Emergence of highly pathogenic avian influenza A(H5N1) virus PB1-F2 variants and their virulence in BALB/c mice.

Ram P. Kamal¹², Amrita Kumar¹², Charles T. Davis¹, Wen-Pin Tzeng¹, Tung Nguyen³, Ruben O. Donis¹, Jacqueline M. Katz¹, and Ian A. York¹,*

¹ Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia; United States of America

² Battelle Memorial Institute, Atlanta, Georgia; United States of America

³ National Centre for Veterinary Diagnostics, Department of Animal Health, Hanoi, Viet Nam

Running title: Prevalence and virulence of H5N1 PB1-F2 variants.

Abstract word count: 235

Text word count: 4557

*Corresponding author. Please direct correspondence to: Ian A York, 1600 Clifton Rd NE, MS-G16, Atlanta, GA 30329, United States of America. Phone: +1 404-639-7286, Email: ITE1@cdc.gov
Influenza A viruses (IAV) express the PB1-F2 protein from an alternate reading frame within the PB1 gene segment. The roles of PB1-F2 are not well understood, but appear to involve modulation of host cell responses. As shown in previous studies, we find that PB1-F2 of mammalian IAV frequently have premature stop codons that are expected to cause truncations of the protein, whereas avian IAV usually express a full-length 90 amino acid PB1-F2. However, in contrast to other avian IAV, recent isolates of highly pathogenic H5N1 influenza viruses had a high proportion of PB1-F2 truncations (15% since 2010; 61% of isolates in 2013) due to several independent mutations that have persisted and expanded in circulating viruses. One natural H5N1 IAV containing a mutated PB1-F2 start codon (i.e., lacking ATG) was 1000-fold more virulent for BALB/c mice than a closely-related H5N1 containing intact PB1-F2. In vitro, we detected expression of an in-frame protein (C-terminal PB1-F2) from downstream ATGs in PB1-F2 plasmids lacking the well-conserved ATG start codon. Transient expression of full-length, truncated (25 amino acids), and PB1-F2 lacking the initiating ATG in mammalian and avian cells had no effect on cell apoptosis or interferon expression in human lung epithelial cells. Full length and C-terminal PB1-F2 mutants co-localized with mitochondria in A549 cells. Close monitoring of alterations of PB1-F2 and their frequency in contemporary avian H5N1 viruses should continue, as such changes may be markers for mammalian virulence.
Although most avian influenza viruses are harmless for humans, some (such as highly pathogenic H5N1 avian influenza viruses) are capable of infecting humans and causing severe disease with a high mortality rate. A number of risk factors potentially associated with adaptation to mammalian infection have been noted. Here we demonstrate that the protein PB1-F2 is frequently truncated in recent isolates of highly pathogenic H5N1 viruses. Truncation of PB1-F2 has been proposed to act as an adaptation to mammalian infection. We show that some forms of truncation of PB1-F2 may be associated with increased virulence in mammals. Our data support the assessment of PB1-F2 truncations for genomic surveillance of influenza viruses.
INTRODUCTION

Highly pathogenic avian influenza viruses (HPAIV) of the H5N1 subtype are now endemic in many countries (1). H5N1 became a public health concern in 1997 when a poultry outbreak in Hong Kong resulted in 18 human infections with six fatalities (2). Subsequent reassortment events with other avian influenza viruses were associated with spread of new genotypes of H5N1 over much of Asia, the Middle East, and Europe, and with localized evolution and antigenic variation of the H5 and N1 surface proteins (3, 4). H5N1 viruses remain capable of infecting humans with high mortality, with over 700 confirmed human cases with a case fatality rate of nearly 60% (5). Unlike seasonal influenza A viruses (IAV) that circulate among humans, H5N1 HPAIVs do not readily transmit from human to human. One of the key reasons for inefficient human-to-human transmission of H5N1 viruses is the preference of their hemagglutinin (HA) for α-2, 3-linked sialic acids that are uncommon in the upper respiratory tract of humans (6). In addition to HA, other IAV genes may also contribute to transmission and virulence.

The PB1 gene segment is of particular interest, since this was the only internal gene from an avian source that reassorted with the surface glycoprotein genes, HA and neuraminidase (NA) in the pandemic virus of 1957 and with the HA in the 1968 pandemic virus (7). One possible factor in the role of PB1 in pathogenicity and transmission is the presence of the PB1-F2 protein, encoded by a +1 reading frame (8), which was present as a full-length protein in the 1918, 1957, and 1968 human pandemic viruses but not in the 2009 H1N1 pandemic virus.

PB1-F2 is a short (87–90 amino acid) IAV protein discovered serendipitously in 2001 during the search for CD8 T cell epitopes encoded by alternative reading frames (8). It was found to localize in mitochondria and cause cell death (8, 9). It is dispensable for viral replication
but has been associated with viral pathogenesis in mice and ducks (8, 10-13), and truncation of PB1-F2 in the 2009 pandemic influenza virus (H1N1pdm09) is hypothesized to be one cause of the less severe pathology and relatively low case fatality rate of this pandemic compared to previous pandemics (14, 15). The effect of PB1-F2 on viral phenotype has been demonstrated to be strain-specific (16-18), particularly among viruses with a PB1 gene of recent avian origin (10). This may be because other pathogenic determinants are able to mask the influence of PB1-F2 (11).

The mechanisms underlying the pathogenic effects of PB1-F2 remain unclear. Several possible mechanisms have been suggested, including mitochondrial targeting and pro-apoptotic activity (8, 9, 19), enhancing PB1 function (20), and inhibition (21-27), or enhancement (28) of the IFN-β response. PB1-F2 has also been linked to bacterial infections following IAV infection (29-32). However, most of these phenotypes have been difficult to generalize to all strains of virus and across host models tested.

