A BIOINFORMATICS-BASED EVOLUTIONARY ANALYSIS ON SOUTHEAST ASIA H1N1 INFLUENZA VIRUS STRAINS

TEH BAN HONG

UNIVERSITI SAINS MALAYSIA

2013

A BIOINFORMATICS-BASED EVOLUTIONARY ANALYSIS ON SOUTHEAST ASIA H1N1 INFLUENZA VIRUS STRAINS

by

TEH BAN HONG

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

November 2013

ACKNOWLEDGEMENT

This thesis would not have been materialized without the support, guidance, and helping hands of many people to whom I really thankful to. First of all, I would like to express my gratitude to my supervisor, Prof. Dr. Habibah A. Wahab who took time off from her busy schedules to go through my work time after time, making valuable remarks and important pointers throughout my postgraduate study in my research. Besides that, the guidance provided by her during my research work was really invaluable which led to the successfulness of the research project.

My heartfelt appreciation goes to Dr. Choi Sy Bing, a post-doctoral researcher in the Pharmaceutical Design and Simulation Laboratory (PhDS) in the School of Pharmaceutical Sciences at Universiti Sains Malaysia. She has acted as my research mentor who has constantly provided me with great advices and motivations during the period of my postgraduate study in the laboratory.

I would like to thank the Ministry of Higher Education for supporting me financially (MyBrain15 Scheme) throughout my Master study. Apart from the financial supports, I am also thankful to my parents Mr. and Mrs. Teh Chin Lai for giving me the moral supports constantly throughout my time doing this research project.

Sincere gratitude goes to all my friends in PhDS for giving me words of encouragements, helps and supports for the project. Finally, I would like to thank everybody who was involved directly or indirectly in supporting me to complete this project. It has been a wonderful and meaningful memory for me to join the research group and I would cherish the memories I have.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D	Three-dimensional
A	Alanine
BLAST	Basic Local Alignment Search Tool
blastp	protein-protein BLAST
C	Cysteine
CDC	Centers for Disease Control and Prevention
D	Aspartic acid
DANA	2-deoxy-2,3-didehydro-D-*N*-acetylneuraminic acid
DOPE	Discrete Optimised Protein Energy
E	Glutamic acid
e.g.	for example
F	Phenylalanine
G	Glycine
H	Histidine
H1	Hemagglutinin subtype-1
H2	Hemagglutinin subtype-2
H3	Hemagglutinin subtype-3
H5	Hemagglutinin subtype-5
H6	Hemagglutinin subtype-6
H16	Hemagglutinin subtype-16
HA	Hemagglutinin
HEF	Hemagglutinin-esterase-fusion
I	Isoleucine

i.e.	that is
IVR	Influenza Virus Resource
K	Lysine
L	Leucine
M	Methionine
M1	Matrix protein 1
M2	Matrix protein 2
ME	Minimum Evolution
MEGA	Molecular Evolutionary Genetic Analysis
ML	Maximum Likelihood
molpdf	MODELLER objective function
MP	Maximum Parsimony
N	Asparagine
NJ	Neighbor-Joining
N1	Neuraminidase subtype-1
N2	Neuraminidase subtype-2
N6	Neuraminidase subtype-6
N9	Neuraminidase subtype-9
NA	Neuraminidase
NCBI	National Center for Biotechnology Information
nm	nanometre
NMR	Nuclear Magnetic Resonance
NP	Nucleoprotein
NS1	Non-structural protein 1

NS2	Non-structural protein 2
P	Proline
PA	Polymerase acidic protein
PB1	Polymerase basic protein 1
PB2	Polymerase basic protein 2
PDB	Protein Data Bank
pdf	probability density function
pH	potential of Hydrogen
PhDS	Pharmaceutical Design and Simulation
PIR	Protein Information Resource
ProSA	Protein Structure Analysis
Q	Glutamine
R	Arginine
RMSD	Root Mean Squared Deviation
RNA	Ribonucleic acid
S	Serine
SCOP	Structural Classification of Proteins
SEA	Southeast Asia
T	Threonine
UPGMA	Unweigthed Pair-Group Method with Arithmetic Mean
USA	United States
USM	Universiti Sains Malaysia
V	Valine
W	Tryptophan

WHO	World Health Organisation
Y	Tyrosine

LIST OF SYMBOLS

α	Alpha
Å	Angstrom
■	H274Y strain
%	Percentage
ϕ	Phi
ψ	Psi
	S31N strain
▲	Vaccine strain
▽	V27A strain

ANALISIS EVOLUSI BERASASKAN BIOINFORMATIK KE ATAS STRAIN VIRUS INFLUENZA H1N1 DI ASIA TENGGARA

ABSTRAK

Influenza virus A (H1N1) dikenali sebagai agen penyebab dalam jangkitan influenza yang serius di serata dunia. Pada bulan April 2009, wabak penyakit influenza muncul buat kali pertama dalam abad 21. Sehingga kini, vaksin dan ubat antiviral adalah antara rawatan yang berkesan untuk memerangi jangkitan influenza. Oleh itu, kajian evolusi virus A (H1N1) di rantau Asia Tenggara (AT) adalah penting untuk lebih memahami tentang kepelbagaian, kemunculan rintangan, dan keberkesanan vaksin. Untuk mendapatkan gambaran dalam hal ini, analisis filogenetik telah dijalankan ke atas rantaian penuh asid amino daripada hemagglutinin (HA), neuraminidase (NA) dan protein matriks 2 (M2) daripada influenza virus A (H1N1) dari AT. Berdasarkan kepada analisis filogenetik, edaran bersama pada kedua-dua strain bermusim dan strain pandemik telah diperhatikan dan dikenalpasti melalui pengasingan rantaian strain kepada dua klad yang berbeza. Hasil kajian tersebut telah mendedahkan bahawa klad "A/California/07/2009-like" adalah penyebab wabak pandemik yang dominan di rantau AT pada 2009. Di samping itu, analisis filogenetik bagi HA, NA dan M2 dengan strain vaksin untuk musim 2009-2011 adalah berkait rapat dengan purata masing-masing sebanyak 96.5%, 98.1% dan 96.9% kesamaan jujukan. Perbezaan yang paling ketara merupakan tiga penggantian asid amino (P100S, S220T dan I338V) di hampir semua HA manakala dua penggantian asid amino (V106I dan N248D) di semua NA daripada virus pandemik influenza A (H1N1) di rantau AT. Antara 24 variasi dalam HA, lima telah dilaporkan

