IDENTIFICATION OF IMMUNOGENIC PEPTIDE IN NUCLEOPROTEIN (NP) OF H1N1 AND H3N2

A Thesis submitted in partial fulfillment of the requirements for the award of the degree of

MASTER OF SCIENCE
IN
BIOTECHNOLOGY

Submitted By
ABHISHEK SINGH RATHORE
Registration No. 301001001

Under the Supervision of
DR. MANOJ BARANWAL

DEPARTMENT OF BIOTECHNOLOGY & ENVIRONMENTAL SCIENCES
THAPAR UNIVERSITY, PATIALA
JULY, 2012

Certificate

This is to certify that the project entitled "**Identification of Immunogenic peptide in Nucleoprotein (NP) of H1N1 and H3N2**" being submitted by Abhishek Singh Rathore, Roll No.301001001 in partial fulfillment of the requirements for the award of degree of Master of Science in Biotechnology in the Department of Biotechnology and Environmental Sciences, Thapar University, Patiala, is a bonafide work carried out under my supervision and guidance. The report has not been submitted for the award of any other degree or certificate in this or any other university.

Dr. Manoj Baranwal
Supervisor
DBTES
Thapar University, Patiala

Dr. M. S. Reddy
Head of Department
DBTES
Thapar University, Patiala

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

Candidate's Declaration

I hereby declare that the work which is being presented in the dissertation entitled "**Identification of Immunogenic peptide in Nucleoprotein (NP) of H1N1 and H3N2**" in partial fulfillment of the requirements for the award of Master in Science in Biotechnology, Department of Biotechnology and Environmental Sciences, Thapar University, Patiala is an authentic record of my own work during a period of six months from January 2012 to June 2012, under the supervision of Dr. Manoj Baranwal, Assistant Professor, Department of Biotechnology and Environmental Sciences, Thapar University, Patiala. The report has not been submitted for the award of any other degree or certificate in this or any other university.

Place: Patiala
Date: 16. 7. 2012

Abhishek Singh

Abhishek Singh Rathore
Roll No. 301001001

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

Dr. Manoj Baranwal
Supervisor
DBTES
Thapar University, Patiala

Dr. M. S. Reddy
Head of Department
DBTES
Thapar University, Patiala

ACKNOWLEDGEMENT

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

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

I feel deep gratitude towards Dr. M.S. Reddy, Head of Department, Department of Biotechnology and Environmental Sciences for his immense concern throughout the project work.

I am thankful to Ms. Neha Lohia, Research scholar for her favour and concern without which it would have been difficult for me to achieve my goals.

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

My special thanks to my friends Rameez, Ranjan, Amit, Ajay, Charu, Rabia, Himangi, Tanya, Sakshi and Shipra ,who have always been there for me through all the tough times to all good times.

I don’t have words to express my thanks to my parents & my sister for their selfless sacrifice, encouragement and heart full blessings that continue to enlighten my life.
In the end, I am thankful to the Almighty for blessing me to complete this work successfully.

Date: Abhishek Singh Rathore

ABSTRACT

Influenza is one of the most infectious diseases confronting the world today; however no effective prevention against influenza has been developed because of the antigenic variation in influenza virus. Epitope based vaccines imparts an efficient strategy for protection against antigenic variations of influenza viruses. Conserved regions of Nucleoprotein (NP) of H1N1 and H3N2 strains of influenza virus A having a conservancy of more than 80% were detected by computational tool based on informational entropy. Epitopes for HLA alleles of MHC class I and class II from the conserved regions were predicted using immunoinformatics tools. Immunogenic peptide were identified and selected by predicted overlapping epitopes. These immunogenic peptides were looked for common immunogenic peptide for both MHC class I and class II. Four immunogenic peptide of H1N1 NP: (i) EIRASVG RMIGGIGRFYIQMCTELKL (26 amino acid), (ii) TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAA (36 amino acid), (iii) AYERMCNILKGKFQTA (16 amino acid), (iv) FLARSALILRGSVAHKSCLPAC VY (24 amino acid) and immunogenic peptide regions of H3N2 NP: (i) KMIDGIGRFYIQMC TELKL (19 amino acid), (ii) TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKGI GTMVMELI (48 amino acid), (iii) NAEIEDLIFLARSALILRGSVAHKS (25 amino acid), (iv) NPAHKSQLVWMACHSAAFEDLRLLSFIRGTKVSPRGKLSTRGVQIASNENMDNMGSST LELRSGYWAIRTRSG (72 amino acid), (v) IIRMMEGAKPEEVSFRGRGVFELSDEKATPNI VPSF(36 amino acid) were found to be interesting as long stretch of immunogenic peptides were found ranging in length from 16 to 72 amino acid residues.

ABBREVIATIONS

RNA	-	Ribonucleic acid
WHO	-	World Health Organization
HLA	-	Human Leukocyte Antigen
MHC	-	Major Histocompatibility Complex
HA	-	Hemagglutinin
NA	-	Neuraminidase
NP	-	Nucleoprotein
M1	-	Matrix protein 1
M2	-	Matrix protein 2
RNP	-	Ribonucleoprotein
NEP	-	Nuclear export protein
NS1	-	Non- Structural protein
IFN	-	Interferon
PA, PB	-	Polymerase protein A, B
HEF	-	Hemagglutinin- esterase-fusion
APC	-	Antigen presenting cells
PBMC	-	Peripheral Blood Mononuclear Cells
MUSCLE	-	Multiple Sequence Comparison By Log Expectation
AVANA	-	Antigen Variability Analyzer
NCBI	-	National Council of Biological Information
MSA	-	Multiple Sequence Alignment
CD	-	Cluster of Differentiation
IEDB	-	Immune Epitope Database
CTL	-	Cytotoxic T- Lymphocytes
BIMAS	-	Bioinformatics and Molecular Analysis Section
ANN	-	Artificial Neural Network
SMM	-	Stabilized Matrix Method
HMM	-	Hidden Markov Model
TAP	-	Transporter Associated with antigen Processing

SVM	-	Support Vector Machine
PFR	-	Peptide Flanking Residue
DMEM	-	Dulbecco's Modified Eagle Medium
PBS	-	Phosphate Buffer Saline
EDTA	-	Ethylenediaminetetraaceticacid
MTT	-	3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide
ELISA	-	Enzyme-linked immunosorbent assay
$T_{(½)}$	-	Half Life
IC_{50}	-	Half of the minimum inhibitory concentration

Table of Contents

INTRODUCTION

Influenza is a major cause of sickness and death around the world and is one of the most infectious diseases confronting the world today. Influenza virus is RNA virus which belongs to *Orthomyxoviridae* family. Influenza pandemics occur when a new influenza virus appears, against which the human population has little or no immunity. While global influenza pandemics have occurred only a few times in the past century, the H1N1 pandemic of 1918–1919 caused 20–50 million deaths and was one of the most serious disease outbreaks in recorded history (Pattreson K.D. and Pyle G.F., 1991).

The 2009 swine flu is a pandemic, and the second one to involve H1N1 influenza virus (the first of them was the 1918 flu pandemic). The virus appeared in April 2009 was a new strain of H1N1 which resulted due to triple reassortment of bird, swine and human flu viruses further combined with a Eurasian pig flu virus, leading to the term "swine flu" to be used for this pandemic (Vladimir T. *et al.*, 2009).

Antigenic variation is the evolutionary mechanism by which viruses evade host immune system. Influenza viruses accomplish this through minor, incremental genetic changes (i.e., antigenic drift), as well as through major genetic alterations (i.e., antigenic shift). Antigenic drift specifically refers to frequent, discrete mutations that occur within the genes (e.g., the hemagglutinin or neuraminidase genes) of a given influenza subtype, leading to new strains that escape host immune surveillance.

Treatments for influenza include a range of medications and therapies that are used in response to influenza. The two main classes of antiviral drugs used against influenza are neuraminidase inhibitors, such as Zanamivir and Oseltamivir, or inhibitors of the viral M2 protein, such as Amantadine and Rimantadine (Nicole S.M. *et al.*, 2006). These drugs are helpful only when they are taken in the early stages of infection. However, viral strains have emerged that show drug resistance to both classes.

Vaccine is considered to be one of the most effective way to control influenza virus. Currently vaccines used against influenza are inactivated influenza vaccine (IIV) and live attenuated influenza vaccine (LAIV), used in monovalent as well as trivalent forms. But currently available vaccines induce immunity against a specific and closely related antigenic viral strain, but do not protect against novel Influenza A viruses which emerged as a result of antigenic variation. So

there is a need of a vaccine which can induce immunity against a broad range of antigenic viral strains and would not need to be updated every year (Stanekova Z. and Vareckova E., 2010).

Epitope based vaccine is a novel approach that can be utilised for the development of a broad range influenza vaccine. Epitope based vaccines are those in which small peptides containing epitopes derived from target proteins are used to provoke an immune response. This will be effective against different serotypes of influenza. Consequently, it is expected to protect against future strains as well.

T-cell epitope identification and selection is the most crucial step in epitope based vaccine development. T cells can recognize antigen only when it is presented by a group of specialized proteins known as HLA or MHC, class I and class II. Experimental methods of epitope prediction are expensive and time consuming hence *in silico* approaches are used for epitope prediction.

