

Accessory Gene Products of Influenza A Virus

Rute M. Pinto,¹ Samantha Lycett, Eleanor Gaunt, and Paul Digard

The Roslin Institute, University of Edinburgh, Midlothian EH25 9RG, United Kingdom

Correspondence: paul.digard@roslin.ed.ac.uk

Influenza A virus has long been known to encode 10 major polypeptides, produced, almost without exception, by every natural isolate of the virus. These polypeptides are expressed in readily detectable amounts during infection and are either fully essential or their loss severely attenuates virus replication. More recent work has shown that this core proteome is elaborated by expression of a suite of accessory gene products that tend to be expressed at lower levels through noncanonical transcriptional and/or translational events. Expression and activity of these accessory proteins varies between virus strains and is nonessential (sometimes inconsequential) for virus replication in cell culture, but in many cases has been shown to affect virulence and/or transmission in vivo. This review describes, when known, the expression mechanisms and functions of this influenza A virus accessory proteome and discusses its significance and evolution.

THE CORE IAV PROTEOME

In total, the influenza A virus (IAV) genome comprises ~13.6 kb of single-stranded, negative-sense RNA. This is split into eight segments (virion RNAs [vRNAs]) with lengths ranging from 0.89 kb to 2.3 kb, which together encode all the viral polypeptide products. The segmented genome facilitates virus evolution through reassortment. The greatest diversity of IAV strains are found in avian host species, but many mammalian species (e.g., humans, swine, horses) have their own reasonably stable IAV lineages, constantly refreshed by spillover from the avian reservoir (Yoon et al. 2014; Lycett et al. 2019). All IAV vRNAs have a generic structure of a long open reading frame (ORF) in antisense flanked both sides by short segment-unique un-

translated regions (UTRs) and then conserved sequences that form a promoter structure for the viral polymerase (Fig. 1). Genomic vRNAs are replicated via low-abundance replicative intermediates (complementary RNAs [cRNAs]) that are not capped or polyadenylated and, as far as is known, are not substrates for translation (Hay et al. 1977, 1982). Instead, viral gene expression is mediated by 5'-capped and polyadenylated messenger RNAs (mRNAs) that are an alternative form of positive-sense RNA transcribed from the vRNA molecules. Caps are derived by “cap snatching”—a process in which the 5'-cap structure and 10-14 downstream nucleotides of mRNA are cleaved from host cell pre-mRNAs by endonucleolytic cleavage and the snatched fragment used to prime viral transcription (Plotch et al. 1981). Polyadenylation is achieved



¹Present address: MRC-University of Glasgow Centre for Virus Research, Glasgow G61 1QH, United Kingdom

Editors: Gabriele Neumann and Yoshihiro Kawaoka

Additional Perspectives on Influenza: The Cutting Edge available at www.perspectivesinmedicine.org

Copyright © 2021 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a038380

Cite this article as *Cold Spring Harb Perspect Med* 2021;11:a038380

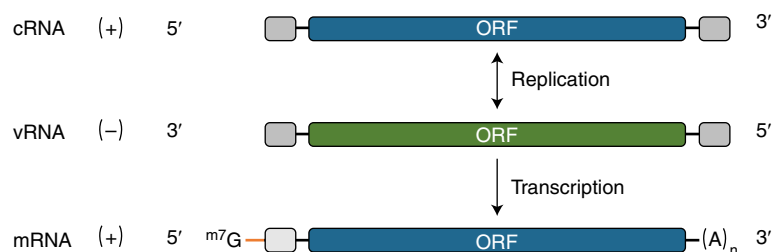


Figure 1. Schematic of the three classes of RNA produced by the influenza A virus (IAV). Colored boxes represent the major open reading frame (ORF) within a segment, gray boxes the segment terminal repeats that form a promoter structure, and lines the untranslated regions (UTRs). Note that mRNA has a 5'-extension derived from host pre-mRNA (orange line) and has a poly(A) tail in place of the 3'-promoter element. (mRNA) Messenger RNA, (cRNA) complementary RNA, (vRNA) virion RNA.

by the polymerase stuttering on a short pol(U) stretch near the 5'-end of the vRNA template (Robertson et al. 1981; Poon et al. 1999).

Each segment contains a principal coding region that is usually translated from the first AUG codon in the mRNA to produce the primary gene product of the segment. This leads to expression of the three elements of the heterotrimeric viral RNA-dependent RNA polymerase (PB2, PB1, and PA from segments 1, 2, and 3, respectively), the surface glycoproteins (HA from segment 4 and NA from segment 6), the single-strand RNA-binding nucleoprotein (NP from segment 5), the virion matrix protein M1, and the primary interferon antagonist NS1 from segments 7 and 8, respectively (Table 1; Fig. 2, dark blue bars). Six of these eight polypeptides are absolutely essential for virus replication, whereas the NA and NS1 proteins are dispensable in tissue culture conditions in which, respectively, neuraminidase activity is supplied exogenously or an intact interferon response is lacking (Liu and Air 1993; García-Sastre et al. 1998). Viruses encoding severely truncated but (with modern hindsight) still partially functional NS1 genes have been isolated (Norton et al. 1987), and there has been a more recent report of human clinical isolates lacking an NA gene (Moules et al. 2010); however, viruses such as these are the exception and the vast majority of naturally occurring IAV strains encode NA and NS1.

Two further viral polypeptides are considered to be part of the core IAV proteome. Unlike

many other RNA viruses, IAV RNA synthesis occurs in the host cell nucleus, giving access to the cellular mRNA splicing machinery. Transcripts from segments 7 and 8 are differentially spliced, giving rise to mRNAs encoding the M2 and NEP (NS2) proteins respectively (Lamb and Lai 1980; Inglis and Brown 1981). These polypeptides are also produced by canonical translation initiating at the first AUG codon in the mRNA, but out-of-frame splicing events direct translation shortly thereafter into an alternative reading frame downstream of the splice acceptor sites (Fig. 2). M2 and NEP are produced by every natural isolate of IAV, and NEP is considered to be essential for virus replication. Viruses lacking an intact M2 gene can be generated in the laboratory but are attenuated (Watanabe et al. 2001; Takeda et al. 2002) and, in some cases, complemented by expression of an alternative isoform of M2 that is considered (in more detail below) to be an accessory gene product (Wise et al. 2012).

These 10 proteins encoded by the 8 IAV segments (PB1, PB2, PA, HA, NP, NA, M1, M2, NS1, and NEP) are all highly expressed in infected cells and, in most cases, were identified as unique polypeptide species before the viral genome sequence became available (Skehel 1972). When the first genomic sequences for IAVs were published in the late 1970s/early 1980s (Sleigh et al. 1979), the coding origins of these polypeptides were obvious. These properties, along with the largely essential nature of the viral gene products, provides reasonable justification for defining them as the core IAV proteome.

Table 1. Influenza A virus (IAV) gene products, expression mechanisms, and functions

Segment	Segment length (nt)	Protein products	Expression mechanism	Length (aa)	Molecular weight (kDa)	Main function(s)
1	2341	PB2	Canonical translation initiation	759	85.7	Component of the heterotrimeric viral RdRp; binds mRNA cap structures to promote cap snatching
2	2341	PB2-S1	mRNA splicing	508	55	Inhibitor of the RIG-I signaling pathway?
		PB1	Canonical translation initiation	757	86.6	Component of the viral RdRp; nucleotide polymerase
3	2233	PB1-N40	Alternative AUG initiation (translation reinitiation)	718	82.4	Unknown; interacts with the viral polymerase
		PB1-F2	Alternative AUG initiation (leaky scanning)	87	10.5	IFN antagonist; pro-apoptotic activity
		PA	Canonical translation initiation	716	84.2	Component of the viral RdRp; viral mRNA endonuclease responsible for cap snatching
4	1778	PA-X	Ribosomal frameshift	252	29	Exhibits endonuclease activity; contributes to host cell shutoff
		PA-N155	Alternative AUG initiation	562	62	Unknown
		PA-N182	Alternative AUG initiation	535	60	Unknown
		HA	Canonical translation initiation	566	61.5	Homotrimeric surface glycoprotein; mediates receptor binding and virus entry
5	1565	NP	Canonical translation initiation	498	56.1	Encapsidation of vRNA (to form RNP complexes); required for transcription activity and vRNP nuclear import
6	1413	eNP	Upstream AUG	504	56.8	Virulence factor for H1N1 IAVs
		NA	Canonical translation initiation	454	50.1	Homotetrameric surface glycoprotein; exhibits sialidase activity allowing release of virus progeny from the cell surface
		NA43	Alternative AUG initiation (leaky scanning?)	440	48.6	Uncertain, plasma-membrane-resident NA variant

Continued

Table 1. *Continued*

Segment	Segment length (nt)	Protein products	Expression mechanism	Length (aa)	Molecular weight (kDa)	Main function(s)
7	1027	M1	Canonical translation initiation	252	27.8	Matrix protein; known to interact with RNPs and viral membrane proteins; involved in RNP nuclear export, virus assembly, and budding
		M2	mRNA splicing	97	11	Acts as proton channel; required during virus entry (uncoating) and for membrane scission at the budding stages of infection
8	890	M42	mRNA splicing and alternative AUG initiation	99	13	Variant of M2 with common functions
		NS1	Canonical translation initiation	230	26.8	Broad-spectrum IFN antagonist
		NEP/NS2	mRNA splicing	121	14.2	Involved in nuclear export of vRNPs
		NS3	mRNA splicing	187	21	Unknown
		NSP	unknown	167	18.1	Unknown; protein expression bioinformatically proposed, but yet not experimentally detected
		tNS1	Alternative AUG initiation	150/ 152	17	Involved in inhibition of IRF3

Segment and protein sizes based on A/Puerto Rico/8/1934 (H1N1).
 (RdRP) RNA-dependent PNA polymerase, (mRNA) messenger RNA, (IFN) interferon, (vRNA) virion RNA, (IAV) influenza A virus, (RNP) ribonucleoprotein.

