

# Secondary Structure of a Conserved Domain in the Intron of Influenza A NS1 mRNA

Salvatore F. Priore<sup>1</sup>, Elzbieta Kierzek<sup>2</sup>, Ryszard Kierzek<sup>2</sup>, Jayson R. Baman<sup>1</sup>, Walter N. Moss<sup>1</sup>, Lumbini I. Dela-Moss<sup>1</sup>, Douglas H. Turner<sup>1\*</sup>

**1** Department of Chemistry and Center for RNA Biology, University of Rochester, Rochester, New York, United States of America, **2** Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

## Abstract

Influenza A virus is a segmented single-stranded (–)RNA virus that causes substantial annual morbidity and mortality. The transcriptome of influenza A is predicted to have extensive RNA secondary structure. The smallest genome segment, segment 8, encodes two proteins, NS1 and NEP, via alternative splicing. A conserved RNA domain in the intron of segment 8 may be important for regulating production of NS1. Two different multi-branch loop structures have been proposed for this region. A combination of *in vitro* chemical mapping and isoenergetic microarray techniques demonstrate that the consensus sequence for this region folds into a hairpin conformation. These results provide an alternative folding for this region and a foundation for designing experiments to probe its functional role in the influenza life cycle.

**Citation:** Priore SF, Kierzek E, Kierzek R, Baman JR, Moss WN, et al. (2013) Secondary Structure of a Conserved Domain in the Intron of Influenza A NS1 mRNA. PLoS ONE 8(9): e70615. doi:10.1371/journal.pone.0070615

**Editor:** Nicole M. Bouvier, Mount Sinai School of Medicine, United States of America

**Received:** January 28, 2013; **Accepted:** June 22, 2013; **Published:** September 2, 2013

**Copyright:** © 2013 Priore et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was primarily funded by NIH grant R01 GM22939. SFP is a trainee in the Medical Scientist Training Program funded by NIH T32 GM07356. EK was funded by NIH grant 1R03TW008739-01 and Polish National Science Center grant N N301 788 440. RK was funded by Polish National Science Center grants UMO-2011/03/B/NZ1/00576 and UMO-2011/03/B/ST5/01098. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: turner@chem.rochester.edu

## Introduction

Influenza A is a segmented, negative-sense RNA virus that causes substantial morbidity and mortality on an annual basis. It is estimated that approximately 25,000 deaths result from influenza infections each year in the United States alone [1]. Influenza A also has the potential to form pandemic strains, which can have devastating public health implications. Because influenza virus uses RNA exclusively throughout its life-cycle, it may contain functionally important RNA secondary structures as observed in other RNA viruses [2–8]. A bioinformatics study revealed widespread predicted structure throughout the coding regions of influenza A, predominantly in the plus-sense (protein coding) orientation [9]. Folding stability is also predicted to be host-specific [10,11]. Thus, secondary structure could also be an important consideration for host range.

The smallest segment of influenza A, segment 8 (NS1/NEP), codes for two proteins via alternative splicing (Figure 1) and is predicted to have wide-spread RNA structure [10]. In a study by Ilyinskii, et al. [12], a set of mutations disrupting the predicted RNA secondary structure of the intronic nucleotides 81–148 (numbering is from the start of the NS1 open reading frame) were linked to the down-regulation of NS1 protein. Despite this important result, no experimental RNA structure determination has been reported for this region. Ilyinskii, et al. proposed a 3-way multi-branch loop using an RNA secondary structure prediction algorithm called Mfold [12,13]. Subsequently, a slightly different multi-branch loop structure was proposed [9]. Here, a hairpin model for the consensus sequence (see Materials and Methods) of the 81–148 nucleotide region is proposed on the basis of chemical

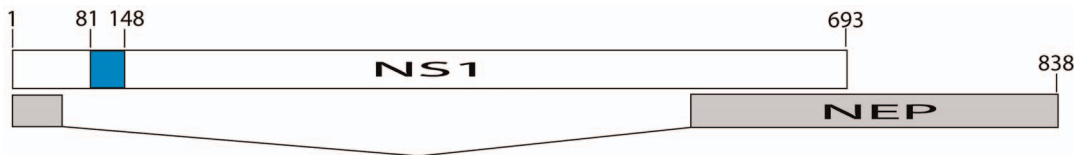
mapping experiments and minimization of free energy (MFE) RNA secondary structure predictions [14,15]. Coincidentally, the consensus sequence for this region is identical to the 1918 pandemic strain, which led to the worst influenza A outbreak in history [16].

Knowledge of RNA secondary structure informs interrogation of function by facilitating design and interpretation of mutational studies. Once function is established, potential therapeutics against RNA structure can be designed. For example, databases that correlate small molecule binding patterns to RNA structural motifs can facilitate rapid identification of lead therapeutic compounds for a given RNA structure [17]. The hairpin reported here for nucleotides 81–148 of the NS1 intron should facilitate further functional analysis of this region.

