ISSN: 0976-3104
REGULAR ISSUE
Tambunan et al.

THE IIOAB JOURNAL

RESEARCH ARTICLE OPEN ACCESS

# IN SILICO GENETIC VARIATION PATHOGENICITY ANALYSIS OF HEMAGGLUTIN, MATRIX 1, AND NON STRUCTURAL 1 PROTEIN OF HUMAN H5N1 INDONESIAN STRAIN

**Usman Sumo Friend Tambunan[1], Danang Febry Witanto[1], and Arli Aditya Parikesit[1]**

[1]*Department of Chemistry, Faculty of Mathematics and Science, University of Indonesia, INDONESIA*

## ABSTRACT

*Avian influenza (H5N1) is a disease caused by type A influenza viruses. Although in general influenza type A is not lethal to humans, several cases did. In silico mutation analysis could be utilized using multiple alignment methods, AminoTrack server, and phylogenetic tree. Mutations were observed for the cleavage site of haemagglutin (HA) protein, while for the Non-Structural 1 (NS1) and Matrix1 (M1) mutations was observed on their entire region. It was followed by prediction of the pro-P (furin) HA specific cleavage, secondary structure, mutation (exposed/buried) prediction, epitope prediction, and 3D structures prediction. Based on the analysis of mutations in the HA cleavage site, the conserved pattern for Indonesia and Hong Kong H5N1 is RXK / RR, while the sequence of subtype H1N1, H1N2, and H3N2 did not have it. Furin prediction showed this pattern causes the HA of H5N1 could be cleaved. Specific mutations in the NS1 control sequences: (A/Indonesia/5/2005 (H5N1), A/Indonesia/CDC1032/2007(H5N1), A/HongKong/156/97(H5N1), A/BrevigMission/1/18(H1N1), A/Mexico/InDRE4487/2009(H1N1)) with the subtype H1N1, H1N2, and H3N2 were found in position 53. Identical specific control mutation on the M1 for the three subtypes was not found. 248 positions have changes in the H1N1 and H3N2. Epitope prediction explains that control sequences NS1 and M1 of H1N1, H1N2, and H3N2 subtypes have similar $IC_{50}$ values below 50 nM. Specific mutations do not occur in epitope recognition. Occurred region-specific mutations for NS1 and M1 did not affect the secondary and tertiary structure of proteins significantly compared with the controls sequences.*

**Received on:** 20-Sep-2011
**Revised on:** 24th-July-2011
**Accepted on:** 12th– Sept-2011
**Published on:** 10th –May-2011

KEY WORDS

*HA; influenza; in silico; M1; NS1*

**Corresponding author:** Email: usman@ui.ac.id ; Tel: +6221 727 00 27; Fax: +6221 786 34 32

## [I] INTRODUCTION

Since 1997, a highly pathogenic avian influenza (HPAI), H5N1 influenza A virus, has caused massive deaths in poultry and humans [1]. This virus has spread in Asia including Japan and Indonesia. The first influenza virus causing the pandemic and many casualties are the subtype H1N1 in 1918 (Spanish Flu) [2]. Recent cases in 2009, the subtype H1N1 that occurred in Mexico (Swine Flu) at least causes hundreds of casualties. Until June 12, 2009, the H1N1 virus caused 145 casualties in various countries. The pandemic outbreak of threat of HPAI H5N1 could be fatal to the human population and poultry. Therefore, the estimation of pandemic occurrence still needs further evaluation because this pandemic occurrence can cause economic and investment losses. To overcome the threat, more insight on H5N1 molecular mechanism is necessary. Genome and Proteome of H5N1 need to be further studied for designing remedy of the disease [1].

Evolution and mutations that occurred in H5N1 viruses are strongly affected by their ability to perform antigenic drift/shift. During its development, H5N1 could be easily transmitted from human to human due to both antigenic processes, so it becomes a real pandemic threat [1].

There are three strains of influenza viruses: type A, B and C. Type A is responsible for the deadly influenza pandemics (worldwide epidemics) that occur every 10 to 40 years [3]. Influenza A virus belongs to the Orthomyxoviridae family. It consists of eight segments of single stranded RNA, that is HA, NA, PB2, PB1, PA, NS, NP, and M. Currently there are 16 known types of HA and 9 types of NA [4].

Viral attachment to sialic acid receptors in the host cell surface is the most decisive phase in the viral infection process. In the avian influenza virus, HA protein can only recognize the [Neu5Ac (α2-3) Gal] receptors that exist in birds. This is different with [Neu5Ac (α2-6) Gal] receptors in humans. Theoretically, the avian influenza virus could not infect humans because these receptors are different. However, subsequent developments indicate a possible change in the amino acid composition as a result of HA antigenic drift/shift. It culminated in inter-human transmission, which marks the beginning of the global pandemic threat of avian influenza [2].

The main aspects in determining the pathogenicity level of influenza virus are (i) the interaction at the mucus layer, (ii) the process of viral entry into host cells, (iii) the process of viral replication, (iv) the influence of viral infection on the immune

THE IIOAB JOURNAL
ISSN: 0976-3104



system, (v) the spread of the virus replicates (vi) the level of cell damage by viral infection, and (vii) the specificity of the host cell [5].

