Upstream start codon in segment 4 of North American H2 avian influenza A viruses

Girard, G.A.O.; Goultiaev, A.P.; Olsthoorn, R.R.C.L.

Upstream start codon in segment 4 of North American H2 avian influenza A viruses

Geneviève Girard [a,1], Alexander P. Gultyaev [a,b], René C.L. Olsthoorn [a,*]

[a] Department of Molecular Genetics, Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands
[b] Department of Virology, Erasmus Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

ABSTRACT

H2N2 influenza A virus was the cause of the 1957 pandemic. Due to its constant presence in birds, the H2 subtype remains a topic of interest. In this work, comparison of H2 leader sequences of influenza A segment 4 revealed the presence of an upstream in-frame start codon in a majority of North American avian strains. This AUG is located seven codons upstream of the conventional start codon and is in a good Kozak context. In vivo experiments, using a luciferase reporter gene fused to leader sequences derived from North American avian H2 strains, support the efficient use of the upstream start codon. These results were corroborated by in vitro translation data using full-length segment 4 mRNA. Phylogenic analyses indicate that the upstream AUG, first detected in 1976, is stably nested in the North American avian lineage of H2 strains nowadays. The possible consequences of the upstream AUG are discussed.

© 2011 Elsevier B.V. Open access under the Elsevier OA license.

1. Introduction

Influenza is a relatively common epidemic disease that became pandemic a few times in the past century. Influenza A viruses can be classified into subtypes based on the antigenicity of hemagglutinin (HA, 16 subtypes) and neuraminidase (NA, 9 subtypes), two glycoproteins found at the surface of the virions. Each HA and NA subtype has been detected in birds, whereas mostly HA types 1, 2 and 3 and NA types 1 and 2 have been found in humans. The influenza A virus genome consists of eight RNA segments of negative polarity encoding at least eleven proteins. Upon infection of a host by two different viral strains, reassortment of the segments can occur and this results into a novel strain. Particularly when this occurs between a human and a non-human strain, the new influenza virus has a higher probability for causing a pandemic in the human population.

In the twentieth century, several subtypes have caused seasonal and pandemic influenza. The 1957 and 1968 pandemics originated from H2N2 and H3N2 reassortants, respectively. Since then, H2 subtype has not been found anymore in the human population, meaning that most humans are immunologically naïve to H2. This subtype, still present in the avian population, is therefore a possible candidate for a future outbreak (Liu et al., 2004; Makarova et al., 1999; Schafer et al., 1993).

Phylogenetic analyses have uncovered two geography-linked lineages of avian H2 HA, Eurasian and North American (Schafer et al., 1993). Studies of American poultry have indicated an increasing prevalence of the H2 subtype, meaning a growing risk of transmission to human (Schafer et al., 1993). The two lineages are not completely isolated, as migratory birds have been shown to transmit Eurasian viruses to North American shorebirds (Makarova et al., 1999) and vice versa (Liu et al., 2004), thereby giving another dimension to the evolution of the virus. Recently, H2N3 influenza viruses were isolated from pigs. These strains showed adaptation to the mammalian host and therefore emphasize the importance of H2 surveillance (Ma et al., 2007).

In addition to the large variation in HA protein sequences there exists also large variation, both in length and composition, in the non-coding regions segment 4. The role of this variation is not known but may be related to the packaging of segment 4 or its translation. Upon comparing 5′-untranslated regions (5′-UTRs) of segment 4 in H2 viruses we noticed that some avian strains of H2 subtype exhibit an upstream AUG, which is in frame with the conventional AUG of segment 4. Here, we show

that this upstream AUG is translationally active in a reporter system and may lead to increased HA expression in H2 viruses.

2. Materials and methods

2.1. Sequence analyses

Influenza sequences were retrieved from the NCBI Influenza Virus Resource (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html) on the 12th of May 2010. Sequences were then aligned using the Mafft algorithm (Katoh and Toh, 2008) and UTRs were extracted with the Extractalign tool of EBioX tools (Barrio et al., 2009). Too incomplete sequences were removed, which allowed proper visualization and analysis of the UTRs alignment. Because some strains were sequenced multiple times, duplicate sequences were removed manually from the alignment.

2.2. Signal sequence and cleavage site prediction

We used the SignalP 3.0 server of the Technical University of Denmark (http://www.cbs.dtu.dk/services/SignalP/), the SOSUIsignal tool (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui-signal/sosuisignal_submit.html) and the SIG-Pred tool (http://bmbpcu36.leeds.ac.uk/prot_analysis/Signal.html). These tools were used to predict signal sequence and cleavage site for all possible additional 7-amino-acid stretches. For each of those stretches, a corresponding strain was selected. The prediction was performed using both the HA sequence starting at the usual start codon and the HA sequence starting at the upstream start codon. It was verified that in both cases, the cleavage site is predicted at the same location.

2.3. Cell culture

HeLa (human cervix cell line) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM). The cultures were supplemented with 10% fetal bovine serum (FBS) and 100 μg/ml penicillin/streptomycin.

