A family of non-classical pseudoknots in influenza A and B viruses

Gultyaev, A.P.; Olsthoorn, R.C.L.

RNA Biology

ISSN: 1547-6286 (Print) 1555-8584 (Online) Journal homepage: https://www.tandfonline.com/loi/krnb20

A family of non-classical pseudoknots in influenza A and B viruses

Alexander P. Gultyaev & René C. L. Olsthoorn

To cite this article: Alexander P. Gultyaev & René C. L. Olsthoorn (2010) A family of non-classical pseudoknots in influenza A and B viruses, RNA Biology, 7:2, 125-129, DOI: 10.4161/rna.7.2.11287

To link to this article: https://doi.org/10.4161/rna.7.2.11287

RNA Biology 7:2, 125-129; March/April 2010; © 2010 Landes Bioscience

A family of non-classical pseudoknots in influenza A and B viruses

Alexander P. Gultyaev[1,2]* and René C.L. Olsthoorn[1]

[1]Department of Molecular Genetics; Leiden Institute of Chemistry; Leiden University; Leiden, The Netherlands; [2]Department of Virology; Erasmus Medical Center; Rotterdam

A very conserved pseudoknot structure has been shown to fold in influenza virus RNA. The pseudoknot encompasses the 3' splice site of segment 8 RNA in both influenza A and B viruses. By sequence comparison of influenza virus strains, we derive a consensus motif that defines a novel RNA pseudoknot family. The orientation of the coaxially stacked stems in the influenza pseudoknot differs from that in classical H-pseudoknots. Apart from the size of the loops, the topology of the influenza pseudoknot resembles that of some long-range pseudoknotted conformations. A seed alignment of the influenza pseudoknot family, containing representative strain sequences together with a consensus structure description, has been submitted to the RNA families (Rfam) database.

Key words: RNA secondary structure, RNA pseudoknot, RNA virus, influenza, splicing

Submitted: 10/17/09

Accepted: 01/25/10

Previously published online: www.landesbioscience.com/journals/rnabiology/article/11287

*Correspondence to: Alexander P. Gultyaev; Email: goultiaevap2@chem.leidenuniv.nl

Introduction

Many functional RNA motifs are formed by RNA pseudoknot structures, in particular, in viral RNA molecules.[1] Recently, a pseudoknot structure has been suggested to fold close to the intron/exon boundaries of segment 8 (or NS segment, coding for nonstructural proteins) RNAs from influenza A and B viruses.[2] These pseudoknots are very similar in the two types of influenza viruses (**Fig. 1**), despite considerable differences between the sequences of these homologous segments in influenza A and B genomes.[3,4] The influenza pseudoknot contains only two stems and thus belongs to the so-called H-type, although the orientation of the two stems differs from the most widespread "classical" H-pseudoknot orientation.[5,6] While in the classical configuration each of the pseudoknot loops bridges the ends of one of the stems, one of the loops in the influenza pseudoknot bridges the ends of a quasi-continuous helix formed by coaxial stacking of the two stems (**Fig. 1**).

Using comprehensive comparisons of influenza pseudoknot sequences from all phylogenetic lineages, accompanied by comparisons with other RNA pseudoknot folds, we show here that the influenza virus pseudoknots define a family of novel structures with a characteristic consensus base-pairing configuration and sequence.

Results

Thousands of NS segment sequences from different influenza strains are known nowadays.[7] Sequence comparison shows that most of the diversity in the pseudoknot regions is determined by differences between influenza A and B viruses and between two main clades (sometimes called "alleles") of influenza A NS segments, clade A and clade B.[2] This diversity is associated with nucleotide covariations in the upper stem of the pseudoknot (**Fig. 1**). The U.G base pair at the pseudoknot junction, mostly observed in clade A segments of influenza A viruses, is mutated to G.A in clade B, suggesting a non-canonical interaction. Two covariations at other positions of this stem are also observed between influenza A and B strains. The lower stems of the pseudoknots have remarkably conserved paired sequences GAGGA/UCCUC, despite their slightly different location in relation to the 3'ends of introns that are present in both virus types and determine the alternative splicing with two encoded proteins, NS1 and NS2.[8]

