Trends in Bioinformatics

ISSN 1994-7941

science alert

ANSInet
an open access publisher
http://ansinet.com

Trends in Bioinformatics 1 (1): 18-24, 2008
ISSN 1994-7941

In silico Mutation Study of Haemagglutinin and Neuraminidase on Banten Province Strain Influenza A H5N1 Virus

Usman S.F. Tambunan, Oksya Hikmawan and Theofilus A. Tockary
Department of Chemistry, Faculty of Mathematics and Natural Sciences,
University of Indonesia, Depok Campus, Depok 16424, Indonesia

Abstract: The aim of this study was to analyse the mutation possibility of influenza A (H5N1) virus of Banten Province strain. The H5N1 amino acid sequences of haemagglutinin (HA) and neuraminidase (NA) were analysed by multiple alignment method using BLAST tool, ClustalW and Bioedit 7.0.1. programs. The results of analysis were followed up by mutation analysis using Amino Track, secondary structure, post translation modification analysis, phylogenetic tree and homology modelling. From the analysis, five point mutations were revealed, namely change of haemagglutinin at position 185, 272 and 309 and neuraminidase at position 244 and 252. Conserved region prediction were determined to be at position 1-552 of HA and at position 1-39, 41-251 and 253-449 of NA. Secondary structure prediction showed a change in HA structure at position 309, namely coil into helix transformation and in NA structure at position 244, namely coil into extended strand transformation. Change of hydrophobicity happened only to HA at position 185 and 272. Mutation did not have any influence on post translation modification. Phylogenetic tree analysis and homology modelling also did not reveal the creation of a new strain.

Key words: Mutation, H5N1, influenza, haemagglutinin, neuraminidase, *in silico*

INTRODUCTION

Avian influenza is caused by viral infection of Orthomyxoviridae family in the influenza virus A genus. This virus was originally recognised only as the causative agent of fowl plague in 1955. However, in the last few years, the occurrence of highly pathogenic avian influenza A (H5N1) virus also began to threaten human safety after fowl to human transfection cases (Kawaoka *et al*., 1989) and was reported to increase (Horimoto and Kawaoka, 2001). This problem arises due to the virus tendency to mutate and recombine with genetical material of other influenza virus (Anwar *et al*., 2006; Ungchusak *et al*., 2005). Very limited human-to-human transmission of the H5N1 strain was documented in healthcare workers and family members with contact (Katz Jacqueline *et al*., 1999; Bridges *et al*., 2000). For the same reason, some experts are now even fearing the worst possibility of another new strain that might be capable of human to human transfection.

In a recent case of avian influenza outbreak in the Indonesian Province of Banten, several members of a certain family were infected by a presumably new strain of this virus. Some suspected who might actually be the first case of human to human transfection as some of the family members could not recall having direct contact with fowls. This present study is an effort to settle this claim by conducting series of *in silico* examination on the Banten Province strain's haemagglutinin and neuraminidase amino acid sequences. It is expected that may reveal the specific mutation sites of the strain and altering of antigenicity, specificity and pathogenicity of the virus (Glaser *et al*., 2005). As

Corresponding Author: Usman Sumo Friend Tambunan, Department of Chemistry, Faculty of Mathematics and Natural Sciences, University of Indonesia, Depok Campus, Depok 16424, Indonesia Tel: +62 21 727 0027 Fax: +62 21 786 3432

is the case of the most viruses, these three altering parameters are strong indications of a new strain. The reasons for choosing haemagglutinin and neuraminidase for this study are (1) both are well-known as antigenic macromolecules, (2) haemagglutinin's participation in binding to the host cell is strongly connected to specificity and (3) neuraminidase is a virulence factor.

Haemagglutinin and neuraminidase amino acid sequences of the Banten Province strain (A/Indonesia/CDC1032/2007/(H5N1)) are readily available at the Los Alamos National Laboratory website. Both were examined for conserved region, mutation sites, secondary structural change, hydrophobicity and post-translational modification behaviour. These were later subjected to homology modelling and a phylogenetic tree analysis was conducted to check the strain's relationship to other strains.

