# PEPTIDES CONTAINING MULTIPLE $CD4^{+}$ AND $CD8^{+}$ T CELL EPITOPES OF H5N1 AND H3N2 INFLUENZA VIRUS NUCLEOPROTEIN

**A thesis submitted in partial fulfillment of the requirements for the Award of degree of**

**MASTER OF SCIENCE IN BIOTECHNOLOGY**


**Submitted By:**

**SUKHJEET KAUR**

**REGISTRATION NO: 301301017**

**Under the Guidance of**

**Dr. MANOJ BARANWAL**

**(Assistant Professor)**

**DEPARTMENT OF BIOTECHNOLOGY**

**THAPAR UNIVERSITY, PATIALA**


**JULY, 2015**

# Candidate's Declaration

I, hereby declare that the work presented in the dissertation entitled **"Peptides containing multiple $CD4^+$ and $CD8^+$ T cell epitopes of H5N1 and H3N2 influenza virus nucleoprotein"** in the partial fulfillment of the requirement for the award of the degree of Master of science in Biotechnology, Department of Biotechnology, Thapar University, Patiala, is an authentic record of my work during the period of six months from January 2015 to June 2015, under the guidance of Dr. Manoj Baranwal, Assistant Professor, Thapar University, Patiala. I have not submitted the matter embodied in this thesis for the award of any other degree or diploma.

Sukhjeet Kaur.

Place: Patiala
Date: 15/7/15

Sukhjeet Kaur
(301301017)

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

Dr. Manoj Baranwal
Supervisor
DBT
Thapar University, Patiala

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

# CERTIFICATE

This is to certify that the thesis entitled **"Peptides containing multiple $CD4^+$ and $CD8^+$ T cell epitopes of H5N1 and H3N2 influenza virus nucleoprotein"** submitted by Sukhjeet Kaur, Registration no. 301301017 in the partial fulfillment of the requirement for the award of Degree of Master in science in Biotechnology to Thapar University, Patiala, is a record of student's own work carried out by her. The report has not been submitted for the award of any other degree or certificate in this or any other University or Institute.

Dr. Manoj Baranwal
Assistant Professor, Supervisor
Department of Biotechnology, TU
Patiala

Dr. Dinesh Goyal
Professor & Head
Department of Biotechnology, TU
Patiala

Dr. S.S. Bhatia
Dean, Academic Affairs
Thapar University, Patiala

# ACKNOWLEDGEMENT

I bow before the almighty for blessing me with his guidance. Thanks to my parents and family for giving me freedom and opportunity to pursue my own interest, even when they appeared incomprehensible. In the completion of the present work, there are feelings of achievement and satisfaction. In moment of happiness I take opportunity to record my sincere gratitude to all those who mattered.

First and foremost, I would like to express my profound gratitude and great indebtedness to my supervisor, Assistant Professor **Dr. Manoj Baranwal**, for his benevolent supervision of my M.Sc. Thesis. I am very grateful to him for giving me the opportunity to conduct this research project within his fascinating research group, for his continuous support and constructive criticism, and for encouraging and challenging me throughout the progress of the study; never accepting less than my best efforts.

I express my sincere thanks to **Dr. Dinesh Goyal**, Professor and Head, Department of Biotechnology (DBT), Thapar University, Patiala, for providing the best laboratory facilities. Special thanks to **Ms. Neha Lohia** research scholar in DBT, for her immense support, concern and valuable suggestions 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 for their love, encouragement and heart full blessings that continuo to enlighten my life.

Last but not the least I am thankful to my wonderful friends for their encouragement. I thank my Friends **Simerdeep Kaur, Simranjeet kaur, Randeep kaur, Ramandeep kaur and Amrit Kaur Virk** and all my classmates for help whenever required. I am forever indebted to my **parents** for their love and support.

Place: Patiala
Date: 15/7/15

Sukhjeet Kaur
Sukhjeet Kaur
301301017

# CONTENTS

# ABSTRACT

Subtypes of influenza virus, H5N1 and H3N2 are the causative of 'avian flu' and 'swine flu' in humans respectively. No single vaccine has been discovered till now, which can protect the world against the ever mutating strain of these influenza A viruses. Considering this fact, conserved regions from nucleoprotein of both the subtypes of influenza virus A isolated from human hosts were identified (≥90% conservation) by computational tool. Epitopes binding to MHC class I and II were predicted from the conserved regions of nucleoprotein by using three different epitope prediction tools for each MHC class I and II. Predicted epitopes were screened for the presence of any similarity with human self proteins. For H5N1, 44 MHC class I and 41 MHC class II epitopes were identified but in case of H3N2, 31 MHC class I and 34 MHC class II epitopes were predicted. Overlapping epitopes were merged to generate putative immunogenic peptide containing both $CD8^+$ and $CD4^+$ T cell predicted epitopes. Population coverage analysis of these epitopes have indicated that they have the capacity to induce a potent immune response among the individuals belonging to different ethnicities around the world. Molecular docking study has shown that eleven nonamers (each for MHC class I and II) common among H5N1, H3N2 and H1N1 have comparative binding energy with naturally bound peptides. All the predicted immunogenic fragments can be considered as the potent candidates for vaccine design against different subtypes of influenza virus.

# LIST OF ABBREVIATIONS

1. RNA – Ribonucleic acid
2. WHO – World health organization
3. HLA – Human Leukocyte Antigen
4. MHC – Major Histocompatibility Complex
5. HA – Hemagglutinin
6. NA – Neuraminidase
7. NP – Nucleoprotein
8. M1 – Matrix Protein 1
9. M2 – Matrix Protein 2
10. vRNPs – Viral Ribonucleoproteins
11. NEP – Nuclear export Protein
12. NS1 – Non-Structural Protein
13. PA, PB – Polymerase Protein A, B
14. APC – Antigen Presenting Cells
15. MUSCLE – Multiple Sequence Comparison by Log Expectation
16. AVANA – Antigen Variability Analyzer
17. NCBI – National Council of Biological Information
18. IEDB – Immune Epitope Database
19. CTL – Cytotoxic T-Lymphocytes
20. BIMAS – Bioinformatics and Molecular Analysis Section
21. SMM – Stabilized Matrix Method
22. QM – Quantitative Matrices
23. LAIV – Live attenuated influenza vaccine
24. RdRp – RNA-dependent RNA polymerase
25. CRM1 – Chromosomal Maintenance 1

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1: Introduction

Influenza virus are the agents of influenza, a contagious, acute and febrile respiratory disease belongs to the *Orthomyxoviridae* family (Ortigoza *et al*., 2012). It infects diverse host species, like pigs, birds and humans. It commonly circulates in wild waterfowl, which is thought to be the natural reservoir of the virus. Infection in humans is initiated in the respiratory tract, where influenza virus first counters the mucus when enters into the host through the oral or nasal cavities. Then it attaches to and invade the respiratory epithelial cells from where, the influenza virus can spread to both non-immune and immune cells (such as macrophages and dendritic cells) in the respiratory tract (Iwasaki *et al*., 2014). Symptoms of influenza infection are high fever, runny or stuffy nose, weakness, chills, sore throat, headache, muscle pain and sometimes diarrhea (vomiting in children). Although more severe than the common cold, influenza is generally a self-limiting disease in healthy adults that lasts about a week, but the cough and lethargy may continue for some time (Barik, 2012).

Influenza virus causes significant morbidity and mortality worldwide due to occasional pandemics, frequent epidemics, and yearly seasonal outbreaks in humans (Yang *et al*; 2013). The pandemic of 1918/1919 called the Spanish flu,  most devastating pandemic in world history was caused by the H1N1 subtype of influenza virus, the genomic sequences of the virus were revealed an avian like H1N1 virus that contains human like signature amino acids in several proteins, resulted in 50 million deaths worldwide (Taubenberger and Morens, 2007). Other influenza pandemics of influenza in $20^{th}$ century were Asian flu (H2N2 virus) in 1957 and Hong Kong flu (H3N2) in 1968.   Spanish flu claimed 70,000 lives in in Southern China , Singapore, Hong Kong, Japan the United States and the United Kingdom  The infection by highly a pathogenic avian influenza virus H5N1 subtype was first reported in Hong Kong in 1997 which resulted in 6 fatalities (Neumann *et al*; 2009).

Antigenic variation is the evolutionary mechanism by which influenza A virus evade the host immune system. There are two types of antigenic variations: antigenic drift (accumulation of point mutations) and antigenic shift (genetic reassortment). While influenza virus are changing by antigenic drift all the time, antigenic shift happens only occasionally

A range of therapies and medications are available to treat influenza i.e. antiviral drugs. At present, two classes of antiviral drugs are approved for influenza therapy: neuraminidase inhibitors (oseltamivir and zanamivir) and M2 Proton Channel blockers (Amantadine and rimantadine). These antiviral drugs are only effective when administered within a certain time period after exposure. Development of resistance against both these classes of antiviral drugs, limits their effectiveness in treating influenza virus infection.

Vaccination is considered a safest and most effective intervention against influenza virus infections. Currently available vaccine induces antibodies against seasonal and closely related antigenic viral strains, but do not protect against antibody-escape variants of seasonal or novel influenza A virus. Also, traditional influenza vaccines cannot respond effectively to an unforeseen epidemic or pandemics caused by a virus due to antigenic variations. Currently vaccines used against influenza are inactivated influenza vaccine (IIV) and live attenuated influenza vaccine (LAIV), used in monovalent as well as trivalent forms. Presently available Influenza vaccine which needs to be updated annually, induces immunity against presently circulating strains. In spite of the availability of antiviral and inactivated trivalent vaccines, which are effective for most recipients, Influenza remains a serious respiratory disease.

The vaccine, activating both the humoral and cellular arms of the immune response, induces long- lasting protection against many strains of the influenza virus. Peptides coupled to various MHC molecules are capable of Inducing $CD8^+$ and $CD4^+$ T cell response and further $CD4^+$ T cell mediated B cell response. Peptide based vaccines appears to attracted considerable attention recently as a potential means of treating infectious diseases. Peptides containing epitopes derived from highly conserved regions of viral proteins of the influenza virus could provide protective immunity against drift variants of influenza A virus (Lee *et al*., 2014). Further, highly conserved epitope common among various subtypes of influenza A virus can be assumed to impart cross protective immunity among the individual.

Considering these facts, present study was undertaken to identify highly conserved epitope of one of the influenza A internal protein i.e. nucleoprotein (NP) of H5N1 and H3N2 subtypes can be further be used as a target for vaccine development using novel approach based on immunoinformatics. Epitopes prediction tools based on various parameters of epitope-HLA interaction were used to identify MHC class I and class II binding epitopes from highly conserved segments of NP. Further, predicted epitopes which were common among H5N1 and H3N2 were docked to one HLA class I and II molecules to calculate their binding affinity.

# Chapter 2: Review of Literature

Influenza, commonly known as 'flu' is a contagious viral infection caused by segmented, pleomorphic, eight single-stranded negative sense RNA virus that belong to the *orthomyxoviridae* family (Somvanshi *et al*., 2008). It is an acute respiratory disease characterized by sudden onset of high fever, cough, malaise, prostration, and inflammation of the upper respiratory tract. Weakness and fatigue may persist for weeks but acute manifestation and febrility often remains for 7 to 10 days. Influenza commonly infects people of all ages but its universality is in school-aged children, severity of disease is greater in the aged and infants (Taubenberger and Morens, 2008).

**2.1 Classification of Influenza virus**

Influenza infections caused by members of *orthomyxoviridae* family that are influenza virus A, influenza virus B, influenza virus C which are capable of infecting humans but, influenza A and B are the common circulating types (Soema *et al*., 2015). Influenza A virus cause epidemics and pandemics globally with severe consequences for humans and various animals like aquatic birds, poultry, horses and sea mammals. Influenza B and C both naturally infect humans but also infects seals and pigs respectively (Noda *et al*.; 2010). Other genera of *orthomyxoviridae*, the Thogotovirus, infectious salmon anaemia (Isavirus) and Quaranjavirus are responsible for distinct clinical or veterinary diseases. The influenza A virus can be subdivided into different serotypes based on the antibody response against the two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA) of influenza virus. Eighteen different types of HA and 11 types of NA forms number of combinations at the time of genetic reassortment giving rise to novel and thus highly virulent subtypes of influenza A virus (Wu *et al*., 2014) A and B type of influenza virus have similar structure containing eight discrete gene segments and covered by three glycoproteins: HA, NA and Matrix 2 (M2). While type C is more divergent consisting of seven gene segments and covered by only one surface glycoprotein (Taubenberger and Morens, 2008).

**2.2 Antigenic variations:**

Viral RNA genome mutated by two mechanisms Influenza virus evades previously acquired immunity by either antigenic drift, in which mutations in the surface glycoprotein limits or prevents antibody binding, or by antigenic shift, in which the virus acquires HA or NA of a new subtype by genetic reassortment between two influenza A viruses. Antigenic drift is the accumulation of the mutations during replication due to lack of proofreading activity, whereas antigenic shift is the replacement of the hemagglutinin and sometimes the neuraminidase coding RNA segments with novel subtypes in the host cell that have not been present in human virus for a long time (Treanor *et al*., 2004).

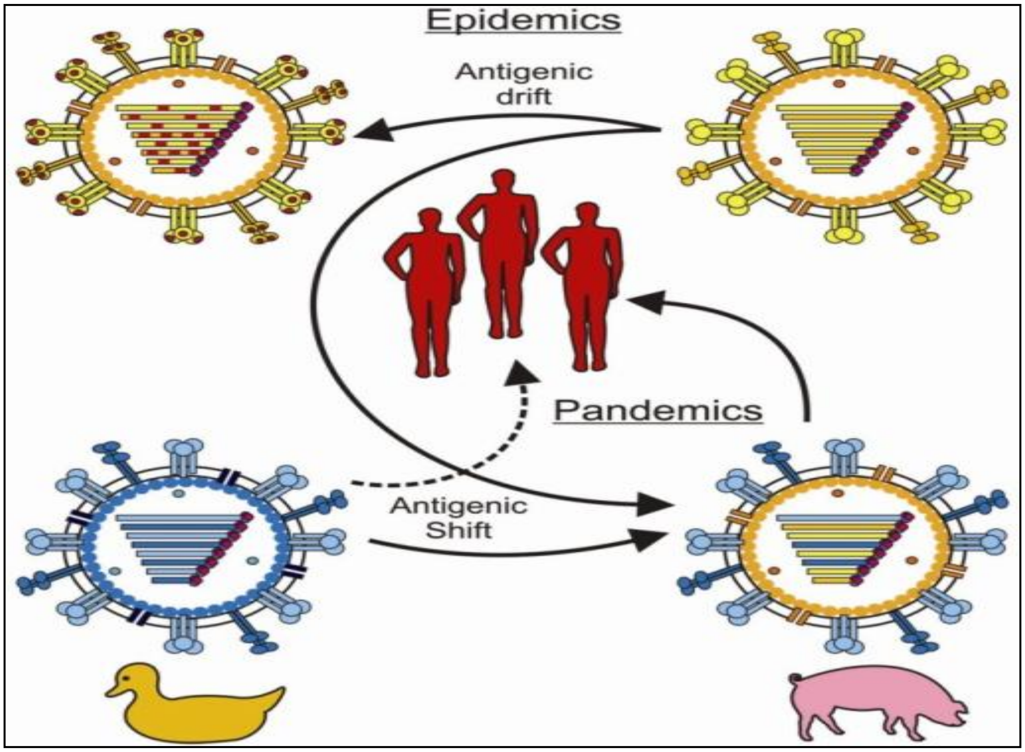



**Figure 1: Antigenic drift and Antigenic shift**

**(Van de Sandt *et al.*, 2012)**

Mutation rates are high in genes of influenza virus approximately $1 \times 10^{-3}$ to $8 \times 10^{-3}$ substitutions per site per year , these mutations in surface glycoproteins allow viral strains to evade pre-existing immunity (Taubenberger and Morens, 2008).

