International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 2, No. 5 (2013): 245-248
Research Article
Open Access

ISSN: 2320-0634

Hemagglutinin: A Comparative *In Silico* Study on Different Strains of Influenza A H5N1 Virus

Manish Kumar*

Shri Venkateshwara University, Uttar Pradesh, India

*Corresponding author: Manish Kumar; e-mail: bioinfomoney@gmail.com

Received: 20 August 2013 Accepted: 25 August 2013 Online: 01 September 2013

ABSTRACT

The importance of influenza viruses as worldwide infectious agents is well recognized. Specific mutations and evolution in influenza viruses is difficult to predict. We studied specific mutations in Hemagglutinin (HA) of H5N1 influenza A virus together with properties associated with it using prediction tools developed in Bioinformatics. Changes in hydrophobicity, polarity and secondary structure at the site of mutations were noticed and documented to gain insight towards its infection.

Keywords: hemagglutinin (HA), mutation, influenza, hydrophobicity

INTRODUCTION

Continuous outbreaks of the highly pathogenic H5N1 avian influenza A virus in Asia has resulted in an urgent effort to improve current diagnostics tools to aid containment of the virus and lower the threat of a influenza pandemic. In recent years, the pathogenic H5N1 subtype of avian influenza A has been reported to cross species barrier and infect humans in Hong Kong during the 1997 and 2003 outbreaks [1-3]. There are 16 different HA subtypes and 9 different NA subtypes together forming different combinations. Among them the highly pathogenic are avian H5N1 viruses, which caused 18 confirmed infections and six deaths in Hong Kong during 1997 and 2 cases and 1 death in 2003 [4]. Thus, the H5N1 avian influenza A virus is a known danger to human health across the globe.

Hemagglutinin protein is the receptor-binding and membrane fusion glycoprotein of influenza virus and the target for infectivity-neutralizing antibodies [5,6]. The entire hemagglutinin protein (HA) from the H5N1 is composed of 568 amino acids, with a molecular weight of 56 kDa. The HA molecule consists of HA1 and HA2 subunits, with the HA1 subunit mediating initial contact with the cell membrane and HA2 being responsible for membrane fusion [5].

Therefore, in addition to containment procedures, detection of the cause of changes in the viral genome plays an important role in controlling the spread of the virus. Mutations in influenza viral proteins not only involve change in polarity or hydrophobicity but also the propensity of each residue to stabilize the secondary structures. In the present study, comparison of sequences of Hemagglutinin Protein (HA) of Influenza A H5N1 virus was carried out to estimate mutations between different strains.

MATERIALS AND METHODS

Sequence analysis

Hemagglutinin (HA) of Influenza A virus subtype (A/Ck/Thailand/1/2004)

Continuous outbreaks of the highly pathogenic H5N1 avian influenza A in Asia has resulted in an urgent effort to improve current diagnostics tools to aid containment of the virus and lower the threat of a influenza (H5N1). The HA protein sequence from GenBank [07] was used in this analysis [GenBank Accession number: AAT73266]. Sequences similar to the HA of A/Ck/Thailand/1/2004(H5N1) were extracted from GenBank file influenza.faa at the FTP site ftp://ftp.ncbi.nih.gov/genomes/INFLUENZA/ (Table 1). We manually identified mutations in these

sequences. Hydropathy: Change in Hydrophobicity values for each mutated position was identified using PROTSCALE at expasy [08] with Kyte & Doolittle hydrophobicity scale [9].

Table 1. HA protein sequence dataset is described with the accession number, source, country and name of the sequence.

S. No	Acc.No.	Country	Source	Name
1	AAW80719	Vietnam	Quail	A/quail/Vietnam/36/04
2	ABO10181	Vietnam	Unknown	A/Viet Nam/JP178/2004
3	ABQ09853	Vietnam	Duck	A/Duck/Viet Nam/367/2005
4	AAZ29974	Thailand	Chicken	A/chicken/Ratchaburi/Thailand/CU-68/04
5	ABQ11015	Sukhothai	Chicken	A/chicken/Sukhothai/NIAH6-3-0018/2005
6	ABQ11259	Thailand	Pigeon	A/pigeon/Thailand/VSMU-25-BKK/2005
7	ABC66582	Vietnam	Mallard Duck	A/Mallard duck/Vietnam/133/2004
8	ABC69224	Thailand	Quail	A/quail/Thailand/Nakhon Pathom/QA-161/2005
9	ABC66578	Vietnam	Duck	A/duck/Vietnam/S654/2005
10	ABA70758	Belgium	Crested Eagle	A/crested eagle/Belgium/01/2004
11	ABE97616	Vietnam	Quail	A/quail/Vietnam/282/2005
12	AAT73281	Vietnam	Chicken	A/Ck/Viet Nam/38/2004
13	ABE97610	Vietnam	Wild Bird	A/wild bird/Vietnam/434/2005
14	ABL67770	Nakhonsawan	Open-billed Stork	A/open-billed stork/Nakhonsawan/BBD0104F/2004

