1 Introduction

1.1 The Mammalian Immune System

Complex organisms have evolved an array of tools to suppress the replication of the viruses that infect them. The immune system of these multicellular organisms can be split into three main branches: the intrinsic, the innate, and the adaptive systems[7]. The first barrier to infection is physical; the skin provides an impermeable barrier that prevents viruses from entering soft tissues, however mucosal membranes such as the mouth, lungs, and eyes are all susceptible to infection. Cells in the trachea and bronchioles beat motile cilia on their surface in order to prevent micro-organisms from colonising the lower respiratory tract[8]. Furthermore, goblet cells in the airway epithelium produce mucin, the main protein of mucus, to trap micro-organisms that do reach the lower respiratory tract, preventing access to the underlying epithelium[8].

If cell infection by a virus occurs, the intrinsic and innate immune responses are initiated. They both function to detect pathogens, initiate cell signalling, and trigger infected cell death (Figure 1). White blood cells (WBCs) are also recruited to the site of infection to phagocytose invading virus and present viral antigens to T lymphocytes[9]. These adaptive immune cells directly kill virus-infected cells as well as provide T-cell help to activate B cells. B cells produce antibodies that are specific to a virus during the adaptive immune response (Figure 1). Although this response is slower to initiate it is vital for immunological memory of an infection, however because it is outside the scope of this thesis, it will not be discussed further.

1.2 The Innate Immune System

The innate arm of the immune system acts as an immediate, but non-specific barrier to infection, thereby developing no immunological memory. Phagocytic cells, including macrophages, neutrophils, and dendritic cells, mediate the innate response. These host cells recognise infection through several families of pattern-recognition receptors (PRRs) (Figure 2) that distinguish evolutionarily-conserved structures, known as pathogen-associated molecular patterns (PAMPs)[10]. These PRRs include families of membrane-bound C-type Lectin and Toll-like receptors (TLRs), as well as cytoplasmic NOD-like receptors and Retinoic Acid-Inducible Gene 1 (RIG-I)-like

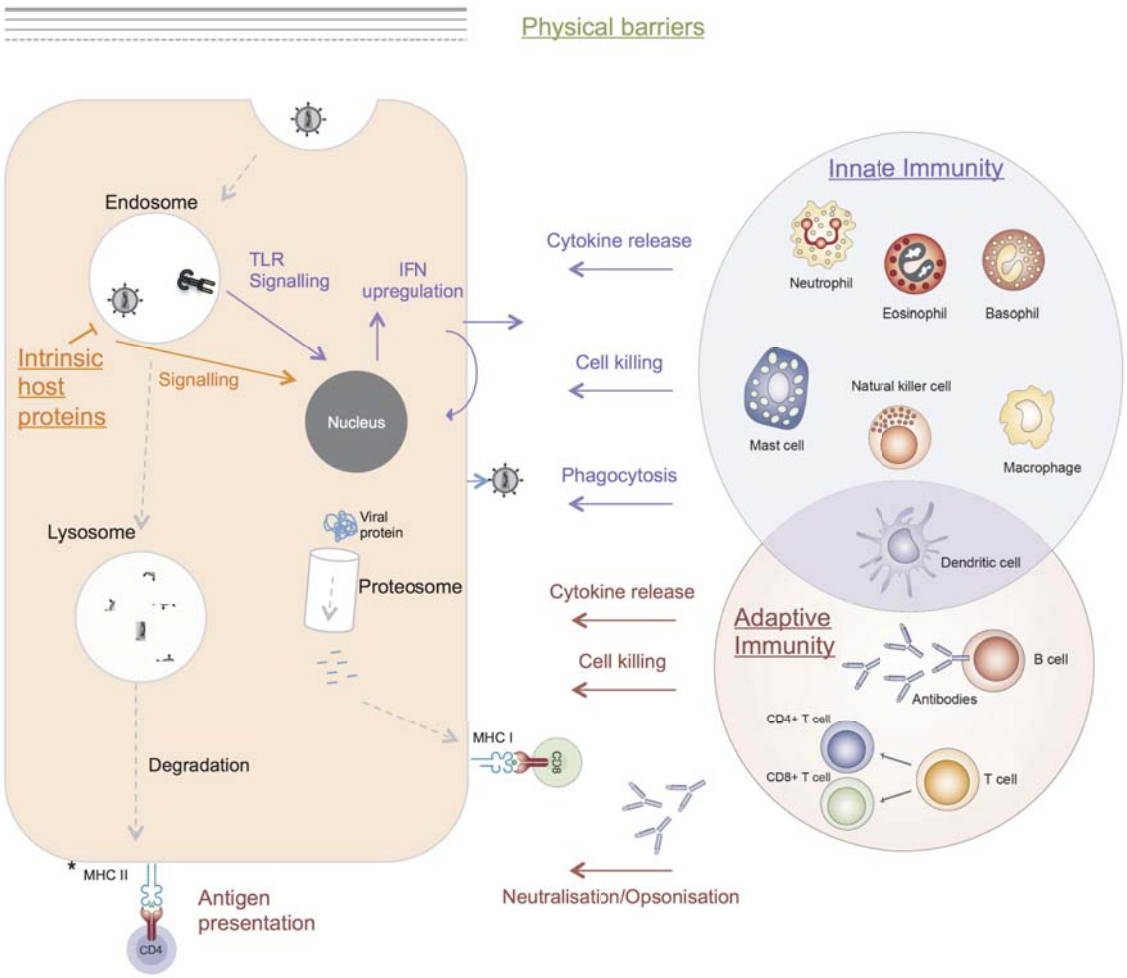


Figure 1: An overview of the mammalian immune systems

Physical barriers, including the skin, mucus, and motile cilia, prevent many micro-organisms from colonising the host. The innate immune cells and intrinsic host proteins mediate an early response against viruses that enter a cell. Virus detection leads to an upregulation of interferon, recruiting more innate cells that phagocytose virus particles or kill infected cells. Specialised phagocytic cells, such as dendritic cells, have MHC II molecules on their surface. The cells degrade the virus particles and present antigens via MHC II receptors to $CD4^+$ T helper cells. These $CD4^+$ cells then activate B cells that are able to produce antibodies specific to the antigen presented. Intracellular particles are processed and presented by MHC I molecules, present on the surface of all nucleated cells, to $CD8^+$ T cells that mediate direct cell killing. *Only on phagocytic cells.

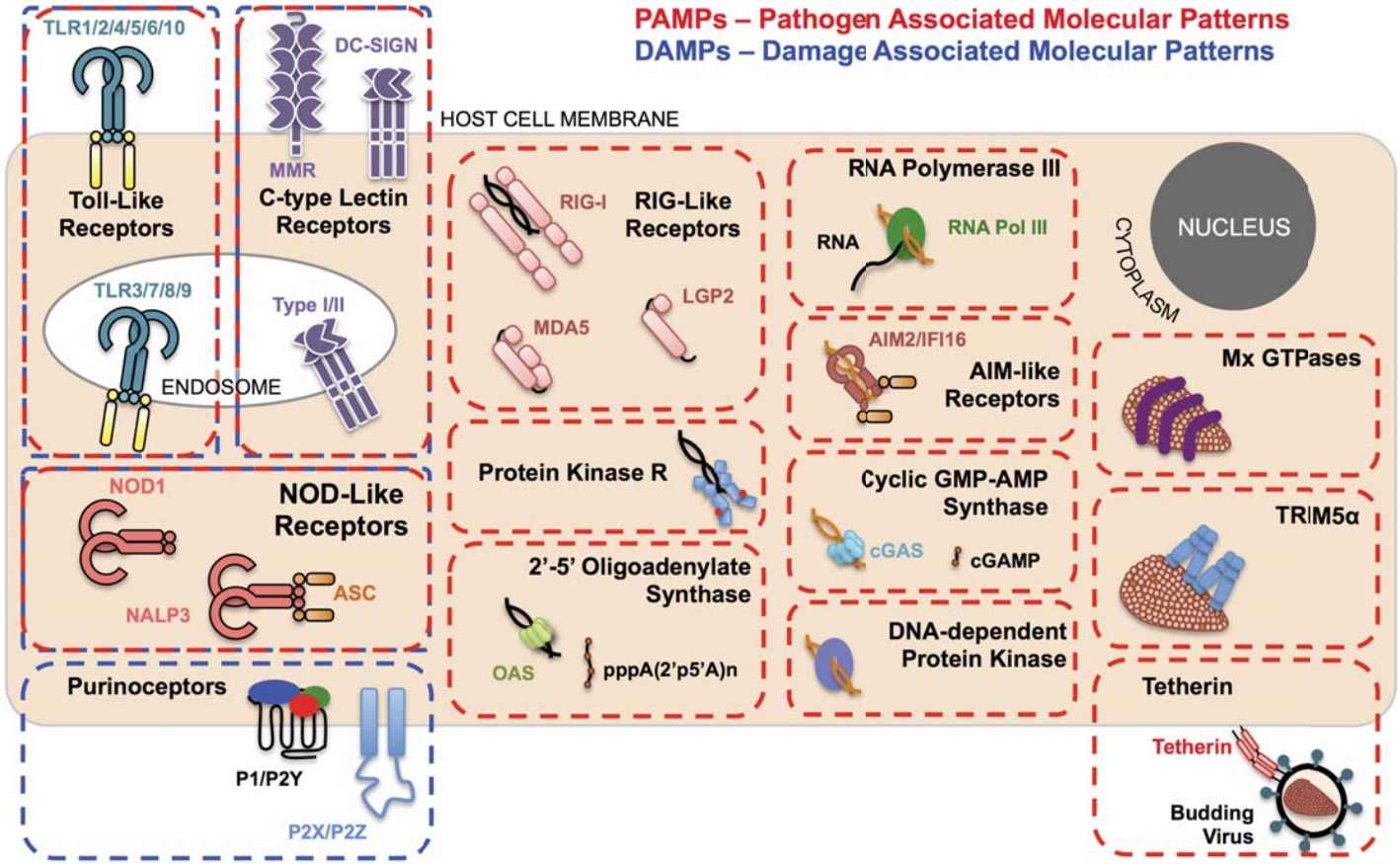


Figure 2: Cellular detection of pathogens by pattern-recognition receptors (PRRs)

The host cell has a vast range of pattern-recognition receptors, both free in the cytoplasm and membrane-bound, which detect pathogen associated molecular patterns (PAMPs) shown as red boxes. The Toll-like receptors, C-type Lectin receptors and NOD-like receptors also detect damage associated molecular patterns (DAMPs), *i.e.* a non-infectious inflammatory response, shown in blue boxes. The host cell-limiting membrane is labelled and the cytoplasm is shown in orange. The nucleus is represented as a black circle and the endosome as a white oval.

Adapted from Tam *et al.* (personal communication)

receptors[10]. Viruses that enter cells by clathrin-mediated endocytosis are detected by TLRs 3, 7 and 9 present in the endosomes, which all recognise viral nucleic acids[11]. These TLRs stimulate type I IFN production via the adapter molecules TRIF and MyD88 that directly interact with IRF3 and IRF7, respectively. Type 1 interferons (IFNs) act as autocrine and paracrine signals that upregulate expression of anti-viral molecules. Numerous studies in *IFNAR* $^{-/-}$ mice (those lacking the receptor for type I IFNs) have shown that the interferon receptor is crucial in mediating sensitivity and severity to pathogen infections[12].

1.2.1 Types of Interferon

IFNs are a large group of cytokines that can be sub-divided into classes: type I (α, β, ε, κ, and ω), II (γ), and III (λ). There are 13 distinct proteins within the IFNα family, two in the IFNβ family, one member in each of the IFNε, κ, ω, and γ families, and three members of the newly-identified IFNλ family. These vertebrate-specific molecules were first discovered in 1957 by Isaacs Lindenmann[13] who found that chicken cells infected with influenza A virus (IAV) secreted a ‘factor’, which prevented virus replication on a plate of previously non-infected cells.

IFN expression is upregulated by phosphorylation of transcription factor IFN response factor 3 (IRF3)[14]. This phosphorylation leads to dimerisation, translocation from the cytoplasm to the nucleus, association with CBP/p300 coactivators, and stimulation of DNA binding. Type I, type II and type III IFNs bind to their cognate receptors (IFNAR, IFNGR, and IFNLR (also called IL28R), respectively), however their signalling pathways partially overlap. Type I IFNs bind to the extracellular region of IFNAR-1 and IFNAR-2 heterodimers, whereas type III IFNs bind to IL10R2 and IFNLR1 heterodimers. However, both receptor dimers associate with Janus-activated kinase 1 (Jak1) and tyrosine kinase 2 (Tyk2). Ligand binding to these receptors results in activation of these protein kinases, which in turn phosphorylate signal transducer and activator of transcription (STAT) proteins[15-17] (Figure 3). For type II IFN signalling, IFNGR-1 and IFNGR-2 are constitutively associated with Jak1 and Jak2 respectively[16].

For type I and III signalling, phosphorylated STAT1 and 2 associate with IRF9 to form the IFN stimulated gene factor 3 (ISGF3) complex[18]. This complex physically binds to the IFN-stimulated response elements (ISRE) in the promoters of ISGs. In contrast,

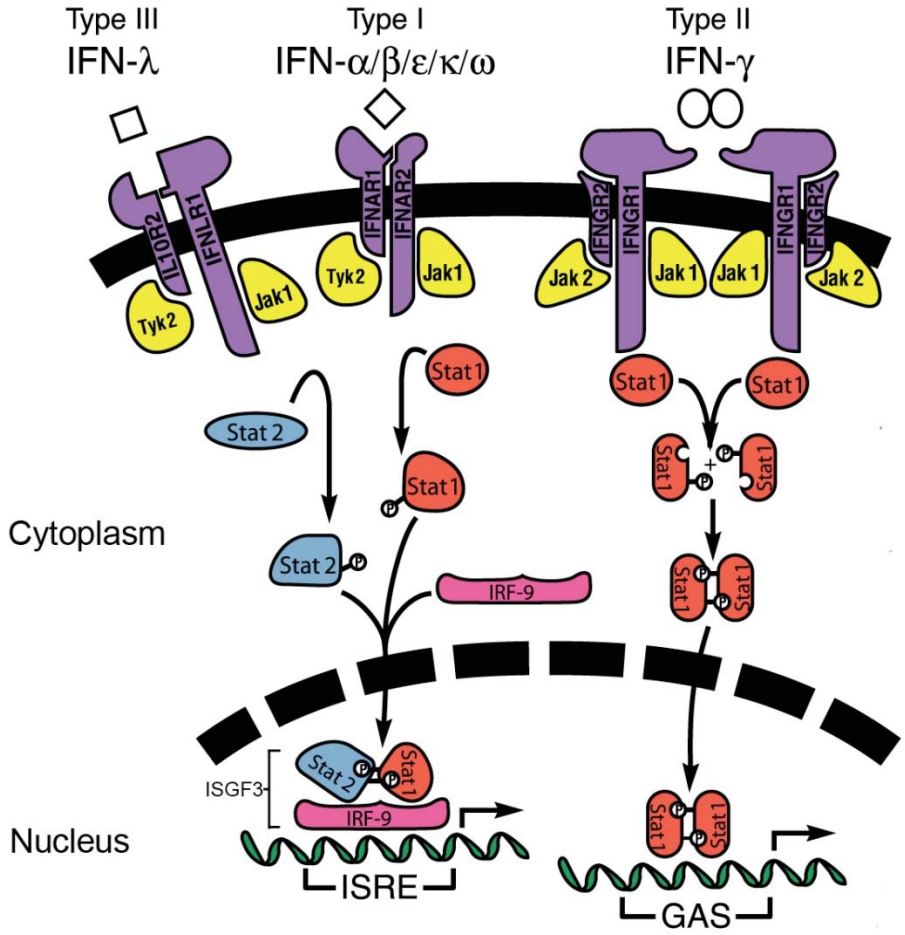


Figure 3: Interferon signalling by the JAK-STAT pathway

Each class of IFN has a distinct receptor molecule on the target cell surface. Ligand binding mediates the activation of overlapping pairs of kinase molecules, Jak1 and Tyk2 in the case of type I and III IFNs, and Jak1 and Jak2 in the case of type II IFN. The downstream events following phosphorylation of STAT1 and/or 2 are dependent on the ISRE or GAS enhancer elements. Proteins are translocated to the nucleus and bind to the upstream regions of ISGs upregulating gene expression.

Adapted from Samuel *et al.* [19]

STAT1 homodimers, which result from type II IFN signalling, bind to IFN-gamma activated sequence (GAS) elements (Figure 3). Since IFNγ does not induce the formation of an ISGF3 complex, IFNγ stimulation does not enhance transcription of genes controlled by an ISRE element[18].

Several types of IFNs have now been approved for therapeutic use. For example, IFNα 2b is used in combination with Ribavirin to treat chronic Hepatitis C Virus (HCV) infection[20], and is effective in 60-80 % of cases, as measured by sustained virological response rates *i.e.* the virus is not detected at the end of therapy. Ge *et al.* found that variation in response to IFNα treatment was associated with single nucleotide polymorphism (SNP) rs12979860, located 3 kb upstream of the gene encoding IFNλ-3[21]. Those with two C alleles at rs12979860 are 2-3 times more likely to have a sustained virological response. The alleles at the SNP are strongly associated with ethnicity; the beneficial C allele is rare in the African population, explaining why the response rate of African-Americans to this HCV treatment is substantially worse than those of European ancestry.

IFNs also have additional functions, primarily activating immune cells, such as natural killer (NK) cells and macrophages, and promoting antigen presentation to T-lymphocytes in order to eliminate an infection.

1.2.2 Cell Types Affected by Interferon Release

Innate immune cells are attracted to the site of infection by chemokines released at the site of infection through IFN signalling. They are responsible for clearing virus infection, either by phagocytosing free pathogens or by destroying pathogen-infected cells. All healthy, nucleated cells express MHC I major histocompatibility complex (MHC) molecules on their cell surface[22]. The cell digests cytosolic proteins, which are subsequently presented by MHC I receptors to $CD8^+$ T cells that mediate direct cell killing. Phagocytic cells also express MHC class II molecules on their cell surface. These present extracellular antigens that enter the cell via the endosomal pathways. The MHC II molecules are trafficked to the cell surface and present to $CD4^+$ cells, which in turn activate B cells[22].