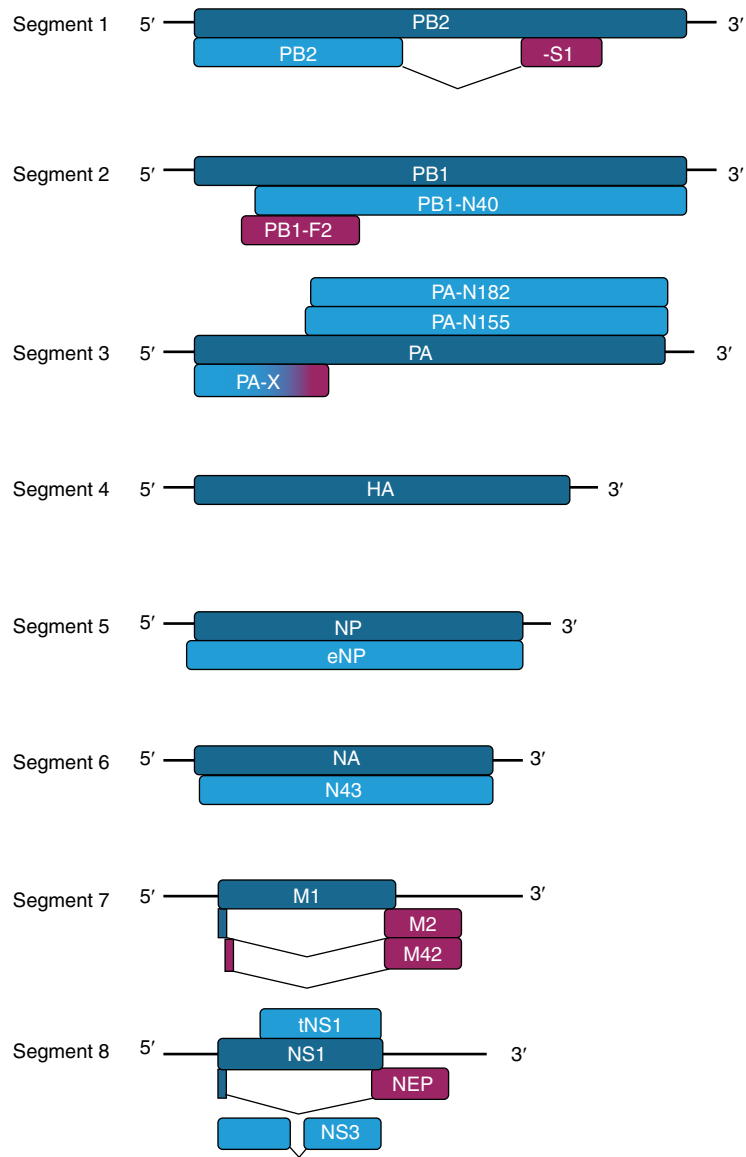


Figure 2. Schematic of the IAV coding strategy. mRNAs from the eight segments are symbolized by horizontal black lines. Coding regions (only approximately to scale) are represented by boxes with colors defining different reading frames (dark blue: primary products, frame 1; light blue: secondary products, frame 1; magenta: secondary products, frame 2). mRNA splicing in segments 1, 7, and 8 is denoted by deflected lines connecting coding regions.

THE IAV ACCESSORY PROTEOME

The last of the core polypeptides encoded by IAV (M2) was identified in 1981 (Lamb and Choppin 1981), and there things mostly rested for 20 years. However, since the turn of the cen-

tury, starting with the identification of the PB1-F2 polypeptide (Chen et al. 2001), it has become evident that the core proteome is further elaborated by the expression of many further minor polypeptide species. In contrast to the core gene products, this second class of products tends to

be expressed in low abundance and a virus strain-dependent manner, and, when expressed, are nonessential for virus replication. Nevertheless, in many cases, expression of these accessory proteins has been shown to affect virus pathogenicity in vivo, although not necessarily in an easily predictable fashion. These “accessory” proteins are produced by various mechanisms, including minor mRNA splicing events and/or examples of noncanonical translation, including leaky ribosomal scanning and ribosomal frameshift events. The following sections will describe (in segment order), these accessory proteins and, when known, their expression mechanisms and function in the IAV replication cycle.

SEGMENT 1: PB2-S1

Identified in 2015 (Yamayoshi et al. 2016), PB2-S1 (“PB2 Spliced-1”) expression results from a spliced mRNA from segment 1 in which the region corresponding to nucleotides 1513–1894 of the PB2 mRNA is excised so that translation of the mRNA produces the first 495 amino acids of PB2 and then continues in the +1 frame from the S1 ORF for a further 13 codons (Fig. 2). First identified in the highly laboratory-passaged A/Wilson-Smith Neurotropic/33 (WSN) H1N1 strain, the splice donor (SD) and splice acceptor (SA) sequences are conserved in the human H1N1 lineage descended from the 1918 pandemic, and expression of PB2-S1 has been demonstrated for a number of strains. However, the splice sites are not conserved in human H3N2 viruses or the more recent swine-derived human H1N1 viruses originating from the 2009 influenza pandemic (H1N1pdm09), and expression of PB2-S1 from segment 1 of these viruses was not detected (Yamayoshi et al. 2016). Localization of PB2-S1 is mainly mitochondrial and occurs via an amino-terminal mitochondrial localization signal shared with PB2 of human lineage viruses (Carr et al. 2006; Graef et al. 2010). Similarly to full-length PB2, PB2-S1 was shown to inhibit the retinoic acid-inducible gene I (RIG-I)-dependent/mitochondrial antiviral-signaling protein (MAVS) interferon signaling pathway (Yamayoshi et al. 2016). It also interacts with the

PB1 subunit of the viral polymerase, but negatively interferes with its activity. Nonetheless, abrogation of expression of PB2-S1 did not alter virus fitness in vitro or virulence in murine in vivo systems (Yamayoshi et al. 2016). The biological significance of this accessory protein thus remains unclear.

SEGMENT 2: PB1-F2

PB1-F2 is a small (typically 57–101 amino acids depending on virus strain) (Zell et al. 2007) polypeptide encoded by an ORF entirely contained within the PB1 coding sequence (Fig. 3A). Translation initiation occurs at the fourth AUG codon in segment 2 mRNA, accessed via leaky ribosomal scanning (Chen et al. 2001; Wise et al. 2009, 2011). This process is facilitated by the suboptimal Kozak (Kozak 1986) translational initiation context of the PB1 AUG codon but also potentially negatively regulated by sequence elements within the 3′-end of the PB1-F2 gene and further downstream in the PB1 gene (Buehler et al. 2013). There is also evidence that amino-terminally truncated PB1-F2 isoforms are expressed from one or more of AUGS 7–9 (Zamarin et al. 2006; Kamal et al. 2015), but how these start codons are accessed by ribosomes is unknown.

PB1-F2 was first identified as a short-half-life protein that interacted with mitochondria, postulated to induce apoptosis in monocytes (Chen et al. 2001). Further studies revealed that PB1-F2 translocated into mitochondria via Tom40 channels, culminating in a loss of mitochondrial membrane potential and apoptosis (Gibbs et al. 2003; Zamarin et al. 2005; Varga et al. 2012; Yoshizumi et al. 2014). Studies have also linked PB1-F2 to inhibition of cellular antiviral responses (potentially related to its mitochondrial function), through interactions with MAVS and the RIG-I and NF-κB pathways (Varga et al. 2011, 2012; Reis and McCauley 2013; Yoshizumi et al. 2014; Leymarie et al. 2017; James et al. 2019). Further postulated functions include pro-inflammatory activity, viroporin function, amyloid behavior, and regulatory interactions with the viral polymerase complex—in short, a plethora of possible roles

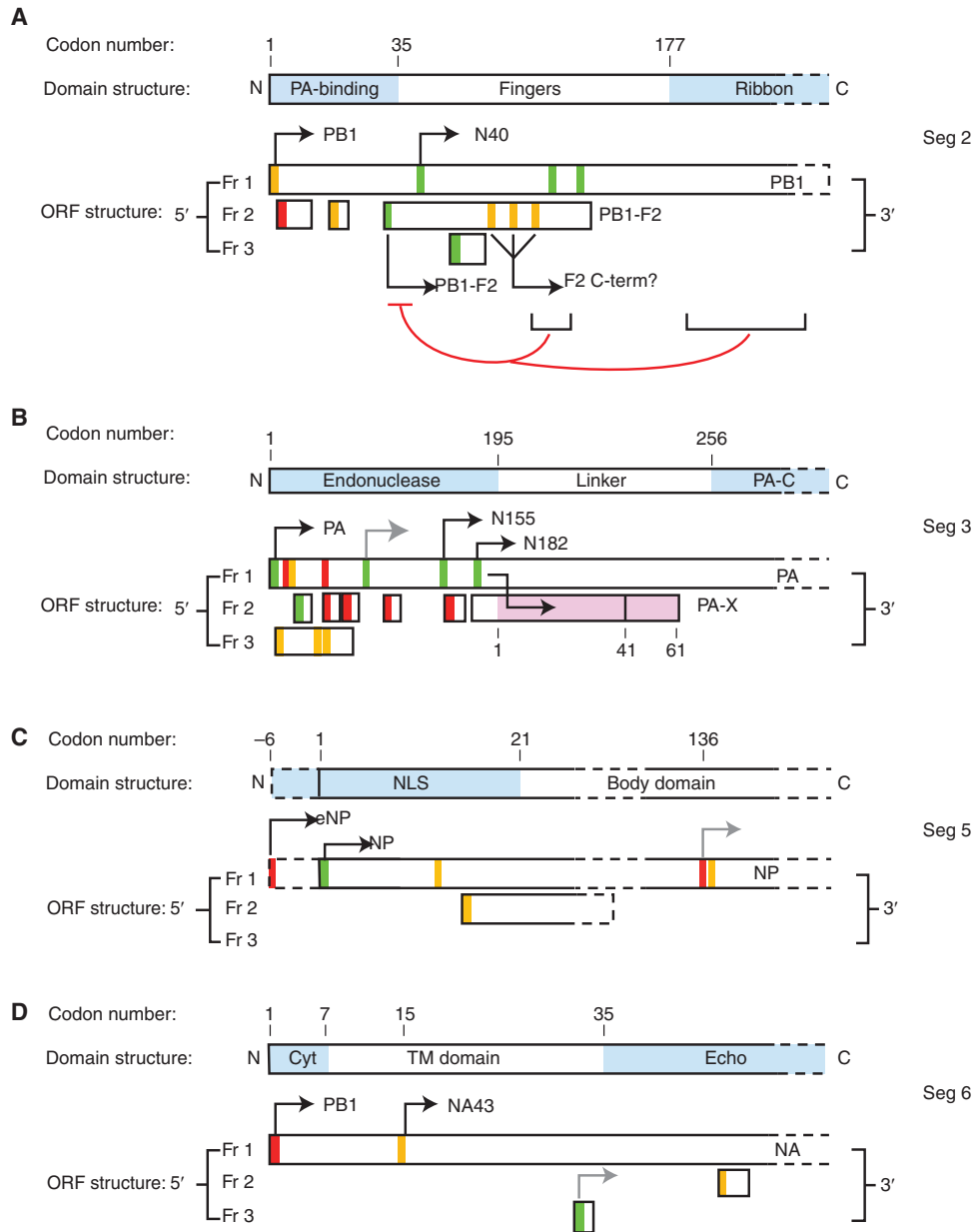


Figure 3. Schematic of the ORFs and encoded protein domain structures of the 5'-end of segment 2, 3, 5, and 6 mRNAs. Boxes represent ORFs in each of the three reading frames (Fr) or the major polypeptide species and their functional domains (*top bars*). AUG codons are represented by vertical lines, colored according to the predicted strength of their Kozak translation initiation context (green, strong; amber, intermediate; red, weak). Black arrows indicate AUGs known to be used for translation initiation for the indicated proteins. Gray arrows indicate AUG codons proposed to fire on the basis of ribosomal profiling data. (A) Represents the first 900 nucleotides of the PR8 sequence. Thin brackets indicate the approximate location of two sequence elements proposed to negatively regulate (blunt arrow) PB1-F2 expression. (Panel adapted from data in Figure 1A of Wise et al. 2009.) (B) Represents the first 900 nucleotides of PR8 sequence; note that in this (and most IAV strains), PA-N155 and -N182 are initiated from the 13th and 15th AUG codons in the segment, respectively. The stepped arrow indicates the +1 ribosomal frameshift event that accesses the X-ORF to produce PA-X. The gray line in the X-ORF indicates a common length polymorphism present in the H1N1pdm09 virus lineage. (C) Represents the first 150 nucleotides of A/England/195/2009 (H1N1) sequence, as well as downstream sequences around 450 nucleotides. (D) Represents the first 150 nucleotides of the WSN sequence.

that are reviewed in more detail elsewhere (Kamal et al. 2018; Cheung et al. 2020).

