Identification of mutations at the antigenic and glycosylation sites in hemagglutinin protein of H5N1 strain

Parveen Salahuddin[1], Asad U Khan[1, 2*]

[1]Distributed Information Sub-Centre, [2]Interdisciplinary Biotechnology Unit, Aligarh Muslim University, Aligarh 202002, India; Asad U Khan - Email: asad.k@rediffmail.com; Phone: +91 571 2723088; Fax: +91 571 2721776; *Corresponding author

Abstract:
Hemagglutinin (HA) is the principal antigen, present on the viral surface. It is the primary target for neutralizing antibodies. In this paper, we have carried out studies on human hemagglutinin protein from H5N1 strain with homologous hemagglutinin from non-human sources of H5N1 strains. In all strains, part of the antigenic site (128-141) predicted by computer program "Antigenic", corresponds to immunodominant site Sa of H1 subtype. In AAF02304 strain, A156→S156 mutation lies at the antigenic subsite of site 2 that corresponds to site B in the H3 subtype. In some strains of non-human origins, there are mutations at the antigenic sites. Interestingly, in AAY56367 strain mutation L138→H138 lies at the receptor binding site, which also overlaps the antigenic site. Therefore, this amino acid substitution may influence both the specificity of receptor recognition and antibody binding. Seven potential glycosylation sites in human HA and in some strains of non-human sources have been predicted by computer program, Scan Prosite. In some strains of HA from non-human sources because of mutation, an additional glycosylation site appeared at the antigenic site. Therefore in these strains the oligosaccharides will mask the surface of HA as well as antigenic site. Hence these strains will not be recognized by host immune system.

Keywords: antigenic site; avian influenza; hemagglutinin

Background:
Influenza viruses pose significant challenge to both human and animal health. Highly pathogenic avian influenza virus infection can cause huge economic loss. Viruses of low virulence are changed to high pathogenicity when they are circulating in the poultry population due to high mutation rate in their genome [1]. The first outbreak of avian influenza A (H5N1) virus in humans occurred in Hong Kong in 1997. Infection was confirmed in 18 individuals, 6 of whom died. Infections were acquired by humans directly from chickens, without the involvement of an intermediate host. The outbreak was halted by a territory-wide slaughter of more than 1.5 million chickens at the end of December 1997. Influenza A virus belongs to the viral family of *Orthomyxoviridae* that have segmented, negative single-stranded sense RNA genome constituting envelop virion [2, 3]. Influenza A virus causes epidemics and pandemics in mammals and birds. However, aquatic birds are natural reservoir of these viruses [4-6]. The H5N1 sub-type is the only highly pathogenic avian viral sub-type that has been documented to cause an outbreak of respiratory disease in humans. Human influenza A virus preferentially binds to sialic acid linked to galactose by alpha-2,6 linkages that are mainly found on the epithelial cells of the human respiratory tract, whereas avian viruses tend to bind to alpha-2,3 linkages that are found on duck intestinal epithelium [7, 8]. These differences may be important in limiting the transfer of viruses between species. There is no evidence of human-to-human transmission to date [9], because they cannot 'hop' easily among humans. However, in future, this strain might acquire ability to spread infection among humans, either by mutation or by reassortment of genetic material with human influenza virus [10].

Hemagglutinin (HA) is the principal antigen present on the viral surface. It is the primary target for neutralizing antibodies and is responsible for viral binding to host receptors, enabling entry into the host cell through endocytosis and subsequent membrane fusion. Earlier *in silico* studies predicted that HA of influenza A virus is highly sensitive to mutations and hence it is significant for its pathogenic nature [11].Therefore, HA is an important target for both drug and vaccine development. Although, 16 avian and mammalian serotypes of HA are known, only three (H1, H2, and H3) have become adapted in the human population. Five antigenic sites designated A to E have been located on the hemagglutinin monomer in H3N1 subtype [12]. Evidence suggests that single amino acid substitution within these five regions results in the ability of flu virus to escape immune surveillance and to spread infections worldwide every year. In addition to minor changes in the antigenic regions, major changes in the antigenic regions have produced the extremely virulent strains that caused the lethal flu pandemics in 1957 and 1968.

