IDENTIFICATION OF PEPTIDE CONTAINING T CELL EPITOPE OF HEMAGGLUTININ PROTEIN IN H5N1 INFLUENZA VIRUS

A thesis submitted in partial fulfillment of the requirements for the award

Of the

degree of

MASTER OF SCIENCE IN MICROBIOLOGY

SUBMITTED BY

NEHA KUMARI

REGISTRATION NO. 301205008

Under the Supervision of
DR. MANOJ BARANWAL

(Assistant Professor)

DEPARTMENT OF BIOTECHNOLOGY

THAPAR UNIVERSITY, PATIALA

JULY, 2014

CERTIFICATE

This is to certify that the thesis entitled "**Identification of peptide of containing T cell epitope of hemagglutinin protien in H5N1 influenza virus**" being submitted by Neha Kumari registration no. 301205008 in the partial fulfillment of the requirements for the award of degree of Master of Science in Microbiology , Department of Biotechnology, Thapar University, Patiala is a bonafide work carried out under my supervision and guidance. The thesis has not been submitted for award of any other degree or certificate in this or any other university.

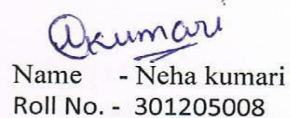
Dr. Manoj Baranwal
Supervisor
DBT
Thapar University, Patiala

Goyal Dr. Dinesh
Head of Department
DBT
Thapar University, Patiala

Dr. S. K. Mohapatra
Dean, Academic Affairs
Thapar University, Patiala

Candidate's Declaration

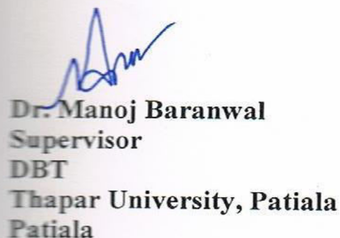
I hereby declare that the work which is being presented in the dissertation entitled – "**Identification of peptide of containing T cell epitope of hemagglutinin protien in H5N1 influenza virus**" in partial fulfillment of the requirements for the award of Master in Science in Microbiology, Department of Biotechnology, Thapar University, Patiala is an authentic record of my own work during a period of six months from January 2014 to July 2014, under the supervision of Dr. Manoj Baranwal, Assistant Professor, Department of Biotechnology, Thapar University, Patiala. The report has not been submitted for the award of any other degree or certificate in this or any other university.



Place: Patiala
Date: 18 July 2014

Name - Neha kumari
Roll No. - 301205008

This is to certify that the above statement given by the candidate is correct and true to the best of our knowledge.



Dr. Manoj Baranwal
Supervisor
DBT
Thapar University, Patiala
Patiala

Dr. Dinesh Goyal
Head Of Department
DBT
Thapar University,

ACKNOWLEDGEMENT

This is a golden opportunity for me to convey my sincere regards to all those people who enabled me to accomplish my Training successfully.

It is my privilege to express my sincere gratitude to Dr. Manoj Baranwal, Assistant Professor, Department of Biotechnology and Environmental Sciences, Thapar University, Patiala for his guidance, encouragement and valuable advice. It is his confidence imbibing attitude and splendid discussions and endless endeavors through which I have gained significant experience.

I feel deep gratitude towards **Dr. Dinesh Goyal**, Head of Department, Department of Biotechnology for his immense concern throughout the project work. I am thankful to **Ms. Neha Lohia**, Research scholar for her favour and concern without which it would have been difficult for me to achieve my goals.

I thank all faculty and staff members, Department of Biotechnology and Environmental Sciences, Thapar University, Patiala, for their constant encouragement and support throughout the Project.

I don't have words to express my thanks to my parents & my sister for their selfless sacrifice, encouragement and heart full blessings that continue to enlighten my life.

In the end, I am thankful to the Almighty for blessing me to complete this work successfully.

Place : Patiala Neha kumari

Date : 18 JULY , 2014

ABBREVIATIONS

RNA - Ribonucleic acid

WHO - World Health Organization

HLA - Human Leukocyte Antigen

MHC - Major Histocompatibility Complex

HA - Hemagglutinin

NA - Neuraminidase

NP - Nucleoprotein

M1 - Matrix protein 1

M2 - Matrix protein 2

RNP - Ribonucleoprotein

NEP - Nuclear export protein

NS1 - Non- Structural protein

IFN - Interferon

PA, PB - Polymerase protein A, B

APC - Antigen presenting cells

MUSCLE - Multiple Sequence Comparison By Log Expectation

AVANA - Antigen Variability Analyzer

NCBI - National Council of Biological Information

MSA - Multiple Sequence Alignment

CD - Cluster of Differentiation

IEDB - Immune Epitope Database

CTL - Cytotoxic T- Lymphocytes

BIMAS - Bioinformatics and Molecular Analysis Section

ANN - Artificial Neural Network

SMM - Stabilized Matrix Method

HMM - Hidden Markov Model

Table of Contents

Sr. No.	Contents	Page No.
1	Introduction	2
2	Review of literature	5
2.1	Influenza Virus	5
2.2	Influenza virus A: Structure	6
2.3	Life Cycle of Influenza	7
2.4	Hemagglutinin protein	9
2.5	H5N1	11
2.6	Vaccine	12
2.7	Peptide based vaccine	12
2.8	Immunoinformatics	13
3	Methodology	15
3.1	Conservency Analysis	15
3.2	T Cell Epitope Prediction	16
3.3	BLAST Analysis	19
4	Objective	21
5	Result	22
6	Discussion	34
7	Summary	36
8	References	38

ABSTRACT

HPAI A(H5N1) highly pathogenic avian influenza virus of type A of subtype H5N1 is a highly pathogenic causative of Avian flu in human . Conserved regions from Hemagglutinin (HA) of H5N1 of influenza virus A strains isolated from human hosts were identified (≥ 80% conservation) by computational tool based on informational entropy. Epitopes for HLA alleles of MHC class I and class II from the conserved regions were predicted using three different epitope prediction tools separately for MHC class I and II. Predicted epitopes were screened for the presence of any similarity with human self-proteins. 27 MHC class I and 27 MHC class II epitope were identified. Overlapping epitopes were merged to generate putative immunogenic peptide containing both MHC class I and Class II predicted epitopes. Seven such sequences were identified, having minimum conservation 81%. One of the sequences RRIENLNKKMEDGFLDVWTYNAELLVLM (CS14) has shown 100% conservation followed by NSSMPFHNIHPLTI (CS12) which was 98.92% conserved. All these sequence can be considered as the potent candidates for vaccine design against H5N1.

1

1.INTRODUCTION

Influenza (commonly referred to as "flu") is a contagious viral infection that affects mainly the upper respiratory organs (nose, throat, bronchi and, occasionally lungs). It is caused by RNA viruses of the family Orthomyxoviridae, the influenza viruses. The virus is transmitted easily from person to person via respiratory droplets produced when people cough, sneeze or spit. Close contact (less than 1 metre) is usually needed to get infected. People usually recover after a few days of what can be severe illness, but it can cause complications, including deaths especially in certain risk groups including pregnant women. Symptoms include high fever, body aches, headache and severe malaise, cough, sore throat and runny nose. Influenza tends to spread rapidly in seasonal epidemics. These flu epidemics occur yearly during autumn and winter in temperate regions and affect ~5% of adults and ~20% of children each year.

H5N1 is a highly pathogenic avian (bird) flu virus that has caused serious outbreaks in domestic poultry in parts of Asia and the Middle East. Although H5N1 does not usually infect humans, nearly 650 cases of human cases of H5N1 have been reported from 15 countries since 2003 (WHO/H5N1 cases/2014). In 1997, Bird flu virus H5N1 is isolated for the first time from a human patient in Hong Kong, but fortunately it failed human-to-human transmission (Claas et al., 1998). Type A influenza strains are classified by the serological subtypes of the primary viral surface proteins, the hemagglutinin (HA) and neuraminidase (NA). There are 18 different HA subtypes (H1 to H18) and 11 different NA subtypes (N1 to N11) (Mänz at el, 2013).

Antigenic variation is the evolutionary mechanism by which viruses evade host immune system. Influenza viruses accomplish this through one of the two mechanism: (1) Antigenic shift (genetic reassortment between a human and non-human virus in a non-human host), and (2) Antigenic drift (accumulation of mutation that facilitate evasion of the host-immune response).

Range of medications and therapies are available to treat Influenza including different classes of antiviral drug. There are two classes of such medicines, adamantanes (amantadine and remantadine), and inhibitors of influenza neuraminidase (oseltamivir and zanamivir). But some influenza viruses develop resistance to the antiviral medicines, limiting the effectiveness of treatment and moreover these antiviral drugs are effective only when administered within a

certain time frame after exposure. Also, antiviral drugs do not impart immunity against influenza.