**2.3 Structure of influenza A virus**

Influenza A virus are most virulent among all the other subtypes of the influenza virus. Type A further classified based on the serological reactivity of surface glycoproteins HA and NA (Webster *et al*., 1992). These surface glycoproteins can make different combinations based on its different subtypes i.e. HA (1-18) and NA (1-11). Influenza is an enveloped virus having eight segments that encode for 11 different viral genes.

The influenza virion is roughly spherical covered with a lipid bilayer, containing three viral transmembrane proteins: HA, NA and M2. This lipid bilayer contains cholesterol enriched lipid rafts and non-lipid rafts, which is derived from plasma membrane of host.

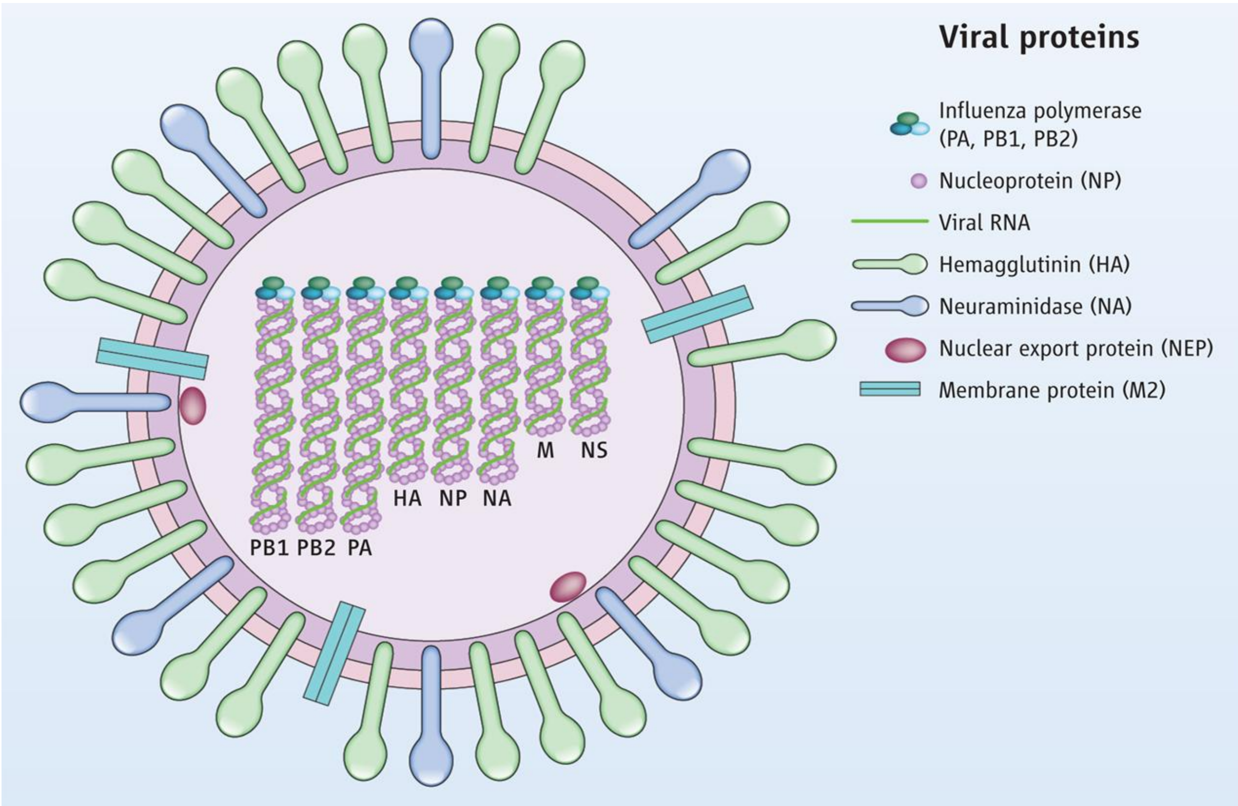



**Figure 2: Structure of influenza virus**

**(Tao and Zheng, 2012)**

Most abundant envelope protein is HA which approximately 80% followed by NA, that makes up 17% of the viral envelope whereas M2 is the minor component with only 16 to 20 molecules per virion. HA and NA are associated with lipid, M1 is present just underneath the lipid membrane having ability to hold the viral ribonucleoproteins (vRNPs). These vRNPs made up of viral negative stranded RNAs and are the core of the virus (Samji, 2009).

Eleven different proteins which encodes by the eight influenza segments:

i. Viral RNA polymerase basic protein 1 (PB1) plays role in RNA transcription and replication activities and PB1F2, which induces cell death whereas polymerase A (PA) in RNA replication. Polymerase basic protein 2 (PB2) is involved in synthesis of capped mRNA by endonuclease activity which cleaves host cell mRNA. Polymerases are coded by RNA segment 1(PB1), 2(PB2) and 3(PA) of influenza virus genome.

ii. Hemagglutinin (HA) surface glycoprotein, coded by RNA segment 4 of influenza virus genome. HA aids the entry of the viral genome into the infected cells by binding to sialic acid-containing receptors on the surface of the host cells, thus causing the fusion of the viral envelope with the host membrane.
iii. Nucleoprotein (NP) is core antigen that forms complex with viral RNA genomes and packages the RNA into a helical ribonucleoprotein core to take part in viral replication and transcription. It is coded by RNA segment 5 of influenza virus genome.
iv. Neuraminidase (NA) glycoprotein, coded by segment 6 of genome, catalyzes the cleavage of terminal sialic acid linked to an array of host membrane glycoproteins and glycolipids. The removal of sialic acid from the surface of the infected cell as well as new virions is considered essential for the release of viral progeny from the host cell
v. Matrix protein 1 (M1) is a bifunctional membrane which is important to mediates the coating of RNA-nucleoprotein during the viral assembly. Matrix protein 2 (M2) transmembrane protein forms ion channel pump to lower or maintain the pH of the endosomal compartment and transports the viral transmembrane protein to the cell surface at the last stage of infection. These 2 proteins are coded by RNA segment 7 of influenza virus genome.
vi. Non-structural protein 1 (NS1) and Nuclear export protein (NEP) are necessary for the nucleocytoplasmic export of viral ribonucleoprotein complex from the nucleus, where they are assembled. They are coded by segment 8 of influenza virus genome.

**2.4 Life cycle of influenza virus**

Life cycle of influenza virus has following stages (Samji, 2009):

I. Entry into the host cell.
II. Entry of vRNPs into the nucleus.
III. Transcription and replication of the viral genome.
IV. Export of vRNPs from the nucleus.
V. Assembly and budding at the host cell’s plasma membrane

Major viral proteins are participate in life cycle of influenza virus, infects host cells via homotrimeric HA of influenza virus which forms spikes bind to sialic acids present on host cell’s membrane and internalizing by endocytosis which is receptor mediated. Low pH of endosome stimulates the fusion of the viral and endosomal membranes, leading to release of vRNPs (viral ribonucleoproteins) into the cytoplasm by the detachment of the M1 coat protein (Samji, 2009). Transfer of RNPs to the nucleus facilitated by nuclear-localized signal found on nucleoprotein. Viral transcription and replication occur in nucleus, mainly dependent on its RNA-dependent RNA polymerase (RdRp) enzyme which comprises of PB1, PA and PB2 subunits (Chang *et al*., 2015). vRNPs serve as active template for the synthesis of viral mRNA as well as cRNAs (complementary RNAs). cDNAare the replication intermediates which direct the synthesis of nascent vRNA (Zheng and Tao, 2013). Replicated RNPs is transported to plasma membrane with the help of NEP and M1 protein from nucleus for assembly with other envelope proteins (HA, NA and M2). Neuraminidase preventing self-aggregation of the viral particles by releasing the newly assembled virions from cell surface after removing sialic acids from sialyl oligosaccharides (Zhang *et al*., 2015).

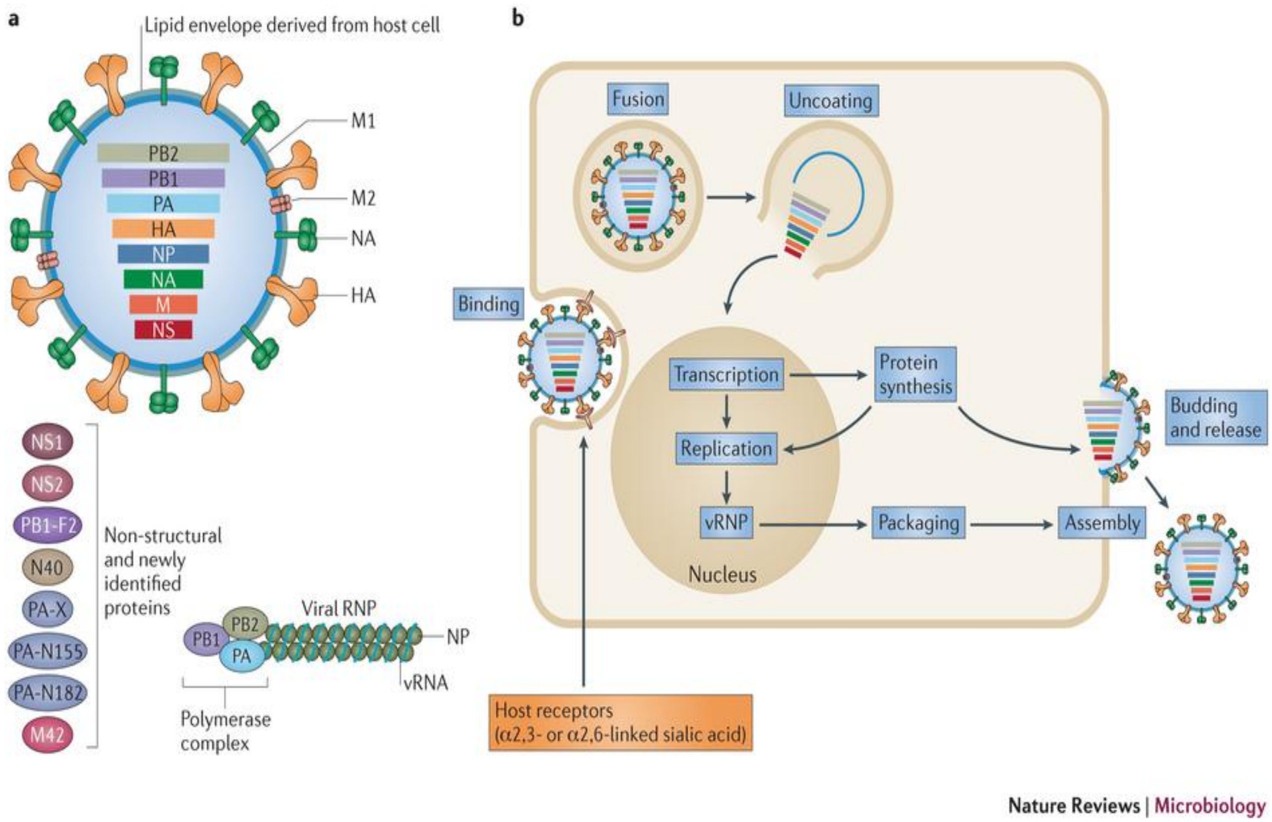



**Figure 3: Life cycle of influenza A virus**

**(Shi Y *et al*., 2014)**

## 2.5 Nucleoprotein: Structure and function

Influenza A virus RNA segment 5 encodes NP which is a 498 amino acids long peptide, rich in serine, arginine and glycine residues. Phylogenetic analysis of virus strains isolated from different hosts reveals that the NP gene is relatively well conserved, with a maximum amino acid difference of less than 11% (Shu et al., 1993). The primary function of NP is to encapsidates the genome of influenza virus thus protecting it from nucleases. Also, it interacts with PB1, PB2 and PA to form viral nucleoprotein complex which are responsible for viral transcription and replication (Ng *et al*., 2008). NP has also been shown to interact with at least four cellular polypeptide families: nuclear import receptors of the importin α class, filamentous (F) actin, the nuclear export receptor CRM1 and a DEAD-box helicase BAT1/UAP56. So it is not only a structural protein for RNA binding but also acts as a multifunctional key adaptor molecule for interactions between virus and host cell.

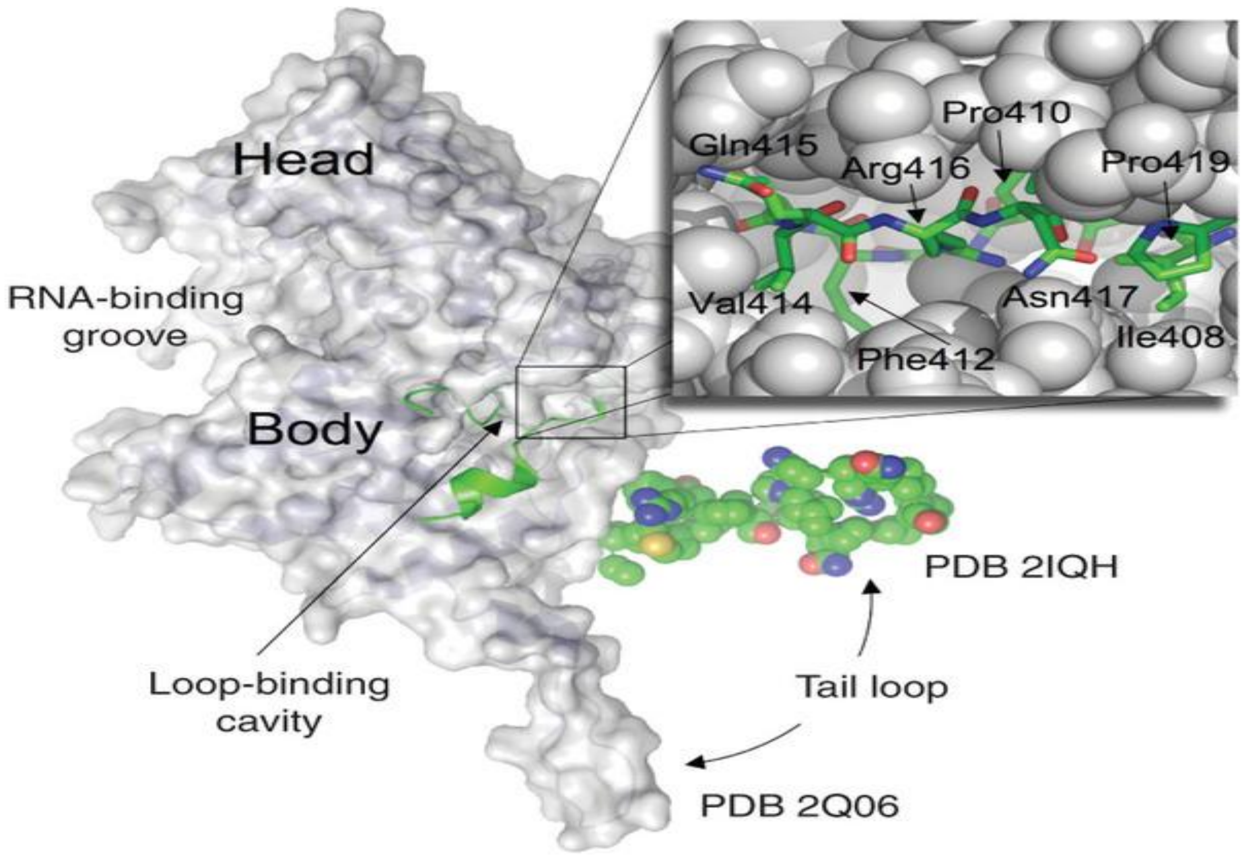



**Figure 4: Structure of nucleoprotein**

**(Das *et al*., 2010)**

Crystal structure of 3.3 Å of nucleoprotein, which is composed of a body domain, a head domain and a tail loop. The head domain is more conserved than the body domain and oligomerization of NP is mediated by the insertion of the non-polymorphic and structurally conserved tail loop of one NP molecule to the groove of another NP. Recombinant electron microscopic model of influenza virus RNP generated has indicated that the molecule has an elongated, curved, 'banana-shape', comprising two domains (Martin- Benito *et al*.; 2001).