The localization and fine structures of two H5 antigenic sites have been described [13]. Site 1 is an exposed loop comprising residues 140 to 145 (H3 numbering) that corresponds to antigenic site A of H3 [14] and Ca2 of H1 [15], and site 2 comprises two subsites, one (HA1 residues 156 and 157) that corresponds to site B in the H3 subtype [14] and one (HA1 residues 129 to 133) that corresponds to immunodominant site Sa of H1 subtype [15]. Thus, HA is an important surface protein. Therefore, in this communication we have compared human hemagglutinin protein of H5N1 influenza virus A of 1997 outbreak in Hong Kong with homologous HA from non-human sources for mutations involved in change of hydropathy-index. Furthermore, in all strains of non-human origins, mutations in antigenic sites, receptor binding site and appearance of additional glycosylation site in HA have been identified which are important source of pathogenicity of viral infection.

Methodology:
Sequence analysis:
The sequence of hemagglutinin of influenza A virus (bird flu virus) A/Hong Kong/ 156/ (H5N1) from human source was analyzed in this report. The protein sequence of HA is available at NCBI [16] with accession number [GenBank: AAC40508.1]. Protein protein Blast (Blast NCBI Server) [17] was performed at NCBI server, and only 97.46 - 99.46 % similar sequences of hemagglutinin of sub-type H5N1 were selected for the sequence analysis. The sequences of HA selected in this study were given in **Table 1 in supplementary material**.

In these sequences, signal peptide segment was predicted using Signal P3.0 server [18]. The signal peptide segment was deleted from all these sequences prior to aligning using ClustalW [19] multiple alignment tools available at Pole Bioinformatique Lyonnais Network Protein Sequence analysis using weight matrix GONNET for proteins. These alignments were then analyzed for differences in their amino acids at specific positions. Hydrophobicity values were obtained from the tool ProtScale at ExPASy choosing Kyte & Doolittle hydrophobicity scale [20]. Antigenic sites in proteins were predicted using program Antigenic [21].

Motif prediction:
Glycosylation motif was predicted in all protein sequences including A/Hong Kong/ 156/ 97 (H5N1) using Scan Prosite at Expasy Server **[22]**.

Discussion:
Sequence analyses and hydrophathy-index:
In this study, sequence analyses of HA of human and non-human sources showed differences in amino acid mutations at specific positions. In all strains like AAF02306 AAF02309, AAF02304, ABD14807, AAY56367 and AAO46801, V178→I178, C363→G363 and S365→A365 mutations were found conserved **(Table 2 in supplementary material)**. The AAF02306 strain showed V178→I178 mutation, which resulted in the substitution of hydrophobic Val by more hydrophobic Ile. The C363→G363 mutation result in the substitution of more hydrophobic Cys by less hydrophobic and small amino acid Gly. Similarly, in S365 →A365 mutation, hydrophilic Ser is replaced by hydrophobic Ala. In AAF02309 strain besides above three mutations, there was one additional mutation R162→K162. This mutation is an example of synonymous mutation. In AAF02304 strain, there are two more mutations, namely A156→S156 and K368→Q368. Here hydrophobic Ala 156 was replaced by hydrophilic Ser 156, contrary to this, in other there was synonymous mutation (K368→Q368). Similarly, the strain ABD14807 has four additional mutations. Mutation K368→Q368 was found common in ABD14807 and AAF02304 strains. The remaining three mutations in ABD14807 strain are A156 → T156, T320→Ala320 and N387→D387. In the first mutation, there was substitution of hydrophobic Ala by hydrophilic Thr and vice-versa in second mutation, therefore hydropathy-index changed. In the last mutation, there was replacement of Asn by Asp causing synonymous mutation. In AAY56367 strain there are eight additional mutations. The A156→T156 and N387→ D387 mutations were found common to both ABD14807 and AAY56367 strains. The R35→K35 and N94→D94 mutations are synonymous mutations. The I71→ T71 and L138→H138 mutations showed substitution of hydrophobic amino acid by hydrophilic amino acid. Conversly, T263→A263 mutation showed substitution of hydrophilic amino acid by hydrophobic amino acid. Finally, N409→S409 mutation showed substitution of less hydrophilic Asn by more hydrophilic Ser. Hence the hydropathy-index decreased. There are eleven additional mutations in AAO46801 strain. The mutations R35→K35, N94→D94, A156→T156, T263→A263 and N387→D387 are common to both AAY56367 and AAO46801 strains. In P217→S217 mutation, there is substitution of hydrophobic Pro to hydrophilic Ser. Thus, the hydropathy-index decreased. The Val 80→G80 mutation shows substitution of more hydrophobic Val by less hydrophobic and small amino acid Gly. Rest of the four mutations is synonymous mutation. From above data, it is evident that V178→I178, C363→G363 and S365→A365 mutations were found conserved among all strains of HA. Further, it was also found that those mutations which changes hydropathy-index of HA, is one important source of pathogenesis of viral infection. The hemagglutinin protein shows species dependent differences.