A striking feature of PB1-F2 among global IAV is the frequent presence of premature stop codons. For example, almost all 2009 H1N1pdm viruses have truncated 11 amino acid PB1-F2 proteins. In 2007, analysis of publicly-available PB1-F2 sequences indicated that a large proportion (19%) of mammalian IAVs had truncated PB1-F2, but such truncations were rare among avian IAVs (only 4%) (33). Similarly, in 2013, Pasricha et al. determined that only 5% of avian IAV analyzed contained truncated PB1-F2, compared to 41% of human and swine IAV (34). These observations suggest that truncation of PB1-F2 may play a role in adaptation of IAVs to mammalian hosts.
Although PB1-F2 truncations have been sporadically observed in avian influenza viruses, until recently it has been unusual for groups of related avian influenza virus subtypes to persistently possess truncated variants. However, since 2009, an increasing number of H5N1 highly-pathogenic avian influenza virus (HPAIV) lineages have maintained PB1-F2 truncations over several years. Based on previous suggestions that viruses with these changes may be better adapted to mammalian replication, we tested these viruses for their ability to cause disease in mice and to alter cellular functions in mammalian and avian cells.
MATERIALS AND METHODS

Database analysis of PB1-F2 truncations. We retrieved full-length PB1 sequences with collection dates prior to 2014 from GISAID (Supplementary Table 1) and the NCBI Influenza databases (as of Dec. 31, 2013), and merged the sets, eliminating duplicates based on virus strain name. The host species (human, swine, or avian) and year of collection was also noted. Highly-pathogenic H5N1 viruses were considered as a separate subset and were excluded from the avian and human subsets. PB1 sequences were aligned using MAFFT (35), and PB1-F2 sequences were translated from each PB1 gene. The numbers of truncated versions present in each subset, per year of collection, were calculated. In cases where no initiating ATG was present at the appropriate position, the length of PB1-F2 was considered to be 0.

Distribution of PB1-F2 truncations in H5N1 viruses. For H5N1 viruses with PB1-F2 truncations, we obtained the corresponding hemagglutinin (HA) sequence from the appropriate database, annotating HA sequences with the length of the truncated PB1-F2. We also included HA sequences from the report of WHO/OIE/FAO H5N1 Evolution Working Group (36) in order to identify clades. We aligned HA sequences using MAFFT and constructed a maximum likelihood tree (generalized time-reversible model of nucleotide evolution, gamma, 10,000 bootstraps) with FastTree (37), using FigTree (38) for annotations.
Identification of related PB1-F2 full-length/truncation H5N1 viruses. We analyzed the set of virus isolates available to us to identify a pair of similar viruses differing in PB1-F2 length. Viral RNAs were isolated by using QIAamp Viral RNA Mini Kit (#52904, QIAGEN Inc. – USA), reverse transcribed and amplified by AccessQuick™ RT-PCR System (#A1702, Promega Corporation, USA). Sequencing reactions were performed by using BigDye® Terminator v3.1 Cycle Sequencing Kit (#4337456, Life Technologies, Grand Island, NY) and run in Applied Biosystems 3730xl DNA Analyzer (Life Technologies). RT-PCR and sequencing primers are available upon request. For each H5N1 virus containing a PB1-F2 truncation, the amino acid sequence of each viral protein was compared to the most closely related virus containing full-length PB1-F2.

Preparation of viral stocks. Clade 2.3.4.2 H5N1 viruses A/chicken/Vietnam/NCVD-281/2009 ("VN/281") (GISAID isolate ID EPI_ISL_80619) and A/chicken/Vietnam/NCVD-296/2009 ("VN/296") (EPI_ISL_80637) were grown in 10-11 days old embryonated chicken eggs incubated at 37°C for 24 hours. Allantoic fluids from eggs, infected with the highest dilution of virus inoculum that gave a hemagglutination inhibition titer of 256 or more, were harvested, clarified by centrifugation and frozen at -80°C as viral stocks.

To determine the 50% egg infectious dose (EID₅₀) of virus stocks, dilutions (10⁻⁵ to 10⁻¹⁰) were made from each virus and inoculated into five 10-11 days old embryonated hen eggs per dilution. Eggs were incubated at 37°C for 24 hours and then chilled overnight at 4°C. Egg infections were detected by standard hemagglutination assays and EID₅₀ was calculated by the method of Reed and Muench (39).
**Plaque assay.** Confluent MDCK-London cells were infected with six dilutions (10^{-5} to 10^{-10}) of each virus, overlaid with 0.8% agarose medium and incubated at two different temperatures (37°C and 40°C). 72 hours post infection, agarose was gently removed from wells and cell monolayers were stained with crystal violet-formalin mix. Plaques were counted in appropriate wells and PFU titers were calculated.

**Pathogenicity in mice.** All animal experiments were done in Animal Biosafety Level 3 (ABSL3E) facility with enhancements required by the U.S. Department of Agriculture and the Select agent program (40) under CDC’s IACUC-approved protocols. 6-8 weeks old female BALB/cJ mice (Jackson Laboratory, Bar Harbor, ME) were used in this study. 5 mice each were intranasally infected with one of 5 virus doses (10^0 to 10^4 PFU in 100 µl volume) of each virus; 5 mice were mock infected with PBS. Mice were observed daily for weight loss and other clinical signs for a period of 16 days. Mice were euthanized if they lost 25% or more of their original body weight, or if neurological signs were observed. Mouse LD50 values were calculated using the Reed and Muench method (39).