2.4. Nucleic acids and cloning

All oligonucleotides were synthesized by Eurogentec (Maastricht, The Netherlands). Enzymes were purchased from Fermentas (St. Leon-Rot, Germany). When necessary, nucleic acids quality and amounts were verified by electrophoresis and using Nanodrop (ThermoScientific, De Meern, The Netherlands). DNA isolations from bacteria were performed using Plasmid Midi and Mini kits from QIAGEN (Venlo, The Netherlands). For DNA purification from agarose gels or after restriction digestion, we used GFX columns from GE Healthcare (Hoevelaken, The Netherlands).

2.5. Construction of the luciferase reporter vectors and mutagenesis

The 1300-bp HindIII fragment of pDual-HIV (Dulude et al., 2006) was cloned into the HindIII unique site of pUC19 in order to obtain a vector (pUCRL) where NheI and BstBI sites were unique. An MfeI site was introduced upstream of the Renilla luciferase ORF of pUCRL by ligating hybridized oligonucleotides (5′-CTAGAGCTCGGTACCCGGGGGATCCAATTGTT-3′ and 5′-CGAACAATTGGATCCCCGGGTACCGAGCT-3′) between the NheI and BstBI sites. Subsequently, the 1300-bp HindIII fragment containing the Renilla luciferase of this vector was ligated with the 7400-bp HindIII fragment of pRLHL (Honda et al., 2000) containing the firefly luciferase. The vector obtained was digested with ApaI and NotI, subjected to Klenow treatment

and religated to create pMRL, containing a CMV promoter, a T7 promoter, a unique MfeI site and the Renilla luciferase ORF.

Various 5′UTR sequences of influenza segment 4 were inserted between the CMV promoter and the Renilla luciferase sequence in pMRL by ligating hybridized oligonucleotides between the HindIII and MfeI sites. This cloning procedure allowed to eliminate as much as possible of non-influenza sequences, including the T7 promoter, in the reporter vector between the CMV transcription start and the influenza 5′-UTR.

2.6. Transfections and luciferase assay

HeLa cells from 80 to 90% confluent cultures in T75 flasks were seeded into 24-well plates to 7×10^4 cells per well in a final volume of 500 μl antibiotic-free DMEM supplemented with 10% FBS. After 24 h incubation at 37 °C, cells were transfected with 50 ng of Renilla-luciferase vector and 50 ng Firefly-luciferase control vector, and 1 μl lipofectamine 2000 (Invitrogen, Breda, The Netherlands) according to the manufacturer's instructions. 24 h after transfection, cells were washed with 500 μl PBS and lysed in 100 μl passive lysis buffer (Promega, Benelux) for 10 min at room temperature. 20 μl of lysate were transferred to a 96-well plate. Luciferase activities were measured with a dual luciferase reporter assay kit (Promega) on a Glomax Multi luminometer (Promega), using a 10 s integration step and 25 μl of each reagent per well.

2.7. In vitro translation

A pUC57 derivative containing the full-length sequence of the hemagglutinin gene of strain A/northern shoveler/California/HKWF1128/2007 (H2N7) (accession number CY033340 in GenBank) downstream of the T7 promoter was purchased from GenScript (Piscataway, NJ, USA). A GCT sequence was introduced just upstream of the hemagglutinin sequence to create an NheI cloning site. In the influenza CY033340 sequence, a G87T mutation was introduced to create an Eco91I cloning site.

The following mutations were introduced by replacing the NheI–Eco91I fragment in pUC57–CY033340 by different pairs of oligonucleotides: a C31T created a stop codon instead of the third codon (arginine) downstream of the upstream start codon, construct AUG–STOP–AUG; this mutation was also introduced in combination with mutation of the upstream start codon (AUA–STOP–AUG); an additional A was introduced upstream of C35 to create a +1 frameshift, construct AUG–FSH–AUG. The latter frameshift was also introduced in an upstream start-codon mutant (AUA–FSH–AUG). As a negative control both start codons were mutated in construct AUA–AUA.