**2.6 Influenza A subtypes H5N1 and H3N2**

H5N1 subtypes of influenza A (avian influenza virus) virus is known to serious outbreaks in domestic poultry (bird flu) Although H5N1 does not usually infect humans, 842 human cases of H5N1 have been reported including 447 deaths in 16 countries (Indonesia, Egypt and Vietnam being the most affected) since 2003 by breaking the species barrier to infect humans (WHO/2015). H5N1 has been reported in Europe, Southeast Asia and Africa associated with high mortality rate i.e. >60%. Human to human transmission of H5N1 has not been confirmed yet. Avian virus based on their pathogenicity categorized into 2 groups: highly pathogenic which results in high mortality rate (90-100%) within 48 hours and low pathogenic which is undetected and usually causes only mild symptoms. The influenza A virus H5N1 has reached endemic levels among poultry in several south Asian countries, and more than 100 reported human infections with high mortality (de Jong and Hien, 2006). However, the probability of a human influenza pandemic resulting from either mutations or avian origin virus reassortments with human strains caused great international concern and pandemic preparedness has been accelerated (Salomon & Webster, 2009).

Influenza H3N2 A viruses circulates in swine and occasionally infect humans, resulting in outbreaks of variant influenza H3N2. It has also not acquired the ability to transmit efficiently among humans. During the period 2011-14, 343 cases of H3N2 cases has been reported including 1 death in USA (CDC/2014). H3N2 were created by reassortments in the polymerase basic protein (PB1) gene segment of influenza A virus in 1968 (Ngai *et al*., 2013). Influenza H3N2 virus with pandemic capacity infects pigs due to potential interspecies transmission of new reassortant strains (Pedersen *et al*., 2014). Sequence analysis of the hypothetical precursor strain, which immediately preceded the pandemic H3N2 virus suggested that fewer than six amino acids in HA had changed during the avian-to-human transition (Bean *et al*., 1992).

### 2.7 Influenza vaccine: Status and Need

Vaccines are the principal defense against influenza, but because it takes time to produce an antigenically appropriate and immunogenic product and deliver it to entire populations, antiviral drugs are an important adjunct to vaccination for influenza prevention and control to reduce the impact of a new pandemic (Monto, 2006; Moscona, 2005). The current armamentarium of licensed anti-influenza medications consists of four drugs: two adamantanes, amantadine and rimantadine, and two Neuraminidase inhibitors, the oral drug oseltamivir (Tamiflu; Roche) and the inhaled medication zanamivir (Relenza; GlaxoSmithKline) (De Clercq, 2006). Adamantanes only active against influenza A but NA inhibitors are against both of A and B type of influenza (CDC, 2006). Many strains of influenza virus develop resistance to these antivirals so vaccines are considered as most effective measure for the control of influenza infection.

Vaccination, in the prevention and control of influenza pathogens which cause high morbidity and mortality, is an effective and cost benefit approach. Due to low vaccine efficacy influenza causes significant health issues because vaccines mainly based on hemagglutinin protein which is highly variable among different strains of influenza circulating in humans and animals. The licensed vaccines against influenza should be updated annually that induces specific immunity based on predicted strains will circulate in upcoming season (Lee *et al*., 2014). It is difficult to develop a vaccine against highly pathogenic avian influenza virus that provide long lasting immunity (Parida *et al*., 2007) because of three main reasons. The virus mutates relatively frequently due to the lack of proof reading activity in the RNA viral genome. Thus it is a selective advantage for virus to be mutated because it develops resistance to existing vaccines and antiviral drugs (Ison *et al*., 2003). Genetic reassortment leads to hybrid strains that are novel and cause infection easily (Webster *et al*., 1992). Also, specific population groups are more prone to influenza infection up to fatality level, such as the elderly and individuals with diabetes or immune deficiency (AIDS). These groups are in greatest need of intensive treatments but they have less tolerance to these aggressive treatments (Metersky *et al*., 2012). Due to continuous emergence of novel strains of influenza virus, vaccines based on predicted strains of the virus may not provide sufficient immunity against future pandemic of the influenza.

Several approaches have been used to overcome these obstacles including use of related pathogenic H5 virus with or without adjuvants, DNA vaccination and reverse genetic techniques

(Webby *et al*., 2004). Currently, vaccine use against influenza consists of formalin inactivated virus strains isolated from embryonated eggs (Katz *et al*., 1989). In November 2013, FDA approved an experimental H5N1 bird flu vaccine AS03 (containing squalene) adjuvant (Glaxosmithkline's Q-Panfor) the U.S. stockpile in 2013 (LaVigne; 2013). A novel vectored vaccine (MVP-NP$^+$M1) i.e. Modified Vaccinia Virus Ankara designed by expressing nucleoprotein and matrix protein that are conserved in all subtypes to boost cross-reactive T-cell responses provided protection against influenza (Berthoud *et al*., 2011).

### 2.8 Peptide based vaccine

Peptide based vaccine or T-cell epitope based vaccine provide minimal but essential information regarding T cell epitopes, for protective immunity, that are critical mediators of cellular immunity (De Groot *et al*., 2013). Peptides cannot serve as vaccine because these are weak immunogens and degrade fast within the body. So, in epitope based vaccine epitopes are being inserted into flagellin serves as carrier as well as adjuvant against influenza (Adar *et al*., 2009). Advantages of peptide based vaccine are; high safety, ability to focus on conserved epitopes to induce immune responses and also have potency due to the opportunity to be rationally engineered (Sette and Fikes, 2013).

Long list of peptide based vaccines are under development, including peptide vaccine against influenza, swine fever, anthrax, human immunodeficiency virus (HIV) and anti-cancer vaccines for pandemic cancer. These vaccines are advantageous as they lack deleterious sequences which may cause autoimmune diseases. It is easy to add organic groups which improve not only the stability but also increase immunogenicity, during the design of a peptide vaccine first consideration which need to be made is to identify the immune-dominant domains of epitopes having ability to induce protective immune response and main focus on the epitopes which induce cytotoxic T cell response in case of vaccines against intracellular pathogens such as virus. It may be necessary to identify highly conserved immunogenic epitopes that confer long-lasting protective immunity against many strains of virus with the activation of both cytotoxic and helper T cells (Yedidia and Arnon, 2007).

## 2.9 Immunoinformatics

Immunoinformatics has emerged as an important field includes the study and design of algorithms for mapping potential B and T cell epitopes. It acts as an intersection between computational approaches and experimental immunology, has the potential to accelerating immunology research which lessens the time and cost required for laboratory analysis of pathogen gene products. Immunoinformatics represents the use of computational methods and resources for the understanding, generation, processing and propagation of immunological information includes study of protein-protein interactions and networks.

Immunoinformatics is capable of identifying virulence genes and surface associated proteins and also in advancement in the genome and proteome sequencing. Immunomics offers new opportunities for future research that combines traditional immunology with mathematics, chemistry and genomics and proteomics for large scale analysis of immune system function. There are different B and T cell epitope prediction tools has been developed and also being with improvement in varied algorithms using by prediction tools (Tong and Ren, 2009; Tomar and De, 2010).

Earlier work done by some authors based on computational immunology by using prediction tools includes prediction of HLA-A24 binding epitopes derived from internal proteins of influenza A virus on H3N2 and H5N1 subtypes by using BIMAS, SYFPEITHI and NetCTL (Ichihashi *et al*., 2011). Another study identified potential binding peptides within protein on H5N1 subtype by using BIMAS (Sun *et al*., 2010). Further, using epitope prediction tools such as NetMHC, NetCTL and IEDB, Gustiananda in 2011 carried out immunoinformatics analysis of H5N1 proteome for designing an epitope derived vaccine (Gustiananda., 2011). The utility of coupling robust *in vitro* immunoassays (such as IFNγ ELISPOT and PBMC) with computational techniques to better understand human immunity, including cross-protective responses, and define $CD4^+$ T cell epitopes for potential vaccination efforts against future influenza viruses and other pathogens (Schanen *et al*., 2011). Validation of predicted $CD4^+$ T-cell epitopes for their ability to bind HLA and to stimulate interferon-γ production in peripheral blood mononuclear cells from a cohort of donors presenting with influenza-like illness during the 2009 pandemic and a separate cohort immunized with trivalent influenza vaccine in 2011 (Moise *et al*., 2013).

A Structural approach gives better understanding for prediction of the HLA-restricted T cell epitopes whose immunogenicity depends on the quality of T-cell receptor interaction with the HLA/peptide complexes. To generate highly conserved cell epitope candidates of current and past pandemic strains, combine affinity-based epitope prediction with molecular docking (Su et al., 2013). Molecular docking is used to predict the HA neutralising epitope of a hemagglutinin with an antibody's variable fragment (Fv) as a preliminary step (Mulyanto and Saleh, 2011). Surflex-Dock is a tool used to dock the octa-peptides and nona-peptides to their corresponding MHC class I molecule receptors. Peptides with a higher docking score and rational conformation were predicted as candidate T cell epitopes for each haplotype (Hou Y *et al*., 2012). Auto Dock Vina is also an open source program for molecular docking and it significantly improves the average accuracy of the binding mode predictions compared to other docking programs.

Computational Modeling became crucial implement to synthesize, assemble and combine various data types and theoretical structure to generate new knowledge and guide *in vivo* experimentation (Carbo *et al*., 2014).

# Chapter 3: Objectives

The main objective of the present study was to carry out the prediction of conserved immunogenic peptide of nucleoprotein of H5N1 and H3N2 influenza virus which can be used as a vaccine against influenza virus.

Work plan of current study is as follows:

1. Finding out the conserved regions of nucleoprotein (H5N1 and H3N2) from all the available sequences.
2. Prediction of peptide containing overlapping T cell epitope in conserved region of nucleoprotein by using different immunoinformatics tools.
3. Population coverage analysis for the capability of the epitopes to induce a potent immune response among individuals from different populations around the world.
4. Molecular docking of predicted peptides with the MHC class I and II structures and analyse their binding ability to the antigen binding pocket of MHC.

# Chapter 4: Methodology

## 4.1 CONSERVANCY ANALYSIS

**Sequence retrieval**

All available sequences of nucleoprotein of H5N1 and H3N2 were retrieved from influenza research database (http://www.fludb.org/) till January, 2015. Full length sequences were downloaded by selected human as host after removing the duplicate sequences using the option available in the database.

**Multiple sequence alignment**

Multiple sequence alignment was carried out using MUSCLE (Edgar, 2004). MUSCLE (Multiple Sequence Comparison by Log-Expectation) is claimed to achieve both better average accuracy and better speed than Clustal W or T-Coffee, depending on the chosen options. Multiple alignments of protein sequences are important in many applications, including phylogenetic tree estimation, secondary structure prediction and critical residue identification (http://www.ebi.ac.uk/Tools/webservices/services/msa/muscle_soap).

**Identification of conserved regions**

Conservancy analysis was carried out to find out the conserved region present in nucleoprotein of both the subtypes of influenza virus i.e. H5N1 and H3N2 by the use of AVANA tool (Antigen Variability Analyzer). The regions of nucleoprotein showing ≥ 90% conservancy were selected. 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. Alignment result obtained from MUSCLE (FASTA format) was used as input for AVANA software. Parameters were set to 90% conservancy and minimum length of 9 amino acids as a threshold value in AVANA. Conserved regions from alignments were merged together to obtain overlapping regions of conserved sequences.

## 4.2 T CELL EPITOPE PREDICTION

An epitope consists of a few amino acid residues that are sufficient to elicit an immune response, predicted as T cell epitope to be included in multiple-subunit vaccine (Desai *et al*., 2014). These epitopes bound to MHC molecules are is presented on surface of antigen presenting cell (APC) to initiate immune response, so binding is the crucial part of prediction methodologies. (Patronov *et al.*, 2013). Further, immunogenicity of an epitope depends on the stability of the MHC peptide complexes (Burg *et al.*, 1996). $CD8^+$ and $CD4^+$ T cell epitope which bind to MHC I and MHC II molecules respectively, were predicted using epitope prediction tools based on various parameters.

### 4.2.1 MHC Class I Epitope Prediction

MHC I molecules bind peptides and present them to $CD8^+$ T cells. Generally, these peptides are derive from endogenous intracellular proteins that are digested in the cytosol. The MHC class I molecules bound peptides are generally nine amino acids long 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.

#### 4.2.1.1 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 subtypes including A1, A2, A3, A24, A26, B7, B8, 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 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 nucleoprotein, H5N1 and also H3N2 identified by AVANA were taken FASTA format and prediction was carried out for all 12 subtypes. 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.

#### 4.2.1.2 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 nucleoprotein, H5N1 and H3N2 were taken as the input sequence and only those epitopes were taken showing score ≥ 20. Threshold for epitope identification was taken greater than or equal to 20.

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

BIMAS stands for Bioinformatics And Molecular Analysis Section (Parker *et al*., 1994). BIMAS locates and ranks 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. 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 nucleoprotein, both H5N1 and H3N2 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 $T_{(1/2)}$ (estimate of half time of dissociation of a molecule containing this subsequence). Different parameters were for epitope identification were predicted $T_{(1/2)} \geq 20$.

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.

**4.2.2 MHC Class II Epitope Prediction**

Class II MHC molecules bind peptides and present these peptides to $CD4^+$ 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.

**4.2.2.1 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 deducted from literature. The predicted binders can be visualized either as peaks in graphical interface or as colored residues in HTML interface (Singh *et al*., 2010). Conserved regions of nucleoprotein, H5N1 and also of H3N2 were taken as the input. Propred predicted MHC class II epitopes for 51 HLA-DR alleles. Epitopes were selected at a threshold of 3%.

**4.2.2.2 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 (Neilsen *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 the half of the maximum inhibitory concentration, is a measure of the effectiveness of a compound. The lower the IC50 if the peptide the higher will be peptide-MHC binding affinity.

Conserved regions of nucleoprotein, H5N1 and H3N2 were taken FASTA format and analysis was carried out for all HLA class II molecules. Threshold for epitope identification was taken as $IC50 \leq 500$.

**4.2.2.3 NetMHCII 2.2** (http://www.cbs.dtu.dk/services/NetMHCII/)

NetMHCII 2.2 server predicts binding of peptides to HLA-DR, HLA-DQ, HLA-DP alleles using artificial neuron networks. The prediction values are given in IC50 values, and percentage rank to a set of 1,000,000 random natural peptides. Strong and weak binding peptides are indicated in the output. NetMHCII (Neilsen *et al.*, 2009) corrects for biases due to replicate binding cores within the training data and is one of the few algorithms that factors in effects of the amino acids that flank the 9 amino acid core of the peptides. It also has added flexibility in that it allows evaluation of peptides of different lengths.

Conserved regions of nucleoprotein, H5N1 and H3N2 were taken FASTA format and analysis was carried out for all alleles of MHC II molecules. Outcome is generated in the form of IC50 value threshold value for epitope identification was IC50≤500. Common epitopes predicted by all the three tools were selected.

### 4.3 Blast screening

In order to avoid any similarity of the peptides with any human protein, blastp analysis (Altschul *et al*., 1990) was performed for epitopes predicted to bind MHC class I and II respectively. The peptides showing similarity in 7 out of 9 amino acids without gap or mismatch were eliminated (Tan *et al*., 2010) thus ruling out any possibility of autoimmune response against any human functional protein. The epitopes obtained after blastp analysis having overlapping sequences were merged together to generate single peptide fragment containing single or multiple epitope containing peptides

### 4.4 Population Coverage Analysis

MHC alleles are highly polymorphic and diversity of alleles were found in the global population. Different individuals of the worldwide population may respond in different way against particular antigen. The IEDB database enables predicting the possible world population capable of responding to particular immunogen. The database contains variety of MHC class I and II alleles frequently found in different world population of geographical areas. It compares the epitopes specific alleles with the alleles in database and find out percentage population capable of responding. In our study, IEDB population coverage analysis was carried out for 16 different geographical areas distributed globally (Bui *et al*., 2006).

### 4.5 Epitope Docking

#### 4.5.1 Peptides Structure Prediction

The structures of common predicted NP epitopes of H5N1, H3N2 (from the current studies) and H1N1 (from the previous work done in our lab) were generated using peptide structure prediction server PEP-FOLD (Thevenet *et al*., 2012). The model generated by PEPFOLD are assorted either by using the coarse grain energy of the PEPFOLD or by predicted Tm score. For the peptides of length 36 residues the coarse grain energy of the PEPFOLD is used to sort the predicted models whereas Tm score for the Peptides with the residues more than 36.