Antigenicity and receptor binding site:
Table 3 (see supplementary material) shows the antigenicity profile of human hemagglutinin (AAC40508). Three potential antigenic sites in non-human hemagglutinin with high score are 171-182,543-549 and 28-141 **(Table 4 in supplementary material)**. Rest of the antigenic sites present in non-human sources are similar to human HA, therefore these are not shown in **Table 4 in supplementary material**. The antigenic segment comprising residues 129 to 133 correspond to immunodominant site SA of the H1 subtype **[15]**. Therefore, antigenic segment (128-141) predicted by computer program (Antigenic) in human HA is an important antigenic site. In all strains of non-human origin, because of two mutations C363→G363 and S365→A365, antigenic segment 361-366 was found missing. In AAF02306 strain, Val 178 → Ile 178 mutation lies at the antigenic site predicted by computer program **(Table 3 in supplementary material)**. In

AAF02304 strain besides V178 → Ile 178 mutation, A156→S156 mutation also occurred that lies at the antigenic subsite of site 2(HA1 residues 156 and 157) which corresponds to site B in the H3 subtype **[14]**. Similarly, in ABD14807 strain V178 → Ile 178 and A156→T156 mutations occurred in antigenic sites as mentioned above. In AAY56367 strain, in addition to two conserved mutations (V178 → Ile 178, A156→S156), two more mutations were observed at the antigenic sites. They are mutations I71→T71 and L138→H138. In AAO46801 strain, three more mutations N45→D45, V80→G80 and R140→K140 were found that lie at the antigenic sites **(Table 3 in supplementary material)**. The above data conclusively showed that there are mutations at the antigenic sites. Therefore, these strains will escape antibody binding and viral neutralization. In HA the receptor-binding site comprises three structural elements, namely an alpha-helix (190-helix, HA1 188 to 190) and two loops (130-loop, HA1 134 to 138, and 220-loop, HA1 221 to 228). Interestingly, in AAY56367 strain mutation L138→H138 lies at the receptor-binding site in 130-loop. Here, the receptor binding site also overlaps the antigenic site. Therefore, this amino acid substitution (L138→H138) will influence both the specificity of receptor recognition as well as antibody binding.

Glycosylation site:
There are seven potential glycosylation sites in human HA **(Table 5 in supplementary material)**. Our sequence analysis data showed that these glycosylation sites varied, both in number and location in different HA subtypes (H1 to H14) and even among variants of the same subtype. Similar conclusions have also been drawn in other studies **[23, 24]**. Experiment on the natural epidemic strain A /VIC/3/75, which has new oligosaccharide at 63, binds antibody produced against 1968 strain (two epidemic earlier) when grown in tunicamycin and lacking the oligosaccharide at 63. This proves that carbohydrate, which is host specific can mask surfaces of the HA from the immune system **[25]**. Thus, the oligoscacharides present at glycosylation sites in human HA **(Table 5 in supplementary material)** will mask portions of the surface of human hemagglutinin; hence these regions will not be recognized by host immune system. Because of mutations, A156→S156 and A156→T156, an additional glycosylation site appeared at the amino acid position 154-157 in AAF02304 **(Table 5 in supplementary material)** and ABD14807, AAY56367, AAO46801 strains respectively. Since this glycosylation site is present at the antigenic site, therefore it will mask the antigenic segment of these strains. Hence, hemagglutinin of these viral strains will be antigenically inactive and therefore will be resistant to antibody binding and viral neutralization.

