

## Modeling of RNA Secondary Structure of Non Structural Gene and Evolutionary Stability of the Influenza Virus Through *In Silico* Methods

Pallavi Somvanshi<sup>1\*</sup>, Vijai Singh<sup>2</sup> and M. Arshad<sup>3</sup>

<sup>1</sup>Bioinformatics Centre, Biotech Park, Sector-G Jankipuram, Lucknow-226021, Uttar Pradesh, India, E-mail: psomvanshi@gmail.com

<sup>2</sup>Aquatic Microbes Section, National Bureau of Fish Genetic Resources, Canal ring road, PO- Dilkusha, Lucknow-226002. India, Email: vijai15@rediffmail.com

<sup>3</sup>Department of Zoology, University of Lucknow, Lucknow, E-mail: arshadm123@rediffmail.com

\*Corresponding author: Pallavi Somvanshi, Bioinformatics Centre, Biotech Park, Sector-G Jankipuram, Lucknow-226021, Uttar Pradesh, India, E-mail: psomvanshi@gmail.com; Tel: +91 522 4012076; Fax: +91 522 4012081

Received July 01, 2008; Accepted July 16, 2008; Published July 17, 2008

**Citation:** Pallavi S, Vijai S, Arshad M (2008) Modeling of RNA Secondary Structure of Non Structural Gene and Evolutionary Stability of the Influenza Virus Through *In Silico* Methods. J Proteomics Bioinform 1: 219-226. doi:10.4172/jpb.1000026

**Copyright:** © 2008 Pallavi S, 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.

### Abstract

An influenza virus is an important pathogen causing disease in the birds and further transmits to humans globally. The genome of Influenza virus encodes highly conserved non structural gene (NS1), which is thermodynamically stable in the evolution. Total 32 NS1 nucleotide sequences of Influenza A virus H5N1 strain varied from 831 to 875 bp were used to construct the phylogeny and nine major clades were obtained. The computational tool was used to model the RNA secondary structure of nine different strains of Influenza A virus. The thermodynamic free energy ranges between -222.90 to -251.10 Kcal/mol of the NS which may provide new insight to understand the evolutionary stability and pathogenesis of Influenza virus.

**Keywords :** Influenza Avirus, NS1; RNA secondary structure; Free energy; Evolution

### Introduction

An outbreak of avian influenza A virus of the H5N1 subtype was spread to poultry in Asian countries. The transmission of influenza virus from birds to humans in Hong Kong with outbreaks H5N1 strains of avian influenza A virus Hiorrimoton and Kawaoka (2005) and highly pathogenic influenza A virus H5N1 was identified among the poultry in republic of Korea Fouchier et al. (2005). An outbreak of highly pathogenic avian influenza (HPAI) H5N1 virus was reported from India in 2006. Phylogenetic analyses revealed that Indian isolates were grouped in the mixed-migratory bird sub-lineage of the Eurasian lineage. In the phylogeny analysis viruses were probably introduced to India from China via Europe because they share a direct ancestral relationship with the Indian isolates Kamal et al. (2007).

Influenza viruses are pleomorphic RNA viruses belonging to the *Orthomyxoviridae* family. The genome of influ-

enza virus is segmented and consists of single stranded negative sense RNA; it encodes the 8 structural proteins and non structural gene (NS1). The Influenza virus proteins such as two surface proteins hemagglutinin (HA) and neuraminidase (NA) Fouchier et al. (2005) and other proteins like three RNA polymerase (PA, PB1 and PB2). Nucleoprotein (NP) and matrix protein (M1 and M2) also play role in the cell cycle. The non structural gene (NS1) provides evolutionary stability and replication of Influenza virus Wan et al. (2007).

Several reports are available on the role of NS of the Influenza virus. The avian influenza virus NS encoding protein induces the apoptosis in the human Lam et al. (2008). The NS protein contributes in the pathogenesis and small fragment have deleted from the NS gene that have been reported as the attenuated vaccine for the chickens Zhu et

al. (2008). The influenza virus subtype H5N1 has raised concerns of a possible human pandemic threat because of its high virulence and mutation rate Landon et al. (2008). Highly pathogenic H5N1 influenza viruses have become endemic in poultry populations throughout Southeast Asia and continue to infect humans with a greater than 50% case fatality rate Neumann et al. (2007).