**Cloning and mutagenesis of PB1-F2:** Wild type PB1-F2 from VN/281 and VN/296 were RT-PCR amplified, using oligonucleotide primers with added KpnI and XbaI restriction sites. Amplified products were cloned into a pTracer-CMV2-GFP vector (Life Technologies, Grand Island, NY). Additional PB1-F2 clones were made with insertion of C-terminal HA epitope tags. Mutants of cloned PB1-F2 (summarized in Figure 1 and Table 1) were made by site-directed mutagenesis using Quick Change Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) and
mutagenesis primers. Cloned PB1-F2 plasmids were amplified in competent *E. coli* cells and were purified using Qiagen plasmid purification kit (QIAGEN Inc, Valencia, CA). The sequences of all plasmid clones were confirmed before use in experiments.

**Confocal microscopy:** Post transfection A549 cells were washed with PBS, fixed in 4% paraformaldehyde for 15 min, permeabilised with 0.5% Tween-20 in PBS for 10 min, washed with PBST (0.2% Tween-20), blocked with 2% BSA in PBST for 2 h. After blocking cells were incubated with HA-Tag (C29F4) Rabbit mAb (1:1000) (Cell signaling) and Anti-MTCO2 antibody (ab91317) (1:2000)(Abcam) overnight at 4°C. Then cells were washed (2.5% FBS in PBST, 5 min ×6), and incubated with Alexa Fluor® 546 Donkey Anti-Mouse IgG (1:1000) and Alexa Fluor® 633 Goat Anti-Rabbit IgG (H+L) (1:1000)(LifeTechnologies), for 2 hours at RT. Non-specific secondary antibodies were removed by washing with PBST (5 min, ×6), followed by a final soak in PBS. Nuclei were counterstained with DAPI and cover slips mounted with Prolong gold antifade mounting media (Invitrogen). Since the plasmids expressing PB1-F2 expressed GFP under a separate promoter, we limited analysis to cells expressing GFP (green). Images were captured using Zeiss invert confocal microscope LSM 710 with X63 objective. Images were processed using Zen 2010 (Zeiss) and Adobe Photoshop (Adobe Inc).

**Quantitative RT-PCR for IFNβ expression:** Human A549 (#CCL-185, ATCC) and chicken DF1 cells (#CRL-12203, ATCC), grown in 24-well cell culture plates were transfected with 1.0 µg of plasmid DNA using TransIT-LT1 transfection reagent (Mirus Bio, Madison, WI). As a positive control, cells were treated with in-vitro transcribed 5’ tri-phosphate RNA (TP-RNA)
Cells were lysed and lysates were used directly for qRT-PCR using CellsDirect One-Step qRTPCR Kit (Life Technologies, Grand Island, NY) and TaqMan gene expression assays for human and avian IFNβ (Life Technologies, Grand Island, NY). The ΔΔCq method in Bio-Rad CFX Manager 2.1 software was used to analyze the expression data. GAPDH expression was measured in each sample as a control.

**Flow cytometry for cell apoptosis:** A549 and DF1 cells were transfected as described above. As a positive control, cells were treated with camptothecin (SIGMA-ALDRICH, St. Louis, MO, USA) at final concentration of 20 µM for 24hr. After harvesting, cells were stained for apoptosis markers using Violet Chromatin Condensation/Dead Cell Apoptosis Kit with Vybrant® DyeCycle™ Violet and SYTOX® AADvanced™ for Flow Cytometry (Life Technologies, Grand Island, NY) according to manufacturer’s instructions. Flow cytometry data were acquired within one hour of staining. Data were acquired on FACSCanto II flow cytometer (BD, Franklin Lakes, NJ), and analyzed using FlowJo software (Tree Star, Ashland, OR).

**Statistical analyses.** To compare weight loss, a linear mixed model with repeated measures was used, using a cutoff of p<0.05 for statistical significance. Kaplan-Meier survival curves were compared by the log-rank test. Effect on apoptosis and IFNβ expression were analyzed by the Kruskal-Wallis test using GraphPad Prism software (La Jolla, California USA).
RESULTS

Evolution of PB1-F2 truncations in avian and mammalian viruses. Previous studies (33, 34) have shown that a large fraction of mammalian influenza viruses express truncated PB1-F2. In contrast, avian influenza viruses usually express full length PB1-F2 of 90 amino acids. To assess changes in this trend since these previous studies, we retrieved 21,984 PB1 sequences that included the full PB1-F2 region from the combined GISAID and NCBI Influenza databases (removing duplicates based on strain name) as of Dec. 31, 2013: 10,498 human, 1,966 HPAIV H5N1, 7,392 avian (excluding H5N1), and 2,128 swine sequences. We identified 23 different truncations of PB1-F2, resulting in PB1-F2 lengths ranging from 0-87aa. A small number of viruses (53, 0.24% of total) had PB1-F2 proteins that were 101aa; most of these were human H3N2 viruses from 1997-1998 (24 viruses) and non-H5N1 avian viruses from the 2000s (16 viruses) (data not shown). 64.3% of all PB1-F2s were considered to be functionally full length (≥79 aa (33)). The prevalence of PB1-F2 truncations (78 amino acids or shorter) was 63.1% for human viruses (H1N1 and H3N2), 6.7% for H5N1, 3.8% for avian viruses excluding H5N1, and 39.7% for swine viruses. Prevalence of PB1-F2 truncations by year of collection, subtype and host is shown in Figure 2, and a summary of the PB1-F2 truncations in H5N1 viruses is shown in Table 2.

The most common truncations were 11 and 57 amino acids (23.2% and 7.6% of total, respectively), with no other truncation exceeding 1% of the total. Swine influenza viruses include multiple subtypes (reviewed in (42)), and PB1-F2 truncations were most common in the H1N1 subtypes (data not shown). Among avian viruses, 57 and 79-residue versions were more
prevalent before 2005, after which variants containing 24 or 25aa became more prevalent. (We
group these together since in H5N1 viruses the 24aa variant arose by adding an extra stop codon
to the 25aa variant.) Similar mutants have arisen sporadically in swine and human viruses since
2001 and 2007, respectively. The prevalence of these variants has been less than 3% in avian,
human, and swine viruses separately and combined.

