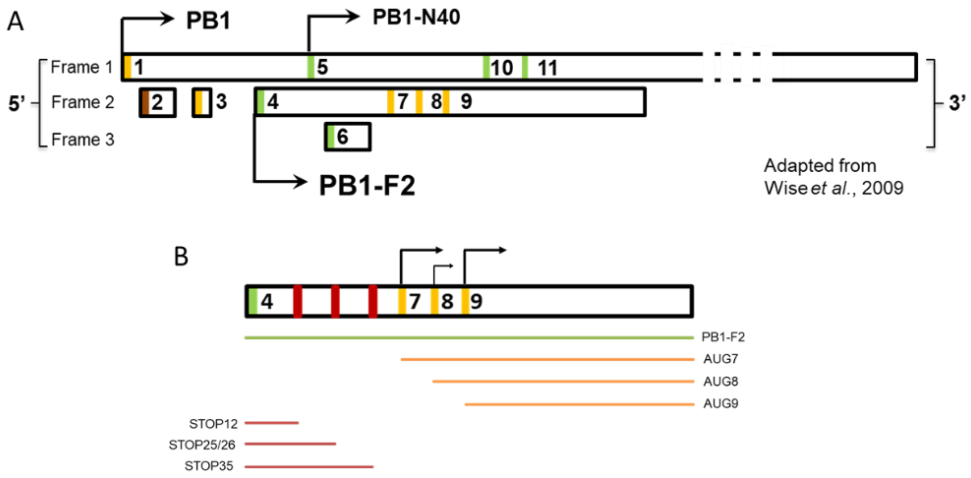

Figure 2.1 Translational Start and Stop Sequences in the 5' End of Segment 2 and the PB1-F2 ORF

A) A schematic view of the 5' end of segment 2. Black boxes indicate ORFs across the three reading frames, with coloured numbered bars for start codons. Kozak signalling strength shown in red for weak, orange for intermediate, and green for strong as defined in Kozak (1984). Black arrows indicate known sites of translation initiation.

B) Within the PB1-F2 ORF (black box) there are a range of stop codons

found in viral isolates, shown in red but unnumbered, and potential alternative start codons 7-9. Coloured lines beneath the ORF indicate predicted PB1-F2 polypeptides that are shortened by early stop codons (red) or are produced from initiation at the C-terminal AUGs (orange), with the full-length product shown in green. Arrows indicate likelihood of C-terminal fragment expression from AUGs 7-9.

IAV segment two has multiple mechanisms of expression to expand the proteome across multiple open reading frames (Figure 2.1) (Vasin et al., 2014; Wise et al., 2011). Full length PB1-F2, generally defined as ≥87aa, is expressed from AUG 4 via leaky ribosomal scanning (Chen et al., 2001; James et al., 2016; Wise et al., 2009, 2011). PB1-F2 expression is host specific, as not all IAV lineages maintain the ORF (James et al., 2016). It has been reported that C-terminal fragments of PB1-F2 expressed from one or more of AUGs 7-9 are produced *in vitro* (Kamal et al., 2015; Zamarin et al., 2006), but it is unclear how these AUGs are accessed by ribosomes. Both AUG 4 and AUG 5 have stronger Kozak signalling than AUGs 7-9 (Kozak, 1984), and should be 'seen' first during ribosome scanning, the mechanism by which PB1-F2 AUG 4 is recognised. Many influenza lineages code early STOP codons in the PB1-F2 ORF (Kamal et al., 2015, 2018). A ribosomal termination-reinitiation mechanism has previously been reported for segment 2, so N-terminal truncations may not entirely ablate PB1-F2 expression (Wise et al., 2011). The PB1-F2 ORF may be capable of producing multiple peptide products which may or may not be functional (Figure 2.1 B).

The Importance of N-terminal truncations and C-terminal Functionality of PB1-F2

An N-terminal truncation of the PB1-F2 gene or loss of segment 2 AUG 4 can affect the expression level of C-terminal fragments of PB1-F2 (Kamal et al., 2015). Functional C-terminal motifs in PB1-F2 include a mitochondrial targeting sequence (MTS) within an α helix, a TQDS motif that enhances pathogenicity in mice, and a ubiquitinated lysine

cluster that affects protein stability (Alymova et al., 2014; Cheng et al., 2017; Košík et al., 2015; Zamarin et al., 2005).

Human pandemic viruses from 1918, 1957, and 1968 all encoded a full length PB1-F2, but virulence markers such as L62, R75, R79, L82 (Alymova et al., 2018; Smith & McCullers, 2013) and expression of the full length ORF have been lost over time and circulation in human hosts. The H1N1 2009 pandemic strain has three stop codons within the PB1-F2 ORF, including STOP12 and STOP58, that prevent expression (Hai et al., 2010). It is theorised that PB1-F2 is an immune antagonist in avian hosts that is either ineffective or detrimental to the virus in mammalian hosts (Kamal et al., 2018). Retention of the full length protein in virus strains that infect chickens can extend the period of virus transmission (James et al., 2016), whereas in mice, a PB1-F2 knock out virus gave reduced pathogenicity and mortality (Zamarin et al., 2006). PB1-F2 is known to be host specific in function, but this may result from host-specific evolution, rather than an innate difference in host-pathogen interactions (Pasricha et al., 2013).

Due to loss of PB1-F2 in mammalian strains during host adaptation, there is concern that reintroduction of the full-length protein into a circulating strain could increase pathogenicity. The circulating virus reservoir in birds remains a potential source of PB1-F2 sequences that could be highly virulent in humans (Kamal et al., 2018). It has been suggested that one of the factors in the 1918 pandemic strain's high pathogenicity was the avian virus-derived source of segment 2 expressing a full length PB1-F2 (Weeks-Gorospe et al., 2012). Correlating PB1-F2 sequences to pathogenicity of IAV may aid in predicting severity of emerging viral strains (Pasricha et al., 2013; Weeks-Gorospe et al., 2012).

Aims

Internal IAV genes are not as intensely surveyed as those encoding the external glycoproteins. PB1-F2 surveillance is further complicated

by the three C-terminal AUGs 7-9 that may produce functional protein fragments. The aims of this chapter were to assess how well conserved the PB1-F2 ORF is in AIVs isolated from wild birds, and to consider whether an N-terminal truncation of the ORF would maintain the functional motifs identified in the C-terminus of the protein. Our goals were;

- Bioinformatic analysis of PB1-F2 ORF structure
- Investigate maintenance of N-terminally truncated PB1-F2 ORFs
- Investigate conservation of AUGs 7-9, and the PB1-F2 C-terminal ORF within the subset of viruses with a truncated PB1-F2 gene
- Investigate whether N-terminal truncations are host or subtype specific

Results

Lineage-Specific Differences in PB1-F2 ORF Structure

To investigate the frequency of N-terminal truncations in IAV PB1-F2, a dataset of 10,279 PB1 sequences was produced from the NCBI Influenza Virus Resource (IVR) (Bao et al., 2008) by our collaborator, Dr Samantha Lycett in July 2018. The dataset was a stratified subsample of virus isolates with complete genome sequences in the NCBI IVR. Sequences from 1902-2018 were collected from avian, swine, human, equine, and canine hosts, and poor-quality sequences (non-full length or with excessive ambiguities) were removed (Lycett et al., 2019).

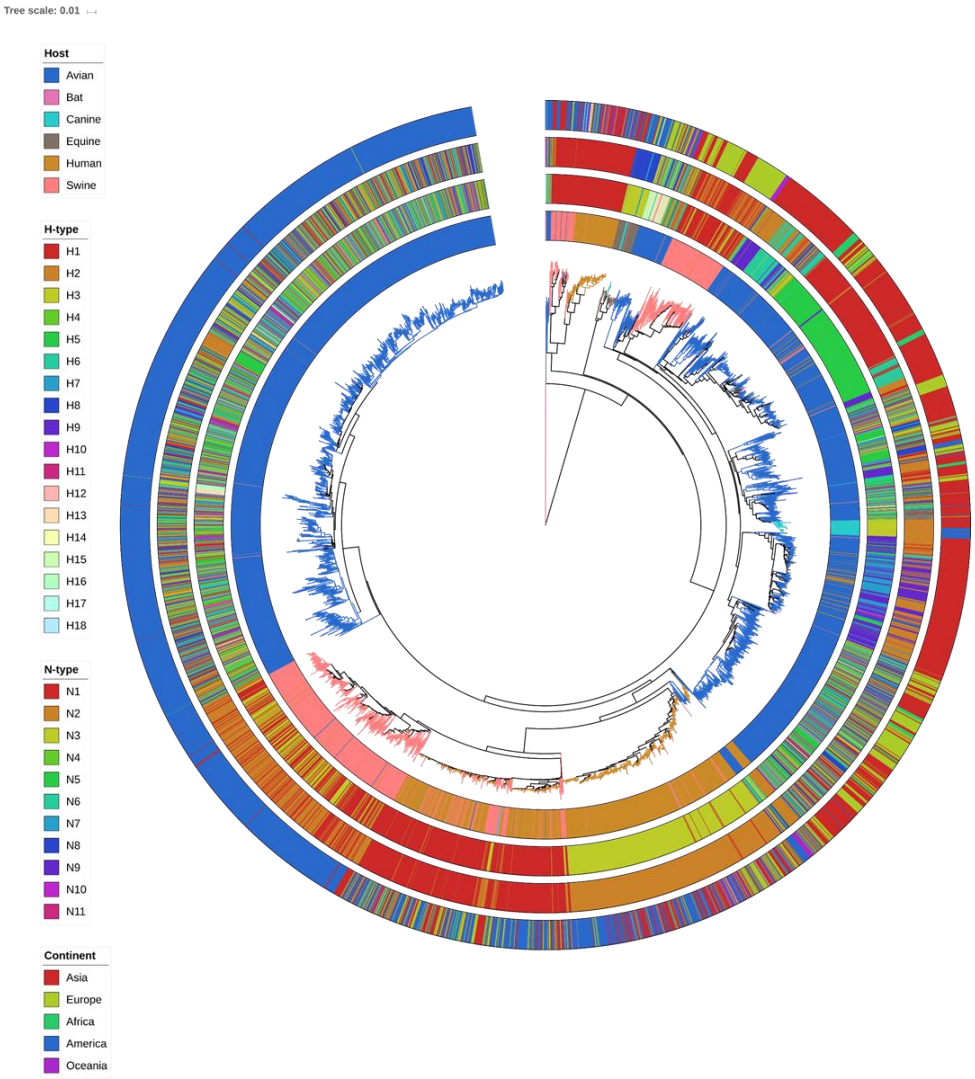

Figure 2.2 Segment 2 Phylogeny of a Stratified Sample of All Influenza Subtypes.

Tips in the circular neighbour-joining tree are coloured by host: blue, avian; pink, swine; orange, human; green, canine; brown, equine; purple, bat. The rings from inner to outer are: host-type, haemagglutinin subtype (H-type), neuraminidase subtype (N-type) and continent of isolation. Figure provided courtesy of Dr Samantha Lycett.

The stratified dataset (Figure 2.2) of 10279 sequences was comprised of a majority of avian (5940 sequences), then human (2673 sequences), and swine (1484 sequences) hosts. A smaller number of isolates from other hosts, such as canines (92 sequences) and equines (86 sequences), were also present. The dataset covered the

full diversity of HA and NA subtypes, including three H17N10 and one H18N11 representatives of bat influenza subtypes. The majority of samples came from Asia and America, followed by Europe (Figure 2.2). Further details of the sampling process are available (Lycett et al., 2019).

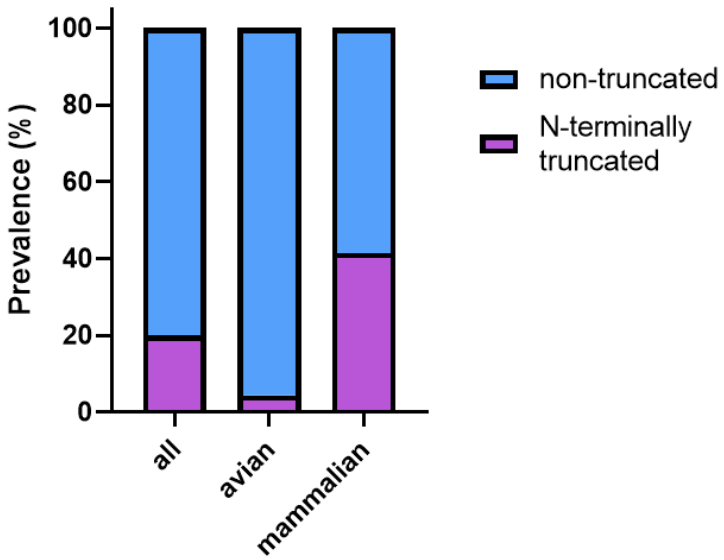

Figure 2.3 Prevalence of N-terminal Truncations in PB1-F2
Influenza A virus segment 2 sequences from multiple virus subtypes and host species from the mixed host dataset had PB1-F2 ORF status (N-terminal truncation or non-truncated) determined for each sequence. The percentage prevalence of N-terminally truncated sequences (defined as the presence of a stop codon before position 39 or the lack of AUG 4) within the dataset was calculated for the complete mixed host dataset, and avian and mammalian hosts subsets.

The prevalence of N-terminal truncations within the PB1-F2 sequences of the mixed host dataset of 10,279 sequences was then calculated as a percentage of the dataset or host subset of sequences (Figure 2.3). N-terminal truncations were defined as any STOP codon before residue 39 (AUG 7), or the loss of AUG 4 through mutation. 20.01% of PB1-F2 sequences within the dataset contain a truncation in the N-terminal region. When the dataset was separated into sequences from avian or mammalian hosts, we found a large

difference in the frequency of N-terminal truncations. In avian isolates, only 4.38% of 5940 sequences were truncated, compared to 41.42% of 4339 mammalian sequences. The increased presence of N-terminal truncations within PB1-F2 sequences from mammalian compared to avian hosts is consistent with previous reports regarding possession of full length (>87aa) sequences (James et al., 2016).

STOP Codon Positions and Prevalence in PB1-F2 ORF

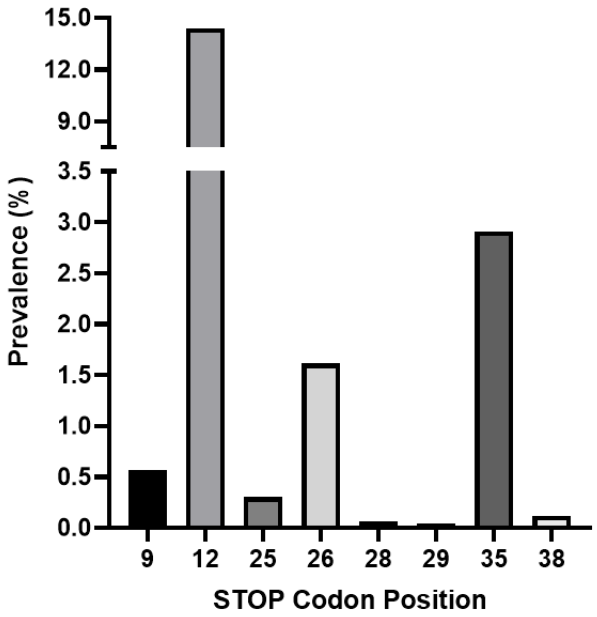

Figure 2.4 PB1-F2 STOP Codon Prevalence in the Mixed Host Dataset
Early STOP codon positions in the mixed host PB1-F2 dataset were annotated for each sequence. The percentage prevalence of each stop codon within the mixed host dataset was then calculated.

Once we confirmed N-terminal PB1-F2 truncations were present in this dataset, we investigated where in the ORF truncations appeared. We were particularly interested in N-terminal truncations that could theoretically allow C-terminus expression from AUGs 7-9. The prevalence of pre-AUG 7 STOP codon positions within the PB1-F2 ORF was calculated (Figure 2.4). Of the eight different STOP codons

present upstream of AUG 7in the 10279 sequences within the dataset, only three (STOP12, STOP26, STOP35) were in >1% of sequences. There was a large range in STOP codon frequencies, from 0.04-14.41% prevalence. The most common truncation position was STOP12, present in 14.41% of sequences, followed by STOP35 with a frequency of 2.91% and STOP26 with a frequency of 1.61%.

STOP Codon Position	Prevalence in avian PB1-F2 % complete dataset (N=10279)	Prevalence in avian PB1-F2 (N=5940)	Prevalence in mammalian PB1-F2 % complete dataset (N=10279)	Prevalence in mammalian PB1-F2 (N=4339)
9	0.12	0.20	0.45	1.06
12	0.92	1.60	13.48	31.94
25	0.06	0.10	0.24	0.58
26	0.38	0.66	1.24	2.93
28	0.01	0.02	0.05	0.12
29	0.03	0.05	0.01	0.02
35	0.95	1.65	1.96	4.63
38	0.06	0.10	0.06	0.14

Table 2.1 Quantification of N-terminal STOP Codon Prevalence in the Mixed Host Dataset

N-terminal STOP codons were annotated for each sequence. The percentage prevalence of each stop codon within the complete mixed host dataset (N=10279) and the host specific subsets was then calculated.

We also investigated host bias in STOP codon positions by quantifying prevalence in avian or mammalian host subsets of the data presented in Figure 2.4. Several STOP codons had very different prevalence's in the host subsets, such as STOP12 which was present in 1.60% of avian sequences compared to 31.94% of mammalian sequences (Table 2.1). Almost all the STOP codons were more prevalent within mammalian sequences than avian sequences, which is consistent with the higher frequency of N-terminally truncated PB1-F2 sequences within viruses from mammalian hosts (Figure 2.3).

AUG Codon Positions and Prevalence in the PB1-F2 ORF

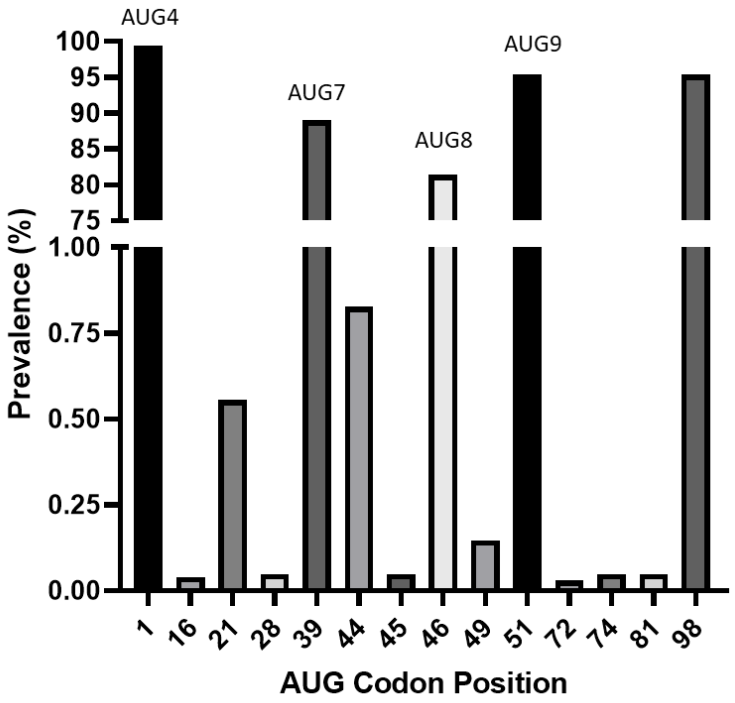

Figure 2.5 AUG Codon Prevalence in the Mixed Host Dataset
AUG codons within the PB1-F2 ORF were annotated for each sequence in the dataset. The percentage prevalence of each AUG codon within the complete mixed host dataset was then calculated.

Having determined the STOP codon landscape of the PB1-F2 ORF, prevalence of AUG codons within the ORF was calculated (Figure 2.5). The most common Met codons within the 10279 sequence dataset were at positions 1, 39, 46, and 51, corresponding to AUGs 4, 7, 8, and 9 respectively in the nomenclature used here (Figure 2.1 A), but we also saw a range of other AUG codons present in <1% of sequences. 99.49% of sequences had AUG 4, indicating some PB1-F2 sequences have a later start site, possibly AUG 7, as the first AUG codon of the ORF. The C-terminal AUGs 7, 8, and 9, were present in 89.03%, 81.39%, and 95.47% of sequences respectively. Several of the rarer AUG codons were not present in all hosts (Table 2.2). Avian sequences did not contain M16, M21 or M74. IAVs isolated from

mammalian hosts were more likely to code AUG 8 (M46), with a prevalence of 97.10% compared to 69.92% in avian hosts (Table 2.2).

AUG Codon Position	Prevalence in avian PB1-F2 % complete dataset (N=10279)	Prevalence in avian PB1-F2 (N=5940)	Prevalence in mammalian PB1-F2 % complete dataset (N=10279)	Prevalence in mammalian PB1-F2 (N=4339)
1	57.58	99.65	41.91	99.29
16	0.00	0.00	0.04	0.09
21	0.52	0.89	0.04	0.09
28	0.00	0.00	0.05	0.12
39	51.44	89.02	37.58	89.03
44	0.54	0.93	0.29	0.69
45	0.03	0.05	0.02	0.05
46	40.40	69.92	40.99	97.10
49	0.10	0.17	0.05	0.12
51	54.57	94.43	40.90	96.89
72	0.03	0.05	0.00	0.00
74	0.00	0.00	0.05	0.12
81	0.02	0.03	0.03	0.07
98	54.89	94.97	40.48	95.90

Table 2.2 Quantification of AUG Codon Prevalence in the PB1-F2 Gene
AUG codons were annotated for each PB1-F2 sequence in the dataset. The percentage prevalence of each AUG codon within the complete mixed host dataset and the host specific subsets was then calculated.

The host-specific variation in AUG and STOP codon positions suggests that ORF structure of PB1-F2 is another lineage specific feature of this protein.

Maintenance of the C-terminal ORF of PB1-F2

Dataset (N=sequences examined)	N-terminally truncated	Non-truncated	Truncated maintain Cterm AUG7 (nº)	Truncated maintain Cterm AUG7 %	Truncated maintain Cterm AUG9 (nº)	Truncated maintain Cterm AUG9 %
10279 (N=10279)	1855	8424	497	26.79	476	26.42
Avian (N=5940)	239	5701	226	94.56	193	80.75
Galliformes (N=1041)	115	926	94	81.74	86	74.78
Anseriformes (N=4021)	117	3904	95	81.20	67	57.26
Charadriiformes (N=536)	10	526	10	100.00	10	100.00
Mammalian (N=3839)	1565	2274	267	17.06	266	17.00
Human (N=2673)	1059	1614	101	9.54	102	9.63
Swine (N=1484)	490	994	166	33.88	183	37.35

Table 2.3 PB1-F2 ORF Maintenance Within Mixed Host Dataset
The table quantifies ORF maintenance in the mixed host dataset. A conserved C-terminus is defined as one without STOP codons from aa39 (AUG 7) – 87 or aa51 (AUG 9) - 87. The percentage prevalence of N-terminally truncated sequences (defined as the presence of a stop codon before position 39 or the lack of AUG 4) within the dataset was calculated for the complete mixed host dataset, then avian and mammalian host subsets. Maintenance of a C-terminal ORF starting from either AUG 7 or AUG 9 was also calculated.

Our starting hypothesis was that the presence of N-terminal (pre-residue 39) truncations could allow expression of functionally significant C-terminal PB1-F2 fragments from AUGs 7-9. In avian PB1-F2 sequences, AUG 8 is the least well conserved and is not present in either of our model viruses A/turkey/England/50-92/1991 H5N1 (5092) and A/chicken/Pakistan/UDL-01/2008 H9N2 (UDL), with which subsequent investigations would be carried out, so we focused on AUGs 7 and 9. In order to assess viability of our C-terminal fragment hypothesis, we calculated how many PB1-F2 sequences with an N-terminal truncation had preserved the C-terminal ORF (Table 2.3). Within the complete mixed host dataset (10279 sequences), 26.79% of sequences preserved the C-terminal ORF from AUG 7 (residues 39-91). This dropped to 17.06% in isolates from mammalian hosts, while

avian hosts preserved the C-terminal ORF in 94.56% of cases. In all avian host Orders, the C-terminal ORF was better conserved beginning from AUG 7 than in those starting from AUG 9. The variation in C-terminal maintenance between hosts indicates that mammalian PB1-F2 sequences are less likely to be full length than avian PB1-F2 sequences (James et al., 2016), not only due to N-terminal truncations, but due to STOP codons throughout the ORF. However, the high levels of C-terminal maintenance within avian hosts supported our original reinitiation hypothesis.

PB1-F2 ORF Structure in Endemic AIVs

We were able to confirm the presence of N-terminal truncations within the PB1-F2 gene of AIVs. From the large multi-host dataset of 10279 sequences, we then chose three endemic avian strains of virus to focus on. H5N1, H7N9, and H9N2 AIV subtypes are all of economic concern, and considered to be zoonotic pandemic risks. These three subtypes were selected for further investigation.
The H5N1 subtype has been endemic in poultry in China for nearly twenty years and remains a high zoonotic risk (C. M. Bui et al., 2017), while continuous circulation of avian H5N1 viruses has caused enormous economic losses and remains a public health concern. H9N2 is a LPAIV (low pathogenicity avian influenza A virus), however it is endemic to much of Asia and the Middle East. There have been several cases of H9N2 viruses making the zoonotic leap into humans, and the lack of effective control methods makes this subtype a human pandemic risk (Pusch & Suarez, 2018; Rahimirad et al., 2016). H9N2 viruses are also a source of much reassortment within bird populations, with transfer of a H9N2 internal gene cassette being common (He et al., 2017; Lam et al., 2015). H7N9 emerged as a novel zoonotic AIV in China in 2013 (C. M. Bui et al., 2017). There is a background level of circulating LPAI H7N9 viruses that are a constant risk for mutating into a HPAI strain (Shi et al., 2017). H7N9 is thought to use H9N2 as a source of reassortment when infected chickens are

in close proximity (Lam et al., 2015). While all three subtypes are of zoonotic and economic concern, studying H9N2 and H7N9 sequences in particular may help us to understand circulation of N-terminally truncated PB1-F2 sequences during reassortment events.

STOP Codon Positions and Prevalence in the PB1-F2 ORF of Endemic AIVs

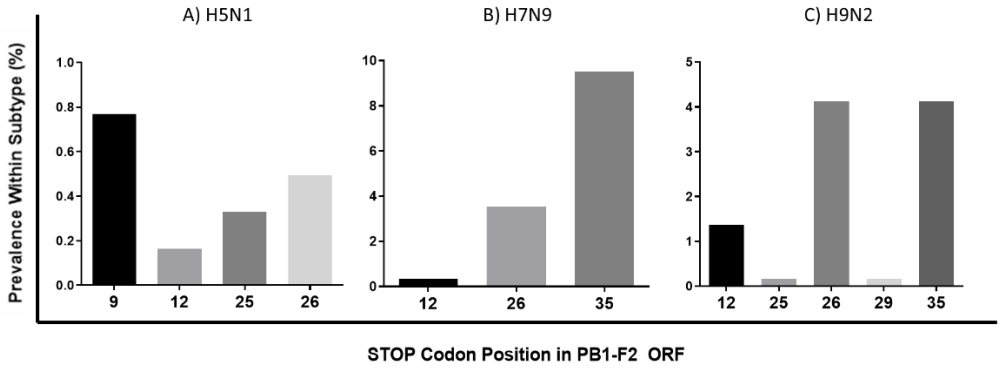

Figure 2.6 STOP Codon Positions of N-terminally Truncated PB1-F2 Sequences

N-terminal STOP codons were annotated for each sequence in the subtype datasets. The percentage prevalence of each stop codon before residue 39 was calculated within each subtype.

To investigate how consistent N-terminal STOP codons were between the major circulating avian subtypes, the prevalence of each STOP codon position before aa-39 (AUG 7) in PB1-F2 was calculated (Figure 2.6). The datasets consisted of 1823 H5N1 sequences, 567 H7N9 sequences, and 1239 H9N2 sequences. STOP codons at positions 12 and 26 appeared in all three subtypes at varying frequencies. H5N1 did not reach a full 1% for any of the four N-terminal STOP codons identified in this dataset, ranging from STOP12 at 0.17% to STOP9 at 0.77%. STOP9 was the most prevalent stop position in H5N1, which was the only subtype to code for it. H7N9 had the highest prevalence of STOP codons among the three subtypes, with a highly prevalent STOP codon at position 35 present in 9.52% of sequences. H9N2 viruses had the most STOP codon positions, and STOP26 and STOP35 were equal in frequency (4.15%

of sequences). H9N2 was the only subtype to code STOP29, but its prevalence is very low, at 0.16%. Although N-terminally truncated sequences are present across all three subtypes, the method of achieving PB1-F2 truncation is highly variable. This analysis also indicated that STOP codon prevalence varies between avian IAV subtypes, not just between sequences from avian or mammalian hosts.

AUG Codon Positions and Prevalence in the PB1-F2 ORF in Endemic AIVs

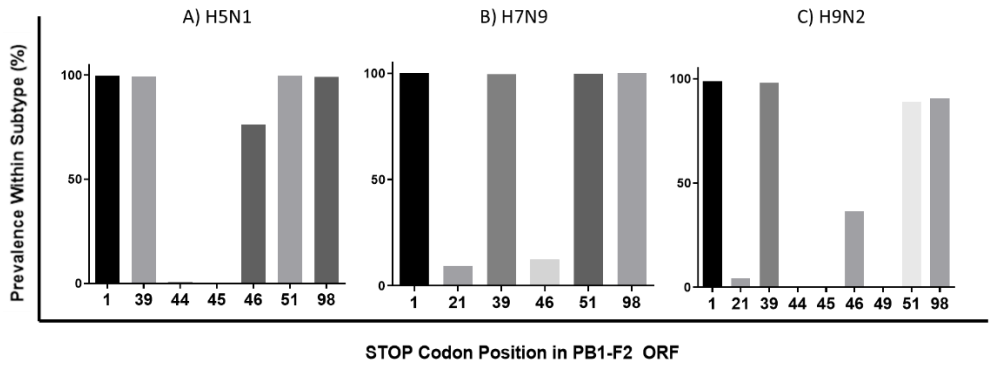

Figure 2.7 Prevalence of PB1-F2 AUG Codons in Endemic AIV Subtypes
AUG codons were annotated for each sequence in the subtype datasets. The percentage prevalence of each AUG codon in the PB1-F2 ORF was calculated within each subtype.

To investigate how consistent AUG codons were between subtypes, prevalence of each AUG codon in the PB1-F2 ORF was calculated (Figure 2.7). The canonical PB1-F2 start codons AUG 4 (position 1), AUG 7 (position 39), AUG 8 (position 46) and AUG 9 (position 51), were amongst the most conserved for all three subtypes, as was expected. H7N9 and H9N2 both show less than 50% prevalence of AUG 8, which was unsurprising as AUG 8 is the least conserved PB1-F2 AUG codon in AIV sequences in general (Table 2.2).

All three subtypes had multiple additional AUG codons within the PB1-F2 ORF. In H5N1 (Figure 2.7 A), the additional AUG codons had very low incidences of 0.67% Met44 and 0.49% Met45, whereas both H7N9

(Figure 2.7 B) and H9N2 (Figure 2.7 C) have an AUG at position 21 that is 9.35% and 4.36% prevalent respectively. It is not unusual for H7N9 and H9N2 to have similar frequencies of the internal PB1-F2 AUG codons, as H7N9 internal genes all originate from H9N2 viruses (Lam et al., 2015). Met21 is the most prevalent additional AUG codon identified in H9N2 and H7N9 avian subtypes. Despite its low prevalence compared to AUGs 7-9, Met21 has similar strength Kozak signalling to the canonical start codons (AUGs 4, 7-9) of PB1-F2, with a purine at -3 and a G at +4 (Kozak, 1984). In all subtypes, AUG 7 and AUG 9 are highly conserved (AUG 7 97.9-99.65%, AUG 9 88.94-99.73%) so could be capable of expressing functional C-terminal fragments of PB1-F2 in multiple strains of virus. AUGs 7-9 may also be conserved due to benefits of Mets within the sequence of the full-length PB1-F2 protein.

Maintenance of the C-terminal ORF of PB1-F2 in Endemic AIVs

Subtype	N-terminally truncated	Non-truncated	Truncated maintain Cterm AUG7 (n°)	Truncated maintain Cterm AUG7 %	Truncated maintain Cterm AUG9 (n°)	Truncated maintain Cterm AUG9 %
H7N9	77	490	77	100.00	77	100.00
H9N2	123	1116	112	94.92	116	100.00
H5N1	40	1783	37	92.50	37	92.50

Table 2.4 PB1-F2 ORF Maintenance in Endemic AIV Subtypes
The table quantifies PB1-F2 ORF maintenance in the H9N2, H7N9, and H5N1 datasets, presented as percentage of the subtype. N-terminal truncation is defined as either a STOP codon pre-aa39 (AUG 7) or lack of AUG 4. A conserved C-terminus is defined as one without STOP codons from aa39 (AUG 7) – 87 or aa51 (AUG 9) - 87.

After confirming N-terminal truncations of the PB1-F2 gene were present in our three AIV subtypes of interest, we investigated whether the C-terminal ORF of PB1-F2 was being maintained as had previously been seen in Table 2.3. As seen in the wider avian subset (Table 2.3), and in Figure 2.6, the prevalence of N-terminal truncations was low in these datasets. However, when truncations did occur, we saw a high level of C-terminus maintenance. All three subtypes

showed >90% of truncated sequences were potentially capable of expressing a full C-terminal fragment of PB1-F2. The H7N9 viruses had no STOP codons in the C-terminal ORF, and although H5N1 lineage strains were least conserved among the three subtypes, they still reached 92.50% maintenance of the C-terminal PB1-F2 coding region (Table 2.4).

Phylogenetics was used to define the clades of viruses with N-terminally truncated PB1-F2 sequences (see below for further description). From each clade, a representative sequence was taken to look at variance of the ORF structure within subtype families. Sequences were aligned, and the potentially expressed protein fragments highlighted in green boxes (Figure 2.8). Across and between subtypes, there was variance in the position of the STOP codon in the N-terminal region of PB1-F2. Some sequences lost N-terminal expression through AUG 4 mutation rather than through the introduction of an early STOP codon.
Of the three representative sequences for H7N9, all retained the start codon for PB1-F2 (AUG 4), two of the three showed a stop codon at position 35 and one at position 26. In contrast the H5N1 viruses showed more variability, with one representative clade mutating the AUG 4 start for PB1-F2 from Met to Val. Four of the representative sequences coded a stop at position 26, one at 25 and one at 12. H9N2 also showed high variability in PB1-F2 truncation mechanism, with two clades mutating AUG 4 from Met to Thr, three clades coding STOP12, five clades coding STOP26 and two clades coding STOP35 (Figure 2.8).

Figure 2.8 Alignment Summary of Truncated PB1-F2 Proteins

One representative PB1-F2 sequence was selected per clade from phylogenetic trees of each subtype and aligned in MEGA. Green boxes represent ORFs, blue box indicates loss of AUG 4. While the position of the N-terminal stop codon (-) varies, all sequences maintain AUG 7, and most sequences express the full C-terminus of PB1-F2.

All sequences maintain AUG 7, which is the first of three potential alternative C-terminal start codons in PB1-F2. AUG 7 may be recognised by ribosomes to begin translation and produce a C-terminal fragment of PB1-F2. All sequences bar H5N1 clade 1 show the possibility of a full C-terminal fragment being expressed. Conservation of the C-terminal ORF could support the theory that loss of N-terminal expression is a means of streamlining PB1-F2 expression and minimising resources wasted on a non-functional section of protein, or removing an unwanted function of the N-terminus. The other interpretation of this data is that mutations within PB1-F2 are evolutionarily neutral (Kimura, 1991). Neutral theory suggests that the majority of evolutionary changes at a molecular level are due to random fixation of selectively neutral mutations due to random genetic drift. If the mutation does not have a negative impact on survival (for IAV, replication and transmission), then sequences will diverge over time as they acquire mutations (Kimura, 1991). Within IAV, neutral mutations can be introduced during infection that are able to be transmitted to the next host (Sigal et al., 2018). However, negative selection of PB1 also affects the variation of PB1-F2, as PB1 polymerase function requires the sequence to be maintained or limited to synonymous mutation (Zell et al., 2007). Synonymous mutations occur at a much higher rate than nonsynonymous mutations as they are likely to be functionally neutral (Kimura, 1991). The majority of PB1-F2 mutations, including STOP12 and STOP35, can be made silently in PB1 due to the +1 overlap of the ORFs.

The high prevalence of AUGs 7 and 9, combined with maintenance of the C-terminal ORF within the clades suggests that expression of C-terminal fragments of PB1-F2 may be a common and conserved strategy within endemic avian viruses.

Maintenance of N-terminal PB1-F2 Truncations in Circulating Avian Viruses

Having shown that N-terminal truncations were present in avian PB1-F2 sequences, we sought to determine if the introduction of truncations into the sequences were individual events or whether they were being maintained in viruses circulating in avian populations. To do this we performed a phylogenetic analysis on the downloaded subtype datasets by using MEGA7 to calculate Maximum Likelihood trees from PB1 sequences, with a bootstrap of 500. Trees were then colour coded to identify sequences truncated through introduction of an early stop codon or the loss of AUG 4 (Figure 2.9). Branches ending in two or more truncated sequences were defined as clades and their phylogeography further investigated to attempt to correlate a species or location with N-terminally truncated sequences.

Phylogenetic Trees of Endemic IAVs with Truncated PB1-F2 Genes

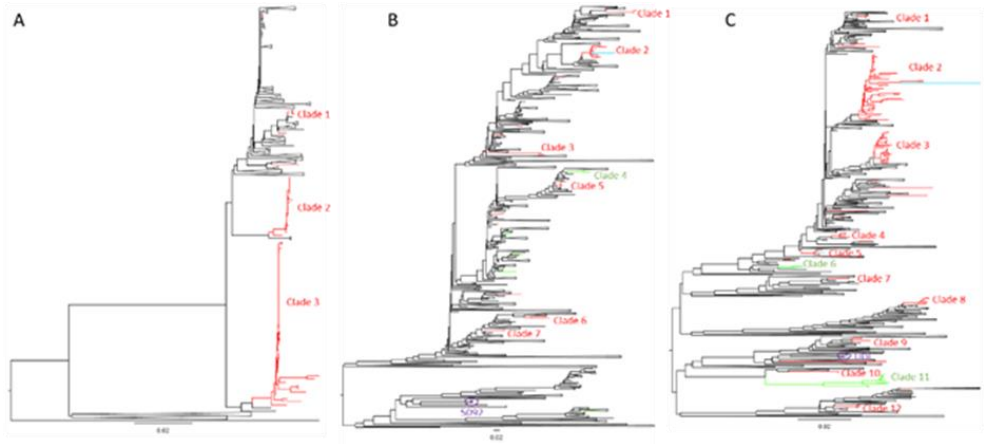

Figure 2.9 Phylogenetic Trees Showing Truncated PB1-F2 Sequences. Maximum Likelihood trees A) H7N9, B) H5N1, C) H9N2 produced using the subtype datasets analysed above in MEGA7 with a bootstrap value of 500. Sequences in red have N-terminal PB1-F2 truncations, in green lack AUG 4, in blue are full length PB1-F2 ORF in an otherwise truncated clade. Any branch ending in 2 or more sequences was defined as a clade for analysis. The model viruses A/turkey/England/50-92/1991 H5N1 (5092) and A/chicken/Pakistan/UDL-01/2008 H9N2 (UDL) are highlighted in purple.

Subtype	Clade number	Number of sequences	Most prevalent truncation	Date range	Circulation length (yrs)	Location
H9N2	1	2	26	2015	1	China
	2	47	35	2012-2017	5	China/Japan
	3	23	26	2012-2013	2	China
	4	4	12	2009-2011	3	China
	5	5	35	2007-2010	3	China
	6	2	ΔAUG4	2001-2002	2	Japan
	7	2	26	2014	1	Poland
	8	6	12	2009-2012	3	Hong Kong/China
	9	3	26	2011-2013	2	Bangladesh
	10	2	12	2006	1	Saudi Arabia
	11	10	ΔAUG4	2015-2017	2	UAE/Morocco/Burkina Faso
	12	3	26	2005	1	Israel
H5N1	1	3	26	2017	1	Vietnam
	2	9	12	2014-2015	2	Bangladesh
	3	2	26	2010	1	Laos
	4	2	ΔAUG4	2014	1	Egypt
	5	4	25	2011-2011	2	Egypt
	6	2	26	2011-2012	2	Vietnam
	7	2	26	2007	1	Cambodia
H7N9	1	2	35	2017	1	China
	2	19	26	2013	1	China
	3	52	35	2014-2017	3	China

Table 2.5 Quantification of Truncated PB1-F2 Clades Identified in Phylogenetic Trees

The table quantifies the clades shown in Figure 2.9 for each subtype of AIV, showing variation of stop codon position between clades and maintenance of truncation over time and approximate geography. H9N2 clades are highlighted in blue, H5N1 in yellow, and H7N9 in green.

Clades of all three subtypes were quantified (Table 2.5) for easier comparison. H9N2 viruses had the most clades, and appeared to be the most geographically mobile. Although H5N1 viruses had seven distinct clades of truncated sequences, the clades were small (median 2 sequences, Table 2.5) and all except clade 2 did not persist over 12 consecutive months. In contrast, H7N9 viruses had only three clades ranging from 2-52 sequences, with clade 3 persisting over three years. H9N2 viruses had 12 clades, with a median size of 3 sequences. The majority persisted over multiple influenza seasons, and were geographically mobile. H9N2 clades 2 and 11 crossed political

borders, and five clades had isolates that spanned distances of >1000km.

The persistence over time and geography in H7N9 and H9N2 subtypes indicate that N-terminal truncations of PB1-F2 are being maintained and we can therefore hypothesise that truncations are unlikely to cause any detrimental effects to the virus. For example, H9N2 clade 11 moved from the Middle East to North Africa, likely through poultry movements (Zecchin et al., 2017), and maintained a truncated PB1-F2 sequence for three years (Table 2.5).

Phylogenetic Trees of Avian H5Nx Viruses

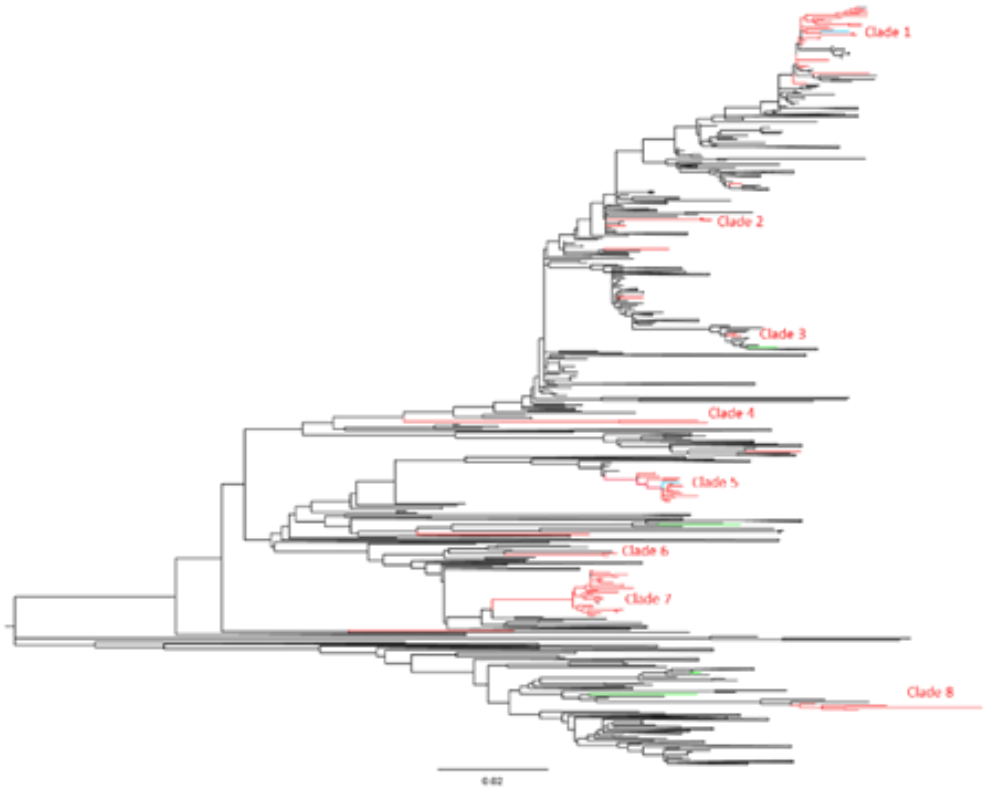

Figure 2.10 Avian H5Nx Viruses Showing Presence of Truncated PB1-F2 Sequences.

Maximum Likelihood trees produced using avian H5Nx sequences from the stratified multi-host dataset in MEGA7 with a bootstrap value of 500. Sequences in red have N-terminal truncation, in green lack AUG 4, in blue have a full length PB1-F2 ORF in an otherwise truncated clade. Any branch

ending in 2 or more sequences was defined as a clade for analysis. Clades are numbered from top of the tree down for analysis.

Subtype	Clade number	Number of sequences	Most prevalent truncation	Date range	Circulation length (yr)	Location
H5N6/H5N2	1	21	12	2013-2016	3	China/HongKong/Vietnam/Laos
H5N1	2	2	26	2010	1	Laos
H5N1	3	2	25	2010-2011	2	Egypt
H5N6	4	2	12	2015	1	China
H5N8/H5N6	5	16	12	2013-2015	2	China
H5N9	6	2	26	2015	1	France
H5N8/H5N6	7	24	12	2016-2018	2	Russia/Belgium/Korea/Egypt/ Cameroon/China/DRC/Uganda
H5N2	8	4	9	2005-2010	5	USA/Mexico

Table 2.6 Quantification of H5Nx Clades Identified in Phylogenetic Tree
The table quantifies the clades shown in Figure 2.10, showing variation of stop codon position between clades and maintenance of truncation over time and geography.

H5N1 showed limited persistence of N-terminally truncated PB1-F2 sequences (Table 2.5), so the analysis was expanded to the avian H5Nx sequences of the stratified multi-host dataset to determine whether all H5 viruses prefer to maintain a full length PB1-F2 ORF or if this is a H5N1 specific phenotype. As above, any branch ending in >2 branches were defined as a clade (Figure 2.10) and clades were quantified in Table 2.6.

H5Nx segment 2 sequences showed three clades containing >10 isolates, as well as several closely related viruses that did not form a clade (Figure 2.10). The median clade size was 4 sequences, similar to the median H9N2 clade size. Although H5N1 did not show maintenance of truncated sequences, several clades of H5Nx showed persistence over multiple influenza seasons (62.5% circulating >1yr) and/or widespread sampling locations, such as Clade 7, which was isolated in Europe, Africa, and Asia. Clades 1, 5, and 7 contained sequences from multiple H5Nx subtypes and all were in circulation for multiple years (Table 2.6). The identification of multiple majority-truncated clades persisting over multiple influenza seasons indicated

that H5Nx viruses are better able to circulate with a truncated PB1-F2 sequence than the H5N1 dataset suggested.

Cluster Analysis of Endemic Avian Viruses

To assess whether the N-terminally truncated PB1-F2 sequences present in the subtype datasets were phylogenetically related and therefore likely to have been maintained in a circulating virus, relationships between segment 2 sequences were analysed using ClusterPicker (CP) and ClusterMatcher (CM) programmes (Ragonnet-Cronin et al., 2013). The PB1 alignments used to generate the phylogenetic trees in Figures 2.8-9 were used to assess phylogenetic relationships using CP. The CP support threshold was set to 0.9, with a genetic distance threshold of 0.45. The support threshold is the minimum support required to define a cluster, and refers to the bootstrap support for the tree. The genetic distance threshold is the maximum genetic distance allowed within a defined cluster. Clusters are defined by high support and low within cluster genetic distance. Setting support threshold to 0.9 gives 90% bootstrap support for clusters and setting genetic distance to 0.45 allows a maximum of 4.5 substitutions/site within clusters. These settings have been successfully used in the literature and were suggested to be suitable for phylogenetic investigation (Ragonnet-Cronin et al., 2013). All subtypes were analysed using the same parameters.

Within this set of parameters, CP identified 158 highly related clusters in the H9N2 dataset, shown in Figure 2.11 B. CM was then used to examine the distribution of truncated PB1-F2 sequences, with a threshold of >3 sequences and at least 1% of sequences being truncated in each cluster. This resulted in 10 clusters that matched the parameters, such as Cluster18 (Figure 2.11 C). In those 10 clusters were 52 PB1-F2 sequences, with varying truncation statuses. 100% of sequences in Clusters 18, 130, and 156 were truncated. Clusters 117,

57, and 62 were all 80% truncated, whereas Clusters 111, 96, and 61 contained 75% full length PB1-F2 genes.