dari epitop A, B dan D (N142D, S160G, I196V, S202T dan A214T) manakala empat daripada jumlah 15 variasi telah ditemui di epitop B dan C (P198S, K217R, S339L dan N369K) dalam NA daripada strain influenza yang muncul semasa wabak influenza 2010-2011. Mutasi H274Y tidak ketara dalam rantau AT. Perolehan mutasi strain influenza ini adalah daripada Thailand dan Singapura. Seterusnya, model jangkaan 3D bagi kedua-dua HA dan NA telah dihasilkan dan difahami struktur perubahan asid amino yang berlaku dengan merujuk kepada strain vaksin. Skor DOPE yang terendah untuk kedua-dua model HA (-53940.43) dan NA (-46962.45) telah dipilih dan sejajar dengan verifikasi yang dihasilkan oleh PROCHECK (iaitu HA: 91.0% dan NA: 89.2% residu-residu di rantau yang digemari), ERRAT (iaitu HA: 74.6% dan NA: 83.6% kawasan dengan lipatan protein yang betul) dan ProSA (iaitu skor-z bagi HA: -.9.19 dan NA: -5.21 dan kedua-dua telah didapati berada di kawasan tenaga yang munasabah).

A BIOINFORMATICS-BASED EVOLUTIONARY ANALYSIS ON SOUTHEAST ASIA H1N1 INFLUENZA VIRUS STRAINS

ABSTRACT

Influenza virus A (H1N1) is known to be the causative agent of severe influenza infections worldwide. In April 2009, a pandemic outbreak of influenza disease had occurred for the first time in 21st century. To date, vaccination and antiviral drugs are the only effective treatments to combat influenza infection. Thus, understanding the evolution of influenza virus A (H1N1) within Southeast Asia (SEA) region is crucial to better understand about the diversification, emergence, resistance and the vaccine efficacy. In order to gain insight into the matter, phylogenetic analyses were conducted on the full-length amino acid sequences of hemagglutinin (HA), neuraminidase (NA) and matrix protein 2 (M2) of influenza virus A (H1N1) obtained from SEA. Based on the phylogenetic analysis, co-circulation of both seasonal and pandemic strains were observed and identified via the segregation of sequences into two different clades. The results had revealed that sequences clustered in A/California/07/2009-like clade were the dominant influenza strains for the 2009 pandemic occurrence in SEA. In addition, phylogenetic analyses for HA, NA and M2 showed they were related to the vaccine strains for the 2009-2011 influenza seasons, with an average of 96.5%, 98.1% and 96.9% sequence identity, respectively. The most notable differences are three amino acid substitutions (P100S, S220T and I338V) in almost all HA while two amino acid substitutions (V106I and N248D) in all NA of influenza A (H1N1) pandemic viruses in SEA. Among the 24 variations in HA, five were reported from epitopes A, B and D

(N142D, S160G, I196V, S202T and A214T). On the other hand, four out of a total of 15 variations had been found in epitopes B and C (P198S, K217R, S339L and N369K) of NA in influenza strains emerged during 2010-2011 influenza season. H274Y mutation was not significantly observed in the viruses in SEA. The influenza strains that acquired this mutation were from Thailand and Singapore. Subsequently, 3D models of both HA and NA were predicted and elucidated structurally for the amino acid changes with reference to the vaccine strain. The lowest DOPE scores for both HA (-53940.43) and NA (-46962.45) models were chosen and they corresponded well with the verifications generated by PROCHECK (i.e. HA: 91.0% and NA: 89.2% residues in the most favoured region), ERRAT (i.e. HA: 74.6% and NA: 83.6% region with the correct protein folding) and ProSA (i.e. z-scores of HA: -9.19 and NA: -5.21 and both were found in feasible range of energy).

1.0 INTRODUCTION

1.1 Problem Statement

Influenza H1N1 disease is not a new disease to the public around the globe. The outbreak of influenza H1N1 disease in March and early April 2009 again represented a significant fatal disease of the 21st century since the last pandemic outbreak of influenza H1N1 disease in 1918. This influenza disease that was first started in Mexico and the United States (USA) (Potdar et al., 2010, Smith et al., 2009) achieved a rapid spread in a short time to about 207 countries all around the world, including countries in the tropics and infected about 622,482 people worldwide resulting more than 7,820 deaths as of November 22, 2009. At the same time, Southeast Asia region reported about 738 death cases, which make it the region with the second highest number of deaths after Americas (WHO, 2009).

Due to the rapid spread of novel influenza A virus (H1N1-subtyped), the World Health Organisation (WHO) was alarmed and raised the worldwide pandemic alert level from Phase 5 to Phase 6 on the 11th June 2009 (Smith et al., 2009). This indicated that the infectious disease brought by the novel influenza A virus (H1N1-subtyped) was underway to be a causative factor for the increased number of reported deaths. Its high incidence and high mortality rate as well as lack of effective treatments have spurred the researchers to carry out extensive researches in discovering the cure for the H1N1 pandemic flu.

With the tropical climate in the Southeast Asia (SEA) region, the people in this region are used to the exposure of being infected with seasonal flu throughout the year. Seasonal flu is also commonly known as the non-pandemic flu, whereby the

influenza A virus is also the causative agent. It is not a new infection to the people in the region, yet many of them have underestimated the fatal consequences brought by the influenza virus, especially when the influenza virus has turned into a pandemic strain.

