

Matrix protein 1: A comparative in silico study on different strains of influenza A H5N1 Virus

Tamanna Anwar¹, Sunil K. Lal³ and Asad U. Khan^{1,2*}

¹Distributed Information Sub-centre; ²Interdisciplinary Biotechnology Unit Aligarh Muslim University, Aligarh 202002, India; ³International Center for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110 067, India; Asad U Khan* - Email: huzzi99@hotmail.com; Phone: +91 571 2723088; Fax: +91 571 2721776;

* Corresponding author

received November 13, 2006; accepted November 21, 2006; published online November 22, 2006

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 matrix protein 1 (M1) 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 mutation were noticed and documented to gain insight towards its infection.

Keywords: matrix protein 1 (M1); mutation; N-myristoylation site

Background:

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 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, 2, 3] There are 16 subtypes of influenza virus A, among them the highly pathogenic are the avian H5N1 viruses which caused 18 confirmed infections and six deaths in Hong Kong during 1997. [4] Thus, the H5N1 avian influenza - A virus is a known danger to human health across the globe.

The influenza A virus M1 protein is the most abundant structural component of the virion, and electron microscopy studies of influenza virus A particles show that M1 forms a shell at the internal surface of the viral membrane. [5, 6, 7] The functions of M1 have been studied extensively, and it has been implicated in a variety of roles in the virus life cycle that include RNA and RNP binding [8-16], transcription inhibition [9, 10, 14, 17, 18], and control of RNP nuclear import and export. [19-24] Additionally, the matrix protein (M1) of influenza virus plays an essential role in viral replication. [25]

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. [26] In the present study, comparison of sequences of Matrix protein (M1) of Influenza A H5N1 virus was carried out to estimate mutations between different strains.

Methodology:

Sequence analysis

Matrix protein (M1) of Influenza A virus sub-type A/Hatay/2004/ (H5N1)

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 M1 protein sequence from GenBank [27] was used in this analysis [GenBank Accession number: CAJ01905]. Sequences similar to the M1 of A/Hatay/2004/(H5N1) were extracted from GenBank file influenza.faa at the FTP site ftp://ftp.ncbi.nih.gov/genomes/INFLUENZA/ (Table 1). We used MAP MUTATION [28] to identify mutations in these sequences. Hydrophobicity: Hydrophobicity values were calculated using PROTSCALE at expasy [29] with the Kyte & Doolittle hydrophobicity scale.

Secondary structure prediction

Secondary structures were predicted using NNpredict. [30] Motif Search: Domains or motifs were searched using ScanProsite at the Expasy Server. [31] Motifs with high probability of occurrence were included in the search.

Results:

Sequence analysis and hydrophobicity

The M1 sequences from different strains were compared using the computer program MAP MUTATION. The mutation (valine to isoleucine) at position 205 was found to be common in almost all strains in Table 1 except for Thailand (ABC69235; AAT70223; ABB43126) and VietNam (AAZ72675; AAZ72687; AAZ72665). In addition, the valine to isoleucine mutation was observed at position 14 in China (ABC66622). This mutation is synonymous and will not change the chemical property at the site. However, non synonymous mutations A83S in Thailand - ABC69235 and S116A in China - AAT70589 were observed. This mutation

changes the chemical property at the respective sites. Similar changes were observed at I168T, I14T and I168T in Thailand - ABB43126, China - AAY53536 and China - AAZ16338, respectively. Likewise, M59I in Russia - ABB43070 and T37A in China - AAT70593 were observed. The V138M mutation in VietNam - AAZ72687 is found at the N-myristoylation site. A S13F mutation was also found in VietNam - AAZ72665.

Secondary structure prediction

We also analysed the secondary changes caused by these mutations. The mutations, I59M in Russia - ABB43070, I14T in China - AAY53536 and V138M in VietNam - AAZ72687 showed a change from coil to extended strand. We have presented our analysis data from secondary structure prediction at the website <http://www.geocities.com/amubioinfo/InfluenzaAVirus.htm>. Domain/motif Search: Domains found in the different strains of M1 protein are given in Table 2. This information is documented for future studies.

Discussion:

Influenza A has the fastest mutation rate at 6.7×10^{-3} mutations per residue per year. [28] Through mutations, antigenic drift enables influenza strains to replicate without

being destroyed by the host defence mechanisms. [32] 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 M1 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/Hatay/2004.

McCullers and colleagues demonstrated that a single point mutation at position 221 in the C-terminal domain of the M1 protein (Asparagine to Serine) was responsible for acquisition of virulence and lethality. [33] It was observed that every mutation may not cause changes in the secondary structure. Moreover, the residue that was itself not undergoing change in secondary structure was capable of changing the conformation of nearby residues. In our previous study, we reported that mutations in predicted secondary structure may be due to changes in side chain conformation. [26] It should be noted that further studies on Influenza H5N1 virus is required to unfold the mystery of mutation.