Despite much effort at molecular characterization of PB1-F2 in vitro, attempts to elucidate in vivo function have given conflicting results. In some virus backgrounds, disruption of PB1-F2 expression (or, conversely, repairing the gene in viruses that have lost it) has given no phenotype in animal experiments (e.g., Zamarin et al. 2006; Hai et al. 2010). In other circumstances, the presence, or precise sequence of PB1-F2 does affect pathogenesis (Zamarin et al. 2006; Conenello et al. 2007; Schmolke et al. 2011; Leymarie et al. 2014), but no overall simple conclusion can be drawn as to whether PB1-F2 expression leads to exacerbation or amelioration of disease. This almost certainly reflects viral strain- and host-dependent effects and is reviewed elsewhere (Kamal et al. 2018; Cheung et al. 2020). Consistent with virus-derived variation, not all IAV strains possess a full-length (defined as ≥ 87 codon) ORF; in general, avian strains of IAV tend to have full-length genes, whereas human and swine viruses (notably including the H1N1pdm09 virus) often do not (Zell et al. 2007; Pasricha et al. 2013; Kamal et al. 2018).

PB1-N40

Like PB1-F2, PB1-N40 is expressed from segment 2 mRNA, but originates from the fifth AUG codon on the message (Fig. 3A) and thus corresponds to an amino-terminally truncated version of the main segment product PB1 (Wise et al. 2009). PB1-N40 was in fact originally identified more than a decade prior to the discovery of PB1-F2 as an unexpected PB1-related polypeptide (Akkina et al. 1991), but the techniques of the day did not permit detailed characterization. As with PB1-F2, ribosomes reach AUG 5 through leaky scanning past the first few codons on the mRNA, probably aided by a termination-reinitiation mechanism following translation of an intervening short ORF started at AUG 3 (Wise et al. 2011). Importantly, expression of PB1-N40 is down-regulated by the presence of the PB1-F2 start codon (Wise et al. 2009), a confounding factor for studies that have used the mutational strategy of altering AUG 4 to re-

move PB1-F2 expression. The amino-terminal 39 amino acids present in PB1, but missing from PB1-N40, include the primary interaction site for the PA polymerase subunit (Fig. 3A; Pérez and Donis 1995), and therefore PB1-N40 lacks heterodimerization function and, consequently, nuclear import and transcriptase functions (Fodor and Smith 2004; Wise et al. 2009). PB1-N40 nevertheless retains some ability to interact with PB2 and the polymerase trimer, as well as with a variety of cellular proteins (Wise et al. 2009; Wang et al. 2019).

The PB1-N40 AUG codon and its favorable Kozak context are very highly conserved in IAV isolates, and the protein is expressed at reasonable levels ($\sim 5\%$ of that of PB1) in a variety of IAV strains that have been tested (Wise et al. 2009). Nevertheless, PB1-N40 function remains to be clarified. Ablating PB1-N40 expression (by mutating its AUG codon, with consequent mutation of the full-length PB1 protein) or inducing its overexpression (by mutating the PB1-F2 AUG codon) slightly reduced viral fitness of the H1N1 laboratory-adapted A/Puerto Rico/8/34 (PR8) and WSN strains in vitro. However, double deletion of PB1-N40 and PB1-F2 resulted in WT-like virus propagation (Wise et al. 2009; Tauber et al. 2012). Thus, as it currently stands, PB1-N40 is an accessory protein in search of a function.

SEGMENT 3: PA-N155 AND PA-N182

PA-N155 and PA-N182 are amino-terminally truncated isoforms of PA translated from the 11th and 13th AUG codons (in the WSN strain), respectively, of segment 3 mRNA (Muramoto et al. 2013). Similar to PB1-N40, these products were noted prior to the advent of reverse genetics and other ready means of further characterization (Akkina et al. 1991). Also like PB1-N40, the missing amino-terminal regions prevent the PA isoforms from supporting viral transcriptase function (Muramoto et al. 2013), as they lack the amino-terminal endonuclease domain necessary for cap snatching (Fig. 3B; Dias et al. 2009; Yuan et al. 2009; Pflug et al. 2014). The mechanism (transcriptional or translational) by which these AUG codons are accessed by ribosomes has



not been elucidated. Inspection of the ORF structure of the 5'-end of segment 3 suggests that it is unlikely to be simple leaky ribosomal scanning, as (unlike the PB1 AUG), the PA AUG has a strong Kozak consensus and there are multiple other methionine codons upstream of those used to make PA-N155 and -N182 (Fig. 3B). However, the PA isoform initiator codons are highly conserved among IAV sequences and are present in isolates from different host species, making it likely that the majority of IAV strains will express them (Akkina et al. 1991; Muramoto et al. 2013). The precise roles of these truncated PA polypeptides within the IAV life cycle have not yet been elucidated. Recent studies on these polypeptides from an H5N1 virus suggested interactions of PA-N155 and PA-N182 with a variety of (chicken) cellular proteins involved in RNA processing, protein transport, and various cellular signaling pathways (Wang et al. 2018). Many, if not most, of these interactions might be common to the interactome of full-length PA protein and their significance is unclear. Nonetheless, mutant WSN viruses unable to express PA-N155 and/or -N182 following mutation of the relevant AUG codons showed delayed virus replication *in vitro* and reduced virulence in mouse models (Muramoto et al. 2013). As with PB1-N40 mutants, however, it must be borne in mind that the strategy used to ablate expression of the accessory proteins also mutates the primary gene product of the segment, making it hard to rule out pleiotropic effects. As with PB1-N40, further work is required to understand the significance of these polypeptides.

PA-X

PA-X expression occurs through a low-level (~1%) +1 ribosomal frameshift event, driven by a rare arginine codon (CGU) next to a sequence that facilitates transfer RNA (tRNA) realignment, leading to the production of a fusion protein containing the first 191 amino acids of PA (corresponding to the endonuclease domain) and a short carboxy-terminal domain translated from reading frame 2 that lacks its own AUG codon (the X-ORF; Fig. 3B; Firth et al. 2012; Jagger et al. 2012). The frameshift

site is highly conserved among IAV strains, suggesting that most IAV strains will express it, but the length of the X-ORF varies, with common polymorphisms being either 41 or 61 codons (Shi et al. 2012; Rash et al. 2014). The best described function of the protein is to contribute to host cell shutoff, through the PA endonuclease domain acting as a broad (but not nonspecific) mRNA endonuclease when liberated from the rest of the viral polymerase complex (Jagger et al. 2012; Desmet et al. 2013; Khaperskyy et al. 2016; Muller and Glaunsinger 2017; Chaimayo et al. 2018; Gaucherand et al. 2019). This is not to say that the sequences from the X-ORF are nonfunctional; polymorphisms in this region affect shutoff activity, possibly by affecting PA-X subcellular localization (Bavagnoli et al. 2015; Oishi et al. 2015; Feng et al. 2016; Hayashi et al. 2016; Nogales et al. 2018).

PA-X functions other than destroying host mRNAs are also possible. The strength of the PA-X shutoff activity varies between IAV strains but even in viruses with low PA-X shutoff activity (e.g., the lab-adapted WSN and PR8 strains), altering PA-X expression has phenotypic effects (Desmet et al. 2013; Hussain et al. 2018; Rigby et al. 2019). Other postulated mechanisms include interactions with stress granules and histone deacetylase 4 (Khaperskyy et al. 2012, 2014; Galvin and Husain 2019). In sum, the picture is of a viral protein that limits cellular responses, including induction of innate immunity, to infection.

Expression of PA-X is not required for virus replication *in vitro*, in mice, in chickens, or *in ovo* (Jagger et al. 2012; Hu et al. 2015; Hussain et al. 2018). Nevertheless, alteration of PA-X expression or sequence has shown phenotypic effects in a variety of animal models of IAV disease. In mice, using H1N1 and H5N1 strains of virus, the absence of PA-X resulted in augmented clinical disease and/or immune responses (Jagger et al. 2012; Gao et al. 2015a; Hu et al. 2015; Rigby et al. 2019). Conversely, comparisons between wild-type and mutant H9N2 avian or human H1N1 2009 pandemic viruses resulted in the opposite effect of reduced virulence in the absence of PA-X, suggesting that viral context is important for the *in vivo* effect of PA-X

on disease outcome (Gao et al. 2015b; Hayashi et al. 2015; Lee et al. 2017; Nogales et al. 2017). Data for other host species are more limited, but in chickens, loss of PA-X expression increased pathogenicity (Hu et al. 2015), whereas in pigs, mutating the frameshift site decreased pathogenicity (Xu et al. 2017). Therefore, as seen for PB1-F2, PA-X exhibits virus strain- and host-specific functions whose interplay and outcome cannot currently be predicted with certainty.

SEGMENT 5: AN AMINO-TERMINALLY EXTENDED ISOFORM OF NP

The UTRs of IAV are generally more conserved than the coding regions and in many cases are disregarded when “full-length” segment sequences are reported on public databases. Nevertheless, the UTRs do exhibit sequence variation (Furuse and Oshitani 2011; Benkaroun et al. 2018), and many H1N1 viruses descended from the 1918 pandemic virus (both human and swine) contain a C28A nucleotide polymorphism in the 5'-UTR of segment 5 that results in the addition of a new start codon upstream of the canonical NP start site (Wanitchang et al. 2011; Wise et al. 2019). This new AUG codon is in a poor Kozak consensus, but in frame with the primary NP AUG so that its use leads to an NP isoform with an extra six amino acids at the amino terminus (Fig. 3C). First noticed in the context of the 2009 pandemic virus, the presence of this upstream AUG (uAUG) was associated with a slight (less than twofold) increase in viral gene expression (minireplicon) activity, but whether or not it was used for translation initiation was not addressed (Wanitchang et al. 2011). Subsequent work has shown that the uAUG is indeed used, but not exclusively, for translation initiation *in vitro* and *in vivo* to produce a mixture of normal and elongated (e)NP. The presence or absence of the uAUG codon had little effect on virus replication *in vitro*, but *in vivo*, in both mice and pigs, expression of eNP correlated with increased viral pathogenicity (Wise et al. 2019). The mechanism behind this phenotypic effect *in vivo* is unknown. The additional six amino acids in eNP lie at the amino terminus of a flexible sequence primarily characterized

as containing an unconventional nuclear localization signal (Wang et al. 1997). Differential interactions between NP and importin α isoforms have been associated with IAV cell tropism (Gabriel et al. 2011), so this represents a potentially testable hypothesis for eNP function.