1.2.2.1 Neutrophils

As with all haematopoietic cells, neutrophils are originally differentiated from

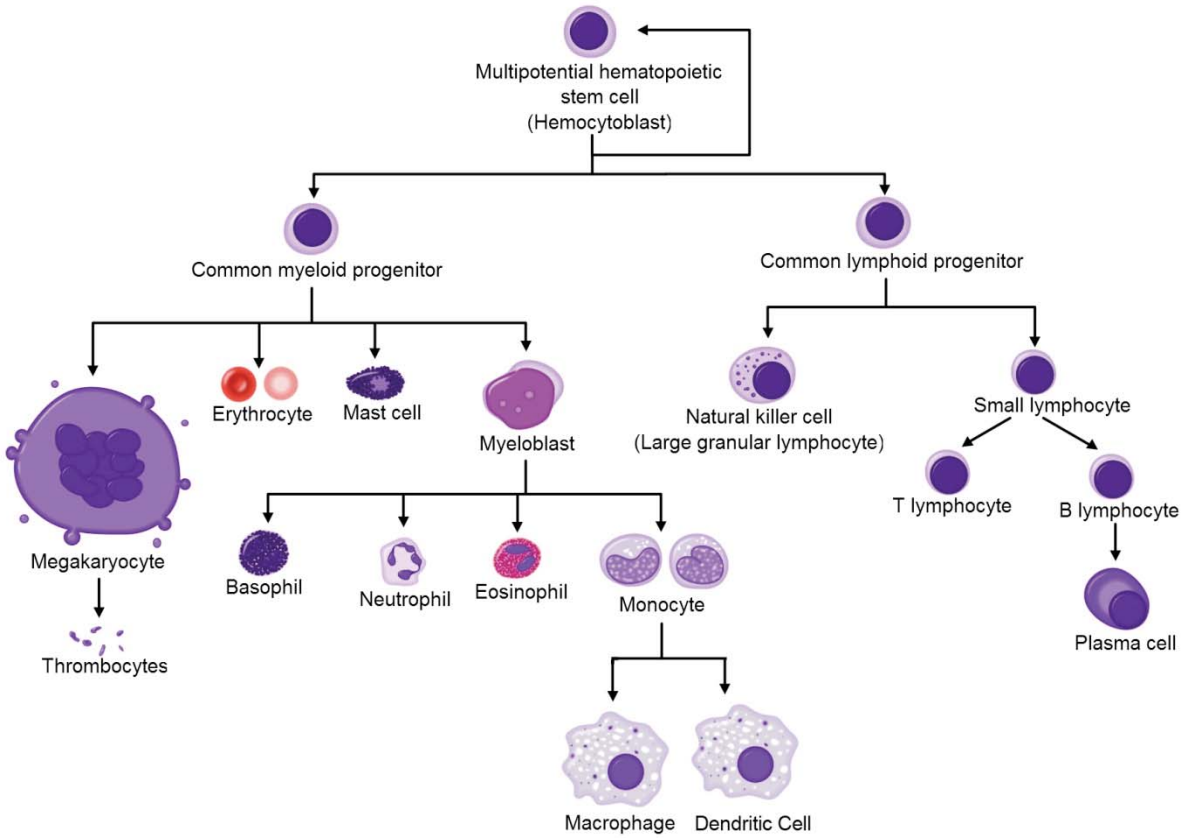


Figure 4: The differentiation of haematopoietic cells

All blood-borne cells are derived from haematopoietic stem cells that reside in the bone marrow. They split broadly into two lineages derived from common myeloid or lymphoid progenitors. These progenitor cells can differentiate into all functional haematopoietic cells that have specialised functions.

Adapted from Wikimedia Commons

multipotent haematopoietic stem cells (Figure 4). They are the most abundant WBC in humans, accounting for approximately 50-70 % of circulating WBCs, and can arrive at the site of infection within minutes, via the blood[9]. However some studies have shown large populations of neutrophils resident in the mammalian lung vasculature, even during non-infected periods[23]. Neutrophils extravasate through the endothelium of the capillaries surrounding the site of infection and recognise micro-organisms opsonised by natural IgM antibodies. These pentameric, low-avidity antibodies are produced by a subset of B cells, called B1 cells. IgM antibodies are produced spontaneously and can bind to diverse antigens or pathogens, even if the host has never previously been exposed to it[24].

The neutrophils bind the exposed Fc-region of the bound antibody, which triggers the cell to engulf the pathogen by phagocytosis, a form of endocytosis. The resultant phagosomes fuse with lysosomes inside the cells that contain superoxide and hypochlorite as well as a number of hydrolases[9], leading to pathogen destruction. Neutrophils are also able to release proteins that have antimicrobial properties from granules into the extracellular space (degranularisation) to further combat infection. Most recently, a third method of killing has been described for neutrophils – the release of DNA and serine proteases known as neutrophil extracellular traps (NETs) to trap microbes extracellularly for phagocytosis[25].

In the mouse model, depletion of neutrophils prior to IAV infection leads to increased mortality[26]. Although this supports the theory of a protective role for neutrophils, disregulation of neutrophil infiltration into the lungs during IAV infection can result in acute lung inflammation and damage due to vascular leakage and high release of NETs[27]. Therefore appropriate control of infiltration dynamics and neutrophil numbers is important for IAV infection control.

1.2.2.2 Macrophages

Macrophages are a large group of cells also derived from the myeloblast cell lineage (Figure 4) and can be broadly split into two categories: tissue resident and inflammatory[28]. Populations of long-lived macrophages reside in almost all tissues in the body, including the bone (osteoclasts), the central nervous system (microglia), and the liver (Kupffer cells)[28]. The inflammatory macrophages can be further divided into classically activated, wound-healing, and regulatory macrophages[28].

The terminal differentiation from monocytes into macrophages occurs when a monocyte is exposed to MCP-1 and extravasates from the bloodstream. Monocytes are attracted to a damaged site by sensing gradients of different chemotactic factors, including growth factors, proinflammatory cytokines, and chemokines[29], often produced from the neutrophils already *in situ*. These recruited inflammatory macrophages, as well as tissue-resident macrophages, detect opsonised microbes, but also express TLRs on their cell surface. Upon pathogen detection, they produce IFN and a number of cytokines, such as IL-1, IL-4, IL-8, and IL-15. These molecules encourage proliferation and differentiation of B and T cells, as well as activation of NK cells and the further migration of neutrophils.

Unlike neutrophils, after degradation of the pathogen, macrophages traffic pathogen peptide fragments to the cell surface, and present them by the class II MHC molecule in a process called antigen presentation. These foreign antigens are then detected by T helper cells or cytotoxic T cells. The former represent a bridge between the innate and the adaptive arms of the immune system, activating B cells to secrete antibodies, while the latter directly kill virus-infected cells by inducing apoptosis.

Although dogma suggests that an atypical adaptive immune response is responsible for chronic inflammation post-infection, some evidence suggests that macrophages can also play a part[30]. Studies in macrophage-depleted mice showed that lung macrophages are a key cellular source of IL-13, which can lead to chronic lung inflammation, and associated diseases such as asthma, long after a viral pathogen has been cleared[30].

1.2.2.3 Natural Killer Cells

Many viruses have evolved mechanisms to inhibit the expression of class I MHC proteins, present on the surface of all nucleated host cells, to avoid detection by cytotoxic T cells. For example, HIV-1 encodes Nef, a protein that prevents *MHC I* gene transcription and blocks transport of MHC I molecules to the cell surface[31]. However, NK cells detect host cells with low class I MHC molecule expression through the Killer-cell immunoglobulin-like receptors (KIRs). These receptors are predominantly inhibitory in their action, *i.e.* recognition of MHC I suppresses the cytotoxic activity of the NK cell (Figure 5). Lack of MHC I detection triggers the NK

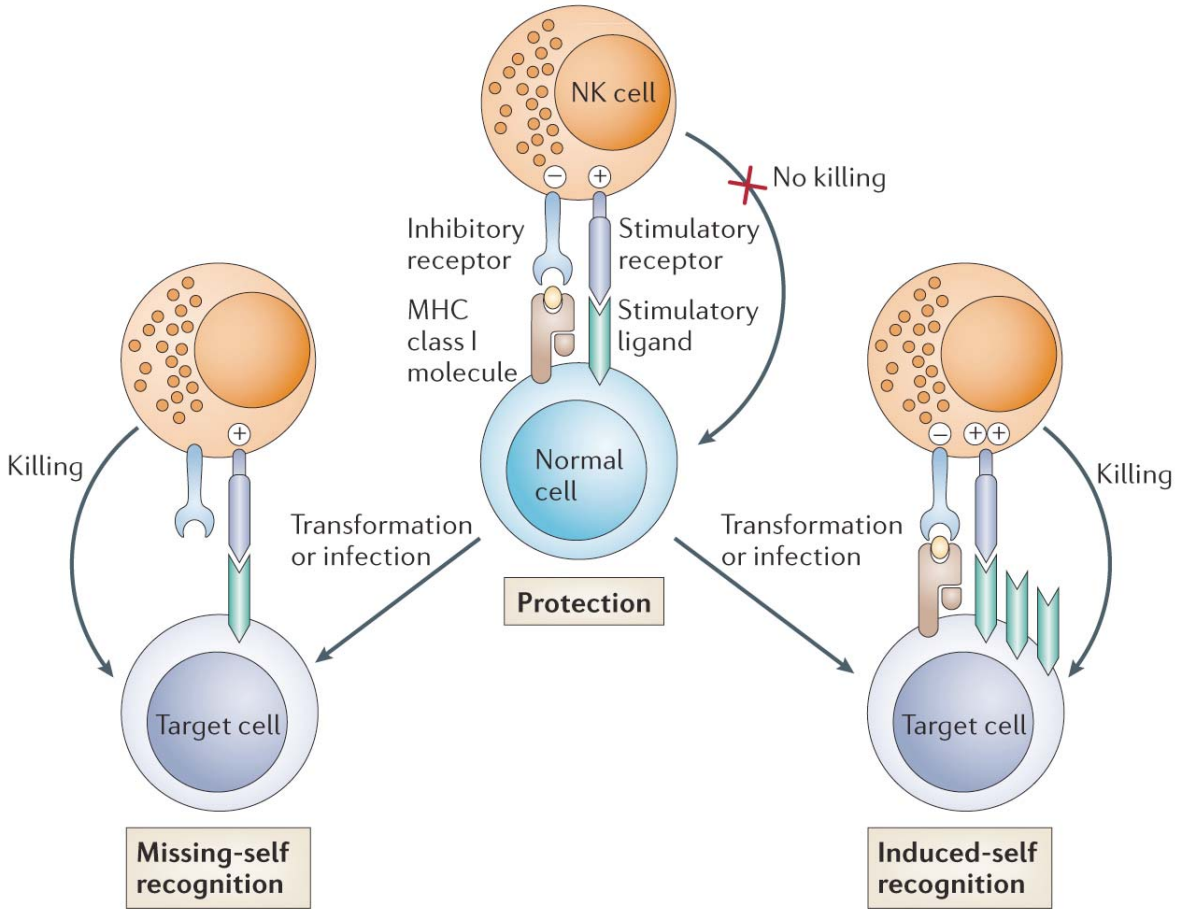


Figure 5: Natural killer cell control of activation

NK cells express inhibitory and stimulatory receptors on their cell surface, where by the ratio of inhibitory to stimulatory signals from the target cell dictates NK cell activation. Depletion of MHC I receptors on the target cell surface (inhibitory signal) results in the stimulatory signals being delivered to the NK cell and ultimately target cell death. Conversely, an over-proliferation of stimulatory signals on the cell surface will also result in cell killing (induced-self recognition).

cell to release perforin, creating pores in the target cell, and inducing apoptosis[9]. These cells are considered ‘lymphocyte-like’ since they are derived from the common lymphoid progenitors.

Depletion of NK cells in mice or hamsters, using anti-GM1 antibodies (a glycosphingolipid expressed on the plasma membrane of NK cells), results in an increased risk of morbidity and mortality during a pulmonary influenza infection[32]. Furthermore, people with severe pandemic H1N1 influenza suffered from drastic NK lymphopenia compared to those with mild symptoms[33]. However, over-expression of IL-15 by NK cells, to maintain and enhance their proliferation, can contribute to lung damage and pathogenesis during IAV infection. *IL-15* $^{-/-}$ mice (depleted of NK cells) were protected from lethal influenza infection[34]. The authors attributed this to the control of neutrophil invasion after infection. Again this highlights the role of a crucial balance in immune cell mediated infection control or enhanced tissue damage and pathology.

1.2.2.4 Dendritic Cells

A subset of dendritic cells (DCs) are also derived from monocytes, but their primary role is as a ‘professional’ antigen presenting cell. DCs are present in primary and secondary lymphoid tissues (classical DCs) and most non-lymphoid tissues (*e.g*. the blood [plasmacytoid DCs] and the skin [Langerhans cells]). DCs detect and internalise pathogens by phagocytosis, macropinocytosis, or receptor-mediated endocytosis, processing the proteins via the proteosome. Following this, the DC migrates to the nearest lymph node whilst undergoing a maturation process that decreases its phagocytic ability and enhances its antigen presenting ability. Once in the lymph node the DC presents the viral antigen to resident T cells by MHC II molecules.

As well as the innate cellular responses described above, there exists another line of defence that is active upon cellular infection and preceeds the adaptive immune response. This intrinsic cellular response is mediated by dozens of proteins, some of which have been termed viral restriction factors, that have a number of general properties in common. Viral restriction factors are mostly constitutively expressed, although they are often upregulated by IFN signalling[7]. These factors are usually conserved across many species, indicating crucial functions, and species-specific

polymorphisms in the factor can result in restriction of different virus species. Thus these factors are thought to evolve under strong positive selection[35]. Viral restriction factors can also be broadly grouped according to their general mechanism of action: protein degradation, mislocalisation/sequestration, or mimicry[36]. They confer anti-viral resistance to a broad range of viruses and block access to crucial regions of the cell required for infection, and therefore play a crucial role in early immune defence against viral infection.

1.3 Host-Virus Interactions

Infectious diseases are traditionally used as examples of illnesses caused purely by exogenous environmental agents. However, it is now clear that this dogmatic view is not true; the clinical phenotype that arises in an infected individual is a result of interactions between the host and virus, and the genetics of both. This idea can be used to explain both the phenotypic variation in the human population after infection by a given pathogen, and the diverse outcomes of an infection in different species[37]. For instance, Albright *et al.* conducted an epidemiological study on families in the state of Utah to investigate heritable susceptibility to severe IAV infection. Using genealogical databases, the authors found that close and distant relatives of people who died of IAV had a significantly higher risk of dying from an IAV infection than the spouses of such individuals[38]. This suggests that genetic susceptibilities may underlie severe responses to IAV infections.

Similarly, differences in species-specific responses are apparent after herpes simian B virus infection. This is an alphaherpesvirus endemic in the macaque population, causing monkeys to develop mild cold sores. Humans infected with the same virus, however, can develop severe encephalitis with a case fatality rate of 80 %[37]. Since the virus genome is very similar in both cases, this suggests that virus genetic differences are unlikely to be the cause of the disparate responses.

Evolutionary biologist Leigh Van Valen eloquently summarised the evolutionary dynamics between host and pathogen in his Red Queen Hypothesis: “For an evolutionary system, continuing development is needed just in order to maintain fitness relative to the systems it is co-evolving with.”[39] Consequently, immune function genes tend to evolve quicker than other genes in the genome[40].

1.3.1 A History of Viral Restriction Factors

In recent years, studies of intrinsic defence mechanisms have identified a number of cellular proteins that interfere with the replication of human and animal viruses. Many of these restriction factors have been most intensively studied for human immunodeficiency virus (HIV-1). TRIM5α[41], APOBEC3G[42], 2',3'-cyclic-nucleotide 3'-phosphodiesterase[43], and tetherin[44] were found to affect uncoating, reverse transcription, HIV-1 assembly, and HIV-1 release respectively. A new addition to this antiviral repertoire is myxoma resistance protein B (MxB/Mx2)[45], which inhibits HIV-1 at a late post-entry step.

However, factors that restrict other viruses have also been identified, including RNA-activated protein kinase (PKR), which restricts HCV and other viruses[46]; Mx1, which restricts IAV and measles virus; and 2'-5' oligoadenylate synthetase (OAS)/Ribonuclease L (RNAse L), which restricts many RNA viruses, including HCV[47]. These viral restriction factors are components of the broad antiviral response that are upregulated by IFNs, collectively known as IFN stimulated genes (ISG)[35]. Although recognised to act at different stages in viral replication cycles, most of the well-characterised restriction factors affect steps following virus entry. Many of these protein families have been studied intensively in both humans and primates, and some have been shown to have direct physical interactions with viral proteins as well as triggering signalling cascades. However, no single restriction factor protects a cell against infection by all viruses. Therefore, layering the activation of these molecules at different time-points during infection, and targeting them to different parts of the virus life-cycle is crucial for host defence.

1.3.1.1 MxA / Mx1

Mx proteins are IFN-induced dynamin-like GTPases found in many vertebrate species, providing significant resistance to a range of viruses in a host species-specific manner. First discovered in A2G mice, Mx1 is a potent restriction factor of orthomyxoviruses, in particular IAV[48]. Its human orthologue, MxA, has been shown to restrict a wider range of viruses, including those in the *Orthomyxoviridae, Paramyxoviridae, Rhabdoviridae, Bunyaviridae, Hepadnaviridae* and *Asfaviridae* families[49]. In physiological salt concentrations, MxA self-assembles into large oligomers in the cytoplasm in a similar fashion to dynamin and other GTPases. It is

most likely that aggregation of Mx proteins prevents their degradation and could explain why MxA is comparatively stable, with a half-life of over 24 hours[50].

Despite being orthologous, the mouse Mx1 protein is located in the nucleus, and inhibits primary transcription of the IAV genome[51], whilst the human MxA protein localises to the cytoplasm[52], recognising the nucleocapsids of invading viruses. Both lead to an early block of the viral replication cycle. Mitchell *et al.* carried out positive selection analysis (the ratio of non-synonymous mutations [d_N] to synonymous mutations [d_S]) on the *MxA* gene of 24 primate species to infer the functional regions of MxA[53]. The average d_N/d_S ratio across the whole protein was close to zero, indicating that purifying selection had occurred to maintain the architecture of the GTPase (Figure 6). However, an exposed region projecting from the stalk domain (the L4 loop) was a hotspot for codons under positive selection (average d_N/d_S=5.08). When comparing this region in human MxA and African green monkey MxA (active and inactive against Thogoto virus, respectively), there were only four amino acid differences. The authors were able to conclude that residue 561 determines the antiviral specificity of MxA (Figure 6ii) against orthomyxoviruses Thogoto virus and IAV. Haller *et al.* discovered an additional protein, MxB/Mx2, with 63 % sequence identity to MxA and of a comparable size, but found that it had no anti-viral activity[50]. However, in 2013 Goujon *et al.* showed that this cytoplasmic protein acts as an antiviral inhibitor of HIV-1 (a virus not restricted by the related MxA protein)[45]. This novel restriction factor was identified by transcriptional profiling of RNA extracted from 15 cell cultures with or without IFNα-stimulation. Further investigation suggested that MxB prevents HIV-1 replication by inhibiting capsid-dependent nuclear import of viral complexes[54]. Gain-of-function experiments in cells over-expressing human MxB showed that HIV-1 could escape restriction by mutating alanine 88 in the viral capsid protein. This mutation prevented the interaction of the viral capsid with the host peptidylprolyl isomerase cyclophilin A (CypA) protein[55], supporting the idea that MxB functions by a CypA-dependent mechanism.