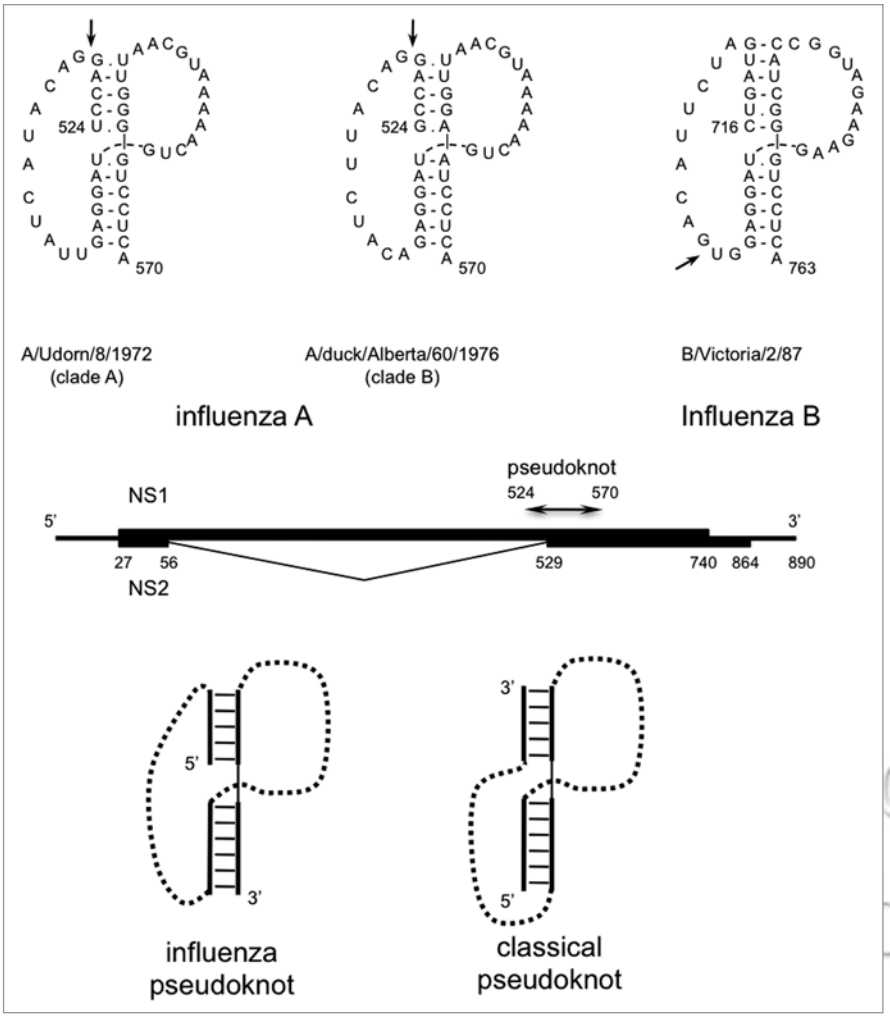

Figure 1. Examples of conserved pseudoknot structures in segment 8 mRNAs of influenza virus strains.[2] The arrows indicate the 3' splice sites. The location of the pseudoknot region in the influenza A segment 8 is also shown in relation to the NS1 and NS2 coding regions. The nucleotide numbering corresponds to the complete lengths of genome segments. The orientation of stems in the influenza pseudoknot is compared schematically with that in common classical H-pseudoknot.[5,6]

The lengths of the influenza pseudoknot stems and loops are very conserved, making the construction of an alignment of structures from both A and B viruses (**Fig. 2**) rather straightforward. There is only one gap in the sequence/structure alignment: insertion of a nucleotide in the loop of influenza B pseudoknot as compared to that of influenza A. Note, however, that this difference seems to be a result of "sliding" of the 5'-proximal part of the stem along the homologous positions rather than of an insertion or deletion of a single nucleotide which would create a shift of reading frame in the unspliced NS1 mRNA. In this way, it is possible to construct a consistent RNA sequence/structure alignment in the so-called Stockholm format, that may serve as a general description of the conserved base pairs and structural motif.

In order to elucidate consensus features of the pseudoknot motif, in the construction of the Stockholm alignment (see Suppl. Material) we aimed to cover the diversity of pseudoknot sequences in the main phylogenetic lineages of NS genes in influenza virus strains[9-12] (**Fig. 3**), while excluding strains where a conformational shift from the pseudoknot to an alternative structure could occur.[2] The selection of representative strains from the lineages (**Table 1**) was assisted by analysis of alignments in order to maximize sequence diversity.