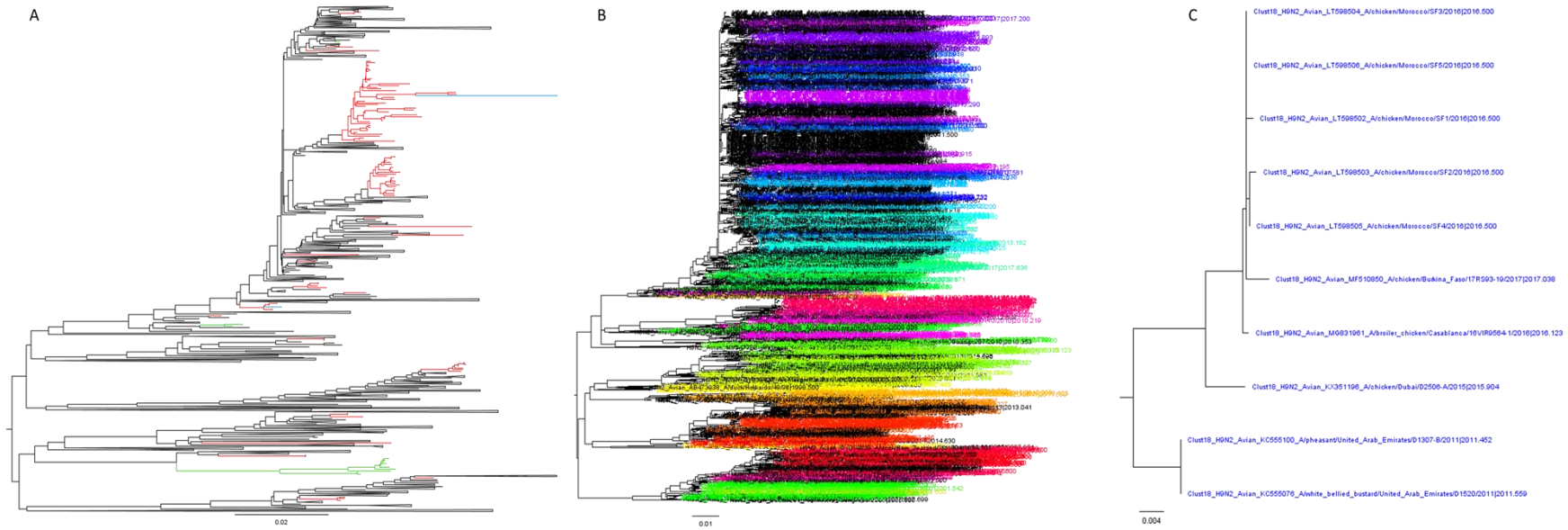

Figure 2.11 Cluster Analysis of H9N2

To determine how closely genetically related PB1-F2 sequences are, newick tree A (previously shown in Figure 2.8) is inputted to ClusterPicker with a FASTA alignment file to output annotated tree B, with clusters of highly related viruses colour coded. Non-clustered virus names remain black. Tree B is inputted to ClusterMatcher with an annotation file of truncation status to correlate phylogenetic relationship with a characteristic of interest. Highly truncated clusters are identified and outputted as separate trees such as C.

Analysis of the H7N9 dataset only identified 11 clusters. There were fewer sequences in the H7N9 dataset than the other subtypes (Table 2.7), so this was not unexpected. The dataset is dominated by samples from chickens in China in 2013-2014, which limited genetic diversity within the dataset due to heavy sampling of few outbreaks. This is visible in Figure 2.12 B as the dominance of yellow and orange sequences. Three clusters were identified by CM as having at least 1% truncated sequences, such as Cluster 5, shown in Figure 2.12 C. Clusters 4 and 5 correlated with clades 2 and 3, respectively. The three matched clusters contained 332 sequences, all of which in Cluster 5 were truncated, with two full length PB1-F2 sequences in Cluster 4, and 5/259 truncated sequences in Cluster 1 (Table 2.7). All three clusters showed a majority of Galliforme isolates (Cluster 1 91.12%, Cluster 4 90.48%, Cluster 5 96.15%), with a further preference for domestic poultry hosts (Cluster 1 96.52%, Cluster 4 90.48%, Cluster 5 100%).

When cluster analysis was performed for H5N1 viruses, CP identified 97 clusters, totalling 233 sequences. In comparison to the other subtypes analysed, the clusters were small (median 2 sequences), with Figure 2.13 clearly showing fewer clustered (coloured) sequences than in Figures 2.10 and 2.11. However, no clusters containing at least 1% truncated sequences were identified by CM. As the H5N1 sequence group appeared to be without highly related groups of truncated PB1-F2 sequences, we concluded that in the H5N1 subtype there was unlikely to be persistence of the N-terminally truncated sequences. This fitted with the limited transmission of only a 12-month period seen for the majority of H5N1 clades and the lack of geographical spread identified in Table 2.6.

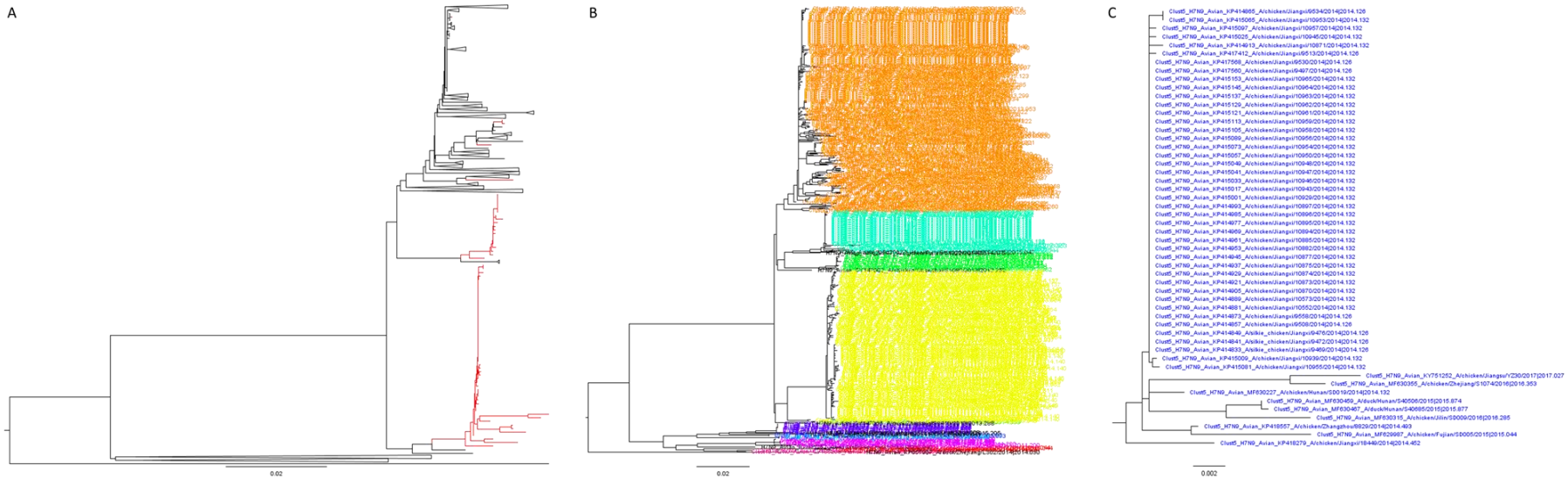

Figure 2.12 Cluster Analysis of H7N9

To determine how closely genetically related PB1-F2 sequences are, phylogenetic tree A (previously shown in Figure 2.8) is inputted to ClusterPicker with a FASTA alignment file to output annotated tree B, with clusters of highly related viruses colour coded. Non-clustered virus names remain black. Tree B is inputted to ClusterMatcher with an annotation file of truncation status to correlate phylogenetic relationship with a characteristic of interest. Highly truncated clusters are identified and outputted as separate trees such as C.

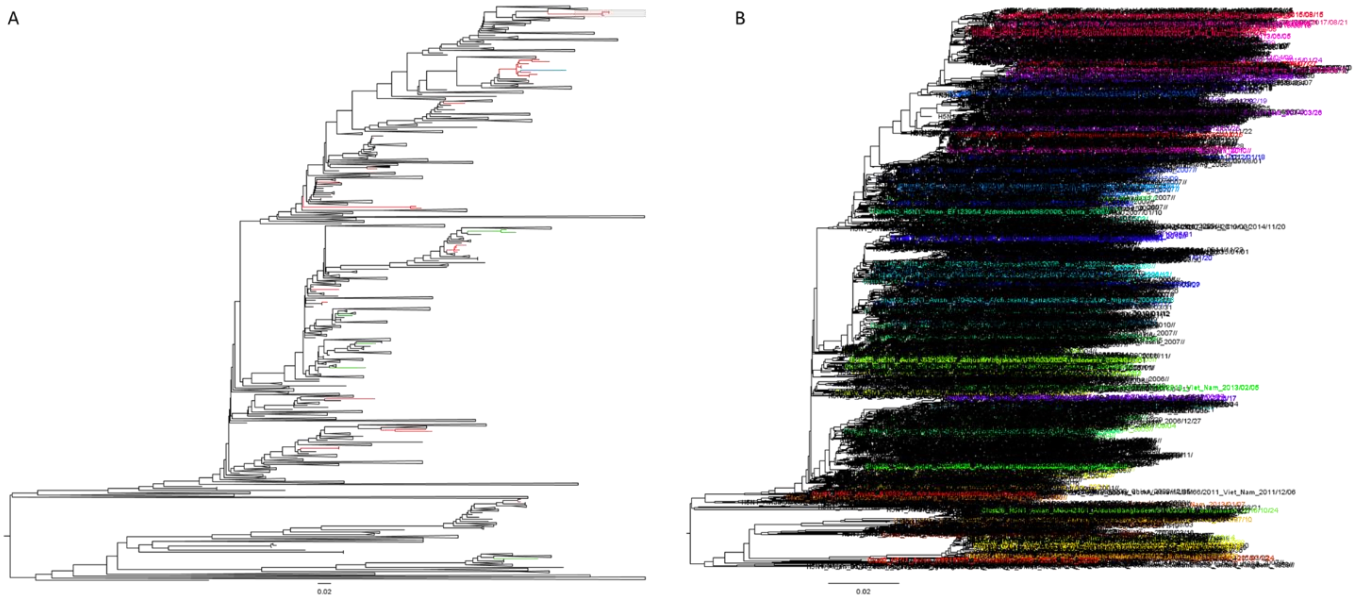

Figure 2.13 Cluster Analysis of H5N1

To determine how closely genetically related PB1-F2 sequences are, phylogenetic tree A (previously shown in Figure 2.8) is inputted to ClusterPicker with a FASTA alignment file to output annotated tree B, with clusters of highly related viruses colour coded. Non-clustered virus names remain black. Tree B is inputted to ClusterMatcher with an annotation file of truncation status to correlate phylogenetic relationship with a characteristic of interest. No clusters matching the parameters were identified, so no output trees were produced.

Cluster analysis suggested N-terminal truncations of PB1-F2 are not supported in H5N1 virus transmission. However, the small average cluster size and low prevalence of truncations may have influenced this analysis. Many of the clustered sequences were collected in the same place at the same time, so are likely to be samples from the same circulating virus. Of the 233 sequences identified by CP, only 39% were isolated from chicken hosts, compared to 74% of 666 H9N2 clustered sequences. The difference in host prevalence may be due to the reservoirs the subtypes circulate in, or it may be an indication of viability of truncated PB1-F2 sequences.

	H9N2	H7N9	H5N1
Number of sequences	1239	567	1823
Number of clusters	158	11	97
Average cluster size	3	10	2
Number of clusters 100% truncated	3	1	0
Number of clusters 81-99% truncated	0	1	0
Number of clusters 51-80% truncated	3	0	0
Number of clusters 11-50% truncated	4	0	0
Number of clusters <10% truncated	0	1	0

Table 2.7 Quantification of Cluster Analysis

The number of sequences and clusters for each subtype identified during ClusterPicker and ClusterMatcher analysis, showing proportion of truncated sequences within each cluster. Average cluster size calculated as median.

Cluster analysis of H7N9, H9N2 and H5N1 was quantified in Table 2.7, showing how few of the clusters identified by CP met the CM threshold of truncation prevalence. H9N2 and H7N9 AIV PB1-F2 genes both had more clusters with a mix of truncated and full length PB1-F2 sequences than 100% truncated sequences. Several of these clusters showed a branch split between sequences coding full length PB1-F2

and the appearance of N-terminal truncations, but we cannot rule out the possibility of some isolates restoring the full length ORF.

Although there are few fully truncated clusters, this analysis did show highly related segment 2 sequences maintaining a truncated PB1-F2. This indicates that the truncations are likely to be persisting during circulation rather than appearing as individual mutation events.

Cluster Analysis of Avian H5Nx Virus PB1-F2 sequences

	H5Nx
Number of sequences	1119
Number of clusters	144
Average cluster size	3
Number of clusters 100% truncated	0
Number of clusters 81-99% truncated	1
Number of clusters 51-80% truncated	1
Number of clusters 11-50% truncated	3
Number of clusters <10% truncated	2

Table 2.8 Quantification of H5Nx Cluster Analysis
The number of sequences and clusters identified during ClusterPicker and ClusterMatcher analysis, showing proportion of truncated sequences within each cluster. Average cluster size calculated as median.

The H5Nx dataset had multiple large clusters (Figure 2.14). The larger H5Nx dataset had more closely related sequences than the original H5N1 dataset (Table 2.7, Table 2.8). Unlike H5N1, CM identified two majority truncated clusters of highly related sequences, Clusters 37 and 40. Unlike H9N2 and H7N9 no H5Nx cluster was 100% truncated (Table 2.8). Although H5Nx had multiple NA lineages which would suggest looser phylogenetic relationships than the H5N1 dataset, the PB1 sequences appeared to be better maintained. Mixed clusters of truncated H5N6/H5N8 PB1-F2 sequences were not unexpected, as it

is well reported that the two subtypes can reassort and swap internal genes (Beerens et al., 2019; Chen et al., 2019; Guo et al., 2017; Yehia et al., 2018). This data gave weight to the suggestion that some H5Nx subtypes are more likely to maintain a truncated PB1-F2 sequence than H5N1.

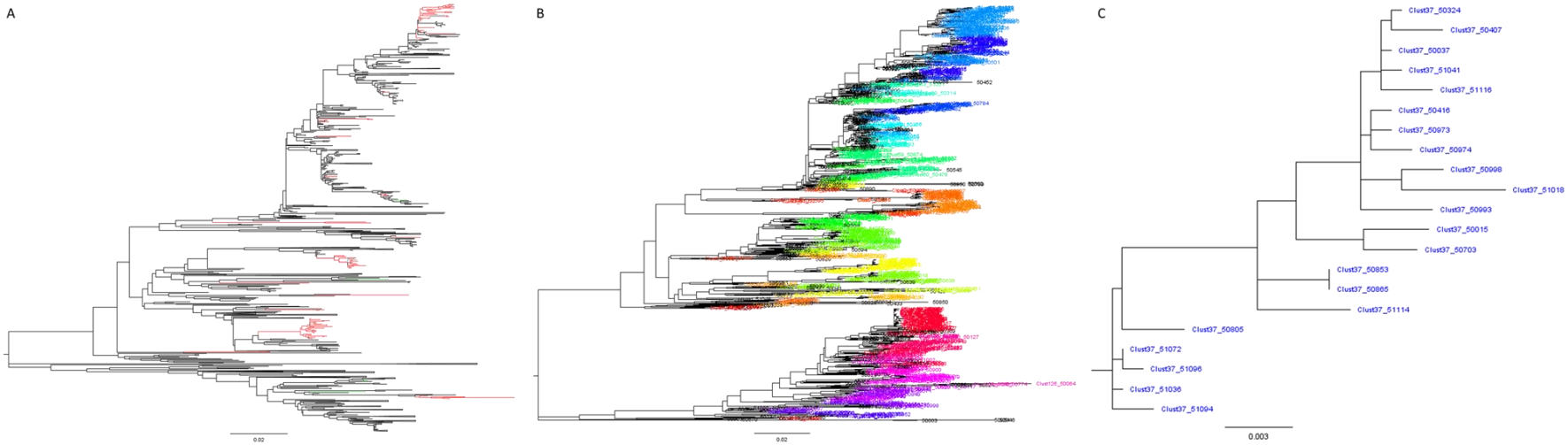

Figure 2.14 Cluster Analysis of H5Nx

To determine how closely genetically related PB1-F2 sequences are, phylogenetic tree A (previously shown in Figure 2.9) is inputted to ClusterPicker with a FASTA alignment file to output annotated tree B, with clusters of highly related viruses colour coded. Tree B is inputted to ClusterMatcher with an annotation file of truncation status to correlate phylogenetic relationship with a characteristic of interest. Highly truncated clusters are identified and outputted as separate trees such as C.

Phylogenetic Analysis of Highly Truncated PB1-F2 Gene Clusters

After identifying highly related clusters of viruses with truncated PB1-F2 ORFs, representative clusters from each subtype were chosen for further phylogeographical analysis. Length of circulation and geographical spread were considered, and attempts were made to determine where a truncation was first generated or identify a likely precursor virus.

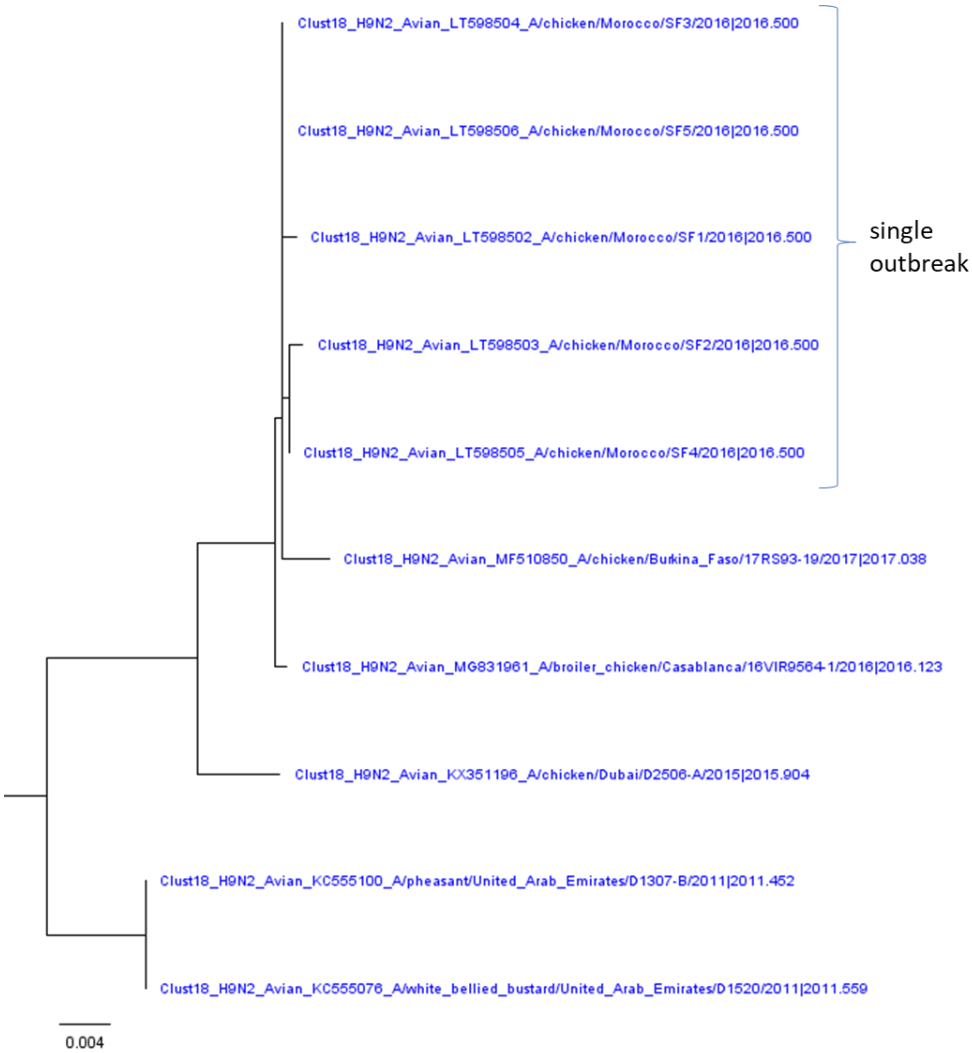

Figure 2.15 H9N2 Cluster 18 Phylogenetic Analysis

Cluster18 was identified from the H9N2 dataset using ClusterPicker and met the truncation threshold in ClusterMatcher. Newick tree outputted by ClusterMatcher. Sequences with the same collection date collected from

the same outbreak highlighted.

After CP and CM analysis, H9N2 Cluster 18 was identified as a set of 10 highly related truncated sequences. Cluster 18 was the largest H9N2 cluster, although cluster size had been inflated by 5 samples from the same outbreak (Figure 2.15). The sequences in this cluster were mostly from chicken hosts, suggesting the viruses are being spread through poultry movements. The earliest truncated sequences were from falconry birds in UAE in 2011, which then spread to northern Africa in 2015-2016. All sequences from chickens in Morocco are multiple samples from the same outbreak, which expands cluster size (Figure 2.15). While the A/chicken/Morocco/x viruses were all sampled at the same time, there was still a clear maintenance of the truncation through poultry movements within the Middle East and northern Africa. However, this did show transmission on a local scale as well as (most likely) through human-mediated transport of birds.

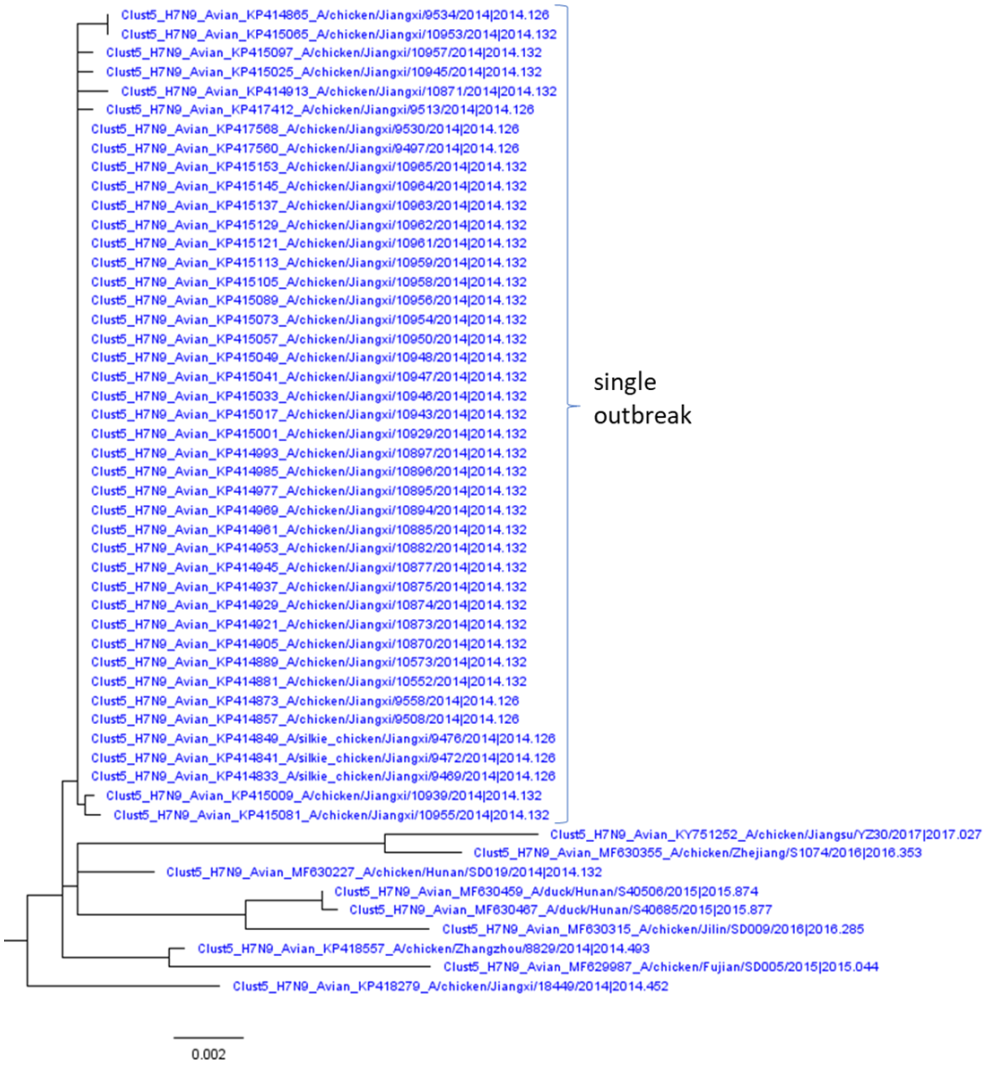

Figure 2.16 H7N9 Cluster 5 Phylogenetic Analysis

Cluster 5 was identified from the H7N9 dataset using ClusterPicker, and met the truncation threshold in ClusterMatcher. Newick tree outputted by ClusterMatcher. Sequences collected during a single outbreak highlighted.

Cluster 5 contained 52 sequences, of which 42 were from the Jiangxi February 2014 outbreak (Figure 2.16). The majority of hosts were chickens, with only two domestic duck hosts from 2015. Although this H7N9 Cluster was also inflated by numerous samples from the same outbreak in February 2014, there was also transmission until 2017. Cluster 5 was limited to China, but it had spread between provinces over several years (Figure 2.16).

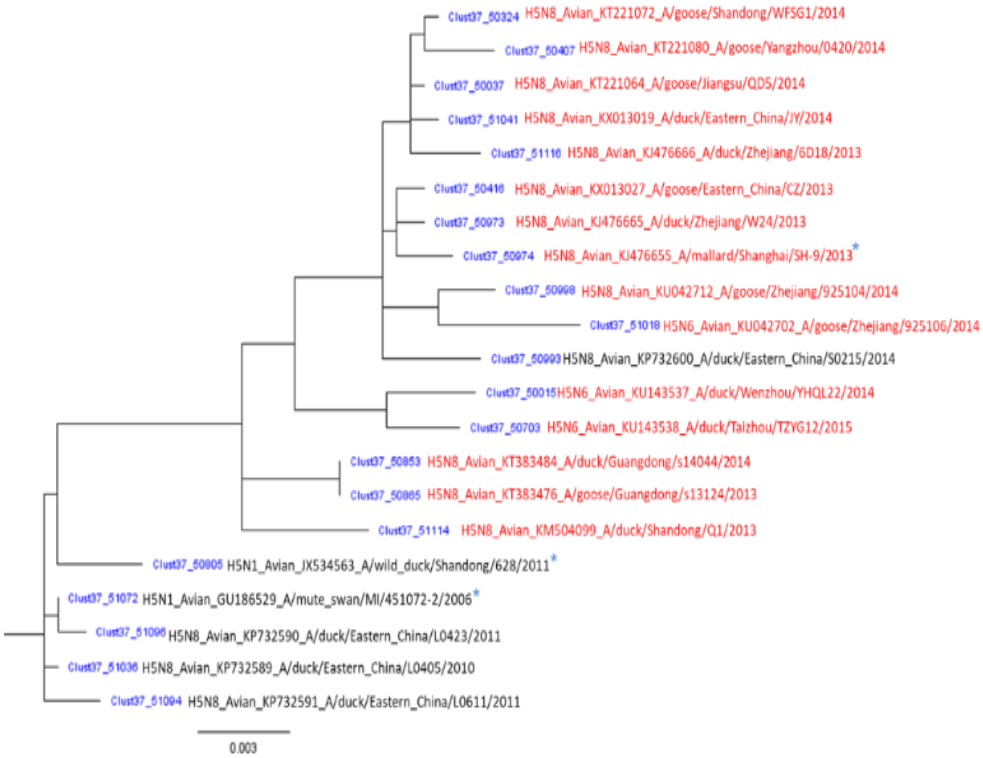

Figure 2.17 H5Nx Cluster 37 Phylogenetic Analysis.
Cluster 37 was identified from the H5Nx dataset using ClusterPicker, and met the truncation threshold in ClusterMatcher. Newick tree outputted by ClusterMatcher. Names in red indicate truncated PB1-F2 sequences, and those marked with a blue asterisk were isolated from a wild bird host.

Unlike Clusters in H9N2 and H7N9 datasets, none of the highly truncated clusters identified in the H5Nx dataset were fully comprised of viruses with truncated PB1-F2 ORFs. Cluster 37 (Figure 2.17) had 71% truncated sequences. The majority of virus isolates (86%) were collected from domestic Anseriformes in China over 2006-2015. This was consistent with the hypothesis that N-terminal truncations are better maintained in domestic birds than in wild populations. Truncations appeared in 2013 and were maintained into 2014/15, through geographical and temporal progress and reassortment events, as there were two instances where the subtype of truncated sequences changes from H5N8 to H5N6. Although all the virus isolates were in China, the viruses had moved between provinces on several occasions. Unfortunately, many samples were listed as having

been sampled only from “Eastern China”, so it was not possible to give an estimate of distance travelled.

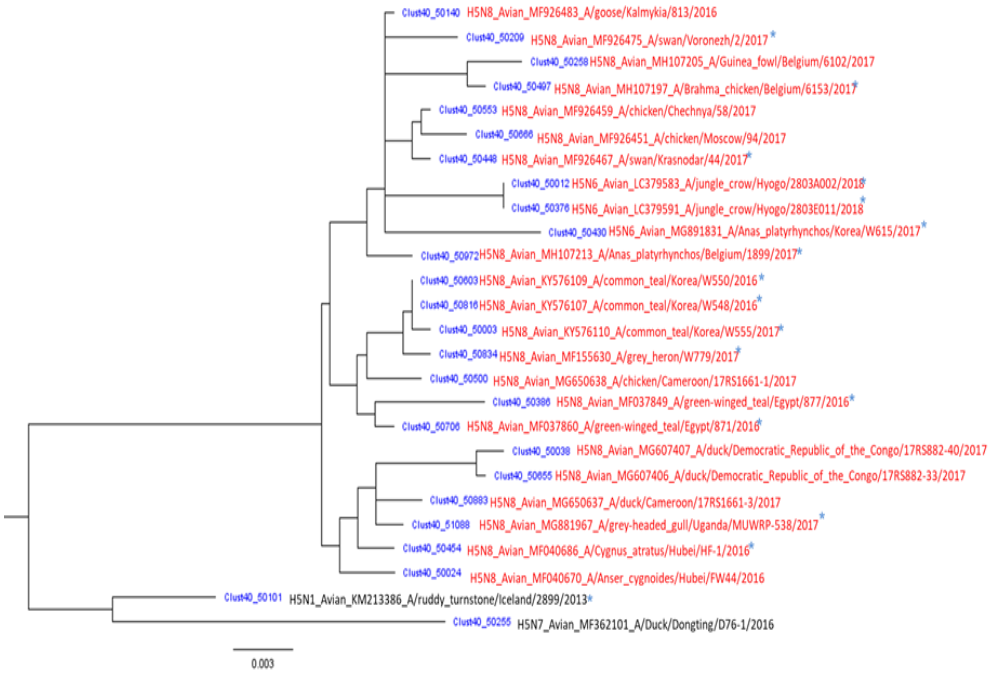

Figure 2.18 H5Nx Cluster 40 Phylogenetic Analysis

Cluster 40 was identified from the H5Nx dataset using ClusterPicker, and met the truncation threshold in ClusterMatcher. Newick tree outputted by ClusterMatcher. Names in red indicate truncated PB1-F2 sequences, and those marked with a blue asterisk were isolated from a wild bird host.

Cluster 40 (Figure 2.18) was the most heavily truncated cluster identified in the H5Nx dataset, reaching 92% of isolates without an intact PB1-F2 ORF, and from a majority (62%) of wild bird hosts. In contrast to the other clusters described, Cluster 40 covered three continents, moving from Europe to Africa and Asia, covering a broad estimate of approximately 30,000-45,000km. The estimate was so broad because it is not always possible to distinguish whether an introduction is a result of internal circulation within a country spilling over to a close neighbour, or an introduction from a non-neighbouring country through migratory birds or domestic birds between. Cluster 40 was mostly comprised of sequences from viruses isolated from wild

Anseriformes, although there were some from domestic Galliformes hosts as well. Similarly, to Cluster 37, the PB1-F2 truncation had been introduced and maintained through time and reassortment events. STOP12 was the only truncation present in this cluster.

From the cluster analysis, we concluded that the N-terminally truncated PB1-F2 sequences were not only present, but persistent in avian isolates. We have shown clusters of highly related viruses with a truncated PB1-F2 gene from several subtypes can circulate for several years, and those viruses can be spread during poultry transports in a similar manner to isolates with a full length PB1-F2 ORF.

Prevalence of N-terminally Truncated PB1-F2 Sequences in Different Avian Hosts

Composition of Endemic Avian Viruses Datasets

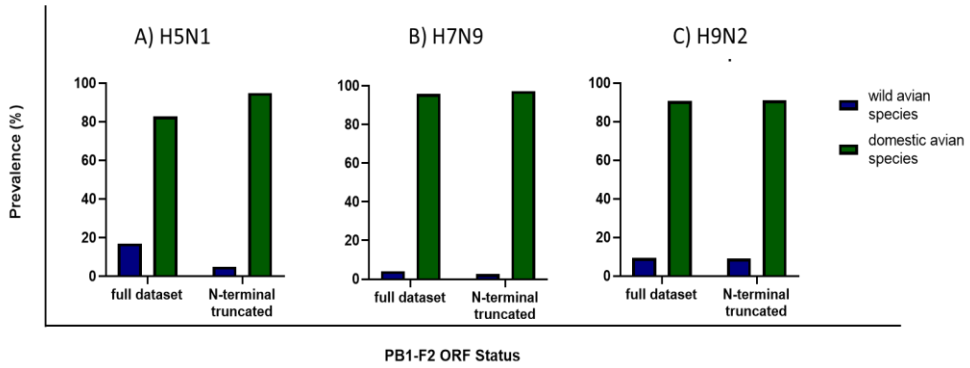

Figure 2.19 Prevalence of Truncated PB1-F2 ORFs from H5N1, H7N9, and H9N2 Subtype Viruses Isolated from Wild or Domestic Avian Hosts.
The prevalence of wild and domestic hosts within full datasets of H5N1 (A), H7N9 (B), and H9N2 (C) datasets and the truncated subsets were calculated for each subtype.

The dataset of H7N9, H5N1, and H9N2 subtypes was not suitable for investigating differences between wild and domestic birds, as most of the isolates in the dataset had been collected from domestic birds. Virus sequences in the full datasets for all three subtypes were from >80% domestic hosts (Figure 2.19) and the subset with truncated PB1-

F2 ORFs ranged from 91 – 98% domestic hosts origin (H7N9 97.40%, H5N1 95.00%, H9N2 91.06%).

Composition of Mixed Avian Host Dataset

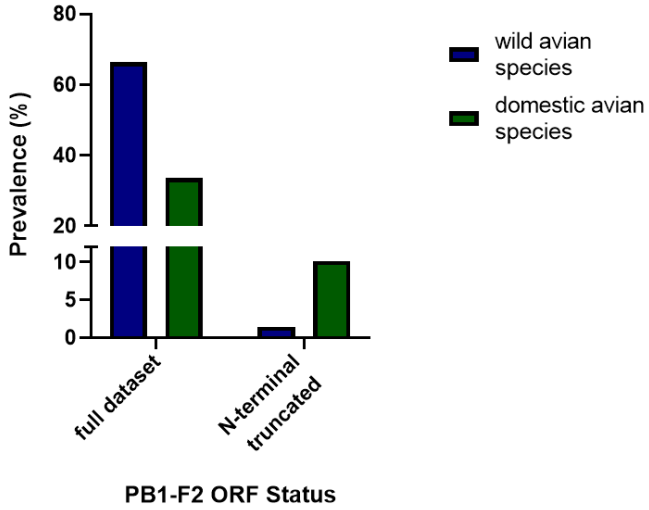

Figure 2.20 Prevalence of Truncated PB1-F2 ORFs from Mixed Subtype Viruses Isolated from Wild or Domestic Avian Hosts.
Avian host type (wild or domesticated), and PB1-F2 ORF status (N-terminal truncated sequence subset or all sequences) was determined for each sequence. The percentage prevalence of each host category of the full avian dataset of 5,940 sequences, and the truncated sequences subset were calculated.

We then analysed host composition of the avian subset of the 10279 sequence dataset provided by Dr Samantha Lycett (Lycett et al., 2019). Stratified sampling reduced bias toward repeated sampling of the same AIV outbreak, by limiting the number of sequences collected to one or two per host-type, subtype, country or state, and year. (Figure 2.20). The dataset had a total sequence number of 5,940; of these, wild hosts accounted for 3,941 (66.35%) and domestic hosts 1,999 (33.65%). With this change in dataset composition, we could have increased confidence that any changes in N-terminal PB1-F2 truncation prevalence between wild and domestic hosts were less likely to be due to host sampling bias.

Prevalence of N-terminal Truncations within the Mixed Avian Host Dataset HA Subtypes

The N-terminal truncated PB1-F2 sequences were considered as a subset of the data. Within each host category, prevalence of N-terminal truncations was calculated. 1.45% of sequences isolated from wild birds had a truncated PB1-F2, while 10.11% of sequences from domestic hosts. These data showed that the majority of truncated PB1-F2 sequences were isolated from domestic birds (Figure 2.20). Although most domestic bird hosts in the dataset were chickens, ducks and geese are also farmed. It is possible that differences in the transmission environment between poultry in close quarters and more distant wild birds could affect selection pressure for either full length or truncated PB1-F2 genes.

HA subtype	Number of sequences	% truncated	% truncated maintain C-terminus
H1	342	4.39	35.71
H2	254	0.79	100.00
H3	676	1.18	62.50
H4	621	0.81	60.00
H5	1119	8.04	65.56
H6	682	8.80	98.28
H7	624	3.52	100.00
H8	66	0.00	0.00
H9	537	8.75	100.00
H10	394	2.03	100.00
H11	329	0.91	100.00
H12	119	0.00	0.00
H13	86	2.33	100.00
H14	19	0.00	0.00
H15	9	0.00	0.00
H16	63	3.17	100.00

Table 2.9 PB1-F2 ORF Maintenance Within Avian Host Dataset
Number of sequences in each HA subtype of the 5940 avian host dataset, showing % truncated sequences, and of those truncated, % maintaining C-terminal ORF.

The 5940 avian sequence dataset was separated into individual HA subtypes (Table 2.9), and the prevalence of PB1-F2 truncations calculated. There was a broad range from 0.00 – 8.80% prevalence of N-terminal truncations between the HA subtypes. H8, H12, H14, and H15 subtypes all had low numbers of sequences (66, 119, 19, and 9 respectively), with 0% containing an N-terminal truncation. H5, H6, and H9 showed the highest prevalence of truncations, with all these HA subtypes having >8% truncated PB1-F2 sequences. These latter subtypes are endemic in many areas of the world including China, south east Asia, and the Middle East, and circulate in domestic poultry (C. M. Bui et al., 2017; Hu et al., 2020; Kandeil et al., 2019; Naguib et al., 2019; Peacock et al., 2019).

This initial analysis suggested that host type may affect prevalence of truncated PB1-F2 sequences, as the highest prevalences were in domestic hosts and HA subtypes that are known to circulate in domestic poultry.

Host Order Prevalence in Mixed Avian Host Dataset

Order	Number sequences	Percentage avian dataset
Galliformes	1041	17.53
Anseriformes	4021	67.70
Charadriiformes	536	9.02
Other	342	5.76

Table 2.10 Host Order Prevalence in Mixed Subtype Avian Dataset
Avian influenza A virus segment 2 sequences were annotated with host Order. The percentage of virus sequences derived from Anseriforme, Galliforme, Charadriiforme and Other host orders were calculated within the full avian dataset of 5,940 sequences. “Other” denotes sequences with either a lack of host information, or a host species outside of the three orders highlighted.

After seeing the variation in PB1-F2 truncation prevalence between HA subtypes, with higher frequencies in subtypes circulating in poultry, we considered whether host species had an impact on the likelihood of gene truncation. The mixed subtype avian dataset of 5,940 sequences was separated into host Order (Table 2.10). The dataset had a bias towards Anseriformes, which make up the host of origin for 67.70% of sequences. Galliforme hosts were 17.53%, Charadriiformes 9.02%, and the remaining 5.76% of sequences from “Other” host species could not be confidently assigned to an Order. The high prevalence of Anseriformes and Charadriiformes was consistent with the majority wild bird source already reported, however not all Anseriformes are wild i.e., domestic ducks.

Order	Number N-terminally truncated sequences (N=259)	Percentage truncated dataset (N=259)	Percentage truncated domestic host (N=202)	Percentage truncated wild host (N=57)
Galliformes	115	44.40	56.93	-
Anseriformes	116	44.79	43.07	50.88
Charadriiformes	11	4.25	-	19.30
Other	17	6.56	-	29.82

Table 2.11 Host Order Prevalence in Truncated PB1-F2 Sequences from the Mixed Subtype Avian Dataset

Host type (wild or domestic), and Order (Anseriformes, Galliformes, Charadriiformes, Other) was determined for each sequence in the avian host truncated subset. The percentage of each order was calculated within the host subsets.

The N-terminally truncated subset of avian PB1-F2 sequences was also separated by host Order (Table 2.11). Unlike in Table 2.10, the truncated subset had a nearly even proportion of Anseriformes and Galliformes (44.79% and 44.40% respectively) (Table 2.11). The difference in host order composition between the full dataset and the N-terminally truncated PB1-F2 gene subset indicated that expression of full length PB1-F2 may be advantageous in Anseriformes, and possibly Charadriiformes, although truncations are common in Galliformes.

The N-terminally truncated subset of sequences was then separated into domestic avian and wild avian host subsets to investigate whether environment or Order had a stronger effect on PB1-F2 gene truncation prevalence (Table 2.11). Within domestic hosts, Anseriformes made up 43.07% of the dataset and Galliformes 56.93% (Table 2.11). This was a more even split than expected, compared to the full dataset shown in Table 2.10, however the global domestic duck industry produces millions of metric tonnes (MT) of meat annually, with China producing 2.801 million MT (64.7% of global production) in 2014 (Batres-Marquez, 2017), so it is reasonable for Anseriformes to be well represented in this set of sequences.
There were no Galliformes samples in the wild avian subset, but Charadriiformes increased from 9.02% of the full dataset (Table 2.10) to 19.30% of wild host truncated sequences (Table 2.11). The Other category also increased from 5.76% to 29.82% of wild truncated PB1-F2 sequences (Table 2.11).

Between HA subtype, host Order, and environment, there was a consistent higher prevalence of N-terminal truncations in PB1-F2 from domesticated poultry. The highest prevalence was with HA subtypes that circulate in poultry, and Galliformes in general. Full length PB1-F2 may be more advantageous for transmission between wild birds, with less close contact than in domesticated birds.

Discussion

In this chapter, evolutionary evidence for the importance of N-terminally truncated PB1-F2 sequences has been presented. N-terminal truncations were present in multiple lineages and have been shown to be maintained in circulating populations of some AIV subtypes (Figures 2.11 – 2.14). Isolates from Galliformes or domestic hosts have a higher prevalence of truncated sequences than wild birds (Figure 2.20, Table 2.11), suggesting a possible link between N-

terminal PB1-F2 truncations and transmission of AIV. PB1-F2 is known to target different host immune modulators, depending on host and viral strain, (James et al., 2019; Leymarie et al., 2014), and has been shown to affect the length of the transmission window in infected chickens (James et al., 2016). A study of white leghorn chickens infected with A/turkey/Turkey/1/2005 (H5N1) found that the probability of infection increased with higher inoculation doses. A higher inoculation dose was also associated with a shorter latent period and earlier shedding (Spekreijse et al., 2011). It is possible that in the more susceptible chicken host, in the close quarters domestic birds live in, expression of a C-terminal fragment of PB1-F2 is sufficient for immune modulation, while the less confined wild birds require the advantage of a full-length PB1-F2 to aid efficient transmission.

Transmission between poultry farms can be mediated through human-mediated transport, infected workers, or shared equipment (Fusaro et al., 2016), whereas transmission between wild birds is dependent on bird contact or contaminated water (Fourment & Holmes, 2015). It should also be noted that flock dynamics can extend the infectious period beyond that of a single bird. The mean infectious period of an LPAI-infected bird is 4-8 days, whereas a chicken flock can remain infectious for up to 2 months (Bergervoet et al., 2019; Gonzales et al., 2014).

There may be additional benefits to maintaining the full length PB1-F2 ORF through functions mediated by the N-terminal region. It has been reported that the N-terminus (aa1-50) of PB1-F2 can stabilise PB1 early in infection, and increase viral RNA polymerase activity (Košík et al., 2011). Also isolated to the N-terminus of PB1-F2 are two protein kinase C (PKC) phosphorylation sites, S35 and T27. Removal of these targets by mutagenesis reduced caspase 3 activity in primary human monocytes (Mitzner et al., 2009). Increased expression of viral proteins and induction of apoptosis in immune cells are both beneficial

during IAV infection (Gao et al., 2015; Pinto et al., 2020; Suttie et al., 2019).

While it has been shown that lack of segment 2 AUG 4 will increase expression of C-terminal fragments of PB1-F2 compared to an N-terminal truncation, the balance of expression between AUG 4 and AUGs 7-9 is still unclear (Kamal et al., 2015). The translational mechanism leading to expression of C-terminal PB1-F2 fragments is currently unknown. The potential stability of C-terminal fragments is also uncertain. Full length PB1-F2 is known to be unstable, with several residues identified in the C-terminus being able to alter stability (James et al., 2016; Košík et al., 2015). There is currently no published comparison of the stability of full length PB1-F2 compared to its fragments.

The total prevalence of N-terminal PB1-F2 truncations across multiple host species and viral subtypes was of interest as avian and mammalian PB1-F2 sequences have been reported to be under very different selective pressures (Schmolke et al., 2011). Although N-terminal gene truncations were present within swine and human sequences, they were rarely accompanied by maintenance of the C-terminus or AUGs 7-9 (Table 2.3). This indicates that in mammalian hosts, N-terminal truncations are merely part of an ongoing genetic drift toward removing PB1-F2 from the IAV proteome. The mammalian PB1-F2 ORF may be in the process of downgrading from a functional accessory protein to a defective ribosomal product (DRiP) produced through non-canonical mistranslation (Pinto et al., 2020; J. Wei & Yewdell, 2018). Degradation of the PB1-F2 ORF is unlikely to impact PB1 or PB1-N40 in their sequence, as negative selection of the PB1 ORF constrains PB1-F2 (Zell et al., 2007). The balance of expression between AUGs 4 (PB1-F2) and 5 (PB1-N40) in segment 2 is likely to be maintained as long as AUG 4 remains present in the sequence. The N-terminus of PB1-F2 could function as a short upstream ORF in the same manner as sORF2 expressed from AUG 3 (Wise et al.,

2011). sORF2 is only two codons long, so translation is almost immediately terminated. Due to the time taken to reassemble a pre-initiation complex, the 40S subunit scans past AUG 4 to reinitiate translation at AUG 5. In this way, translational initiation at AUG 2 can help balance expression of PB1-F2 and PB1-N40 (Wise et al., 2011).

The presence of N-terminal truncations in AIV sequences was elucidated through bioinformatic analysis of sequence datasets obtained from the NCBI IVR and curated, collated and annotated with a custom R code, by Dr Samantha Lycett. Having shown PB1-F2 gene truncations were present in viruses with a variety of HA subtypes, H5N1, H7N9, and H9N2 lineages were selected as subtypes of interest for further investigation. All three AIV subtypes are burdens to the global poultry industry and considered to be zoonotic pandemic risks.

Although there was a variety in prevalence and position of N-terminal PB1-F2 truncations, all three subtypes of interest were found to maintain the C-terminal ORF of PB1-F2 (Figure 2.8). Functional motifs of PB1-F2 are concentrated in the C-terminal sequence (Conenello et al., 2007; Košík et al., 2015; Yamada et al., 2004), so this indicated N-terminal truncations might not prevent immune antagonism or mitochondrial targeting of the protein. However, the N-terminus of the protein may be important *in vivo* as a STOP12 mutation shortened the viral transmission window in chickens (James et al., 2016). In addition to maintenance of the C-terminal ORF, we found a high prevalence of AUGs 7-9 within truncated sequence subsets, indicating a strong possibility that C-terminal fragments are able to be expressed from truncated sequences (Figure 2.7).