The phylogenetic and proteome analysis of influenza A virus subtype H5N1 have been earlier reported. The study furnish a understanding of the whole proteome function, gene regulation and may be also supportive to vaccine and antiviral drug target to inhibit the functioning of influenza at the specific position of predicted motifs Somvanshi et al. (2008a). The antigenic epitopes were also reported in two highly virulence surface proteins HA and NA of influenza A virus. The host specific epitopes and conserved epitopes have been identified. These results could help in development of immunodiagnostic kit and also designing of vaccine candidate Somvanshi et al. (2008b).

There are limited reports available on RNA secondary structure of the NS in influenza virus. This conformational shift may consequences for splicing regulation of segment mRNA. This study suggest that besides changes at the protein level, changes in RNA secondary structure should be seriously considered when attempting to explain influenza virus evolution Gulyaev et al. (2007).

A varied number of bioinformatics tools were reported to generate the RNA secondary structure of viral gene like RNAdraw, RNAfold, Mfold etc. Mfold uses the nearest neighbor energy rules to calculate the energy of the RNA secondary structure. RNA structure plays an important role in the life cycle of RNA viruses. Many functional viral RNA structure are known and evolution of virus RNA genome is subjected to various structure constrains Simmonds et al. (2004). In the present study, we have predicted the secondary structure of non structural gene of Influenza A virus subtype H5N1. The free energy of the module of non structural gene may predict the evolutionary stability of different host specific strains of Influenza A virus.

## Material and Methods

### Selection of Sequence Data Set

The complete nucleotide sequences of non-structural proteins from different goose, chicken, turkey, swine, swan and duck of Influenza A virus H5N1 strains were retrieved from the biological database such as National Centre for Biotechnology Information (NCBI) cited at <http://www.ncbi.nlm.nih.gov>

### Construction and Analysis of Phylogenetic Tree

All the sequences were aligned with Clustal X2. The computed alignment was manually checked and corrected. Pair-wise evolutionary distances were computed using the Jukes and Cantor equation implemented in the MEGA 3.1 program and a phylogenetic tree was constructed by neighbor-joining method which comprise DNA weight matrix for nucleotide. Bootstrapped values of 100 were sampled to determine a measure of support for each node on the consensus tree.

