# Detection of structural constraints and conformational transitions in the influenza virus RNA genome using structure predictions and mutual information calculations

Alexander P. Gultyaev*, Rene C.L. Olsthoorn, Monique I.J. Spronken, and Ron A.M. Fouchier

Erasmus Medical Center, Department of Viroscience,
P.O. Box 2040, 3000 CA Rotterdam, The Netherlands and
Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden,
The Netherlands
{a.goultiaev,m.spronken,r.fouchier}@erasmusmc.nl
{olsthoor}@chem.leidenuniv.nl

**Abstract.** Influenza A virus is characterised by remarkable genome diversity in multiple virus strains. New strains are continuously evolved, occasionally leading to dangerous outbreaks and pandemics. Local RNA secondary structures were predicted in several regions of the influenza A virus genome. The conserved structures may be functional in either genomic negative-sense RNA (e.g. involved in vRNA packaging) or positive-sense mRNA, regulating gene expression. As thousands of genomic sequences are available, it is important to estimate the extent of structure conservation and statistically significant structural constraints. Such constraints are detected using calculations of mutual information and entropy values for the paired nucleotide positions in the conserved structures. Furthermore, these calculations demonstrate the presence of evolutionarily conserved alternative conformations such as pseudoknot/hairpin equilibria or alternative hairpins. Remarkably, conformational transitions in some of the influenza virus lineages that determined major virus outbreaks can also be predicted. The analysis of influenza virus RNA structure evolution shows that it has an epochal character resembling the evolution of its antigenic properties.

## 1 Introduction

Influenza A virus is an important pathogen infecting various animals, including humans. Wild aquatic birds are the main reservoir of the virus, that continuously produces new strains able to cause a disease in other species [1]. Especially dangerous are the outbreaks of novel virus variants in humans and in poultry.

The virus genome consists of eight negative-sense RNA segments. The influenza A viruses are classified according to the subtypes of two segments encoding the surface glycoproteins, haemagglutinin (HA) and neuraminidase (NA).

The main known subtypes of human viruses are H1N1, circulated in 1918-1957 and since 1977, H2N2 (1957-1968) and H3N2 (since 1968). All these subtype changes resulted in pandemics, and in 2009 a pandemic was caused by a novel H1N1 variant. More subtypes (H1-H16 and N1-N9, with various combinations) circulate in birds, occasionally leading to major outbreaks in poultry and presenting continuous threat for humans [1]. Various virus subtypes are able to exchange their genomic segments upon infection of host cells, leading to generation of new gene constellations via reassortment.

Similar to what is observed in other RNA virus genomes, it has been suggested that influenza A virus genome, including its protein-coding parts, has regions of functional higher-order RNA structures [2-5]. Existence of such structures may be an important factor in virus evolution. Here, we attempt to determine potential contribution of structural constraints in the virus genome evolution using structural models that present the evidence of coordinated variation in the paired nucleotides (covariations). The patterns of influenza A virus genome variation can be studied using available data on thousands of different strains, circulated in different periods of time in various hosts [6]. In order to detect statistically significant correlations, we use the calculations of mutual information (MI) and entropy values for the sequence alignment positions. In this work, we focus on two smallest segments of influenza A virus genome, that were suggested to have conserved structured domains, in particular, because of regulation of alternative splicing [2-5, 7-8].

## 2 Methods

The sequences of M segment (segment 7, 1027 nucleotides) and NS segment (segment 8, 890 nucleotides) were retrieved from the NCBI Influenza Virus Resource [6]. Conserved structures were predicted using RNAalifold program [9] with various datasets of representative strains. Repeated prediction of local structural elements by RNAalifold in different datasets and the presence of compensatory changes (covariations) was used to identify potential conserved structures. It should be emphasized that the actual number of functional structures could be higher, but here we restricted our analysis only to those that could be reliably supported by covariations and thus reveal the evolutionary constraints imposed by RNA folding. The alternative structure formation was also studied by the free energy minimization program Mfold [10].

The covariations detected in the representative datasets were further evaluated using all available sequences. The segment sequences were aligned using the multiple alignment program of the NCBI Influenza Virus Resource [6], using the options to sort the strains according to virus subtype, host, time of isolation, geographic origin etc. This allowed us to detect important evolutionary transitions. The numbers of database sequence entries containing specific nucleotide combination at the putatively paired positions were counted and used for mutual information calculations. In order to minimize the influence of the database

bias on the computation, the overrepresented 2009 pandemic H1N1 strains were filtered out.