Cytotoxic T lymphocytes (CTLs) play an important role in the control of virus infections, including those caused by influenza viruses. Targets for CTL responses, such as the virus nucleoprotein (NP) and matrix protein, are considered as candidate vaccines because of their conserved nature. The use of conserved proteins could provide protective immunity against drift variants or viruses with novel subtypes (Townsend A.R.M. *et al.*, 1986).

Present study is carried out to predict immunogenic peptide containing overlapping epitope of Nucleoprotein (NP) in H1N1 and H3N2 strains of Influenza Virus A which can be further be used as a target for vaccine development. To develop a vaccine which will be effective against a broad range of influenza strains, the first necessary step is to determine the conserved regions of the protein from different Influenza strains. Next step is the identification of potential epitopes recognized by MHC class I and class II using various immunoinformatics tools. Overlapping epitopes are used to predict immunogenic peptides. Final step is to determine immunogenicity of predicted immunogenic peptide using cell proliferation assay of PBMC (Peripheral Blood Mononuclear Cell).

REVIEW OF LITERATURE

2.1 Influenza Virus

Influenza is a viral infection caused by segmented negative sense single stranded RNA viruses of the family Orthomyxoviridae that affects mainly the nose, throat, bronchi and, occasionally, lungs. Infection usually lasts for about a week, and is characterized by sudden onset of high fever, aching muscles, headache and severe malaise, non-productive cough, sore throat and rhinitis (http://www.who.int/topics/ influenza/en/). Influenza is transmitted through the air by coughs or sneezes, creating aerosols containing the virus. Error prone RNA-dependent RNA polymerases and segmented genome of influenza viruses allows virus to undergo minor (antigenic drift) as well as major (antigenic shift) antigenic changes which permit the virus to evade adaptive immune response in a variety of mammalian and avian species (Stanekova Z. and Vareckova E., 2010).

2.1.1 Taxonomy of Influenza Virus

Influenza is a part of Group V of RNA viruses. Orthomyxoviridae is a family of Group V of RNA virus that includes five generas Influenzavirus A, Influenzavirus B, Influenzavirus C, Isavirus and Thogotovirus. Influenza viruses make up three of the five genera of the family Orthomyxoviridae.

TABLE 1: Taxonomy of Influenza Virus

Group	Family	Genus	Species	Subtype or Serotype
Group V (-)ssRNA	Orthomyxoviridae	Influenzavirus A	Influenza A Virus	H1N1, H1N2, H2N2, H3N2, H5N1, H7N1,H9N2, H10N7
		Influenzavirus B	Influenza B Virus	
		Influenzavirus C	Influenza C Virus	

Influenza Type A

The type A viruses are the most virulent pathogens among the three influenza types and cause the most severe disease. Influenza A shows both antigenic drift and antigenic shift. Influenza type A viruses are divided into subtypes based on two surface proteins hemagglutinin (HA) and neuraminidase (NA). There are 17 different HA subtypes and 9 different NA subtypes. Many different combinations of HA and NA proteins are possible. Only some influenza A subtypes (i.e., H1N1, H1N2, and H3N2) are currently circulating among people.

TABLE 2: Serotypes of Influenza A involved in humans Epidemics

Serotypes of Influenza A	Epidemics
H1N1	Spanish Flu in 1918 Swine Flu in 2009
H2N2	Asian Flu in 1957
H3N2	Hong Kong Flu in 1968
H5N1	Bird Flu in 2004
H1N2	Endemic in humans, pigs and birds

Influenza Type B

Unlike influenza A viruses, these viruses are not classified according to subtype. Although influenza type B viruses can cause human epidemics, they do not cause pandemics. This type of influenza mutates at a rate 2–3 times slower than type A and consequently is less genetically diverse. The reduced rate of antigenic change, combined with its limited host range (inhibiting cross species antigenic shift), ensures that pandemics of influenza B do not occur.

Influenza Type C

Influenza type C viruses cause mild illness in humans and do not cause epidemics or pandemics. These viruses are not classified according to subtype. The major influenza C virus envelope glycoprotein is called HEF (hemagglutinin-esterase-fusion) because it has the functions of both the HA and the NA. Therefore the influenza virion contains 7 RNA segments, not 8 RNAs like influenza A and B viruses (www.virology.ws/2009/09/22/ the-a-b-and-c-of-influenza-virus/).

TABLE 3: Comparison of Different Types of Influenza Virus

Features	Influenza Type A	Influenza Type B	Influenza Type C
Membrane Protein	HA, NA, M1	HA, NA, NB, BM2	HEF, CM2
RNA Segment	8	8	7
Ion Channel	M2	BM2	CM2
Antigenic Variation	Antigenic shift and antigenic drift	Antigenic drift	-
Epidemiology	Pandemic	Epidemic	Endemic
Host	Human, pig, bird, horse, seals, whales	Human, Seal	Human, pig

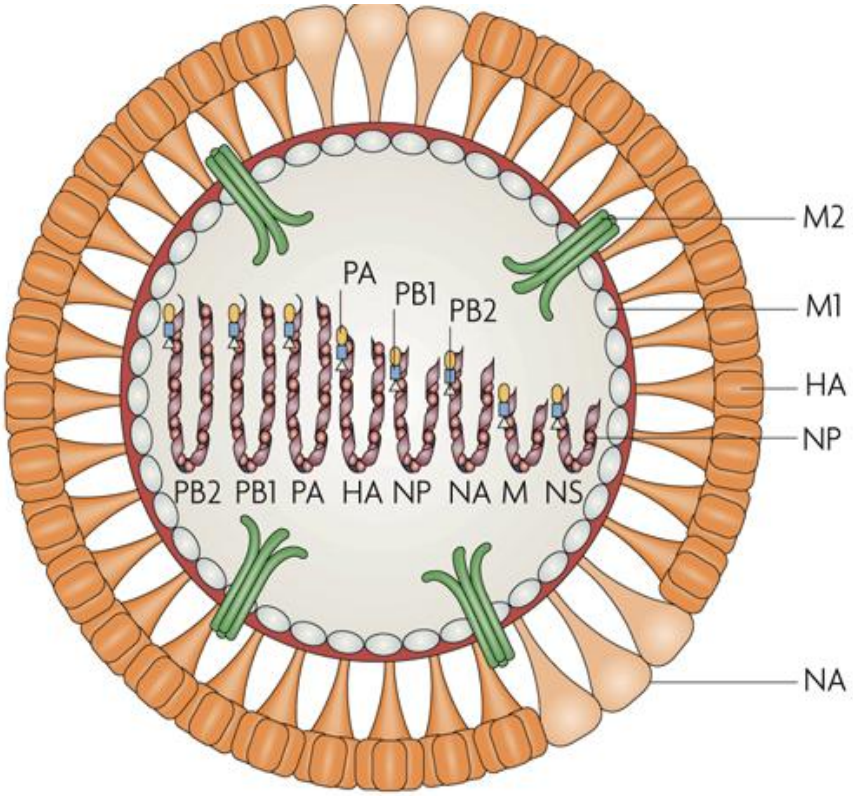
2.1.2 Structure of Influenza Virus

The virus particle is 80–120 nanometers in diameter and usually roughly spherical. The viral envelope is made of lipid bilayer which contains two main types of glycoproteins hemagglutanin (HA) and nuraminidase (NA), wrapped around a central core. The central core contains the viral RNA genome and viral proteins that package and protect this RNA. Influenza virus genome is segmented single stranded negative-sense RNA.

It is an enveloped virus i.e. the outer layer is lipid membrane which is taken from the host cell in which the virus multiplies. Inserted in the lipid membrane are glycoproteins (HA and NA). These are the proteins that determine the subtypes of influenza virus.

Influenza viruses are subject to high rates of mutation. Because they lack proof reading enzymes that maintain the fidelity of DNA replication. They show high frequency gene reassortment during mixed infections, during replication of their single stranded segmented genome.

The Influenza A viruses have eight segments that encode for the 11 viral genes: hemagglutinin (HA), neuraminidase (NA), matrix 1 (M1), matrix 2 (M2), nucleoprotein (NP), non-structural protein 1 (NSP1), non-structural protein 2 (NS2; also known as nuclear export protein, NEP), polymerase acidic protein (PA), polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2) and polymerase basic protein 1 – F2 (PB1-F2) (Samji T. *et al.*,2009).



Structure of Influenza Virus A

The influenza A virus genomes consist of 8 separate segments and each segment codes for a functionally important protein:

1) PB 2 – It is a viral polymerase component involved in synthesis of capped messenger RNA by endonuclease activity which cleaves host cell messenger RNA.
2) PB 1 – It is a viral polymerase component with RNA transcription and replication activities.
3) PA - It is a viral polymerase compound involved in RNA replication.
4) HA – It is surface glycoprotein and major antigenic determinant. HA attachment is required for infection, it attaches to cell receptors. It plays an important role in release of viral RNA into cells by causing fusion of viral and cellular membrane.
5) NA – It is major nucleocapsid structural complex and type specific antigen. It facilitates release of viruses. In the viral infection sialic acid is required as receptor. NA removes

sialic acid from cell surface of the infected cell. So progeny virus will not stick to the cell. They will therefore be able to diffuse away until they meet an uninfected cell. Also at the time of release of virus, the HA bound to sialic residue and NA is required to hydrolyze the glycosidic linkages between sialic acid residue and HA.