SEGMENT 6: NA43

NA43 is an amino-terminally truncated form of the NA polypeptide that lacks the first 14 amino acids of the canonical protein and thus the entire amino-terminal cytoplasmic domain and a portion of the transmembrane domain (Fig. 3D). It was discovered in a PR8 reassortant virus with WSN segment 6 as an unexpected peak in a ribosome profiling (Ribo-seq) data set generated using the inhibitor lactimidomycin to bias reads toward sites of translation initiation (Machkovech et al. 2019). Initiation at the second AUG present in the mRNA (codon 15 in the NA gene, starting at the 43rd nucleotide downstream of the primary NA AUG, hence the protein name chosen by the authors) seems likely to occur via leaky scanning; the primary AUG, serving as the initiation site for NA translation, is in a poor Kozak consensus and the second (NA43) AUG is in a moderate Kozak consensus (Fig. 3D). Most, but not all, N1 NAs have this methionine codon, suggesting that NA43 expression may be widespread. The existence of the truncated NA polypeptide is well supported by plasmid-based experiments for WSN and the early H1N1pdm09 strain A/California/04/2009, but has not been unequivocally demonstrated in actual virus infection (Machkovech et al. 2019). Despite the partial loss of the transmembrane domain, exogenously expressed NA43 still accumulates on the plasma membrane. However, probing its importance through reverse genetics gave no indication of functional importance for replication *in vitro* or in mice (Machkovech et al. 2019), so its significance remains unclear.

SEGMENT 7: M42 AND OTHER SEGMENT 7 GENE PRODUCTS

Segment 7 is known to produce a primary unspliced mRNA and up to three other transcripts

via alternative splicing: mRNAs 2–4. These splice variants use a common splice acceptor site but use different splice donor sequences (for review, see Dubois et al. 2014). Virus strain-dependent variations in the sequences of these splice donor sites leads to variability in their use. All IAVs produce the unspliced transcript in abundance, from which the M1 polypeptide is translated from the first AUG codon in the transcript (Fig. 4A).

Spliced mRNA2 is produced by virtually all strains of IAV and encodes the ion channel, M2. The structure of the spliced transcript is such that the M2 ORF has the initiator codon and the first nine codons in common with M1 but thereafter encodes the carboxy-terminal 88 amino acids from what was the distal end of the +2 reading frame before the splice event (Fig. 4B; Lamb et al. 1981).

Spliced mRNA3 is produced using the most 5'-proximal splice donor site, but because this is at the boundary of the conserved terminal promoter element and segment-unique UTR, its use removes the M1/M2 start codon and the resulting short mRNA lacks major protein-coding potential. The first potential ORF encodes a putative nine amino acid peptide from the carboxyl terminus of M1 (Fig. 4C) that has been named M3. Production of mRNA3 has been proposed to negatively regulate segment 7 protein expression at early stages of infection (Shih et al. 1995). However, its presence is nonessential for virus growth in vitro, and protein expression from the mRNA has also not been detected (Chiang et al. 2008; Jackson and Lamb 2008). Production of mRNA3 has become enshrined in influenza textbooks through its early discovery (Lamb et al. 1981), but it is not clear if all strains of IAV produce it; for example, the PR8 strain accumulates far lower quantities of mRNA3 than other human strains of IAV (Wise et al. 2012).

A small fraction of IAV strains produce appreciable amounts of a third splice variant, mRNA4 (Shih et al. 1998; Wise et al. 2012; Dubois et al. 2014). This mRNA is produced using the most 3' of the known splice donor sites and is predicted to express a 54 amino acid-long internally deleted version of M1 from the first

AUG codon (Fig. 4D), but no such protein product has been detected. However, in the PR8 strain of IAV, mRNA4 has been shown to encode an M2 isoform, named M42 (Wise et al. 2012). M42 translation is initiated from the second AUG codon in the +2 frame, to produce a polypeptide that is identical to M2 downstream of the splice junction but which has a unique amino-terminal ectodomain sequence (Fig. 4D). M42 can functionally replace M2 during virus replication, despite accumulating less at the plasma membrane and more toward Golgi localization than M2 (Wise et al. 2012).

SEGMENT 8: tNS1 AND NS3

The main polypeptide species made from segment 8 is NS1, the primary viral IFN antagonist (Hale et al. 2008). NS1 is translated from the first AUG in the unspliced mRNA transcript, to produce a protein around 230 amino acids long (the length varies between IAV strains) that folds into two domains: an amino-terminal RNA-binding domain and a carboxy-terminal “effector” domain, responsible for binding multiple cellular polypeptides (Fig. 5A). Like segment 7, segment 8 encodes two core IAV polypeptides. Differential splicing of the primary transcript using a splice donor site near the 5'-end of the mRNA leads to the production of an 121 amino acid NEP protein that shares its amino-terminal 10 residues with NS1 (Lamb et al. 1978; Inglis et al. 1979) but contains a carboxyl terminus expressed from the +2 reading frame of the segment (Fig. 5B). All strains of IAV contain this 5' splice donor sequence and produce NEP, as it is essential for RNP nuclear export and completion of the virus life cycle (O'Neill et al. 1998; Neumann et al. 2000).

Recently, amino-terminally truncated versions of NS1 have been identified (Kuo et al. 2016). These “tNS1” polypeptides are expressed from the 5th and/or 6th AUG codons of segment 8 and thus lack the first 78 or 80 amino acids of NS1, effectively producing the NS1 effector domain in isolation (Fig. 5A). A mutant PR8 virus lacking these AUG codons induced higher levels of IRF3 phosphorylation and hence greater IFN- β induction (Kuo et al. 2016).

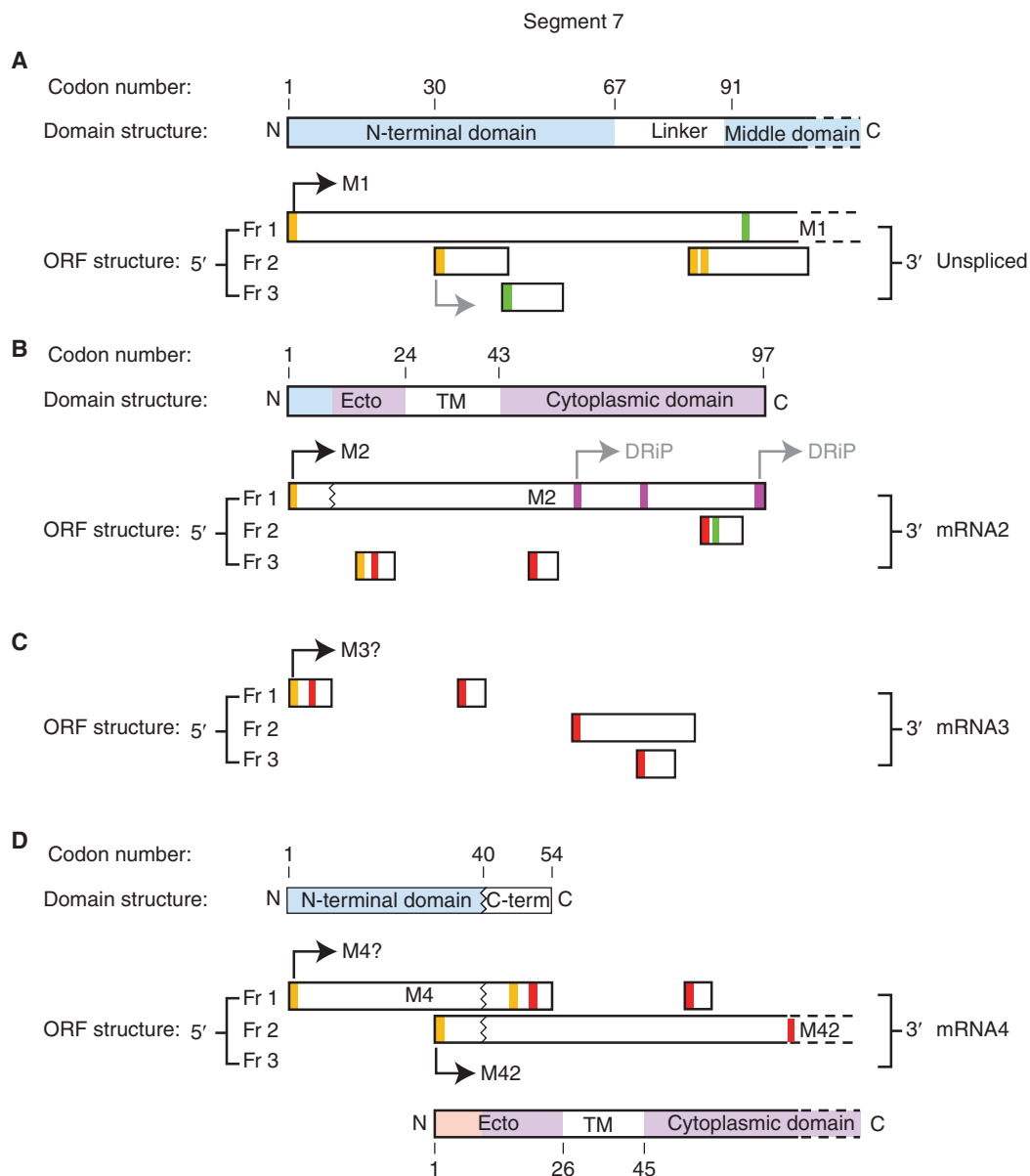


Figure 4. Schematic of the open reading frames (ORFs) and encoded protein domain structures of segment 7 mRNAs (PR8 strain). Boxes represent ORFs in each of the three reading frames (Fr) or the major polypeptide species and their functional domains (partially shaded bars). Zigzag lines represent splice junctions. AUG codons are represented by vertical lines, colored according to the predicted strength of their Kozak translation initiation context (green, strong; amber, intermediate; red, weak). Purple bars indicate CUG codons thought to be used for translation initiation. Black arrows indicate AUGs known to be used for translation initiation for the indicated proteins. Gray arrows indicate AUG codons proposed to fire on the basis of ribosomal profiling data or (DRiPs) from T-cell epitope data. (A) Represents the first 350 nucleotides of the unspliced transcript. (B) Represents mRNA2 (M2). (C) Represents mRNA3. (D) Represents mRNA4.