Vaccines is considered to be one of the most effective way to control influenza. Although antiviral vaccines are available, which are effective against limited strains of influenza virus, so influenza remains a serious respiratory disease. Therefore, there is an urgent call for the development of a universal vaccine, which could be protective against the existing and future virus strains of H5N1. Out of various types of vaccines being designed, epitope based vaccine is a novel approach since it offers appropriate choice of well-defined antigenic epitopes with desired MHC restrictions and relative ease of production. Unlike attenuated vaccines, these vaccine lacks the risk of reversion and also lacks the risk of genetic integration or recombination like DNA based Vaccine. This approach can be adopted for the development of cross strain influenza.

Current study was undertaken to identify conserved immunogenic peptides of hemagglutinin (HA) protein (H5N1) containing MHC class I and Class II restricted epitope using various immunoinformatics tools .

2. REVIEW OF LITERATURE

Influenza is a viral infection caused by segmented negative sense single stranded RNA viruses of the family Orthomyxoviridae that affects mainly the nose, throat, bronchi and, occasionally, lungs. Infection usually lasts for about a week, and is characterized by sudden onset of high fever, aching muscles, headache and severe malaise, non-productive cough, sore throat and rhinitis (http://www.who.int/topics/ influenza/en/). The highly pathogenic avian influenza (HPAI) H5N1 virus, which is panzootic in poultry, continues to spread and pose a major challenge to animal and human health (Peiris *et al.,* 2007).

2.1 Influenza Virus

Orthomyxoviridae is a family of Group V of RNA virus that includes five generas which are Influenza virus A, Influenza virus B, Influenza virus C, Isavirus and Thogotovirus. Out of three types of influenza viruses, only type A and B are associated with major outbreaks of influenza (Zambon,1999). Influenza C virus commonly infects children, but does not cause serious disease. Influenza virus undergoes minor (antigenic drift) as well as major antigenic changes (antigenic shift) which allows the virus to overcome the adaptive immune response of mammalian and avian species. (Zambon, 1999).

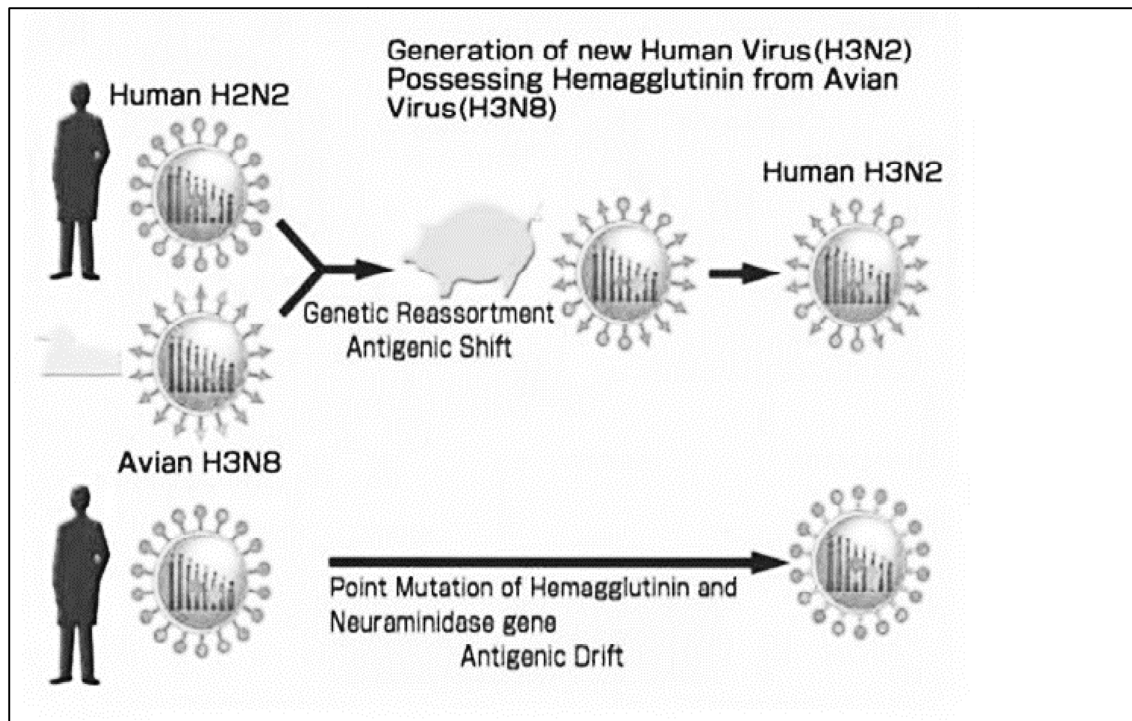


Figure 1: Antigenic shift and antigenic drift

Antigenic drift involves point mutation, each time of virus replicates, leading to the rapid spread of new viral strain among the populationcirculated decades before. Antigenic drift occurs in all strains of A and B viruses whereas antigenic shift is observed only in case in influenza of influenza A virus. Antigenic shift is the process of genetic reassortment (mixing of genetic material between different viral strains), giving rise to a new virus never been present in human circulation or last

Antigenic drift occurs in all strains of A and B viruses whereas antigenic shift is observed only in case in influenza of influenza A virus. Antigenic shift is the process of genetic reassortment (mixing of genetic material between different viral strains), giving rise to a new virus never been present in human circulation or last. The phenomenon of antigenic shift is only observed in influenza A viruses, which makes them most dangerous among all the influenza viruses. Thus influenza A viruses are responsible for seasonal as well as pandemic outbreaks.

2.2 Influenza virus A: Structure and lifecycle

The type A viruses are the most virulent pathogens among the three influenza types and cause the most severe disease. Influenza type A viruses are divided into subtypes based on two surface

proteins hemagglutinin (HA) and neuraminidase (NA). There are 18 different HA subtypes and 10 different NA subtypes (Tong et al., 2013). Many different combinations of HA and NA proteins are possible.

Influenza viruses are roughly spherical, although somewhat pleomorphic, particles, ranging from 80 to 120 nm in diameter (Webster et al., 1998). It is an enveloped virus, carrying layer is a lipid membrane from the host cell in which the virus multiplies. A characteristic feature of influenza virus particles is their external layer of approximately 500 spike-like projections (Figure 2). These spikes represent the envelope glycoproteins HA (which has a rod-like shape) and NA which is mushroom-shaped (Webster et al., 1998). The HA spike is a trimer, consisting of three individual HA monomers, while the NA spike is a tetramer (Nicholson et al., 1998). The HA and NA are important in the immune response against the virus; antibodies against these spikes may protect against infection.

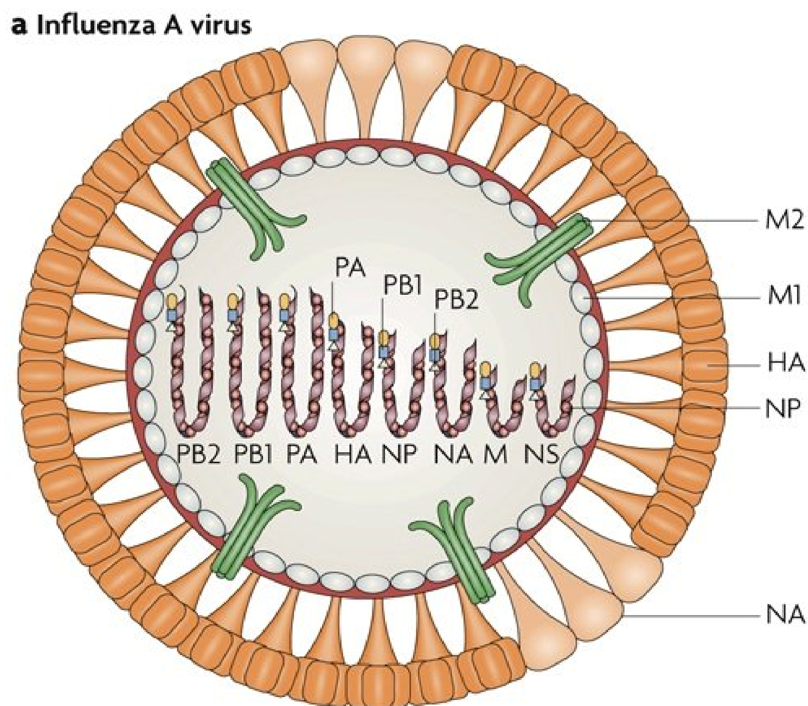


Figure 2: Structure of Influenza virus

***Nature Reviews Immunology* 7, 267-278 (April 2007)**

Beneath the lipid membrane of influenza virus there is a viral protein called M1, i.e matrix protein. M1 protein, which forms a shell, gives strength and rigidity to the lipid enveloped.