Of recent years, a lot of studies have been researched on the evolutionary of the pandemic outbreak of 2009 H1N1 influenza virus that was first reported in Mexico and also the discovery of the potential neuraminidase inhibitors in combating the influenza disease. Researchers found that mutation and reassortment caused the antigenic drift while antigenic shift in different hosts of influenza virus have led to the outbreak of novel influenza A virus strain. Existence of the novel strains prompted the researchers into the discovery of potential inhibitors on neuraminidase protein as the current existing antiviral drugs, such as oseltamivir, is no longer effective for the inhibitory effects, especially for the drug-resistant strains.

Among the various strains of influenza viruses from the 2009 influenza outbreak discovered in SEA region, the evolutionary relationship among these influenza viruses in the region are yet to be studied. It is crucial to get the relationship among these viruses elucidated by phylogenetic means in order to guide the disease prophylaxis before probing for potential inhibitory compounds on hemagglutinin (HA), neuraminidase (NA) and matrix protein 2 (M2). The evolutionary findings will be of benefits in monitoring the diversification and distribution of the HA, NA and M2 proteins from the influenza A viruses in the region.

Phylogenetic study is the main interest for this research since there is no previous study has carried out on the evolutionary relationship of all influenza A/H1N1-subtyped viruses in SEA. Examination of a phylogeny allows the

determination of the relatedness of a sequence of interest to a group of sequences whose three-dimensional (3D) structures have been solved experimentally and their functions have already been discovered. Comparatively, protein structure is more conserved than the protein sequence. Thus, a 3D protein structure would be more informative in the binding affinity studies especially in searching for the potential inhibitory compounds.

However, studies have revealed that M2 protein has mutated rapidly and developed resistance to the existing adamantine-based antiviral drugs. Since HA and NA proteins of influenza virus are relatively evolving at a slower rate than M2 and are yet to be determined experimentally in SEA, both HA and NA are worthwhile targets to be modelled to obtain the 3D structures. Hence, structural prediction via comparative modelling approach were used in this research to obtain the 3D protein structures which can lead to the initiation of the drug discovery and drug design processes to combat the influenza disease effectively.

1.2 Objectives of the study

The objectives of this study were:

- to study the evolutionary relationship of the HA, NA, and M2 proteins of influenza viruses A (H1N1) in SEA;
- to investigate the amino acid changes in antigenic determinants for HA, NA and M2 proteins of influenza viruses A (H1N1) from SEA in order to determine the efficacy of vaccine recommended by the WHO;
- to investigate if the mutations conferring resistance to oseltamivir and adamantane-based antiviral drugs were present in the influenza viruses A (H1N1) in the region;
- to determine the 3D model of both glycoproteins (i.e. HA and NA) of influenza virus A (H1N1) from Malaysia (A/Malaysia/2142295/2009).

2.0 LITERATURE REVIEW

2.1 Background of influenza viruses

In general, there are three types of influenza viruses in the Orthomyxoviridae family, namely influenza type A, type B and type C (Bouvier and Palese, 2008, ElHefnawi et al., 2011, Rendell, 2006).

Among the three main types of influenza viruses, influenza A viruses have a broad range of ability to infect, humans, birds, pigs and other animals. However, birds, pigs and humans are the most common hosts to get infected, with the pigs being the intermediate host for the influenza A viruses to be spread between the birds and the humans (Brown, 2001, Scholtissek, 1994). Influenza A viruses are further characterized on the basis of their surface glycoproteins subtypes, i.e. hemagglutinin (HA) and neuraminidase (NA). The ratio of both the studded glycoprotein spikes (HA and NA) are approximately 4:1. Besides having two glycoproteins on the surface, a smaller number of matrix protein 2 (M2), which are the ion channels found integrated through the viruses membrane (Bouvier and Palese, 2008).

Influenza B viruses have a narrow range of susceptible hosts compared to influenza A viruses. These viruses are usually found only in the human population infections. Thus far, infections by influenza type B viruses are associated with less severe epidemics and have not caused any pandemics (Bouvier and Palese, 2008, CDC, 2005). With the aid of electron microscopy, they are indistinguishable from the influenza A viruses, through the shape or morphology they are sharing. As with influenza A, influenza B viruses have both the HA and NA glycoproteins embedded

in the membrane and instead of M2, both NB and BM2 substituted the M2 channel of type A viruses in influenza B viruses (Bouvier and Palese, 2008).

In nature, both influenza A and B viruses envelope a segmented genome made up of eight negative-sense single-stranded RNA (as shown in Figure 2.1). Segmented genome benefits the influenza virus survivability as the nature of the genome is able to increase the recombination of different virus genome. These eight single-stranded RNA encode for a total of ten proteins, which are named as hemagglutinin (HA), neuraminidase (NA), polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2), polymerase acidic protein (PA), matrix protein 1 (M1), matrix protein 2 (M2), non-structural protein 1 (NS1), non-structural protein 2 (NS2) and nucleoprotein (NP) (Prakash et al., 2011, Vazquez, 2008).

Similar to type B viruses, influenza C viruses are also commonly found in human infection with a much lesser severity. To date, they have not caused any epidemics or pandemics (CDC, 2005, Rendell, 2006). For influenza C viruses, they have only seven segmented RNA genome and express only a single glycoprotein on the influenza membrane, which are structurally diverged from both type A and type B influenza. The studded surface glycoprotein is identified as hemagglutinin-esterase-fusion, which is functionally replacing both the HA and NA glycoproteins of influenza A and B viruses (Rosenthal et al., 1998, Vazquez, 2008). Unlike influenza A strains, both influenza B and C viruses have yet to be further characterized into distinct subtypes according to the WHO system of nomenclature. As stated in the WHO memorandum published in 1980, the existing antigenic variation among influenza B and C viruses are well established and thus the divison into subtypes are not warranted (WHO, 1980).