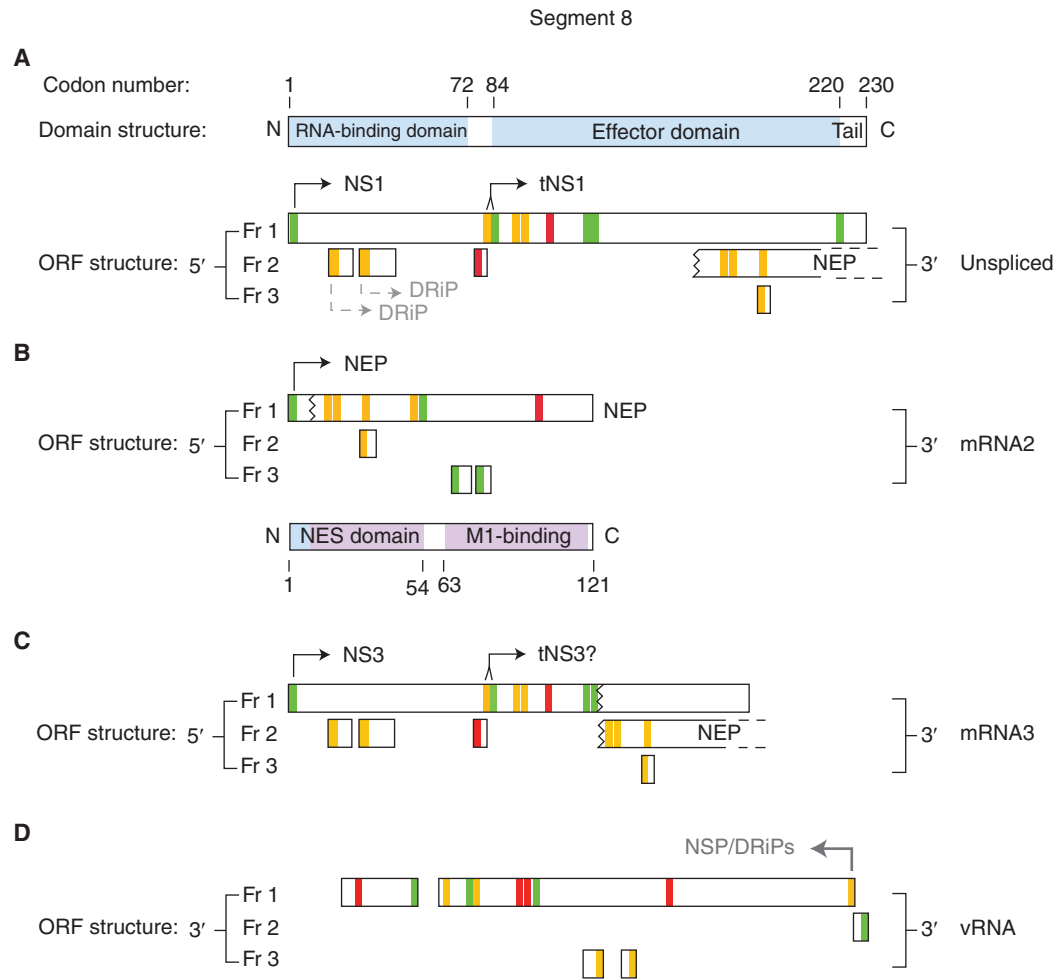


Figure 5. Schematic of the ORFs and encoded protein domain structures of segment 8 mRNAs and vRNA (PR8 strain). Zigzag lines represent splice junctions. Boxes represent ORFs in each of the three reading frames (Fr) or the major polypeptide species and their functional domains (partially shaded bars). AUG codons are represented by vertical lines, colored according to the predicted strength of their Kozak translation initiation context (green, strong; amber, intermediate; red, weak). Black arrows indicate AUGs known to be used for translation initiation for the indicated proteins. Gray arrows indicate AUG codons proposed to fire on the basis of T-cell epitope data (DRiPs). (A) Represents the unspliced transcript. (B) Represents mRNA2 (NEP). (C) Represents mRNA3 (NS3). (D) Represents vRNA (the putative NSP protein).

A priori, this could be due to nonsynonymous mutation of NS1 and/or failure to express the tNS1 polypeptides. However, exogenously expressed tNS1 polypeptides displayed differential localization (cytoplasmic versus nuclear) to the WT protein and were able to antagonize IRF3 activation, supporting the suggestion that loss of their expression contributed to the phenotype. The mechanism underlying their expression has

not been characterized—the main NS1 AUG codon is in a seemingly strong Kozak consensus, so leaky ribosomal scanning is perhaps not the explanation.

A third variant of the NS1 protein (“NS3”) has also been identified. The NS3 protein is encoded by a splice variant mRNA in which an intron between nucleotides 400 and 528 has been removed to produce an internally truncat-

ed version of NS1 (Fig. 4C), missing amino acids 126-168 (Selman et al. 2012). NS3 was discovered in the process of characterizing the phenotypic effects of a D125G change in the NS1 ORF following adaptation of the prototype virus of the 1968 H3N2 pandemic (A/Hong Kong/1/1968) to growth in mice (Brown et al. 2001). This revealed that increased virus propagation in murine cells actually resulted from the mutation introducing a novel splice donor site, rather than the nonsynonymous change in NS1 (Selman et al. 2012). As with segment 7 mRNA4, production of the NS3 mRNA is likely to occur in only a small minority of IAV strains, because most do not contain the necessary splice donor sequence.

DRiPs, UFOs, AND NONSENSE PEPTIDES

The IAV gene products described in the sections above are all bona fide proteins—within the size range expected for polypeptides capable of folding into a discrete domain (Garbuzynskiy et al. 2013) and, in all cases, visualized at some point as a “band on a gel.” However, there is also evidence for a further understratum of viral gene expression that occurs in infected cells, largely seen at the level of peptide species. Much of the data showing the existence of these products stem from work aimed at understanding the rapid kinetics of MHC Class I peptide presentation of virus-derived peptides—seemingly too fast to be solely explained through the natural turnover of stably folded viral polypeptides (Esquivel et al. 1992). Instead, the defective ribosomal product (DRiP) hypothesis holds that the relatively rare mistranslation events that produce defective polypeptides provide a source of more rapidly turned-over polypeptides that feed the antigenic processing pathway (Yewdell et al. 1996). These DRiPs could be miscoded products from bona fide viral ORFs or products from viral ORFs that have not been selected to encode functional proteins. The sensitive assay methods developed in the course of testing this hypothesis (reviewed elsewhere [Wei and Yewdell 2019]) have not only shed light on fundamental processes of cell biology but have identified a variety of unexpected translation events that occur dur-

ing IAV replication. As well as prompting the discovery of PB1-F2 (promptly reclassified as a genuine accessory gene) (Chen et al. 2001), specific DRiPs that provide T cell epitopes originating from both canonical (AUG) and noncanonical (CUG) internal translation initiation in the M1 and NS1 genes (Figs. 4B and 5A, respectively) have been identified (Yang et al. 2016; Zanker et al. 2019).

In addition, T-cell epitope presentation provides the best evidence to date for protein expression from the negative-strand RNA of IAV; from the enigmatic long ORF (167 codons in PR8 but up to 216 codons in some strains; Fig. 5D) present in segment 8 vRNA (Baez et al. 1980; Zhirnov et al. 2007; Clifford et al. 2009; Hickman et al. 2018). Expression of a full-length negative-strand polypeptide (NSP) has not been detected, however (Hickman et al. 2018).

Other evidence for noncanonical gene expression events in IAV has come from the technique of ribosomal profiling. This technique, which maps fragments of mRNA protected from RNase digestion by ribosomes back to the genome to (among other things) identify sites of translation initiation (Ingolia 2016), has identified at least four previously unrecognized ribosome start sites in the IAV genome (Machkovech et al. 2019). These are internal methionine codons at positions 86 and 136 of the PA and NP genes (Fig. 3B,C), respectively, and out-of-frame ORFs in the NA (Fig. 3D) and M1 genes (Fig. 4A). No matching polypeptides have yet been defined for these initiation events, in contrast to the NA43 polypeptide described above, whose identification also stemmed from the same study.

A final layer of complexity in IAV gene expression comes from the demonstration that the cap-snatching mechanism itself can generate novel translational start sites on viral mRNAs, when the leader sequence generated from the host cell pre-mRNA contains an AUG codon or non-AUG initiation site (Sloan et al. 2019; Ho et al. 2020). The evidence for this comes from a variety of methods, including deep sequencing of the capped leader sequences, ribosomal profiling, T-cell epitope assays, and mass-spectrometric detection of novel peptides.

This mechanism can in theory produce amino-terminal extensions to the core IAV polypeptides, as well as providing means to drive expression of otherwise cryptic ORFs near the 5'-end of the viral segments. However, given the level of variability in the cellular sequences (upward of 4 million unique leader sequences generated during the course of infection [Clohisey et al. 2020]), this mechanism is unlikely to generate large amounts of a discrete polypeptide species, and its ultimate significance remains to be determined.

CONCLUDING REMARKS

Nearly 40 years ago, IAV was characterized as encoding 10 polypeptides, initiated from eight AUG codons elaborated by two major mRNA splice isoforms. Since then, a further 10 polypeptide species have been defined in some detail, and as many again potential protein/peptide products suggested. These novel products are produced from multiple alternative translation initiation sites (both AUG and CUG codons), minor spliced mRNAs, and one example of ribosomal frameshifting. So, what does it all mean? A negative characterization could be gene expression mayhem in a colonized and dying cell, producing random junk. We think that a more useful way of considering the data is in the context of the current theories of how cellular genes are birthed de novo. In this hypothesis, genes do not just originate by diversification following gene duplication but can evolve step by step from noncoding sequence, through acqui-

sition of an ORF, that, once translated, allows selection (if the protein product is useful) to drive increases in expression, ORF length, protein domain structure, and stability (Carvunis et al. 2012; Schmitz and Bornberg-Bauer 2017).

Applying this model of gene evolution to IAV (or any other virus) gene structure provides a simple framework for conceptualizing the plethora of “new” polypeptide species defined in recent years (Fig. 6). At least a subset of the many ORFs in the IAV genome is translated to at least some extent. Some of these have been defined as DRiPs in the narrow sense of fortuitously being detected as a T-cell epitope in a mouse, as well as small size and/or low expression level and rapid turnover. Others, such as the amino-terminally truncated isoforms of the polymerase proteins and NA, in which no function has been shown, might also fall into a broader DRiP category of being translational noise of no importance to viral fitness, although their evident stability and domain structure perhaps argue against this. Further IAV polypeptides, in which unambiguous evidence of in vivo importance exists (e.g., PB1-F2, PA-X), can be firmly classified as accessory proteins. However, it is important to note that this model of gene birth also includes death, in which (lack of) selection can lead to gene loss (Carvunis et al. 2012). Indeed, the PB1-F2 gene might be following this backward trajectory in swine and/or humans (Zell et al. 2007; Trifonov et al. 2009). An admitted problem with the de novo gene birth hypothesis is the difficulty of observing the process in motion, given the evolutionary timescales animal

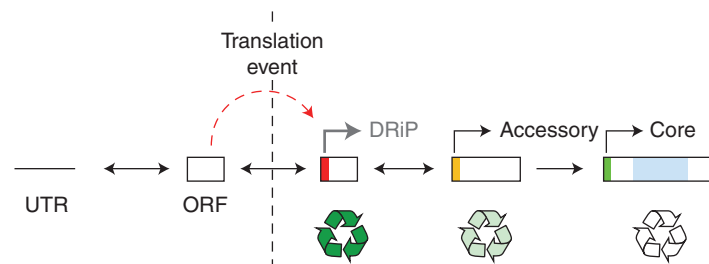


Figure 6. A model for gene birth and death in IAV. Boxes represent ORFs. Arrows and colored bars indicate approximate expression levels, whereas darker shading of the recycling symbol indicates faster turnover of the polypeptide species. (Figure based on data in Carvunis et al. 2012.)