Virion consists of eight segments of single stranded, negative sense RNAs, of which, each segment codes for one or two proteins as listed below.

- Segment 1, 2 and 3 codes for viral RNA polymerase Polymerase Basic 1(PB1), Polymerase Basic 2 (PB2) and Polymerase A (PA) respectively. PB2 is involved in synthesis of capped messenger RNA by endonuclease activity which cleaves host cell messenger RNA. PB1 viral polymerase plays role in RNA transcription and replication activities whereas PA in RNA replication.
- Segment 4 codes for Hemagglutinin(HA) which is surface glycoprotein involved in attachment and internalization of the viral particle in the host cell.
- Nuceoprotein (NP) is coded by segment 5. It forms a complex with the viral RNA genome and packages the RNA into a helical ribonucleoprotein core.
- Neuraminidase(NA) is another glycoprotein involved in the release of newly synthesized viral particles, thus playing role in the spread of viral infection.
- Segment 7 codes for two matrix proteins M1 and M2. M1 is a bifunctional membrane/RNA-binding protein that mediates the encapsulation of RNA-nucleoprotein cores into the membrane envelope. M2 is a surface protein which form an ion channel pump to lower or maintain the pH of the endosome
- Two other proteins, Non-structural protein (NS1) and Nuclear export protein (NEP/ NS2) are coded by Segment 8. NEP which mediates the export of influenza virus ribonucleoprotein (RNP) complexes from the nucleus, where they are assembled.

2.3 Life Cycle of Influenza

The influenza virus life cycle can be divided into the five stages (Samji T. *et al.*,2009):

1. Entry of virus particle into the host cell.

2. Entry of viral Ribonucleoprotein (RNP) into the nucleus.

3. Transcription and replication of the viral genome.

4. Export of the vRNPs from the nucleus.

5. Assembly and budding at the host cell plasma membrane.

HA homotrimer of Influenza virus bind to sialic acid found on the surface of the host cell's membrane (Figure 3). After binding to the sialic acid residues, receptor-mediated endocytosis occurs and the virus enters the host cell in an endosome. The endosome has a low pH, which triggers the fusion of the viral and endosomal membranes. The low pH induces a conformational change in HA leading to formation of fusion peptide. This fusion peptide inserts itself into the endosomal membrane, bringing both the viral and endosomal membranes into contact with each other (Huang Q. *et al.*, 2003). The acidic environment of the endosome opens up the M2 ion channel. Opening the M2 ion channels acidifies the viral core. This acidic environment in the virion releases the vRNP from M1 such that vRNP is free to enter the host cell's cytoplasm (Pinto L.H. *et al.*, 2006). The viral proteins that make up the vRNP are NP, PA, PB1, and PB2. All of these proteins have known nuclear localization signals (NLSs) that can bind to the cellular nuclear import machinery and, thus, enter the nucleus.

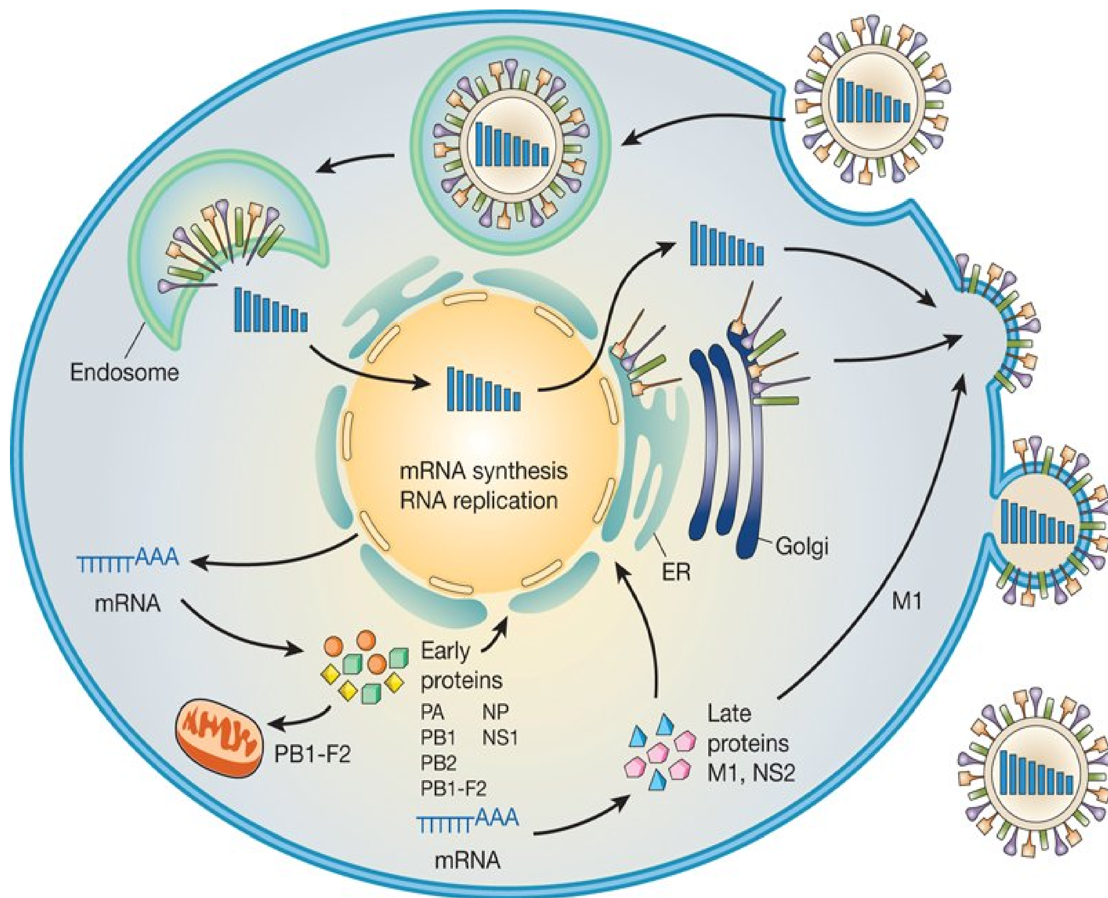


Figure 3: Life Cycle of Influenza Virus

***Nature Reviews Immunology* 459, 931-939 (April 2009)**

In the nucleus, the viral polymerase complexes transcribe and replicate the viral RNAs. Influenza viral genome is negative sense RNA which is first converted into a positive sense RNA. Positive sense RNA is now used as a template for RNA replication which is carried out by viral protein PA. After replication viral RNA is transcribed to mRNA of different viral proteins which is carried out by PB1 protein. Viral mRNA does not have 5' cap so PB2 functions as endonuclease and cleaves 5' methylated caps of cellular mRNA (10 to 15 nucleotides). This cellular capped RNA fragment is used as a primer for viral transcription (Li M.L. *et al.,* 2001). Viral mRNAs migrate to cytoplasm where they are translated. HA, NA and M2 proteins are transferred to cell membrane while other proteins like NP, M1, NS1 and NEP (nuclear export protein) move to the nucleus where they bind to viral RNA forming Ribonucleoprotein complex (RNP). This RNP migrates into the cytoplasm in a NEP-dependent process and eventually interact via M1 with a region of the cell membrane, where HA, NA and M2 have been inserted. Then the newly synthesized virions bud from infected cell. NA destroys the sialic acid moiety of cellular receptors, thereby releasing the progeny virions.

2.4 Hemagglutinin protein

HA is a homotrimeric integral membrane glycoprotein, cylindrical in shape, and is approximately 13.5 nanometres long (Wharton *et al.,* 1989). HA monomers are synthesized as precursors that are then glycosylated and cleaved into two smaller polypeptides: the HA1 and HA2 subunits (Figure 4). Each HA monomer consists of a long, helical chain anchored in the membrane by HA2 and topped by a large HA1 globule (Seto*et al.,* 1966). Hemagglutinin is essential for the entry of virus in the target cell and also it is the primary target of neutralizing antibodies. HA has two functions, firstly, it allows the recognition of target vertebrate cells, accomplished through the binding to these cell's sialic acid-containing receptors. Secondly, once bound it facilitates the entry of the viral genome into the target cells by causing the fusion of host endosomal membrane with the viral membrane (Varghese *et al.,* 1983).