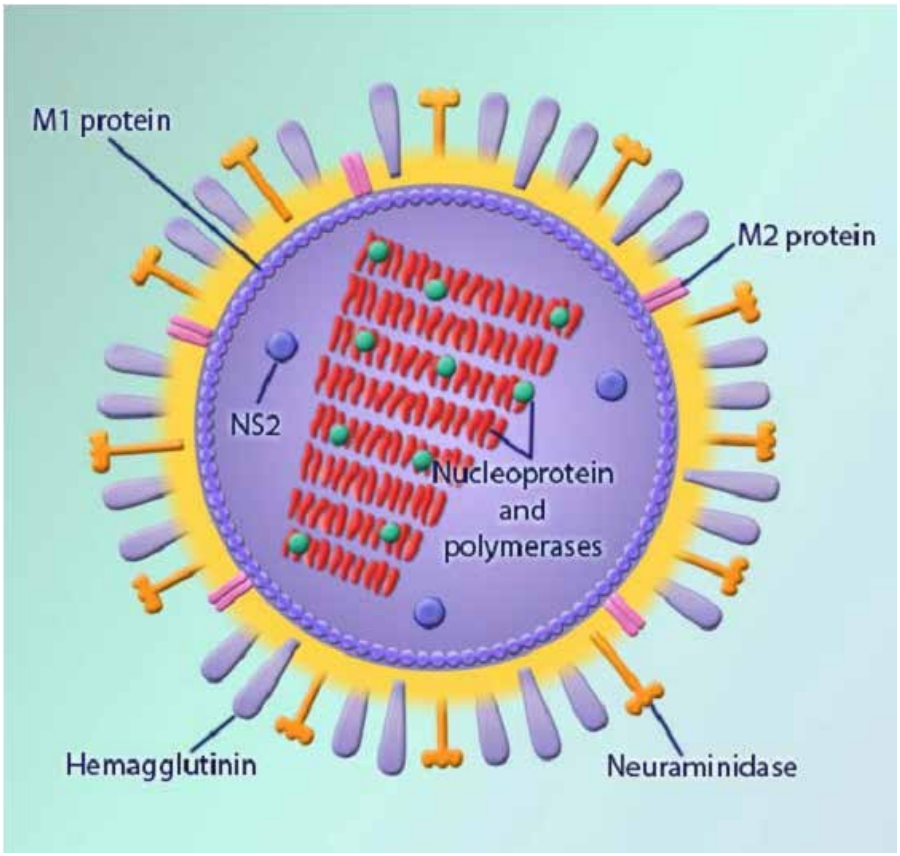

Figure 2.1: The structure of influenza A viruses (Taken from: Vazquez, 2008).

2.1.1 Standard nomenclature for influenza A viruses

According to the WHO, the standard nomenclature for all three types of influenza (A, B and C) includes the virus type; species from which it was isolated (if non-human); location of virus isolation; isolate number; year of isolation (Bouvier and Palese, 2008). In addition, an index describing both the surface glycoproteins (HA and NA) subtypes are followed in parentheses in identifying the strains of influenza A viruses. For instance, A/Malaysia/820/2009 (H1N1) is a human influenza A virus from Malaysia in 2009 with the isolate number of 820, and it has both HA and NA of subtype 1 (Bouvier and Palese, 2008, WHO, 1980).

2.1.2 Reproductive cycle of influenza A viruses

In nature, the influenza survivability is favoured by their segmented genome. Influenza viruses replicate themselves by using the host cellular machinery for the synthesis of their genome and the related components (Smith and Helenius, 2004).

To gain access into a living host, recognition and attachment between the HA surface glycoproteins of influenza viruses and the N-acetylneuraminic acid (also known as sialic acid) is of prime importance to initiate the infection. Sialic acid is a cell receptor with nine-carbon acidic monosaccharide, which is usually seen at the terminal end of surface glycoproteins of animal species. As mentioned earlier, pigs are being the intermediate host for the influeza A viruses to spread between the avians and the humans. The logical explanation for this phenomenon could be the presence of both sialic acid receptors (α-2,3 and α-2,6) for avian and human respectively in pigs (Brown, 2001, Scholtissek, 1994).

As shown in Figure 2.2, once the HA is attached to the host sialic acid receptor, the influenza virus particles will penetrate into the host by endocytosis. The penetrated virions will be released from the endosome when the pH of the surrounding environment inside the host is decreased to approximately pH 5.0. At the acidic environment, the viral M2 ion channels will be activated and allow the influx of protons into the virions. The influx of the protons will bring the acidic condition to the virions and thus disrupt the protein-protein interactions. As a consequence, the virions will be released into the infected cellular cytoplasm (Sidorenko and Reichl, 2004, Sieczkarski and Whittaker, 2002a, Sieczkarski and Whittaker, 2002b, Vazquez, 2008) before the uncoated virions being transported into the cellular nucleus. Using the host cellular nucleus, these virions are transcribed and

translated into all the ten structural proteins respectively (Sidorenko and Reichl, 2004).

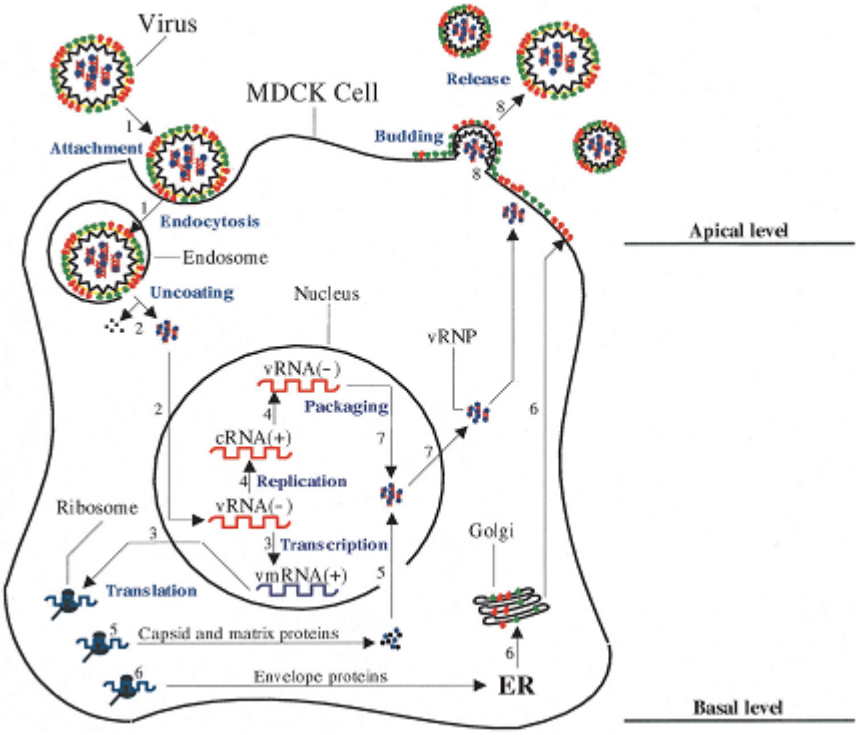

Figure 2.2: Reproductive cycle of influenza A viruses (Taken from: Sidorenko and Reichl, 2004).

