BMC
Research Notes

SHORT REPORT   **Open Access**

Identification of conserved RNA secondary structures at influenza B and C splice sites reveals similarities and differences between influenza A, B, and C

Lumbini I Dela-Moss, Walter N Moss and Douglas H Turner[*]

Abstract

Background: Influenza B and C are single-stranded RNA viruses that cause yearly epidemics and infections. Knowledge of RNA secondary structure generated by influenza B and C will be helpful in further understanding the role of RNA structure in the progression of influenza infection.

Findings: All available protein-coding sequences for influenza B and C were analyzed for regions with high potential for functional RNA secondary structure. On the basis of conserved RNA secondary structure with predicted high thermodynamic stability, putative structures were identified that contain splice sites in segment 8 of influenza B and segments 6 and 7 of influenza C. The sequence in segment 6 also contains three unused AUG start codon sites that are sequestered within a hairpin structure.

Conclusions: When added to previous studies on influenza A, the results suggest that influenza splicing may share common structural strategies for regulation of splicing. In particular, influenza 3' splice sites are predicted to form secondary structures that can switch conformation to regulate splicing. Thus, these RNA structures present attractive targets for therapeutics aimed at targeting one or the other conformation.

Keywords: Influenza, RNA, Secondary structure, Splice sites, Bioinformatics, Splicing

Findings

Background

Influenza virus causes more than 200,000 hospitalizations and about 3000 – 49,000 deaths per year in the United States alone [1,2]. Influenza A, B, and C viruses belong to the family *Orthomyxoviridae* and are characterized by segmented, single-stranded, negative-sense (−) RNA genomes. These viruses share a common ancestry but are also genetically distant, such that segment reassortment does not occur between each group [3]. Each of the (−) RNA segments is used as a template to produce two types of positive-sense (+) RNA with distinct functions: mRNA for protein production and complementary RNA (cRNA) for viral replication. Influenza B has eight genome segments that encode at least eleven proteins and influenza C has seven genome segments that encode at least nine proteins. Influenza A infects avian, human, swine, and many other mammalian species, whereas influenza B and C infect primarily humans [4-6]. Influenza B and C do not undergo pandemic-causing antigenic shifts (reassortment of segments from different subtypes) like influenza A, because both viruses contain only one antigenic subtype and have limited host specificity [7,8]. All influenza viruses are able to undergo antigenic drift, which occurs as a result of accumulation of mutations in the antigenic sites [3,7,8]. Concern, however, has been growing, as two lineages of influenza B (Yamagata and Victoria) have been co-circulating in the human population [7,9]. This has led to a novel formulation of a quadrivalent vaccine: against two strains of influenza A and two strains of influenza B [10,11], rather than the previous trivalent vaccine.

* Correspondence: turner@chem.rochester.edu
Department of Chemistry and Center for RNA Biology, University of Rochester, Rochester, New York 14627-0216, USA

BioMed Central

RNA structure plays important roles in many viruses. For example, internal ribosome-entry sites (IRES) in viral mRNAs are heavily structured regions, which initiate cap-independent translation by directly binding to the ribosome [12,13]. RNA structure is also used for start codon selection and viral replication [14], for packaging signals [15], for RNA editing [16], and for many more functions. RNA secondary structure also plays an important role in viral mRNA splicing regulation [17-19]. A relatively rare type of RNA structure, pseudoknots, often plays important roles in biology [20-22]. In particular, pseudoknots are important in the regulation of viral gene expression and genome replication [23,24].

RNA structure is also important in influenza. The 5′ and 3′ ends of each genome segment of influenza A, B, and C are highly conserved, partially complementary, and base pair to form a promoter region that can be either in a panhandle or corkscrew conformation [25-27]. This structure is essential in vRNA transcription, replication, and viral packaging [28-30].

A variety of de novo methods exist to predict conserved secondary structure in genomes [31-35]. RNA structure can occur in protein coding regions and has many potential functional roles [34,36]. A survey for conserved secondary structure in the (+) and (−) RNAs of influenza A was carried out [37] by scanning for thermodynamically stable and conserved regions with the program RNAz [38-40] and coupling this with evidence of suppression of synonymous codon usage (SSCU), which identifies possible constraints of secondary structure acting on codon diversity [34,41,42]. Twenty conserved, thermodynamically stable regions were identified. Secondary structure is strongly favored in the (+) RNA. Of these predicted regions, five occur at or near functionally relevant sites [37]. Two of these, occurring in the segment 8 (+) RNA, were previously proposed [43-45].

This paper extends the search for influenza RNA structure in coding regions to influenza B and C, where conserved and thermodynamically stable regions are predicted to occur at splice sites. The secondary structures of these splice sites are modeled here. The results suggest that influenza RNA splicing may share common structural strategies between the three viral species.

Methods
Influenza B and C sequences
The sequences used in this study were obtained from the National Center for Biotechnology Information (NCBI) Influenza Virus Resource [46]. All non-redundant sequences for each segment of influenza B and C were downloaded for the prediction of conserved secondary structure.

Predicting conserved, thermodynamically stable regions
All non-redundant sequences for segments coding a single protein were translated into amino acid sequences via Seaview 4.3.0 [47,48] and aligned with ClustalW [49]. The aligned sequences were then converted back to nucleotides. Non-redundant sequences of segments that code multiple proteins were aligned according to nucleotides via MAFFT with FFT-NS-i strategy and default parameters [50,51].

Alignments were split into windows of 120 nucleotides (nt) with a step size of 10 nt. Between 6 and 50 sequences, with an average pairwise identity of 80%, were selected for scoring by RNAz 2.1 using the RNAz dinucleotide-shuffling model [39,40]. For a given alignment, RNAz calculates a z-score as an estimate of normalized difference in thermodynamic stability of native versus dinucleotide randomized sequences, and a structure conservation index (SCI), which measures the conservation of the minimum free energy of the consensus RNA fold in the alignment. RNAz then uses these as features in a support vector machine (SVM) to output an RNA class probability (p-class), which classifies the RNA fragment as structured or not.

RNA secondary structure modeling
Five regions within or overlapping RNAz predicted windows with high thermodynamic stability/conservation and/or that contain splice sites were structurally modeled. These regions were extracted as alignments and submitted to RNAalifold [52]. RNAalifold predicts structure via thermodynamic energy minimization [53] coupled with a scoring model for evolutionary conservation. The resulting consensus sequence was also submitted to RNAstructure [54], which utilizes a revised set of nearest neighbor energy parameters to fold single sequences [55]. The minimum free energy (MFE) structure and suboptimal structures [56] were analyzed based on folding free energy and base pairing probability from the calculated base pair partition function [57] and compared to the RNAalifold results. Fragments within these predicted structural regions were extracted for further analysis based on having higher probability pairs (from the partition function) than surrounding structure and on their conservation in the alignment of all non-redundant influenza B or C sequences (paying special attention to evidence of consistent and compensatory mutations). The resulting structural models were used to constrain MC-Fold [58] calculations to suggest possible non-canonical base pairing interactions in predicted loop regions. MC-Fold utilizes high resolution RNA structural information from the Protein Data Bank (PDB) to estimate non-canonical base pairing energies.

DotKnot [59], which folds single sequences, was used to predict pseudoknots in sequences containing splice

sites. DotKnot extracts possible stem regions from RNA secondary structure partition function dot plots and assembles pseudoknots according to free energy parameters. Free energies of the pseudoknots were computed with experimentally based thermodynamic energy models [53,60] and loop entropy parameters derived from a diamond lattice model [61,62].

Results and discussion

The bioinformatics survey for structured RNAs in influenza B and C revealed multiple regions with putative conserved RNA structure (Table 1). Many more high probability prediction windows were identified in influenza C, but these are likely false positives due to the lack of diversity in the input sequences. In several cases there were as few as two influenza C sequences with which to base predictions, versus hundreds in influenza B, and thousands in influenza A. Nevertheless, significant predictions were made in segments 6 and 7 of influenza C, where 28 and 50 sequences were available, respectively. In general, fewer sequence variants for B and C are available versus influenza A due to their lower mutation

rate [63-67] and fewer resources for acquiring sequence data for these groups.

Similar to results for influenza A [37], where predicted conserved structure appears at or near splice sites, influenza B and C splice sites show evidence for having stable and/or conserved RNA secondary structure (Figure 1). Structural modeling in these regions reveals RNA structures with similarities between influenza A, B, and C, suggesting common strategies for regulation of splicing. In influenza C segment 7, the region near the 3′ splice site is not predicted to have strong structure (p-class range of 0.01-0.12), but a pseudoknot is predicted to occur in this region using the DotKnot program, which is a type of motif forbidden in the RNAz folding algorithm.