The MI and entropies in the alignment columns were calculated as described by Gutell et al. [11]. The mutual information at two nucleotide positions $x$ and $y$ was calculated as follows:

$$M(x,y) = \Sigma f(b_x b_y) * log_4 \frac{f(b_x b_y)}{f(b_x) * f(b_y)}, \quad (1)$$

where $f(b_x)$ and $f(b_y)$ are nucleotide occurrence frequencies ($b_x, b_y$ = A, G, C, or U) and $f(b_x b_y)$ are the frequencies of nucleotide combinations at the positions $x$ and $y$. For given $x$ and $y$, the summation is taken over all observed $b_x b_y$ combinations. This value is maximum M(x,y) = 1 for the perfect correlation with f(AU) = f(UA) = f(GC) = f(CG) = 1/4, and the minimum M(x,y) = 0 if no correlation is seen, that is, if for all combinations $f(b_x b_y) = f(b_x) * f(b_y)$.

As M(x,y) may have low values even at correlated positions due to strong sequence biases, the ratios $R_1(x,y)$ and $R_2(x,y)$ [11], that take the entropy terms for each of the positions into account, were calculated. The formula for M(x,y) can be rewritten in the following way:

$$M(x,y) = H(x) + H(y) - H(x,y), \quad (2)$$

where H is an entropy term

$$H = \Sigma f(b) * log_4[f(b)] \quad (3)$$

that describes variation at a nucleotide position or in a base pair. The entropy values allow to estimate the correlation at two (probably biased) positions using two ratios: $R_1(x,y) = M(x,y)/H(x)$ and $R_2(x,y) = M(x,y)/H(y)$. Both ratios are between 0 and 1, higher values correspond to more significant correlations. In general, the values are not equal to each other.

## 3 Conserved structures in the M segment

Conserved structures have previously been predicted in several regions of the M segment [4, 5, 7, 8]. These structures are located in the regions of splice sites and in the packaging signal region. A host-specific covariation has been suggested to occur in the hairpin predicted in the packaging signal region [4]. In this hairpin, the base pair at positions 970 and 991 is predominantly UA in avian strains, while human viruses have mostly CG (Fig. 1).

The study of all currently available M segment sequences confirms statistical significance of the 970/991 covariation. Indeed, the overwhelming majority of avian and swine strains contain UA base pair and human viruses have predominantly CG (Fig. 1). The number of exceptions from this rule is very small: 204 out of 20894 sequences, that is, less than 1%. The total number of sequences with a non-Watson-Crick pair at this position is even smaller: 86 (0.4%). This is reflected in the mutual information value (0.47), which can be considered as rather

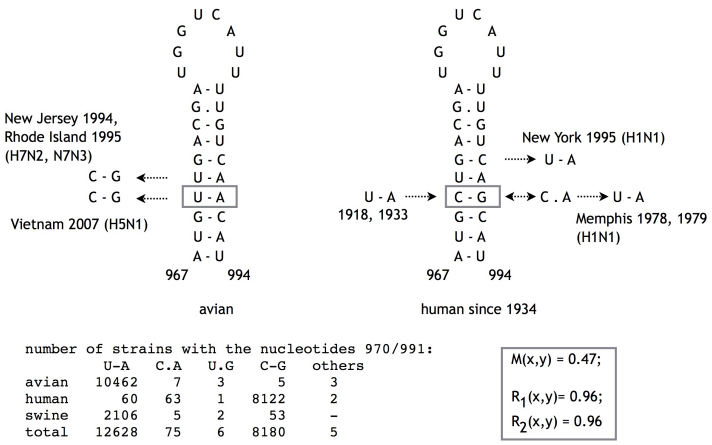


**Fig. 1.** Conserved hairpin in the packaging signal region of the influenza A virus M segment. Host-specific covariation 970/991 is boxed.

high, taking into account that the nucleotides at both positions mostly fluctuate between only two possibilities. Both ratios $R_1(x, y)$ and $R_2(x, y)$ between mutual information and entropy values at the two positions (see Methods), that take this bias into account, reach the value of 0.96, thus close to the maximum of 1 corresponding to a perfect correlation.