1.3.1.2 TRIM5α

TRIM5α was discovered as a cytoplasmic retrovirus restriction factor in 2004, during a screen for HIV-1 resistance factors in rhesus monkeys[41]. HIV-1 causes acquired immunodeficiency syndrome (AIDS) in humans. Although HIV-1 can effectively enter

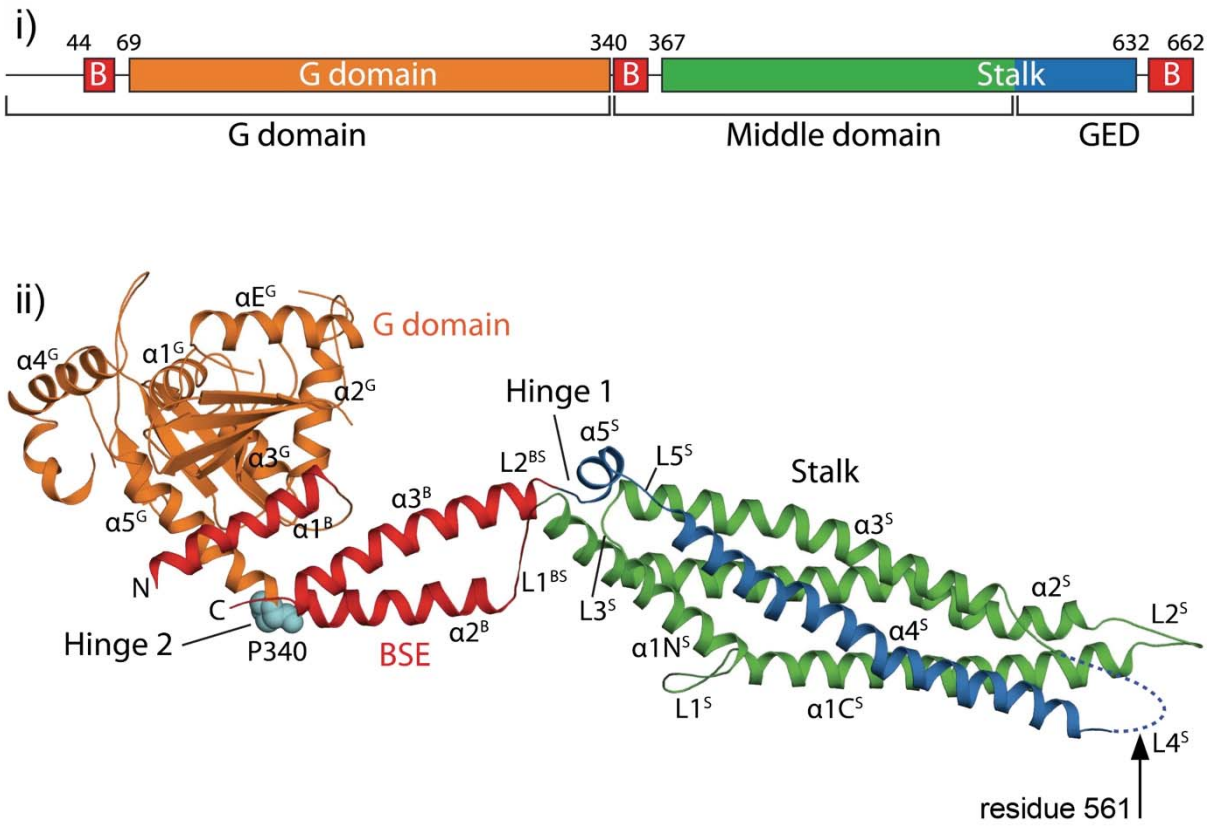


Figure 6: Secondary and tertiary structure of the MxA monomer

i) Structure-based domain representation of human MxA showing B (the bundle-signalling element [BSE]), the GTPase (G) domain, the middle domain, and the GTPase effector domain (GED). A ribbon-type representation of an MxA monomer is shown in ii), colour-coded as in i). The invariant Pro340-linking the G domain and BSE (hinge 2) is shown as van der Waals spheres in cyan. The L4 loop of the stalk domain (under positive selection) is shown as a blue dotted line and residue 561 is highlighted with an arrow.

Adapted from Gao *et al.* (2011)[56]

the cells of genetically-similar primates, such as rhesus monkeys, it encounters a cell intrinsic block to replication. TRIM5α is part of a large family containing 70 genes, all of which contain several conserved domains: a really interesting new gene (RING) domain (which confers E3 ubiquitin ligase activity), a B-box 2 domain, and a coiled-coil domain[57]. TRIM5α also features a C-terminal SPla and the RYanodine Receptor (SPRY) domain that recognises and binds to motifs within the capsid proteins of invading viruses. Although this finding is supported by several independent groups, the downstream mechanism of TRIM5α activity remains ambiguous. The accelerated-uncoating model suggests that TRIM5α promotes dissociation of the capsid from the viral ribonucleoprotein complex[58]. The proteasome dependent model suggests that after binding the capsid, TRIM5α induces the proteasome-dependent disassembly of the virus, thereby preventing successful reverse transcription[59,60]. The third model, the proteasome-independent model, suggests that additional cellular proteases are utilised by TRIM5α. Supporters of this model showed that proteasome inhibitors do not fully protect cytosolic capsid degradation studies[61].

TRIM5α is also able to self-assemble into dimers and higher-order multimerisation, and, like MxA, abrogation of specific amino acids controlling this process (like Cys-96) can severely reduce the antiviral capabilities of the protein[58,62,63]. In addition to having a physical interaction with HIV-1, TRIM5α also activates TAK1 kinase as part of a cell signalling cascade. The downstream effect of this is stimulation of AP-1 and NF-κB (proinflammatory) signalling[64], leading to the upregulation of other antiviral genes.

Several groups have investigated whether or not SNPs in human or rhesus monkey *TRIM5α* influence susceptibility to HIV-1 or disease progression. Although no associations have been found between common SNPs in human *TRIM5α* and progression to AIDS[65], several studies have shown that a SNP that causes an H43Y change, highly prevalent in Central Americans, results in a protein with impaired antiviral activity[65,66]. Therefore these individuals may be more likely to have a faster progression to AIDS. A number of other common SNPs in *TRIM5α* have been tested experimentally, but were shown to have a neutral effect on protein function.

Alternative splicing of *TRIM5α* was identified in owl monkey kidney cells, a species of new world monkey known to restrict HIV-1 post-entry[67]. Expression of a fusion

protein consisting of full length TRIM5α and C-terminal cyclophilin A (CypA) connected by 11 amino acids encoded by the CypA 5' UTR (TRIMCyp) was shown by Northern blot. Mammalian genomes contain many CypA transposons that have been randomly inserted into the genome as pseudogenes, but in this case it was inserted into an active splice region[67].

Studies in rhesus monkeys have revealed *TRIM5α* polymorphisms that enhance the antiviral activity against SIVmac239 – a simian immunodeficiency virus strain used to induce an AIDS-like disease in monkeys. Full length *TRIM5α* was sequenced from 79 rhesus monkeys, revealing 14 SNPs and a two-amino acid deletion[68]. The majority of non-synonymous SNPs found in rhesus monkey *TRIM5α* occur primarily in the SPRY and coiled-coil domains, which are known to dictate antiviral activity[68]. Furthermore, Sawyer *et al.* identified that *TRIM5α* was under positive selection, specifically at a 13-amino acid patch in the SPRY domain[69]. This evidence suggests that TRIM5α is a crucial, evolutionarily-conserved member of the innate intrinsic immune response.

1.3.1.3 Tetherin

Tetherin was first investigated in 1995 as a factor potentially involved in pre-B-cell growth[70] and in 2003 it was identified during a large luciferase-activation screen to identify proteins that activated NF-κB[71]. However, its antiviral effect on HIV-1 was not established until five years later[44].

Tetherin (also known at BST-2 and CD317) is an IFN-inducible gene, whose encoded protein is capable of inhibiting the release of virus particles from the host plasma cell membrane. As the name suggests, it does this by 'tethering' the virus particle to the cell membrane[44]. Retained virions may be internalised by endocytosis and subsequently degraded, or remain on the cell surface. Tetherin is a type II transmembrane protein (a single-pass protein targeted to the endoplasmic reticulum (ER) lumen with its C-terminal domain [CTD]) (Figure 7). Its expression is strongly induced by type I IFN, but it is also upregulated by upstream TLR3 activation and IRF3 expression[72].

Tetherin exists as a dimer on the surface of cells. Mutation of three conserved cystine residues (C53, C63 and C91) prevents tetherin dimerisation thereby reducing its antiviral function[73]. Therefore it is likely that the intermolecular disulphide bonds

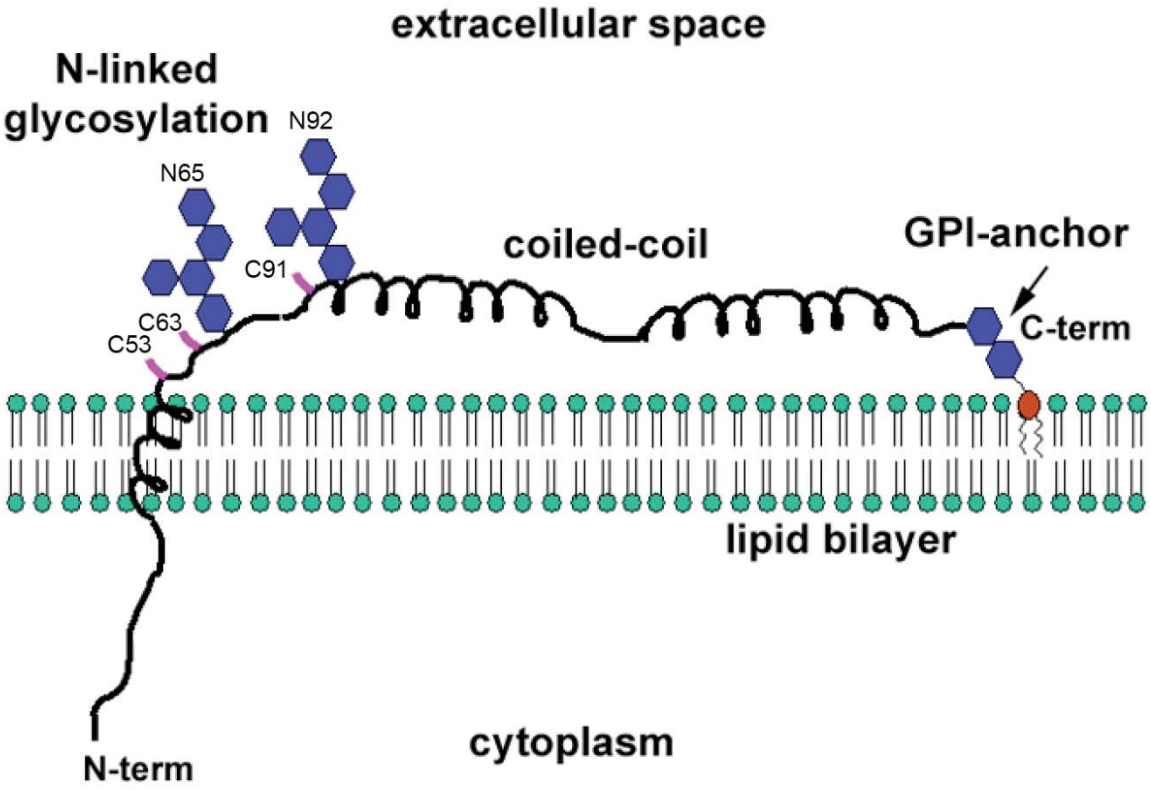


Figure 7: Tetherin is a type II transmembrane protein

The N-terminus of tetherin lies in the cytoplasm. The protein crosses the cell membrane once, with its coiled-coil domain residing in the extracellular space, whilst the C-terminus (modified with a glycosyl-phosphatidylinositol [GPI] membrane anchor) reattaches the protein to the outer leaflet of the cell membrane. Conserved cysteines are highlighted with a pink line, and glycogen molecules are shown as blue hexagons.

Adapted from Evans *et al.* (2010)[74]

between the cysteine residues stabilise the ectodomain *in vivo*; the temperature at which denaturation occurs drops from 65 °C to 35 °C after breaking the disulphide bridges[75]. Two sites are also commonly glycosylated (N65 and N92), but mutation of these sites does not inhibit cell surface expression of tetherin, nor eliminate its inhibitory effect on HIV-1 particle release[73]. Furthermore, several sites in the transmembrane domain were found to be under positive selection – a threonine to isoleucine change at position 45 in the human tetherin protein conferred resistance to the HIV antagonistic Vpu protein[76,77]. Although most research on tetherin has been conducted on retroviruses, Winkler *et al.* provide some evidence suggests that influenza induces IFN-dependent tetherin expression in infected cells[78] and several groups show that tetherin prevents the formation of influenza-like particles[79,80]. However, additional studies using infectious influenza could not show restriction by tetherin. This suggests that influenza virions may encode an antagonistic protein[81], although this is controversial[78].

Like TRIM5α, it has been suggested that the human isoform of tetherin not only has a physical interaction with the virus envelop, but also causes a signal cascade that results in an NF-κB inflammatory response[82]. It has been shown that both the extracellular and cytoplasmic domains are required for signalling, although it is independent of virion endocytosis. Furthermore, recruitment of an auxiliary factor, TRAF6, and activation of TAK1 are critical for signalling.

1.3.1.4 IFITs

Four members of the IFN-induced protein with tetratricopeptide repeats (IFIT) family have been characterised in humans (*IFIT1, 2, 3* and *5*) although homologues have been identified in several mammals, birds, fish, and amphibians[83]. Constitutive expression is very low, but IFITs are induced by type I IFN stimulation and ligation of PRRs with PAMPs. These proteins exist cytoplasmically and function through a variety of distinct mechanisms. Human IFIT1 and 2 inhibit translation initiation of virus proteins, including HCV. IFIT1 and 2 bind to eukaryotic initiation factor 3E (eIF3E), preventing binding of eIF3 to eIF2-GTP-met-tRNA[84]. Human IFIT2 also binds to eIF3C preventing the pre-initiation complex from binding to mRNA. IFIT1 has also been shown to recognise viral RNAs by their lack of 2'-O methylation of the 5' cap, preventing binding to the pre-initiation complex[85]. Further evidence suggests that

IFIT1, in complex with IFIT2 and IFIT3, recognises uncapped viral RNAs, and sequesters them in the cytoplasm by an unknown mechanism[86].

1.3.1.5 APOBEC3G

APOBEC3G, part of the *APOBEC3* gene family, is unique to primates. It was first discovered as a factor that was expressed by T cells non-permissive to HIV-1 infection. During lentiviral infection, APOBEC3G is packaged into newly-produced viral particles. Upon infection of a new target cell, the retroviral genome is released and replicated. APOBEC3G deaminates cytidines in the negative strand of retroviral genomes (Figure 8), thereby catalysing 'G' to 'A' mutations in the proviral genome, preventing efficient replication and disrupting the conserved coding function of the HIV genome[87]. The action of APOBEC3G is therefore the cause of G-to-A hypermutation patterns commonly found in clinical HIV samples[88]. APOBEC3G is well-conserved across primate species, except for localised regions of strong positive selection[89]. The residues W127 and Y124 of APOBEC3G were found to be important for encapsidation of the protein into HIV-1 virions[90], and later studies showed that a direct interaction with HIV-1 Gag nucleocapsid protein facilitated this process[91].

1.3.1.6 Restriction Factors that Require Activation

ISG15 is a 15 kDa ubiquitin-like protein (ubiquitin homologue) that is one of the proteins most highly-induced following type I IFN release. The 165 amino acid pre-protein is processed to expose a C-terminal 'LRLRGG' sequence[92]. ISG15 is then activated by the E1-like ubiquitin-activating enzyme (UBE1L), and conjugated to a number of other enzymes including UBCH8 and HERC5. This process allows ISG15 to bind to proteins in a process called ISGylation. Unlike ubiquitination, ISGylation does not mediate protein degradation but instead has an activating effect. More than 160 putative ISG15 targets have been identified, both cellular components and viral proteins[93]. For example, ISG15 has been shown to prevent virus-mediated degradation of IRF3, increasing transcriptional activity and thus increasing the expression of IFNβ[94]. Furthermore, ISG15 ISGylates HIV-1 protein Gag, preventing its ubiquitination, which is needed to release the virions from host cells[95].

Mice deficient in ISG15 had an increased susceptibility to IAV, influenza B virus, Sindbis virus, and herpes simplex virus 1 (HSV-1). Conversely, knock-out of de-

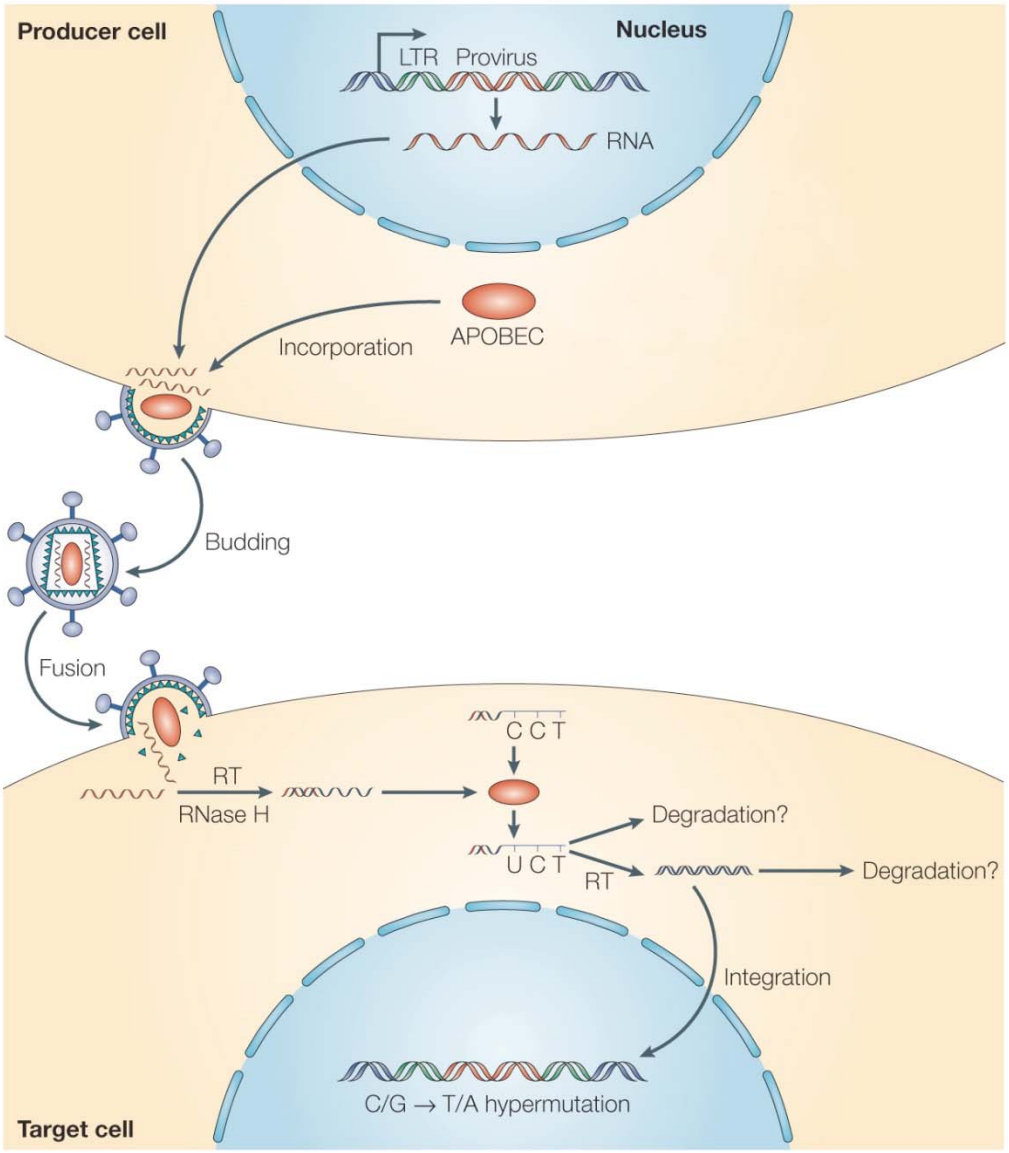


Figure 8: APOBEC3G is a cytidine deaminase

During lentiviral infection, APOBEC3G is incorporated into the newly-produced virus particle. During infection of the next cell, APOBEC3G is released at the same time and deaminates cytidines to uracil in the negative strand of newly-replicated genomes. This causes an accumulation of G to A mutations in the newly-synthesised viral genome that can disrupt coding function.