Structures predicted at 5′ splice sites of NEP mRNA in influenza B and C

Segment 8 in influenza B and segment 7 in influenza C both encode the nuclear export protein, NEP (NS2), involved in vRNP export [68] and in viral transcription and replication regulation [69]. NEP is expressed late in

Table 1 Summary of RNAz scans of predicted structured regions in influenza B and C

	Segment	Protein coded[a]	Region	z-score[b]	SCI[b]	p-class[b]
Influenza B	8	NS1, *NEP*	30–220	−1.44 (−2.49)	0.78 (0.86)	0.30 (0.94)
			450–620	−2.27 (−3.29)	0.83 (0.91)	0.81 (0.99)
			720–840	−0.98	0.94	0.37
	7	M1, BM2	0–170	−2.12 (−2.63)	0.66 (0.78)	0.28 (0.56)
	5	NP	1040–1160	−2.29	0.86	0.93
	4	HA	0–160	−2.22 (−3.21)	0.78 (0.87)	0.76 (0.96)
			780–950	−0.81 (−1.4)	0.80 (0.89)	0.20 (0.67)
	3	PA	630–750	−2.19	0.91	0.94
			1870–1990	−2.25	0.66	0.31
			2010–2130	−0.85	0.96	0.46
	2	PB1	650–770	−1.67	0.81	0.39
			1100–1220	−0.57	0.97	0.31
	1	PB2	210–350	−2.82 (−3.12)	0.68 (0.69)	0.85 (0.94)
			670–810	−1.49 (−1.85)	0.83 (0.86)	0.38 (0.62)
			1320–1450	−2.50 (−2.51)	0.76 (0.81)	0.77 (0.91)
Influenza C	7	NS1, *NEP*	190–400	−2.16 (−3.55)	0.70 (0.82)	0.41 (0.99)
			640–800	−1.39 (−2.04)	0.83 (0.87)	0.31 (0.79)
	6	*M1*, CM2	0–160	−2.12 (−2.87)	0.86 (0.91)	0.72 (0.96)
			160–310	−1.13 (−1.36)	0.92 (0.93)	0.33 (0.54)
			350–470	−0.68	0.99	0.37
			450–680	−1.54 (−2.65)	0.96 (0.99)	0.65 (0.98)
			670–790	−1.05	0.97	0.54
			990–1133	−0.97 (−1.19)	0.98	0.54 (0.71)

[a]Proteins coded via splicing are italicized.
[b]Average values for overlapping windows with p-class >0.3. When the region is defined by more than one 120 nt window and the most favorable value for a 120 nt window differs from the average value of overlapping windows, it is presented in parentheses.
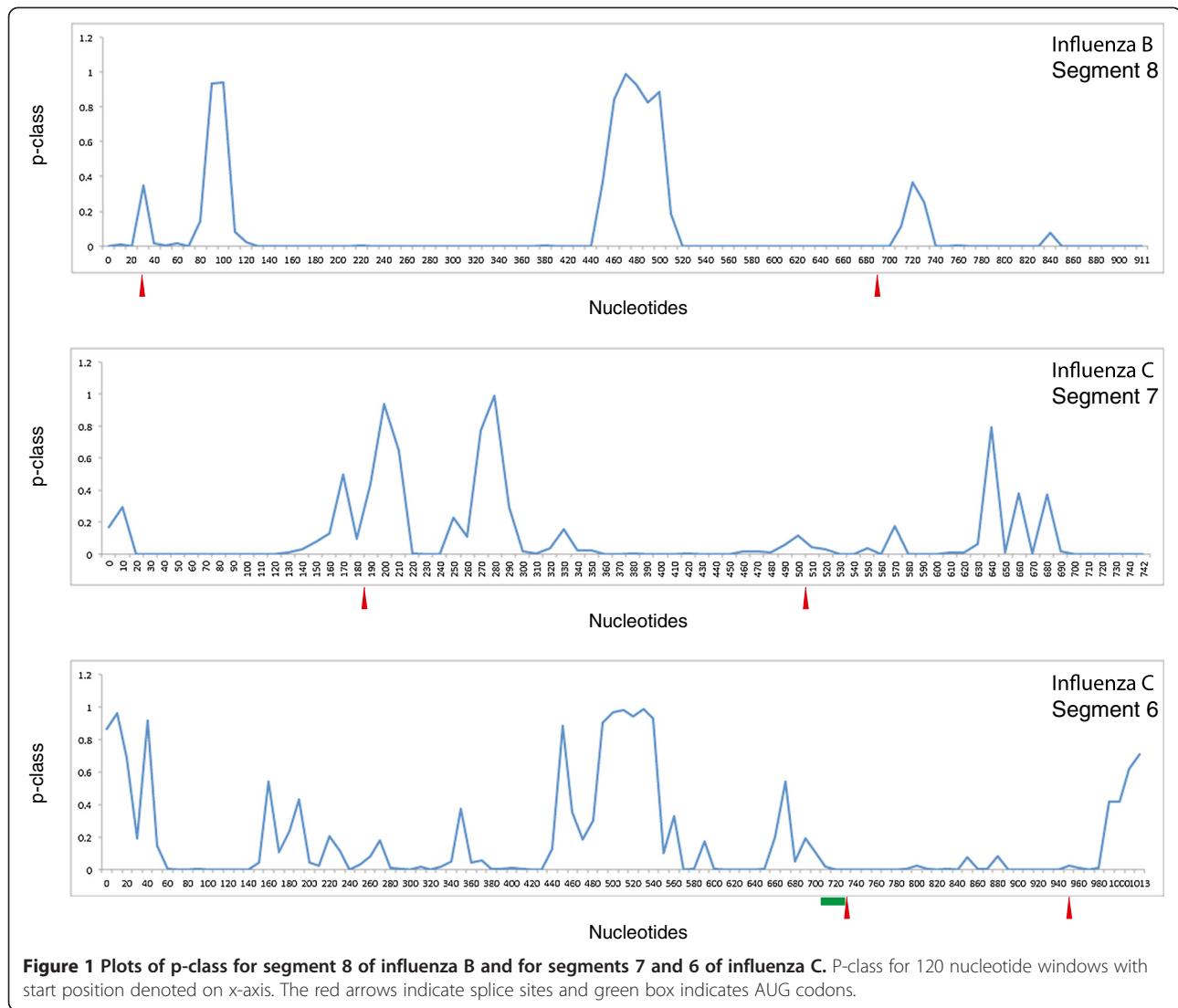
Figure 1 Plots of p-class for segment 8 of influenza B and for segments 7 and 6 of influenza C. P-class for 120 nucleotide windows with start position denoted on x-axis. The red arrows indicate splice sites and green box indicates AUG codons.

viral infection [8,69]. In influenza B and C, the mature NEP mRNA is generated via alternative splicing [8]. A conserved hairpin is predicted at the 5′ splice site in both influenza B (Figure 2) and C (Figure 3). In each case the 5′ splice site is contained within a helix. Sequestering a splice site in a helical region is a mechanism for regulating splicing [70]. For example, sequestering the 5′ splice site in a helix down-regulates splicing in the hnRNP A1 and SMN2 pre-mRNAs [71,72]. In rat calcitonin/CGRP pre-mRNA, a splice site appears near a 1x1 nucleotide homo-purine internal loop and mutations that change the loop into a Watson-Crick base pair inhibit in vitro splicing [73]; interestingly, in each of the predicted influenza structures containing a 5′ splice site, it also occurs near a homo-purine 1x1 nucleotide internal loop (Figures 2 and 3). Non-Watson-Crick pairs are important in RNA-protein interactions [74] and homo-purine pairs can increase protein binding affinity [75].

