PROJECT FINAL REPORT

Publishable Summary

Grant Agreement number: 201607

Project acronym: RNAFLU

Project title: Effect of natural sequence variation on influenza virus RNA function

Funding Scheme: Collaborative project – Small- or medium-scale focused research project

Period covered: from February 1, 2008 to January 31, 2010

Name of the scientific representative of the project's co-ordinator, Title and Organisation:

Stefan Schwartz, PhD
Professor in Molecular Virology
Department of Medical Biochemistry & Microbiology (IMBIM),
Uppsala University, Sweden

Tel: : +46 184714239

Fax: +46 184714763

E-mai: Stefan.Schwartz@imbim.uu.se

Project website address: http://www.imbim.uu.se/RNAFLU/index/html

Final publishable summary report

In this project we have investigated if mutations introduced in the influenza virus RNA genome could affect the function of the viral RNA genome itself, irrespectively of the effect on the viral proteins. Mutations in the viral RNA genome may for example affect the stability of the viral RNA, the splicing efficiency or the translation efficiency. Such mutations may indirectly affect the ability of the virus to replicate in human cells, thereby affecting the pathogenic properties of the influenza virus, and/or their ability to adapt to a new host species. We speculated that there are mutations in influenza virus that affect the properties of the viral RNA genome and we designed experiments to identify such mutations in naturally occurring influenza virus strains. The identification of such mutations would be important as they could serve as markers for the appearance of novel influenza virus strains with altered pathogenic properties. To investigate if naturally occurring mutations affect the function of the influenza virus RNA itself, we selected a number of important properties and functions of the viral RNA, such as mRNA structure, mRNA splicing, binding to cellular microRNAs and mRNA translation that could be monitored in various assays.

RNA molecules derived from influenza viruses with different pathogenic properties or with different species tropism, had different properties. The secondary structure of the RNA molecules from different influenza viruses was different. Such differences in RNA structure may contribute to viral replication efficiency. Well-conserved, putative bindings sites for cellular microRNAs were identified in the influenza virus mRNAs and experimental data indicated that such cellular microRNAs inhibited influenza virus expression. Processing of the influenza viral mRNAs, revealed substantial differences in mRNA splicing efficiencies, which may affect the ration between the NS1 and NS2 proteins and the M1 and M2 proteins. Furthermore, we show that the translation of influenza virus mRNAs was modulated by the NS1 protein. Differences in NS1 expression and function are of particular interest as it has been implicated in influenza virus pathogenesis.

RNA sequence variation seen in naturally occurring influenza virus strains, with widely different pathogenic properties, can affect the structure and function of the influenza virus mRNAs. We found that influenza RNA sequence differences affected influenza virus RNA secondary structure and RNA splicing efficiency in a virus-strain specific manner. Therefore, variation in influenza virus RNA sequence may affect the pathogenic properties of the virus without affecting viral protein sequences.

This is the first evidence to show that influenza virus sequence polymorphism affects the function of the viral RNA. Differences in viral mRNA function were found in influenza virus isolates with widely different pathogenic properties, leading to the testable hypothesis that pathogenicity of influenza virus may be affected by differences in viral RNA sequences that do not affect the function of the viral proteins. If this hypothesis is correct, identification of RNA sequence markers for viral pathogenicity will aid in early detection of highly pathogenic influenza viruses that suddenly appear in the avian and/or human populations. Early detection of pathogenic influenza virus strains may limit the transmission of such strains in the human population. The results of this research may therefore be important for disease control around the world, in particular for strategic decisions taken at centers of disease control.

List of all beneficiaries:

Uppsala Universitet (Coordinator)
Prof. Stefan Schwartz
Uppsala Biomedicinska Centrum
Box 596
751 24 Uppsala, Sweden

Leiden University
Ass. Prof. Rene Olsthorn
Leiden Institute of Chemistry,
Molecular Genetics Unit
Gorlaeus Laboratories
Einsteinweg 55
2333 CC Leiden
The Netherlands

Centro Nacional de Biotecnologia
Prof. Amelia Nieto
Darwin 3, Cantoblanco
28049 Madrid, Spain

Address of website:

http://www.imbim.uu.se/RNAFLU/index.html

A summary description of project context and objectives

Pathogenicity of influenza virus can range from respiratory infections that are cleared within weeks to pneumonia and lower respiratory tract infections with high mortality such as Spanish flu or highly pathogenic bird flu. The sequences of the RNA genomes of the various influenza viruses show a high degree of variability due to influenza virus inherent ability to evolve at a high rate. Since it has not been clarified why some influenza virus types are highly pathogenic in spite of many studies on differences in influenza virus protein sequence, we have investigated if sequence differences between low and high-pathogenic influenza viruses affected the function of the viral RNA rather than viral protein. We found that RNA molecules derived from influenza virus strains with different pathogenic properties have different structural and functional properties. These results open up for further research on the role of influenza virus RNA structure and function in virus pathogenesis.

The Orthomyxoviridae family of RNA viruses contains three human viruses named influenza A, B and C. These three viruses cause acute infections of the respiratory tract, normally giving rise to symptoms that are typical for influenza virus infection: head-ache, high fever, myalgia and coughing. Influenza virus type C appears to be less pathogenic than A and B. While all three viruses can cause local, seasonal epidemics of influenza, influenza virus A can also give rise to pandemics that involve the entire world, spreading across the globe within a year. Although influenza virus infections have been described many times in historical records, the most infamous pandemic we are aware of today is the Spanish flu that occurred in 1918, as well as the 1957 H2N2 and 1968 H3N2 pandemics. Inbetween pandemics, strains from the last pandemic circulate in the population and give rise to local epidemics and annual outbreaks of influenza. Different strains of influenza virus therefore display an enormous difference in transmission efficiency.

Time has also shown that there is huge variation in the pathogenicity of various influenza virus strains, ranging from relatively mild symptoms caused by today's circulating H3N2 strain to the extreme mortality of the 1918 H1N1 strain, which gave rise to the infamous Spanish flu pandemic, in which an estimated 30-50 million people succumbed. The H1N1 virus apparently caused local infections of the respiratory tract, as all other influenza virus strains that have been isolated from humans, as respiratory failure was the main cause of death. However, these examples highlight the ability of influenza virus to change its pathogenic properties over a wide range, from the characteristic symptoms of influenza with fever, coughing and head-ache, to severe, lower respiratory tract infections and pneumonia, in some cases with an extremely high mortality rate.

Recent decades have also seen examples of influenza virus spreading from animals directly to humans, demonstrating the zoonotic properties of influenza A. One example is the transmission of the highly pathogenic H5N1 bird flu to humans in Hong Kong in 1997, killing 6 of 18 infected humans. Although it was generally believed that influenza virus from mammals such as pigs could be transmitted to humans, and vice versa, it was not until recently we obtained evidence that influenza virus can also spread directly from avian hosts to humans, sometimes with devastating results. This was highly unexpected since influenza infection in birds is primarily an enteric infection, targeting the mucosal epithelium of the gut, whereas influenza virus infects respiratory epithelium in humans. Fortunately these viruses were poorly transmitted between humans and did not adopt a more efficient mode of transmission during the epidemics.

At the molecular level, influenza virus is a negative strand RNA virus, with a segmented genome. Mutations are frequently introduced (10^{-5}) in the viral genomes as a result of lack of proof-reading of the viral RNA-dependent RNA polymerase. An additional mechanism for generating variability in the influenza virus population is genetic shift in which different influenza virus strains exchange one or more segments with a co-infecting strain. This mechanism is of particular importance for influenza virus type A that has an extraordinary large gene pool available for genetic shifts that includes both avian and various mammalian influenza virus strains.

As described above, influenza A virus has an amazing ability to rapidly change its properties. Three properties of influenza A is of particular interest: antigenicity, transmission efficiency, and pathogenicity. Although the antigenic properties, and to a large extent, the transmission efficiency are determined primarily by genetic changes of influenza virus that cause changes in the amino acid sequence of the viral H and N proteins, it has been more difficult to identify a single viral proteins that is responsible for the pathogenic properties of highly pathogenic Spanish flu or bird flu. We speculated that pathogenic properties of an influenza virus could also be determined by differences in influenza virus RNA sequence that do not affect the protein sequence of the virus genome. In other words, we hypothesized that sequence variation in the influenza virus genome could also affect structure and/or function of the viral RNA, thereby contributing to the pathogenic properties of the influenza A virus, as well as its ability to adopt to a new host.

All RNAs in a cell are associated with proteins and RNAs are dependent on these interactions to function efficiently. Interactions of RNA with proteins depend on the RNA sequence and secondary structure. The exact RNA sequence is therefore of paramount importance since it affects secondary structure and function and utilisation efficiencies of the viral RNAs. RNA sequence has a direct effect on influenza virus RNA splicing, mRNA stability and translation. This will affect the replication efficiency of the influenza virus strain. It is thus reasonable to speculate that RNA sequence variation itself can affect virus pathogenicity.

The immediate goal of this short, 2-year project was to investigate if naturally occurring influenza virus RNA sequences displayed differences in virus RNA structure and function, such RNA splicing and translation. If so, these results would open up for further research on the effect of these effects on pathogenicity and tropism of various influenza virus strains.

Description of the main S & T results

WP1
Prediction of RNA structures in influenza genomes derived from virus isolates with different pathogenic properties

Development and adjustment of a computer-aided approach to detect conserved RNA structures in the influenza A virus genome.

A special procedure has been developed for predictions of conserved RNA structures encoded in influenza virus genome. The approach allowed us to deal with enormous numbers of sequences isolated from various virus strains and stored in the databases such as GenBank and Influenza Virus Resource (thousands of sequences for each genome segment). The general scheme of the approach is the following:
(a) for each genome segment, identification of clusters of related sequences/strains;
(b) prediction of local consensus structures conserved in different clusters, using representative strains;
(c) in-depth analysis of local consensus structure conservation within all clusters, using all sequences available.
The step (a) has been performed using two non-exclusive approaches: analysis of evolutionary trees available in the literature and building of evolutionary trees using the available phylogenetic packages. The trees have been used to select the main clusters of sequences. In principle, for each segment these clusters are different: for instance, for segments 4 and 6 (coding for surface antigens) the clustering follows well-known subtypes such as H1, H2 etc. or N1, N2 etc., for segment 5 (nucleoprotein) the clustering is mostly host-specific, i.e. human, avian, swine strains, for segment 8 the main clades are determined by hosts and geography. The step (b) has been performed using predictions of RNA structures for single sequences, followed by their comparisons, or by means of the program (Alifold, Hofacker et al.) allowing a prediction of consensus structures on the basis of sequence alignment. The second option turned out to be faster and has been used extensively to build initial models. The step (c) has been carried out using multiple alignments of sequences and analysis of nucleotide covariations that support the models.

For analysis of covariations and their statistical significance in thousands of homologous sequences from different strains, we have developed several dedicated algorithms. This allows us to identify structural elements that are the most conserved in different subtypes of influenza A virus.

Identification of conserved RNA structures in influenza virus RNA segments (deliverable D1.1).
The segments of influenza A virus genome have been analyzed as described above. In addition to predictions of structures in positive-sense mRNAs, predictions of consensus structures in vRNA molecules (negative-sense) have been carried out. These vRNA consensus models were also compared with the predictions for defective-interfering RNAs (DI RNAs), that are supposed to contain the essential vRNA signals required for the packaging in the virions. The most conserved structures have been detected in mRNAs and/or vRNAs from segments 2 (PB1), 5 (NP), 7(M) and 8 (NS) (Figures 1.1-1.4). In mRNAs encoded by segments 8 and 7, the putative conserved structure elements are spread over the whole lengths of the segments, being located in both terminal and interior areas of the coding regions (Figures 1.1 and 1.2). In segments 5 and 2, the conserved structures have been detected in the regions close to the terminal panhandle structures, both in (+)RNA and in vRNA (Figures 1.3 and 1.4). Several putative conserved local structures have been predicted in other segments as well.