6) NP – It forms a complex with the viral RNA genome and packages the RNA into a helical ribonucleoprotein core.
7) M - Membrane matrix protein is type specific antigen. M1 constructs the matrix and M2 acts as an ion channel pump to lower or maintain the pH of the endosome.
8) NS –NS is non structural protein. Its unique post translational regulator inhibits the nuclear transport of poly(A) containing mRNAs and inhibit pre mRNAs splicing by binding to a specific region of small nuclear RNA.

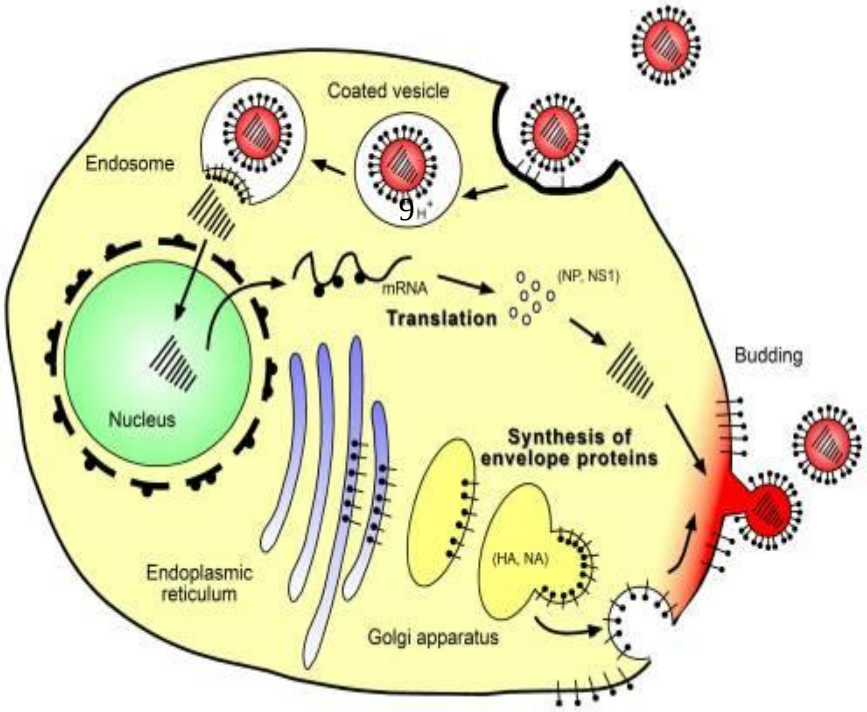
2.1.3 Life Cycle of Influenza

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

1. Entry of virus particle into the host cell.
2. Entry of viral Ribonucleoprotein (RNP) into the nucleus.
3. Transcription and replication of the viral genome.
4. Export of the vRNPs from the nucleus.
5. Assembly and budding at the host cell plasma membrane.

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

proteins have known nuclear localization signals (NLSs) that can bind to the cellular nuclear import machinery and, thus, enter the nucleus.



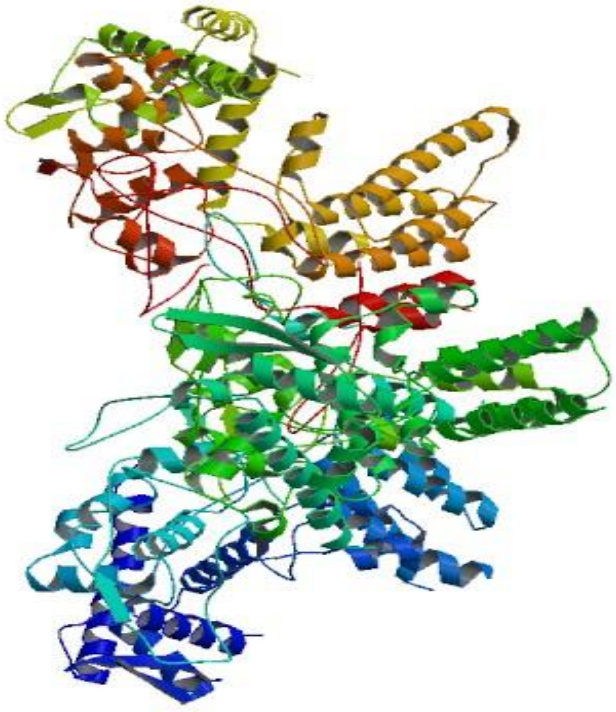
Life Cycle of Influenza Virus

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

synthesized virions bud from infected cell. NA destroys the sialic acid moiety of cellular receptors, thereby releasing the progeny virions.

2.1.4 Nucleoprotein

Influenza A virus RNA segment 5 encodes NP (a polypeptide of 498 amino acids in length), which is rich in arginine, glycine and serine residues and has a net positive charge at neutral pH. However, the majority of the polypeptide has a preponderance of basic amino acids and an overall predicted pI of 9·3, the C-terminal 30 residues of NP are, with a pI of 3·7, markedly acidic (Londo D.R. *et al.*, 1983). 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 L.L. *et al.*, 1993).



Structure of Influenza A (H1N1) Nucleoprotein

PDB ID: 2IQH

NP is able to interact with a variety of other macromolecules, of both viral and cellular origins. As well as binding to ssRNA, NP is able to self-associate to form large oligomeric complexes. It also binds the PB1 and PB2 subunits of the polymerase and the matrix protein M1. 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.

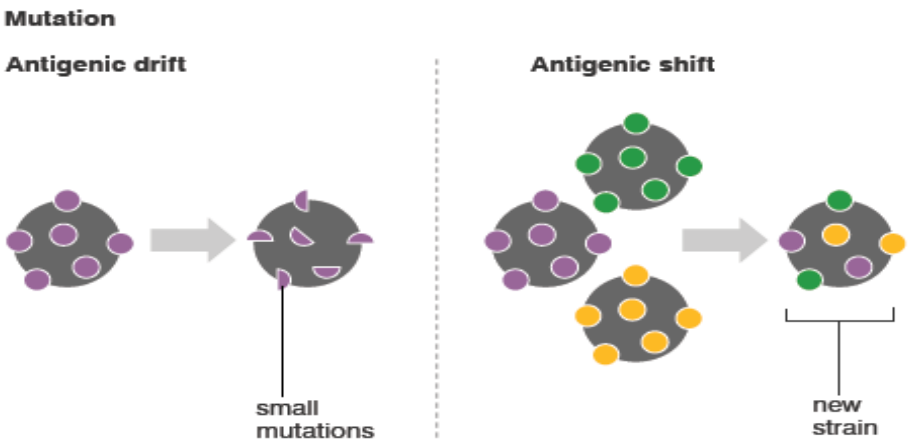
TABLE 4: Nucleoprotein interaction with various viral and cellular proteins and their function

Macromolecule	Function
Viral	
SsRNA	RNP Structure
NP	RNP Structure
PB1	Transcription
PB2	Transcription
M1	Transcription Inhibitor Nuclear Export
Cellular	
Importin α	Nuclear Import
F-actin	Cytoplasmic Retention
CRM 1/exportin-I	Nuclear Export
BAT 1 / UAP56	Transcription

2.1.5 Antigenic Drift and Antigenic Shift

Influenza virus shows remarkable antigenic variation which helps Influenza virus to evade the host immune response. These antigenic variations are due to two different mechanisms, antigenic drift and antigenic shift.

Antigenic drift involves a series of spontaneous point mutations that occur gradually resulting in evolution of viral genome and leads to formation of novel antigen which cannot be recognized by host defense system. Antigenic drift mainly occurs due to error prone RNA dependent RNA polymerase which has no proof reading activity and this leads to mutation.



Antigenic shift involves sudden change in viral antigens. This occurs when two different strains of viruses infect the same cell at the same time. Due to reassortment of segmented genome of Influenza virus, genomes of the two strains mix and rearrange to form a novel viral strain having character of both parent viruses. Animals such as pigs and birds harbor an archive of well preserved antigenic types. So if the two strains exchange the right genetic information they, might produce a more versatile and dangerous offspring.

2.2 Treatment for Influenza

Treatment of influenza is done by two different approaches. First one is by the use of anti-viral drug therapy and second approach is prevention by the use of vaccine.

2.2.1 Anti-viral Drug Therapy

Antiviral drugs directly target the influenza viruses. Anti-viral drug works optimally if taken in early stage of infection. Two main classes of drug interfere with influenza virus infection. The first class of drug is the ion channel blocker group of anti-influenza drugs which involves Rimantadine and Amantadine . The second class of agents consists of the viral neuraminidase inhibitors, Zanamivir (Relenza) and Oseltamivir (Tamiflu) (Laver W.G. *et al., 1999*).

Ion Channel Blockers

These drugs inhibit the function of the M2 protein and thus stop the replication process. These drugs is only effective against influenza A, and not influenza B, because influenza B does not

have an M2 protein, but a substitute protein called NB that is not affected by ion channel blocker drugs (Betakova T. *et al.* 1996). Example of ion channel blocker drugs are Rimantadine and Amantadine .

Neuraminidase Inhibitors

Neuraminidase cleaves the neuraminic acid component of sialic acid residue to which hemagglutinin is bound. This cleavage is necessary for the virus to get released from the host cell. The neuraminidase inhibitor drugs, Zanamivir and Oseltamivir, bind to the active site on the viral neuraminidase, blocking its activity. Thus, virus particles cannot exit the cells as easily, and they tend to clump and not disperse. This impedes their ability to infect more cells and attenuates the patient's infection.