When avian H5N1 viruses were found to have a much lower prevalence of PB1-F2 truncations than the wider mixed host H5 dataset, analysis was widened to avian H5Nx sequences to determine if this was a result of truncations being less common in an avian host,

or a H5N1 specific factor. Within H5Nx viruses, there was evidence that some H5 subtypes were maintaining N-terminally truncated PB1-F2 sequences during circulation, along with conservation of AUGs 7-9 (Table 2.8).

H5Nx clusters were heavily made up of sequences that were collected from Anseriformes hosts, whereas H9N2 and H7N9 isolates are almost all from Galliformes. Galliformes are generally more susceptible to severe disease and have higher mortality rates following HPAI infection than Anseriformes. This may be influencing the composition of the clusters, particularly for H5Nx. However, any domestic outbreak of HPAI in the UK or EU will result in culling of birds, regardless of their Order.

Phylogenetic relationships between truncated sequences were examined using ClusterPicker and ClusterMatcher programmes to cluster highly related sequences and correlate clusters with level of PB1-F2 ORF truncation. Setting aside truncation status, the initial CP run identified smaller clusters for H5N1 than either H7N9 or H9N2 (Table 2.7). The phylogenetic relationships between the sequences in the dataset will have affected the likelihood of truncated clusters. Clusters with a high proportion of truncated PB1-F2 genes were identified in both H9N2 and H7N9 datasets, but none were identified in the H5N1 sequences. This suggested a lack of support for transmission of H5N1 viruses with a truncated PB1-F2. The H5Nx dataset did have some clusters of mixed truncated and full length PB1-F2 sequences, but the sequences were likely collected during single outbreaks and sampled the same circulating strain. H9N2 and H7N9 both had clusters of viruses with truncated PB1-F2 genes that travelled long distances or persisted for several years.

The greater number of clusters of truncated sequences identified in H9N2 than H7N9 and the high likelihood of reassortment between subtypes (Lam et al., 2015), raises the possibility that N-terminal

truncations of PB1-F2 are novel only in H9N2 viruses and are simply being maintained in H7N9 strains when PB1 segments are obtained. However, the truncated PB1-F2 sequences are still circulating in H7N9 subtypes, so are unlikely to be detrimental to the virus' successful transmission.

There were more sequences within the H7N9 clusters, but this is likely due to multiple samples of the same virus outbreak inflating cluster size. The greater range of clusters identified in H9N2 indicates a larger circulating population of viruses coding truncated PB1-F2s. Looking at the majority truncated clades for each subtype (H7N9 Clusters 4 & 5, H9N2 Clusters 18, 130, 156, 57, 117, & 62) gives a dataset of 73 H7N9 sequences and 35 H9N2 sequences. This is somewhat mismatched, however when we consider that 43 of the H7N9 sequences are from the same outbreak, the subtype numbers are more comparable. Within these clusters, H7N9 was only circulating within China, whereas H9N2 Cluster 117 moved into Japan and Cluster 18 was intercontinental (Table 2.7). H7N9 truncated PB1-F2 sequences were travelling approximately 2,830km in Cluster 4 and 6,203km in Cluster 5. Half of the H9N2 clusters examined here were restricted to a single province of China (C130, C57, C62). Cluster 18 and 117 however, travelled approximately 8,413km and 4,197km respectively (data not shown). Truncated PB1-F2 sequences are not geographically restricted, and appear to be transported during human poultry movements in a similar manner to viruses encoding the full length gene.

Due to the mostly domestic host composition of the H5N1/H7N9/H9N2 datasets (Figure 2.19), the mixed avian dataset of 5,940 sequences was used to investigate how host Order and environment affected N-terminal PB1-F2 truncation prevalence. There is a high level of variation in the level of host information available in the NCBI IVR database, so analysis was restricted to Anseriformes, Galliformes, and Charadriiformes. A proportion of the dataset had insufficient host data to be classified further than 'Other' (Table 2.10). Many species of wild

birds are likely to be infected yearly either at, or shortly after the return from, breeding grounds (Maxted et al., 2016; Verhagen et al., 2015). Autumn migrations disperse IAVs, with increased viral reassortment and interspecies transmission in winter compared to more local circulation over summer months (Hill et al., 2016). Domestic birds are likely to infected by interaction with wild birds on migration (Elbers & Gonzales, 2020; Lu et al., 2017), but will not have memory of previous infection so will be immunologically vulnerable.

Many of the domestic sequences that are truncated are isolated from domestic Anseriformes, raising a question of whether host environment, such as water availability, is more important than host species for N-terminal truncations. Differences in proximity/range between wild and farmed birds may also affect transmission rate. Poultry on free-range farms are at higher risk of AIV transmission from wild birds than those in less open farming environments. Visits by dabbling ducks to pools can be a source of infection through indirect contact with wild birds and poultry (Elbers & Gonzales, 2020). Wild birds are also unlikely to be receiving veterinary care or under biosecurity surveillance.

The truncations were more prevalent in domestic hosts. Although Anseriformes made up ~40% more of the dataset than Galliformes, within the truncated subset of sequences, the two orders were almost equal (Table 2.11). The enrichment of truncated sequences within the domestic hosts indicates that loss of the full length PB1-F2 ORF is more tolerable in a more IAV susceptible host, or may have less effect in birds contained in a domestic setting.

Overall, the data described in this chapter shows the evolving ORF structure of PB1-F2 within AIV subtypes of economic importance and public health concern. Confirmation of C-terminal expression and functional studies of N-terminally truncated PB1-F2 proteins will be explored in the following chapters.

Chapter 3: Expression and Stability of PB1-F2

Chapter 3.1 Expression Within the PB1-F2 Open Reading Frame

Introduction

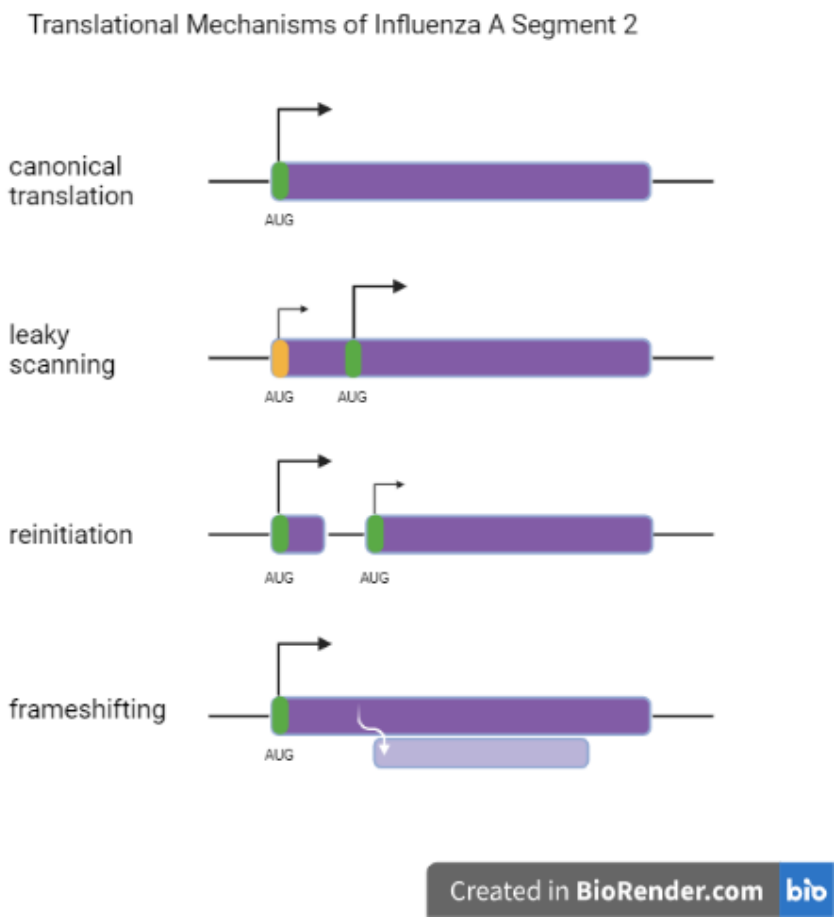

Figure 3.1.1 Translational Mechanisms of Influenza A Virus Segments
Purple boxes indicate ORFs, with coloured bars for strength of Kozak sequencing around AUG codons (green for strong, orange for weak). Black arrows indicate initiation of protein synthesis, and the white arrow indicates trans-frame movement of the ribosome during frameshifting.

PB1, the essential protein product of segment 2, is expressed through canonical translation initiation (Figure 3.1.1). A pre-initiation complex forms of the 40S ribosomal subunit, eukaryotic initiation factors (eIFs), and a $tRNA^{Met}$. Using secondary structures in the 5' UTR and Kozak consensus recognition, the preinitiation complex dissociates and the

60S ribosome subunit binds to begin protein synthesis (Jackson et al., 2010).

IAVs utilise multiple translation mechanisms from overlapping reading frames to maximise the proteome available from the small 13kb RNA genome (Figure 3.1.1) (Firth & Brierley, 2012; Pinto et al., 2020; Vasin et al., 2014). Non-canonical translation mechanisms used in segment 2 include leaky ribosomal scanning and translation termination-reinitiation (Wise et al., 2009, 2011).

Leaky ribosomal scanning is an important mechanism of proteome expansion, where variations in the Kozak consensus allows manipulation of ribosomal recognition of AUG codons to express overlapping reading frames (Kozak, 1984). In IAV segment 2, leaky scanning occurs when the preinitiation complex skips over the weaker Kozak consensus sequences of the first three AUGs in the segment in favour of binding AUGs 4 or 5 that have a stronger consensus (Wise et al., 2011). The most efficient translation initiation motif is CACCAUGG, where the -3/+4 positions have the greatest impact. Initiating AUGs have a strong preference for a purine in the -3 position (79% A, 18% G), while AUG codons upstream of the initiation site have a preference for a pyrimidine at -3 (Kozak, 1981, 1984). In IAV segment 2, both AUG 1 and AUG 2 have a pyrimidine at -3, increasing the likelihood of leaky scanning, although AUG 1 has a G at +4 and is more likely to be expressed by virtue of being the first AUG codon in the segment (Kozak, 1981). AUGs 3 and 5 both have an A in the -3 position while AUG 4 has a G. This minor difference in Kozak consensus may be involved in the balance of expression between PB1-N40 and PB1-F2.

Reinitiation is defined as the ribosome reaching a stop codon, halting translation, and then recommencing translation from the same mRNA in the same reading frame further along the sequence to produce additional protein products (Figure 3.1.1) (Kozak, 2001). The process

is most efficient with a short upstream ORF of less than 13 codons, due to loss of protein factors required for reinitiation (Hronová et al., 2017; Kozak, 2001). In viral systems, the ORFs typically overlap in different reading frames (Powell et al., 2008). In IAV segment 2, reinitiation is used to balance the expression of PB1-F2 and PB1-N40. Leaky scanning allows initiation at AUG 3 in frame 2. AUG 3 initiation is immediately followed by termination and continued scanning of the 40S subunit. Due to proximity (34 bases from the AUG 3 ORF stop codon to AUG 4), the 40S subunit scans past AUG 4 before re-acquiring a full pre-initiation complex, leading to expression of PB1-N40 from AUG 5 in frame 1 (Kozak, 1984; Wise et al., 2011).

PB1-F2 is expressed from AUG 4 in a +1 reading frame by leaky scanning (Chen et al., 2001) (Figure 3.1.1). There have also been reports of expression from internal PB1-F2 AUGs 7-9 producing C-terminal fragments of the protein. C-terminal expression of PB1-F2 was first identified in 2006, although multiple molecular species in western blots had been noted previously (Chen et al., 2001; Zamarin et al., 2006). Zamarin et al., found two major species of PR8 PB1-F2 in the lysates of infected cells, and were able to remove the faster migrating species by mutating all three C-terminal AUGs 7-9 to threonines, but they did not attempt to elucidate which of AUGs 7-9 was initiating translation. Although they determined via immunoprecipitation that the C-terminal fragments were able to interact with the full length PB1-F2, minimal phenotypic investigation of the C-terminus was performed (Zamarin et al., 2006). Kamal et al. noted an increase in N-terminal truncations of avian IAV PB1-F2 proteins since 2005, resulting in 24 or 25aa PB1-F2 proteins in H5N1 strains. Mimicking these truncations by introducing two STOP codons at positions 25 and 26 in two avian H5N1 PB1-F2 genes (A/chicken/Vietnam/NCVD-281/2009 and A/chicken/Vietnam/NCVD-296/2009) had no effect on apoptosis, and did not increase IFNβ expression during transient transfection, but in confocal microscopy,

the N-terminally truncated proteins were not visible. However, PB1-F2 was still detected from a ΔAUG 4 mutant, suggesting possible expression from C-terminal AUGs (Kamal et al., 2015).

There is a clear gap in our understanding of IAV segment 2 protein expression. To our knowledge, there has not been a focused attempt to match N-terminally truncated PB1-F2 peptides to a specific C-terminal AUG codon of PB1-F2. There is also little known about possible phenotypic impacts of C-terminal fragment expression, despite common use of N-terminal truncation (James et al., 2016; Mazel-Sanchez et al., 2018; McAuley et al., 2017) or ΔAUG mutations as a PB1-F2 knock out strategy (Leymarie et al., 2014; Schmolke et al., 2011; Tauber et al., 2012; Wise et al., 2009). If N-terminal truncation allows for expression of functional C-terminal fragments, as originally suggested by Zamarin and colleagues (Zamarin et al., 2006) then many 'PB1-F2 knockouts' are not actually knock outs. The aim of this chapter was to investigate expression of C-terminal fragments, test what effect N-terminal truncations have on C-terminus expression, and determine their functionality compared to full length PB1-F2.

Results

Expression of WT PB1-F2 Proteins

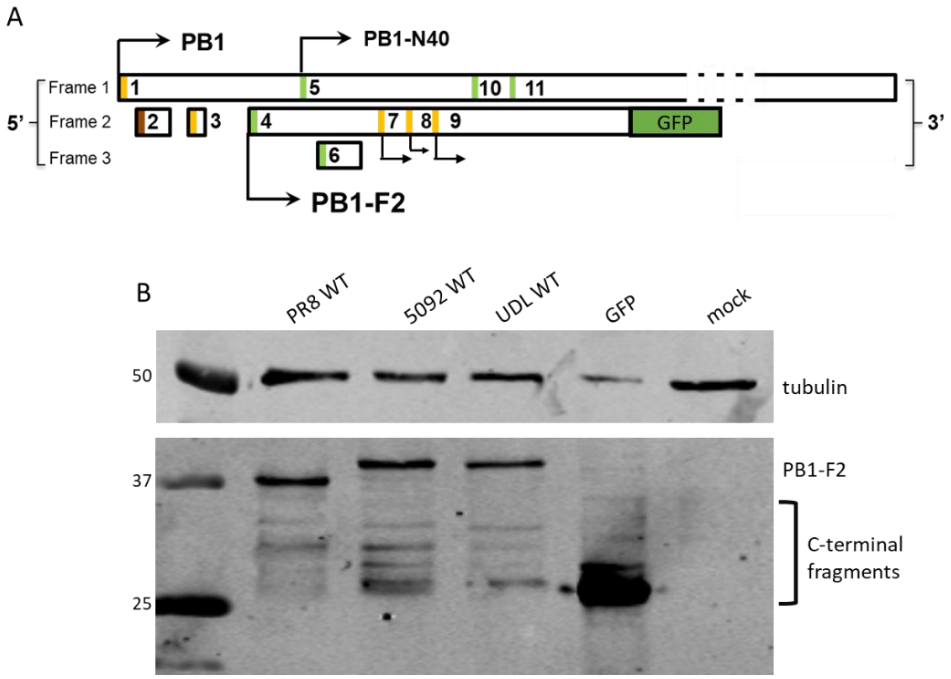

Figure 3.1.2 Expression of WT PB1-F2 Proteins

Panel A; schematic diagram of GFP tagged constructs used, in the context of the 5' ORF structure of IAV segment 2. Black boxes indicate ORFs, with coloured bars for AUG codons. The strength of Kozak signalling is indicated by traffic light colours, and black arrows indicate initiation of protein expression. Adapted from (Wise et al., 2009).

Panel B; GFP-tagged PB1-F2 plasmids were transfected into 293T cells and 48h later, cell lysates were run on a 4-20% gradient SDS-PAGE gel before transfer to nitrocellulose. Anti-β-tubulin was used a loading control (top panel), and anti-GFP antibody used to probe for PB1-F2 proteins (lower panel). 7µl PB1-F2-transfected cell lysates were run, with 3µl GFP-transfected sample run to avoid oversaturation of the GFP band.

Many studies have reported that PB1-F2 is better conserved in avian viruses than in mammalian, with 93% of PB1-F2 sequences from avian hosts coding ≥87 amino acids compared to only 43% of human host sequences (James et al., 2016). This is also supported by the

bioinformatic analysis performed in Chapter 2. Our hypothesis therefore was that C-terminal fragments were likely to be expressed and have a phenotype in avian subtypes. To begin answering this question, we first had to test whether PB1-F2 fragments were expressed from our model viruses A/turkey/England/50-92/1991 (H5N1) 5092 and A/chicken/Pakistan/UDL-01/2008 (H9N2) UDL. There is a lack of broadly binding antibodies for PB1-F2 because the protein has high sequence variability, therefore all expression assays used tagged proteins to allow comparison between virus strains (Figure 3.1.2 A). 293T cells were transfected with GFP tagged PB1-F2 plasmids, and the lysates visualised by western blotting for GFP and tubulin. The latter loading control confirmed roughly equal loading of the cell lysates, while an untransfected control ("mock") showed no reactivity with anti-GFP serum (Figure 3.1.2 B). Lysate from cells transfected with a plasmid encoding only GFP produced large amounts of an ~ 26 kDa product (the expected size for GFP) as well as trace amounts of higher molecular weight products. The three plasmids encoding fusions of IAV PB1-F2 to GFP all produced readily detectable amounts of polypeptides migrating at the expected size (~ 37 kDa) for a PB1-F2-GFP fusion protein, although the 5092 and UDL proteins ran slightly slower than the PR8 one. All three PB1-F2 proteins also showed some smaller products that might represent expression of C-terminal PB1-F2 fragments, although it was not clear which, if any, are expressed from AUGs 7-9 or were degradation products of the full length protein (Figure 3.1.2 B). Expression of these smaller products was strongest in 5092.

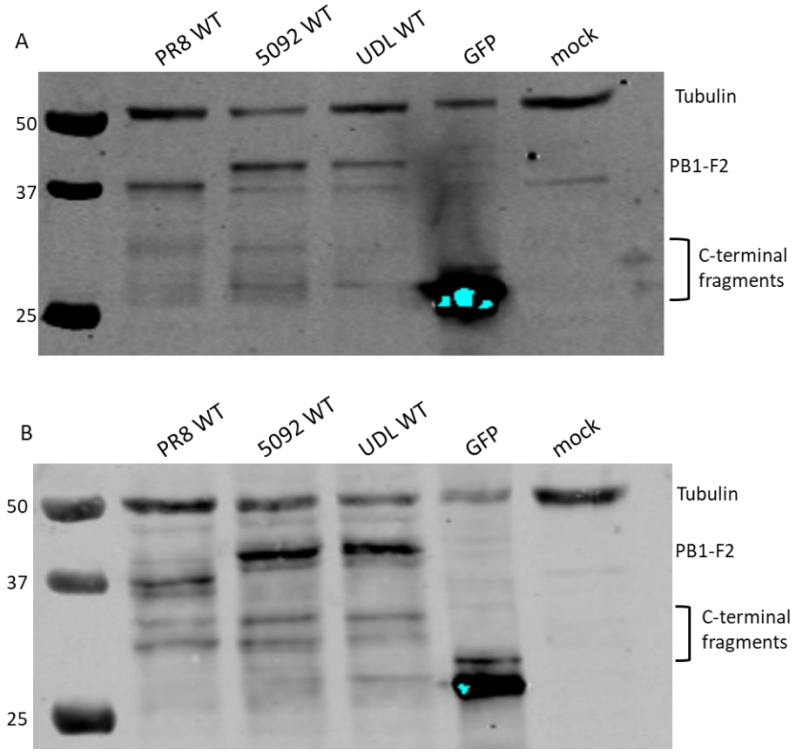

Figure 3.1.3 Effects of MG132 on Detection of PB1-F2 Polypeptides
293Ts were transfected with GFP tagged PB1-F2 plasmids, and in panel B, incubated with 10μM proteasome inhibitor MG132 16hrs before lysis. At 48h post transfection, cell lysates were run on a 4-20% gradient SDS-PAGE gel. 7μl PB1-F2-transfected cell lysates were run, with 3μl GFP-transfected sample to avoid oversaturation of the GFP band. Blots were simultaneously probed for β-tubulin as a loading control, and anti-GFP antibody was used to probe for PB1-F2 proteins.

Due to the inherent instability of PB1-F2 (Chen et al., 2001; James et al., 2019), and the possibility this would carry over to PB1-F2-GFP fusion proteins, including any C-terminal fragments, the proteasome inhibitor MG132 was used to improve visualisation (Figure 3.1.3). 293T cells were transfected in duplicate, and incubated with or without MG132. MG132 treatment in Figure 3.1.3 B led to stronger full length PB1-F2 bands for 5092 and UDL PB1-F2 peptides compared to 2A, and appeared to have stabilised the multiple possible C-terminal fragments into two stronger bands. Due to this improved clarity,

MG132 was used for subsequent PB1-F2-GFP expression experiments.

Expression of multiple protein products from the PB1-F2 ORF confirmed our hypothesis that avian sequences were likely to express C-terminal fragments, and confirmed that 5092 and UDL were viable models to characterise the fragments.

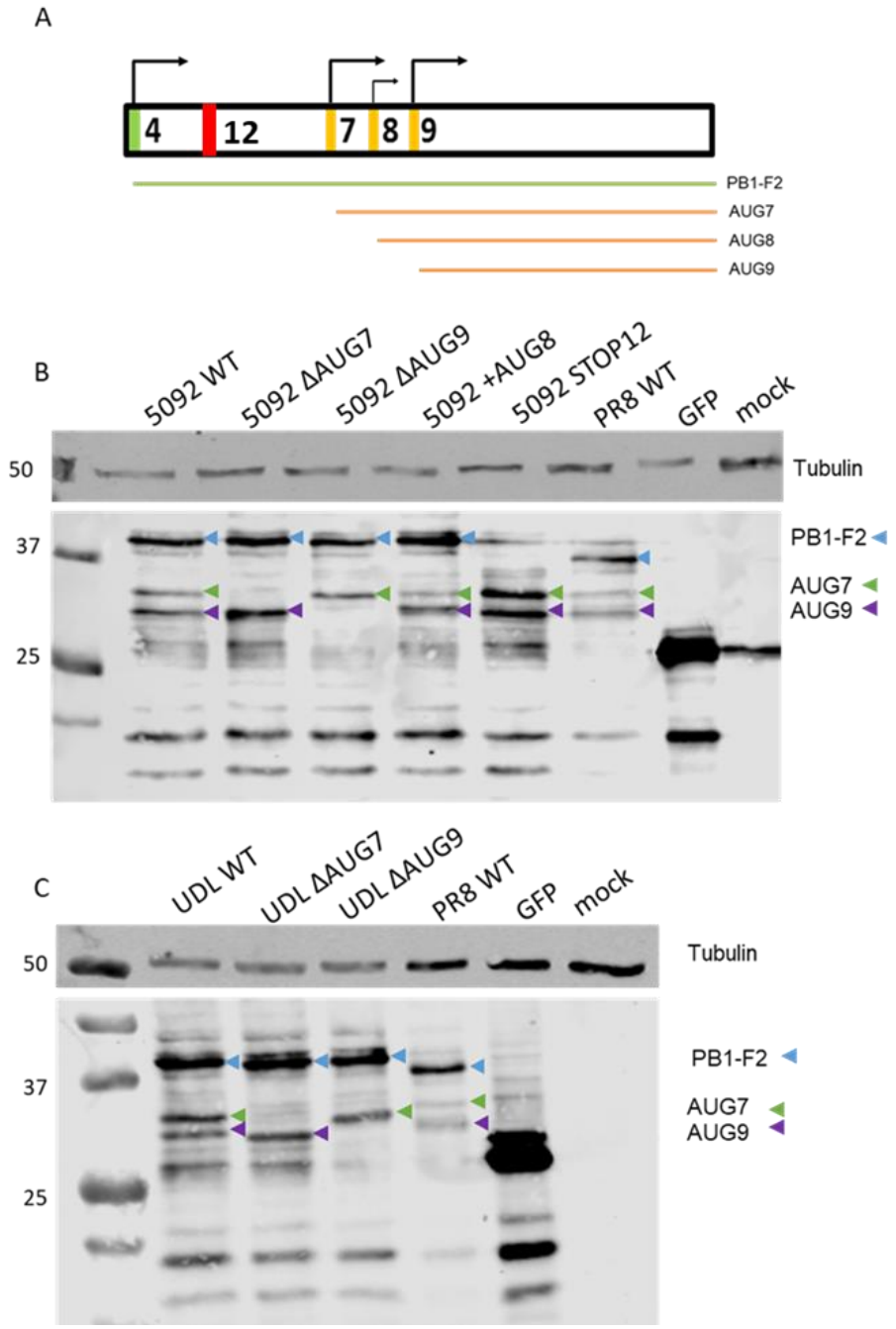

Figure 3.1.4 Expression of PB1-F2 ΔAUG Mutants

GFP-tagged PB1-F2 plasmids were transfected into 293T cells, with 15.5hr 10μM MG132 treatment and cell lysates were run on a 4-20% gradient SDS-PAGE gel 48h post transfection. Panel A shows expected PB1-F2 proteins produced, including the position of the STOP12 mutation. Panel B shows 5092 proteins and panel C shows UDL proteins. β-tubulin was used a loading control (top panels), and anti-GFP antibody used to probe for

PB1-F2 proteins (lower panels). 7µl PB1-F2-transfected cell lysates were run, with 3µl GFP-transfected samples run to avoid oversaturation of the GFP band.

To firmly identify which protein product was produced by which AUG codon within PB1-F2, a series of GFP tagged ΔAUG mutants were made (Figure 3.1.4). Each ΔAUG mutant had a Met>Thr change to prevent translation initiation. In both UDL and 5092 contexts, loss of either AUG 7 or AUG 9 resulted in the loss of a single polypeptide product, indicating the provenance of each. In 5092 PB1-F2, which in the WT genes only codes AUGs 7 and 9, introduction of AUG 8 appeared to have produced an additional band migrating just above the AUG 9 product (Figure 3.1.4 A). However, these species were hard to separate as the two AUG codons are only five amino acids apart. Reproducing the common N-terminal truncation STOP12 in the 5092 PB1-F2 gene still retained strong expression of both C-terminal PB1-F2 fragments.

These data supported the hypothesis that internal AUGs 7-9 of the PB1-F2 gene are recognised by ribosomes and can produce polypeptides. We were able to link specific polypeptide products to individual AUG codons, and showed that a STOP12 N-terminal truncation did not prevent continued C-terminal expression.

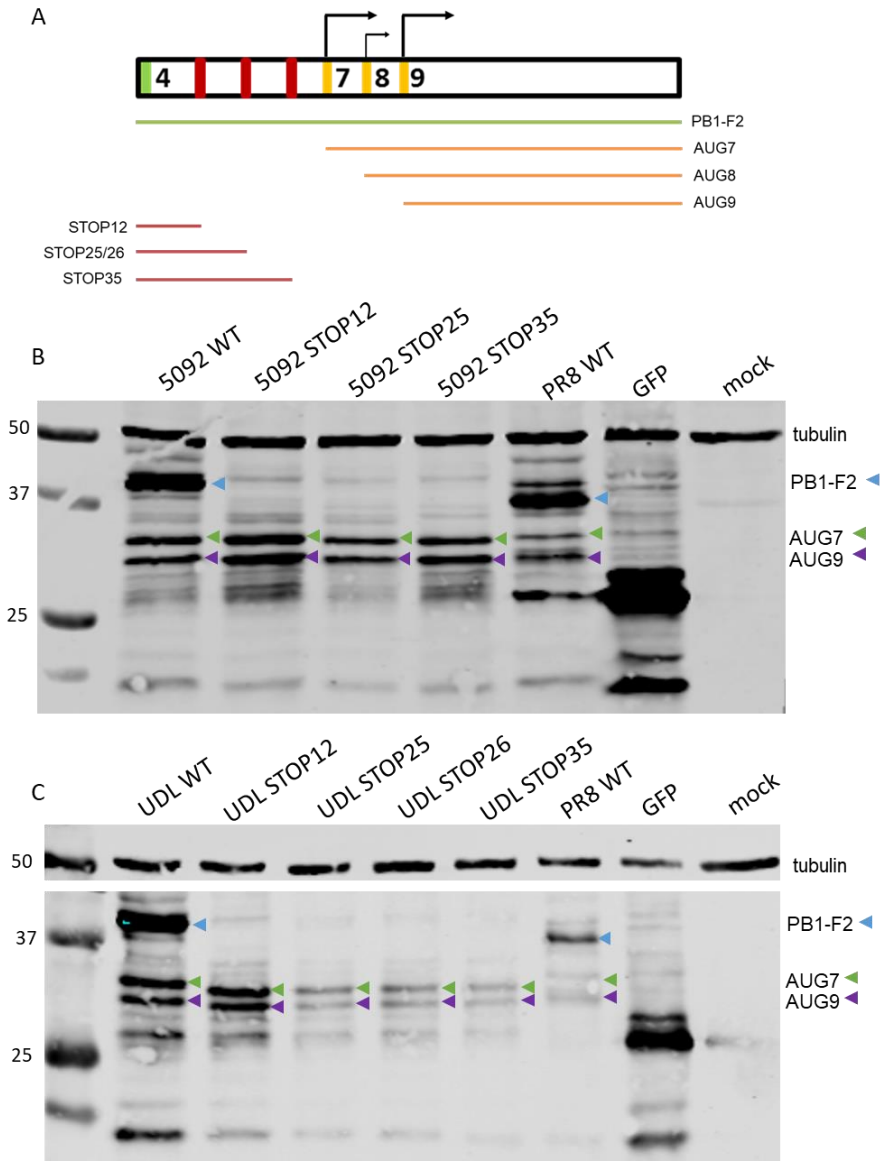

Figure 3.1.5 Expression of PB1-F2 STOP Mutants

GFP-tagged PB1-F2 plasmids were transfected into 293T cells, with 15.5hr 10μM MG132 treatment and cell lysates were run on a 4-20% gradient SDS-PAGE gel 48h post transfection. Panel A shows the relative position of STOP codons to AUGs 7-9, and possible PB1-F2 protein products. Panel B shows 5092 proteins and panel C shows UDL proteins. β-tubulin was used a loading control (top panels), and anti-GFP antibody used to probe for PB1-F2 proteins (lower panels). 7μl of PB1-F2-transfected cell lysates were run, with 3μl GFP-transfected sample run to avoid oversaturation of the GFP band.

Following the bioinformatic analysis detailed in Chapter 1, the most common N-terminal truncations (STOP12, STOP25, STOP26, STOP35) of PB1-F2 in avian viruses were introduced into GFP tagged constructs and the expressed polypeptides visualised by western blot as before (Figure 3.1.5). In both 5092 (H5N1) and UDL (H9N2) PB1-F2 proteins, truncation prevented expression of full length PB1-F2-GFP, but did not prevent expression of C-terminal fragments produced from either AUG 7 or AUG 9. However, STOP35 did result in the loss of a minor species above AUG 7 that was present as a doublet in all other 5092 lanes. Interestingly, both PB1-F2 backgrounds appeared to show variation in the strength of C-terminal fragment expression from the different STOP mutants. STOP12 constructs in particular produced more of the AUG 7 and AUG 9 peptides.

This is the first analysis we are aware of to identify protein products of individual C-terminal AUGs and to confirm this individual expression in the presence of N-terminal truncations.

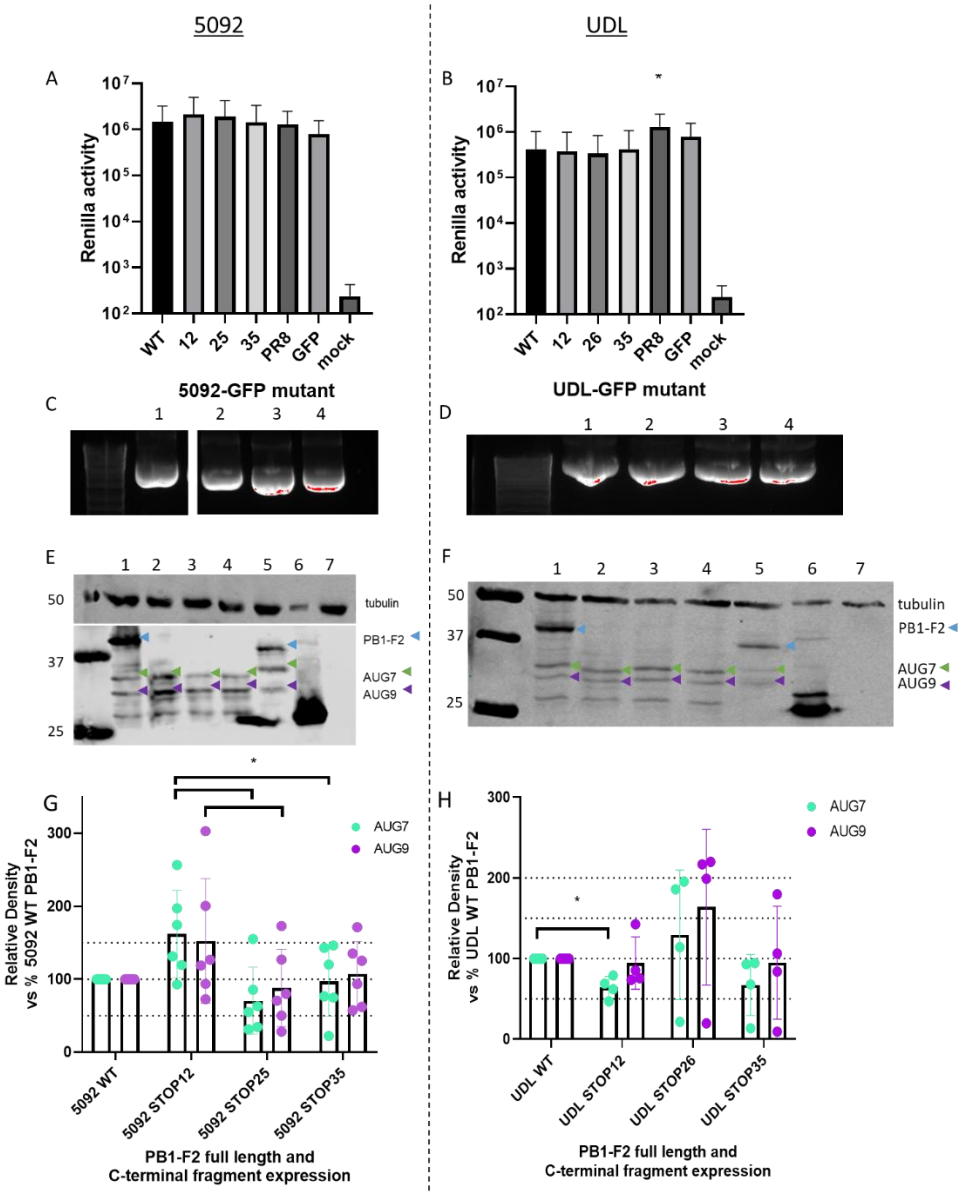

Figure 3.1.6 Quantification of C-terminal Expression from PB1-F2 STOP Mutant Transfection

293Ts were co-transfected with a Renilla luciferase reporter and GFP-tagged PB1-F2 plasmids to determine whether apparent differences in strength of C-terminal expression were due to inconsistent amounts of plasmid. All samples had 15.5hr 10μM MG132 treatment. Plasmid stocks were also run on an agarose gel to ensure equal concentration. Cell lysates were run on 4-20% gradient SDS-PAGE gels. β-tubulin was used a loading control, and anti-GFP antibody used to probe for PB1-F2 proteins.

Panel A: Renilla activity from co-transfection with 5092 PB1-F2. Panel C, lane order: 1, 5092 WT; 2, 5092 STOP12; 3, 5092 STOP25; 4, 5092

STOP35. 2% agarose gel of plasmid stocks. Panel D, lane order: 1, 5092 WT; 2, 5092 STOP12; 3, 5092 STOP25; 4, 5092 STOP35; 5, PR8 WT; 6, GFP; 7, mock. Representative western blot of co-transfected samples. Panel G: Relative density of 5092 PB1-F2 vs tubulin from panel C, quantified in ImageJ, analysed by two-way ANOVA corrected for multiple comparisons using a Tukey test.
Panel B: Renilla activity from co-transfection with UDL. Panel D, lane order: 1, UDL WT; 2, UDL STOP12; 3, UDL STOP25; 4, UDL STOP35. 2% agarose gel of plasmid stocks. Panel F, lane order: 1, UDL WT; 2, UDL STOP12; 3, UDL STOP26; 4, UDL STOP35; 5, PR8 WT; 6, GFP; 7, mock. Representative western blot of co-transfected samples. Panel H: Relative density of UDL PB1-F2 vs tubulin from panel G, quantified in Image J, analysed by two-way ANOVA corrected for multiple comparisons using a Tukey test.

To test if the difference in strength of C-terminal fragment visualisation between mutants was due to the introduction of STOP codons and not an artefact of the amount of plasmid introduced by transfection, co-transfections with a plasmid encoding Renilla luciferase were performed (Figure 3.1.6). The Renilla luciferase gene is under a constitutively active RNA pol II promoter, and given equal transfection volume and efficiencies should give equal expression and activity in 293T cells, under the assumption PB1-F2 does not have an impact on global protein expression. One-way ANOVA vs WT of Renilla activity showed no significant changes in activity between PB1-F2 mutants for either 5092 (Figure 3.1.6 A) or UDL (Figure 3.1.6 B), suggesting variation in PB1-F2 C-terminus expression was not due to human error. The PB1-F2-GFP plasmid stocks were also run on a 2% agarose gel to ensure equal concentrations were being used; this confirmed that human error in preparing the plasmid stocks was unlikely to be the source of the variation in expression (Figure 3.1.6 C, Figure 3.1.6 D). When western blots of replicate experiments were quantified by densitometry, 5092 STOP12 saw significant differences in the relative density of AUG 7 and 9 products compared to the other mutants

(Figure 3.1.6 G, Table 3.1.1). 5092 STOP12 AUG 7 had 162% of WT expression, with AUG 9 expressing at 153% of WT. STOP12 AUG 7 was significantly increased compared to expression from both STOP25 and STOP35 mutants, with similar p values of 0.0364 and 0.0328 respectively. STOP12 AUG 9 was only significant against STOP25 AUG 9 (p = 0.0270). UDL only saw a significant difference between WT and STOP12 AUG 7 (p = 0.0391) (Figure 3.1.6 H). The UDL STOP26 mutant did show a non-significant increase in expression from WT, with AUG 7 expressing 129% and AUG 9 expressing 164% of WT. Both STOP12 and STOP35 showed a decrease in expression compared to WT (Table 3.1.1).

Mutation	5092		UDL	
	AUG7	AUG9	AUG7	AUG9
STOP12	162.0	152.6	64.3	94.4
STOP25	70.7	88.1	-	-
STOP26	-	-	129.4	164.0
STOP35	97.1	107.4	67.3	94.9

Table 3.1.1 Relative Density of PB1-F2 C-terminal Fragment Expression
Mean relative density of C-terminal fragments produced from N-terminally truncated PB1-F2 sequences, normalised to tubulin to 1 decimal place, expressed as percentage WT expression.

Therefore, introducing N-terminal STOP codons into the PB1-F2 gene affected expression from AUGs 7 and 9 in a virus strain dependent manner. One possible explanation for the increased expression from AUG 9 rather than AUG 7 for most of the STOP mutants (Figure 3.1.6, Table 3.1.1), is that the additional 12 codons between STOP codons and AUG 9 (Figure 3.1.7) allows more time for the ribosome to reassemble the pre-initiation complex, if a termination-reinitiation mechanism is occurring.

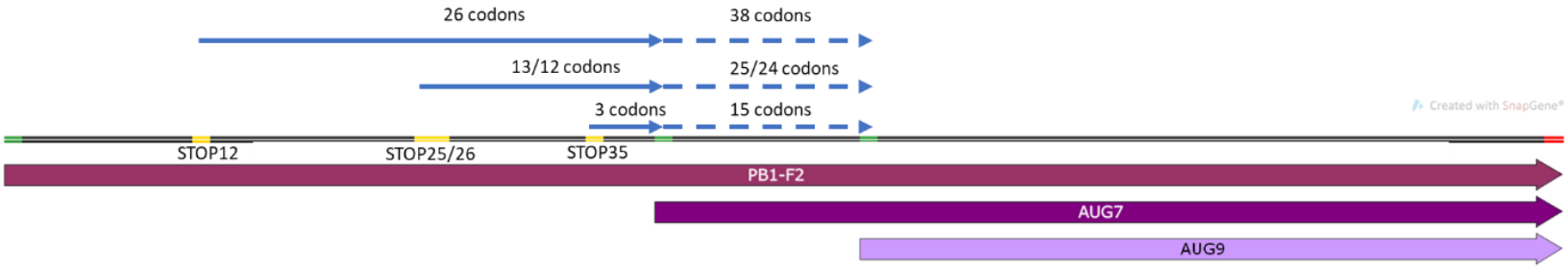

Figure 3.1.7 Reinitiation in PB1-F2 ORF
Schematic of possible products within the PB1-F2 ORF. Coloured arrows indicate ORFs, with green highlight for AUG codon and orange for N-terminal STOP codons. Solid blue arrows show distances between STOP codons and AUG 7, dotted blue arrows the continuation to AUG 9, with the total codon distances between STOP and AUG given.

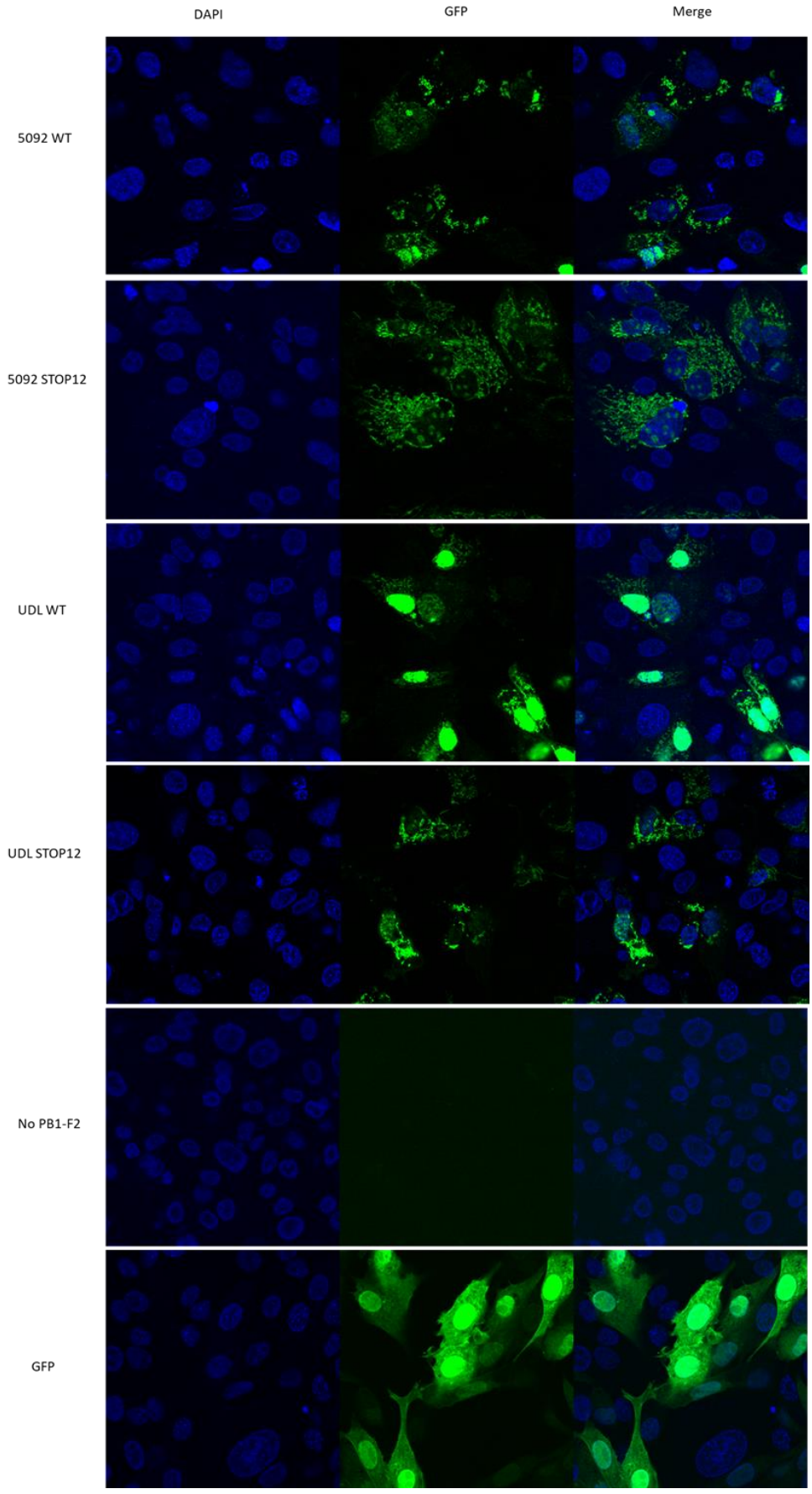
DAPI
GFP
Merge
5092 WT
5092 STOP12
UDL WT
UDL STOP12
No PB1-F2
GFP

Figure 3.1.8 Expression of WT and STOP12 PB1-F2-GFP Proteins
DF1 cells were transfected with 1µg GFP tagged PB1-F2 using Xfect, and slides fixed with 10% formalin 24h post transfection to avoid protein degradation. Cells were permeabilised with 2% Triton X-100 and mounted with Vectashield mounting medium with DAPI. GFP is a fluorophore, so no antibodies were required. Slides were imaged on a Leica Stellaris 5 confocal microscope. DAPI counterstain brightness was adjusted.

Intracellular Localisation of Full-length and N-terminally Truncated PB1-F2 Proteins

To assess how N-terminal truncation may affect cellular localisation of PB1-F2 polypeptides, plasmids encoding the WT and STOP12 PB1-F2-GFP proteins were transfected into DF1 chicken fibroblast cells and visualised by confocal microscopy (Figure 3.1.8). STOP12 was chosen as it consistently showed good expression of C-terminal fragments and would allow a comparison of C-terminal only and full length PB1-F2 localisation. Non-transfected and GFP only conditions were used as negative controls. GFP showed good expression levels and was present throughout the cell. The non-transfected control cells looked healthy in the DAPI channel and had no signal in the GFP channel. 5092 (H5N1) WT PB1-F2-GFP was present in the cytoplasm in a reticular speckled pattern previously associated with mitochondrial localisation (McAuley et al., 2010) and in some nuclei. 5092 STOP12 also showed a cytoplasmic reticular pattern consistent with the WT, with very limited nuclear localisation (Figure 3.1.8). UDL (H9N2) WT PB1-F2-GFP showed strong nuclear localisation and some cytoplasmic associated structures. The UDL STOP12 mutant had far less nuclear localisation than WT, but retained the cytoplasmic speckled structure.

While these localisation patterns were not entirely as expected (James et al., 2019), we were using a different protein tag, and there was some evidence of subtype-specific localisation.