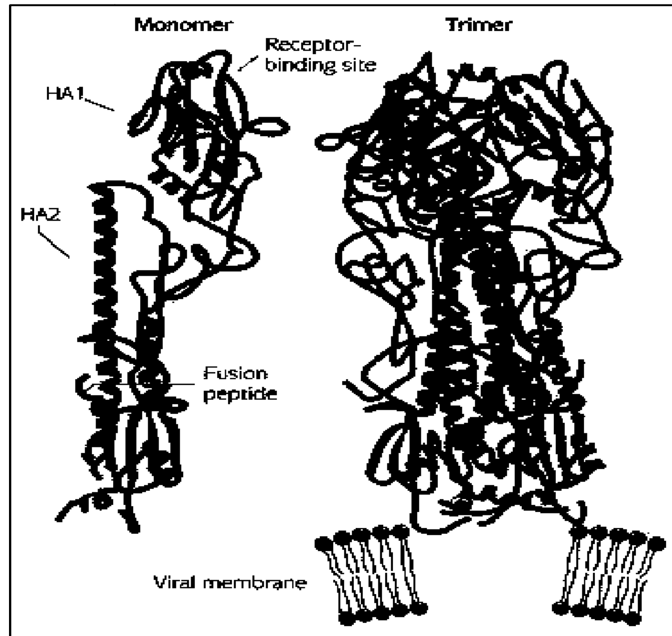


Figure 4: Structure of Hemagglutinin protein

HA binds to the monosaccharide sialic acid which is present on the surface of its target cells, which causes the viral particles to stick to the cell's surface. The cell membrane then engulfs the virus and the portion of the membrane that encloses it pinches off to form a new membrane-bound compartment within the cell called an endosome, which contains the engulfed virus (Kielian*et al.*, 1990). The cell begin digesting the contents of the endosome by acidifying its interior and transforming it into a lysosome and when the pH within the endosome drops to about 6.0, the original folded structure of the HA molecule becomes unstable, causing it to partially unfold and release a very hydrophobic portion of its peptide chain that was previously hidden within the protein, called the "fusion peptide" acts like a hook by inserting itself into the endosomal membrane. When the rest of the HA molecule refolds into a new structure (which is more stable at the lower pH), it "retracts the hook" and pulls the endosomal membrane right up next to the virus particle's own membrane, causing the two to fuse together. Once this has happened, the contents of the virus, including its RNA genome, are free to pour out into the cell's cytoplasm.

2.5 H5N1

H5N1 is a highly pathogenic avian flu virus that has caused serious outbreaks in domestic poultry in parts of Asia and the Middle East. Highly pathogenic refers to the virus's ability to produce disease. Although H5N1 does not usually infect humans, nearly 650 cases of human cases of H5N1 have been reported from 15 countries since 2003 (WHO/ 2014). Most human cases of "highly pathogenic" H5N1 virus infection have in people who had recent contact with sick or dead poultry that were infected with H5N1 viruses. About 60% of people infected with the virus died from their illness. H5N1 usually does not spread between people.There have been no reported infections with these viruses in birds, poultry, or people in the United States. H5N1 is an avian (bird) flu virus that has caused outbreaks in domestic poultry in parts of Asia and the Middle East. Because H5N1 is so deadly to poultry, it is considered "highly pathogenic," or highly disease causing. (Nguyen DH.et al, 2008).

In the recent past (2011), 62 human H5N1 cases and 34 deaths were reported from five countries—Bangladesh, Cambodia, China, Egypt, and Indonesia(http://www.flu.gov/planning-preparedness/index.html). Six countries— Bangladesh, China, Egypt, India, Indonesia, and Vietnam—have widespread and ongoing infections in their poultry. Poultry outbreaks have occurred in other countries recently as well.

In the symptoms of H5N1 infection are fever and cough, acute respiratory distress, shortness of breath/difficulty breathing, abdominal pain, Diarrhea. A person infected with H5N1 may develop complications like Pneumonia,Respiratory failure,Shock,Altered mental state,Seizures,Failure of multiple organs (e.g. kidney failure),Death.

Oseltamivir and zanamivir, two well-known antiviral drugs are recommended for treatment (and prevention) of human infection with avian influenza A viruses by CDC and WHO. Analyses of available HPAI (highly pathogenic avian flu virus)/H5N1 viruses circulating worldwide suggest that most viruses are susceptible to oseltamivir and zanamivir. However, some evidence of resistance to oseltamivir has been reported in HPAI H5N1 viruses isolated from some human HPAI H5N1 cases (Puthavathana et al., 2005).

2.6 Status and Need for vaccine

Vaccination is the most effective public health intervention strategy and this must be supported by enhanced surveillance. Although the use of antiviral drugs is also an important public health countermeasure for preventing and treating influenza, the emergence of drug-resistant strains of avian H5N1 viruses (Puthavathana et al., 2005) underscores the need for new drugs and other novel preventive and therapeutic strategies. While vaccines are the most effective tool for the control of influenza (Nichol and Treanor, 2006), the combined production capacity of global vaccine suppliers is not sufficient to meet the demand during a pandemic, so a vaccine shortage is expected. Any strategy that can maximize vaccine coverage will be valuable in a pandemic. Development of effective vaccines against HPAI H5N1 viruses presented several challenges. Since, H5N1 has high pathogenicity for poultry, handling of the wild-type (wt) viruses requires enhanced biosafety level 3 (BSL-3) containment (Peiris*et al.,* 2007). Also, these viruses are embryo-lethal so it is difficult to propagate the wt viruses efficiently and safely in eggs.

Several strategies have been explored in attempts to overcome these obstacles including the use of a low pathogenicity H5N1 avian influenza virus that is antigenically related to circulating strains (Nicholson et al., 2001), the use of recombinant H5 HA expressed in baculovirus (Treanor et al., 2001), the use of recombinant adenoviruses expressing the H5 HA (Hoelscher et al., 2006), and the production of attenuated seed viruses with the H5 HA modified by reverse genetics (Webby et al., 2004). Recently in November 2013, the FDA approved an experimental H5N1 bird flu vaccine with reactive AS03 (containing squalene) adjuvant (GlaxoSmithKline's Q-Panfor) the U.S. stockpile in 2013 (LaVigne, 2013).

Also, due to continuous mutation in H5N1, vaccines based on current samples of avian H5N1 may not provide immunity against in the case of a future pandemic of H5N1.

2.6 Peptide based vaccine

Peptides based vaccine development has gained pace in recent past over the conventional whole protein vaccine. Peptide based vaccines are attractive because of their stability, cost effectiveness and easy manufacturing (Ovsyannikova., 2007). The peptide based vaccines are advantageous as they are devoid of deleterious sequences which may give rise to some autoimmune diseases. The addition of different organic groups is easy in case of peptide based vaccines which not only improve the stability but also immunogenicity. The absence of infectious material, which can

compromise many live or attenuated vaccines. At the same time many pathogens can be difficult or impossible to culture by conventional methods. The addition of different natural adjuvants with the peptide based vaccines has been widely tested and has given good clinical results. Mainly GM-CSF has been used as adjuvant with the peptide based cancer vaccine that has given promising results for evoking immune response (Purcell et al., 2007). The peptide based vaccines provides a more rational approach towards the vaccine development with several additive advantages, thus it is a viable option for improved vaccine development in modern therpeuutics.

2.7 Immunoinformatics

The advancement in the genome and proteome sequencing has resulted in the deposition of huge database different biomolecular sequences. The knowledge of variety of B-cell and T-cell epitopes has given a new direction to the immunological research. As a result of which immunoinformatics has emerged as an important area for the development of varied tools for epitope prediction and vaccine designs with the help of these databases. A variety of T-cell and B-cell epitope prediction tools has been developed and are being developed with continuous improvement using varied algorithms.

2.7.1 Epitope Prediction Using Machine Learning Methodologies

Machine learning methodologies are being widely used by the researcher for the identification of epitope characteristics and prediction algorithm development. For example in ABCPREDartificial neural networks (ANNs) is used for B-cell epitope prediction (Saha et al., 2006), similarly Sweredoski and Baldi presented COBEPRO using a support vector machine (SVM) (Sweredoski et al., 2009). The above algrithms has been mainly used for the B-cell epitope prediction.

2.7.2 Epitope prediction by using Matrix driven method

The method uses the BLOSSOM matrix with the amino acid indicator for the direct prediction of the epitopes. The MMBPredserver (Bhasin et al., 2003) predicts the mutated promiscuous and high-affinity MHC binding peptide. It uses the matrix data in a linear prediction model whereas the structure of the peptide is not taken into consideration

2.7.3 Prediction through ANN (Artifical Neural Network)

Artificial neural network has been used in the development of several T-cell epitope prediction serevers. NetCTL server the is based on the ANN it uses a method to integrate the prediction of peptide MHC class I binding, proteasomal C-terminal cleavage and TAP transport efficiency(Larsen et al., 2007). Similarly the NETMHC uses the ANN along with weight matrix for the prediction algorithms (Lundegaard., 2008).