**4.5.2 Docking using Auto dock vina:**

The structure of peptide bound MHC class I (HLA-A2, 3MRK) and MHC class II (HLA-DR3, PDB id) were retrieved RCSB Protein Data Bank. In each case, the ligands and receptor of the PDB structure were separated using the discovery studio visualization tool 3.5. These naturally bound ligands (peptides) were docked to the respective MHC molecules (as test sets for comparison) using Auto dock Vina (Trott *et al*., 2009). Further, the predicted epitope were docked to these MHC molecules. Steps involved in docking are listed below:

**A. Ligand Separation**

1. Open the PDB structure of HLA molecule in discovery studio visualization tool 3.5
2. Go to the scripts option and select the water molecules
3. Go to edit and delete
4. Go to scripts and select protein chains
5. Go to edit and delete
6. Save as ligand.pdb file

**B. Receptor Separation**

1. Open the PDB structure of HLA molecule in discovery studio visualization tool 3.5
2. Go to the scripts option and select the water molecules
3. Go to edit and delete
4. Go to scripts and select ligands
5. Go to edit and delete
6. Save as the receptor.pdb file

**C. Receptor preparation for docking**

1. Open the PDB structure of receptor molecule in MGL tools
2. Go to the edit and add hydrogen atoms
3. Go to grid and choose and select the macromolecule i.e. receptor molecule
4. Go to grid box and select the grid where the ligand should be bound
5. Save the file as receptor.pdbqt

**D. Ligand preparation for docking**

1. Open the ligand.pdb file in the MGL tools
2. Go to ligand and choose the molecule
3. Select the ligand molecule
4. Go to output and save the molecule as ligand.pdbqt

**F. Preparation of configuration file**

To perform the docking in autodock vina a text file named as config was generated, containing the following commands. The config file was created for the each molecule separately depending upon their grid dimensions. Grid dimensions were defined using MGL tools.

Receptor = receptor.pdbqt
Ligand = ligand.pdbqt
Out = out.pdbqt
center_x = -15.833 (molecules specific obtained by grid box selection)
center_y = 11.601(molecules specific obtained by grid box selection)
center_z = 14.053(molecules specific obtained by grid box selection)
size_x = 40(molecules specific obtained by grid box selection)
size_y = 68(molecules specific obtained by grid box selection)
size_z = 34(molecules specific obtained by grid box selection)
Exhaustiveness = 16.

**G. Docking**

The separated naturally bound peptides of the MHC molecules were redocked with the MHC peptide free MHC using AutoDock Vina (Trott *et al*., 2009) tool. The binding energy of these naturally bound ligands of MHC molecules was determined. The epitopes were docked with the MHC class I and II structure and binding energy was compared with the binding energy of the redocked naturally bound peptides. The docking was performed at exhaustiveness 16 for all the common epitopes.

# Chapter 5: Results

### 5.1 Conserved sequences of nucleoprotein

Using the sequence data obtained from influenza research database (www.fludb.org), conserved sequences were identified for H3N2 and H5N1 using MUSLE alignment tool and AVANA.

#### 5.1.1 Conserved sequences of nucleoprotein, H5N1

Eighty eight sequences of NP protein (H5N1) available till January, 2015 were taken from influenza research database. Twelve conserved sequences were obtained as given in table 1. Sequence length of conserved peptides varies from 15 to 114.

#### 5.1.2 Conserved sequences of nucleoprotein, H3N2

Seven hundred fifty three sequences of NP protein (H3N2) were taken from influenza research database. Nineteen conserved sequences were obtained as given in table 2. Sequence length of conserved peptides varies from 9 to 60.

**Table 1: Conserved sequences of nucleoprotein of H5N1 virus**

| Conserved sequences | Sequences | Length |
|---|---|---|
| NP1 | MASQGTKRSYEQMETGGERQNATEIRASVGRMV | 33 |
| NP2 | GIGRFYIQMCTELKLSD | 17 |
| NP3 | EGRLIQNSITIERMVLSAFDERRNRYLEEHPSAGKDPKKTGGPIYRRRDG KWVRELILYDKEEIRRIWRQANNGEDATAGLTH | 83 |
| NP4 | MIWHSNLNDATYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAA | 46 |
| NP5 | KGVGTMVMELIRMIKRGINDRNFWRGENGRRTRIAYERMCNILKGKFQ TAAQ | 52 |
| NP6 | AMMDQVRESRNPGNAEIEDLIFLARSALILRGSVAHKSCLPACVYGLAV ASGYDFEREGYSLVGIDPFRLLQNSQVFSLIRPNENPAHKSQLVWMACH SAAFEDLRVSSFIRGT | 114 |
| NP7 | PRGQLSTRGVQIASNEN | 17 |
| NP8 | TLELRSRYWAIRTRSGGNTNQQ | 22 |
| NP9 | ASAGQISVQPTFSVQRNLPFER | 22 |
| NP10 | GNTEGRTSDMRTEII | 15 |
| NP11 | MMESARPEDVSFQGRGVFELSDEKATNPIVPSFDM | 35 |
| NP12 | NEGSYFFGDNAEEYDN | 16 |

**Table 2: Conserved sequences of nucleoprotein of H3N2 virus**

| Conserved sequences | Sequences | Length |
|---|---|---|
| NP1 | MASQGTKRSYEQMETDG | 17 |
| NP2 | EGRLIQNSLTIE | 12 |
| NP4 | MVLSAFDERRN | 11 |
| NP5 | YLEEHPSAGKDPKKTGGPIY | 20 |
| NP6 | WMRELVLYDKEEIRRIWRQANNG | 23 |
| NP7 | MIWHSNLND | 9 |
| NP8 | TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKG | 39 |
| NP9 | GTMVMELIRM | 10 |
| NP10 | KRGINDRNFWRGENGRKTRSAYERMCNILKGKFQTAAQRAM | 41 |
| NP11 | DQVRESRNPGNAEIEDLIF | 19 |
| NP12 | ARSALILRGSVAHKSCLPAC | 20 |
| NP13 | SGYDFEKEGYSLVGIDPFKLLQNSQ | 25 |
| NP14 | YSLIRPNENPAHKSQLVWMACHSAAFEDLRLLSFIRGTKVSPRGKLSTRG VQIASNENMD | 60 |
| NP15 | YWAIRTRSGGNTNQQRASAGQ | 21 |
| NP16 | FSVQRNLPF | 9 |
| NP17 | MAAFTGNTEGRTSDMRAEIIRMME | 24 |
| NP18 | KPEEVSFRGRGVFELSDEKA | 20 |
| NP19 | MSNEGSYFFGDNAEEYDN | 18 |

## 5.2 Identification of $CD8^+$ T cell epitopes of nucleoprotein of H5N1 and H3N2 virus

Influenza nucleoprotein (NP) is an internal protein which encapsidates the genome of the virus and it is more conserved than surface glycoproteins. BIMAS, NetCTL and SYFPEITHI tools were used to identify $CD8^+$ T cell epitopes from the conserved peptides of NP, H5N1 (Table 1) and H3N2 (Table 2). All the predicted epitopes were compared, then a set of epitopes was selected that was predicted by all the three tools.

In case of H5N1, initially 48 epitopes were identified (data not shown). These epitopes were analyzed for their homology with human proteome using BLASTp tool. Four of the predicted epitopes were found to be homologous to human self-proteins, thus eliminated from further studies. Thus, final set of 44 $CD8^+$ T cell epitopes were obtained (Table 3). Epitopes along with the number of MHC class I alleles predicted to bind these epitopes and score range are given in table 3. Overlapping epitopes were merged to generate peptides having single or multiple epitopes, and thus sixteen peptides were generated (Table 4).

In case of H3N2, initially 33 epitopes were identified (data not shown). Two of the predicted epitopes were found to be homologous to human self-proteins, thus eliminated from further studies. Thus, final set of 31 $CD8^+$ T cell epitopes were obtained (Table 5). Epitopes along with the number of MHC class I alleles predicted to bind these epitopes and score range are given in table 5. Overlapping epitopes were merged to generate 16 peptides having single or multiple epitopes (Table 6).

**Table 3: CD8$^+$ T cell epitopes of nucleoprotein of H5N1 which are commonly predicted by different immunoinformatics tools**

| Sequence | Epitope | BIMAS (cut off 20) | | NetCTL ( cut off 0.75) | | SYFPEITHI (cut off ≥20) | |
|---|---|---|---|---|---|---|---|
| | | No of alleles | Score range | Supertype | Score range | No of alleles | Score range |
| NP2 | FYIQMCTEL | 3 | 20-330 | 2 | 0.912-1.2497 | 1 | 22 |
| | GRFYIQMCT | 2 | 100-1000 | 1 | 1.0283 | 1 | 20 |
| | IQMCTELKL | 2 | 20-200 | 2 | 0.8774-0.9265 | 1 | 24 |
| NP3 | AFDERRNRY | 1 | 49.5 | 1 | 0.7641 | 2 | 22-28 |
| | FDERRNRYL | 2 | 20-40 | 2 | 1.0782-1.6614 | 3 | 20-27 |
| | IERMVLSAF | 1 | 90 | 3 | 0.8376-1.0802 | 3 | 20-23 |
| | KTGGPIYRR | 2 | 30-100 | 1 | 1.0131 | 3 | 20-21 |
| | RLIQNSITI | 1 | 27 | 3 | 0.7525-0.9627 | 6 | 20-22 |
| | RRDGKWVRE | 1 | 60 | 1 | 1.0621 | 1 | 20 |
| | RRIWRQANN | 2 | 30-600 | 1 | 1.2699 | 1 | 21 |
| | RRNRYLEEH | 2 | 60-600 | 1 | 0.9834 | 1 | 24 |
| | RRRDGKWVR | 1 | 3000 | 1 | 1.4652 | 1 | 29 |
| | VRELILYDK | 1 | 2000 | 1 | 0.9973 | 1 | 24 |
| NP4 | ATYQRTRAL | 2 | 20-50 | 4 | 0.8163-1.4061 | 2 | 20-22 |
| | GMDPRMCSL | 2 | 20-50 | 2 | 0.8659-1.0691 | 2 | 20-23 |
| | TRALVRTGM | 3 | 20-600 | 1 | 1.3824 | 1 | 22 |
| | VRTGMDPRM | 2 | 20-600 | 1 | 0.8374 | 2 | 21-24 |
| NP5 | AYERMCNIL | 3 | 150-360 | 2 | 1.0628-1.2864 | 1 | 20 |
| | KRGINDRNF | 2 | 180-900 | 1 | 1.0368 | 3 | 20-25 |
| | RRTRIAYER | 1 | 3000 | 1 | 1.7565 | 1 | 28 |
| NP6 | AEIEDLIFL | 7 | 21.544-640 | 1 | 1.9599 | 6 | 20-28 |
| | CLPACVYGL | 2 | 30-49.134 | 1 | 1.097 | 1 | 26 |
| | FEDLRVSSF | 1 | 120 | 3 | 0.7941-0.8618 | 3 | 22-31 |
| | FEREGYSLV | 1 | 40 | 1 | 1.3138 | 1 | 20 |
| | FLARSALIL | 3 | 21-40.289 | 4 | 0.9845-1.8955 | 3 | 20-24 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | IFLARSALI | 1 | 33 | 1 | 1.5085 | 1 | 21 |
| | ILRGSVAHK | 2 | 30-90 | 1 | 1.5717 | 3 | 20-30 |
| | KSCLPACVY | 1 | 20 | 4 | 0.8152-1.286 | 3 | 20-27 |
| | LIFLARSAL | 3 | 25-50 | 2 | 0.8587-1.0427 | 1 | 22 |
| | NPAHKSQLV | 3 | 40-242 | 1 | 0.9564 | 1 | 20 |
| | REGYSLVGI | 1 | 27 | 1 | 1.4817 | 2 | 20-22 |
| | RLLQNSQVF | 2 | 44-45 | 4 | 0.8651-1.3863 | 2 | 20-24 |
| | YSLVGIDPF | 1 | 48 | 2 | 1.141-1.3523 | 2 | 22 |
| NP8 | ELRSRYWAI | 1 | 80 | 1 | 2.1401 | 1 | 29 |
| | LRSRYWAIR | 1 | 100 | 1 | 0.9507 | 1 | 21 |
| | SRYWAIRTR | 1 | 1500 | 1 | 1.6628 | 1 | 26 |
| NP9 | FSVQRNLPF | 1 | 44 | 5 | 0.818-1.4207 | 2 | 20 |
| | GQISVQPTF | 4 | 20-176 | 5 | 0.8304-1.3337 | 2 | 20-21 |
| NP10 | TSDMRTEII | 1 | 24.2 | 1 | 1.2652 | 2 | 20-22 |
| NP11 | ATNPIVPSF | 1 | 144 | 4 | 0.7926-1.4748 | 3 | 21-27 |
| | FQGRGVFEL | 4 | 20-200 | 4 | 0.8249-2.192 | 1 | 22 |
| | GVFELSDEK | 3 | 45-240 | 1 | 0.8596 | 3 | 20-25 |
| | SARPEDVSF | 1 | 40 | 3 | 0.9045-1.1591 | 1 | 20 |
| | VSFQGRGVF | 3 | 25-56.25 | 2 | 1.2695-1.3546 | 1 | 20 |

**Table 4: Peptide containing overlapping CD8$^+$ T cell epitopes of nucleoprotein of H5N1**

| Sequence | Epitopes | Overlapping sequences |
|---|---|---|
| NP2 | FYIQMCTEL | GRFYIQMCTELKL |
| | GRFYIQMCT | |
| | IQMCTELKL | |
| NP3 | RLIQNSITI | RLIQNSITIERMVLSAFDERRNRYLEEH |
| | IERMVLSAF | |
| | AFDERRNRY | |
| | FDERRNRYL | |
| | RRNRYLEEH | |
| | RRDGKWVRE | KTGGPIYRRRDGKWVRELILYDK |
| | RRRDGKWVR | |
| | KTGGPIYRR | |
| | VRELILYDK | |
| | RRIWRQANN | RRIWRQANN |
| NP4 | ATYQRTRAL | ATYQRTRALVRTGMDPRMCSL |
| | GMDPRMCSL | |
| | TRALVRTGM | |
| | VRTGMDPRM | |
| NP5 | RRTRIAYER | RRTRIAYERMCNIL |
| | AYERMCNIL | |
| | KRGINDRNF | KRGINDRNF |
| NP6 | IFLARSALI | AEIEDLIFLARSALILRGSVAHKSCLPACVYGL |
| | FLARSALIL | |
| | LIFLARSAL | |
| | AEIEDLIFL | |
| | ILRGSVAHK | |
| | KSCLPACVY | |
| | CLPACVYGL | |
| | FEREGYSLV | FEREGYSLVGIDPFRLLQNSQVF |
| | REGYSLVGI | |
| | YSLVGIDPF | |
| | RLLQNSQVF | |
| | FEDLRVSSF | FEDLRVSSF |
| | NPAHKSQLV | NPAHKSQLV |
| NP8 | ELRSRYWAI | ELRSRYWAIRTR |
| | LRSRYWAIR | |
| | SRYWAIRTR | |
| NP9 | FSVQRNLPF | GQISVQPTFSVQRNLPF |
| | GQISVQPTF | |
| NP10 | TSDMRTEII | TSDMRTEII |
| NP11 | FQGRGVFEL | SARPEDVSFQGRGVFELSDEK |
| | GVFELSDEK | |
| | VSFQGRGVF | |
| | SARPEDVSF | |
| | ATNPIVPSF | ATNPIVPSF |

