Evolutionary Conservation of the PA-X Open Reading Frame in Segment 3 of Influenza A Virus

Citation for published version:

Digital Object Identifier (DOI):
10.1128/JVI.01677-12

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published in:
Journal of Virology

Publisher Rights Statement:
Copyright © 2012, American Society for Microbiology. All Rights Reserved.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Evolutionary Conservation of the PA-X Open Reading Frame in Segment 3 of Influenza A Virus

Mang Shi,a Brett W. Jagger,b Helen M. Wise,c Paul Digard,c Edward C. Holmes,a,d and Jeffery K. Taubenbergerb

Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania, USA,a; Viral Pathogenesis and Evolution Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA,b; The Roslin Institute, The University of Edinburgh, Midlothian, Scotland, United Kingdom,c; and Fogarty International Center, National Institutes of Health, Bethesda, Maryland, USA,d

PA-X is a fusion protein of influenza A virus encoded in part from a +1 frameshifted X open reading frame (X-ORF) in segment 3. We show that the X-ORFs of diverse influenza A viruses can be divided into two groups that differ in selection pressure and likely function, reflected in the presence of an internal stop codon and a change in synonymous diversity. Notably, truncated forms of PA-X evolved convergently in swine and dogs, suggesting a strong species-specific effect.

It was recently determined that segment 3 of influenza A virus (IAV) encodes a second protein, made in part from a +1 open reading frame (ORF) embedded within the PA gene. This ORF is accessed via ribosomal frameshifting to produce a fusion protein, denoted PA-X, whose N-terminal 191 amino acids are derived from the PA ORF, while the C-terminal sequence, most commonly 61 amino acids long, is derived from the PA ORF. This phylogenetic pattern suggests that swine CS viruses with the +1 X-ORF occurred in a cluster of viruses sampled between 1985 and 2006 (Fig. 2). These truncated PA-X proteins were uniquely, twice in swine and twice in dogs. Uniquely, the stop codon mutation at X-ORF codon 42 was fixed at least four times due to nonsense mutations, of which 2,279 (99%) occurred at X-ORF codon 42 (Fig. 2). These truncated PA-X proteins were associated with seven IAV groups: human H1N1pdm, swine TR H1N1, one cluster of swine CS H1N1 viruses, equine H7N7, canine H3N8, canine H3N2, and bat influenza virus (italics in Table 1), although the truncated proteins in the very small equine H7N7 and bat influenza virus data sets are due to stop codon mutations other than those at codon 42. The CS H1N1 viruses are particularly noteworthy, as the truncated form of PA-X occurs in a cluster of viruses sampled between 1985 and 2009, which fall within a group of generally older (1930 to 2006) swine CS viruses that possess a full-length PA-X (Fig. 2). This phylogenetic pattern suggests that swine CS viruses with the truncated form of PA-X were directly derived from those with full-length PA-X sequences. In addition, as the origins of the PA segment in human H1N1pdm lie with the TR swine influenza virus (2), the stop codon mutation at X-ORF codon 42 in this virus was clearly inherited from swine.

Overall, our phylogenetic analysis suggests that the nonsense mutation at X-ORF codon 42 was fixed at least four times independently—twice in swine and twice in dogs. Uniquely,
FIG 1 Influenza A virus segment 3 showing open reading frames for PA (in frame 0) and for the PA-X ORF (in frame 1), with the frameshift motif shown. Yellow shading indicates the region of the X-ORF that is translated; blue shading indicates structural domains of PA (1, 4, 10). The PA-X open reading frame encodes either 61 or 41 amino acids as indicated. Note that the X-ORF product lies largely within a linker region between the PA N- and C-terminal domains. (Reprinted from reference 6 with permission of the publisher).