## Results

### Comparison of Structural Models

Figure 2 shows three different predictions for the 81–148 nucleotide region of NS1 mRNA. Structure A is the model from Ilyinskii, et al. [12] based on the sequence of influenza strain A/PR/8/34 (H1N1). Structure B was predicted for the consensus sequence with RNAalifold [18] as part of a genome-wide survey for influenza A RNA secondary structures [9]. Structure C is the MFE prediction by RNAstructure [14,15] for the consensus sequence. All three structures exhibit over 90% base pair conservation, largely because the stem below base pair 99/131 has high conservation in all three models (see Figure S1 for complete co-variation data). However, the region above base pair



**Figure 1. Diagram of the segment 8 (NS1/NEP) coding region.** Numbering begins at the start of the NS1 open reading frame (ORF). White and shaded bars depict NS1 and NEP ORFs, respectively. Diagonal lines indicate the segment 8 intron and the blue box represents the 81–148 nt region. doi:10.1371/journal.pone.0070615.g001

94/136 of each structure is folded quite differently with free energy of folding for sequences and structures A, B, and C predicted by RNAstructure to be  $-7.1$ ,  $-8.2$ , and  $-12.9$  kcal/mol, respectively, at  $37^{\circ}\text{C}$ . Partition function calculations [19] provide the base pairing probabilities annotated on each structure. Structure A has two changes from the consensus sequence for this region. The G $\rightarrow$ A transition at nucleotide 123 disfavors prediction of structure C because Mfold and RNAstructure assign an unfavorable free energy to CA pairs, although they do form *in vivo* and *in vitro* [20–22]. Thus, sequence A has a much higher probability than the consensus sequence for the shared helix that begins at nucleotide 102 in structures A and B (see Figure 2). Because nucleotide 123 is a G in the consensus sequence, RNAstructure favors structure C and predicts low base pair probabilities for the multi-branch loop in structure B. Thus, based on MFE predictions structure C is expected to be the favored conformation for the consensus sequence at  $37^{\circ}\text{C}$ .

### Experimental Secondary Structure Determination

To better define the folding of this region, *in vitro* mapping experiments with small molecules (DMS, NMIA, and CMCT) and isoenergetic microarrays were performed on the consensus sequence as determined from all available unique influenza A strains. Small molecules probe the structure of folded RNA by modifying the base or sugar of unpaired, or weakly paired, nucleotides. Universal isoenergetic microarrays contain 861 different modified pentamer and hexamer oligonucleotides, thus providing complementarity to most sequences. The oligonucleotides are designed to bind with roughly equivalent stability to unpaired regions of a target RNA in order to minimize the effect of sequence dependence on binding. Folded RNA is hybridized to the array and binding to a probe reveals unpaired, or weakly paired, regions of the molecule. This technique is helpful for the design of potential therapeutics [23]. In summary, strong reactivity with small molecules or strong binding to microarray probes typically indicates unstructured (non-base paired) regions of an RNA structure.

The experimental results are overlaid on structural models B and C in Figure 2. The pattern of chemical reactivity and probe accessibility clearly favors structure C. Many reactive sites occur in predicted structured regions of model B, but are clustered in the hairpin loop of structure C. Specifically in support of structure C, positions 112 and 113 are reactive, but are inside a helix in structure B (Figures 2, S2, and S3). Positions 98, 106, and 124 are also reactive, although predicted to be double-stranded in both B and C. This is not inconsistent with either structure B or C, however, as chemical reactivity at or next to a G-U pair or at the end of helix is commonly observed [14]. Microarray binding sites only occur in either a hairpin or internal loop in structure C. (Table 1, Figures 2 and S4). Analysis of hybridization results in 10 mM  $\text{MgCl}_2$ , 300 mM KCl showed that probes bind strongly to sites 114, 115, and 116 in the hairpin loop of structure C, whereas these sites are involved in base pairing in structure B. Interpre-

tation of probes complementary to sites 127 (strong binding), 128 and 91 (medium binding) is more complicated because they have potential alternative binding sites. However, sites 127 and 128 are likely accessible because this region also shows chemical mapping reactivity (Figure 2). Interestingly, the hairpin loop of structure C also had strong or medium binding of five probes not completely Watson-Crick paired (i.e. forming G-C or A-U pairs) to the target RNA (Table 1). Their binding is well predicted to region 115–118 assuming G-U pairing. Probe CUC<sup>1</sup>UC<sup>1</sup>G<sup>1</sup> is also well predicted to site 114 because the last five nucleotides of the probe are completely complementary to the target RNA. Hybridization in 1 M KCl, 10 mM  $\text{MgCl}_2$  reveals some additional binding sites and overall confirms structure C. The 1 M KCl conditions are less specific and only the new binding to sites 99 and 109 are unambiguous since there are not alternative binding sites for these probes. Higher KCl concentrations apparently make RNA more accessible in internal loop regions, perhaps by competing with  $\text{Mg}^{2+}$  binding sites.