2.2.2 Vaccines

Resistant mutants to both classes of antiviral agent have been detected. Also, costs, occasional side effects and the likely limited availability of such drugs during major outbreaks highlight the role of vaccination as the primary preventive measure against influenza (WHO, 2002). Vaccination is the administration of antigenic material to stimulate the immune system of an individual to develop adaptive immunity to a disease. HA and NA glycoprotein of influenza virus are used in influenza vaccine.

Current available influenza vaccines contain antigens from two A subtypes, H3N2 and H1N1, and one type B virus. These vaccines are of four types:

Whole virus vaccines

Whole virus vaccines are those vaccines in which complete virus particle is used for vaccination. Whole virus vaccine can be divided into two types.

Inactivated Influenza Vaccine

Inactivated influenza vaccine (IIV) was one of the first kind of vaccine used against influenza virus. IIV is synthesized by egg based culturing method and after culturing the viruses are inactivated by chemical treatment and used as a vaccine. Infecting virus shows antigenic variation which will make IIV less effective.

Live Attenuated Influenza Vaccine

Live attenuated influenza vaccine (LAIV) is the most used and most effective vaccine at present time. In these vaccines cold adapted and heat sensitive strain of influenza virus is used. Immunogenicity of this vaccine is very high. The only problem with this type of vaccine is that attenuation of the strain is a time consuming process. Currently available vaccines require annual revision of viral strains.

Split virus vaccines

Because of the incidence of reversion seen in whole virus vaccines, attempts have been made to produce a vaccine which is less antigenic but conserving immunogenicity. Split virus vaccines were prepared by disrupting virus particles with detergent treatment. Among split vaccine, the surface antigen vaccine contains predominantly purified HA and NA. These vaccines have been shown to induce fewer side effects and are as immunogenic as whole virus vaccine.

Subunit vaccines

Subunit vaccines consists only viral protein (haemagglutinin and neuraminidase) from which other virus components have been removed.

2.2.3 Epitope Based Vaccine

The resistance in vaccine during the past decade led to several new approaches of vaccine development. One of these approach is synthetic peptides for eliciting protective immunity against infection (Arnon R. *et al.*, 2006). Focusing on Influenza, several epitopes stimulating T helper cells and cytotoxic T lymphocytes (CTLs), provide protective immunity against several strains of influenza. These results are indicative of the potential of such epitope-based vaccines as long-range and broad-spectrum vaccines.

In epitope based vaccine immune response induced by an immunogen is not directed against the whole molecule, but it is against a specific region (sequence) of protein called epitopes. Epitopes are localized region of an antigen that is capable of eliciting an immune response by binding to membrane-bound antibody or T-cell receptor.

However epitope-based vaccines can be long-range and broad-spectrum vaccines, but with the use of small molecules such as peptides, a problem of low immunogenicity is encountered, as compared to multi-epitope protein antigens, or the entire pathogen, that is used for immunization in conventional vaccines. Antibody- mediated immunity in epitope based vaccine is currently not

feasible because the structural determinants of B-cell immunity are highly complex and there is no effective means for predicting the antibody epitope structure (Khan A.M. *et al.*, 2006).

2.3 Immunoinformatics

Immunoinformatics is a field of science that encompasses high-throughput genomic and bioinformatics approaches to immunology. The field's main aim is to convert immunological data into computational problems, solve these problems using mathematical and computational approaches and then convert these results into immunologically meaningful interpretations.

Vaccine informatics is an emerging research area in Immunoinformatics that focuses on development and applications of bioinformatics methods that can be used to facilitate every aspect of vaccine development. Many immunoinformatics algorithms and resources have been developed to predict T cell and B cell immune epitopes for epitope vaccine development and protective immunity analysis (Yongqun He *et al.*, 2010).

2.3.1 MHC Polymorphism and Vaccine Design

The major histocompatibility complex is a collection of genes arrayed within a long continuous stretch of DNA on chromosome 6 in humans. The MHC is referred to as the HLA complex in humans. MHC genes are organized into regions encoding three classes of molecules. Class I MHC genes encode glycoprotein expressed on the surface of nearly all nucleated cells; the major function of the class I MHC is presentation of intracellular antigens to T cytotoxic cells. Class II MHC genes encode glycoprotein expressed primarily on antigen-presenting cells (macrophages, dendritic cells, and B cells), where they present processed extracellular antigenic peptides to TH cells. Class III MHC genes encode, in addition to other products, various secreted proteins that have immune functions, including components of the complement system and molecules involved in inflammation (Kuby Immunology, Chapter 6).

The loci constituting the MHC are highly polymorphic; that is, many alternative forms of the gene, or alleles, exist at each locus among the population. The genes of the MHC loci lie close together. MHC class I gene complex contains three different HLA loci A, B and C each of which codes of α chain polypeptides of MHC class I molecule. The class II gene complex contains three major HLA loci, DP, DQ and DR; each of these loci codes α and β chain polypeptides of MHC class II molecule. Most individuals inherit the alleles encoded by these closely linked loci

as two sets, one from each parent. Each set of alleles is referred to as a haplotype. The alleles are codominantly expressed; that is, both maternal and paternal gene products are expressed in the same cells.

The MHC polymorphism within a species generates a diversity of binding specificities, and thus different patterns of responsiveness to antigens. So T cell epitopes which bind to multiple HLA molecules will be an effective vaccine target as it can provide better coverage of the population.

2.3.2 T-Cell Epitope Prediction

The major histocompatibility complex (MHC) plays a pivotal role in regulating immune response. During T cell activation, peptides derived from protein antigens are presented by MHC molecules. Only a small fraction of the derived peptides are involved in eliciting host immune response (Viret C. and Janeway C.A., 1999). The short antigenic peptides, derived from parent molecules have been proposed as interesting targets of vaccine design (Ishioka G.Y. *et al.*, 1999). Thus, a new generation of high efficiency, low toxicity epitope-based vaccines are in development.

The primary objective in these prediction methodologies is the calculation of MHC peptide binding because high affinity binding often correlates with immunogenicity and this depends on the stability of MHC peptide complexes (van der Burg *et al.*, 1996). A crucial step towards the rational design of peptide vaccines is the identification of T-cell epitopes from disease causing antigen proteins. Experimentally determined affinities data have formed the basis of many peptide-MHC binding prediction methods, able to effectively discriminate binding from nonbinding peptides (Dimitrov I *et al.* 2010). Prediction models are either based on sequence data or structure data.

Sequence-based methods rely on the primary sequence of peptides that are known to bind specific MHC allele using binding assays. The information on sequence anchors that are deterministic in binding are encoded into a binding motif, a position-dependent matrix, Hidden Markov Model (HMM), Support Vector Machine (SVM), stepwise discriminant analysis, or an Artificial Neural Network (ANN).These models are developed using peptide data in large databases derived from naturally bound peptides or synthetic peptide. In these approaches, the binding affinity of each peptide residue is scored using a matrix taking into account the relative contribution of residues at other positions. The sum of contributions by all the residues gives

predicted binding value. It should be noted that such matrices are to be generated for each known HLA alleles (ZhaoB *et al.*, 2007).

OBJECTIVE

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

Work plan of current study is as follows:

1. Finding out conserved regions of Nucleoprotein, from all available Nucleoprotein sequences of H1N1 and H3N2 serotype of Influenza virus.

2. Prediction of peptide containing overlapping T cell epitope in the conserved region of Nucleoprotein using different immunoinformatics tools which can act as a vaccine target.

3. Optimization of MTT assay to estimate cell proliferation of peripheral blood mononuclear cells.

MATERIAL AND METHODS

4.1 Conservancy analysis of NP

4.1.1 Sequence Retrieval

The protein sequence of NP of H1N1 and H3N2 were downloaded from Influenza virus resource database of NCBI (http://www.ncbi.nlm.nih.gov/genomes/FLU). All sequences which were collected between January 1918 and 31 January 2012 were downloaded. Sequences of H1N1 and H3N2 were downloaded separately.

First of all different criteria of searching were put in Influenza virus resource search engine and all the identical sequences were collapsed together. The sequences were downloaded in FASTA format in same order as they were in influenza virus recourse. This FASTA file is opened in Notepad and the data was saved in Microsoft office word. Some sequences had an invalid letter code "J" which does not represent any amino acid, and so "J" was replaced with "X".

TABLE 5: Criteria of searching in Influenza Virus Resource Database and number of sequence retrieved

Protein	Subtype	Host	Collection Date		Total Sequence	No. of non-redundant sequence
			From	To		
NP	H1N1	Human	Jan 1918	31 Jan 2012	4922	643
NP	H3N2	Human	Jan 1918	31 Jan 2012	3386	511

4.1.2 Multiple sequence alignment

Multiple sequence alignment is carried out to align multiple related sequences to achieve optimal matching of the sequences. Related sequences are identified through the database similarity searching (JIN XIONG, 2006). Multiple sequence alignment was carried out by MUSCLE tool.

MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/)

MUSCLE stands for **MU**ltiple **S**equence **C**omparison by **L**og- **E**xpectation. MUSCLE is a computer program for creating multiple alignments of protein sequences. Elements of the

algorithm include fast distance estimation using kmer counting, progressive alignment using the log-expectation score, and refinement using tree dependent restricted partitioning. MUSCLE is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee.

The limitation of MUSCLE is that it can only align 500 sequences of more than 350 amino acid length at a time. The number of sequence retrieved is more than 500 and MUSCLE (Multiple sequence alignment tool) can only align only 500 sequences at a time. So sequences were divided into two equal sets according to their order accession number in influenza virus resource for multiple sequence alignment.

