

## Analysis of PB2 protein from H9N2 and H5N1 avian flu virus

Danishuddin<sup>1</sup> and Asad Ullah Khan<sup>1,2,\*</sup>

<sup>1</sup>Bioinformatics distributed Information sub-centre, AMU, Aligarh-202002, India; <sup>2</sup>Interdisciplinary Biotechnology Unit, Aligarh Muslim University, Aligarh-202002, India; Asad Ullah Khan\* - E-mail: asad.k@rediffmail.com; Phone: 91-571-2723088; Fax: 91-571-2721776; \* Corresponding Author

received April 26, 2008; revised June 10, 2008; accepted July 06, 2008; published September 22, 2008

### Abstract:

Influenza A viruses of subtype H9N2 are wide spread among poultry and other mammalian species. Crossing the species barrier from poultry to human occurred in recent years creating a pandemic of H9N2 virus. It is known that the pathogenicity of H9N2 is lower than H5N1. Nonetheless, it is important to establish the molecular functions of H9N2 viral proteins. We studied mutations in the polymerase protein PB2 of H9N2 from different strains and compared it with the highly pathogenic H5N1. The mutation M294T was found to be important in the N-myristoylation domain of Ck/UP/2573/India/04(H9N2) isolate. Prediction of secondary structures and PROSITE motif assignments were performed for PB2 to gain functional insight. Subsequently, the effect of mutations in secondary structures among strains is discussed.

**Keywords:** PB2 protein; H5N1; H9N2; mutations; N-myristoylation domain

### Background:

Influenza viruses belong to the member of orthomyxo-viridae family, having a single-standard, negative sense, segmented RNA genome in an enveloped virion [1-6]. The genome of influenza viruses have 8 RNA segments encoding 10 proteins including two surface glycoprotein, haemagglutinin (HA), neuraminidase (NA), nucleoproteins (NP), three polymerase proteins (PA,PB1,PB2) two matrix (M1,M2) and nonstructural proteins (NS1,NS2). Influenza viruses are classified as types A, B and C based on the antigenic properties of nucleoprotein (NP) and matrix (M1) [4, 5]. Avian influenza is caused by a type-A virus and it is further classified into subtype based on two surface glycoprotein, haemagglutinin (HA) and neuraminidase (NA). There are about 16 known HA and 9 known NA in type-A viruses [7].

Viral genes from the aquatic bird are thought to be the source of all influenza-A viruses in other animal species [8]. Recently, subtypes H5N1, H9N2, H7N7 of influenza A have been identified in humans, raising the possibility of a new influenza pandemic among the uninfected populations. Avian influenza viruses are categorized as highly pathogenic avian influenza (HPAI) and low pathogenic avian influenza (LPAI) viruses based on virulence. The HPAI virus causes systemic lethal infection and can kill birds within 24 hours of post-infection. However, the LPAI viruses show rare outbreak of pandemics. Therefore, their morbidity and mortality rates are low [8]. Surveillance of live poultry in December 1997 revealed that 4% of infected poultry was caused by H9N2 subtype [9, 10].

The polymerase protein PB2 is 759 residues long influenza-A viruses. The polymerase complex is formed protein subunit interactions with the PB1 subunit [11]. PB2 interacts with NP and not with PA [12, 13]. This subunit plays an important role in transcription of mRNA by binding with the 5' methylated cap of pre-mRNA in host cells for providing primers to viral mRNA synthesis [14]. Two hypothetical domains (residues 242-280 and 538-577) identified in PB2 are hypothesized for viral binding to host cells [15]. Here, we describe the effect of mutations in PB2 from different strains of H9N2. The analysis is extended to that in H5N1 for comparison.

### Methodology:

#### Data set

The PB2 protein sequences of subtypes H5N1 and H9N2 were downloaded from Genome directory INFLUENZA at NCBI [23] and summary of sequence data given in Table 1 (see supplementary material). The dataset consists of PB2 sequences from 3 H9N2 strains and 2 H5N1 strains. We used CLUSTALW [16] for generating a multiple sequence alignment MSA of 5 PB2 sequences from H9N2 strains and H5N1 strains.

