RNA structural constraints in the evolution of the influenza A virus genome NP segment

Goultiaev, A.P.; Tsyganov-Bodounov, A.; Spronken, M.I.; Kooij, S. van der; Fouchier, R.A.M.; Olsthoorn, R.R.C.L.

RNA structural constraints in the evolution of the influenza A virus genome NP segment

Alexander P Gultyaev, Anton Tsyganov-Bodounov, Monique IJ Spronken, Sander van der Kooij, Ron AM Fouchier & René CL Olsthoorn

RNA structural constraints in the evolution of the influenza A virus genome NP segment

Alexander P Gultyaev[1,2,*], Anton Tsyganov-Bodounov[3,†], Monique IJ Spronken[1], Sander van der Kooij[1,‡], Ron AM Fouchier[1], and René CL Olsthoorn[3]

[1]Department of Viroscience, Erasmus Medical Center, The Netherlands; [2]Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Niels Bohrweg 1, The Netherlands;; [3]Leiden Institute of Chemistry, Leiden University, P.O.Box 9502, 2300 RA Leiden, The Netherlands;

[†]Current address: Illumina UK Ltd., Chesterford Research Park, Little Chesterford, Essex, UK;

[‡]Current address: BaseClear B.V., Einsteinweg, The Netherlands

Conserved RNA secondary structures were predicted in the nucleoprotein (NP) segment of the influenza A virus genome using comparative sequence and structure analysis. A number of structural elements exhibiting nucleotide covariations were identified over the whole segment length, including protein-coding regions. Calculations of mutual information values at the paired nucleotide positions demonstrate that these structures impose considerable constraints on the virus genome evolution. Functional importance of a pseudoknot structure, predicted in the NP packaging signal region, was confirmed by plaque assays of the mutant viruses with disrupted structure and those with restored folding using compensatory substitutions. Possible functions of the conserved RNA folding patterns in the influenza A virus genome are discussed.

Introduction

Higher-order RNA structures play an important role in the replication of RNA viruses. In addition to protein-coding, RNA virus genomes encode RNA conformations that determine replication, transcription, translation and packaging of virus RNA.[1] Apart from specific functional motifs, many viruses also posess a genome-scale ordered RNA structure (GORS) that is suggested to contribute to viral fitness.[2] Functional requirements in virus RNA folding are expected to constrain virus genome evolution and have a considerable effect on sequence diversity.[3-5]

Evolution of the influenza A virus genome demonstrates a remarkable diversity of genomic sequences (for a review see ref. 6). Novel virus strains are continuously evolving, infecting various host species, sometimes with dangerous outbreaks in humans and domesticated animals. Such outbreaks are frequently determined by changes in the virus host specificity. One of the major problems in the development of intervention strategies against influenza is virus escape from protective immunity or upon drug treatment, that quickly generates new virus variants.

A role of RNA folding in influenza A virus evolution is unclear. The influenza A virus has a segmented genome, consisting of eight RNA segments of negative-sense polarity. In general,

negative-sense RNA viruses seem to have less ordered RNA structures as compared with positive-sense RNA genomes.[2] In part, this can be explained by the replication mechanism of negative-sense RNA genomes, with formation of ribonucleoprotein (RNP) complexes, where most of the genomic virus RNA (vRNA) structure is expected to melt.[7] On the other hand, some RNA structures can be observed in the RNP complexes as well.[8] Furthermore, global folding patterns were detected in the influenza A virus genome.[9,10] However, apart from well-studied universally conserved structures at the ends of each of the RNA segments, only few conserved structures were revealed in the influenza A virus RNA genome.[11-16]

Here, we describe the analysis of potential folding patterns in the NP segment (segment 5) of the influenza A virus genome, using the available sequences from various strains. The evolution of this segment resulted in several distinct host-specific clusters.[17,18] In order to detect the structural elements that may considerably affect genome sequence evolution, we restricted the analysis to those structures that exhibit patterns of correlated nucleotide variations (covariations) and/or deletions/insertions in the predicted single-stranded loops. The significance of covariations was estimated by calculations of mutual information in the variation at the presumably paired genome positions.

The results suggest the presence of universally conserved structural elements in both coding and non-coding regions of NP segment RNA.

Results

Predictions of conserved local structures in segment 5

Initial RNA secondary structure predictions for segment 5 were produced using three data sets, each containing five different representative NP RNA sequences from five main NP clades corresponding to different hosts (see Materials and Methods). Such an approach allowed us to obtain structural models while minimising possible sequence data bias. For each of the data sets, a consensus structure prediction was calculated using the RNAalifold program.[19] Using these three partially different structural models, we selected local structural elements that were recurrently predicted (in at least two out of three alternative RNAalifold predictions). The same procedure was performed for negative-sense RNA sequences. For both positive- and negative-sense RNA sequences together, 39 local domains satisfied this criterion. Searching further for those local structures that could present significant constraints on the sequence evolution, the structures with nucleotide covariations in the representative sequences or those containing indels in the loops were selected. As RNAalifold does not include pseudoknot prediction,[19] possibilities for pseudoknots involving the RNAalifold-predicted hairpins were explored separately (see Materials and Methods).

The stringent criteria for evolutionary conservation of predicted structures left 6 candidate local structures over the whole full-length (1565 nucleotides) NP segment (**Fig. 1**). One structure (positions 16–39) was predicted in the mRNA 5′ untranslated region (5′UTR), the other five in the coding part. Four out of six suggested structures (16–39; 89–105; 1436–1475; 1476–1530) are located in regions known to contain the NP segment packaging signals.[20] A pseudoknot in the region 1436–1475 (**Fig. 1F**) is possible only in negative-sense vRNA, with one nucleotide (position 1469) in the deep groove of the crossed helix of 7 base pairs (1468–1462/1442–1436), while its positive-sense "mirror" with one nucleotide in the shallow groove is stereochemically prohibited.[21] However, both pseudoknot stems can be separately folded in both positive- and negative-sense RNAs, and the other five structures (**Fig. 1**), in principle, can be folded in both polarities. Although none of the structures exhibited an unusual thermodynamic stability, all of them have negative values of free energy in the overwhelming majority of influenza A viruses. Calculations with two different sets of pseudoknot thermodynamic parameters[22,23] yielded rather close free energy values for the pseudoknot 1475–1436-folding: -13.6 and -15.7 kcal/mol, respectively, in A/Puerto Rico/8/34(H1N1) strain (**Fig. 1F**), both indicating to a stable pseudoknot conformation in vRNA.

Three of the predicted structures contain nucleotides that were deleted in some strains. The 16–39 hairpin in the mRNA 5′UTR is strongly conserved in the NP segments of all viruses with H1-H16 HA subtypes, but its two base pairs closing the loop are deleted in the structure predicted for a distinct lineage[24,25] of

bat H17N10 and H18N11 viruses (**Fig. 1B**). Such a deletion of paired nucleotides, that keeps the hairpin topology essentially the same, does not preclude the hairpin formation and is consistent with the model. This hairpin has been previously suggested to function in translational regulation,[26,27] but other functions, for instance the folding of its complement in the vRNA packaging signal, cannot be excluded.

Two hairpins in the NP coding region (89–105 and 577–593) exhibit deletions in the loops. Remarkably, in both cases the deletions are just single glycine codons with their first nucleotides involved in the loop-closing pairs (**Fig. 1C,D**). Both deletions have been previously described and are unlikely to be sequencing errors: the 94–96 deletion occurs in H7N1 viruses isolated in Italy in 1999–2000,[28] the 583–585 deletion is observed in a H1N1 strain isolated in Mongolia in 1991.[29]

Covariation analysis

The covariations suggested by the structures in 15 representative strains, used for RNAalifold predictions, were further analyzed by alignments of all NP sequences available in GenBank. Almost all covariations involve silent substitutions at "wobble" positions of NP codons. Remarkably, despite being silent in terms of protein-coding, the paired nucleotide positions demonstrate a presence of strong evolutionary pressure, apparently determined by RNA structure.