Conclusion:
These results suggest that several factors contribute to the pathogenesis of viral infection. For example mutations in HA which changes hydropathy-index is one important source of pathogenicity. Other important sources of antigenic variations are mutations at the antigenic sites of surface viral protein hemagglutinin, addition of glycosylation site at the antigenic site and mutation at the receptor-binding site. Synthetic peptides vaccines, corresponding to these antigenic sites should be synthesized to combat bird flu infection. Thus, *in silico* approach can serve as an important tool in delineating the disease pathogenesis.

References:
[1] S Padhi *et al.*, *Indian J. Med. Microbiol.* **22**:143 (2004) [PMID: 17642720]
[2] HK Pradhan *et al.*, *Arch. Virol.* **152:**1637 (2007) [PMID: 17558541]

[3]   Y Suzuki, M Nei, *Mol. Biol. Evol.* **19:**501 (2002) [PMID: 11919291]

[4]   RD Slemons *et al.*, *Avian. Dis.* **18:**119 (1974) [PMID: 4205344]

[5]   RG Webster *et al.*, *Virology* **84**:268 (1978) [PMID: 23604]

[6]   VS Hinshaw *et al.*, *Can. J. Microbiol.* **26**:622 (1980) [PMID: 7397605]

[7]   JN Couceiro *et al.*, *Virus. Res.* **29:**155 (1993) [PMID: 8212857]

[8]   T Ito *et al.*, *J. Virol.* **72:**7367 (1998) [PMID: 9696833]

[9]   AU Khan, *Bioinformation* **1:**132 (2006) [PMID: 17597864]

[10]   H Pearson *et. al.,* *Nature* 424:121 (2003) [PMID:12853923]

[11]   A Tamanna, *In Silico Biol.* **6:** 161 (2006) [PMID: 16922680]

[12]   JJ Skehel *et al.*, *Proc. Natl. Acad. Sci. USA* **81:**1779 (1984) [PMID: 6584912]

[13]   YA Smirnov *et al.*, *J. Gen. Virol.* **83:**2497 (2002) [PMID: 12237433]

[14]   DC Wiley et *al., Nature* **289:** 373 (1981) [PMID: 6162101]

[15]   AJ Caton *et al., Cell* **31:**417 (1982) [PMID: 6186384]

[16]   http://www.ncbi.nlm.nih.gov/

[17]   http://www.ncbi.nlm.nih.gov/BLAST/

[18]   http://www.cbs.dtu.dk/services/SignalP/

[19]   http://npsa-pbil.ibcp.fr/

[20]   http://www.expasy.org/tools/protscale.html

[21]   http://liv.bmc.uu.se/cgi-bin/emboss/antigenic

[22]   http://www.expasy.ch/tools/scanprosite/

[23]   YS Kawaoka *et al.*, *Virology* **179:**759 (1990) [PMID: 2238469]

[24]   E Nobusawa *et al.*, *Virology* **182:**475 (1991) [PMID: 2024485]

[25]   DC Wiley, JJ Skehel, *Ann. Rev. Biochem* **56:**365 (1987) [PMID: 3304138]