#### Mutational analyses

The MSA for 5 PB2 sequences from H9N2 strains and H5N1 strains was used to identify mutations in PB2 with reference to Ck/UP/India/04. We used the amino acids sequences 151 to 690 for the analyses.

HADLSAKEAQQDVIMEVFPNEVGARILTSQSQTITKEKKEELQDCKIAP	200	Ck/UP/Ind/04 H9N2
.....		Gs/MM/80 H9N2
.....		Dk/HD/9/99 H9N2
.....		Ck/Hu/w1/97 H5N1
.....		Hk/213/03 H5N1
LMVAYMLERELVRKTRFLPVAGGTSVYIEVHLTQGTWQMYTPGGEV	250	Ck/UP/Ind/04 H9N2
.....		Gs/MM/80 H9N2
.....		Dk/HD/9/99 H9N2
.....		Ck/Hu/w1/97 H5N1
.....		Hk/213/03 H5N1
RNDDVDQSLIIAARSI VRRATVSADPLASLLEMCHSTQIGGVRTVDILRQ	300	Ck/UP/Ind/04 H9N2
.....N.....I.M.....		Gs/MM/80 H9N2
.....N.....I.M.....		Dk/HD/9/99 H9N2
.....N.....I.M.....		Ck/Hu/w1/97 H5N1
.....N.....I.M.....		Hk/213/03 H5N1
NPTEEQAQVDICKAAMGLRISSSFSGGFTEKRTSGSSVKKEEVL TGNLQ	350	Ck/UP/Ind/04 H9N2
.....R.....		Gs/MM/80 H9N2
.....R.....		Dk/HD/9/99 H9N2
.....R.....		Ck/Hu/w1/97 H5N1
.....R.....		Hk/213/03 H5N1
TLKIRVHEGYEEFTMVGRRATAILRKATRRLIQLIVSGRDEQSI AERLIV	400	Ck/UP/Ind/04 H9N2
.....		Gs/MM/80 H9N2
.....		Dk/HD/9/99 H9N2
.....		Ck/Hu/w1/97 H5N1
.....		Hk/213/03 H5N1
AMVFSQEDCMIKAVRGDLNFVNRATQRLNPMHQLLRHFQRDAKVL FQNWG	450	Ck/UP/Ind/04 H9N2
.....N.....K.....		Gs/MM/80 H9N2
.....N.....K.....		Dk/HD/9/99 H9N2
.....N.....K.....		Ck/Hu/w1/97 H5N1
.....N.....K.....		Hk/213/03 H5N1
IEPIDNVMGIGILPDMTPEMSLRGVRSKMGVDEYSSTERVVVSDR	500	Ck/UP/Ind/04 H9N2
.....I.....		Gs/MM/80 H9N2
.....I.....		Dk/HD/9/99 H9N2
.....I.....		Ck/Hu/w1/97 H5N1
.....I.....		Hk/213/03 H5N1
QWLRNWTVKIQWSQDPTILYKMEFEPFQSLVPKAARGQYSGFARTLF	600	Ck/UP/Ind/04 H9N2
.....M.....V.....		Gs/MM/80 H9N2
.....M.....V.....		Dk/HD/9/99 H9N2
.....M.....V.....		Ck/Hu/w1/97 H5N1
.....M.....V.....		Hk/213/03 H5N1
QQMRDVLGFTDVTQIKLLPFARAPPEQSRMQFSSLTVNVRGSGMRILVR	650	Ck/UP/Ind/04 H9N2
.....		Gs/MM/80 H9N2
.....		Dk/HD/9/99 H9N2
.....		Ck/Hu/w1/97 H5N1
.....		A/HK/213/03 H5N1
GNSPVFNHVKVKRLTVLGGKAGALTEDPDEGTAGVESAV	690	Ck/UP/Ind/04 H9N2
.....A.....		Gs/MM/80 H9N2
.....A.....		Dk/HD/9/99 H9N2
.....A.....		Ck/Hu/w1/97 H5N1
.....A.....I.....		Hk/213/03 H5N1

**Figure 1:** Multiple sequence alignment of PB2 protein sequences (residue positions 151 to 690) in H9N2 and H5N1. Identical residues to Ck/UP/2543/India/04 H9N2 are indicated by dots (.).