Clade A of influenza A viruses can be subdivided into 8 lineages with obvious host and/or geographic attributes (**Fig. 3**).[9,10] One representative strain was selected from each of these lineages but one. A group of equine strains (A/equine/Prague/1956-like) was excluded, because these sequences contained a mutation (substitution of guanine by adenine at position 562 in the full-length genome segment) disrupting the second base pair in the upper stem, probably making the pseudoknot conformation less likely than an alternative hairpin fold. Such a conformational transition was also suggested to occur in currently dominant H5N1 strains because of a unique C563 substitution at the neighboring position.[2] Thus, a different representative from the H5N1-containing lineage (Eurasian avian strains) was selected.

Influenza A clade B sequences, found only in avian strains, can be subdivided into two major lineages, North American and Eurasian, with further minor divergence in more recent isolates as compared to the older strains.[11] Four sequences were selected: two from North American cluster and two from Eurasian one.

Four influenza B representatives were selected, following the main clustering of the NS-genes (**Fig. 3**).[12] Because representatives from two of the main four groups of influenza B have identical pseudoknot sequences, no representative from one of them was taken. Instead, a sequence of the pseudoknot from the oldest known isolate (B/Lee/40), with two unique substitutions in the pseudoknot loop, was added to the alignment.

The sequences of the representative strains in the Stockholm alignment (**Fig. 2**) define the consensus of the pseudoknot motif in both influenza A and B viruses, with similar stem/loop topology and conserved sequence GAGGA…UCCUC in the "bottom" stem. It should be noted, however, that the strong conservation of this sequence may be, at least partially, determined by a strong pressure to preserve amino acid sequences of NS1 and/or NS2 proteins in the overlapping reading frames.

Discussion

An interesting aspect of the influenza pseudoknot is the unusual orientation

```
# STOCKHOLM 1.0
A/Udorn/8/1972     UUCCAGGA-CAUACUAUUGAGGAUGUCAAAAAUGCAAUUGGGGUCCUCA
A/sw/Iowa/1930     UUCCAGGA-CAUACUGAUGAGGAUGUCAAAAAUGCAGUUGGGGUCCUCA
A/ch/Brescia/1902  UUCCAGGA-CAUACUGAUGAGGAUGUCAAAAAUGCAAUUGGGGUCCUCA
A/budg/Hokk/1977   UUUCAGGA-CAUACUAAUGAGGAUGUCAAAAAUGCAAUUGGGAUCCUCA
A/mlrd/NY/1978     UUCCAGGA-CAUACUGAUGAGGAUGUCAAAAAUGCAAUUGGGGUCCUCA
A/eq/Font/1979     UUCCAGGA-CAUACUAAUGAGGAUGUCAAAAAUGCAAUUGGGGUCCUCA
A/gull/MD/1977     UUCCAGGA-CAUACUAGUGAGGAUGUCAAAAAUGCAAUUGGAAUCCUCA
A/duck/Alb/1976    UGCCAGGA-CAUUCUACAGAGGAUGUCAAAAAUGCAAUUGGAAUCCUCA
A/ch/NY/2005       UGCCAGGG-CAUUCUGGAGAGGAUGUCAAAAAUGCAAUUGGAAUCCUCA
A/ch/Germany/1949  UGCCAGGA-CAUUCUACAGAGGAUGUCAAAAAUGCAAUUGGAAUCCUCA
A/turkey/Ita/2001  UGCCAGGA-CAUUCUACAGAGGAUGUCAAAAAUGCAGUUGGAAUCCUCA
B/Vic/1987         ACUGAUGAUCUUACAGUGGAGGAUGAAGAAGAUGGCCAUCGGAUCCUCA
B/Lee/40           ACUGAUGAUCGGACAGUGGAGGAUGAAAAAGAUGGCCAUCGGAUCCUCA
B/GL/54            ACUGAUGAUCUUACAGUGGAGGAUGAAAAAGAUGGCCAUCGGAUCCUCA
B/Argentina/97     ACUGAUGAUCUUACAGUGGAGGAUGAAGAAGAUGGUCAUCGGAUCCUCA
#=GC SS_cons       .<<<<<............AAAAAA.............>>>>>aaaaaa.
//
```