For instance, the nucleotides at wobble positions 582 and 588 of two alanine codons, predicted to pair in the 577–593 hairpin (**Fig. 1D**), clearly change in a coordinated way (**Table 1**). The predominant base pair in the avian viruses is U-A, whereas the overwhelming majority of human strains have A-U. Swine viruses have both main alternatives, actually reflecting the avian origin of many swine strains.[30,31] The 582/588 covariation between avian and human strains involves two transversions, the deviations within these clades are mostly transitions, with predominant R-Y pairs in human viruses and Y-R in avian ones (R denotes purines A and G, Y denotes pyrimidines C and U). In the equine viruses, including related H3N8 canine strains, the pressure to have Watson-Crick base-pairing seems to disappear, with predominant Y-Y combinations (**Table 1**). Such a pressure seems to arise again in South American avian strains with equine-like NP segments.[32,33] However, in contrast to U-A, dominant in other avian viruses, these strains contain G-U or A-U pairs, demonstrating an independent covariation event.

Other suggested covariation sites in the predicted structures (**Fig. 1**) demonstrate a different extent of pressure on the nucleotides occupying putatively paired positions (**Tables S1-S6**). The predominant base pairs are mostly host-specific, consistent with the covariations initially detected in the representative strains. However, the alignments of all available sequences show that the covariation events do not always coincide with speciation between the main host-specific lineages.

Thus, all sequenced human H1N1 strains before 1948 still have the avian-like GC base pair at the positions 923/937 (**Fig. 1E**), whereas A-U becomes predominant in later human viruses (**Table S2**). A relatively small number of human H1N1 and H3N2 strains with the 923/937 GC pair after 1948 originates from either preservation of earlier sequences (e.g., ref. 29) or human infections
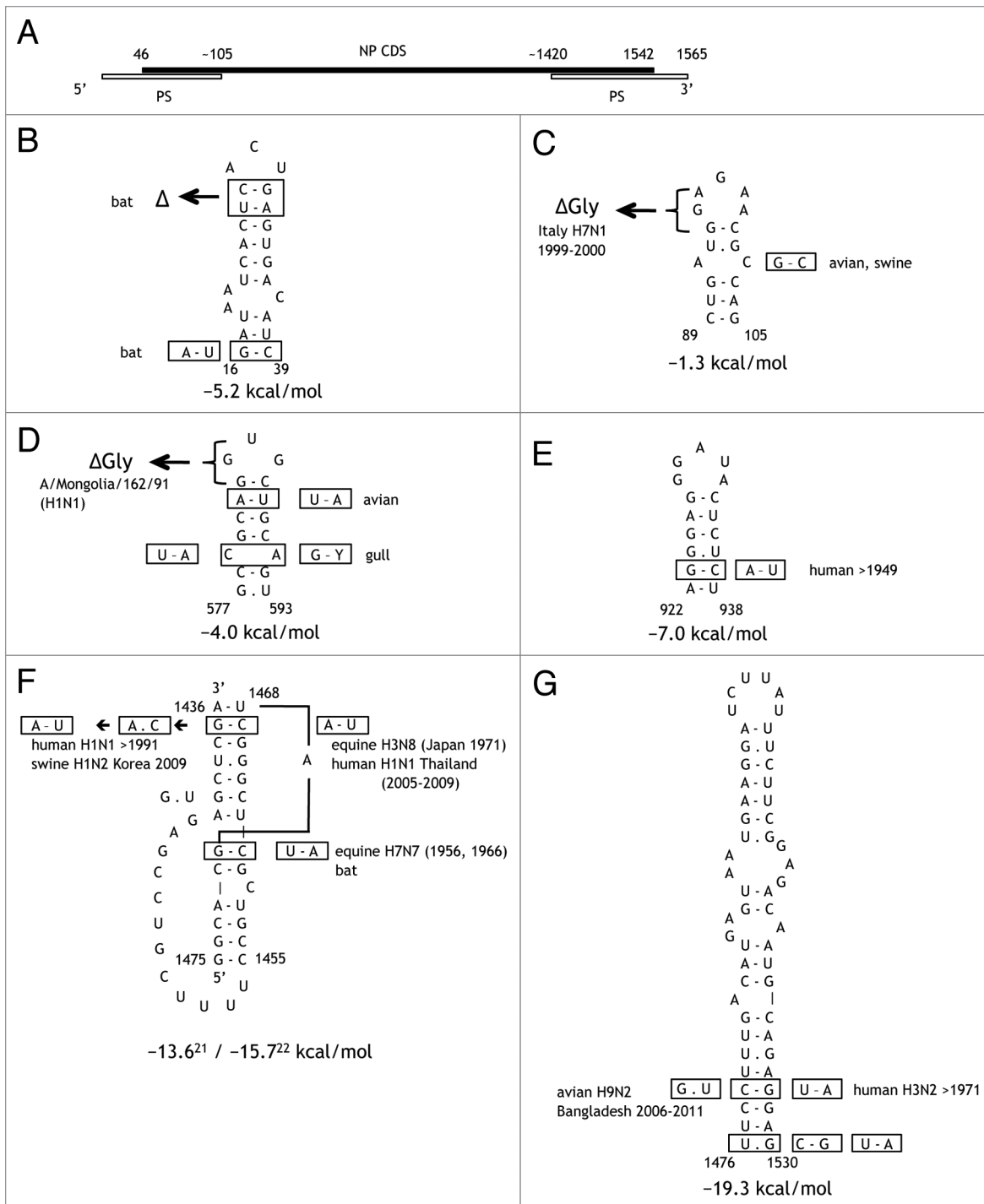
Figure 1. (**A**) Positions of the nucleoprotein coding region (NP CDS) and packaging signals (PS)[20] in the NP segment of the influenza A virus genome. The numbering of the (+) sense RNA is used. (**B-G**) Predicted conserved structures in the NP segment. All structures except the vRNA pseudoknot are shown for (+) RNA polarity. The structures are shown for the sequence from the A/Puerto Rico/8/34 (H1N1) strain (accession EF467822); the most frequent base pairs and covariations in specific lineages and strains are boxed. Statistics of covariations is shown in **Table 1** and **Tables S1-S6**. Deletions of glycine codons are denoted as ΔGly. Folding free energies are given for A/Puerto Rico/8/34, calculated as described in Materials and Methods; pseudoknot thermodynamic parameters are from Gultyaev et al.[22] or Cao et al.[23]

with swine or triple reassortant viruses (e.g., ref. 34). Swine-origin NP segments have predominantly G-C pairs (**Table S2**). In equine viruses, the hairpin 922–938 seems to be disrupted, as many mismatches are observed at multiple stem positions.

The 1467/1437 pair in one of the NP vRNA pseudoknot stems (**Fig. 1F**) is dominated by the CG combination in all lineages, with the exception of human H1N1 viruses (**Table S3**). In all NP sequences of this lineage, available in GenBank, the

Table 1. Number of strains in the GenBank with various combinations of nucleotides 582–588 in the NP segment

strains	AU	AC	GU	UA	UG	CA	CG	UU	CU	UC	AA	AG	GA
human	5783	386	7	7	-	-	-	46	2	-	8	2	-
avian	41	-	4	7001	661	2395	31	161	11	12	400	5	38
swine	1441	1	2	374	2	4	-	23	4	4	11	1	1
equine	-	-	-	8	-	-	-	105	70	-	-	-	-
total	7265	387	13	7390	663	2399	31	335	87	16	419	8	39

Table 2. Correlation values for nucleotide covariations

positions (x,y)	M(x,y)	R_1(x,y)	R_2(x,y)	NP changes
579/591	0.008	0.017	0.104	silent
582/588	0.362	0.498	0.556	silent
923/937	0.373	0.800	0.801	Arg/Lys/silent
1437/1467	0.026	0.146	0.430	silent
1461/1470	0.002	0.494	0.462	silent
1476/1530	0.020	0.043	0.088	silent
1479/1527	0.313	0.710	0.754	Glu/Asp/silent

Mutual information M(x,y) and ratios R_1(x,y) and R_2(x,y) were calculated as described in Materials and Methods. NP amino acid changes are given for the most abundant covariations shown in **Figure 1**.