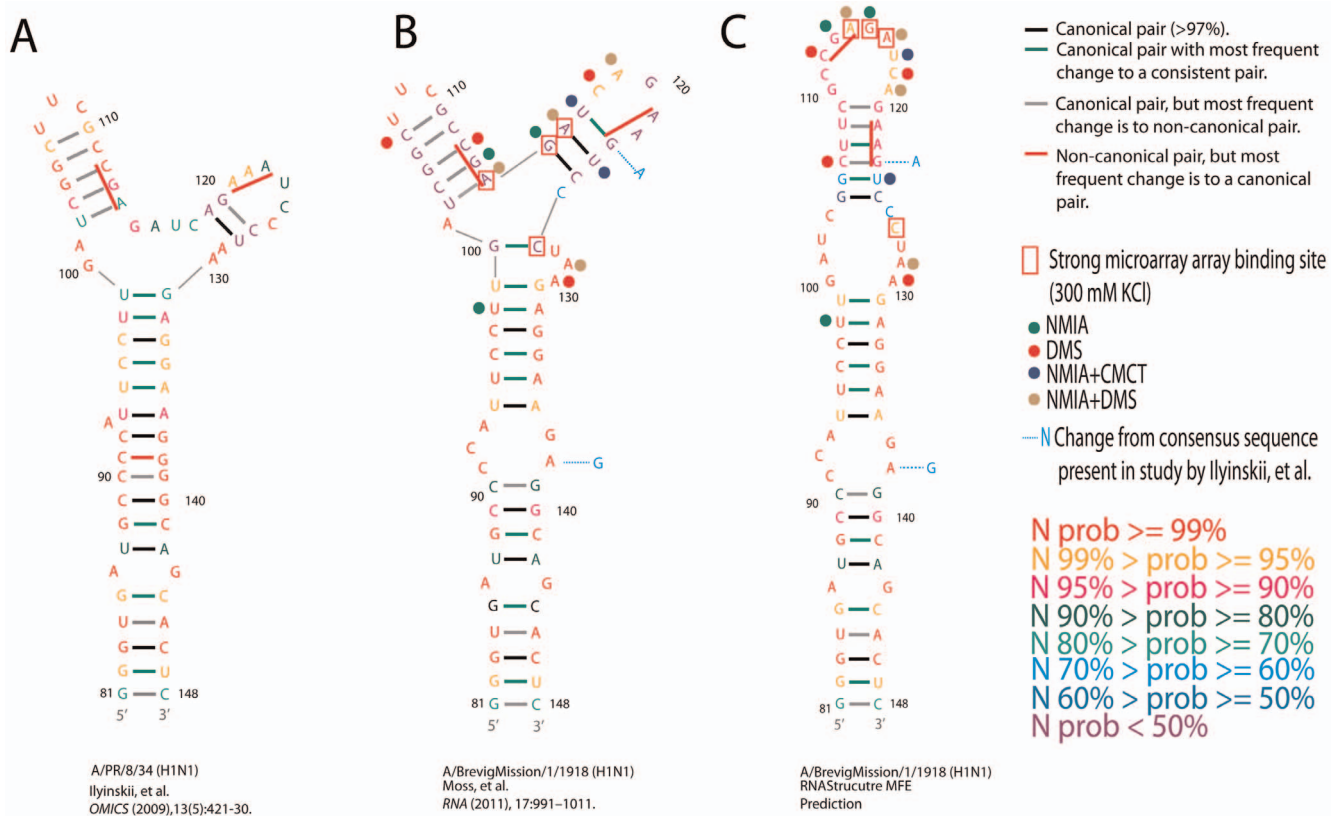
### Discussion

#### Influenza A Nucleotides 81–148 Fold into a Hairpin In Vitro

As seen in Figure 2, the chemical mapping and microarray data support structure C over structure B. In particular, the hairpin loop of structure C is very reactive to chemicals and binds strongly to three microarray probes. In contrast, the reactive nucleotides and probe binding sites are in paired regions of structure B. Evidently, the consensus sequence for this region folds into a hairpin *in vitro* and not a multi-branch loop structure as previously predicted [9]. The results of this study have two important implications: First, in order to improve RNA structure prediction algorithms, the factors that led to multiple previous structure predictions for this region should be considered. Secondly, because functional experiments conducted on this RNA secondary structure assumed a multi-branch loop, new sets of mutations, based on the consensus hairpin structural model, should be designed to further probe the cellular structure and functional role of this domain in host cells.

#### Structural Models and The Potential Importance of CA Base Pairs

Defining the functional role of RNA requires knowledge of its likely structure. Based on Mfold [12,13] and RNAalifold [9,18] predictions, nucleotides 81–148 of segment 8 were thought to fold into one of two different multi-branch loop structures (Figure 2 and S1A and S1B), one of which may be important for the production of NS1 protein [12]. Both structures are consistent with the known general topology of three-way multi-branch loops [24]. However, the consensus sequence for this region folds into a hairpin structure *in vitro*. The discrepancy between the different structural models may be related to non-canonical base pairs present in some influenza strains. As can be seen from Table C in