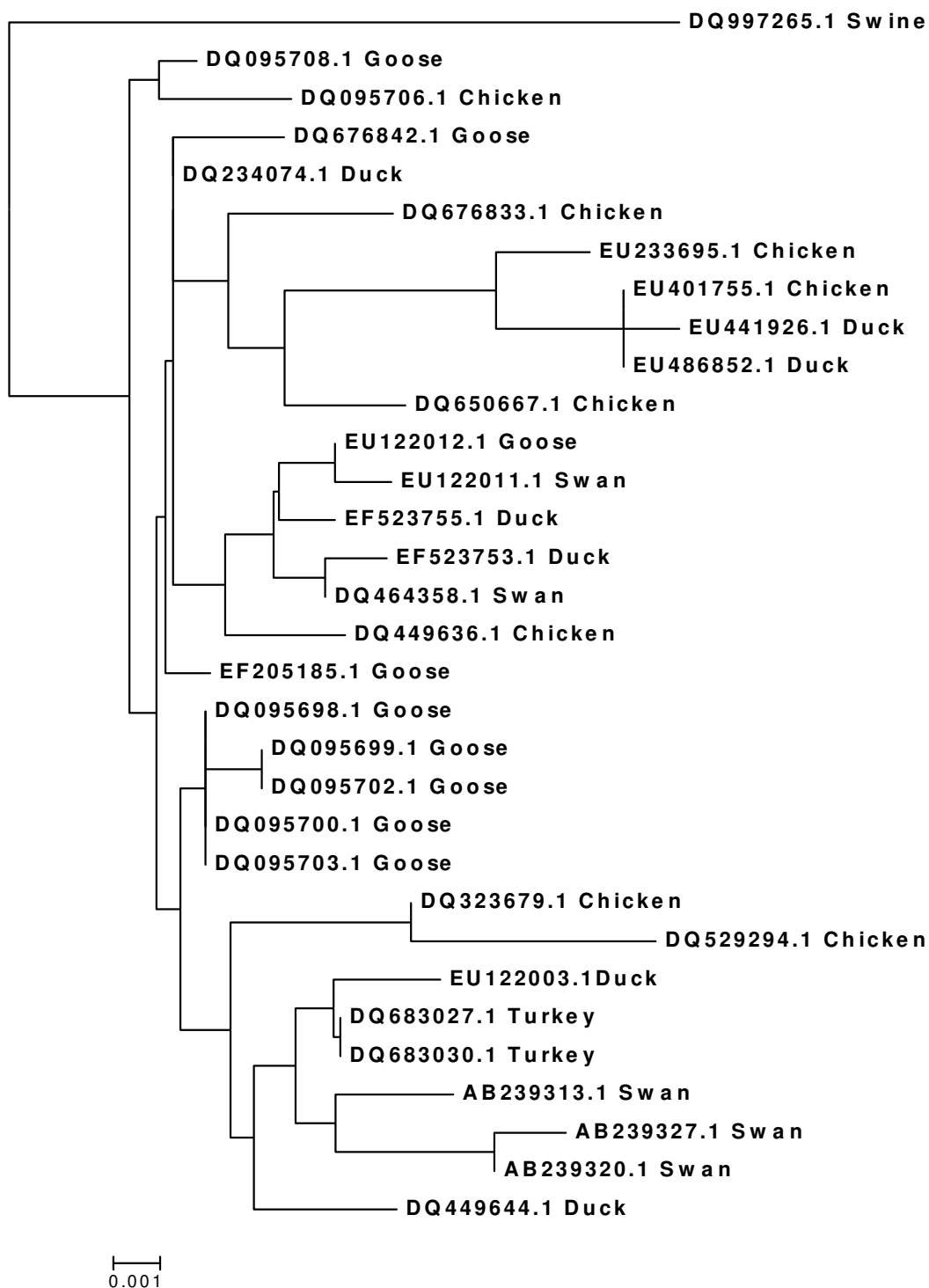
### RNA Secondary Structure Prediction

Prediction of the possible folding of the non structural protein of influenza virus was done with the online Mfold package. The most widely used algorithms for RNA secondary structure prediction, which are based on a search for the minimal free energy state Zuker (1989). The genetic algorithm (GA) simulates the natural folding pathway which takes place during RNA synthesis. This is not only enables new stems added to growing RNA chain but also allows structures to be removed at later stages of the simulation if other pairings are found to be more favourable. The GA also allows the prediction of certain tertiary interactions, including RNA pseudoknots. The minimum free energy was obtained from the secondary structure of RNA.

## Results and Discussion

The size of non structural gene diverge from 831 to 875 bp was used to construct the phylogenetic relationship. All the 32 strains of influenza A virus subtype H5N1 were isolated from a diverse series of animal hosts like Goose, Chicken, Turkey, Swine, Swan and Duck. In this investigation, nine major clades were obtained from the six host strains of influenza A virus (Figure 1). The four major clades in the phylogenetic tree based on the surface HA and NA proteins of influenza A virus H5N1 has been reported Somvanshi et al. (2008a).

The RNA secondary structure of NS gene of influenza A virus showed the lowest free energy -251.10 Kcal/mol from the host swine. The highest free energy in the analysis of the duck strain was -221.70 Kcal/mol. The free energy value of all these NS of different strains of influenza A virus was given (Table 2). The prediction of evolutionarily conserved RNA structures important for elucidating the potential functions of RNA sequences and the mechanisms by which these functions are exerted but it also lies at the core of RNA gene prediction. An accurate prediction of the conserved RNA structure needs a high-quality sequence alignment and an evolutionary tree relating several evolutionarily related sequences.



**Figure 1:** Unrooted phylogenetic tree based on NS1 gene of different strains of Influenza A virus H5N1. The bar represents 0.001 base changes per site.