Avian influenza can be found in two forms namely the acute form of HPAI and low pathogenic form of mild avian influenza (LPAI). In HPAI, morbidity and mortality can reach 100% [2].

Pathogenicity Level of H5N1 viruses is influenced partly by the termination site of HA (cleavage site) comprising amino polybasic region that allows the HA to be highly sensitive to endogenous proteases/cellular host. As a result, the virus can spread in a wide range of cells and not limited only in the respiratory tract. HA addition and mutation of other genes (i.e., NA, PB1, PB2, NS, and M) also affect the level of pathogenicity of the virus [2].

Our research group has conducted research on HA mutation of H1N1 [6]. The question is whether there is a common genetic variation of H5N1 Indonesia, which has a high level of pathogenicity of influenza cases or there are other genetic differences that determine the degree of pathogenicity of the virus.

The purpose of this study is to analyze the in silico genetic variation of HA protein (i.e., NS1) and the M1 protein of Indonesian H5N1; this variation may affect the level of human pathogenicity compared with other subtypes of influenza. The effect of mutation on the change in secondary and tertiary structures of protein as well as the introduction of the epitope will be expected.

## [II] MATERIALS AND METHODS

The data for Influenza A protein were obtained from the GenBank database. It was issued by the National Center for Biotechnology Information (NCBI) through the Influenza Virus Resource.

### 2.1. Search sequence database of H5N1 proteins

H5N1 protein sequences were taken from the website http://ncbi.nlm.nih.gov/ which issued by the NCBI. The downloaded protein data were the HA, NS, and M1 Indonesian H5N1 virus isolates from humans, H5N1 from Hong Kong, H1N1 from 1918 outbreak, and H1N1 from outbreak at Mexico in 2009. They have high degree of malignancy characteristics. These data were also supported by other types of influenza, which had relatively low level of malignancy, such as H1N1, H1N2, and H3N2.

### 2.2. Performing Multiple Sequence Alignment

Having obtained the desired protein sequence, multiple alignments were then performed to determine protein sequences similarity in each strain of H5N1 influenza subtype with others that have relatively low levels of malignancy. Program used in this step was the CLC Main Workbench 5.0. The results of this alignment would be useful for the analysis of mutations of the H5N1 virus.

### 2.3. Generating Phylogenetic Tree

Making phylogenetic tree was intended to look for sequences that have a close kinship with virus HPAI type. Phylogenetic tree-making process was done using software CLC Main Workbench 5.0 with the input alignment of obtained sequences.

### 2.4. Mutation Analysis

This step was performed to look for areas where there were differences and similarities of protein sequences of Indonesian H5N1 influenza virus subtype to another. It is expected to find areas that determine the nature of viral pathogenicity by comparing with the LPAI cases. Literature study was necessary to look for possible areas that decisive for the pathogenicity of influenza A viruses, especially H5N1.
The program used to analyze of mutation is called the Amino Track on a server, which can be accessed through http://apps.sbri.org/AminoTrack/. This program would detect the amino acid mutation that occurred in the protein. The data input is from the result of previous sequence alignment, and the output is their specific mutations.

### 2.5. Site cleavage Prediction HA0 by furin

One of the pathogenicity determinants is the pattern of HA intracellular protein breakdown by furin. Patterns that occur by furin can be predicted using the furin online server, which can be accessed through the following Web site http://cbs.dtu.dk/services/ProP.

### 2.6. Secondary Structure Prediction

The data of amino acid changes from the Amino Track analysis were utilized to predict whether there are changes in protein secondary structure. Secondary structure prediction was evaluated using Jpred server 3, which was accessed through http://compbio.dundee.ac.uk/ by entering amino acid sequences of the isolates in FASTA format.

### 2.7. Prediction of Protein Sequences Buried / Exposed Location

Prediction whether a specific mutation occurred in buried or exposed areas was done using the Conseq server. It could be accessed through the Web Site http://conseq.bioinfo.tau.ac.il

### 2.8. Prediction epitope HA and NA

Programs used for epitope prediction was Immuneepitope. This program was accessed through the website http://tools.immuneepitope.org/analyze/html/mhc-binding.html. From the server analysis, it is expected to know the changes that occur from the introduction of the epitope. The changes in amino acid composition could affect the epitopes.

### 2.9. Protein 3D structure Prediction.

The depiction of the proteins 3D structure could help to show the physiological structure similarity of the existing protein sequence. Online courses were accessed via http://ps2.life.nctu.edu.tw.

## [III] RESULTS

### 3.1. Protein Sequence Search

Utilized sequences of influenza A viruses were H1N1 subtype,

THE IIOAB JOURNAL
ISSN: 0976-3104



H1N2, H3N2, and H5N1 from human host. The selected influenza protein sequences in this study were the HA, NS1, and M1. The search towards entire sequence of the desired protein was done by accessing the web site of NCBI with the following address: http://ncbi.nlm.nih.gov/genomes/FLU/Database.html.