**Figure 2. Structural models for nucleotides 81–148 of influenza A segment 8 (NS1/NEP).** Numbering begins at the start of the NS1 open reading frame (ORF). Structural model A comes from Ilyinskii, et al. [12] and the sequence is from strain A/PR/8/34 (H1N1), B is modeled [9] with RNAalifold [18], and C is the MFE prediction from RNAstructure [14,15]. B and C have the consensus sequence for this region based on an alignment of 1017 unique sequences and can be found in many strains, including A/BrevigMission/1918 (H1N1), which caused the most deadly pandemic in history [16]. RNAstructure free energy predictions at 37°C for sequences and structures A, B, and C are  $-7.1$ ,  $-8.2$ , and  $-12.0$  kcal/mol, respectively [15]. These free energy predictions exclude the GC pair at 100/127 in structure B, because it is predicted not to form on the basis of thermodynamic parameters for secondary structure [14]. Tertiary interactions may allow this pair to form or, alternatively, a C127/G131 base pair to form closing a UAA triloop. UAA triloops are common [42,43] and can form tertiary interactions [24,44,45]. Individual nucleotides are colored based on probabilities from RNAstructure partition function calculations as shown in the key [19]. Colors of lines between nucleotides indicate type of conservation of pairing (see tables in Figure S1). Strong modification sites for NMIA, CMCT, and DMS are indicated by colored-dots next to each reactive nucleotide. Red boxes indicate the center nucleotide of strongly binding iso-energetic microarray probes in 10 mM Tris-HCl, pH 7.0, 300 mM KCl, and 10 mM  $MgCl_2$ . Probe 127 could bind strongly to site 91, but it binds stronger than probe 91, indicating that site 127 is a true binding site for probe 127. Medium binding sites are not shown, but are centered at nucleotides 91 and 128. Light blue nucleotides next to structures B and C indicate positions in which sequence A differs from the consensus sequence. Solid red bars indicate the codons that were changed to GCG by Ilyinskii, et al. and led to down-regulation of NS1 protein [12]. The chemical mapping data, particularly the strong reactivity of nucleotides 112 and 113, are consistent with structure C but not with structure B.  
 doi:10.1371/journal.pone.0070615.g002

Figure S1, CA pairs are quite common at positions 106/123, 108/121, and 109/120. Programs, like RNAalifold [18] and Multilign [25], that consider multiple sequences will disfavor structure C because formation of CA pairs is penalized in these algorithms. In the case of Multilign, a base pair is predicted only if all the sequences can form a canonical pair at that position in the developing alignment. Thus, any alignment that contains a CA pair at the positions described above will disfavor prediction of structure C. Similarly, RNAalifold folds the consensus sequence for an alignment and penalizes alignment columns that cannot form a canonical pair above a user-defined threshold. This explains why the hairpin fold (Figure 2C) was not identified in a previous study [9]. However, single-sequence prediction software, including Mfold [13], RNAstructure [15] and RNAfold [26] unanimously predict structure C as the MFE fold for the consensus sequence. In contrast, single sequence folding of the sequence in Figure 2 (panel A) [12] predicts the three-way multi-branch loop because of two variations from the consensus sequence, G123→A

creating a CA pair and A138→G changing a potential CA to CG pair (See changes marked on Figure 2B and 2C). CA pairs were found to be between 2 and 45 times more common than other non-canonical base pairs predicted in alignments throughout influenza A coding regions [9]. Until recently it was thought that CA pairs could only form at low pH because the N1 of the adenine needed to be protonated to form a hydrogen bond. However, new studies reveal that adenines within helices can have  $pK_a$ 's shifted to physiological ranges (i.e.  $pH \sim 7.0-8.0$ ) [27]. Thus, future investigations into the general occurrence of CA pairs in coding regions and integration of the sequence dependence of energetics for CA pairs may improve predictions using single and multiple sequences. For example, rather than penalizing all non-canonical pairs equally in programs using sequence comparison, it may be better to weight penalties by the individual thermodynamic stabilities of non-canonical pairs. The results for segment 8 nucleotides 81–148 illustrate the importance of comparing

**Table 1.** Binding of nucleotides 81 to 148 of influenza A NS1 mRNA to isoenergetic probes on microarray.

Binding sites <sup>a</sup>	Modified probe 5' to 3' <sup>b</sup>	Base pair with G <sup>L</sup> at 3' end	RNA binding <sup>c</sup>		$\Delta G^{\circ}_{37}$ predicted (kcal/mol) for complement and alternative binding site <sup>f</sup>	Predicted $\Delta G^{\circ}_{37}$ of modified probe/RNA for complementary binding site <sup>g</sup>
			Buffer I <sup>d</sup>	Buffer II <sup>e</sup>		
91; [127]	UGG <sup>L</sup> GG <sup>L</sup>		M		-8.2; [-6.0]	-9.72
99	U <sup>L</sup> CD <sup>L</sup> DG <sup>L</sup> G <sup>L</sup>	C-G <sup>L</sup>		<u>M</u>	-6.9	-8.68
102; [115–118]	C <sup>L</sup> GD <sup>L</sup> UC <sup>L</sup> G <sup>L</sup>	U-G <sup>L</sup>		<u>M</u>	-7.2; [-5.2]	-9.66
109	GC <sup>L</sup> GD <sup>L</sup> AG <sup>L</sup>	C-G <sup>L</sup>		<u>M</u>	-8.1	-9.87
114	UC <sup>L</sup> UC <sup>L</sup> GG <sup>L</sup>	C-G <sup>L</sup>	<u>S</u>	<u>S</u>	-8.8	-9.35
115	D <sup>L</sup> UC <sup>L</sup> UC <sup>L</sup> G <sup>L</sup>	C-G <sup>L</sup>	<u>S</u>	<u>S</u>	-6.2	-9.17
116; [100–103]	GD <sup>L</sup> UC <sup>L</sup> UG <sup>L</sup>	G-G <sup>L</sup>	<u>S</u>	M	-6.3; [-4.7]	-9.02
127; [89–92]; [96–100]	UD <sup>L</sup> GG <sup>L</sup> GG <sup>L</sup>	U-G <sup>L</sup>	S	M	-7.7; [-7.7]; [-4.7]	-10.38
128; [90–92]; [96–101]	U <sup>L</sup> UD <sup>L</sup> GG <sup>L</sup> G <sup>L</sup>	C-G <sup>L</sup>	M	M	-7.2; [-4.5]; [-4.4]	-8.65
132; 138	CC <sup>L</sup> UC <sup>L</sup> UG <sup>L</sup>	A-G <sup>L</sup>		S	-7.5; -6.8	-10.09
[114]; [131–134]	CUC <sup>L</sup> UC <sup>L</sup> G <sup>L</sup>		S	S	[-5.8]; [-4.1]	-11.51
[116]	GG <sup>L</sup> UC <sup>L</sup> UG <sup>L</sup>		S	S	[-5.7]	-12.10
[114/115]	G <sup>L</sup> UC <sup>L</sup> UC <sup>L</sup> G <sup>L</sup>		S	S	[-6.5]	-9.21
[115–118]	GG <sup>L</sup> UC <sup>L</sup> DG <sup>L</sup>		S	S	[-4.6]	-7.73
[115–118]; [99–103]	GG <sup>L</sup> UC <sup>L</sup> GG <sup>L</sup>		M	M	[-4.6]; [-4.8]	-7.73