R.M. Pinto et al.

and plant genomes evolve at (Schmitz and Bornberg-Bauer 2017). The far higher (perhaps a million-fold) mutation rate of RNA viruses (Belshaw et al. 2008) may make them a useful model system for observing gene evolution in action.

ACKNOWLEDGMENTS

P.D. and S.L. are supported by Institute Strategic Programme (BB/P013740/1). P.D. is also supported by project grant (BB/S00114X/1) funding from the U.K. Biotechnology and Biological Sciences Research Council, as well as the European Union's Horizon 2020 research and innovation programme under grant no. 727922 (DELTA-FLU). S.L. is also supported by the European Union's Horizon 2020 research and innovation programme under Grant No. 874835 (VEO). E.G. is supported by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship (211222/Z/18/Z). R.M.P. was supported by a PhD studentship from the University of Edinburgh.

This article has been made freely available online courtesy of TAUNS Laboratories.

REFERENCES

- Akkina RK, Richardson JC, Aguilera MC, Chi-Ming Y. 1991. Heterogeneous forms of polymerase proteins exist in influenza A virus-infected cells. *Virus Res* **19**: 17–30. doi:10.1016/0168-1702(91)90091-9
- Baez M, Taussig R, Zazra JJ, Young JF, Palese P, Reifeld A, Skalka AM. 1980. Complete nucleotide sequence of the influenza A/PR/8/34 virus NS gene and comparison with the NS genes of the A/Udorn/72 and A/FPV/Rostock/34 strains. *Nucleic Acids Res* **8**: 5845–5858. doi:10.1093/nar.8.23.5845
- Bavagnoli L, Cucuzza S, Campanini G, Rovida F, Paolucci S, Baldanti F, Maga G. 2015. The novel influenza A virus protein PA-X and its naturally deleted variant show different enzymatic properties in comparison to the viral endonuclease PA. *Nucleic Acids Res* **43**: 9405–9417. doi:10.1093/nar/gkv926
- Belshaw R, Gardner A, Rambaut A, Pybus OG. 2008. Pacing a small cage: mutation and RNA viruses. *Trends Ecol Evol* **23**: 188–193. doi:10.1016/j.tree.2007.11.010
- Benkaroun J, Robertson GJ, Whitney H, Lang AS. 2018. Analysis of the variability in the non-coding regions of influenza A viruses. *Vet Sci* **5**: 76. doi:10.3390/vetsci5030076
- Brown EG, Liu H, Kit LC, Baird S, Nesrallah M. 2001. Pattern of mutation in the genome of influenza A virus on adaptation to increased virulence in the mouse lung: identification of functional themes. *Proc Natl Acad Sci* **98**: 6883–6888. doi:10.1073/pnas.111165798
- Buehler J, Navi D, Lorusso A, Vincent A, Lager K, Miller CL. 2013. Influenza A virus PB1-F2 protein expression is regulated in a strain-specific manner by sequences located downstream of the PB1-F2 initiation codon. *J Virol* **87**: 10687–10699. doi:10.1128/JVI.01520-13
- Carr SM, Carnero E, Garcia-Sastre A, Brownlee GG, Fodor E. 2006. Characterization of a mitochondrial-targeting signal in the PB2 protein of influenza viruses. *Virology* **344**: 492–508. doi:10.1016/j.virol.2005.08.041
- Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirir MA, Simonis N, Charloteaux B, Hidalgo CA, Barrette J, Santhanam B, et al. 2012. Proto-genes and de novo gene birth. *Nature* **487**: 370–374. doi:10.1038/nature11184
- Chaimayo C, Dunagan M, Hayashi T, Santoso N, Takimoto T. 2018. Specificity and functional interplay between influenza virus PA-X and NS1 shutoff activity. *PLoS Pathog* **14**: e1007465. doi:10.1371/journal.ppat.1007465
- Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, Basta S, O'Neill R, Schickli J, Palese P, et al. 2001. A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med* **7**: 1306–1312. doi:10.1038/nm1201-1306
- Cheung PHH, Lee TWT, Chan CP, Jin DY. 2020. Influenza A virus PB1-F2 protein: an ambivalent innate immune modulator and virulence factor. *J Leukoc Biol* **42**: 2–9.
- Chiang C, Chen GW, Shih SR. 2008. Mutations at alternative 5' splice sites of M1 mRNA negatively affect influenza A virus viability and growth rate. *J Virol* **82**: 10873–10886. doi:10.1128/JVI.00506-08
- Clifford M, Twigg J, Upton C. 2009. Evidence for a novel gene associated with human influenza A viruses. *Virol J* **6**: 198. doi:10.1186/1743-422X-6-198
- Clohisey S, Parkinson N, Wang B, Bertin N, Wise H, Tomoiu A, FANTOM5 Consortium, Summers KM, Hendry RW, Carninci P, et al. 2020. Comprehensive characterization of transcriptional activity during influenza A virus infection reveals biases in cap-snatching of host RNA sequences. *J Virol* **94**: 2401. doi:10.1128/JVI.01720-19
- Conenello GM, Zamarin D, Perrone LA, Tumpey T, Palese P. 2007. A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. *PLoS Pathog* **3**: e141. doi:10.1371/journal.ppat.0030141
- Desmet EA, Bussey KA, Stone R, Takimoto T. 2013. Identification of the N-terminal domain of the influenza virus PA responsible for the suppression of host protein synthesis. *J Virol* **87**: 3108–3118. doi:10.1128/JVI.02826-12
- Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RWH. 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**: 914–918. doi:10.1038/nature07745
- Dubois J, Terrier O, Rosa-Calatrava M. 2014. Influenza viruses and mRNA splicing: doing more with less. *MBio* **5**: e00070-14. doi:10.1128/mBio.00070-14
- Esquivel F, Yewdell J, Bennink J. 1992. RMA/s cells present endogenously synthesized cytosolic proteins to class I-restricted cytotoxic T lymphocytes. *J Exp Med* **175**: 163–168. doi:10.1084/jem.175.1.163



- Feng KH, Sun M, Iketani S, Holmes EC, Parrish CR. 2016. Comparing the functions of equine and canine influenza H3N8 virus PA-X proteins: suppression of reporter gene expression and modulation of global host gene expression. *Virology* **496**: 138–146. doi:10.1016/j.virol.2016.06.001
- Firth AE, Jagger BW, Wise HM, Nelson CC, Parsawar K, Wills NM, Naphthine S, Taubenberger JK, Digard P, Atkins JF. 2012. Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UC-C_UUU_CGU and is in the +1 direction. *Open Biol* **2**: 120109. doi:10.1098/rsob.120109
- Fodor E, Smith M. 2004. The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J Virol* **78**: 9144–9153. doi:10.1128/JVI.78.17.9144-9153.2004
- Furuse Y, Oshitani H. 2011. Evolution of the influenza A virus untranslated regions. *Infect Genet Evol* **11**: 1150–1154. doi:10.1016/j.meegid.2011.04.006
- Gabriel G, Klingel K, Otte A, Thiele S, Hudjetz B, Arman-Kalcek G, Sauter M, Shmidt T, Rother F, Baumgarte S, et al. 2011. Differential use of importin- α isoforms governs cell tropism and host adaptation of influenza virus. *Nat Commun* **2**: 156. doi:10.1038/ncomms1158
- Galvin HD, Husain M. 2019. Influenza A virus-induced host caspase and viral PA-X antagonize the antiviral host factor, histone deacetylase 4. *J Biol Chem* **294**: 20207–20221. doi:10.1074/jbc.RA119.010650
- Gao H, Sun Y, Hu J, Qi L, Wang J, Xiong X, Wang Y, He Q, Lin Y, Kong W, et al. 2015a. The contribution of PA-X to the virulence of pandemic 2009 H1N1 and highly pathogenic H5N1 avian influenza viruses. *Sci Rep* **5**: 8262. doi:10.1038/srep08262
- Gao H, Xu G, Sun Y, Qi L, Wang J, Kong W, Sun H, Pu J, Chang KC, Liu J. 2015b. PA-X is a virulence factor in avian H9N2 influenza virus. *J Gen Virol* **96**: 2587–2594. doi:10.1099/jgv.0.000232
- Garbuzynskiy SO, Ivankov DN, Bogatyreva NS, Finkelstein AV. 2013. Golden triangle for folding rates of globular proteins. *Proc Natl Acad Sci* **110**: 147–150. doi:10.1073/pnas.1210180110
- García-Sastre A, Egorov A, Matassov D, Brandt S, Levy DE, Durbin JE, Palese P, Muster T. 1998. Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. *Virology* **252**: 324–330. doi:10.1006/viro.1998.9508
- Gaucherand L, Porter BK, Levene RE, Price EL, Schmalzing SK, Rycroft CH, Kevorkian Y, McCormick C, Khapersky DA, Gaglia MM. 2019. The influenza A virus endoribonuclease PA-X usurps host mRNA processing machinery to limit host gene expression. *Cell Rep* **27**: 776–792.e7. doi:10.1016/j.celrep.2019.03.063
- Gibbs JS, Malide D, Hornung F, Bennink JR, Yewdell JW. 2003. The influenza A virus PB1-F2 protein targets the inner mitochondrial membrane via a predicted basic amphipathic helix that disrupts mitochondrial function. *J Virol* **77**: 7214–7224. doi:10.1128/JVI.77.13.7214-7224.2003
- Graef KM, Vreede FT, Lau YF, McCall AW, Carr SM, Subbarao K, Fodor E. 2010. The PB2 subunit of the influenza virus RNA polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting expression of beta interferon. *J Virol* **84**: 8433–8445. doi:10.1128/JVI.00879-10
- Hai R, Schmolke M, Varga ZT, Manicassamy B, Wang TT, Belser JA, Pearce MB, García-Sastre A, Tumpey TM, Palese P. 2010. PB1-F2 expression by the 2009 pandemic H1N1 influenza virus has minimal impact on virulence in animal models. *J Virol* **84**: 4442–4450. doi:10.1128/JVI.02717-09
- Hale BG, Randall RE, Ortín J, Jackson D. 2008. The multifunctional NS1 protein of influenza A viruses. *J Gen Virol* **89**: 2359–2376. doi:10.1099/vir.0.2008/004606-0
- Hay AJ, Lomniczi B, Bellamy AR, Skehel JJ. 1977. Transcription of the influenza virus genome. *Virology* **83**: 337–355. doi:10.1016/0042-6822(77)90179-9
- Hay AJ, Skehel JJ, McCauley J. 1982. Characterization of influenza virus RNA complete transcripts. *Virology* **116**: 517–522. doi:10.1016/0042-6822(82)90144-1
- Hayashi T, MacDonald LA, Takimoto T. 2015. Influenza A virus protein PA-X contributes to viral growth and suppression of the host antiviral and immune responses. *J Virol* **89**: 6442–6452. doi:10.1128/JVI.00319-15
- Hayashi T, Chaimayo C, McGuinness J, Takimoto T. 2016. Critical role of the PA-X C-terminal domain of influenza A virus in its subcellular localization and shutoff activity. *J Virol* **90**: 7131–7141. doi:10.1128/JVI.00954-16
- Hickman HD, Mays JW, Gibbs J, Košik I, Magádan JG, Takeda K, Das S, Reynoso GV, Ngudiankama BF, Wei J, et al. 2018. Influenza A virus negative strand RNA is translated for CD8⁺ T cell immunosurveillance. *J Immunol* **201**: 1222–1228. doi:10.4049/jimmunol.1800586
- Ho JSY, Angel M, Ma Y, Sloan E, Wang G, Martinez-Romero C, Alenquer M, Roudkko V, Chung L, Zheng S, et al. 2020. Hybrid gene origination creates human-virus chimeric proteins during infection. *Cell* **181**: 1502–1517.e23. doi:10.1016/j.cell.2020.05.035
- Hu J, Mo Y, Wang X, Gu M, Hu Z, Zhong L, Wu Q, Hao X, Hu S, Liu W, et al. 2015. PA-X decreases the pathogenicity of highly pathogenic H5N1 influenza A virus in avian species by inhibiting virus replication and host response. *J Virol* **89**: 4126–4142. doi:10.1128/JVI.02132-14
- Hussain S, Turnbull ML, Wise HM, Jagger BW, Beard PM, Kovacicova K, Taubenberger JK, Vervelde L, Engelhardt OG, Digard P. 2018. Mutation of influenza A virus PA-X decreases pathogenicity in chicken embryos and can increase the yield of reassortant candidate vaccine viruses. *J Virol* **93**: e01551-18. doi:10.1128/JVI.01551-18
- Inglis SC, Brown CM. 1981. Spliced and unspliced RNAs encoded by virion RNA segment 7 of influenza virus. *Nucleic Acids Res* **9**: 2727–2740. doi:10.1093/nar/9.12.2727
- Inglis SC, Barrett T, Brown CM, Almond JW. 1979. The smallest genome RNA segment of influenza virus contains two genes that may overlap. *Proc Natl Acad Sci* **76**: 3790–3794. doi:10.1073/pnas.76.8.3790
- Ingolia NT. 2016. Ribosome footprint profiling of translation throughout the genome. *Cell* **165**: 22–33. doi:10.1016/j.cell.2016.02.066
- Jackson D, Lamb RA. 2008. The influenza A virus spliced messenger RNA M mRNA3 is not required for viral replication in tissue culture. *J Gen Virol* **89**: 3097–3101. doi:10.1099/vir.0.2008/004739-0