This study utilized the influenza A H1N1, H1N2, H3N2, and H5N1 subtypes from human hosts. H1N1 and H3N2 sequences were isolated from Asia, while the H1N2 sequences were taken from the entire region in the database. This was due to the scarce availability of the Asian subtype. The H5N1 subtype sequences came from Indonesia. The derived protein sequences were HA, NS1, and M1 as well as selected full-length sequences in order to facilitate the subsequent analysis **[Supplementary Table– 1]**.

### *3.2. Multiple Sequence Alignment*

This study was aimed to determine the differences in Indonesian H5N1 human host protein sequences from the other cases of influenza virus subtypes. To facilitate the causal factors determination of the Indonesian H5N1 human high pathogenicity, two sequences of Indonesian H5N1 human were retrieved as the representative for the HA protein, NS1, and M1. Selected sequences were based on the first case of Indonesian H5N1 human infection (A/Indonesia/5/2005) and the last case reported to the NCBI database (A/Indonesia/CDC1032/2007).

Sequences that have a high level of human malignancy were also retrieved for comparison. They were subtype H1N1 in 1918 (A / Brevig Mission/1/18) and 2009 (A/Mexico/InDRE4487/2009) and the Hong Kong H5N1 strain of the human host (A / Hong Kong / 156/97). Sequences that have a high level of pathogenicity were subsequently used as a control to see the occurred protein sequences differences when compared with other relatively low pathogenicity human influenza subtypes cases.
The sequences with high level of pathogenicity were then carried out by sequence alignment with H1N1, H1N2, and H3N2.

### *3.3. Generating Phylogenetic Tree*

The results of the created phylogenetic tree were interesting. It was found that the control strains with high pathogenicity are on adjacent branches. It could be concluded that they have a close kinship. This trend is found on each sequence for low pathogenicity H1N1, H1N2, and H3N2; they have relatively pretty close kinship with the controls **[Supplementary Table– 2]**. Sequences of these proteins can be used as a comparison and can be searched for specific mutations that affect the pathogenicity level of a virus.

### 3.4. Mutation Analysis

Amino Track is a server that could identify the occurred mutations such as deletion, substitution, or insertion an amino acid to the resulted multiple sequence alignment in the form of comparisons between one sequence as a reference with others based on statistical programming [7].

The result of multiple sequence alignment must be converted into FASTA format. It required Jalview software that could convert the *.aln format alignment into FASTA. The output of Amino Track was in zip format that included the Microsoft Excel format. This server makes it easy to see changes in position in the protein sequences of each isolated virus.

#### *3.4.1. Mutation HA*

The results of mutation analysis using Amino Track for HA protein sequences on the HA0 cleavage site are shown in **[Supplementary Table– 3]**. In addition to its role in receptor binding, HA also plays a role in viral fusion process to release its genetic material into host cells. This process requires activation by the HA precursor molecule that cuts into HA1 and HA2. It was done by host cell proteases [8].

It can be seen that the Indonesian human H5N1 sequences and Hong Kong have a polybasic amino acid insertion region RXK / RR while the sequences of H1N1, H1N2, and H3N2 have only one arginine in the cleavage area **[Supplementary Table– 3]**. The pattern of the Indonesian H5N1 sequences is in accordance with the A/HongKong/156/97 avian influenza cases. They are classified as HPAI (H5N1).

#### *3.4.2. Mutations in NS1*

NS1-specific mutation sequence is shown in **Supplementary Table– 4**. The Table shows that specific mutations with the H1N1 control sequences occurred in position 3, 47, 51, 53, 137, 145, 164, 173, 176, and 216. Control with H1N2 specific mutations occurred at positions 53, 56, 57, 66, 73, 180,183, and 211. While specific mutations in the H3N2 control occurred at position 53, 56, 66, 73, 98, 180, 183, and 185. The interesting thing is found in the deletion occurrence at positions 80, 81, 82, and 83 for Indonesian H5N1 isolates sequences. Mutations occurred at position 53 to either the control sequences with the H1N1, H1N2, and H3N2. D53N mutation happened in the H1N1 and H1N2 D53E / H3N2.

Pathogenicity Level of influenza viruses is strongly influenced by the work of the NS1 protein, which is involved in innate immune response inhibition process mediated by type 1 interferons (IFN). The presence of glutamic acid at position 92 of NS1 causes the virus to be resistant to IFN. If the aspartic acid residue is at position 92, the influenza virus becomes sensitive to IFN [9].

Glutamic acid was found at position 92 in the sequence

THE IIOAB JOURNAL
ISSN: 0976-3104



A/HongKong/156/97 (H5N1), whereas the Indonesian H5N1 as well as H1N1, H1N2, and H3N2 isolates have the aspartic acid residue at position 92. The difference is that the origin of Indonesian H5N1 has a deletion area; this is different with other viral sequences. This deletion region could possibly exert influence on the pathogenicity level of influenza viruses, especially H5N1 [8].

The sequences of A/Brevig Mission/1/18 (H1N1) and A/Mexico/InDRE4487/2009 (H1N1) also have aspartic acid at position 92. However, both of these sequences have high level of pathogenicity toward humans. This may illustrate that the D92E mutation on the NS1 is not the sole determinant of influenza virus A pathogenicity.
The results of the specific mutations analysis were observed on the changes in secondary structure and epitope prediction.