<sup>a</sup>- Center of binding site, where center is target RNA nucleotide complementary to third nucleotide from 5'-end of probe. Square brackets indicate possible alternative binding sites (if 5 or 6 nucleotides of probe bind) or region of binding (if predicted duplex has less than 5 canonical base pairs). Data in *italics* represent probes that bind in buffer II, but not in buffer I;

<sup>b</sup>- LNA nucleotides are marked with superscript <sup>L</sup>, D represents 2,6-diaminopurine, nucleotides without superscript are 2'-O-methyl-nucleotides;

<sup>c</sup>- S -strong binding, M - medium binding; underline - binding site with no ambiguity;

<sup>d</sup>- buffer composition is 300 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM Tris-HCl pH 7.0;

<sup>e</sup>- buffer composition is 1 M KCl, 10 mM MgCl<sub>2</sub>, 10 mM Tris-HCl pH 7.0;

<sup>f</sup>- calculated by RNAstructure 4.6 program (for 1 M NaCl, assuming no structure of target RNA and unmodified probe), values correspond to binding sites listed in the first column;

<sup>g</sup>- calculated for 100 mM NaCl buffer according to published equation [37,38].

Probes listed underneath the double line have no perfect Watson-Crick match of the first five probe nucleotides to sequences in the target, but bind to target with at least one predicted GU wobble pair or in the case of site 114 have perfect complementarity to the target with only last five nucleotides of the probe. There are no complementary probes for sites: 129–131 and 137 on universal microarrays. Probes complementary to all other sites not listed in table do not bind strongly or moderately.

doi:10.1371/journal.pone.0070615.t001

predictions from different approaches and testing them experimentally.

### Determining Function for the 41–148 Nucleotide Region

Determining the function for this region will require cell culture and/or *in vivo* assays. The mutations reported by Ilyinskii, et al. [12] that led to NS1 protein down-regulation (NS1mut3841) were designed on expectations that mutation of codon CGA→GCG starting at nucleotide 112 is in a helix that would be disrupted and codon AAA(AAG in consensus) →GCG starting at 121 is in a loop. While chemical mapping shows that the consensus sequence folds into a hairpin *in vitro*, which flips the context of these two codons (Red bars in Figure 2A and 2C), the hairpin structure would also be destabilized by NS1mut3841. In fact, two mutations tested by Ilyinskii, et al. (NS1mut 3841 and NS1mut3540) are expected to perturb both the multi-branch loop and new hairpin model, but only NS1mut3841 caused down-regulation of NS1 protein [12]. Thus, the function of RNA secondary structure in this region is not clear. As previously reported [9], all the codon mutations tested by Ilyinskii, et al. [12] occur in natural sequences except for the alanine introduced by the codon change at residue 121 in NS1mut3841. The combined mutations in NS1mut3841

produce an AXXA amino acid motif that has been reported to cause destruction-box mediated protein degradation [28]. Further mutational studies should focus on making mutations that change RNA secondary structure without changing amino acid sequence and should be conducted on the consensus sequence for which this study provides a confident structural model based on experimental techniques. Such experiments should be able to separate the importance of amino acid effects from those caused by changes in RNA secondary structure. Reverse genetic techniques provide a method for readily introducing mutations into influenza virus, and function can be assayed for a variety of factors including: NS1 protein production, segment 8 splicing, and influenza replication efficiency [29].

An understanding of conserved structure of influenza mRNA can provide new insights into the life cycle of the virus and lead to new therapeutic targets for the treatment of acute influenza infections. For example, if the disruption of a structured domain by site-directed mutagenesis decreases the virulence of a virus (by slowing replication efficiency, etc.), but provides enough viral protein production to elicit an adapted immune response, then these strains may make ideal candidates for live attenuated vaccine development. Because conserved, structured RNA domains are



likely fundamental to a wide variety of influenza strains, they could be universal mutation sites for the generation of attenuated viruses.