Such high values indicate a significant correlation between two nucleotides, suggesting a considerable constraint on variability of two nucleotides, imposed by RNA structure. Interestingly, the early human strains (isolated in 1918 and 1933) still contain the avian-like UA pair, presumably inherited from an avian precursor of these viruses [e.g. 1]. In the later human viruses, an UA pair is observed only in two Memphis strains of 1978 and 1979, apparently formed via CA mismatch (Fig. 1), other UA-containing M segments of human viruses have swine origin [e.g. 12, 13]. Several covariations at the positions 970-991 and two other base pairs in the hairpin stem have occurred in avian and human viruses (Fig. 1). Being independent evolutionary events, these covariations further confirm functionality of the predicted structure and its influence on the sequence variability. However, bearing in mind the number of analyzed sequences, population size and circulation time, the number of independent covariation events is rather small. It should be noted that the presence of C970-G991 pairs in swine strains in contrast to the predominant UA pair is determined by human virus origin of some swine virus lineages [e.g. 14].

Small number of compensatory changes preserving the base-pairing was also observed in the structures suggested in the splice site regions of M segment, where single mutations outnumber covariations [5, 8]. This conclusion is consistent with our calculations of mutual information values showing that structural constraints are considerably weaker in these conformations than those in the

967-994 hairpin. Thus, the best values for covariations from the data of Moss et al. [8], are M(x,y)=0.08; $R_1(x,y) = 0.17$ and $R_2(x,y) = 0.15$, showing no significant correlation between paired positions.

## 4 Conserved structures in the NS segment

### 4.1 Alternative structures in the NS 3' splice site region

Two alternative structures, a pseudoknot and a hairpin (Fig. 2A,B) have previously been suggested in the 3' splice site region of NS segments of influenza viruses [2, 3]. Both structures are supported by covariations between sequences from two distinct clusters of influenza A virus NS segments (clade A and clade B) and influenza B virus NS segments. Furthermore, the folding of these alternatives in vitro was shown by experiments with model oligonucleotides [2].

A conformational transition shifting the equilibrium between two alternatives have been suggested to occur in the lineage of avian influenza A H5N1 viruses that have spread over the world and remain endemic in Asia. The transition is caused by a single G-to-C mutation at position 563 (Fig. 2), that destabilises the pseudoknot in favour of more stable hairpin, and can be induced in in vitro experiments [2]. The C563 mutation is stably nested in several lineages of highly pathogenic H5N1 viruses since 2001 and is present in all available NS sequences from H5N1 infections of humans since then.

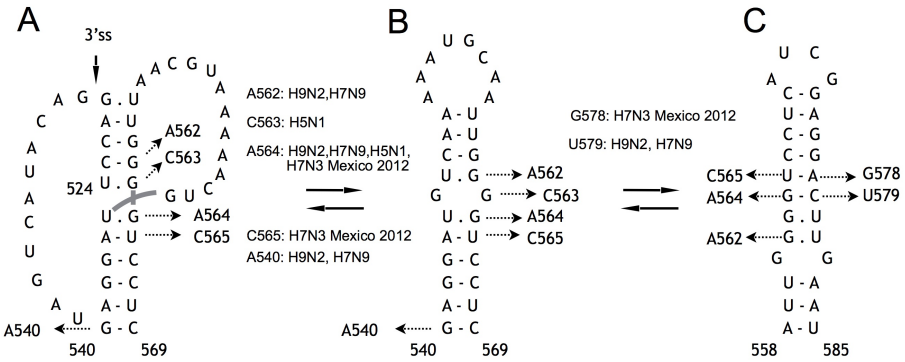


**Fig. 2.** Alternative structures in the 3' splice site region of influenza A virus NS mRNA. The substitutions in the strains of avian influenza outbreaks as compared to the avian clade A consensus are indicated.

Analysis of recent outbreaks of avian influenza infecting humans indicates that similar conformational shifts may be an important feature of virus adaptation, and a folding ensemble in the 3' splice site region may be even more complicated. Thus, highly pathogenic H7N3 viruses that caused an outbreak in Mexico in 2012 [15] contain three mutations in this region as compared to the

consensus for avian viruses: A564, C565 and G578 (Fig. 2). Remarkably, the C565 and G578 substitutions constitute a covariation stabilising another hairpin competing with both pseudoknot and the hairpin sharing the stem with the pseudoknot (Fig. 2C). The alternative hairpin, folded between nucleotides 558-585, is conserved in all influenza A NS segments, suggesting that it may be important player in the competition between folding alternatives. Recent H7N9 viruses that caused a severe outbreak in China in 2013 [16], also exhibit compensatory changes at positions 564 and 579 in this hairpin (Fig. 2C). Furthermore, the substitutions A540 and A562, observed in these H7N9 viruses and in the H9N2 strains from which they inherited NS segments by reassortment [16], are expected to considerably destabilise the pseudoknot (Fig. 2A).