Supplementary material

Table 1: HA protein sequence dataset

ID	Sequence name	Gene Bank AC	Protein sequence	Place of isolate	Source
1	Hemagglutinin [Influenza A virus(A/Duck/Hong Kong/p46/97 (H5N1))]	AAF02306	552	Hong Kong	Duck
2	Hemagglutinin [Influenza A virus (A/Silky Chicken/ HongKong/p17/97(H5N1))]	AAF02309	552	Hong Kong	Silky chicken
3	Hemagglutinin [Influenza A virus (A/Chicken/Hong Kong/y385/97 (H5N1))]	AAF02304	552	Hong Kong	Chicken
4	Hemagglutinin [Influenza A virus(A/goose/Vietnam/ 3/05(H5N1))]	ABD14807	552	Vietnam	Goose
5	Hemagglutinin [Influenza A virus (A/chicken/China/ 1/02(H5N1))]	AAY56367	552	China	Chicken
6	Hemagglutinin [Influenza A virus (A/Pheasant/Hong Kong/FY155/01-MB(H5N1))]	AAO46801	552	Hong Kong	Pheasant

Table 2: Amino acid mutations in hemagglutinin of different strains at specific positions and their physico-chemical properties

Residue position	Residue change	Chemical property change	Hydrophobicity change (Kyte Doolittle)
A/Duck/Hong Kong/p46/97			
178	V→I	Hydrophobic→Hydrophobic	-1.700→-0.033
363	C→G	Hydrophobic→Hydrophobic	-0.433→-1.089
365	S→A	Hydrophilic→Hydrophobic	-0.478→-1.133
A/Silky Chicken/Hong Kong/ p17/97			
162	R→K	Hydrophilic→Hydrophilic	-0.900→-1.633
178	V→I	Hydrophobic→Hydrophobic	-1.700→-0.033
363	C→G	Hydrophobic→Hydrophobic	-0.433→-1.089
365	S→A	Hydrophilic→Hydrophobic	-0.478→-1.133
A/Chicken/Hong Kong/y385/97			
156	A→S	Hydrophobic→Hydrophilic	-0.000→-1.33
178	V→I	Hydrophobic→Hydrophobic	-1.700→-0.03
363	C→G	Hydrophobic→Hydrophobic	-0.433→-1.09
365	S→A	Hydrophilic→Hydrophobic	-0.478→-1.13
368	K→Q	Hydrophilic→Hydrophilic	-1.11→-1.467
A/goose/Vietnam/3/05			
156	A→T	Hydrophobic→Hydrophilic	-0.000→-1.32
178	V→I	Hydrophobic→Hydrophobic	-1.700→-0.033
320	T→A	Hydrophilic→Hydrophobic	-0.967→-1.767
363	C→G	Hydrophobic→Hydrophobic	-0.433→-1.089
365	S→A	Hydrophobic→Hydrophobic	-0.478 → 1.133
368	K→Q	Hydrophilic→Hydrophilic	-1.11→-1.467
387	N→D	Hydrophilic→Hydrophilic	-1.500→-0.556
A/chicken/China/1/02			
35	R→K	Hydrophilic→Hydrophilic	-0.889→-1.156
71	I→T	Hydrophobic→Hydrophilic	+0.844→0.756
94	N→D	Hydrophilic→Hydrophilic	-0.756→ -1.089
138	L→H	Hydrophobic→Hydrophilic	-2.11→ -0.922
156	A→T	Hydrophobic→Hydrophilic	-0.00→-1.322
178	V→I	Hydrophobic→Hydrophobic	-1.70→-0.033
263	T→A	Hydrophilic→Hydrophobic	-0.467→-0.567
363	C→G	Hydrophobic→Hydrophobic	-0.433→-1.089
365	S→A	Hydrophilic→Hydrophobic	-0.478→-1.133