MATERIALS AND METHODS

This *in silico* study was conducted in 2007 at the Laboratory of Bioinformatics, Department of Chemistry, Faculty of Science, University of Indonesia.

Haemagglutinin and Neuraminidase of Influenza A Virus Sub-Type H5N1 of A/Indonesia/CDC1032/2007 Strain

Haemagglutinin and neuraminidase amino acid sequence of influenza A virus sub-type H5N1 of A/Indonesia/CDC1032/2007 was downloaded in GenBank Flat File (GBFF) format at Los Alamos National Laboratory website (http://www.ncbi.nlm.nih.gov). Other haemagglutinin and neuraminidase amino acid sequences from Indonesian influenza A virus sub-type H5N1 isolates were also collected from this source.

Database Similarity Screening

Amino acid sequences of all Indonesian influenza A (H5N1) virus isolates were screened for 99-100% homology with A/Indonesia/CDC1032/2007 strain using online Basic Local Aligment Search Tool (BLAST) for proteins at National Centre for Biotechnology Information (NCBI) website (http://www.ch.embnet.org/software/BLASTp.html).

Conserved Region Prediction and Mutation Analysis of Influenza A Virus Sub-Type H5N1 of A/Indonesia/CDC1032/2007 Strain

Conserved region is a similar or identical region of sequences. In order to find the conserved region, we used a multiple sequence alignment (MSA) method which aligned three or more biological sequences (protein, DNA, or RNA) using ClustalW and BioEdit 7.0.1 programs. ClustalW program is a multiple sequence alignment program which is created by the European Molecular Biology Laboratory (http://www.ebi.ac.uk/Tools/clustalW/index/html) and BioEdit 7.0.1 program is a biological sequence alignment editor which is created by Tom Hall Ibis Biosciences (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html). ClustalW and BioEdit 7.0.1 programs calculate the best match for the selected sequences and line them up, so that the identities, similarities and differences can be seen. The similar of each amino acid or nucleotide are marked by asteriks (*) and the difference sequences are marked by gab (-). Mutation analysis was conducted using AminoTrack™ toolbox at http://apps.sbri.org/AminoTrack/. AminoTrack™ is a web based tool designed to increase the efficiency with which sequence data is recorded for further analysis. This program is used to identify mutations in viral proteins as these proteins evolve during infection (Mahalanabis *et al*., 2006).

Secondary Structure, Hydrophobicity and Post-Translational Modification Prediction

Secondary structures were predicted using NNPREDICT at http://www.cmpharm.ucsf.edu~nomi/nnpredict.html. Hydrophocity prediction was conducted online using ProtScale at

http://expasy.org/tools/ProtScale.html. Post-translational modifications were searched using ScanProsite at http://expasy.org/tools/ScanProsite.html.

Homology Modeling

Homology modeling was carried out by comparing influenza A virus sub-type H5N1 of A/Indonesia/CDC1032/2007 strain protein with this available database at Protein Data Bank.

Phylogenetic Analysis

Phylogene of influenza virus type A (H5N1) was presented using TreeView program.

RESULTS AND DISCUSSION

Sequence Analysis

Amino acid sequences of influenza A virus sub-type H5N1 of A/Indonesia/CDC1032/2007 strain was downloaded by the GenBank Accession number of ABM90478 for haemagglutinin sequences and ABM90480 for neuraminidase sequences. The 99-100% homology screening yielded 3 homologs of ABM90478 and 4 homologs for ABM90480 (Table 1).

Conserved Region Prediction and Mutation Analysis of Influenza A Virus Sub-Type H5N1 of A/Indonesia/CDC1032/2007 Strain

Conserved regions are determined to be at position 1-552 for haemagglutinin and at position 1-39, 41-251 and 253-449 for neuraminidase. Point mutations are observed at position 183, 184, 185, 272 and 309 for haemagglutinin (Table 2a) and at position 40, 63, 239, 244 and 252 for neuraminidase (Table 2b), however specific mutations occurring only to A/Indonesia/CDC1032/2007 strain are at position 185, 272 and 309 for haemagglutinin and 244 and 252 for neuraminidase. The mutation analysis using AminoTrack™ also extracted information on potential n-glycosylation (PNG and PNG AA) sites. However, none of the above mentioned specific mutation occurred at any potential n-glycosylation site. In addition, all mutation, except at position 40 and 252 for neuraminidase, occurred outside the conserved region.