In conclusion, on the basis of experimental results, the 81–148 region of segment 8 folds into a hairpin structure. However, it should be noted that these results do not preclude the possibility that natural mutations at positions 108, 120, and 123 could favor a reorganization of nucleotides at the top of the structure as is seen in structure A or B, or that this region exists in a structural equilibrium with multiple conformations inside the cell. Other factors present *in vivo*, such as protein binding, can also elicit structural changes. The function of this domain is unclear, but its proximity to the 5' splice site makes it a possible intronic splicing enhancer/inhibitor [30]. Its location in the segment 8 intron could also serve as a tag to distinguish un-spliced NS1 mRNA from spliced NEP mRNA. Structure C in Figure 2 provides a secondary structure model useful for designing mutation and chemical genetics experiments to test for function and eventually to design potential therapeutics.

## Materials and Methods

### Alignments and Prediction of RNA Secondary Structure

An alignment was created for nucleotides 81 to 148 of the NS1 gene (numbering does not include the UTR) by downloading all available segment 8 sequences (approximately 11,000) from the NCBI Influenza Virus Resource [31]. Sequences were aligned using the MAFFT web server [32]. The 81–148 region was excised from this alignment and the final alignment of 1017 unique sequences was generated by removing all redundant sequences and phylogenetically distinct avian Clade B sequences [33,34]. Average base pair conservation was calculated from this alignment for each of the three structural models in Figure 2. Structures A and B in Figure 2 are described in Ilyinskii et al. [12] and Moss et al. [9], respectively. Structure C is the minimum free energy structure predicted by RNAstructure [14,15]. RNAstructure was also used to calculate partition functions for the two sequences in Figure 2 [19].

### Experimental Constructs

DNA oligonucleotide templates were obtained from IDT containing a T7 RNA polymerase site, the consensus sequence for the 81 to 148 nt region, and EcoRI and HindIII restriction sites for cloning (underlined below). The sense DNA construct had the following sequence: 5'-GAATTCTAATACGACTCACTA-TAGGGTGTATGCCCAT TCCTTGATCGGCTTCGCCGA-GATCAGAAGTCCCTAAGAGGAAGAGGCAGCACTCAA-GCTT-3'. For chemical mapping studies, an additional nine consensus nucleotides, 5'-TTGGTCTGG-3', were added at the 3' end of the consensus sequence, before the HindIII site, to serve as a primer binding site for reverse transcription. The DNA oligonucleotides were annealed and ligated into a digested pUC17 vector. The resulting plasmid was cloned into DH5-alpha cells [Invitrogen]. Plasmid was isolated from cloned cells using a Qiagen mini-prep kit and the sequence was confirmed by sequencing. PCR primers were designed to amplify only the T7 RNA polymerase promoter and the consensus sequence and the resulting product was used for *in vitro* RNA production. RNA was synthesized *in vitro* using a T7 RNA polymerase kit (Ambion). RNA products were purified on a denaturing 8% PAGE gel and passively eluted overnight in 0.3 M sodium acetate. After ethanol precipitation, the RNA was re-suspended in 0.5 X TE (5 mM Tris, pH 7.5, 0.5 mM EDTA) buffer.

### Chemical Mapping

For all chemical mapping experiments, 10  $\mu$ L samples were prepared as outlined below. All final concentrations listed are for the final 10  $\mu$ L mapping samples. Ten pmol of RNA was combined with buffer (final concentration: 10 mM HEPES or potassium borate (CMCT), 100 mM KCl) with a pH of 7.5 (DMS) or 8.0 (NMIA and CMCT). Samples were heated to 95°C for 2 min and then placed on ice for 2 min. MgCl<sub>2</sub> was added to each sample for a final concentration of 10 mM and the RNA was folded at 37°C for 10 min. As assayed by native gel electrophoresis, this folding protocol produced a single conformation, which runs identically to the single conformation generated with a slow-cooling method [9]. For DMS mapping, 9  $\mu$ L samples were treated with either 1  $\mu$ L of 100% ethanol or 1  $\mu$ L of 1:30, 1:60, or 1:120 dilution of DMS (Sigma-Aldrich) in 100% ethanol for 5 min. NMIA samples were treated with 1  $\mu$ L of DMSO or 1  $\mu$ L NMIA (Invitrogen) in DMSO to a final concentration of 12, 6 or 3 mM for a total of 45 min. For CMCT, samples were treated with either 1  $\mu$ L of water or 1  $\mu$ L of 50, 25, or 12.5 mg/mL CMCT in water for 5 min. All 10  $\mu$ L reactions were carried out at 37°C. After incubation, each sample was ethanol precipitated and re-suspended in 10  $\mu$ L of 0.5 X TE buffer. Primer extension and gel analysis were carried out as described by Wilkinson et al. [35]. Nucleotide positions were identified using dideoxy sequencing ladders. Primer extension was identical to the experimental samples, but 1  $\mu$ L of A, C, G, or T ddNTP (GE Life Sciences) was added to each reaction.