Thus, in these three important outbreaks of avian influenza the equilibrium is expected to shift from pseudoknot to one or both alternative hairpins (Fig. 2). Moreover, analysis of available sequences shows that the substitutions leading to these shifts are rare in other avian strains. The C563 is unique for highly pathogenic H5N1 viruses. The C565-G578 compensatory change, besides three sequenced NS segments from Mexican 2012 H7N3 outbreak, is observed in only one other avian clade A NS sequence (out of 9138 available ones), and each of these substitutions is a quite uncommon one. The A564-U579 combination is almost unique for H7 and H9 viruses: it occurs in 63 and 153 out of total of 710 and 1289 strains, respectively (8.9% and 11.9%, respectively), while it is present in only 13 out of 7139 strains (0.18%) from all other avian clade A lineages. The pseudoknot-destabilising mutation G540 is abundant only in H7 and H9 viruses: 97 and 241 strains, respectively, while only 12 cases are seen in other viruses. Similarly, A562 occurs in 94 H7 strains, 51 H9 strains and 36 viruses from other groups.

Interestingly, although the A564-U579 covariation is more frequent in H9 than in H7 strains, the 564/579 base pair seems to be more important for H7 viruses. The nucleotide substitutions at these positions do not seem to correlate in H9 strains (M(x,y) = 0.017; $R_1(x, y) = 0.058$; $R_2(x, y) = 0.035$), but the correlation values are higher in H7 : M(x,y) = 0.08; $R_1(x, y) = 0.30$; $R_2(x, y) = 0.20$.

Similar structures can be folded in the clade B NS segments of influenza A viruses, albeit the 558-585 hairpin is considerably less stable thermodynamically as compared to that in clade A. Remarkably, highly pathogenic H5N1 viruses seem to destabilise the clade B pseudoknot as well. All 20 clade B NS segments of this lineage contain A at position 524, apparently destabilising the pseudoknot junction, because the overwhelming majority of clade B NS segments (2254 out of 2293 available sequences) contain a GA non-canonical pair at positions 524/563 that is a covariation as compared to UG pair in clade A (Fig. 2A) [2].

### 4.2 Alternative structures in the NS 5' splice site region

A structured domain, predicted in the 5'-proximal part of NS mRNA intron [17], has been shown to be conserved in both clade A and B segments [4, 5]. The predicted folding in the region 107-174 (or 81-148 in the numbering from

the start of the NS1 protein reading frame) has been evidenced by structure probing *in vitro* [18]. Here we show that the 107-174 domain is a part of larger structured domain that encloses the 5' splice site located upstream at positions 56/57. This conclusion is based on the analysis of possible structures in this region and covariation patterns, given below.

The suggested larger domain (46-182) as predicted by free energy minimization is shown in Fig. 3. It consists of an extended stem-loop 107-174 till nucleotides 94-182 and an adjacent stem-loop structure between positions 46-93. The closing stem of the 94-182 is conserved in clade A NS segments of human and avian strains. Moreover, a strong indication to importance of this stem is obtained from the covariation at the terminal base pair 94-182 (Table 1).

The avian strains have predominantly the CG pair at these positions. The CG pair is also present in the majority of the human H1N1 viruses. However, during circulation of human H2N2 strains (1957-1968) the pair undergoes a covariation transition: in 1961-1962 it becomes an UG wobble pair, and since 1963 it is predominantly CG. The CG pair is also inherited by H3N2 human viruses that substituted the H2N2 subtype in 1968. Mutual information calculations, in particular, the M(x,y) ratios to position entropies, show that the nucleotides at positions 94 and 182 do correlate in avian and human strains: M(x,y) = 0.37; $R_1(x, y) = 0.86$ and $R_2(x, y) = 0.79$. However, this constraint disappears in swine strains (Table 1), and mutual information values are not significant.