The presence of a full-length PA-X protein in avian influenza virus, human H3N2, human 1918 pandemic and seasonal H1N1, swine EA, and equine H3N8 was associated with a reduction in $d_{s}$ in the region of the PA gene (frame 0) overlapping the +1 X-ORF compared to the rest of the PA gene (reflected in the “$d_{s}$/X-ORF/$d_{s}$ rest” column in Table 1). For example, in the case of the largest data set, representing avian influenza sequences, $d_{s}$ in the region of PA gene covering the X-ORF is 0.25 times that of $d_{s}$ in the remainder of the PA gene. In contrast, there was no such reduction in $d_{s}$ in those groups that possess truncated PA-X proteins, as was also the case for influenza B virus. For example, in human H1N1pdm, $d_{s}$ in the X-ORF was 1.19 times that of $d_{s}$ in rest of the PA gene (Table 1). The CS H1N1 viruses were again noteworthy; although most of the earlier-sampled CS viruses possess a full-length PA-X, there was little reduction in $d_{s}$ in the region of the PA gene covering the X-ORF. This, coupled with the fact that most recent CS sequences harbor a truncated PA-X, suggests that this protein may be of less functional importance in classical swine influenza viruses. Finally, it was also striking that the overall $d_{s}/d_{t}$ ratio for the +1 X-ORF was substantially higher than that for the PA gene encoded in frame 0 (Table 1). Although accurately estimating selection pressures in sequences with dual reading frames is notoriously difficult (7), this observation indicates that PA-X is able to tolerate a high number of amino acid changes. As synonymous mutations in frame 0 will either be nonsynonymous or nonsense mutations in frame 1, this is likely to be the case for many viral proteins encoded by +1 ORFs (5).

The conservation of the decanucleotide frameshift motif (Fig. 1) and the maintenance of the full-length +1 X-ORF in the majority of IAV genomes infecting diverse host species (6) suggest that PA-X is important for influenza A virus biology. Interestingly,

### TABLE 1 Frequency of stop codon mutations and the numbers of synonymous and nonsynonymous nucleotide substitutions per site for different influenza viruses

<table>
<thead>
<tr>
<th>Virus</th>
<th>n</th>
<th>No. truncated</th>
<th>$d_{s}$ for:</th>
<th>$d_{s}$ for:</th>
<th>$d_{s}$ for:</th>
<th>$d_{s}$/X-ORF/ $d_{s}$ rest</th>
<th>$d_{s}$/X-ORF</th>
<th>$d_{s}$/X-ORF + 1 X-ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avian influenza</td>
<td>4,361 (80)</td>
<td>19 (2)</td>
<td>0.50</td>
<td>0.52</td>
<td>4.08</td>
<td>16.44</td>
<td>0.25</td>
<td>0.12</td>
</tr>
<tr>
<td>Human H3N2</td>
<td>2,296 (49)</td>
<td>8 (5)</td>
<td>0.09</td>
<td>0.09</td>
<td>0.53</td>
<td>1.23</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>Human H1N1s</td>
<td>853 (50)</td>
<td>3 (1)</td>
<td>0.10</td>
<td>0.11</td>
<td>0.93</td>
<td>1.62</td>
<td>0.57</td>
<td>0.10</td>
</tr>
<tr>
<td>Human H1N1pdm</td>
<td>1,916 (56)</td>
<td>1,914 (1,914)</td>
<td>0.09</td>
<td>0.10</td>
<td>0.68</td>
<td>0.57</td>
<td>1.19</td>
<td>0.13</td>
</tr>
<tr>
<td>Swine CS</td>
<td>121 (52)</td>
<td>6 (3)</td>
<td>0.17</td>
<td>0.17</td>
<td>2.18</td>
<td>2.18</td>
<td>1.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Swine CS–stop</td>
<td>99 (53)</td>
<td>98 (98)</td>
<td>0.15</td>
<td>0.16</td>
<td>1.95</td>
<td>1.67</td>
<td>1.17</td>
<td>0.08</td>
</tr>
<tr>
<td>Swine EA</td>
<td>152 (60)</td>
<td>0 (0)</td>
<td>0.33</td>
<td>0.28</td>
<td>3.18</td>
<td>3.60</td>
<td>0.88</td>
<td>0.10</td>
</tr>
<tr>
<td>Swine TR</td>
<td>248 (73)</td>
<td>225 (225)</td>
<td>0.46</td>
<td>0.30</td>
<td>3.18</td>
<td>2.76</td>
<td>1.15</td>
<td>0.14</td>
</tr>
<tr>
<td>Equine H3N8</td>
<td>80 (35)</td>
<td>0 (0)</td>
<td>0.12</td>
<td>0.09</td>
<td>0.25</td>
<td>0.73</td>
<td>0.34</td>
<td>0.49</td>
</tr>
<tr>
<td>Equine H7N7</td>
<td>2 (2)</td>
<td>2 (0)</td>
<td>0.01</td>
<td>0.005</td>
<td>0.23</td>
<td>0.19</td>
<td>1.21</td>
<td>1.21</td>
</tr>
<tr>
<td>Bat influenza</td>
<td>2 (2)</td>
<td>2 (0)</td>
<td>0.01</td>
<td>0.005</td>
<td>0.14</td>
<td>0.13</td>
<td>1.08</td>
<td>0.05</td>
</tr>
<tr>
<td>Canine H3N8</td>
<td>26 (6)</td>
<td>25 (24)</td>
<td>0.03</td>
<td>0.01</td>
<td>0.07</td>
<td>0.05</td>
<td>1.40</td>
<td>0.41</td>
</tr>
<tr>
<td>Canine H3N2</td>
<td>8 (6)</td>
<td>8 (7)</td>
<td>0.03</td>
<td>0.01</td>
<td>0.16</td>
<td>0.05</td>
<td>3.20</td>
<td>0.18</td>
</tr>
<tr>
<td>Influenza B</td>
<td>190 (34)</td>
<td>190 (AU)</td>
<td>0.05</td>
<td>0.06</td>
<td>1.07</td>
<td>1.11</td>
<td>0.96</td>
<td>0.05</td>
</tr>
</tbody>
</table>