Figure 2. The Stockholm sequence/structure alignment of pseudoknots from representative strains. The full strain names are given in Table 1. The base-pairing is indicated with combinations of brackets, <<<…>>>, and capital/lower case characters, AAA…aaa. Sequences conserved in all representatives are shown in bold type.

of the two pseudoknotted stems (**Fig. 1**). It corresponds to one of the three stem stacking topologies proposed for the so-called H-type (hairpin) pseudoknots.[5,6,13,14] Due to the handedness of the RNA double helix and polarity of the polynucleotide chain, different orientations of coaxially stacked stems in pseudoknots define different stereochemical locations of the loops bridging the stem ends. The stem configuration in the influenza pseudoknot differs from the most widespread orientation of classical H-pseudoknot.

In the classical H-pseudoknot (**Fig. 1**), the loop bridging the ends of the "upper" stem crosses the deep groove and the other loop spans the shallow groove of the "lower" stem.[15,16] In the suggested influenza pseudoknot, the first pseudoknot loop should span both stems. The 3D-modeling of such a configuration[6,13,14] suggests that this loop does not cross any groove, while the second loop should cross the deep groove of the upper stem. To the best of our knowledge, the pseudoknot loops in similar orientation have been observed only in structural context of long-range tertiary interactions, e.g., in various ribozymes[13,17-20] and in long-range pseudoknot in bacteriophage Qβ RNA.[21] In

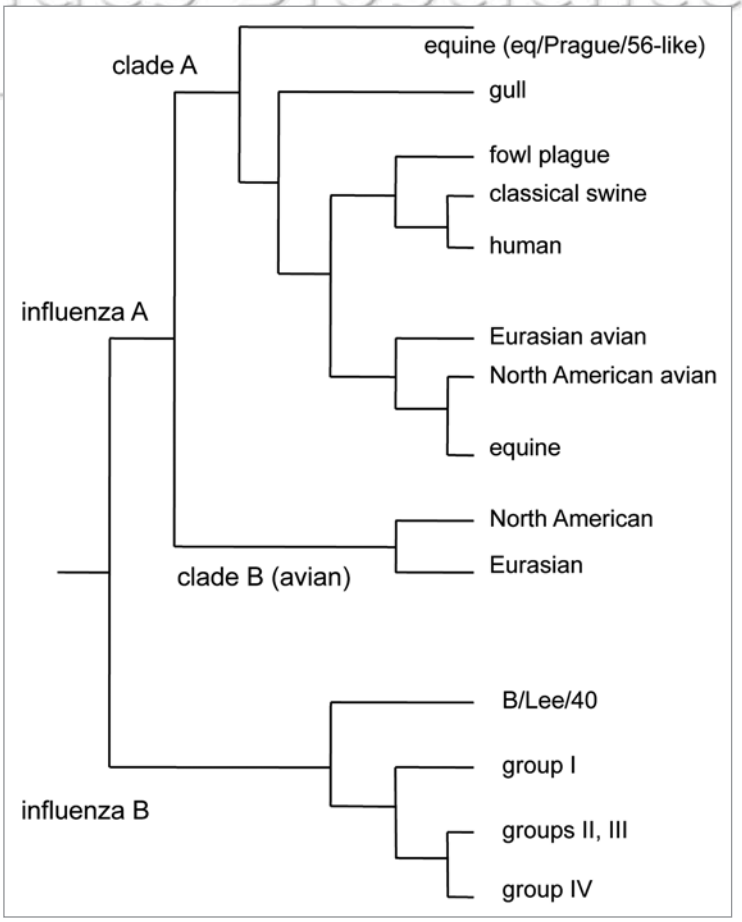

Figure 3. Phylogenetic relationships between the main lineages of NS genes in influenza A and B viruses.[9-12]

Table 1. Representative virus strains selected for the SEED alignment of the influenza RNA pseudoknot family