#### *3.4.3. Mutation in M1*

Amino Track servers obtained several mutations in M1 sequence. From these results then sought a position in which specific mutations occurred with the selected H1N1, H1N2, and H3N2 control sequences **[Supplementary Table– 5]**.

Specific mutations that occurred between the H1N1 and H3N2 control occurs at only one position. They are positions 248 and 193 for H1N1 and H3N2. While specific mutations among control with H1N2 occurred in position 9, 10, 11, 12, 13, 14, 201, 208, and 248. The mutation position was not identical for all three subtypes. Only position 248 is experiencing changes in the H1N1 and H3N2.

The specific mutations analysis would be observed for its influence on changes in secondary structure and epitope prediction for protein NS1.

## [IV] DISCUSSION

### *4.1. Prediction Site Cutting HA0 by furin*

LPAI usually has a single arginine residue in the cleavage region. HPAI viruses have the polybasic amino acids that could be cleaved by many intracellular proteases such as furin. This would cause a systemic infection on the host [8]. The results of multiple alignment and mutation analysis using the Amino tracks on the utilized sequences show that Indonesian H5N1 human isolates already have an RXK/RR pattern similar to that of A/Hongkong/156/97. It was seen that the entire sequence of the Indonesian H5N1 human host virus has the pattern, and it belongs to the HPAI class.

Moreover, for proving that this pattern affects the ease of cleavage by using intracellular proteases such as furin, the online server furin was employed to determine the cleavage site(s) on HA. Scores are given to indicate the most likely position of the HA cleavage process. Here, the results of pro-P prediction server for HA sequences are shown in **Supplementary Table– 6**. The Table shows that the addition of H5N1 polybasic amino acid sequences could be easily broken down by furin. While the subtype H1N1, H1N2, and H3N2 did not fit with furin pattern.

1N1 in 1918 (A/SouthCarolina/1/18) and H1N1 in 2009 (A/Mexico/InDRE4487/2009) are strains that cause substantial mortality for humans as well as H5N1 case. However, the two sequences do not have a pattern such as those owned by H5N1. This illustrates that the virulence factors or pathogenicity of human influenza viruses are not merely influenced by the cleavage sites of HA alone. Moreover, there are other genetic factors influencing the level of malignancy of influenza virus. Although the H1N1 influenza virus of 1918 has no polybasic amino acids at the HA-cutting areas, it has trait for ease of HA cleavage. This proved that the virus could also be cleaved by the absence of trypsin. Until now, the mechanism of ease of cleavage for HA in the 1918 H1N1 has not been clearly identified [8].

### *4.2. Secondary Structure Prediction*

Determination of secondary structure of influenza virus was predicted. The results were obtained in the form of secondary structure prediction for each protein.

The result for NS1 protein was that the amino acid change at a specific position of mutation did not significantly affect the coherence of secondary structure. It was found that there is only one position (position of 51) that acts as a specific mutation site in the H1N1 versus control. It changes the secondary structure of the coil into an extended strand.

The secondary structure prediction on a specific mutation on M1 found that changes in secondary structure occurred only in H1N2 sequences. However, in position 11 and 12 there were changes from a helix and extended strand, and at position 14 changes were found from the extended strand to coil. Mutations in the H1N1 and H3N2 did not change the secondary structure of M1 protein. To this end, it can be inferred that the structure of M1 sequences of Indonesian H5N1, H5N1 Hong Kong, 1918 H1N1, and H1N1 Mexico 2009 are more homologous to the sequences of H1N1 and H3N2. They have not had a high level of pathogenicity.

### *4.3. Mutation Prediction Burried / Exposed*

Specific mutation prediction that occurs in areas that are buried or exposed can be evaluated using the Conseq server. FASTA-formatted protein sequences were used as input, and the protein

THE IIOAB JOURNAL
ISSN: 0976-3104



sequence A/Indonesia/5/2005 (H5N1) was employed as a comparison.

Specific mutations among controls were carried on H1N1 in the exposed region (positions 3, 53, 164, 173, 176, 216) and the buried region (positions 47, 51, 137, 145). Specific mutations among controls were carried out to H1N2 in the exposed region (positions 53, 56, 66, 211) and in the buried regions (positions 57, 73, 180, 183). There are specific mutation among controls which exposed to H3N2 in positions 53, 56, and 66; whereas in buried regions occurred in positions 73, 98, 180, 183, and 185. Mutations region that occur in the NS1 protein could detect both the exposed and buried area. Position 53 is the point of specific mutations among the controls within H1N1, H1N2, and H3N2. They are located in exposed areas, and it could influence the specific activity of the protein.

Controls with H1N1 specific mutations occurred in the buried area (position 248). Specific mutation between controls with H1N2 occurred in exposed regions (position 9, 201, 208), while in buried regions, they are some different positions. They are position 10, 11, 12, 13, 14, and 248). Specific mutation between controls with H3N2 occurred in exposed regions (position 93).