The six pUC57-derivatives were isolated using the Plasmid Midi kit from Qiagen. Six micrograms of DNA was then digested with SacI and HindIII and purified. One microgram of digested DNA was used as template for transcription using a RiboMax Large Scale T7 Kit (Promega) in the presence of cap analog (Promega). After the DNase treatment, RNA was purified over Micro Bio-Spin 6 columns (BioRad). RNA was used at a final concentration of 30 nM in 10 μl volume as template for in vitro translation using Rabbit Reticulocyte Lysate (Promega) with 1 μl [^{35}S]-methionine (1175 Ci/mmol, 10 mCi/ml) for 1 h. The firefly luciferase gene control provided by the manufacturer was used as positive control for translation. After translation, 2× Laemmli buffer was added and the samples were heated for 4 min at 95 °C, and electrophoresed in an 8% SDS-polyacrylamide-gel. The gel was dried and exposed to a phosphorimager screen (BioRad). Relative band intensities were calculated using Quantity One software (Biorad).
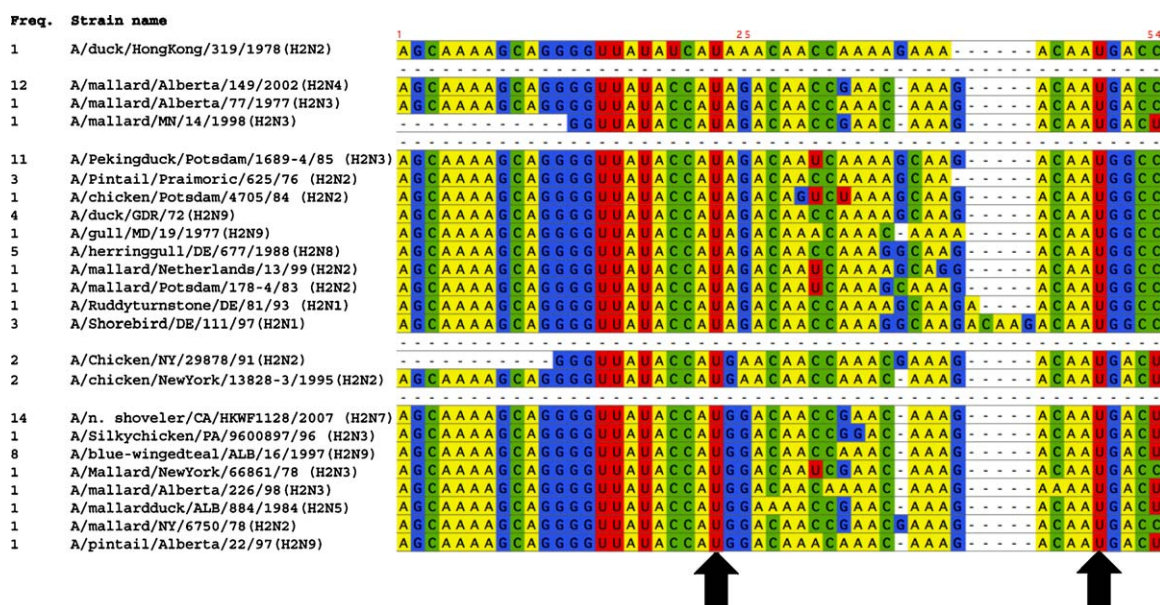
Fig. 1. Alignment of avian H2 segment 4 sequences. The frequency (first column) refers to the number of times the corresponding 5′UTR sequence was found among the 78 different strains for which 5′UTR sequence was available on GenBank. For each type of 5′UTR sequence, one representative strain is given (strain name is indicated). The two black arrows indicate the positions of the two possible start codons.

3. Results

3.1. Segment 4 harbors an additional upstream in-frame start codon in most North American avian H2 strains

Alignment of segment 4 sequences of avian H2 strains (Fig. 1) revealed that on a total of 78 strains for which the 5′UTR is available in GenBank, 29 exhibit an AUG located 21 nucleotides upstream of the regular start codon. This upstream start codon (−21) is in frame with the HA ORF and may therefore also be used for translation initiation. In three strains, A/Guinea/NJ/3070/91 (H2N2), A/Chicken/NY/29878/91 (H2N2), and A/Mallard/NY/6750/78 (H2N2), an upstream start codon is also present at the same position, but not in frame. In the remaining 46 strains AUA is found at the upstream location.

The optimal context for initiation of translation is known to be GCCRCCaugG in vertebrates (Kozak, 2005). Among these elements, the -3 purine and the +4 G are the most important. For influenza strains of the avian H2 group, the [−7; −4] region is optimal neither for the usual nor for the upstream start codon when present. ACA is found upstream of the main AUG, whereas the slightly more optimal ACC is present upstream of the alternative AUG. The +4 G is found downstream of the main start codon in 31 out of the 46 strains that have no upstream start codon. Remarkably, when an upstream start codon is present, the +4 G at the main start is lost (29 strains) while at the same time the upstream start does possess a +4 G (27 strains). This almost perfect correlation strongly suggests that the upstream start is indeed functional. In only 2 cases both up and downstream starts have a +4 A.

3.2. Phylogeny of North American avian H2 strains

Using the known H2 phylogeny (Liu et al., 2004; Makarova et al., 1999; Schafer et al., 1993), which is based on the coding region of segment 4, the phylogenetic distribution of strains harboring the upstream start codon was investigated. The recent classification of the main H2 sublineages as proposed by Liu et al. (2009) was used to map these strains (Fig. 2). It appears that none of the avian strains of the branch h2.2 display an upstream AUG and all have a +4 G. This group corresponds to the Eurasian lineage of avian H2 viruses. Strains possessing an upstream AUG cluster in the North American lineage (group h2.1.2), most of them having upstream AUGG and downstream AUGA (Fig. 2, "AUGG/AUGA"). Correlatively, strains isolated in Delaware (USA) are known to be part of the Eurasian lineage and do not exhibit any upstream AUG. However, not all strains of the North-American lineage h2.1 possess the upstream AUG: several strains having AUAG/AUGA (in h2.1.2) or AUAG/AUGG (in h2.1.1) cluster with the North American branch.