### Isoenergetic Microarrays

Microarray experiments were carried out as previously described [11,23]. Isoenergetic microarrays were designed to have modified probes with similar free energies of binding to unstructured RNA [36–38]. A universal microarray with 861 different sequences, printed in triplicate, was used. The library of short, modified 2'-O-methyl pentamer and hexamer oligonucleotide probes had LNA and 2,6-diaminopurine substitutions to provide binding independent of sequence of an unstructured RNA. Internal universal negative controls were U, UUUUU, and also spotted buffer. Probes with little complementarity to the 81–148 nucleotide region of the NS1 mRNA also served as negative controls. Microarrays were printed in the European Center of Bioinformatics and Genomics in Poznan, Poland. Prior to hybridization experiments, 5'-end <sup>32</sup>P labeled RNA was folded as described above. Final folding buffer was 300 mM (or 1 M) KCl, 10 mM Tris-HCl, pH 7.0, and 10 mM MgCl<sub>2</sub>. Microarrays require higher monovalent ion concentrations than used in the chemical mapping experiments to achieve consistent binding [23,39]. Folded RNA was hybridized to the isoenergetic microarrays for 18 h at 4°C. Microarrays were washed in the same buffer solution, dried by centrifugation and exposed to phosphor-imager screen. Obtained data were analyzed using ArrayGauge version 2.1 program as previously described [40,41]. Intensity of binding was assessed relative to the strongest hybridized probe. Strong and medium binding have  $\geq 1/3$  and  $\geq 1/9$  of the strongest integrated intensity, respectively. Alternative binding sites were analyzed using RNA-RNA thermodynamics with RNAstructure and prediction of modified probe/RNA free energies [38]. The binding of target RNA to complementary probes is listed in Table 1.

### Supporting Information

**Figure S1 Co-variation analysis of structural models from Figure 2.** (A, B, and C) Base pair counts in the tables are

based on an alignment of 1017 unique sequences. Tables A, B, and C correspond to their respective structures in Figure 2. Base pair conservation and mutations are denoted by different colors on the table, which correspond to the color of the bar between each base pair on the structural models in Figure 2. Canonical pairs are to the left of the vertical line separating the UG and GA columns. A consistent change is a change from a Watson-Crick pair to a GU pair or vice versa. (D) The table at the bottom of the figure gives the amino acid sequence, composition, and percent conservation of each nucleotide position for this region based on the alignment described above.

(TIF)

**Figure S2 Denaturing 8% PAGE analysis for NMIA, DMS and CMCT mapping.** Dideoxy ladders are shown in the first four lanes. NMIA, DMS, and CMCT lanes are marked as being a mock treated control or the amount of reagent used. Strong sites of modification are labeled to the right of the gel. These labels correspond to the site of modification, which is one nucleotide longer than the adjacent cDNA fragment.

(TIF)

**Figure S3 Denaturing 8% PAGE analysis for DMS and CMCT mapping.** Dideoxy ladders are shown in the first four

lanes. DMS and CMCT lanes are marked as being a mock treated control or the amount of reagent used. Strong sites of modification are labeled to the right of the gel. These labels correspond to the site of modification, which is one nucleotide longer than the adjacent cDNA fragment.

(TIF)

**Figure S4 Hybridization results for consensus sequence nucleotides 81–148 of influenza A NS1 mRNA.** Bar graph represents relative probe binding intensities after hybridization in the following conditions: 300 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 7.0 (Blue) and 1 M KCl, 10 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 7.0 (Red) at 4°C. The average intensities are plotted on the y-axis and the center of probe binding to the target RNA is shown on the x-axis.

(TIF)

## Author Contributions

Conceived and designed the experiments: SFP EK RK JB WNM LIDM DHT. Performed the experiments: SFP EK RK. Analyzed the data: SFP EK RK DHT. Contributed reagents/materials/analysis tools: SFP EK RK JB WNM LIDM DHT. Wrote the paper: SFP EK RK JB WNM LIDM DHT.