These predicted hairpins in influenza B (Figure 2) and C (Figure 3) have >99% base pair conservation. When mutations occurred in stems, they most often were consistent with base pairing. For example, in the nucleotides bordering the splice sites, mutations preserve base pairing: G36-C63 to a G36-U63 in B (Figure 2) and U186-G246 to U186-A246 in C (Figure 3). When mutations in helixes did not maintain canonical pairing, they most often resulted in CA pairs. Protonated C-A$^+$ pairs are isosteric with GU pairs and maintain A form helixes [76]. In DNA helixes, C-A$^+$ pairs can have pKa's as high as 7.6 [77] suggesting a similar possibility for RNA. In the 5′ hairpin structure of influenza C segment 7, the A208-U228 pair occurs below the C209 bulge loop and mutates to an AC pair (Figure 3). The flanking GC pair (G207-C229) can mutate to an AA pair. In some contexts, protonated C-A$^+$ base pairs adjacent to AA pairs can stabilize internal loops [78] and may be important in RNA-protein interactions

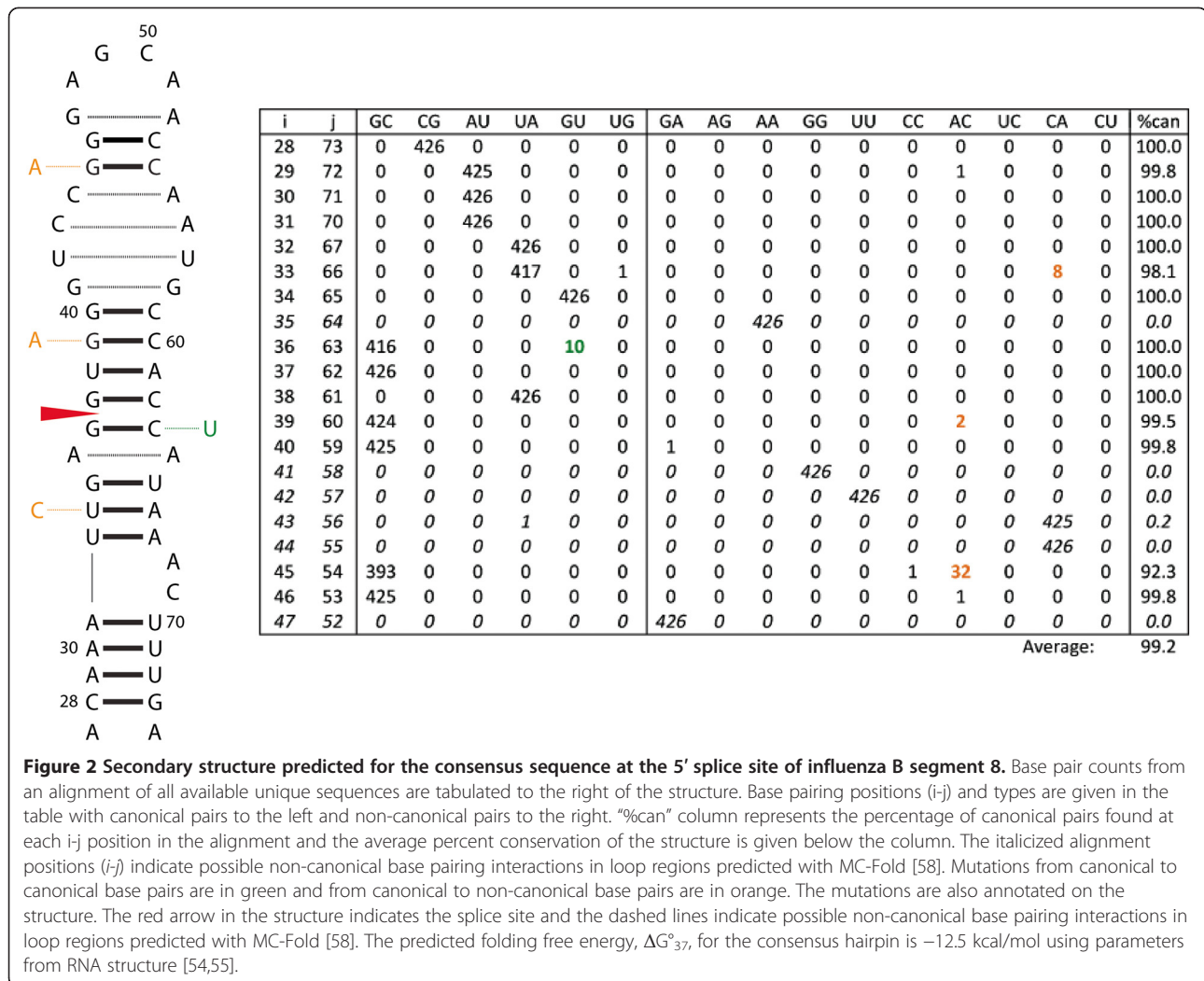| i	j	GC	CG	AU	UA	GU	UG	GA	AG	AA	GG	UU	CC	AC	UC	CA	CU	%can
28	73	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
29	72	0	0	425	0	0	0	0	0	0	0	0	0	1	0	0	0	99.8
30	71	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
31	70	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
32	67	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	100.0
33	66	0	0	0	417	0	1	0	0	0	0	0	0	0	0	8	0	98.1
34	65	0	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	100.0
35	64	0	0	0	0	0	0	0	0	426	0	0	0	0	0	0	0	0.0
36	63	416	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	100.0
37	62	426	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
38	61	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	100.0
39	60	424	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	99.5
40	59	425	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	99.8
41	58	0	0	0	0	0	0	0	0	0	426	0	0	0	0	0	0	0.0
42	57	0	0	0	0	0	0	0	0	0	0	426	0	0	0	0	0	0.0
43	56	0	0	0	1	0	0	0	0	0	0	0	0	0	0	425	0	0.2
44	55	0	0	0	0	0	0	0	0	0	0	0	0	0	426	0	0	0.0
45	54	393	0	0	0	0	0	0	0	0	0	0	1	32	0	0	0	92.3
46	53	425	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	99.8
47	52	0	0	0	0	0	0	426	0	0	0	0	0	0	0	0	0	0.0
																Average:		99.2

Figure 2 Secondary structure predicted for the consensus sequence at the 5′ splice site of influenza B segment 8. Base pair counts from an alignment of all available unique sequences are tabulated to the right of the structure. Base pairing positions (i-j) and types are given in the table with canonical pairs to the left and non-canonical pairs to the right. "%can" column represents the percentage of canonical pairs found at each i-j position in the alignment and the average percent conservation of the structure is given below the column. The italicized alignment positions (*i-j*) indicate possible non-canonical base pairing interactions in loop regions predicted with MC-Fold [58]. Mutations from canonical to canonical base pairs are in green and from canonical to non-canonical base pairs are in orange. The mutations are also annotated on the structure. The red arrow in the structure indicates the splice site and the dashed lines indicate possible non-canonical base pairing interactions in loop regions predicted with MC-Fold [58]. The predicted folding free energy, $\Delta G°_{37}$, for the consensus hairpin is −12.5 kcal/mol using parameters from RNA structure [54,55].

[79]. Simultaneous AA (207−229) and AC (208−228) mutations occur in one influenza C sequence (GenBank accession: AB034159).
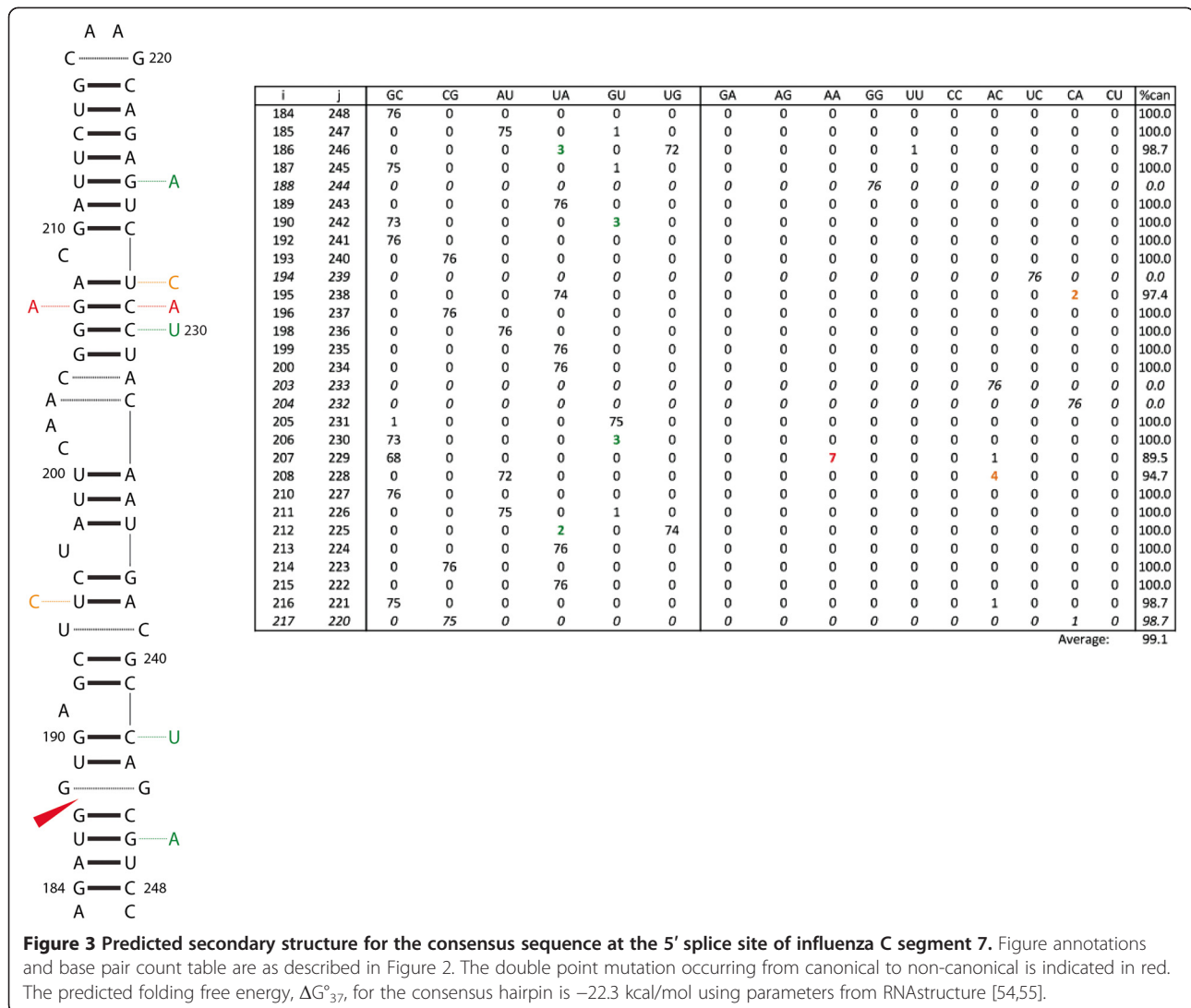
In contrast to influenza B and C, structures in influenza A are predicted to occur near (79 and 51 nts downstream from), but not overlap with, the 5′ splice sites in segments 7 and 8. Segment 8 of influenza A is homologous to segments 8 and 7 of B and C, respectively, and also produces mRNA for the NEP protein via alternative splicing. The structure in segment 8 of influenza A has been predicted to fold into an extended stem capped by a multibranch [37,45] or hairpin loop, where the hairpin is strongly favored in an avian enriched clade [37]. In vitro mapping experiments, however, reveal a hairpin structure for the consensus sequence that includes the human clade [80]. The 5′ structure in influenza A segment 7 is predicted to form an extended stem topped with a multibranch loop and in vitro mapping is consistent with this structure (Jiang T, Kierzek E, Moss WN, and Turner DH unpublished experiments). Thus,

