Infection, Genetics and Evolution 11 (2011) 489–495

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

# 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

ARTICLE INFO

*Article history:*
Received 31 August 2010
Received in revised form 22 December 2010
Accepted 24 December 2010
Available online 11 January 2011

*Keywords:*
Segment 4
Influenza A virus
Hemagglutinin
Start codon
Kozak consensus
Untranslated region

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.


## 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

*Abbreviations:* FL, firefly luciferase; HA, hemagglutinin; NA, neuraminidase; N-ter, N-terminus; RL, Renilla luciferase; UTR, untranslated region.

* Corresponding author. Tel.: +31 715274586; fax: +31 715274357.
*E-mail address:* olsthoor@chem.leidenuniv.nl (René C.L. Olsthoorn).
[1] Present address: Department of Molecular Systems Biotechnology, Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands.

1567-1348/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2010.12.014

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′-CTAGAGCTCGGTACCCGGGGATCCAATTGTT-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 $7 \times 10^4$ 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).