Adapted from Harris *et al.* (2008)[96]

ISGylating enzymes (such as USP18) *in vivo* increased the resistance to viral infection in mice[92].

OAS and PKR are two more potent IFN-inducible restriction factors, but both require activation through binding of dsRNA to become functional. After dsRNA binding OAS generates 2'-5' oligoadenylates that act as a co-factor for Ribonuclease L (RNAseL), an enzyme that cleaves both cellular and viral RNA[97]. This prevents viral replication but can also induce cell apoptosis, preventing viral spread. Similarly, post-dsRNA binding, PKR phosphorylates eukaryotic initiation factor 2α (eIF2α), resulting in a general translational block[46]. Deletion of *PKR* in mice allows NS1 deficient IAV to replicate[98].

1.4 The Interferon-Inducible Transmembrane (IFITM) Family

The IFITMs are another family of anti-viral proteins, but are the first restriction factors discovered to act specifically on virus entry. The *IFITM* gene family was initially identified more than 20 years ago[99], but the transcripts were originally named 9-27, 1-8D and 1-8U. Several groups investigated the role of IFITM1 (9-27) in homotypic adhesion of leukemic B cells (IFITM1 was found to interact with B cell receptors CD19, CD21 and CD81[100]) and the role of Ifitm1 and Ifitm3 in germ cell maturation in the mouse embryo[101]. Ifitm3 is an important marker for germ cell competence in mice, however these roles were called into question when a mouse with a deletion of the IFITM locus (*IfitmDel* $^{-/-}$) developed normally[102].

However the antiviral properties encoded by IFITM proteins were only established in 2009 during an RNAi screen for host factors that influence IAV replication. Knock-down of *IFITM3 in vitro* led to enhanced viral replication whilst, conversely, over-expression of *IFITM1*, *2* or *3* inhibited early viral replication[1]. Subsequent genome analyses have indicated that the *IFITM* genes are likely to have arisen by gene duplication very early in vertebrate evolution[103], since 'lower' vertebrates, such as lampreys, possess at least one *IFITM*-like gene[104]. To date, five *IFITM* genes have been identified in humans, of which *IFITM1, 2, 3* and *5* are clustered within a 26 Kb region towards the telomere on the short arm of chromosome 11. *IFITM5* is not IFN inducible and is involved in bone mineralisation[105]. The fifth gene, *IFITM10*, is

located 1.4 Mb towards the centromere of chromosome 11, but little is known about its function. *IFITM6* is not present in humans, but is located close *to Ifitm1, 2, 3* and *5* on chromosome 7 of the mouse genome[106]. *Ifitm10* is located towards the centromere as is the case in humans, but the family has been further expanded to include *Ifitm7* on chromosome 16, resulting in a total of seven murine *Ifitm* genes. Orthologous and paralogous genes have also been found in other mammals[103], including marsupials[104].

Although the molecular function of these proteins has largely been studied in cell culture systems, studies in mice and humans suggest IFITM proteins, and IFITM3 in particular, restrict IAV infection *in vivo. Ifitm3* $^{-/-}$ mice fail to control infection by mildly-pathogenic strains of IAV compared to their wild-type littermates, developing fatal fulminant viral pneumonia[3,107].

1.4.1 Broad-Spectrum Antiviral Function

Initial investigations into the inhibitory activities of IFITM1, 2, and 3 *in vitro* using both normal or pseudotyped viruses demonstrated that, in addition to IAV, entry and infection by representatives of multiple virus families was also inhibited by over-expression of IFITMs, particularly IFITM3. These families include: filoviruses, rhabdoviruses, coronaviruses, and flaviviruses (see Table 1). Interestingly, these restricted viruses are all enveloped, with ssRNA genomes, and considered to enter cells by membrane fusion following endocytosis. However, most recently, evidence has been published suggesting that IFITM3 can also restrict a non-enveloped reovirus[108], and a respiratory syncytial virus (RSV), which fuses at the cell membrane[109]. This evidence suggests the range of viruses influenced by the IFITM proteins is not limited to those with an envelope or those using the endosomal pathway.

However, some retroviruses (*e.g.* moloney leukemia virus [MLV]), several arenaviruses, and two DNA viruses (human cytomegalovirus [HCMV] and adenovirus type 5 [Ad5]) are apparently not restricted by IFITMs[110]. Although restriction of HIV-1 infection was not initially detected[1], several more recent studies have reported some restriction of cell infection[111-113]. Contrary to these examples, it has been found that over-expression of IFITM3 causes an increase in infection by human coronavirus HCoV-OC43[114] and that over-expression of IFITM1 and IFITM3 modestly enhanced

Table 1: A summary of the viruses IFITM proteins have been tested against

Enveloped							
Family	**Virus**	**pH dependent**	**Restricts Infectivity**	**Prevents cell-cell fusion**	**Pseudotyped virions (P) or live virus (L)**	**Restriction status**	**Reference**
Orthomyxoviridae	Influenza A virus	✓✓	✓	✓	P L	M1-3	Brass *et al.* (2009)
	Influenza B virus	✓✓	✓		L	M1-3	Everitt *et al* (2012)
Flaviviridae	West Nile virus	✓	✓		P	M1-3	Brass *et al.* (2009)
	Dengue virus	✓✓	✓		P	M1-3	Brass *et al.* (2009)
	Hepatitis C virus	✓	✓ / ×		P L	M3 – No M1 – Yes	Brass *et al.* (2009), Wilkins *et al.* (2013)
Rhabdoviridae	Vesicular stomatitis virus	✓	✓	✓	P L	M1-3	Weidner *et al.* (2010)
Filoviridae	Marburg virus	Δ	✓		P L	M1-3	Huang *et al.* (2011)
	Ebola virus	Δ	✓		P L	M1-3	Huang *et al.* (2011)
Coronaviridae	SARS coronavirus	Δ	✓		P L	M1-3	Huang *et al.* (2011)
	HCoV-OC43	Δ	×		P L	M2 and 3 enhanced	Zhao *et al.* (2014)
Retroviridae	HIV-1	×	✓ / ×		P L	Mixed results	Brass *et al.* (2011), Lu *et al.* (2011), Jia *et al.* (2012)
	Murine leukemia virus	×	×		P L	No	Brass *et al.* (2009), Huang *et al.* (2011)
	Jaagsiekte sheep retrovirus	✓	✓	✓	P	M1 best	Li *et al.* (2013)
Arenaviridae	Lassa virus	✓	×		P	No	Brass *et al.* (2009)
	Machupo virus	✓	×		P	No	Brass *et. al.* (2009)
	Lymphocytic choriomeningitis virus	✓	×		P	No	Brass *et al.* (2009)
Alphaviridae	Semliki Forest virus	✓	✓	✓	L	M2 and M3 best	Li *et al.* (2013)
Bunyaviridae	La Crosse virus	✓✓	✓		L	M1-3	Mudhasani *et al.* (2013)
	Hantaan virus	✓✓	✓		L	M1-3	Mudhasani *et al.* (2013)
	Andes virus	✓✓	✓		L	M1-3	Mudhasani *et. al.* (2013)
	Rift Valley fever virus	✓✓	✓		L-attenuated	M2 and M3	Mudhasani *et al.* (2013)
	Crimean-Congo Haemorrhagic fever virus	✓✓	×		L	No	Mudhasani *et. al.* (2013)
Herpesviridae	Human cytomegalovirus[§]	✓	×		P	No	Warren *et al. (2014)*
Paramyxoviridae	Respiratory syncytial virus	×	✓		L	M3	Everitt *et al.* (2013)
Non-Enveloped							
Family							
Reoviridae	Reovirus	✓✓	✓		L	M3	Anafu *et al.* (2013)
Papillomaviridae	Human papillomavirus 16[§]	✓	×		P L	M1 and 3 enhanced	Warren *et al. (2014)*
Adenoviridae	Adenovirus 5[§]	✓	×		P	No	Warren *et al. (2014)*

✓: fuses at pH >6, ✓✓: fuses at pH <6, ×: does not require fusion, Δ: requires cathespin L in lysosome, §: DNA virus

Adapted from Smith *et al.* (2014)

human papillomavirus 16 (HPV16) infection[110]. Further studies in *Ifitm3* $^{-/-}$ mice have also shown that IFITM3 is a viral-specific restriction factor – this protein does not prevent infection by intracellular bacteria (*Salmonella typhimurium, Citrobacter rodentium, Mycobacterium tuberculosis*) or protozoa (*Plasmodium berghei*)[109].

Using a pseudotype virus carrying the Jaagsiekte sheep retrovirus (JSRV) envelope protein (Env), for which fusion requires initial Env priming by receptor binding and subsequent exposure to pH 6.3, IFITM1 restricted replication of JSRV more potently than IFITM2 and 3[115]. As IFITM1 appears to be located earlier in the endocytic pathway, where the pH is higher[116], these data suggest that the cellular location of different IFITM proteins determines the range of viruses that each restricts. Although not strictly pH-related, virus restriction correlates with the cellular compartment where cellular membrane penetration occurs *e.g.* the endosome. Differential restriction of viruses in the vector-borne *Bunyaviridae* family has also been found[117] (Table 1); only IFITM2 and 3 were capable of restricting Rift Valley fever virus (RVFV), and none of the IFITM proteins prevented replication of Crimean-Congo haemorrhagic fever virus (CCHFV). The reason(s) underlying this difference in susceptibility are unclear as the bunyaviruses share similar morphologies and glycoproteins on their envelopes, although CCHFV is from a different genus (nairovirus) than the other susceptible viruses.

1.4.2 Protein Structure, Cellular Distribution, and Trafficking

In mice, Ifitm3 is expressed constitutively in cells of the upper airway and visceral pleura[107], but its expression, and that of IFITM1 and 2, in humans *in vivo* is poorly understood. However, immunohistochemistry on organs harvested from C57BL/6 mice showed that expression of Ifitm3 was strong at the predominant sites of pathogen infection, including the lymph nodes, lungs, spleen, liver, and intestines[109].

In both cell lines and primary cells *ex vivo*, IFITM protein expression is upregulated by type I IFN. Of the three IFN-inducible human IFITM proteins, IFITM3 and IFITM2 share 90.2 % sequence similarity at the amino-acid level, and IFITM3 and IFITM1 share 73.7 % sequence similarity (excluding the N-terminal deletion). They are all membrane located, though their topologies remain to be clearly established. Initially proposed as transmembrane proteins (Figure 9), with both the N- and C-termini located externally, subsequent studies suggested both the N- and C-termini, as well

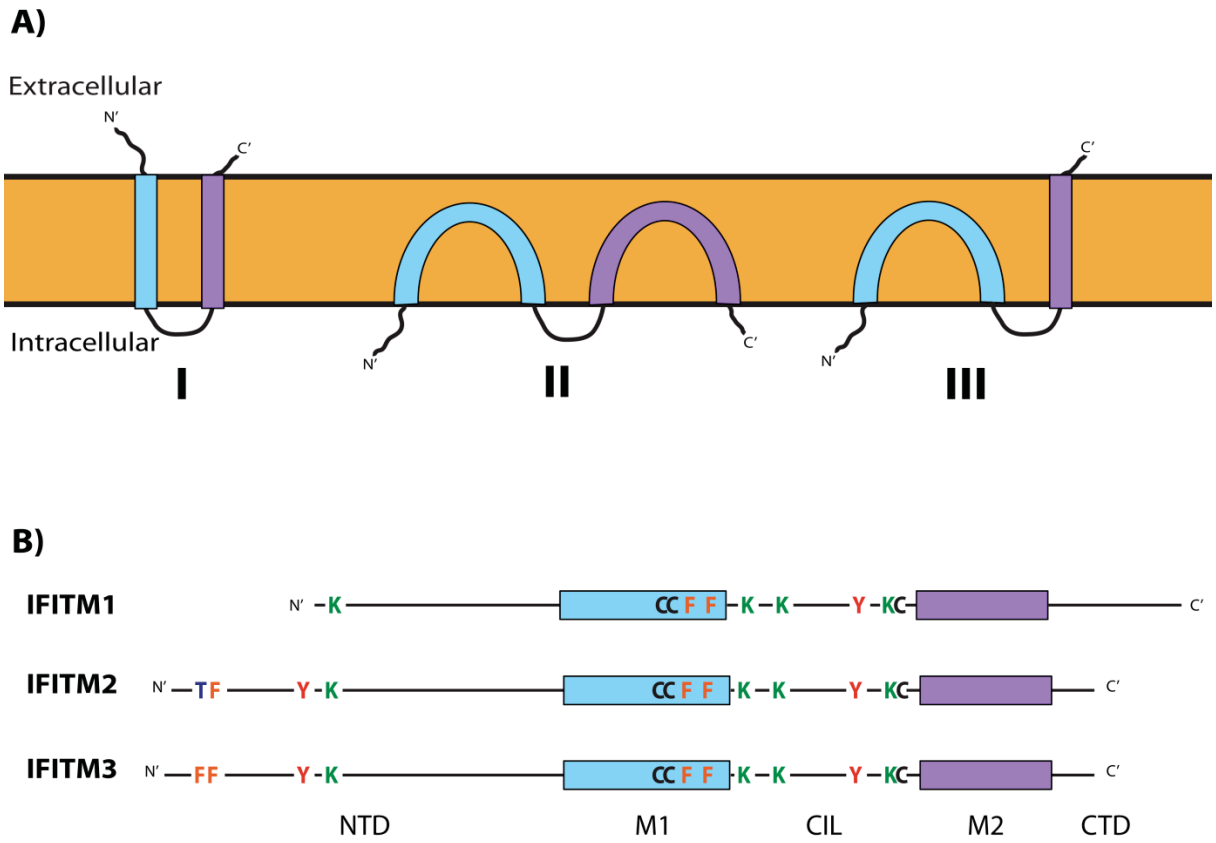


Figure 9: IFITM protein topology and domain organisation

Panel A. Topological models for IFITM proteins. (I) Represents an initial model for the proteins as transmembrane molecules with both the N- and C-terminal domains (NTD and CTD) extracellular, and the conserved intracellular loop (CIL) facing the cytoplasm [118]. Subsequently, an alternative model (II) was proposed with the NTD, CTD and CIL all positioned intracellularly, and neither membrane domain (M1 and M2, blue and purple respectively) spanning the bilayer[119]. The most recent model (III) combines models (I) and (II), positioning the NTD and CIL in the cytoplasm and the CTD extracellularly. Currently, the topology represented by (III) is only established for murine Ifitm3[120].

Panel B. Linear representation of human IFITM1, 2 and 3 showing key amino acids. In all cases, modifications and functional activities have only been established for these amino acids in IFITM3, but conserved residues in IFITM1 and 2 are shown.

Adapted from Smith *et al.* (2014)[121]

as the hydrophobic domains interacting with the membrane but not spanning it[119]. More recently, a model for IFITM3 in which the N-terminal domain (NTD) and CIL domain are located in the cytoplasm, and the CTD is extracellular, supports a type II transmembrane topology for the second hydrophobic domain[120]. However, it is possible that all three models are correct, and that the proteins move dynamically between each configuration to achieve function. All three IFN-inducible IFITM proteins contain conserved cysteine residues at the junctions of the CIL domain and the putative membrane interacting domains. These cysteines (C71, 72 and 105 in human IFITM3, Figure 9B) are palmitoylated, with this modification required for full viral restriction[118]. Substitution of the cysteines for alanines inhibits IFITM3 clustering in membranes and reduces its antiviral function[118].IFITM3 can also be ubiquitinated on any of four lysines in the NTD and CIL domain. Ubiquitination enhances IFITM3 turnover[119], thus substitution of the lysines with alanines slows the protein's degradation and increases its antiviral activity[116]. John *et al.*[5] showed that IFITM3 can interact with itself, as well as IFITM1 and 2, and that phenylalanine residues (F75 and F78) are required for this interaction. Although the significance of this association is unclear, the formation of homo- and/or hetero-oligomers might also influence the distribution and functional activities of these proteins.

The NTD of IFITM1 is short (35 amino acids) compared to IFITM2 and 3, which are 20 and 21 amino acids longer, respectively (Figure 9). These N-terminal extensions include a key tyrosine residue (IFITM3_Y20) that appears to control the cellular distributions of the two longer IFITMs[3,5,111] (Figure 9B). Thus IFITM1 is predominantly at the plasma membrane, while IFITM2 and 3 are located mainly in intracellular compartments. IFITM3 is reported to reside primarily in endosomal organelles, identified by co-labelling with endosomal markers, including Lamp1, Rab7, and CD63[4,5,122], (Figure 10) but the location of IFITM2 remains to be clearly established.