structure is proposed to play roles in splicing of segments 7 and 8 of influenza A [37].

Structure predicted at the 5′ splice site of influenza C segment 6 contains in-frame start codons

A conserved hairpin is also predicted to form in segment 6 of influenza C (Figure 4), which codes for the M1 and P42 (M1′/CM2) proteins. The essential M1 matrix protein of influenza C is produced by splicing of segment 6 mRNA, whereas in influenza A and B, M1 is produced via un-spliced mRNA [81]. The 58 nucleotides surrounding the 5′ splice site of influenza C are predicted to fold into a hairpin, with the 5′ splice site contained within the apical tetra-loop (Figure 4). Splicing elements presented in single-stranded regions of hairpin loops can be more accessible to trans-acting splicing factors [82-85].

CM2 is an ion channel protein, which is also involved in packaging of vRNPs during virus assembly, and release of vRNPs during virus uncoating [86,87]. CM2 is

i	j	GC	CG	AU	UA	GU	UG	GA	AG	AA	GG	UU	CC	AC	UC	CA	CU	%can
184	248	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
185	247	0	0	75	0	1	0	0	0	0	0	0	0	0	0	0	0	100.0
186	246	0	0	0	3	0	72	0	0	0	0	0	0	1	0	0	0	98.7
187	245	75	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	100.0
188	244	0	0	0	0	0	0	0	0	0	76	0	0	0	0	0	0	0.0
189	243	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
190	242	73	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	100.0
192	241	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
193	240	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
194	239	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0	0.0
195	238	0	0	0	74	0	0	0	0	0	0	0	0	0	0	2	0	97.4
196	237	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
198	236	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
199	235	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
200	234	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
203	233	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0	0	0.0
204	232	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0.0
205	231	1	0	0	0	75	0	0	0	0	0	0	0	0	0	0	0	100.0
206	230	73	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	100.0
207	229	68	0	0	0	0	0	0	0	7	0	0	0	1	0	0	0	89.5
208	228	0	0	72	0	0	0	0	0	0	0	0	0	4	0	0	0	94.7
210	227	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
211	226	0	0	75	0	1	0	0	0	0	0	0	0	0	0	0	0	100.0
212	225	0	0	0	2	0	74	0	0	0	0	0	0	0	0	0	0	100.0
213	224	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
214	223	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
215	222	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
216	221	75	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	98.7
217	220	0	75	0	0	0	0	0	0	0	0	0	0	0	0	1	0	98.7
															Average:			99.1

Figure 3 Predicted secondary structure for the consensus sequence at the 5′ splice site of influenza C segment 7. Figure annotations and base pair count table are as described in Figure 2. The double point mutation occurring from canonical to non-canonical is indicated in red. The predicted folding free energy, $\Delta G°_{37}$, for the consensus hairpin is −22.3 kcal/mol using parameters from RNAstructure [54,55].

believed to be structurally and functionally equivalent to proteins M2 of influenza A and NB of influenza B [88,89]. CM2 was hypothesized to be produced by one of the three in-frame AUG start codons (Figure 4), especially the start codon at position 705–707 that occurs within a predicted strong ribosome initiation site sequence (RNNAUGG) [90]. It was later found that CM2 is instead produced by proteolytic cleavage of an internal signal peptide in the P42 (M1′/CM2) protein [91,92]. The precise mechanism for the lack of translation initiation at these cryptic start sites is unknown. Interestingly, all three AUG codons occur in the helices of the predicted hairpin (Figure 4). Cryptic AUG start codons sequestered in helices are also found in Polio and Coxsackieviruses [93,94]. Translation initiation is commonly reduced when start codons are embedded in RNA secondary structure [95,96], especially when the mRNA folding free energy near the start codon is more

favorable than roughly −12 kcal/mol [97]. Notably, the influenza C segment 6 hairpin has a predicted folding free energy of −13.9 kcal/mol at 37°C suggesting this hairpin may suppress the use of these cryptic start codons in addition to influencing splicing of M1.

The hairpin structure has 99.8% base pair conservation, with two consistent mutations (Figure 4); U706-A745 and A719-U734 change to UG and GU, respectively, preserving base pairing.