Subsequently, all the transcribed and translated virion components are packaged in the host cellular nucleus and transported into the cellular cytoplasm. In the cytoplasm, accumulation of the M1 matrix proteins initiates the influenza virus budding at the cell membrane (Portela and Digard, 2002). The virus progeny can only be released upon the present of sialidase activity from the NA surface glycoprotein (Bouvier and Palese, 2008, Vazquez, 2008).

2.2 Glycoproteins of influenza A viruses

2.2.1 Hemagglutinin

Hemagglutinin (HA) is one of the two major glycoproteins that are expressed and embedded on the viral membrane. To date, HA exists in a total of 16 serologically distinct subtypes (designated as H1-H16). However, only three HA subtypes (H1, H2, and H3) are found to be the circulating strains in human population (Astrahan and Arkin, 2010, Bouvier and Palese, 2008). H5 subtype on the other hand has been found to be circulating among the avians (Anwar et al., 2006, Cheung et al., 2006).

Structurally HA is composed as a trimeric protein (as shown in Figure 2.3). HA is further subdivided into two structurally distinct regions, where there are a stem comprising a triple-stranded coiled-coil of alpha-helices, and a globular head of antiparallel beta-sheet being positioned atop the stem. The globular head contains the sialic acid receptor binding domain and is surrounded by the predicted variable antigenic determinants (Bouvier and Palese, 2008).

HA protein is responsible to initiate the attachment of virus to cellular sialic acid receptors and subsequently mediating the virus to gain entry into the infected cell via fusion of viral and cellular membranes (Astrahan and Arkin, 2010). In addition, HA also plays a structural role in budding and particle formation (Furuse et al., 2010).

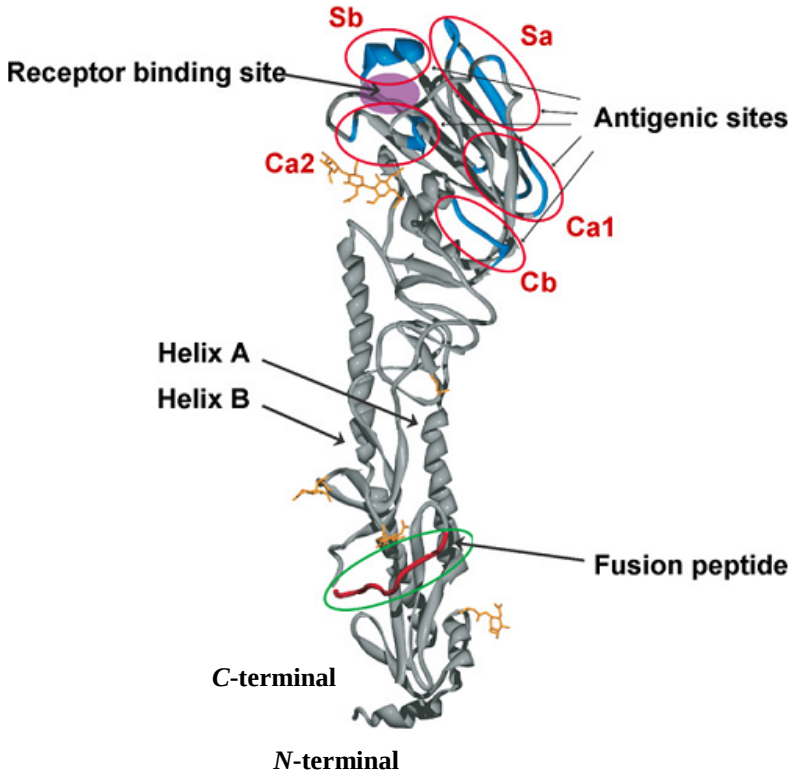

Figure 2.3: Diagram of an uncleaved HA monomer in ribbon representation of influenza virus A (H1N1). The structure contains the sialic acid receptor-binding site, which is surrounded by the 5 predicted antigenic sites (Sa, Sb, Ca1, Ca2, and Cb). As shown, the stem is composed of the Helix A and B, and the fusion peptide (Taken from: Bouvier and Palese, 2008).

2.2.2 Neuraminidase

Neuraminidase (NA) being the other major glycoproteins, are expressed on the influenza A virus surface (Duvvuri et al., 2009, Webster et al., 1982). Together with HA, both HA and NA form the basis in identifying the virus. To date, there are a total of nine serologically distinct subtypes (designated as N1-N9) of NA are known (Astrahan and Arkin, 2010, Bouvier and Palese, 2008). Across these subtypes,

only N1 and N2 subtypes of NA are more commonly found to cause an influenza infectious disease in human populations thus far (Reid et al., 2000).

The NA proteins are anchored in the viral membrane as a tetramer, which are composed of four identical monomeric structures (as shown in Figure 2.4). Each of the NA monomers is structurally divided into four main regions: an N-terminal conserved hydrophilic hexapeptide, a hydrophobic segment across the transmembrane, a thin stalk which is stabilized by carbohydrate and intermolecular disulphide bonds, and the enzymatically and antigenically active "head" which is positioned at the top of protein structure (Blok and Air, 1982, Guangxiang et al., 1993, Reid et al., 2000, Sylte and Suarez, 2009).