**Table 1.** Number of clade A NS segment sequences from human, avian and swine influenza A virus strains with various nucleotide combinations at positions 94-182

| strains | CG | CA | UG | UA | CC | CU | UC | UU | GG | GA |
|---|---|---|---|---|---|---|---|---|---|---|
| avian | 8683 | 269 | 15 | 1 | 35 | 6 | - | - | 1 | - |
| human H1N1 | 1512 | 6 | 1 | 2 | 15 | 5 | - | - | - | - |
| human H2N2 1957-1960 | 55 | - | 1 | 1 | - | - | - | - | - | - |
| human H2N2 1961-1962 | - | - | 6 | 1 | - | - | - | - | - | - |
| human H2N2 1963-1968 | 1 | - | - | 54 | - | - | - | - | - | - |
| human H3N2 1968-now | 3 | 10 | 2 | 4210 | 14 | 9 | 4 | - | - | 3 |
| total avian+human | 10254 | 285 | 25 | 4269 | 64 | 20 | 4 | - | 1 | 3 |
| swine | 375 | 165 | 1 | 66 | 1172 | 259 | 3 | 1 | - | - |

The remarkable conservation of the 94-182 pair in avian and human viruses and its disappearance in swine strains turn out to correlate with alternative foldings in the 5'-proximal stem-loop 46-93 (Fig. 3). In avian and human viruses, the 94-182 base pair contributes to the so-called coaxial stacking [19] at the junction of two helices in the predicted domain structure. Predictions using free energy minimization by Mfold program [10] clearly demonstrate that without this coaxial stacking the bottom helix of the 46-93 stem-loop is not formed at all because of its positive free energy contribution. However, upon the folding of the 94-182 structure the formation of the 46-93 stem is favoured thermodynam-

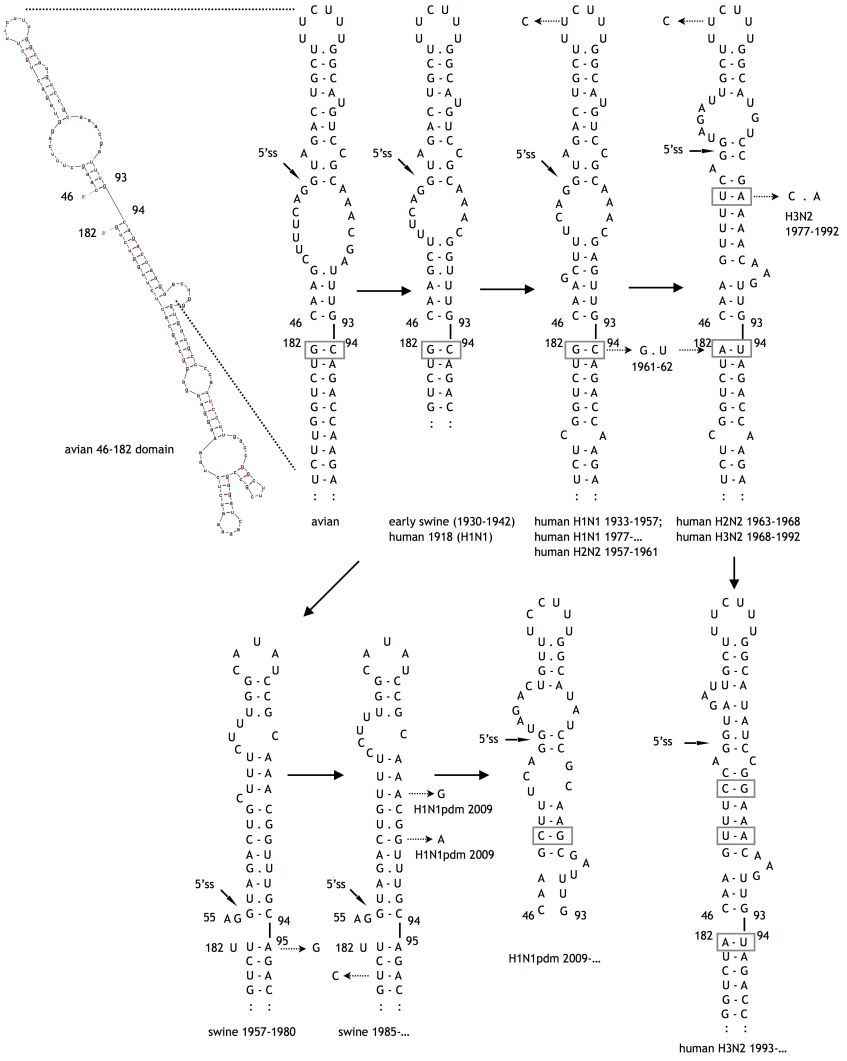


**Fig. 3.** The predicted structure in the 46-182 region of influenza A virus NS mRNA. The minimum free energy prediction yielded by Mfold [10] for the avian consensus sequence and the main changes in the 5' splice site region in human and swine viruses are shown. The 5' splice site (5'ss) is indicated by arrow. Covariations are boxed.

ically: in avian strains the free energy gain is 2.2 kcal/mol according to Mfold predictions.