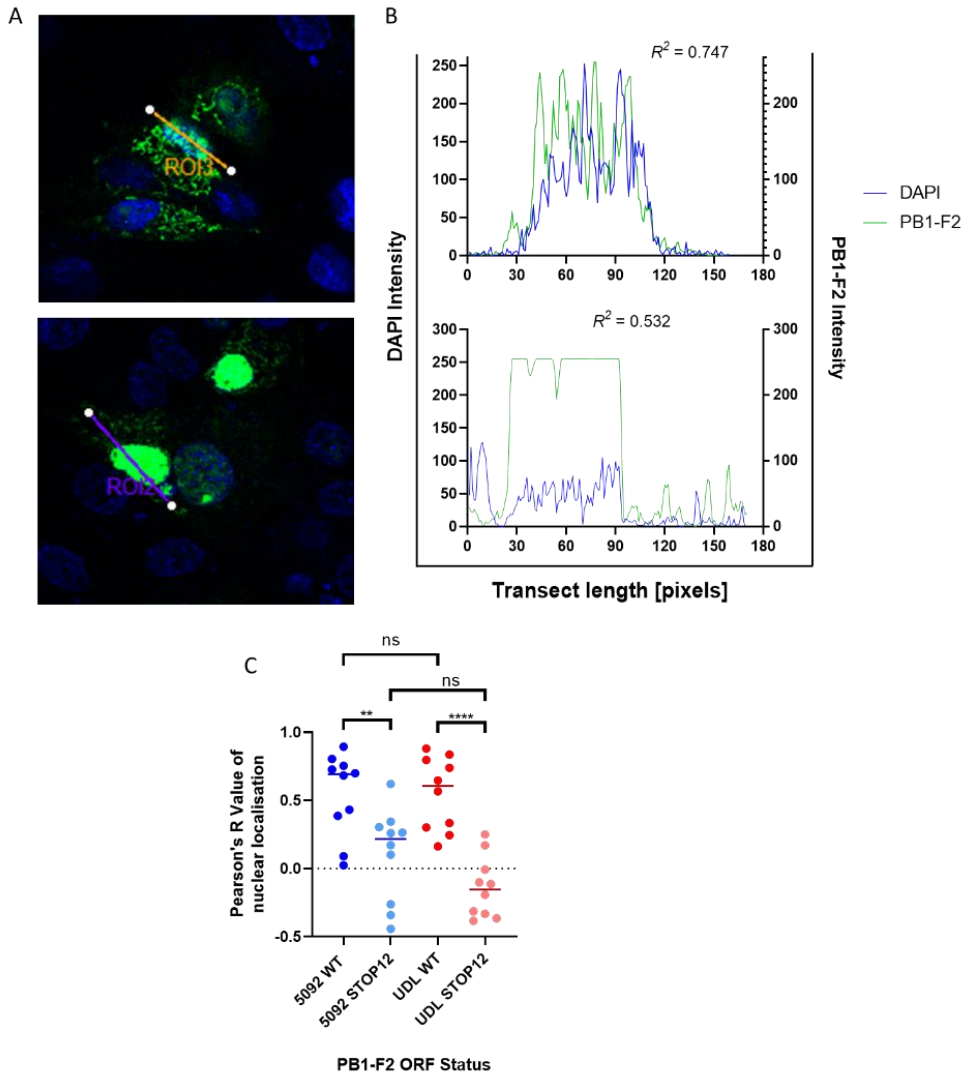

Figure 3.1.9 Localisation of WT and STOP12 PB1-F2-GFP Proteins

A) DF1 cells were transfected with 1µg GFP tagged PB1-F2 plasmids (top 5092 WT, bottom UDL WT) and slides fixed with 10% formalin 24h post transfection. Slides were imaged on a Leica Stellaris 5, as previously described. B) Transects drawn in both the DAPI and PB1-F2 channel were compared via Pearson's correlation coefficient (top 5092 WT, bottom UDL WT). C) The Pearson's correlation coefficient (R^2) values of 3 transects per cell from 10 cells for each PB1- F2 construct were averaged and plotted. One-way analysis of variance (ANOVA) was used to determine whether significant differences in the correlations existed.

After observing variation between cytoplasmic and nuclear subcellular localisation in Figure 3.1.8, we decided to quantify this using Pearson's correlation coefficient to measure correlation of DAPI and PB1-F2-GFP as a proxy for nuclear localisation (Figure 3.1.9). Three transects per cell for ten cells of both UDL and 5092 PB1-F2s were quantified and the R^2 values averaged for each cell. The mean Pearson's coefficient from each cell were then analysed by one-way ANOVA. There were no significant differences between subtypes for either WT or STOP12 PB1-F2 proteins. However, within subtype, both STOP12 proteins were significantly less nuclear than the corresponding WT proteins. 5092 PB1-F2 showed a decrease of $p = 0.0069$, while UDL had an even greater decrease of $p = <0.0001$ from WT to STOP12. Full length and C-terminal fragments of PB1-F2 have different levels of nuclear localisation. The C-terminus of the protein is significantly less present in the nucleus for both 5092 and UDL proteins.

Discussion

In this chapter, we assessed PB1-F2 peptide expression from all AUGs within the ORF. Previous studies have reported that C-terminal fragments of PB1-F2 can be expressed (Chen et al., 2001; Kamal et al., 2015; Zamarin et al., 2006) without identifying the specific AUG codons that trigger translation initiation. There are currently a wide range of PB1-F2 knock out strategies being used, with N-terminal truncations (James et al., 2016, 2019; Mazel-Sanchez et al., 2018), ΔAUG mutations (Le Goffic et al., 2010; Wise et al., 2009, 2011), or a combination (Cheung, Lee, et al., 2020; Cheung, Ye, et al., 2020; Leymarie et al., 2013, 2014; Schmolke et al., 2011).

Following reports of PB1-F2 C-terminal fragment expression (Zamarin et al., 2006), expression of full length PB1-F2 and AUG 7-9 products were validated for the model genes from UDL (H9N2) and 5092 (H5N1) viruses, using western blotting experiments of GFP-tagged

proteins (Figure 3.1.2). We were able to correlate protein products with specific AUG codons by sequential introduction of ΔAUG mutations (Figure 3.1.4), to show that both AUG 7 and AUG 9 expressed a peptide product in both PB1-F2 backgrounds. We also showed that expression of these C-terminal products is not abrogated by N-terminal truncations at different positions within the ORF (Figure 3.1.5). While expression from N-terminally truncated PB1-F2 proteins has been shown before (Kamal et al., 2015; Zamarin et al., 2006), this is the first time specific protein products have been identified for the individual C-terminal AUGs.

Given the reported variability in PB1-F2 stability (Cheng et al., 2017; James et al., 2019), it was not surprising to find variation extended to the strength of C-terminal fragment expression (Figure 3.1.5). Having confirmed the apparent variation in C-terminal fragment accumulation from PB1-F2 genes with differing STOP codon positions was not due to human error (Figures 3.1.6), we saw significant variation in fragment expression among 5092 mutants. Visualisation of peptides on western blots is the sum of expression from the relative AUG codon and the relative stability of the peptide. Data from the UDL constructs only reached significance when comparing AUG 7 expression between WT and the STOP12 mutant.

One possible explanation for the increased expression of STOP12 C-terminal fragments compared to other truncations of 5092, is efficiency of translation reinitiation. Reinitiation is known to occur on segment 2, to balance expression of AUGs 4 and 5 (Wise et al., 2011), and has been found to be most efficient with short ORFs of 13 codons or less (Kozak, 2001; Powell et al., 2008). A greater distance between the stop codon and the downstream AUG allows more time for the scanning 40S subunit to reacquire the trappings of the pre-initiation complex (Kozak, 2001). If translation reinitiation is behind the expression of the C-terminal fragments, the variation in expression according to STOP codon position could be explained by this

dependency on distance. This hypothesis would explain the non-significant trend of expression from AUG 9 (residue 51) having higher mean relative density compared to WT PB1-F2 than expression from AUG 7 (residue 39) when STOP25, 26 or 35 codons were introduced in both UDL and 5092 (Figures 3.1.6, Table 3.1.1, Figure 3.1.7). Wise and colleagues describe a termination-reinitiation mechanism for the expression of PB1-N40 from AUG 5 with a distance of 63 nucleotides between STOP and AUG codons (Wise et al., 2011). Within the PB1-F2 ORF, the only distances of >60 nucleotides between STOP and AUG codons are STOP25/26 to AUG 9, or STOP12 to either C-terminal AUG (Figure 3.1.7). However, this does not perfectly match the relative density data as STOP12 and STOP25/26 mutants did not always most strongly express C-terminal fragments (Table 3.1.1), so more work is needed to fully elucidate the mechanism. In future, this hypothesis could be tested by increasing or decreasing the distance between STOP and AUG codons along the RNA through mutagenesis and quantification of western blots.

Besides the retention of C-terminal expression from N-terminally truncated PB1-F2 mutants, another interesting question was whether the C-terminal fragments would have different intracellular localisation patterns than the full length WT protein (Figure 3.1.8, Figure 3.1.9). Previous reports have indicated that sequence polymorphisms in the C-terminus can affect localisation of the full length PB1-F2 through the presence of the mitochondrial targeting sequence (MTS), or minimal sequence motif (James et al., 2019). The N-terminus of PB1-F2 is thought to be disordered (Bruns et al., 2007), so we wanted to investigate how expression of only the smaller, more targeted C-terminus would affect localisation.

We were able to confirm C-terminus expression for both 5092 and UDL (Figure 3.1.8), and found that full length PB1-F2 had a greater nuclear localisation than C-terminus fragments (Figure 3.1.9). However, due to the use of eGFP as a C-terminal tag, these results may not be

reflective of the behaviour of PB1-F2 during infection. PB1-F2 is a small protein, with an estimated molecular weight of 10-11kDa, whereas eGFP is a large 29kDa protein. eGFP has previously been noted to affect localisation of PB1-F2 in comparison to a smaller tag (Yamada et al., 2004), and the nuclear/mitochondrial localisation of 5092 and UDL PB1-F2s seen here is quite different to that seen in previous experiments that used smaller N-terminal tags to detect the proteins (James et al., 2019). Unfortunately, a C-terminal V5 tagged PB1-F2 could not be reliably detected (data not shown), and there were concerns about the stability of such small protein fragments (5092 AUG 7 6.30kDa, AUG 9 4.92kDa; UDL AUG 7 6.37kDa, AUG 9 4.98kDa; calculated in Expasy). Therefore, the PB1-F2-eGFP was used to increase both the stability (and hence accumulation), and size of fragments for easier detection on SDS-PAGE gels, and to ensure visibility in confocal experiments.

Full length and C-terminal fragments of PB1-F2 may have different localisations due to the inherent disorder of the N-terminus of the protein. The N-terminus has previously been associated with structural change in PB1-F2 (Ajjaji et al., 2016; Bruns et al., 2007) that could impact efficiency of the MTS by masking during protein folding. Previous reports of the N-terminus interacting with PB1 in the nucleus (Chen et al., 2010; McAuley et al., 2010), suggest there could be a PB1 binding site in the N-terminus of PB1-F2, as PB1/PB1-F2 colocalisation during PR8 infection was only seen with a full length protein, suggesting a binding interaction is required (McAuley et al., 2010). Structural predictions generally suggest that the C-terminus is largely helical while the N-terminus is not (Bruns et al., 2007; Solbak et al., 2013). Without the N-terminus, C-terminal fragments may be able to insert into membranes more efficiently or interact with different binding partners as smaller helical proteins.

The data presented in this chapter provide an analysis of the full range of peptide products from the PB1-F2 ORF. In a plasmid-based

expression system, translation from AUGs 7 and 9 was detected in cells regardless of the introduction of N-terminal truncations, and C-terminal fragments showed different intracellular localisation patterns than the full length 5092 and UDL PB1-F2 proteins. Investigations of how changes in the PB1-F2 expression repertoire may affect the proteins' functions will be explored in the following Chapter 4.

Chapter 3.2 Variable Stability of PB1-F2 Proteins

Introduction

PB1-F2 is an unstable protein with a short half-life (Zamarin et al., 2006). Multiple studies have found that differences in PB1-F2 expression levels occur at the level of protein turnover and stability within the cell rather than due to differences in mRNA production (Chen et al., 2010; Park et al., 2019). PB1-F2 proteins are rapidly degraded, but the mechanism is unclear. Stability is highly sequence dependent and varies between virus strains (Cheng et al., 2017). In a panel of eight PB1-F2 proteins from different virus subtypes only the Hong Kong 156 (H5N1) PB1-F2 was stabilised by addition of the proteasome inhibitor MG132, suggesting multiple mechanisms of degradation (Chen et al., 2010). In our results above, we saw that MG132 increased protein stability (Figure 3.1.3), so both 5092 and UDL PB1-F2s appear to be subject to proteasomal degradation. It is thought that different intracellular localisations of PB1-F2 proteins can affect the proteins stability. Stability is known to affect protein function, and in PB1-F2, stability and localisation appear to be linked (Cheng et al., 2017).

Amino acids 68-71 of PB1-F2 affect its stability (Cheng et al., 2017). PR8 PB1-F2, with a 68-71 TQDS motif, shows higher stability than PB1-F2 from an H5N1 virus (A/Hong Kong/156/1997) with an ILVF motif at the same position (Cheng et al., 2017). Alteration of amino acids 68-71 affected both protein half-life and subcellular localisation, resulting in instability in the cytoplasm and a longer half-life at the mitochondria for re-directed mutant proteins (Cheng et al., 2017). Residues 68 and 69 were also found to alter PB1-F2 stability in the 1918 strain protein (Park et al., 2019). Altering residues 68 and 69 swapped intracellular localisation of PR8 and 1918 strain PB1-F2 proteins between punctate structures in the mitochondria and a diffuse cytoplasmic localisation. Changes in PB1-F2 stability appeared in a

cumulative manner, with the double mutant causing a stronger change in phenotype than single changes (Park et al., 2019). PB1-F2 half-life may also be affected by ubiquitination of lysines at positions 73, 78, and 85 (Cheng et al., 2017). The lysines form a major ubiquitination cluster which is 86-89% conserved in the avian consensus, rising to 90-96% conserved in avian H5, H7 consensus (Košík et al., 2015). Immunoprecipitation assays showed that both PR8 and 1918 PB1-F2 proteins were heavily ubiquitinated (Park et al., 2019). Ubiquitin is directly involved in protein degradation via proteasome targeting, MHC (major histocompatibility complex) peptide presentation, and viral recognition (Michalek et al., 1993). The 1918 PB1-F2 is sensitive to proteasomal degradation and may be able to use this as an IFN antagonism mechanism by causing co-degradation of host immune modulators in the induction pathway (Park et al., 2019).

Mutating the lysine residues to arginines increases PB1-F2 stability, inducing protein accumulation and increasing the antibody response to the protein. Inhibition of ubiquitination also causes accumulation of the protein (Košík et al., 2015). Not all PB1-F2 proteins correlate stability with the number of lysine residues in the sequence, but whether this is due to alternate degradation mechanisms or specific intracellular localisation patterns protecting against ubiquitination is unclear (James et al., 2019).

The nuclear localisation of the Hong Kong 1997 H5N1 PB1-F2 protein appears to protect it from degradation (Košík et al., 2015). Other polymorphisms also affect localisation and stability of PB1-F2. Disrupting mitochondrial localisation of another H5N1 PB1-F2 through mutation of residues 60, 62, 66, and 68 resulted in a significant decrease in protein stability compared to WT (James et al., 2019). Sub-cellular localisation has been associated with determination of PB1-F2 function due to proximity with interaction partners (James et al., 2019). It is possible that localisation in a membrane-bound

organelle provides protection from degradation by preventing trafficking to the proteasome.

There is limited literature examining PB1-F2 protein stability compared to its other properties. As always with this protein, results must be considered in the context of virus strain sequence specificity and the host specific nature of PB1-F2. However, a link between localisation and stability has been hypothesised in several studies using different virus strains (Cheng et al., 2017; James et al., 2019; Kamal et al., 2018). Multiple PB1-F2 proteins have been noted to become stabilised by the addition of the proteasome inhibitor MG132, indicating that at least a proportion of PB1-F2 variants will be degraded via this route (Buehler et al., 2013; Cheng et al., 2017; James et al., 2016; Pasricha et al., 2018).

Previous work from the Shelton lab has suggested a link between localisation and stability (James et al., 2019). In this study, we assessed the stability of a panel of PB1-F2 proteins reported to have different intracellular localisation patterns to test the hypothesis, and suggest other possible factors affecting PB1-F2 stability, such as structural differences.

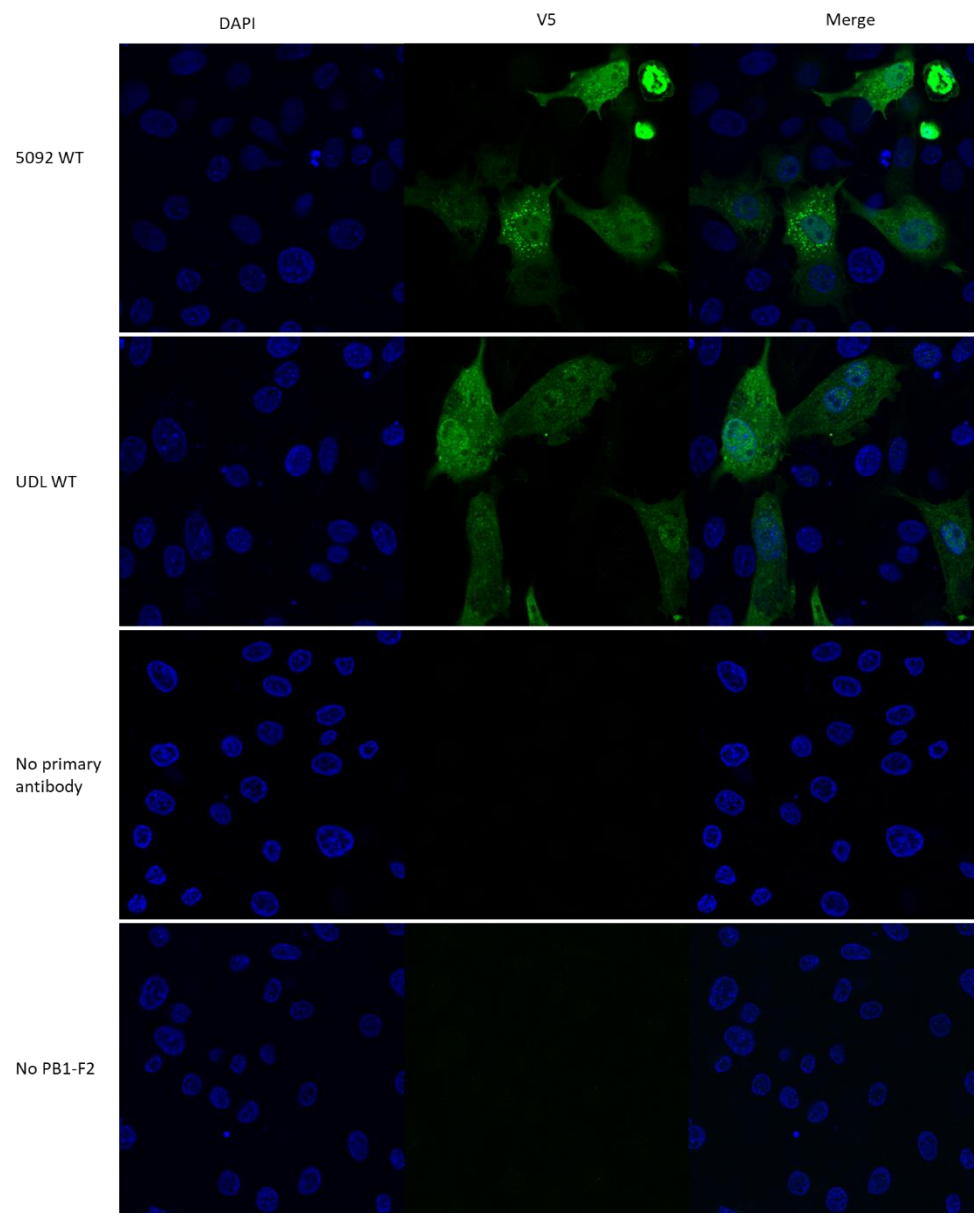

Figure 3.2.1 Variable Intracellular Localisation of V5-PB1-F2 Proteins
DF1 cells were transfected with 1µg V5 tagged PB1-F2 and slides fixed 24h post transfection to avoid protein degradation. Cells were fixed, permeabilised, and stained with anti-V5 tag primary antibody followed by anti-mouse 488nm secondary antibody. Nuclei were counterstained with DAPI. Coverslips were mounted and imaged on a Leica Stellaris 5 confocal microscope.

To investigate the intracellular localisations of PB1-F2 proteins from 5092 (H5N1), and UDL (H9N2) viruses, DF1 cells were transfected with 1µg of plasmids encoding the V5 tagged PB1-F2s (Figure 3.2.1). The V5 tag is a short sequence of 14 residues that can be detected using a specific antibody (James et al., 2019). Preliminary experiments showed that the addition of MG132 had resulted in aggregation of PB1-F2 within the cells so was not used (data not shown). The slides were fixed with 10% formalin at 24h to avoid protein degradation affecting visualisation. The 5092 PB1-F2 showed areas of punctate structures associated previously with mitochondrial localisation (McAuley et al., 2010), while UDL PB1-F2 appeared more evenly cytoplasmic. Both strains of PB1-F2 also localised partially to nuclei, though this was more marked for the UDL protein. Immunostaining for the V5 tag allows detection of PB1-F2 proteins (Figure 3.2.1), although the localisation appears to differ from the GFP tagged proteins, with a more generally cytoplasmic localisation for UDL (Figure 3.1.8).

Stability of PB1-F2 Proteins

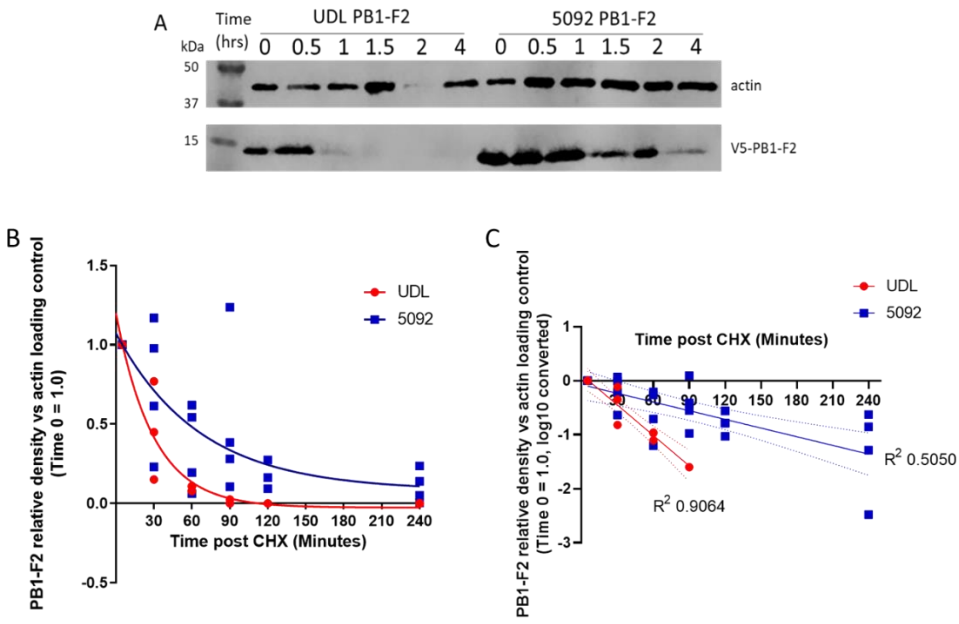

Figure 3.2.2 Degradation Curves of Cycloheximide Chase Assays of PB1-F2 Proteins

DF1 cells were transfected with V5 tagged PB1-F2 plasmids (5092 n=4,

UDL n=3). At 24hpt, CHX at 10µg/ml was added, and samples collected over a 4hr timecourse. Western blots of CHX assays probed with anti-V5 antibody and using actin as a loading control (Panel A) were quantified by densitometry using ImageJ. V5-PB1-F2 values were made relative to the corresponding actin loading controls, with time 0 set to 1, and a non-linear regression fit for a degradation curve performed (Panel B), which was then log_{10} transformed (Panel C). Dashed lines indicate 95% CI bands.

Following microscopy to confirm sub-cellular localisation of the V5-tagged PB1-F2 proteins, stability was assessed using cycloheximide (CHX) chase assays for both 5092 (H5N1) and UDL (H9N2) model proteins (Figure 3.2.2). DF1 cells were transfected with V5 tagged PB1-F2 plasmids, and CHX applied at 24 hours post transfection (hpt), with samples taken over 4h to measure protein degradation. The actin loading control was maintained throughout the time course as it has a long half-life (James et al., 2019). Western blots were quantified using ImageJ. For each timepoint, the band of interest was made relative to the loading control. By comparing all subsequent timepoints to time 0, we were able to visualise degradation rate more effectively (Figure 3.2.2 B, Figure 3.2.2 C). In order to better quantify and compare the degradation of PB1-F2 proteins from different strains, a non-linear regression was performed to estimate their half lives (Figure 3.2.2 C, Table 3.2.2). Both viral proteins degraded during the time course, however 5092 PB1-F2 protein was more stable than the PB1-F2 of UDL (Figure 3.2.2 A, Table 3.2.2). UDL PB1-F2 degraded below the limit of detection within 1.5h, whereas 5092 PB1-F2 was still detectable at the final 4hr timepoint.

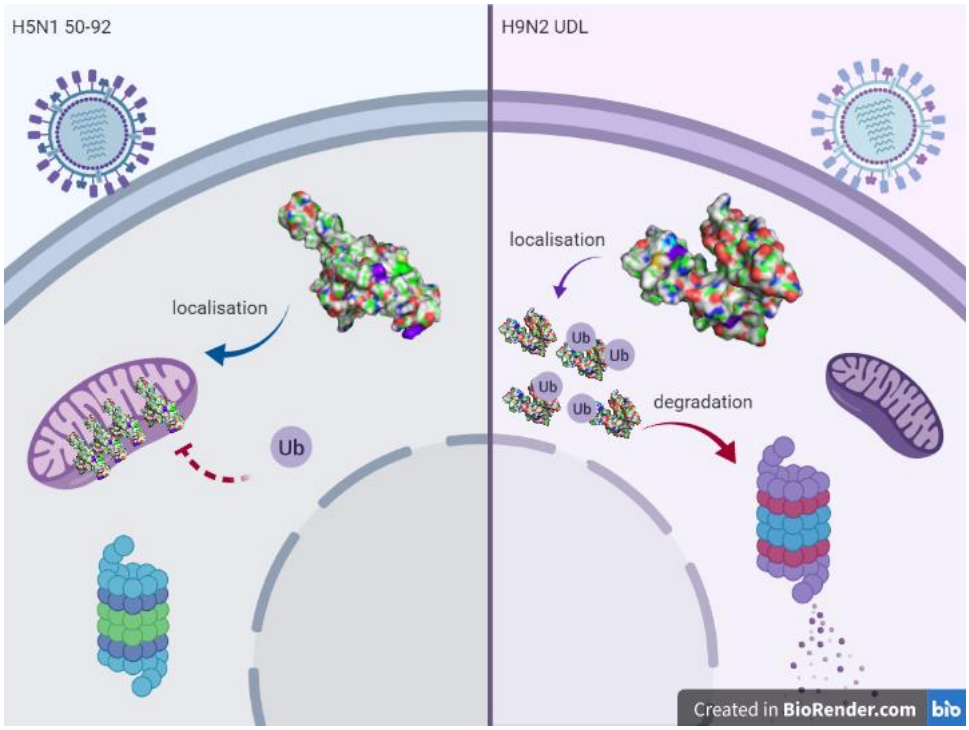

Figure 3.2.3 Localisation and Stability Hypothesis

Graphical representation of our hypothesis. 5092 PB1-F2 is associated with mitochondrial localisation, where membrane insertion may protect the protein from cellular degradation mechanisms. UDL PB1-F2 is associated with cytoplasmic localisation, which may leave it vulnerable to ubiquitin targeting and proteasomal degradation.

The different degradation rates of 5092 and UDL PB1-F2s (Figure 3.2.2) appeared to support a link between differing intracellular localisation and variation in stability, as described in the literature. We began working under the hypothesis that mitochondrial localisation inhibited PB1-F2 degradation, possibly through mitochondrial membrane insertion hindering ubiquitin conjugation (Figure 3.2.3). Stability assays were then expanded to the rest of the PB1-F2 panel of proteins from different subtypes assessed in James et al. (2019) to test the colocalization hypothesis. We chose to use these proteins as localisation had already been assessed (Table 3.2.1).

Full name	Abbreviation	Reported Localisation
A/turkey/England/5092/91	5092	Mitochondrial
A/mallard/Alberta/156/2001	H3N8a	Mitochondrial
A/chicken/Pakistan/UDL01/08	UDL	Cytoplasmic
A/duck/Italy/6103/V07	H11N9	Cytoplasmic
A/duck/Italy/3139/V06	H3N8b	Cytoplasmic

Table 3.2.1 V5-PB1-F2 Panel

V5-tagged PB1-F2 proteins used in CHX assays and reported intracellular localisations from James et al. (2019).

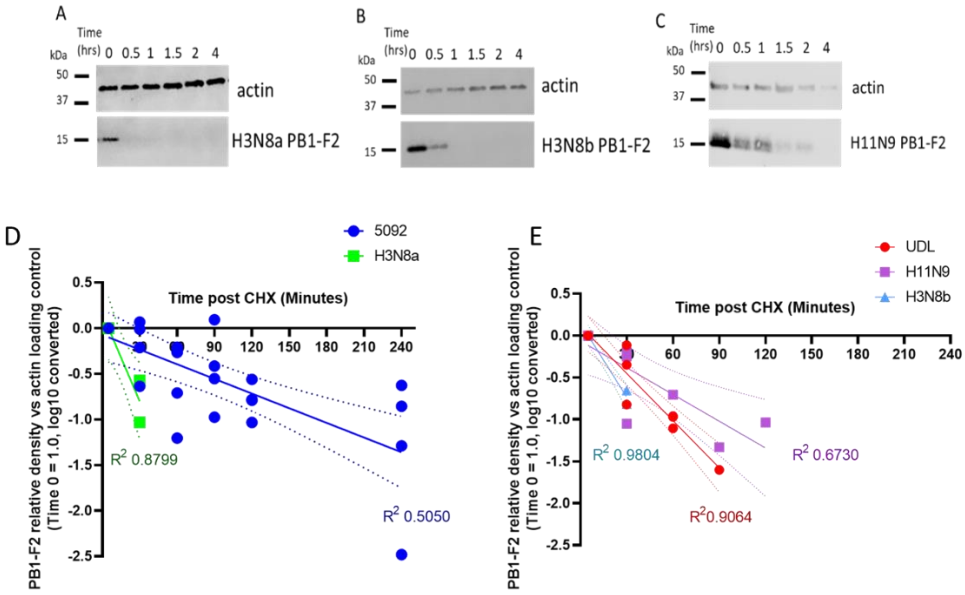

Figure 3.2.4 Degradation Curves of Cycloheximide Chase Assays of Mitochondrially vs Cytoplasmically Localised PB1-F2 Proteins

Representative western blots of H3N8a, H11N9 and H3N8b PB1-F2s are shown, all to n=3 (Panels A-C). PB1-F2 proteins reported to have mitochondrial (Panel D) or cytoplasmic (Panel E) localisations were assessed by CHX chase assay as described previously, and a nonlinear regression of $\log_{10}$ relative density performed. 5092 and UDL PB1-F2 data repeated from Figure 3.2.2.

Across this protein panel, there was a variation in PB1-F2 stability that was not entirely explained by their reported localisations (Table 3.2.1)

(James et al., 2019). H3N8a was reported to have a mitochondrial localisation but could not be reliably visualised past 0.5h in these CHX assays (Figure 3.2.4 A). H3N8b was reported to have cytoplasmic localisation and was also unable to be reliably detected beyond 0.5h of CHX treatment (Figure 3.2.4 B). H11N9 PB1-F2 appeared to be more stable, as it could still be clearly visualised at 1h (Figure 3.2.4 C).

Replicate experiments showed consistent rapid degradation of H3N8 PB1-F2 proteins (Figure 3.2.4). 5092 PB1-F2 was more stable than the H3N8a protein, and also showed more variability in the CHX assays. H3N8a expressed well at the 0h timepoint but degraded rapidly whereas 5092 PB1-F2 could be visualised throughout the timecourse (Figure 3.2.4 D). H11N9 and UDL proteins had more similar degradation curves to each other than to H3N8b, which degraded within 1hr while H11N9 and UDL were both able to be visualised up to 90 minutes, although UDL had less variance and a higher R^2 value than H11N9 PB1-F2 (Figure 3.2.4 E).

Such different degradation curves (Figure 3.2.4 D, Table 3.2.2) of the two proteins with mitochondrial localisation suggests other factors affect PB1-F2 protein stability. 5092 was the most stable but also the most variable PB1-F2 assessed, as shown by its low R^2 value. This variance appeared to be inherent rather than due to a single run. Although it is possible that there is a difference in degradation between the H3N8 proteins that was missed due to the choice of time points, the consistent instability of H3N8 PB1-F2s suggests there may be a separate virus strain specific reason for their rapid degradation rather than intracellular localisation alone.

Abbreviation	Mean half life (minutes)	Standard deviation of half-life (minutes)	Number of repeats	Predicted structure
5092	41	7	4	Linear
H3N8a	7	6	3	Linear
UDL	20	10	3	Globular
H11N9	21	12	3	Globular
H3N8b	7	5	3	Globular

Table 3.2.2 Half lives of PB1-F2 Proteins

V5-PB1-F2 densitometry values were made relative to the corresponding actin loading controls, with time 0 set to 1, and a non-linear regression curve fitted for a degradation curve. From this non-linear regression, a half-life in minutes was calculated for each replicate. Arithmetic mean and standard deviation were calculated for the half-life, and one-way ANOVA used to test for significant differences between mean half-lives of each PB1-F2 protein.

Using the non-linear regression, a mean half-life in minutes was calculated for each protein to the nearest whole minute (Table 3.2.2). This allowed a direct comparison of PB1-F2 proteins. 5092 was the most stable, with a mean half-life of 41 minutes. As expected from their predicted localisations, UDL and H11N9 had on average shorter half-lives than 5092 of 20 and 21 minutes respectively. Despite this similarity in degradation rate, only the mean half-life of H11N9 was significantly different to 5092 ($p=0.0127$). Both H3N8 proteins had shorter mean half-lives of seven minutes, with comparatively large standard deviations (Table 3.2.2). Although H3N8a was predicted to have a similar localisation to 5092 (James et al., 2019), the proteins half lives were significantly different ($p=0.0124$).

This data appeared to contradict the original colocalization hypothesis (Figure 3.2.3) as the reportedly mitochondrial H3N8a was highly unstable, and suggested PB1-F2 stability is likely to be a multifactor property. One possible additional factor for the variation in stability was that there could be differences in protein structure.

Structural Predictions of PB1-F2 Proteins

To investigate possible structural causes for the incompleteness of our hypothesis, the I-TASSER structural prediction server was used. The WT amino acid sequence for each protein was submitted to the server for predictions to be generated through iterative threading (Roy et al., 2010). These predictions suggested the 5092 PB1-F2 is mostly linear in structure with parallel alpha helices giving a long narrow structure, whereas UDL PB1-F2 has more unstructured loops linking three helices into a globular shape, with two termini around a central cleft (Figure 3.2.5). Both proteins were predicted to have a disordered N-terminus, however 5092 seems to be the more 'classical' PB1-F2 structure, with a long helix in the C-terminus. All current structural predictions (PR8, 2009 H1N1, 1918 H1N1, and avian 2006 H5N1) support long linear helices in the C-terminus rather than the disjointed helices seen in UDL PB1-F2 (Figure 3.2.5), but there is very limited data on 3D structure of PB1-F2 (Bruns et al., 2007; Solbak et al., 2013).

To improve confidence in the different predicted models, a sequence alignment of UDL and 5092 was performed (Figure 3.2.5 C). While the proteins had very similar sequences in some areas previously noted to affect stability, such as the TQDS motif and the lysine cluster in the C-terminal region, several polymorphisms within the putative mitochondrial targeting sequence (MTS) had previously been noted to influence localisation and stability.

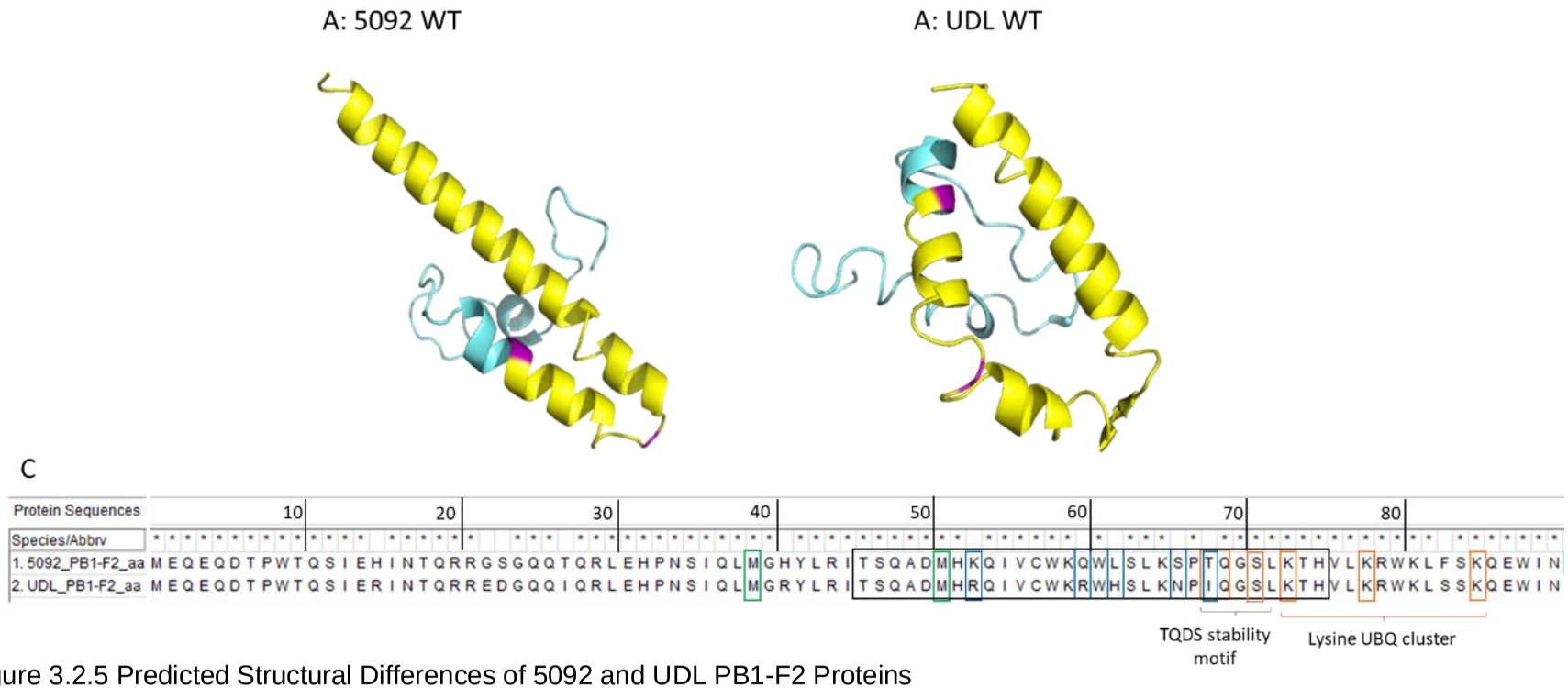

Figure 3.2.5 Predicted Structural Differences of 5092 and UDL PB1-F2 Proteins
(A, B) Structural predictions from I-TASSER were visualised in Pymol. N-terminus (aa1-38) in blue, C-terminus (aa39-90) in yellow. AUGs 7 and 9 highlighted in purple. C) Amino acid alignment of PB1-F2 proteins, with C-terminal AUGs highlighted in green. Previously reported stability markers are highlighted in orange. The black box demarcates aa46-75, which are generally accepted to contain the MTS (James et al., 2019). Polymorphisms within the MTS, which are mostly likely to affect localisation, are highlighted in blue.

After discovering such a large difference in predicted structure between 5092 and UDL PB1-F2s, the WT amino acid sequences of the rest of the PB1-F2 panel were put through I-TASSER to determine if there was a commonality in structure across intracellular localisation or rate of degradation. The main difference between stable 5092 and unstable H3N8a PB1-F2s (Figure 3.2.6) is the break in the C-terminal helix. The predicted helix break appears to alter the tightly parallel organisation of the C-terminus, and may affect membrane insertion and therefore stability, explaining why the two proteins have such different degradation rates despite both being mitochondrially localised. 5092 and H3N8a have a difference in sequence within the MTS at residue 70, which corresponds to the break in the long C-terminal helix, however most of the polymorphisms are focused around the link between the short and the long helices (Figure 3.2.6).

In a departure from the reported PR8 and avian H5N1 or H9N2 PB1-F2 structures (Pasricha et al., 2018; Solbak et al., 2013), all three cytoplasmically localised proteins were predicted to form a triangle of shorter helices in the C-terminal (Figure 3.2.7). Some previous structural predictions have reported shorter helices in the C-terminus, although their 3D orientation was not predicted (Pasricha et al., 2018). The structure of the three cytoplasmically localised proteins is more globular than the mitochondrial localised proteins and therefore less likely to facilitate membrane insertion, which may leave the proteins more vulnerable to degradation by cellular proteases. Although UDL and H11N9 PB1-F2s have very similar half-lives and predicted structures, the H3N8a sequence has a high similarity to UDL within the putative MTS (Figure 3.2.7 D).

The predicted structures of the PB1-F2 panel do not predict protein stability, but we did see consistent patterns between linear or globular predicted structures and the reported intracellular localisations (Table 3.2.2).

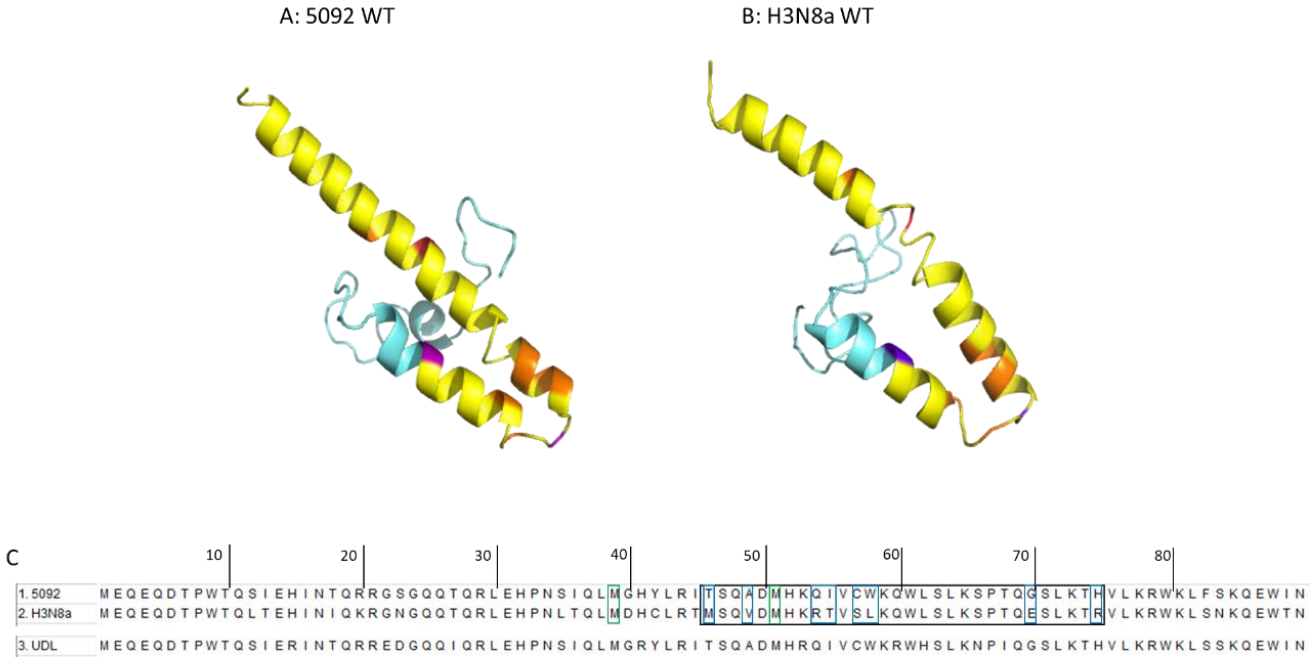

Figure 3.2.6 Ribbon Structure and Amino Acid Alignment of Mitochondrially Localised PB1-F2 Proteins
(A, B) Structural predictions of mitochondrially localised PB1-F2 proteins from I-TASSER were visualised in Pymol. N-terminus (aa1-38) in blue, C-terminus (aa39-90) in yellow. AUGs 7 and 9 highlighted in purple. Polymorphisms between aa46-75 highlighted in orange, with residue 70 in red. (C) Protein sequences were aligned in MEGA7, with C-terminal AUGs highlighted in green. The black box demarcates aa46-75, which are generally accepted to contain the MTS (James et al., 2019). Polymorphisms within the MTS, which are mostly likely to affect localisation, are highlighted in blue. UDL sequence is also shown for comparison between reported intracellular localisations.

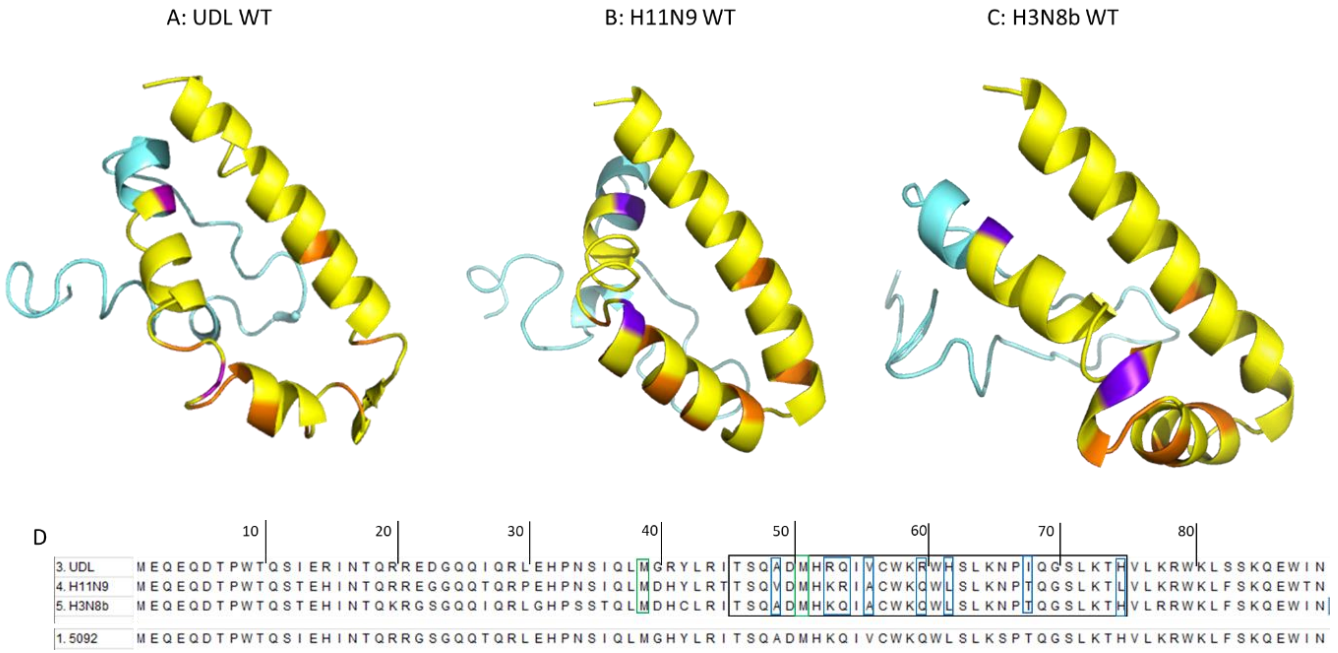

Figure 3.2.7 Ribbon Structure and Amino Acid Alignment of Cytoplasmically Localised PB1-F2 Proteins
(A – C) Structural predictions of cytoplasmically localised PB1-F2 proteins from I-TASSER were visualised in Pymol. N-terminus (aa1-38) in blue, C-terminus (aa39-90) in yellow. AUGs 7 and 9 highlighted in purple. Polymorphisms between aa46-75 highlighted in orange. (D) Protein sequences were aligned in MEGA7, with C-terminal AUGs highlighted in green. The black box demarcates aa46-75, which are generally accepted to contain the MTS (James et al., 2019). Polymorphisms within the MTS, which are mostly likely to affect localisation, are highlighted in blue. 5092 sequence is also shown for comparison between reported intracellular localisations.

Although H3N8a and H3N8b PB1-F2 proteins are from the same IAV subtype, their sequences were not highly similar, with only 75.56% identity (Figure 3.2.8). There is a run of mismatched sequence between residues 54-58 of the H3N8 proteins that is not consistent with either intracellular localisation as represented by the protein panel used here. This series of polymorphisms is within the second helix for both H3N8 PB1-F2 proteins rather than the long C-terminal helix associated with membrane insertion, so may not impact stability.

After identifying possible structural explanations for varying stability across the protein panel, the possible effect of AUG mutations on protein structure of 5092 (H5N1) and UDL (H9N2) were assessed. As previously described, structural predictions were calculated using the I-TASSER server by submitting the mutated amino acid sequence, and AUG positions highlighted in the resulting models.