Evolution of PB1-F2 truncations in H5N1 viruses. HPAIV H5N1 viruses have been
categorized into multiple clades based on their HA sequences (36). We constructed phylogenetic
trees, based on HA sequence (Figure 3 and Supplemental Figure 1), to analyze the distribution of
PB1-F2 truncations in H5N1 clades. Until 2009, PB1-F2 truncations in H5N1 viruses were rare
(~2% of sequences, Figure 2 and Table 2), with only a few instances in which truncations
persisted in a particular lineage for more than several months (e.g., clade 1 in Thailand from
2007-2008; see below). However, in 2009, 15.6% of the H5N1 viruses had truncated PB1-F2
(Fig. 1 and Table 2). In particular, clade 2.3.4.1 viruses isolated in 2009 and 2010 in Vietnam
and China consistently had a PB1-F2 truncation of 24 or 25 amino acids. These viruses included
both poultry isolates and human cases (A/Guizhou/1/2009, A/Hunan/2/2009). In 2010 and 2011,
clade 2.3.4.1 was effectively replaced in Vietnam by clade 2.3.2.1 and has not been isolated from
Vietnam since 2010. In 2011/2012 a new PB1-F2 truncation of 25 amino acids arose in the clade
1.1.1 lineage in Vietnam. Since PB1 in the clade 1.1.1 viruses is phylogenetically distinct from
that of the 2.3.4.1 lineage (data not shown) this mutation appears to have arisen independently
and not via reassortment. To date, these truncations in the clade 1.1.1 lineage have only been
detected in viruses from poultry. A number of other viruses containing a 24/25 aa truncation of
PB1-F2 have been detected (Table 2), but these appear to have been sporadic mutants with short-lived circulation.

A PB1-F2 truncation of 57 amino acids arose in the clade 2.3.2.1b lineage in China and Vietnam in 2010 and this variant has persisted through 2013, with the majority of 2.3.2.1b viruses containing a truncated PB1-F2. Furthermore, in 2012 a reassortment event occurred among 2.3.2.1 viruses circulating in Vietnam, in which viruses with HA from 2.3.2.1c acquired the internal genes including PB1 from 2.3.2.1b viruses (43). Since the 2.3.2.1b PB1 included the 57-aa truncation of PB1-F2, most H5N1 viruses isolated in Vietnam in 2012 and 2013 contained this PB1-F2 variant (Figure 3 and Table 2). Other sporadic instances of 57-aa truncations of PB1-F2 arose independently in several years, including in 2 of the 1997 viruses, but were not detected over multiple years.

Viruses in which the initiating ATG for PB1-F2 was mutated were detected sporadically from 2006 to 2012, representing 0.3-3% of sequenced H5N1 PB1-F2 (Table 2). Although limited persistence of these variants was observed (e.g. in Clade 2.3.4 [Figure 3]) no lineages of viruses containing these variants were detected for multiple years.

Finally, a number of less common PB1-F2 truncations have been isolated in HPAIV H5N1, including variants that are 63, 34, 11, and 8 amino acids in length. However, since none of these variants have been isolated in a subgroup of viruses for more than one year, we consider them to be sporadic mutations.

**Construction of PB1-F2 truncation mutants.** To assess the effects of H5N1 PB1-F2 truncations *in vitro*, we cloned PB1-F2 from A/chicken/Vietnam/NCVD-296/2009 (“VN/296”),
which lacks the initiating ATG, called henceforth as F2NO. We made a mutant by restoring the
initiating ATG to express a full-length (90 aa) protein (F2FL), and another mutant by introducing
mutations into this gene to produce a 24 aa protein (F2TR) (Figure 1 and Table 1). Each variant
was constructed with and without an HA tag at the C-terminus (Figure 1 and Table 1) to
facilitate detection by antibodies. In the case of the 24aa truncation, variants were constructed in
which the HA tag was placed at the C-terminus of the anticipated truncated protein (F2HA-TR),
or at the C-terminus of the full-length (90aa) protein (F2TR-HA) to detect expression from in-
frame initiation sites (Figure 1 and Table 1). Similar PB1-F2 constructs were made from the
paired virus VN/281 (Table 1).

**PB1-F2 mutants had no effect on apoptosis.** Apoptosis has been associated with expression of
PB1-F2 from PR8 and 1918 H1N1 strains (10, 44), but not with PB1-F2 from H5N1, 1957 or
1968 pandemic viruses (10). We transfected A549 (human) and DF1 (chicken) cells with PB1-F2
variants, harvested cells at 12, 24, 36 and 48 hours post transfection, stained for apoptosis and
cell death, and analyzed the cells by flow cytometry. Neither the full-length nor the truncated or
ATG-less versions of PB1-F2 affected the proportion of apoptotic or dead cells compared to
control cells (p > 0.05, Kruskal-Wallis test) (Figure 4A).

**PB1-F2 truncations and IFNβ expression.** PB1-F2 from mouse-adapted viruses alters type I
interferon production in some cell types (22-24, 28). We transfected A549 and DF1 cells with
the PB1-F2 plasmids and measured IFNβ expression at various time points using the TaqMan
assay qRT-PCR. We used empty plasmid vector as a negative control and in-vitro transcribed 5’
tri-phosphate RNA (TP-RNA) as a positive control for IFNβ induction (41). IFNβ expression levels were normalized to GAPDH and expressed as fold induction over empty vector transfection. Treatment with TP-RNA consistently induced IFNβ expression at least 100-fold over CMV2 vector transfection. However, transfection with full-length or truncated PB1-F2 did not significantly alter IFNβ expression A549 cells and DF1 cells (p > 0.05, Kruskal-Wallis test) (Figure 4B).