### Secondary structure prediction

We used GOR IV to assign secondary structures to PB2 sequences [17].

### Domains/motifs search

We used PROSITE scan at the EXPASY to identify PROSITE motifs in PB2 proteins [18].

### Results:

#### Sequence analyses

The PB2 protein sequence of A/Ck/UP/2543/India/04(H9N2) isolate was compared with other homologous sequences as were given in Table 1 (see supplementary material). The PB2 protein sequences obtained from GenBank have at least 97% similarity among themselves and a multiple sequence alignment is obtained using the software ClustalW (Figure 1). Mutations are summarized in Table 2 (under supplementary material) with changes in corresponding predicted secondary structures.

On comparing the data we found that all the mutations except 340(R→K) and 478(V→I) in A/goose/MN/5733-1/80(H9N2), and 667(V→I) in A/HK/213/03(H5N1). All the mutations that were observed are synonymous and will not change the chemical and structure properties of this protein. Only one non synonymous mutation, M294T in Ck/UP/2543/India was found to change the chemical property from Hydrophobic to Hydrophilic at the respective site. The most important finding, observed, was the identification of N-myristoylation domain. This domain (GGvrTV) was found at positions 290-295 due to mutation at 294 (M→T) in A/Ck/UP/2543/India/04 (H9N2) as shown in Table 3 (supplementary material). Moreover, mutations at position N265S and M570I in A/Ck/UP/2543/India/04(H9N2) were observed in binding regions of PB2 protein.

### Secondary structure prediction

Secondary structure changes due to mutations were also found. Mutations, M294T in Ck/UP/2543/India/04 and I478V in Gs/Mn/5733-1/80 responsible for change in secondary structure from helix to strand, mutation M570I in U.P isolate altered the secondary structure from strand to helix (Table 2 in supplementary material). No other significant difference in the predicted regions in all the sequences were observed except in A/Ck/UP/2543/India/04(H9N2) strain, region 293 to 298 was predicted to be extended strand while all the other selected strains have helix at this position and region 569 to 574 was predicted to be helix while all the other strains have extended strand at this position.

### Domain/Motif search

Different domains/Motifs that were found in PB2 protein of A/Ck/UP/2543/India/04 (H9N2) were listed in Table 3 (under supplementary material). One of the most important finding that we observed was the identification of a new N-

myristoylation domain at position 290-295 in A/Ck/UP/2543/India/04(H9N2). This domain was found due to mutation at M294T. Amino acids (290-295) code for a new N-myristoylation domain (GGvrTV).

### Sequence alignment

Amino acid sequences of PB2 protein of H9N2 and H5N1 subtypes were aligned by using ClustalW program. Sequence analyses of PB2 protein (AA from 151 to 690) of H9N2 and H5N1 isolates revealed that PB2 protein of these two subtypes were similar by 97% (Figure 1).

### Discussion:

PB2 protein is 759 amino acids long segment on the influenza A genome. This subunit bind to the 5' methylated cap of pre-mRNA of host cells, before they cleaved to provide primers for viral mRNA synthesis. Honda, et al showed that two domains (residues, 242-280 and 538-577) were identified as a binding region in PB2 protein. These binding regions are conserved among different subtypes. Our analyses showed that mutations (N265S and M570I) were identified in binding region of PB2 of Ck/UP/2543/India/04(H9N2) isolate. The most important finding that we observed was identification of a new N-myristoylation domain found at position 290-295 in A/Ck/UP/2543/India/04(H9N2). It is post translation domain, plays a vital role in membrane targeting and signal transduction. N-myristoylation domain is found in higher eukaryotes, fungi and viruses which are formed as a result of post-translation protein modification. In myristoylation domain, myristic acid is covalently attached via an amide bond to the alpha-amino group of a glycine residue at N-terminal. This modification is catalyzed by N-myristoyl transferase enzyme. It plays a vital role in membrane targeting and signal transduction [19]. Myristoylation have ability to influence the conformational stability of individual proteins. Moreover, it also interacts with membranes or the hydrophobic domains of other proteins [20].