### *4.4. Predicted epitope*

The immuoepitope server predicts the epitope recognition using the FASTA-formatted protein sequences input [10]. In this server, MHC class selection (Major Histocompatibility Complex) will be used in epitope prediction. This study is using MHC class I-related peptide fragment that is activated by intracellular signaling and CD8 + [11].

The output prediction is given in immune-epitope units ($IC_{50}$ nM). From these predictions if there are peptides that have the $IC_{50}$ value (inhibitory concentration) of less than 50 nM, then the peptide is having high affinity. $IC_{50}$ value of less than 500 nM has intermediate affinity and the $IC_{50}$ value of less than 5000 nM has a low affinity (http://tools.immuneepitope.org/analyze /html/MHC-binding.html, accessed May 18, 2009, 10:45 GMT).

By comparing the NS1 and M1 epitope between controls with subtype H1N1, H1N2, and H3N2, we can see if there are changes in the predicted epitope recognition and changes the level of virus pathogenicity.

NS1 epitope predictions detected LKANFSVVF peptide sequence that has the lowest $IC_{50}$ value of 0.98 nM. This protein sequence presents in isolates of A/Thailand/271/2005 (H1N1) at position 130-138 with recognition by HLA B * 1503. This recognized position also occurs in isolates of A/HongKong/156/97 (H5N1), A / BrevigMission / 1 / 18 (H1N1), A/Mexico/InDRE4487/2009 (H1N1), A/Philippines/344/2004 (H1N2), A/HongKong/1774/99 (H3N2) with peptide sequence LKANFSVIF and $IC_{50}$ of 1.2 nM. The predicted epitope of Indonesian sequence that has $IC_{50}$ of 1.0 nM at position 196-204, which has ETIQRFAWR sequence with the recognition of HLA A * 6801. $IC_{50}$ value is less than 50 nM, suggesting that that all isolates of the virus are easily inhibited by the immune system. It can be concluded that specific mutations that occur have less effect on this epitope prediction [12].

Epitope predictions for M1 obtained RKLKREITF sequence that has the lowest $IC_{50}$ value of 1.1 nM. This sequence was found in isolates of A/ Brevig_Mission/1/8 (H1N1) at position 101-109 with recognition by HLA B * 1503. This position also occurs in isolates of A/Mexico/InDRE4487/2009 (H1N1), A/Philippines/344/2004 (H1N2), A/HongKong/1774/99 (H3N2) with KKLKREITF sequence and has $IC_{50}$ of 1.4 nM. The A/HongKong/156/97 isolates (H5N1) also have the epitope region with this position and have KKLKREMTF sequence with $IC_{50}$ of 1.3 nM. For Indonesian isolates, the RSHRQMATI predicted epitope has $IC_{50}$ of 1.3 nM at position 160-168 and it is recognize by HLA B * 1517. Just as predicted epitope for the NS1, all isolates had $IC_{50}$ values below 50 nM so they can be recognized by HLA and easy to inhibit. Specific mutations occurred did not affect the change of epitope prediction and the obtained $IC_{50}$ values.

### *4.5. Structure Prediction*

Predicted 3D structures of NS1 and M1 proteins were determined and the extent of changes occurred due to mutation was evaluated. This prediction method is known as the homology modeling. This process uses an online server that can be accessed through http://ps2.life.nctu.edu.tw. Input of the predicted protein sequences was in FASTA format. Then, Blast was performed to search for template that has the highest homology level.

The search results of NS1 protein template have similarity for all viral isolates. Template to the RNA Binding Domain is 2ZoaA and to the effectors domain is 2gx9A. M1 template for the whole virus isolates is 1aa7A. **Figures– 1 and –2** show the 3D structures of protein NS1 of the selected sequences and superimposed with the protein from A/Indonesia/5/2005. **Figure– 3** shows the 3D structure of the selected M1 protein sequences and superimposed with protein from A/Indonesia/5/2005.

Value of RMSD shows a very close resemblance to each structure prediction results. So it can be concluded that the specific mutations that occurred did not affect the tertiary structure of proteins significantly. The **Supplementary Table– 7 and– 8** show the RSMD data for M1 and NS1 protein. These results could be scaled up with molecular docking and modeling, with the example of our previous research [13]

THE IIOAB JOURNAL
ISSN: 0976-3104

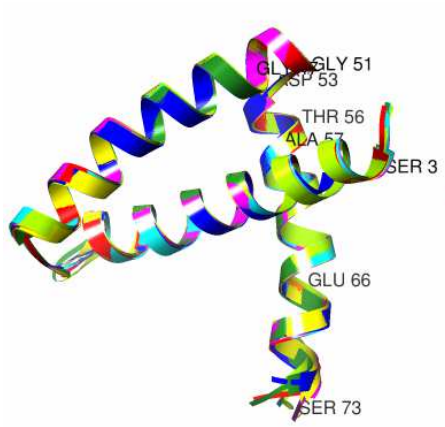


Fig: 1.

Fig: 2.