- Jagger BW, Wise HM, Kash JC, Walters K-A, Wills NM, Xiao YL, Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL, et al. 2012. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* **337**: 199–204. doi:10.1126/science.1222213
- James J, Smith N, Ross C, Iqbal M, Goodbourn S, Digard P, Barclay WS, Shelton H. 2019. The cellular localization of avian influenza virus PB1-F2 protein alters the magnitude of IFN2 promoter and NFκB-dependent promoter antagonism in chicken cells. *J Gen Virol* **100**: 414–430. doi:10.1099/jgv.0.001220
- Kamal RP, Kumar A, Davis CT, Tzeng WP, Nguyen T, Donis RO, Katz JM, York IA. 2015. Emergence of highly pathogenic avian influenza A(H5N1) virus PB1-F2 variants and their virulence in BALB/c mice. *J Virol* **89**: 5835–5846. doi:10.1128/JVI.03137-14
- Kamal RP, Alymova IV, York IA. 2018. Evolution and virulence of influenza A virus protein PB1-F2. *Int J Mol Sci* **19**: 96. doi:10.3390/ijms19010096
- Khaperskyy DA, Hatchette TF, McCormick C. 2012. Influenza A virus inhibits cytoplasmic stress granule formation. *FASEB J* **26**: 1629–1639. doi:10.1096/fj.11-196915
- Khaperskyy DA, Emara MM, Johnston BP, Anderson P, Hatchette TF, McCormick C. 2014. Influenza A virus host shutoff disables antiviral stress-induced translation arrest. *PLoS Pathog* **10**: e1004217. doi:10.1371/journal.ppat.1004217
- Khaperskyy DA, Schmaling S, Larkins-Ford J, McCormick C, Gaglia MM. 2016. Selective degradation of host RNA polymerase II transcripts by influenza A virus PA-X host shutoff protein. *PLoS Pathog* **12**: e1005427. doi:10.1371/journal.ppat.1005427
- Kozak M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283–292. doi:10.1016/0092-8674(86)90762-2
- Kuo RL, Li LH, Lin SJ, Li ZH, Chen GW, Chang CK, Wang YR, Tam EH, Gong YN, Krug RM, et al. 2016. Role of N terminus-truncated NS1 proteins of influenza A virus in inhibiting IRF3 activation. *J Virol* **90**: 4696–4705. doi:10.1128/JVI.02843-15
- Lamb RA, Choppin PW. 1981. Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus. *Virology* **112**: 729–737. doi:10.1016/0042-6822(81)90317-2
- Lamb RA, Lai CJ. 1980. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. *Cell* **21**: 475–485. doi:10.1016/0092-8674(80)90484-5
- Lamb RA, Etkind PR, Choppin PW. 1978. Evidence for a ninth influenza viral polypeptide. *Virology* **91**: 60–78. doi:10.1016/0042-6822(78)90355-0
- Lamb RA, Lai CJ, Choppin PW. 1981. Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. *Proc Natl Acad Sci* **78**: 4170–4174.
- Lee J, Yu H, Li Y, Ma J, Lang Y, Duff M, Henningson J, Liu Q, Li Y, Nagy A, et al. 2017. Impacts of different expressions of PA-X protein on 2009 pandemic H1N1 virus replication, pathogenicity and host immune responses. *Virology* **504**: 25–35. doi:10.1016/j.virol.2017.01.015
- Leymarie O, Embury-Hyatt C, Chevalier C, Jouneau L, Moroldo M, Da Costa B, Berhane Y, Delmas B, Weingartl HM, Le Goffic R, et al. 2014. PB1-F2 attenuates virulence of highly pathogenic avian H5N1 influenza virus in chickens. *PLoS One* **9**: e100679. doi:10.1371/journal.pone.0100679
- Leymarie O, Meyer L, Tafforeau L, Lotteau V, Costa BD, Delmas B, Chevalier C, Le Goffic R. 2017. Influenza virus protein PB1-F2 interacts with CALCOCO2 (NDP52) to modulate innate immune response. *J Gen Virol* **98**: 1196–1208. doi:10.1099/jgv.0.000782
- Liu C, Air GM. 1993. Selection and characterization of a neuraminidase-minus mutant of influenza virus and its rescue by cloned neuraminidase genes. *Virology* **194**: 403–407. doi:10.1006/viro.1993.1276
- Lycett SJ, Duchatel F, Digard P. 2019. A brief history of bird flu. *Philos Trans R Soc Lond B Biol Sci* **374**: 20180257. doi:10.1098/rstb.2018.0257
- Machkovech HM, Bloom JD, Subramaniam AR. 2019. Comprehensive profiling of translation initiation in influenza virus infected cells. *PLoS Pathog* **15**: e1007518. doi:10.1371/journal.ppat.1007518
- Moules V, Ferraris O, Terrier O, Giudice E, Yver M, Rolland JP, Bouscambert-Duchamp M, Bergeron C, Ottmann M, Fournier E, et al. 2010. In vitro characterization of naturally occurring influenza H3NA- viruses lacking the NA gene segment: toward a new mechanism of viral resistance? *Virology* **404**: 215–224. doi:10.1016/j.virol.2010.04.030
- Muller M, Glaunsinger BA. 2017. Nuclease escape elements protect messenger RNA against cleavage by multiple viral endonucleases. *PLoS Pathog* **13**: e1006593. doi:10.1371/journal.ppat.1006593
- Muramoto Y, Noda T, Kawakami E, Akkina R, Kawaoka Y. 2013. Identification of novel influenza A virus proteins translated from PA mRNA. *J Virol* **87**: 2455–2462. doi:10.1128/JVI.02656-12
- Neumann G, Hughes MT, Kawaoka Y. 2000. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J* **19**: 6751–6758. doi:10.1093/emboj/19.24.6751
- Nogales A, Rodriguez L, DeDiego ML, Topham DJ, Martínez-Sobrido L. 2017. Interplay of PA-X and NS1 proteins in replication and pathogenesis of a temperature-sensitive 2009 pandemic H1N1 influenza A virus. *J Virol* **91**: e00720-17.
- Nogales A, Martínez-Sobrido L, Chiem K, Topham DJ, DeDiego ML. 2018. Functional evolution of the 2009 pandemic H1N1 influenza virus NS1 and PA in humans. *J Virol* **92**: 115. doi:10.1128/JVI.01206-18
- Norton GP, Tanaka T, Tobita K, Nakada S, Buonagurio DA, Greenspan D, Krystal M, Palese P. 1987. Infectious influenza A and B virus variants with long carboxyl terminal deletions in the NS1 polypeptides. *Virology* **156**: 204–213. doi:10.1016/0042-6822(87)90399-0
- Oishi K, Yamayoshi S, Kawaoka Y. 2015. Mapping of a region of the PA-X protein of influenza A virus that is important for its shutoff activity. *J Virol* **89**: 8661–8665. doi:10.1128/JVI.01132-15
- O'Neill RE, Talon J, Palese P. 1998. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribo-