2.7.4 Structure-based prediction methods

Three dimensional QSAR based technology CoMSIA has been used for the epitope prediction. It align the three dimensional structure of the peptide to predict the interaction and binding ability. TEPITOPE (Bian et al., 2004) uses the structure based prediction algorithm for class to binding ability with peptide.

Various immunoinformatics tools have been employed to identify epitopes which can be potent candidates for peptide base vaccines against HPAI A (H5N1) highly pathogenic avian influenza virus. Cheung and co- workers in 2012 indentified, two novel HLA-A2.1 specific H5N1 nucleoprotein epitopes (NP373-381 AMDSNTLEL and NP458-466 FQGRGVFEL) capable of activating cytotoxic T-cells, in vitro. Somvanshi and colleagues in 2008 made an attempt to identify immunogenic epitopes of HA and NA (H5N1) using Propred I (MHC class I) and Propred (MHC class II) tools.

3. METHODOLOGY

1) CONSERVANCY ANALYSIS

a) Sequence retrieval

The sequences of hemagglutinin protein of H5N1 were retrieved from NCBI Influenza database (http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi) from January 1997 to December 2013. Full length sequence were downloaded after removing the duplicate sequences using the option available in the database. Some sequences had an invalid letter code 'J' which does not represent any amino acid, and so 'J' was replaced with 'X'.

b) Multiple Sequence alignment

Multiple sequence alignment was carried out using Clustal Omega (Sievers et. al., 2011). Clustal Omega in an iterative method of multiple sequence alignment which can align virtually any number of protein sequences quickly and that delivers accurate alignments

c) Identification of conserved Regions

Conservancy analysis was carried out to find out the conserved region present in Hemagglutinin Protein of H5N1 subtype of influenza virus. The regions of Hemagglutinin showing ≥ 80% conservancy were selected. AVANA tool (Antigen Variability ANAlyzer) was used for conservancy analysis.

AVANA (Miotto et al., 2008) tool uses information entropy to measure variability in protein sequence alignments. It also compares alignments using mutual information, identifying the mutations that characterize specific sequence sets.

Information entropy is a measure of the uncertainty associated with a random variable or in case of protein mutations occurring in the protein sequence. AVANA uses Shannon Entropy analysis (Shannon C.E., 1948) in account to measure the informational entropy. Shannon entropy is measured by the following formula.

$$H(x) = -\sum_{i=1}^{I} P_i(x) \log_2(P_i(x))$$

Where H is the entropy, x is the position in the Multiple Sequence Alignment, i represents a given individual amino acid at position x, I is the number of different amino acids on position x, and $Pi(x)$ is the frequency of amino acid i at position x (Olsen L.R. *et al.*, 2011). The conservation of a given position is defined as the frequency of the consensus amino acid (most frequent at a given position) present at that position.

Alignment result of Clustal Omega (FASTA format) was used as input for AVANA software. Parameters were set to 80% conservancy and minimum length of 9 amino acids as a threshold value in AVANA. Conserved regions were searched in the alignments. The overlapping regions found in the conserved regions were merged together to obtain final set of conserved sequences.

2) T CELL EPITOPE PREDICTION

In epitope based vaccine, binding of the immunogenic epitope to the MHC molecule is the most crucial part for activation of antigen specific T-cell. So primary objective in these prediction methodologies is the calculation of MHC peptide binding because high affinity binding often correlates with higher immunogenicity and this depends on the stability of MHC peptide complexes (Burg et al., 1996). There are two different type of epitope one which binds to MHC class I and other to MHC class II. Various algorithms based online tools were used for prediction of both MHC Class I and MHC Class II binding peptide.

a) MHC Class I Epitope Prediction

Class I MHC molecules bind peptides and present them to $CD8^+$ T cells. In general, these peptides are derived from endogenous intracellular proteins that are digested in the cytosol. The MHC class I molecules bound peptides commonly have a length of nine amino acids and they contain specific amino acid residues that appear to be essential for binding to a particular MHC molecule. Three different immunoinformatics tools having different epitope prediction algorithms were used to identify MHC class I epitope.

NetCTL 1.2 (http://www.cbs.dtu.dk/services/NetCTL/)

NetCTL 1.2 server predicts CTL epitopes in protein sequences. The current version 1.2 can do MHC class I binding prediction for 12 MHC supertypes including A1, A2, A3, A24, A26, B7, B8, B27, B39,B44, B58, B62. NetCTL predicts the epitope on the basis of a multi-step algorithm. The method integrates prediction of peptide MHC class I binding proteasomal C terminal cleavage and TAP transport efficiency. MHC class I binding and proteasomal cleavage is performed using artificial neural networks. TAP transport efficiency is predicted using weight matrix. The scores from the three individual prediction methods are integrated as a weighted sum with a relative weight on peptide/MHC binding of 1(Larsen *et al*., 2007).

Conserved regions of Hemagglutinin, H5N1 were taken FASTA format and analysis was carried out for all 12 supertypes. Epitope prediction was carried out at default parameters i.e. weight on C terminal cleavage = 0.15, weight on TAP transport efficiency = 0.05 and threshold for epitope identification = 0.75. Only those epitopes showing score more than the threshold (≥ 0.75) were taken as epitopes.

SYFPEITHI (http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.htm)

SYFPEITHI contains a collection of MHC class I and class II ligands and peptide motifs of humans and other species (Rammensee et al., 1999). The prediction is based on motif matrices of published motifs (pool sequencing, natural ligands) and takes into consideration the amino acids in the anchor and auxiliary anchor positions, as well as other frequent amino acids.

Length of epitope was selected as 9 amino acid (nonamer) and epitopes were searched for all MHC class I molecules. Conserved regions of Hemagglutinin, H5N1 were taken as the input sequence and only those epitopes were taken showing score ≥ 13. Threshold for epitope identification was taken greater than or equal to13.

BIMAS (http://www-bimas.cit.nih.gov/molbio/hla_bind/)

BIMAS stands for Bioinformatics and Molecular Analysis Section (Parker et al., 1994). BIMAS locate and rank 8-mer, 9-mer, or 10-mer peptides that contain peptide-binding motifs for HLA class I molecules. BIMAS works on the principle of Quantitative matrices. In this method, the contribution to binding from each amino acid at each peptide position within the binding groove is quantified (Parker K.C. *et al.*, 1994). It is assumed that each position within the peptide contributes independently in binding to an MHC molecule, and a residue located at a given

peptide position contributes an equal amount to binding, even within different peptides. This method involves producing a matrix in which every entry (X, Y) represents a score associated with amino acid residue X at position Y. The position-specific amino acid values reflect the structural properties of HLA alleles, therefore representing a fingerprint for HLA binding domains.

Conserved regions of Hemagglutinin, H5N1 were taken as the input. Conserved regions were taken FASTA format and analysis was carried out for all HLA class I molecules. Score was selected in the form of T(½) (estimate of half time of disassociation of a molecule containing this subsequence). Different parameters were for epitope identification were predicted T(½) ≥ 5.

Peptides were only defined as epitopes and selected for further analysis when they were commonly predicted by all three tools i.e. NetCTL, SYFPEITHI and BIMAS.

b) MHC Class II Epitope Prediction

Class II MHC molecules bind peptides and present these peptides to CD4+ T cells. Peptides of class II MHC.peptide complexes is generally 13.18 amino acid residues long, somewhat longer than the nonameric peptide of class I molecules but the core sequence is still 9 amino acid in length. The peptide-binding cleft in class II molecules is open at both ends, allowing longer peptides to extend beyond the ends. Three different immunoinformatics tools having different epitope prediction algorithms were used to identify MHC class II epitope.

ProPred (http://www.imtech.res.in/raghava/propred/)

ProPred is a graphical online tool for predicting MHC class II binding regions in antigenic protein sequences. The server implement Quantitative Matrix based prediction algorithm, employing amino-acid / position coefficient table deduced from literature. The predicted binders can be visualized either as peaks in graphical interface or as colored residues in HTML interface (Singh et al., 2001). Conserved regions of Hemagglutinin, H5N1 (amino acid sequence in single letter code) were taken as the input. ProPred carried out MHC class II epitope prediction in 51 HLA-DR alleles. Epitopes were selected at a threshold of 3%.

ProPred measures the percentage score with respect to the best score for that particular allele. In other words if score of a epitope is less than 3% of the best score for that specific HLA allele then it will be considered as a negative result.

IEDB SMM-Align (http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html)

IEDB stands for Immune Epitope Database. IEDB SMM-Align is a tool of IEDB analysis resource which predicts MHC Class II binding epitopes. It follows SMM-QM (Stabilized Matrix Method-Quantitative Matrices) algorithm for prediction which is a modification of original Quantitative Matrices (Nielsen M. et al., 2007). IEDB SMM-Align identifies the MHC class II binding motif in terms of a position specific weight matrix. The output of the SMM-align method is IC50 binding affinity values, enabling direct readout of the peptide-MHC binding affinity. IC50 is half of the maximum inhibitory concentration, is a measure of the effectiveness of a compound. The lower the IC50 of the peptide the higher the higher will be peptide-MHC binding affinity.