Like all type II transmembrane proteins, *IFITM* mRNA is bound to a free ribosome in the cytosol, but translation is paused upon detection of an alpha helix, characteristic of membrane proteins, by the signal recognition particle (SRP)[9]. The SRP traffics the ribosome and bound mRNA to the ER where translation continues and the protein is folded correctly into the ER membrane. The folded proteins are trafficked from the ER to the Golgi network via COPI- and COPII-coated vesicles[123]. It is thought that the 20-YEML-23 sequence in the N-terminus of IFITM2 and 3 may be a component of a

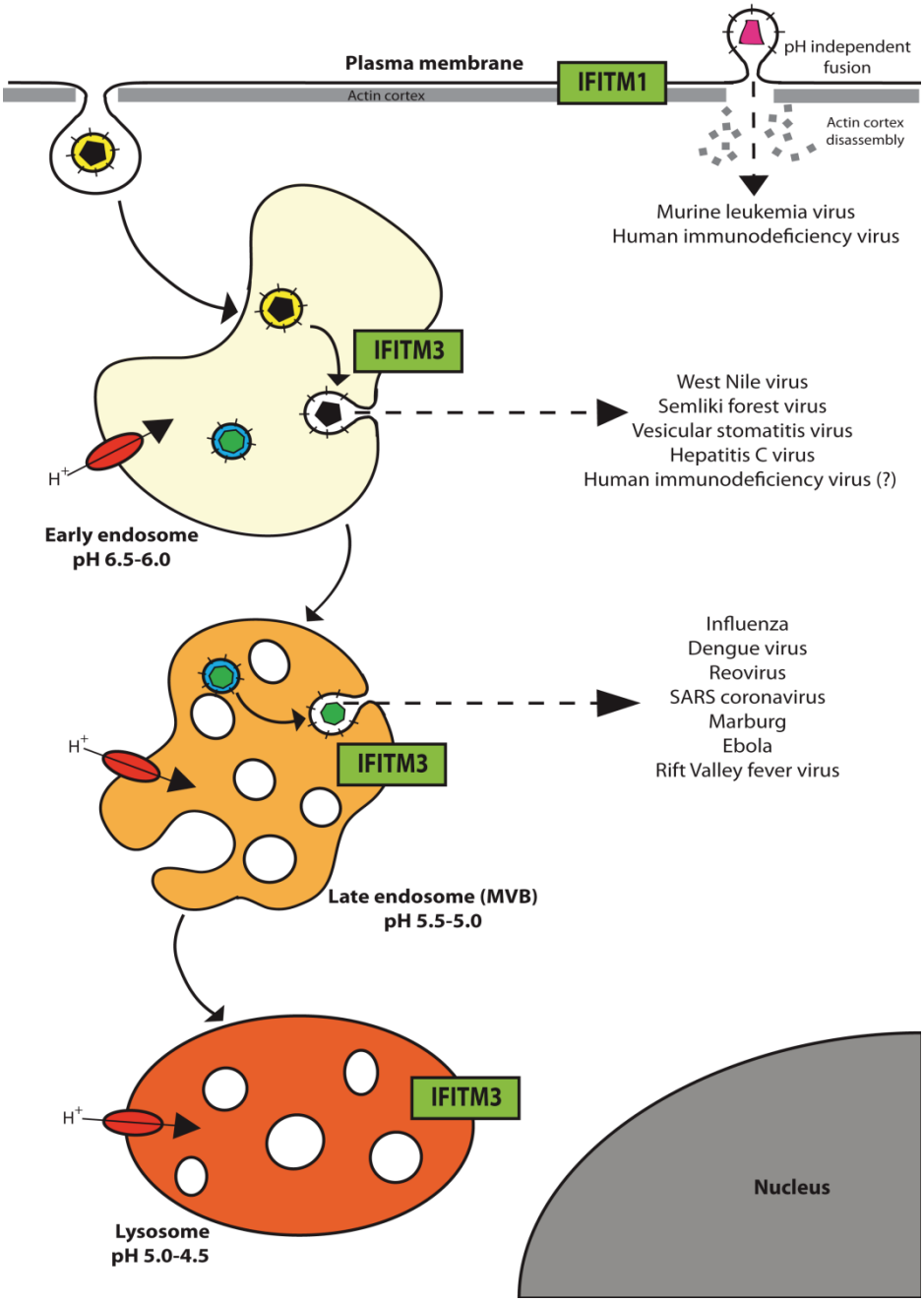


Figure 10: IFITM proteins inhibit virus entry at different stages of cell trafficking

Viruses enter cells by fusing with or penetrating a limiting cellular membrane. For most enveloped viruses, fusion occurs either at the cell surface or, following uptake by endocytosis, from within endosomes. Acid-dependent viruses require acidification of the endosomal lumen by the membrane-associate vacuolar proton ATPase (shown in red) for fusion. Trafficking through the endocytic system, from early to late endosomes, exposes virions to increasingly acidic environments. IFITM proteins (green) can inhibit entry and infection by a number of viruses that fuse/penetrate at the cell surface or from within endosomes. IFITM1 is expressed primarily at the cell surface, while IFITM2 and 3 are primarily intracellular. IFITM3 has been localised to endosomal compartments, but the distribution of IFITM2 still needs to be clearly established. Adapted from Smith *et al.* (2014)[121].

YxxØ-type sorting signal for clathrin-mediated trafficking[111], where x is any amino acid and Ø is valine, leucine, or isoleucine. The N-terminal cytosolic sorting signal on the protein identifies it for trafficking to endosomes in clathrin-coated vesicles, deviating from the 'default' pathway to the cell surface[123]. The YxxØ motif has been shown to interact with the μ-subunit of the adapter protein 2 (AP-2) complex[124], which associates with clathrin proteins in lipid membrane. The AP-2 complex is known to be involved in internalisation and lysosomal targeting[125]. Significantly, Y20 has also been identified as a target for Fyn-mediated phosphorylation, suggesting that perhaps the activity of this motif as a trafficking signal can be regulated[111,126]. However, Chesarino *et al.* showed that phosphorylation by Fyn was not required for IFITM3's antiviral activity, but that the N-terminal tyrosine may have a dual-function of regulating ubiquitination and endocytosis[127]. It is important to note, however, that studies of the subcellular location of the IFITMs to date have been based on HA- or Myc-tagged proteins, where tagging and/or over-expression (in transient systems) may have an impact on protein localisation and/or detection.

1.4.2.1 Mode of Action

Experiments with reovirus subvirus particles (ISVPs), which do not require endosomal acidification for entry and are not inhibited by IFITM3 expression, have suggested that IFITM3 may perturb endosomal acidification[108]. However, studies with various enveloped viruses suggest a different mode of action. Morphological analysis of IFITM3-restricted IAV in cells showed the accumulation of viral particles in acidified endosomal compartments, suggesting IFITM3 has no effect on receptor-binding, endocytosis or acidification[2,4].

Studies using cell-cell fusion assays suggest that IFITM3 blocks enveloped virus entry by preventing fusion of the viral membrane with a limiting membrane of the host cell, either the plasma membrane and/or endosomal membranes[115]. Fusion is an essential step in enveloped virus entry, and results in the transfer of viral capsids into the cytoplasm of a target cell. This process is extremely well-characterised for a number of viruses, in particular IAV. Low pH in the endosomal lumen triggers conformational changes in one of the viral envelope proteins – haemagglutinin (HA) – that results in fusion of the outer leaflet of the viral membrane with the inner leaflet of luminal endosomal membrane, forming a short-lived hemifusion intermediate.

Resolution of the hemifusion intermediate allows fusion of the viral membrane inner leaflet with the cytoplasmic leaflet of endosomal membranes and the opening of a stable fusion pore[128].

Although often not a reflection of the pathway of infectious virus entry, a commonly-used approach to studying viral fusion mechanisms is the formation of syncytia by cell-cell fusion. This requires the presence of viral fusion proteins in the plasma membrane of cells and appropriate signals, such as receptor-bearing cells and/or a transient change in the pH of the medium. Using JSRV Env (discussed previously), the IFITMs had no effect on either priming or pH-induced conformational changes[115]. Moreover, syncytia formation induced by representatives of all three classes of viral fusion proteins[129] could be blocked by IFITM1. Using cold to arrest fusion at the hemifusion state, and chlorpromazine, which resolves cold-arrested hemifusion intermediates, IFITM proteins were found to inhibit the initial stages of fusion leading to the formation of hemifusion intermediates[115].

The mechanism(s) through which the IFITMs inhibit the early stages of fusion is unclear. Two-photon laser scanning and fluorescence lifetime imaging (FLIM) of Laurdan-labelled cells, together with the effects of oleic acid treatment on cell-cell fusion, suggest that IFITM proteins may reduce membrane fluidity and increase spontaneous positive curvature in the outer leaflet of membranes[115]. Such changes might be expected to impact on fusion, but how IFITMs affect membrane fluidity, and whether or not this has consequences for other membrane functions in the absence of infection, is unknown.

However, one mechanism has been suggested from experiments on IFITM3. Amini-Bavil-Olyaee *et al.* showed IFITM3 interacts with vesicle membrane protein associated protein A (VAPA) and disrupts its association with an oxysterol binding protein (OSBP) that regulates the cholesterol content of endosomal membranes (Figure 11). Therefore, over-expression of IFITM3 increases endosomal cholesterol, which may impact viral fusion through a corresponding decrease in endosomal membrane fluidity[122]. Further support for this paradigm is given by Lin *et al.* who show that amphotericin B (an anti-fungal treatment that forms complexes with sterols) rescues IAV infection in IFITM3 over-expressing cells[130]. However, Desai *et al.* show that although IFITM3 prevents the formation of fusion pores,

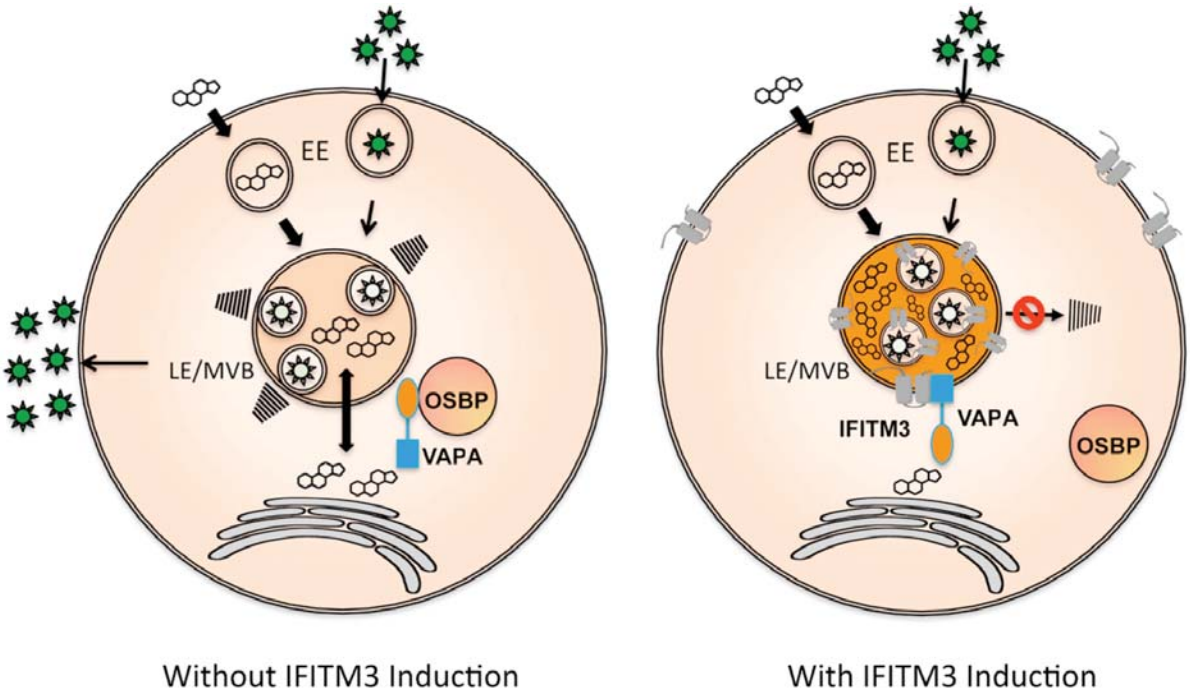


Figure 11: Schematic model of IFITM3-VAPA anti-viral activity

Amini-Bavil-Olyaee *et al.* present a model of infection in the absence of IFITM3 (left) and with IFITM3 (right). VAPA interacts with OSBP in the cytoplasm, and virus particles enter the cell via the late endosomal pathway. The viral and cellular membranes fuse allowing viral escape into the cytoplasm. Expression of IFITM3 causes a disruption in the VAPA-OSBP interaction and thus alters cholesterol homeostasis. Cholesterol accumulates in the endosomes and prevents fusion of the virus and cellular membranes.

Adapted from Amini-Bavil-Olyaee *et al.* (2013)

cholesterol-laden endosomes are permissive for virus fusion[6]. Overall, evidence to date suggests that IFITM3 stabilises the membranes through a physical change in membrane properties, either directly or indirectly, and this results in an anti-viral effect.

1.4.2.2 Evolution

Orthologues of the human *IFITM* genes were shown, by RT-PCR, to be expressed in several mammalian species, including mice[107], rats[131], and marsupials[104]. However, more in-depth bioinformatic approaches have been used to identify orthologues in species where robust laboratory reagents (*e.g.* immortalised cell lines) are not established. Zhang *et al.* used tBLASTn searches to identify 286 *IFITM*-like sequences from 27 vertebrate genomes[132]. The species selected included all of the major evolutionary lineages of vertebrate: five species of fish, one amphibian, one reptile, three birds, and 17 mammals (including representatives from the primates, glires, metatherians and prototherians). The authors identified 29 *IFITM*-like genes in humans, 18 of which they classified as pseudogenes. However, it is difficult to know whether or not the greatest numbers of pseudogenes were identified in humans simply because it is the most well-annotated genome. Also, because of high sequence-similarity between human *IFITM2* and *3*, it is hard to interpret the results of these BLAST searches. In addition, although Zhang *et al.* used bioinformatic algorithms to identify functional genes based on conserved sequence motifs, this data has not been corroborated *in vitro* or *in vivo*.

Zhang *et al.* carried out phylogenetic analysis of the *IFITM* gene family across different species and showed two separate clusters of *IFITM1* and *IFITM2/3* sequences within the primate sub-clade. The *IFITM1* cluster was located at the basal position of this sub-clade, which indicates that *IFITM1* diverged earlier than *IFITM2* and *IFITM3* during primate evolution. The *IFITM2* and *3* cluster contains sequences from three of the six primates (human, chimpanzee, and gorilla), suggesting that the duplication events that gave rise to *IFITM2* and *3* occurred prior to the speciation of these three hominids.

Several viral restriction factors, including tetherin, Zinc-finger antiviral protein (ZAP), APOBEC3G, and TRIM5α, have been shown to evolve under positive selection[69,77,89,133]. As discussed previously, positive selection is indicated when the ratio of non-synonymous mutations (d_N) to synonymous mutations (d_S) at a codon, or

averaged over a specific region, is greater than one. It was hypothesised that IFITM3, being similar in structure and function to these proteins, could also be under positive selection in the genome. However, Zhang *et al.* undertook pairwise comparisons between all of the *IFITM-like* sequences and did not detect any positively-selected sites in the primate *IFITM1-3* genes[132]. Zhang *et al.* employed bioinformatic algorithms to identify *IFITM* genes, which can result in mis-identification of pseudogenes as functional genes, making phylogenetic analysis difficult to interpret. Moreover, although the central CD225 domain of IFITM proteins are highly conserved, the N- and C-termini are very diverse, which makes aligning the sequences, and therefore also the phylogenetic conclusions, unreliable. It is also difficult to determine whether or not these analyses should include all the members of a gene family as Zhang *et al.* did, or just to include the orthologous genes. Since the *IFITM* gene family arose by gene duplication, each gene could be evolving independently, and therefore a signal of positive selection may have been lost.

1.5 Viral Antagonism of the Innate Immune Response

The Red Queen Hypothesis states that viruses co-evolve with their hosts in order to maintain fitness with them. To this end, viruses have evolved to evade restriction factors by expressing antagonistic proteins as a countermeasure, allowing continued viral replication.

1.5.1 HIV Immune Antagonist Proteins

1.5.1.1 Vif Protein

APOBEC3G is a cytidine deaminase that targets the negative strand of HIV genomes, causing an accumulation of G-to-A mutations. Cell lines permissive to HIV-1 replication were not permissive for viruses that lacked the viral infectivity factor (Vif) protein, as Vif antagonises APOBEC3G activity by targeting it for polyubiquitination and degradation by the 26S proteasome[134] (Figure 12). The neutralising relationship between Vif and APOBEC3G is species-specific – Vif from SIV is unable to abrogate the effect of human APOBEC3G. This specificity is due to a single amino acid; an aspartic acid in the human APOBEC3G protein is replaced with a lysine in the simian protein at position 128[135].

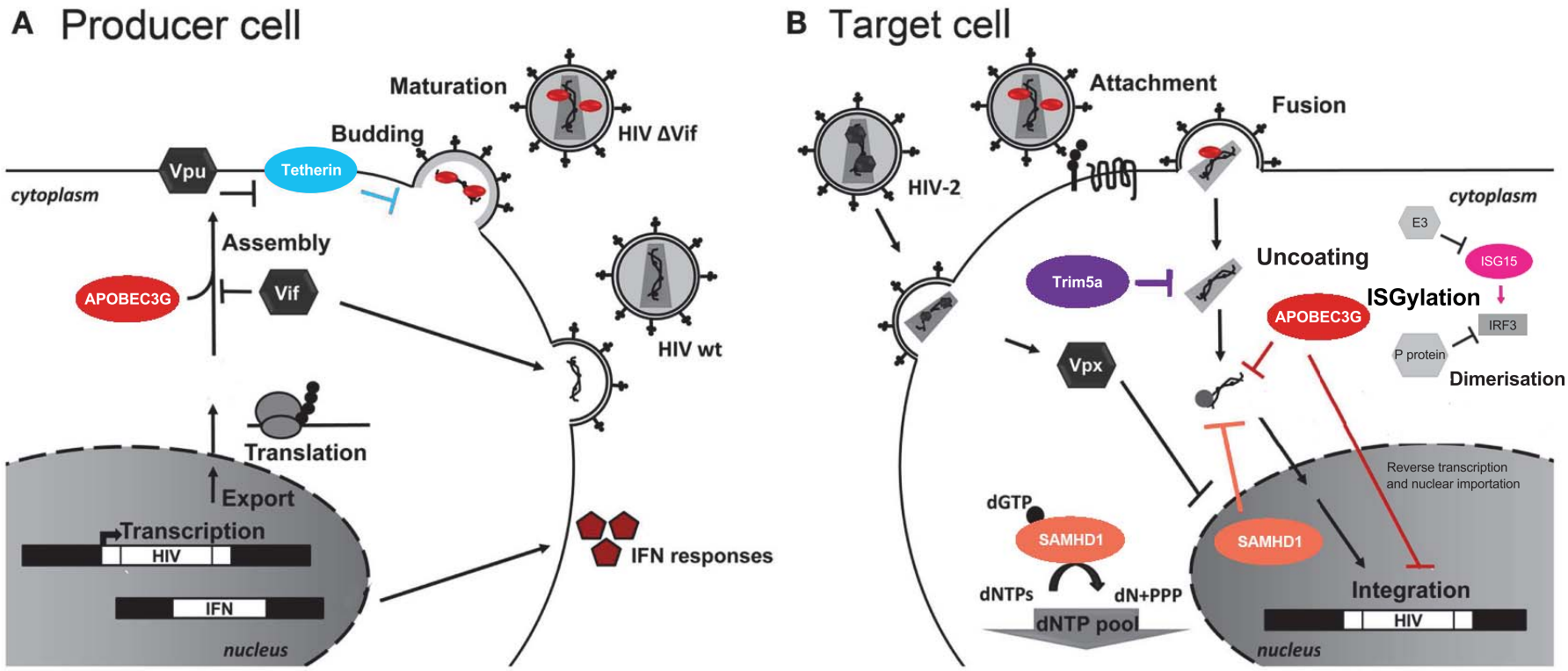


Figure 12: Antiviral restriction factors and their antagonists

A schematic to show the location of important cellular restriction factors (coloured ovals) and their broad mechanism of action, alongside their antagonistic viral proteins (grey hexagons). Black arrows represent the course of viral replication, flat arrows represent inhibition.

Adapted from Santa-Marta *et al.* (2013)[136]

As Vif orthologues are present in many related lentiviruses[137], Vif may have arisen in the progenitor lentivirus before the divergence of the host species. Currently, the structure of Vif in complex with APOBEC3G and the E3 ligase has not been established. The 3D structure of this interaction could reveal vital amino-acid interactions that could be targeted for anti-viral drug therapy.

1.5.1.2 Vpx Protein

HIV primarily replicates in $CD4^+$ T cells, and patients with AIDS usually die from an opportunistic infection due to profound T-cell immune deficiency. HIV replication in other cell types such as macrophages follows different kinetics because of SAMHD1, a dNTPase, expressed in the myeloid cells[138,139]. SAMHD1 expression depletes the availability of nucleotides required for HIV-1 reverse transcription. However, the Vpx protein of HIV-2 also directs the degradation of SAMHD1 by targeting it for ubiquitination, thereby retaining the nucleotide pools for efficient reverse transcription of the viral genome (Figure 12).

1.5.1.3 Vpu Protein

Vpu, encoded by HIV-1, is a single-pass type I transmembrane protein, which contains two serine residues in close proximity (S52 and S53). These residues are phosphorylated to recruit auxiliary factors, including β-TrCP (β-Transducin Repeat Containing E3 Ubiquitin Protein Ligase). This complex allows Vpu to downregulate and degrade tetherin, as well as displace it from virions at the cell surface. Residues 2-4 and 25-27 in the transmembrane domain of Vpu interact strongly with adjacent residues in the transmembrane domain of tetherin. However it is the first α-helix of Vpu that mediates the degradation of tetherin, and the second α-helix that mediates displacement[140]. It is likely the interactions in the transmembrane domain facilitate recruitment of the cytoplasmic domain of Vpu to tetherin. High-level expression of Vpu has been found to cause the ER-associated degradation of tetherin,[141] although the mechanism has not been clearly established.