Table 2. Comparison of mutations in the HA protein in A/Ck/Thailand/1/2004(H5N1) with other strains in the dataset giving its specific position and changes in secondary structure and associated properties. H - Alpha helix, E - Extended strand and C - Random coil.

Strain	Base position of mutation	Change of residue in (A/Ck/Thailand/1/2004)	Change in properties	Change in secondary structure	Change in Hydrophobicity using ProtScale (Kyte & Doolittle)
A/quail/Vietnam/36/04	191	H→L	Hydrophilic→ Hydrophobic	E	1.133→ 1.344
A/Viet Nam/JP178/2004	191	H→L	Hydrophilic→ Hydrophobic	E	1.133→ 1.344
A/Duck/Viet Nam/367/2005	149	A→S	Hydrophobic→ Hydrophilic	C	1.111→ 0.822
	191	H→L	Hydrophilic→Hydrophobic	E	1.133→ 1.344
A/chicken/Ratchaburi/Thailand/CU-68/04	87	V→I	Hydrophobic	C	-0.211→ -0.178
	133	F→I	Hydrophilic→ Hydrophobic	C→ E	-0.300→ -0.111
	341	K→R	Hydrophilic	C	-3.756→ -3.822
A/chicken/Sukhothai/NIAH6-3-0018/2005	102	I→V	Hydrophobic	C	-0.378→ -0.411
	387	S→T	Hydrophilic	H	-1.511→ -1.500
A/pigeon/Thailand/VSMU-25-BKK/2005	102	A→V	Hydrophilic→ Hydrophobic	C	-0.678→ -0.411
	198	K→N	Hydrophilic	C	-1.200→ -1.156
	341	K→R	Hydrophilic	C	-3.756→ -3.822
A/Mallard duck/Vietnam/133/2004	191	H→L	Hydrophilic→ Hydrophobic	E	1.133→ 1.344
A/quail/Thailand/Nakhon Pathom/QA-161/2005	102	A→V	Hydrophilic→ Hydrophobic	C	-0.678→ -0.411
	233	P→S	Hydrophilic	C	-0.456P→ -0.367
	341	K→R	Hydrophilic	C	-3.756→ -3.822
A/duck/Vietnam/S654/2005	149	A→S	Hydrophobic→ Hydrophilic	C	1.111→ 0.822
	191	H→L	Hydrophilic→ Hydrophobic	E	1.133→ 1.344
A/crested eagle/Belgium/01/2004	157	P→S	Hydrophilic	C	-1.156→ -1.067
	189	V→L	Hydrophobic	E	0.467→ 0.422
	341	K→R	Hydrophilic	C	-3.756→ -3.822
	375	K→E	Hydrophilic	C	-2.189→ -2.144
A/quail/Vietnam/282/2005	5	H→V	Hydrophilic→ Hydrophobic	H → E	1.456→ 1.711
	191	H→L	Hydrophilic→ Hydrophobic	E	1.133→ 1.344
A/Ck/Viet Nam/38/2004	191	H→L	Hydrophilic→ Hydrophobic	E	1.133→ 1.344
	291	D→N	Hydrophilic	C	-1.311→ -1.311
	135	Q→K	Hydrophilic	C	-0.322→ -0.367
A/wild bird/Vietnam/434/2005	149	A→S	Hydrophobic→ Hydrophilic	C	1.111→ 0.822
	157	P→S	Hydrophilic	C	-1.156→ -1.067
	191	H→L	Hydrophilic→ Hydrophobic	E	1.133→ 1.344
	527	V→I	Hydrophobic	C	-0.100→ -0.067
A/open-billed stork/Nakhonsawan/BBD0104F/2004	341	K→R	Hydrophilic	C	-3.756→ -3.822