Virus type, clade and strain names	Lineage	Accession
Influenza A, clade A:		
A/Udorn/8/1972 (H3N2)	human	V01102
A/swine/Iowa/15/1930 (H1N1)	classical swine	M80965
A/chicken/Brescia/1902 (H7N7)	fowl plague	L37798
A/budgerigar/Hokkaido/1/1977 (H4N6)	Eurasian avian	M80942
A/mallard/New York/6750/1978 (H2N2)	North American avian	M80945
A/equine/Fontainebleau/1979 (H3N8)	equine	M80953
A/gull/Maryland/704/1977 (H13N6)	gull	M80959
Influenza A, clade B:		
A/duck/Alberta/60/1976 (H12N5)	North American	J02105
A/chicken/New York/23164-9/2005 (H7N2)	North American	CY031703
A/chicken/Germany/n/1949 (H10N7)	Eurasian	CY014675
A/turkey/Italy/1351/2001 (H7N1)	Eurasian	CY021425
Influenza B:		
B/Victoria/02/1987	III	CY018761
B/Lee/40	I	J02096
B/GL/54	I	M19797
B/Argentina/218/97	IV	AF100403

The lineages are given according to the published phylogenetic trees.[9-12] Influenza A virus subtypes are indicated in parentheses.

all these cases, the loop bridging the ends of two coaxially stacked stems is relatively large and contains one or more structured domains (**Fig. 4**).

For instance, in the so-called P11/P7 pseudoknot, conserved in the subgroup IA1 group I introns,[13] the 5'region of the pseudoknot loop is folded into a hairpin (P6b) which is coaxially stacked on the pseudoknot stem P11 (**Fig. 4**). The remaining part of the pseudoknot loop is involved in pairing to other regions resulting in stems P6a, P6 and P7. Structured regions and pseudoknot loop interactions with distant RNA regions are also formed in other pseudoknots of this kind.[17-21] In principle, the formation of such long-distance pseudoknots is equivalent to base-pairing between 3'-proximal nucleotides of hairpin loops with distant upstream regions of the RNA.[13] In phage Qβ and ribozyme pseudoknots, these upstream regions are located rather far from the hairpins they bind, and are separated by the regions involved in the interactions elsewhere, so that the highly-ordered region connecting the ends of two coaxially stacked pseudoknotted stems can hardly be called a loop. In contrast, in the suggested influenza pseudoknot the upstream paired region is rather close to the hairpin, resulting in a relatively short loop of only 11 nucleotides.

Although additional interactions of the influenza pseudoknot loops with distant RNA regions cannot be excluded, these loops themselves do not seem to form internal secondary structures.[2] The loop connecting the ends of a putative quasi-continuous helix of two stems with 11–12 base pairs in total has to span approximately one turn of A-form helix and thus can be located on one side of RNA duplex without wrapping it. Apparently, the loop length of 11 nucleotides, as seen in both influenza A and B pseudoknots (**Fig. 1**) would not present any restriction here. The other loop, crossing the deep groove of the top stem, is actually significantly longer than what is required to connect the ends of the stem with 5–6 base pairs across the deep groove. Such a loop can be as small as just one nucleotide in a classical pseudoknot,[15,22] and indeed, at least 66 entries out of 309 pseudoknot structures stored in the pseudoknot database[23,24] do have such a loop. Furthermore, the non-classical configuration of the influenza pseudoknot should not make any difference in the constraints for this type of loop:[13] for instance, a single nucleotide in the loop can span the P11 stem in the group I intron pseudoknot as well (**Fig. 4**). Therefore, the loop of 12–13 nucleotides in the influenza pseudoknots (**Figs. 1 and 4**), with relatively long polynucleotide chain spanning a short distance, may have a loose structure or be involved in other interactions. For instance, a topologically similar loop of 11 nucleotides in the glmS ribozyme pseudoknot interacts with the other loop of the pseudoknot, resulting in a complex double pseudoknot structure.[19,20]

Thus the pseudoknot structures, folded in the NS segments of influenza A and B viral genomes, can be classified as a separate family of conserved RNA motifs. Their function is yet to be elucidated. The close proximity to splicing sites (**Fig. 1**) suggests their probable involvement in the regulation of RNA processing, but other roles cannot be excluded.

The Stockholm alignment, shown in **Figure 2**, can be used as a seed alignment in the search of other sequences satisfying this motif, using the procedures of the Rfam database.[25,26] This SEED alignment should be considered as an update and revision of the one used in building the Rfam family (Rfam Accession RF01099) for the influenza A pseudoknot, with the addition of influenza B motif and the selection of representative pseudoknot-containing virus strains. Two representative pseudoknot sequences of influenza A and B viruses are also submitted to the pseudoknot database PseudoBase[23,24] (Accessions PKB273, PKB274).