- nucleoproteins. *EMBO J* **17**: 288–296. doi:10.1093/emboj/17.1.288
- Pasricha G, Mishra AC, Chakrabarti AK. 2013. Comprehensive global amino acid sequence analysis of PB1F2 protein of influenza A H5N1 viruses and the influenza A virus subtypes responsible for the 20th-century pandemics. *Influenza Other Respir Viruses* **7**: 497–505. doi:10.1111/j.1750-2659.2012.00400.x
- Pérez DR, Donis RO. 1995. A 48-amino-acid region of influenza A virus PB1 protein is sufficient for complex formation with PA. *J Virol* **69**: 6932–6939. doi:10.1128/JVI.69.11.6932-6939.1995
- Pflug A, Guilligay D, Reich S, Cusack S. 2014. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**: 355–360. doi:10.1038/nature14008
- Plotch SJ, Bouloy M, Ulmanen I, Krug RM. 1981. A unique cap(m⁷GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**: 847–858. doi:10.1016/0092-8674(81)90449-9
- Poon LL, Pritlove DC, Fodor E, Brownlee GG. 1999. Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *J Virol* **73**: 3473–3476. doi:10.1128/JVI.73.4.3473-3476.1999
- Rash A, Woodward A, Bryant N, McCauley J, Elton D. 2014. An efficient genome sequencing method for equine influenza [H3N8] virus reveals a new polymorphism in the PA-X protein. *Virology* **11**: 159. doi:10.1186/1743-422X-11-159
- Reis AL, McCauley JW. 2013. The influenza virus protein PB1-F2 interacts with IKK β and modulates NF- κ B signaling. *PLoS One* **8**: e63852. doi:10.1371/journal.pone.0063852
- Rigby RE, Wise HM, Smith N, Digard P, Rehwinkel J. 2019. PA-X antagonises MAVS-dependent accumulation of early type I interferon messenger RNAs during influenza A virus infection. *Sci Rep* **9**: 7216. doi:10.1038/s41598-019-43632-6
- Robertson JS, Schubert M, Lazzarini RA. 1981. Polyadenylation sites for influenza virus mRNA. *J Virol* **38**: 157–163. doi:10.1128/JVI.38.1.157-163.1981
- Schmitz JF, Bornberg-Bauer E. 2017. Fact or fiction: updates on how protein-coding genes might emerge de novo from previously non-coding DNA. *F1000Res* **6**: 57. doi:10.12688/f1000research.10079.1
- Schmolke M, Manicassamy B, Pena L, Sutton T, Hai R, Varga ZT, Hale BG, Steel J, Pérez DR, García-Sastre A. 2011. Differential contribution of PB1-F2 to the virulence of highly pathogenic H5N1 influenza A virus in mammalian and avian species. *PLoS Pathog* **7**: e1002186. doi:10.1371/journal.ppat.1002186
- Selman M, Dankar SK, Forbes NE, Jia JJ, Brown EG. 2012. Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerg Microbes Infect* **1**: e42. doi:10.1038/emi.2012.38
- Shi M, Jagger BW, Wise HM, Digard P, Holmes EC, Taubenberger JK. 2012. Evolutionary conservation of the PA-X open reading frame in segment 3 of influenza A virus. *J Virol* **86**: 12411–12413. doi:10.1128/JVI.01677-12
- Shih SR, Nemeroff ME, Krug RM. 1995. The choice of alternative 5' splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex. *Proc Natl Acad Sci* **92**: 6324–6328. doi:10.1073/pnas.92.14.6324
- Shih SR, Suen PC, Chen YS, Chang SC. 1998. A novel spliced transcript of influenza A/WSN/33 virus. *Virus Genes* **17**: 179–183. doi:10.1023/A:1008024909222
- Skehel JJ. 1972. Polypeptide synthesis in influenza virus-infected cells. *Virology* **49**: 23–36. doi:10.1016/S0042-6822(72)80004-7
- Sleigh MJ, Both GW, Brownlee GG. 1979. The influenza virus haemagglutinin gene: cloning and characterisation of a double-stranded DNA copy. *Nucleic Acids Res* **7**: 879–893. doi:10.1093/nar/7.4.879
- Sloan E, Alenquer M, Chung L, Clohisey S, Dinan AM, Gifford R, Gu Q, Irigoyen N, Jones JD, van Knippenberg I, et al. 2019. Upstream translation initiation expands the coding capacity of segmented negative-strand RNA viruses. bioRxiv doi:10.1101/795815
- Takeda M, Pekosz A, Shuck K, Pinto LH, Lamb RA. 2002. Influenza A virus M2 ion channel activity is essential for efficient replication in tissue culture. *J Virol* **76**: 1391–1399. doi:10.1128/JVI.76.3.1391-1399.2002
- Tauber S, Ligertwood Y, Quigg-Nicol M, Dutia BM, Elliott RM. 2012. Behaviour of influenza A viruses differentially expressing segment 2 gene products in vitro and in vivo. *J Gen Virol* **93**: 840–849. doi:10.1099/vir.0.039966-0
- Trifonov V, Racaniello V, Rabadán R. 2009. The contribution of the PB1-F2 protein to the fitness of influenza A viruses and its recent evolution in the 2009 influenza A (H1N1) pandemic virus. *PLoS Curr* **1**: RRN1006. doi:10.1371/currents.RRN1006
- Varga ZT, Ramos I, Hai R, Schmolke M, García-Sastre A, Fernandez-Sesma A, Palese P. 2011. The influenza virus protein PB1-F2 inhibits the induction of type I interferon at the level of the MAVS adaptor protein. *PLoS Pathog* **7**: e1002067. doi:10.1371/journal.ppat.1002067
- Varga ZT, Grant A, Manicassamy B, Palese P. 2012. Influenza virus protein PB1-F2 inhibits the induction of type I interferon by binding to MAVS and decreasing mitochondrial membrane potential. *J Virol* **86**: 8359–8366. doi:10.1128/JVI.01122-12
- Wang P, Palese P, O'Neill RE. 1997. The NPI-1/NPI-3 (karyopherin α) binding site on the influenza A virus nucleoprotein NP is a nonconventional nuclear localization signal. *J Virol* **71**: 1850–1856. doi:10.1128/JVI.71.3.1850-1856.1997
- Wang Q, Li Q, Liu T, Chang G, Sun Z, Gao Z, Wang F, Zhou H, Liu R, Zheng M, et al. 2018. Host interaction analysis of PA-N155 and PA-N182 in chicken cells reveals an essential role of UBA52 for replication of H5N1 avian influenza virus. *Front Microbiol* **9**: 936. doi:10.3389/fmicb.2018.00936
- Wang Q, Liu R, Li Q, Wang F, Zhu B, Zheng M, Cui H, Wen J, Zhao G. 2019. Host cell interactome of PB1 N40 protein of H5N1 influenza A virus in chicken cells. *J Proteomics* **197**: 34–41. doi:10.1016/j.jprot.2019.02.011
- Wanitchang A, Patarasirin P, Jengarn J, Jongkaewwattana A. 2011. Atypical characteristics of nucleoprotein of pandemic influenza virus H1N1 and their roles in reassortment restriction. *Arch Virol* **156**: 1031–1040. doi:10.1007/s00705-011-0947-6



R.M. Pinto et al.

- Watanabe T, Watanabe S, Ito H, Kida H, Kawaoka Y. 2001. Influenza A virus can undergo multiple cycles of replication without M2 ion channel activity. *J Virol* **75**: 5656–5662.
- Wei J, Yewdell JW. 2019. Flu DRiPs in MHC class I immunosurveillance. *Viol Sin* **34**: 162–167. doi:10.1007/s12250-018-0061-y
- Wise HM, Foeglein A, Sun J, Dalton RM, Patel S, Howard W, Anderson EC, Barclay WS, Digard P. 2009. A complicated message: identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. *J Virol* **83**: 8021–8031. doi:10.1128/JVI.00826-09
- Wise HM, Barbezange C, Jagger BW, Dalton RM, Gog JR, Curran MD, Taubenberger JK, Anderson EC, Digard P. 2011. Overlapping signals for translational regulation and packaging of influenza A virus segment 2. *Nucleic Acids Res* **39**: 7775–7790. doi:10.1093/nar/gkr487
- Wise HM, Hutchinson EC, Jagger BW, Stuart AD, Kang ZH, Robb N, Schwartzman LM, Kash JC, Fodor E, Firth AE, et al. 2012. Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. *PLoS Pathog* **8**: e1002998. doi:10.1371/journal.ppat.1002998
- Wise HM, Gaunt E, Ping J, Holzer B, Jasim S, Lycett SJ, Murphy L, Livesey A, Brown R, Smith N, et al. 2019. An alternative AUG codon in segment 5 of the 2009 pandemic influenza A virus is a swine-derived virulence motif. bioRxiv doi:10.1101/738427
- Xu G, Zhang X, Liu Q, Bing G, Hu Z, Sun H, Xiong X, Jiang M, He Q, Wang Y, et al. 2017. PA-X protein contributes to virulence of triple-reassortant H1N2 influenza virus by suppressing early immune responses in swine. *Virology* **508**: 45–53. doi:10.1016/j.virol.2017.05.002
- Yamayoshi S, Watanabe M, Goto H, Kawaoka Y. 2016. Identification of a novel viral protein expressed from the PB2 segment of influenza A virus. *J Virol* **90**: 444–456. doi:10.1128/JVI.02175-15
- Yang N, Gibbs JS, Hickman HD, Reynoso GV, Ghosh AK, Bennink JR, Yewdell JW. 2016. Defining viral defective ribosomal products: standard and alternative translation initiation events generate a common peptide from influenza A virus M2 and M1 mRNAs. *J Immunol* **196**: 3608–3617. doi:10.4049/jimmunol.1502303
- Yewdell JW, Antón LC, Bennink JR. 1996. Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules? *J Immunol* **157**: 1823–1826.
- Yoon S-W, Webby RJ, Webster RG. 2014. Evolution and ecology of influenza A viruses. *Curr Top Microbiol Immunol* **385**: 359–375.
- Yoshizumi T, Ichinohe T, Sasaki O, Otera H, Kawabata S-I, Mihara K, Koshihara T. 2014. Influenza A virus protein PB1-F2 translocates into mitochondria via Tom40 channels and impairs innate immunity. *Nat Commun* **5**: 4713–4714. doi:10.1038/ncomms5713
- Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T, et al. 2009. Crystal structure of an avian influenza polymerase PA_N reveals an endonuclease active site. *Nature* **458**: 909–913. doi:10.1038/nature07720
- Zamarin D, García-Sastre A, Xiao X, Wang R, Palese P. 2005. Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS Pathog* **1**: e4. doi:10.1371/journal.ppat.0010004
- Zamarin D, Ortigoza MB, Palese P. 2006. Influenza A virus PB1-F2 protein contributes to viral pathogenesis in mice. *J Virol* **80**: 7976–7983. doi:10.1128/JVI.00415-06
- Zanker DJ, Oveissi S, Tschärke DC, Duan M, Wan S, Zhang X, Xiao K, Mifsud NA, Gibbs J, Izzard L, et al. 2019. Influenza A virus infection induces viral and cellular defective ribosomal products encoded by alternative reading frames. *J Immunol* **202**: 3370–3380. doi:10.4049/jimmunol.1900070
- Zell R, Krumbholz A, Eitner A, Krieg R, Halbhuber K-J, Wutzler P. 2007. Prevalence of PB1-F2 of influenza A viruses. *J Gen Virol* **88**: 536–546. doi:10.1099/vir.0.82378-0
- Zhirnov OP, Poyarkov SV, Vorob'eva IV, Safonova OA, Malyshev NA, Klenk H-D. 2007. Segment NS of influenza A virus contains an additional gene NSP in positive-sense orientation. *Dokl Biochem Biophys* **414**: 127–133. doi:10.1134/S1607672907030106