(a) **(b)**

Figure 2.4: **(a)** The NA molecule with its "square boxlike head" consists of 4 coplanar subunits connected to the "stalk" with a hydrophobic knob at the end. **(b)** A rosette of 6 tetrameric NA molecules which have "heads" of about 4 x 8 nm and stalks of about 10 nm (Taken from: Blok and Air, 1982).

NA protein serves as an enzyme to remove the terminal sialic acids from both the viral and host receptors (Sylte and Suarez, 2009). With this cleavage enzymatic functionality, NA facilitates the elution of newly formed viruses during budding from host cells to initiate the next infection cycle. Cleaving at the terminal sialic acids can also prevent the agglutination of HA proteins to the host sialic acid receptors (Blok and Air, 1982, Guangxiang et al., 1993, Reid et al., 2000, Sylte and Suarez, 2009).

2.2.3 Matrix protein 2

Besides HA and NA, the third type of membrane protein, called matrix protein 2 (M2), is abundantly expressed on the plasma membrane of infected cells. Structurally M2 is a small integral membrane protein with a total of 97 amino acid residues, which forms a pH-gated proton channels in the viral transmembrane (Schnell and Chou, 2008, Sansom et al., 1997). M2 protein (as shown in Figure 2.5) is composed of 3 domains: a short extracellular N-terminal domain with 24 amino acid residues; a transmembrane domain with 19 amino acid residues, and a long intracellular C-terminal cytoplasmic domain with 54 amino acid residues (Zhong et al., 1998). The ion channel is a homotetrameric membrane protein in its native state (Holsinger and Alams, 1991, Sugrue and Hay, 1991) which is stabilized by disulphide bridges (Sansom et al., 1997, Zhong et al., 1998).

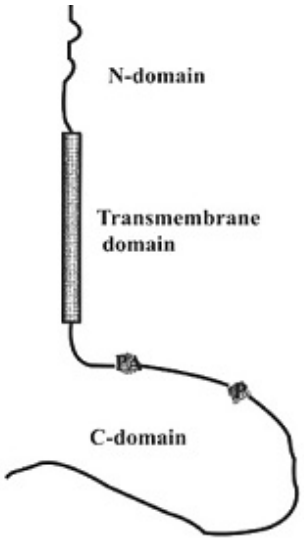

Figure 2.5: Scheme of M2 ion channel. M2 protein has a small N-terminal domain (24 residues), transmembrane domain (19 residues), and long C-terminal domain (54 residues) (Taken from: Betakova, 2007).

The M2 protein functions as a gated ion channel in the replicative cycle of influenza A virus, which will be activated at low acidic pH. At low acidic pH, it enables acidification of the interior of endocytosed virions (Holsinger et al., 1994) and thus facilitates the protein-protein interactions between the matrix protein and virion components. Thus, M2 protein plays a determining role in regulating virus morphogenesis, virus assembly and virus budding (Betakova, 2007, Sansom et al., 1997).

2.3 Influenza pandemics

Influenza virus A is associated with its deadly infection that could bring the sudden increase in the mortality rate. They had been the causative agents for several pandemic flu outbreaks, such as the Spanish flu (1918), the Asian flu (1957), the Hong Kong flu (1968) and the recent Mexican flu (2009). Of all these pandemics, they were informally identified in accordance to their presumed sites of origin (Kilbourne, 2006). Back to the 1957 and 1968, around the time of the Asian and Hong Kong pandemics, some postulated that the influenza pandemic will recur as frequent as every 10-11 years (Gaydos, 2006, Morens and Fauci, 2007). However, according to the influenza expert, Edwin Kilbourne, Sr., he concluded that there is no predictable cyclicity of the major influenza outbreaks, with all of the occurrences differ from one another (Kilbourne, 2006).

Historically, the Spanish pandemic (H1N1-subtyped) that occurred in 1918-1919 was among the most deadly pandemics which had killed about 50-100 million people worldwide. Thus far, the outbreak was characterized by its exceptionally high mortality in the 20th century, especially among the young adults aged in between 15-45 years old (Morens and Fauci, 2007, Reid et al., 2000). The lack of a complete explanation for the origin of 1918 Spanish pandemic has always been disputed. However, according to Kilbourne (2006), epidemiologic-epizootiologic evidence strongly suggested that the causative influenza virus of the pandemic was actually transmitted from humans to swine rather than the reverse direction. On the other hand, Taubenberger and colleagues reported contrary to the previous research group. Instead of the 1918 virus arose by gene reassortment between human and animal

virus, they claimed the whole process occurred by genome adaptation (Morens and Fauci, 2007, Taubenberger et al., 1997, Taubenberger et al., 2005).

After a few decades, another pandemic outbreak of Asian flu occurred in 1957. When substantial amino acid changes had accumulated in the circulating H1N1 influenza virus from the very first pandemic outbreak in 1918, the influenza virus was therefore drifted to become a novel emergent of H2N2-subtyped influenza virus. The disease achieved a rapid global spread to the public with the exception of the elderly people aged 70-year-old and above. The 1957 pandemic had called for the different vaccine to respond on the novel HA and NA antigens of H2N2-subtyped influenza. Unlike the 1918 pandemic, Asian flu viruses (H2N2-subtyped) emerged for only 11 years in the human population before being substituted by the Hong Kong flu viruses (H3N2-subtyped) (Kilbourne, 2006).

Another novel pandemic Hong Kong flu, of H3N2-subtyped influenza virus came into existence in the human population in 1968. This pandemic was named in accordance to its site of emergence and it was brought by the H3N2 strain influenza viruses that arose by gene reassortment. However, widespread immunity to N2 from the past exposure to H2N2 flu had helped to lessen the severity of 1968 pandemic flu. It has been estimated that 700,000 people died from this pandemic (Reid et al., 2000).

Among all the pandemics, only the pandemic Mexican flu happened in the 21st century. Mexican flu was initiated by the H1N1-subtyped influenza viruses, which was similar to the influenza subtype that caused 1918 pandemic. However, the 2009 outbreak was again announced as the pandemic rather than epidemic due to the emergence of the resistant H1N1 strains. Evidence suggested that pandemic flu outbreak as the result of genetic reassortment and recombination with different

strains of influenza A viruses to produce a novel surface protein, either the HA or NA or both happened to be novel antigen variants on the viral surface.