Materials and Methods

The genomic sequences of influenza virus strains were retrieved from the Influenza Virus Resource database.[7] The available options of this database were also used for the selections of sequences from particular virus groups and multiple sequence alignments. The database of RNA pseudoknots Pseudobase,[23] with its option to sort pseudoknot structures, was used for the statistics of loop lengths in classical pseudoknots. In the database estimates of the number of pseudoknot loops with one nucleotide,

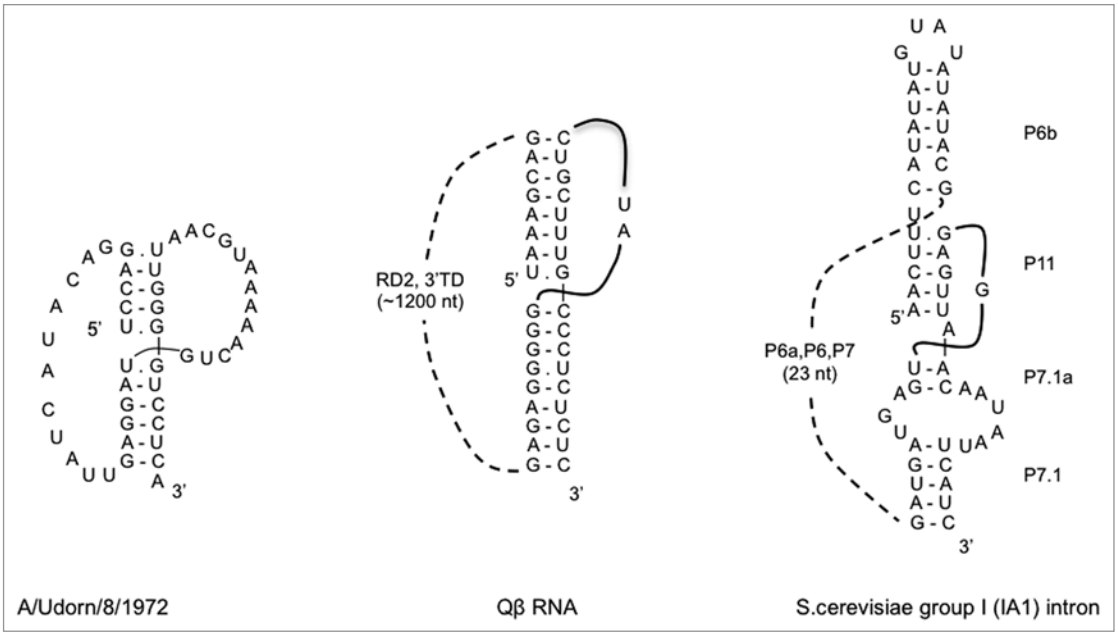

Figure 4. Comparison of stem-loop topologies in the influenza pseudoknot (strain A/Udorn/8/1972) with those in the phage Qβ long-distance pseudoknot[21] and in the P11/P7 pseudoknot of the group I intron (subgroup IA1) in the large ribosomal RNA of *S. cerevisiae* mitochondria according to the model of Jaeger et al.[13] For simplicity, large pseudoknot loops involved in other interactions and domains[13,21] are depicted schematically with dashed lines.

only the pseudoknots with no free nucleotides at the stem junction were taken into account, in order to ensure exclusive selection of classical pseudoknots.

Acknowledgements

This work was supported by the European Commission grant no. 201607 (also known under the acronym RNAFLU). The authors thank Maarten de Smit for critical reading of the manuscript.