The upstream AUG is found in all type of birds, including mallard, green-winged and blue-winged teals, sanderling, laughing gull, northern shoveler, pintail, semi-palmated sandpiper, guinea fowl, silky chicken and chicken. There is no obvious correlation with the NA type as the upstream AUG is found in N1, N2, N3, N5, N7, N8, and N9 subtypes (N6 has not been found yet in American avian H2 strains, and of H2N4 only one sequence is known).

3.3. The upstream AUG is used in the context of an in vivo reporter fusion

In order to investigate whether this upstream AUG is functional in vivo, we selected the 5′UTR region of strain A/northern shoveler/California/HKWF1128/2007 (H2N7) (accession number CY033340), which displays the AUGG/AUGA combination predominantly found in h2.1.2 strains (Fig. 2). The entire 5′-UTR, including the second AUG and one adjacent codon, was cloned upstream of Renilla luciferase in our pMRL reporter plasmid. Mutations were introduced that affected the context of the two starts (Fig. 3). As reference, the 5′UTR sequence of strain A/gull/MD/19/1977 (H2N9) (accession number EU742644) from the other major group (h2.1.1 with AUAG/AUGG) was also cloned in the reporter construct. HeLa cells were transfected with these constructs together with a reference construct producing firefly luciferase. Results are shown in Fig. 3.

Fig. 3 shows that the upstream AUG is functional in translation (compare AUAG/AUAA with AUGG/AUAA, bars 3 and 9). In the presence of a +4 G, translation from leaders with an upstream AUG was only mildly enhanced by introducing the downstream AUG
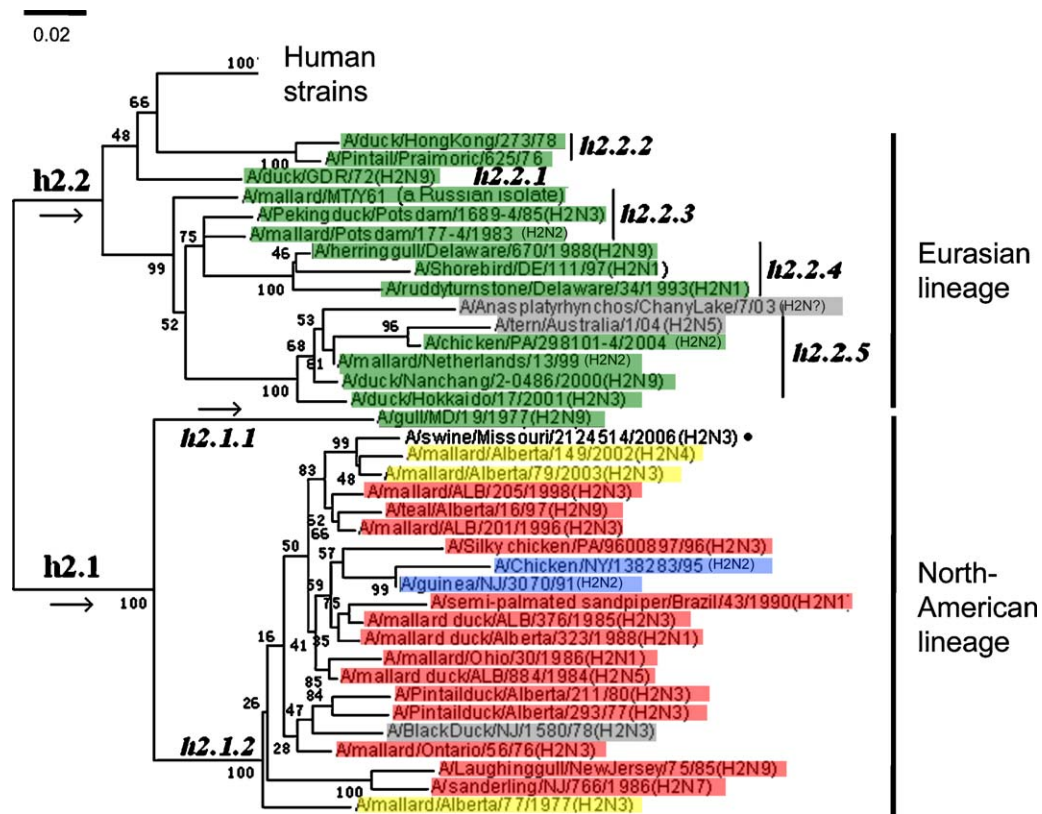
Fig. 2. Phylogenetic tree of the avian H2 segment 4 sequences. The colors indicate the sequence that is found at positions −21/+1 (+1 representing the conventional start codon and −21 the upstream one). Gray: no 5′UTR sequence available. Green: AUAG/AUGG. Red: AUGG/AUGA. Yellow: AUAG/AUGA. Blue: AUGA/AUGA.Phylogenetic tree adapted from Liu et al. (2009).