Structures predicted at 3′ splice sites

Stable and conserved structures are predicted for the nucleotides surrounding the 3′ splice sites in segment 8 of influenza B (Figure 5) and segment 7 of influenza C (Figure 6). In segment 8 of influenza B, a 53 nt region including the 3′ splice site can fold into either a hairpin or a pseudoknot [43,44]. In the hairpin (Figure 5), the 3′ splice site occurs in a helical region near a 1x1, homo-
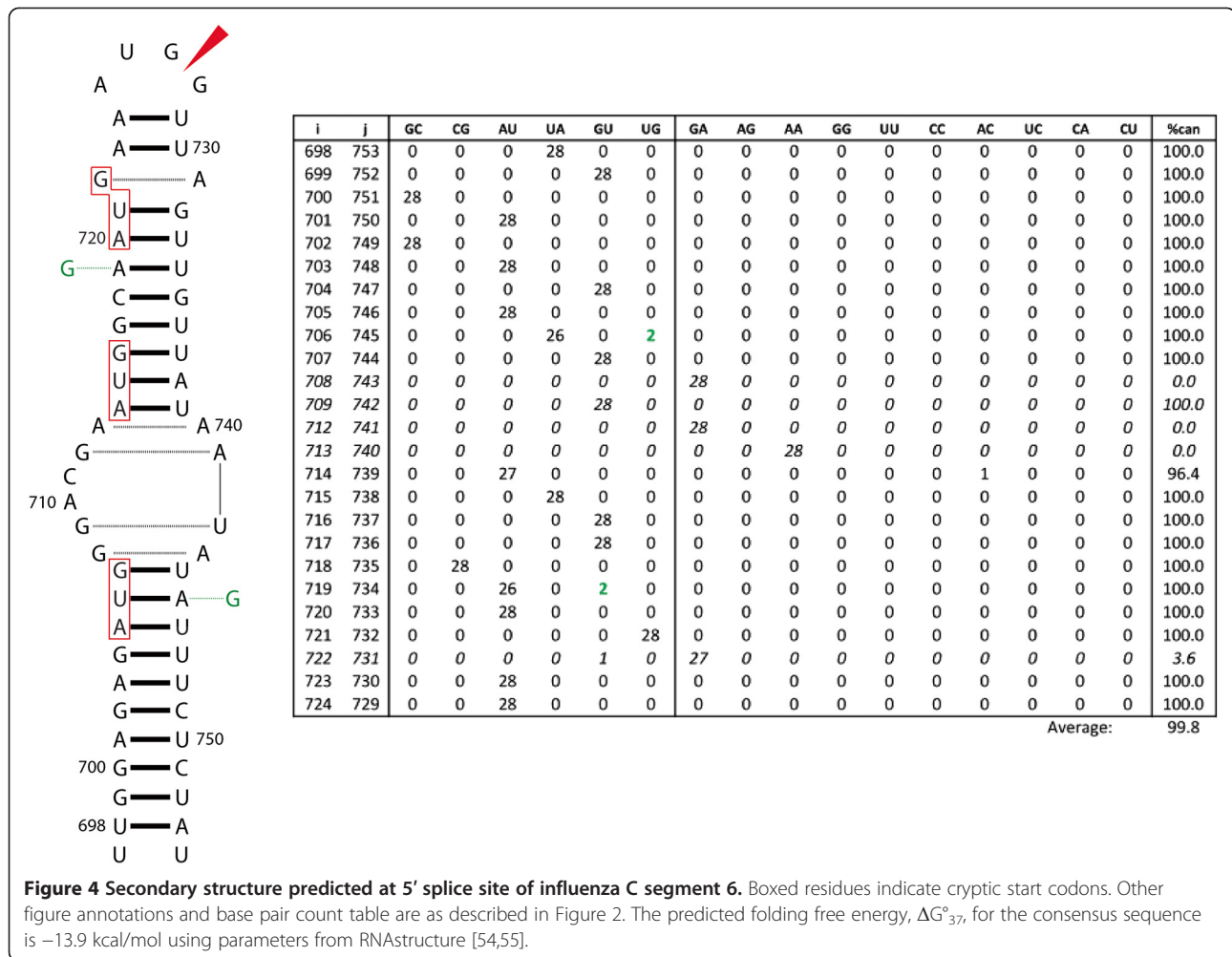
i	j	GC	CG	AU	UA	GU	UG	GA	AG	AA	GG	UU	CC	AC	UC	CA	CU	%can
698	753	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	100.0
699	752	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	100.0
700	751	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
701	750	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
702	749	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
703	748	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
704	747	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	100.0
705	746	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
706	745	0	0	0	26	0	2	0	0	0	0	0	0	0	0	0	0	100.0
707	744	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	100.0
708	743	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0.0
709	742	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	100.0
712	741	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0.0
713	740	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0.0
714	739	0	0	27	0	0	0	0	0	0	0	0	0	1	0	0	0	96.4
715	738	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	100.0
716	737	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	100.0
717	736	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	100.0
718	735	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
719	734	0	0	26	0	2	0	0	0	0	0	0	0	0	0	0	0	100.0
720	733	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
721	732	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	100.0
722	731	0	0	0	0	1	0	27	0	0	0	0	0	0	0	0	0	3.6
723	730	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
724	729	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0

Average: 99.8

Figure 4 Secondary structure predicted at 5' splice site of influenza C segment 6. Boxed residues indicate cryptic start codons. Other figure annotations and base pair count table are as described in Figure 2. The predicted folding free energy, $\Delta G°_{37}$, for the consensus sequence is −13.9 kcal/mol using parameters from RNAstructure [54,55].

purine internal loop, similar to the 5′ splice sites. In the pseudoknot model proposed by Gultyaev et al. [43,44], the 3′ splice site is situated in the loop that spans both helixes. The lower stem can be extended by 4 additional base pairs, which would sequester the splice site and also place it near the 1x1 homo-purine loop (Figure 5). The basal stem encompassing the 3′ splice site is common to both the hairpin and extended pseudoknot models, thus transitioning between the two folds would require a modest structural rearrangement. In segment 7 of influenza C, a 36 nt region encompassing the 3′ splice site is also predicted to fold into a hairpin or a pseudoknot (Figure 6). In the hairpin, the splice site is located in the apical loop and in the pseudoknot it is located in the 3 nt loop joining the two helices.

Three of the 3′ structures in influenza B (Figure 5) and C (Figure 6) are 100% conserved, and the pseudoknot in influenza B is 99.9% conserved. The extended pseudoknot and hairpin folds in segment 8 in B have a frequent consistent mutation, G694-C729 to GU, in the base pair bordering the splice site. The pseudoknotted

fold has another single mutation to a non-canonical pair: U684-A716 to a CA pair.

The 3′ splice sites of segments 7 and 8 of influenza A can fold into a hairpin or pseudoknot in a manner similar to that predicted for influenza B and C [37,43,44,98,99]. For segment 7 of influenza A (alternatively spliced to produce M2), native gel analysis showed that the 3′ splice site could form an equilibrium between the pseudoknot and the hairpin [98]. Chemical and enzymatic mapping, as well as oligonucleotide binding support both pseudoknot and hairpin conformations [98]. In the pseudoknot conformation, the 3′ splice site is sequestered in a helix and in the hairpin conformation the splice site is exposed in a 2x2 nucleotide internal loop. The pseudoknot/hairpin conformations predicted at the 3′ splice site of segment 8 were experimentally verified by Gultyaev et al. [43]. These conformational switches are proposed to regulate splicing. A similar switch may regulate splicing of segments 8 and 7 of influenza B and C, respectively.

Influenza B segment 6 and influenza C segment 6 encode ion channel proteins NB and CM2, respectively,
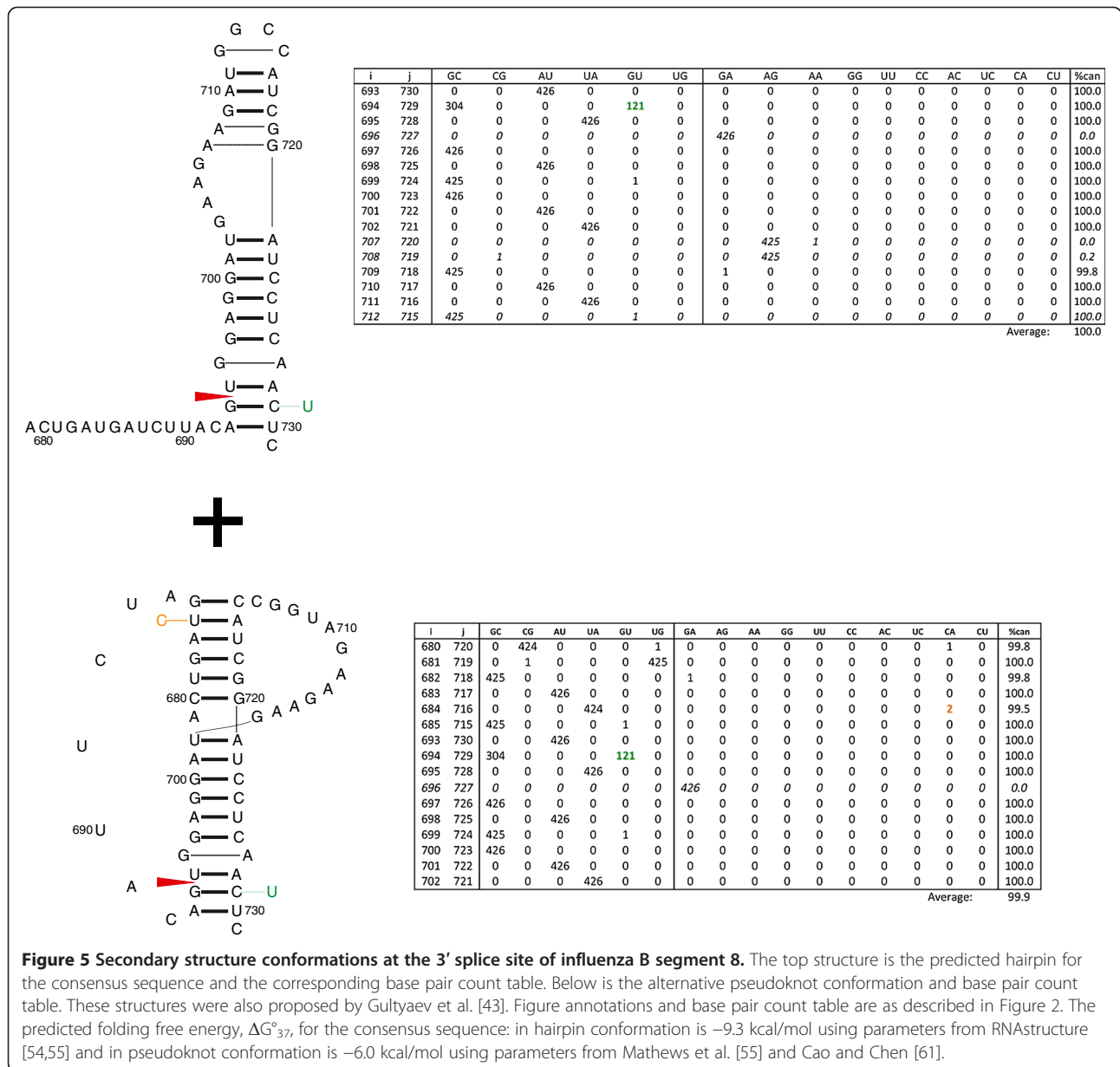
i	j	GC	CG	AU	UA	GU	UG	GA	AG	AA	GG	UU	CC	AC	UC	CA	CU	%can
693	730	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
694	729	304	0	0	0	121	0	0	0	0	0	0	0	0	0	0	0	100.0
695	728	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	100.0
696	727	0	0	0	0	0	0	426	0	0	0	0	0	0	0	0	0	0.0
697	726	426	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
698	725	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
699	724	425	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	100.0
700	723	426	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
701	722	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
702	721	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	100.0
707	720	0	0	0	0	0	0	0	425	1	0	0	0	0	0	0	0	0.0
708	719	0	1	0	0	0	0	0	425	0	0	0	0	0	0	0	0	0.2
709	718	425	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	99.8
710	717	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
711	716	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	100.0
712	715	425	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	100.0
																Average:		100.0