5092 ΔAUG 7 retains the most similar predicted structure to 5092 WT PB1-F2 (Figure 3.2.9). Although the small helix is almost entirely disrupted by the mutation, losing the tight parallel structure, the long C-terminal helix has minimal structural change (Figure 3.2.9 B). However, the ΔAUG 9 mutations causes the 5092 structure to resemble more closely that of UDL WT PB1-F2, with the C-terminus forming the globular triangle of smaller helices seen in cytoplasmically localised PB1-F2 proteins (Figure 3.2.9 C).

UDL is also likely to undergo structural changes from the introduction of ΔAUG mutations (Figure 3.2.10). Both ΔAUG 7 and ΔAUG 9 mutations cause disruption to the triangular conformation of helices, causing a large change to the globular structure of the WT protein. Although both ΔAUG 7 and ΔAUG 9 preserve some portion of a C-terminal helix, they show much more disorder in their C-termini than the WT structure (Figure 3.2.10 B, Figure 3.2.10 C). Disordered loops previously linking helices are expanded as helical structures are disrupted.

ΔAUG mutations are predicted to affect the structures of both UDL and 5092 PB1-F2 proteins. Both the proteins saw an increase in disorder when the ΔAUG mutations were introduced. This may be due to the use of a Met>Thr mutation strategy, as Thr can act as a helix breaker (Alymova et al., 2014). Possible functional impacts of the predicted structural changes will be assessed in Chapter 4.

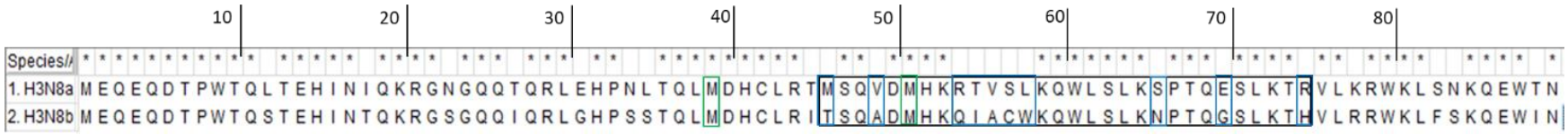

Figure 3.2.8 H3N8 PB1-F2 Alignment
Protein sequences were aligned in MEGA7, with C-terminal AUGs highlighted in green. The black box demarcates aa46-75, which are generally accepted to contain the MTS (James et al., 2019). Polymorphisms within the MTS, which are mostly likely to affect localisation, are highlighted in blue.

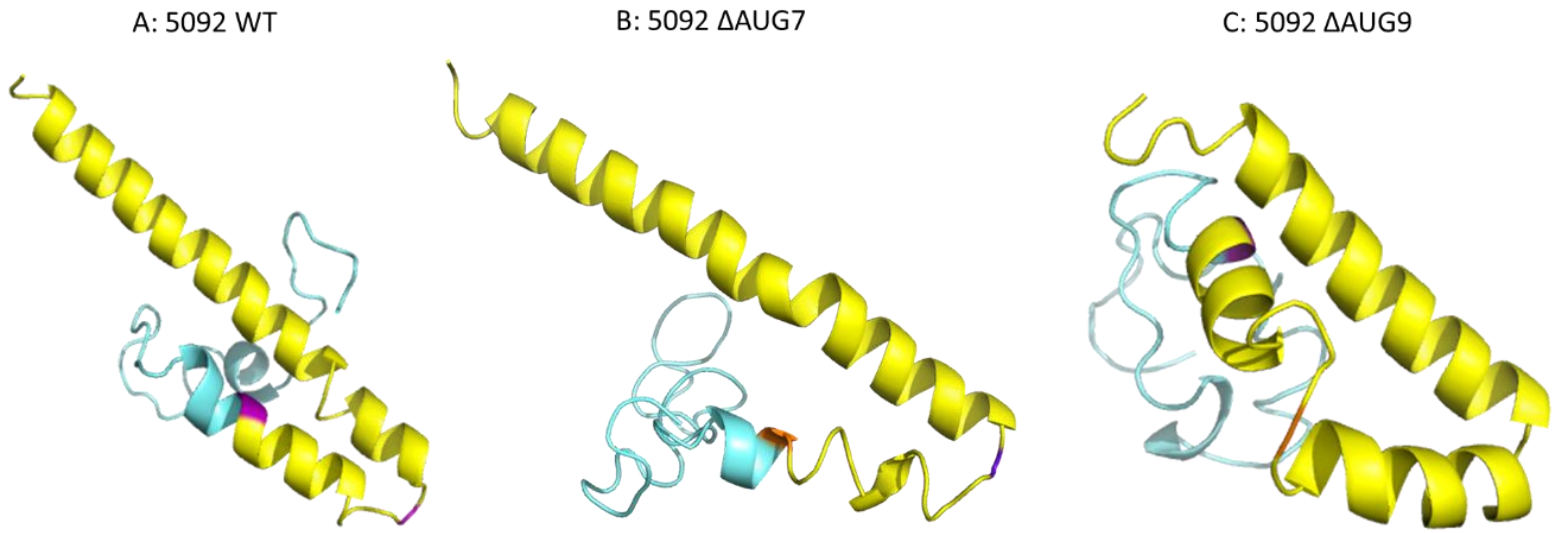

Figure 3.2.9 Effects of ΔAUG Mutations on Predicted Structure of 5092 PB1-F2

Structural predictions of mutant 5092 PB1-F2 proteins from I-TASSER were visualised in Pymol. N-terminus (aa1-38) in blue, C-terminus (aa39-90) in yellow. AUGs present in the sequence are highlighted in purple, AUGs altered by mutagenesis are highlighted in orange.

Figure 3.2.10 Effects of ΔAUG Mutations on Predicted Structure of UDL PB1-F2
Structural predictions of mutant UDL PB1-F2 proteins from I-TASSER were visualised in Pymol. N-terminus (aa1-38) in blue, C-terminus (aa39-90) in yellow. AUGs present in the sequence are highlighted in purple, AUGs removed by mutagenesis highlighted in orange.

Discussion

In this chapter, we assessed the variability of PB1-F2 stability. Previous studies have associated PB1-F2 stability with subcellular localisation (James et al., 2019) or sequence motifs (Alymova et al., 2014; Cheng et al., 2017; James et al., 2019; Košík et al., 2015). We were able to confirm different localisation (Figure 3.2.1) and stability (Figure 3.2.2) for our model proteins, 5092 (H5N1) and UDL (H9N2). This supported the hypothesis that localisation could affect stability (James et al., 2019), possibly through membrane insertion preventing trafficking to the proteasome (Figure 3.2.3). However, when we expanded the stability analysis to additional PB1-F2 proteins to test this hypothesis, we found that not all variation in stability could be explained by localisation alone (Figure 3.2.4). Instead, we find it probable that protein structure is also a factor in determining protein stability.

5092 PB1-F2 showed a punctate localisation that has previously been noted as indicative of mitochondrial association (McAuley et al., 2010), whereas UDL was more diffusely cytoplasmic (Figure 3.2.1). Due to the formation of PB1-F2 inclusion bodies within the cell when proteasome inhibitor MG132 was used, 1µg of PB1-F2 plasmids were transfected and cells were fixed at 24h. We chose a 24h fixation time point as during stability assays, 24hpt was taken as the 0h timepoint for the addition of CHX, and under these conditions PB1-F2 proteins had been visible on western blots. Unfortunately, due to the extreme instability of the H3N8 PB1-F2 proteins, we were unable to confirm their subcellular localisation by confocal microscopy. The proteins were visible within cells, but expression was low, and their localisation was inconclusive (data not shown).

We would have liked to obtain confirmation of localisation for the full PB1-F2 panel (Table 3.2.1) using MitoTracker to quantify colocalization with mitochondria. Despite attempting to optimise this

stain with different concentrations (1:200, 1:500, 1:2000), different batches of DF1 cells, growing cells in different media, and different batches of MitoTracker, as well as attempting immunofluorescent staining with a cytochrome c antibody, we were unable to get usable images of mitochondria. This meant we have had to rely on previously published colocalization analyses (James et al., 2019). We were able to produce similar results to this prior study for our main model PB1-F2s 5092 and UDL, so have had to assume the rest of the panel would also behave similarly to what was reported in James et al, 2019.

CHX assays show decay from a steady state of protein production and degradation (Eldeeb et al., 2019). The compound occupies the E-site of the 60S subunit of the ribosome to prevent protein elongation. If the protein has not accumulated sufficiently to be visible on a western blot at 0hr then its subsequent degradation cannot be measured with this technique. Not every PB1-F2 protein described in James et al, 2019 could be assessed with the CHX assay. Within the PB1-F2 panel assessed there was a lack of mitochondrially localised proteins. As a result, we have less stability data for mitochondrial PB1-F2s than for cytoplasmic proteins, and the data we do have is highly variable.

The 5092 PB1-F2 is relatively stable, with a mean half-life of 41 minutes, whereas H3N8a has a mean half-life of 7 minutes (Table 3.2.2). This was unexpected, as our hypothesis had been that mitochondrial localisation would result in membrane insertion to protect the proteins against cellular degradation mechanisms. Finding that both H3N8 subtype PB1-F2s were highly unstable suggested that there was a subtype-specific factor affecting stability. If there is a H3N8 specific stability motif, it is not obvious in the sequences (Figure 3.2.8), as the H3N8 PB1-F2s did not show high sequence identity to one another. However, we cannot rule out that the PB1 segments came from different sources through reassortment (Lu et al., 2014). The CHX assays revealed a range of protein stabilities, with the H3N8 PB1-F2s least stable, UDL and H11N9 PB1-F2s showing very similar

degradation patterns, and 5092 the most stable of the panel (Table 3.2.2). The similarity of degradation curves between UDL and H11N9 (Figure 3.2.4 B, Table 3.2.2), both reported to have cytoplasmic localisations, suggests that localisation is not entirely separate from stability. UDL, H11N9, and both H3N8 PB1-F2's mean half-lives appear to overlap within a single standard deviation of each other (Table 3.2.2), indicating that stability of PB1-F2 is more likely to be a gradient than a sharply defined property of the protein.

We have hypothesised that cytoplasmic and mitochondrially localised PB1-F2 proteins may have different structures that can impact their stability (Figure 3.2.6, Figure 3.2.7). We used I-TASSER to generate structural predictions. I-TASSER uses an iterative threading technique to predict a structure for the query based on comparison with already solved structures in the PDB structure library (Roy et al., 2010). There is only a single partial crystal structure of PR8 PB1-F2 in the PDB database (Bruns et al., 2007), although structural predictions (Pasricha et al., 2018) and some NMR spectroscopy have been performed (Solbak et al., 2013). Most of this analysis has focused on the length and position of α helices rather than producing a 3D structural model, so the orientation of helices in the C-terminus is mostly unknown.

I-TASSER uses a C-score based on the quality of threading alignments and the convergence (cluster density) of the I-TASSER simulations to measure prediction accuracy. C-score ranges from [-5, 2], with a higher score being an indication of a better quality model (Roy et al., 2010). The C-scores for the models shown in Figures 3.2.5, 3.2.6, and 3.2.7 were in the range of [-2.60, -1.86] with H3N8b PB1-F2 being the most confident (-1.86), and 5092 (-2.56) and H3N8a (-2.60) being the least confident. This is likely due to the lack of good templates in the PDB library, although I-TASSER has shown improvements in modelling proteins with few templates over other modelling tools (Roy et al., 2010). The predictions were made before

AlphaFold became available. It would be interesting to see if the new gold standard gave similar predictions. As a control, partial 5092 and UDL sequences the same length as the partial PR8 crystal structure were submitted to I-TASSER (data not shown). For these short sequences, the PR8 structure was one of the top ten templates used in the prediction, and the C-scores improved dramatically from -2.56 full length to 0.54 for short 5092 and -2.39 for full length to 0.61 for short UDL. This indicates that I-TASSER is capable of producing high quality structural predictions for PB1-F2 proteins with the appropriate templates.

Due to the low C-scores of our predictions, we cannot confirm that structural differences between PB1-F2 proteins with different subcellular localisations will look like the models presented here. However, it is highly likely that cellular localisation will have an impact on PB1-F2 structure (Thornton et al., 1999; Uversky, 2019). PB1-F2 structure has previously been seen to be affected by a membranous environment (Chevalier et al., 2010). Proteins with different cellular environments and interacting partners are likely to have different structures as a consequence of differing functional requirements (James et al., 2019; Thornton et al., 1999).

More structural information of full length PB1-F2 proteins of different subtypes is needed to help understand not only stability, but varying functionality of the proteins. Unfortunately, obtaining crystal structures of the PB1-F2 panel used here was outside the scope of this project. Protein structure and function are tightly related, as structure helps determine binding affinity for protein-protein interactions.

Chapter 4: Phenotypes Induced by PB1-F2 C-terminus Expression

Introduction

PB1-F2 function is known to be viral strain and host dependent (Deventhiran et al., 2016; James et al., 2019; Leymarie et al., 2014; Mettier et al., 2021). This specificity includes varying phenotypes of PB1-F2 proteins in reassortant viruses with different backbones, suggesting that simple reassortment will not transmit duplicate PB1-F2 functions (Mettier et al., 2021). PB1-F2 is known to function as an immune antagonist in avian hosts, but may be either ineffective or detrimental to the virus as a full length protein in a mammalian host, which may be indicative of host-specific evolution (Kamal et al., 2018; Pasricha et al., 2013).

PB1-F2 Modulation of Viral Polymerase Activity

Evidence on the ability of PB1-F2 to modulate the polymerase activity is an ongoing question, with contradictory results reported. At least three mechanisms have been proposed: direct effects of selection pressures acting on PB1-F2 resulting in changes to PB1 sequence; indirect effects on expression of the PB1-related PB1-N40 accessory protein; and direct effects arising from interactions between the PB1-F2 protein and the viral polymerase complex

PB1-F2 is expressed from the +1 reading frame of segment 2, which also expresses the viral polymerase subunit PB1 (Chen et al., 2001). Due to the overlapping reading frames, mutations in PB1-F2 can impact the PB1 sequence. However, study of the molecular characteristics of H7N9 PB1-F2 proteins showed that within PB1-F2 codons the first and second positions had significantly higher evolutionary rates than the third position. The third position in a codon normally allows synonymous mutations, but in this case it allows the PB1 ORF to maintain its sequence (Wei et al., 2015). Due to the

overlapping open reading frames in segment 2, codon position 2 of PB1-F2 overlaps with position 3 of PB1. The offset ORFs allow a nonsynonymous change at position 2 within PB1-F2 to be silent in the PB1 ORF (Figure 4.1). Silent mutations in the PB1 sequence are less likely to induce changes in polymerase function.

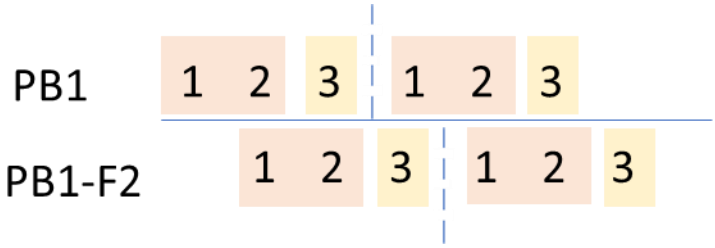

Figure 4.1 Overlapping Open Reading Frames
Overlapping ORFs of PB1 and PB1-F2 within segment 2, showing codons and nucleotide positions. Nonsynonymous codon positions highlighted in orange; synonymous codon positions highlighted in yellow.

However, analysis is made complicated by the interdependent nature of segment 2 polypeptide expression (Wise et al., 2009), and use of polymerase segments from different viruses to study a strain specific protein (Mettier et al., 2021). Wise et al. found that mutating the fourth AUG codon of segment 2 (AUG 4) to prevent PB1-F2 expression from its first AUG resulted in an increase in PB1-N40 accumulation, but nonsignificant changes in polymerase activity. However, this work was done in the PR8 strain of virus, where PB1-F2 activity is minimal and ablation of expression has no effect on replication or immune antagonism (Wise et al., 2009; Zamarin et al., 2006). As PB1-F2 function is known to be highly strain specific, conflation of changes in polymerase activity, or viral replication rate resulting from changes in expression of either PB1-F2 or PB1-N40 cannot be ruled out.

Mazur et al, using a ΔAUG 4 STOP12 double knock out strategy, reported protein-protein interactions between PB1-F2 and PB1, including an effect on PB1 localisation at late timepoints (Mazur et al., 2008). When Wise et al. investigated this phenomenon, they only saw

a change in PB1 localisation when the PB1-F2 ΔAUG 4 had affected PB1-N40 expression and not with a STOP9 strategy (Wise et al., 2009). These authors proposed that the apparent change in PB1 localisation was in fact detection of upregulated expression of the cytoplasmic PB1-N40 protein.

McAuley et al, also using the double knock out strategy, reported a lack of changes to polymerase activity through altered PB1-N40 expression levels, but found a modest delay and decrease in PB1 accumulation when a ΔAUG 4 mutation was introduced. They also found colocalization between PB1 and PB1-F2 was strain specific and not confined to the nucleus (McAuley et al., 2010). Viral polymerase activity can be affected by PB1-F2 expression in a cell and strain dependent manner, but polymerase phenotypes do not always translate into changes in overall viral replication kinetics (Chen et al., 2010; McAuley et al., 2010). Any effect of PB1-F2 on polymerase activity is likely to be small, but this is hard to confirm, as there are few studies that assess polymerase activity using a knock out strategy that does not affect PB1-N40 expression (Tauber et al., 2012).

PB1-F2 and Viral Replication Kinetics

As an accessory protein, PB1-F2 is non-essential for virus replication (Chen et al., 2001). Although some PB1-F2 sequence polymorphisms (i.e. N66S) have been seen to affect virus growth kinetics (Schmolke et al., 2011), PB1-F2's effect on viral growth appears to be limited and strain specific. Several studies report a difference in viral pathogenesis, including weight loss in mice, with no change in replication kinetics in cell lines (Alymova et al., 2018; Conenello et al., 2007; Kamal et al., 2015; Mettier et al., 2021; Zamarin et al., 2006).

The PB1-F2 N66S polymorphism is known to increase virulence in mouse models (Conenello et al., 2007; Varga et al., 2011), and a virus with this change can reach significantly higher titres than WT PB1-F2 viruses in a multi-cycle growth curve on mouse lung epithelial cell line

LA-4 (Schmolke et al., 2011). However, in A549 cells and the swine alveolar macrophage cell line 3D4/31, or duck embryonic fibroblasts, N66S caused no significant changes in replication (Deventhiran et al., 2016; Schmolke et al., 2011). The effect of PB1-F2 N66S has been attributed to its impact on viral pathogenesis allowing uncontrolled viral replication rather than a direct alteration of virus replication kinetics (Conenello et al., 2011). PB1-F2 is well known to function differently in avian and mammalian hosts. N66S' effect on replication kinetics is likely to be host specific, and limited to mammals (Deventhiran et al., 2016).

PB1-F2 and Immune Antagonism

PB1-F2 was initially described as a promoter of apoptosis (Chen et al., 2001), but as our understanding of the protein has developed, more potential functions have been revealed. Several immune antagonism methods of PB1-F2 are associated with mitochondria (Kamal et al., 2018; Xiao et al., 2020; Yoshizumi et al., 2014; Zamarin et al., 2005).

PB1-F2 proteins from multiple subtypes have been found to target MAVS, an immune adaptor protein that promotes production of interferon (Seth et al., 2005; Varga et al., 2011). Inhibition of this by preventing MAVS interactions with other proteins (Cheung, Ye, et al., 2020) or encouraging degradation of host immune factors (Cheung, Lee, et al., 2020; R. Wang et al., 2020) can prevent a robust interferon-based anti-viral response. Unfortunately, both recent studies from Cheung et al. used a ΔAUG 4 and STOPs knock out strategy and do not detail the placement of their stop codons. Several reports of PB1-F2 innate antagonism have focused on pandemic risks, so there is limited information on activities of avian PB1-F2 proteins in avian cells (Park et al., 2019; Pinar et al., 2017). A recent study of avian H5N1 and H9N2 PB1-F2 proteins in primary chicken cells found that while both proteins acted as antagonists to expression of interferon stimulated genes (ISGs), virus strain determined the specialism of

PB1-F2 immune antagonism (James et al., 2019). The mitochondrially localised H5N1 PB1-F2 was a stronger inhibitor of the IFNβ promoter, probably due to increased colocalization with MAVS. The H9N2 PB1-F2 was cytoplasmically localised, and interacted more strongly with IκκB. PB1-F2 binding IκκB may prevent phosphorylation of IκB, leading to inhibition of the NFκB response (James et al., 2019).

James et al. identify alterations to the sequence of the C-terminus of PB1-F2 as having the greatest impact on immune antagonism (James et al., 2019), but there has been no assessment on the impact of removing expression of C-terminal fragments of PB1-F2 by mutating AUGs 7-9 on overall PB1-F2 function. Here, using this strategy, we assess PB1-F2's impact on virus replication kinetics without affecting PB1-N40 expression, and test how the balance of PB1-F2 expression affects immune antagonism. We wanted to explore whether the non-synonymous (in PB1-F2) mutations introduced by removing PB1-F2 C-terminal AUGs affected the function of full length PB1-F2, and how alterations in immune antagonism may affect viral replication. The aim of this chapter was to determine whether the full ORF is needed for PB1-F2 to be functional, or if a C-terminal fragment containing functional motifs is sufficient.

In this chapter, we focused on possible phenotypes within polymerase activity, replication kinetics, and innate immune antagonism.

Results

Effect of PB1-F2 Mutations on Reassortant Virus Replication Kinetics

To assess the impact of PB1-F2 mutations in a viral context, reassortant viruses were made with WT or mutant H5N1 5092 segment 2 in an otherwise H1N1 PR8 backbone. 5092 is a highly pathogenic virus that requires high containment to study safely. PR8 is a mouse-adapted lab strain frequently used for PB1-F2 research, so there are data for comparison. These reassortant viruses allowed study of 5092 PB1-F2 in BSL2 conditions. We rescued viruses that had common avian PB1-F2 N-terminal truncations introduced (STOP12, STOP25, STOP35), and inserted (+AUG 8) or mutated C-terminal AUG codons (ΔAUG 7, ΔAUG 9, ΔAUG 7,9). The STOP12 ΔAUG 9 double mutant was created to allow PB1-F2 expression from only AUG 7, which was the best conserved C-terminal AUG codon according the bioinformatic analysis presented in Chapter 2.

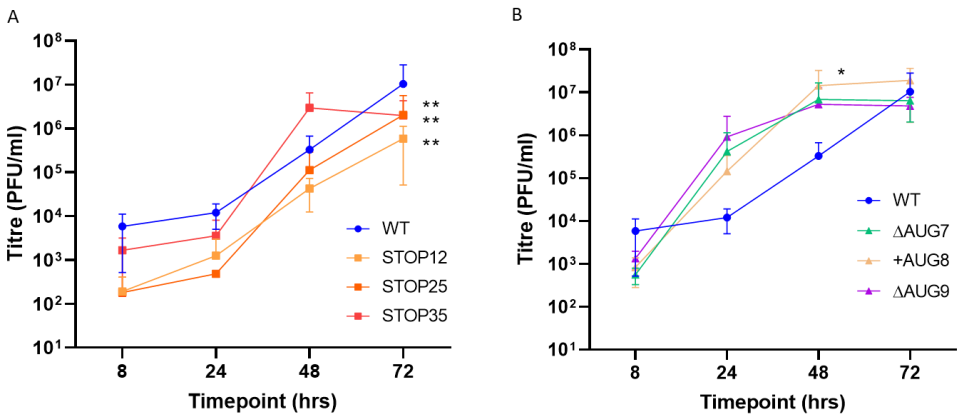

Figure 4.2 Effect of PB1-F2 Mutations on Viral Replication in MDCK cells Cells were inoculated with an MOI 0.001 of either STOP (Panel A) or AUG (Panel B) mutants, in three independent experiments. All mutants were 7:1 reassortants, with 7 PR8 segments and the appropriate 5092 PB1 segment. Supernatants were collected at the relevant timepoint and titrated by plaque assay on MDCK cells. Data are mean +/- SD, analysed by two way ANOVA with multiple comparisons against WT as the control condition (*: $p<0.05$, **: $p<0.01$).

Initial multi-cycle growth curves were performed in MDCKs, a cell line highly permissive to IAV infection (Figure 4.2). STOP mutants (Figure 4.2 A) were not significantly different compared to WT virus in a two way ANOVA between 8 and 48h timepoints. However, at 72h, all three mutants had significantly lower titres, with p values for STOP12, 25, and 35 being 0.0018, 0.0091, and 0.0084 respectively. The AUG mutants (Figure 4.2 B) had noticeably different kinetics of replication than the WT virus, but differences were nonsignificant by two-way ANOVA except for the 48h timepoint with the +AUG 8 mutant, which had a p value of 0.0150. The AUG mutants showed more rapid growth kinetics at 24h and 48h but did not have higher titres than WT virus at 72h. Significance at late timepoints suggests the STOP mutants were affecting the host innate response or cell viability, rather than a consequence of affecting viral replication directly, i.e., through polymerase activity.

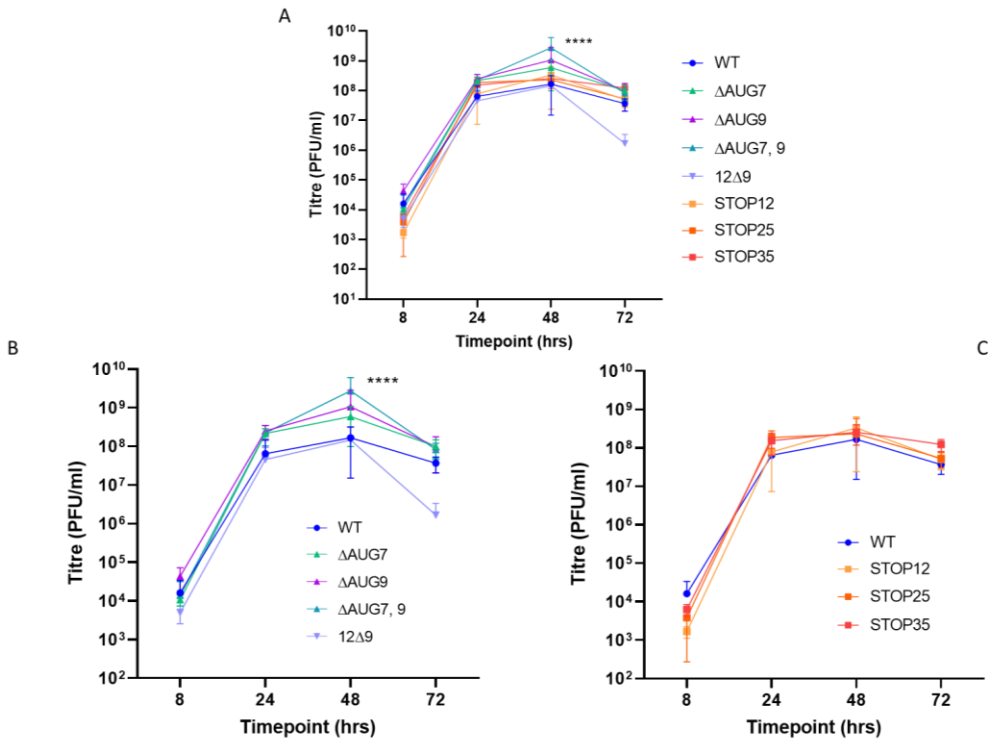

Figure 4.3 Effect of PB1-F2 Mutations on Viral Replication *in ovo*

The allantoic cavities of 10-day-old embryonated chicken eggs were

inoculated with 100 PFU of virus. All mutants were 7:1 reassortants, with 7 PR8 segments and the appropriate 5092 PB1 segment. Each virus was inoculated in three eggs for technical repeats. Eggs were refrigerated under schedule 1 at the appropriate timepoint, allantoic fluid harvested and titrated by plaque assay on MDCK cells. Data are mean +/- SD, analysed by two-way ANOVA with multiple comparisons against WT as the control condition (****: $p<0.0001$). Panel A shows the full panel, which was separated into AUG mutants and STOP mutants in panels B and C for clarity.

5092 is an avian virus, and PB1-F2 function is known to be host specific, so growth curves were performed in eggs as an avian model (Figure 4.3). Despite apparent increases in titre of ΔAUG 9 above WT at all timepoints, and ΔAUG 7 and ΔAUG 7,9 from 24h onwards, the only AUG mutant that showed a significant difference compared to WT by two-way ANOVA was the ΔAUG 7, 9 double mutant at 48h (Figure 4.3 B). In contrast to the MDCK cell growth curve (Figure 4.2), in eggs the STOP mutants were not significantly different compared to WT at any timepoint (Figure 4.3 C). During the growth curve, eggs infected with ΔAUG 7 and ΔAUG 9 appeared unhealthy, with loss of eggshell veins and lack of embryo movement, when candled at 48h and 72h, but this did not seem to have impacted virus titre. In an avian system, introduction of ΔAUG mutations resulted in higher titres than loss of full-length expression by N-terminal truncation.

Effect of PB1-F2 Mutations on Viral Polymerase Activity

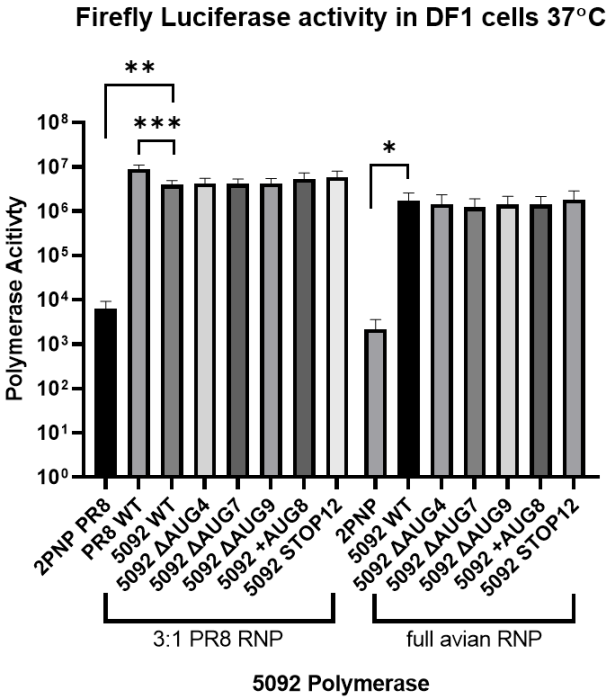

Figure 4.4 Effect of ΔAUG mutations on polymerase activity
Chicken DF1 cells were transfected with PA, PB2, and NP polymerase subunit expression plasmids from either PR8 or 5092 viruses. A 5092 segment 2 (PB1) of interest, and a firefly luciferase reporter under the control of an influenza promoter were also transfected. Assays were performed in triplicate to n=4. Each triplicate read was used to calculate a mean which was then analysed by one-way ANOVA with multiple comparisons against 5092 WT control for each polymerase context (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

To test whether introduction of ΔAUG mutations affected IAV polymerase activity, either a reassortant RNP that matched the viruses used in the growth curves (PR8 NP, PA, PB2 with 5092 PB1), or a fully avian 5092 H5N1 RNP were reconstituted in DF1 cells alongside a reporter construct where firefly luciferase transcription is driven by the influenza virus noncoding regions that act as a promoter for the viral polymerase (Figure 4.4). In both contexts, 5092 WT showed a significant increase in luciferase activity compared to a “2PNP” negative control, which was transfected with plasmids expressing the

luciferase reporter and subunits NP, PA, PB2, but lacking PB1. None of the AUG mutants, or STOP12, showed any significant changes in activity as either a reassortant or avian polymerase.

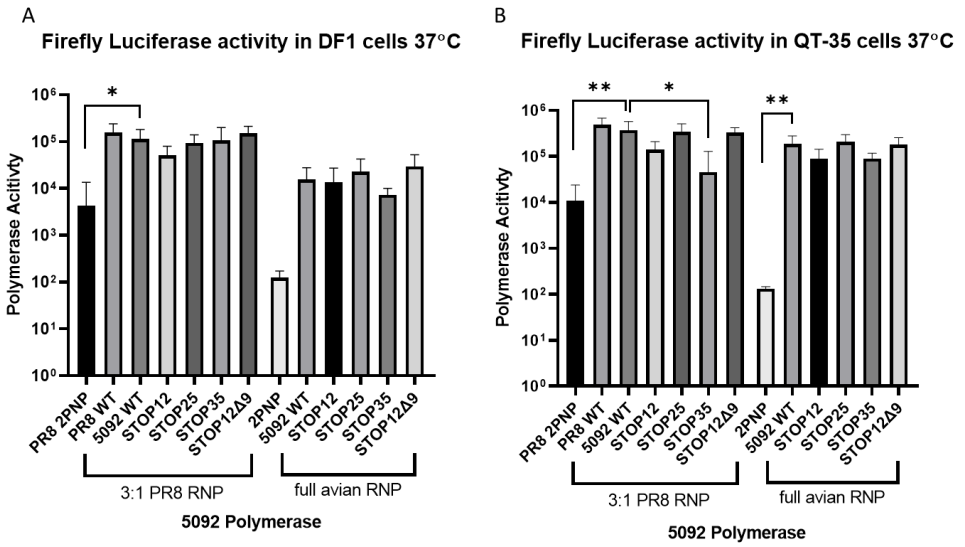

Figure 4.5 Effect of STOP codons on polymerase activity
Chicken DF1 and quail QT-35 cells were transfected with PA, PB2, and NP polymerase subunit expression plasmids from either PR8 or 5092 viruses. A 5092 segment 2 (PB1) of interest, and a firefly luciferase reporter under the control of an influenza promoter were also transfected. Assays were done in triplicate to n=4. Each triplicate read was used to calculate a mean which were then analysed by one way ANOVA with multiple comparisons against WT as a control for each polymerase context. (*: $p<0.05$, **: $p<0.01$).

Following the lack of significant impact on polymerase activity by AUG mutants, the analysis was expanded to the STOP mutants (Figure 4.5). In DF1s, we saw no significant changes in polymerase activity for any of the STOP mutants. Polymerase activity was also assessed in QT35 quail fibroblasts (Figure 4.5 B). STOP35 was the only mutant which showed a significant change in activity, with a p value of 0.012.

Changes in polymerase activity do not explain the significant changes in overall virus replication kinetics. In MDCK cells (Figure 4.2), all three

STOP codons showed significant differences in titre compared to WT at 72h, while only STOP35 had a significant effect on polymerase activity (Figure 4.5B). None of the ΔAUG mutants showed any significant changes in polymerase activity (Figure 4.4), although some showed a significant change in viral titre compared to WT (Figure 4.2). As PB1-F2 is known to act as an innate immune antagonist, we next investigated its effect on IFN promoter activity in avian cells.

Effect of PB1-F2 Mutations on Innate Immune Antagonism

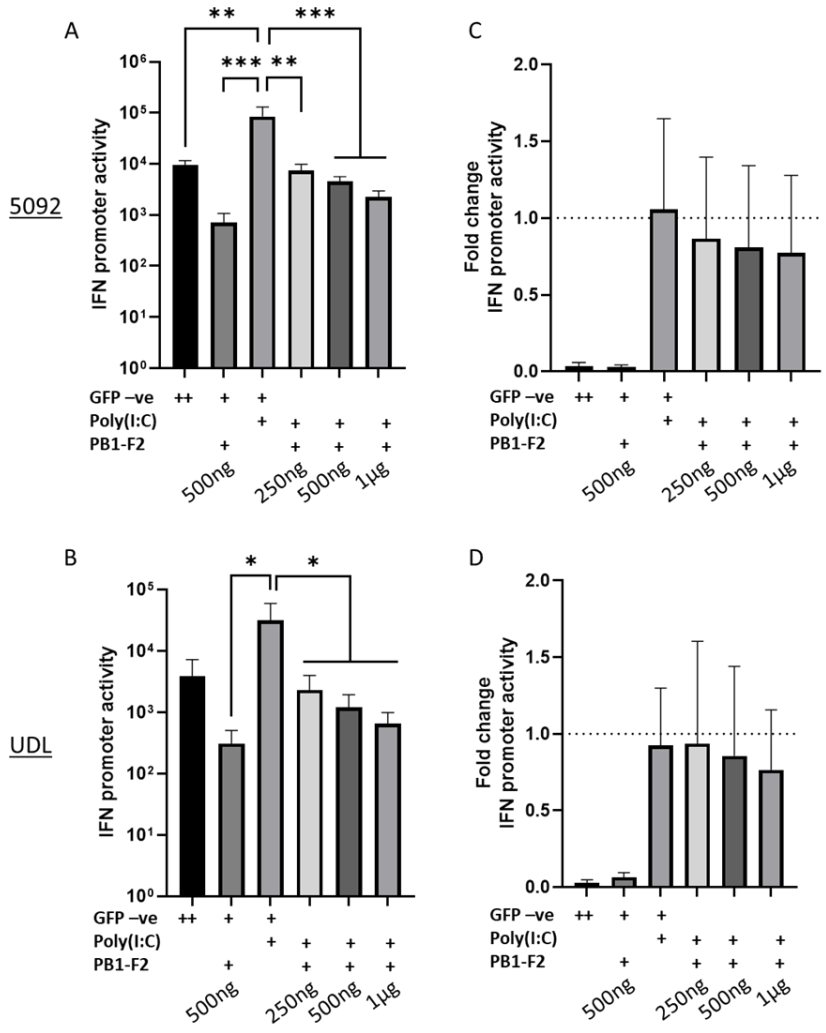

Figure 4.6 Dose dependency of PB1-F2 Interferon Antagonism.
DF1 cells were transfected with plasmids encoding a luciferase reporter gene under control of the chicken IFNβ promoter, a Renilla luciferase gene under the control of a constitutive RNA polymerase II promoter, and either

GFP tagged PB1-F2 variants (Panels A, C, 5092; Panels B, D UDL) or empty GFP vector as a negative control. At 24hpt, cells were stimulated with either poly(I:C) or empty GFP vector as a negative control. 16hrs post stimulation, cells were lysed, and luciferase activity measured. Panel A, B: IFN promoter activity, measured with firefly luciferase. Panel C, D: Fold change of IFN promoter activity. Firefly luciferase activity was divided by Renilla luciferase activity to normalise IFN promoter activity to background cellular expression levels. N=3, each replicate performed in triplicate and mean calculated. Mean IFN promoter activity was analysed by one way ANOVA with multiple comparisons to the PB1-F2 -ve control (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$).

To find conditions under which poly(I:C) gave strong IFN promoter induction, and PB1-F2 was able to show significant inhibition, DF1s were transfected with three different concentrations of 5092 (H5N1) or UDL (H9N2) PB1-F2 plasmids. Cells were stimulated with poly(I:C) to induce the chicken IFNβ promoter, and firefly luciferase activity read as a reporter (Figure 4.6 A, B). Transfection with poly(I:C) strongly induced IFN promoter activity above non-stimulated controls. The IFN response was then significantly reduced by both PB1-F2 proteins, with small non-significant trends towards increased antagonism with higher doses of plasmid by UDL PB1-F2. 5092 PB1-F2 showed less significant antagonism by 250ng PB1-F2 than other conditions (Figure 4.6A). However, when the IFN response was normalised to the constitutive Renilla activity (Figure 4.6 C, D), there was no significant antagonism by either PB1-F2 protein, so PB1-F2 inhibition may not be specific to the IFN promoter. Due to the strong induction by poly(I:C) and lack of significant PB1-F2 dose dependency (Figure 4.6 A, B), we continued our investigations using 250ng poly(I:C) and 500ng PB1-F2.

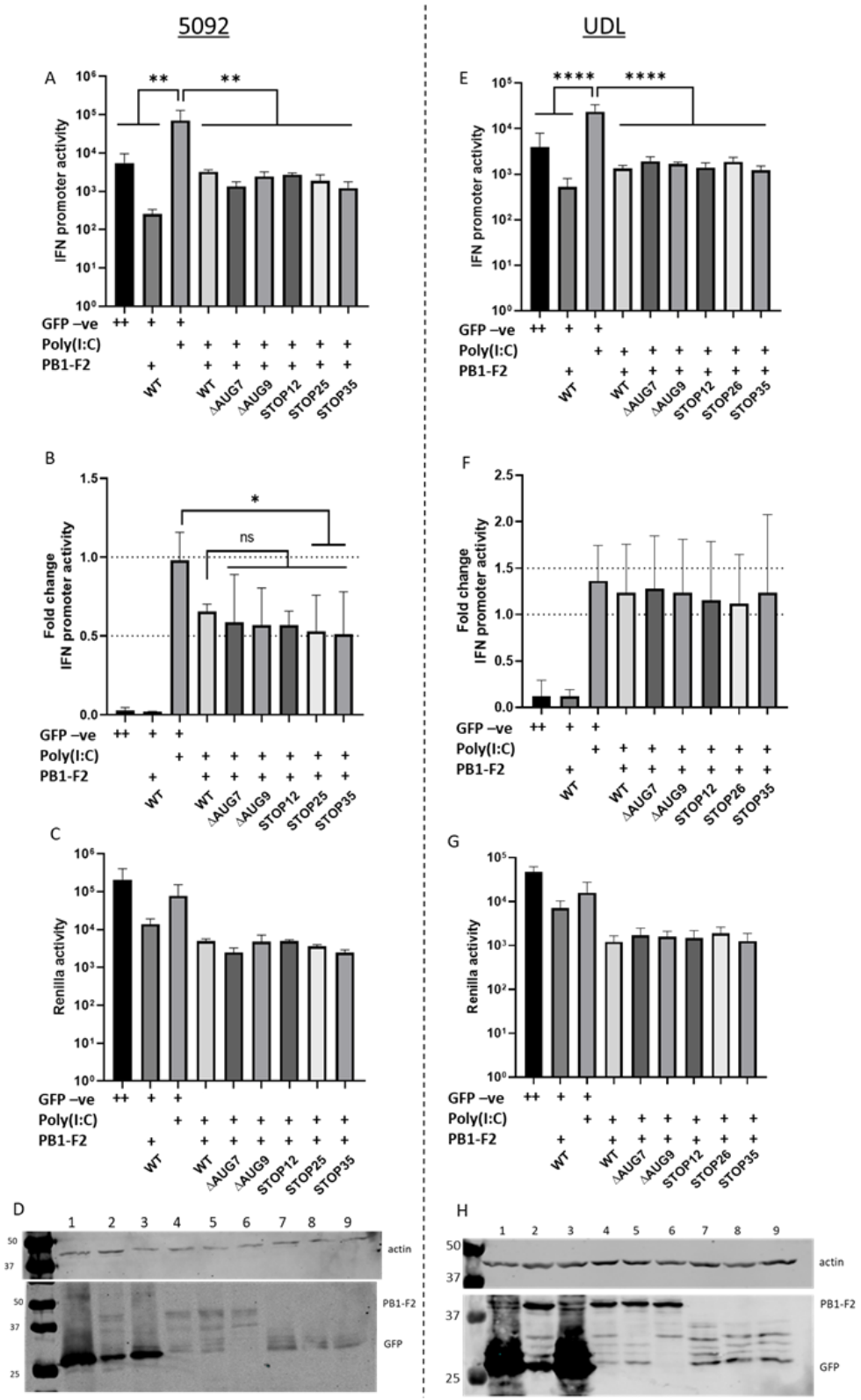

Figure 4.7 Interferon Antagonism of PB1-F2 Mutants

DF1 cells were transfected with luciferase reporters and either a PB1-F2 mutant (5092 A-D, UDL E-H) or empty GFP vector, then stimulated with

poly(I:C). Panel A, E: IFN promoter activity, measured with firefly luciferase. Panel B, F: Fold change of IFN promoter activity, normalised to constitutive Renilla activity as above. Panel C, G: Renilla luciferase activity. Panel D: western blot of 5092 cell lysates. Lane order; 1, GFP ++; 2 PB1-F2 + GFP +; 3, GFP + poly(I:C) +; 4, WT; 5, ΔAUG 7; 6, ΔAUG 9; 7, STOP12; 8, STOP25; 9, STOP35. Panel H: western blot of cell lysates, lysed in SDS lysis buffer due to difficulty visualising UDL from passive lysis buffer. Lane order; 1, GFP ++; 2 PB1-F2 + GFP +; 3, GFP + poly(I:C) +; 4, WT; 5, ΔAUG 7; 6, ΔAUG 9; 7, STOP12; 8, STOP25; 9, STOP35. Luciferase activity was read in triplicate to n=3 5092 and n=4 UDL, and the means analysed by one way ANOVA compared to either PB1-F2 -ve or WT PB1-F2 conditions as control (*: $p<0.05$,**: $p<0.01$, ****: $p<0.0001$).

We then focused on the effect of AUG and STOP mutations on PB1-F2 inhibitory function. DF1 cells were transfected with 500ng PB1-F2 plasmids and the previously described luciferase reporters, then stimulated with poly(I:C) as above. When analysing only IFN promoter activity by measuring firefly luciferase activity (Figure 4.7A), all 5092 PB1-F2 variants significantly antagonised the IFN response compared to the PB1-F2 negative condition. When comparing mutants to the WT PB1-F2, none were significant (Figure 4.7A).

However, when normalised activity of the mutants were compared to the PB1-F2 -ve condition, only STOP25 and STOP35 significantly reduced activity, with p values of 0.0483 and 0.0370 respectively. This was unexpected considering the strong inhibition shown by all the PB1-F2 proteins in panel A. Aliquots of the lysed samples were also used for western blots to confirm PB1-F2-GFP expression (Figure 4.7 D). Overall, the data show that C-terminal 5092 PB1-F2 expression was sufficient to antagonise the IFN promoter when full-length PB1-F2 was not expressed.

IFN antagonism by UDL PB1-F2 was also assessed in the same manner (Figure 4.7 E-H). Analysing IFN promoter activity alone showed significant inhibition of the IFN response by all PB1-F2

proteins tested (Figure 4.7 E). Unlike 5092, normalising Firefly luciferase activity to the constitutive Renilla values did not show any specific inhibition of the immune response by UDL PB1-F2 proteins (Figure 4.7 F). Due to UDL PB1-F2's lower stability, cells were transfected as previously but had to be lysed in SDS lysis buffer to confirm PB1-F2-GFP expression (Figure 4.7 H). 5092 PB1-F2 appears to be a more effective IFN antagonist than UDL PB1-F2, which did not show specific IFNβ promoter inhibition. Neither PB1-F2 protein required expression of the full-length protein for inhibitory activity supporting the hypothesis that the C-terminal region of the protein is functionally significant.

Discussion

This is the first time a study on the impact of PB1-F2 C-terminal expression has been performed. Previous works have focused on the protein's immune antagonism functions over polymerase activity or viral replication kinetics (Alymova et al., 2018; Chen et al., 2001; Kamal et al., 2015). We are also the first to examine the effect of individual ΔAUG mutations, as previous studies have accepted or removed AUGs 7-9 of PB1-F2 as a grouping rather than as individual sources of translation initiation (Kamal et al., 2015; Leymarie et al., 2013; Zamarin et al., 2006).

In this chapter, we found that PB1-F2 effects on viral replication kinetics appear at later timepoints, implying that this is due to a host-pathogen interaction rather than a direct impact of PB1-F2 on viral polymerase activity or replication kinetics (Figure 4.2). Altered polymerase activity is unlikely to explain the change in replication kinetics seen here. Previous studies show viral polymerase activity can affect replication rate (Wanitchang et al., 2010) and inhibition of polymerase activity begins to affect viral growth within 12hpi at MOI 0.02 (Ghanem et al., 2007). STOP35 did show a significant decrease in activity in QT-35 cells, but it was the only mutant to impact

polymerase activity, compared to multiple significant impacts on viral replication (Figure 4.2, Figure 4.5). While for 5092 C-terminus expression without full length PB1-F2 was sufficient to antagonise the IFN promoter, including when normalised to Renilla activity, UDL did not show any specific inhibition of the IFN promoter over the constitutive Renilla luciferase (Figure 4.7). As previously reported (Cheung, Lee, et al., 2020; James et al., 2019; McAuley et al., 2010), several phenotypes of PB1-F2 were subtype or cell line specific.