Conserved regions of Hemagglutinin, H5N1 were taken as the input. Conserved regions were taken FASTA format and analysis was carried out for all HLA class II molecules. Threshold for epitope identification was taken as IC50≤500.

MHCPred (http://www.ddg-pharmfac.net/mhcpred/MHCPred/help.htm**)**

MHCPred uses the additive method to predict the binding affinity of major histocompatibility complex (MHC) class II molecules and also to the Transporter associated with Processing (TAP). Allele specific Quantitative Structure Activity Relationship (QSAR) models were generated using partial least squares (Guan et al, 2003). A peptide of length≤ 1000 amino acid can be used as input and outcome is generated in the form of IC50 value. Threshold for epitope identification was IC50≤500.

Peptides were only defined as epitopes and selected for further analysis when they were commonly predicted by all three tools i.e. IEDB SMM align, ProPred and MHCpred.

3) BLAST Analysis

The similarity of the host self-protein may lead to an immunogenic response against the self-antigen, which may lead to autoimmunity. Hence, in order to obtain nonself epitopes, epitopes predicted for class I and II MHC molecules were analyzed for their homology with annotated human proteins using BLAST. Predicted epitopes (nonamers) sharing a minimum of 7/9 amino acids identity with the human peptides, without gaps or mismatches, were

eliminated. The final conserved epitopes showing an overlap were merged into a single peptide fragment containing multiple epitopes.

4. OBJECTIVES

The main objective of the present study carried out was to predict an immunogenic peptide of hemagglutinin of H5N1 Influenza virus which can be used as a vaccine against Influenza virus.

Work plan of current study is as follows:

1. Finding out conserved regions of Hemagglutinin, from all available hemagglutinin sequences of H5N1 serotype of Influenza virus.
2. Prediction of peptide containing overlapping T cell epitope in the conserved region of hemaggluitinin using different immunoinformatics tools which can act as a vaccine target.

5. RESULTS

1) Conserved Regions of Hemagglutinin H5N1(HA) virus

One hundred eighty sequences of HA protein (H5N1) were taken from NCBI influenza virus resource Database. These sequences were used to find out the conserved regions by using two tools: AVANA and Clustal Omega. Seventeen conserved sequences were obtained as given in table1. Sequence length of conserved peptides varies from 9 to 47.

Table1: Conserved sequences of hemagglutinin protein in H5N1 virus

Conserved sequence	Sequences	Length
CS1	SLVKSDQICIGYHANNSTEQVDTITHNGKLC	31
CS2	LDGVKPLILRDCSVAGWLLGNPMCDEF	27
CS3	NVPEWSYIVEK	11
CS4	FNDYEELKHLLSRINHFEKIQIIPK	27
CS5	SFFRNVVWL	9
CS6	SYNNTNQEDLL	11
CS7	LYQNPTTYISVGTSTLNQRLVP	22
CS8	IATRSKVNGQSGRM	14
CS9	NDAINFESNGNFIAPE	16
CS10	AYKIVKKGDS	10
CS11	EYGNCNTKCQTP	12
CS12	GAINSSMPFHNIHPLTIGECPKYVKSN	29
CS13	LVLATGLRNSP	11
CS14	KRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAA DKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRI ENLNKKMEDGFLDVWTYNAELLVLMENERTLDFHDSNVK NLYDKVRLQLRDNAKELGNGCFEFYH	47
CS15	YPQYSEEARLKREEISGVKLESIG	24
CS16	YQILSIYSTVASSLALAIMVAGL	23
CS17	LWMCSNGSLQCRICI	15

2)Identification of $CD8^+$ T cell epitopes of hemagglutinatin protein in H5N1 virus

hemagglutinin (HA) is a glycoprotein found on the surface of the influenza viruses. BIMAS, NetCTL and Syfphythi tools were used to identify $CD8^+$ T cell epitopes from the conserved peptides (Table 1) of hemagglutinin protein . All the predicted epitopes were compared, and then a set of epitopes was obtained that was predicted by all the tools. Epitopes which are predicted in common by all the three tools were selected. Initially, 27 epitopes were identified (Table 2). These epitopes were analyzed for their homology with human proteome using BLASTp tool. Fortunately, none of the predicted epitope was found to be homologous to human self-proteins. Number of MHC class I alleles predicted to bind the epitopes and score range are given in table 3. Overlapping epitopes were merged to generate peptides having single or multiple epitopes (Table 3).

Table 2 : . CD8+T cell epitopes of hemagglutinin in H5N1 which are commonly predicted by different immunoinformatics tools

Sequence	Epitope	BIMAS(cutt off 5)		NetCTL(cutt off 0.75)		SYFPEITHI(cutt off 13)	
		No of allele	Score range	Supertype	Score range	No of allele	Score range
S1	KSDQICIGY	4	29.7- 187.5	1	2.76	1	30
S2	RDCSVAGWL	2	7 to 22	1	0.75	1	16
S5	DYEELKHLL	4	5-43.2	1	0.98	1	24
	SRINHFEKI	5	5.80- 600	1	0.87	2	14-17
	NHFEKIQII	4	5.32- 180	1	0.84	1	16
S7	LYQNPTTYI	3	5.32- 75	1	1.38	1	13
	MPFHNIHPL	8	8.7- 665.5	2	0.81- 1.61	1	16
S14	KMNTQFEAV	3	6-16.681	1	1.07	1	21
	VLMENERTL	5	9-110.18	1	0.89	1	26
	KMEDGFLDV	4	6- 87.65	1	1.02	1	14
	TYNAELLVL	1	300	1	1.19	1	16
	LYDKVRLQL	2	7.8- 280	1	1.16	1	15
	GLFGAIAGF	2	15-67.5	1	0.87	1	21
	GMVDGWYGY	3	5 to 54	1	1.32	1	16
	ELGNGCFEF	1	160	1	0.81	1	17
	WQGMVDGWY	3	20-160	2	0.86 - 1.09	1	15
	RRIENLNKK	2	12 to 6000	1	1.71	1	13
	RLQLRDNAK	2	20-90	1	0.86	1	28
	HDSNVKNLY	2	22.5- 60	1	0.88	1	16
	FHDSNVKNL	2	90-405	1	2.19	1	14
	MENERTLDF	2	15-60	1	1.10	1	14
	IAGFIEGGW	1	90	1	1.39	1	16
S15	REEISGVKL	5	9-352	2	0.88- 1.89	1	13
	EARLKREEI	4	18 – 160	1	1.45	1	30
	KREEISGVK	2	12 to 6000	1	1.47	1	30
	QILSIYSTV	3	6.6-53.07	1	1.09	1	22

	IYSTVASSL	1	280	1	1.75	1	14

Table3 : Peptide containing overlapping CD8+T cell epitopes of hemagglutinin in H5N1

conserved sequences	**Epitopes**	**overlapping sequence**
CS1	KSDQICIGY	KSDQICIGY
CS2	RDCSVAGWL	RDCSVAGWL
CS4	DYEELKHLL	DYEELKHLLSRINHFEKIQII
	SRINHFEKI	
	NHFEKIQII	
CS7	LYQNPTTYI	LYQNPTTYI
CS12	MPFHNIHPL	MPFHNIHPL
CS14	KMNTQFEAV	KMNTQFEAV
	WQGMVDGWY	GLFGAIAGFIEGGWQGMVDGWYGY
	GMVDGWYGY	
	GLFGAIAGF	
	IAGFIEGGW	
	KMEDGFLDV	RRIENLNKKMEDGFLDV
	RRIENLNKK	
	VLMENERTL	TYNAELLVLMENERTLDFHDSNVKNLYDKVRL QLRDNAKELGNGCFEF
	TYNAELLVL	
	LYDKVRLQL	
	ELGNGCFEF	
	RLQLRDNAK	
	HDSNVKNLY	
	FHDSNVKNL	
	MENERTLDF	
CS15	REEISGVKL	EARLKREEISGVKL
	EARLKREEI	
	KREEISGVK	
	QILSIYSTV	QILSIYSTVASSL
	IYSTVASSL	

3) Identification of $CD4^+$ T cell epitopes of hemagglutinatin protein in H5N1 virus

NetMHCII 2.2, MHCPred, and IEDB SMM align tools were used to identify $CD4^+$ T cell epitopes from the conserved peptides (Table 1) of H5 hemagglutinin protein. Since all the tools predicted the binding of epitopes based on IC50 value, the cutoff selected for prediction was IC50≤ 500. Epitope having IC ≤ 50 are considered to be stronger binders whereas, epitopes with IC ≤ 500 are categorized medium weak binders. All the predicted epitopes were compared, and then a set of epitopes was obtained that was predicted by all the tools. Epitopes which are predicted in common by all the three tools were selected. Initially, 27 epitopes were identified (Table 4). These epitopes were analyzed for their homology with human proteome using BLASTp tool. Fortunately, none of the predicted epitope was found to be homologous to human self-proteins. Number of MHC class I alleles predicted to bind the epitopes and score range are given in table 4. Overlapping epitopes were merged to generate peptides having single or multiple epitopes (Table 5).