2.4 Evolution of influenza A viruses

The hallmark of influenza viruses is antigenic variation, which they evolve continually over time by two different means: antigenic drift and antigenic shift (as shown in Figure 2.6). Evolution of the influenza is credited to the mutability of the segmented RNA genome and thus contributes to the recurrent annual epidemics of the infectious disease. Accumulation of the wide antigenic diversity will cause the emergence of novel influenza A viruses in human population that has little immunity towards the disease (Hay et al., 2001, McDonald et al., 2007, Morens and Fauci, 2007).

2.4.1 Antigenic drift

According to Treanor (2004), antigenic drift is explained as a subtler process than antigenic shift that accumulates point mutations within the antibody-binding sites in the HA or / and NA that abrogate the binding of some antibody. The amino acid replacements will result in the viruses cannot be inhibited well by the antibodies induced by the previous infectious strains. In addition to mutation, Scholtissek (1994) also associated the phenomenon as a selection of influenza variants under the selection pressure of the host immunity. It is thus easier for the viruses to intrude on a partially immune population. Thus, it is necessary to update the vaccine

components consistently to confer maximum immunity to the human populations against the circulating strains (McDonald et al., 2007, Plotkin et al., 2002).

2.4.2 Antigenic shift

This phenomenon of amino acid replacements produce a new strain of influenza A viruses that totally escape from the host immunity. The mutations are mainly found in the epitope regions of the surface glycoproteins (HA and NA). As influenza A viruses have a broad range of susceptible host, reassortment and recombination of genetic materials from different animal species are commonly detected. The resulted changes will thus create novel subtypes that have not been present in human viruses. Consequently, the introduction of a novel strain into human population is usually a pandemic or a worldwide epidemic that will result in hundreds of thousands or millions of influenza-related deaths (Scholtissek, 1994, Treanor, 2004).

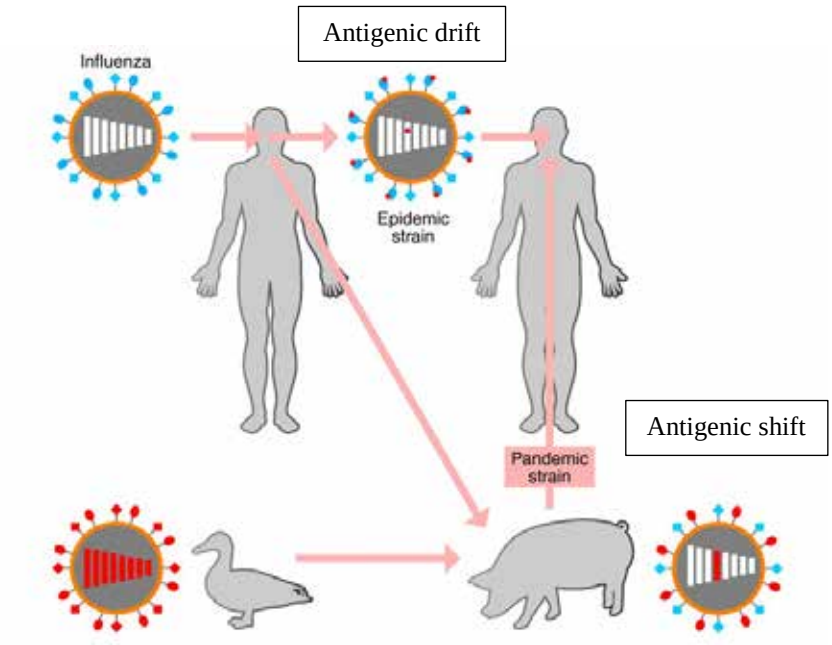

Figure 2.6: Phenomenon of antigenic drift that occurred within the same animal species while antigenic shift involved combination of genomes from different animal species (Taken from: Turner et al., 2010).

2.5 Anti-influenza therapy

Influenza continues to be a major threat to people around the globe. Evolution by antigenic drift caused the regular recurrence of influenza epidemics and numerous scientific studies have revealed that sufficient changes will produce drifted virus to reinfect the same host (Carrat and Flahault, 2007). To avoid and recover from the infection, influenza vaccine and the use of antiviral agents, respectively, are recommended.

2.5.1 Influenza vaccine

Vaccine remains the most effective defense for the prophylaxis and prevention from influenza infections (Gust et al., 2001). However, due to the regular emergence of drifted viruses, a review on the influenza vaccine content is needed consistently (McDonald et al., 2007, Plotkin et al., 2002). The strain selection is an ongoing and complex process whereby it involves the concerted and coordinated efforts from national and international organisations. A year round surveillance of new influenza strains is conducted by the Centers for Disease Control and Prevention and the WHO (Treanor, 2004). As a result, antigenic mismatch can be avoided and maximum vaccine effectiveness can be delivered (Wood, 2002).

Anti-HA antibody provides a better protection against the drifted influenza variants. They are playing the protective role in shielding the recognition and attachment between viral HA and host sialic acid receptors. Thus, the influenza infection becomes impossible when the first stage of the viral infection is disturbed (Virelizier, 1975).

However, anti-NA antibody in host is triggered by the vaccine strains and they serve to inhibit the release of the progeny virus particles from infected cells. The inhibited NA proteins of influenza virus are unable to remove the sialic acid receptors attached to the HA and thus causing self-aggregation at the surface of infected cells (Bouvier and Palese, 2008, Palese et al., 1974, Webster et al., 1982).

In a nutshell, antibody against the HA is neutralising and is very protective against infection and illness while antibody against the NA serves to mitigate the severity of illness (Treanor, 2004).

2.5.2 Antiviral agents

The unique replication mechanism of influenza virus has allowed investigators to identify a number of potential molecular targets for drug design. However, there are only two established antiviral agents that are available in the current marketplace. Both types of antiviral agents have two different kinds of inhibitory mechanisms, either targeting on M2 ion channel or neuraminidase glycoprotein, respectively.