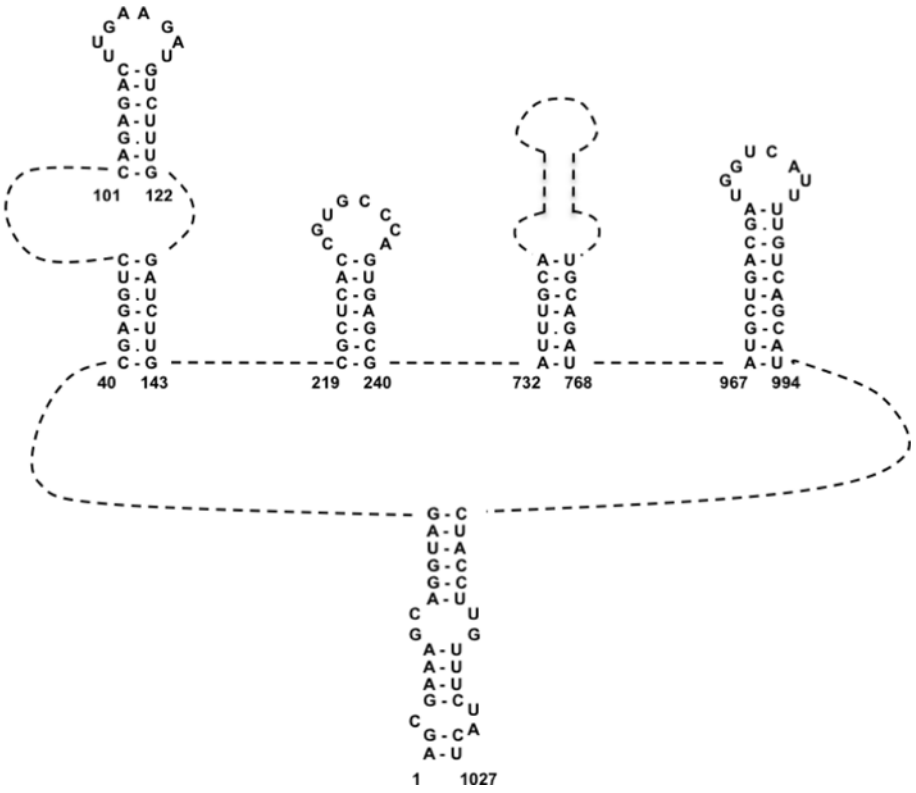


Figure 1.1. Schematic overview of the most conserved structures predicted in the segment 7 (M) (+)RNA of influenza A virus genome. The sequence from the strain A/Puerto Rico/8/34 (EF467824) is shown.

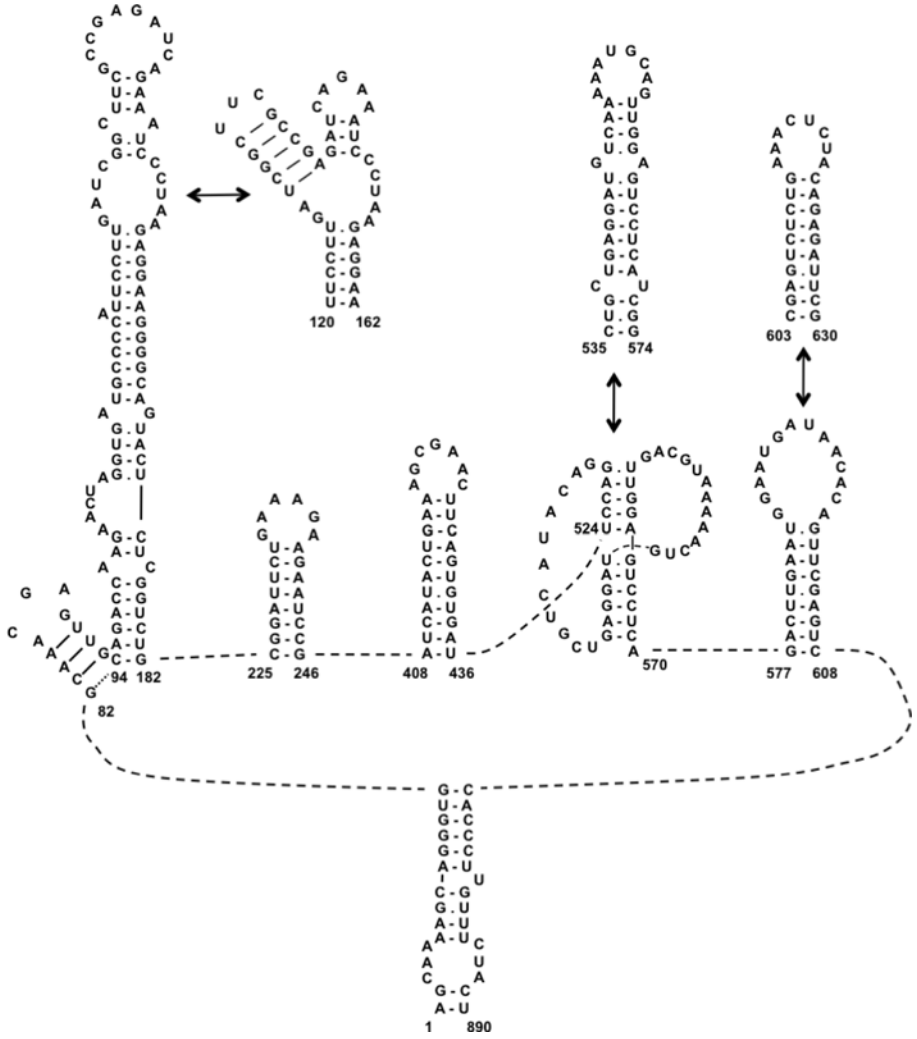


Figure 1.2. Schematic overview of the most conserved structures predicted in the segment 8 (NS) (+)RNA of influenza A virus genome. The sequence from the strain A/Puerto Rico/8/34 (EF467817) is shown. Double arrows indicate putative conformational transitions between conserved structures that are mutually exclusive.

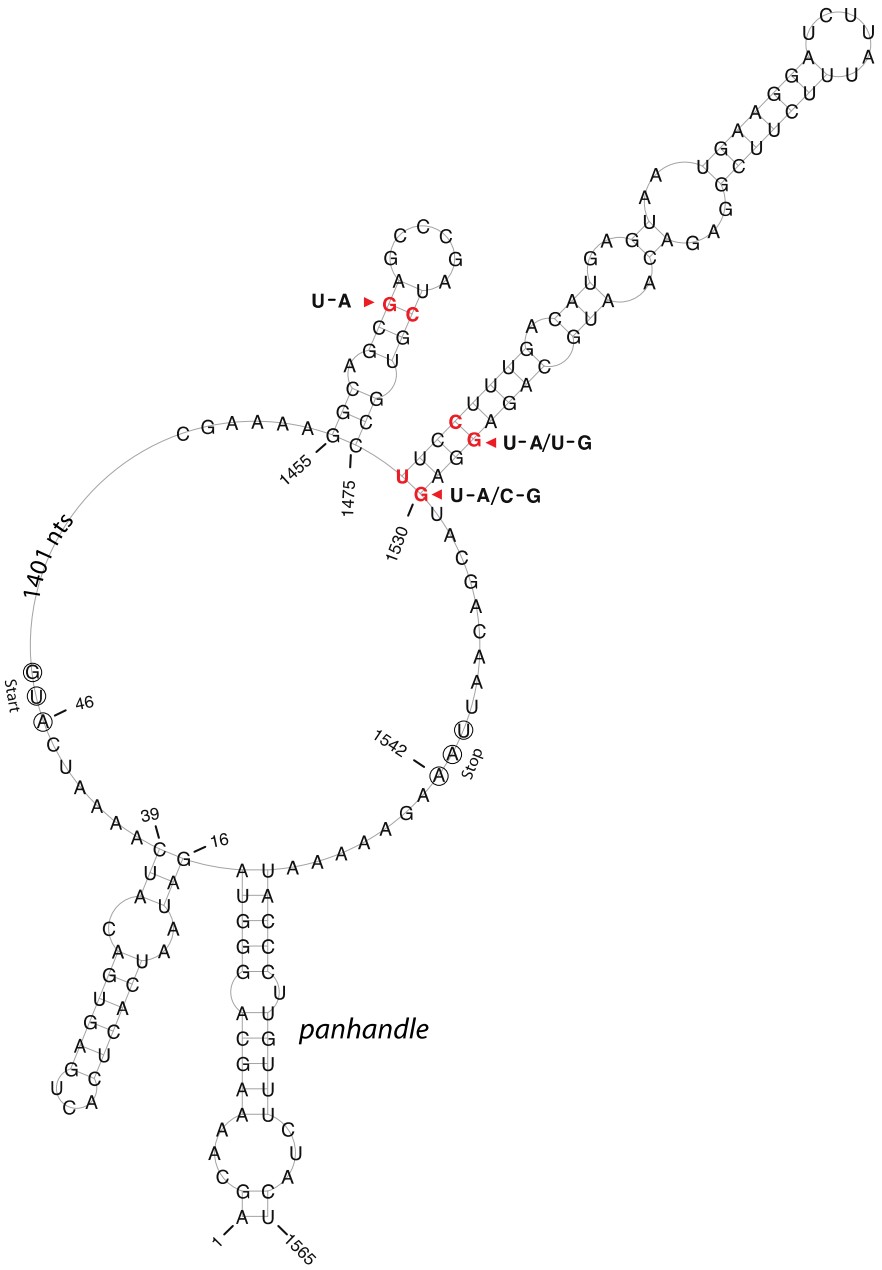


Figure 1.3. Schematic overview of the most conserved structures predicted in the segment 5 (NP) mRNA of influenza A virus genome. The sequence from the strain A/Puerto Rico/8/34 (J02147) is shown. Arrows indicate nucleotide co variations in various strains.