Full length and C-terminal PB1-F2 from H5N1 viruses localizes to mitochondria. We evaluated the expression and cellular co-localization of each of the PB1-F2 variants containing a C-terminal HA epitope tag after transient transfection in A549 cells and using immunofluorescence. Since the plasmids expressing PB1-F2 also expressed GFP under a separate promoter, we limited analysis to cells expressing GFP (green). Confocal microscopy revealed that the full-length PB1-F2 proteins were expressed and readily detected (Figure 5). The full length PB1-F2 localized with cell mitochondria in A549 cells (Figure 5, f-j). The truncated 24-amino acid version of PB1-F2 was not detectable (Figure 5, k-o) even after treating the cells with the proteasome inhibitor MG-132 (Figure 5o). Interestingly, the presence of an HA-tagged protein was detectable by confocal microscopy in cells transfected with the ATG-less version of PB1-F2; this was expressed at comparable levels to that of full length PB1-F2, and also localized to mitochondria (Figure 5, u-y). Since the HA epitope tag is in-frame and at C-terminal (Figure 1), this protein expression presumably originated from an in-frame ATG (residue number 39, 46, or 51; Figure 1) downstream of the authentic start site, leading to expression of an N-terminally truncated version of the protein (c-terminal PB1-F2), as previously described (13). Alternatively, expression of full-length PB1-F2 using ACG as an initiation codon is possible. However,
western blots did not detect expression of a full-length PB1-F2 in cells transfected with this plasmid (data not shown). No protein expression was detected in two truncated variants in which the initiating ATG was intact, two stop codons were inserted at positions 25 and 26, and one had HA tag at position 25 (F2HA-TR) and another at the c-terminal (F2TR-HA).

Paired viruses for PB1-F2 studies. Although several studies have tested the effect of PB1-F2 truncations on mouse virulence, these studies introduced stop codons into the PB1-F2 ORF instead of, or as well as, mutating the initiating ATG (10, 11, 13, 25). Since confocal microscopy suggested that mutating the start codon, but not introducing stop codons, allowed abundant expression of a protein from the PB1-F2 ORF, we tested the virulence of H5N1 viruses with full-length or ATG-less PB1-F2 in mice. VN/281 and VN/296 are closely-related clade 2.3.4.2 H5N1 viruses, isolated from chickens in Vietnam, with disparate PB1-F2 sequences. VN/281 has full length PB1-F2, while VN/296 has an ATG to ACG mutation in the PB1-F2 start codon. Aside from PB1-F2, these viruses differ by only 16 of 4,440 total amino acids across the viral proteome (0.34%) (summarized in Table 3). None of these amino acid differences are known determinants of pathogenicity in H5N1 viruses (45).

Both the viruses VN/281 and VN/296 replicated to similar titers in eggs and cells, as measured by EID₅₀/ml (10⁹.⁵ and 10⁸.⁸, respectively) and plaque assay (10⁸.⁹ and 10⁸.⁴ PFU/ml, respectively).

Absence of PB1-F2 initiating ATG is associated with increased mice virulence. Mice were infected with various doses (10⁰ to 10⁴ pfu) with either VN/281 or VN/296. The two viruses
showed marked differences in their virulence in mice. Mice infected with VN/281 (full-length PB1-F2) showed weight loss only at the highest dose (10^4 pfu) and yielded a MLD50 of 10^{4.3} pfu (Figure 6). In contrast, a majority of mice infected with VN/296 virus (lacking the start codon for PB1-F2) showed marked weight loss at all doses >10^0. VN/296 was much more lethal than VN/281, with the LD50 of virus VN/296 (10^{1.4} PFU) being approximately 1000-fold lower than VN/281 virus.
DISCUSSION

Previous studies have shown that the great majority of avian influenza viruses contain a full-length PB1-F2 protein, whereas mammalian viruses frequently have premature stop codons resulting in truncation (33, 34). Incorporating more recent sequences obtained from influenza sequence databases, we found a similar picture. In mammalian viruses (swine and human) PB1-F2 truncations (≤78aa) are common (approximately 63% of sequences in the database). In contrast, in avian viruses over 95% of PB1-F2 sequences are full length. An important exception to this is in the highly pathogenic H5N1 virus group, in which the frequency of PB1-F2 truncations in databases increased to 61% in 2013. In addition, non-H5N1 avian viruses in general also showed an increase in PB1-F2 truncation frequency in 2013, to about 20%. This increase was mainly due to H7N9 and H9N2 AIV in China, in which 24% and 21% of PB1-F2, respectively, were truncated.

It is important to note that these percentages are inevitably skewed due to sampling bias. For example, human H1N1pdm viruses (almost all of which have truncated PB1-F2) were intensively sequenced in 2009. Similarly, the published sequences of H5N1 internal genes are dominated by those from Vietnam and China, where PB1-F2 truncations are increasingly common in endemic, circulating viruses. Nevertheless, the frequent appearance and evolutionary success of viruses bearing truncations in PB1-F2 appears to be a recent phenomenon.

Although multiple studies have demonstrated a role for PB1-F2 in viral virulence (10, 11, 13, 25), the molecular mechanisms involved are not clear. Most studies on the effect of PB1-F2 on virulence have used mouse-adapted strains of H1N1 viruses (A/Puerto Rico/08/1934(H1N1) (PR8) and WSN/33). The effects of PB1-F2 are dependent on genomic context and the presence
of other virulence factors (10, 11, 16-18). Only two studies have looked at the effect of PB1-F2 on H5N1 pathogenicity, and both found that PB1-F2 truncations slightly reduced lethality in mice (12, 46).