This study also revealed that the change in secondary structure occurred, is due to mutations as discussed in Table 2 (supplementary material). These mutations not only change the secondary structure conformation at the respective site but also alter the nearest side chain conformation as predicted by GORIV methods. Sequence of PB2 protein of these two subtypes was showing similarity (Figure 1) among each other, raising the possibility that exchange of internal segments between these two subtypes have been occurred [21, 22].

### Conclusion:

The PB2 subunit plays an important role in transcription regulation in viral mRNA synthesis. We showed mutational difference in PB2 in different strains of H9N2. Mutations are seen in the predicted secondary structures of PB2. These findings have implications in understating PB2 function in viral multiplication and infection.

**Acknowledgement:**

We are thankful to Prof. M. Saleemuddin, coordinator, IBU for providing facility for the work. AUK acknowledges DBT grant BT/PR7507/BID/07/201/2006.

**References:**

- [01] T. Anwar, *et al.*, *In Silico Biology*, 6: 0015 (2006) [PMID: 16922680]
- [02] A. U. Khan, *Bioinformatics*, 1: 132 (2006) [PMID: 1891670]
- [03] T. Anwar and A. U. Khan, *Bioinformatics*, 1: 253 (2006) [PMID: 1891697]
- [04] H. K. Pradhan *et al.*, *Arch Virol.*, 152: 1637 (2007) [PMID: 17558541]
- [05] Y. Suzuki and M. Nei, *Mol Biol Evol.*, 19: 501 (2002) [PMID: 11919291]
- [06] I. Ahn *et al.*, *Eur J Epidemiol.*, 21: 519 (2006) [PMID: 16858618]
- [07] R. A. Fouchier *et al.*, *J. Virol.*, 79: 2814 (2005) [PMID: 15709000]
- [08] T. Horimoto and Y. Kawaoka, *Nat Rev Microbiol.*, 8: 591 (2005) [PMID: 16064053]
- [09] S. Li *et al.*, *J Virol.*, 77: 6988 (2003) [PMID: 12768017]
- [10] Y. Guan *et al.*, *J. Virol.*, 74: 9372 (2000) [PMID: 11000205]
- [11] P. Digard *et al.*, *Virol.*, 171: 162 (1989) [PMID: 2741339]
- [12] S. K. Biswas *et al.*, *J. Gen. Virol.*, 72: 5493 (1998) [PMID: 9621005]
- [13] T. Zurcher *et al.*, *J.Gen. Virol.*, 77: 1745 (1996) [PMID: 8760421]
- [14] Y. Nakagawa *et al.*, *J Virol.*, 70: 6390 (1995) [PMID: 8709268]
- [15] A. Honda *et al.*, *Genes Cells*, 4: 475 (1999) [PMID: 10526235]
- [16] <http://www.ebi.ac.uk/Tools/clustalw>
- [17] <http://npsa-pbilibcp.fr>
- [18] <http://ca.expasy.org/tools/scanprosite/>
- [19] S. Podell and M. Gribskov, *BMC Genomics*, 5: 37 (2004) [PMID: 15202951]
- [20] J. Zheng *et al.*, *Protein Sci.*, 2: 1559 (1993) [PMID: 8251932]
- [21] Y. Guan *et al.*, *Proc. Natl. Acad. Sci. USA*, 96: 9363 (1999) [PMID: 10430948]
- [22] M. Peiris *et al.*, *J Clin Microbiol.*, 37: 3426 (1999) [PMID: 10488228]
- [23] <ftp://ftp.ncbi.nih.gov/genomes/INFLUENZA>

Edited by P. Kanguane

Citation: Danishuddin & Khan, *Bioinformatics* 3(1): 41-46 (2008)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