In MDCKs, all three 5092 (H5N1) STOP mutants showed significantly lower titres than WT at 72hpi by at least tenfold, while +AUG 8 showed a significantly higher titre than WT at 48h ($1.4x10^7$ vs $3.3x10^5$ PFU/ml) (Figure 4.2). We performed a growth curve in MDCKs because they are a highly permissive cell line and an established IAV plaque assay system. MDCKs are susceptible to IAV strains from multiple lineages, including avian strains. A homogenous population of susceptible cells allows consistent plaque production (Gaush & Smith, 1968; Tobita et al., 1975). Unfortunately, they are a mammalian cell line and PB1-F2 is known to have host and cell specific effects (Chang et al., 2015; Chen et al., 2010; Krumbholz et al., 2011). To our knowledge, we are the first to assess the impact of individual internal start and stop codons of PB1-F2 on viral replication kinetics. Perhaps due to PB1-F2's status as an accessory protein with immune antagonism functions, there have been comparatively few studies assessing its effect on viral replication kinetics. However, our results were somewhat surprising, as previous studies had not seen a phenotype for a recombinant WSN virus encoding PR8 PB1, with a PB1-F2 STOP9 mutant in MDCK, MDBK or eggs (Chen et al., 2001), or had reported a consistent decrease in titre from 12-48h for a STOP12 STOP58 double mutant of WSN in MDCKs (Tauber et al., 2012). In contrast, our STOP12 single mutant of a reassortant H5N1-PR8 virus only showed a significant decrease in titre at 72h. As we were removing fragments of a non-essential protein, we did not expect a

significant impact on viral replication in vitro, although introducing a STOP12 mutation into UDL did reduce the transmission window in infected chickens (James et al., 2016).

In eggs, an avian system, no STOP mutants displayed significantly different replication kinetics at any timepoint (Figure 4.3). The double mutant ΔAUG 7, 9 showed a significantly higher titre than WT at 48h ($2.75x10^9$ vs $1.67x10^8$ PFU/ml), although titres for all three ΔAUG mutants could be seen to increase from 24hpi. Despite a very low titre at 72h, the STOP12 ΔAUG 9 mutant was not significant at any timepoint. ΔAUG 7, 9 reached higher titres than either single ΔAUG mutant at 48h (ΔAUG 7 $5.92x10^8$ PFU/ml, ΔAUG 9 $1.06x10^9$ PFU/ml, ΔAUG 7, 9 $2.75x10^9$ PFU/ml), suggesting the effect of C-terminus fragment loss is cumulative. It is important to note that the ΔAUG mutants are still able to produce full length PB1-F2 protein, albeit mutated at the various methionine codons. The effect of ΔAUG mutations may be to do with the volume of C-terminal fragments within the cell. The N-terminal region of PB1-F2 is thought to be disordered (Bruns et al., 2007), and has been associated with the formation of amyloid fibres (Ajjaji et al., 2016; Chevalier et al., 2010). It is possible that without the N-terminus, the C-terminal fragments can more effectively bind interaction partners, or can interact with a wider range of proteins due to their smaller size. Loss of the disordered N-terminal may also affect how efficient the MTS is, or allow easier trafficking of a more soluble protein within the cell. In both UDL and 5092, at least one STOP codon results in increased expression of C-terminal fragments over the WT (Chapter 3, Figure 3.1.6).

Polymerase assays were performed to confirm that our PB1-F2 mutants, the majority of which were silent in PB1, were not affecting polymerase activity. Possible interactions between PB1 and PB1-F2 are a subject of debate (Košík et al., 2011; Mazur et al., 2008; McAuley et al., 2010; Wise et al., 2009). Interestingly, the one mutation that was not able to be introduced silently in the PB1 ORF was STOP25, which

did not significantly alter polymerase activity in any assay performed. STOP25 causes a T56I mutation in PB1, which has not previously been investigated. T56I has been seen in wild isolates with a STOP25 in PB1-F2, but is a rarer polymorphism than T56A or T56K in samples of H9 or H5N1 sequences within the NCBI IVR between 2003-2020 (data not shown). Residue 56 of PB1 is within the finger region of the right-handed RdRp fold, but does not appear to be in any fingertip loops (Pflug et al., 2014). Polymerase assays in DF1 cells showed no effect of ΔAUG mutants (Figure 4.4). Although one STOP mutant did have significant impact on polymerase activity in QT-35 (Figure 4.5), this was unlikely to be responsible for the significant change in titre at 72hpi in MDCKs seen in Figure 4.2, as a defect in the polymerase would likely affect replication kinetics at multiple earlier timepoints (Muratore et al., 2012).

However, most polymerase activity studies have been performed with PR8 or WSN in mammalian 293T cells (Tauber et al., 2012; Wise et al., 2009), whereas our assay was performed with an avian 5092 PB1 segment in avian DF1 and QT35 cells. As PB1-F2 effects on polymerase activity have previously been reported as strain and cell type specific (McAuley et al., 2010), and *in vivo* experiments have shown strain-dependent phenotypes of PB1-F2 (James et al., 2019) this could account for variation between our results and others.

PB1-F2 is well known as an immune antagonist (Kamal et al., 2015; Xiao et al., 2020; Zamarin et al., 2006). We used a plasmid-based system to assess the IFN antagonism of our model viruses 5092 (H5N1) and UDL (H9N2). Poly(I:C) stimulation has previously been seen to induce similar immune responses to H5N1 infection in chickens (Karpala et al., 2008). The dsRNA analogue is recognised by MDA5, which then binds MAVS to trigger transcription factors such as IRF3 and IRF7 to stimulate the IFN response (Kato et al., 2008). During IAV infection, recognition of viral RNA by RIG-I-like receptors (RLRs) triggers activation of the IFN pathway (Evseev & Magor, 2019).

In ducks, RIG-I recognises the “panhandle” structure of complementary ends of genome segments. In chickens, which lack RIG-I, recognition is mediated by MDA5 to partially compensate (Hayashi et al., 2014). Toll-like receptors (TLRs) can also recognise viral RNA and activate the IRFs (Barjesteh et al., 2020; Evseev & Magor, 2019).

Unfortunately, due to a lack of available antibodies, we would not be able to confirm PB1-F2 full length and fragment expression if viral infection was used to induce IFN promoter activity. Using IAV would also make it harder to directly correlate changes in IFN promoter activity to PB1-F2 mutants, as IAV has several other immune antagonists, such as NS1 (Hatada & Fukuda, 1992; Krug, 2015) and PA-X (Gao et al., 2015; Jagger et al., 2012). Modifying NS1 has previously been suggested as a vaccine strategy, because mutating NS1 often results in attenuation of IAVs. NS1 deletion mutants grow to significantly lower titres than WT in IFN competent systems (Donelan et al., 2003; García-Sastre et al., 1998; Richt & García-Sastre, 2009). Mutating the frameshift sequence to reduce PA-X expression can also result in viral attenuation and also alter the host immune response (Hayashi et al., 2015).
A non IFN antagonising IAV mutant would already be attenuated and alter IFN expression compared to a WT virus. Using an IFN antagonism mutant IAV and complementing with PB1-F2 would also be an inaccurate model due to the changes during replication. While using a whole virus may trigger a more rounded immune response than transient transfection with poly(I:C), it would require a larger investment of time to produce and would still have pre-existing effects on IFN production that are not due to PB1-F2.

We did not find a significant dose dependency for IFN antagonism with either PB1-F2 protein, although there was a small trend (Figure 4.6). Previous work using higher concentrations of PB1-F2 and a shorter time between transfections (James et al., 2019), suggests that this

trend could have been pushed further with a change in the protocol. However, using 500ng PB1-F2 and an overnight poly(I:C) stimulation achieved antagonism vs induced PB1-F2 -ve and retained good cell viability.

5092 (H5N1) did not see any significant changes compared to WT PB1-F2 for STOP or ΔAUG mutants (Figure 4.7). When assessing IFN promoter activity only compared to the PB1-F2 -ve control by one way ANOVA (Figure 4.7 A), all PB1-F2 proteins had p values <0.01. After normalising to Renilla, with one way ANOVA compared to PB1-F2 -ve control the extent of IFN promoter antagonism appeared to vary among the PB1-F2 proteins, with only STOP25 and STOP35 showing significant reduction in IFN promoter activity.

UDL (H9N2) did not show antagonism of the IFN promoter when normalised to constitutive Renilla expression, even though when assessing IFN promoter activity only, all PB1-F2 proteins showed significant antagonism, with p=<0.0001 (Figure 4.7). This suggests that it had no specific effect on IFN promoter induction but instead was generally inhibitory to cellular gene expression.

In both assays, Renilla activity was not as expected (Figure 4.7 C, G). Renilla had previously been used as a transfection control for normalising expression (Chapter 3), and in polymerase assays (data not shown). In those instances, transfection of PB1-F2 mutants had had minimal impact on the constitutive expression of Renilla. In both the IFN assays, Renilla activity peaked in the GFP double -ve and PB1-F2 -ve conditions, mimicking the Firefly IFN promoter reporter activity. It is possible that something about the protocol or the DF1 immune response caused this unusual activity. Lipofectamine 2000 can induce cytotoxicity, and using a double transfection protocol may have encouraged any negative effect on the health of the cells. It is possible that using Xfect, a non-liposomal transfection reagent, would have reduced the impact of the repeated transfections (Chong et al., 2021). When IFN promoter activity was normalised to Renilla, the

uninduced controls show no excess induction of the IFN promoter, while the induced PB1-F2 -ve condition shows strong induction (Figure 4.7 B, F). As these three controls performed as expected, we were happy to display this data.

UDL PB1-F2 has 66N, which is associated with lower levels of immune antagonism than 66S (Conenello et al., 2011; Schmolke et al., 2011; Varga et al., 2011), the sequence which is present in 5092. In PR8 in 293T cells, an N66S mutation was found to enhance IFN antagonism activity (Varga et al., 2011). Infected mice lost weight earlier when infected with an N66S WSN/Hong Kong/156/97 PB1 reassortant compared to WT, and had increased cytokine production in the lung. The inverse was seen when a 1918 pandemic virus was changed to S66N. Mice required higher inoculums to cause severe disease, and produced lower lung titres (Conenello et al., 2007). N66S can cause increased cellular filtration and cytokine expression in infected mouse lungs, resulting in higher titres. The pathogenic phenotype of N66S is associated with suppression of the early antiviral response (Conenello et al., 2011). UDL being less effective as an IFN promoter antagonist is consistent with reports of 66N as less pathogenic than 66S. N66S is slightly unusual within PB1-F2, as it retains its phenotype across multiple subtypes. It has been mainly studied in mammalian systems to better understand pandemic risk, but was found to slightly increase disease progression in Peking ducks (Schmolke et al., 2011). In Figure 4.7 D, H, 5092 was able to be visualised on a western blot directly from passive lysis buffer lysed samples, whereas UDL had to be lysed with SDS lysis buffer to confirm expression. Considering the differences in protein stability described in Chapter 3, it is likely that UDL was degrading faster than 5092 during the assay. UDL PB1-F2 may have been less effective an antagonist in part because there was less protein present in the cell as time progressed. Although the same amount of plasmid was transfected, it may not be possible to do a like for like comparison without stabilising UDL PB1-F2 in some way.

Overall, this data has agreed with previous reports that PB1-F2 is strain and host specific in its functions. We are able to report a cumulative effect of ΔAUG mutations on viral replication kinetics for the first time, which is not matched by any impact on polymerase activity. No mutant of 5092 or UDL PB1-F2 proteins had a significant difference from the WT IFN antagonism activity. As reported previously, we found variation in immune antagonism activity between our H5N1 and H9N2 models. We have confirmed our hypothesis that introducing an N-terminal truncation to PB1-F2 does not fully inhibit its function.

Chapter 5: Discussion

This project aimed to investigate alternative AUG usage within segment 2 to determine which, if any, PB1-F2 C-terminal AUGs were recognised by ribosomes and, if so, to determine the functions of the resulting polypeptides compared to the full length PB1-F2 protein. We also wanted to continue previous work in the group started in James et al. (2019) and further assess the relationship between PB1-F2 intracellular localisation and protein stability. The work can be considered through a set of linked hypotheses.

Hypothesis 1; C-terminal AUGs express fragments of PB1-F2, and are functionally equivalent to a full length PB1-F2 protein in vitro

The data presented in this thesis showed that N-terminal (pre-residue 39) truncations were present and maintained in AIV sequences. The majority of these truncated sequences also maintained a full C-terminal ORF, with the potential to express functional fragments from AUGs 7, 8, or 9. N-terminal truncations have been maintained over time, through poultry movements, and during reassortment events. This strongly suggested there was no significant loss of fitness from a reduced PB1-F2 ORF in these populations. Poultry transmission is partly dependent on human intervention (Fusaro et al., 2016), whereas transmission between waterfowl is more dependent on contaminated water sources (Fourment & Holmes, 2015). Our bioinformatic analysis found greater transmission of viruses coding an N-terminally truncated PB1-F2 gene among poultry than in wild bird populations. Having confirmed the maintenance of the C-terminal PB1-F2 ORF in circulating avian strains, we showed that C-terminal fragments can be expressed from PB1-F2 proteins of multiple viral subtypes. We were also able to identify which AUG codon produced which protein on western blots for the first time. Identification allowed an analysis of the levels of expression from AUGs 7 and 9 in both WT and N-terminally truncated contexts. Detection of continued expression of C-terminal

fragments from an N-terminally truncated PB1-F2 sequence confirmed that N-terminal truncation alone (James et al., 2016; Mazel-Sanchez et al., 2018; McAuley et al., 2017) was an ineffective knock out strategy for PB1-F2. Previous work also showed independent expression of C-terminal PB1-F2 fragments (Zamarin et al., 2006). Their cautious strategy of removing all internal AUG codons has been verified in our work, which showed all three of AUGs 7-9 can be used, to varying extents.

Any activity of the C-terminal PB1-F2 fragments was then investigated. Due to the mutation strategy, each ΔAUG mutant caused a Met>Thr mutation in the full length PB1-F2. The resulting phenotypes could be due to either loss of C-terminal fragment expression or changes to the PB1-F2 sequence. ΔAUG mutations to prevent C-terminal fragment expression had no effect on viral polymerase activity assays in avian cell lines, whereas N-terminal truncation STOP35 had a modest cell and polymerase context dependent impact on activity. In QT-35 cells, STOP35 significantly reduced polymerase activity in a 3:1 polymerase context. The viral polymerase activity data was not an effective predictor of viral replication kinetics. *In ovo*, N-terminal truncations had no effect on viral replication kinetics, whereas a ΔAUG 7, 9 double mutant reached significantly higher titres than WT at 48hpi. When the impact of disrupting the PB1-F2 ORF on innate immune antagonism was assessed, none of the mutants tested had a significant difference in IFN promoter activity compared to the WT protein. PB1-F2 C-terminus expression was sufficient to antagonise IFN activity. IFN promoter inhibition by expression of C-terminal PB1-F2 fragments from STOP25 and STOP35 mutants suggests that the ΔAUG phenotypes seen could indeed be due to loss of expression rather than the Met>Thr mutations in the full-length protein.

We were able to prove our hypothesis that functional C-terminal fragments of PB1-F2 can be expressed. More work is needed to determine how equivalent the fragments are to a full-length protein,

but in some areas expression of fragments alone appears to be sufficient for function. We would like to confirm PB1-F2 C-terminus expression during virus infection, as our work was limited to using expression plasmids. There is also scope to investigate at which stage of the type 1 IFN pathway PB1-F2 is active, and to use techniques such as co-immunoprecipitation to determine if the full-length protein and C-terminus of PB1-F2 have differing interactomes.

PB1-F2 is not the only viral protein to express a functional fragment. PB1-N40 is expressed from an internal AUG of PB1 by reinitiation. PB1-N40 has lost the transcriptase function of PB1, but its expression still has an impact on viral replication, although the exact mechanism is unknown (Tauber et al., 2012; Wise et al., 2009, 2011). A number of potential start sites downstream of canonical AUG codons have been identified across PB1, M, NP, PA, and NA segments within the IAV genome, indicating that this is a widespread strategy for expanding the IAV proteome (Machkovech et al., 2019; Pinto et al., 2020). PB1-N40 is known to affect viral replication kinetics (Tauber et al., 2012; Wise et al., 2009), while NA43 is able to support low levels of viral growth in the absence of its full length counterpart (Machkovech et al., 2019).

There has been a general lack of investigation of PB1-F2 C-terminal fragment functionality. While fragment expression has been previously reported (Kamal et al., 2015; Zamarin et al., 2006), minimal phenotypic analysis has been performed. Kamal and colleagues did investigate the effect of N-terminal truncations of PB1-F2 (Kamal et al., 2015), and it has been shown that a STOP9 truncation caused no attenuation of PR8 in mice, whereas a complete ablation of PB1-F2 expression caused reduced pathogenicity and delayed clearance of the virus from the lungs (Zamarin et al., 2006). However, the internal AUG codons of PB1-F2 have not been individually investigated. To our knowledge, this is the first work to attempt to fill this gap in our understanding. While we have answered some questions, there is further work

required. For example, it is not clear how AUGs 7-9 are reached by an initiation-competent ribosome, and our investigation of innate antagonism by the C-terminus of PB1-F2 was not as broad as originally planned. It would be interesting to see if changing the strength of the Kozak consensus around the AUGs 7-9 would give a similar phenotype to the ΔAUG mutants without completely removing the AUG codon. Methods such as ribosome profiling or altering distance between STOP and AUG codons could be used to improve our understanding of segment 2 translation (Ingolia et al., 2019).

PB1-F2 is a highly strain specific protein, and the expression and phenotypic results discussed were obtained from only two model viruses: 5092 (H5N1) and UDL (H9N2). We could be confident that N-terminal truncations were present in multiple subtypes, and that the C-terminal ORF was maintained from the sequence data, but we cannot confirm expression from all sequences.
Due to time all work was performed *in vitro*, and the majority with tagged proteins rather than whole virus due to a lack of available antibodies. The large difference in size between untagged PB1-F2 and PB1-F2-GFP may also have affected intracellular localisation due to steric hindrance of PB1-F2 by the large GFP tag (Yamada et al., 2004). It is possible that using expression plasmids rather than viral infection masked some activities of PB1-F2, such as interactions with other viral proteins, and we have been unable to test possible effects of the PB1-F2 mutations on properties such as virus infectivity and transmission without an *in vivo* experiment.

Hypothesis 2; PB1-F2 stability is linked to localisation

Our initial data showed that PB1-F2 proteins from H5N1 A/turkey/England/50-92/1991 (5092) and H9N2 A/chicken/Pakistan/UDL-01/2008 (UDL) had very different intracellular localisation and stability, as previously reported (James et al., 2019). 5092 and UDL proteins showed varying reticular or cytoplasmic

intracellular localisation. As expected, 5092 PB1-F2 had a much longer mean half-life (of 41 minutes) than UDL, with a mean half-life of 20 minutes. PR8 has previously been reported to have a half-life of 1hr, but this was calculated from infected MDCKs using ^{35}S-amino acid labelling rather than CHX chase of transfected avian source PB1-F2 proteins (Zamarin et al., 2006).

While we were unable to confirm the intracellular localisation of H3N8a, H3N8b, and H11N9 PB1-F2 proteins, we did find that both H3N8 proteins were highly unstable. This was unexpected, as H3N8a PB1-F2 was reported to have mitochondrial localisation (James et al., 2019) and therefore, under our hypothesis, expected to be stable. The other mitochondrial PB1-F2 protein examined in James et al., 2019, H10N7, was not able have its stability quantified by CHX assays, and showed very weak expression from transfected cells, even when incubated with MG132 (data not shown). While we did produce structural predictions of PB1-F2 proteins that reported different intracellular localisations in James et al., 2019 and that we had stability data for using the I-TASSER servers, a lack of available structural data resulted in low confidence predictions. Without more structures in the PDB, predictions will remain highly speculative, especially as the N-terminus of PB1-F2 is thought to be highly disordered. Unfortunately, producing our own crystal structures was outside the scope of this thesis.

This work was performed before AlphaFold became available, but it would be interesting to repeat the predictions with the new gold standard. AlphaFold uses an AI system to predict 3D structures from amino acid sequences, and recently won CASP14. AlphaFold works entirely from the amino acid sequence, without requiring a homologous sequence (Jumper et al., 2021). Despite the lack of existing structural data, it is highly likely that PB1-F2 proteins with different intracellular localisations, that interact with different binding

partners, will have differences in structure to complement and facilitate the variation in environment and function.

As we were unable to confirm intracellular localisation for the wider panel of PB1-F2 proteins, we cannot conclusively prove or disprove this hypothesis, though the data available do not support a direct correlation. However, we consider PB1-F2 stability to be a multi-factor property, affected by both intracellular localisation and structure.

Several papers have correlated stability of PB1-F2 with specific sequence motifs (Alymova et al., 2014; Cheng et al., 2017; Košík et al., 2015), with fewer reports linking stability and localisation (Cheng et al., 2017; James et al., 2019). Localisation has been reported to vary with subtype (James et al., 2019; McAuley et al., 2010). PB1-F2 has been previously reported to have variable stability (Cheng et al., 2017; Park et al., 2019). Pulse chase experiments with PR8 PB1-F2 report a half-life of one hour (Zamarin et al., 2006), which is longer than any strain tested in this thesis. However, this was in the context of viral infection, in comparison to our CHX chase assays using V5-tagged expression plasmids. There may be variation in expression levels or efficiency of proteasome targeting during viral infection that is not mimicked by transient transfection.

There is limited data on PB1-F2 structure. Several papers have attempted secondary structure predictions using software such as RaptorX (Pasricha et al., 2018), calculations from NMR spectroscopy data for sections of PB1-F2 that were then pieced together using the consecutive segment approach (Solbak et al., 2013), or circular dichroism (CD) spectroscopy in various environments and dynamic light scattering data to estimate mean hydrodynamic radius as a measure of oligomerisation of PB1-F2 (Chevalier et al., 2010). There is only one entry for PB1-F2 in the protein data bank, produced using NMR and CD spectroscopy. Structures were generated using CNS-solve 1.0 software, and the 20 predictions with the lowest energy were

superimposed to produce a final structure of residues 50-87 of PR8 PB1-F2 (Bruns et al., 2007). The majority of PB1-F2 structural papers have calculated structures for fragments rather than the full-length protein.

As we were unable to confirm intracellular localisation for the PB1-F2 panel, we cannot fully determine the relationship between intracellular localisation and protein stability. We have had to rely on previously published reports of intracellular localisation (James et al., 2019) to form hypotheses on the effect of intracellular localisation. We were limited in the CHX data we were able to obtain for mitochondrially-localised proteins in the James 2019 panel. PB1-F2 is a short protein with relatively few methionine residues, and preliminary radioactive pulse-chase assays, performed by a colleague in the Digard lab, using ^{35}S-methionine labelling had not produced high quality data (data not shown), so we used CHX chase assays. CHX chase assays are not as sequence reliant and do not require the use of radioactivity, so are safer to work with. Unfortunately, CHX chasing does rely on a certain level of expression at time zero. If the protein has failed to accumulate, its degradation cannot be measured. Not all the PB1-F2 proteins in the panel could be visualised in CHX assays, which biased our data collection.

Concluding Remarks/Summary

In this thesis we have shown N-terminal truncations of PB1-F2 are maintained in clinical isolates, particularly in poultry hosts. While we found some subtype variation in the prevalence of truncations, there was no suggestion that expressing only the C-terminus of PB1-F2 is detrimental to viral fitness in poultry. The C-terminal fragments appeared to be highly functional *in vitro*, and we cannot discount a further phenotype *in vivo*. The decreased prevalence of truncated sequences in wild hosts suggested a possible effect on the viral replication cycle that we were unable to test. For example, full length

PB1-F2 has been seen to affect transmission in chickens (James et al., 2016). Within poultry hosts, N-terminal truncations of PB1-F2 may be a method of focusing translation on the terminal containing the functional domains. It is possible that the different termini of PB1-F2 are under varying selection pressures. Previous work has suggested that PB1-F2 is constrained by the PB1 sequence (Zell et al., 2007). The combination of constraint by PB1 and the concentration of functional motifs in the C-terminus of avian PB1-F2 proteins may reduce the pressure to maintain the N-terminus. PB1-F2 may be continuing to evolve as a gene.

We presented evidence that C-terminal fragments of PB1-F2 are functional and can have a phenotypic impact on virus replication. We confirmed that introduction of an N-terminal STOP codon is not sufficient to knock out PB1-F2 expression. As ΔAUG mutations can impact viral replication kinetics, a STOP12 STOP58 double mutant may be more suitable to ablate PB1-F2 expression. This improved understanding of the ORF structure of PB1-F2 suggests a change needs to be made in the way segment 2 sequences are reported, as an early STOP codon does not prove a lack of PB1-F2 expression. A minimal PB1-F2 ORF may also impact pandemic risk assessments, as we have proven that a full length protein is not required for innate immune antagonism.

Strain variation of PB1-F2 is likely to include variable structure for binding variable interaction partners, and to account for different cellular environments. All these factors will affect the stability and half-life of PB1-F2 proteins from different viruses. We hope the data presented here support the consideration of stability as a multi-factor property of PB1-F2.

Future Work

To further explore the relationship between intracellular localisation, structure, and stability, confirmation of the intracellular localisation of

the PB1-F2 panel is required. Ideally this would be done through confocal microscopy of the full PB1-F2 panel with a mitochondrial stain to allow quantification of intracellular colocalization. We would also have liked to have performed mitochondrial membrane insertion assays by assessing the quantity of PB1-F2 in mitochondrial vs cytoplasmic fractions of transfected cells. It would have been interesting to determine whether membrane insertion correlates with stability, and could help support the variation in predicted structures, particularly the predicted difference in C-terminal helix structure between H3N8a and 5092 PB1-F2 proteins. Cellular fractionation could also have been an additional control for the colocalization work. Creating 5092 PB1-F2 mutants to mimic the H3N8a sequence and assessing impact on colocalization and membrane insertion would have helped confirm whether protein structure is affecting protein stability. We had hoped to perform more CHX assays with the H3N8 proteins, using a shorter time course. A one hour time course with samples taken every fifteen minutes may have caught smaller differences in degradation between proteins that were not visible over the original 4hr assay, and given more confidence in the data.

The key to confirming protein intracellular localisation varies according to structure and function is to obtain 3D structures of PB1-F2 proteins from multiple subtypes. Due to its small size and disordered N-terminus, obtaining crystal structures of PB1-F2 could be highly challenging. Other methods of obtaining structural information include NMR (Bruns et al., 2007) and cryo-EM, a form of electron microscopy (Hebert, 2019). Without more PB1-F2 structures, future studies will continue to be hampered by a lack of information. Structures of full length PB1-F2 proteins of different subtypes are needed to help understand not only stability, but varying functionality of the proteins. Protein structure and function are tightly related, as structure helps determine binding affinity for protein-protein interactions. At a minimum, structures for 5092 and UDL would have been highly

informative, and possibly confirm our hypothesis that 5092 is better suited to membrane insertion than UDL. Ideally the full panel of PB1-F2 proteins would be crystallised to allow for a high confidence comparison of structures from differently localised PB1-F2 proteins.

We were unable to fully assess the effect of truncations or ΔAUG mutations on PB1-F2 immune antagonism. In future, NFκB activity should be assessed, as UDL has been reported to antagonise the NFκB pathway to a higher level than the IFN pathway (James et al., 2019). We would also like to assess antagonism of different members of the activation pathway to see at which stage PB1-F2 inhibits the immune responses. This work could include co-immunoprecipitation to identify binding partners.

To fully assess the impact of interrupting the PB1-F2 ORF on viral infection, an *in vivo* experiment is needed. In our original plans, we hoped to assess differences in shedding and transmission between WT, STOP12, ΔAUG 7, and a STOP12 ΔAUG 7 double mutant. We would also have taken tissue samples from culled birds for plaque assays, RT-PCR, and cytokine measurements to attempt to answer the question of how much difference in viral shedding and immune antagonism there is between full length PB1-F2 and C-terminal expression only during infection. Hopefully this work will be performed in the future.

Chapter 6: Materials and Methods

6.1 Materials

6.1.1 General Reagents

The Roslin Institute and Pirbright Institute Central Services Units (CSU) prepared and provided sterile water and phosphate-buffered saline (PBS). Additional specific reagents and kits and their respective suppliers are listed below.

30% acrylamide:bisacrylamide (37.5:1) (BioRad)

Agarose (Eurogentec)

Dimethyl sulphoxide (DMSO) (Sigma)

DNA molecular weight markers, HyperLadder 1kb (BioLine)

Lipofectamine 2000 (ThermoFisher Scientific)

Nitrocellulose membrane (GE Healthcare)

Precision plus molecular weight marker (BioRad)

Polyinosinic:polycytidylic acid (poly I:C) (Invitrogen)

QIAamp Viral RNA Kit (Qiagen)

QIAGEN Plasmid Midi Kit (Qiagen)

QIAprep Spin Miniprep Kit (Qiagen)

QIAquick Gel Extraction Kit (Qiagen)

QIAquick PCR Purification kit (Qiagen)

QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies)

DUAL Glo Luciferase Assay kit (Promega)

GelRed Nucleic Acid stain (Biotium)

Tetramethylethylenediamine (TEMED) (BioRad)

Tween20 (Sigma)

Transblot Turbo Transfer System (Biorad)

Odyssey CLx (LI- COR)

Image Studio™ Software (LI-COR)

6.1.2 Enzymes

DNA restriction endonucleases and their respective reaction buffers were supplied by Promega and New England Biolabs and used according to manufacturer's instructions, unless otherwise stated.

6.1.3 Bacterial Cells

DH5α

6.1.4 Eukaryotic Cells

Cell line	Source	Reference
Madin-Darby Canine Kidney Cells (MDCK)	The Roslin Institute/The Pirbright Institute CSU	(Gaush et al., 1966)
Human Embryonic Kidney 293T Cells (293T)	The Roslin Institute/The Pirbright Institute CSU	(DuBridge et al., 1987)
Chicken Embryonic Fibroblasts (DF1)	The Roslin Institute/The Pirbright Institute CSU	(Himly et al., 1998)
Japanese Quail Fibrosarcoma Cells (QT-35)	The Roslin Institute/The Pirbright Institute CSU	(Moscovici et al., 1977)

Table 6.1 Cell Lines

6.1.5 Solutions and Media

6.1.5.1 Eukaryotic cell culture media and cell passage solutions

0.25% Trypsin-EDTA (Sigma)

2.5% trypsin, Versene (Gibco)

Foetal bovine serum (FBS) (Sigma)

Dulbecco's modified Eagle Medium (DMEM) (Invitrogen)

L-glutamine (Gibco)

Penicillin/Streptomycin (Gibco)

Opti-MEM Reduced Serum Medium (Gibco)

The previously listed supplies were used to prepare the following media:

Complete medium; DMEM, 10% (v/v) FBS, 1% penicillin/streptomycin (P/S) (initial concentration 10,000 units/ml)

Avian medium; DMEM, 10% FBS, 1% P/S, 1% L-Glu (initial concentration 200 mM)

Transfection medium; DMEM

Viral Growth Medium; DMEM, 0.15% BSA, 0.01% TPCK trypsin, 1% P/S

Serum Free Media; DMEM, 1% P/S

Plaque Overlay; 0.6% agarose, Eagle's Minimum Essential Medium, 0.21% Bovine Serum Albumin, 1 mM L-Glu, 0.15% Sodium Bicarbonate, 10 mM HEPES, 1x P/S, 0.01% Dextran DEAE, 2 µg/ml TPCK treated trypsin.

6.1.5.2 Bacterial Media

Luria-Bertani (LB) broth and LB-agar were prepared and provided by the Roslin Institute and Pirbright Institute CSUs.

Ampicillin (Sigma) 100 µg/ml

Kanamycin (Sigma) 50 µg/ml

6.1.5.3 Nucleic Acid Gel Electrophoresis Buffers

Tris/Borate/Ethylenediaminetetraacetic acid (EDTA) (TBE) buffer (Invitrogen)

Gel loading dye Purple 6x (NEB)

1kb DNA ladder (NEB)

6.1.5.4 Protein Buffers and Solutions

6.1.5.4.1 Lysis Buffers

4x Laemmli's sample buffer (BioRad)

SDS Lysis buffer

5x Passive Lysis buffer (Promega)

6.1.5.4.2 Acrylamide Gel Electrophoresis

SDS-PAGE running buffer; 10x Tris/Glycine/SDS (National Diagnostics)

6.1.5.4.2 Western Blotting

Protein transfer buffer; 1.5g Tris, 7.2g Glycine, 100ml methanol

Blocking solution; PBS/0.1% Tween20, 5% skimmed milk

Washing solution; PBS/0.1% Tween20 (PBST)

6.1.6 Viruses and Reverse Genetics Systems

Name	Accession number
PR8 PB1	EF467819
PR8 PB2	EF467818
PR8 PA	EF467820
PR8 NP	EF467822

PR8 NS1	EF467817
PR8 M	EF467824
PR8 NA	EF467823
PR8 HA	EF467821
5092 PB1	CY015125
5092 PB2	GU052517
5092 PA	GU052515
5092 NP	GU052513
UDL PB1	CY038456

Table 6.2: Sequence Accession Numbers for Viral cDNA Inserts in Reverse Genetics and Expression Plasmids

6.1.7 Plasmids

Name	Description	Selection	Source	Application
pEGFP-N1	Express eGFP under control of CMV IE promoter. Multiple cloning site upstream allows cloning of C-terminally tagged proteins	Kanamycin	Clontech	Expression GFP tagged proteins
V5	Express V5 under control of human elongation factor-1 α promoter. Multiple cloning site allows cloning of tagged proteins	Ampicillin	Shelton lab	Expression V5 tagged proteins
Avian pPol I vRNA FFLuc	Reporter for RNP reconstitution assays. Contains firefly luciferase reporter gene in the reverse (-) orientation, flanked by the UTRs of PR8 segment 8, under the control of an avian pol I promoter.	Ampicillin	Digard lab (Gift from Dr Laurence Tiley (The University of Cambridge, UK))	Luciferase reporter assays
pPol I vRNA FFLuc	Reporter for RNP reconstitution assays. Contains firefly luciferase reporter gene in the reverse (-) orientation, flanked by the UTRs of PR8 segment 8, under the control of a pol I promoter	Ampicillin	Shelton lab	Luciferase reporter assays
Ck IFN-β FFLuc	Reporter for IFN inhibition assays. Contains firefly luciferase reporter gene under control of chicken IFN-β promoter	Ampicillin	Shelton lab	Luciferase reporter assays
Renilla	Constitutively expressed Renilla luciferase under the control of the SV40 promoter	Ampicillin	Shelton lab	Luciferase reporter assays
pDUAL/pHW2000	Reverse genetics plasmid. Bidirectional pol I and pol II promoters either side of insert lead to mRNA and vRNA-like RNA synthesis. Used for Pr8, 5092, UDL.	Ampicillin	Digard lab (Gift from Prof Ron Fouchier (Erasmus University Medical Center, Rotterdam, The Netherlands))	Influenza reverse genetics
pCAGGs	Pol II promoter leads to constitutive high protein expression. Used for UDL PB1.	Ampicillin	Shelton lab	Protein expression
Pol1	Reverse genetics plasmid. Pol1 promoter leads to vRNA-like RNA synthesis. Used for UDL PB1.	Ampicillin	Shelton lab	Influenza reverse genetics

Table 6.3 Expression and Reverse Genetics Plasmids

6.1.8 Oligonucleotides

6.1.8.1 Oligonucleotides used for Sequencing of Constructs and Viruses

Primer	Sequence	Application
UDL newseq Rv	GGCCAAAGCAGTTGCAGCTG	Rev to sequence 5' PB1-F2
5092 newseq Rv	AGTGTTAGCCAAAGCAGTTGC	Rev to sequence 5' PB1-F2
5092 PB1 f1	CCATTTGAATGGATGTC	Fwd primer to sequence segment 2
PR8 PB1 Fwd	GGAACAGGATACACCATGG	Fwd primer to sequence segment 2
5092 PB1 f2	GGTCACACAAAGAACAATAGG	Fwd primer to sequence segment 2
5092 PB1 f3	CCTGAAATACTTCAACGAATCAACG	Fwd primer to sequence segment 2
5092 PB1 r1	GGACAAGCTAAATTCACTG	Rev primer to sequence segment 2
5092 PB1 r3	GGTATTTGTGTCCGTATCTTC	Rev primer to sequence segment 2
5092 PB1 r4	GTCATGTTGTCCCTTACTC	Rev primer to sequence segment 2
UDL PB1 Inseq4	CATCTTCTTGGTCATGCTCTGG	Rev primer to sequence segment 2
UDL PB1 Inseq2	GGAATGAGAATCAGAATCCCCG	Fwd primer to sequence segment 2
UDL PB1 Inseq5	GTATCCCTTCCCTAATCTCGC	Rev primer to sequence segment 2
UDL PB1 Inseq3	GAAGTCTTACATAAATCGGACAGGG	Fwd primer to sequence segment 2
UDL PR8 PB1 Rv	GGAACAGATCTTCATGATCTCAG	Rev primer to sequence segment 2
Pol1 Cterm	GGAATTTCCAGCATGGTGGAGGCCATGGTT	Fwd primer to confirm plasmid vector
CMV Fwd	CGCAAATGGGCGGTAGGCGTG	Fwd primer to sequence insert

[1]

[1] Fwd = forward primer. Rv/Rev = Reverse Primer

Table 6.4 Sequencing Oligonucleotides

6.1.8.2 Oligonucleotides used to Subclone Viral Sequences into the Indicated Vectors

Primer	Sequence	Application
GFP IF Fw long	CCGGTCGCCACCGGTGAGCAAGGGCG	Remove GFP ATG and bring tag into frame 2
GFP IF Rv long	CGCCCTTGCTCACCGGTGGCGACCGG	Remove GFP ATG and bring tag into frame 2
1-27 KpnI Fw	CCCGGTACCAGCGAAAGCAGGCAAACCATTTGAATG	Amplify and add restriction site
AgeI 283-261 Rv	GGGCACCGGTAGTTTATCCACTCTTGTTTGC	Amplify and add restriction site

Table 6.5 Cloning Oligonucleotides

6.1.8.3 Oligonucleotides used for Site-Directed Mutagenesis

Primer	Sequence	Mutation
UDL ΔAUG4 Fw	GACCCTCCCTACAGCCACGGAACAGGAACA	T120C (M1T)
UDL ΔAUG4 Rv	TGTTCCTGTTCCGTGGCTGTAGGGAGGGTC	T120C (M1T)
UDL ΔAUG7 Fw	ACCCCAACTCAATCCAATTGACGGGCCGCTACC	T234C (M39T)
UDL ΔAUG7 Rv	GGTAGCGGCCCGTCAATTGGATTGAGTTGGGGT	T234C (M39T)
UDL ΔAUG9 Fw	GGATAACGAGCCAAGCGGATACGCACAGACAGA	T270C (M51T)
UDL ΔAUG9 Rv	TCTGTCTGTGCGTATCCGCTTGGCTCGTTATCC	T270C (M51T)
UDL +AUG8 Fw	GGCCGCTACCTGAGGATAATGAGCCAAGCGG	C255T (T46M)
UDL +AUG8 Rv	CCGCTTGGCTCATTATCCTCAGGTAGCGGCC	C255T (T46M)
UDL STOP12 Fw	GGATACACCATGGACACAGTAAATAGAACGCATCAATACTC	C153A (S12STOP)
UDL STOP12 Rv	GAGTATTGATGCGTTCTATTTACTGTGTCCATGGTGTATCC	C153A (S12STOP)
UDL STOP25 Fw	GGGAAGATGGATAACAAATACAGAGACTGG	C191T (Q25STOP)
UDL STOP25 Rv	CCAGTCTCTGTATTTGTTATCCATCTTCCC	C191T (Q25STOP)
UDL STOP26 Fw	GGAAGATGGACAATAAATACAGAGACTGG	C194T (Q26STOP)
UDL STOP26 Rv	CCAGTCTCTGTATTTATTGTCCATCTTCC	C194T (Q26STOP)
UDL STOP35 Fw	GGAGCACCCCAACTGAATCCAATTGATGGG	C222A (S35STOP)
UDL STOP35 Rv	CCCATCAATTGGATTCAGTTGGGGTGCTCC	C222A (S35STOP)
5092 ΔAUG4 Fw	GATCCTCCATACAGCCACGGAACAGGAACA	T120C (M1T)
5092 ΔAUG4 Rv	TGTTCCGTGGCTGTATGGAGGATC	T120C (M1T)
5092 ΔAUG7 Fw	CCAATTGACGGGCCATTACC	T234C (M39T)
5092 ΔAUG7 Rv	GGTAATGGCCCGTCAATTGG	T234C (M39T)
5092 ΔAUG9 Fw	GATAACGAGCCAAGCGGATACGCACAAACAGATTG	T270C (M51T)
5092 ΔAUG9 Rv	CAATCTGTTTGTGCGTATCCGCTTGGCTCGTTATC	T270C (M51T)
5092 +AUG8 Fw	GGATAATGAGCCAAGC	C255T (T46M)
5092 +AUG8 Rv	GCTTGGCTCATTATCC	C255T (T46M)
5092 STOP12 Fw	GGATACACCATGGACACAGTAAATAGAACACATCAATACTC	C153A (S12STOP)
5092 STOP12 Rv	GAGTATTGATGTGTTCTATTTACTGTGTCCATGGTGTATCC	C153A (S12STOP)
5092 STOP25 Fw	GGGGAAGTGGATAACAAACACAGAGACTGG	C191T (Q25STOP)
5092 STOP25 Frv	CCAGTCTCTGTGTTTGTTATCCATTCCCC	C191T (Q25STOP)
5092 STOP26 Fw	GGAAGTGGACAATAAACACAGAGACTGG	C194T (Q26STOP)
5092 STOP26 Rv	CCAGTCTCTTGTGTTTATTGTCCACTTCC	C194T (Q26STOP)
5092 STOP35 Fw	GGAGCACCCCCCATGAATCCAATTGATGGGCC	C222A (S35STOP)
5092 STOP35 Rv	GGCCCATCAATTGGATTCAGTTGGGGTGCTCC	C222A (S35STOP)

Table 6.6 Oligonucleotides used to Introduce Mutations into the PB1-F2 ORF

[2]

6.1.9 Immunological Reagents and Dyes

Antibody	Supplier	Applied dilution

[2] Fw = forward primer. Rv = reverse primer. Mutations refers to nucleotide change (amino acid in brackets).

IRDye® goat anti mouse 680	Li-Cor C90910-21	1:10 000
IRDye® goat anti rabbit 800	Li-Cor C90723-19	1:10 000
Anti mouse 488	Abcam ab150113	1:1000
Rabbit pAb α tubulin	Abcam ab15246	1:2000
Mouse anti-V5 tag	The Pirbright Institute	1:1000
Mouse anti-NP	The Pirbright Institute	1:1000/1:3000
β actin ab8226	Abcam	1:2000

Table 6.7 Antibodies

6.2 Molecular Techniques and Nucleic Acid Handling

6.2.1 Polymerase Chain Reaction (PCR)

PCR was used for cloning applications. Different size fragments were amplified using forward and reverse primers from section 6.1.8. Each PCR mix contained 100 ng each primer in a total volume of 50 µL. The PCR was carried out in a BioRad T100 thermocycler. PCR conditions included a denaturation step performed at 94ºC for 2 minutes, followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 55-60°C and 1-6 minutes at 72°C (1 minute per 1Kb), finishing with 10 minute incubation at 72°C. Annealing temperature and extension time varied depending on melting temperature of primers and amplicon length, respectively. PCR product length was confirmed by 1% agarose gel electrophoresis.

2 x PfuUltra II Fusion HS DNA Polymerase master mix (Agilent)

6.2.2 DNA Gel Electrophoresis

DNA electrophoresis was performed with agarose gels. 1% agarose gels were prepared by dissolving agarose in a microwave oven in TBE buffer, with the addition of 1x GelRed gel stain. Melted agarose was poured into a gel tray with a comb containing the desired

number of wells. Once set, the gel was immersed in 1× TBE buffer. 5µl HyperLadder 1kb (BioLine) DNA ladder was loaded alongside samples containing 1x DNA loading buffer. Gels were run at 80V for approximately 45min and imaged using a BioRad GelDoc EZ Imager.

6.2.3 Purification of DNA Fragments

PCR amplified or restriction digested fragments were purified using the QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions.
Briefly, a DNA-containing sample was diluted in binding buffer which allows binding to a silica-gel membrane. Throughout a series of centrifugation steps, contaminants were removed by an ethanol-based buffer wash and silica matrix-retained DNA was eluted with nuclease free water. Purified DNA was analysed by spectrophotometry and electrophoresis.

6.2.4 Restriction Enzyme Digestion

During subcloning, restriction sites were added to the end of insert sequences for digestion. Age1 and Kpn1 enzymes were used to digest both insert and plasmid vectors, providing differing sticky ends. Digested plasmid DNA was then treated with rSAP to prevent self-ligation.

6.2.5 Extraction of DNA Fragments from Agarose Gels

When subcloning, restriction digested vector DNA fragments were separated by electrophoresis, visualised in an ultraviolet (UV) transilluminator (Gel DOC EZ System, Bio-Rad) and excised with a scalpel. The agarose slice was extracted using the QIAquick Gel Extraction Kit in a bind-wash-elute procedure. Concisely, gel slices were dissolved at 50°C in a high salt, neutral pH buffer and the nucleic acid-containing mixture was applied to a silica membrane. Impurities were sequentially washed away using an ethanol-based solution and pure DNA was eluted in 30 µL of water. DNA was

assessed by spectrophotometry for its content and purity and further used in subsequent applications such as restriction digestions or ligations.

6.2.6 Ligation of DNA fragments

During subcloning, ligation reactions were made at 1:3, 1:5, and 1:7 vector:insert ratios. Reactions were incubated with T4 ligase overnight in an ice bucket to allow the reaction to reach a temperature where the ligase was most efficient.

6.2.7 Preparation of Competent Bacterial Cells

Each new batch of competent bacterial cells was prepared from a previously competent bacterial cell liquid culture. After overnight culture, in antibiotic free broth, *E. coli* DH5α were diluted 1:400 and incubated to an OD_{600} of ~0.6. Cells were centrifuged at 4000rpm, 20 minutes at 4ºC. Supernatant was discarded and the pellet resuspended in 1/5 original volume ice cold 100mM $CaCl_2$. Cells were incubated on ice for 30 minutes, then harvested by centrifuged at 4000rpm for 5 minutes at 4ºC. Supernatant was discarded and pellets resuspended in 1/5 starting volume ice cold 100mM $CaCl_2$. Sterile glycerol was added and cell suspensions were aliquoted into 50µl dosages. Aliquots were snap frozen in a dry ice methanol bath and stored at -80ºC until required for plasmid transformation.

6.2.8 Transformation of Competent Bacterial Cells

Plasmid DNA was transformed into chemically competent *E. coli* using the heat shock method. Briefly, 500 ng DNA was incubated with 30 µL competent bacteria on ice 30 minutes. Cells were heated at 42ºC for 45 seconds, allowing the opening of bacterial membrane pores and the entry of plasmid DNA. Cells were returned to ice for 2 minutes, then had 200 µL LB broth added and recovered in a 37ºC shaker (180rpm) for 45 minutes. The suspension was then plated on the appropriate selective LB agar medium.

6.2.9 Bacterial culture

E. coli strains were grown (from single colonies or previously amplified mini cultures) in LB broth supplemented with 100 µg/ml ampicillin or 50 µg/ml kanamycin (initial concentrations) and incubated overnight at 37 ºC in a 180rpm shaker.

6.2.10 Plasmid DNA extraction and quantification

Plasmid DNA was prepared in small or large scales. Small-scale preparations (mini-prep) was performed from 3 mL of overnight *E. coli* cultures, whilst large- scale plasmid DNA preparations (midi-preps) used 50 mL of overnight culture. Mini- and midi-preps were performed using Qiagen's Plasmid DNA Mini Prep and Plasmid Plus Midi kits, respectively. DNA concentration was determined by measuring the absorbance at 260nm using a Nanodrop spectrophotometer. Purity of the same preparations was assessed through the analysis of absorbance ratios A260/A280 and A260/A230 which indicate protein and ethanol/phenol/EDTA contaminations, respectively.

6.2.11 Site-directed mutagenesis

Introduction of single and double nucleotide mutations in expression and reverse genetics plasmids was performed using the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent) according to manufacturer's instructions with minor changes. Briefly, a PCR was performed using synthetic oligonucleotides containing the desired mutation in a total volume of 25 µL followed by the addition of 1 µL of DpnI restriction enzyme and incubation at 37ºC for 1 hour. 5 µL PCR purified reaction was used to transform competent cells which were plated on antibiotic resistance selective LB agar plates. Following an overnight 37°C incubation, individual colonies were selected, regrown in broth for a following overnight period and used to extract plasmid DNA.

6.2.12 DNA sequencing

DNA was sequenced using Eurofins Supreme Run. All plasmid DNA used in this work was sequence confirmed. For sequencing of plasmid DNA, 20 µL of 100 ng plasmid was sent with 20 µL 100 ng/µL of the appropriate primer. For sequencing of PCR products, 20 µL of 20 ng/µL purified PCR product was sent with the appropriate primer.