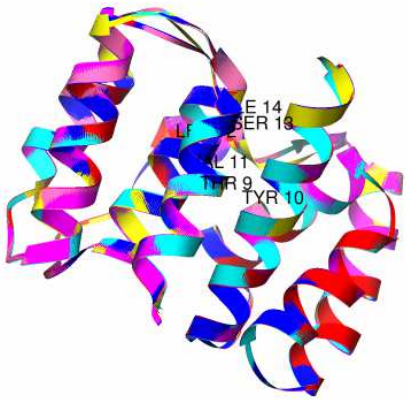


Fig: 3.

**Fig: 1.** Superimposed RNA Binding Domain NS1. **Fig: 2.** Superimposed NS1 effector domain. **Fig: 3.** Superimposed M1

## [V] CONCLUSION

Mutation analysis found specific mutations in the HA sequences of Indonesian and Hong Kong H5N1 isolates in the site which cleaved HA0 into HA1 and HA2. The results concluded that avian influenza H5N1 virus could cause a systemic infection on the host compared with other isolates. Epitope prediction shows that the sequences NS1 and M1 of H1N1, H1N2, and H3N2 subtypes have similar $IC_{50}$. Henceforth, it still could be recognized by the host immune system. The results of these specific mutations that occurred for both NS1 and M1 did not affect any significant changes in secondary structure and tertiary structure. The next step will be conducting molecular docking and modeling of our designs.

## ACKNOWLEDGEMENT

The authors would like to express our gratitude to Ridla Bakrie PhD, the chief of Chemistry Department, Faculty of Mathematics and Science, University of Indonesia, for his support toward this research. We would also like to express our gratitude to Professor Teruna Siahaan from Department of Pharmaceutical Chemistry, University of Kansas for proofreading this manuscript.

## CONFLICT OF INTEREST

Authors declare no conflict of interest.

## FINANCIAL DISCLOSURE

The work is not supported by any grant.

## REFERENCES

[1] Peiris JSM, de Jong MD, Guan Y. [2007] Avian Influenza Virus (H5N1): a Threat to Human Health. *Clin. Microbiol.*20: 243–267.
[2] Asmara WW, Michael H, Tabbu, CR. [2005] Identification of Hemagglutinin of Avian Influenza subtypes on poultry by using RT-PCR. *Veteriner Science Journal* 2005, XXIII(1)
[3] Baigent SJ, John WM. [2003] Influenza Type A in Humans Mammals and Birds: Determinants of Virus Virulence, Host range and interspecies Transmission. *BioEssays* 25: 657–671.
[4] Cox N, Neumann G, Donis RO, Kawaoka Y. [2005] *Orthomyxoviridae. Sem. Viro*l. 6: 359–370.
[5] Sweet C, Smith H. [1980]. Pathogenicity of Influenza Virus. *Microbiological Review* 44: 303–330
[6] Tambunan U.S.F, Ramdhan. [2010] Identification of sequence mutations affecting hemagglutinin specificity to sialic acid receptor in influenza A virus subtypes. *Bioinformation* 5(6): 244–249
[7] Mahalanabis M, Blue J, Haigwood NL. [2006] AminoTrack™ : Automating the Entry and Analysis of Mutations in Multiple Protein Sequences Using a Spreadsheet Format. In H. R. Arabnia (ed.), Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology (BIOCOMP'06). CSREA Press, Las Vegas, NV. pp. 549–555.
[8] Korteweg C, Gu J. [2008] Pathology, Molecular Biology, and Pathogenesis of Avian Influenza A (H5N1) Infection in Humans. *Am J Pathol* 172(5): 1155–1170.
[9] Zambon MC. [2001] The Pathogenesis of Influenza in Humans. *Rev.Med. Virol.* 11: 227–241
[10] Kulkarni-Kale U, Bhosle B, Kolakar AS. [2005] CEP: a Conformational Epitope Prediction Server. *Nuc. Acid Research*. Vol. 33.
[11] Ulmer J. B. [2006] Novel Vaccines: Bridging Research, Development and Production-CHI's *Inaugural Conference.Drugs.*9 (10).

[12] Zhang GL, Khan AM, Srinivasan KN, Angust JT, Brusic V. [2005] MULTIPRED: a Computational System for Prediction of Promiscuous HLA Binding Peptides. *Nuc. Acids Res.*
[13] Tambunan USF, Limanto A, Parikesit AA. [2010] In Silico Analysis of Hemagglutinin, Neuraminidase, and Matrix2 of H5N1 Virus Indonesian Strain Related to Its High Pathogenicity. *IIOAB Journal*.Vol.1; Issue 3.