CG pair is conserved only before 1991, when it becomes mostly a C-A mismatch in the suggested vRNA structure (corresponding to the U-G combination in positive sense RNA). The C-A mismatch is subsequently converted to an U-A pair that becomes predominant in 2009. A similar transition occurs in one H1N2 swine virus, isolated in Korea (A/swine/Korea/VDS2/2009, accession JN043443). Interestingly, two other covariation events from C-G to U-A are observed at these positions: one in two equine H3N8 strains from Japan (A/equine/Tokyo/2/1971 and A/equine/Sachiyama/1/1971, accessions CY096918 and CY034937, respectively), and the other in seven related swine H1N1 viruses from Thailand, isolated in 2005, 2008 and 2009, e.g., A/swine/Saraburi/NIAH100761–22/2009 (H1N1), accession AB620164.[35]

A relatively slow transition between C-G and U-A pairs via an intermediate wobble U-G at position 1479–1527 occurs in the human H2N2 and H3N2 lineages (**Table S6**). The wobble U-G pair is already predominant in H2N2 strains circulating in 1957–1968, and this is inherited by early (1968–1970) H3N2 viruses. The NP segments with the U-A pair appear for the first time in 1971 and since 1977 substitute those containing the C-G pair. After 1977, the C-G pair is seen almost exclusively in those human H3N2 strains that have swine-origin NP segments (e.g., ref. 34), according to BLAST alignments and/or GenBank annotation. Both swine strains and human H1N1 viruses have mostly the avian-like C-G pair at this position (**Table S6**). However, two cases of reversion from the 1479/1527 U-A pair back to C-G were found in GenBank NP segment sequences: in A/Nepal/921/2006 (H3N2) (accession CY047404) and in A/Mexico/InDRE2118/2005 (H3N2) (accession CY100555). In both these cases BLAST alignments clearly show that these C-G pair - containing sequences are close relatives of the U-A

pair - containing contemporary segments rather than of previously circulating ones or those of swine strains and thus represent independent covariation events. A unique covariation with two transversions at the same positions, from C-G to G-U pair (**Fig. 1G**) is observed in 41 Bangladesh H9N2 strains that apparently constitute a monophyletic group.[36]

A covariation involving two transversions also occurs at positions 1470–1461 in the vRNA pseudoknot structure (**Fig. 1F, Table S4**). This base pair is a remarkably conserved G-C pair in all lineages, with exception of early equine H7N7 viruses isolated in 1956 and 1966, where U-A is observed. The same covariation is also observed in recently identified bat H17N10 and H18N11 strains that represent a distinct lineage.[24,25] Yet, this covariation seems to be a single event. The influenza A NP segment trees, rooted to influenza B virus NP, locate A/equine/Prague/56(H7N7) near the root of H1-H16 strains and the H17N10/H18N11 branch between the H1-H16 cluster and influenza B.[17,24,25]

The constraints imposed by RNA secondary structure on the paired nucleotides x and y can be quantified using calculations of mutual information M(x,y) and the ratios of M(x,y) in relation to variability entropies H(x), H(y) at each of the paired positions: $R_1(x,y) = M(x,y)/H(x)$ and $R_2(x,y) = M(x,y)/H(y)$.[37] In principle, the values M(x,y), $R_1(x,y)$ and $R_2(x,y)$ may be in the range between 0 and 1, with higher values corresponding to more significant correlations. However, M(x,y) can reach high values only if both positions x and y are highly variable, otherwise it is low even in case of a perfect correlation. If one or both positions are biased, the values of $R_1(x,y)$ and $R_2(x,y)$ are more informative. They take the nucleotide bias into account by measuring the ratios between M(x,y) and variabilities H(x) and H(y) at both positions (see Materials and Methods for calculations). Thus a relatively high value of at least one of the ratios may indicate a significant correlation (in other words, indicating that nucleotide variation at one of the positions depends on the bias at the other position, presumably due to base-pairing constraints). Low values of all three parameters do not allow one to make any conclusions.

The values M(x,y), $R_1(x,y)$ and $R_2(x,y)$ for putative covariations in predicted structures (**Fig. 1B-G**) are given in **Table 2**. The highest values are calculated for the pair 923/937 (**Fig. 1E**). The value of M(x,y) is 0.373, which is rather high in this case, taking into account that both paired nucleotides vary only between two possibilities each: A or G at position 923 and U or C at position 937 (**Table S2**). The maximum value of M(x,y) for such a variability, in case of perfect covariation with 50% A-U and 50% G-C would be 0.5. Indeed, a strong pressure of

base-pairing is seen in high values of $R_1(x,y)$ and $R_2(x,y)$ (**Table 2**). Interestingly, the substitution at position 923 is not silent: the substitution of G to A changes an arginine codon AGR into a lysine codon AAR. This is accompanied by a silent change of a CUN leucine codon into UUN, leading to the covariation. In this case, coding restrictions on both paired positions seem to explain the absence of any combination other than four observed purine-pyrimidine pairs (**Table S2**).

The values calculated for other base pairs are in a rather broad range (**Table 2**). All three values $M(x,y)$, $R_1(x,y)$ and $R_2(x,y)$ are low only for 579/591 and 1476/1530 base pairs. Obviously, this reflects relatively high frequencies of mismatches at these positions (**Tables S1 and S5**). A low value of $M(x,y)$ for the 1461/1470 pair is determined by extreme biases at both positions (**Table S4**), but the covariation pressure is obvious from higher $R_1(x,y)$ and $R_2(x,y)$ values. Thus, in each of the structures with covariations (**Fig. 1**) there is at least one base pair demonstrating considerable secondary structure constraints.

Conserved structures in the NP packaging signal region are functional

The most extended conserved structured region, occupying almost 100 nucleotides, was predicted between positions 1436–1530 of the NP segment (**Fig. 1F,G**; **Figure 2A**). The whole domain, consisting of a pseudoknot and a stem-loop structure, comprises sequences contributing to the NP vRNA incorporation into virions: deletion experiments identified 120 proximal nucleotides of the coding region as sufficient for efficient incorporation.[20] Here we provide an argument in favor of the role of the predicted structure in the NP vRNA packaging; as this fold is stereochemically prohibited in the reverse complement it is likely only forming in the vRNA.

Furthermore, silent substitutions introduced in this region by two research groups were previously shown to affect the NP vRNA packaging.[38,39] These results seem to support a functional role of the structures predicted in this work (**Fig. 2A**), but interpretation of structure-function relationships is not straightforward. Mutations introduced in the conserved codons[38] indicate the pseudoknot as the most essential element, because substitutions at codons F464-L466, destabilizing one of its stems, showed the most significant packaging defect. Mutations in the stem or hairpin loop of the 1476–1530 stem-loop structure did not impair packaging. Two mutations (WSN-NP-2x) in different pseudoknot stems, converting two C-G pairs in less stable U-G wobble pairs (**Fig. 2A**), contributed to an attenuated phenotype of A/WSN/33 (H1N1) virus, but only in
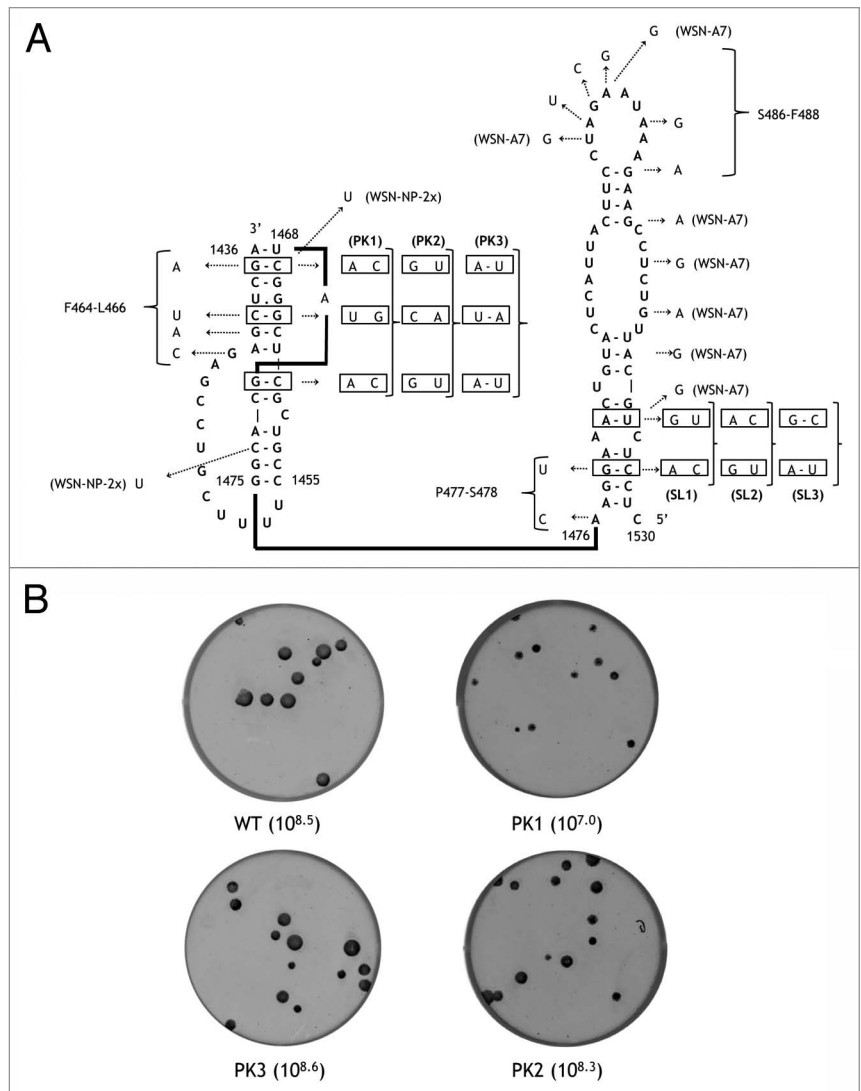