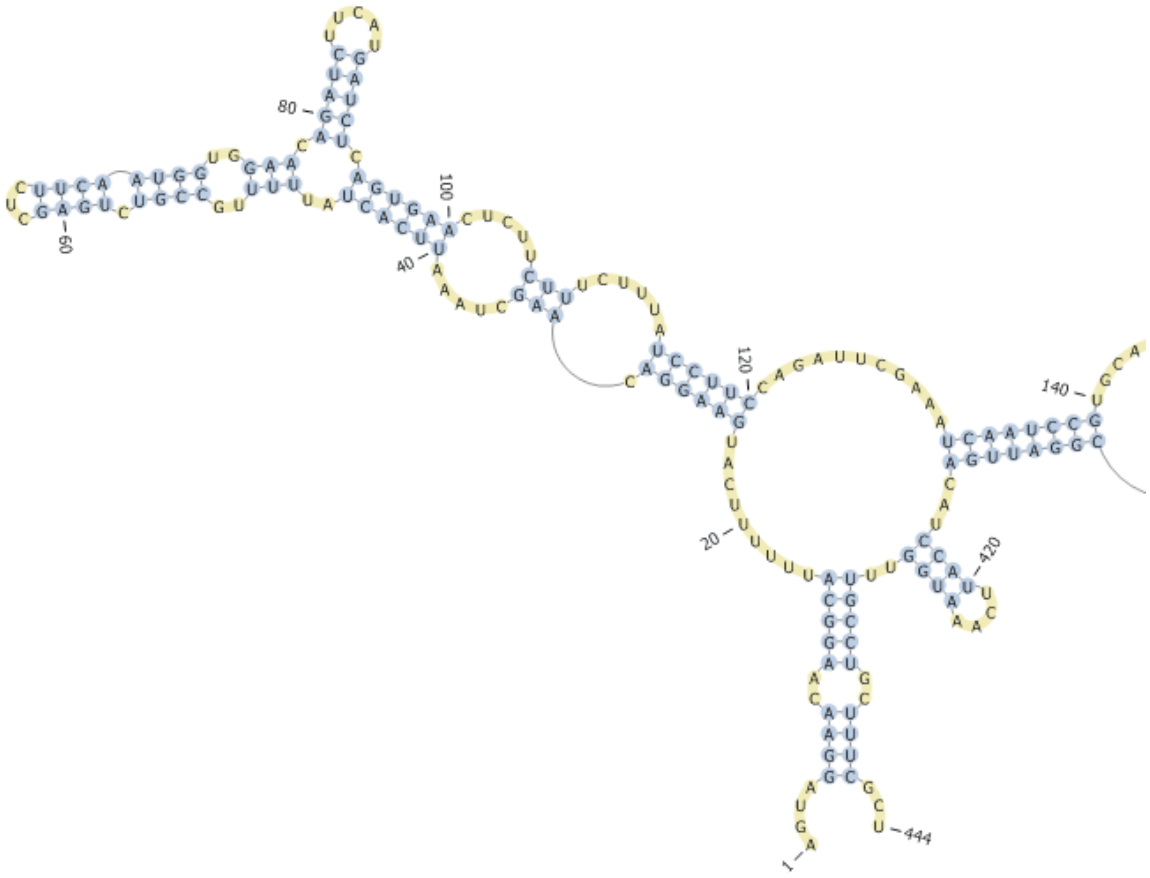


Figure 1.4. Schematic overview of the conserved structures predicted in the segment 2 (PB1) vRNA of influenza A virus genome. Here the part of the structure predicted for the defective interfering (DI) RNA of 444 nucleotides (Accession K00866) derived from the strain A/PR/8/34 is shown.

Structural and evolutionary analysis of found structures (deliverable D1.2).
Following the identification of putative consensus structures in segments of influenza A virus genome, these structures have been analyzed using evolutionary and structural comparisons. This analysis shows that influenza genome segments have a potential to encode very conserved RNA structures. In all analyzed segments the regions encoding for conserved folds have been identified, although the segments 4 and 6 seem to have less conserved RNA folding. This may be explained by the strong evolutionary pressure in these segments to encode novel antigenic signatures of encoded proteins.

On the other hand, a potential for the formation of conserved local structural elements have been detected in segments 4 and 6 as well. These local structures (mostly short hairpins) are mostly conserved within the antigenic subtype groups (H1, H2 etc. or N1, N2, etc.) and frequently host-specific. However, one structure in segment 4 vRNA turned out to be conserved in both H1 and H2 subtypes from different hosts (human, avian and swine strains).

The sequences of the segments encoding polymerase subunits PB2, PB1 and PA (segments 1-3) are more conserved as compared to other segments. Although this leads to prediction of rather conserved putative structural elements, such predictions should be interpreted with caution, because low sequence diversity is not sufficient for a strong comparative support. Nevertheless, on the basis of structure and sequence comparisons, several structures can be supported with more certainty. For instance, a small hairpin, predicted in segment 3 vRNA, is supported by a unique covariation from a G-C to A-U base pair, occurring in the human strains of 1954 (Figure 1.5). Also, it is interesting to

note that the terminal regions of positive- and negative-sense RNAs from segment 2 (Figure 1.4) are predicted to fold into consensus “mirror” structures, suggesting that these regions form well-defined conformations without a potential for alternatives.

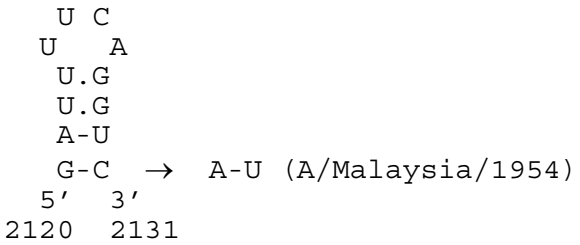


Figure 1.5. A covariation in the conserved structure in the region 2120-2131 of segment 3 (PA) vRNA.

Also, putative conformational transitions in conserved structures have been detected. Interestingly, after the sequences of the new swine-origin pandemic strains have become available, we have revealed that their segments 8 mRNAs could have a structure that differs from the one typically found in human viruses. This is due to different trends in the evolution of alternative mutually exclusive structures, that could be traced back in the available sequences of human and swine strains. One of these structures is strongly conserved in human viruses, while it is gradually disrupted in swine strains, making an alternative more favourable. This shift becomes even more pronounced in the swine lineage from which the new H1N1 pandemic viruses inherited the segment 8.

In summary, the following putative structures, conformational transitions and specific signatures have been predicted in the segments of influenza A virus genome:

Segments 1-3 (PB2, PB1 and PA)

A number of conserved local structures may be suggested, but a strong sequence conservation does not allow to build a definitive model without additional data. Neverheless, rather conserved structure can be predicted in the terminal regions of segment 2 (Figure 1.4) and a conserved hairpin folded in segment 3 vRNA is supported by a covariation (Figure 1.5). Presumably, these structures may be functional in the packaging process.

Segments 4 (HA) and 6 (NA)

In general, there is no universal structure conserved in all subtypes. Nevertheless, one stem-loop structure (region 1689-1716) seems to be conserved in the segment 4 vRNAs of H1 and H2 subtypes. It is supported by covariations, but is not preserved in the sequences of the H5 subtype, which is most closely related to H1/H2 branch in the phylogenetic tree.

Segment 5 (NP)

Highly conserved structures are detected in the regions near the terminal panhandle (Figure 1.3). Remarkably, some covariations in the stem regions exhibit host specificity. The location of structures suggests that they are likely to be involved in the vRNA packaging and/or specific nuclear export of mRNA.

Segment 7 (M)

Several conserved hairpins can be predicted in various regions of M segment, in both positive- (Figure 1.1) and negative-sense RNA. One of these hairpins 967-994 demonstrates a remarkable pattern of host-specific covariation: the base-pair 970 is typically C-G in all human strains (with exception of early strains before 1934), but it is predominantly U-A in avian viruses (Figure 1.6). The hairpin is presumably functional in vRNA and is involved in the packaging, being located in the region that includes the packaging signal(s).

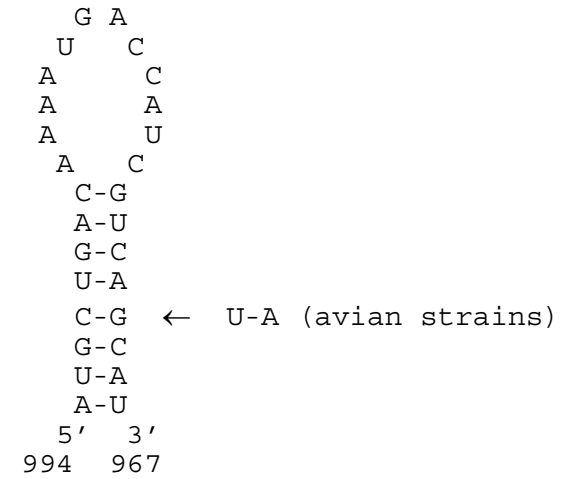


Figure 1.6. A conserved hairpin exhibiting host-specific covariation in the region 967-994 of segment 7 (M) vRNA. Numbering corresponds to the (+)-sense RNA, as in Figure 1.1 with the structure for the full-length segment.

Segment 8 (NS)

Several conserved structures can be predicted, some of them may also undergo conformational transitions. In particular, a unique G563-C substitution, observed in recent H5N1 viruses, apparently favors a pseudoknot/hairpin shift in the region 524-574 (Figure 1.2). The C563 mutation seems to be unique for H5N1 viruses, and it is possible to trace its first occurrence in China in 2001. It is stably inherited in the dominant H5N1 lineages spreading over the world, isolated as far from China as in Africa. It has been previously noted that the C563 is present in all H5N1 viruses isolated from humans since 2003 [Gultyaev et al., 2007] - this is also true for all sequences available in the Influenza Virus Resource at the moment of writing (March 2010). This means that the suggested structures are functional, and a specific balance between them is important for the virus fitness. The location of these structures near the 3'splice site suggests that they may be involved in the regulation of splicing.

The structure in the region 94-182 (Figure 1.2) is remarkably conserved, and supported by covariations in two major clades of influenza A NS sequences, A and B. The interior part of this stem-loop structure may be folded in the alternative conformations.

Upstream from the stem-loop structure 94-182 in segment 8, a small hairpin is predicted in many strains at positions 82-94 (Figure 1.2). In segments of classical swine origin, its folding seems to be further stabilized by an extension at the expense of the bottom stem of the 94-182 structure. Notably, the 94-182 stem is considerably destabilized in the pandemic 2009 H1N1 human strains, which also have segment 8 of swine origin. On the other hand, in the previously circulating human strains the hairpin 82-94 is gradually destabilized, suggesting that NS segments of human and classical swine origin are under opposite structural constraints in this area.

We believe that the structures identified so far clearly show that RNA folding play an important role the influenza virus life cycle. The obtained data demonstrate that experimental studies of influenza virus RNA structure are absolutely necessary for understanding of the virus functioning and development of novel means of influenza prevention and control.

WP2
Validation of differences in RNA secondary structure between influenza virus genomes with different pathogenic properties

Generation of T7-promoter driven expression plasmids producing influenza virus segment 7 and 8 mRNAs of different influenza isolates with different pathogenic properties and production of their full-length mRNA (deliverable D2.1)

Using plasmids carrying the respective cDNAs behind a T7 promoter (a gift from R. Fouchier, Rotterdam), we produced (+)-strand viral RNA of segments 8 and 7 from H1N1, H7N7, H5N1, and H3N2 strains. For practical purposes, these RNAs carry short 5' and 3' extensions, but these do not appear to affect the structure of the viral RNA significantly, as phylogenetically supported structures near both the 5' (nt 94-182) and the 3' (nt 811-834) showed up clearly in the probing.

Determination of the overall secondary structure of influenza virus segment 7 and 8 mRNAs and identification of major structural changes among different influenza virus subtypes (deliverable D2.2 and D2.3)

For comparative purposes, we chose four human influenza A strains to probe the structures of the viral RNA from segments 8 (NS) and 7 (M). The frequent reassortment of the viral segments implies that the evolution of the NS and M genes does not coincide with that of HA and NP, which are generally used to define the viral subtypes. Therefore, segment-specific phylogenetic trees were constructed from all (more than 5000) publicly available Influenza A sequences. These were found to agree well with trees that had been published earlier, based on only a few dozen sequences. The four strains that were chosen from different sections of the trees are: A/Puerto Rico/8/34(H1N1), A/Netherlands/219/03 (H7N7), A/Netherlands/213/03 (H3N2) and A/Viet Nam/1194/2004(H5N1).

Structure probing using enzymes and chemicals was started on segments 8 of all four strains and initially analyzed by conventional primer extension reaction using radioactive labeling followed by polyacrylamide gelectrophoresis. It was soon realized that this conventional method would not achieve complete secondary structure determination of segments 7 and 8 of these four different strains within the time frame of the project. We therefore switched to a new approach that was recently developed by the labs of Weeks and Wilkinson. This method abolished the use of radioactivity and polyacrylamide gels and allowed for larger sets of data to be analyzed simultaneously. The transition to this new method took some time,

So the four RNAs from segment 8 were treated with the chemical structure probes DMS, DEPC and CMCT and with the enzymes RNase 1, RNase T1 and RNase V1. The results of the limited modifications and digestions were examined by primer extension using fluorescently labeled primers, followed by capillary electrophoresis and visual inspection of the electropherograms using dedicated software. An example of structure probing of A/Neth/219/03[H7N7] RNA is shown for single-stranded probes in Figure 2.1.