***Prof Usman Sumo Friend Tambunan*** *is currently working as a permanent professor in the chair of Bioinformatics, Department of Chemistry, Faculty of Mathematics and Science, University of Indonesia. Previously, he worked at the Indonesian Agency of Technology Assessment and Application, as a senior scientist, and graduated his Phd in Tohoku University, Japan. He was the former vice dean of Faculty of Mathematics and Science, University of Indonesia. He has conducted research on Bioinformatics for biomedics, in the topics of HPV, Avian Influenza, and Dengue Virus. He had secured a copyright patent for HPV vaccine design and published many bioinformatics-related articles in peer-reviewed international scientific journals. He is elected several times as the best lecturer and researcher in University of Indonesia. Indonesian ministry of national education grant currently supports his research.*

***Danang Febry Witanto****, Bsc has worked as junior assistant and researcher of Professor Usman Sumo Friend Tambunan at Bioinformatics Laboratory, Departement of Chemistry, Faculty of Mathematics and Science, University of Indonesia. He has involved especially in analyzing Influenza type A research. Now he currently works as purchasing executive in Kao Indonesia Inc.*

***Arli Aditya Parikesit****, Msc is currently working as assistant of Prof Usman Sumo Friend Tambunan, lecturer, and researcher in the chair of Bioinformatics, Department of Chemistry, Faculty of Mathematics and Science, University of Indonesia. He has conducted research on Bioinformatics for biomedics, in the topics of HPV, Avian Influenza, and Dengue Virus. He finished and published his master thesis about HPV, with the support of Indonesian ministry of Education Graduate grant (hibah pasca). Now, he is a doctorate/Phd candidate at the chair of Bioinformatics, Department of Computer Science, Faculty of Computer Science and Mathematics, University of Leipzig, Germany with the support of DAAD fellowship. The theme of his doctorate research is 'Domain Co-occurrences Distribution of Genetic Regulators from an Evolutionary Perspective'. Some parts of his Phd research are already published.*

THE IIOAB JOURNAL
ISSN: 0976-3104







## SUPPLEMENTARY TABLES (As supplied by author)

**Supplementary Table: 1.** Number of Obtained Influenza A Protein Sequences

| Subtype | Protein | | |
|---|---|---|---|
| | HA | NS1 | M1 |
| H1N1 | 79 sequence | 34 sequence | 69 sequence |
| H1N2 | 30 sequence | 32 sequence | 34 sequence |
| H3N2 | 430 sequence | 130 sequence | 510 sequence |
| H5N1 Indonesia | 97 sequence | 85 sequence | 81 sequence |

**Supplementary Table: 2.** Sequences of Influenza A Virus H1N1, H1N2, and H3N2 that Has Kinship Close to Control

| Subtype | Protein | | |
|---|---|---|---|
| | HA | NS1 | M1 |
| H1N1 | A/Thailand/271/2005 | A/Thailand/271/2005 | A/Thailand/271/2005 |
| H1N2 | A/Philipine/344/2004 | A/Philipine/344/2004 | A/Philipine/344/2004 |
| H3N2 | A/HongKong/1774/99 | A/HongKong/1774/99 | A/HongKong/1774/99 |

**Supplementary Table: 3.** Mutations that occurred in the Regional HA Cleavage Site

| Sample | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A/Indonesia/5/2005 (H5N1) | N | S | P | Q | R | E | S | R | R | K | K | R | G |
| A/Indonesia/CDC1032 /2007(H5N1) | N | S | P | Q | R | E | S | R | R | K | K | R | G |
| A/HongKong/156/97 (H5N1) | N | T | P | Q | R | E | R | R | R | K | K | R | G |
| A/SouthCarolina/1/18 (H1N1) | N | I | P | S | I | Q | S | R | - | - | - | - | G |
| A/Mexico/InDRE4487/ 2009 (H1N1) | N | V | P | S | I | Q | S | R | - | - | - | - | G |
| A/Thailand/271/2005 (H1N1) | N | I | P | S | I | Q | S | R | - | - | - | - | G |
| A/Philippines/344/ 2004(H1N2) | N | I | P | S | I | Q | S | R | - | - | - | - | G |
| A/HongKong/1774/99 (H3N2) | N | I | P | E | K | Q | T | R | - | - | - | - | G |

THE IIOAB JOURNAL
ISSN: 0976-3104

**Supplementary Table: 4.** Specific mutations NS1

| Sample | 3 | 47 | 51 | 53 | 80 | 81 | 82 | 83 | 84 | 137 | 145 | 164 | 173 | 176 | 216 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A/Indonesia/5/2005 (H5N1) | S | G | G | D | - | - | - | - | - | I | I | P | D | N | P |
| A/Indonesia/CDC1032/ 2007 (H5N1) | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| A/Brevig_Mission/1/ 18(H1N1) | - | - | - | - | T | I | A | S | V | - | - | - | - | - | - |
| A/HongKong/156/97 (H5N1) | - | - | - | - | T | I | A | S | V | - | - | - | - | - | - |
| A/Mexico/InDRE4487/ 2009(H1N1) | - | - | - | - | T | I | A | S | V | - | - | - | - | - | - |
| A/Thailand/271/2005 (H1N1) | P | S | S | N | T | I | A | S | V | V | T | L | N | D | S |

a. Control vs H1N1

| Sample | 53 | 56 | 57 | 66 | 73 | 80 | 81 | 82 | 83 | 84 | 180 | 183 | 211 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A/Indonesia/5/2005 (H5N1) | D | T | A | E | S | - | - | - | - | - | V | G | R |
| A/Indonesia/CDC1032/2007 (H5N1) | - | - | - | - | - | - | - | - | - | - | - | - | - |
| A/Brevig_Mission/1/18 (H1N1) | - | - | - | - | - | T | I | A | S | V | - | - | - |
| A/HongKong/156/97 (H5N1) | - | - | - | - | - | T | I | A | S | V | - | - | - |
| A/Mexico/InDRE4487/2009 (H1N1) | - | - | - | - | - | T | I | A | S | V | - | - | - |
| A/Philipine/344/2004 (H1N2 | E | P | S | K | Y | T | I | A | S | V | T | K | G |