**Table 5: CD8$^+$ T cell epitopes of nucleoprotein of H3N2 which are commonly predicted by different immunoinformatics tools.**

| Sequence | Epitope | BIMAS (cut off 20) | | NetCTL ( cut off 0.75) | | SYFPEITHI (cut off ≥20) | |
|---|---|---|---|---|---|---|---|
| | | No of alleles | Score range | Supertype | Score range | No of alleles | Score range |
| NP2 | FYIQMCTEL | 3 | 20-330 | 2 | 0.912-1.2497 | 1 | 22 |
| | GRFYIQMCT | 2 | 100-1000 | 1 | 1.0283 | 1 | 20 |
| | IQMCTELKL | 3 | 20-200 | 2 | 0.8774-0.9265 | 1 | 24 |
| | KMIDGIGRF | 1 | 75 | 3 | 1.2387-1.4813 | 1 | 21 |
| | MIDGIGRFY | 1 | 25 | 2 | 0.9043-2.3523 | 1 | 25 |
| NP3 | EGRLIQNSL | 2 | 40-48.4 | 1 | 0.9136 | 1 | 21 |
| | RLIQNSLTI | 1 | 27 | 4 | 0.8344-1.0003 | 3 | 22-24 |
| NP6 | MRELVLYDK | 1 | 2000 | 1 | 1.0723 | 1 | 24 |
| | RRIWRQANN | 2 | 30-600 | 1 | 1.2913 | 1 | 21 |
| NP8 | GMDPRMCSL | 2 | 20-50 | 2 | 0.8659-1.0691 | 2 | 20-23 |
| | TRALVRTGM | 3 | 20-600 | 1 | 1.3824 | 1 | 22 |
| | VRTGMDPRM | 2 | 20-600 | 1 | 0.8374 | 2 | 21-24 |
| NP10 | AYERMCNIL | 2 | 220-360 | 2 | 1.0651-1.2887 | 1 | 20 |
| | KRGINDRNF | 2 | 180-900 | 1 | 1.0319 | 3 | 20-25 |
| NP12 | ILRGSVAHK | 2 | 30-90 | 1 | 1.5719 | 3 | 20-30 |
| NP13 | FEKEGYSLV | 1 | 40 | 1 | 1.2397 | 1 | 20 |
| | KEGYSLVGI | 1 | 27 | 1 | 1.2092 | 2 | 21 |
| | SLVGIDPFK | 2 | 30-67.5 | 1 | 1.074 | 2 | 20-23 |
| | YSLVGIDPF | 1 | 48 | 2 | 1.1409-1.3521 | 2 | 22 |
| NP14 | FEDLRLLSF | 1 | 60 | 1 | 1.1415 | 3 | 23-32 |
| | IRGTKVSPR | 1 | 1000 | 1 | 1.0286 | 1 | 24 |
| | KLSTRGVQI | 2 | 27-36.515 | 1 | 0.837 | 3 | 21-25 |
| | LLSFIRGTK | 1 | 30 | 1 | 1.4445 | 1 | 26 |
| | NPAHKSQLV | 3 | 40-242 | 1 | 0.9564 | 1 | 20 |
| NP16 | FSVQRNLPF | 1 | 44 | 5 | 0.7973-1.3999 | 2 | 20 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NP17 | MRAEIIRMM | 1 | 180 | 1 | 1.2612 | 1 | 23 |
| | TSDMRAEII | 1 | 22 | 1 | 1.1064 | 2 | 21-22 |
| NP18 | FRGRGVFEL | 4 | 22.5-2000 | 2 | 0.9393-2.313 | 4 | 21-26 |
| | GVFELSDEK | 3 | 45-240 | 1 | 0.858 | 3 | 21-25 |
| | VSFRGRGVF | 3 | 25-56.25 | 2 | 1.328-1.3818 | 1 | 20 |
| NP19 | MSNEGSYFF | 1 | 88 | 3 | 1.1623-1.4455 | 1 | 21 |

**Table 6: Peptide containing overlapping CD8$^+$ T cell epitopes of nucleoprotein of H3N2**

| Sequence | Epitopes | Overlapping sequences |
|---|---|---|
| NP2 | FYIQMCTEL | KMIDGIGRFYIQMCTELKL |
| | GRFYIQMCT | |
| | IQMCTELKL | |
| | KMIDGIGRF | |
| | MIDGIGRFY | |
| NP3 | EGRLIQNSL | EGRLIQNSLTI |
| | RLIQNSLTI | |
| NP6 | MRELVLYDK | MRELVLYDK |
| | RRIWRQANN | RRIWRQANN |
| NP8 | GMDPRMCSL | TRALVRTGMDPRMCSL |
| | TRALVRTGM | |
| | VRTGMDPRM | |
| NP10 | AYERMCNIL | AYERMCNIL |
| | KRGINDRNF | KRGINDRNF |
| NP12 | ILRGSVAHK | ILRGSVAHK |
| NP13 | FEKEGYSLV | FEKEGYSLVGIDPFK |
| | KEGYSLVGI | |
| | SLVGIDPFK | |
| | YSLVGIDPF | |
| NP14 | FEDLRLLSF | FEDLRLLSFIRGTKVSPR |
| | LLSFIRGTK | |
| | IRGTKVSPR | |
| | KLSTRGVQI | KLSTRGVQI |
| | NPAHKSQLV | NPAHKSQLV |
| NP16 | FSVQRNLPF | FSVQRNLPF |
| NP17 | MRAEIIRMM | TSDMRAEIIRMM |
| | TSDMRAEII | |
| NP18 | FRGRGVFEL | VSFRGRGVFELSDEK |
| | GVFELSDEK | |
| | VSFRGRGVF | |
| NP19 | MSNEGSYFF | MSNEGSYFF |

### 5.3 Identification of $CD4^+$ T cell epitopes of nucleoprotein of H5N1 and H3N2 virus

NetMHCII 2.2, Propred, and IEDB SMM align tools were used to identify $CD4^+$ T cell epitopes from the conserved peptides of nucleoprotein of H5N1 (Table 1) and H3N2 (Table 2). The cutoff selected for prediction was IC50≤ 500 (for IEDB-SMM align and NetMHCII 2.2). 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 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.

In case of H5N1, initially, 44 epitopes were identified. Three of the predicted epitope was found to be homologous to human self-proteins, thus eliminated from further studies. Final set of 41 $CD4^+$ T cell binding epitopes. Epitope along with the number of MHC class II alleles to which they are predicted to bind and score range are given in table 7. Overlapping epitopes were merged to generate 18 peptides having single or multiple epitopes (Table 8).

In case of H3N2, initially 35 epitopes were identified. One of the predicted epitope was found to be homologous to human self-proteins, thus eliminated. Final set of 34 $CD4^+$ T cell binding epitopes, the number of MHC class II alleles to which they are predicted to bind and score range are given in table 9. Overlapping epitopes were merged to generate 15 peptides having single or multiple epitopes (Table 10).

**Table 7: CD4[+] T cell epitopes of nucleoprotein of H5N1 which are commonly predicted by different immunoinformatics tools**

| Sequence | Epitopes | IEDB-SMM Align (cut off IC50≤500) | | NETmhc2 (cut off IC50≤500) | | Propred |
|---|---|---|---|---|---|---|
| | | No of Alleles | IC50 value | No of Alleles | IC50 value | No of Alleles |
| NP2 | FYIQMCTEL | 1 | 325 | 18 | 27.6-438.7 | 11 |
| | IGRFYIQMC | 1 | 452 | 3 | 91.9-346 | 10 |
| | IQMCTELKL | 7 | 220-446 | 4 | 29.7-144.4 | 3 |
| | YIQMCTELK | 19 | 35-443 | 7 | 10.3-259.8 | 15 |
| NP3 | ILYDKEEIR | 10 | 37-481 | 27 | 7.6-371.1 | 1 |
| | IQNSITIER | 5 | 214-463 | 11 | 78.7-457.7 | 15 |
| | IRRIWRQAN | 11 | 95-500 | 17 | 49.5-477 | 20 |
| | ITIERMVLS | 12 | 173-472 | 35 | 30.2-409.7 | 2 |
| | LILYDKEEI | 4 | 203-219 | 5 | 441.5-498.9 | 24 |
| | VRELILYDK | 1 | 262 | 7 | 138.7-477.8 | 2 |
| | WVRELILYD | 6 | 201-500 | 30 | 55.4-483.2 | 13 |
| NP4 | LMQGSTLPR | 9 | 111-457 | 15 | 160.9-478.5 | 6 |
| | YQRTRALVR | 34 | 15-491 | 58 | 9.8-469.7 | 11 |
| NP5 | IRMIKRGIN | 1 | 313 | 21 | 8.4-453.5 | 26 |
| | LIRMIKRGI | 39 | 12-499 | 27 | 7.4-498.5 | 29 |
| | MELIRMIKR | 4 | 12-348 | 14 | 45.6-414.6 | 2 |
| | MVMELIRMI | 14 | 74-481 | 34 | 13.2-489.4 | 26 |
| | VMELIRMIK | 3 | 124-336 | 3 | 68.4-420.9 | 4 |
| | WRGENGRRT | 4 | 217-248 | 4 | 164.4-358.1 | 1 |
| | YERMCNILK | 34 | 53-430 | 29 | 28.6-495.9 | 11 |
| NP6 | FLARSALIL | 32 | 14-356 | 24 | 27.5-482.2 | 7 |
| | FSLIRPNEN | 38 | 12-437 | 46 | 28.7-481.2 | 3 |
| | IFLARSALI | 65 | 7-466 | 71 | 5-486.6 | 13 |
| | ILRGSVAHK | 3 | 73-458 | 24 | 30.5-498.9 | 7 |
| | IRPNENPAH | 6 | 198-357 | 8 | 59.4-337.4 | 11 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | LIFLARSAL | 15 | 13-321 | 15 | 18.5-396.4 | 6 |
| | LILRGSVAH | 1 | 386 | 11 | 58.5-484.1 | 11 |
| | LRGSVAHKS | 23 | 61-499 | 25 | 8.8-478.9 | 33 |
| | LRVSSFIRG | 1 | 270 | 9 | 74.3-397.2 | 2 |
| | LVGIDPFRL | 34 | 5-500 | 29 | 19.9-444.1 | 34 |
| | LVWMACHSA | 5 | 66-400 | 6 | 54.2-361.9 | 38 |
| | VFSLIRPNE | 3 | 99-294 | 8 | 35.9-373.2 | 5 |
| | VYGLAVASG | 11 | 48-454 | 28 | 11.6-424.5 | 22 |
| | WMACHSAAF | 20 | 114-489 | 28 | 49.8-487.3 | 2 |
| | YGLAVASGY | 23 | 69-497 | 24 | 13.1-489.6 | 4 |
| NP8 | LRSRYWAIR | 4 | 165-380 | 4 | 28.6-54.4 | 2 |
| | YWAIRTRSG | 19 | 56-497 | 40 | 5.4-482.4 | 1 |
| NP9 | FSVQRNLPF | 16 | 94-349 | 38 | 8.5-452.8 | 3 |
| | VQPTFSVQR | 3 | 104-207 | 13 | 138.7-449.9 | 3 |
| NP11 | FQGRGVFEL | 14 | 129-434 | 9 | 131.7-470.3 | 13 |
| NP12 | FFGDNAEEY | 10 | 100-451 | 10 | 16-418.7 | 5 |

**Table 8: Peptide containing Overlapping $CD4^+$ T cell epitopes of nucleoprotein of H5N1**

| Sequence | Epitopes | Overlapping Sequences |
|---|---|---|
| NP2 | FYIQMCTEL | IGRFYIQMCTELKL |
| | IGRFYIQMC | |
| | IQMCTELKL | |
| | YIQMCTELK | |
| NP3 | IQNSITIER | IQNSITIERMVLS |
| | ITIERMVLS | |
| | ILYDKEEIR | WVRELILYDKEEIRRIWRQAN |
| | LILYDKEEI | |
| | VRELILYDK | |
| | WVRELILYD | |
| | IRRIWRQAN | |
| NP4 | LMQGSTLPR | LMQGSTLPR |
| | YQRTRALVR | YQRTRALVR |
| NP5 | IRMIKRGIN | MVMELIRMIKRGIN |
| | LIRMIKRGI | |
| | MELIRMIKR | |
| | MVMELIRMI | |
| | VMELIRMIK | |
| | WRGENGRRT | WRGENGRRT |
| | YERMCNILK | YERMCNILK |
| NP6 | FLARSALIL | LIFLARSALILRGSVAHKS |
| | IFLARSALI | |
| | LIFLARSAL | |
| | ILRGSVAHK | |
| | LILRGSVAH | |
| | LRGSVAHKS | |
| | LRVSSFIRG | LRVSSFIRG |
| | LVGIDPFRL | LVGIDPFRL |
| | LVWMACHSA | LVWMACHSAAF |
| | WMACHSAAF | |
| | VYGLAVASG | VYGLAVASGY |
| | YGLAVASGY | |
| | FSLIRPNEN | VFSLIRPNENPAH |
| | VFSLIRPNE | |
| | IRPNENPAH | |
| NP8 | LRSRYWAIR | LRSRYWAIRTRSG |
| | YWAIRTRSG | |
| NP9 | FSVQRNLPF | VQPTFSVQRNLPF |
| | VQPTFSVQR | |
| NP11 | FQGRGVFEL | FQGRGVFEL |
| NP12 | FFGDNAEEY | FFGDNAEEY |

**Table 9: CD4$^{+}$ T cell epitopes of nucleoprotein of H3N2 which are commonly predicted by different immunoinformatics tools.**

| Sequence | Epitopes | IEDB-SMM Align (cut off IC50≤500) | | NETmhc2 (cut off IC50≤500) | | Propred |
|---|---|---|---|---|---|---|
| | | No of Alleles | IC50 value | No of Alleles | IC50 value | No of alleles |
| NP2 | FYIQMCTEL | 4 | 41-325 | 35 | 27.6-438.7 | 11 |
| | IGRFYIQMC | 2 | 81-452 | 21 | 91.9-358 | 10 |
| | IQMCTELKL | 7 | 220-446 | 12 | 29.7-406.8 | 5 |
| | YIQMCTELK | 19 | 35-443 | 25 | 10.3-356.9 | 13 |
| NP6 | IRRIWRQAN | 4 | 95-500 | 6 | 66.3-477 | 20 |
| | LVLYDKEEI | 3 | 316-339 | 1 | 478.7 | 34 |
| | VLYDKEEIR | 6 | 53-176 | 15 | 10.9-436.3 | 1 |
| NP8 | LMQGSTLPR | 9 | 111-457 | 15 | 160.9-478.5 | 6 |
| | YQRTRALVR | 8 | 46-273 | 11 | 16.1-431.3 | 11 |
| NP10 | LKGKFQTAA | 3 | 246-393 | 8 | 200.6-476.1 | 1 |
| | MCNILKGKF | 1 | 232 | 9 | 53.5-442 | 3 |
| | WRGENGRKT | 5 | 202-381 | 3 | 278.2-466.8 | 1 |
| | YERMCNILK | 34 | 55-459 | 26 | 29.5-410 | 11 |
| NP12 | ILRGSVAHK | 3 | 73-458 | 19 | 70.1-418.3 | 8 |
| | LILRGSVAH | 1 | 386 | 9 | 58.5-484.1 | 11 |
| | LRGSVAHKS | 22 | 61-499 | 22 | 8.8-478.9 | 33 |
| NP13 | LVGIDPFKL | 26 | 59-494 | 32 | 28.6-494.4 | 34 |
| NP14 | FEDLRLLSF | 26 | 9-393 | 55 | 5.6-441.5 | 2 |
| | FIRGTKVSP | 4 | 250-390 | 12 | 15.3-298 | 2 |
| | IRGTKVSPR | 12 | 156-465 | 41 | 18.5-485.8 | 8 |
| | IRPNENPAH | 3 | 198-390 | 4 | 53-298.5 | 11 |
| | LLSFIRGTK | 4 | 126-183 | 15 | 13.4-355.8 | 2 |
| | LRLLSFIRG | 14 | 25-483 | 17 | 17.5-242.4 | 18 |
| | LSFIRGTKV | 44 | 9-481 | 51 | 4.6-480.9 | 1 |
| | LSTRGVQIA | 1 | 471 | 2 | 432.3-487.4 | 4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | LVWMACHSA | 5 | 66-400 | 6 | 54.2-361.9 | 38 |
| | MACHSAAFE | 1 | 257 | 10 | 154.7-474.2 | 1 |
| | VQIASNENM | 3 | 331-406 | 8 | 85.6-497.8 | 12 |
| | WMACHSAAF | 20 | 114-490 | 28 | 49.8-487.3 | 2 |
| | YSLIRPNEN | 2 | 67-329 | 2 | 328.2-472.4 | 3 |
| NP15 | YWAIRTRSG | 2 | 176-467 | 1 | 26.7 | 1 |
| NP17 | MRAEIIRMM | 2 | 269-458 | 11 | 109.7-470.8 | 12 |
| NP18 | FRGRGVFEL | 7 | 251-428 | 11 | 170.1-474.6 | 29 |
| NP19 | FFGDNAEEY | 10 | 100-451 | 14 | 16-418.7 | 5 |