Secondary Structure, Hydrophobicity and Post-Translational Modification Prediction

In order to better understand what sort of changes this mutations possibly might have on the A/Indonesia/CDC1032/2007 strain, secondary structure, hydrophobicity and post-translational modification in this mutation site are examined. Secondary structures of haemagglutinin and neuraminidase amino acids are presented at Table 3. For haemagglutinin, helix changes into coil structure occur at position 309, while for position 185 and 272, no secondary structure change is detected. For neuraminidase, coil changes into extended strand structure at position 244, while for 252, no secondary structure change is detected. Hydrophobicity is evaluated using ProtScale software. This software displays the polarity of the initial amino acids and amino acids after the mutation (Table 4).

Table 1: A/Indonesia/CDC1032/2007 (H5N1) strain

Antigens	No. GenBank	Isolates
Haemagglutinin	ABI36275	A/Indonesia/CDC523E/2006/(H5N1)
	ABI36307	A/Indonesia/CDC610/2006/(H5N1)
	ABM90434	A/Indonesia/CDC1031/2007/(H5N1)
Neuraminidase	ABI36457	A/Indonesia/CDC699/2006/(H5N1)
	ABL07021	A/Indonesia/CDC938/2006/(H5N1)
	ABM90436	A/Indonesia/CDC1031/2007/(H5N1)
	ABL31768	A/Indonesia/CDC836T/2006/(H5N1)

Table 2a: Mutation analysis of A/Indonesia/CDC1032/2007(H5N1) strain haemagglutinin using AminoTrack

AA Seq Changes								
Sample	183	184	185	272	309			
CDC1032(REF)	N	E	E	S	S			
CDC523E	-	A	A	G	N			
CDC610	D	A	A	G	N			
CDC1031	-	-	A	G	N			
MutMatrix								
Sample	N183D	E184A	E185A	S272G	S309N			
CDC1032(REF)								
CDC523E		1	1	1	1			
CDC610	1	1	1	1	1			
CDC1031			1	1	1			
PNG								
Sample	10	11	23	154	165	286	484	543
CDC1032(REF)	1	1	1	1	1	1	1	1
CDC523E	1	1	1	1	1	1	1	1
CDC610	1	1	1	1	1	1	1	1
CDC1031	1	1	1	1	1	1	1	1
PNG AA								
Sample	10	11	23	154	165	286	484	543
CDC1032(REF)	NNST	NSTE	NVTV	NSTY	NNTN	NSSM	NGTY	NGSL
CDC523E	NNST	NSTE	NVTV	NSTY	NNTN	NSSM	NGTY	NGSL
CDC610	NNST	NSTE	NVTV	NSTY	NNTN	NSSM	NGTY	NGSL
CDC1031	NNST	NSTE	NVTV	NSTY	NNTN	NSSM	NGTY	NGSL

Table 2b: Mutation analysis of A/Indonesia/CDC1032/2007 (H5N1) strain neuraminidase using AminoTrack

AA Seq Changes					
Sample	40	63	239	244	252
CDC1032(REF)	K	V	K	I	T
CDC699	-	-	E	V	P
CDC938	-	I	-	V	P
CDC1031	-	-	-	V	P
CDC836T	T	-	E	V	P
MutMatrix					
Sample	K40T	V63I	K239E	I244V	T252P
CDC1032(REF)					
CDC699			1	1	1
CDC938		1		1	1
CDC1031				1	1
CDC836T	1		1	1	1
PNG					
Sample	68	126	215		
CDC1032(REF)	1	1	1		
CDC699	1	1	1		
CDC938	1	1	1		
CDC1031	1	1	1		
CDC836T	1	1	1		
PNG AA					
Sample	68	126	215		
CDC1032(REF)	NSSL	NGTV	NGSC		
CDC699	NSSL	NGTV	NGSC		
CDC938	NSSL	NGTV	NGSC		
CDC1031	NSSL	NGTV	NGSC		
CDC836T	NSSL	NGTV	NGSC		