Table 4. CD4+T cell epitopes of hemagglutinin in H5N1 which are commonly predicted by different immunoinformatics tools

Conserved sequences	Epitopes	IEDB-SMM Align (cut off		NETmhc2 (cut off ≤500)		MHCpred (cut off ≤500)	
		No. of allele	IC50 value	No. of allele	IC50 value	No. of allele	IC50 value
CS2	LDGVKPLIL	1	81	2	256-400	2	256-400
	PLILRDCSV	1	437	4	345-347	2	345-347
	ILRDCSVAG	4	221-448	8	265-485	4	265-485
	DCSVAGWLL	2	229-356	9	120-350	3	120-350
	AGWLLGNPM	1	314-355	6	200-256	5	203-467
	WLLGNPMCD	1	348	12	180-395	6	180-395
	VKPLILRDC	46	504-2413	5	155-400	2	155-400
CS4	LKHLLSRIN	11	47 -311	7	155-467	5	155-467
	LLSRINHFE	1	319	6	256-358	6	84-203
	HLLSRINHF	2	397 -460	2	155-585	2	155-585
	INHFEKIQI	6	184-422	1	180-562	5	180-562
	LSRINHFEK	1	490	2	130-280	4	130-280
CS7	LYQNPTTYI	2	320 - 495	2	130-356	2	130-356
	YISVGTSTL	6	15-347	2	150-349	3	150-349
	ISVGTSTLN	1	270-333	1	150-245	7	150-245
CS9	NFESNGNFI	2	142-447	9	250-459	4	250-459
CS12	FHNIHPLTI	10	37-475	4	155-426	8	155-426
	SMPFHNIHP	1	62	2	250-346	6	250-346
	INSSMPFHN	5	38 - 482	8	120-485	9	120-485
CS14	LFGAIAGFI	4	22-343	7	110-356	8	110-356
	FLDVWTYNA	3	139 - 487	24	220-456	12	220-456
	YNAELLVLM	7	28- 448	34	210-356	15	210-356
	WTYNAELLV	6	123 - 440	16	258-354	4	258-354
CS16	IYSTVASSL	6	14-352	19	50-290	3	50-290
	YSTVASSLA	9	5 -485	15	100-456	12	100-456
	VASSLALAI	9	30 - 488	21	250-365	16	36-226
	ILSIYSTVA	6	7 - 258	24	155-485	15	45-229

Table 5.

Peptide containing overlapping CD4+T cell epitopes of hemagglutinin in H5N1

conserved sequences	Epitopes	overlapping sequences
CS2	LDGVKPLIL	LDGVKPLILRDCSVAGWLLGNPMCD
	PLILRDCSV	
	ILRDCSVAG	
	DCSVAGWLL	
	AGWLLGNPM	
	WLLGNPMCD	
	VKPLILRDC	
CS4	LKHLLSRIN	LKHLLSRINHFEKIQI
	LLSRINHFE	
	HLLSRINHF	
	INHFEKIQI	
	LSRINHFEK	
CS7	LYQNPTTYI	LYQNPTTYISVGTSTLN
	YISVGTSTL	
	ISVGTSTLN	
CS9	NFESNGNFI	NFESNGNFI
CS12	FHNIHPLTI	INSSMPFHNIHPLTI
	SMPFHNIHP	
	INSSMPFHN	
CS14	LFGAIAGFI	LFGAIAGFI
	FLDVWTYNA	FLDVWTYNAELLVLM
	YNAELLVLM	
	WTYNAELLV	
CS16	IYSTVASSL	ILSIYSTVASSLALAI
	YSTVASSLA	
	VASSLALAI	
	ILSIYSTVA	

4) Prediction of common immunogenic epitopes/ peptides of HA (H5N1) :

Putative immunogenic peptides as given in table 3 and 5 were merged to generate peptide fragment containing both, MHC class I and MHC class II epitopes. Eight immunogenic peptides were identified containing both CD8+ and CD4+ epitopes which is shown in Table 6. These peptides may be considered for target of vaccine design for H5N1 virus.

Table 6: Prediction of common immunogenic epitopes/ peptides of HA (H5N1) :

Sequence	Overlapping sequence of MHC class I epitopes	No. of epitopes	Overlapping sequence of MHC class epitopes II	No. of epitopes	Overlapping peptiesd
CS1	KSDQICIGY	1			
CS2	RDCSVAGWL	1	LDGVKPLILRDCSVAG WLLGNPMCD	7	LDGVKPLILRDCSVAGW LLGNPMCD
CS4	DYEELKHLLSRINHFEKIQII	3	LKHLLSRINHFEKIQI	5	DYEELKHLLSRINHFEKI QII
CS7	LYQNPTTYI	1	LYQNPTTYISVGTSTLN	3	LYQNPTTYISVGTSTLN
CS9			NFESNGNFI	1	
CS12	MPFHNIHPL	1	INSSMPFHNIHPLTI	3	INSSMPFHNIHPLTI
CS14	KMNTQFEAV	1			
	GLFGAIAGFIEGGWQGMVDG WYGY	4	LFGAIAGFI	1	GLFGAIAGFIEGGWQGM VDGWYGY
	RRIENLNKKMEDGFLDV	2	FLDVWTYNAELLVLM	3	RRIENLNKKMEDGFLDV WTYNAELLVLM
	TYNAELLVLMENERTLDFHD SNVKNLYDKVRLQLRDNAKE	8			
CS15	EARLKREEISGVKL	3			
CS16	QILSIYSTVASSL	2	ILSIYSTVASSLALAI	4	ILSIYSTVASSL

Prediction of conservancy from common epitopes:

Final set of peptides of H5N1 Hemagglutinin containing both MHC class I and Class II epitopes of H5 were analysed for their conservancy in original set of sequences taken initially. It was observed that the peptide derived from CS14 ie. RRIENLNKKMEDGFLDVWTYNAELLVLM was 100% conserved followed by NSSMPFHNIHPLTI (CS12) which was 98.92% conserved. Conservation of these peptide ranged from 81-100% (Table 7).

Table 7: Prediction of conservancy from common epitopes:

Conserved sequence	Common peptide	Conservancy %
CS2	LDGVKPLILRDCSVAGWLLGNPMCD	81.66
CS4	DYEELKHLLSRINHFEKIQII	96.23
CS7	LYQNPTTYISVGTSTLN	82.77
CS12	INSSMPFHNIHPLTI	98.92
CS14	GLFGAIAGFIEGGWQGMVDGWYGY	95.16
	RRIENLNKKMEDGFLDVWTYNAELLVLM	100
CS16	ILSIYSTVASSL	97.3

Conclusion

Eight immunogenic peptides of hemagglutinin (HA) containing both CD4+ and CD8+ T cell epioptes were found ranging in length from 9 to 16 amino acid residues. These peptides also bind large number of Major Histocompatibility complex alleles which shows they have potential to cover large population considering the fact that MHC is higly polymorphic. Hence it is inferred that these immunogenic peptides can be used as potential targets for epitope based vaccines against H5N1 virus . These peptides can be used for further study to assess the affinity of immunogenic peptide to MHC molecule by structural analysis and molecular modeling. Further study can be carried out to assess the potential of these peptides for immunogenic response in the PBMC by T-cell proliferation assay (MTT assay) and cytokine production assay.

6. Summary

Influenza is one of the most infectious diseases confronting the world today; however no effective prevention against influenza has been developed because of the antigenic variation in influenza virus. Error prone RNA-dependent RNA polymerases and segmented genome of influenza viruses allows virus to undergo antigenic drift as well as antigenic shift which are the major reason for antigenic variation in influenza virus.