To facilitate the interpretation of the probing data, model structures were devised for each of the probed RNAs. For this purpose, various selections of sequences from the phylogenetic clade to which each RNA belongs were analyzed using the RNAAlifold web server, which employs both energetic and phylogenetic parameters to predict common structure elements. The most conserved structures appearing from these predictions were then forced in an mfold run on each of the four RNAs, allowing mfold to make predictions for the less conserved regions. The probing data were graphically entered into the structure models and the resulting figures were examined stem by stem for support or disproof of the predicted structures. Potential structures found in predictions from other

clades of the tree, predicted earlier using different approaches or seen by eye were used to interpret the regions were the probing data disagreed with the original models.

Overall, the results indicate that regions of sharply defined structure are interspersed with stretches were interpretation of the probing data is difficult if not impossible. Remarkably, this pattern appears to coincide with the prediction results, in that the unclear regions largely correspond to the regions were no conserved structures could be found. This correspondence is not obvious on beforehand, as a non-conserved structure formed by a single RNA species could in principle yield an unambiguous probing result. We therefore tentatively interpret this finding as the result of evolutionary pressures that avoid the possible formation of competing structures where specific, functional structures are needed, whereas sections without such structures are allowed to fold in multiple ways.

Various well-defined structures have been found in the probing results, mostly agreeing closely with structures predicted using energetic and phylogenetic criteria. One hairpin, near the 3' end of segment 8 (nt 811-834) had not been previously predicted, but turned out to be well supported by the presence of two covarying base pairs within a stem of eight pairs. Both pairs are encountered as C-G, U-A and C·A, suggesting that this structure is functional in the (-)-strand, where C·A becomes U·G. Indeed, the hairpin did turn up in subsequent structure predictions based on the (-)-strand sequences. Remarkably, the availability of sequences from an extended period of time has allowed us to pinpoint the moment of change of one of these pairs from C-G to U-A as the 1940s, with variants C·A and U·G being encountered in human viruses almost exclusively during this transitory period. About the possible function of this structure we can only speculate, but if it is indeed functional in the (-)-strand, replication, transcription and packaging are the most obvious options to look into.

Another interesting structure was found both as a clade-specific prediction and as a probing result specific to one of our four RNAs. This concerns a hairpin structure (nt 47-89) that is only stable in a subgroup of the human H3N2 viruses, which includes A/Netherlands/213/03 (H3N2). Since the splice donor site of segment 8 is located in this region, the stabilization of this structure and the apparent refolding observed in our probing results may be expected to have significant consequences for splicing, unless the virus has ways of fixing the functional structure, *e.g.*, through protein binding. Unfortunately, no experimental data are yet available that provide more information.

The analyses so far have produced a number of interesting structures that may be of importance to the virus either as functional or as potentially interfering elements. We anticipate that these data can serve as a valuable basis for interpreting experimental results as well as for designing new mutational experiments, both *in vitro* and *in vivo*.

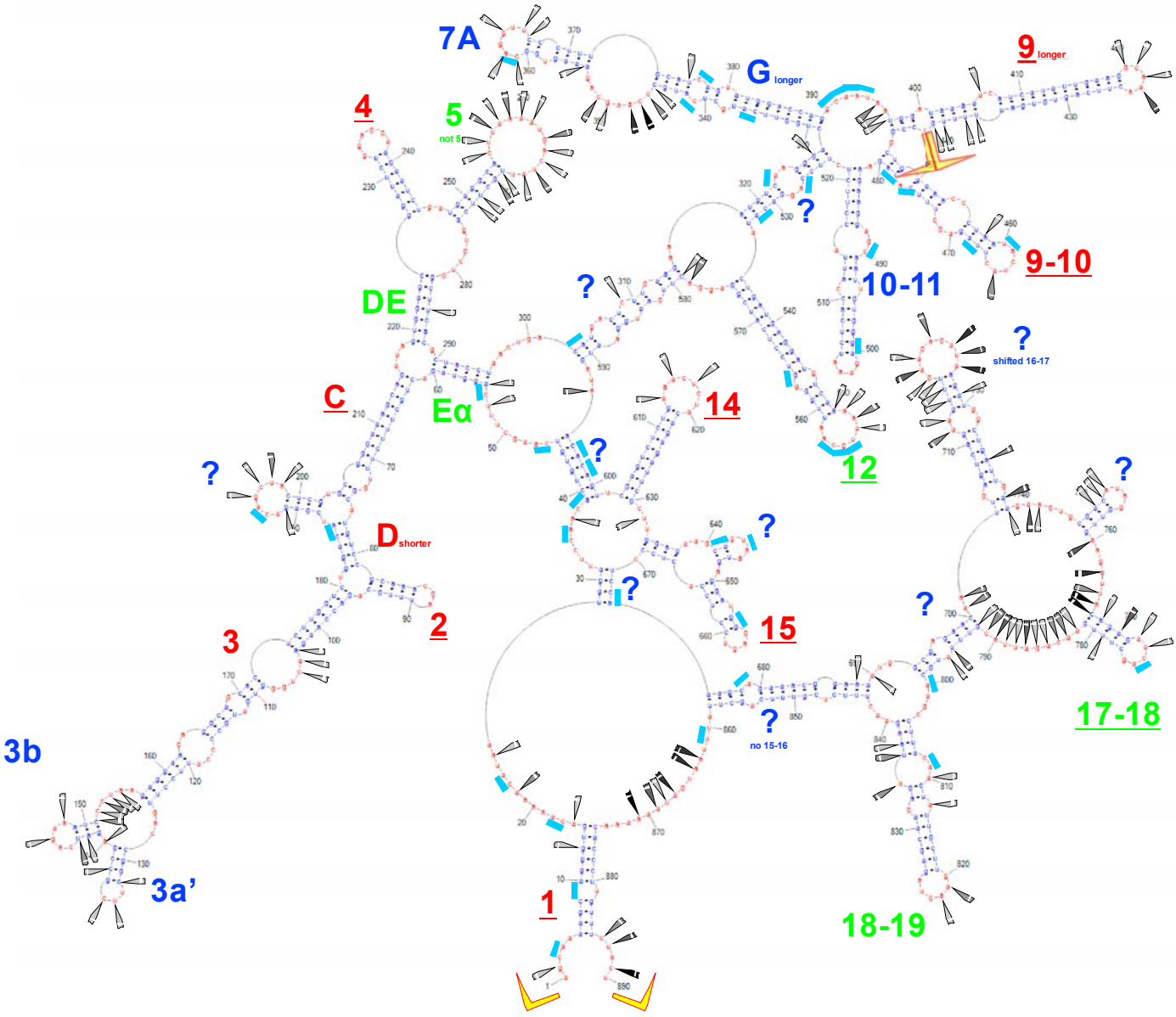


Figure 2.1. RNA: A/Neth/219/03[H7N7] segment 8(+) strand, model: conserved + Mfold, probe: RNase I (ssRNA).

WP3
Computer-assisted detection of miRNA targets in influenza genome and experimental validation

A list of potential miRNA binding sites on influenza virus RNAs (deliverable D3.1)
We first looked for perfect complementarity in all influenza sequences (> 20.000) to all currently known miRNAs from *H. sapiens* (866), and *G. gallus* (512) by BLAST searches using NCBI website. No single sequence was found to have a perfect match with any of the miRNAs. This could be an indication of counter-selection. Perfect complementarity would trigger the siRNA pathway /induce cleavage of the mRNA.
When miRNAs do not have perfect complementarity their binding can result in translational repression of the influenza mRNA but only when they bind to untranslated regions of the influenza mRNA. For this repression to occur 7 or 8 bp complementarity ("seed") can be sufficient. This makes the prediction of functional target sites extremely challenging. Algorithms have been developed by others to predict miRNA binding sites. We have used the miRanda algoritm to predict putative target sites of human miRNAs in the 5' and 3' UTRs of all influenza mRNAs.
During our analysis, we encountered several technical problems:

1. The input files (all human microRNAs and all flu sequences) were unusually large and therefore the miRanda algorithm was very slow to produce the output file.
2. The output files were extremely large (text files of up to 4 gigabytes) and analysis of the results was in practice impossible on such files.
3. When the results were analyzed, it was difficult to estimate the proportion of flu sequences containing putative miRNA target sites, since many of the flu sequences in the input files were incomplete. Therefore, it was impossible to judge whether negative results were due to the absence of miRNA target sites or to the lack of flu sequences to be recognized by the miRanda algorithm.
4. Finally, if an miRNA was predicted to have target sites in several flu sequences, it was very time-consuming to verify whether these sites were at an identical location among different flu sequences or not.

Solutions were found to answer each of these problems:

1. More powerful hardware (new MacBook Pro) was purchased, which decreased the time of running the miRanda algorithm from 3 days to 1 hour.
2. The miRacle script was written that allows to clean the miRanda output file from all the negative results and to categorize the positive hits into three major groups (human, swine, avian).
3. The Undasher script was written that allows removing sequences that are judged too incomplete and therefore susceptible to bias the results. With this script it is possible to define to what degree the length of a sequence is acceptable.
4. Prior to miRanda analysis, the flu sequences were aligned using the Mafft algorithm so that a particular region of a flu sequence would have the same localization through all similar sequences. Then only the 5'UTR or 3'UTR was extracted and an alignment of 5'UTRs or 3'UTRs was used as input into miRanda. The miRage script was written that allows visualization of the location of the putative miRNA target regions among all flu sequences.

The final number of potential miRNA target sites is rather small since most segments have only very short 5' and 3' UTRs. Exceptions are the 50-nt 5' UTR of segment 5 and the 3' UTRs of the unspliced segment 7 and 8 RNAs which are 252 nt and 219 nt, respectively.
A list of miRNAs with putative targets is shown in table 3.1. miRNAs targeting the 3' UTR of segment 7 and 8 are shown in separate tables 3.2 and 3.3, respectively.

The analysis of sequence diversity of the predicted binding sites: selection of promising microRNA target sites on influenza virus RNAs and mutational analysis of potential microRNA targets on influenza virus RNAs Deliverable 3.2-3.3 LU-1

As can be seen in table 3.1 several miRNAs had targets in more than one segment**.** miR-1468 has targets in 5'UTR of segment 1 (40%), 2 (60%), 3 (53%), 7 (21%) and 8 (5%). However, due to strong target site conservation in humans, swines and avians it is probably not related to pathogenicity. 27 miRNAs were predicted to target the 5'UTR of segment 5. The majority of these miRNAs were directed against avian strains and not to human strains. However of only 3 miRNAs some (low) expression has been detected in HeLa or A549 cells. We therefore focused on miR-433, 597, and 574-5p.

To test the effect of these miRNAs on translation, the 5'UTR of segment 5 from an H5N1 avian strain was cloned upstream of a luciferase reporter gene (Flu_A). Mutations were made in Flu_A to create perfect complementarity to miR-597 (Flu_B), to miR-433 (Flu-E), and to miR-574-5p (Flu-M). As a positive control Flu_K was constructed which has perfect complementarity to let-7 miRNA, of which we know is present in high amounts in HeLa cells. Since the putative target sites for these miRNAs are located in a base-paired region (identified in WP1 and 2) we also investigated the effect of mutation in the hairpin. In Flu_D, the hairpin was destabilized without changing the complementarity to miR597 and 433. In Flu_C, complementarity to 597 was extended without changing the stability of the hairpin. Data after transfection to Hela cells showed an 1.5 to 2-fold increase in expression of Flu_B, C, D and E relative to Flu_A. Flu-K experienced a 6-fold decrease as expected for let-7 targets (Flu-M was tested later, see also D3.4. Similar results were obtained with A549 lung cells. From this we concluded that miRNAs 597,433 were not present in sufficient amounts in Hela and A549 cells to repress translation of our reporter construct. Therefore to test the effect of these miRNAs we had to overexpress them in our cell lines (see D3.4 below).