S.No.	Acc. No.	Country	Source	Name
i.	AAV48543	China	Chicken	A/chicken/Hubei/489/2004(H5N1)
ii.	ABC66622	China	Duck	A/duck/Shantou/4610/2003(H5N1)
iii.	ABC69235	Thailand	Chicken	A/chicken/Thailand/Nontaburi/CK-162/2005(H5N1)
iv.	AAZ72675	VietNam	Chicken	A/chicken/VietNam/LA024/2004(H5N1)
v.	AAT70223	Thailand	Leopard	A/leopard/Suphanburi/Thailand/Leo-1/04(H5N1)
vi.	AAT70589	China	Duck	A/Dk/HN/5806/2003(H5N1)
vii.	ABB43070	Russia	Duck	A/duck/Novosibirsk/56/2005(H5N1)
viii.	AAT12062	China	Duck	A/duck/Zhejiang/52/2000(H5N1)
ix.	AAZ16345	China	Duck	A/Duck/Fujian/1734/05(H5N1)
x.	ABB43126	Thailand	Cat	A/cat/Thailand/KU-02/04(H5N1),
xi.	AAT70593	China	Chicken	A/Ck/ST/4231/2003(H5N1),
xii.	AAZ53536	China	Chicken	A/chicken/Yunnan/K001/2004(H5N1),
xiii.	AAZ16338	China	Chicken	A/Chicken/Yunnan/493/05(H5N1),
xiv.	ABC74395	China	Chicken	A/chicken/Hebei/326/2005(H5N1),
xv.	AAZ72687	VietNam	DN-33	A/Viet Nam/DN-33/2004(H5N1),
xvi.	AAZ72665	VietNam	Chicken	A/chicken/VietNam/VL008/2004(H5N1)

Table 1: M1 protein sequence dataset is described with accession number, source, country and name of the sequence

Conclusion:

Mutations in the M1 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 predicted secondary structures were also observed at the site of few mutations. The locations of mutations at different PROSITE motifs were documented in this study.

Acknowledgement:

The authors are grateful to Prof. M Saleemuddin for providing facilities to carryout this work. We appreciate the assistance of staff of the Distributed information sub-centre for their technical help. Department of Biotechnology, Ministry of Science and Technology, Government of India is acknowledged for the financial support.

Strain	Base Position of Mutation	Change of residue in Hatay/2004	Change in Properties	Change in Secondary Structure	Change in Hydrophobicity (Kyte & Doolittle)
A/chicken/Hubei/489/2004	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/duck/Shantou/4610/2003	14	V → I	Hydrophobic	E	0.978 → 1.011
	205	V → I	Hydrophobic	H	-0.278 → -0.244
A/chicken/Thailand/Nontaburi/CK-162/2005	83	S → A	Hydrophilic → Hydrophobic	C	-0.489 → -0.2
A/chicken/Viet Nam/LA-024/2004	109	Y → F	Hydrophobic	H	-1.244 → -0.789
A/leopard/Suphanburi/Thailand/Leo-1/04	100	F → Y	Hydrophobic	H	0.089 → -0.367
A/Dk/HN/5806/2003	116	S → A	Hydrophilic → Hydrophobic	E	-0.144 → 0.144
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/duck/Novosibirsk/56/2005	59	M → I	Hydrophobic	C → E	1.233 → 1.522
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/duck/Zhejiang/52/2000	27	R → K	Hydrophilic	H	-0.467 → -0.4
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/Duck/Fujian/1734/05	174	K → R	Hydrophilic	H	-1.711 → -1.778
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/cat/Thailand/KU-02/04	168	T → I	Hydrophilic → Hydrophobic	C	-0.356 → 0.222
A/Ck/ST/4231/2003	37	A → T	Hydrophobic → Hydrophilic	C	-0.622 → -0.9
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/chicken/Yunnan/K001/2004	14	T → I	Hydrophilic → Hydrophobic	C → E	0.767 → 1.344
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/Chicken/Yunnan/493/05	168	T → I	Hydrophilic → Hydrophobic	C	-0.356 → 0.222
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/chicken/Hebei/326/2005	85	S → N	Hydrophilic	C	-1.111 → -1.411
	205	V→I	Hydrophobic	H	-0.278 → -0.244
A/Viet Nam/DN-33/2004	109	Y → F	Hydrophobic	H	-1.244 → -0.789
	138	M → V	Hydrophobic	C → E	-0.278 → -0.022
A/chicken/Viet Nam/VL-008/2004	13	F → S	Hydrophobic → Hydrophilic	E	1.711 → 1.311

Table 2: Comparison of mutations in the M1 protein of A/Hatay/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