with a +4 A or G (compare bars 1–3), indicating that the vast majority of scanning ribosomes do not reach past the first AUG. Interestingly, expression from the upstream AUG is actually much higher than that from the second AUG (compare bars 3 and 7). In
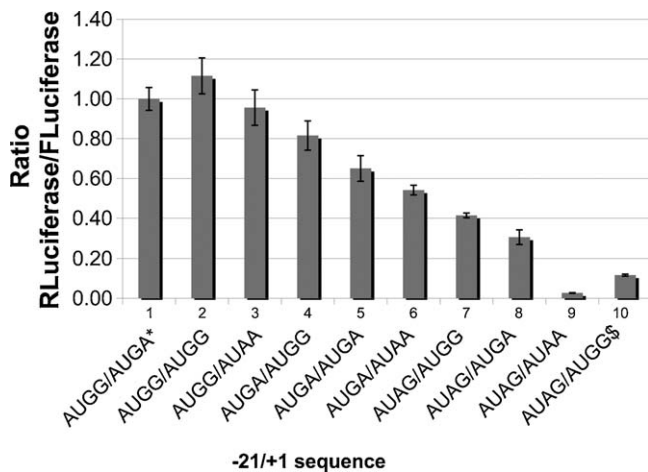

Fig. 3. Effect of changes in number and context of start codons in the H2 5′ noncoding region on translation. Various influenza 5′UTR sequences and derivatives were cloned upstream of Renilla luciferase (RL) in our reporter pMRL vector. A vector expressing firefly luciferase (FL) in a strong context was co-transfected with the RL reporter as a transfection control. All the mutants derive from the 5′UTR of strain A/northern shoveler/California/HKWF1128/2007 (H2N7) (wild-type sequence indicated by an asterisk). The only modifications are at the usual start codon (+1) and the upstream one (−21), as indicated on the horizontal axis. The dollar sign indicates a reference 5′UTR (A/gull/MD/19/1977 (H2N9)). The experiment was performed five times independently with each time four replicates for each strain. One representative experiment is shown here.

fact, the presence of the upstream AUG, irrespective of the identity of the +4 nucleotide of any of the two start codons, in all cases seemed to enhance expression by a factor of two (compare bars 5 and 8, or bars 4 and 7). We note that the 5′UTR of A/gull/MD/19/1977 (AUAG/AUGG, bar 10) gave a 3-fold lower RL expression than the supposedly equivalent AUAG/AUGG mutant of the A/northern shoveler/California/HKWF1128/2007 (H2N7) strain (bar 7). This may be related to three nucleotide differences in the remainder of the UTR between these two constructs. Finally it is noteworthy that translation from the A/northern shoveler/California/HKWF1128/2007 strain leader is almost 10-fold higher than that from the A/gull/MD/19/1977 strain leader.

3.4. The upstream AUG in the context of the full-length segment 4 is active in vitro

As the above data were obtained with only a short region of segment 4, i.e. the 5′UTR, we could not rule out that other features of segment 4 could play a role in translation. To examine this possibility, we tested the translational efficiency of the full-length segment 4 of the A/northern shoveler/California/HKWF1128/2007 (H2N7) strain in vitro. To investigate to what extent the first start codon was used, a frameshift mutation was introduced downstream of the first start codon. Translation from the first start codon would result in a product of 38 aa (not visible on the gel in Fig. 4) whereas translation from the second start codon would yield a product of 63 kDa. As can be seen in Fig. 4 this mutation resulted in a 5-fold decrease in the yield of the full-length product (lane 2). The remaining 22% probably derives from translation from the second start codon. Introduction of a stop codon downstream of the first start codon resulted in only a two-fold decrease in full-length product (lane 3). This could be due to the fact that
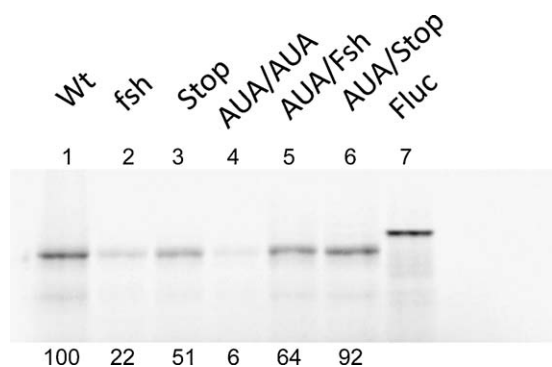
Fig. 4. *In vitro* translation results for full-length segment 4 of strain A/northern shoveler/California/HKWF1128/2007 (H2N7) and mutant derivatives. Lane 1: wild type, lane 2: AUG–FSH–AUG, lane 3: AUG–STOP–AUG, lane 4: AUA–AUA, lane 5: AUA–FSH–AUG, lane 6: AUA–STOP–AUG, and lane 7: firefly luciferase gene. The numbers at the bottom of the picture indicate relative intensities. See text Sections 3 and 2 for more details.

ribosomes are very close the second AUG when dissociating at the introduced stop codon, which makes re-initiation more probable than when ribosomes terminate 38 codons further downstream in the case of the frame-shift mutant. Knockdown of the first start codon in the frameshift mutant largely restored translation of the full-length product from the second start codon (lane 5), indicating that the reduced yield in lane 2 was not due to adverse effects of the frameshift mutation on translation from the 2nd start codon, but was the result of ribosomes initiating mainly at the first start codon. Lane 6 shows that in the absence of the first AUG ribosomes efficiently initiate translation at the 2nd AUG. In the absence of both start codons translational yield is close to background levels (lane 4). These results show that also in the context of the entire segment 4 mRNA of A/northern shoveler/California/HKWF1128/2007 (H2N7) translation is initiated mainly at the upstream AUG.