1.5.2 Influenza Virus NS1

Non-structural protein 1 (NS1) of IAV is a more broadly-acting viral protein that has several mechanisms to reduce the host immune response to infection (Figure 13). NS1 can bind to dsRNA, masking it from the OAS/RNAse L[142]. NS1 also inhibits

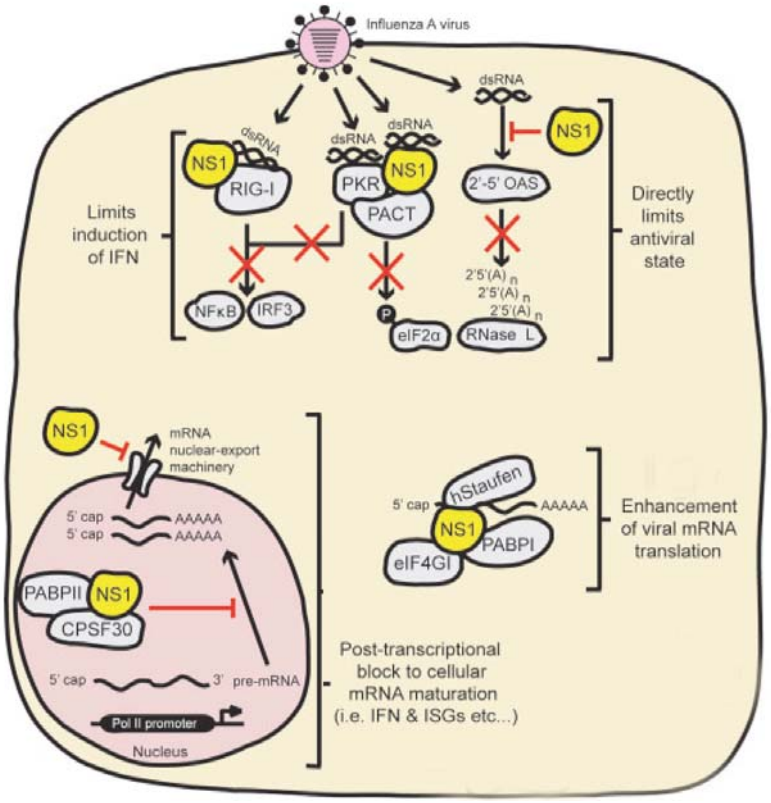


Figure 13: Influenza's NS1 has multiple functions within an infected cell

NS1 can function to block pre-transcriptional IFN induction by sequestering RIG-I and PKR. It can prevent processing and export of cellular mRNA, whilst enhancing vRNA translation. Some studies have also shown that NS1 may have an impact on apoptosis.

activation of RIG-I (section 1.7.2.2), the cytosolic PRR that detects viral RNAs[143], by binding and sequestering TRIM25, a ubiquitin ligase required for RIG-I activation. Furthermore, NS1 also decreases transcription of cellular mRNAs, including other restriction factors, by binding to cleavage and polyadenylation specificity factor 30 (CPSF30) and thus preventing 3' polyadenylation[144]. NS1 is also reported to have anti-apoptotic properties, preventing the infected host cell from dying quickly and limiting viral replication. NS1 also targets the epigenome of an infected cell; NS1 of influenza A H3N2 encodes a histone-like sequence that the virus uses to target the human PAF1 transcription elongation complex (hPAF1C). Loss of transcription of anti-viral restriction factors results in greater susceptibility to virus infection[145].

NS1-deleted IAV were non-pathogenic in wild-type mice, but virus replication was rescued in STAT1$^{-/-}$ mice[146], *i.e.* those that are lacking an interferon response. These mutant viruses have been shown to be effective as vaccines in wild-type mice, providing immunity against NS1-competent virus challenge four weeks later[147].

1.5.3 Rabies P Protein

The rabies virus only encodes five proteins, one of which is the phosphoprotein (P protein). The P protein functions to prevent detection of infection by inhibiting the activation (dimerisation) of IRF3 by TANK-binding kinase 1 (TBK1)[148]. It also interacts with STAT1 and STAT2 to form an inactive complex, thus preventing the dimers from interacting with IRF9 and stimulating ISG expression[149].

1.5.4 Vaccinia E3 Protein

The E3 protein of vaccinia virus has multiple functions. The CTD of E3 has been shown to bind to ISG15, which inhibits its ISGylation function[150], thereby encouraging the degradation of cellular proteins such as IRF3 (section 1.3.1.6). E3 also sequesters dsRNA using the dsRNA-binding domain at its amino terminus, preventing detection by PKR and RNAseL (two IFN-inducible proteins that, upon activation, trigger inhibition of virus replication[151]).

A lot is known about HIV-1 antagonistic proteins, but it is likely that many viruses have alternative strategies to abrogate viral restriction factors. Understanding how these viral proteins interact with host proteins is important for drug and vaccine development.

1.6 Influenza A Virus

IAV, a member of the *Orthomyxoviridae* family, causes an infectious disease in birds and mammals. IAV is primarily spread between mammals via the aerosolisation of virus particles in the droplets expelled during sneezing and coughing, and predominantly via the faeco-oral route in birds. IAV has a single-stranded negative-sense RNA genome divided into eight segments[152]. Each nucleic acid segment encodes between one and three proteins, and is closely associated with viral nucleoproteins (NP) and the viral RNA polymerase complex (PA, PB1, and PB2) to form a viral ribonucleoprotein complex (vRNP) (Figure 14). Influenza viruses are encased within a host-derived lipid-membrane envelope containing two transmembrane proteins, haemagglutinin (HA) and neuraminidase (NA), which are transported to the plasma membrane during virus budding. The antigenicity of the HA and NA proteins are used to categorise influenza A viruses into 18 HA and 11 NA subtypes[153].

1.6.1 Influenza A Haemagglutinin

HA is a cylindrical type I transmembrane glycoprotein, which is 135 Å long and self-assembles into homo-trimers in the lipid membrane (Figure 15). There are approximately 500 HA spikes on the surface of each capsular-shaped viral envelope, with more present on filamentous viruses[153]. HA is comprised of two main domains – the globular head domain and the stem domain. The globular head includes the receptor-binding domain (RBD) (Figure 15), used to facilitate cell entry[154] and the stem domain incorporates the fusion region that encodes a hydrophobic fusion peptide necessary for membrane fusion (1.6.2). HA is initially translated into an inactive precursor protein (HA0) that must be glycosylated and proteolytically cleaved into its polypeptides (HA1 and HA2) to become functional. These two subunits are covalently linked by disulphide bridges. Cleavage of HA0 is essential for membrane fusion and infectivity, as the N-terminus of HA2 functions as a fusion peptide[153].

The majority of influenza subtypes have a single arginine at the cleavage site of their haemagglutinin. This motif is cleaved at the cell surface by secreted enzymes such as tryptase Clara found in human airway epithelial cells. However some H5 and H7 subtypes contain a multi-basic cleavage site allowing cleavage to be

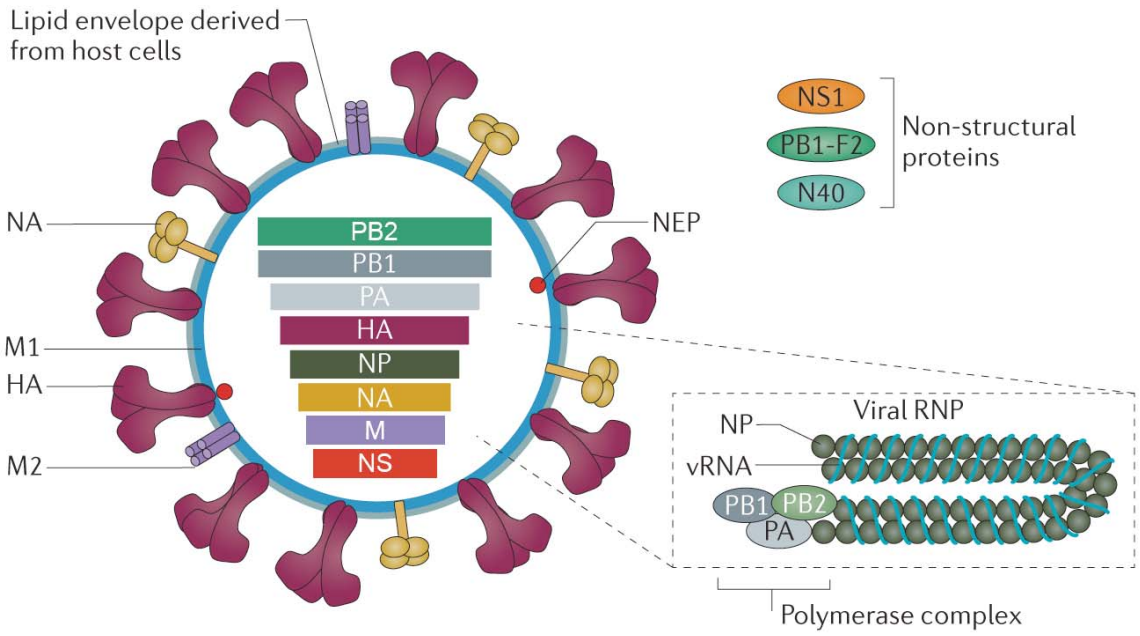


Figure 14: Components of the influenza A virion

The IAV genome is made up of eight single-stranded RNAs that encode up to 13 proteins. PB1, PB2, and PA make up the RNA polymerase complex, and the vRNA is wrapped around nucleoprotein (NP). Together, this RNA-protein complex is known as vRNP. The surface-expressed proteins include the sialic-acid-binding protein haemagglutinin (HA), and the sialic-acid-cleaving enzyme neuraminidase (NA), as well as the ion channel M2. Pro-apoptotic protein PB1-F2 and non-structural protein 1 (NS1) are both involved in reducing the host antiviral response. The matrix protein 1 (M1) makes up the capsid just below the lipid-membrane. Additional proteins include the nuclear export protein (NEP), and PB1-N40, which is currently of unknown function.

Adapted from Medina and Garcia-Sastre (2011)[155].

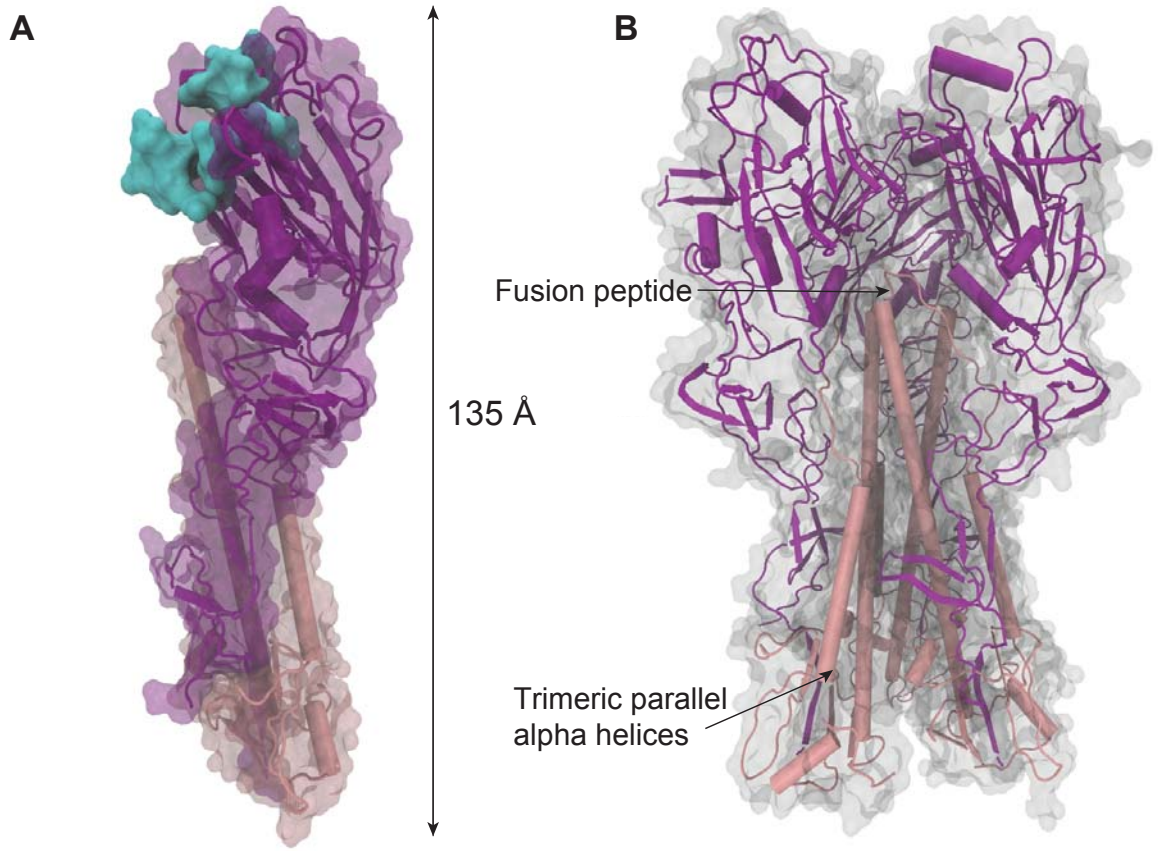


Figure 15: Three-dimensional structure of haemagglutinin

A single monomer (A) and a homo-trimer (B) of influenza haemagglutinin is shown. HA1 is shown in purple and HA2 in peach. The receptor binding domain (RBD) is shown in cyan. The protein backbone is shown by ribbon representation, with alpha-helices shown as rods, beta-sheets shown as arrows and beta-turns shown as tubes. The solvent-accessible area is shown as a transparent surface representation. The image was created in Visual Molecular Dynamics using the Protein Data Bank code 1RUZ.

mediated in the trans Golgi network by ubiquitous proteases such as furin. As cleavage occurs readily, H5 and H7 viruses cause systemic infections in the host and are therefore considered 'highly pathogenic'[153]. Cleaved HA at neutral pH is more stable than HA0, which is reflected in an increase in denaturation temperature from 50 °C to 62 °C[153]. However, inhibitors of membrane fusion such as *Tert*-butylhydroquinone (TBHQ), stabilise the neutral pH conformation of cleaved HA[156]. TBHQ was shown to bind at the interface of two HA monomers, cross-linking the trimer through three identical interactions, thus reducing infectivity.

Viral fusion proteins can be categorised into one of three types according to the tertiary structure of the 'post-fusion' peptide. HA belongs to the class I fusion proteins, which are characterised by a repeated motif of seven non-polar amino acids. This repeated pattern results in parallel trimeric α-helices that make up a rod-shaped molecule (Figure 15)[129].

Of the two external influenza proteins, HA is the primary target recognised by neutralising antibodies in the host[157]. These proteins are heavily N-glycosylated whilst in the endoplasmic reticulum, which is thought to mask antigenic sites, protecting the virus from antibody recognition[157]. An increase in the number of glycosylation sites in the globular head is associated with human seasonal IAV as opposed to avian IAV.

1.6.2 Mechanism of Cell Entry

HA binds to sialic acids linked to galactose by α2-3 or α2-6 linkages on the surface of host cells via the RBD of HA1. The binding site of H1 IAVs relies upon residues 190-198, 135-138, and 221-228 that produce three secondary structures (two loops and a helix) forming a binding pocket[154] (Figure 15). Mutation of residues in this pocket revealed that only three amino acids were essential for sialic acid binding: Y98F, H183F and L194A[158]. Y98 is known to be involved in hydrogen bonding to the 8-hydroxyl group of sialic acids. H183 forms a hydrogen bond with Y98 and L194 forms a non-polar contact to N-acetyl methyl group. In addition, substitution of W153A prevented cell surface expression of HA[159].

Binding of HA to sialic acids is thought to initiate *de novo* formation of clathrin-coated pits (CCPs) at the site of binding, resulting in invagination of the membrane

and the formation of clathrin-coated vesicles (CCVs) facilitating endocytosis of the virus particle[160]. Using electron microscopy, these virus particles have also been detected inside smooth, uncoated surface invaginations suggesting that influenza can also exploit clathrin-independent endocytic pathways[161].

The CCV is trafficked along the endocytic pathway, fusing with late endosomes that have an acidic environment (pH 5-6). Matrix protein 2 (M2) is an ion channel in the viral envelope, allowing protons to pass through the viral envelope into the virus interior. In addition, exposure to low pH causes conformational changes in HA1, exposing HA2 to water, which causes the N-terminal region of the domain to rearrange like opening a flick knife. The fusion peptide is now distal to the viral membrane and can insert into the endosomal membrane avoiding the hydrophilic environment[162] (Figure 16). The low pH environment of the endosome causes the fusion peptide to undergo an additional structural rearrangement, which is more energetically favourable; the denaturation temperature increases from 62 °C to 90 °C at fusion pH. This rearrangement leads to an intermediate – the pre-hairpin – which brings the viral and endosomal membranes into close proximity[163] (Figure 16). Ca^{2+} binding to negatively charged lipids allows the two lipid-bilayers to approach one another by overcoming the hydrophobicity between the two layers. The outer leaflet of each bi-layer touch and lipid mixing occurs, creating a lipid stalk. Whilst the internal contents remain separated this process is known as hemifusion[164]. It is estimated that 4–6 HA molecules are required for membrane fusion, and that the proteins form a ring at the edge of the eventual fusion pore. The hemifusion diaphragm then grows, increasing the tension between the two bi-layers until it is more energetically favourable for the two outer leaflets to also fuse, creating a fusion pore.