Such coaxial stacking contributions may be very considerable, because stacking free energies at the junctions of two helices are even lower than those between base pairs in the regular double helices. Indeed, according to the coaxial stacking measurements with model oligonucleotides [19], the stabilisation of the 46-93 stem by C94-G182 pair is 4.3 kcal/mol. Thus, the presence of the 94-182 base pair is expected to affect the RNA conformation in the 5' splice site region (nucleotides 56/57).

Predictions of possible alternative foldings further confirm a correlation between 94-182 base-pairing and the 5' splice site conformation. While the folding in the region 46-93 of the NS segment of the first known human strain of 1918 is still identical to the conserved avian structure, this structure is changed in later human viruses (Fig. 3). Nevertheless, a coaxially stacked conformation at the 46-93/94-182 junction is still possible in all human strains. Two substitutions (U50 and A83), occurring simultaneously with CG transition to UA at position 94-182, rearrange the interior of the 46-93 folding. This conformation is conserved in subsequent human strains and is supported by a covariation at positions 53-83, undergoing a slow transition during 1977-1992 from UA via CA mismatch to CG pair. Due to a large number of CA mismatches, only one of correlation values is relatively high for this base pair: $R_1(x, y) = 0.6$, while $R_2(x, y) = 0.36$ and M(x,y) = 0.1.

The absence of correlation between nucleotides 94 and 182 in swine viruses also correlates with the folding of the splice site region. In contrast to early swine strains with avian-like structure, in later classical swine viruses the 46-93 stem-loop structure is disrupted in favour of more stable alternative 57-94 stem-loop (Fig. 3). Apparently, a conformation of the 5' splice site is changed. This conformation is thermodynamically stable in the absence of coaxial stacking and, consistently, a pressure to keep the 94-182 pairing disappears. Moreover, several mutations disrupt the whole closing stem of domain 94-182, while the 57-94 stem-loop is retained in subsequent swine viruses. However, two synonymous mutations (G86 and A89) in the 2009 H1N1 pandemic human strains, which received their NS segments from classical swine viruses by reassortment [20], destabilize this conformation in favour of a structure more similar to that in human viruses (Fig. 3). Interestingly, this structure in the 2009 pandemic strains contains a convergence covariation at positions 50-86 as compared to this pair in the human H3N2 strains.

## 5 Discussion

The presented results show that influenza A virus genome segments encode RNA structures that impose significant constraints on sequence evolution. These constraints are seen in high values of mutual information M(x,y) and/or of M(x,y) ratios to entropy values at the alignment positions of paired nucleotides. It should be noted that many covarying positions are biased (for instance, fluc-

tuating mostly between only two nucleotide types, unequally represented in the available data), leading to relatively low M(x,y) values. However, this problem can be solved by using ratios of M(x,y) to entropies that take variations at the paired positions into account. The detected constraints suggest that predicted structures perform important functions in the virus replication.

Similar to what is observed in other RNA viruses, these functions may be diverse. For instance, the location of the M segment hairpin (Fig. 1) in the region known to include packaging signals, strongly suggests its involvement in the packaging of viral RNA in the virion. Moreover, this hairpin is thermodynamically stable in negative-sense vRNA. However, it is also stable in positive-sense mRNA, thus some other regulatory role in gene expression cannot be excluded.

The presence of conserved conformations in the regions of both 5'- and 3'-splice sites of two intron-containing segments of the influenza virus genome strongly suggests involvement of RNA structure in splicing regulation. Furthermore, a number of specific conformational changes indicate that such a regulation is an important feature of virus adaptation. Thus, the shift from pseudoknot to alternative hairpin conformations at the NS gene 3' splice site is predicted to accompany several important outbreaks of avian influenza affecting humans (Fig. 2). The regulatory structures at the 5' splice site, suggested by thermodynamic and comparative analysis, undergo host-specific evolution, evolving a variety of topologies (Fig. 3). Interestingly, the swine-origin NS segments of human 2009 H1N1 pandemic viruses seem to return to human-like conformation, presumably adapting their splicing to human cells. Adaptation of the influenza virus RNA splicing to different environments has also been suggested previously [21-23]. In particular, alterations in the NS mRNA structure were suggested to correlate with the reduced splicing in neurovirulent strains [21].