S. No.	Host	Accession no.	Base	$\delta G$ value (kcal/mol)
A	Goose	DQ095699	831	-222.90
B	Swan	EF523757	833	-226.00
C	Chicken	EU401755	850	-231.10
D	Turkey	DQ683027	867	-234.20
E	Duck	DQ449644	850	-231.90
F	Chicken	DQ095706	833	-225.50
G	Goose	DQ095708	830	-222.40
H	Duck	DQ321138	831	-221.70
I	Swine	DQ997265	875	-251.10

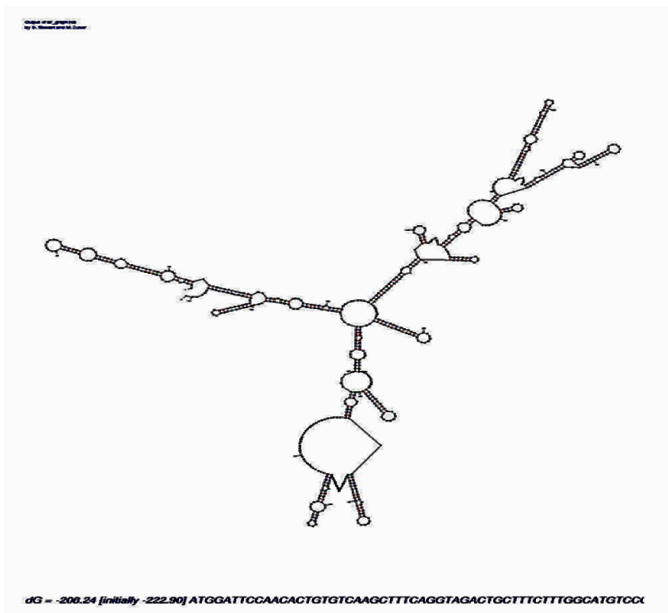
**Table 1:** The thermodynamic free gibbs energy ( $\delta G$ ) of Influenza A virus H5N1 nonstructural gene (NS1).

Computational tool was used to model the RNA secondary structure in coding region of hepatitis C virus (HCV) includes thermodynamic prediction, calculation of free energy on folding, and a newly developed method to scan sequences for covariant sites and associated secondary structures using a parsimony-based algorithm. Total six evolutionary conserve stem-loop structures in the NS5B-encoding region and two in the core gene. The virus most closely relate to HCV, GB virus-B (GBV-B) also showed evidence for similar internal base pairing in its coding region, although predictions of secondary structures were limited by the absence of comparative sequence data for this virus Tuplin et al. (2002).

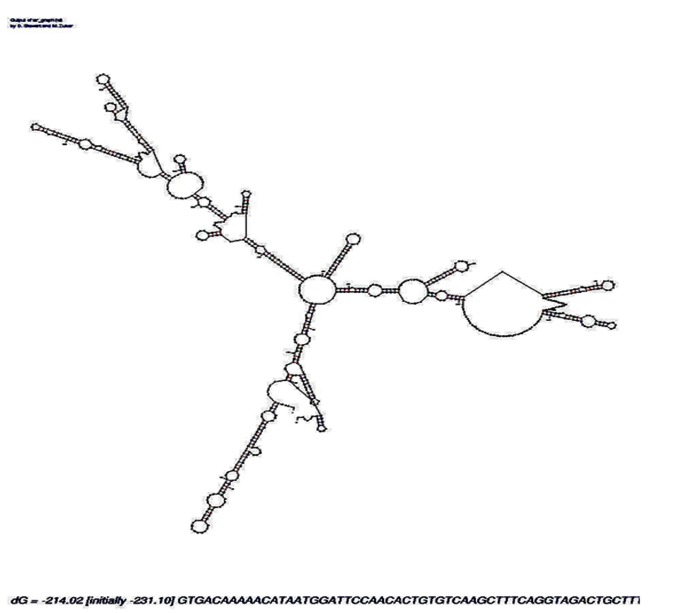
The secondary structure RNA non structural gene of influenza A virus provides stability of the genome during evolution and adaptation in the infection to various hosts. This gene is highly conserved in the genome of influenza A virus and slight variation in the size. The confirmation of secondary structure was assorted due to mutation in the NS gene. The mutation of the NS region possibly will be reducing the pathogenesis of influenza virus. The secondary structure of RNA of NS of influenza A virus was given (Figure 2A-2I). In 2001 and 2003, two influenza A virus H5N1, A/swine/Fujian/1/01 (SW/FJ/01) and A/swine/Fujian/1/03 (SW/FJ/03) isolated from pigs in Fujian Province, southern China was studied. Both the virus were genetically similar, although the NS gene of the SW/FJ/03 virus has a 15-nucle-