Within the virus interior, the vRNP is tethered to another viral protein, matrix protein 1 (M1). The change in pH (allowed by acidification of the endosome and M2) causes M1 to dissociate from vRNPs. Once the fusion pore is formed, the vRNP is released into the cytoplasm of the host cell. The four proteins that compose the vRNP contain nuclear localisation signals that target the complex to the host cell nucleus[165]. The viral complexes are actively transported through the nuclear pores by importins. Once inside the nucleus, the negative-sense RNA is transcribed into mRNA ready for translation. The vRNA-dependent RNA polymerase (RdRp) uses a

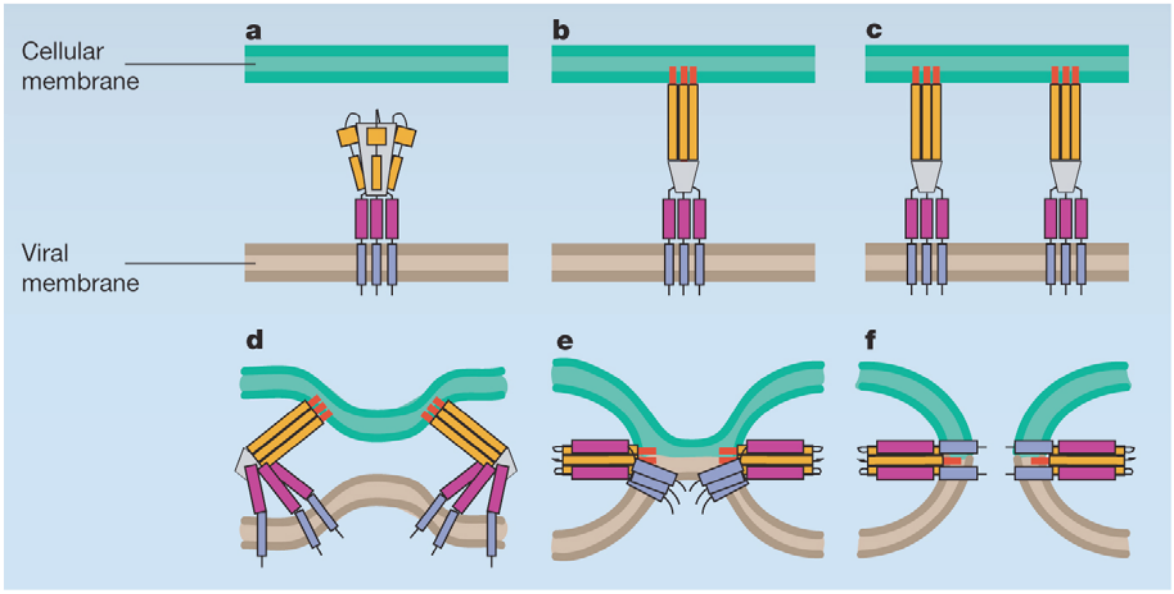


Figure 16: Membrane Fusion by Class I fusion peptides

HA changes conformation during pH changes. Helical domain A is shown in orange, helical domain B in pink, and the transmembrane domain in purple. At neutral pH (a) the protein is in a stable conformation. After exposure to low pH in the endosome the globular head of HA dissociates allowing refolding and exposure of a hydrophobic fusion peptide (HA2), which is inserted into the cellular membrane (b). Several trimers associate together (c) and the proteins re-fold (d). The energy released causes the membranes to bend towards one another (e). Hemifusion (fusion of the outer leaflets) occurs first, followed by full fusion and mixing of the content (f).

Adapted from Jardetzky *et al.* (2004)[166]

‘cap-snatching’ method to initiate transcription of the vRNA; the 5’ cap from cellular mRNAs is cleaved and used to prime viral mRNA synthesis[167]. The mRNAs are exported from the nucleus and translated by cellular ribosomes attached the to ER (HA, NA and M2) or free ribosomes (all other proteins). HA, NA and M2 have cleavable signal peptides used to facilitate transport to the cell surface and are inserted into the host membrane.

The IAV genome is also replicated using RdRp, and vRNPs leave the nucleus via the CRM1-dependent nuclear export pathway. The vRNPs assemble at the host membrane, where the membrane proteins have been inserted, and bud from the apical side of polarised cells[162] to leave the cell as an enveloped virion.

1.6.3 Influenza Infections

Seasonal IAV has an incubation period of approximately 1-2.4 days[168] post-exposure, and virus shedding normally begins one day before the onset of symptoms. Symptoms can include a fever, muscle pain, and a cough lasting 5-12 days. However, some studies suggest that about a third of seasonal influenza infections are asymptomatic[169]. Conversely, there are several established ‘at-risk’ groups for influenza infection, whose symptoms may require hospitalisation and the administration of antiviral drugs[170]. People in this group include pregnant women, the obese, those with cardiovascular disease, and those with compromised lung function.

Unlike influenza B and C, IAV infects many avian species, as well as humans and several other mammals, including pigs and horses. Widespread annual epidemics and strong host selection pressure result in rapid genome change, with an average of 2.6×10^{-3} nucleotide substitutions per site per year[171]. This is faster than several other RNA viruses, including dengue virus, HCV and Japanese encephalitis virus[172]. The influenza genome is approximately 13.4 kb, therefore this mutation rate equates to an average of 34 nucleotide substitutions per genome per year. This process is also known as genetic drift and allows IAV to evade the human immune system, changing epitopes that are recognised by specific antibodies, in a process called ‘antigenic drift’. Specifically Koel *et al.* found that the major antigenic changes in IAV H3N2 between 1968 and 2003 could be caused by a single amino acid substitution at one of seven positions adjacent to the receptor binding sites of HA[173].

Co-infection of a cell by genotypically-distinct virions can result in the reassortment of the two genomes, producing a mosaic virus with segments from each parent virus. If this reassortment results in a phenotypic change in the virus, it is known as 'antigenic shift'. This is an important evolutionary process that can result in an altered host range, or an increase in virulence. An interspecies transmission into an immunologically-naïve population can result in a pandemic outbreak, such as occurred in the 2009 H1N1 pandemic[174].

1.6.4 The Influenza A 2009 Pandemic

There have been four influenza A pandemics in the last century: H1N1 "Spanish 'flu" (1918), H2N2 "Asian 'flu" (1957), H3N2 "Hong Kong 'flu" (1968), and H1N1 "swine 'flu" (2009). The pandemics of 1918, 1957, and 1968 are all thought to be caused by reassortment of avian viruses with those of circulating human viruses[175,176]. The most recent pandemic virus emerged in 2009 from swine in Mexico[177], hence its colloquial name of "swine 'flu". This A/H1N1/09 virus was unusual because it primarily infected otherwise-healthy young-to-middle-aged adults. The virus had HA and NA segments from swine-adapted H1N1 viruses, and a 'triple reassortant internal gene cassette' whose segments were derived from avian, swine, and human-seasonal influenza viruses (Figure 17). Although seasonal H1N1 had been circulating in humans since the 1970s, the HA from classical swine was divergent enough to not elicit an effective cross-reactive immune response. Moreover, mutations in the nucleoprotein (NP) of pandemic A/H1N1/09 (Asp53, Ile/Val101, and Val313) were associated with evasion of human MxA[178].

The first cases in humans occurred in April 2009 and, as with the previous pandemics in the late 20th century, quickly spread around the world due to global travel[179]. However, the swine-adapted NS1 was inefficient at binding to CPSF30[180]. Thus it is possible that the replication and transmission rate in humans was attenuated, and therefore the severity of the A/H1N1/09 was altered compared to the 1918 H1N1 pandemic. Although the statistics varied depending on country of origin, in general older adults fared relatively well, with children and younger adults being worse affected[181]; the median patient age ranged from seven years in Japan to 38 years in Spain[182]. The total number of pandemic H1N1/09 influenza-related deaths worldwide was broadly estimated between 105,700 to 395,600 deaths[183], which is

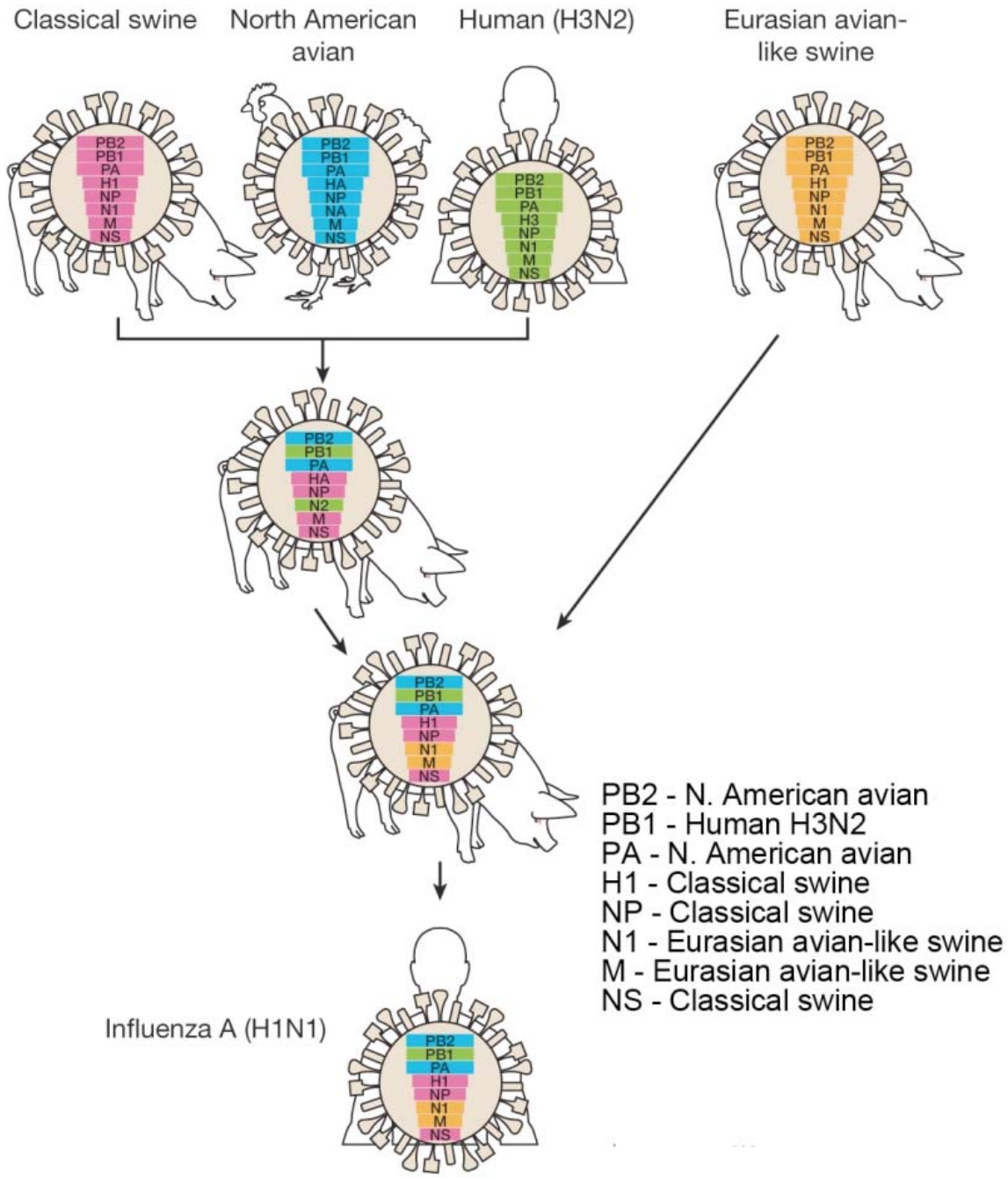


Figure 17: The genesis of pandemic A/H1N1/09 influenza

A triple reassortant virus with segments from viruses commonly circulating in pigs, birds and humans has existed in swine for many years. A separate reassortment occurred in pigs, resulting in the triple-reassortant virus acquiring the neuraminidase (N) and matrix protein (M) segments of a Eurasian avian-like swine virus.

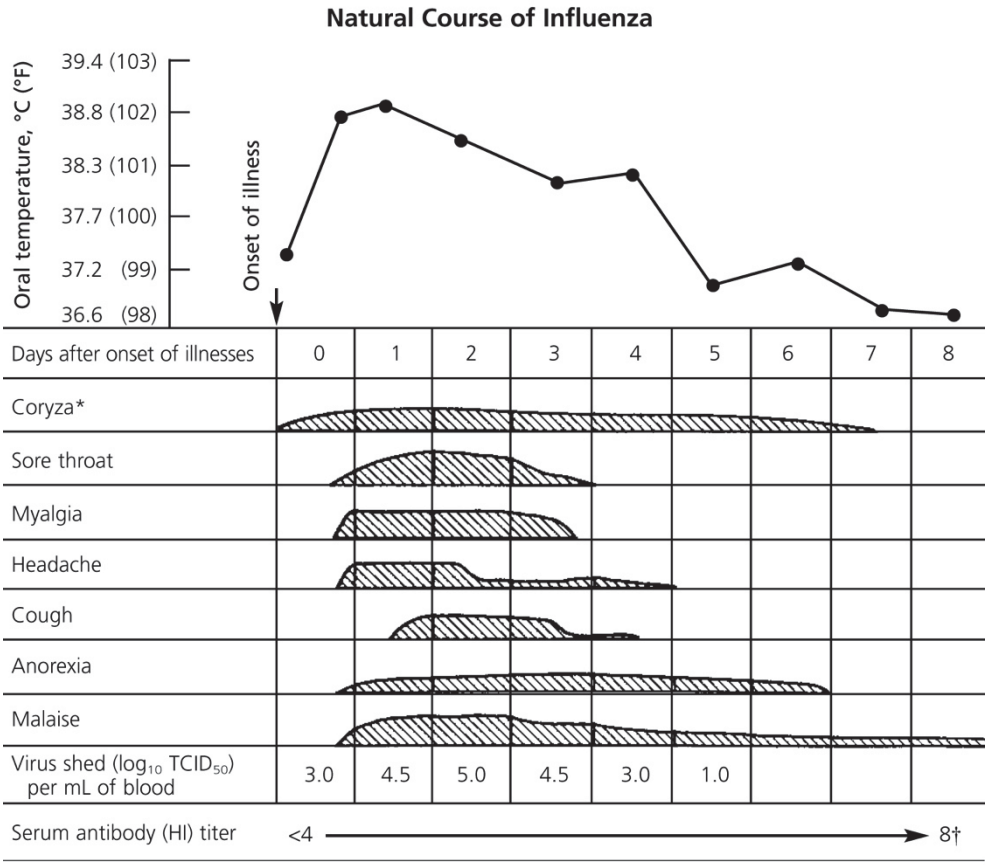
Adapted from Neumann *et al.* (2009)[174]

similar to the number of deaths in a relatively mild year of seasonal influenza[181]. Nonetheless, because of the proportionately higher mortality among children, the severity in terms of years of potential life lost (YPLL) was greater than in a typical influenza year[184].

1.6.5 Clinical Signs of Influenza Virus Infection in Humans

The annual number of cases of mild IAV infections is difficult to determine, since patients do not present to healthcare systems for mild disease[185]. These mild cases are often caused by infection of contemporaneous strains that have been circulating in humans for a while. Furthermore, respiratory symptoms are non-specific and of those patients that do present at hospital, few are routinely investigated for diagnostic evidence of influenza infection[185]. Symptoms of an uncomplicated, mild respiratory viral infection include a cough, sore throat, fever, malaise, a runny rose, and a headache[168], and these symptoms can present for between 3 and 7 days (Figure 18). Some strains of IAV have the capacity to cause severe symptoms in people, either because of the virulence of the virus itself or an exacerbated host immune response to the virus. For example, avian H5N1 viruses are known to penetrate deeper into the lung and cause more severe alveolar damage, compared to other IAV subtypes that replicate in the upper respiratory tract[186]. Moreover, proinflammatory cytokines such as TNFα and IL-6 have been shown to be at a higher concentration in people with a high-pathogenicity infection than in a low-pathogenicity infection[187]. These proinflammtory cytokines cause greater infiltration by macrophages and neutrophils into the alveoli, resulting in congestion and acute lung injury[188]. A prolonged high fever, viral pneumonia, and secondary bacterial infections during IAV infection can all result in hospitalisation. Hypoxemia, defined as an oxygen saturation less than 90 %[189], or abnormalities on a chest x-ray, such as regions of opacity, can lead to intubation in severe cases of influenza.

A global meta-analysis carried out by the World Health Organisation showed that a third of people with a severe A/H1N1/09 infection had a pre-existing chronic illness[182]. This proportion increased to 52.3 % of people admitted to intensive care, and 61.8 % of fatal cases. Cardiac disease, chronic respiratory disease, and diabetes were highlighted as high risk-factors[182]. Pregnant women were less likely than non-pregnant women to have respiratory distress on admission during the 2009 H1N1



*—*Coryza is an acute inflammatory condition of the nasal mucous membranes with a profuse discharge from the nose.*
†—*Serum antibody titer was 64 at day 21.*

Figure 18: Clinical symptoms and disease progression with uncomplicated IAV infection

The symptoms of a mild respiratory infection include a sore throat, headache, cough, and malaise, which can present for between 3 and 7 days after the onset of illness.

Adapted from Montalto *et al.* (2003)[190].

pandemic, but severe outcomes were equally likely in both groups[191]. Further rare complications can include neurological disorders such as encephalopathy[192] or encephalitis. The case fatality rate (CFR) of A/H1N1/09 was calculated to be 0.4 % at the end of April 2009[193], although this was difficult to assess because the total number of infections was an estimate. The 2009 H1N1 pandemic provided the opportunity to study IAV infection in a 'naïve' population; the majority of the population did not have immunity to this antigenically-shifted reassortant virus because the external glycoproteins had originated in a swine-adapted virus. Therefore, immunological memory can be eliminated as a factor affecting the variation in symptoms across the population. Sridhar *et al.* also utilised these conditions and showed that $CD8^{+}IFN\text{-}\gamma^{+}IL\text{-}2^{-}$ cross-reactive T cells were associated with a decreased severity in symptoms associated with A/H1N1/09 infection[194].

Second-generation sequencing technologies for viral genomes were also being developed around this period[177]. This allowed the sequencing and assembly of full viral genomes from infected patients to be matched with detailed patient information. This technique gave a more informed picture of the outcome of host-virus interactions.

1.7 Viral Zoonoses

Zoonotic infections are those that can be transmitted in either direction between vertebrate animals and humans, via direct contact or an invertebrate vector. Animals, both wild and domesticated, are therefore an important reservoir for viruses that could have the potential to infect humans. For instance, wild ducks and geese are the natural host of IAV, but zoonotic events can occur between them and other birds or mammals they have contact with, such as humans, chickens or pigs (Figure 19). Other species are also susceptible to IAV infections, but contact between them is more limited. As described earlier (1.6), the HA protein of IAV binds to two types of sialic acid on host cell membranes, which thereby limits the virus' host range. Humans predominately have sialic acids linked to galactose by α2-6 linkages in the upper respiratory tract, birds use α2-3 linkages, and pigs have both receptor types[174]. Humans do also express α2-3 linkages in the lower respiratory tract, but viruses that use these receptors are less likely to be spread because they are replicating deeper

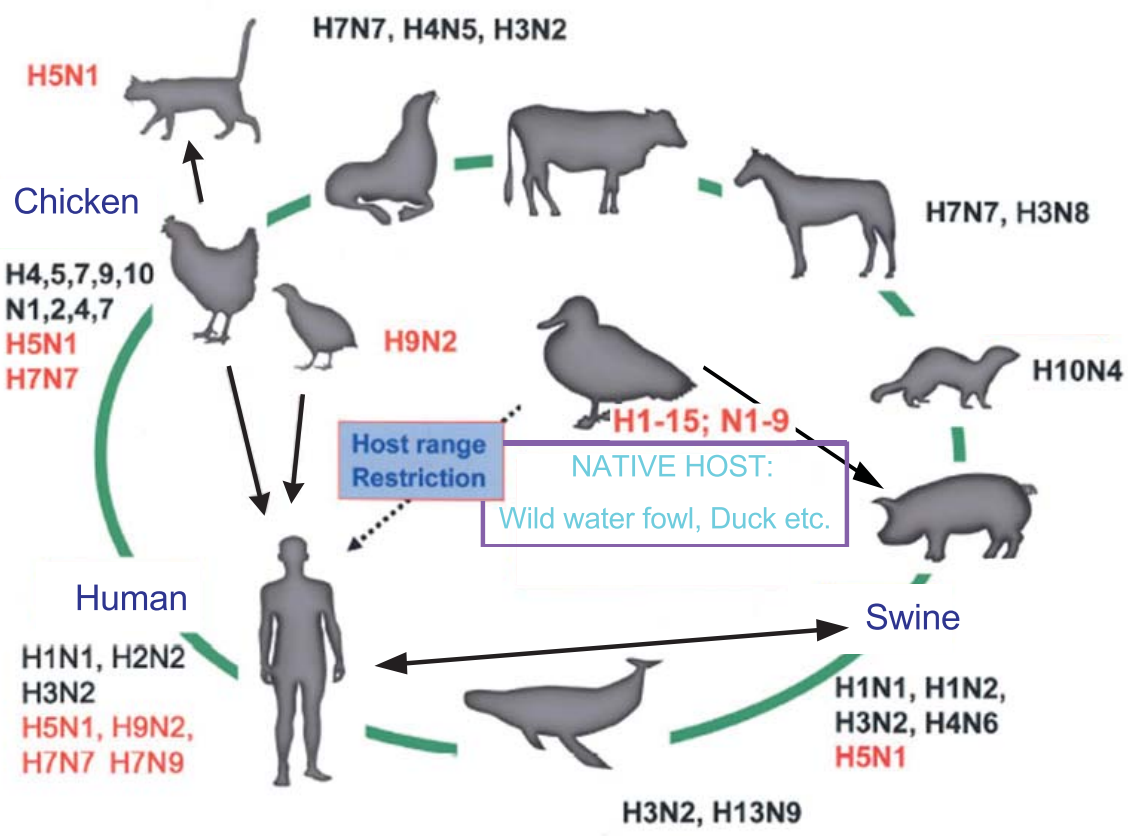


Figure 19: Transmission of influenza A between different animal species

Although wild fowl are the natural hosts of IAV, particular subtypes are able to infect a number of other birds and mammals. Zoonotic events can lead to the transfer of a virus from one species to another.

in the lung. Therefore bird-adapted viruses, such as avian H5N1, would need to evolve the ability to recognise α2-6 sialic acids in order to spread efficiently between people. Since pigs have both receptors on their cell membranes, they are an ideal mixing vessel in which the virus can adapt and evolve. Pigs co-infected with two viruses containing a human-adapted and avian-adapted HA segment can allow reassortment of the segments, producing an avian-adapted virus that is able to bind to α2-6 sialic acids.