2.5.2(a) M2 ion channel blockers

M2 ion channel is a specific target for anti-influenza drugs amantadine and rimantadine (Chizhmakov et al., 1996, Holsinger et al., 1994). Amantadine and Rimantadine (as shown in Figure 2.7) are two drugs belong to the Adamantanes group (Schnell and Chou, 2008, Zaraket et al., 2010) which has been introduced for prophylaxis and treatment of influenza A disease for many years. Both drugs are proven inactive against influenza B viruses which lack of the M2 membrane-spanning proteins (Sansom et al., 1997).

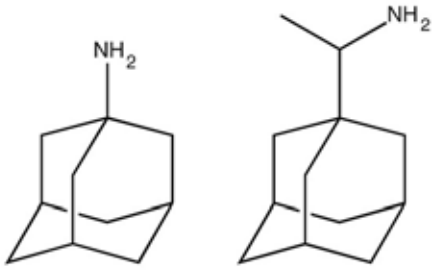

Figure 2.7: The structure of M2 ion channel blockers: amantadine (left) and rimantadine (right) (Taken from: Astrahan and Arkin, 2010).

In designing drugs to exert inhibitory effects on the wild-type M2 (WT-M2) ion channel, two mechanisms of action were proposed: the drug that blocks the channel by binding at the opening of the pore (Astrahan et al., 2004), and the drug that binds firmly to the side chains of H37 in a deeper location of the channel (Pinto and Lamb, 2007, Zhong et al., 2000). These proposed mechanisms were subsequently proven in the experimental and theoretical studies (Chen et al., 2007, Hu et al., 2006, Intharathep et al., 2008, Kass and Arkin, 2005, Laohpongspaisan et al., 2009).

However, studies revealed that their use has been restricted as the adverse effects on the intestinal and central nervous systems arose and the emergence of resistance towards the drugs (Englund et al., 1998, Hayden et al., 1989, McKimm-Breschkin et al., 2003). The resistance to these drugs is genetically associated with amino acid substitutions at positions L26, V27, A30, S31, G34, L38 and W41 in the membrane-spanning domain of the protein (Englund et al., 1998, Holsinger et al., 1994, Saito et al., 2006). Among these mutations, A30T and S31N are the most common mutations for influenza A viruses from Vietnam, Malaysia, Cambodia and Thailand in SEA region that confer resistance towards amantadine (Hurt et al., 2007, Ilyushina et al., 2005).

For V27A and S31N mutations, the replacement of valine to alanine had caused the widening of channel's pore size while the latter mutation led to the reduction in channel's pore size. The size of the channel's pore is associated to the changes in the H^+ influx (Astrahan and Arkin, 2010). Laohpongspaisan and coworkers (2009) had also studied the proton conductance in terms of the water density across the channel. However, zero water density was reported associated to the L26I mutation while a slight degree of water density was traced in the presence

of A30T mutation (Laohpongspaisan et al., 2009). The pH of M2 channel will ultimately determine the opening of the channel, of which the protonated H37 residue and the rotation of W41 side chain to a conformation parallel to the pore's axis will allow the influx of H^+ (Betakova, 2007).

2.5.2(b) Neuraminidase inhibitors

Prior to the existence of both zanamivir and oseltamivir, 2-deoxy-2,3-didehydro-D-*N*-acetylneuraminic acid (DANA, as shown in Figure 2.8), an unsaturated transition-state sialic acid analog was first designed and synthesized to inhibit the NA enzyme activity. However, its inhibitory activity on NA is reported to be weak *in vitro* and vice versa when administered *in vivo* (McKimm-Breschkin, 2000). As a result, a review on DANA drug design is demanded in order to produce a NA antiviral drug with better inhibitory activity.

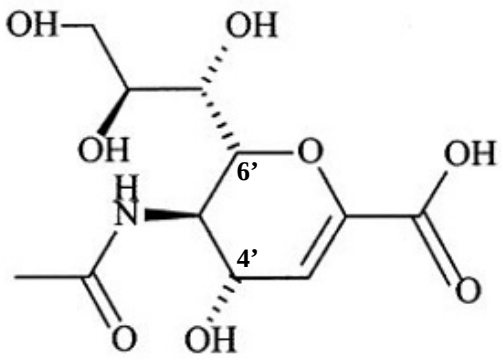

Figure 2.8: Chemical structure of unsaturated sialic acid analog (DANA) (Taken from: Varghese et al., 1998).

Drug design based on the complex structure of viral NA protein and its cell surface receptor (sialic acid), such as zanamivir (GG167) (as shown in Figure 2.9) is efficacious for inhibiting the influenza infection. Zanamivir has a different conformation than the previous NA inhibitor where the 4' position on the sugar ring of DANA is replaced by a guanidinium group. There are two ways of administration, either through oral inhalation or intranasal instillation. High concentration of the inhaled drug will be delivered to the respiratory tract (Gubareva et al., 2000), where the virus is replicating (McKimm-Breschkin et al., 2003, McKimm-Breschkin, 2000, Smith et al., 2001).

Subsequently, oseltamivir (as shown in Figure 2.9) was developed based on the need for antiviral to be efficacious as an orally administered NA inhibitor. Instead of a single substitution on the sugar ring in zanamivir, oseltamivir has a cyclohexene ring with a substitution of an amino group at the 4' position in the ring and more importantly the glycerol side chain at 6' position is replaced by a bulky hydrophobic pentyl ether group (McKimm-Breschkin et al., 2003, McKimm-Breschkin, 2000, McKimm-Breschkin et al., 2007). Oseltamivir exists in two forms: oseltamivir carboxylate (GS4071) and oseltamivir phosphate (GS4104) (Lew et al., 2000). Usually, it is taken orally in the form of ethyl ester prodrug oseltamivir phosphate, which is subsequently converted by hepatic esterases to the active form of oseltamivir, called oseltamivir carboxylate (McKimm-Breschkin et al., 2003, McKimm-Breschkin, 2000, Smith et al., 2001).