Analysis of the longer 3'UTRs of segment 7 and 8 identified a few dozen miRNAs of which some had more targets in avian isolates than in human isolates, but also vice versa. Of thes miRNA, only the let-7 miRNAs are known to be expressed in HeLa and A549 lung cells (our laboratory cell lines). Also since more avian than human strains appeared to have targets for the let-7 family (even though highly conserved in birds as well, that is chicken) we constructed a Renilla luciferase reporter plasmid containing the 3' UTRs of 4 different isolates (two avian, H5N1 and H7N7, and two human, H1N1 and H3N2). Transfection of these plasmids into HeLa cells showed no significant differences in expression of any of the four constructs. Co-transfection of an antisense LNA oligo to "inactivate" let-7 miRNAs also did not change expression levels. Control transfections with known targets of let-7 did respond to addition of the LNA antisense, indicating that the targets in the 3'UTR of segment 8 are either not accessible or not good enough. Inspection of the base-pairing scheme of let7 miRNAs revealed that the first nt of the seed (nt 2 of let7) is actually a mismatch. This may interfere with binding of the miRNA. Another possibility is the accessibility of the site. Looking at the models we have developed in WP1 and 2 for segment 8 we can see that the target site for let-7 miRNAs is involved in the formation of hp18-19, thereby probably rendering the RNA resistant to let-7 miRNAs (figure 3.1).

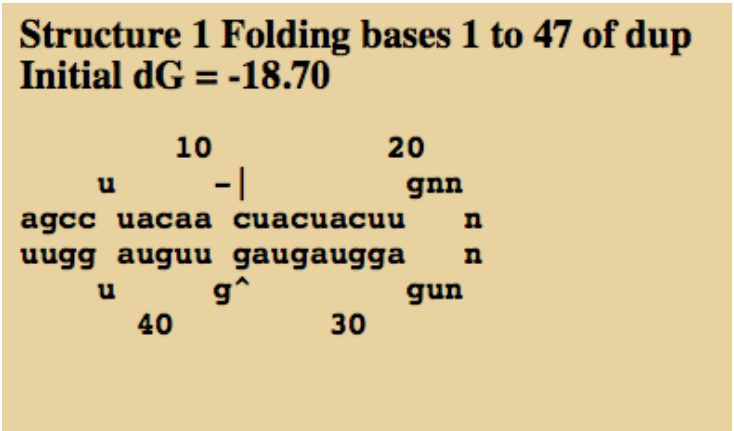


Figure 3.1. Duplex of Let7b and 3'utr segment 8, artificially linked by N5 for MFOLD application.

Deliverable 3.4-3.6 UU
Generation of plasmids expressing microRNA potentially binding to influenza virus RNAs, the effect of microRNA overexpression on influenza virus RNA levels and sensitivity to microRNA overexpression of influenza virus RNAs derived from different isolates with different pathogenic properties (month 9-18).

This part of the work package was entirely carried out by LU-1. Overexpression over miRNAs 433, 574-5p, and 597 was achieved by co-transfection of the reporter constructs with pre-miRNA precursors. Figure 3.2 shows that adding pre-miRNA 574-5p caused a ten-fold reduction in translation. This effect was specific as an antisense LNA against 574-5p could restore expression completely, whereas an unrelated LNA (MS2 LNA) could not. Similarly, pre-miRNA 597 knocked down the expression of Flu_B completely but had no effect on Flu_E. Vice versa, pre-miRNA 433 knocked down expression of Flu_E but not of Flu_B. The endogeneous levels of these miRNAs in Hela cells are apparently too low to cause any effect on our reporter constructs. In the next experiments we investigated the effect of these three miRNAs on the 5'UTR of an avian H5N1 strain (Flu_A) and a human strain (Flu_F). All three miRNAs had significantly more effect on translation from the avian 5'UTR than from the human 5'UTR. A computer predicted analysis of the putative duplexes formed by the 5'UTRs and the miRNAs indicated that in all cases the avian sequence resulted in more stable duplex formation. Interestingly, mutant Flu_C which could form the most stable duplex with miRNA-597 also had the lowest expression when co-transfected with this miRNA (figure 3.3).

We conclude that the 5'UTR of avian H5N1 strains is more susceptible for miRNAs 433, 596, and 574-5p than that of human influenza strains. Whether these miRNAs will also affect the replication of the whole virus and in which (relevant) tissues these miRNAs are active remains to be investigated.

Table 3.1 miRNAs targeting UTRs of influenza segments

miRanda results - HUMAN DATABASE OF miRNAs

	TOTAL	AVIAN	HUMAN	SWINE	REGION LENGTH	CONS. IN Chicken	CONS. IN SWINE
S1U5	**1050**	**645**	**321**	**84**	**33 nt**		
miR-1468	420 (40%)	392 (61%)	19 (6%)	9 (11%)		N	N
S1U3	**1499**	**1027**	**380**	**92**	**42 nt**		
No miR							
S2U5	**670**	**501**	**120**	**69**	**30 nt**		
miR-1468	399 (60%)	345 (69%)	37 (31%)	17 (35%)		N	N
miR-638	48 (7%)	45 (9%)	2 (2%)	1 (2%)		N	N
miR-26a-1*	43 (6%)	41 (8%)	0	2 (4%)		N	N
miR-196b	16 (2%)	16 (3%)	0	0		Y	Y
miR-584	562 (84%)	403 (80%)	113 (94%)	46 (94%)		N	N
S2U3	**1579**	**860**	**600**	**119**	**49 nt**		
miR-658	137 (9%)	120 (14%)	11 (2%)	6 (5%)		N	N
miR-1269	90 (6%)	51 (6%)	5 (1%)	34 (29%)		N	N
miR-921	251 (16%)	75 (9%)	139 (23%)	37 (31%)		N	N
miR-593*	47 (3%)	9 (1%)	3 (0%)	35 (29%)		N	N
miR-555	190 (12%)	50 (6%)	134 (22%)	6 (5%)		N	N
miR-423-5p	24 (2%)	20 (2%)	3 (0%)	1 (1%)		N	N
miR-887	97 (6%)	0	97 (16%)	0		N	N
miR-223*	940 (60%)	565 (66%)	342 (57%)	33 (28%)		N	N
S3U5	**778**	**549**	**189**	**40**	**30 nt**		
miR-1468	416 (53%)	380 (69%)	16 (8%)	20 (50%)		N	N
miR-302c*	406 (52%)	344 (63%)	34 (18%)	28 (70%)		P	N
miR-106b*	18 (2%)	5 (1%)	12 (6%)	1 (2%)		N	N
miR-15b*	148 (19%)	1 (0%)	144 (76%)	3 (8%)		N	N
S3U3	**2316**	**1434**	**744**	**138**	**70 nt**		
miR-1228*	75 (3%)	63 (4%)	2 (0%)	10 (7%)		N	N
let-7d	65 (3%)	32 (2%)	22 (3%)	11 (8%)		Y	?
miR-638	1432 (62%)	1087 (76%)	235 (32%)	110 (80%)		N	N
miR-658	332 (14%)	40 (3%)	287 (39%)	5 (4%)		N	N
miR-103	72 (3%)	31 (2%)	37 (5%)	4 (3%)		P	P
miR-34a	185 (8%)	19 (1%)	107 (14%)	59 (43%)		P	P
miR-608	55 (2%)	25 (2%)	28 (4%)	2 (1%)		N	N
let-7g	77 (3%)	38 (3%)	24 (3%)	15 (11%)		P	Y
miR-517*	352 (15%)	10 (1%)	339 (46%)	3 (2%)		N	N
miR-512-3p	391 (17%)	15 (1%)	369 (50%)	7 (5%)		N	N

1

miRanda results - HUMAN DATABASE OF miRNAs

	TOTAL	AVIAN	HUMAN	SWINE	REGION LENGTH	CONS. IN Chicken	CONS. IN SWINE
miR-593*	150 (6%)	68 (5%)	78 (10%)	4 (3%)		N	N
miR-675	123 (5%)	48 (3%)	64 (9%)	11 (8%)		N	N
miR-107	443 (19%)	133 (9%)	284 (38%)	26 (19%)		P	P
miR-509-5p	59 (3%)	54 (4%)	0	5 (4%)		N	N
miR-639	244 (11%)	0	243 (33%)	1 (1%)		N	N
S4U5	**2603**	**1456**	**964**	**183**	**23 to 50 nt**		
miR-453	84 (3%)	66 (4%)	16 (2%)	2 (1%)			
miR-557	40 (2%)	37 (2%)	2 (0%)	1 (1%)		N	N
miR-542-3p	128 (5%)	119 (8%)	5 (1%)	4 (2%)		N	N
miR-28-3p	97 (4%)	93 (6%)	1 (0%)	3 (2%)		N	N
miR-147b	47 (2%)	36 (2%)	10 (1%)	1 (1%)		P	N
miR-608	67 (3%)	60 (4%)	1 (0%)	6 (3%)		N	N
miR-940	41 (2%)	38 (2%)	1 (0%)	2 (1%)		N	N
miR-331-5p	49 (2%)	32 (2%)	15 (2%)	2 (1%)		N	N
miR-612	89 (3%)	81 (5%)	4 (0%)	4 (2%)		1 to 13	N
miR-659	379 (14%)	69 (5%)	288 (31%)	22 (13%)		N	N
miR-614	77 (3%)	69 (5%)	4 (0%)	4 (2%)		N	N
miR-296-3p	86 (3%)	78 (5%)	0	8 (5%)		N	N
miR-662	54 (2%)	54 (4%)	0	0		N	N
miR-1293	58 (2%)	53 (3%)	0	5 (3%)		N	N
miR-1225-3p	94 (4%)	85 (6%)	5 (1%)	4 (2%)		N	N
S4U3	**4230**	**1685**	**2334**	**211**	**33 to 54 nt**		
miR-939	270 (6%)	6 (0%)	264 (11%)	0		N	N
miR-769-3p	380 (9%)	0	377 (16%)	3 (1%)		N	N
miR-149*	1108 (26%)	751 (45%)	329 (14%)	28 (13%)		N	N
miR-330-5p	489 (12%)	3 (0%)	479 (21%)	7 (3%)		N	N
miR-17*	359 (8%)	6 (0%)	352 (15%)	1 (0%)		P	N
miR-1299	740 (17%)	1 (0%)	738 (32%)	1 (0%)		N	N
miR-99b*	65 (2%)	7 (0%)	6 (0%)	52 (25%)		N	N
miR-944	233 (6%)	0	233 (10%)	0		N	N
S5U5	**3585**	**1433**	**1931**	**221**	**51 nt**		
miR-124*	200 (6)	163 (11%)	8 (0%)	29 (13%)		N	N
miR-132*	735 (21%)	112 (8%)	466 (24%)	157 (71%)		N	N
miR-453	487 (14%)	486 (34%)	1 (0%)	0		N	N
miR-510	1880 (52%)	1010 (70%)	785 (41%)	85 (38%)		N	N

2

Table 3.2 miRNAs targeting the 3' UTR of segment 7

Segment 7: 3'utr long				
miRNA	**All hosts**	**Avian**	**Human**	**Swine**
miR-194*	54	77	10	41
miR-453	12	2	38	6
miR-557	13	5	32	14
miR-665	14	20	7	5
miR-637	19	4	47	34
miR-877	56	82	12	23
miR-1266	75	86	63	44
miR-645	41	58	7	26
miR-1254	7	1	24	3
miR-431*	19	5	53	22
miR-1180	16	2	38	43
miR-1909*	59	86	12	30
miR-1915	61	59	69	54
miR-766	47	61	23	30
miR-650	8	1	27	6
miR-125a-5p	28	33	18	25
miR-18a*	89	90	91	83
miR-193b	22	20	16	43
miR-28-3p	10	12	1	19
miR-324-3p	16	3	49	8
miR-93	51	62	32	34
miR-96	7	2	6	37
LEGEND	1-24%	25-49%	50-74%	75-100%