TABLE 6: Grouping of NP sequences according to the order of their Accession No. in Influenza Virus Resource

Influenza Subtype	Group	Accession No.	Number of Sequences
H1N1	H1N1 ALN 1	AAA51491-AEW46954	300
	H1N1 ALN 2	ACR20063-AEM60008	337
H3N2	H3N2 ALN 1	ABV30419- ADR78616	240
	H3N2 ALN 2	ABB96502- AAC63462	271

The results of MUSCLE were saved in FASTA format.

4.1.3 Conservancy Analysis

Conservancy analysis was carried out to find out the conserved region present in nucleoprotein of H1N1 and H3N2 subtype of influenza virus. The regions of nucleoprotein showing ≥ 80% conservancy were selected. AVANA tool was used for conservancy analysis.

AVANA

The Antigen Variability ANAlyzer (AVANA) 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 (http://avana.sourceforge.net/).

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

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

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

Alignment result of MUSCLE (FASTA format) was used as input for AVANA software. Parameters were set to 80% conservancy and minimum length of 9 amino acids as a threshold value in AVANA. Conserved regions were searched in the alignments. The overlapping regions found in the conserved regions were merged together. After that conserved regions of different sets (ALN1and ALN2) of H1N1 and H3N2 were compared manually and conserved region common in both sets were considered as the final conserved region.

4.2 T-Cell Epitope Prediction

In epitope based vaccine, binding of the immunogenic epitope to the MHC molecule is the most crucial part for activation of antigen specific T-cell. So primary objective in these prediction methodologies is the calculation of MHC peptide binding because high affinity binding often correlates with higher immunogenicity and this depends on the stability of MHC peptide complexes (van der Burg S.H. *et al.*, 1996). There are two different type of epitope one which

binds to MHC class I and other to MHC class II. Various algorithms based online tools were used for prediction of both MHC Class I and MHC Class II binding peptide.

4.2.1 MHC Class I Epitope Prediction

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

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

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

Conserved regions of both H1N1 and H3N2 were taken as the input. Conserved regions were taken FASTA format and analysis was carried out for all 12 supertypes. Different parameters were set as follows:

1. Weight on C terminal cleavage = 0.15
2. Weight on TAP transport efficiency = 0.05
3. Threshold for epitope identification = 0.75

Only those epitopes showing score more than the threshold (≥ 0.75) were taken as epitopes.

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

SYFPEITHI contains a collection of MHC class I and class II ligands and peptide motifs of humans and other species. 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 both H1N1 and H3N2 were taken as the input sequence and only those epitopes were taken showing score ≥ 13. Different parameters selected for analysis were set as follows:

1. Length of epitope = 9 amino acid (Nonamer)
2. Threshold value for score ≥ 13

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

BIMAS stands for Bioinformatics and Molecular Analysis Section. BIMAS locate and rank 8-mer, 9-mer, or 10-mer peptides that contain peptide-binding motifs for HLA class I molecules. BIMAS works on the principle of Quantitative matrices. In this method, the contribution to binding from each amino acid at each peptide position within the binding groove is quantified (Parker K.C. *et al.*, 1994). It is assumed that each position within the peptide contributes independently in binding to an MHC molecule, and a residue located at a given peptide position contributes an equal amount to binding, even within different peptides. This method involves producing a matrix in which every entry (X, Y) represents a score associated with amino acid residue X at position Y. The position-specific amino acid values reflect the structural properties of HLA alleles, therefore representing a fingerprint for HLA binding domains.

Conserved regions of both H1N1 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 $T_{(½)}$ (estimate of half time of disassociation of a molecule containing this subsequence). Different parameters were set as follows:

1. Length of Epitope = 9 amino acid (Nonamer)
2. Predicted $T_{(½)} \geq 5$

4.2.2 MHC Class II Epitope Prediction

Class II MHC molecules bind peptides and present these peptides to $CD4^+$ T cells. Peptides of class II MHC–peptide complexes is generally 13–18 amino acid residues long, somewhat longer than the nonameric peptide of class I molecules but the core sequence is still 9 amino acid in length. The peptide-binding cleft in class II molecules is open at both ends, allowing longer peptides to extend beyond the ends. Three different online software having different epitope prediction algorithms were used to identify MHC class II epitope.

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

ProPred is a graphical online tool for predicting MHC class II binding regions in antigenic protein sequences. The server implement Quantitative Matrix based prediction algorithm, employing amino-acid / position coefficient table deduced from literature. The predicted binders can be visualized either as peaks in graphical interface or as colored residues in HTML interface (Singh H. *et al., 2001*).

Conserved regions of both H1N1 and H3N2 (amino acid sequence in single letter code) were taken as the input. ProPred carried out MHC class II epitope prediction in 51HLA-DR alleles. Different parameters were set as follows:

1. Threshold for score = 3%

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

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

IEDB stands for Immune Epitope Database. IEDB SMM-Align is a tool of IEDB analysis resource which predicts MHC Class II binding epitopes. It follows SMM-QM (Stabilized Matrix Method-Quantitative Matrices) algorithm for prediction which is a modification of original Quantitative Matrices (Nielsen M. *et al.,* 2007). IEDB SMM-Align identifies the MHC class II binding motif in terms of a position specific weight matrix. The output of the SMM-align method is IC_{50} binding affinity values, enabling direct readout of the peptide-MHC binding affinity. IC_{50} is half of the maximum inhibitory concentration, is a measure of the effectiveness of a

compound. The lower the IC_{50} of the peptide the higher the higher will be peptide-MHC binding affinity.

Conserved regions of both H1N1 and H3N2 were taken as the input. Conserved regions were taken FASTA format and analysis was carried out for all HLA class II molecules. Different parameters selected for analysis were set as follows:

1. Prediction Method = SMM-Align
2. Threshold for $IC_{50} \leq 500$

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 MHC class II alleles using artificial neural network. The artificial neural network method includes explicit encoding of the peptide flanking residues in terms of amino acid composition and length, as well as a novel scheme for neural network training that deals with the data redundancy inherent in the peptide data due to multiple examples of identical binding cores (Morten Nielsen *et al.*, 2009). The prediction values are given in nM IC_{50} values, and as a percentage rank to a set of 1,000,000 random natural peptides. Strong and weak binding peptides are indicated in the output. The accuracy of the peptide binding core identification has been improved using a neural network alignment procedure. This has made obsolete the need for P1 amino acids encoding.

Conserved regions of both H1N1 and H3N2 were taken as the input. Conserved regions were taken FASTA format and analysis was carried out for all HLA class II molecules. Different parameters selected for analysis were set as follows:

1. Threshold for IC50 $\leq$ 500
2. Turn on P1 amino acid preference = No

4.2.3 Epitope Selection

- For selection of putative epitope those epitopes having score equal to or more than threshold value were selected and results of all different online tools were compiled together.

- The epitopes were observed for maximum scores and showing positive results in maximum number of HLA Alleles. The epitope showing positive results in all three online tools of both MHC classes were compiled for different conserved regions.
- Overlapping epitopes were searched in each of the conserved region to finally get the immunogenic peptides for Nucleoprotein.

4.3 Optimization of the protocol for MTT cell proliferation assay

TABLE 7: Reagents for optimization of MTT cell proliferation assay

Requirements	Company
Powdered DMEM Media with Phenol Red	HIMEDIA
Powdered DMEM Media without Phenol Red	HIMEDIA
Sodium bicarbonate	HIMEDIA
L-glutamine	HIMEDIA
Foetal Bovine Serum	HIMEDIA
Penicillin	HIMEDIA
Streptomycin	HIMEDIA
Hisep Ficoll LSM	HIMEDIA
Trypan Blue	HIMEDIA
MTT Reagent	HIMEDIA
Dimethyl sulfoxide (DMSO)	SRL

4.3.1Preparation of Powdered DMEM media-

8.3 g of powder DMEM media was suspended in 900 ml double distilled water and constantly, stirred gently until the powder was completely dissolved. 49.3 ml of 7.5% Sodium bicarbonate solution and20 ml of 200 mM L-glutamine solution were added for 1 liter of medium and stirred until dissolved. pH was adjusted 0.2-0.3 pH units below the desired pH using 1N HCl or 1N NaOH since the pH tends to rise during filtration. The final volume was made up to 1000 ml with double distilled water. The medium was immediately sterilized by filtering through a sterile membrane filter with porosity of 0.22 micron or less, using positive pressure rather than vacuum to minimize the loss of carbon dioxide. Liquid medium was stored at 2-8°C and in dark till use.

10% Heat inactivated fetal bovine serum (57^{o}C for 30 minutes) and filter sterilized antibiotics (Penicillin and Streptomycin) are added to media before culturing of cells.

4.3.2 Preparation of PBS-

For 1 liter of 1X PBS, 8 g of NaCl, 0.2 g of KCl. 1.44 g of Na_2HPO_4. 0.24 g of KH_2PO_4.were added in 800 ml of distilled water. pH was adjusted to 7.4 using HCl and NaOH. Volume was made up to 1 liter by distilled water. PBS was autoclaved for 20 min at 121°C. After autoclaving PBS was stored at 4^{o}C temperature.