combination with seven substitutions in the stem-loop structure (WSN-A7) that were otherwise neutral.[39] In order to elucidate potential contributions of separate predicted structural elements, we studied the replication of viruses with synonymous mutations that disrupted the stems in the structures and restored them in double mutants (**Fig. 2A**). The mutants with disrupted structures contained mutations on either the 5′-proximal or the 3′-proximal side of stems, and double mutants had the same mutations on both sides reconstructing the stems.

Plaque assays of mutant viruses (**Fig. 2B**) confirmed a role of the pseudoknot structure in virus replication. Disruption of three base pairs in both pseudoknot stems in the PK1 mutant resulted in a significant decrease of virus titer and smaller plaque

Figure 2. (**A**) The predicted vRNA structure in the region 1530–1436 and mutations introduced in this domain. The mutations described previously are denoted as in the original publications: F464-L466; P477-S478; S486-F488[38] and WSN-NP-2x; WSN-A7.[39] The mutations made in this work are boxed. Although the structure is shown for the negative sense vRNA, the sequence numbering of the (+) sense RNA is used. (**B**) plaque assays of the WT virus (PR/8/34) and the pseudoknot mutants (PK1, PK2 and PK3). Virus titers (TCID$_{50}$/ml) are shown in brackets.

sizes. Notably, three compensatory mutations in the PK3 mutant were able to revert the fitness of the virus back to wild-type level. These compensatory substitutions by themselves (mutant PK2) did not significantly affect virus replication (**Fig. 2B**). No considerable effects were detected when any of the six nucleotide substitutions was introduced separately (not shown). Also, no deviation from wild-type replication was observed upon disruption of the bottom stem of the 1530/1476 structure, located upstream of the pseudoknot in vRNA, in the SL series of mutants (**Fig. 2A**; **Fig. S1**).

Discussion

The presence of universally conserved structures in the influenza A virus NP segments from different clades, together with sequence diversity, results in a number of nucleotide positions exhibiting covariation patterns. The covariations serve as evidence for both functional importance of the predicted structures and significant effects of RNA folding on virus genome evolution.

Not surprisingly, the covariations mostly involve wobble positions in the codons, consistent with the previously hypothesized correlation between RNA folding and codon conservation.[14,40,41] Furthermore, a relatively large number of codons with (nearly) absolute conservation of wobble positions suggests the actual number of base-pairing interactions to be larger than described in this work, which considers only the structures reliably supported by covariation analysis. On the other hand, the locations of structures (**Fig. 1**) remarkably correlates with structural regions detected in the NP-coding part by search for unusual thermodynamic stabilities and suppression of synonymous codon usage.[14] We did not find any covariation-supported structure only in one of four such regions. Moreover, only one of our five predicted structures in the coding part (positions 922–938, **Figure 1E**) did not correspond to any of the structured regions suggested by Moss et al.[14] The domain 1436–1530 (**Fig. 2A**) is almost equivalent to the structured region 1426–1539 (1381–1494 in the numbering from the start codon).

A high extent of sequence conservation in the predicted structures results in a relatively small number of covariations and low value of mutual information at some positions. Mostly the diversity of paired nucleotides is restricted by two dominating base types at each of the paired positions. This means that even a perfect covariation might have a maximum $M(x,y)$ value of 0.5, with the formula for mutual information taking the nucleotide alphabet size of 4 into account (see Materials and Methods). Furthermore, a strong bias toward certain bases leads to a very small $M(x,y)$ value even in cases of an obvious covariation. For instance, the base pair 1470/1461 in the vRNA pseudoknot (**Fig. 1**) is almost invariantly G-C, with an apparent covariation U-A, observed only in a few strains (**Table S4**). As expected, such a covariation is not reflected in the $M(x,y)$ value (**Table 2**). On the other hand, $M(x,y)$ is much higher for covariations that have a more balanced distribution of different base-pair types, such

as 582/588; 923/937 and 1479/1527 (**Table 2**), despite a higher flexibility to accept mismatches (**Tables 1, S2, S6**).

Thus, estimates of constraints based only on the mutual information value of $M(x,y)$ are expected to yield a lot of false-negative results simply due to the sequence bias and/or over-representation of certain phylogenetic groups. Such a problem in the analysis of conserved RNA structures may be solved by taking into account the entropy values describing the sequence diversity at the covarying positions.[37] Indeed, the relatively high ratios between $M(x,y)$ and at least one of two entropies $H(x)$ and $H(y)$ do show significant covariations even in cases of strong sequence bias (**Table 2**).

Of course, high values of $M(x,y)$, $R_1(x,y)$ and $R_2(x,y)$ may also arise at arbitrary pairs of positions in related RNA sequences due to a common ancestry rather than to structural constraints, leading to a well-known problem of false-positives in mutual information estimates of coevolution mechanisms (e.g., ref. 41). Unfortunately, the low number of covariation events detected in influenza virus RNA excludes an application of statistics based on counting of covariations occurring in multiple branches of an evolutionary tree in order to filter out spurious coincidences upon speciation.[42,43] In contrast to considerable numbers of covariation events, recorded in large data sets of structural RNAs such as rRNAs,[43] only few covariations are seen in the NP segment structures. Although several base pairs exhibit covariations in large numbers of strains, this diversity is mostly determined by single covariation events specific for particular clades. Similarly, clade-specific covariations were found in the structures suggested for NS and M segments of the influenza virus genome.[11,13] Presumably, such a pattern is determined by restrictions imposed by protein coding, accompanied by a remarkable codon conservation with a number of lineage-specific codons.[40] Nevertheless, some base-paired positions in the predicted NP segment structures (**Fig. 1**) underwent a covariation event more than once, providing more support for existence of constraints imposed by RNA folding.

Some putative paired positions exhibit low correlation values despite apparent covariation patterns. For instance, a sublineage of gull strains has unique base pairs G-U, G-C and A-U at positions 579–591 (**Fig. 1D**), while the majority of other viruses has U-A or C-A combinations at these positions (**Table S1**). Nevertheless, not only $M(x,y)$ has a low value, but so do $R_1(x,y)$ and $R_2(x,y)$ (**Table 2**). In principle, it is possible that this reflects a low pressure to keep this base pair. Alternatively, it is possible that a structural constraint disappears only in a subset of viruses: low values of $R_1(x,y)$ and $R_2(x,y)$ at positions 579–591 are mostly determined by a large number of A-A mismatches in just one lineage of human H1N1 strains, while other viruses have clear preference for base-pairing (**Table S1**). This may occur due to a loss of function, for instance, because of lineage-specific conformational transition, like e.g., in the influenza virus NS segment of highly pathogenic Asian H5N1 influenza A viruses.[11] Interestingly, the Asian H5N1 and H9N2 viruses together with the recent 2013 outbreak H7N9 strains[44] yield the majority of mismatches observed at positions 582–588 (**Table 1**): 2018 out of total 2399 C-A mismatches and 344 out of 419 A-A combinations. An over-representation of sequences from these groups in the database (3499 out of 10760 avian NP segments) lowers

the correlation values. Not surprisingly, exclusion of these clades from calculations leads to higher values, in particular, raising $R_1(x,y)$ from 0.498 (**Table 2**) to 0.675. The $M(x,y)$ and $R_2(x,y)$ are increased as well: from 0.362 to 0.403 and from 0.556 to 0.599, respectively.