6.3 Eukaryotic cell culture, isolation, and manipulation

6.3.1 Cell passage

Madin-Darby canine kidney (MDCK), and human embryonic kidney (293T) cell lines were cultured in complete media. Chicken embryonic fibroblast (DF1) cell line was cultured avian media. All cell lines were grown at 37°C 5% CO_2. Cell passage was performed twice weekly. Complete medium was removed, the cells washed once in PBS, and incubated with 0.25% trypsin-EDTA (Sigma) until cells detached. Cells were resuspended in fresh complete medium and 10-50% of the cells were transferred to a new flask containing complete medium (total volumes: 15 mL for 75 cm^2 and 30 mL for 150 cm^2 flasks).

6.3.2 Cell counting

Cell lines were counted either by hand with a haemocytometer or using a Countess II (Invitrogen). For both methods, 20 µL resuspended cells were added to 20 µL 0.4% tryphan blue (Gibco) and 10 µL supplied to the chamber under a cover slip. With a haemocytometer cells were counted under a light microscope in a squared area equivalent to 0.1 mm^3 ($1x10^{-4}$ml). Cell concentration was estimated to be number of counted cells x $1x10^4$ cells/mL. With the Countess, cell concentration was automatically calculated.

6.3.3 Plasmid transfection of mammalian and avian cells

6.3.3.1 Lipofectamine transfection

Mammalian and avian cells were routinely transfected with Lipofectamine 2000 reagent following manufacturer's instructions. Briefly, cells were seeded the day before in order to get 70-80% confluency on the day of transfection. The desired concentration of DNA was diluted in OptiMEM (Gibco). In a separate tube, the desired concentration of lipofectamine 2000 (Invitrogen) was added to the same volume of OptiMEM and incubated at RT for 5mins. Equal volumes DNA-OptiMEM and lipofectamine-OptiMEM were combined and incubated at RT for 20 mins. During this incubation, complete media was removed from cells and either OptiMEM or transfection media was added. 100-200 µL transfection mix was added dropwise to each well depending on assay and well size.

6.3.3.2 Xfect transfection

DF1 cells were transfected with Xfect reagent (Takara Bio) following manufacturer's instructions. Briefly, cells were seeded the day before in order to get 70-80% confluency on the day of transfection. The desired concentration of DNA was diluted in Xfect reaction buffer, and the required amount of Xfect polymer added. Reactions were incubated at RT for 5 minutes. During this time, the media in wells was reduced by half. 25 µL transfection mix was added dropwise to each well.

6.3.4 Influenza minireplicon reporter assays

Monolayers of DF1 or QT-35 cells were seeded to 80% confluency, then co-transfected in triplicate with 100 ng each pDUAL plasmid encoding PB2, PA, and NP, in addition to 50 ng of a PolI plasmid containing a reverse oriented firefly luciferase reporter flanked by the 5 and 3' UTRs of segment 8, 60 ng of a Renilla luciferase plasmid, and 100 ng of various PB1 plasmids. As a negative control, transfections lacking the PB1 plasmid (empty vector was used to

balance plasmid intake) were also performed (2PNP). Two days post transfection, medium was removed, and cells were lysed according to manufacturers instructions (DUAL Glo Luciferase Assay Kit, Promega). Plates were frozen at -80ºC for 30 minutes, and thawed at RT on a rocker. Luminescence was measured using 10 µL lysate, and 25 µL substrate in opaque white 96-well plates (Pierce). 25 µL STOP&GLO was then added and luminescence measured again as a transfection control. Luciferase activity was analysed with a one-way ANOVA vs WT, and presented as % WT activity. Renilla activity was measured to confirm equal transfection but was not used for analysis of multiple repeats.

6.3.5 IFN-β promoter reporter assays

Monolayers of DF1 cells were seeded to 80% confluency, then co-transfected with 500 ng IFN reporter plasmid, 100 ng Renilla luciferase, 500 ng PB1-F2-GFP, or 500 ng GFP as a negative control. 24 hours post transfection, cells were either stimulated with 250 ng poly(I:C) or transfected with 250 ng GFP as a negative control. 16 host post stimulation, cells were lysed with passive lysis buffer according to manufacturers instructions, and luciferase activity measured as in section 6.3.4.

6.4 Virological assays

6.4.1 Generation of P0 viral stocks

PR8 wild-type and PR8-5092 reassortant viruses were rescued using a previously described reverse genetics system (De Wit et al., 2004; Hoffmann & Webster, 2000). Briefly, monolayers of 293T cells in 6-well plates were transfected with each PR8 segment cDNA cloned into individual pDUAL plasmids (250 ng of each) along with 4 µL of Lipofectamine2000 in a total volume of 800 µL of Opti-MEM. After overnight incubation at 37ºC, the transfection medium was replaced with fresh virus growth medium (DMEM, 2 mM L-glutamine, 100

U/mL penicillin, 100 µg/mL streptomycin, 0.14% BSA, 1 µg/mL TPCK trypsin). Cells were incubated for 48 hours, after which supernatants were harvested, cleared by centrifugation and stored at -80°C until future use. For reassortant viruses, seven PR8 segments with 5092 PB1 were transfected.

6.4.2 Generation of cell-grown P1 viral stocks

A first passage stock was grown in confluent monolayers of MDCKs where cells were infected with an estimated MOI of 0.01 using the P0 stock. Following a minimum of 48 hour incubation, or when infected cells showed cytopathic effect (CPE), cell supernatants were harvested, cleared by centrifugation (1500rpm, 5 minutes), aliquoted and stored at -80ºC until titrated by plaque assay.

6.4.3 Generation of egg-grown P1 viral stocks

A first passage stock was grown in VALO leghorn embryonated chicken eggs. 200 µL P0 stock was used to infect the egg at 10 days old. Eggs were incubated at 37 °C and 50 % relative humidity until either 14 days old or death of the embryo. Allantoic fluid was harvested, clarified by centrifugation (3000rpm, 20 minutes), aliquoted and stored at -80ºC until titrated by plaque assay.

6.4.4 Quantification of viral stocks and samples by plaque assay

MDCKs were seeded in 12 well plates in DMEM/10% FBS and incubated overnight at 37°C/5% CO_2. Media was removed and cells washed twice with PBS to remove residual serum. Virus samples were serially diluted 10-fold and 200 µL applied to each well. Plates were incubated at 37°C for one hour to allow adsorption, then the inoculum removed and cells covered with 1 mL/well agarose overlay. Plates were returned to the incubator for 48h and then fixed with 500 µL/well methanol-acetone before immunostaining.

6.4.5 Immunostaining of plaque assays

Cells were fixed with ice cold methanol:acetone (50:50) for 15 mins, before removal and replacement with PBS. Cells were washed with fresh PBS for 5 mins RT with shaking before being blocked with 5% BSA/PBS for 1hr RT with shaking. The blocking solution was removed, and cells washed with PBS. Cells were incubated with 1:3000 primary NP antibody in 1% BSA/PBS for 1hr RT with shaking. Cells were washed 3x 5min with PBS/T RT with shaking. Cells were incubated with 1:1000 secondary antibody in 1% BSA/PBS for 1hr RT with shaking. Cells were washed 3x 5min with PBS/T RT with shaking. Plates were then scanned on the LiCor Odessy CLx.

6.4.6 Quantification of viral stocks and samples by heamagglutination assay

Virus stocks grown in eggs were 2-fold serial diluted in PBS in a final volume of 100 µL after which 50 µL of 1% chicken blood cells were added to each well. After a 30 minute incubation at room-temperature, haemagglutination-positive wells were recorded. Each plate contained a negative (PBS only) and a positive sample for haemagglutination comparison.

6.4.7 Viral RNA isolation and sequencing

First passage viral stocks, and growth curve samples, had RNA extracted using the QIAamp Viral RNA Kit via the spin protocol according to manufacturer's guidelines. Briefly, 140 µL of P1 stock or virus sample was mixed with a lysis buffer to release RNA from the virions which was then bound to a silica-based membrane. After a series of alcohol-based washes, RNA was eluted from the column using 40 µL of milliQ water.

cDNA was produced via RT-PCR using the Verso cDNA kit according to manufacturer's instructions, with a total reaction volume of 20 µL. The reverse transcription was performed at 48ºC for 30 minutes. Samples were then denatured at 94ºC for 1 minute,

followed by 28 cycles of 15 seconds 94ºC, 30 seconds 60ºC, and 3 minutes 68ºC, finishing with 5 minutes at 68ºC. cDNA was then amplified with 100 ng PB1 or HA specific primers in 50 µL reactions for sequencing, using 35 cycles of the amplification stage of RT-PCR programme.

6.4.8 Viral multicycle growth kinetic analysis

Confluent monolayers of MDCKs were infected with PR8-5092 reassortant viruses at MOI 0.001 in 200 µL serum free DMEM in 12 well plates. After a 1-hour incubation at 37°C (which allowed virus attachment to the cells) inoculum was replaced by 1 mL of fresh virus growth medium. At different times post infection, supernatants were frozen, and samples were stored at -80°C until titrated by plaque assay.

6.5 Protein purification and detection

6.5.1 SDS polyacrylamide gel electrophoresis

Polypeptides were separated using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Polyacrylamide gels were cast using approximately 5 mL per gel of the resolving gel with 1mm spacer. Acrylamide:bisacrylamide content was routinely 12% resolving gels and 4% stacking gels.
Commercial gradient gels 4-20% were also used (BioRad mini protean TGX).

Lysates were generated by adding 100-200 µL Laemmli/SDS lysis/passive lysis buffer to cells depending on the assay. Samples in passive lysis buffer had SDS loading dye added. Samples were transferred into clean 1.5 mL tubes, boiled at 95°C for 10 minutes, vortexed for 15 seconds and centrifuged at 10,000rpm for 5 minutes. Each lane was loaded with 3-20 µL of lysate, 5 µL ladder was added to the first lane and gels were run at 120V.

6.5.2 Western blot

After separation by SDS-PAGE according to the previous section 6.5.1, proteins were transferred to 45 µm nitrocellulose membrane in a BioRad TransBlot Turbo machine using Tris Glycine Methanol (TGM) transfer buffer. Membranes were blocked in PBS/T with 5% milk for 1 hour and stained with primary antibodies diluted in blocking buffer overnight at 4ºC. After three 5 minute PBS/T washes, membranes were incubated with secondary antibodies diluted in blocking buffer for 1 hour at room temperature in the dark. Membranes were washed again, and imaged on the LiCor Odessy CLx with the ImageStudio Lite software.

6.5.3 Densitometry

Densitometric analysis of polypeptide abundance was performed using Image J software (Schneider et al., 2012). Concisely, ImageStudio acquired images were compiled and saved as TIFF files. Sample lanes were defined using the gel analysing tool, including both loading control and bands of interest in each lane. Once selected, pixel intensity of each band in the lane was plotted and the area under each peak curve above background was calculated.

For each timepoint, time zero was made equal to 1.0, and all other timepoints made relative to time zero. The sample band was then made relative to the loading control. A $\log_{10}$ transformation was performed, and half-lives calculated from a non-linear regression of a straight line in GraphPad.

6.6 Flourescent imaging and staining

6.6.1 Immunoflourescence staining

DF1s were seeded on coverslips in 24 well plates and Xfect transfected with 1 µg of the appropriate plasmid. Cells were fixed at 24h post-transfection with 10% formalin (Sigma) for 20 mins, washed 3 times for 5 minutes in PBS/1% BSA with rocking, permeabilised

with 1% Triton X-100 (Sigma) PBS for 5 mins, and washed again. GFP tagged protein slides not requiring antibodies were then mounted on slides with VectaShield mounting medium with DAPI. Coverslips for V5 tagged protein slides were incubated with 100 µL primary Ab for 1hr RT with shaking, washed as before, incubated with 100 µL 488nm secondary Ab, and washed again. Coverslips were mounted with Vectashield mounting medium with DAPI (Vector Laboratories Inc), and left at RT overnight. Coverslips were then sealed with nail varnish.

6.6.2 Confocal microscopy

Confocal fluorescent imaging was carried out on a Leica Stellaris 5 confocal microscope, with a 63x oil immersion objective lens. Images were taken using Leica Application Suite (LAS) X, and analysed for colocalisation using LAS AF Lite.

6.9 Bioinformatic analyses

6.9.1 Sequence collection and annotation

Bioinformatic analyses were performed in collaboration with Dr Samantha Lycett (The Roslin Institute, University of Edinburgh). PB1 nucleotide sequences from 1902-2018 were acquired from the NCBI Influenza Virus Database by our collaborator, Dr Samantha Lycett. Sequences were curated, collated, and annotated with a custom R code by Dr Lycett. Further details of the sampling process are available (Lycett et al., 2019).

AUG and STOP codon positions were annotated using R scripts and further analysed in Microsoft Excel and FigTree. All R scripts were written and kindly provided by Dr Samantha Lycett.

6.9.2 Phylogenetic analysis

To assess persistence of N-terminal truncations in PB1-F2 sequences, phylogenetic relationships between sequences were

assessed using ClusterPicker and ClusterMatcher programmes (Ragonnet-Cronin et al., 2013).
A FigTree phylogenetic tree and BioEdit sequence alignment file were inputted to ClusterPicker, under the parameters 0.9 support threshold and 0.045 genetic distance. Clusters of closely related sequences were outputted (Ragonnet-Cronin et al., 2013).
The ClusterPicker output phylogenetic tree and an Excel annotation file with a binary choice of full length or truncated PB1-F2 were inputted to ClusterMatcher, under the parameters 3 sequences in a cluster, 1% truncated sequences. Closely related clusters with a high rate of truncation were outputted. Fully truncated clusters were further investigated by literature search, geographical journey, and precursor virus investigation. Possible precursor or ancestral viruses were investigated to attempt to discover a non-truncated ancestor and possible source or timing of the introduction of the truncation to the PB1-F2 sequence.

6.10 Statistical analysis

All statistic tests were performed in GraphPad Prism 8 Software. Specific statistical tests were chosen for different analyses and are stated in each figure legend.

ns: non-significant

*: $p<0.05$

**: $p<0.01$

***: $p<0.001$

****: $p<0.0001$

6.11 Structure modelling

Protein structure predictions were performed through the I-TASSER online server, using the pre-set parameters (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al. 2011;

Yang et al. 2015). All proteins models were analysed and visualised using PyMOL Molecular graphics system using the default settings (https://pymol.org/2/).

Bibliography

Ajjaji, D., Richard, C. A., Mazerat, S., Chevalier, C., & Vidic, J. (2016). N-terminal domain of PB1-F2 protein of influenza A virus can fold into amyloid-like oligomers and damage cholesterol and cardiolipid containing membranes. *Biochemical and Biophysical Research Communications*, *477*(1), 27–32. https://doi.org/10.1016/j.bbrc.2016.06.016

Akkina, R. K. (1990). Antigenic reactivity and electrophoretic migrational heterogeneity of the three polymerase proteins of type A human and animal influenza viruses. *Archives of Virology*, *111*, 187–197. https://doi.org/10.1007/BF01311053

Alexander, D. J. (2007). An overview of the epidemiology of avian influenza. *Vaccine*, *25*(30 SPEC. ISS.), 5637–5644. https://doi.org/10.1016/j.vaccine.2006.10.051

Alymova, I. V., Samarasinghe, A., Vogel, P., Green, A. M., Weinlich, R., & McCullers, J. A. (2014). A Novel Cytotoxic Sequence Contributes to Influenza A Viral Protein PB1-F2 Pathogenicity and Predisposition to Secondary Bacterial Infection. *Journal of Virology*, *88*(1), 503–515. https://doi.org/10.1128/jvi.01373-13

Ampomah, P. B., & Lim, L. H. K. (2020). Influenza A virus-induced apoptosis and virus propagation. *Apoptosis: An International Journal on Programmed Cell Death*, *25*(1–2), 1–11. https://doi.org/10.1007/s10495-019-01575-3

Arranz, R., Coloma, R., Chichón, F. J., Conesa, J. J., Carrascosa, J. L., Valpuesta, J. M., Ortín, J., & Martín-Benito, J. (2012). The structure

of native influenza virion ribonucleoproteins. *Science*, *338*(6114), 1634–1637. https://doi.org/10.1126/science.1228172

Artarini, A., Meyer, M., Shin, Y. J., Huber, K., Hilz, N., Bracher, F., Eros, D., Orfi, L., Keri, G., Goedert, S., Neuenschwander, M., von Kries, J., Domovich-Eisenberg, Y., Dekel, N., Szabadkai, I., Lebendiker, M., Horváth, Z., Danieli, T., Livnah, O., … Karlas, A. (2019). Regulation of influenza a virus mRNA splicing by CLK1. *Antiviral Research*, *168*, 187–196. https://doi.org/10.1016/j.antiviral.2019.06.003

Barber, M. R. W., Aldridge, J. R., Webster, R. G., & Magor, K. E. (2010). Association of RIG-I with innate immunity of ducks to influenza. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(13), 5913–5918. https://doi.org/10.1073/pnas.1001755107

Barjesteh, N., Dowd, K. O., & Milad, S. (2020). Antiviral responses against chicken respiratory infections: Focus on avian influenza virus and infectious bronchitis virus. *Cytokine*, *127*. https://doi.org/10.1016/j.cyto.2019.154961

Batres-Marquez, S. P. (2017, June 29). *U.S. Duck Production and Exports*. Iowa Farm Bureau. https://www.iowafarmbureau.com/Article/US-Duck-Production-and-Exports

Bazzigher, L., Schwartz, A., & Staeheli, P. (1993). No enhanced influenza virus resistance of murine and avian cells expressing cloned duck Mx protein. *Virology*, *195*(1), 100–112. https://doi.org/10.1006/VIRO.1993.1350

Belshaw, R., Pybus, O. G., & Rambaut, A. (2007). The evolution of genome compression and genomic novelty in RNA viruses. *Genome Research*, *17*(10), 1496–1504. https://doi.org/10.1101/gr.6305707

Benfield, C., Lyall, J., Kochs, G., & Tiley, L. (2008). Asparagine 631 variants of the chicken Mx protein do not inhibit influenza virus replication in primary chicken embryo fibroblasts or in vitro surrogate assays. *Journal of Virology*, *82*(15), 7533–7539. https://doi.org/10.1128/JVI.00185-08

Bergervoet, S. A., Heutink, R., Bouwstra, R., Fouchier, R. A. M., & Beerens, N. (2019). Genetic analysis identifies potential transmission of low pathogenic avian influenza viruses between poultry farms. *Transboundary and Emerging Diseases*, *66*(4), 1653–1664. https://doi.org/10.1111/TBED.13199

Bernasconi, D., Schultz, U., & Staeheli, P. (1995). The Interferon-Induced Mx Protein of Chickens Lacks Antiviral Activity. *Journal of Interferon and Cytokine Research*, *15*(1), 47–53. https://doi.org/10.1089/JIR.1995.15.47

Bian, Q., Lu, J., Zhang, L., Chi, Y., Li, Y., & Guo, H. (2017). Highly pathogenic avian influenza A virus H5N1 non-structural protein 1 is associated with apoptotic activation of the intrinsic mitochondrial pathway. *Experimental and Therapeutic Medicine*, *14*(5), 4041–4046. https://doi.org/10.3892/etm.2017.5056

Bogdanow, B., Wang, X., Eichelbaum, K., Sadewasser, A., Husic, I., Paki, K., Budt, M., Hergeselle, M., Vetter, B., Hou, J., Chen, W., Wiebusch, L., Meyer, I. M., Wolff, T., & Selbach, M. (2019). The dynamic proteome of influenza A virus infection identifies M segment splicing as a host range determinant. *Nature Communications*. https://doi.org/10.1038/s41467-019-13520-8

Brown, J. D., Swayne, D. E., Cooper, R. J., Burns, R. E., & Stallknecht, D. E. (2007). Persistence of H5 and H7 Avian Influenza Viruses in

Water. *Avian Diseases*, *51*(s1), 285–289. https://doi.org/10.1637/7636-042806R.1

Bruns, K., Studtrucker, N., Sharma, A., Fossen, T., Mitzner, D., Eissmann, A., Tessmer, U., Röder, R., Henklein, P., Wray, V., & Schubert, U. (2007). Structural characterization and oligomerization of PB1-F2, a proapoptotic influenza A virus protein. *Journal of Biological Chemistry*, *282*(1), 353–363. https://doi.org/10.1074/jbc.M606494200

Buehler, J., Navi, D., Lorusso, A., Vincent, A., Lager, K., & Miller, C. L. (2013). Influenza A Virus PB1-F2 Protein Expression Is Regulated in a Strain-Specific Manner by Sequences Located Downstream of the PB1-F2 Initiation Codon. *Journal of Virology*, *87*(19), 10687–10699. https://doi.org/10.1128/jvi.01520-13

Bui, M., Whittaker, G., & Helenius, A. (1996). Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *Journal of Virology*, *70*(12), 8391–8401. https://doi.org/10.1128/jvi.70.12.8391-8401.1996

Bullough, P. A., Hughson, F. M., Skehel, J. J., & Wiley, D. C. (1994). Structure of influenza HA at the pH of membrane fusion. *Nature*, *371*(September), 37–43. https://doi.org/10.1038/371037a0

Calder, L. J., Wasilewski, S., Berriman, J. A., & Rosenthal, P. B. (2010). Structural organization of a filamentous influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(23), 10685–10690. https://doi.org/10.1073/pnas.1002123107

Cardona, C. J., Xing, Z., Sandrock, C. E., & Davis, C. E. (2009). Avian influenza in birds and mammals. *Comparative Immunology,*

Microbiology and Infectious Diseases, *32*, 255–273. https://doi.org/10.1016/j.cimid.2008.01.001

Carnaccini, S., & Perez, D. R. (2020). H9 Influenza Viruses: An Emerging Challenge. *Cold Spring Harbor Perspectives in Medicine*, *10*(6). https://doi.org/10.1101/CSHPERSPECT.A038588

Chaimayo, C., Hayashi, T., Underwood, A., Hodges, E., & Takimoto, T. (2017). Selective incorporation of vRNP into influenza A virions determined by its specific interaction with M1 protein. *Virology*, *505*, 23–32. https://doi.org/10.1016/j.virol.2017.02.008

Chang, P., Kuchipudi, S. V., Mellits, K. H., Sebastian, S., James, J., Liu, J., Shelton, H., & Chang, K. C. (2015). Early apoptosis of porcine alveolar macrophages limits avian influenza virus replication and pro-inflammatory dysregulation. *Scientific Reports*, *5*. https://doi.org/10.1038/srep17999

Chen, C.-J., Chen, G.-W., Wang, C.-H., Huang, C.-H., Wang, Y.-C., & Shih, S.-R. (2010). Differential Localization and Function of PB1-F2 Derived from Different Strains of Influenza A Virus. *Journal of Virology*, *84*(19), 10051–10062. https://doi.org/10.1128/jvi.00592-10

Chen, H., Smith, G., Zhang, S., Qin, K., Wang, J., Li, K., Webster, R. G., Peiris, J., & Guan, Y. (2005). H5N1 virus outbreak in migratory waterfow. *Nature*, *436*, 191–192. https://doi.org/10.1038/nature03974

Chen, W., Calvo, P. A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., Henklein, P., Bennink, J. R., & Yewdell, J. W. (2001). A novel influenza A virus mitochondrial protein that induces cell death. *Nature Medicine*, *7*(12), 1306–1312. https://doi.org/10.1038/nm1201-1306

Chen, Z., & Rand, R. P. (1997). The influence of cholesterol on phospholipid membrane curvature and bending elasticity. *Biophysical Journal*, *73*(1), 267–276. https://doi.org/10.1016/S0006-3495(97)78067-6

Cheng, Y.-Y., Yang, S.-R., Wang, Y.-T., Lin, Y.-H., & Chen, C.-J. (2017). Amino Acid Residues 68–71 Contribute to Influenza A Virus PB1-F2 Protein Stability and Functions. *Frontiers in Microbiology*, *8*(APR), 692. https://doi.org/10.3389/fmicb.2017.00692

Cheung, P. H. H., Lee, T. W. T., Kew, C., Chen, H., Yuen, K. Y., Chan, C. P., & Jin, D. Y. (2020). Virus subtype-specific suppression of MAVS aggregation and activation by PB1-F2 protein of influenza A (H7N9) virus. *PLoS Pathogens*, *16*(6). https://doi.org/10.1371/journal.ppat.1008611

Cheung, P. H. H., Ye, Z. W., Lee, T. W. T., Chen, H., Chan, C. P., & Jin, D. Y. (2020). PB1-F2 protein of highly pathogenic influenza A (H7N9) virus selectively suppresses RNA-induced NLRP3 inflammasome activation through inhibition of MAVS-NLRP3 interaction. *Journal of Leukocyte Biology*, *108*(5), 1655–1663. https://doi.org/10.1002/JLB.4AB0420-694R

Chevalier, C., Al Bazzal, A., Vidic, J., Février, V., Bourdieu, C., Bouguyon, E., Le Goffic, R., Vautherot, J. F., Bernard, J., Moudjou, M., Noinville, S., Chich, J. F., Da Costa, B., Rezaei, H., & Delmas, B. (2010). PB1-F2 influenza A virus protein adopts a β-sheet conformation and forms amyloid fibers in membrane environments. *Journal of Biological Chemistry*, *285*(17), 13233–13243. https://doi.org/10.1074/jbc.M109.067710

Chirico, N., Vianelli, A., & Belshaw, R. (2010). Why genes overlap in viruses. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1701), 3809–3817. https://doi.org/10.1098/rspb.2010.1052

Chizhmakov, I. V., Geraghty, F. M., Ogden, D. C., Hayhurst, A., Antoniou, M., & Hay, A. J. (1996). Selective proton permeability and pH regulation of the influenza virus M2 channel expressed in mouse erythroleukaemia cells. *The Journal of Physiology*, *494*(Pt 2), 329. https://doi.org/10.1113/JPHYSIOL.1996.SP021495

Chlanda, P., Schraidt, O., Kummer, S., Riches, J., Oberwinkler, H., Prinz, S., Kräusslich, H.-G., & Briggs, J. A. G. (2015). Structural Analysis of the Roles of Influenza A Virus Membrane-Associated Proteins in Assembly and Morphology. *Journal of Virology*, *89*(17), 8957–8966. https://doi.org/10.1128/JVI.00592-15

Chong, Z. X., Yeap, S. K., & Ho, W. Y. (2021). Transfection types, methods and strategies: A technical review. *PeerJ*, *9*. https://doi.org/10.7717/PEERJ.11165

Ciminski, K., Chase, G. P., Beer, M., & Schwemmle, M. (2021). Influenza A Viruses: Understanding Human Host Determinants. *Trends in Molecular Medicine*, *27*(2), 104–112. https://doi.org/10.1016/j.molmed.2020.09.014

Conenello, G. M., Tisoncik, J. R., Rosenzweig, E., Varga, Z. T., Palese, P., & Katze, M. G. (2011). A Single N66S Mutation in the PB1-F2 Protein of Influenza A Virus Increases Virulence by Inhibiting the Early Interferon Response In Vivo. *Journal of Virology*, *85*(2), 652–662. https://doi.org/10.1128/JVI.01987-10

Conenello, G. M., Zamarin, D., Perrone, L. A., Tumpey, T., & Palese, P. (2007). A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918

influenza A viruses contributes to increased virulence. *PLoS Pathogens*, *3*(10), 1414–1421. https://doi.org/10.1371/journal.ppat.0030141

Connor, R. J., Kawaoka, Y., Webster, R. G., & Paulson, J. C. (1994). Receptor Specificity in Human, Avian, and Equine H2 and H3 Influenza Virus Isolates. *Virology*, *205*(1), 17–23. https://doi.org/10.1006/viro.1994.1615

Copeland, C. S., Doms, R. W., Bolzau, E. M., Webster, R. G., & Helenius, A. (1986). Assembly of influenza hemagglutinin trimers and its role in intracellular transport. *The Journal of Cell Biology*, *103*(4), 1179–1191. https://doi.org/10.1083/jcb.103.4.1179

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics*, *209*(4), 1235. https://doi.org/10.1534/GENETICS.118.301058

Dadonaite, B., Gilbertson, B., Knight, M. L., Trifkovic, S., Rockman, S., Laederach, A., Brown, L. E., Fodor, E., & Bauer, D. L. V. (2019). The structure of the influenza A virus genome. *Nature Microbiology*, *4*(11), 1781–1789. https://doi.org/10.1038/s41564-019-0513-7

Dadonaite, B., Vijayakrishnan, S., Fodor, E., Bhella, D., & Hutchinson, E. C. (2016). Filamentous Influenza Viruses. *The Journal of General Virology*, *97*(8), 1755. https://doi.org/10.1099/JGV.0.000535

Dasso, M., & Jackson, R. (1989). Efficient initiation of mammalian mRNA translation at a CUG codon. *Nucleic Acids Research*, *17*(16), 6485–6496. https://doi.org/10.1093/NAR/17.16.6485

De Arce, A. J. D., Noderer, W. L., & Wang, C. L. (2018). Complete motif analysis of sequence requirements for translation initiation at non-

AUG start codons. *Nucleic Acids Research*, *46*(2), 985–994. https://doi.org/10.1093/nar/gkx1114

Department for Environment, Food & Rural Affairs, & Animal and Plant Health Agency. (2020, May 15). *Avian influenza (bird flu)—GOV.UK*. Guidance Avian Influenza (Bird Flu). https://www.gov.uk/guidance/avian-influenza-bird-flu

Deventhiran, J., Kumar, S. R. P., Raghunath, S., Leroith, T., & Elankumaran, S. (2016). PB1-F2 Protein Does Not Impact the Virulence of Triple-Reassortant H3N2 Swine Influenza Virus in Pigs but Alters Pathogenicity and Transmission in Turkeys. *Journal of Virology*, *90*(1), 222–231. https://doi.org/10.1128/jvi.01551-15

Dillon, D., & Runstadler, J. (2010). Mx gene diversity and influenza association among five wild dabbling duck species (Anas spp.) in Alaska. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *10*(7), 1085–1093. https://doi.org/10.1016/J.MEEGID.2010.07.004

Donelan, N. R., Basler, C. F., & García-Sastre, A. (2003). A Recombinant Influenza A Virus Expressing anRNA-Binding-Defective NS1 Protein Induces High Levels of BetaInterferon and Is Attenuated inMice. *Journal of Virology*, *77*(24), 13257–13266. https://doi.org/10.1128/jvi.77.24.13257-13266.2003

Dou, D., Revol, R., Östbye, H., Wang, H., & Daniels, R. (2018). Influenza A virus cell entry, replication, virion assembly and movement. *Frontiers in Immunology*, *9*(JUL), 1. https://doi.org/10.3389/fimmu.2018.01581

Dubois, C., Terrier, J., & Rosa-Calatrava, O. (2014). Influenza viruses and mRNA splicing: Doing more with less. *MBio*, *5*(3), 70–84. https://doi.org/10.1128/mBio.00070-14

Elbers, A. R. W., & Gonzales, J. L. (2020). Quantification of visits of wild fauna to a commercial free-range layer farm in the Netherlands located in an avian influenza hot-spot area assessed by video-camera monitoring. *Transboundary and Emerging Diseases*, *67*(2), 661–677. https://doi.org/10.1111/tbed.13382

Eldeeb, M. A., Siva-Piragasam, R., Ragheb, M. A., Esmaili, M., Salla, M., & Fahlman, R. P. (2019). A molecular toolbox for studying protein degradation in mammalian cells. *Journal of Neurochemistry*, *151*(4), 520–533. https://doi.org/10.1111/JNC.14838

European Centre for Disease Prevention and Control. (2021). *First identification of human cases of avian influenza A (H5N8) infection*. *24 February*.

European Commission. (2022, February 23). *Market overview by sector* [Text]. https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/markets/overviews/market-overview-sector_en

Evseev, D., & Magor, K. E. (2019). Innate immune responses to avian influenza viruses in ducks and chickens. *Veterinary Sciences*, *6*(1). https://doi.org/10.3390/VETSCI6010005

Firth, A. E., & Brierley, I. (2012). Non-canonical translation in RNA viruses. *Journal of General Virology*, *93*(PART 7), 1385–1409. https://doi.org/10.1099/vir.0.042499-0

Firth, A. E., Jagger, B. W., Wise, H. M., Nelson, C. C., Parsawar, K., Wills, N. M., Napthine, S., Taubenberger, J. K., Digard, P., & Atkins, J. F. (2012). Ribosomal frameshifting used in influenza A virus

expression occurs within the sequence UCC-UUU-CGU and is in the +1 direction. *Open Biology*, *2*(OCT). https://doi.org/10.1098/rsob.120109

Fourment, M., & Holmes, E. C. (2015). Avian influenza virus exhibits distinct evolutionary dynamics in wild birds and poultry. *BMC Evolutionary Biology*, *15*(1), 1–6. https://doi.org/10.1186/s12862-015-0410-5

Fritsch, C., Herrmann, A., Nothnagel, M., Szafranski, K., Huse, K., Schumann, F., Schreiber, S., Platzer, M., Krawczak, M., Hampe, J., & Brosch, M. (2012). Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Research*, *22*(11), 2208. https://doi.org/10.1101/GR.139568.112

Fusaro, A., Tassoni, L., Milani, A., Hughes, J., Salviato, A., Murcia, P. R., Massi, P., Zamperin, G., Bonfanti, L., Marangon, S., Cattoli, G., & Monne, I. (2016). Unexpected Interfarm Transmission Dynamics during a Highly Pathogenic Avian Influenza Epidemic. *Journal of Virology*, *90*(14), 6401–6411. https://doi.org/10.1128/jvi.00538-16

Gale, M., & Katze, M. G. (1998). Molecular mechanisms of interferon resistance mediated by viral-directed inhibition of PKR, the interferon-induced protein kinase. *Pharmacology & Therapeutics*, *78*(1), 29–46. https://doi.org/10.1016/s0163-7258(97)00165-4

Gao, H., Xu, G., Sun, Y., Qi, L., Wang, J., Kong, W., Sun, H., Pu, J., Chang, K.-C., & Liu, J. (2015). PA-X is a virulence factor in avian H9N2 influenza virus. *Journal of General Virology*, *96*(9), 2587–2594. https://doi.org/10.1099/JGV.0.000232

García, M. A., Gil, J., Ventoso, I., Guerra, S., Domingo, E., Rivas, C., & Esteban, M. (2006). Impact of protein kinase PKR in cell biology: From antiviral to antiproliferative action. *Microbiology and Molecular Biology Reviews: MMBR*, *70*(4), 1032–1060. https://doi.org/10.1128/MMBR.00027-06

García-Sastre, A., Egorov, A., Matassov, D., Brandt, S., Levy, D. E., Durbin, J. E., Palese, P., & Muster, T. (1998). Influenza A Virus Lacking the NS1 Gene Replicates in Interferon-Deficient Systems. *Virology*, *252*(2), 324–330. https://doi.org/10.1006/VIRO.1998.9508

Gaush, C. R., & Smith, T. F. (1968). Replication and Plaque Assay of Influenza Virus in an Established Line of Canine Kidney Cells. *Applied Microbiology*, *16*(4), 588–594.

Ghanem, A., Mayer, D., Chase, G., Tegge, W., Frank, R., Kochs, G., García-Sastre, A., & Schwemmle, M. (2007). Peptide-Mediated Interference with Influenza A Virus Polymerase. *Journal of Virology*, *81*(14), 7801. https://doi.org/10.1128/JVI.00724-07

Gonzales, J. L., Boender, G. J., Elbers, A. R. W., Stegeman, J. A., & de Koeijer, A. A. (2014). Risk based surveillance for early detection of low pathogenic avian influenza outbreaks in layer chickens. *Preventive Veterinary Medicine*, *117*(1), 251–259. https://doi.org/10.1016/j.prevetmed.2014.08.015

Gu, W., Gallagher, G. R., Dai, W., Liu, P., Li, R., Trombly, M. I., Gammon, D. B., Mello, C. C., Wang, J. P., & Finberg, R. W. (2015). Influenza A virus preferentially snatches noncoding RNA caps. *RNA*, *21*(12), 2067. https://doi.org/10.1261/RNA.054221.115

Guan, K., Zheng, Z., Song, T., He, X., Xu, C., Zhang, Y., Ma, S., Wang, Y., Xu, Q., Cao, Y., Li, J., Yang, X., Ge, X., Wei, C., & Zhong, H.

(2013). MAVS Regulates Apoptotic Cell Death by Decreasing K48-Linked Ubiquitination of Voltage-Dependent Anion Channel 1. *Molecular and Cellular Biology*, *33*(16). https://doi.org/10.1128/MCB.00030-13

Guerra, S., López-Fernández, L. A., García, M. A., Zaballos, A., & Esteban, M. (2006). Human Gene Profiling in Response to the Active Protein Kinase, Interferon-induced Serine/threonine Protein Kinase (PKR), in Infected Cells. Involvement of the transcription factor ATF-3 IN PKR-induced apoptosis. *Journal of Biological Chemistry*, *281*(27), 18734–18745. https://doi.org/10.1074/jbc.M511983200

Hai, R., Schmolke, M., Varga, Z. T., Manicassamy, B., Wang, T. T., Belser, J. A., Pearce, M. B., García-Sastre, A., Tumpey, T. M., & Palese, P. (2010). PB1-F2 Expression by the 2009 Pandemic H1N1 Influenza Virus Has Minimal Impact on Virulence in Animal Models. *Journal of Virology*, *84*(9), 4442–4450. https://doi.org/10.1128/JVI.02717-09

Haimov, O., Sinvani, H., Martin, F., Ulitsky, I., Emmanuel, R., Tamarkin-Ben-Harush, A., Vardy, A., & Dikstein, R. (2017). Efficient and Accurate Translation Initiation Directed by TISU Involves RPS3 and RPS10e Binding and Differential Eukaryotic Initiation Factor 1A Regulation. *Molecular and Cellular Biology*, *37*(15), e00150-17. https://doi.org/10.1128/MCB.00150-17

Halvorson, D., Kelleher, C., & Senne, D. (1985). Epizootiology of avian influenza: Effect of season on incidence in sentinel ducks and domestic turkeys in Minnesota. *Applied and Environmental Microbiology*, *49*(4), 914–919. https://doi.org/10.1128/AEM.49.4.914-919.1985

Han, X., Bertzbach, L. D., & Veit, M. (2019). Mimicking the passage of avian influenza viruses through the gastrointestinal tract of chickens. *Veterinary Microbiology*, *239*, 108462. https://doi.org/10.1016/j.vetmic.2019.108462

Hanson, B. A., Luttrell, M. P., Goekjian, V. H., Niles, L., Swayne, D. E., Senne, D. A., & Stallknecht, D. E. (2008). Is the occurrence of avian influenza virus in Charadriiformes species and location dependent? *Journal of Wildlife Diseases*, *44*(2), 351–361. https://doi.org/10.7589/0090-3558-44.2.351

Hao, W., Wang, L., & Li, S. (2020). Roles of the non-structural proteins of influenza a virus. *Pathogens*, *9*(10), 1–19. https://doi.org/10.3390/pathogens9100812

Hatada, E., & Fukuda, R. (1992). Binding of influenza A virus NS1 protein to dsRNA in vitro. *Journal of General Virology*, *73*(12), 3325–3329. https://doi.org/10.1099/0022-1317-73-12-3325

Hay, A. J. (1974). Studies on the Formation of the Influenza Virus Envelope. *Virology*, *60*, 398–418. https://doi.org/10.1016/0042-6822(74)90335-3

Hayashi, T., MacDonald, L. A., & Takimoto, T. (2015). Influenza A Virus Protein PA-X Contributes to Viral Growth and Suppression of the Host Antiviral and Immune Responses. *Journal of Virology*, *89*(12), 6442. https://doi.org/10.1128/JVI.00319-15

Hayashi, T., Watanabe, C., Suzuki, Y., Tanikawa, T., Uchida, Y., & Saito, T. (2014). Chicken MDA5 Senses Short Double-Stranded RNA with Implications for Antiviral Response against Avian Influenza Viruses in Chicken. *Journal of Innate Immunity*, *6*(1), 58–71. https://doi.org/10.1159/000351583

Hill, N. J., Ma, E. J., Meixell, B. W., Lindberg, M. S., Boyce, W. M., & Runstadler, J. A. (2016). Transmission of influenza reflects seasonality of wild birds across the annual cycle. *Ecology Letters*, *19*(8), 915–925. https://doi.org/10.1111/ELE.12629

Hinshaw, V. S., Olsen, C. W., Dybdahl-Sissoko, N., & Evans, D. (1994). Apoptosis: A mechanism of cell killing by influenza A and B viruses. *Journal of Virology*, *68*(6), 3667–3673. https://doi.org/10.1128/jvi.68.6.3667-3673.1994

Hirst, J. C., & Hutchinson, E. C. (2019). Single-particle measurements of filamentous influenza virions reveal damage induced by freezing. *The Journal of General Virology*, *100*(12), 1631–1640. https://doi.org/10.1099/jgv.0.001330

Ho, J. S. Y., Angel, M., Ma, Y., Sloan, E., Wang, G., Martinez-Romero, C., Alenquer, M., Roudko, V., Chung, L., Zheng, S., Chang, M., Fstkchyan, Y., Clohisey, S., Dinan, A. M., Gibbs, J., Gifford, R., Shen, R., Gu, Q., Irigoyen, N., … Marazzi, I. (2020). Hybrid Gene Origination Creates Human-Virus Chimeric Proteins during Infection. *Cell*, *181*(7), 1502-1517.e23. https://doi.org/10.1016/j.cell.2020.05.035

Honda, A., Ueda, K., Nagata, K., & Ishihama, A. (1988). RNA Polymerase of Influenza Virus: Role of NP in RNA Chain Elongation 1. *J. Biochem*, *104*(6), 1021–1026. https://doi.org/10.1093/oxfordjournals.jbchem.a122569

Hronová, V., Mohammad, M. P., Wagner, S., Pánek, J., Gunišová, S., Zeman, J., Poncová, K., & Valášek, L. S. (2017). Does eIF3 promote reinitiation after translation of short upstream ORFs also in

mammalian cells? *RNA Biology*, *14*(12), 1660. https://doi.org/10.1080/15476286.2017.1353863

Hsiang, T., Zhou, L., & Krug, R. M. (2012). Roles of the phosphorylation of specific serines and threonines in the NS1 protein of human influenza A viruses. *Journal of Virology*, *86*(19), 10370–10376. https://doi.org/10.1128/JVI.00732-12

Hutchinson, E. C., Charles, P. D., Hester, S. S., Thomas, B., Trudgian, D., Martínez-Alonso, M., & Fodor, E. (2014). Conserved and host-specific features of influenza virion architecture. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms5816

Ingolia, N. T., Hussmann, J. A., & Weissman, J. S. (2019). Ribosome Profiling: Global Views of Translation. *Cold Spring Harbor Perspectives in Biology*, *11*(5). https://doi.org/10.1101/CSHPERSPECT.A032698

Jackson, R. J., T Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, *10*. https://doi.org/10.1038/nrm2838

Jagger, B. W., Wise, H. M., Kash, J. C., Walters, K.-A. A., Wills, N. M., Xiao, Y.-L. L., Dunfee, R. L., Schwartzman, L. M., Ozinsky, A., Bell, G. L., Dalton, R. M., Lo, A., Efstathiou, S., Atkins, J. F., Firth, A. E., Taubenberger, J. K., & Digard, P. (2012). An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science*, *337*(6091), 199–204. https://doi.org/10.1126/science.1222213

James, J., Howard, W., Iqbal, M., Nair, V. K., Barclay, W. S., & Shelton, H. (2016). Influenza a virus PB1-F2 protein prolongs viral shedding in

chickens lengthening the transmission window. *Journal of General Virology*, *97*(10), 2516–2527. https://doi.org/10.1099/jgv.0.000584

James, J., Smith, N., Ross, C., Iqbal, M., Goodbourn, S., Digard, P., Barclay, W. S., & Shelton, H. (2019). The cellular localization of avian influenza virus PB1-F2 protein alters the magnitude of IFN2 promoter and NFκB-dependent promoter antagonism in chicken cells. *Journal of General Virology*, *100*(3), 414–430. https://doi.org/10.1099/jgv.0.001220

Jaworska, J., Coulombe, F., Downey, J., Tzelepis, F., Shalaby, K., Tattoli, I., Berube, J., Rousseau, S., Martin, J. G., Girardin, S. E., McCullers, J. A., & Divangahi, M. (2014). NLRX1 prevents mitochondrial induced apoptosis and enhances macrophage antiviral immunity by interacting with influenza virus PB1-F2 protein. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), E2110–E2119. https://doi.org/10.1073/pnas.1322118111

Jeong, O. M., Kim, M. C., Kim, M. J., Kang, H. M., Kim, H. R., Kim, Y. J., Joh, S. J., Kwon, J. H., & Lee, Y. J. (2009). Experimental infection of chickens, ducks and quails with the highly pathogenic H5N1 avian influenza virus. *Journal of Veterinary Science*, *10*(1), 53–60. https://doi.org/10.4142/jvs.2009.10.1.53

Jones, I. M., Reay, P. A., & Philpott, K. L. (1986). Nuclear location of all three influenza polymerase proteins and a nuclear signal in polymerase PB2. *The EMBO Journal*, *5*(9), 2371–2376. https://doi.org/10.1002/j.1460-2075.1986.tb04506.x

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A.,

Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kamal, R. P., Alymova, I. V., & York, I. A. (2018). Evolution and virulence of influenza A virus protein PB1-F2. *International Journal of Molecular Sciences*, *19*(1). https://doi.org/10.3390/ijms19010096

Kamal, R. P., Kumar, A., Davis, C. T., Tzeng, W.-P., Nguyen, T., Donis, R. O., Katz, J. M., & York, I. A. (2015). Emergence of Highly Pathogenic Avian Influenza A(H5N1) Virus PB1-F2 Variants and Their Virulence in BALB/c Mice. *Journal of Virology*, *89*(11), 5835–5846. https://doi.org/10.1128/jvi.03137-14

Karpala, A. J., Lowenthal, J. W., & Bean, A. G. (2008). Activation of the TLR3 pathway regulates IFNβ production in chickens. *Developmental and Comparative Immunology*, *32*(4), 435–444. https://doi.org/10.1016/j.dci.2007.08.004

Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T. S., Fujita, T., & Akira, S. (2008). Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid–inducible gene-I and melanoma differentiation–associated gene 5. *The Journal of Experimental Medicine*, *205*(7), 1601. https://doi.org/10.1084/JEM.20080091

Kawaoka, Y., Chambers, T. M., Sladen, W. L., & Gwebster, R. (1988). Is the gene pool of influenza viruses in shorebirds and gulls different from that in wild ducks? *Virology*, *163*(1), 247–250. https://doi.org/10.1016/0042-6822(88)90260-7

Kayed, A. S., Kandeil, A., Gomaa, M. R., El-Shesheny, R., Mahmoud, S., Hegazi, N., Fayez, M., Sheta, B., McKenzie, P. P., Webby, R. J., Kayali, G., & Ali, M. A. (2019). Surveillance for avian influenza viruses in wild birds at live bird markets, Egypt, 2014-2016. *Influenza and Other Respiratory Viruses*, *13*(4), 407–414. https://doi.org/10.1111/irv.12634

Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent evidence. *Idengaku Zasshi*, *66*(4), 367–386. https://doi.org/10.1266/jjg.66.367

Klemm, C., Boergeling, Y., Ludwig, S., & Ehrhardt, C. (2018). Immunomodulatory Nonstructural Proteins of Influenza A Viruses. *Trends in Microbiology*, *26*(7), 624–636. https://doi.org/10.1016/j.tim.2017.12.006

Košík, I., Práznovská, M., Košíková, M., Bobišová, Z., Hollý, J., Varečková, E., Kostolanský, F., & Russ, G. (2015). The ubiquitination of the influenza a virus PB1-F2 protein is crucial for its biological function. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0118477

Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the aug initiator codon by eukaryotic ribosomes. *Nucleic Acids Research*, *9*(20), 5233–5252. https://doi.org/10.1093/nar/9.20.5233

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. In *Nucleic Acids Research* (Vol. 12, Issue 2, pp. 857–872). https://doi.org/10.1093/nar/12.2.857

Kozak, M. (2001). Constraints on reinitiation of translation in mammals. *Nucleic Acids Research*, *29*(24), 5226–5232. https://doi.org/10.1093/nar/29.24.5226

Krug, R. M. (2015). Functions of the Influenza A Virus NS1 Protein In Antiviral Defense. *Current Opinion in Virology*, *12*, 1. https://doi.org/10.1016/J.COVIRO.2015.01.007

Krumbholz, A., Philipps, A., Oehring, H., Schwarzer, K., Eitner, A., Wutzler, P., & Zell, R. (2011). Current knowledge on PB1-F2 of influenza A viruses. *Med Microbiol Immunol*, *200*, 69–75. https://doi.org/10.1007/s00430-010-0176-8

Kuchipudi, S. V., Dunham, S. P., Nelli, R., AWhite, G., Coward, V. J., Slomka, M. J., Brown, I. H., & Chang, K. C. (2012). Rapid death of duck cells infected with influenza: A potential mechanism for host resistance to H5N1. *Immunology and Cell Biology*, *90*(1), 116–123. https://doi.org/10.1038/icb.2011.17

Kuo, R.-L., Li, L.-H., Lin, S.-J., Li, Z.-H., Chen, G.-W., Chang, C.-K., Wang, Y.-R., Tam, E.-H., Gong, Y.-N., Krug, R. M., & Shih, S.-R. (2016). *Role of N Terminus-Truncated NS1 Proteins of Influenza A Virus in Inhibiting IRF3 Activation*. https://doi.org/10.1128/JVI.02843-15

Lakadamyali, M., Rust, M. J., & Zhuang, X. (2004). Endocytosis of influenza viruses. *Microbes and Infection*, *6*(10), 929–936. https://doi.org/10.1016/j.micinf.2004.05.002

Lam, T. T. Y., Zhou, B., Wang, J., Chai, Y., Shen, Y., Chen, X., Ma, C., Hong, W., Chen, Y., Zhang, Y., Duan, L., Chen, P., Jiang, J., Zhang, Y., Li, L., Poon, L. L. M., Webby, R. J., Smith, D. K., Leung, G. M., … Zhu, H. (2015). Dissemination, divergence and establishment of H7N9 influenza viruses in China. *Nature*, *522*(7554), 102–105. https://doi.org/10.1038/nature14348

Lamb, R. A., Lai, C. J., & Choppin, P. W. (1981). Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: Colinear

and interrupted mRNAs code for overlapping proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(7 I), 4170–4174. https://doi.org/10.1073/pnas.78.7.4170

Landolt, G. A. (2014). Equine influenza virus. *Veterinary Clinics of North America - Equine Practice*, *30*(3), 507–522. https://doi.org/10.1016/j.cveq.2014.08.003

Le Goffic, R., Bouguyon, E., Chevalier, C., Vidic, J., Da Costa, B., Leymarie, O., Bourdieu, C., Decamps, L., Dhorne-Pollet, S., & Delmas, B. (2010). Influenza A virus protein PB1-F2 exacerbates IFN-beta expression of human respiratory epithelial cells. *Journal of Immunology (Baltimore, Md. : 1950)*, *185*(8), 4812–4823. https://doi.org/10.4049/JIMMUNOL.0903952

Lee, M.-H., Arrecubieta, C., Martin, F. J., Prince, A., Borczuk, A. C., & Lowy, F. D. (2010). A Postinfluenza Model of Staphylococcus aureus Pneumonia. *The Journal of Infectious Diseases*, *201*(4), 508–515. https://doi.org/10.1086/650204

Leiding, T., Wang, J., Martinsson, J., DeGrado, W. F., & Årsköld, S. P. (2010). Proton and cation transport activity of the M2 proton channel from influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(35), 15409–15414. https://doi.org/10.1073/pnas.1009997107

Leymarie, O., Embury-Hyatt, C., Chevalier, C., Jouneau, L., Moroldo, M., Da Costa, B., Berhane, Y., Delmas, B., Weingartl, H. M., & Le Goffic, R. (2014). PB1-F2 attenuates virulence of highly pathogenic avian H5N1 influenza virus in chickens. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0100679

Leymarie, O., Jouvion, G., Hervé, P. L., Chevalier, C., Lorin, V., Lecardonnel, J., Da Costa, B., Delmas, B., Escriou, N., & Le Goffic, R. (2013). Kinetic Characterization of PB1-F2-Mediated Immunopathology during Highly Pathogenic Avian H5N1 Influenza Virus Infection. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0057894

Liang, L., Jiang, L., Li, J., Zhao, Q., Wang, J., He, X., Huang, S., Wang, Q., Zhao, Y., Wang, G., Sun, N., Deng, G., Shi, J., Tian, G., Zeng, X., Jiang, Y., Liu, L., Liu, J., Chen, P., … Li, C. (2019). Low polymerase activity attributed to PA drives the acquisition of the PB2 E627K mutation of H7N9 avian influenza virus in mammals. *MBio*, *10*(3). https://doi.org/10.1128/mBio.01162-19

Lindenmann, J., Lane, C. A., & Hobson, D. (1963). The Resistance of A2G mice to Myxoviruses. *Journal of Immunology (Baltimore, Md.: 1950)*, *90*, 942–951.