otide deletion at coding positions 612 to 626. The small fragment deleted from the NS gene that have used as the attenuated vaccine for the chickens Zhu et al. (2008). Secondary structure models exhibited three pairs of small subunit ribosomal RNA molecules. These are the 16S rRNA from *E. coli* cytoplasmic and *Z. mays* chloroplast ribosomes, the 18S rRNA from *S. cerevisiae* and *X. laevis* cytoplasmic ribosomes, and the 12S rRNA from human and mouse mitochondrial ribosomes. The model supports the concept that secondary structure of ribosomal RNA has been extensively conserved throughout the evolution Zwiebel et al. (1981).

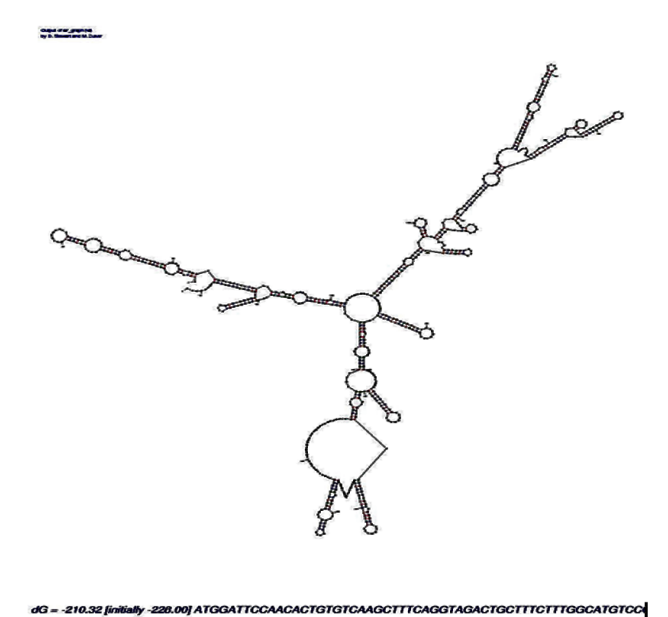
The 5S rRNA gene from *Sphingobium chungbukense* DJ77 was identified. The secondary structure of the 199-base-long RNA was proposed. The two-base-long D loop was the shortest among all of the known 5S rRNAs. The U19-U64 non-canonical pair in the helix II region was uniquely found in strain DJ77 among all of the sphingomonads Kwon and Kim (2007). The nucleotide sequence of *Pinus silvestris* 5S rRNA was determined using two independent methods and compared with other plant 5S rRNAs. It shows more than 90% sequence homology with gymnosperm 5S RNAs. The free energy ( $\delta G$ ) analysis of 5S rRNAs from gymnosperms, angiosperms and the other higher plants revealed that the free energy of this ribosomal RNA decreases with evolution Mashkova et al. (1990).



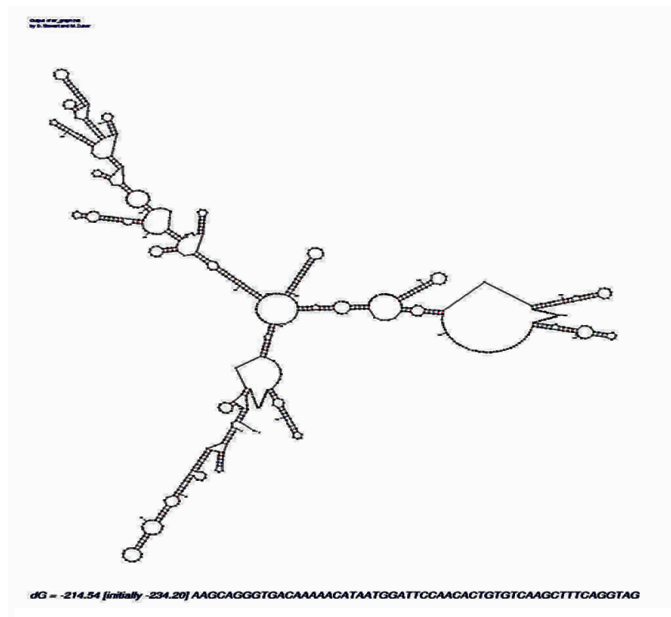
Goose (A)