Although PB1-F2 has been shown to affect apoptosis, most studies on PB1-F2 effects have used the gene from the mouse-adapted strains PR8 or WSN/33, and the limited studies on PB1-F2 from H5N1 virus have not supported a role in apoptosis (44). Similarly, PB1-F2 in H5N1 was not found to affect IFN levels in most cell types (12). In full-length PB1-F2, the presence of serine rather than arginine at position 66 (N66S) is associated with increased virulence and inhibition of IFN (21, 24, 47), while the PB1-F2 used in our experiments has 66N. It is important to note that the proapoptotic and IFNβ regulatory roles of PB1-F2 are specific for both virus strain and cell type. These effects have been most consistently (though not solely) seen with PB1-F2 from lab-adapted viruses such as A/Puerto Rico/8/1934 and A/WSN/1933 cells of immune origin. However, PB1-F2 from H5N1 viruses has shown no effect either on IFNβ expression (12) or on apoptosis (44), even in cells of immune origin. Consistent with these observations, in our experiments neither full-length, truncated, nor ATG-mutated versions of PB1-F2 from H5N1 affected apoptosis or IFNβ expression in human or chicken cells.

Previous studies on the effect of PB1-F2 truncation or deletion have modified PB1-F2 by mutating multiple downstream ATGs along with the 1\textsuperscript{st} start codon (12, 28, 46), by introducing a stop codon after the last in-frame ATG (13), or by both mutating the 1\textsuperscript{st} start codon and inserting a stop codon at position 11 (10, 11, 20, 25). Importantly, we observed abundant expression of an in-frame C-terminal fragment of PB1-F2 (as previously described by Zamarin et al (13)) from a plasmid lacking the initiating ATG but without downstream stop codons, as seen in natural H5N1 isolates. In contrast, only very low level, or no, expression of in-frame proteins...
were detected from PB1-F2 plasmids containing the initiating ATG with premature stop codons (Figure 5p-t). This C-terminal fragment presumably was not expressed in the previous studies that inserted one or more stop codons into the open reading frame. This C-terminal fragment, like the full-length PB1-F2, localized to mitochondria, consistent with the mitochondrial targeting signal in PB1-F2 being entirely present in the C-terminal region (9, 48).

To test the effect on virulence of the start codon mutation, we used a pair of naturally-occurring viruses, VN/281 and VN/296, which differ by only 0.34% (16 amino acid differences out of 4,440 in the proteome). Although the viruses replicated similarly in eggs and in MDCK cells, their 50% mouse lethal dose (MLD50) differed markedly, with VN/281 (full length PB1-F2) being 1000-fold less pathogenic than VN/296 (ATG-less PB1-F2). Thus, while complete deletion of PB1-F2 was found to reduce H5N1 virulence (12, 46), expression of the C-terminal fragment instead of the full-length protein is correlated with increased lethality in mice.

Previous studies have raised the concern that avian viruses containing truncated PB1-F2 may be “pre-adapted” to mammalian infection. Despite the increased prevalence of PB1-F2 truncated H5N1 mutants (i.e., clade 2.3.2.1 viruses in Vietnam and China), these viruses continue to circulate in poultry populations and cause outbreaks suggesting viral replication is not dramatically altered in the avian host. Expression of a C-terminal version of PB1-F2 may be associated with increased virulence. Further experiments will be needed to determine the mechanism(s) underlying the differential virulence identified, since no effects on apoptosis and Type I IFN induction were observed. The study justifies continued monitoring of changes in the prevalence of H5N1 viruses carrying truncated PB1-F2 proteins, including those containing start codon deletions, for their public health significance.
ACKNOWLEDGEMENTS

The authors thank the CDC Animal Resource Branch for excellent animal care. We thank Dr. Vilma Yuzbasiyan-Gurkan and administrators of the Comparative Medicine and Integrative Biology at Michigan State University for their unfailing support of RPK during his Ph.D. program and ORISE for supporting RPK as a guest researcher at CDC.

The findings and conclusions in this report are those of the authors and do not necessarily reflect the views of the funding agency.
REFERENCES


**Figure legends**

**Figure 1. Schematic diagrams of PB1-F2 mutants.** Seven PB1-F2 variants from VN/296 (as shown here) and VN/281 (not shown) were cloned into expression plasmids. Red text indicates stop codons or mutation of the initiating start codon. “HA” indicates an in-frame C-terminal HA epitope tag. All in-frame ATGs in the PB1-F2 ORF are indicated. “F2FL”, full-length PB1-F2, 90aa. “F2NO”, 1st ATG is mutated to ACG as in wt VN/296. “F2TR”, truncated (24 amino acids) due to stop codons at position 25 and 26. Please note that there are two variants of F2TR with HA tag. “F2TR-HA” has an HA tag at the end of the protein so can only detect C-terminal protein expressed from in-frame downstream ATGs. F2HA-TR has the tag immediately upstream to the double stops (25, 26), thus will be used to detected N-terminal protein (25aa).

**Figure 2. Prevalence of PB1-F2 truncations.** PB1 sequences from human, swine, and avian (not including H5N1) influenza viruses, and from highly-pathogenic H5N1 influenza viruses, were retrieved from NCBI, GISAID and CDC databases. Human H1N1 viruses include both seasonal H1N1 and 2009 H1N1pdm viruses. Sequences were aligned using MAFFT and PB1-F2 sequence lengths were inferred from the sequences. Sequences were sorted by year of isolation after 2000. Because sequences were less abundant pre-2000, sequences from 1980-1999 were pooled in 5-year groups, and all sequences pre-1980 were pooled. Data are shown as percent of truncated PB1-F2 among all sequences for that host and year.