**a** Virus groups with truncated PA-X proteins and associated data are in italics.

**b** Human H1N1s, seasonal H1N1; human H1N1pdm, 2009 pandemic H1N1; swine CS, “classical” swine H1N1; swine EA, Eurasian avian H1N1-like swine H1N1; swine TR, triple-reassortant swine H1N1.

**c** Reading frame encoding the PA protein.

**d** First number is the total data set. The representative sample size used for the analysis of $d_{s}$ and $d_{t}$ is shown in parentheses.

**e** First number is the number of viruses from the total data set that are truncated because of premature stop codons. The number of viruses with stop codon mutations at codon 42 is shown in parentheses.

**f** The classical swine viruses can be divided into two groups representing those with and without the stop codon mutation at codon 42. Note that there has been a single reversion to the non-stop codon state in the stop codon group.

**g** Because the very small sample size these sequences were analyzed with MEGA5 rather than PAML.

**h** AU, sequence alignment is uncertain in this region.

**i** For the full-length X-ORF the analysis of $d_{s}$/X-ORF was performed on all 61 amino acids present in the protein, while for sequences encoding the truncated protein this analysis was restricted to the 41 amino acids prior to the position corresponding to the stop codon.

**j** NA, contains too many stop codons for meaningful analysis.
there are other possible synonymous mutations in the PA gene that would result in a PA-X truncated more severely than the commonly observed 41-amino-acid product of X-ORF. That these are not often seen suggests that stop codon mutations at X-ORF codon 42 renders the protein functionally distinct from products of 61-codon X-ORF isolates. As there is no evidence for decreased synonymous variability in the overlapping 0 frame of these truncated isolates, there are likely to be few selective constraints on X-ORF in these particular lineages. It is therefore probable that the protein domains encoded by the truncated X-ORFs have lost or altered functionality compared to the PA-Xs encoded by full-length X-ORFs. This hypothesis will need to be evaluated experimentally, especially in the context of particular host species. Indeed, as X-ORF protein truncation appears to be associated with IAV lineages circulating in particular hosts, i.e., pigs and dogs, there may be some species specificity to the evolution of a truncated PA-X protein. That a 41-amino-acid X-ORF protein evolved convergently in both IAV subtypes that infect dogs (H3N2 and H3N8) is particularly noteworthy and suggests that the truncation of this protein may be associated with the adaptation and emergence of influenza virus in this host species.

REFERENCES