i	j	GC	CG	AU	UA	GU	UG	GA	AG	AA	GG	UU	CC	AC	UC	CA	CU	%can
680	720	0	424	0	0	0	1	0	0	0	0	0	0	0	0	1	0	99.8
681	719	0	1	0	0	0	425	0	0	0	0	0	0	0	0	0	0	100.0
682	718	425	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	99.8
683	717	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
684	716	0	0	0	424	0	0	0	0	0	0	0	0	0	0	2	0	99.5
685	715	425	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	100.0
693	730	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
694	729	304	0	0	0	121	0	0	0	0	0	0	0	0	0	0	0	100.0
695	728	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	100.0
696	727	0	0	0	0	0	0	426	0	0	0	0	0	0	0	0	0	0.0
697	726	426	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
698	725	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
699	724	425	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	100.0
700	723	426	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
701	722	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
702	721	0	0	0	426	0	0	0	0	0	0	0	0	0	0	0	0	100.0
																Average:		99.9

Figure 5 Secondary structure conformations at the 3′ splice site of influenza B segment 8. The top structure is the predicted hairpin for the consensus sequence and the corresponding base pair count table. Below is the alternative pseudoknot conformation and base pair count table. These structures were also proposed by Gultyaev et al. [43]. Figure annotations and base pair count table are as described in Figure 2. The predicted folding free energy, $\Delta G°_{37}$, for the consensus sequence: in hairpin conformation is −9.3 kcal/mol using parameters from RNAstructure [54,55] and in pseudoknot conformation is −6.0 kcal/mol using parameters from Mathews et al. [55] and Cao and Chen [61].

but are not predicted to form a pseudoknot/hairpin switch. Unlike M2 in influenza A, NB and CM2 are not produced from mRNA alternative splicing and thus, would not be expected to maintain this structural switch that has apparent importance for splicing regulation.

Conclusions

This study predicts regions of conserved secondary structure in the coding regions of influenza B and C (+) RNA, which allows comparisons to be made with RNA structures in influenza A. In influenza B and C, regions of high thermodynamic stability and/or base pair conservation are found at splice sites. Similarly, influenza A also has conserved structure at or near splice sites

[37,43,45]. In the alternatively spliced influenza A, B, and C RNA segments, structure is predicted at or near the 5′ splice site. In contrast to influenza A, however, the 5′ splice sites in influenza B and C are predicted to be part of hairpins. Four of five 3′ splice sites are predicted to have a pseudoknot/hairpin structural switch. The exception is segment 6 in influenza C. This segment differs from other spliced segments of influenza: it splices to form a UGA stop codon at the splice junction.

Similar to segment 8 of influenza B and segment 7 of influenza C, structure is proposed to occur at the 5′ splice site of influenza C segment 6 (Figure 4). This structure, in addition to containing the 5′ splice site, also buries cryptic start codons in its strong secondary structure.

i	j	GC	CG	AU	UA	GU	UG	GA	AG	AA	GG	UU	CC	AC	UC	CA	CU	%can
487	512	0	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	100.0
488	511	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
489	510	0	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	100.0
490	509	0	0	0	75	0	1	0	0	0	0	0	0	0	0	0	0	100.0
491	508	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
492	506	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
493	505	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
494	504	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
495	503	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
496	502	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
																Average:		100.0

i	j	GC	CG	AU	UA	GU	UG	GA	AG	AA	GG	UU	CC	AC	UC	CA	CU	%can
480	499	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
481	498	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
482	497	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
483	496	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
488	511	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
489	510	0	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	100.0
490	509	0	0	0	75	0	1	0	0	0	0	0	0	0	0	0	0	100.0
491	508	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
492	506	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
493	505	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
494	504	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	100.0
495	503	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	100.0
																Average:		100.0

Figure 6 Predicted secondary structure conformations at the 3' splice site of influenza C segment 7. The top structure is the predicted hairpin for the consensus sequence and the corresponding base pair count table. Below is the DotKnot [59] predicted, alternative pseudoknot conformation and base pair count table. Figure annotations and base pair count table are as described in Figure 2. The predicted folding free energy, ΔG°_{37}, for the consensus sequence: in hairpin conformation is −3.8 kcal/mol using parameters from RNAstructure [54,55] and in pseudoknot conformation is −9.9 kcal/mol from DotKnot [59].

This structural model provides a possible mechanism by which these cryptic start codons are suppressed.

RNA secondary structure is known to play an important role in regulating splicing by hiding or revealing splice sites and protein binding sites, or by changing the distance between regulatory elements [70]. Splicing can also be regulated via protein-induced RNA conformational switching [100,101] or small molecule binding [102,103]. Previous studies have postulated roles for RNA secondary structure in the regulation of splicing in influenza virus [104,105]. The predicted structures in Figures 2, 3, 4, 5 and 6 provide further evidence for the importance of RNA structure in influenza splicing. These results suggest that these RNA structures may be attractive targets for therapeutics as the targeting of RNA splicing with drugs is a growing area of research [106]. Knowing the structure/function relationships of influenza RNAs may be useful in designing therapeutics that specifically target these structures: with small molecules [107-111], oligonucleotides [112-114], or aptamers [115], for example.