3.5. Use of the upstream AUG is predicted not to affect signal peptide cleavage and HA secretion

HA is known to be translocated into the endoplasmic reticulum of host cells (Sabatini et al., 1982). This happens after recognition of an N-terminal signal sequence by the cellular machinery. Cleavage of the signal sequence is necessary for production of the mature HA protein and also involves molecular recognition within the amino-acid sequence (McQueen et al., 1984; Sekikawa and Lai, 1983). One can rightfully wonder whether the seven additional amino acids in the HA of North American strains could interfere with the translocation and cleavage signaling in the HA polypeptide. Algorithms are available that allow to predict the presence of a signal sequence and the location of the cleavage site. We used three of them and the same results were obtained; when an alternative start codon is present, a 24-amino acid signal peptide and a cleavage site can be predicted in both versions of the HA protein (starting at the conventional start codon or seven amino acids upstream) at the same location. This was checked for all the different stretches of seven amino acids predicted in the strains with upstream AUG: MNNQTKT, MDNRTKT, MENRTKT, MDNQTKT, MDNQTKK and MDNKTKK.

4. Discussion

Various studies suggest that the amounts of viral proteins produced during influenza infection are primarily under transcriptional control, but that post-transcriptional control plays an important role in fine-tuning (Hatada et al., 1989; Ortin, 1998; Tekamp and Penhoet, 1980; Yamanaka et al., 1988, 1991),

particularly for late proteins, which include HA (Hatada et al., 1989). It is therefore relevant to pay attention to possible alternative translation initiation events in segment 4.

Here we have discovered a previously unrecognized alternative start codon in the leader of segment 4 of most North American avian H2 strains. This start codon is located seven codons upstream of the conventional start codon (Fig. 1) and is in a relatively strong Kozak context of ACCaugG, while the downstream Kozak consensus is weaker, suggesting that the upstream start is functional. Indeed, our *in vivo* data with the leader of one of the North American avian strains, having this upstream AUG (A/northern shoveler/California/HKWF1128/2007(H2N7)) fused to a luciferase reporter gene, supported the efficient use of the upstream start codon (Fig. 3). These results were corroborated by *in vitro* translation data using the full-length segment 4 of A/northern shoveler/California/HKWF1128/2007(H2N7) (Fig. 4).

What could be the consequences of the use of the upstream codon for the virus? Use of the upstream AUG leads to a longer N-terminus, but this is not retained in the mature version of the HA protein if cleavage occurs normally. By using several different algorithms we predicted no adverse effects on cleavage of the signal peptide, whose size increases to 24 amino-acids when translation starts at the upstream AUG. Besides, many proteins are known to have a signal sequence of 24 amino acids or more (von Heijne, 1983). Finally, in the case of HA itself, we verified that the position of the cleavage site can be variable between strains (data not shown and von Heijne, 1983), up to 21 amino acids from the N-terminus. All together, these observations suggest that the use of the alternative upstream start codon is unlikely to interfere with proper translocation and secretion of HA.

The data of the *in vitro* expression with full-length segment 4 mRNA showed that by leaky scanning the downstream start codon can be used as well. In the *in vivo* experiments the presence of the downstream start codon resulted in about 10% higher expression of the luciferase (Fig. 3, compare columns 2 and 3) suggesting that also *in vivo* a small percentage of ribosomes start translation from the second start codon.

The use of alternative start codons is not without precedent in influenza virus A mRNAs. In segment 2 (PB1), leaky scanning of ribosomes past the first AUG, which is not in an optimal Kozak context (AUUUGAaugG), is postulated to produce the PB1-F2 product, which is in the +1 frame with respect to the PB1 gene (Chen et al., 2001). An N-terminally truncated form of PB1, the so-called N40 product, is also postulated to result from leaky scanning past the first four AUGs (Wise et al., 2009).

There exist at least two models for describing ribosome behavior around a eukaryotic start codon. Whether the ribosome scans from 5′ to 3′ and starts translation at the first AUG with reasonably good Kozak context (Kozak, 2005) or that it is 'fluttering' as suggested by Matsuda and Dreher (2006), we are fairly confident that the context of the 5′UTR of this segment 4 is in favor of the upstream AUG when present, also in the whole virus situation. Indeed, its context is better than that of the conventional AUG, which is what matters for the first model. Concerning the second model, the 18-nucleotide distance between the two AUGs is larger than the postulated distance on which the ribosome can travel back and forth (Matsuda and Dreher, 2006).