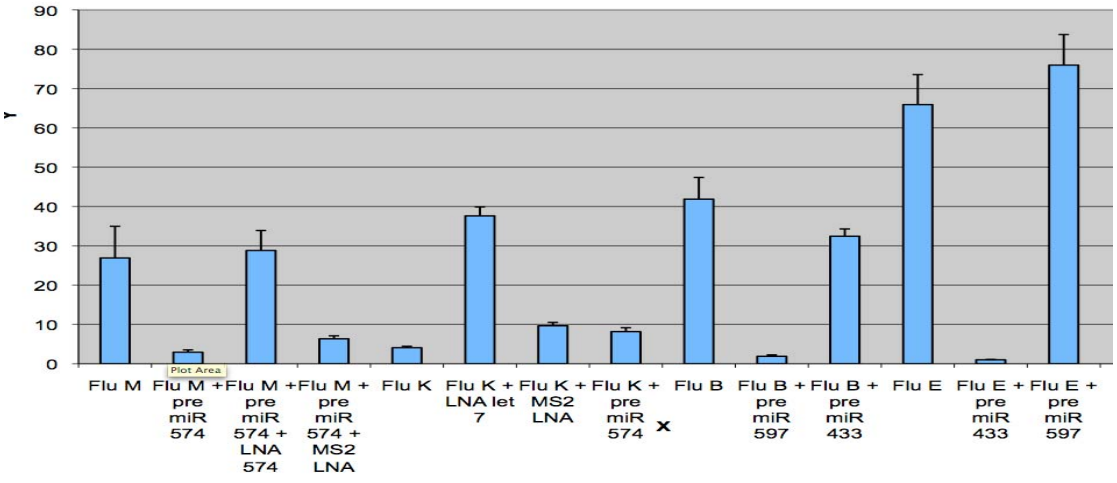


Figure 3.2. Effect of overexpression of miRNAs on translation of reporter constructs with perfectly complementary targets.

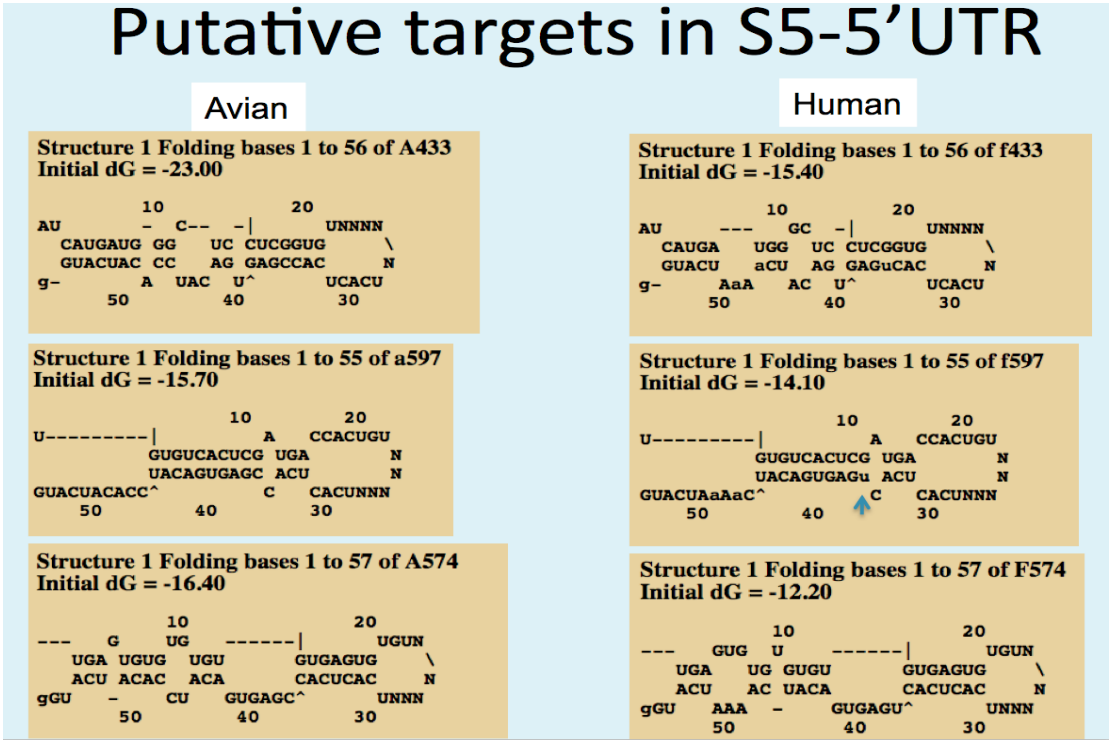


Figure 3.3. Computer stimulated prediction of duplex stability between miRNAs 433 (top), miR-597 (middle), miR-574-5p (bottom), and avian (left0 or human (right) 5'UTRs of segment 5.

WP4

Nucleotide polymorphism in influenza virus segment 7 and 8 mRNAs that cause differences in influenza virus mRNA splicing

We cloned and expressed segment 7 and 8 mRNAs of low and high pathogenic properties (Fig. 4.1) and we have expressed these mRNAs in human cells using transient transfections. We found that segment 7 and 8 mRNAs of highly pathogenic influenza virus (H1N1 Spanish flu) were inefficiently spliced compared to those from influenza viruses with low pathogenic properties (H3N2 epidemic strain) (Fig. 4.2). Since this was a major finding the same segments were analysed in an RNA polI based system in which the segment 7 and 8 antisense RNAs are produced by RNA polymerase I in order to allow transcription from the influenza virus own polymerase, coexpressed with segment 7 and 8 plasmids. These results confirmed differences in RNA expression and splicing between H1N1 and H3N2 (Fig. 4.3).

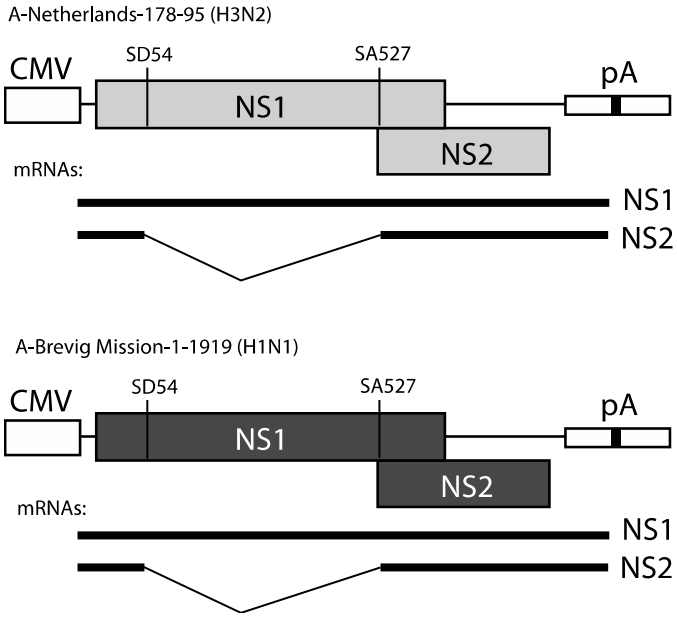


Figure 4.1. Schematic drawing of mammalian expression plasmids of influenza virus segment 8 (NS) from influenza virus strain A-Netherlands-178-95 (H3N2), a strain with relatively low pathogenicity, and A-Brevig Mission-1-1919 (H1N1), a highly pathogenic influenza virus strain. Predicted unsliced (NS1) and spliced mRNAs (NS2) are indicated. 5'-splice site is indicated as SD54 and 3'-splice site as SA527.

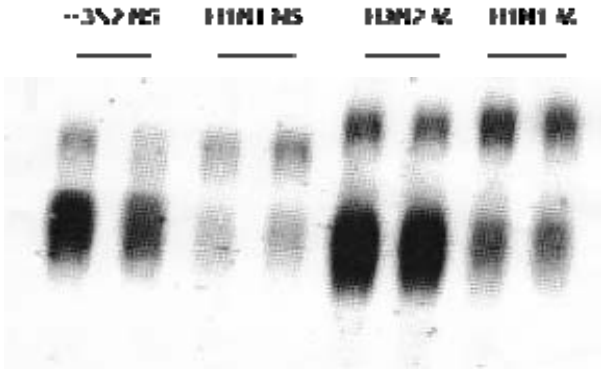


Figure 4.2. Northern blot of influenza virus RNA. HeLa cells were transfected with CMV driven expression plasmids encoding either influenza virus H3N2 or H1N1 (Spanish flu), segment 7 (M) or segment 8 (NS) cDNA. RNA levels were monitored by Northern blotting. Upper band represents unspliced M1 or NS1 mRNA and lower band represent spliced M2 or NS2 mRNA.

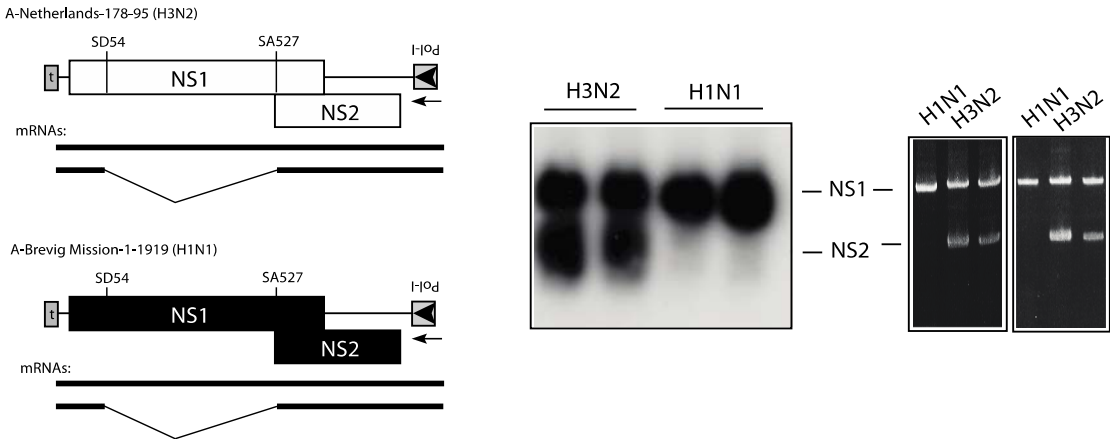


Figure 4.3. Schematic structure of RNA polI plasmids, Northern blot and RT-PCR on RNA obtained from cells transfected with the RNA polI plasmids in the presence of plasmids expressing influenza virus RNA polymerase. HeLa cells were transfected with the indicated expression plasmids encoding either influenza virus H3N2 or H1N1 (Spanish flu), segment 8 (NS) cDNAs. Upper bands in the Northern blot or RT-PCR gels represented unspliced NS1 mRNA and the lower bands represented spliced NS2 mRNA.

Segment 7 is 982 nucleotides and 93% identical at the nucleotide level. Therefore, there are a total of 68 nucleotide differences between segment 7 of the two viruses that could contribute to the difference in splicing efficiency. With the help of the hybrids (Fig. 4.4) we have mapped one region that affects splicing between the two viruses. This region is located between nucleotide positions 300 and 425 (where the A in ATG is +1). This sequence encodes a total number of 9 nucleotide differences, which result in 2 amino acid differences. The majority are G to A or C to T changes.

Segment 8 hybrids have been constructed and analysed. These results demonstrated that hybrids produced mRNAs that were spliced with different efficiency. The inefficient splicing of H1N1 was due to sequences in the last 2/3 of the H1N1 segment 8.