Haemagglutinin sequence at position 185, hydrophobic alanine is replaced by hydrophilic glutamic acid, resulting in a non-polar into polar transition, at position 272, glysine is replaced by serine, resulting in non-polar to polar transition and at position of 309, asparagine is replaced by serine, resulting the insignificant polar into very slightly more polar transition. For neuraminidase, both mutations (244 and 252) cause insignificant change of hydrophobicity. Post-translation modification analysis is shown in Table 5, where no modification is observed to occur at any of the mutation sites (Table 5).

Table 3: Secondary structure prediction of A/Indonesia/CDC1032/2007 (H5N1) strain haemagglutinin and neuraminidase

Position	Isolate	Secondary structure
Haemagglutinin		
185	CDC1032 (Ref)	Coil
	CDC523E	Coil
	CDC610	Coil
	CDC1031	Coil
272	CDC1032 (Ref)	Coil
	CDC523E	Coil
	CDC610	Coil
	CDC1031	Coil
309	CDC1032 (Ref)	Helix
	CDC523E	Coil
	CDC610	Coil
	CDC1031	Coil
Neuraminidase		
244	CDC1032 (Ref)	Extended strand
	CDC699	Coil
	CDC938	Coil
	CDC1031	Coil
	CDC836T	Coil
252	CDC1032 (Ref)	Coil
	CDC699	Coil
	CDC938	Coil
	CDC1031	Coil
	CDC836T	Coil

Table 4: Hydrophobicity prediction of A/Indonesia/CDC1032/2007 (H5N1) strain

Position	Amino acid conversion	Hydrophobicity	Polarity
Haemagglutinin			
185	alanine → glutamic acid	Hydrophobic (-1,911) →Hydrophilic (-3,089)	Non polar →Asam polar
272	glycine → serine	Hydrophobic (-1,122) →Hydrophilic (-1,167)	Non polar →Polar
309	asparagine → serine	Hydrophilic (0,22) →Hydrophilic (0,522)	Polar → Polar
Neuraminidase			
Position	Amino acid conversion	Hydrophobicity	Polarity
244	valine → isoleucine	Hydrophobic (-0,422) → Hydrophobic (-0,389)	Non polar →Non polar
252	proline → threonine	Hydrophilic (-1,367) →Hydrophilic (-1,267)	Polar →Polar

Table 5: Post-translational modification prediction of A/Indonesia/CDC1032/2007 (H5N1) strain

	Haemagglutinin		Neuraminidase	
Site	Position	Amino acid	Position	Amino acid
N-glycosylation	10-13	NNST		
	11-14	NSTE		
	23-26	NVTV	68-71	NSSL
	154-157	NSTY	126-129	NGTV
	165-168	NNTN	215-218	NGSC
	286-289	NSSM		
	484-487	NGTY		
	543-546	NGSL		
Casein kinase II	18-21	TimE	90-93	SkgD
phosphorylation	94-97	SfnD	105-108	ShlE
	121-124	SwsD	128-131	TvkD
	123-126	SdhE	152-155	Srfe
	167-170	TnqE	176-179	SgpD
	267-270	SelE	349-352	SgfE
	298-301	TigE	361-364	TgtD
	384-387	SiiD	393-396	TglD
	391-394	TqfE	436-439	SwpD