Lohmeyer, J., Talens, L. T., & Klenk, H.-D. (1979). Biosynthesis of the Influenza Virus Envelope in Abortive Infection. *Journal of General Virology*, *42*(1), 73–88. https://doi.org/10.1099/0022-1317-42-1-73

Long, J. S., Giotis, E. S., Moncorgé, O., Frise, R., Mistry, B., James, J., Morisson, M., Iqbal, M., Vignal, A., Skinner, M. A., & Barclay, W. S. (2016). Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature*, *529*(7584), 101–104. https://doi.org/10.1038/nature16474

Long, J. S., Mistry, B., Haslam, S. M., & Barclay, W. S. (2019). Host and viral determinants of influenza A virus species specificity. *Nature Reviews Microbiology*, *17*(2), 67–81. https://doi.org/10.1038/s41579-018-0115-z

Lu, L., Leigh Brown, A. J., & Lycett, S. J. (2017). Quantifying predictors for the spatial diffusion of avian influenza virus in China. *BMC Evolutionary Biology*, *17*(1). https://doi.org/10.1186/s12862-016-0845-3

Lu, L., Lycett, S. J., & Leigh Brown, A. J. (2014). Reassortment patterns of avian influenza virus internal segments among different subtypes. *BMC Evolutionary Biology*, *14*, 1–15. https://doi.org/10.1186/1471-2148-14-16

Lycett, S. J., Duchatel, F., & Digard, P. (2019). A brief history of bird flu. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1775). https://doi.org/10.1098/rstb.2018.0257

Ma, C., Cui, S., Sun, Y., Zhao, J., Zhang, D., Zhang, L., Zhang, Y., Pan, Y., Wu, S., Duan, W., Zhang, M., Yang, P., & Wang, Q. (2019). Avian influenza A (H9N2) virus infections among poultry workers, swine workers, and the general population in Beijing, China, 2013-2016: A serological cohort study. *Influenza and Other Respiratory Viruses*, *13*(4), 415–425. https://doi.org/10.1111/irv.12641

Machkovech, H. M., Bloom, J. D., & Subramaniam, A. R. (2019). Comprehensive profiling of translation initiation in influenza virus infected cells. *PLoS Pathogens*, *15*(1). https://doi.org/10.1371/journal.ppat.1007518

Maeda, T., & Ohnishi, S. ichi. (1980). Activation of influenza virus by acidic media causes hemolysis and fusion of erythrocytes. *FEBS Letters*, *122*(2), 283–287. https://doi.org/10.1016/0014-5793(80)80457-1

Mahardika, G. N., Suartha, N., Yuniati Kencana, G. A., Kade Suardana, I. B., Mahardika, W. W., & Budayanti, N. S. (2019). Biochemistry and computer-generated graph comparison of the structural and

nonstructural proteins of Spanish-1918 Influenza, pandemic-2009, and bird flu viruses. *Acta Biochimica Polonica*, *66*(3), 329–336. https://doi.org/10.18388/abp.2019_2795

Marjuki, H., Scholtissek, C., Franks, J., Negovetich, N. J., Aldridge, J. R., Salomon, R., Finkelstein, D., & Webster, R. G. (2010). Three amino acid changes in PB1-F2 of highly pathogenic H5N1 avian influenza virus affect pathogenicity in mallard ducks. *Archives of Virology*, *155*(6), 925–934. https://doi.org/10.1007/s00705-010-0666-4

Martin, K., & Helenius, A. (1991). Transport of incoming influenza virus nucleocapsids into the nucleus. *Journal of Virology*, *65*(1), 232–244. https://doi.org/10.1128/jvi.65.1.232-244.1991

Martin, V., Sims, L., Lubroth, J., Kahn, S., Domenech, J., & Begnino, C. (2006). History and evolution of HPAI viruses in Southeast Asia. *Annals of the New York Academy of Sciences*, *1081*, 153–162. https://doi.org/10.1196/annals.1373.017

Martin-Loeches, I., Van Someren Gréve, F., & Schultz, M. J. (2017). Bacterial pneumonia as an influenza complication. *Current Opinion in Infectious Diseases*, *30*(2), 201–207. https://doi.org/10.1097/QCO.0000000000000347

Maxted, A. M., Sitters, H. P., Luttrell, M. P., Dey, A. D., Kalasz, K. S., Niles, L. J., & Stallknecht, D. E. (2016). Spring Migration Stopover Ecology of Avian Influenza Virus Shorebird Hosts at Delaware Bay. *Avian Diseases*, *60*(1 Suppl), 394–405. https://doi.org/10.1637/11079-040515-Reg

Mazel-Sanchez, B., Boal Carvalho, I., Silva, F., Dijkman, R., & Schmolke, M. (2018). H5N1 influenza A virus PB1-F2 relieves HAX-1-mediated restriction of avian virus polymerase PA in human lung cells.

Journal of Virology, *1*(March), JVI.00425-18. https://doi.org/10.1128/jvi.00425-18

Mazur, I., Anhlan, D., Mitzner, D., Wixler, L., Schubert, U., & Ludwig, S. (2008). The proapoptotic influenza A virus protein PB1-F2 regulates viral polymerase activity by interaction with the PB1 protein. *Cellular Microbiology*, *10*(5), 1140–1152. https://doi.org/10.1111/j.1462-5822.2008.01116.x

McAuley, J., Deng, Y. M., Gilbertson, B., Mackenzie-Kludas, C., Barr, I., & Brown, L. (2017). Rapid evolution of the PB1-F2 virulence protein expressed by human seasonal H3N2 influenza viruses reduces inflammatory responses to infection. *Virology Journal*, *14*(1), 162. https://doi.org/10.1186/s12985-017-0827-0

Mcauley, J. L., Chipuk, J. E., Boyd, K. L., Van De Velde, N., Green, D. R., & Jonathan, A. M. C. (2010). PB1-F2 proteins from H5N1 and 20 century pandemic influenza viruses cause immunopathology. *PLoS Pathogens*, *6*(7), 1–12. https://doi.org/10.1371/JOURNAL.PPAT.1001014

McAuley, J. L., Zhang, K., & McCullers, J. A. (2010). The Effects of Influenza A Virus PB1-F2 Protein on Polymerase Activity Are Strain Specific and Do Not Impact Pathogenesis. *Journal of Virology*, *84*(1), 558–564. https://doi.org/10.1128/jvi.01785-09

Mehrbod, P., Ande, S. R., Alizadeh, J., Rahimizadeh, S., Shariati, A., Malek, H., Hashemi, M., Glover, K. K. M., Sher, A. A., Coombs, K. M., & Ghavami, S. (2019). The roles of apoptosis, autophagy and unfolded protein response in arbovirus, influenza virus, and HIV infections. *Virulence*, *10*(1), 376–413. https://doi.org/10.1080/21505594.2019.1605803

Meineke, R., Rimmelzwaan, G. F., & Elbahesh, H. (2019). Influenza Virus Infections and Cellular Kinases. *Viruses*, *11*(2). https://doi.org/10.3390/V11020171

Mettier, J., Marc, D., Sedano, L., Da Costa, B., Chevalier, C., & Le Goffic, R. (2021). Study of the host specificity of PB1-F2-associated virulence. *Virulence*, *12*(1), 1647–1660. https://doi.org/10.1080/21505594.2021.1933848

Meurs, E., Chong, K., Galabru, J., Thomas, N. S. B., Kerr, I. M., Williams, B. R. G., & Hovanessian, A. G. (1990). Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell*, *62*(2), 379–390. https://doi.org/10.1016/0092-8674(90)90374-N

Michalek, M. T., Grant, E. P., Gramm, C., Goldberg, A. L., & Rock, K. L. (1993). A role for the ubiquitin-dependent proteolytic pathway in MHC class l-restricted antigen presentation. *Nature*, *363*(6429), 552–554. https://doi.org/10.1038/363552a0

Mitzner, D., Dudek, S. E., Studtrucker, N., Anhlan, D., Mazur, I., Wissing, J., Jänsch, L., Wixler, L., Bruns, K., Sharma, A., Wray, V., Henklein, P., Ludwig, S., & Schubert, U. (2009). Phosphorylation of the influenza a virus protein PB1-F2 by PKC is crucial for apoptosis promoting functions in monocytes. *Cellular Microbiology*, *11*(10), 1502–1516. https://doi.org/10.1111/j.1462-5822.2009.01343.x

Monne, I., Fusaro, A., Nelson, M. I., Bonfanti, L., Mulatti, P., Hughes, J., Murcia, P., Schivo, A., Valastro, V., Moreno, A., Holmes, E. C., & Cattoli, G. (2014). Emergence of a highly pathogenic avian influenza virus from a low-pathogenic progenitor. *Journal of Virology*, *88*(8), 4375–4388. https://doi.org/10.1128/JVI.03181-13

Mor, A., White, A., Zhang, K., Thompson, M., Esparza, M., Muñoz-Moreno, R., Koide, K., Lynch, K. W., García-Sastre, A., & Fontoura, B. M. A. (2016). Influenza virus mRNA trafficking through host nuclear speckles. *Nature Microbiology*, *1*. https://doi.org/10.1038/NMICROBIOL.2016.69

Moriyama, M., Nagai, M., Maruzuru, Y., Koshiba, T., Kawaguchi, Y., & Ichinohe, T. (2020). Influenza Virus-Induced Oxidized DNA Activates Inflammasomes. *IScience*, *23*(7). /pmc/articles/PMC7293844/

Mosley, V. M., & Wyckoff, R. W. G. (1946). Electron micrography of the virus of influenza [2]. *Nature*, *157*(3983), 263. https://doi.org/10.1038/157263a0

Muramoto, Y., Noda, T., Kawakami, E., Akkina, R., & Kawaoka, Y. (2013). Identification of Novel Influenza A Virus Proteins Translated from PA mRNA. *Journal of Virology*, *87*(5), 2455–2462. https://doi.org/10.1128/jvi.02656-12

Muratore, G., Mercorelli, B., Goracci, L., Cruciani, G., Digard, P., Palù, G., & Loregian, A. (2012). Human Cytomegalovirus Inhibitor AL18 Also Possesses Activity against Influenza A and B Viruses. *Antimicrobial Agents and Chemotherapy*, *56*(11), 6009. https://doi.org/10.1128/AAC.01219-12

Neumann, G., Hughes, M. T., & Kawaoka, Y. (2000). Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO Journal*, *19*(24), 6751–6758. https://doi.org/10.1093/emboj/19.24.6751

Nguyen, D., Uyeki, T., Jadhao, S., Maines, T., Shaw, M., Matsuoka, Y., Smith, C., Rowe, T., Lu, X., Hall, H., Xu, X., Balish, A., Klimov, A.,

Tumpey, T., Swayne, D., Huynh, L., Nghiem, H., Nguyen, H., Hoang, L., … Katz, J. (2005). Isolation and characterization of avian influenza viruses, including highly pathogenic H5N1, from poultry in live bird markets in Hanoi, Vietnam, in 2001. *Journal of Virology*, *79*(7), 4201–4212. https://doi.org/10.1128/JVI.79.7.4201-4212.2005

Nicholls, J. M., Chan, R. W. Y., Russell, R. J., Air, G. M., & Peiris, J. S. M. (2008). Evolving complexities of influenza virus and its receptors. *Trends in Microbiology*, *16*(4), 149–157. https://doi.org/10.1016/j.tim.2008.01.008

Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R. H., & Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature*, *439*(7075), 490–492. https://doi.org/10.1038/nature04378

Nuñez, I. A., & Ross, T. M. (2019). A review of H5Nx avian influenza viruses. *Therapeutic Advances in Vaccines and Immunotherapy*, *7*, 251513551882162. https://doi.org/10.1177/2515135518821625

OIE - World Organisation for Animal Health. (2020). *Prevention and Control: OIE - World Organisation for Animal Health*. https://www.oie.int/en/animal-health-in-the-world/avian-influenza-portal/prevention-and-control/

Pappas, C., Aguilar, P. V., Basler, C. F., Solórzano, A., Zeng, H., Perrone, L. A., Palese, P., García-Sastre, A., Katz, J. M., & Tumpey, T. M. (2008). Single gene reassortants identify a critical role for PB1, HA, and NA in the high virulence of the 1918 pandemic influenza virus. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(8), 3064. https://doi.org/10.1073/PNAS.0711815105

Park, E., Byun, Y. H., Park, S., Jang, Y. H., Han, W., Won, J., Cho, K. C., Kim, D. H., Lee, A. R., Shin, G., Park, Y. K., Kang, H. S., Sim, H., Ha, Y. N., Jae, B., Son, A., Kim, P., Yu, J., Lee, H., … Kim, K. (2019). Co-degradation of interferon signaling factor DDX3 by PB1-F2 as a basis for high virulence of 1918 pandemic influenza. *The EMBO Journal*, *38*(10). https://doi.org/10.15252/embj.201899475

Pasricha, G., Mishra, A. C., & Chakrabarti, A. K. (2013). Comprehensive global amino acid sequence analysis of PB1F2 protein of influenza A H5N1 viruses and the influenza A virus subtypes responsible for the 20th-century pandemics. *Influenza and Other Respiratory Viruses*, *7*(4), 497–505. https://doi.org/10.1111/j.1750-2659.2012.00400.x

Pasricha, G., Mukherjee, S., & Chakrabarti, A. K. (2018). Apoptotic and Early Innate Immune Responses to PB1-F2 Protein of Influenza A Viruses Belonging to Different Subtypes in Human Lung Epithelial A549 Cells. *Advances in Virology*, *2018*, 5057184–5057184. https://doi.org/10.1155/2018/5057184

Pavesi, A., Vianelli, A., Chirico, N., Bao, Y., Blinkova, O., Belshaw, R., Firth, A., & Karlin, D. (2018). Overlapping genes and the proteins they encode differ significantly in their sequence composition from non-overlapping genes. *PLoS ONE*, *13*(10), e0202513. https://doi.org/10.1371/journal.pone.0202513

Petersen, H., Matrosovich, M., Pleschka, S., & Rautenschlein, S. (2012). Replication and adaptive mutations of low pathogenic avian influenza viruses in tracheal organ cultures of different avian species. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0042260

Pflug, A., Guilligay, D., Reich, S., & Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. *Nature 2014 516:7531*, *516*(7531), 355–360. https://doi.org/10.1038/nature14008

Piantadosi, C. A. (2020). Mitochondrial DNA, oxidants, and innate immunity. *Free Radical Biology & Medicine*, *152*, 455–461. https://doi.org/10.1016/j.freeradbiomed.2020.01.013

Pinar, A., Dowling, J. K., Bitto, N. J., Robertson, A. A. B., Latz, E., Stewart, C. R., Drummond, G. R., Cooper, M. A., McAuley, J. L., Tate, M. D., & Mansell, A. (2017). PB1-F2 peptide derived from avian influenza A virus H7N9 induces inflammation via activation of the NLRP3 inflammasome. *Journal of Biological Chemistry*, *292*(3), 826–836. https://doi.org/10.1074/jbc.M116.756379

Ping, J., Dankar, S. K., Forbes, N. E., Keleta, L., Zhou, Y., Tyler, S., & Brown, E. G. (2010). PB2 and Hemagglutinin Mutations Are Major Determinants of Host Range and Virulence in Mouse-Adapted Influenza A Virus. *Journal of Virology*, *84*(20), 10606–10618. https://doi.org/10.1128/jvi.01187-10

Pinto, R. M., Lycett, S., Gaunt, E., & Digard, P. (2020). Accessory Gene Products of Influenza A Virus. *Cold Spring Harbor Perspectives in Medicine*, a038380. https://doi.org/10.1101/cshperspect.a038380

Plotch, S. J., Bouloy, M., Ulmanen, I., & Krug, R. M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*, *23*(3), 847–858. https://doi.org/10.1016/0092-8674(81)90449-9

Poon, L. L. M., Pritlove, D. C., Fodor, E., & Brownlee, G. G. (1999). Direct Evidence that the Poly(A) Tail of Influenza A Virus mRNA Is Synthesized by Reiterative Copying of a U Track in the Virion RNA Template. *Journal of Virology*, *73*(4), 3473. https://doi.org/10.1128/JVI.73.4.3473-3476.1999

Powell, M. L. (2010). Translational termination-reinitiation in RNA viruses. *Biochemical Society Transactions*, *38*(6), 1558–1564. https://doi.org/10.1042/BST0381558

Powell, M. L., Brown, T. D. K., & Brierley, I. (2008). Translational termination-re-initiation in viral systems. *Biochemical Society Transactions*, *36*(4), 717–722. https://doi.org/10.1042/BST0360717

Prasad, S., Starck, S. R., & Shastri, N. (2016). Presentation of cryptic peptides by MHC I is enhanced by inflammatory stimuli. *Journal of Immunology (Baltimore, Md. : 1950)*, *197*(8), 2981. https://doi.org/10.4049/JIMMUNOL.1502045

Racaniello, V. (2014, June 12). Origin of segmented RNA virus genomes. *Virology Blog*. http://www.virology.ws/2014/06/12/origin-of-segmented-rna-virus-genomes/

Ramey, A. M., Reeves, A. B., Drexler, J. Z., Ackerman, J. T., De, S., Cruz, L., Lang, A. S., Leyson, C., Link, P., Prosser, D. J., Robertson, G. J., Wight, J., Youk, S., Spackman, E., Pantin-Jackwood, M., Poulson, R. L., & Stallknecht, D. E. (2020). Influenza A viruses remain infectious for more than seven months in northern wetlands of North America. *Proceedings of the Royal Society B*, *287*. https://doi.org/10.1098/rspb.2020.1680

Ramos, S., Maclachlan, M., & Melton, A. (2017). *Impacts of the 2014-2015 Highly Pathogenic Avian Influenza Outbreak on the U.S. Poultry Sector* (pp. 282–284). www.ers.usda.gov

Ren, X., Yu, Y., Li, H., Huang, J., Zhou, A., Liu, S., Hu, P., Li, B., Qi, W., & Liao, M. (2019). Avian influenza A virus polymerase recruits cellular RNA helicase eIF4A3 to promote viral mRNA splicing and spliced mRNA nuclear export. *Frontiers in Microbiology*, *10*(JULY). https://doi.org/10.3389/fmicb.2019.01625

Richt, J. A., & García-Sastre, A. (2009). Attenuated Influenza Virus Vaccines with Modified NS1 Proteins. *Current Topics in Microbiology and Immunology*, *333*(1), 177–195. https://doi.org/10.1007/978-3-540-92165-3_9

Rimondi, A., Gonzalez-Reiche, A. S., Olivera, V. S., Decarre, J., Castresana, G. J., Romano, M., Nelson, M. I., van Bakel, H., Pereda, A. J., Ferreri, L., Geiger, G., & Perez, D. R. (2018). Evidence of a fixed internal gene constellation in influenza A viruses isolated from wild birds in Argentina (2006–2016). *Emerging Microbes and Infections*, *7*(1), 1–13. https://doi.org/10.1038/s41426-018-0190-2

Rossman, J. S., & Lamb, R. A. (2011). Influenza virus assembly and budding. *Virology*, *411*(2), 229–236. https://doi.org/10.1016/j.virol.2010.12.003

Rossman, J. S., Leser, G. P., & Lamb, R. A. (2012). Filamentous influenza virus enters cells via macropinocytosis. *Journal of Virology*, *86*(20), 10950–10960. https://doi.org/10.1128/JVI.05992-11

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J., Hurt, A. C., de Jong, J. C., Kelso, A., Klimov, A. I., Kageyama, T., Komadina, N., Lapedes, A. S., Lin, Y. P., Mosterin, A., Obuchi, M., … Smith, D. J. (2008). The Global Circulation of Seasonal Influenza A (H3N2) Viruses. *Science*, *320*(5874), 340–346. https://doi.org/10.1126/science.1154137

Rynda-Apple, A., Robinson, K. M., & Alcorn, J. F. (2015). Influenza and Bacterial Superinfection: Illuminating the Immunologic Mechanisms of Disease. *Infection and Immunity*, *83*(10). https://doi.org/10.1128/IAI.00298-15

Sabath, N., Morris, J., & D, G. (2011). Is there a twelfth protein-coding gene in the genome of influenza A? A selection-based approach to the detection of overlapping genes in closely related sequences. *Journal of Molecular Evolution*, *73*(5–6), 305–315. https://doi.org/10.1007/S00239-011-9477-9

Sarkar, M., Chanda, S., Chakrabarti, S., Mazumdar, J., Ganguly, A., Chadha, M. S., Mishra, A. C., & Chawla-Sarkar, M. (2012). Surveillance in eastern India (2007-2009) revealed reassortment event involving ns and PB1-F2 gene segments among co-circulating influenza a subtypes. *Virology Journal*, *9*. https://doi.org/10.1186/1743-422X-9-3

Schmidt, N. W., Mishra, A., Wang, J., Degrado, W. F., & Wong, G. C. L. (2013). Influenza virus A M2 protein generates negative gaussian

membrane curvature necessary for budding and scission. *Journal of the American Chemical Society*, *135*(37), 13710–13719. https://doi.org/10.1021/ja400146z

Schmolke, M., Manicassamy, B., Pena, L., Sutton, T., Hai, R., Varga, Z. T., Hale, B. G., Steel, J., Pérez, D. R., & García-Sastre, A. (2011). Differential contribution of pb1-f2 to the virulence of highly pathogenic h5n1 influenza a virus in mammalian and avian species. *PLoS Pathogens*, *7*(8). https://doi.org/10.1371/journal.ppat.1002186

Schwab, S. R., Shugart, J. A., Horng, T., Malarkannan, S., & Shastri, N. (2004). Unanticipated Antigens: Translation Initiation at CUG with Leucine. *PLoS Biology*, *2*(11). https://doi.org/10.1371/JOURNAL.PBIO.0020366

Sealy, J. E., Peacock, T. P., Sadeyen, J.-R., Chang, P., Everest, H. J., Bhat, S., & Iqbal, M. (2020). Adsorptive mutation and N-linked glycosylation modulate influenza virus antigenicity and fitness. *Emerging Microbes & Infections*, *9*(1), 2622. https://doi.org/10.1080/22221751.2020.1850180

Sediri, H., Schwalm, F., Gabriel, G., & Klenk, H. D. (2015). Adaptive mutation PB2 D701N promotes nuclear import of influenza vRNPs in mammalian cells. *European Journal of Cell Biology*, *94*(7–9), 368–374. https://doi.org/10.1016/j.ejcb.2015.05.012

Selman, M., Dankar, S. K., Forbes, N. E., Jia, J. J., & Brown, E. G. (2012). Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerging Microbes & Infections*, *1*(11). https://doi.org/10.1038/EMI.2012.38

Seth, R. B., Sun, L., Ea, C. K., & Chen, Z. J. (2005). Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-κB and IRF3. *Cell*, *122*(5), 669–682. https://doi.org/10.1016/j.cell.2005.08.012

Shao, Q., Xu, W., Guo, Q., Yan, L., Rui, L., Liu, J., Zhao, Y., & Li, Z. (2015). RIG-I from waterfowl and mammals differ in their abilities to induce antiviral responses against influenza A viruses. *Journal of General Virology*, *96*, 277–287. https://doi.org/10.1099/vir.0.069914-0

Shaw, M. L., & Stertz, S. (2018). Role of host genes in influenza virus replication. In *Current Topics in Microbiology and Immunology* (Vol. 419, pp. 151–189). Springer Verlag. https://doi.org/10.1007/82_2017_30

Shaw, M. L., Stone, K. L., Colangelo, C. M., Gulcicek, E. E., & Palese, P. (2008). Cellular proteins in influenza virus particles. *PLoS Pathogens*, *4*(6). https://doi.org/10.1371/journal.ppat.1000085

Shih, S. R., Nemeroff, M. E., & Krug, R. M. (1995). The choice of alternative 5’ splice sites in influenza virus Ml mRNA is regulated by the viral polymerase complex. *Biochemistry*, *92*, 6324–6328. https://doi.org/10.1073/pnas.92.14.6324

Shinya, K., Watanabe, S., Ito, T., Kasai, N., & Kawaoka, Y. (2007). Adaptation of an H7N7 equine influenza A virus in mice. *Journal of General Virology*, *88*, 547–553. https://doi.org/10.1099/vir.0.82411-0

Sigal, D., Reid, J. N. S., & Wahl, L. M. (2018). Effects of Transmission Bottlenecks on the Diversity of Influenza A Virus. *Genetics*, *210*(3), 1075–1088. https://doi.org/10.1534/genetics.118.301510

Sikora, D., Rocheleau, L., Brown, E. G., & Pelchat, M. (2014). Deep sequencing reveals the eight facets of the influenza

A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Scientific Reports*, *4*. https://doi.org/10.1038/SREP06181

Skehel, J. J., & Wiley, D. C. (2000). Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. *Annual Review of Biochemistry*, *69*, 531–569. https://doi.org/10.1146/annurev.biochem.69.1.531

Smith, A. M., & McCullers, J. A. (2013). Molecular signatures of virulence in the PB1-F2 proteins of H5N1 influenza viruses. *Virus Research*, *178*(1), 146–150. https://doi.org/10.1016/j.virusres.2013.02.012.Molecular

Solbak, S. M. Ø., Sharma, A., Bruns, K., Röder, R., Mitzner, D., Hahn, F., Niebert, R., Vedeler, A., Henklein, P., Henklein, P., Schubert, U., Wray, V., & Fossen, T. (2013). Influenza A virus protein PB1-F2 from different strains shows distinct structural signatures. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1834*(2), 568–582. https://doi.org/10.1016/j.bbapap.2012.11.009

Sonneveld, S., Verhagen, B. M. P., & Tanenbaum, M. E. (2020). Heterogeneity in mRNA Translation. *Trends in Cell Biology*, *30*(8), 606–618. https://doi.org/10.1016/J.TCB.2020.04.008

Spackman, E., Gelb, J., Preskenis, L. A., Ladman, B. S., Pope, C. R., Pantin-Jackwood, M. J., & McKinley, E. T. (2010). The pathogenesis of low pathogenicity H7 avian influenza viruses in chickens, ducks and turkeys. *Virology Journal*, *7*, 331. https://doi.org/10.1186/1743-422X-7-331

Spekreijse, D., Bouma, A., Stegeman, J. A., Koch, G., & de Jong, M. C. M. (2011). The effect of inoculation dose of a highly pathogenic avian influenza virus strain H5N1 on the infectiousness of chickens.

Veterinary Microbiology, *147*(1–2), 59–66. https://doi.org/10.1016/j.vetmic.2010.06.012

Stallknecht, D. E., Shane, A. C. S. M., Kearney, A. M. T., & Zwank, P. J. (1990). Persistence of Avian Influenza Viruses in Water. *Avian Diseases*, *34*(2), 406–411. https://doi.org/10.2307/1591428

Subbarao, E. K., London, W., & Murphy, B. R. (1993). A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *Journal of Virology*, *67*(4), 1761–1764. https://doi.org/10.1128/jvi.67.4.1761-1764.1993

Suttie, A., Deng, Y. M., Greenhill, A. R., Dussart, P., Horwood, P. F., & Karlsson, E. A. (2019). Inventory of molecular markers affecting biological characteristics of avian influenza A viruses. *Virus Genes*, *55*(6), 739–768. https://doi.org/10.1007/s11262-019-01700-z

Sutton, T. C. (2018). The pandemic threat of emerging H5 and H7 avian influenza viruses. *Viruses*, *10*(9), 1–21. https://doi.org/10.3390/v10090461

Szewczyk, B., Bieńkowska-Szewczyk, K., & Król, E. (2014). Introduction to molecular biology of influenza A viruses. *Acta Biochimica Polonica*, *61*(3), 397–401. https://doi.org/10.18388/abp.2014_1857

Taubenberger, J. K., & Kash, J. C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host and Microbe*, *7*(6), 440–451. https://doi.org/10.1016/j.chom.2010.05.009

Tauber, S., Ligertwood, Y., Quigg-Nicol, M., Dutia, B. M., & Elliott, R. M. (2012). Behaviour of influenza A viruses differentially expressing segment 2 gene products in vitro and in vivo. *Journal of General Virology*, *93*(4), 840–849. https://doi.org/10.1099/vir.0.039966-0

Taylor, J. M., Illmensee, R., Litwin, S., Herring, L., Broni, B., & Krug, R. M. (1977). Use of specific radioactive probes to study transcription and replication of the influenza virus genome. *Journal of Virology*, *21*(2), 530–540. https://doi.org/10.1128/jvi.21.2.530-540.1977

Tchatalbachev, S., Flick, R., & Hobom, G. (2001). The packaging signal of influenza viral RNA molecules. *RNA*, *7*(7), 979. https://doi.org/10.1017/S1355838201002424

Te Velthuis, A. J. W. W., & Fodor, E. (2016). Influenza virus RNA polymerase: Insights into the mechanisms of viral RNA synthesis. *Nature Reviews Microbiology*, *14*(8), 479–493. https://doi.org/10.1038/nrmicro.2016.87

Thattai, M., & Oudenaarden, A. van. (2001). Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(15), 8614–8619. https://doi.org/10.1073/PNAS.151588598

The Business Research Company. (2021, January). *Poultry Global Market Report 2021: COVID-19 Impact and Recovery to 2030*. https://www.researchandmarkets.com/reports/5240275/poultry-global-market-report-2021-covid-19?utm_source=GNOM&utm_medium=PressRelease&utm_code=xjkrtf&utm_campaign=1502792+-+Global+Poultry+Market+Report+2021-2030%3A+COVID-19+Impact+and+Recovery+-+Market+is

Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., Pflug, A., Hengrung, N., El Omari, K., Baudin, F., Hart, D. J., Beck, M., & Cusack, S. (2016). Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2

Domains. *Molecular Cell*, *61*(1), 125–137. https://doi.org/10.1016/j.molcel.2015.11.016

Thornton, J. M., Orengo, C. A., Todd, A. E., & Pearl, F. M. G. (1999). Protein folds, functions and evolution. *Journal of Molecular Biology*, *293*(2), 333–342. https://doi.org/10.1006/jmbi.1999.3054

Tobita, K., Sugiura, A., Enomote, C., & M, F. (1975). Plaque assay and primary isolation of influenza A viruses in an established line of canine kidney cells (MDCK) in the presence of trypsin. *Medical Microbiology and Immunology*, *162*(1), 9–14. https://doi.org/10.1007/BF02123572

Tsurudome, M., Gluck, R., Graf, R., Falchetto, R., Schaller, U., & Brunner, J. (1992). Lipid interactions of the hemagglutinin HA2 NH2-terminal segment during influenza virus-induced membrane fusion. *Journal of Biological Chemistry*, *267*(28), 20225–20232. https://doi.org/10.1016/s0021-9258(19)88690-8

Urbaniak, K., & Markowska-Daniel, I. (2014). In vivo reassortment of influenza viruses. *Acta Biochimica Polonica*, *61*(3), 427–431. https://doi.org/10.18388/abp.2014_1860

Uversky, V. N. (2019). Protein intrinsic disorder and structure-function continuum. In *Progress in Molecular Biology and Translational Science* (Vol. 166). Elsevier Inc. https://doi.org/10.1016/bs.pmbts.2019.05.003

Vandegrift, K. J., Sokolow, S. H., Daszak, P., & Kilpatrick, A. M. (2010). Ecology of avian influenza viruses in a changing world. *Annals of the New York Academy of Sciences*, *1195*, 113–128. https://doi.org/10.1111/j.1749-6632.2010.05451.x

Varga, Z. T., & Palese, P. (2011). The influenza A virus protein PB1-F2: Killing two birds with one stone? *Virulence*, *2*(6), 542–546. https://doi.org/10.4161/viru.2.6.17812

Varga, Z. T., Ramos, I., Hai, R., Schmolke, M., García-Sastre, A., Fernandez-Sesma, A., & Palese, P. (2011). The influenza virus protein PB1-F2 inhibits the induction of type i interferon at the level of the MAVS adaptor protein. *PLoS Pathogens*, *7*(6). https://doi.org/10.1371/journal.ppat.1002067

Varghese, J., Laver, W., & Colman, P. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature*, *303*(5912), 35–40. https://doi.org/10.1038/303035A0

Vasin, A. V., Temkina, O. A., Egorov, V. V., Klotchenko, S. A., Plotnikova, M. A., & Kiselev, O. I. (2014). Molecular mechanisms enhancing the proteome of influenza A viruses: An overview of recently discovered proteins. *Virus Research*, *185*, 53–63. https://doi.org/10.1016/j.virusres.2014.03.015

Verhagen, J. H., Höfle, U., van Amerongen, G., van de Bildt, M., Majoor, F., Fouchier, R. A. M., & Kuiken, T. (2015). Long-Term Effect of Serial Infections with H13 and H16 Low-Pathogenic Avian Influenza Viruses in Black-Headed Gulls. *Journal of Virology*, *89*(22), 11507–11522. https://doi.org/10.1128/jvi.01765-15

Verhelst, J., Hulpiau, P., & Saelens, X. (2013). Mx Proteins: Antiviral Gatekeepers That Restrain the Uninvited. *Microbiology and Molecular Biology Reviews : MMBR*, *77*(4), 551–566. https://doi.org/10.1128/MMBR.00024-13

Vidic, J., Richard, C. A., Péchoux, C., Da Costa, B., Bertho, N., Mazerat, S., Delmas, B., & Chevalier, C. (2016). Amyloid assemblies of

influenza a virus PB1-F2 protein damage membrane and induce cytotoxicity. *Journal of Biological Chemistry*, *291*(2), 739–751. https://doi.org/10.1074/jbc.M115.652917

Vreede, F. T., & Brownlee, G. G. (2007). Influenza Virion-Derived Viral Ribonucleoproteins Synthesize both mRNA and cRNA In Vitro. *Journal of Virology*, *81*(5), 2196–2204. https://doi.org/10.1128/jvi.02187-06

Wang, F., Liu, G., Lu, Y., Hlasny, M., Liu, Q., & Zhou, Y. (2020). Acquisition of Avian-Origin PB1 Facilitates Viral RNA Synthesis by the 2009 Pandemic H1N1 Virus Polymerase. *Viruses*, *12*(3). https://doi.org/10.3390/V12030266

Wang, R., Zhu, Y., Ren, C., Yang, S., Tian, S., Chen, H., Jin, M., & Zhou, H. (2020). Influenza A virus protein PB1-F2 impairs innate immunity by inducing mitophagy. *Autophagy*, 1–16. https://doi.org/10.1080/15548627.2020.1725375

Wanitchang, A., Kramyu, J., & Jongkaewwattana, A. (2010). Enhancement of reverse genetics-derived swine-origin H1N1 influenza virus seed vaccine growth by inclusion of indigenous polymerase PB1 protein. *Virus Research*, *147*(1), 145–148. https://doi.org/10.1016/j.virusres.2009.10.010

Wasik, B. R., De Wit, E., Munster, V., Lloyd-Smith, J. O., Martinez-Sobrido, L., & Parrish, C. R. (2019). Onward transmission of viruses: How do viruses emerge to cause epidemics after spillover? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1782). https://doi.org/10.1098/rstb.2019.0017

Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M., & Kawo. (1992). Evolution and Ecology of Influenza A Viruses.

Microbiological Reviews, *56*(1), 152–179. https://doi.org/10.1128/mr.56.1.152-179.1992

Webster, R. G., & Govorkova, E. A. (2014). Continuing challenges in influenza. *Annals of the New York Academy of Sciences*, *1323*(1), 115–139. https://doi.org/10.1111/nyas.12462

Weeks-Gorospe, J. N., Hurtig, H. R., Iverson, A. R., Schuneman, M. J., Webby, R. J., McCullers, J. A., & Huber, V. C. (2012). Naturally Occurring Swine Influenza A Virus PB1-F2 Phenotypes That Contribute to Superinfection with Gram-Positive Respiratory Pathogens. *Journal of Virology*, *86*(17), 9035–9043. https://doi.org/10.1128/jvi.00369-12

Wei, P., Li, W., Zi, H., Cunningham, M., Guo, Y., Xuan, Y., Musa, T. H., & Luo, P. (2015). Epidemiological and molecular characteristics of the PB1-F2 proteins in H7N9 influenza viruses, Jiangsu. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/804731

Wendel, I., Rubbenstroth, D., Doedt, J., Kochs, G., Wilhelm, J., Staeheli, P., Klenk, H.-D., & Matrosovich, M. (2015). The Avian-Origin PB1 Gene Segment Facilitated Replication and Transmissibility of the H3N2/1968 Pandemic Influenza Virus. *Journal of Virology*, *89*(8), 4170. https://doi.org/10.1128/JVI.03194-14

White, M. C., & Lowen, A. C. (2018). Implications of segment mismatch for influenza A virus evolution. *Journal of General Virology*, *99*(1), 3–16. https://doi.org/10.1099/jgv.0.000989

Whitel, J., Kartenbeck, J., & Heleniuse, A. (1982). Membrane fusion activity of influenza virus. *The EMBO Journal*, *1*(2), 217–222.

Wille, M., & Holmes, E. C. (2020). The ecology and evolution of influenza viruses. *Cold Spring Harbor Perspectives in Medicine*, *10*(7), 1–19. https://doi.org/10.1101/cshperspect.a038489

Wilson, I., Skehel, J., & Wiley, D. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature*, *289*(5796), 366–373. https://doi.org/10.1038/289366A0

Wise, H. M., Barbezange, C., Jagger, B. W., Dalton, R. M., Gog, J. R., Curran, M. D., Taubenberger, J. K., Anderson, E. C., & Digard, P. (2011). Overlapping signals for translational regulation and packaging of influenza A virus segment 2. *Nucleic Acids Research*, *39*(17), 7775–7790. https://doi.org/10.1093/nar/gkr487

Wise, H. M., Foeglein, A., Sun, J., Dalton, R. M., Patel, S., Howard, W., Anderson, E. C., Barclay, W. S., & Digard, P. (2009). A Complicated Message: Identification of a Novel PB1-Related Protein Translated from Influenza A Virus Segment 2 mRNA. *Journal of Virology*, *83*(16), 8021–8031. https://doi.org/10.1128/jvi.00826-09

Wise, H. M., Gaunt, E., Ping, J., Holzer, B., Jasim, S., Lycett, S. J., Murphy, L., Livesey, A., Brown, R., Smith, N., Morgan, S., Clark, B., Kudryavtseva, K., Beard, P. M., Nguyen-Van-Tam, J., Salguero, F. J., Tchilian, E., Dutia, B. M., Brown, E. G., & Digard, P. (2019). An alternative AUG codon in segment 5 of the 2009 pandemic influenza A virus is a swine-derived virulence motif. *BioRxiv*. https://doi.org/10.1101/738427

Wise, H. M., Hutchinson, E. C., Jagger, B. W., Stuart, A. D., Kang, Z. H., Robb, N., Schwartzman, L. M., Kash, J. C., Fodor, E., Firth, A. E., Gog, J. R., Taubenberger, J. K., Digard, P., & Pekosz, A. (2012). Identification of a Novel Splice Variant Form of the Influenza A Virus

M2 Ion Channel with an Antigenically Distinct Ectodomain. *PLoS Pathogens*, *8*(11). https://doi.org/10.1371/journal.ppat.1002998

World Health Organisation. (2021, October 10). *Global Influenza Surveillance and Response System (GISRS)*. https://www.who.int/initiatives/global-influenza-surveillance-and-response-system

Wu, W. W., Sun, Y.-H. B., & Panté, N. (2007). Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. *Virology Journal*, *4*(49). https://doi.org/10.1186/1743-422X-4-49

Xiao, Y., Evseev, D., Stevens, C. A., Moghrabi, A., Miranzo-Navarro, D., Fleming-Canepa, X., Tetrault, D. G., & Magor, K. E. (2020). Influenza PB1-F2 inhibits avian MAVS signaling. *Viruses*, *12*(4). https://doi.org/10.3390/v12040409

Xu, W., Shao, Q., Zang, Y., Guo, Q., Zhang, Y., & Li, Z. (2015). Pigeon RIG-I Function in Innate Immunity against H9N2 IAV and IBDV. *Viruses*, *7*, 4131–4151. https://doi.org/10.3390/v7072813

Yamada, H., Chounan, R., Higashi, Y., Kurihara, N., & Kido, H. (2004). Mitochondrial targeting sequence of the influenza A virus PB1-F2 protein and its function in mitochondria. *FEBS Letters*, *578*(3), 331–336. https://doi.org/10.1016/j.febslet.2004.11.017

Yamayoshi, S., Watanabe, M., Goto, H., & Kawaoka, Y. (2016). Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. *Journal of Virology*, *90*(1), 444–456. https://doi.org/10.1128/jvi.02175-15

Yang, J., Cui, H., Teng, Q., Ma, W., Li, X., Wang, B., Yan, D., Chen, H., Liu, Q., & Li, Z. (2019). Ducks induce rapid and robust antibody

responses than chickens at early time after intravenous infection with H9N2 avian influenza virus. *Virology Journal*, *16*(1). https://doi.org/10.1186/s12985-019-1150-8

Yeganeh, B., Ghavami, S., Rahim, M. N., Klonisch, T., Halayko, A. J., & Coombs, K. M. (2018). Autophagy activation is required for influenza A virus-induced apoptosis and replication. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1865*(2), 364–378. https://doi.org/10.1016/j.bbamcr.2017.10.014

Yesylevskyy, S. O., Demchenko, A. P., Kraszewski, S., & Ramseyer, C. (2013). Cholesterol Induces Uneven Curvature of Asymmetric Lipid Bilayers. *The Scientific World Journal*, *2013*, e965230. https://doi.org/10.1155/2013/965230

Yoon, S. W., Webby, R. J., & Webster, R. G. (2014). Evolution and ecology of influenza a viruses. *Current Topics in Microbiology and Immunology*, *385*, 359–375. https://doi.org/10.1007/82_2014_396

Yoshizumi, T., Ichinohe, T., Sasaki, O., Otera, H., Kawabata, S. I., Mihara, K., & Koshiba, T. (2014). Influenza a virus protein PB1-F2 translocates into mitochondria via Tom40 channels and impairs innate immunity. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms5713

Zamarin, D., García-Sastre, A., Xiao, X., Wang, R., & Palese, P. (2005). Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS Pathogens*, *1*(1), 0040–0054. https://doi.org/10.1371/journal.ppat.0010004

Zamarin, D., Ortigoza, M. B., & Palese, P. (2006). Influenza A Virus PB1-F2 Protein Contributes to Viral Pathogenesis in Mice. *Journal of Virology*, *80*(16), 7976–7983. https://doi.org/10.1128/jvi.00415-06

Zecchin, B., Minoungou, G., Fusaro, A., Moctar, S., Ouedraogo-Kaboré, A., Schivo, A., Salviato, A., Marciano, S., & Monne, I. (2017). Influenza A(H9N2) Virus, Burkina Faso. *Emerging Infectious Diseases*, *23*(12), 2118–2119. https://doi.org/10.3201/eid2312.171294

Zell, R., Krumbholz, A., Eitner, A., Krieg, R., Halbhuber, K. J., & Wutzler, P. (2007). Prevalence of PB1-F2 of influenza A viruses. *Journal of General Virology*, *88*(2), 536–546. https://doi.org/10.1099/vir.0.82378-0

Zell, R., Scholtissek, C., & Ludwig, S. (2012). Genetics, Evolution, and the Zoonotic Capacity of European Swine Influenza Viruses. *Current Topics in Microbiology and Immunology*, *370*, 29–55. https://doi.org/10.1007/82_2012_267

Zhang, J., Ruan, T., Sheng, T., Wang, J., Sun, J., Wang, J., Prinz, R. A., Peng, D., Liu, X., & Xu, X. (2019). Role of c-Jun terminal kinase (JNK) activation in influenza A virus-induced autophagy and replication. *Virology*, *526*, 1–12. https://doi.org/10.1016/J.VIROL.2018.09.020

Zheng, W., Li, J., Wang, S., Cao, S., Jiang, J., Chen, C., Ding, C., Qin, C., Ye, X., Gao, G. F., & Liu, W. (2015). Phosphorylation Controls the Nuclear-Cytoplasmic Shuttling of Influenza A Virus Nucleoprotein. *Journal of Virology*, *89*(11), 5822–5834. https://doi.org/10.1128/JVI.00015-15

Zhirnov, O. P., & Klenk, H.-D. (2007). Control of apoptosis in influenza virus-infected cells by up-regulation of Akt and p53 signaling. *Apoptosis: An International Journal on Programmed Cell Death*, *12*(8), 1419–1432. https://doi.org/10.1007/s10495-007-0071-y