The location of relatively large structured domains (**Fig. 2A**) in the NP segment packaging signal regions suggests that they may be involved in the vRNA packaging in the virions. Furthermore, the previously characterized nucleotide substitutions affecting packaging[38,39] destabilize the structures predicted in this work. A functional importance of the predicted 1436–1475 pseudoknot is evidenced by compensatory mutagenesis: reduced virus replication caused by three substitutions disrupting the pseudoknot (PK1 mutant) is rescued by three substitutions at the paired positions that restore the structure in the PK3 virus (**Fig. 2B**). The loss of fitness in the PK1 mutant is consistent with significant packaging disruption upon synonymous changes in codons F464-L466,[38] as two out of three PK1 substitutions were also used in the F464-L466 cluster of mutations (**Fig. 2A**). The reversion to the wild type phenotype in the PK3 mutant shows that low variability of these codons, apparently dictated by packaging,[38,40] is determined by structural constraints rather than by a specific sequence motif.

It seems that only extensive mutations at multiple positions in the structures predicted in the 1436–1530 region of the NP segment (**Fig. 2**) distort the packaging process. Our experiments and those performed by others[38,39] suggest that these conserved conformations are not absolutely required for virus replication. None of the tested disruptions of single base pairs produced any effect on virus replication. Furthermore, three mismatches, introduced in the pseudoknot stems in the mutant PK2, did not reduce the virus growth. This can be explained either by insufficient pseudoknot destabilization or by redundancy in the functioning of RNA structural elements. Indeed, the pseudoknot free energy estimates of about -15 kcal/mol (**Fig. 1F**) indicate that it can tolerate some mismatches in stems. Possible triple interactions in the helical grooves, not taken into account by pseudoknot thermodynamic models[22,23] may lead to even lower free energy. Interestingly, mapping the mutations, introduced by Anhlan et al.[39] into the A/WSN/33 NP segment, on the structures predicted in this work (**Fig. 2A**) suggests that only disruption of both pseudoknot (1436–1475) and stem-loop structure (1476–1530) results in virus attenuation. Multiple mutations, introduced in the 1476–1530 region alone by Hutchinson et al.,[38] Anhlan et al.[39] and in this work did not induce significant change of virus phenotype. It has been previously noted that the packaging signals of the influenza A virus genome, the NP segment included, can accomodate diverse mutations.[38,45] The NP packaging signal has been recently suggested to consist of an incorporation signal, absolutely required for the NP incorporation into virions, and a bundling signal, located in the protein-coding part, essential for interactions with other segments.[5] Probably, such interactions are established via multiple partially redundant motifs.

Apart from packaging signal regions, the structures predicted in all regions of the influenza A virus NP segment (**Fig. 1**) clearly impose considerable constraints on the genome evolution,

suggesting that they do contribute to viral fitness. Such contributions may be determined by various functions. Location of the structures outside the main packaging signals does not exclude their function in RNA packaging. Recent electron microscopy data on the influenza virus virion structure indicated that multiple RNA-RNA contacts between the middle parts of vRNPs may be involved in the specific packaging of eight genomic segments.[46,47] These contacts could be established via small conserved RNA hairpins, like those predicted in the NP segment (**Fig. 1**). Such "kissing" loop-loop interactions between small RNA hairpins are consistent with the observations of short string-like linkers between RNPs inside the virions, with the fibril shape resembling that of RNA.[46] Comparison of NP-RNA binding stoichiometry with the crystal structure of NP from another orthomyxovirus, infectious salmon anemia virus (ISAV), suggests the existence of structured NP-free RNA regions in the ISAV RNP and other orthomyxoviruses with similar NP 3D topologies, such as influenza A viruses.[48] An interaction between two "kissing" hairpins in segments PB1 and NS, only partially conserved in the H5N2 subtype strains, has recently been identified.[49] The presence of lineage-specific intersegmental interactions, important for genetic reassortment of influenza A viruses,[50,51] may also explain the disappearance of base-pairing constraints in some of the predicted NP segment hairpins, as observed only in certain strain clusters (see e.g., **Table 1**, **Table S1**).

Other functions of RNA structures in influenza virus replication can also be a reason for structure conservation. Apart from the panhandle motifs at the genome ends, higher-order structures have previously been suggested to regulate influenza virus RNA translation[26,27] and adaptation to specific temperatures.[9,52] Extensive structured regions in the coding parts of virus RNA may also modulate cotranslational protein folding, presumably guiding native structure formation by translational pausing that allows more time for the folding of important protein domains.[53,54] In the HIV-1 RNA genome, the locations of many stable structures correlate with those of interdomain regions in the encoded proteins.[54] In principle, similar effects can be envisaged for the influenza virus RNA.[13] However, no clear correlation is seen between the locations of predicted NP segment RNA structures and the nucleoprotein domains.[55,56] Furthermore, moderate free energy values of the conserved NP segment RNA structures make their regulatory role in translation rather unlikely. One of the possible functions of structured regions in RNA genomes is to interfere with the host innate systems recognizing a pathogen RNA.[2,57-59]

Future experimental studies will elucidate the structure-function relationships in the influenza virus RNA genome. Apparently, RNA folding is an important factor contributing to the virus evolution.

Materials and Methods

Prediction of conserved structures

Candidate conserved RNA structures in the NP segment RNA were predicted for representative strain sequences using the algorithm RNAalifold.[19] The representative strains were selected

according to the phylogenetic tree of NP sequences with five major branches corresponding to human, swine, avian, gull and equine viruses, respectively.[17,18] For RNAalifold predictions, three sequence data sets, each consisting of five strains from different clades, were constructed. The data sets were constructed with an attempt to have as much sequence variability as possible, using available full-length NP segment sequences. The sequences from the following strains were selected: human viruses - A/Victoria/3/75 (H3N2), accession AF072545; A/Puerto Rico/8/34 (H1N1), EF467822; A/Singapore/1/1957 (H2N2), M63752; swine viruses - A/swine/Tennessee/24/1977 (H1N1), M30748; A/swine/Iowa/15/1930 (H1N1), M30747; A/swine/Wisconsin/1/1961 (H1N1), M63763; avian viruses - A/goose/Guangdong/1/96 (H5N1), AF144303; A/duck/Pennsylvania/1/1969 (H6N1), M63775; A/duck/Hong Kong/7/1975 (H3N2), CY005555; gull viruses - A/gull/Maryland/704/1977 (H13N6), M27521; A/gull/Astrakhan/227/1984 (H13N6), M30753; A/herring gull/DE/712/1988 (H16N3), CY004563; equine viruses - A/equine/Miami/1/1963 (H3N8), M22575; A/equine/Prague/1956 (H7N7), M22572; A/equine/Kentucky/2/1986 (H3N8), M30751. A separate lineage of bat influenza A H17N10 and H18N11 viruses[24,25] was not included because of considerable sequence differences that were expected to lead to major differences in the global NP segment folding.

The sequences in each of the data sets were aligned using ClustalW and used as an input for RNAalifold. RNAalifold predicts the best possible consensus structure for a given data set. The local structural elements present in at least two out of three RNAalifold predictions were further inspected for the presence of nucleotide covariations within the representative sequences and/or known deletions in the loops. The whole procedure was performed for both positive-sense sequences (corresponding to mRNA and cRNA molecules) and negative-sense vRNA. The statistical significance of covariations was estimated further by mutual information calculations for covarying positions, using all sequences available in GenBank. The sequences were retrieved and aligned using the NCBI's Influenza Virus Resource,[60] accessed in February, 2014.

RNA folding free energies (ΔG^0_{37}) of pseudoknot-free secondary structures were calculated at the Mfold server.[61] Upon identification of conserved hairpins, possible pseudoknotted hairpin loop interactions were explored manually, searching for pseudoknot stems of at least 3 base pairs that form loops not larger than 20 nucleotides with negative pseudoknot free energy values. For the calculation of pseudoknot free energy, the nearest-neighbor base-pair stacking parameters of the Mfold program were used,[62] taking into account additional stabilization by 1 kcal/mol due to coaxial stacking of two helices.[63] Destabilizing free energies of the pseudoknot loops were estimated using two different approximations of conformational entropies.[22,23]

Mutual information calculations

Mutual information (MI) and the *R* values (ratios of MI and entropies in the alignment columns) were calculated as described by Gutell et al.[37] The MI at two nucleotide positions *x* and *y* was calculated as follows:

$$M(x,y) = \Sigma\ f(b_x b_y) \cdot \log_4 [\ f(b_x b_y)\ /\ f(b_x) \cdot f(b_y)\ ]$$
$$(b_x, b_y \in [A, G, C, U]),$$

where $f(b_x)$, $f(b_y)$ are nucleotide occurence frequencies 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 the frequencies of all combinations are determined by the independent variations at both positions: $f(b_x b_y) = f(b_x) \cdot f(b_y)$.