One could argue that the upstream AUG is located too close to the 5′ end and when the influenza polymerase complex is bound to the 5′ end, as some models suggest (Burgui et al., 2007), such an AUG may not be used or recognized by the 40S subunit. However other HA segments have similarly-sized or even shorter 5′UTRs. So there is no reason why the H2 HA upstream AUG should not be used by the virus.

The presence of the upstream AUG resulted in a two to three-fold higher translation efficiency as compared to leaders having
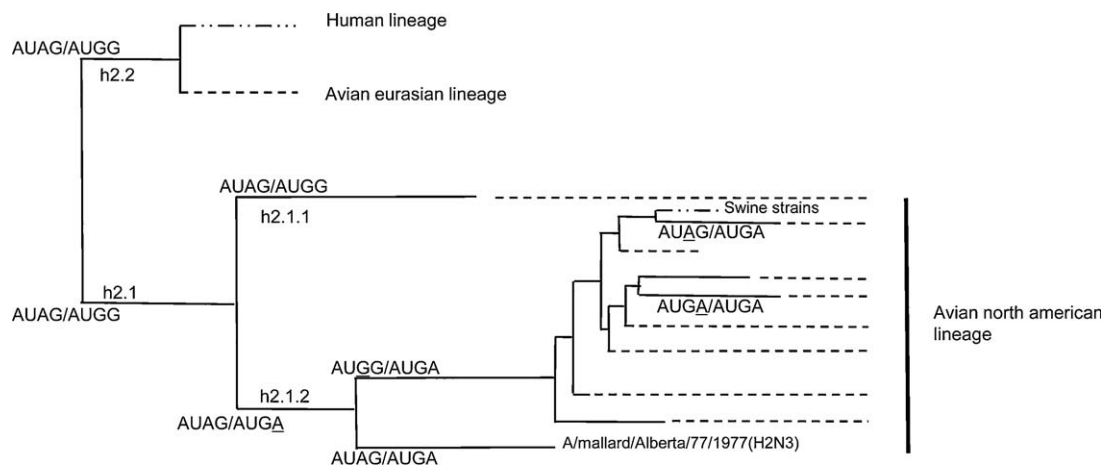
Fig. 5. Schematic phylogenetic distribution of the −21 and +1 nucleotides. Simple dashes indicate that subclusters are present but not drawn. Alternation of dashes and dots indicate clusters that are not avian (human or swine). Underlined nucleotides indicate single mutations.

only the downstream start codon. It is not clear whether this would occur in a viral infection. It has been shown that unknown mechanisms operating in influenza A virus mRNAs are able to overrule classical translation-regulation rules (Maeda et al., 2004), probably because maximum expression of all viral proteins is not necessary or not even favorable. Some influenza viral proteins such as NS1 and NEP (Salvatore et al., 2002) or host factors such as GRSF-1 (Kash et al., 2006) are able to influence the expression of viral proteins by binding the 5′UTR. These factors may attenuate translation from the upstream AUG in these North American H2 strains.

In view of evolution of avian influenza H2 virus, the distribution of the various start-codon contexts in the phylogeny (Fig. 2) suggests a scenario schematically described in Fig. 5. The presence of AUGA/AUGA and AUAG/AUGA strains in the AUGG/AUGA-dominated subcluster within the North American lineage h2.1.2, shows that occasionally the upstream AUG may be weakened (+4 A instead of +4 G) or lost. However, the fact that in 2007 (A/northern shoveler/California/HKWF1128/2007 (H2N7)) and 2008 (A/bald eagle/2008/mixed subtype) H2 strains with the upstream AUG were still isolated, indicates that the upstream AUG, first detected in 1976, is stably nested in the North American avian lineage of H2 strains.

Concerning possible reassortment with human strains and/or transmission to human, the emergence and persistence of an HA segment with a putative longer N-terminus is of interest. Since the N-terminus of HA protein is cleaved off, it is not expected to have an impact on the antigenicity of the virus. However, our data suggest that HA expression may be elevated with possible consequences for virus fitness and its ability to compete with other strains. It should be noted that upstream AUGs are absent in mammalian H2 strains or other HA types, with the exception of a few H5 strains (data not shown). Interestingly, human H2 strains cluster with the Eurasian avian strains, which lack the upstream AUG, while swine H2 strains seem to cluster with Alberta mallard strains, which are among the few North-American strains lacking the upstream AUG (Fig. 2). We do not know whether this means that only strains that lose the upstream AUG are able to infect mammalian hosts.

Finally, we would like to point out that none of the GenBank annotations of these strains make mention of an alternative start codon. Apparently, by comparison with previously sequenced strains of human H2, one assumes that the downstream start codon is the active one, even when this one is in a weaker Kozak context than the upstream one. Inspection of other segments also reveals a number of strains that have an unannotated upstream in-frame start codon in addition to the conventional one (R.C.L.O., *unpublished data*). In the future, a more systematic investigation of effective start codons may complicate our view of influenza translation even more.

References

Barrio, A.M., Lagercrantz, E., Sperber, G.O., Blomberg, J., Bongcam-Rudloff, E., 2009. Annotation and visualization of endogenous retroviral sequences using the Distributed Annotation System DAS and eBioX. BMC Bioinform. 10, S18.