4.3.3 Isolation of Peripheral Blood Mononuclear Cells

Blood was drawn from a healthy human being with the help of vacutainer system (EDTA coated, Becton Dickinson). Blood was diluted in 1:1 ratio with PBS. Now blood sample was layered carefully over equal volume of ficoll and it was centrifuged at 600×g for 30 min. at 25^{o}C. Plasma was removed and then the buffy coat layer was taken out with help of a dropper. Buffy coat layer was diluted 1:1 with PBS and centrifuged at 500×g for 10 min. at 25^{o}C. Supernatant was discarded and pellet was again washed with 5ml of PBS and centrifuged at 500×g for 10 min. at 25^{o}C. Supernatant was discarded and pellet of PBMC was suspended in cell culture medium [DMEM + 10% FBS + Penicillin (100 units/ml) + Streptomycin (100 µg/ml)].

4.3.4 Cell Counting and viability testing

Cell counting was done with the help of hemocytometer using trypan blue as a stain. Trypan blue is a stain which penetrates cell membrane of dead cells and dead cells are stained blue while live cells remain unstained. 30µl of PBMC cell suspension was taken and equal volume of trypan blue was added to it. Now cell suspension containing trypan blue was loaded on hemocytometer. Hemocytometer was focused on using the 10X objective of the microscope and cells were counted in all 4 sets of squares of hemocytometer using 40X objective of the microscope.
Cell count and percent cell viability were calculated using following formula

$$Cell\ Count = \frac{Total\ number\ of\ cells\ counted}{\text{Number of chambers counted}} \times \text{Dilution Factor}$$

$$Percent\ Cell\ Viability = \frac{Total\ number\ of\ viable\ cells}{Total\ number\ of\ cells\ counted} \times 100$$

4.3.5 Optimization of protocol for PBMC Proliferation Assay at preliminary step (MTT Assay)

Cell proliferation was tested using a 3-(4, 5-dimethylthiazol-2-yl) - 2, 5-diphenyl tetrazolium bromide (MTT) assay. For the assay, lymphocytes were freshly isolated and plated in 96-well flat bottom microtiter plate containing 100 µl of cell culture medium[DMEM + 10% FBS + Penicillin (100 units/ml) + Streptomycin (100 µg/ml)]. Plate was incubated CO_2 incubator (eppendorf) at 37^oC and 5% CO_2 concentration for 48 hours. After incubation 10µl of MTT was added to each well and an incubation of another 4 hours was given for reduction of MTT to formazan. After incubation formazan crystals were dissolved in 100 µl of DMSO and O.D. was taken at 570 nm by microtiter plate reader (Thermo Scientific).

For preliminary optimization two types DMEM media were used one with phenol red and other was without phenol red; and two types of DMSO were used one laboratory grade and other cell freezing DMSO.

TABLE 8: Micro titer plate arrangement for Optimization of protocol for Media and DMSO

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		M_1+D_L	M_2+D_L		M_1+D_C	M_2+D_C						
C		M_1+D_L	M_2+D_L		M_1+D_C	M_2+D_C						
D		M_1+D_L	M_2+D_L		M_1+D_C	M_2+D_C						
E												
F												
G												
H												

M_1-Medium with phenol red, M_2-Medium without phenol red,
D_L-Laboratory grade DMSO, D_C – Cell Freezing DMSO

In another process of preliminary optimization to optimize cell concentration and the procedure of cell proliferation assay cell concentration gradient was taken.

TABLE 9: Micro titer plate arrangement for optimization of MTT assay

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B	P	M	C1	C2	C3	C4	C5	C6	C7	C8	M	P
C	P	M	C1	C2	C3	C4	C5	C6	C7	C8	M	P
D	P	M	C1	C2	C3	C4	C5	C6	C7	C8	M	P
E	P	M	C1	C2	C3	C4	C5	C6	C7	C8	M	P
F												
G												
H												

Serial No.	Number of cells per well
C1	$4x10^5$
C2	$3x10^5$
C3	$2x10^5$
C4	$1.5x10^5$
C5	$1x10^5$
C6	$0.75x10^5$
C7	$0.5x10^5$
C8	$0.25x10^5$

P-PBS, M-Media without phenol red, C- PBMC cell concentration

RESULT AND DISCUSSION

5.1 Conserved regions of NP in H1N1 and H3N2

Sequences were aligned with MUSCLE and conserved regions were obtained for both groups of H1N1 and H3N2 with help of AVANA. Due to different alignment (MSA) in ALN1 and ALN2 groups the position of same conserved region came out to be different.

Seventeen final conserved regions having a conservancy range 80%-99.33% were found in NP of H1N1 strain of influenza and their length varied from 9-42 amino acid residues (TABLE 10). In NP of H3N2 13 final conserved regions having a conservancy range 80%-99.33% and their length varied from 9-93 amino acid residues (TABLE 11). This shows that nucleoprotein is highly conserved which has already reported. It is reported that in comparision to surface protein, internal protein are more conserved (Epstein S.L. *et al*., 2005).

TABLE 10: NP conserved region in H1N1

Serial No.	Sequence	Length	Conservancy (%)
H1N1 CS 1	MASQGTKRSYEQMET	15	≥95.76%
H1N1 CS 2	ATEIRASVGRMIGGIGRFYIQMCTELKL	28	≥80.00%
H1N1 CS 3	TIERMVLSAFDERRNKYLEEHPSAGKDPKK TGGPIY	36	≥92.31%
H1N1 CS 4	LYDKEEIRR	9	≥96.39%
H1N1 CS 5	TAGLTHIMIWHSNLND	16	≥80.00%
H1N1 CS 6	TYQRTRALVRTGMDPRMCSLMQGSTLPRRS GAAGAAVKGVGT	42	≥93.47%
H1N1 CS 7	ELIRMIKRGINDRNFWRGENGR	22	≥94.67%
H1N1 CS 8	AYERMCNILKGKFQTAAQ	18	≥96.33%
H1N1 CS 9	AMMDQVRESRNPGNAEIEDL	20	≥92.56%
H1N1 CS 10	FLARSALILRGSVAHKSCLPACVYG	25	≥95.67%
H1N1 CS 11	RPNENPAHKSQLVWMAC	17	≥97.33%

H1N1 CS 12	SAAFEDLRVSSFIRG	15	≥92.00%
H1N1 CS 13	PRGKLSTRGVQIASNEN	17	≥92.88%
H1N1 CS 14	TLELRSRYWAIRTRSGGNTNQQ	22	≥95.67%
H1N1 CS 15	QPTFSVQRNLPF	12	≥99.33%
H1N1 CS 16	SFQGRGVFELSDE	13	≥95.99%
H1N1 CS 17	ATNPIVPSFDMSNEGSYFFGDNAEEYD	27	≥80.33%

TABLE 11: NP conserved region in H3N2

Serial No.	Sequence	Length	Conservancy (%)
H3N2 CS 1	MASQGTKRSYEQMETDG	17	≥95.00%
H3N2 CS 2	SVGKMIDGIGRFYIQMCTELKLSD	24	≥80.83%
H3N2 CS 3	EGRLIQNSLTIEKMVLSAFDERRN	24	≥85.24%
H3N2 CS 4	YLEEHPSAGKDPKKTGGPIY	20	≥96.67%
H3N2 CS 5	WMRELVLYDKEEIRRIWRQANNG	23	≥80.00%
H3N2 CS 6	MIWHSNLND	9	≥99.63%
H3N2 CS 7	TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKGIGTMVMELIRM	50	≥88.19%
H3N2 CS 8	KRGINDRNFWRGENGRKTRSAYERMCNILKGKFQTAAQRAM	41	≥84.17%
H3N2 CS 9	DQVRESRNPGNAEIEDLIFLARSALILRGSVAHKSCLPAC	41	≥80.83%
H3N2 CS 10	SGYDFEKEGYSLVGIDPFKLLQNSQ	25	≥88.56%
H3N2 CS 11	YSLIRPNENPAHKSQLVWMACHSAAFEDLRLLSFIRGTKVSPRGKLSTRGVQIASNENMDNMGSSTLELRSGYW	93	≥80.33%

	AIRTRSGGNTNQQRASAGQ		
H3N2 CS 12	SVQPTFSVQRNLPF	14	≥80.83%
H3N2 CS 13	MAAFTGNTEGRTSDMRAEIIRMME GAKPEEVSFRGRGVFELSDEKATPN IVPSFDMSNEGSYFFGDNAEEYD	72	≥80.00%

5.2 Immunogenic peptide prediction from putative epitopes of NP for H1N1

Three different programs were used to predict epitopes for each MHC Class I and Class II. NetCTL 1.2, SYFPEITHI and BIMAS were used for the prediction of MHC class I epitope. ProPred, IEDB-SMM Align and NetMHC II 2.2 were used for MHC class II epitope prediction. All the epitopes that gave score more than threshold value and gave significant result for all three programs used for their MHC type were selected as epitopes. Immunogenic peptides were identified by selecting the overlapping epitopes.

5.2.1 Immunogenic peptide prediction for MHC Class I of H1N1 strain

16 MHC class I specific immunogenic detected were detected in 17 conserved regions of NP in H1N1. Out of 17 conserved regions 2 conserved regions (H1N1 CS 4 and H1N1 CS 9) did not have any immunogenic peptide while H1N1 CS 3 have two immunogenic peptides. The length of immunogenic peptide varied from 9-36 amino acid residues. Conserved region H1N1 CS 6 had the longest immunogenic peptide (36 a.a.) which was formed by 7 different MHC class I epitope (TABEL 13). A representative example of finding putative epitopes for MHC class I is shown in table 12.