**Table 10: Peptide containing Overlapping CD4+ T cell Epitopes of nucleoprotein of H3N2**

| Sequence | Epitopes | Overlapping sequences |
|---|---|---|
| NP2 | FYIQMCTEL | IGRFYIQMCTELKL |
| | IGRFYIQMC | |
| | IQMCTELKL | |
| | YIQMCTELK | |
| NP6 | LVLYDKEEI | LVLYDKEEIRRIWRQAN |
| | VLYDKEEIR | |
| | IRRIWRQAN | |
| NP8 | LMQGSTLPR | LMQGSTLPR |
| | YQRTRALVR | YQRTRALVR |
| NP10 | LKGKFQTAA | YERMCNILKGKFQTAA |
| | MCNILKGKF | |
| | YERMCNILK | |
| | WRGENGRKT | WRGENGRKT |
| NP12 | ILRGSVAHK | LILRGSVAHKS |
| | LILRGSVAH | |
| | LRGSVAHKS | |
| NP13 | LVGIDPFKL | LVGIDPFKL |
| NP14 | FIRGTKVSP | LVWMACHSAAFEDLRLLSFIRGTKVSPR |
| | IRGTKVSPR | |
| | FEDLRLLSF | |
| | LLSFIRGTK | |
| | LRLLSFIRG | |
| | LSFIRGTKV | |
| | LVWMACHSA | |
| | MACHSAAFE | |
| | WMACHSAAF | |
| | YSLIRPNEN | YSLIRPNENPAH |
| | IRPNENPAH | |
| | LSTRGVQIA | LSTRGVQIASNENM |
| | VQIASNENM | |
| NP15 | YWAIRTRSG | YWAIRTRSG |
| NP17 | MRAEIIRMM | MRAEIIRMM |
| NP18 | FRGRGVFEL | FRGRGVFEL |
| NP19 | FFGDNAEEY | FFGDNAEEY |

## 5.4 Identification of immunogenic peptides containing MHC class I and II epitopes of nucleoprotein (H5N1 and H3N2)

Putative immunogenic peptides of H5N1 as given in tables 4 and 8, and of H3N2 as given in tables 6 and 10 were merged to generate peptide fragments containing both, MHC I and MHC II epitopes. Ten putative immunogenic peptides of H5N1 were identified containing both $CD8^+$ and $CD4^+$ epitopes (Table 11), Whereas 12 putative immunogenic peptides were found in case of H3N2 (Table12). These peptides may be considered for good target of vaccine design for H5N1 and H3N2 virus as they may be capable of generating both, MHC class I and II mediated immune response.

**Table 11: Prediction of common immunogenic epitopes/peptides of NP (H5N1)**

| Sequence | Overlapping sequence of MHC class I epitopes | No of epitopes | Overlapping sequence of MHC class II epitopes | No of epitopes | Overlapping peptides |
|---|---|---|---|---|---|
| NP2 | GRFYIQMCTELKL | 3 | IGRFYIQMCTELKL | 4 | IGRFYIQMCTELKL |
| NP3 | RLIQNSITIERMVLSAFDERRNRYLEEH | 5 | IQNSITIERMVLS | 2 | RLIQNSITIERMVLSAFDERRNRYLEEH |
| | KTGGPIYRRRDGKWVRELILYDK | 4 | WVRELILYDKEEIRRIWRQAN | 5 | KTGGPIYRRRDGKWVRELILYDKEEIRRIWRQANN |
| | RRIWRQANN | 1 | | | |
| NP4 | ATYQRTRALVRTGMDPRMCSL | 4 | LMQGSTLPR | 1 | ATYQRTRALVRTGMDPRMCSLMQGSTLPR |
| | | | YQRTRALVR | 1 | |
| NP5 | RRTRIAYERMCNIL | 2 | MVMELIRMIKRGIN | 5 | MVMELIRMIKRGINDRNFWRGENGRRTRIAYERMCNILK |
| | KRGINDRNF | 1 | WRGENGRRT | 1 | |
| | | | YERMCNILK | 1 | |
| NP6 | AEIEDLIFLARSALILRGSVAHKSCLPACVYGL | 7 | LIFLARSALILRGSVAHKS | 6 | AEIEDLIFLARSALILRGSVAHKSCLPACVYGLAVASGY |
| | | | VYGLAVASGY | 2 | |
| | FEREGYSLVGIDPFRLLQNSQVF | 4 | LRVSSFIRG | 1 | FEREGYSLVGIDPFRLLQNSQVFSLIRPNENPAHKSQLVWMACHSAAFEDLRVSSFIRG |
| | FEDLRVSSF | 1 | LVGIDPFRL | 1 | |
| | NPAHKSQLV | 1 | LVWMACHSAAF | 2 | |
| | | | VFSLIRPNENPAH | 3 | |
| NP8 | ELRSRYWAIRTR | 3 | LRSRYWAIRTRSG | 2 | ELRSRYWAIRTRSG |
| NP9 | GQISVQPTFSVQRNLPF | 2 | VQPTFSVQRNLPF | 2 | GQISVQPTFSVQRNLPF |
| NP11 | SARPEDVSFQGRGVFELSDEK | 4 | FQGRGVFEL | 1 | SARPEDVSFQGRGVFELSDEKATNPIVPSF |
| | ATNPIVPSF | 1 | | | |

**Table 12: Prediction of common immunogenic epitopes/peptides of NP (H3N2)**

| Sequence | Overlapping sequence of MHC class I epitopes | No of epitopes | Overlapping sequence of MHC class II epitopes | No of epitopes | Overlapping peptides |
|---|---|---|---|---|---|
| NP2 | KMIDGIGRFYIQMCTELKL | 5 | IGRFYIQMCTELKL | 4 | KMIDGIGRFYIQMCTELKL |
| NP6 | MRELVLYDK | 1 | LVLYDKEEIRRIWRQAN | 3 | MRELVLYDKEEIRRIWRQANN |
| | RRIWRQANN | 1 | | | |
| NP8 | TRALVRTGMDPRMCSL | 3 | LMQGSTLPR | 1 | YQRTRALVRTGMDPRMCSLMQGSTLPR |
| | | | YQRTRALVR | 1 | |
| NP10 | AYERMCNIL | 1 | YERMCNILKGKFQTAA | 3 | AYERMCNILKGKFQTAA |
| | KRGINDRNF | 1 | WRGENGRKT | 1 | KRGINDRNFWRGENGRKT |
| NP12 | ILRGSVAHK | 1 | LILRGSVAHKS | 3 | LILRGSVAHKS |
| NP13 | FEKEGYSLVGIDPFK | 4 | LVGIDPFKL | 1 | FEKEGYSLVGIDPFKL |
| NP14 | FEDLRLLSFIRGTKVSPR | 3 | LVWMACHSAAFEDLRLLSFIRGTKVSPR | 9 | YSLIRPNENPAHKSQLVWMACHSAAFEDLRLLSFIRGTKVSPR |
| | NPAHKSQLV | 1 | YSLIRPNENPAH | 2 | |
| | KLSTRGVQI | 1 | LSTRGVQIASNENM | 2 | KLSTRGVQIASNENM |
| NP17 | TSDMRAEIIRMM | 2 | MRAEIIRMM | 1 | TSDMRAEIIRMM |
| NP18 | VSFRGRGVFELSDEK | 3 | FRGRGVFEL | 1 | VSFRGRGVFELSDEK |
| NP19 | MSNEGSYFF | 1 | FFGDNAEEY | 1 | MSNEGSYFFGDNAEEY |

## 5.5 Identification of unique T cell epitopes from epitopes of H5N1 and H3N2

Nucleoprotein of influenza A virus shares similarity among various subtypes. Therefore, it was interesting to find some of the epitopes which were unique to H5N1 and H3N2. Out of total 85 epitopes predicted for H5N1 (44 for MHC Class I and 41 for MHC class II), 44 epitopes were found to be unique (Table 13). Similarly, in case of H3N2, 22 epitopes were found to be unique out of 65 epitopes predicted (Table 14).

**Table 13: Unique epitopes in nucleoprotein of H5N1**

| Conserved sequences | MHC I epitopes | MHC II epitopes |
|---|---|---|
| NP3 | AFDERRNRY<br>KTGGPIYRR<br>RLIQNSITI<br>RRDGKWVRE<br>RRNRYLEEH<br>RRRDGKWVR | ILYDKEEIR<br>IQNSITIER<br>ITIERMVLS<br>LILYDKEEI<br>WVRELILYD |
| NP4 | ATYQRTRAL | |
| NP5 | RRTRIAYER | MELIRMIKR<br>MVMELIRMI<br>VMELIRMIK<br>WRGENGRRT |
| NP6 | AEIEDLIFL | FRLLQNSQV |
| | CLPACVYGL | FSLIRPNEN |
| | FEDLRVSSF | IFLARSALI |
| | FEREGYSLV | LIFLARSAL |
| | FRLLQNSQV | LQNSQVFSL |
| | IFLARSALI | LRVSSFIRG |
| | LIFLARSAL | LVGIDPFRL |
| | LQNSQVFSL | VFSLIRPNE |
| | REGYSLVGI | VYGLAVASG |
| | RLLQNSQVF | YGLAVASGY |
| NP9 | GQISVQPTF | FSVQRNLPF<br>VQPTFSVQR |
| NP11 | ATNPIVPSF<br>SARPEDVSF<br>VSFQGRGVF | FQGRGVFEL |

**Table 14: Unique epitopes in nucleoprotein of H3N2**

| Conserved sequences | MHC I epitopes | MHC II epitopes |
|---|---|---|
| NP2 | KMIDGIGRF | |
| | MIDGIGRFY | |
| NP3 | EGRLIQNSL | |
| | RLIQNSLTI | |
| NP6 | MRELVLYDK | LVLYDKEEI |
| | | VLYDKEEIR |
| NP10 | | LKGKFQTAA |
| | | MCNILKGKF |
| | | WRGENGRKT |
| NP13 | FEKEGYSLV | LVGIDPFKL |
| | KEGYSLVGI | |
| | SLVGIDPFK | |
| NP14 | | LSFIRGTKV |
| | | FIRGTKVSP |
| | | LRLLSFIRG |
| | | LSTRGVQIA |
| | | MACHSAAFE |
| | | VQIASNENM |
| | | YSLIRPNEN |
| NP18 | VSFRGRGVF | |

## 5.6 Identification of epitopes capable of binding both the MHCs

MHC class I and II, both play an important role in the elimination of viral infection. Nine epitopes of H5N1 were predicted to bind both MHC class I and II. Similarly, seven epitopes of H3N2 were reported to bind both the MHCs (Table 15).

**Table 25: Common epitopes between MHC I and MHC II of H5N1 and H3N2**

| H5N1 subtype | | H3N2 subtype | |
|---|---|---|---|
| **Sequence** | **epitopes** | **Sequence** | **epitopes** |
| NP2 | FYIQMCTEL | NP2 | FYIQMCTEL |
| | IQMCTELKL | | IQMCTELKL |
| NP3 | VRELILYDK | NP12 | ILRGSVAHK |
| NP6 | FLARSALIL | NP14 | IRGTKVSPR |
| | IFLARSALI | | FEDLRLLSF |
| | LIFLARSAL | NP17 | MRAEIIRMM |
| NP8 | LRSRYWAIR | NP18 | FRGRGVFEL |
| NP9 | FSVQRNLPF | | |
| NP11 | FQGRGVFEL | | |

### 5.7 Identification of epitopes common among H5N1 and H3N2

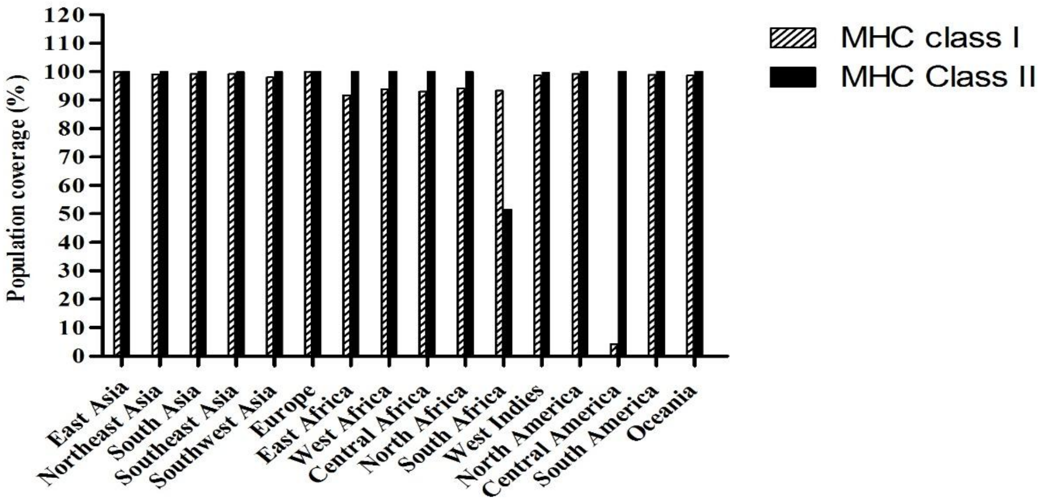
Twenty eight predicted epitopes were found to be common among H5N1 and H3N2 subtypes of influenza A virus. These epitopes can be assumed to be the stronger candidates for vaccine development since, they may impart hetrosubtypic immunity against H3N2 and H5N1 (Table 16).

| **Table 16: Common epitopes between H5N1 and H3N2 subtypes** | | |
|---|---|---|
| FYIQMCTEL | GVFELSDEK | FFGDNAEEY |
| GRFYIQMCT | RRIWRQANN | YIQMCTELK |
| IQMCTELKL | TSDMRAEII | VRTGMDPRM |
| AYERMCNIL | YSLVGIDPF | LILRGSVAH |
| FSVQRNLPF | IGRFYIQMC | LMQGSTLPR |
| GMDPRMCSL | IRPNENPAH | LRGSVAHKS |
| ILRGSVAHK | IRRIWRQAN | YERMCNILK |
| KRGINDRNF | LVWMACHSA | YQRTRALVR |
| NPAHKSQLV | WMACHSAAF | |
| TRALVRTGM | YWAIRTRSG | |

### 5.8 Population Coverage Analysis

It is important to find out the expected immune response generation in the global population by the predicted epitopes which was done by population coverage analysis. Peptides containing multiple $CD4^+$ and $CD8^+$ T cell specific epitopes of H5N1 and H3N2 were subjected to population coverage analysis. These epitopes and corresponding MHC restricted alleles obtained by prediction servers were the input data in the IEDB population coverage tool. Then the tool calculates the expected response of these epitopes in different population based on these input data. The average population coverage for immunogenic response of predicted peptides were found out to be 91.28 % (MHC class I) and 96.94 % (MHC class II) for H5N1 (Figure 5A) and 89.28% (MHC class I) and 96.94% (MHC class I) for H3N2 (Figure 5B).