Avian influenza virus splicing efficiency

The splicing efficiency of two avian H5N1 influenza virus segment 8 mRNAs were compared. The two H5N1 clones were selected as they displayed big differences in the sequence immediately upstream of the major 5'-splice site of segment 8. There were 8 nucleotide differences causing two amino acid differences in both NS1 and NS2 proteins. Additional mutations were constructed to find out whether the effect on splicing was at the RNA or protein levels. This is not a trivial question for segment 8 as NS1 protein has been shown to be involved in mRNA splicing. Inactivation of the start codon and introduction of premature stop codons in NS1 did not seem to affect splicing, suggesting that the difference in splicing was caused by the nucleotide sequence rather than the amino acid sequence difference.

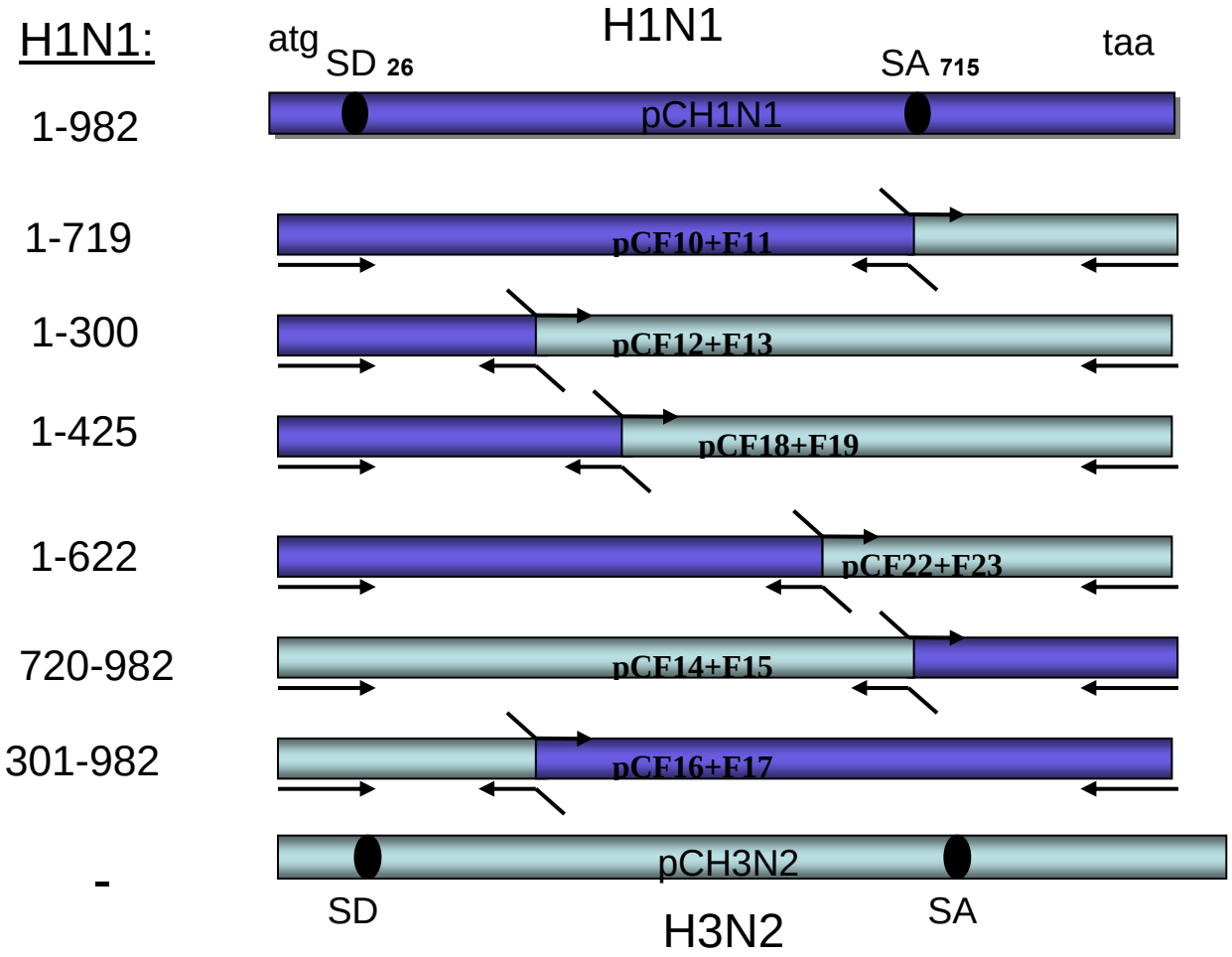


Figure 4.4. Hybrids between H1N1 Spanish flu and H3N2 segment 7.

Cellular proteins involved in avian influenza virus RNA splicing

To identify proteins that are involved in influenza virus mRNA splicing, cDNAs for all proteins in the important splicing factor family of SR proteins were acquired and the effect on influenza virus mRNAs splicing was determined in transient transfection experiments. Expression plasmids producing SR-proteins SRp20, SRp30c, SC35, ASF/SF2, SRp40, SRp55 and SRp75 were cotransfected with virus plasmids expressing avian H5N1 influenza segment 8 mRNAs. The results revealed that overexpression of ASF/SF2, SRp30c and SRp40 had a strong inhibitory effect on influenza mRNA splicing (Fig. 4.5), demonstrating that these cellular splicing factors are involved in the regulation of influenza virus mRNA splicing. Of particular interest was SRp30c, a modular splicing factor with an RNA-binding domain and an activation domain. Overexpression of an inactive mutant SRp30c protein that contains the RNA-binding domain, but lacks activation domain (and therefore is believed to interfere with endogenous SRp30c protein by competing for RNA binding sites), not only prevented mRNA splicing, but reduced the levels of unspliced influenza virus mRNA molecules in the cytoplasm. SRp30c may therefore contribute to both splicing regulation and nuclear export of unspliced influenza virus mRNAs.

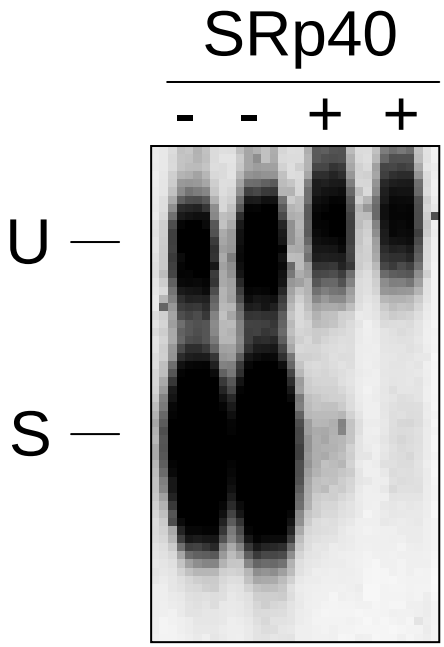


Figure 4.5. Northern blot showing unspliced (U) and spliced (S) segment 8 mRNAs from H5N1 in the absence or presence of SRp40 overexpresssion.

Summary

1) There is a difference in splicing efficiency between H1N1 and H3N2 segments 7 mRNAs. This difference is largely attributed to a 125-nucleotide region in the M1 coding region which displays 9 nucleotide and 2 amino acid differences. The effect on splicing is believed to be due to one or more of the nucleotide differences as the M1 protein itself is not involved in mRNA splicing.

2) Nucleotide differences between two segment 8 mRNAs derived from either a highly pathogenic H5N1 virus isolated from a human or from a low pathogenic H5N1 isolated from a goose, caused differences in mRNA splicing.

3) ASF/SF2, SRp30c and SRp40 are involved in regulation of splicing of influenza virus segment 8 mRNAs. SRp30c is possibly involved in nuclear export of unspliced influenza virus mRNAs.

Conclusion:
Influenza virus isolates with different pathogenic properties display differences in splicing efficiency of segment 7 and 8 mRNAs.

WP5. Effect of natural RNA sequence variation on influenza virus RNA function

The general aim of our part of the project ***"Effect of natural RNA sequence variation on influenza virus RNA function"*** *was* to characterize the role of 5´-untranslated regions (5'-UTRs) sequences of influenza virus mRNAs in the translational enhancement mediated by the viral NS1 protein and their correlation to viral pathogenic properties.

This part had three milestones:

(M5.1) Does the 5´UTRs of influenza virus mRNAs control the translation of their coding regions? MONTH 3-18.
(M5.2) Is the NS1 translational enhancement of the different RNA segments, mediated by sequence variation inside these 5´UTRs? MONTH 3-18.
(M5.3) Do the 5´UTRs has a role on viral pathogenesis? MONTH 12-24.

Our working model to study the contribution of the 5'UTR sequences on the translation of influenza virus mRNAs was based on a transfection/infection protocol. We expressed under the RNA polymerase I promoter, an antisense chloramphenicol acetyl transferase gene (CAT) containing the 5' and 3' UTR sequences of different influenza virus segments, in such a way that its expression was dependent on influenza virus polymerase activity. This allowed us to transfect different plasmids that gave raise to viral mRNAs with mutations on the 5'UTR and to infect the transfected cells with different influenza viruses that express wt or mutants NS1 proteins.

In the previous report we showed that we had set up the protocol and that we had obtained a collection of different mutants in this 5'UTR sequence, corresponding to the NS and M segments of influenza virus.

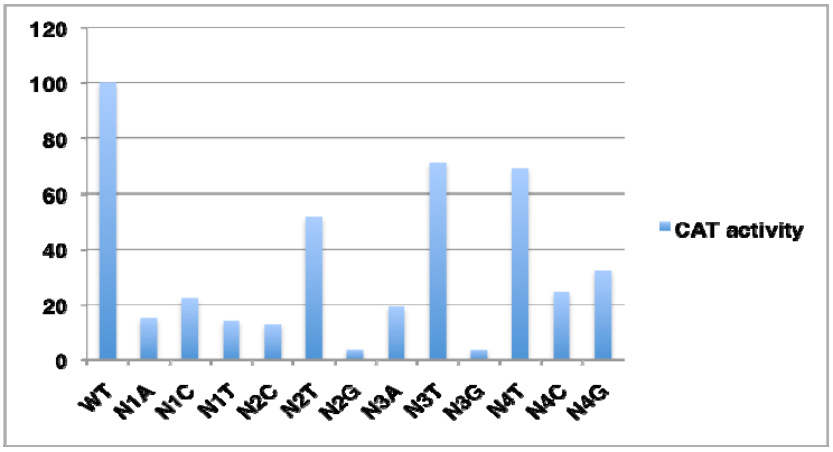
This collection provided us a way to get the answers to the questions that we proposed in our project:

M5.1 Does the 5´UTRs of influenza virus mRNAs control the translation of their coding regions?