**Figure 3. Evolutionary relationship of PB1-F2 truncations in H5N1 viruses.** HA gene sequences of H5N1 viruses were aligned using MAFFT and a maximum-likelihood tree was determined using FastTree. The virus strains having PB1-F2 truncations of 24/25, 57 and 8 aa are colored blue, green and fuchsia, respectively. Viruses with mutated initiating ATG are
colored red. Clades were defined by taking representative sequences from the WHO/OIE/FAO H5N1 Classification Working Group.

**Figure 4. PB1-F2 mutants do not affect cell apoptosis or interferon response.** A549 cells were transfected with either empty vector or PB1-F2 mutant plasmids. Cells were harvested at 12 hr (a), 24 hr (b), 36 hr (c) and 48 hr (d) post transfection. (A) Cells were stained using dyes specific for apoptosis and necrosis. Data represent dead + apoptotic cells as fold change over CMV2 vector (control). Significant apoptosis induction was detected in positive control cells (treated with camptothecin) at all time points. (B) IFNβ expression was measured using TaqMan assay and qRT-PCR. IFNβ expression is normalized to GAPDH (internal control) and presented as fold increase relative to CMV2-vector transfection. For positive control, cells were transfected with tri-phosphate RNA (TP-RNA) and harvested at the indicated time points. Statistical significance was determined by Kruskal-Wallis test (all groups) and Mann-Whitney U test (mock and positive control pair). Bar graphs represent means with standard deviation as error bars.

**Figure 5. PB1-F2 expression and cellular localization.** A549 cells were transfected with either control plasmids ("Vector") or PB1-F2-HA plasmids. In some cases (e,j,o,t,y), MG132 (proteasome inhibitor) was added to the medium at 30 hr post-transfection, for 6 hr. 36 hr post transfection, cells were fixed, permeabilized and stained using anti-HA antibody (yellow), anti-MTCO2 (mitochondrially encoded cytochrome c oxidase II) (red, to indicate mitochondria), or DAPI (blue, nucleus) and examined by confocal microscopy. Transfected cells were identified by GFP expression (green), expressed under a separate promoter on the plasmid vector. a-e, untransfected cells as negative control; f-j, F2FL-HA; k-o, F2HA-TR (N-term 24aa); p-t, F2TR-
HA (C-term); u-y, F2NO-HA (initiating ATG to ACG). The first column shows the merge of all stains at low resolution. The 2\textsuperscript{nd} and 3\textsuperscript{rd} columns show mitochondria and HA (marker for PB1-F2), respectively. The 4\textsuperscript{th} and 5\textsuperscript{th} columns show the merge of mitochondria and PB1-F2 in the absence and presence of MG132, respectively.

**Figure 6. Mouse virulence of paired viruses with and without PB1-F2.** Five mice each were infected intranasally with VN/281 or VN/296 at various doses as indicated on the left, or with PBS as a control. Average survival (A-E) and body weight (F-J) of mice are shown as percent of day 0 values. Statistical significance of the difference between the VN/281 and VN/296 groups (survival: log-rank test; body weight: linear mixed model with repeated measures) is indicated on each chart.
Fig. 2.

Truncated PB1-F2 (Percent of host total)

Date of collection

- Pre-1980
- 1980-1984
- 1985-1989
- 1990-1994
- 1995-1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013

- Avian (excluding H5N1)
- Swine
- Human H1N1
- HPAIV H5N1
Fig. 6

**A** 10^0 PFU  
Percent survival  
100  
80  
60  
40  
20  
0  

**B** 10^1 PFU  
Percent survival  
100  
80  
60  
40  
20  
0  

**C** 10^2 PFU  
Percent survival  
100  
80  
60  
40  
20  
0  

**D** 10^3 PFU  
Percent survival  
100  
80  
60  
40  
20  
0  

**E** 10^4 PFU  
Percent survival  
100  
80  
60  
40  
20  
0  

**F**  
Weight (% of day 0)  
110  
100  
90  
80  
70  
60  

**G**  
Weight (% of day 0)  
110  
100  
90  
80  
70  
60  

**H**  
Weight (% of day 0)  
110  
100  
90  
80  
70  
60  

**I**  
Weight (% of day 0)  
110  
100  
90  
80  
70  
60  

**J**  
Weight (% of day 0)  
110  
100  
90  
80  
70  
60  

**Legend:**  
- Mock  
- VN/281  
- VN/296  

**P-values:**  
- **A**: p=0.317  
- **B**: p=0.015  
- **C**: p=0.0625  
- **D**: p=0.0585  
- **E**: p=0.00231  
- **F**: p=0.2284  
- **G**: p<0.0001  
- **H**: p=0.0544  
- **I**: p=0.0005  
- **J**: p=0.0302
TABLE LEGENDS

Table 1. Details of PB1-F2 mutants made in this study. A total of 6 mutants were made from each virus: 3 with C-terminal HA tag and 3 without HA tag. See Figure 3 for schematic diagrams.

* Stop codons were added at positions 25 and 26; and the plasmid is expected to express a truncated protein (24aa). Half of these mutants have HA tag immediately upstream to the stop codons.

** In this variant, only the first ATG is mutated to ACG, leaving the downstream in-frame ATGs intact.

***This variant has stops at positions 25 and 26 but HA tag is at C-terminal.

Table 2. Summary of PB1-F2 truncations in H5N1 viruses. PB1 sequences from highly-pathogenic H5N1 influenza viruses were retrieved from NCBI, GISAID and CDC databases. Sequences were aligned using MAFFT and PB1-F2 sequence lengths were inferred from the sequences. Sequences were sorted by year of isolation and PB1-F2 protein lengths. The sequences expressing a protein of 78aa or smaller were annotated as truncated.
Table 3. Amino acid differences between viruses VN/281 and VN/296. Amino acid differences for each protein were calculated from nucleotide sequence of the respective ORFs using “Cubit” application in “BioEdit” software.
Table 1. Details of PB1-F2 mutants made in this study.