A




B

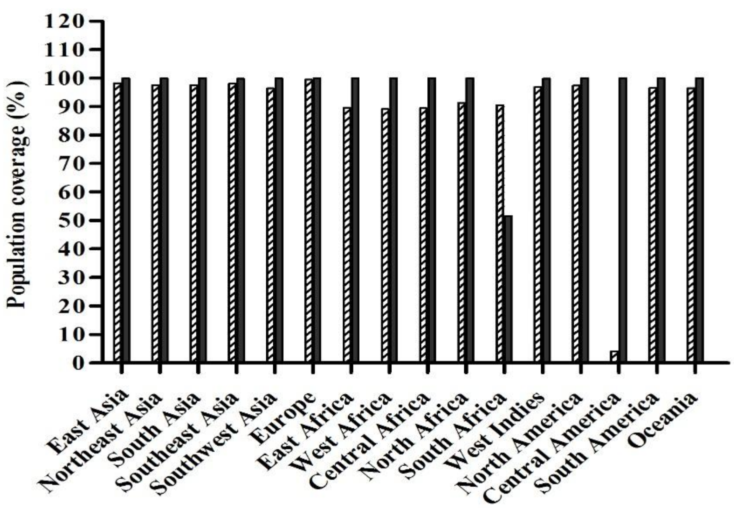



**Figure 5: Population Coverage Analysis of predicted epitopes of MHC I and II of (A) H5N1 (B) H3N2**

### 5.9 Epitope structure prediction

PEP-FOLD predict the structure of small peptides from the amino acid sequence. It predicts the structure of four consecutive amino acids and then combine the series of structural alphabet letter to generate the structure (Thevenet *et al*., 2012). Structure of 11 epitopes which were found to be commonly predicted among three subtypes of influenza A virus i.e. H5N1, H3N2 (from the current study) and H1N1 (from the previous work done in our lab), were generated by the server. Further, these structures were used for epitope docking (Table 17).

**Table 17: List of epitopes common among H1N1, H5N1 and H3N2**

| **MHC class I epitopes** | | | **MHC class II epitopes** | | |
|---|---|---|---|---|---|
| AYERMCNIL | GRFYIQMCT | NPAHKSQLV | FFGDNAEEY | ILRGSVAHK | YERMCNILK |
| FSVQRNLPF | ILRGSVAHK | TRALVRTGM | FYIQMCTEL | LILRGSVAH | YIQMCTELK |
| FYIQMCTEL | IQMCTELKL | VRTGMDPRM | IGRFYIQMC | LMQGSTLPR | YQRTRALVR |
| GMDPRMCSL | KRGINDRNF | | YWAIRTRSG | LRGSVAHKS | |

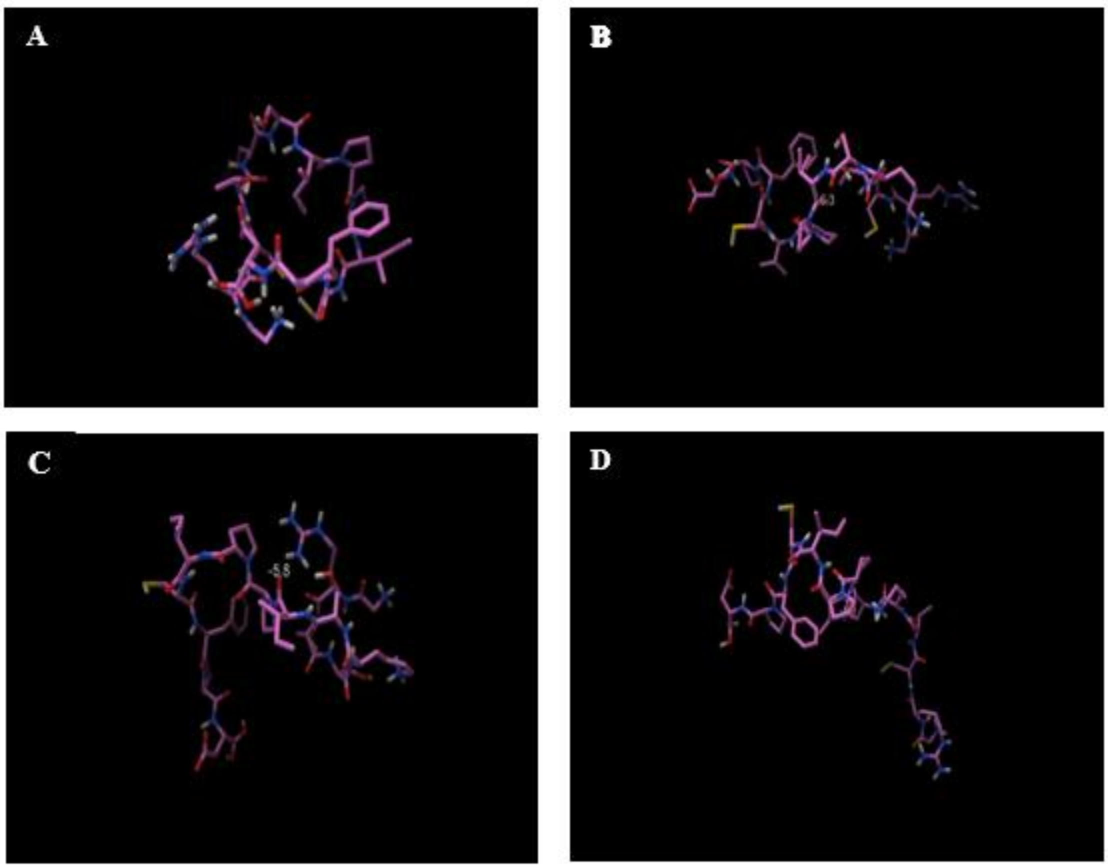



**Figure 6: PEP-FOLD predicted structures of some common epitopes among three subtypes of influenza, that are of MHC class I (A) AYERMCNIL, (B) ILRGSVAHK, and MHC class II (C) FFGDNAEEY (D) LMQGSTLPR.**

## 5.10 Epitope Docking with Class I and II MHC molecules

The immunogenic potential of the predicted epitopes depends upon the fact that how efficiently it can bind to peptide binding pocket of the MHC molecule. The auto dock vina tool has been used to assess the binding affinity of the predicted epitope with MHC class I and class II molecules. Two peptide MHC complex for Class I (HLA-A2, 3MRK) and class II (HLA-DR3, 1A6A) were retrieved from PDB database for docking purpose. Peptides were extracted from the MHC complex and then re-docked with corresponding MHC molecules using auto dock vina. The docking poses of the bound epitopes are given in figure 6 (representative results) for selective MHC class I and class II alleles to demonstrate the docking. Some peptides were found to bind outside the peptide binding groove of MHC were designated as non-binders.

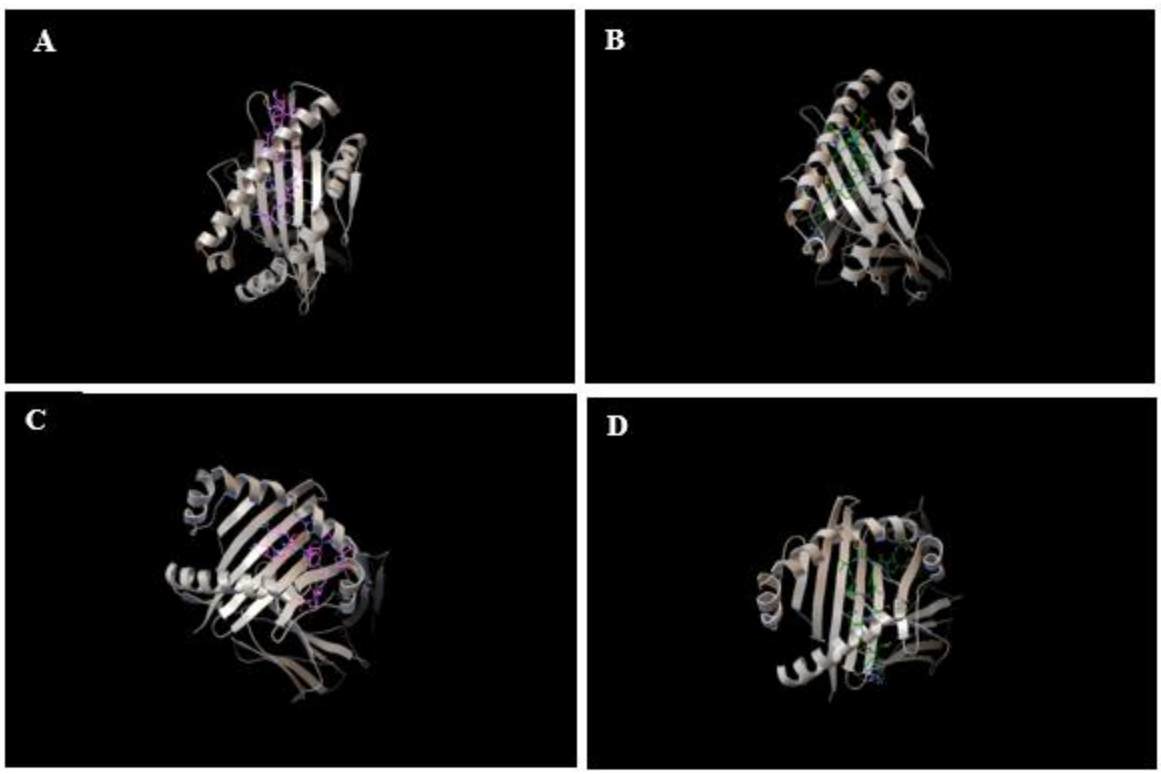



**Figure 7: The epitopes A and B are docked to the epitope binding pocket of MHC class I molecule HLA-A2 (PDB ID: 3MRK), epitopes C and D are docked to epitope binding pocket of MHC class II molecule HLA-DR3 (PDB ID: 1A6A).**

Eleven epitopes for both CD4$^+$ and CD8$^+$ T cell epitopes which are common among three subtypes i.e. H5N1, H3N2 and H1N1 were docked with MHC Class I and II molecules. The binding energies of the epitopes capable of binding to MHC molecule were compared with the naturally bound peptide (Fig 7A and 7B). The average binding energy of MHC class II epitopes (-5.9 Kcal/mol) was found to be higher than the naturally bound peptides (-5.2 Kcal/mol) . Thus, it can be concluded that the predicted epitopes had similar binding affinity towards the MHC Molecules as that of the naturally bound peptide.

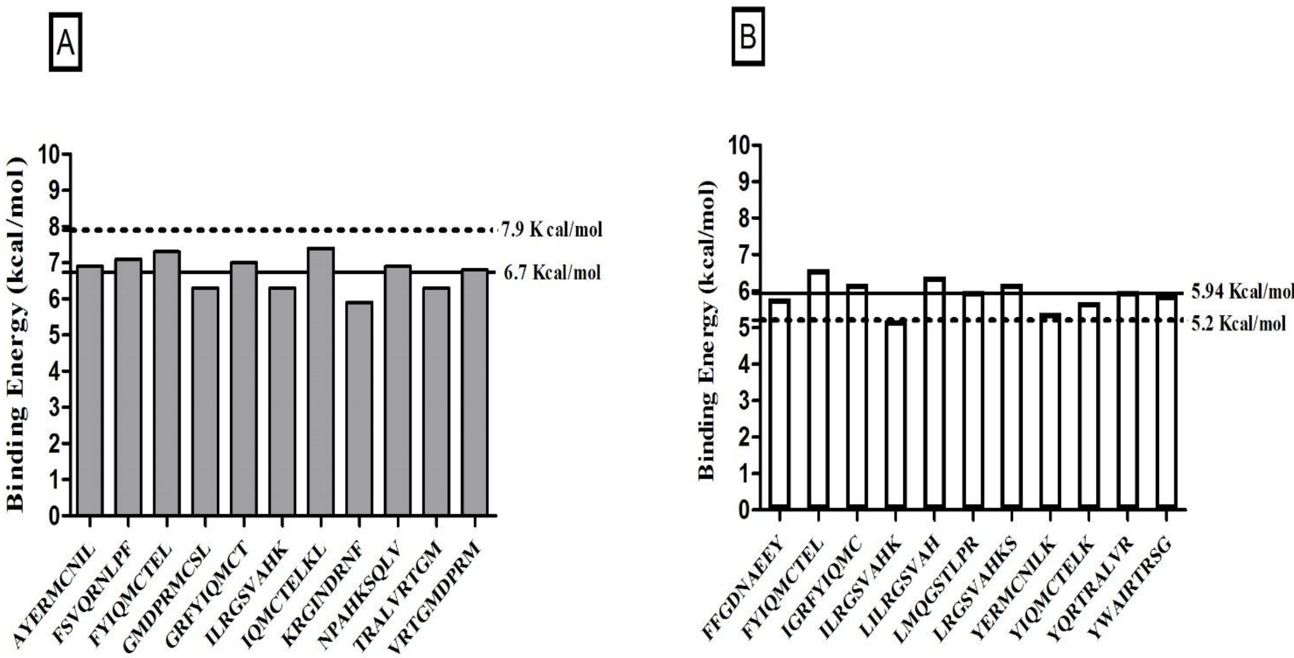



**Figure 8: Binding Energy* in Kcal/mol (A) MHC class I binding epitopes (B) MHC class II binding epitopes.**

*Dotted line represents the binding energy of naturally bound peptide to the respective MHC molecule. Solid line represent the average binding energy of the epitopes to the given MHC. Binding energy is negative**.**

# Chapter 6: Discussion

Immunoinformatics based prediction of peptides containing T-cell epitopes can significantly boost the selection of targets for vaccine design. The conventional technique of experimentally identification immunogenic peptides is a laborious and uneconomical process. Hence, *in silico* screening of immunogenic peptides may be a practical alternative for it. Earlier studies by Torjan et al. (2003) and Assarsson et al. (2008) indicated that epitopes predicted by various bioinformatics tools have proven to be immunogenic in *in vitro* and *in vivo* experiments. Due to error prone replication, influenza A virus undergoes frequent mutations, thus ruling out the potency of predefined epitopes against all the variants.  Also, cell mediated immune response against viral peptides which are likely to be shared between different virus strains, may impart immunity against various strains and subtypes of influenza virus. Hence, we focused on conserved peptide sequences to predict the epitopes.

H5N1 and H3N2 are the subtypes of influenza A virus infecting humans, others being H1N1, H2N2 and H7N9 (Liu et al., 2013). Nucleoprotein in an internal influenza protein and is found in association with vRNA in the form of ribonucleoprotein complex. Since, internal viral proteins often display, low rate of mutation as compared to the surface proteins, therefore nucleoprotein appears to be a good candidate for vaccine development. Current study considered, all the available sequences of nucleoprotein from H5N1 and H3N2 till January, 2015. There have only been a few studies undertaken on epitope prediction in the H5N1 and H3N2 virus based on single (Gustiananda, 2011) or multiple epitope prediction algorithms (Ichihashi et al., 2011). But performance of the consensus prediction approach of epitope prediction has been reported to be superior to single predictive strategy. Six different algorithms in our study was to ensure that immunogenic potential of the predicted epitopes is analyzed based on the MHC-peptide bonding and other factors determining the immune response. Further the selection of common epitopes enriched peptide fragments for both class I and II MHC may give rise to overall T-cell specific immune response.

Owing to MHC polymorphism, its distribution varies among different ethnic groups around the world. Genetic polymorphisms of MHC significantly influence the variation in viral vaccine response in the human population. Hence, it is important to consider peptide enriched epitopes, which could be effective in evoking an immune response in large populations. Ten immunogenic peptides of nucleoprotein (NP) in H5N1 and twelve that of H3N2, contains both $CD8^+$ and $CD4^+$ T cell epitopes were found ranging in length from 14 to 59 and 11 to 43 amino acid residues respectively. Further, 28 epitopes were identified to be common among H5N1 and H3N2 and also 44 and 22 epitopes were predicted to be unique among H5N1 and H3N1 nucleoprotein, respectively. These peptides can capable of inducing immune response in population distributed world, irrespective of their HLA type. Population coverage analysis has also indicated that majority of populations distributed around the world are expected to respond to these epitope containing peptides.

Further to analyze MHC-peptide binding, molecular docking approach was used, by taking MHC class I (HLA-A2) and MHC class II (HLA-DR3). These HLA alleles are expressed in high frequency in human population. Docking results have indicated that the epitopes identified to be common among H1N1, H5N1 and H3N2 binds to these MHCs with relative similar binding energy as compared to the naturally bound peptides. Hence, these epitopes can be strongly presented by MHC molecules present on antigen presenting cells to induce T cell response.