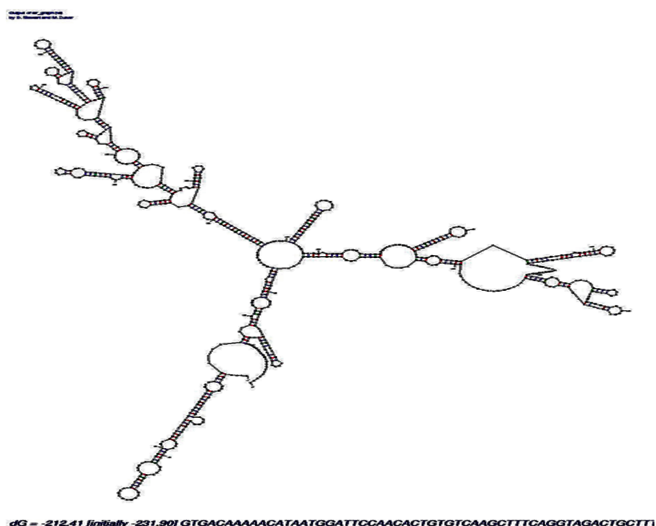
(C) Chicken



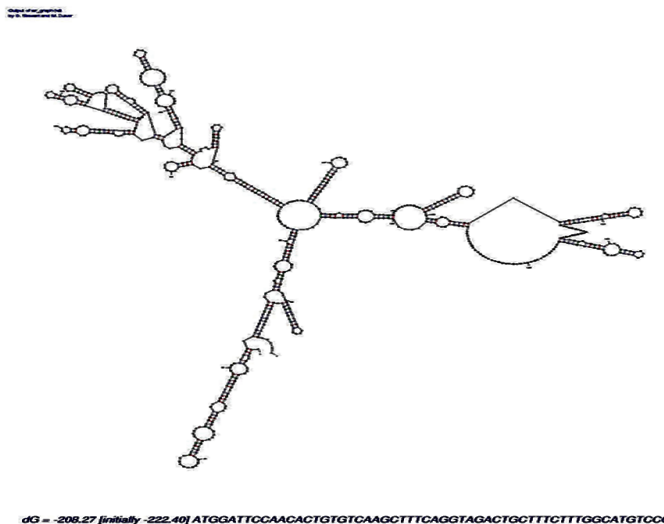
Swan (B)