387	N→D	Hydrophilic→Hydrophilic	-1.500→-0.56
409	N→S	Hydrophilic→Hydrophilic	-0.10→-1.811
A/Pheasant/HongKong/FY155/01-MB			
35	R→K	Hydrophilic→Hydrophilic	-0.889→-1.156
45	N→D	Hydrophilic→Hydrophilic	+0.300→+0.16
80	V→G	Hydrophobic→Hydrophobic	+0.233→-0.589
94	N→D	Hydrophilic→Hydrophilic	-0.756→-1.089
126	D→E	Hydrophilic→Hydrophilic	-0.844-→-1.344
140	R→K	Hydrophilic→Hydrophilic	-1.389→+0.07
156	A→T	Hydrophobic→Hydrophilic	-0.00→-1.322
178	V→I	Hydrophobic→Hydrophobic	-1.700→-0.033
212	E→K	Hydrophilic→Hydrophilic	-0.378→-0.100
217	P→S	Hydrophobic→Hydrophilic	+0.867→-0.367
263	T→A	Hydrophilic→Hydrophobic	-0.467→-0.567
363	C→G	Hydrophobic→Hydrophobic	-0.433→-1.089
365	S→A	Hydrophilic→Hydrophobic	-0.478→-1.133
387	N→D	Hydrophilic→Hydrophilic	-1.500→ -0.556

Table 3: Predicted antigenic sites in human hemagglutinin (A/Hong Kong/156/97)

Amino acid Position	Antigenic Segment	Total no of amino acid residues	Score
43->549	NGSLQCR	7	1.181
128->141	SSGVSSACPYLGRS	14	1.158
514->541	YQILSIYSTVASSLALAIMVAGLSLWMC	28	1.148
291->315	FHNIHPLTIGECPKYVKSNRLVLAT	28	1.14
39->60	GKLCDLNGVKPLILRDCSVAGW	22	1.135
144->151	FRNVVWLI	8	1.132
426->432	AELLVLM	7	1.125
249->259	APEYAYKIVKK	11	1.112
23->32	NVTVTHAQDI	10	1.11
446->455	KNLYDKVRLQ	10	1.102
206->213	NQRLVPEI	8	1.096
100->108	ELKHLLSRI	9	1.096
468->476	FEFYHKCDN	9	1.092
86->93	ANDLCYPG	8	1.09
197->203	YISVGTS	7	1.084
418->424	FLDVWTY	7	1.079
70->83	FINVPEWSYIVEKA	14	1.079
361->366	GSCYSA	6	1.079
113->120	KIQIIPKS	8	1.068
504->509	SGVKLE	6	1.065
488->493	DYPQYS	6	1.063
333->339	FGAIAGF	7	1.053
275->281	NTKCQTP	7	1.052
187->193	QTKLYQN	7	1.026

Table 4: Predicted antigenic sites in hemagglutinin of non-human sources

Strain	Acc.No	Location	Antigenic Segment	Score
A/Duck/HongKong/p46/97	AAF02306	Hong Kong	543->549	1.181

			128->141	1.158
			171->178	1.15
			361->366	(missing)
A/SilkyChicken/HongKong/p17/97	AAF02309	Hong Kong	543->549	1.181
			128->141	1.158
			171->178	1.15
			361->366	(missing)
A/Chicken/Hong Kong/y385/97	AAF02304	Hong Kong	543->549	1.181
			128->141	1.158
			171->178	1.15
			361->366	(missing)
A/goose/Vietnam/3/05	ABD14807	Vietnam	543->549	1.181
			128->141	1.158
			171->178	1.15
			361->366	(missing)
A/chicken/China/1/02	AAY56367	China	543->549	1.181
			128->141	1.158
			171->178	1.15
			361->366	(missing)
A/Pheasant/HongKong/FY155/01-MB	AAO46801	Hong Kong	543->549	1.181
			128->141	1.158
			171->178	1.15
			361->366	(missing)

Table 5: Predicted glycosylation site in human hemagglutinin (A/Hong Kong/156/97(H5N1) and in non-human source(A/Chicken/Hong Kong /y385/97)

Glycosylation Position	Glycosylation motif
A/Hong Kong/156/97	
10 – 13	NNST
11 – 14	NSTE
23 – 26	NVTV
165 – 168	NNTN
286 – 289	NSSM
484 – 487	NGTY
543 – 546	NGSL
A/Chicken/Hong Kong /y385/97	
10 – 13	NNST
11 – 14	NSTE
23 – 26	NVTV
154– 157	NSSY
165 – 168	NNTN
286 – 289	NSSM
484 – 487	NGTY
543 – 546	NGSL