Table 3. Domains/motifs in Hemagglutinin protein are given. Domains and motifs in the HA protein of A/Ck/Thailand/1/2004 (H5N1) showing the site name, its position on the sequence and the domain directing the site

HA Protein		
Site	**Position**	**Domain**
N-glycosylation site Potential N-glycosylation sites are specific to the consensus sequence Asn-Xaa-Ser/Thr. It must be noted that the presence of the consensus tripeptide is not sufficient to conclude that an asparagine residue is glycosylated, due to the fact that the folding of the protein plays an important role in the regulation of N-glycosylation	26 – 29 27 – 30 39 – 42 170 – 173 181 – 184 302 – 305 500 – 503 559 – 562	NNST NSTE NVTV NSTY NNTN NSSM NGTY NGSL
Casein kinase II phosphorylation site Casein kinase II (CK-2) is a protein serine/threonine kinase whose activity is independent of cyclic nucleotides and calcium. CK-2 phosphorylates many different proteins.	34 – 37 139 – 142 183 – 186 283 – 286 314 – 317 400 – 403 407 – 410	TimE SshE TnqE SelE TigE SiiD TqfE
N-myristoylation site A number of eukaryotic proteins are acylated by the covalent addition of myristate (a C14-saturated fatty acid) to their N-terminal residue via an amide linkage. The sequence specificity of the enzyme responsible for this modification, myristoyl CoA:protein N-myristoyl transferase (NMT), has been derived from the sequence of known N-myristoylated proteins and from studies using synthetic peptides.	79 – 84 146 – 151 288 – 293 299 – 304 347 – 352 350 – 355 358 – 363 362 – 367 377 – 382 560 – 565	GNpmCD GVssAC GNcnTK GAinSS GLfgAI GAiaGF GGwqGM GMvdGW GSgyAA GSlqCR
ATP/GTP-binding site motif A (P-loop) An appreciable proportion of proteins that bind ATP or GTP share a number of more or less conserved sequence motifs. The best conserved of these motifs is a glycine-rich region, which typically forms a flexible loop between a β-strand and an α-helix. This loop interacts with one of the phosphate groups of the nucleotide. This sequence motif is generally referred to as the 'A' consensus sequence or the 'P-loop'.	150 – 157	AcpyqGKS
cAMP- and cGMP-dependent protein kinase phosphorylation site cAMP- and cGMP-dependent protein kinases, both types of kinases appear to share a preference for the phosphorylation of serine or threonine residues found close to at least two consecutive N-terminal basic residues. It is important to note that there are quite a number of exceptions to this rule.	168 – 171	KKnS
Protein kinase C phosphorylation site In vivo, protein kinase C exhibits a preference for the phosphorylation of serine or threonine residues found close to a C-terminal basic residue. The presence of additional basic residues at the N- or C-terminal of the target amino acid enhances the Vmax and Km of the phosphorylation reaction.	175 – 177 239 – 241 324 – 326 387 – 389 395 – 397 497 – 499	TiK SgR SnR TqK TnK SvR
Prenyl group binding site (CAAX box) A number of eukaryotic proteins are post-translationally modified by the attachment of either a farnesyl or a geranyl-geranyl group to a cysteine residue. The modification occurs on cysteine residues that are three residues away from the C-terminal extremity; the two residues that separate this cysteine from the C-terminal residue are generally aliphatic. This Cys-Ali-Ali-X pattern is generally known as the CAAX box.	564 – 567	CRIc

Secondary structure prediction

Secondary structures were predicted using Consensus prediction from DSC, GOR4, PHD, Predator & SOPMA [10-14] at NPS@ (Network Protein Sequence Analysis) server [15]. Motif Search: Domains or motifs were searched using ScanProsite at the Expasy Server [16]. Motifs with high probability of occurrence were included in the search.