b. Control vs H1N2

| Sample | 53 | 56 | 66 | 73 | 80 | 81 | 82 | 83 | 84 | 98 | 180 | 183 | 185 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A/Indonesia/5/2005 (H5N1) | D | T | E | S | - | - | - | - | - | M | V | G | L |
| A/Indonesia/CDC1032/ 2007 (H5N1) | - | - | - | - | - | - | - | - | - | - | - | - | - |
| A/Brevig_Mission/1/18 (H1N1) | - | - | - | - | T | I | A | S | V | - | - | - | - |
| A/HongKong/156/97 (H5N1) | - | - | - | - | T | I | A | S | V | - | - | - | - |
| A/Mexico/InDRE4487/2009 (H1N1) | - | - | - | - | T | I | A | S | V | - | - | - | - |
| A/HongKong/1774/99 (H3N2) | E | P | K | Y | T | I | A | S | V | I | I | K | F |

c. Control vs H3N2

**Supplementary Table: 5.** M1 Specific mutations

| Sample | 248 |
|---|---|
| A/Indonesia/5/2005 (H5N1) | M |
| A/Indonesia/CDC1032/2007 (H5N1) | - |
| A/HongKong/156/97 (H5N1) | - |
| A/Brevig_Mission/1/18 (H1N1) | - |
| A/Mexico/InDRE4487/2009(H1N1 | - |
| A/Thailand/271/2005(H1N1) | I |

| Sample | 193 |
|---|---|
| A/Indonesia/5/2005 (H5N1) | A |
| A/Indonesia/CDC1032/2007 (H5N1) | - |
| A/HongKong/156/97 (H5N1) | - |
| A/Brevig_Mission/1/18 (H1N1) | - |
| A/Mexico/InDRE4487/2009(H1N1 | - |
| A/HongKong/1774/99(H3N2) | V |

**a. Control vs H1N1**

c. Control vs H3N2

THE IIOAB JOURNAL
ISSN: 0976-3104



| Sample | 9 | 10 | 11 | 12 | 13 | 14 | 201 | 208 | 248 |
|---|---|---|---|---|---|---|---|---|---|
| A/Indonesia/5/2005 (H5N1) | T | Y | V | L | S | I | E | Q | M |
| A/Indonesia/CDC1032/2007 (H5N1) | - | - | - | - | - | - | - | - | - |
| A/HongKong/156/97 (H5N1) | - | - | - | - | - | - | - | - | - |
| A/Brevig_Mission/1/18 (H1N1) | - | - | - | - | - | - | - | - | - |
| A/Mexico/InDRE4487/2009(H1N1) | - | - | - | - | - | - | - | - | - |
| A/Philippines/344/2004(H1N2) | N | V | C | S | L | Y | D | K | I |

b. Control vs H1N2

**Supplementary Table: 6.** Prediction results of pro-P

| Sample | *Furin-type Cleavage Site Prediction* |
|---|---|
| A/Indonesia/5/2005 (H5N1) | ESRRKKR\|GL |
| A/Indonesia/CDC1032 /2007(H5N1) | ESRRKKR\|GL |
| A/HongKong/156/97 (H5N1) | ERRRKKR\|GL |
| A/SouthCarolina/1/18 (H1N1) | *none* |
| A/Mexico/InDRE4487/2009 (H1N1) | *none* |
| A/Thailand/271/2005(H1N1) | *none* |
| A/Philippines/344/2004(H1N2) | *none* |
| A/HongKong/1774/99(H3N2) | *none* |

**Supplementary Table: 7.** NS1 RMSD data

| Sequence | RMSD (Å) | |
|---|---|---|
| | RNA Binding domain | Effector Domain |
| A/HongKong/156/97 (H5N1) | 0,361 | 0.222 |
| A/SouthCarolina/1/18 (H1N1) | 0,27 | 0,135 |
| A/Mexico/InDRE4487/2009 (H1N1) | 0,356 | 0,152 |
| A/Thailand/271/2005(H1N1) | 0,427 | 0,17 |
| A/Philippines/344/2004(H1N2) | 0,379 | 0,235 |
| A/HongKong/1774/99(H3N2) | 0,433 | 0,117 |

**Supplementary Table: 8.** RMSD Data M1

| Sequence | RMSD (Å) |
|---|---|
| A/HongKong/156/97 (H5N1) | 0,139 |
| A/SouthCarolina/1/18 (H1N1) | 0,191 |
| A/Mexico/InDRE4487/2009 (H1N1) | 0,201 |
| A/Thailand/271/2005(H1N1) | 0,134 |
| A/Philippines/344/2004(H1N2) | 0,148 |
| A/HongKong/1774/99(H3N2) | 0,134 |