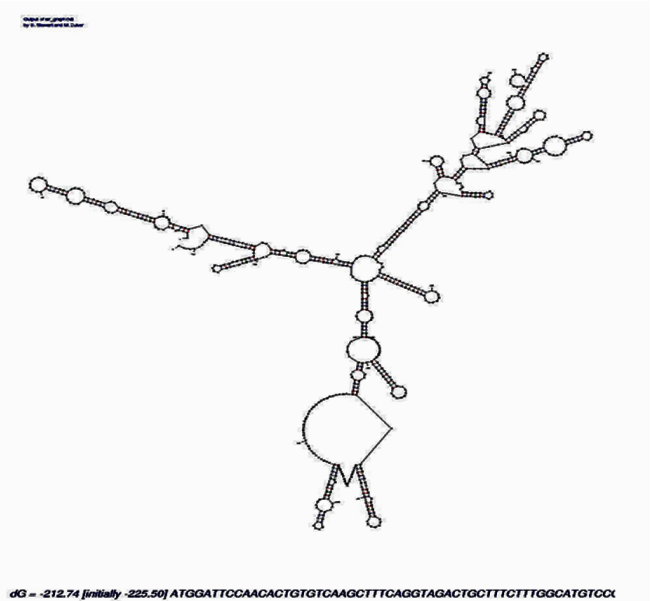
(D) Turkey



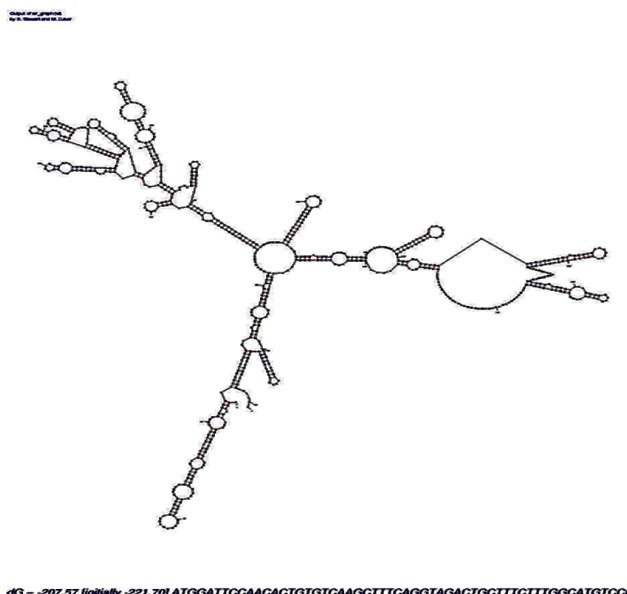
(E) Duck



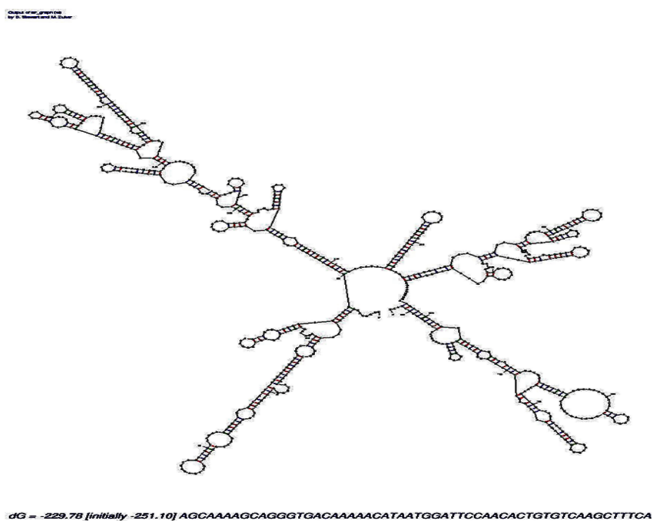
Goose (G)



(F) Chicken



Duck (H)



Swine (I)

Figure 2: RNA secondary structure of Influenza A virus H5N1 strains from different animals.

Hepatitis C virus (HCV) possesses extensive RNA secondary structure in the core and NS5B-encoding regions of the genome. A program was developed (STRUCTUR\_DIST) that analyses multiple RNA-folding patterns predicted by MFOLD to determine the evolutionary conservation of predicted stem-loop structures by a new method, to analyze frequencies of covariant sites in predicted RNA folding between HCV genotypes Tuplin et al. (2004).

RNA secondary structure prediction was combined with comparative sequence analysis to construct models of folding for the distal 380 nucleotides of the 3'-untranslated region (3'-UTR) of yellow fever virus (YFV). A number of structural elements that are thermodynamically stable, conserved in shape, and confirmed by compensatory mutations were revealed. At the same time structural polymorphisms were observed among strains of YFV. The observation of a strong association between secondary structure of the 3'-UTR and virulence of YFV may help elucidate the molecular mechanisms of virus attenuation and lead to new strategies of vaccine development directed towards rational modification of secondary structure of the 3'-UTR. In this study, the free energy value of pathogenic YFV was higher and non pathogenic YFV was lower Proutski et al. (1997).

In conclusion, this study was carried out for the modeling of RNA secondary structure and phylogenetic analysis of influenza A virus H5N1. The prediction of RNA structure in conjunction with structure known to exist within the virus untranslated region may facilitate further understanding of virus translation, replication and packaging. The free energy of the conserved NS gene may help to understand the stability of influenza A virus through out the evolution.