The predicted lineage-specific conformational shifts in various regions of the influenza virus genome suggest an epochal character of virus RNA structure evolution. Similar to evolution of influenza virus antigenic properties [24], the evolution of RNA folding occurs with relatively long periods (epochs) of stable phenotypic features (in this case RNA topologies), punctuated by rare conformational shifts.

The number of observed nucleotide covariation events is also relatively small. Partially it is explained by virus RNA protein-coding properties: many of these covariations lead to amino acid changes that may affect virus fitness. On the other hand, high mutual information values show that RNA folding dictates correlated substitution patterns at multiple positions in the viral genome. Apart from importance of the covariations in providing evidence for structure functionality, they may have implications in the virus phylogeny reconstruction, as many phylogenetic approaches are based on an assumption of uncorrelated mutations that may be invalid [25]. The used method for estimating significance of mutual information with a correction for sequence bias can help in understanding influenza virus adaptation and evolution.

**Acknowledgments.** We thank Maarten de Smit for useful discussions. This work was supported by a VICI grant 91896613 of the Netherlands Organisation for Scientific Research (NWO) to R.A.M.F. and by NIAID/NIH contract HHSN266200700010C.

## References

1. Horimoto, T., Kawaoka, Y.: Influenza: lessons from past pandemics, warnings from current incidents. Nature Rev. Microbiol. 3, 591-600 (2005)
2. Gultyaev, A.P., Heus, H.A., Olsthoorn, R.C.L: An RNA conformational shift in recent H5N1 influenza A viruses. Bioinformatics 23, 272-276 (2007)
3. Gultyaev, A.P., Olsthoorn R.C.L. A family of non-classical RNA pseudoknots in influenza A and B viruses. RNA Biol. 7, 125-129 (2010)
4. Gultyaev, A.P., Fouchier, R.A.M., Olsthoorn, R.C.L.: Influenza virus RNA structure: unique and common features. Int. J. Immunol. 29, 533-556 (2010)
5. Moss, W.N., Priore, S.F., Turner, D.H.: Identification of potential conserved RNA secondary structure throughout influenza A coding regions. RNA 17, 991-1011 (2011)
6. Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J., Lipman, D..:The influenza virus resource at the National Center for Biotechnology Information. J Virol. 82, 596-601 (2008)
7. Moss, W.N., Dela-Moss, L.I., Kierzek, E., Kierzek, R., Priore, S.F., Turner, D.H.: The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. PLoS One 7, e38323 (2012)
8. Moss, W.N., Dela-Moss, L.I., Priore, S.F., Turner, D.H.: The influenza A segment 7 mRNA 3' splice site pseudoknot/hairpin family. RNA Biol. 9, 1305-1310 (2012)
9. Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., Stadler, P.F.: RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 9, 474 (2008)
10. Zuker, M.: Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406-3415 (2003)
11. Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J., Stormo, G.D.: Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucleic Acids Res. 20, 5785-5795 (1992)
12. Gregory, V., Lim, W., Cameron, K., Bennett, M., Marozin, S., Klimov, A., Hall, H., Cox, N., Hay, A., Lin, Y.P.: Infection of a child in Hong Kong by an influenza A H3N2 virus closely related to viruses circulating in European pigs. J Gen. Virol. 82, 1397-1406 (2001)
13. Olsen,C.W., Karasin,A.I., Carman,S., Li,Y., Bastien,N., Ojkic,D., Alves,D., Charbonneau,G., Henning,B.M., Low,D.E., Burton,L., Broukhanski,G.: Triple reassortant H3N2 influenza A viruses, Canada, 2005. Emerg. Infect. Dis. 12, 1132-1135 (2006)
14. de Jong, J.C., Smith, D.J., Lapedes, A.S., Donatelli, I., Campitelli, L., Barigazzi, G., Van Reeth, K., Jones, T.C., Rimmelzwaan, G.F., Osterhaus, A.D., Fouchier, R.A.: Antigenic and genetic evolution of swine influenza A (H3N2) viruses in Europe. J Virol. 81, 4315-4322 (2007)
15. Kapczynski, D.R., Pantin-Jackwood, M., Guzman, S.G., Ricardez, Y., Spackman, E., Bertran, K., Suarez, D.L., Swayne, D.E.: Characterization of the 2012 highly