As $M(x,y)$ may have low values even at correlated positions due to strong sequence biases, the ratios $R_1(x,y)$ and $R_2(x,y)$,[37] 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:

$$M(x,y) = H(x) + H(y) - H(x,y),$$

where *H* is an entropy term $H = -\ \Sigma\ f(b) \cdot \log_4[f(b)]$ 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. The values are asymmetric, that is, in general, $R_1(x,y) \neq R_2(x,y)$. The main advantage of using these parameters, tested on tRNA and rRNA data sets,[37] is that they may detect a correlation at biased positions (for instance, if the majority of sequences has a G-C pair, and only a minor fraction contains A-U). Also, a higher value of one of these ratios may detect a significant constraint at one of the paired positions, determined by the other nucleotide, while that other nucleotide is less constrained by the pairing. It should be noted that other covariation measures are also possible, with good discriminative power noted for those favoring canonical base pairs stacked on the neighboring pairs,[64] e.g., the covariation measure of the RNAalifold algorithm. As RNAalifold program was used here for the identification of conserved structural elements, MI values and their ratios to position entropies provided independent estimate of constraints in the evolution of the presumably paired nucleotides, including possible non-canonical interactions.

In order to minimize the influence of the database bias on the computation, the massively sequenced 2009 pandemic H1N1 strains were excluded from MI calculations. We did, however, check that this exclusion did not violate our conclusions, as the suggested structures predicted for these viruses did not exhibit any unexpected features. The sequences of a recently discovered distinct lineage of bat H17N10 and H18N11 strains[24,25] were not taken into account in the MI calculations due to obvious strong sequence and structure differences, but the configurations of the suggested structures in bat strains were analyzed as well.

Cells

293T cells were cultured in DMEM (Lonza, Breda, The Netherlands) supplemented with 10% FCS, 100 IU/ml penicillin (Lonza), 100 μg/μl streptomycin (Lonza), 2 mM glutamine (Lonza), 1 mM sodiumpyruvate (Gibco, Leusden, The Netherlands), and non-essential amino acids (MP biomedicals

Europe). Madin-Darby Canine Kidney (MDCK) cells were cultured in EMEM (Lonza) supplemented with 10% FCS, 100 IU/ml penicillin, 100 µg/µl streptomycin, 2 mM glutamine, 1,5 mg/ml sodiumbicarbonate (Lonza), 10 mM Hepes (Lonza), and non-essential amino acids.

Plasmids

Mutations were introduced into the NP gene segment of the A/PR/8/34 influenza virus. Plasmids containing the gene segments of A/PR/8/34 have been described previously.[65] Primers to introduce the desired mutations were designed using the web based quick change primer design program (Agilent technologies, Amstelveen, The Netherlands). A forward and reverse primer containing the desired mutation(s) was used in a PCR reaction, using 10 pmol of each primer, 5 µl of pfu ultra II buffer, 1.25 µl of (10 mM each) dNTP (Roche, Woerden, The Netherlands), 1 µl of pfu ultra II fusion (Agilent), and 100 ng of DNA in a total reaction volume of 50 µl. The PCR reaction was performed in a C1000 thermal cycler (Biorad, Veenendaal, The Netherlands), using the following program: 2 min 95 °C, followed by 18 cycles of 1 min 95 °C, 1 min 50 °C, 7 min 68 °C, and a final extension of 7 min at 68 °C. The PCR product was digested with 20 Units of DpnI (New England biolabs) and incubated for 1 h at 37 °C. A small fraction of the digested PCR product was transformed into XL-10 gold competent cells (Agilent) according to the instructions of the manufacturer.

293T cell transfections

293T cells were transiently transfected using the calcium phosphate method as described previously.[62] One day before transfection, gelatinized 100 mm dishes were plated with $3x10^6$ cells. The next day 50–70% confluent monolayers were transfected with 5 µg of each of the A/PR/8/34 gene segments; for NP 5 µg of either the wild type or the desired mutant was used. Approximately 16 h after transfection, cells were washed once with PBS and refreshed with DMEM containing 2% FCS for virus production. Cells were incubated for 72 h at 37 °C and 5% CO2, and supernatants were harvested by low-speed centrifugation at 1000 rpm for 10 min. Supernatants were used to inoculate a confluent monolayer of MDCK cells.

Inoculation of MDCK cells

293T cell supernatants were used to inoculate a confluent monolayer of MDCK cells in T25 flasks. MDCK cells were washed twice with PBS and inoculated with 3 ml of 293T supernatant. After incubation for 1 h at 37 °C and 5% CO2, cells were washed with PBS twice and 7 ml of infection medium was added to the cells to allow for virus replication. Infection medium consisted of EMEM (Lonza) supplemented with 100 IU/ml penicillin, 100 µg/µl streptomycin, 2 mM glutamine, 1,5 mg/ml sodiumbicarbonate, 10 mM Hepes, non-essential amino acids, and 1 µg/ml TPCK treated trypsin (Sigma, Zwijndrecht, The

Netherlands). Three days after inoculation, supernatants were harvested and tested for hemagglutination activity using turkey red blood cells as an indicator of virus production.

Virus titrations

The MDCK supernatants containing the recombinant viruses were used to assess infectious virus titers, essentially as described previously.[66] Virus supernatants were 10-fold serially diluted in infection medium. Cells were washed twice with PBS and 100 µl of the diluted virus was used to inoculate confluent monolayers of MDCK cells in 96-wells plates. After incubation for 1 h at 37 °C and 5% CO2 cells were washed with PBS twice and 200 µl of infection medium was added to each well. At 3 d post inoculation, the presence of virus was determined by testing for hemagglutination activity with turkey red blood cells and the $TCID_{50}$ was calculated from 10 replicates according to the method of Spearman-Karber.[67]

Plaque assay

The assay was performed as described previously.[68] In brief, MDCK supernatant from the recombinant wild type and mutant viruses used to inoculate a ~90% monolayer of MDCK cells in 6-well plates. Cells were washed twice with PBS and 1 ml of infection medium was added to each well. Recombinant virus supernatants were diluted and 100 µl of the selected dilution, to obtain a plaque density of ~10, was added to each well. After incubation for one hour at 37 °C and 5% CO2 cells were washed with PBS once and 2 ml of an overlay containing 2x EMEM (Lonza) and avicel (FMC BioPolymer, Newark, US) in a 1:1 ratio was added. Plates were incubated at 37 °C and 5% CO2. After 30 h, cells were washed with PBS twice and 1 ml of 80% acetone was added. Plates were incubated at -20 °C overnight and virus infection was determined by NP antibody staining. Briefly, NP monoclonal antibody (IgG2a, clone Hb65, American Type Culture Collection, Wesel, Germany) and goat-anti-mouse Ig FITC (BD biosciences, USA) antibody were used to detect NP positive cells. The plaques were scanned using a typhoon scanner (GE Healthcare, Diegem, Belgium).

Supplemental Materials

Supplemental materials may be found here:
www.landesbioscience.com/journals/rnabiology/article/29730/

References

1. Liu Y, Wimmer E, Paul AV. Cis-acting RNA elements in human and animal plus-strand RNA viruses. Biochim Biophys Acta 2009; 1789:495-517; PMID:19781674; http://dx.doi.org/10.1016/j.bbagrm.2009.09.007

2. Simmonds P, Tuplin A, Evans DJ. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. RNA 2004; 10:1337-51; PMID:15273323; http://dx.doi.org/10.1261/rna.7640104

3. Simmonds P, Smith DB. Structural constraints on RNA virus evolution. J Virol 1999; 73:5787-94; PMID:10364330

4. Holmes EC. Error thresholds and the constraints to RNA virus evolution. Trends Microbiol 2003; 11:543-6; PMID:14659685; http://dx.doi.org/10.1016/j.tim.2003.10.006

5. Goto H, Muramoto Y, Noda T, Kawaoka Y. The genome-packaging signal of the influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal. J Virol 2013; 87:11316-22; PMID:23926345; http://dx.doi.org/10.1128/JVI.01301-13

6. Nelson MI, Holmes EC. The evolution of epidemic influenza. Nat Rev Genet 2007; 8:196-205; PMID:17262054; http://dx.doi.org/10.1038/nrg2053