Putative immunogenic peptides of NP identified in the current study are thus good candidates for vaccine design against H5N1 and H3N2. These peptides can be used for further study to assess the potential of these peptides for immunogenic response in the PBMC by T-cell proliferation assay and cytokine production assay.

# Chapter 7: Summary

Influenza is one 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 polymerase 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, form 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 *et al*., 2004). Over the past 150 years, atleast four pandemics of influenza occurred at irregular intervals, including three in the 20$^{th}$ century. These have caused high attack rates in all susceptible age groups, with high morbidity and mortality. The most lethal influenza A 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 vaccines, 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.

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 variations. Therefore, 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 for vaccine development, as it is internal protein and it shows high level of conservancy as compared to the surface glycoprotein.

Our approach was to find out a stretch of immunogenic peptide from conserved sequences of (NP) nucleoprotein of H5N1 and H3N2 using immunoinformatics tools. Twelve conserved regions were found in H5N1 and 19 were found in H3N2. In these conserved regions of H5N1, 16 MHC I specific and 18 MHC II specific immunogenic peptides were found. In case of H3N2, 16 MHC I specific and 15 MHC II specific immunogenic peptides were found. Population coverage analysis was done. 91.28% and 96.94% of the individuals are expected to respond to the predicted MHC class I and II nucleoprotein (H5N1) peptides, respectively. Similarly, 89.28% and 96.94% of the individuals are expected to respond to the predicted MHC class I and II nucleoprotein (H3N2) peptides, respectively. Molecular docking is an approach to find out the binding affinity of predicted peptides to their respective MHC molecules. Docking was done using MHC class I (HLA A2) and MHC class II (HLA DR3). 11 epitopes of MHC class I and 11 epitopes of MHC class II (commonly predicted for H1N1, H3N2 and H5N1) were docked to class I and Class II HLA respectively and binding energy was found to be comparable to their native peptide.

Further study can be carried out for these immunogenic peptides to assess their immunogenic response by T-cell proliferation assay and cytokine production assay.

# Chapter 8: References


1. Adar Y, Singer Y, Levi R, Tzehoval E, Perk S, Banet-Noach C, Nagar S, Arnon R, Ben-Yedidia T. A universal epitope-based influenza vaccine and its efficacy against H5N1. *Vaccine* (2009); 27(15): 2099-2107.
2. Altschul SF, Gish W, Miller W, Myers EW, and Lipman D. Basic local alignment search tool. *J Mol Biol* (1990); 215: 403–10.
3. Assarsson E, Bui HH, Sidney J, et al. Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans. *J Virol* (2008); 82: 12241–12251.
4. Barik S. New treatments for influenza. *BMC Med* (2012); 10: 104.
5. Bean W.J, Schell M, Katz J, Kawaoka Y, Naeve C, Gorman O, *et al.* Evolution of the H3 influenza virus hemagglutinin from human and nonhuman hosts. *J Virol* (1992); 66: 1129–1138.
6. Berthoud TK, Hamill M, Lillie PJ, Hwenda L, Collins KA, Ewer KJ, Milicic A, Poyntz HC, Lambe T, Fletcher HA, Hill AV, Gilbert SC. Potent CD8$^{+}$ T-cell immunogenicity in humans of a novel heterosubtypic influenza A vaccine, MVA-NP$^{+}$M1. *Clin Infect Dis* (2011); 52(1):1–7.
7. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope based diagnostics and vaccines. *BMC Bioinformatics* (2006); 7: 153.
8. Burg SH, Visseren MJ, Brandt RM, Kast WM, Melief CJ. Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *J Immunol* (1996); 156(9): 3308-14.
9. Carbo A, Hontecillas R, Andrew T, Eden K, Mei Y, Hoops S and BassaganyaRiera J. Computational modeling of heterogeneity and function of CD4$^{+}$ T cells. Front *Cell Dev Biol* (2014); 2: 31.
10. CDC/2014: http://www.cdc.gov/flu/swineflu/h3n2v-case-count.htm
11. Centers for Disease Control and Prevention (CDC). High levels of adamantane resistance among influenza A (H3N2) viruses and interim guidelines for use of antiviral agents--

United States, 2005-06 influenza season. *MMWR Morb Mortal Wkly Rep* (2006); 55(2): 44-6.

12. Chang S, Sun D, Liang H, Wang J, Li J, Guo L, Wang X, Guan C, Boruah B, Yuan L, Feng F, Yang M, Wang L, Wang Y, Wojdyla J, Li L, Wang J, Wang M, Cheng G, Wang H, Liu Y. Cryo-EM Structure of Influenza Virus RNA Polymerase Complex at 4.3 Å Resolution. *Molecular cell* (2015); 57(5): 925-935.
13. Das K1, Aramini JM, Ma LC, Krug RM, Arnold E. Structures of influenza A proteins and insights into antiviral drug targets. *Nat Struct Mol Biol* (2010);17(5):530-8..
14. De Clercq E. Antiviral agents active against influenza A viruses. *Nat. Rev. Drug Discov* (2006); 5: 1015–1025.
15. De Groot A.S, Einck L, Moise L, Chambers M, Ballantyne J, Malone R.W, Ardito M and Martin W. Making vaccines "on demand" A potential solution for emerging pathogens and biodefense? *Hum Vaccin Immunother* (2013); 9(9): 1877–1884.
16. De Jong MD, Hien TT. Avian influenza A (H5N1). *J Clin Virol* (2006); 35(1): 2-13.
17. Desai DV, Kulkarni-Kale U. T-cell epitope prediction methods: an overview. *Methods Mol Biol* (2014); 1184: 333-64.
18. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* (2004); 32(5):1792-1797.
19. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ. The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* (2004); 303(5665):1838-42.
20. Gustiananda M. Immunoinformatics analysis of H5N1 proteome for designing an epitope derived vaccine and predicting the prevalence of preexisting cellular mediated immunity toward bird flu virus in Indonesian population. *Immunome Res* (2011); 7 (3): 111.
21. Hou Y, Guo Y, Wu C, Shen N, Jiang Y, Wang J. Prediction and Identification of T Cell Epitopes in the H5N1 Influenza Virus nucleoprotein in Chicken. *PLoS One*. (2012); 7(6): e39344.
22. Ichihashi T, Yoshida R, Sugimoto C, Takada A, and Kajino K. CrossProtective Peptide Vaccine against Influenza A Viruses Developed in HLAA* 2402 Human Immunity Model. *PLoS One*(2011); 6(9): e24626.

23. Ison MG, Gnann JW Jr, NagyAgren S, Treannor J, Paya C, Steigbigel R, Elliott M, Weiss HL, Hayden FG. Safety and efficacy of nebulized zanamivir in hospitalized patients with serious influenza. *Antivir Ther* (2003); 8: 183–190.
24. Iwasaki A and Pillai P.S. Innate immunity to influenza virus infection. *Nat Rev Immunol* (2014); 14(5): 315–328.
25. Katz JM, Webster RG. Efficacy of inactivated influenza A virus (H3N2) vaccines grown in mammalian cells or embryonated eggs. *J Infect Dis* (1989); 160:191–8.
26. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Neilsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* (2007); 8: 424.
27. 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.
28. Lee Y.T, Kim K, Ko E, Lee Y, Kim M, Kwon Y.M, Tang Y, Cho M, Lee Y.J, and Kang S.M. New vaccines against influenza virus. *Clin Exp Vaccine Res* (2014); 3(1): 12–28.
29. Liu Q1, Liu DY, Yang ZQ. Characteristics of human infection with avian influenza viruses and development of new antiviral agents. *Acta Pharmacol Sin* (2013); 34(10): 1257-69.
30. Martín-Benito J, Area E, Ortega J, Llorca O, Valpuesta J, Carrascosa J and Ortín J. Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep* (2001); 2(4): 313–317.
31. Metersky ML, Masterton RG, Lode H, File TM Jr, Babinchak T. Epidemiology, microbiology, and treatment considerations for bacterial pneumonia complicating influenza. *Int J Infect Dis* (2012); 16: 321–331.
32. Miotto O, Heiny A, Tan TW, August JT, and Brusic V. Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC Bioinformatics* (2008); 9(1):18.
33. Moise L, Tassone R, Latimer H, Terry F, Levitz L, Haran J.P, Ross T.M, Boyle C, Martin W.D, and De Groot A.S. Immunization with crossconserved H1N1 influenza CD4 Tcell epitopes lowers viral burden in HLA DR3 transgenic mice. *Hum Vaccin Immunother* (2013); 9(10): 2060–2068.

34. Monto A.S. Vaccines and antiviral drugs in pandemic preparedness. *Emerg. Infect. Dis* (2006); 12: 55–60.
35. Moscona A. Neuraminidase inhibitors for influenza. *N. Engl. J. Med* (2005); 353: 1363–1373.
36. Mulyanto C and Saleh R. Prediction of a neutralizing epitope of a H5N1 virus hemagglutinin complexed with an antibody variable fragment using molecular dynamics simulation. *JBPC* (2011); 2(3).
37. Neilsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* (2009); 10: 296.
38. Neilsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* (2007); 8: 238.
39. Neumann G, Noda T, and Kawaoka Y. Emergence and pandemic potential of swine origin H1N1 influenza virus. *Nature* (2009); 459(7249): 931–939.
40. Ng K, Zhang H, Tan K, Li Z, Liu J, Chan P, Li S, Chan W, Au S.W, Joachimiak A, Walz T, Wang J, and Shaw P. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J* (2008); 22(10): 3638–3647.
41. Ngai K, Chan M and Chan P. Replication and Transcription Activities of Ribonucleoprotein Complexes Reconstituted from Avian H5N1, H1N1pdm09 and H3N2 Influenza A Viruses. *PLoS One* (2013); 8(6): e65038.
42. Noda T and Kawaoka Y. Structure of influenza virus ribonucleoprotein complexes and their packaging into virions. *Rev. Med. Virol* (2010); 20: 380–391.
43. Ortigoza M.B, Dibben O, Maamary J, MartinezGil L, LeyvaGrado V.H, Abreu, Jr., Ayllon J, Palese P, and Shaw M.L. A Novel Small Molecule Inhibitor of Influenza A Viruses that Targets Polymerase Function and Indirectly Induces Interferon. *PLoS Pathog* (2012); 8(4): e1002668.
44. Parida R, Shaila M.S, Mukherjee S, Chandra N.R, Nayak R. Computational analysis of proteome of H5N1 avian influenza virus to define T cell epitopes with vaccine potential. *Vaccine* (2007); 25: 7530–7539.

45. Parker KC, Bednarek MA, and Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *Immunol* (1994); 152: 163.
46. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biol* (2013); 3(1): 120139.
47. Pedersen L, Breum S, Riber U, Larsen L and Jungersen G. Identification of swine influenza virus epitopes and analysis of multiple specificities expressed by cytotoxic T cell subsets. *Virology* (2014); 11:163.
48. Rammensee H.G, Bachmann J, Emmerich N, Bachor O. A, Stevanović S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* (1999); 50(3): 213-219.
49. Salomon R, Webster RG. The influenza virus enigma. Cell. 2009 Feb 6; 136(3):402-10
50. Samji T. Influenza A: Understanding the Viral Life Cycle. *Yale J Biol Med* (2009); 82(4): 153–159.
51. Schanen B, De Groot, Moise L, Ardito M, McClaine E, Martin W, Wittman V, Warren W, Drake III D. Coupling sensitive *in vitro* and *in silico* techniques to assess cross-reactive $CD4^+$ T cells against the swine-origin H1N1 influenza virus. *Vaccine* (2011); 29(17): 299-309.
52. Sette A, Fikes J. Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. *Immunology* (2003); 15(4): 461-470.
53. Shi Y, Wu Y, Zhang W, Qi J and Gao G. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nature Reviews Microbiology* (2014); 12: 822-831.
54. Shu L. L., Bean W. J. and Webster R. G. Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *Virology* (1993); 67: 2723-2729.
55. Singh H and Raghava G. Propred: Prediction of HLA-DR binding sites. *Bioinformatics* (2001); 17(12): 1236-37.
56. Soema PC, van Riet E, Kersten G and Amorij JP. Development of cross-protective influenza a vaccines based on cellular responses. *Front Immunol* (2015); 6: 237.

57. Somvanshi P, Singh V and Seth P.K. Prediction of Epitopes in Hemagglutinin and Neuraminidase Proteins of Influenza A Virus H5N1 Strain: A Clue for Diagnostic and Vaccine Development. *OMICS A Journal of Integrative Biology* (2008); 12.

58. Su CT, Schonbach C and Kwoh C. Molecular docking analysis of 2009-H1N1 and 2004-H5N1 influenza virus HLA-B*4405-restricted HA epitope candidates: implications for TCR cross-recognition and vaccine development. *BMC Bioinformatics* (2013); 2: 21.

59. Sun Y, Liu J, Yang M, Gao F, Zhou J, Kitamura Y, Gao B, Tien P, Shu Y, Iwamoto A, Chen Z and Gao G. Identification and structural definition of H5 specific CTL epitopes restricted by HLAA*0201 derived from the H5N1 subtype of influenza A viruses. *J Gen Virol* (2010); 91(4): 919–930.

60. Tan P.T, Heiny A.T, Miotto O, Salmon J, Marques Jr., Lemonnier F, and August J. Conservation and Diversity of Influenza A H1N1 HLA Restricted T Cell Epitope Candidates for Epitope Based Vaccines. *PLoS One* (2010); 5(1): e8754.

61. Tao YJ1, Zheng W. Biochemistry. Visualizing the influenza genome. *Science* (2012); 338(6114): 1545-6.

62. Taubenberger J.K and Morens D.M. The Pathology of Influenza Virus Infections. *Annu Rev Pathol* (2008); 3: 499–522.

63. Thevenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tuffery P. *Nucleic Acids Research* (2012); 40:288–93.

64. Tomar N and De R.K. Immunoinformatics: an integrated scenario. *Immunology* (2010); 131(2): 153–168.

65. Tong J, Ren E. Immunoinformatics: Current trends and future directions. *Drug Discovery* (2009); 14(13): 684-689.

66. Treanor J. Influenza vaccine--outmaneuvering antigenic shift and drift. *N Engl J Med* (2004); 350(3):218-20.

67. Trojan A, Urosevic M, Hummerjohann J, Giger R, Schanz U, and Stahel RA. Immune reactivity against a novel A3-restricted influenza virus peptide identified by predictive algorithms and interferon-gamma quantitative PCR. *J Immunother* (2003); 26:41–46.

68. Trott O, Olson AJ. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Computational Chemistry* (2010); 31:455–61.

69. Van de Sandt CE1, Kreijtz JH, Rimmelzwaan GF. Evasion of influenza A viruses from innate and adaptive immune responses. *Viruses* (2012); 4(9): 1438-76.

70. Webby R.J, Lipatov A.S, Govorkova E.A, Ozaki H, Peiris M, Guan Y, Poon L and Webster R.G. Influenza: Emergence and Control. *J Virol* (2004); 78(17): 8951–8959.

71. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y (1992) Evolution and ecology of influenza A viruses. *Microbiol Rev* (1992); 56: 152–79.

72. WHO/2015:http://www.who.int/influenza/human_animal_interface/EN_GIP_20150623cumulativeNumberH5N1cases.pdf?ua=1

73. Wu Y1, Wu Y2, Tefsen B3, Shi Y4, Gao GF5. Bat-derived influenza-like viruses H17N10 and H18N11. *Trends Microbiol* (2014);22(4):183-91.

74. Yang S, Niu S, Guo Z, Yuan Y, Xue K, Liu S, and Jin H. Cross-protective immunity against influenza A/H1N1 virus challenge in mice immunized with recombinant vaccine expressing HA gene of influenza A/H5N1 virus. *Virol J* (2013); 10: 291.

75. Yedidia T.B and Arnon R. Epitope based vaccine against influenza. *Expert review of vaccines* (2007), 6(6): 939-948.

76. Zhang N, Zheng B, Lu L, Zhou Y, Jiang S, Du L. Advancements in the development of subunit influenza vaccines. *Microbes and Infection* (2015); 17(2): 123-134.

77. Zheng W, Tao Y.J. Structure and assembly of the influenza A virus ribonucleoprotein complex. *FEBS Letters* (2013); 587: 1206–1214.