M1		
Site	Position	Domain
Casein kinase II phosphorylation site	5 - 8	TevE
(CK-2 is a protein serine/threonine kinase whose activity is independent of cyclic nucleotides and calcium. CK-2 phosphorylates many different proteins)	37 - 40	TdIE

Protein kinase C phosphorylation site	70 - 72	SeR
(In vivo, protein kinase C exhibits a preference for the phosphorylation of serine or threonine residues found close to a C-terminal basic residue)	161 - 163	ShR
	185 - 187	TaK
N-myristoylation site	122 - 127	GAlaSC
(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)	129 - 134	GLIyNR
	136 - 141	GTvtTE
	145 - 150	GLvcAT
	220 - 225	GThpNS
	228 - 233	GLrdNL
	224 - 227	NSSA
N-glycosylation site		
(Potential N-glycosylation sites are specific to the consensus sequence Asn-Xaa-Ser/Thr. 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)		

Table 3: Domains/motifs in matrix protein 1 are given. Domains and motifs in the M1 protein of A/Hatay/2004/ (H5N1) showing the site name, its position on the sequence and the domain directing the site

References:

- [01] P. K. Chan, *Clin. Infect. Dis.*, 34:S58 (2002) [PMID: 11938498]
- [02] J. S. M. Peiris, *et al., Lancet*, 363:617 (2004) [PMID: 14987888]
- [03] K. Y. Yuen, *et al., Lancet*, 351:467 (1998) [PMID: 9482437]
- [04] K. Y. Yuen & S. S. Y. Wong, *Hong Kong Med. Jour.*, 11:189 (2005) [PMID: 15951584]
- [05] F. P. Booy, *et al., J. Mol. Biol.*, 184:667 (1985) [PMID: 4046029]
- [06] M. V. Nermut, *J. Gen. Virol.*, 17:317 (1972) [PMID: 4119596]
- [07] I. T. Schulze, *Virology*, 47:181 (1972) [PMID: 4110126]
- [08] F. Baudin, *et al., EMBO J.*, 13:3158 (1994) [PMID: 8039508]
- [09] F. Baudin, *et al., Virology*, 281:102 (2001) [PMID: 11222100]
- [10] C. Elster, *et al., J. Gen. Virol.*, 78:1589 (1997) [PMID: 9225034]
- [11] S. Patterson, *et al., J. Gen. Virol.*, 69:1859 (1998) [PMID: 3404117]
- [12] P. J. Rees & N. J. Dimmock, *J. Gen. Virol.*, 53: 125 (1981) [PMID: 7276909]
- [13] L. Wakefield & G. G. Brownlee, *Nucleic Acids Res.*, 17:8569 (1989) [PMID: 2479906]
- [14] K. Watanabe, *et al., J. Virol.*, 70:241 (1996) [PMID: 8523532]
- [15] Z. Ye, *et al., J. Virol.*, 73:7467 (1999) [PMID: 10438836]
- [16] Z. P. Ye, *et al., J. Virol.*, 63:3586 (1989) [PMID: 2474671]
- [17] Z. P. Ye, *et al., J. Virol.*, 61:239 (1987) [PMID: 2433462]
- [18] A. Y. Zvonarjev & Y. Z. Ghendon, *J. Virol.*, 33: 583 (1980) [PMID: 6893343]
- [19] M. Bui, *et al., J. Virol.*, 70:8391 (1996) [PMID: 8970960]
- [20] X. Huang, *et al., Virolog*, 287:405 (2001) [PMID: 11531417]
- [21] K. Martin & A. Helenius, *Cell*, 67:117 (1991) [PMID: 1913813]
- [22] K. Martin & A. Helenius, *J. Virol.*, 65:232 (1991) [PMID: 1985199]
- [23] G. Whittaker, *et al., J. Virol.*, 70:2743 (1996) [PMID: 8627748]
- [24] G. Whittaker, *et al., J. Virol.*, 69:439 (1995) [PMID: 7983740]
- [25] T. Liu & Z. Ye, *J. Virol.*, 79:1918 (2005) [PMID: 15650216]
- [26] T. Anwar, *et al., In Silico Biology*, 6:0015 (2006) [PMID: 16922680]
- [27] <http://www.ncbi.nlm.nih.gov/>
- [28] T. Anwar & A. U. Khan, *Bioinformatics*, 1:92 (2006)
- [29] <http://www.expasy.org/cgi-bin/protscale.pl>
- [30] <http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html>
- [31] <http://ca.expasy.org/tools/scanprosite/>
- [32] J. T. M. Voeten, *et al., J. of Virology*, 74:6800 (2000) [PMID: 10888619]
- [33] J. A. McCullers, *et al., Virology*, 336:318 (2005) [PMID:15892972]

Edited by P. Kanguane

Citation: Anwar *et al.*, *Bioinformatics* 1(7): 253-256 (2006)

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.