This receptor specificity means that intra-human transmission is uncommon after zoonotic events occur. As of January 2014, H5N1 has infected 650 people worldwide (386 deaths), but most of the patients had known contact with live or dead poultry[195]. Some limited clusters of suspected human-to-human transmission were reported in China and Indonesia, but these were confined to familial cases and were not sustained[196,197]. Several groups embarked on gain-of-function experiments on H5N1 in ferrets to determine the likelihood of H5N1 adapting to aerosol transmission in mammals[198,199]. This identified four mutations that allowed efficient aerosol transmission in ferrets, but at a significant cost in fitness to the virus – none of the ferrets died after airborne infection with the mutant A/H5N1 viruses. This suggests that although H5N1 has a high mortality rate in humans, a severe pandemic is currently unlikely because sustained human-to-human transmission has not been detected, and mutations that facilitate aerosol transmission decrease the lethality of the virus.

However, if a virus can evolve and adapt in the human host, these zoonotic events can lead to pandemics, as was the case for A/H1N1/09. Several other factors can either reduce or increase the likelihood of a pandemic occurring. Firstly, the original animal host may express a restriction factor that keeps viral replication to a low level, or isolated to a single organ, for instance TRIM5α in Old World Monkeys prevents replication of HIV-1[41]. Secondly, the population is immunologically naïve after a zoonotic event, as was the case when SIV jumped from monkeys into humans. In these cases the host cannot rely on the support of the adaptive immune response, increasing the time that the virus is shed for.

1.7.1 Influenza Circulating in Avian Species

Waterbirds of the orders Anseriformes and Charadriiformes are the natural hosts of IAVs (Figure 19). Infected wildfowl display very mild or sub-clinical intestinal tract infections. These low-pathogenic avian influenza viruses (LPAIV) can be transmitted between birds by the faeco-oral route, facilitated by the environmental reservoirs the birds gather at, such as the surface water of lakes[200]. LPAIV has been known to persist in wild birds for several months under these conditions. LPAIV can be transmitted to poultry (chickens and turkeys), in which it also causes mild respiratory tract infections[201]. However, once established as an infection in poultry LPAIVs can evolve into viruses that cause significant systemic infections, characterised as highly-pathogenic avian influenza viruses (HPAIV)[202]. At a molecular level there are significant differences in the cleavage sites of the HA proteins of LPAIV and HPAIV. The former have a single arginine at the cleavage site, which is activated by trypsin-like proteases found in the enteric or respiratory epithelia. Whereas HPAIV usually exhibit a multibasic cleavage site that is activated by furin and related proteases found systemically, allowing the rapid dissemination of the virus throughout the organism, rather than it being isolated to one organ[203]. Consequently, the spread of HPAIV is somewhat different – transmission can occur via the faeco-oral or respiratory routes in poultry. However transmission of HPAIV from poultry to other species is uncommon and sustained transmission in a different host is very rare[202], as replication occurs in the lower respiratory tract of poultry reducing the chances that the virus will be spread by fomites.

1.7.2 Influenza Infections in Domesticated Poultry

In the last decade, the frequency of detected HPAIV outbreaks in poultry has increased, with 12 outbreaks occurring worldwide between 1994 and 2005[204]. For each of the HPAIV epidemics in Europe since 1997, a closely-related LPAIV virus was detected in mallards[204]. This suggests a continuous spill-over of virus from wildfowl into domesticated birds, and monitoring circulating lineages in wild birds, particularly ducks, could provide an opportunity for pandemic-preparedness.

The clinical signs of avian IAV infection depend not only on the strain of virus but also on the age of the birds and husbandry practices *i.e.* how closely confined the birds are. Clinical signs in poultry can include ruffled feathers, soft-shelled eggs, a drop in

egg production, and discolouring of the wattles and comb. In severe cases, birds can die rapidly without showing any previous sign of infection.

Domesticated chickens seem to be more susceptible to severe IAV infections than wildfowl, which could be due to selective breeding. Chickens have been bred for particular physical qualities for over 8000 years[205], such as feather colour, size, or egg-laying rates. This intensive selective breeding could also have led to a number of deletions in immunity genes[206] that protect wildfowl against IAV infection, including Mx and RIG-I.

1.7.2.1 Chicken Mx Proteins Are Not Anti-Viral

The Mx proteins have been found in all vertebrate species investigated and usually confer a broad anti-viral function (see 1.3.1.1). However, the existence of an anti-viral Mx protein in chickens has been contentious – several groups have provided evidence for this, but many others have been unable to replicate the results.

The Mx protein from the White Leghorn chicken strain was initially investigated in 1995, and found to lack anti-viral activity[207]. Analysis of polymorphisms in this gene revealed that the *Mx* alleles of other chicken breeds did confer anti-viral activity against some viruses[208]. Specifically, the presence of an asparagine or a serine at residue 631 increased resistance to IAV and vesicular stomatitis virus (VSV), respectively, *in vitro*. However, this result was not replicated in chicken challenge studies[209] or 293-T cell culture models. Relocalisation of the mutant proteins (Δ631N or Δ631S) into the nucleus of 293-T cells also failed to restore any antiviral activity to chicken Mx[210]. However, subsequent experiments over-expressing Mx Δ631N or Δ631S in Cos-I cells, Chicken Embryonic Fibroblasts (CEFs), or NIH 3T3 cells showed that the Δ631N chicken Mx variant did provide resistance to Newcastle disease virus (NDV) and VSV[211]. Conversely, Schusser *et al.*[212] showed that CEFs stimulated with IFN before an influenza infection had a significantly-reduced viral load compared to control cells, regardless of whether or not an siRNA specific to Mx was used. This suggests Mx is not an essential component of the type I IFN response, and that some other IFN-induced factors must contribute to the inhibition of IAV in chicken cells. The authors also found that both mutant isoforms of chicken Mx appeared to lack GTPase activity. In conclusion, the role of chicken Mx has been controversial for nearly two decades, with different groups finding different results.

The reasons for this are unclear, however differences in chicken breeds, cell types, assays, and infection models may all contribute to these inconsistencies.

1.7.2.2 Chickens Lack RIG-I Proteins

Retinoic-acid-inducible gene 1 (*RIG-I*) is one of the RIG-I-like family of viral cytoplasmic dsRNA sensors, which plays a major role in host protection against influenza infection[213]. Furthermore, mice lacking the *RIG-I* orthologue are more susceptible to RNA virus infection[214].

RIG-I has been identified in ducks, which has 53 % sequence identity to the human orthologue and 78 % identity to the zebra finch orthologue[215]. However using these sequences in BLAST searches did not reveal any orthologous sequences in chickens. Chickens have been shown to express melanoma-differentiation-associated gene 5 (*MDA5*), another potent dsRNA sensor, which thus far has been thought to compensate for the lack of *RIG-I*[216]. However, other studies have shown that MDA5 and RIG-I recognise different types of dsRNAs: MDA5 detects poly I:C (a synthetic dsRNA analogue) whereas RIG-I detects *in vitro* transcribed dsRNAs[214]. Furthermore, RIG-I is essential for the production of IFNs in response to RNA viruses, including paramyxoviruses, IAV and Japanese encephalitis virus, whereas MDA5 is critical for picornavirus detection[214]. Whether all chicken breeds lack *RIG-I*, or if more extensive genome analysis will reveal *RIG-I* family members, is not known. However, as chickens are not hyper-susceptible to all RNA virus infections, most aspects of intrinsic antiviral defence should be functional in chickens.

1.8 Host Genetics

The Human Genome project was an international effort that aimed to sequence and map all of the genes in the genome. Subsequent ventures, including the HapMap project, initially catalogued all the genetic similarities and differences between 270 people from diverse parts of the world[217]. This project, and others, have identified over 15 million common variants, most of which are SNPs[218]. Now that the HapMap project has entered the third phase, the goal is to increase the numbers of individuals and populations studied. These datasets have provided insight into the occurrences

of linkage disequilibrium (LD), the prevalence of structural variation, and genes that are under selection.

1.8.1 How Host Genetics Influences Human Infectious Disease

Clinical heterogeneity in response to infection has been noted since Charles Nicolle's discovery of asymptomatic infections in 1933[219], although the reasons behind this were unclear. Since then, twin and adoptee studies have been some of the first to suggest a genetic component may influence susceptibility to infectious disease, including cases of tuberculosis, leprosy, and poliomyelitis[220]. More recently, genome wide association studies (GWAS) have been used to study diseases and traits with genetic components across the whole genome, rather than confined to a limited number of candidate genes. The major strength of this technique lies in not needing prior knowledge of which regions of the genome may be implicated, and is therefore relatively unbiased.

1.8.1.1 Genetic Susceptibility During HIV-1 Infection

The first GWAS of an infectious disease was carried out on patients with HIV-1. The authors compared the genomes of patients with variable viral load 'set points' (stable viral load during the asymptomatic period of infection), which is known to be predictive for disease progression[221]. An association was found between SNPs in the *human leukocyte antigen* (*HLA*) *-B* and *-C* genes. These genes are located in the most diverse part of the human genome, on chromosome 6. The most likely causal SNPs were deemed to be HLA-B*5701 allele and the HLA-C promoter SNP-35C. This analysis was repeated in patients of African descent, and HLA-B*5701 was also found to determine viral load ($p=5.6x10^{-10}$). From these studies investigators estimated that 22 % of the variability in HIV-1 load could be due to human genetic variation. HLA-B*5701 was also found to be associated with 'long-term non-progressors', whereas HLA-B35 is associated with faster progression to AIDS[222].

CCR5 is also an important co-receptor for HIV-1 cell entry, and deletion of this gene was discovered by a candidate gene approach[223]. One GWAS confirmed a known association between a 32 base pair deletion in CCR5 (CCR5Δ32) and protection against HIV-1 infection[224]. Furthermore, heterozygotes for this deletion express half

the normal amount of this receptor and thus are often delayed in progression to AIDS after infection[225].

As well as susceptibility to HIV-1 and progression to AIDS, associations have also been found between HLA-B*5701 and tolerance to an HIV-1 drug, Abacavir[226]. Approximately 5 % of patients treated with this nucleoside analogue reverse-transcriptase inhibitor suffer from multisystem symptoms, including fever, rash, vomiting, diarrhoea, and in some cases oedema and renal failure that can result in death. In a cohort of 18 hypersensitive individuals, 14 carried the HLA-B*5701 SNP.

1.8.1.2 Hepatitis C and IL28B

HCV is a chronic viral infection that can lead to liver cirrhosis and hepatocellular carcinoma, although some infected individuals spontaneously clear the virus. Polymorphisms in *IL28B,* which encodes IFN λ3, are associated with both spontaneous clearance of HCV and response to treatment[21]. In particular rs12979860, located 3 kb 3' to the start codon of *IL28B*, was associated with both spontaneous clearance and better drug response. The protective C allele is most prevalent in people of Asian descent and least common in those of African descent[21].

1.8.1.3 Host Genetic Determinants on Susceptibility to Influenza

The WHO prioritised identifying the role of host genetic factors on the outcome of influenza infection in 2009. Since then several studies have investigated a range of host immune genes in both mice and humans (Table 2). Bottomly *et al.* used expression quantitative loci mapping to identify 21 genes in mice that were previously unknown to be important in influenza A pathogenesis[227]. Furthermore, Srivastava *et al*. compared influenza A PR8 (H1N1) infections in different strains of inbred mice. They showed that although both strains DBA/2J and C57BL/6J were deficient in Mx1, DBA/2J were highly susceptible to influenza infection; after a dose of $2x10^3$ focus forming units (FFU) of PR8 75 % of C57BL/6J mice survived after 14 days, compared to 100 % mortality in DBA/2J mice 7 days post-infection[228]. These strains vary at several positions in the genome, so further work must be done to identify the genomic regions responsible for poor control of influenza A.

In 2009, Zhang *et al.* proposed a list of 100 systems-based candidate genes for future study into the genetic influence on the outcome of IAV infection, based on

Table 2: Studies on host genetic susceptibility to influenza infection

Author (Year)	Study/Investigation	Main reported findings
Ferdinands, J. M. *et al.* 2011	SNPs in *TNF* and *MBL* genes in 105 children and young adults with fatal influenza compared with population controls	No differences in genotype or allele frequency between case and control groups. Fatal influenza cases with methicillin-resistant *Staphylococcus aureus* (MRSA) co-infection had a higher prevalence of a low-producing *MBL2* genotype compared with fatal cases without MRSA co-infection.
Chan, J. F. *et al.* 2011	IgHG2 and Fc gamma receptor IIa (FcYRIIa) genotype in 38 severe A/H1N1/09 cases	Severe A/H1N1/09 cases had lower levels of IgG2 than mild cases, but IgHG2 allotype was not associated with IgG2 levels and FcYRIIa genotype frequencies did not differ from population controls. The authors concluded that relative IgG2 suppression in this sample is probably the result of cytokine dysregulation rather than genetic factors.
Antonopoulou, A. *et al.* 2012	Analysis of tumour necrosis factor gene (*TNF*) in 109 A/H1N1/09 cases and 108 controls	The minor allele (A) at position 238 of *TNF* (SNP rs361525) was more common in cases (frequency=0.064) compared with controls (frequency=0.019; p=0.016). A diagnosis of pneumonia was more common in cases with a least one copy of the minor allele (7/13) compared with cases with no copies of the minor allele (20/96).
Boivin, G. A. *et al.* 2012	Response to infection with a mouse-adapted H3N2 in 29 recombinant congenic mouse strains	Genomic areas of interest identified by co-localisation of clinical quantitative trait loci (cQTL). The most significant loci identified were *Hc* on chromosome 2, and *Pla2 g7* and *Tnfrsf21* on chromosome 17.
Bottomly, D. *et al.* 2012	Animals with high and low response phenotypes following infection with H1N1 (A/PR/8/34) were identified	Twenty-one genes were identified that may be involved in genetic control of RNA expression at 4 days post-infection.
Keynan, Y. *et al.* 2010	CCR5Δ32 allele identified in 20 patients with severe pandemic (H1N1) 2009	The CCR5Δ32 was not found in 10 non-white cases, and was present in 5/9 white cases. The proportion of white cases with the CCR5Δ32 allele was higher than has been reported for healthy Caucasian controls.
Hidaka, F. *et al.* 2006	Toll-like receptor 3 recognises dsRNA	The F303S mutation of TLR3 was found to be associated with IAE, and caused decreased NF-kB and IFNb receptor functions *in vitro*.
Esposito, S. *et al.* 2012	Toll-like receptor 3 recognises dsRNA	SNP (rs5743313, genotype C/T) was found in all patients with pneumonia (18 cases) but in a significantly lower number of those with milder H1N1-induced disease (p<0.0001).
Zuniga, J. *et al.* 2012	RPAIN facilitates nuclear localisation of RPA C1QBP inhibits complement activation FCGR2A plays a role in phagocytosis and clearance of immune complexes	Four disease-outcome-associated SNPs were identified on chromosome 17 (RPAIN and C1QBP), chromosome 1 (FCGR2A), and chromosome 3 (unknown gene). C1QBP and GCGR2A play roles in the formation of immune complexes and complement activation, suggesting that the severe disease outcome of H1N1 infection may result from an enhanced host immune response.

Adapted from Horby *et al.* (2013)[229] and Lin *et al. (2013)*[230]

evidence in the published literature and localisation[231]. These include genes involved in preventing viral attachment (*MUC1*), those involved in endocytosis (*V-ATPase*), transcription (*POLR2A*), and translation (*EIF4G1*).

Zuniga *et al.* took a more practical gene-discovery approach by carrying out a case-control association study on 91 individuals with severe pneumonia as a result of A/H1N1/09 infection[232]. They found four risk SNPs that were significantly associated (p=0.0001) with severe symptoms: rs1801274 (Fc fragment of immunoglobulin G, low-affinity IIA, receptor [*FCGR2A*] gene); rs9856661 (gene unknown); rs8070740 (RPA interacting protein [*RPAIN*] gene); and rs3786054 (complement component 1, q subcomponent binding protein [*C1QBP*] gene). It is important to note that FCGR2A and C1QBP are both involved with the handling of immune complexes and complement activation. Furthermore a missense mutation (F303S) and a SNP in the intron of *TLR3* have been shown to be associated with influenza-associated encephalopathy[233] and severe pneumonia[234], respectively. A further two studies found an association between pneumonia and encephalopathy as a result of IAV and polymorphisms in the *carnitine palmitoyltransferase II* (*CPT2)* gene, involved in the oxidation of long chain fatty acids[230]. This is a mitochondrial gene and is therefore passed from mother to offspring. The authors found that the mutant enzyme was less active (intracellular ATP had 48-79 % activity compared to controls) and had weaker thermal stability.

All of these examples indicate that host genetic variants can be associated with susceptibility to infectious disease, progression through disease, and predict treatment outcomes. Thus analysis of genetic determinants of other infectious diseases is important for improving prevention and treatment of infectious disease.

1.9 Thesis Aims

This thesis aims to explore the function of IFITM3, a potent broad-acting restriction factor, in humans and in chickens (a species in which zoonotic events occur regularly). Specifically I will be investigating the following:

Chapter 1: Variants

What is the degree of allelic variation of *IFITM3* in people? Are any SNPs associated with the development of severe influenza and what effect on IFITM3 expression or function do these SNPs have?

Chapter 2: Importance in other species

Chickens are known to be deficient in a number of viral restriction factors, such as Mx and RIG-I. Do avian species such as chickens encode *IFITM3* orthologues and if so are they anti-viral?

Chapter 3: Mechanism

Which proteins does IFITM3 interact with? Does IFITM3 signal to the innate immune system?