A highly pathogenic H5N1 avian influenza virus was first isolated in 1996, from a goose in China and in 1997 it caused death in poultry in Asia. In general, avian influenza A viruses do not cause disease in humans and prior to 1997 only two cases of natural human infections by animal influenza virus had been documented(Gamblin S.J et.al., 2004). Over the past 150 years at least four pandemics of influenza occurred at irregular intervals, including three in the 20th century . These have caused high attack rates in all susceptible age groups, with high morbidity and mortality. The most lethal influenza pandemic in modern history was the H1N1 Spanish flu, which killed approximately 100 million people around the world between 1918 and 1919. The origin of the 1918 pandemic remains an enigma, but it is now clear that the virus had features of an avian virus , and it appears that an intermediate host, such as swine was involved . Swine are known to be susceptible to both avian and human viruses, and could have served as hosts for additional drift resulting in the accumulation of changes from the original avian strain . The pandemics of "Asian flu" (H2N2) in 1957 and the "Hong Kong flu" (H3N2) in 1968 caused an estimated 1 to 3 million deaths (Lamb R.A et.al.,2001).

Current treatment for influenza involves antiviral drug therapy and vaccination. Antiviral drug therapy involves two antiviral drugs, ion channel blockers and neuraminidase inhibitor. These drugs are effective in early stages of infection and viral strains have emerged that show drug resistance to both classes.

Conventional influenza virus vaccines protect against one particular strain of influenza and thus it is not effective against novel Influenza viruses which emerged as a result of antigenic variation. So there is a requirement for a broad range vaccine. Epitope based vaccines imparts an efficient strategy for protection against antigenic variations of influenza viruses. Immune response in epitope based vaccine is directed against a specific stretch of protein sequence called

epitopes. So if we find out the epitope conserved in different strains of influenza virus then it can be used as a vaccine target which will be effective against a broad range of influenza strains. Consequently, it will also protect against future (novel) strains as well. Nucleoprotein is considered to be a good candidate as it is a internal protein and it shows high level of conservancy.

Our approach was to find a stretch of immunogenic peptide from conserved peptide sequences of (HA) Hemagllutinin protein of H5N1 using immunoinformatics tools. Seventeen conserved regions were found in HA of H5N1. In these conserved regions of H1N1 twenty seven(27) MHC class I specific and fifteen (15) MHC class II specific immunogenic peptides were found. In these immunogenic peptides eight(8) immunogenic peptides were common for both classes of MHC.

Further study can be carried out for these immunogenic peptides to assess their affinity to MHC molecule by structural analysis and molecular modeling and immunogenic response of these peptides can be checked by T-cell proliferation assay and cytokine production assay.

REFERENCES

1. Varghese JN, Laver WG and Colman PM (1983); Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature* 303, 35 -40,
2. Kielian M and Jungerwirth S (1990).; Mechanisms of enveloped virus entry into cells. *Mol Biol Med.*7 (1):17-31,
3. LaVigne, Patrice (2013).; "FDA Approves Experimental H5N1 Bird Flu Vaccine with Reactive AS03 Adjuvant for U.S. Stockpile". National Vaccine Information Center 2013. Retrieved 23 April 2014.
4. Bhasin M, Raghava GPS(2013).; Prediction of promiscuous and high-affinity mutated MHC binders. *Hybrid Hybridomics*. 2003;22:229–34.
5. Bian, Hammer H (2004).; Discovery of promiscuous HLA restricted T cell epitope with TEPITOPE. Methods. 2004;34:468–75.
6. Cheung Y.K., Cheng S.C., Ke Y. and Xie Y. (2010);. Two novel HLA-A*0201 T-cell epitopes in avian H5N1 viral nucleoprotein induced specific immune responses in HHD mice. *Vet Res.* ;41(2):24.
7. Claas EC[1], Osterhaus AD, van Beek R, De Jong JC, Rimmelzwaan GF, Senne DA, Krauss S, Shortridge KF, and Webster RG (1998). Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet.*;351(9101):472-7.
8. Guan P, Doytchinova IA, Zygouri C, Flower DR (2003). MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. Appl Bioinformatics. 2:63-66.
9. Hoelscher, M.A., Garg, S., Bangari, D.S., Belser, J.A., Lu, X., Stephenson, I., Bright, R.A.,Katz, J.M., Mittal, S.K., Sambhara, S., (2006). Development of adenoviral-vector-based pandemic influenza vaccine against antigenically distinct human H5N1strains in mice. *Lancet* 367 (9509), 475–481.
10. http://www.flu.gov/planning-preparedness/index.html

11. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. (2007): Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. Oct 31;8:424.

12. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M (2008). NetMHC-3.0: accurate web accessible predictions of human mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res*. 2008; 36:509–12

13. Mänz B[1], Schwemmle M, and Brunotte L. (2013) Adaptation of avian influenza A virus polymerase in mammals to overcome the host species barrier. *J Virol.*;87(13):7200-9

14. Miotto O, Heiny A, Tan TW, August JT, and Brusic V (2008). Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC Bioinformatics*; 9 Suppl 1:S18.

15. Nichol, K.L., Treanor, J.J., 2006. Vaccines for seasonal and pandemic influenza. Journalof Infectious Diseases 194 (Suppl. 2), S111–S118.

16. Nicholson, K.G., Colegate, A.E., Podda, A., Stephenson, I., Wood, J., Ypma, E., Zambon,M.C., 2001. Safety and antigenicity of non-adjuvanted and MF59-adjuvantedinfluenza A/Duck/Singapore/97 (H5N3) vaccine: a randomised trial of twopotential vaccines against H5N1 influenza. *Lancet* 357 (9272), 1937–1943.

17. Nielsen M., Lundegaard C. and Lund O. (2007) ―Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method *BMC Bioinformatics*, 8:238

18. Pallavi S, Vijai S, M Arshad (2008). Modeling of RNA Secondary Structure of Non Structural Gene and Evolutionary Stability of the Influenza Virus Through *In Silico* Methods. *J Proteomics Bioinform;* 1: 219-226.

19. Parker K C, Bednarek M A, and Coligan J E. (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol. 152:163.

20. Peiris JS1, de Jong MD, Guan Y. (2007) Avian influenza virus (H5N1): a threat to human health. Clin Microbiol Rev. 2007 Apr;20(2):243-67.

21. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design 2007. *Nature Reviews Drug Discovery* 6, 404-414.

22. Puthavathana, P., Auewarakul, P., Charoenying, P.C., Sangsiriwut, K., Pooruk, P., Boonnak, K., Khanyok, R., Thawachsupa, P., Kijphati, R., Sawanpanyalert, P., (2005). Molecular characterization of the complete genome of human influenza H5N1virus isolates from Thailand. *Journal of General Virology* 86 (Pt 2), 423–433.
23. Rammensee H G, Bachmann J, Emmerich N N, Bachor O A, Stevanovic S (1999): SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*: 50: 213-219
24. Saha S, Raghava GPS(2006).; Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. 2006;65:40–8
25. Seto JT and Rott R (1966); Functional significance of sialidase during influenza virus multiplication. *Virology* 30, 731–737, (1966).
26. Shannon CE. (1948). "A Mathematical Theory of Communication". *Bell System Technical Journal* **27** (3): 379–423.
27. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, and Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*.:11;7:539.
28. Singh H , and Raghava GP (2001). ProPred: prediction of HLA-DR binding sites. *Bioinformatics*.;17(12):1236-7.
29. Sweredoski MJ, Baldi P.(2009).; COBEpro: a novel system for predicting continuous B-cell epitopes. *Protein Eng Des Sel*. 2009;22:113–20.
30. Tong S1, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J, Chen LM, Johnson A, Tao Y, Dreyfus C, Yu W, McBride R, Carney PJ, Gilbert AT, Chang J, Guo Z, Davis CT, Paulson JC, Stevens J, Rupprecht CE, Holmes EC, Wilson IA, Donis RO. (2013). New world bats harbor diverse influenza A viruses. PLoS Pathog. 2013;9(10):e1003657. doi: 10.1371/journal.ppat.1003657. Epub 2013 Oct 10.
31. Webby, R.J., Perez, D.R., Coleman, J.S., Guan, Y., Knight, J.H., Govorkova, E.A.,McClain-Moss, L.R., Peiris, J.S., Rehg, J.E., Tuomanen, E.I., Webster, R.G., (2004) .Responsiveness to a pandemic alert: use of reverse genetics for rapid develop-ment of influenza vaccines. *Lancet 363* (9415), 1099–1103.

32. Wharton SA, Weis W, Skehel, JJ and Wiley DC (1989); Structure Function, and Antigenicity of the Hemagglutinin in The Influenza Viruses New York, *Plenum Press* 153-174, (1989).
33. WHO/2014 http://www.who.int/influenza/human_animal_interface/Influenza_Summary_IRA_HA_interface_24January14.pdf
34. **WHO/H5N1cases/2014** http://www.who.int/influenza/human_animal_interface/EN_GIP_20140124CumulativeNumberH5N1cases.pdf?ua=1)
35. Zambon M.C. (1999); Epidemiology and pathogenesis of influenza. J Antimicrob Chemother; 44 Suppl B: 3-9.