Note

Supplementary materials can be found at: www.landesbioscience.com/supplement/GultyaevRNA7-2-Sup01.pdf

References

1. Brierley I, Pennell S, Gilbert RJ. Viral RNA pseudoknots: versatile motifs in gene expression and replication. Nature Rev Microbiol 2007; 5:598-610.
2. Gultyaev A, Heus HA, Olsthoorn RCL. An RNA conformational shift in recent H5N1 influenza A viruses. Bioinformatics 2007; 23:272-6.
3. Briedis DJ, Lamb RA. Influenza B virus genome: sequences and structural organization of RNA segment 8 and the mRNAs coding for the NS1 and NS2 proteins. J Virol 1982; 42:186-93.
4. Wang W, Krug RM. The RNA-binding and effector domains of the viral NS1 protein are conserved to different extents among influenza A and B viruses. Virology 1996; 223:41-50.
5. Pleij CWA. Pseudoknots: a new motif in the RNA game. Trends Biochem Sci 1990; 15:143-7.
6. Westhof E, Jaeger L. RNA pseudoknots. Curr Opin Struct Biol 1992; 2:327-33.
7. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, et al. The Influenza Virus Resource at the National Center for Biotechnology Information. J Virol 2008; 82:596-601.
8. Lamb RA, Horvath CM. Diversity of coding strategies in influenza viruses. Trends Genet 1991; 7:261-6.
9. Kawaoka Y, Gorman OT, Ito T, Wells K, Donis RO, Castrucci MR, et al. Influence of host species on the evolution of the nonstructural (NS) gene of influenza A viruses. Virus Res 1998; 55:143-56.
10. Basler CF, Reid AH, Dybing JK, Janczewski TA, Fanning TG, Zheng H, et al. Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. PNAS 2001; 98:2746-51.
11. Spackman E, Stallknecht DE, Slemons RD, Winker K, Suarez DL, Scott M, Swayne DE. Phylogenetic analyses of type A influenza genes in natural reservoir species in North America reveals genetic variation. Virus Res 2005; 114:89-100.
12. Lindstrom SE, Hiromoto Y, Nishimura H, Saito T, Nerome R, Nerome K. Comparative analysis of evolutionary mechanisms of the hemagglutinin and three internal protein genes of influenza B virus: multiple cocirculating lineages and frequent reassortment of the NP M, and NS genes. J Virol 1999; 73:4413-26.
13. Jaeger L, Westhof E, Michel F. Function of P11, a tertiary base pairing in self-splicing introns of subgroup IA. J Mol Biol 1991; 221:1153-64.
14. Mans RMW, Pleij CWA. RNA pseudoknots. In: Eckstein F, Lilley DMJ, eds. Nucleic Acids and Molecular Biology. Berlin Heidelberg: Springer-Verlag 1993; 7:250-70.
15. Pleij CWA, Rietveld K, Bosch L. A new principle of RNA folding based on pseudoknotting. Nucleic Acids Res 1985; 13:1717-31.
16. Kolk M, van der Graaf M, Wijmenga SS, Pleij CWA, Heus HA, Hilbers CW. NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. Science 1998; 280:434-8.
17. Costa M, Michel F, Westhof E. A three-dimensional perspective on exon binding by a group II self-splicing intron. EMBO J 2000; 19:5007-18.
18. Bergman NH, Lau NC, Lehnert V, Westhof E, Bartel DP. The three-dimensional architecture of the class I ligase ribozyme. RNA 2004; 10:176-84.
19. Cochrane JC, Lipchock SV, Strobel SA. Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. Chem Biol 2007; 14:97-105.
20. Klein DJ, Wilkinson SR, Been MD, Ferré-D'Amaré AR. Requirement of helix P2.2 and nucleotide G_1 for positioning the cleavage site and cofactor of the glmS ribozyme. J Mol Biol 2007; 373:178-89.
21. Klovins J, van Duin J. A long-range pseudoknot in Qβ RNA is essential for replication. J Mol Biol 1999; 294:875-84.
22. Michiels PJA, Versleijen AAM, Verlaan PW, Pleij CWA, Hilbers CW, Heus HA. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. J Mol Biol 2001; 310:1109-23.
23. van Batenburg FHD, Gultyaev AP, Pleij CWA. PseudoBase: structural information on RNA pseudoknots. Nucleic Acids Res 2001; 29:194-5.
24. Taufer M, Licon A, Araiza R, Mireles D, van Batenburg FHD, Gultyaev AP, Leung M-Y. PseudoBase++: an extension of PseudoBase for easy searching, formatting and visualization of pseudoknots. Nucleic Acids Res 2009; 37:127-35.
25. Griffitths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. Nucleic Acids Res 2003; 31:439-41.
26. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, et al. Rfam: updates to the RNA families database. Nucleic Acids Res 2009; 37:136-40.