### Supplementary material

GenBank accession	Strain	Subtype	Source
ABI53725	Ck/UP/India/04	H9N2	Chicken
BAE96030	Ck/HD/9/99	H9N2	Chicken
ABB88400	Gs/MN/5733-1/80	H9N2	Goose
ABI96740	Ck/Hu/w1/97	H5N1	Chicken
BAE07197	HK/213/03	H5N1	Human

**Table 1:** Strains selected for comparison study.

A/Gs/MN/5733/-1/80 (H9N2)			
Base change	Change in Ck/2543/UP/04	Property change	Secondary structure change
265	N→S	Hydrophilic	H
292	I→V	Hydrophobic	C
294	M→T	Hydrophobic → Hydrophilic	H→E
340	R→K	Hydrophilic	H
425	N→T	Hydrophilic	H
440	K→R	Hydrophilic	H
478	I→V	Hydrophobic	H→E
570	M→I	Hydrophobic	E→H
596	V→A	Hydrophobic	H
661	A→V	Hydrophobic	C
A/DK/HD/9/99 (H9N2)			
265	N→S	Hydrophilic	H
292	I→V	Hydrophobic	C
294	M→T	Hydrophobic → Hydrophilic	H→E
425	N→T	Hydrophilic	H
440	K→R	Hydrophilic	H
570	M→I	Hydrophobic	E→H
596	V→A	Hydrophobic	H
661	A→V	Hydrophobic	C
A/Ck/Hubei/W1/97 (H5N1)			
265	N→S	Hydrophilic	H
292	I→V	Hydrophobic	C
294	M→T	Hydrophobic → Hydrophilic	H→E
425	N→T	Hydrophilic	H
440	K→R	Hydrophilic	H
570	M→I	Hydrophobic	E→H
596	V→A	Hydrophobic	H
661	A→V	Hydrophobic	C
A/HK/213/03 (H5N1)			
265	N→S	Hydrophilic	H
292	I→V	Hydrophobic	C
294	M→T	Hydrophobic → Hydrophilic	H→E
425	N→T	Hydrophilic	H
440	K→R	Hydrophilic	C
570	M→I	Hydrophobic	E→H
596	V→A	Hydrophobic	H
661	A→V	Hydrophobic	C
667	I→V	Hydrophobic	C

**Table 2:** Residue level mutations in PB2 for different strains compared to A/Ck/UP/2543/India/04(H9N2). (H = helix, E = extended strand, C = random coil).

Site	Position	Domain	Function	
Casein kinase II phosphorylation site	155-158	SakE	Involve in budding process of influenza A virus.	
	238-241	TcwE		
	279-282	SIEE		
	387-390	Sgrd		
	393-396	SiaE		
	405-408	SqeD		
	489-492	SstE		
cAMP and cGMP-dependent protein kinase phosphorylation site	514-517	SstE	Involve in Phosphorylation of Serine and Threonine residue found close to a N-terminal basic residue.	
	268-271	Rrat		
	331-334	KRtS		
	368-371	RRaT		
	375-378	RKat		
Protein kinase C phosphorylation site	663-666	KRIT	Involve in Phosphorylation of Serine and Threonine residue found close to a C-terminal basic residue.	
	155-157	SaK		
	329-331	TfK		
	337-339	SvK		
	351-353	TlK		
	378-380	TrR		
	387-389	SgR		
	425-427	TqR		
	474-476	SIR		
	491-493	TeR		
	524-526	TeK		
	559-561	TvK		
	662-664	TkR		
N-myristoylation Site	222-227	GGtsSV	Post translation protein modification domain.	
	<b>290-295*</b>	<b>GGvrTV*</b>		
	316-321	GLriSS		Plays a vital role in membrane targeting and signal transduction. Myristoylation occur in Cytoplasm.
	347-352	GN1qTL		
	477-482	GVrvSK		
	590-595	GQysGF		
	608-613	GTfdTV		
	685-690	GVesAV		
Amidation Site	366-369	vGRR	Required for C-terminus amidation site Occurs in Secretory vesicles and granules.	
Cell attachment Sequence	415-417	RGD	It plays an important role in virus attachment and virus entry.	

**Table 3:** PROSITE motif assignments in PB2 protein of A/Ck/UP/2543/India/04 (H9N2)