## References

- Thompson MG, Shay DK, Zhou H, Bridges CB, Cheng PY, et al. (2010) Estimates of deaths associated with seasonal influenza — United States, 1976–2007. *MMWR Morb Mortal Wkly Rep* 59: 1057–1062.
- Weiner AM, Maizels N (1987) tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc Natl Acad Sci* 84: 7383–7387.
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460: 711–716.
- Kuo MY, Sharmeen L, Dinter-Gottlieb G, Taylor J (1988) Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J Virol* 62: 4439–4444.
- Kieft JS (2008) Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci* 33: 274–283.
- Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55: 447–458.
- Clever J, Sasseti C, Parslow TG (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol* 69: 2101–2109.
- Schroeder SJ, Stone JW, Bleckley S, Gibbons T, Mathews DM (2011) Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints. *Biophys J* 101: 167–175.
- Moss WN, Priore SF, Turner DH (2011) Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA* 17: 991–1011.
- Priore SF, Moss WN, Turner DH (2012) Influenza A virus coding regions exhibit host-specific global ordered RNA structure. *PLoS One* 7: e35989.
- Moss WN, Dela-Moss LI, Kierzek E, Kierzek R, Priore SF, et al. (2012) The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. *PLoS One* 7: e38323.
- Ilyinskii PO, Schmidt T, Lukashov D, Meriin AB, Thoidis G, et al. (2009) Importance of mRNA secondary structural elements for the expression of influenza virus genes. *OMICS* 13: 421–430.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406–3415.
- Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, et al. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci* 101: 7287–7292.
- Reuter JS, Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC bioinformatics* 11: 129.
- Johnson NP, Mueller J (2002) Updating the accounts: global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bull Hist Med* 76: 105–115.
- Velagapudi SP, Seedhouse SJ, Disney MD (2010) Structure-activity relationships through sequencing (STARTS) defines optimal and suboptimal RNA motif targets for small molecules. *Angew Chem Int Ed Engl* 49: 3816–3818.
- Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9: 474.
- Mathews DH (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA* 10: 1178–1190.
- Lima S, Hildenbrand J, Korostelev A, Hattman S, Li H (2002) Crystal structure of an RNA helix recognized by a zinc-finger protein: an 18-bp duplex at 1.6 Å resolution. *RNA* 8: 924–932.
- Leontis NB, Stombaugh J, Westhof E (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* 30: 3497–3531.
- Jang SB, Hung LW, Chi YI, Holbrook EL, Carter RJ, et al. (1998) Structure of an RNA internal loop consisting of tandem C-A+ base pairs. *Biochemistry* 37: 11726–11731.
- Kierzek E, Kierzek R, Moss WN, Christensen SM, Eickbush TH, et al. (2008) Isoenergetic penta- and hexanucleotide microarray probing and chemical mapping provide a secondary structure model for an RNA element orchestrating R2 retrotransposon protein function. *Nucleic Acids Res* 36: 1770–1782.
- Lescoute A, Westhof E (2006) Topology of three-way junctions in folded RNAs. *RNA* 12: 83–93.
- Xu Z, Mathews DH (2011) Multalign: an algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics* 27: 626–632.
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, et al. (1994) Fast folding and comparison of RNA secondary structures. *Monatshfte Fur Chemie* 125: 167–188.
- Wilcox JL, Bevilacqua PC (2013) A simple fluorescence method for pK(a) determination in RNA and DNA reveals highly shifted pK(a)'s. *J Am Chem Soc* 135: 7390–7393.
- Yamano H, Gannon J, Hunt T (1996) The role of proteolysis in cell cycle progression in *Schizosaccharomyces pombe*. *The EMBO journal* 15: 5268–5279.
- Pleschka S, Jaskunas R, Engelhardt OG, Zurcher T, Palese P, et al. (1996) A plasmid-based reverse genetics system for influenza A virus. *J Virol* 70: 4188–4192.
- Buratti E, Baralle FE (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* 24: 10505–10514.
- Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, et al. (2008) The influenza virus resource at the National Center for Biotechnology Information. *J Virol* 82: 596–601.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059–3066.
- Kawaoka Y, Gorman OT, Ito T, Wells K, Donis RO, et al. (1998) Influence of host species on the evolution of the nonstructural (NS) gene of influenza A viruses. *Virus research* 55: 143–156.
- Basler CF, Reid AH, Dybing JK, Janczewski TA, Fanning TG, et al. (2001) Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. *Proc Natl Acad Sci* 98: 2746–2751.
- Wilkinson KA, Merino EJ, Weeks KM (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 1: 1610–1616.

36. Kierzek E, Ciesielska A, Pasternak K, Mathews DH, Turner DH, et al. (2005) The influence of locked nucleic acid residues on the thermodynamic properties of 2'-O-methyl RNA/RNA heteroduplexes. *Nucleic Acids Res* 33: 5082–5093.
37. Kierzek E, Mathews DH, Ciesielska A, Turner DH, Kierzek R (2006) Nearest neighbor parameters for Watson-Crick complementary heteroduplexes formed between 2'-O-methyl RNA and RNA oligonucleotides. *Nucleic Acids Res* 34: 3609–3614.
38. Pasternak A, Kierzek E, Pasternak K, Turner DH, Kierzek R (2007) A chemical synthesis of LNA-2,6-diaminopurine riboside, and the influence of 2'-O-methyl-2,6-diaminopurine and LNA-2,6-diaminopurine ribosides on the thermodynamic properties of 2'-O-methyl RNA/RNA heteroduplexes. *Nucleic Acids Res* 35: 4055–4063.
39. Kierzek E, Christensen SM, Eickbush TH, Kierzek R, Turner DH, et al. (2009) Secondary structures for 5' regions of R2 retrotransposon RNAs reveal a novel conserved pseudoknot and regions that evolve under different constraints. *J Mol Biol* 390: 428–442.
40. Frateczak A, Kierzek R, Kierzek E (2011) Isoenergetic microarrays to study the structure and interactions of DsrA and OxyS RNAs in two- and three-component complexes. *Biochemistry* 50: 7647–7665.
41. Kierzek E (2009) Binding of short oligonucleotides to RNA: studies of the binding of common RNA structural motifs to isoenergetic microarrays. *Biochemistry* 48: 11344–11356.
42. Lee JC, Cannone JJ, Gutell RR (2003) The lonepair triloop: a new motif in RNA structure. *J Mol Biol* 325: 65–83.
43. Thulasi P, Pandya LK, Znosko BM (2010) Thermodynamic characterization of RNA triloops. *Biochemistry* 49: 9058–9062.
44. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289: 905–920.
45. Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, et al. (2000) Structure of the 30S ribosomal subunit. *Nature* 407: 327–339.