7. Coloma R, Valpuesta JM, Arranz R, Carrascosa JL, Ortín J, Martín-Benito J. The structure of a biologically active influenza virus ribonucleoprotein complex. PLoS Pathog 2009; 5:e1000491; PMID:19557158; http://dx.doi.org/10.1371/journal.ppat.1000491

8. Yamanaka K, Ishihama A, Nagata K. Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. J Biol Chem 1990; 265:11151-5; PMID:2358455

9. Brower-Sinning R, Carter DM, Crevar CJ, Ghedin E, Ross TM, Benos PV. The role of RNA folding free energy in the evolution of the polymerase genes of the influenza A virus. Genome Biol 2009; 10:R18; PMID:19216739; http://dx.doi.org/10.1186/gb-2009-10-2-r18

10. Priore SF, Moss WN, Turner DH. Influenza A virus coding regions exhibit host-specific global ordered RNA structure. PLoS One 2012; 7:e35989; PMID:22558296; http://dx.doi.org/10.1371/journal.pone.0035989

11. Gultyaev AP, Heus HA, Olsthoorn RC. An RNA conformational shift in recent H5N1 influenza A viruses. Bioinformatics 2007; 23:272-6; PMID:17090581; http://dx.doi.org/10.1093/bioinformatics/btl559

12. Gultyaev AP, Olsthoorn RC. A family of non-classical pseudoknots in influenza A and B viruses. RNA Biol 2010; 7:125-9; PMID:20200490; http://dx.doi.org/10.4161/rna.7.2.11287

13. Gultyaev AP, Fouchier RA, Olsthoorn RC. Influenza virus RNA structure: unique and common features. Int Rev Immunol 2010; 29:533-56; PMID:20923332; http://dx.doi.org/10.3109/08830185.2010.507828

14. Moss WN, Priore SF, Turner DH. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. RNA 2011; 17:991-1011; PMID:21536710; http://dx.doi.org/10.1261/rna.2619511

15. Moss WN, Dela-Moss LI, Kierzek E, Kierzek R, Priore SF, Turner DH. The 3′ splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. PLoS One 2012; 7:e38323; PMID:22685560; http://dx.doi.org/10.1371/journal.pone.0038323

16. Moss WN, Dela-Moss LI, Priore SF, Turner DH. The influenza A segment 7 mRNA 3′ splice site pseudoknot/hairpin family. RNA Biol 2012; 9:1305-10; PMID:23064116; http://dx.doi.org/10.4161/rna.22343

17. Gorman OT, Bean WJ, Kawaoka Y, Webster RG. Evolution of the nucleoprotein gene of influenza A virus. J Virol 1990; 64:1487-97; PMID:2319644

18. Reid AH, Fanning TG, Janczewski TA, Lourens RM, Taubenberger JK. Novel origin of the 1918 pandemic influenza virus nucleoprotein gene. J Virol 2004; 78:12462-70; PMID:15507633; http://dx.doi.org/10.1128/JVI.78.22.12462-12470.2004

19. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 2008; 9:474; PMID:19014431; http://dx.doi.org/10.1186/1471-2105-9-474

20. Ozawa M, Fujii K, Muramoto Y, Yamada S, Yamayoshi S, Takada A, Goto H, Horimoto T, Kawaoka Y. Contributions of two nuclear localization signals of influenza A virus nucleoprotein to viral replication. J Virol 2007; 81:30-41; PMID:17050598; http://dx.doi.org/10.1128/JVI.01434-06

21. Pleij CW, Rietveld K, Bosch L. A new principle of RNA folding based on pseudoknotting. Nucleic Acids Res 1985; 13:1717-31; PMID:4000943; http://dx.doi.org/10.1093/nar/13.5.1717

22. Gultyaev AP, van Batenburg FH, Pleij CW. An approximation of loop free energy values of RNA H-pseudoknots. RNA 1999; 5:609-17; PMID:10334330; http://dx.doi.org/10.1017/S135583829998189X

23. Cao S, Chen SJ. Predicting RNA pseudoknot folding thermodynamics. Nucleic Acids Res 2006; 34:2634-52; PMID:16709732; http://dx.doi.org/10.1093/nar/gkl346

24. Tong S, Li Y, Rivailler P, Conrardy C, Castillo DA, Chen LM, Recuenco S, Ellison JA, Davis CT, York IA, et al. A distinct lineage of influenza A virus from bats. Proc Natl Acad Sci U S A 2012; 109:4269-74; PMID:22371588; http://dx.doi.org/10.1073/pnas.1116200109

25. Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J, et al. New world bats harbor diverse influenza A viruses. PLoS Pathog 2013; 9:e1003657; PMID:24130481; http://dx.doi.org/10.1371/journal.ppat.1003657

26. Enami K, Sato TA, Nakada S, Enami M. Influenza virus NS1 protein stimulates translation of the M1 protein. J Virol 1994; 68:1432-7; PMID:7508995

27. Park YW, Wilusz J, Katze MG. Regulation of eukaryotic protein synthesis: selective influenza viral mRNA translation is mediated by the cellular RNA-binding protein GRSF-1. Proc Natl Acad Sci U S A 1999; 96:6694-9; PMID:10359774; http://dx.doi.org/10.1073/pnas.96.12.6694

28. Di Trani L, Bedini B, Cordioli P, Muscillo M, Vignolo E, Moreno A, Tollis M. Molecular characterization of low pathogenicity H7N3 avian influenza viruses isolated in Italy. Avian Dis 2004; 48:376-83; http://dx.doi.org/10.1637/7088; PMID:15283425

29. Anhlan D, Ludwig S, Nymadawa P, Mendsaikhan J, Scholtissek C. Previous H1N1 influenza A viruses circulating in the Mongolian population. Arch Virol 1996; 141:1553-69; PMID:8856033; http://dx.doi.org/10.1007/BF01718254

30. Schultz U, Fitch WM, Ludwig S, Mandler J, Scholtissek C. Evolution of pig influenza viruses. Virology 1991; 183:61-73; PMID:2053297; http://dx.doi.org/10.1016/0042-6822(91)90118-U

31. Guan Y, Shortridge KF, Krauss S, Li PH, Kawaoka Y, Webster RG. Emergence of avian H1N1 influenza viruses in pigs in China. J Virol 1996; 70:8041-6; PMID:8892928

32. Suarez DL, Senne DA, Banks J, Brown IH, Essen SC, Lee CW, Manvell RJ, Mathieu-Benson C, Moreno V, Pedersen JC, et al. Recombination resulting in virulence shift in avian influenza outbreak, Chile. Emerg Infect Dis 2004; 10:693-9; PMID:15200862; http://dx.doi.org/10.3201/eid1004.030396

33. Rimondi A, Xu K, Craig MI, Shao H, Ferreyra H, Rago MV, Romano M, Uhart M, Sutton T, Ferrero A, et al. Phylogenetic analysis of H6 influenza viruses isolated from rosy-billed pochards (*Netta peposaca*) in Argentina reveals the presence of different HA gene clusters. J Virol 2011; 85:13354-62; PMID:21976652; http://dx.doi.org/10.1128/JVI.05946-11

34. Olsen CW, Karasin AI, Carman S, Li Y, Bastien N, Ojkic D, Alves D, Charbonneau G, Henning BM, Low DE, et al. Triple reassortant H3N2 influenza A viruses, Canada, 2005. Emerg Infect Dis 2006; 12:1132-5; PMID:16836834; http://dx.doi.org/10.3201/eid1207.060268

35. Takemae N, Parchariyanon S, Ruttanapumma R, Hiromoto Y, Hayashi T, Uchida Y, Saito T. Swine influenza virus infection in different age groups of pigs in farrow-to-finish farms in Thailand. Virol J 2011; 8:537; PMID:22166074; http://dx.doi.org/10.1186/1743-422X-8-537

36. Shanmuganatham K, Feeroz MM, Jones-Engel L, Smith GJ, Fourment M, Walker D, McClenaghan L, Alam SM, Hasan MK, Seiler P, et al. Antigenic and molecular characterization of avian influenza A(H9N2) viruses, Bangladesh. Emerg Infect Dis 2013; 19:1393-402; PMID:23968540; http://dx.doi.org/10.3201/eid1909.130336

37. Gutell RR, Power A, Hertz GZ, Putz EJ, Stormo GD. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucleic Acids Res 1992; 20:5785-95; PMID:1454539; http://dx.doi.org/10.1093/nar/20.21.5785