## Acknowledgement

The authors are thankful to Indian Council of Medical Research, New Delhi for providing financial support and also appreciate critical suggestions provided by Mr. Indramani and Mr. D. K. Chaudhary during the preparation of manuscript.

## References

1. Fouchier RA, Munster V, Wallensten A, Bestebroer TM, Herfst S, et al. (2005) Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J virol* 79: 2814-22. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
2. Gulyaev AP, Heus HA, Olsthoorn RC (2007) An RNA conformational shift in recent H5N1 influenza A viruses. *Bioinformatics* 23: 272-6. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
3. Hiorrimoton T, Kawaoka Y (2005) Influenza: Lessons from past pandemic warnings from current incidents. *Nature Rev Microbiol* 3: 591-600. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
4. Kamal RP, Tosh C, Pattnaik B, Behera P, Nagarajan S, et al. (2007) Analysis of the PB2 gene reveals that Indian H5N1 influenza virus belongs to a mixed-migratory bird sub-lineage possessing the amino acid lysine at position 627 of the PB2 protein. *Arch Virol* 152: 1637-44. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
5. Kwon HR, Kim YC (2007) Nucleotide sequence and secondary structure of 5S rRNA from *Sphingobium chungbukense* DJ77. *J Microbiol* 45: 79-82. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
6. Lam WY, Tang JW, Yeung AC, Chiu LC, Sung JJ, et al. (2008) Avian influenza virus A/HK/483/97(H5N1) NS1 protein induces apoptosis in human airway epithelial cells. *J Virol* 82: 2741-51. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
7. Landon MR, Amaro RE, Baron R, Ngan CH, Ozonoff D, et al. (2008) Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. *Chem Biol Drug Des* 71: 106-16. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
8. Mashkova TD, Barciszewska MZ, Joachimiak A, Nalaskowska M, Barciszewski J (1990) Molecular evolution of plants as deduced from changes in free energy of 5S ribosomal RNAs. *Int J Biol Macromol* 12: 247-50. » [Pubmed](#) » [Google Scholar](#)
9. Neumann G, Shinya K, Kawaoka Y (2007) Molecular pathogenesis of H5N1 influenza virus infections. *Antivir Ther* 12 : 617-26. » [Pubmed](#) » [Google Scholar](#)
10. Proutski V, Michael W, Ernest A, Holmes EC (1997) Secondary structure of the 3'- untranslated region of yellow fever virus: implications for virulence, attenuation and vaccine development. *J Gen Virol* 78: 1543-1549. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
11. Simmonds P, Tuplin A, Evans D J (2004) Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA* 10: 1337-1351. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
12. Somvanshi P, Singh V, Seth PK (2008a) Phylogenetic and computational proteomes analysis of Influenza A virus subtype H5N1. *Int J Genomics Proteomics* 3. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
13. Somvanshi P, Singh V, Seth PK (2008b) Prediction of epitopes in Hemagglutinin and Neuraminidase proteins of Influenza A virus H5N1 strain: A clue for diagnostic and vaccine development. *Journal of Integrative Biology* 12: 61-69. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
14. Tuplin A, Evans DJ, Simmonds P (2004) Detailed map-

- ping of RNA secondary structures in core and NS5B-ncoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J Gen Virol* 85: 3037-3047. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
15. Tuplin A, Wood J, Evans DJ, Patel AH, Simmonds P (2002) Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* 8: 824-841. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
17. Wan XF, Wu X, Lin G, Holton SB, Desmone RA, et al. (2007) Computational identification of reassortments in avian influenza viruses. *Avian Dis* 51: 434-9. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
18. Zhu Q, Yang H, Chen W, Cao W, Zhong G, et al. (2008) Naturally occurring deletion in its NS gene contributes to the attenuation of an H5N1 swine influenza virus in chickens. *J Virol* 82: 220-8. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
19. Zuker M (1989) On Finding All Suboptimal Foldings of an RNA Molecule. *Science* 244: 48-52. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)
20. Zwieb C, Glotz C, Brimacombe R (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. *Nucleic Acids Res* 9: 3621-40. » [CrossRef](#) » [Pubmed](#) » [Google Scholar](#)