pathogenic avian influenza H7N3 virus isolated from poultry in an outbreak in Mexico: pathobiology and vaccine protection. J. Virol. 87, 9086-9096 (2013)
16. Gao, R., Cao, B., Hu, Y., Feng, Z., Wang, D., Hu, W., Chen, J., Jie, Z., Qiu, H., Xu, K., Xu, X., Lu, H., Zhu, W., Gao, Z., Xiang, N., Shen, Y., He, Z., Gu, Y., Zhang, Z., Yang, Y., Zhao, X., Zhou, L., Li, X., Zou, S., Zhang, Y., Li, X., Yang, L., Guo, J., Dong, J., Li, Q., Dong, L., Zhu, Y., Bai, T., Wang, S., Hao, P., Yang, W., Zhang, Y., Han, J., Yu, H., Li, D., Gao, G.F., Wu, G., Wang, Y., Yuan, Z., Shu, Y.: Human infection with a novel avian-origin influenza A (H7N9) virus. New Engl. J. Med. 368, 1888-1897 (2013)
17. Ilyinskii, P.O., Schmidt, T., Lukashev, D., Meriin, A.B., Thoidis, G., Frishman, D., Shneider, A.M.: Importance of mRNA secondary structural elements for the expression of influenza virus genes. OMICS 13, 421-430 (2009)
18. Priore, S.F., Kierzek, E., Kierzek, R., Baman, J.R., Moss, W.N., Dela-Moss, L.I., Turner, D.H.: Secondary structure of a conserved domain in the intron of influenza A NS1 mRNA. PLoS One 8, e70615 (2013)
19. Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H., Zuker, M.: Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. Proc. Natl. Acad Sci. USA 91, 9218-9222 (1994)
20. Garten, R.J., Davis, C.T., Russell, C.A., Shu, B., Lindstrom, S., Balish, A., Sessions, W.M., Xu, X., Skepner, E., Deyde, V., Okomo-Adhiambo, M., Gubareva, L., Barnes, J., Smith, C.B., Emery, S.L., Hillman, M.J., Rivailler, P., Smagala, J., de Graaf, M., Burke, D.F., Fouchier, R.A., Pappas, C., Alpuche-Aranda, C.M., Lpez-Gatell, H., Olivera, H., Lpez, I., Myers, C.A., Faix, D., Blair, P.J., Yu, C., Keene, K.M., Dotson, P.D. Jr, Boxrud, D., Sambol, A.R., Abid, S.H., St George, K., Bannerman, T., Moore, A.L., Stringer, D.J., Blevins, P., Demmler-Harrison, G.J., Ginsberg, M., Kriner, P., Waterman, S., Smole, S., Guevara, H.F., Belongia, E.A., Clark, P.A., Beatrice, S.T., Donis, R., Katz, J., Finelli, L., Bridges, C.B., Shaw, M., Jernigan, D.B., Uyeki, T.M., Smith, D.J., Klimov, A.I., Cox, N.J.: Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325, 197-201 (2009)
21. Ward, A.C., Azad, A.A., McKimm-Breschkin, J.L.: Changes in the NS gene of neurovirulent strains of influenza affect splicing. Virus Genes 10, 91-94 (1995)
22. Backstrom Winquist, E., Abdurahman, S., Tranell, A., Lindstrom, S., Tingsborg, S., Schwartz, S.: Inefficient splicing of segment 7 and 8 mRNAs is an inherent property of influenza virus A/Brevig Mission/1918/1 (H1N1) that causes elevated expression of NS1 protein. Virology 422, 46-58 (2012)
23. Selman, M., Dankar, S.K., Forbes, N.E., Jia, J.J., Brown, E.G.: Adaptive mutation in influenza A virus non-structural gene is linked to host-switching and induces a novel protein by alternative splicing. Emerg. Microbes Infect. 1, e42 (2012)
24. Smith, D.J., Lapedes, A.S., de Jong, J.C., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D., Fouchier, R.A.: Mapping the antigenic and genetic evolution of influenza virus. Science 305, 371-376 (2004)
25. Shapiro, B., Rambaut, A., Pybus, O.G., Holmes, E.C.: A phylogenetic method for detecting positive epistasis in gene sequences and its application to RNA virus evolution. Mol. Biol. Evol. 23, 1724-1730 (2006)