RESULTS AND DISCUSSION

Sequence analysis and hydropathy

The HA sequences from different strains were compared manually. From Table 2, the mutation (Histidine→Leucine) at position 191 was found to be common in most of the strains in Table 1 except for Thailand (AAZ29974; ABQ11259; ABC69224), Sukhothai (ABQ11015), Belgium (ABA70758) and Nakhonsawan (ABL67770). In addition, this mutation is non-synonymous and changes the chemical property at

the respective site. Similar changes were observed at H5V (Vietnam- ABE97616), A102V (Thailand- ABQ11259; ABC69224, A149S (Viet Nam- ABQ09853; ABC66578; ABE97610) and F133I (Thailand- AAZ29974). The second most common mutation observed (Lysine→Arginine) at position 341 and is observed in strains Table 1 for Thailand (AAZ29974, ABQ11259, ABC69224), Belgium (ABA70758) and Nakhonsawan (ABL67770). This mutation is synonymous and will not change the chemical property at the site. Likewise, in Thailand- AAZ29974 (V87I); ABQ11259 (K198N); ABC69224 (P233S), in Belgium- ABA70758 (P157S, V189L, K375E), in Vietnam- AAT73281 (D291); ABE97610 (Q135K, P157S, V527I) and in Sukhothai- ABQ11015 (I102V, S387T) synonymous mutations were observed (Table 2).

Secondary structure prediction

We also analysed the secondary structure changes caused by these mutations. The mutation F133I in Thailand- AAZ29974 showed a change from coil to extended strand.

Domain/Motif Search

Domains, found in the strain (A/Ck/Thailand/1/2004) of HA protein, are given in Table 3. The mutation A149S in Viet Nam- ABQ09853, ABC66578 and ABE97610 and D291N in Viet Nam- AAT73281are found at N-myristoylation site. Further, the mutation P157S in Belgium- ABA70758 and in Vietnam- ABE97610 is found at ATP/GTP-binding site motif A (P-loop) and the mutation S387T in Sukhothai- ABQ11015 is found at Protein kinase C phosphorylation site.

Influenza A has the fastest mutation rate at $6.7x10^{-3}$ mutations per residue per year. Through mutations, antigenic drift enables influenza strains to replicate without being destroyed by the host defence mechanisms. [17] It is important to predict mutations in influenza virus to know specific mutations for increased transmissibility of the virus among humans. In the present communication, we analysed mutations in protein HA of Influenza A virus from chicken (H5N1) and studied changes associated with it. At least one mutation was noticed in each strain available in the dataset when compared with the reference sequence of A/Ck/Thailand/1/2004.

CONCLUSION

Mutations in the HA protein from different strains were studied to understand their relationship with susceptibility and infection by the virus. The analysis helped to document the location of specific mutations and the changes in properties related with it. Changes in the predicted secondary structures were also observed at the site of a few mutations. The locations of mutations at different PROSITE motifs were documented in this study.

Acknowledgements

The author is grateful to Vandna Chawla, Ph.D Scholar, Studio of Structural and Computational Biology, Palampur, Himachal Pradesh for her help and support to carry out this work.

REFERENCES

1. P. K. Chan, (2002) *Clin. Infect. Dis.,* 34:S58
2. J. S. M. Peiris, *et al.,* (2004) *Lancet*, 363:617
3. K. Y. Yuen, *et al.,* (1998) *Lancet*, 351:467
4. K. Y. Yuen and S. S. Y. Wong, (2005) *Hong Kong Med. Jour.*, 11:189
5. Chizmadzhev YA, (2004) Bioelectrochemistry, 63:129–136
6. Skehel JJ, Wiley DC, (2000) 69:531-69
7. http://www.ncbi.nlm.nih.gov/
8. http://www.expasy.ch/tools/protscale.html
9. J. Kyte and R. Doolittle, (1982) *J. Mol. Biol.* 157: 105-132
10. R. D.King and M.J.E Sternberg, (1996) *Protein Sci*, 5, 2298-2310
11. J. Garnier, D. ~J. Osguthorpe and B. Robson, (1978) *Journal of Molecular Biology*, 120, 97-120
12. B. Rost, and C. Sander, (1993) *Journal of Molecular Biology*, 232, 584-599
13. D. Frishman and P. Argos (1996) Protein Eng, 9(2): 133-142
14. C. Geourjon, and G. Deleage, (1994) *Protein Engineering*, 7, 157-16
15. C. Combet., C. Blanchet, C. Geourjon and G. Deléage, (2000) TIBS Vol. 25, No 3 [291]:147-150
16. http://ca.expasy.org/tools/scanprosite/
17. J. T. M. Voeten, *et al.,* (2000) *J. of Virology,* 74:6800

*****