TABLE 12: Representation of results for H1N1 MHC Class I epitope prediction

Most Conserved Region (PEPTIDE)	NONAMER	CLASS I					
		SYFPEITHI		NetCTL		BIMAS	
		No. of Alleles	Score Range	No. of Alleles	Score Range	No. of Alleles	Score Range
H1N1 CS 1	ASQGTKRSY	5	13-20	2	0.7674-1.2114	2	6-9
H1N1 CS 2	EIRASVGRM	5	13-19	1	1.569	1	10
	FYIQMCTEL	4	13-22	2	0.9114-1.2492	1	6
	GRFYIQMCT	2	14-20	1	1.0411	2	5-1000
	GRMIGGIGR	4	14-26	1	14-26	1	1000
	IQMCTELKL	9	13-18	2	0.8782-0.9273	6	6.6-200
	MIGGIGRFY	5	13-19	4	0.8203-1.068	1	48
	RMIGGIGRF	7	13-25	4	0.8271-1.5202	2	6-75
H1N1 CS 3	DPKKTGGPI	3	19-26	2	0.9297-1.0212	3	8-968
	FDERRNKYL	3	14-27	2	0.9263-1.6124	2	20-40
	IERMVLSAF	6	14-23	3	0.8315-1.0741	2	15-90
	RRNKYLEEH	1	26	1	1.0227	1	600

TABLE13: Epitopes and Immunogenic peptide of NP H1N1 strain (MHC Class I)

Sr. No.	Immunogenic Peptide	Nonamer
H1N1 CS 1	MASQGTKRSYEQMET	ASQGTKRSY
H1N1 CS 2	ATEIRASVGRMIGGIGRFYIQMCTELKL	EIRASVGRM
		MIGGIGRFY
		GRMIGGIGR
		RMIGGIGRF
		GRFYIQMCT
		FYIQMCTEL
		IQMCTELKL
H1N1 CS 3	TIERMVLSAFDERRNKYLEEHPSAGKDPKKTGGPIY	IERMVLSAF
		FDERRNKYL
		RRNKYLEEH
		DPKKTGGPI
H1N1 CS 4	LYDKEEIRR	-
H1N1 CS 5	TAGLTHIMIWHSNLND	AGLTHIMIW
		HIMIWHSNL

H1N1 CS 6	TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKGVGT	TYQRTRALV YQRTRALVR TRALVRTGM VRTGMDPRM GMDPRMCSL LMQGSTLPR RRSGAAGAA
H1N1 CS 7	ELIRMIKRGINDRNFWRGENGR	KRGINDRNF RGINDRNFW GINDRNFWR
H1N1 CS 8	AYERMCNILKGKFQTAAQ	AYERMCNIL RMCNILKGK ILKGKFQTA
H1N1 CS 9	AMMDQVRESRNPGNAEIEDL	-
H1N1 CS 10	FLARSALILRGSVAHKSCLPACVYG	FLARSALIL ILRGSVAHK KSCLPACVY
H1N1 CS 11	RPNENPAHKSQLVWMAC	NPAHKSQLV PAHKSQLVW
H1N1 CS 12	SAAFEDLRVSSFIRG	FEDLRVSSF
H1N1 CS 13	PRGKLSTRGVQIASNEN	KLSTRGVQI
H1N1 CS 14	TLELRSRYWAIRTRSGGNTNQQ	ELRSRYWAI
H1N1 CS 15	QPTFSVQRNLPF	FSVQRNLPF
H1N1 CS 16	SFQGRGVFELSDE	FQGRGVFEL
H1N1 CS 17	ATNPIVPSFDMSNEGSYFFGDNAEEYD	ATNPIVPSF NPIVPSFDM FDMSNEGSY MSNEGSYFF

5.2.2 Immunogenic peptide prediction for MHC Class II of H1N1 strain

10 MHC class II specific immunogenic peptides were detected in 17 conserved regions of NP in H1N1. Out of 17 conserved regions 9 conserved regions did not have any immunogenic peptide while 2 conserved regions (H1N1 CS 2 and H1N1 CS 6) have two immunogenic peptides each. The length of immunogenic peptide varied from 9-17 amino acid residues (TABEL 15). A representation of finding class II MHC putative epitopes for H1N1 is shown in Table 14.

TABLE 14: Representation of results for H1N1 MHC Class II epitope prediction

Most Conserved Region (PEPTIDE)		CLASS II					
	NONAMER	PROPRED		IEDB (Smm Alin.)		NET MHC II	
		No. of Alleles	Score Range	No. of Alleles	Score Range	No. of Alleles	Score Range
H1N1 CS 1	_	_	_	_	_	_	_
H1N1 CS 2	FYIQMCTEL	12	20.45-51.5	3	41-325	1	306
	IRASVGRMI	16	10.99-53.45	9	17-455	3	5.8-238.2
	YIQMCTELK	17	7.5-44.19	6	36-443	2	50.1-334.8
H1N1 CS 3	_	_	_	_	_	_	_
H1N1 CS 4	_	_	_	_	_	_	_
H1N1 CS 5	IMIWHSNLN	26	9.2-51.6	2	46-166	1	104.4
H1N1 CS 6	LMQGSTLPR	26	3.41-32.63	3	111-426	1	304.2
	YQRTRALVR	28	3.19-54.08	4	46-267	3	7.9-281.4
H1N1 CS 7	LIRMIKRGI	43	6.5-64.77	5	32-443	2	27.9-322.5
H1N1 CS 8	YERMCNILK	19	3.14-52.04	5	150-265	2	206.8-383.2
H1N1 CS 9	_	_	_	_	_	_	_

TABLE 15: Epitopes and Immunogenic peptide of NP H1N1 strain (MHC Class II)

Sr. No.	Immunogenic Peptide	Nonamer
H1N1 CS 1	MASQGTKRSYEQMET	-
H1N1 CS 2	ATEIRASVGRMIGGIGRFYIQMCTELKL	IRASVGRMI
		FYIQMCTEL
		YIQMCTELK
H1N1 CS 3	TIERMVLSAFDERRNKYLEEHPSAGKDPKKTGGPIY	-
H1N1 CS 4	LYDKEEIRR	-
H1N1 CS 5	TAGLTHIMIWHSNLND	IMIWHSNLN
H1N1 CS 6	TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKGVGT	YQRTRALVR
		LMQGSTLPR

H1N1 CS 7	ELIRMIKRGINDRNFWRGENGR	LIRMIKRGI
H1N1 CS 8	AYERMCNILKGKFQTAAQ	YERMCNILK
H1N1 CS 9	AMMDQVRESRNPGNAEIEDL	-
H1N1 CS 10	FLARSALILRGSVAHKSCLPACVYG	FLARSALIL
		ILRGSVAHK
		LRGSVAHKS
H1N1 CS 11	RPNENPAHKSQLVWMAC	-
H1N1 CS 12	SAAFEDLRVSSFIRG	-
H1N1 CS 13	PRGKLSTRGVQIASNEN	-
H1N1 CS 14	TLELRSRYWAIRTRSGGNTNQQ	LRSRYWAIR
		YWAIRTRSG
H1N1 CS 15	QPTFSVQRNLPF	-
H1N1 CS 16	SFQGRGVFELSDE	-
H1N1 CS 17	ATNPIVPSFDMSNEGSYFFGDNAEEYD	FFGDNAEEY

5.2.3 Common immunogenic peptide selection of NP from both Classes of MHC molecule for H1N1

Common immunogenic peptides were found by identifying and selecting the common region of both MHC class I and MHC class II immunogenic peptides.

Six common immunogenic peptides were found in NP of H1N1. The length of common immunogenic peptide varied from 9-17 amino acid residues. . If the common region between MHC class I and MHC class II immunogenic peptides is less than 9 amino acids then it cannot act as an immunogenic peptide (Table 16).

TABLE 16: Common Immunogenic peptide of NP for H1N1

Sr. No.	MHC Class I Immunogenic Peptide	MHC Class II Immunogenic Peptide	COMMON IMMUNOGENIC PEPTIDE
H1N1 CS 1	ASQGTKRSY	_	_
H1N1 CS 2 (i)	EIRASVGRMIGGIGRFYIQMCTELKL	IRASVGRMI	IRASVGRMI
H1N1 CS 2 (ii)		FYIQMCTELK	FYIQMCTELK
H1N1 CS 3	IERMVLSAFDERRNKYLEEH	_	_
	DPKKTGGPI		_
H1N1 CS 4	_	_	_
H1N1 CS 5	AGLTHIMIWHSNL	IMIWHSNLN	_
H1N1 CS 6 (i)	TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAA	YQRTRALVR	YQRTRALVR
H1N1 CS 6 (ii)		LMQGSTLPR	LMQGSTLPR
H1N1 CS 7	KRGINDRNFWR	LIRMIKRGI	
H1N1 CS 8	AYERMCNILKGKFQTA	YERMCNILK	YERMCNILK
H1N1 CS 9	_	_	_
H1N1 CS 10	FLARSALILRGSVAHKSCLPACVY	FLARSALILRGSVAHKS	FLARSALILRGSVAHKS
H1N1 CS 11	NPAHKSQLVW	_	_
H1N1 CS 12	FEDLRVSSF	_	_
H1N1 CS 13	KLSTRGVQI	_	_
H1N1 CS 14	ELRSRYWAI	LRSRYWAIRTRSG	_
H1N1 CS 15	FSVQRNLPF	_	_
H1N1 CS 16	FQGRGVFEL	_	_

H1N1 CS 17	ATNPIVPSFDMSNEGS YFF	FFGDNAEEY	_

5.2.4 Verification of predicted common immunogenic peptide of NP (H1N1) from literature

Three out of six common immunogenic peptides for H1N1 were found to be already reported in IEDB as shown in Table 17. Hence it is interesting to consider all these peptide for evaluation of immunogenic response in the *in vitro* approach. These peptides may be considered for target of vaccine design for H1N1 virus.