<table>
<thead>
<tr>
<th>Source virus</th>
<th>Name</th>
<th>PB1-F2</th>
<th>Mutagenesis Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>VN/281</td>
<td>VN/281/PB1-F2FL</td>
<td>Full-length (90aa)</td>
<td>Kpn1PB1F76-94 and Xba1PB1R384-67</td>
</tr>
<tr>
<td></td>
<td>VN/281/PB1-F2TR</td>
<td>Truncated* (24aa)</td>
<td>PB1C167T_C170T_F</td>
</tr>
<tr>
<td></td>
<td>VN/281/PB1-F2NO</td>
<td>Deleted**</td>
<td>PB1T96C_F</td>
</tr>
<tr>
<td></td>
<td>VN/281/PB1-F2FL-HA</td>
<td>Full-length (90aa)</td>
<td>Kpn1PB1F68-90 and Xba1HAtagPB1R364-43</td>
</tr>
<tr>
<td></td>
<td>VN/281/PB1-F2TR-HA</td>
<td>Truncated***</td>
<td>PB1C167T_C170T_F and Xba1HAtagPB1R364-43</td>
</tr>
<tr>
<td></td>
<td>VN/281/PB1-F2NO-HA</td>
<td>Deleted**</td>
<td>PB1T96C_F</td>
</tr>
<tr>
<td>VN/296</td>
<td>VN/296/PB1-F2FL</td>
<td>Full-length (90aa)</td>
<td>PB1C96T_F</td>
</tr>
<tr>
<td></td>
<td>VN/296/PB1-F2TR</td>
<td>Truncated (24aa)</td>
<td>PB1C167T_C170T_F</td>
</tr>
<tr>
<td></td>
<td>VN/296/PB1-F2NO</td>
<td>Deleted**</td>
<td>Kpn1PB1F76-94 and Xba1PB1R384-67</td>
</tr>
<tr>
<td></td>
<td>VN/296/PB1-F2FL-HA</td>
<td>Full-length (90aa)</td>
<td>PB1C96T_F</td>
</tr>
<tr>
<td></td>
<td>VN/296/PB1-F2HA-TR</td>
<td>Truncated* (24aa)</td>
<td>PB1F151-185 (HA_C167T_C170T)</td>
</tr>
<tr>
<td></td>
<td>VN/296/PB1-F2TR-HA</td>
<td>Truncated***</td>
<td>PB1C167T_C170T_F and Xba1HAtagPB1R364-43</td>
</tr>
<tr>
<td></td>
<td>VN/296/PB1-F2NO-HA</td>
<td>Deleted**</td>
<td>Kpn1PB1F68-90 and Xba1HAtagPB1R364-43</td>
</tr>
</tbody>
</table>
Table 2. Summary of PB1-F2 truncations in H5N1 viruses.

<table>
<thead>
<tr>
<th>Year of Isolation</th>
<th>Putative lengths (amino acids) of PB1-F2 protein</th>
<th>Percent truncated (aa≤78)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Truncated (aa≤78)</td>
<td>Non truncated (aa≥79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 8 11 24 25 34 57 63 79 81 87 90 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1997</td>
<td>2</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>1998</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1999</td>
<td>11</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>2000</td>
<td>1</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>2001</td>
<td>39</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>2002</td>
<td>3</td>
<td>31</td>
<td>34</td>
</tr>
<tr>
<td>2003</td>
<td>61</td>
<td>1</td>
<td>62</td>
</tr>
<tr>
<td>2004</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2005</td>
<td>2</td>
<td>242</td>
<td>243</td>
</tr>
<tr>
<td>2006</td>
<td>1 2 1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2007</td>
<td>3 1 2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2008</td>
<td>1 2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2009</td>
<td>1 16</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2010</td>
<td>1 12 7</td>
<td>1</td>
<td>1 2</td>
</tr>
<tr>
<td>2011</td>
<td>1 2 3</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>2012</td>
<td>2 3 7</td>
<td>2</td>
<td>52</td>
</tr>
</tbody>
</table>
Table 3. Amino acid differences between viruses VN/281 and VN/296.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Size (aa)</th>
<th>No. of Variations</th>
<th>Variant Residues (VN/281→VN/296)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA</td>
<td>567</td>
<td>1</td>
<td>Q31R</td>
</tr>
<tr>
<td>M1</td>
<td>252</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>97</td>
<td>2</td>
<td>G21V, L86V</td>
</tr>
<tr>
<td>NA</td>
<td>449</td>
<td>4</td>
<td>I63V, G85S, I203T, T244I</td>
</tr>
<tr>
<td>NP</td>
<td>498</td>
<td>1</td>
<td>A373T</td>
</tr>
<tr>
<td>NS1</td>
<td>225</td>
<td>1</td>
<td>D120G</td>
</tr>
<tr>
<td>NS2</td>
<td>121</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>716</td>
<td>2</td>
<td>S66G, A323V</td>
</tr>
<tr>
<td>PB1</td>
<td>756</td>
<td>3</td>
<td>I181M, P369S, V667I</td>
</tr>
<tr>
<td>PB1-F2</td>
<td>90</td>
<td>3</td>
<td>M1T, Q54R, P67L</td>
</tr>
<tr>
<td>PB2</td>
<td>759</td>
<td>2</td>
<td>D304E, A588V</td>
</tr>
<tr>
<td>Total</td>
<td>4530</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Total (excluding PB1-F2)</td>
<td>4440</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

1 Both the viruses have a C-terminal truncation in NS1 that renders it 215aa long.
2 PB1-F2 is 90aa in virus VN/281, while initiating ATG is mutated to ACG in virus VN/296. The 3 differences mentioned in the table are counted since a protein may either express from alternating start codons or from downstream in-frame ATGs.