References
1. Thompson WW, Shay DK, Weintraub E, Brammer L, Bridges CB, Cox NJ, Fukuda K: Infleunza-associated hospitilizations in the United States. *JAMA* 2004, **292**:1333–1340.
2. Thompson MG, Shay DK, Zhou H, Bridges CB, Cheng PY, Burns E, Bresee JS, Cox NJ: Centers for Disease Control and Prevention. Estimates of deaths associated with seasonal influenza - United States, 1976–2007. *MMWR* 2010, **59**:1057–1062.
3. Bouvier NM, Palese P: The biology of influenza viruses. *Vaccine* 2008, **26**:D49–D53.
4. McCauley JW, Hongo S, Kaverin NV, Kochs G, Lamb RA, Matrosovich MN, Perez DR, Palese P, Presti RM, Rimstad E: In *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. 1st edition. Edited by King AMQ, Lefkowitz E, Adams MJ, Carstens EB. San Diego: Elsevier Academic Press; 2012:749–761.
5. Ito T, Kawaoka Y: Host-range barrier of influenza A viruses. *Vet Microbiol* 2000, **74**(1–2):71–75.
6. Lamb RA: **Influenza.** In *Encyclopedia of Virology*. 3rd edition. Edited by Mahy BWJ, van Regenmortel MHV. Academic Press; 2008:95–104.
7. Hay AJ, Gregory V, Douglas AR, Lin YP: The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci* 2001, **356**(1416):1861–1870.
8. Shaw ML, Palese P: *Orthomyxoviruses: Molecular Biology*. In: Encyclopedia of Virology. Third Edition edn; 2008:483–489.
9. Rota PA, Wallis TR, Harmon MW, Rota JS, Kendal AP, Nerome K: Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. *Virology* 1990, **175**(1):59–68.
10. Lee BY, Bartsch SM, Willig AM: The economic value of a quadrivalent versus trivalent influenza vaccine. *Vaccine* 2012, **30**(52):7443–7446.
11. Ambrose CS, Levin MJ: The rationale for quadrivalent influenza vaccines. *Human Vaccines & Immunotherapeutics* 2012, **8**(1):81–88.
12. Kieft JS: Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci* 2008, **33**(6):274–283.
13. Martinez-Salas E, Ramos R, Lafuente E, Lopez de Quinto S: Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *J Gen Virol* 2001, **82**:973–984.
14. Clyde K, Harris E: RNA secondary structure in the coding region of dengue virus type 2 directs translation start codon selection and is required for viral replication. *J Virol* 2006, **80**(5):2170–2182.
15. Clever J, Sassetti C, Parslow TG: RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol* 1995, **69**:2101–2109.
16. Linnstaedt SD, Kasprzak WK, Shapiro BA, Casey JL: The role of a metastable RNA secondary structure in hepatitis delta virus genotype III RNA editing. *RNA* 2006, **12**(8):1521–1533.
17. Abbink TE, Berkhout B: RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J Virol* 2008, **82**(6):3090–3098.
18. Cabello-Villegas J, Giles KE, Soto AM, Yu P, Mougin A, Beemon KL, Wang YX: Solution structure of the pseudo-5' splice site of a retroviral splicing suppressor. *RNA* 2004, **10**(9):1388–1398.
19. Zychlinski D, Erkelenz S, Melhorn V, Baum C, Schaal H, Bohne J: Limited complementarity between U1 snRNA and a retroviral 5' splice site permits its attenuation via RNA secondary structure. *Nucleic Acids Res* 2009, **37**(22):7429–7440.
20. Brierley I, Gilbert RJ, Pennell S: RNA pseudoknots and the regulation of protein synthesis. *Biochem Soc Trans* 2008, **36**(Pt 4):684–689.
21. Liu B, Mathews DH, Turner DH: RNA pseudoknots: folding and finding. *F1000 Biology Reports* 2010, **2**:8.
22. Staple DW, Butcher SE: Pseudoknots: RNA structures with diverse functions. *PLoS Biol* 2005, **3**(6):e213.
23. Brierley I, Pennell S, Gilbert RJ: Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Microbiol* 2007, **5**(8):598–610.
24. Giedroc DP, Cornish PV: Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res* 2009, **139**(2):193–208.
25. Desselberger U, Racaniello VR, Zazra JJ, Palese P: The 3' and 5'-terminal sequences of influenza A, B, and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene* 1980, **8**:315–328.
26. Flick R, Neumann G, Hoffman E, Neumeier E, Hobom G: Promoter elements in the influenza vRNA terminal structure. *RNA* 1996, **2**:1046–1057.
27. Noble E, Mathews DH, Chen JL, Turner DH, Takimoto T, Kim B: Biophysical analysis of influenza A virus RNA promoter at physiological temperatures. *J Biol Chem* 2011, **286**(26):22965–22970.
28. Cheong H, Cheong C, Choi B: Secondary structure of the panhandle RNA of influenza virus A studied by NMR spectroscopy. *Nucleic Acids Res* 1996, **24**:4197–4201.
29. Brownlee GG, Sharps JL: The RNA polymerase of influenza A virus is stabilized by interaction with its viral RNA promoter. *J Virol* 2002, **76**(14):7103–7113.
30. Crow M, Deng T, Addley M, Brownlee GG: Mutational analysis of the influenza virus cRNA promoter and identification of nucleotides critical for replication. *J Virol* 2004, **78**(12):6263–6270.
31. Mathews DH, Moss WN, Turner DH: Folding and finding RNA secondary structure. *Cold Spring Harb Perspect Biol* 2010, **2**(12):a003665.
32. Schroeder S: Advances in RNA structure prediction from sequence: new tools for generating hypotheses about viral RNA structure-function relationships. *J Virol* 2009, **83**:6326–6334.
33. Gorodkin J, Hofacker IL, Torarinsson E, Yao Z, Havgaard JH, Ruzzo WL: De novo prediction of structured RNAs from genomic sequences. *Trends Biotechnol* 2010, **28**(1):9–19.
34. Pedersen JS, Meyer IM, Forsberg R, Simmonds P, Hein J: A comparative method for finding and folding RNA secondary structures within protein-coding regions. *Nucleic Acids Res* 2004, **32**(16):4925–4936.
35. Lange SJ, Maticzka D, Mohl M, Gagnon JN, Brown CM, Backofen R: Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res* 2012, **40**(12):5215–5226.
36. Findeiss S, Engelhardt J, Prohaska SJ, Stadler PF: Protein-coding structured RNAs: A computational survey of conserved RNA secondary structures overlapping coding regions in drosophilids. *Biochimie* 2011, **93**(11):2019–2023.
37. Moss WN, Priore SF, Turner DH: Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA* 2011, **17**(6):991–1011.
38. Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF: Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* 2005, **23**(11):1383–1390.
39. Washietl S, Hofacker IL, Stadler PF: Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci USA* 2005, **102**(7):2454–2459.
40. Gruber AR, Findeiss S, Washietl S, Hofacker IL, Stadler PF: RNAz 2.0: improved noncoding RNA detection. *Pac Symp Biocomput* 2010, **15**:69–79.
41. Simmonds P, Smith DB: Structural constraints on RNA virus evolution. *J Virol* 1999, **73**(7):5787–5794.
42. Tuplin A, Evans DJ, Simmonds P: Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J Gen Virol* 2004, **85**(Pt 10):3037–3047.
43. Gultyaev AP, Heus HA, Olsthoorn RC: An RNA conformational shift in recent H5N1 influenza A viruses. *Bioinformatics* 2007, **23**(3):272–276.
44. Gultyaev AP, Olsthoorn RC: A family of non-classical pseudoknots in influenza A and B viruses. *RNA Biol* 2010, **7**(2):125–129.
45. Ilyinskii PO, Schmidt T, Lukashev D, Meriin AB, Thoidis G, Frishman D, Shneider AM: Importance of mRNA secondary structural elements for the expression of influenza virus genes. *OMICS* 2009, **13**(5):421–430.
46. 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**(2):596–601.
47. Gouy M, Guindon S, Gascuel O: SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010, **27**(2):221–224.
48. Galtier N, Gouy M, Gautier C: SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 1996, **12**(6):543–548.
49. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al: Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, **23**(21):2947–2948.

50. Katoh K, Kuma K, Toh H, Miyata T: MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 2005, 33(2):511–518.

51. Katoh K, Misawa K, Kuma K, Miyata T: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002, 30(14):3059–3066.

52. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF: RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 2008, 9:474.

53. 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(5):911–940.

54. Reuter JS, Mathews DH: RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 2010, 11:129.

55. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH: Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A 2004, 101(19):7287–7292.

56. Mathews DH: Revolutions in RNA secondary structure prediction. J Mol Biol 2006, 359(3):526–532.

57. Mathews DH: Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. RNA 2004, 10(8):1178–1190.

58. Parisien M, Major F: The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 2008, 452(7183):51–55.

59. Sperschneider J, Datta A: DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. Nucleic Acids Res 2010, 38(7):e103.

60. Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH: Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry 1998, 37(42):14719–14735.

61. Cao S, Chen SJ: Predicting RNA pseudoknot folding thermodynamics. Nucleic Acids Res 2006, 34(9):2634–2652.

62. Cao S, Chen SJ: Predicting structures and stabilities for H-type pseudoknots with interhelix loops. RNA 2009, 15(4):696–706.

63. Buonagurio DA, Nakada S, Fitch WM, Palese P: Epidemiology of influenza C virus in man: multiple evolutionary lineages and low rate of change. Virology 1986, 153(1):12–21.

64. Nobusawa E, Sato K: Comparison of the mutation rates of human influenza A and B viruses. J Virol 2006, 80(7):3675–3678.

65. Yamashita M, Krystal M, Fitch WM, Palese P: Influenza B virus evolution: co-circulating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. Virology 1988, 163(1):112–122.

66. Air GM, Gibbs AJ, Laver WG, Webster RG: Evolutionary changes in influenza B are not primarily governed by antibody selection. Proc Natl Acad Sci U S A 1990, 87(10):3884–3888.

67. Suzuki Y, Nei M: Origin and evolution of influenza virus hemagglutinin genes. Mol Biol Evol 2002, 19(4):501–509.

68. Paragas J, Talon J, O'Neill RE, Anderson DK, Garcia-Sastre A, Palese P: Influenza B and C virus NEP (NS2) proteins possess nuclear export activities. J Virol 2001, 75(16):7375–7383.

69. Robb NC, Smith M, Vreede FT, Fodor E: NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. J Gen Virol 2009, 90(Pt 6):1398–1407.

70. Warf MB, Berglund JA: Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem Sci 2010, 35(3):169–178.

71. Blanchette M, Chabot B: A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. RNA 1997, 3(4):405–419.

72. Singh NN, Singh RN, Androphy EJ: Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. Nucleic Acids Res 2007, 35(2):371–389.