Burgui, I., Yanguez, E., Sonenberg, N., Nieto, A., 2007. Influenza virus mRNA translation revisited: is the eIF4E cap-binding factor required for viral mRNA translation? J. Virol. 81, 12427–12438.

Chen, W., Calvo, P.A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., Henklein, P., Bennink, J.R., Yewdell, J.W., 2001. A novel influenza A virus mitochondrial protein that induces cell death. Nat. Med. 7, 1306–1312.

Dulude, D., Berchiche, Y.A., Gendron, K., Brakier-Gingras, L., Heveker, N., 2006. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. Virology 345, 127–136.

Hatada, E., Hasegawa, M., Mukaigawa, J., Shimizu, K., Fukuda, R., 1989. Control of influenza virus gene expression: quantitative analysis of each viral RNA species in infected cells. J. Biochem. 105, 537–546.

Honda, M., Kaneko, S., Matsushita, E., Kobayashi, K., Abell, G., Lemon.F S. M., 2000. Cell cycle regulation of hepatitis C virus IRES-directed translation. Gastroenterology 118, 152–162.

Kash, J.C., Goodman, A.G., Korth, M.J., Katze, M.G., 2006. Hijacking of the host-cell response and translational control during influenza virus infection. Virus Res. 119, 111–120.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 9, 286–298.

Kozak, M., 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361, 13–37.

Liu, J.H., Okazaki, K., Bai, G.R., Shi, W.M., Mweene, A., Kida, H., 2004. Interregional transmission of the internal protein genes of H2 influenza virus in migratory ducks from North America to Eurasia. Virus Genes 29, 81–86.

Liu, S., Ji, K., Chen, J., Tai, D., Jiang, W., Hou, G., Li, J., Huang, B., 2009. Panorama phylogenetic diversity and distribution of type A influenza virus. PLoS One 4, e5022.

Ma, W., Vincent, A.L., Gramer, M.R., Brockwell, C.B., Lager, K.M., Janke, B.H., Gauger, P.C., Patnayak, D.P., Webby, R.J., Richt, J.A., 2007. Identification of H2N3 influenza A viruses from swine in the United States. Proc. Natl. Acad. Sci. U.S.A. 104, 54–20949.

Maeda, Y., Goto, H., Horimoto, T., Takada, A., Kawaoka, Y., 2004. Biological significance of the U residue at the -3 position of the mRNA sequences of influenza A viral segments PB1 and NA. Virus Res. 100, 153–157.

Makarova, N.V., Kaverin, N.V., Krauss, S., Senne, D., Webster, R.G., 1999. Transmission of Eurasian avian H2 influenza virus to shorebirds in North America. J. Gen. Virol. 80, 3167–3171.

Matsuda, D., Dreher, T.W., 2006. Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. RNA 12, 1338–1349.

McQueen, N.L., Nayak, D.P., Jones, L.V., Compans, R.W., 1984. Chimeric influenza virus hemagglutinin containing either the NH_2 terminus or the COOH terminus of G protein of vesicular stomatitis virus is defective in transport to the cell surface. Proc. Natl. Acad. Sci. U.S.A. 81, 395–399.

Ortin, J., 1998. Multiple levels of posttranscriptional regulation of influenza virus gene expression. Semin. Virol. 8, 335–342.

Sabatini, D.D., Kreibich, G., Morimoto, T., Adesnik, M., 1982. Mechanisms for the incorporation of proteins in membranes and organelles. J. Cell Biol. 92, 1–22.

Salvatore, M., Basler, C.F., Parisien, J.P., Horvath, C.M., Bourmakina, S., Zheng, H., Muster, T., Palese, P., Garcia-Sastre, A., 2002. Effects of influenza A virus NS1 protein on protein expression: the NS1 protein enhances translation and is not required for shutoff of host protein synthesis. J. Virol. 76, 1206–1212.

Schafer, J.R., Kawaoka, Y., Bean, W.J., Suss, J., Senne, D., Webster, R.G., 1993. Origin of the pandemic 1957 H2 influenza A virus and the persistence of its possible progenitors in the avian reservoir. Virology 194, 781–788.

Sekikawa, K., Lai, C.J., 1983. Defects in functional expression of an influenza virus hemagglutinin lacking the signal peptide sequences. Proc. Natl. Acad. Sci. U.S.A. 80, 3563–3567.

Tekamp, P.A., Penhoet, E.E., 1980. Quantification of influenza virus messenger RNAs. J. Gen. Virol. 47, 449–459.

von Heijne, G., 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. 133, 17–21.

Wise, H.M., Foeglein, A., Sun, J., Dalton, R.M., Patel, S., Howard, W., Anderson, E.C., Barclay, W.S., Digard, P., 2009. A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. J. Virol. 83, 8021–8031.

Yamanaka, K., Ishihama, A., Nagata, K., 1988. Translational regulation of influenza virus mRNAs. Virus Genes 2, 19–30.

Yamanaka, K., Nagata, K., Ishihama, A., 1991. Temporal control for translation of influenza virus mRNAs. Arch. Virol. 120, 33–42.