38. Hutchinson EC, Wise HM, Kudryavtseva K, Curran MD, Digard P. Characterisation of influenza A viruses with mutations in segment 5 packaging signals. Vaccine 2009; 27:6270-5; PMID:19840659; http://dx.doi.org/10.1016/j.vaccine.2009.05.053

39. Anhlan D, Hrincius ER, Scholtissek C, Ludwig S. Introduction of silent mutations into the NP gene of influenza A viruses as a possible strategy for the creation of a live attenuated vaccine. Vaccine 2012; 30:4480-9; PMID:22575164; http://dx.doi.org/10.1016/j.vaccine.2012.04.070

40. Gog JR, Afonso EdosS, Dalton RM, Leclercq I, Tiley L, Elton D, von Kirchbach JC, Naffakh N, Escriou N, Digard P. Codon conservation in the influenza A virus genome defines RNA packaging signals. Nucleic Acids Res 2007; 35:1897-907; PMID:17332012; http://dx.doi.org/10.1093/nar/gkm087

41. Anhlan D, Grundmann N, Makalowski W, Ludwig S, Scholtissek C. Origin of the 1918 pandemic H1N1 influenza A virus as studied by codon usage patterns and phylogenetic analysis. RNA 2011; 17:64-73; PMID:21068184; http://dx.doi.org/10.1261/rna.2395211

42. Dutheil JY. Detecting coevolving positions in a molecule: why and how to account for phylogeny. Brief Bioinform 2012; 13:228-43; PMID:21949241; http://dx.doi.org/10.1093/bib/bbr048

43. Shang L, Xu W, Ozer S, Gutell RR. Structural constraints identified with covariation analysis in ribosomal RNA. PLoS One 2012; 7:e39383; PMID:22724009; http://dx.doi.org/10.1371/journal.pone.0039383

44. Lam TT, Wang J, Shen Y, Zhou B, Duan L, Cheung CL, Ma C, Lycett SJ, Leung CY, Chen X, et al. The genesis and source of the H7N9 influenza viruses causing human infections in China. Nature 2013; 502:241-4; PMID:23965623; http://dx.doi.org/10.1038/nature12515

45. Fujii K, Ozawa M, Iwatsuki-Horimoto K, Horimoto T, Kawaoka Y. Incorporation of influenza A virus genome segments does not absolutely require wild-type sequences. J Gen Virol 2009; 90:1734-40; PMID:19297607; http://dx.doi.org/10.1099/vir.0.010355-0

46. Noda T, Sugita Y, Aoyama K, Hirase A, Kawakami E, Miyazawa A, Sagara H, Kawaoka Y. Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. Nat Commun 2012; 3:639; PMID:22273677; http://dx.doi.org/10.1038/ncomms1647

47. Sugita Y, Sagara H, Noda T, Kawaoka Y. Configuration of viral ribonucleoprotein complexes within the influenza A virion. J Virol 2013; 87:12879-84; PMID:24067952; http://dx.doi.org/10.1128/JVI.02096-13

48. Zheng W, Olson J, Vakharia V, Tao YJ. The crystal structure and RNA-binding of an orthomyxovirus nucleoprotein. PLoS Pathog 2013; 9:e1003624; PMID:24068932; http://dx.doi.org/10.1371/journal.ppat.1003624

49. Gavazzi C, Yver M, Isel C, Smyth RP, Rosa-Calatrava M, Lina B, Moulès V, Marquet R. A functional sequence-specific interaction between influenza A virus genomic RNA segments. Proc Natl Acad Sci U S A 2013; 110:16604-9; PMID:24067651; http://dx.doi.org/10.1073/pnas.1314419110

50. Essere B, Yver M, Gavazzi C, Terrier O, Isel C, Fournier E, Giroux F, Textoris J, Julien T, Socratous C, et al. Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. Proc Natl Acad Sci U S A 2013; 110:E3840-8; PMID:24043788; http://dx.doi.org/10.1073/pnas.1308649110

51. Gavazzi C, Isel C, Fournier E, Moules V, Cavalier A, Thomas D, Lina B, Marquet R. An *in vitro* network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. Nucleic Acids Res 2013; 41:1241-54; PMID:23221636; http://dx.doi.org/10.1093/nar/gks1181

52. Chursov A, Kopetzky SJ, Leshchiner I, Kondofersky I, Theis FJ, Frishman D, Shneider A. Specific temperature-induced perturbations of secondary mRNA structures are associated with the cold-adapted temperature-sensitive phenotype of influenza A virus. RNA Biol 2012; 9:1266-74; PMID:22995831; http://dx.doi.org/10.4161/rna.22081

53. Guisez Y, Robbens J, Remaut E, Fiers W. Folding of the MS2 coat protein in Escherichia coli is modulated by translational pauses resulting from mRNA secondary structure and codon usage: a hypothesis. J Theor Biol 1993; 162:243-52; PMID:8412226; http://dx.doi.org/10.1006/jtbi.1993.1085

54. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr., Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 2009; 460:711-6; PMID:19661910; http://dx.doi.org/10.1038/nature08237

55. Ye Q, Krug RM, Tao YJ. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. Nature 2006; 444:1078-82; PMID:17151603; http://dx.doi.org/10.1038/nature05379

56. Ng AK, Zhang H, Tan K, Li Z, Liu JH, Chan PK, Li SM, Chan WY, Au SW, Joachimiak A, et al. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. FASEB J 2008; 22:3638-47; http://dx.doi.org/10.1096/fj.08-112110; PMID:18614582

57. Knoepfel SA, Berkhout B. On the role of four small hairpins in the HIV-1 RNA genome. RNA Biol 2013; 10:540-52; PMID:23535706; http://dx.doi.org/10.4161/rna.24133

58. McFadden N, Arias A, Dry I, Bailey D, Witteveldt J, Evans DJ, Goodfellow I, Simmonds P. Influence of genome-scale RNA structure disruption on the replication of murine norovirus--similar replication kinetics in cell culture but attenuation of viral fitness *in vivo*. Nucleic Acids Res 2013; 41:6316-31; PMID:23630317; http://dx.doi.org/10.1093/nar/gkt334

59. Hyde JL, Gardner CL, Kimura T, White JP, Liu G, Trobaugh DW, Huang C, Tonelli M, Paessler S, Takeda K, et al. A viral RNA structural element alters host recognition of nonself RNA. Science 2014; 343:783-7; PMID:24482115; http://dx.doi.org/10.1126/science.1248465

60. 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 2008; 82:596-601; PMID:17942553; http://dx.doi.org/10.1128/JVI.02005-07

61. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003; 31:3406-15; PMID:12824337; http://dx.doi.org/10.1093/nar/gkg595

62. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 1999; 288:911-40; PMID:10329189; http://dx.doi.org/10.1006/jmbi.1999.2700

63. Walter AE, Turner DH, Kim J, Lyttle MH, Müller P, Mathews DH, Zuker M. Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. Proc Natl Acad Sci U S A 1994; 91:9218-22; PMID:7524072; http://dx.doi.org/10.1073/pnas.91.20.9218

64. Lindgreen S, Gardner PP, Krogh A. Measuring covariation in RNA alignments: physical realism improves information measures. Bioinformatics 2006; 22:2988-95; PMID:17038338; http://dx.doi.org/10.1093/bioinformatics/btl514

65. de Wit E, Spronken MI, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. Virus Res 2004; 103:155-61; PMID:15163504; http://dx.doi.org/10.1016/j.virusres.2004.02.028

66. Rimmelzwaan GF, Baars M, Claas EC, Osterhaus AD. Comparison of RNA hybridization, hemagglutination assay, titration of infectious virus and immunofluorescence as methods for monitoring influenza virus replication *in vitro*. J Virol Methods 1998; 74:57-66; PMID:9763129; http://dx.doi.org/10.1016/S0166-0934(98)00071-8

67. Karber G. Beitrag zur kollektiven Behandlung pharmakologischer Reiheversuche. Naunyn-Schmiedeberg. Arch Exp Pathol Pharmacol 1931; 162:480-3; http://dx.doi.org/10.1007/BF01863914

68. Matrosovich M, Matrosovich T, Garten W, Klenk HD. New low-viscosity overlay medium for viral plaque assays. Virol J 2006; 3:63; PMID:16945126; http://dx.doi.org/10.1186/1743-422X-3-63