Table 5: Continued

	Haemagglutinin		Neuraminidase	
Site	Position	Amino acid	Position	Amino acid
N-myristoylation	63-68	GnpmCD		
	130-135	GvssAC	27-32	GnmiSI
	283-286	GainSS	117-122	GallND
	331-336	GlfgAI	216-221	GScfTV
	334-339	GaiaGF	311-316	GTgsCG
	342-347	GgwqGM	336-341	GvwiGR
	346-351	GmvdGW	362-367	GTdsSF
	361-366	GsgyAA	420-425	GssiSF
	544-549	GslqCR		
CAMP-phospho-site	152-155	KKnS		
Protein kinase C	159-161	TiK	56-58	TeV
phosphorylation	223-225	SgR	128-130	TvK
	308-310	SsR	195-197	TiK
	325-327	SrR	198-200	SwR
	371-373	TqK	232-234	SyK
	379-381	TnK	279-281	SnR
	481-483	SiR	330-332	SfK
			368-370	SvK

Homology Modeling

Homology modelling at Swiss-model server yielded PDB 2fk0 for haemagglutinin and PDB 2hty for neuraminidase. Protein Data Bank database searching revealed that this virus isolate is still similar in structure to Vietnam isolate (A/Vietnam/1203/2004 (H5N1)).

Phylogenetic Analysis

Phylogenetic tree analysis places avian influenza A virus sub-type H5N1 of A/Indonesia/CDC1032/2007 strain in separate cluster to human influenza.

CONCLUSION

Present study revealed that all mutation occur outside the conserved region, except at position 40 and 252 for neuraminidase. In addition, there is no post-translational modification occur at any of the mutation sites. The *in silico* study cannot prove that A/Indonesia/CDC1032/2007 (H5N1) strain is not a totally new strain. We conclude that the existing mutations in haemagglutinin and neuraminidase might only be a case of antigenic drift. Based on the phylogenetic tree analysis and 3-dimensional homology modelling, the mutation is not significant and an influenza A (H5N1) virus of Banten Province strain can not spread from human to human.

ACKNOWLEDGMENTS

We would like to thank Dr. Ridla Bakri for his support and comments on the manuscript. This research was supported by grants from Ministry of Education (Grant Hibah A3 No. 60/DK-A3/06/2007).

REFERENCES

Anwar, T., K. Lal Sunil and U. Khan, 2006. *In silico* analysis of genes Nucleoprotein, Neuramidase and Hemaguggltinin: A comparative study on different strains of influenza A (Bird Flu) virus Sub-type H5N1. J. Silico Biol., 6: 161-168.

Bridges, C.B., J.M. Katz, W.H. Seto, P.K.S. Chan, D. Tsang and W. Ho *et al*., 2000. Risk of influenza A (H5N1) infection among health care workers exposed to patients with influenza A (H5N1), Hong Kong. J. Infect Dis., 181: 344-348.

Glaser, L., J. Stevens, D. Zamarin, A.I. Wilson, G.A. Sastre, M.T. Tumpey, F.C. Basler, K.J. Taubenberger and P. Palese, 2005. A single amino acid substitution in 1918 influenza virus haemagglutinin changes receptor binding specificity. J. Virol., 79: 11533-11536.

Horimoto, T. and Y. Kawaoka, 2001. Pandemic threat posed by avian influenza A viruses. Clin. Microbiol. Rev., 14: 129-149.

Katz Jacqueline, M., W. Lim, C.B. Bridges, T. Rowe, J.H. Primmer and X. Lu *et al*., 1999. Antibody response in individuals Infected with avian influenza A (H5N1) viruses and detection of anti-H5 antibody among household and social contacts. J. Infect. Dis., 180: 1763-1770.

Kawaoka, Y., S. Krauss and R.G. Webster, 1989. Avian-to-human transmission of the *PB1* gene of influenza A viruses in the 1957 and 1968 pandemics. Virology, 63: 4603-4608.

Mahalanabis, M., J. Blue and N.L. Haigwood, 2006. AminoTrack™: Automating the entry and analysis of mutations in multiple protein sequences using a spreadsheet format. Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology (BIOCOMP'06). November 2006, CSREA Press, Las Vegas, pp: 549-555.

Ungchusak, K., P. Auewarakul, S.F. Dowell, R. Kitphati, W. Auwanit and P. Puthavathana, 2005. Probable person to person transmission of avian influenza (H5N1). N. Engl. J. Med., 352: 333-340.