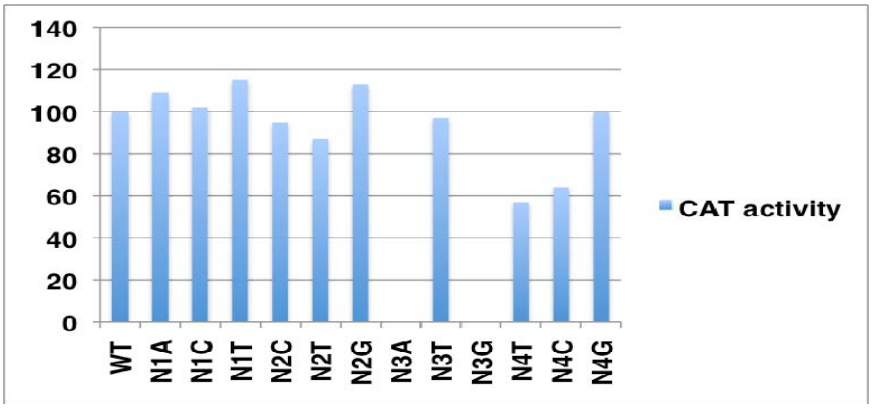
To answer that we have transfected HEK293T cells with the following plasmids:

M-CAT $^{5'}m^7G_{ppp}X^m(N_{10\text{-}13})$AGCGAAAGCAGGU**A^{1}G^{2}A^{3}U^{4}AUUGAAAG***AUG*
NS-CAT $^{5'}m^7G_{ppp}X^m(N_{10\text{-}13})$AGCAAAAGCAGGGU**G^{1}A^{2}C^{3}A^{4}AAAACAUA***AUG*

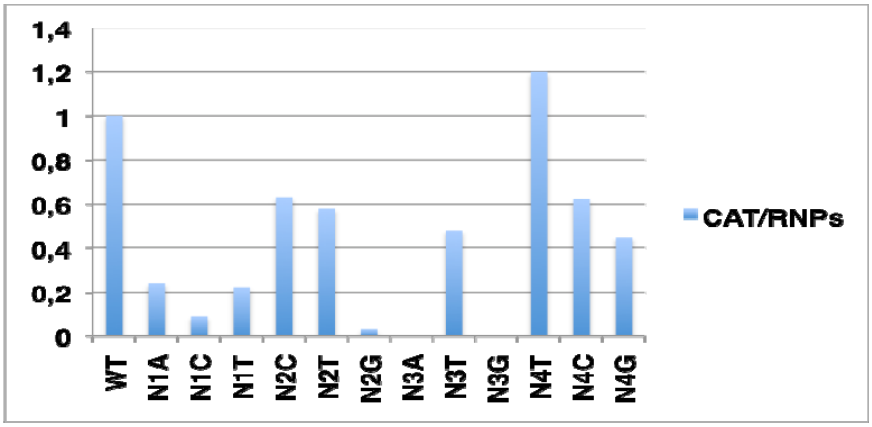
that after influenza virus infection gave raise to viral mRNAs, where positions 1, 2, 3, and 4 were changed to the three different possible nucleotides. After that, the CAT activity of the different **NS-CAT** mutants was measured and the data obtained were the following:



To monitor the translational efficiency it was necessary to know the amount of the corresponding mRNAs generated to obtain the CAT activity/CAT mRNA ratio. To that aim we first had to make a reconstitution of viral RNPs using the different NS and M-CAT mutants and assay the CAT activity. This was a measure of the viral polymerase activity including CAT transcription. The resukts obtained were the following:

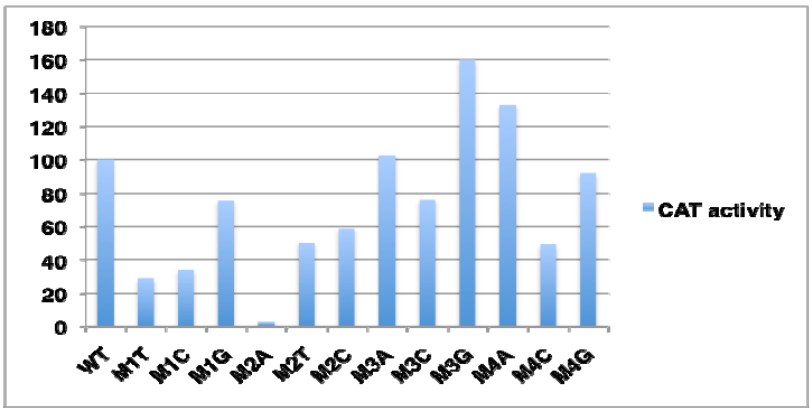


The results indicated that there are no major differences in the transcriptional efficiency of the different NS-CAT mutants, with the exception of mutants N3A and N3G. We had to determine the amount of cytosolic mRNA to quantitate the CAT mRNAs by RT-PCR. The results showed that no big differences were found. We had to set the RT-PCR conditions to determine the cytosolic CAT-mRNA levels. To distinguish between the two different CAT RNAs of positive polarity, the mRNAs and the cRNAs, a collection of different primers had to be tested to make sure that we specifically quantified the mRNAs. Doing the ratios of the CAT activities obtained in the transfection/infection system and in reconstituted RNPs, we could estimate the translational efficiency. These ratios were the following:

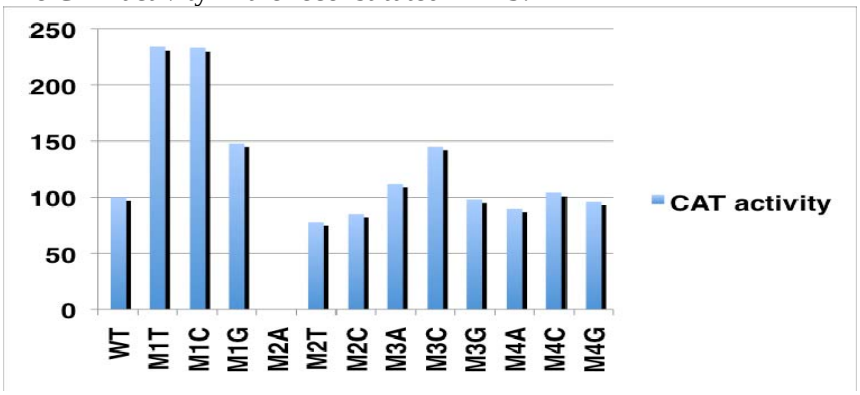


The results indicated that, with the exception of mutants N3A and N3G that had a very reduced activity in the reconstituted RNPs, the mutants were very good candidates to be defective in translation, with the exception of mutant N4T. Mutants N1A, N1C, N1T and N2G were the most affected ones. Similar studies have been done using the **M-CAT** collection of mutants and the results were the following:

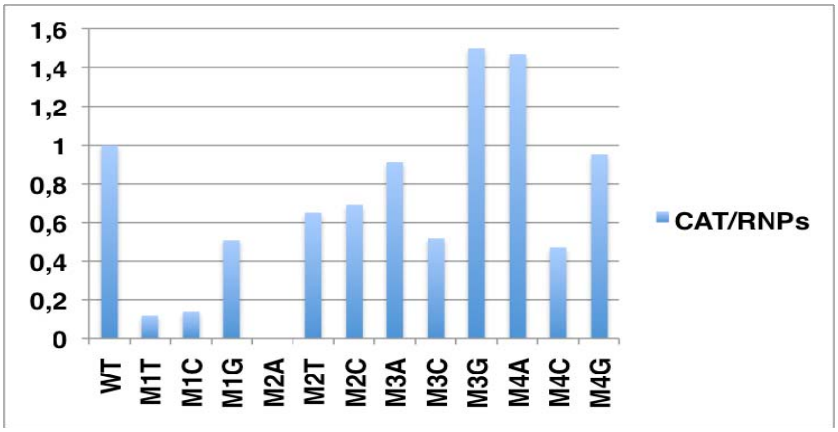
The CAT activity in the transfection/infection system:



The CAT activity in the reconstituted RNPS:



and the ratios:



Similarly to the results described for the NS-CAT mutants, mutant M2A could be discarded since it was a mutant of polymerase activity. In the case of these constructs, most of them should not be defective for translation with the exception of mutants M1T and M1C. We also approached

milestone 2: **(M5.2)** Is the NS1 translational enhancement of the different RNA segments, mediated by sequence variation inside these 5´UTRs?
We found that in the transfection/infection system, the CAT activity obtained with NS-CAT and M-CAT wild type was much lower when infection was carried out with a virus lacking NS1. However, after evaluation of different NS-CAT mutants it seemed that the differences between wt and mutants were similar when infected with wild type or the virus lacking NS1. In contrast, in the case of M-CAT mutants it seemed that NS1 coul differentially modulate the translation of the different mutants.

Summary

1. - With these data we can concluded that the 5'UTR sequences that lie outside the promoter can affect the translational efficiency of influenza virus mRNAs, being the first position the most important for translational regulation. In addition some positions also have a profound effect on viral polymerase activity.

2. - In agreement with that, these positions are absolutely conserved in NS and M segments of different influenza A viruses, such as the 1918 pandemic virus, the avian and human H5N1 or the pandemic H1N1 virus of 2009. These data are in agreement with their role in a proper viral life cycle since changes in these sequences would produce attenuated viruses.

3. – NS1 could modulate translation of M segment mediated by sequences present in the 5'UTR of that segment.

Potential impact

Due to the high variability of influenza virus and its ability to cross species barriers, new influenza virus types with different pathogenic properties may arise quickly. It is of paramount importance to quickly identify influenza viruses that are highly pathogenic and/or are efficiently transmitted between humans at an early stage of an epidemic. Knowledge of viral properties that determine its pathogenesis is therefore necessary. It is still not clear why the Spanish flue virus or why highly pathogenic H5N1 bird flue were so deadly. It is generally believed that differences in pathogenicity are caused by mutations in the viral genome that affect structure and functions of the viral proteins. Here we have investigated if mutations in the viral genome can affect the structure and functions of the viral RNAs. If so, such mutations could potentially affect the pathogenic properties of the influenza virus.

We have investigated if RNA sequence differences between highly pathogenic influenza viruses and low pathogenic influenza viruses affect virus RNA structure (primary structure and microRNA binding sites and secondary structure) and function (translation, stability, splicing), and we have investigated if these sequence differences, correlate with the pathogenic properties of the virus.

We have found that differences in RNA sequence between highly pathogenic and low pathogenic influenza virus strains affect both structure and function of the viral RNAs. We speculate that such differences in viral RNA function may affect the pathogenic properties of the virus. The research results obtained here, pave the way for experiments designed to test if mutations that affect viral RNA structure and function also affect the pathogenic properties of the virus. The contribution of the viral RNA sequence to the virus pathogenic properties has not been studied previously and may have been overlooked.

This knowledge may lay the ground for new diagnostic tools that can quickly identify a highly pathogenic influenza virus strain that has entered the human population, thereby contributing to measures that stop the transmission of the virus in the population.

Main dissemination activities

Scientific articles:

1. A family of non-classical pseudoknots in influenza A and B viruses. Gultyaev AP, Olsthoorn RC. RNA Biol. 2010 Mar 25;7(2). [Epub ahead of print]

2. Attenuated strains of influenza A viruses do not induce degradation of RNA polymerase II". A. Rodriguez, A. Pérez-González, M. Jaber Hossain, L-M. Chen,T. Rolling, P. Pérez-Breña, R. Donis, G. Koch and A. Nieto. J. Virol (2009) 83, 11166-11174.

Results have also been presented at international and national scientific meetings:

Presentations at conferences:

Contribución de los extremos 5'UTR de los mRNAs del virus de la gripe en la activación traduccional mediada por la proteína NS1. P. Rodriguez, E. Yángüez and A. Nieto. X-Congreso Nacional de Virología. Salmanca. Spain 2009.

Splicing in high and low pathogenic influenza A viruses. Ellenor Backström, Susanne Tingsborg, Samir Abdurahman, Sofia Lindström, Anna Tranell and Stefan Schwartz. Oxford Virology Symposium - From Theory to Therapeutics, Christ Church College, Oxford, U.K. 14th - 16th April, 2010.

Splicing in high and low pathogenic influenza A viruses. Ellenor Backström, Susanne Tingsborg, Samir Abdurahman, Sofia Lindström, Anna Tranell and Stefan Schwartz. RNA 2010 annual meeting, Seattle, WA, USA, 22th - 66th June 2010.

Contribution of the 5'UTR of the influenza virus mRNAs in the translational activation mediated by NS1 protein. Rodríguez P., Yánguez E. and Nieto A. Negative Strand Virus meeting 2010. Brugge. Belgium, June, 2010