73. Coleman TP, Roesser JR: RNA secondary structure: an important cis-element in rat calcitonin/CGRP pre-messenger RNA splicing. Biochemistry 1998, 37(45):15941–15950.

74. Hermann T, Westhof E: Non-Watson-Crick base pairs in RNA-protein recognition. Chem Biol 1999, 6(12):R335–R343.

75. Giver L, Bartel DP, Zapp ML, Green MR, Ellington AD: Selection and design of high-affinity RNA ligands for HIV-1 Rev. Gene 1993, 137(1):19–24.

76. Jang SB, Hung LW, Chi YI, Holbrook EL, Carter RJ, Holbrook SR: Structure of an RNA internal loop consisting of tandem C-A + base pairs. Biochemistry 1998, 37(34):11726–11731.

77. Wilcox JL, Bevilacqua PC: A Simple Fluorescence Method for pK(a) Determination in RNA and DNA Reveals Highly Shifted pK(a)'s. J Am Chem Soc 2013, 135(20):7390–7393.

78. Chen G, Kennedy SD, Turner DH: A CA(+) pair adjacent to a sheared GA or AA pair stabilizes size-symmetric RNA internal loops. Biochemistry 2009, 48(24):5738–5752.

79. Bink HH, Hellendoorn K, van der Meulen J, Pleij CW: Protonation of non-Watson-Crick base pairs and encapsidation of turnip yellow mosaic virus RNA. Proc Natl Acad Sci USA 2002, 99(21):13465–13470.

80. Priore SF, Kierzek E, Kierzek R, Baman JR, Moss WN, Dela-Moss LI, Turner DH: Secondary structure of a conserved domain in the intron of influenza A NS1 mRNA. PloS ONE 2013, 8(9):e70615.

81. Yamashita M, Krystal M, Palese P: Evidence that the matrix protein of influenza C virus is coded for by a spliced mRNA. J Virol 1988, 62(9):3348–3355.

82. Sidrauski C, Walter P: The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell 1997, 90(6):1031–1039.

83. Muro AF, Caputi M, Pariyarath R, Pagani F, Buratti E, Baralle FE: Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. Mol Cell Biol 1999, 19(4):2657–2671.

84. Buratti E, Muro AF, Giombi M, Gherbassi D, Iaconcig A, Baralle FE: RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. Mol Cell Biol 2004, 24(3):1387–1400.

85. Hiller M, Zhang Z, Backofen R, Stamm S: Pre-mRNA secondary structures influence exon recognition. PLoS Genet 2007, 3(11):e204.

86. Stewart SM, Pekosz A: The influenza C virus CM2 protein can alter intracellular pH, and its transmembrane domain can substitute for that of the influenza A virus M2 protein and support infectious virus production. J Virol 2012, 86(2):1277–1281.

87. Furukawa T, Muraki Y, Noda T, Takashita E, Sho R, Sugawara K, Matsuzaki Y, Shimotai Y, Hongo S: Role of the CM2 protein in the influenza C virus replication cycle. J Virol 2011, 85(3):1322–1329.

88. Hongo S, Sugawara K, Muraki Y, Kitame F, Nakamura K: Characterization of a second protein (CM2) encoded by RNA segment 6 of influenza C virus. J Virol 1997, 71(4):2786–2792.

89. Pekosz A, Lamb RA: The CM2 protein of influenza C virus is an oligomeric integral membrane glycoprotein structurally analogous to influenza A virus M2 and influenza B virus NB proteins. Virology 1997, 237(2):439–451.

90. Hongo S, Sugawara K, Nishimura H, Muraki Y, Kitame F, Nakamura K: Identification of a second protein encoded by influenza C virus RNA segment 6. J Gen Virol 1994, 75(Pt 12):3503–3510.

91. Pekosz A, Lamb RA: Influenza C virus CM2 integral membrane glycoprotein is produced from a polypeptide precursor by cleavage of an internal signal sequence. Proc Natl Acad Sci USA 1998, 95(22):13233–13238.

92. Hongo S, Sugawara K, Muraki Y, Matsuzaki Y, Takashita E, Kitame F, Nakamura K: Influenza C virus CM2 protein is produced from a 374-amino-acid protein (P42) by signal peptidase cleavage. J Virol 1999, 73(1):46–50.

93. Slobodskaya OR, Gmyl AP, Maslova SV, Tolskaya EA, Viktorova EG, Agol VI: Poliovirus neurovirulence correlates with the presence of a cryptic AUG upstream of the initiator codon. Virology 1996, 221(1):141–150.

94. Verma B, Ponnuswamy A, Gnanasundram SV, Das S: Cryptic AUG is important for 48S ribosomal assembly during internal initiation of translation of coxsackievirus B3 RNA. J Gen Virol 2011, 92(Pt 10):2310–2319.

95. Pavlakis GN, Lockard RE, Vamvakopoulos N, Rieser L, RajBhandary UL, Vournakis JN: Secondary structure of mouse and rabbit alpha- and beta-globin mRNAs: differential accessibility of alpha and beta initiator AUG codons towards nucleases. Cell 1980, 19(1):91–102.

96. Kozak M: Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 2005, 361:13–37.

97. Kudla G, Murray AW, Tollervey D, Plotkin JB: Coding-sequence determinants of gene expression in Escherichia coli. Science 2009, 324(5924):255–258.

98. 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**(6):e38323.

99. 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**(11):1305–1310.

100. Warf MB, Diegel JV, von Hippel PH, Berglund JA: **The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing.** *Proc Natl Acad Sci USA* 2009, **106**(23):9203–9208.

101. Honig A, Auboeuf D, Parker MM, O'Malley BW, Berget SM: **Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72.** *Mol Cell Biol* 2002, **22**(16):5698–5707.

102. Cheah MT, Wachter A, Sudarsan N, Breaker RR: **Control of alternative RNA splicing and gene expression by eukaryotic riboswitches.** *Nature* 2007, **447**(7143):497–500.

103. Winkler W, Nahvi A, Breaker RR: **Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.** *Nature* 2002, **419**(6910):952–956.

104. Nemeroff ME, Utans U, Kramer A, Krug RM: **Identification of cis-acting intron and exon regions in influenza virus NS1 mRNA that inhibit splicing and cause the formation of aberrantly sedimenting presplicing complexes.** *Mol Cell Biol* 1992, **12**(3):962–970.

105. Plotch SJ, Krug RM: **In vitro splicing of influenza viral NS1 mRNA and NS1-beta-globin chimeras: possible mechanisms for the control of viral mRNA splicing.** *Proc Natl Acad Sci U S A* 1986, **83**(15):5444–5448.

106. Havens MA, Duelli DM, Hastings ML: **Targeting RNA splicing for disease therapy.** *Wiley Interdiscip Rev RNA* 2013.

107. Mei HY, Cui M, Heldsinger A, Lemrow SM, Loo JA, Sannes-Lowery KA, Sharmeen L, Czarnik AW: **Inhibitors of protein-RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules.** *Biochemistry* 1998, **37**(40):14204–14212.

108. Sucheck SJ, Wong CH: **RNA as a target for small molecules.** *Curr Opin Chem Biol* 2000, **4**(6):678–686.

109. Wilson WD, Li K: **Targeting RNA with small molecules.** *Curr Med Chem* 2000, **7**(1):73–98.

110. Disney MD, Labuda LP, Paul DJ, Poplawski SG, Pushechnikov A, Tran T, Velagapudi SP, Wu M, Childs-Disney JL: **Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners.** *J Am Chem Soc* 2008, **130**(33):11185–11194.

111. Labuda LP, Pushechnikov A, Disney MD: **Small molecule microarrays of RNA-focused peptoids help identify inhibitors of a pathogenic group I intron.** *ACS Chem Biol* 2009, **4**(4):299–307.

112. Childs JL, Disney MD, Turner DH: **Oligonucleotide directed misfolding of RNA inhibits Candida albicans group I intron splicing.** *Proc Natl Acad Sci USA* 2002, **99**(17):11091–11096.

113. Disney MD, Childs JL, Turner DH: **New approaches to targeting RNA with oligonucleotides: inhibition of group I intron self-splicing.** *Biopolymers* 2004, **73**(1):151–161.

114. Kierzek E: **Binding of short oligonucleotides to RNA: studies of the binding of common RNA structural motifs to isoenergetic microarrays.** *Biochemistry* 2009, **48**(48):11344–11356.

115. Watrin M, Dausse E, Lebars I, Rayner B, Bugaut A, Toulme JJ: **Aptamers targeting RNA molecules.** *Methods Mol Biol* 2009, **535**:79–105.