TABLE 17: Predicted peptides verification (of H1N1) from Literature (Source: IEDB, Immune Epitope Database)

Sr. No.	Common Immunogenic Peptide	Reference Epitope	Epitope ID	References
H1N1 CS 6 (ii)	LMQGSTLPR	ALVRTGMDPR MCSLMQGSTL	1542406	Junbao Yang 2009
H1N1 CS 8	YERMCNILK	IAYERMCNILK GKFQTAA	1835708	Chao Wu; Proc Natl Acad Sci U S A 2011
H1N1 CS 10	FLARSALILRGSVA HKS	FLARSALILRGS VAHKS	1766069	Jenny Aurielle B Babon; Hum Immunol 2009

5.3 Immunogenic peptide prediction from putative epitopes of NP for H3N2

Immunogenic peptides and epitopes of NP in H3N2 virus were predicted in same way as discussed in section 5.2 for H1N1.

5.3.1 Immunogenic peptide prediction for MHC Class I of H1N1 strain

14 MHC class I specific immunogenic peptides were detected in 13 conserved regions of NP in H3N2. Out of 13 conserved regions one conserved regions (H3N2 CS 6) did not have an immunogenic peptide while other two conserved regions (H3N2 CS 8 and H3N2 CS 13) have two immunogenic peptides each. The length of immunogenic peptide varied from 9-69 amino

acid residues. Conserved region H1N1 CS 11 had the longest immunogenic peptide (69 a.a.) which was formed by 17 different MHC class I epitope (Table 19). A representative example to find putative epitopes for each class I MHC is shown in Table 18.

TABLE 18: Representation of results for H3N2 MHC Class I epitope prediction

Most Conserved Region (PEPTIDE)	NONAMER	CLASS I					
		SYFPEITHI		NetCTL		BIMAS	
		No. of Alleles	Score Range	No. of Alleles	Score Range	No. of Alleles	Score Range
H3N2 CS 1	ASQGTKRSY	5	13-20	2	0.7906-1.2346	2	6-9
H3N2 CS 2	FYIQMCTEL	4	13-22	2	0.912-1.2497	2	6-330
	GRFYIQMCT	2	14-20	1	1.0283	2	5-1000
	IQMCTELKL	9	13-18	2	0.8774-0.9265	6	6.6-200
	KMIDGIGRF	8	13-21	3	1.2387-1.4813	2	7.2-75
	MIDGIGRFY	5	14-25	2	0.9043-2.3523	2	7.2-25
H3N2 CS 3	EGRLIQNSL	8	13-21	1	0.8778	3	6.72-48.4
	IEKMVLSAF	5	14-23	3	1.2414	2	15-90
	IQNSLTIEK	2	14-16	1	1.2299	1	200
	RLIQNSLTI	6	13-24	6	0.8106-0.9765	2	10.433-27
	SLTIEKMVL	5	14-21	1	0.7737	2	8.759-30

TABLE 19: Epitopes and Immunogenic peptide of NP H3N2 strain (MHC Class I)

Sr. No.	Immunogenic Peptide	Nonamer
H3N2 CS 1	MASQGTKRSYEQMETDG	ASQGTKRSY
H3N2 CS 2	SVGKMIDGIGRFYIQMCTELKLSD	KMIDGIGRF
		MIDGIGRFY
		GRFYIQMCT
		FYIQMCTEL
		IQMCTELKL
H3N2 CS 3	EGRLIQNSLTIEKMVLSAFDERRN	EGRLIQNSL
		RLIQNSLTI
		IQNSLTIEK
		SLTIEKMVL
		IEKMVLSAF

H3N2 CS 4	YLEEHPSAGKDPKKTGGPIY	DPKKTGGPI
H3N2 CS 5	WMRELVLYDKEEIRRIWRQANNG	MRELVLYDK
H3N2 CS 6	MIWHSNLND	-
H3N2 CS 7	TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAA VKGIGTMVMELIRM	TYQRTRALV
		YQRTRALVR
		TRALVRTGM
		VRTGMDPRM
		GMDPRMCSL
		LMQGSTLPR
		RRSGAAGAA
		AAVKGIGTM
H3N2 CS 8	KRGINDRNFWRGENGRKTRSAYERMCNILKGKFQTA AQRAM	KRGINDRNF
		RGINDRNFW
		GINDRNFWR
		AYERMCNIL
		RMCNILKGK
		ILKGKFQTA
H3N2 CS 9	DQVRESRNPGNAEIEDLIFLARSALILRGSVAHKSCLP AC	NAEIEDLIF
		AEIEDLIFL
		LIFLARSAL
		IFLARSALI
		FLARSALIL
		ILRGSVAHK
H3N2 CS 10	SGYDFEKEGYSLVGIDPFKLLQNSQ	FEKEGYSLV
		KEGYSLVGI
		YSLVGIDPF
		SLVGIDPFK
H3N2 CS 11	YSLIRPNENPAHKSQLVWMACHSAAFEDLRLLSFIRG TKVSPRGKLSTRGVQIASNENMDNMGSSTLELRSGY WAIRTRSGGNTNQQRASAGQ	NPAHKSQLV
		PAHKSQLVW
		LVWMACHSA
		WMACHSAAF
		CHSAAFEDL
		SAAFEDLRL
		FEDLRLLSF
		LLSFIRGTK
		IRGTKVSPR
		SPRGKLSTR
		KLSTRGVQI
		VQIASNENM
		MDNMGSSTL
		NMGSSTLEL
		STLELRSGY
		ELRSGYWAI
		LRSGYWAIR
H3N2 CS 12	SVQPTFSVQRNLPF	FSVQRNLPF

H3N2 CS 13	MAAFTGNTEGRTSDMRAEIIRMMEGAKPEEVSFRGR GVFELSDEKATPNIVPSFDMSNEGSYFFGDNAEEYD	GAKPEEVSF
		VSFRGRGVF
		FRGRGVFEL
		GVFELSDEK
		ATPNIVPSF
		MSNEGSYFF

5.3.2 Immunogenic peptide prediction for MHC Class II of H1N1 strain

13 MHC class II specific immunogenic peptides were detected in 13 conserved regions of NP in H3N2. Out of 13 conserved regions 5 conserved regions did not have any immunogenic peptide while other two conserved regions (H3N2 CS 7 and H3N2 CS 11) have three immunogenic peptides each. The length of immunogenic peptide varied from 9-19 amino acid residues (Table 21). A representative example to find putative epitopes for each class II MHC is shown in Table 20.

TABLE 20: Representation of results for H3N2 MHC Class II epitope prediction

Most Conserved Region (PEPTIDE)		CLASS II					
	NONAMER	PROPRED		IEDB (Smm Alin.)		NET MHC II	
		No. of Alleles	Score Range	No. of Alleles	Score Range	No. of Alleles	Score Range
H3N2 CS 1	_	_	_	_	_	_	_
H3N2 CS 2	YIQMCTELK	22	7.5-44.19	7	35-443	2	50.1-334.8
	FYIQMCTEL	15	14.29-51.5	4	41-325	1	306
	IQMCTELKL	26	7.14-40.82	4	220-446	1	204.4
H3N2 CS 3	_	_	_	_	_	_	_
H3N2 CS 4	_	_	_	_	_	_	_
H3N2 CS 5	IRRIWRQAN	37	7.5-41.86	2	95-481	1	453.1
	VLYDKEEIR	1	5.56	1	53	1	200.5
H3N2 CS 6	_	_	_	_	_	_	_
H3N2 CS 7	IGTMVMELI	7	2.27-22.41	4	101-322	2	148.3-460.3
	LMQGSTLPR	26	3.41-32.63	3	111-426	1	304.2
	YQRTRALVR	28	3.19-54.08	4	46-267	3	7.9-281.4
H3N2 CS 8	YERMCNILK	30	3.3-52.04	7	55-459	2	206.6-383.2

TABLE 21: Epitopes and Immunogenic peptide of NP H3N2 strain (MHC Class II)

Sr. No.	Immunogenic Peptide	Nonamer
H3N2 CS 1	MASQGTKRSYEQMETDG	-
H3N2 CS 2	SVGKMIDGIGRFYIQMCTELKLSD	FYIQMCTEL YIQMCTELK IQMCTELKL
H3N2 CS 3	EGRLIQNSLTIEKMVLSAFDERRN	-
H3N2 CS 4	YLEEHPSAGKDPKKTGGPIY	-
H3N2 CS 5	WMRELVLYDKEEIRRIWRQANNG	VLYDKEEIR IRRIWRQAN
H3N2 CS 6	MIWHSNLND	-
H3N2 CS 7	TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAA VKGIGTMVMELIRM	YQRTRALVR LMQGSTLPR IGTMVMELI
H3N2 CS 8	KRGINDRNFWRGENGRKTRSAYERMCNILKGKFQTAA QRAM	YERMCNILK
H3N2 CS 9	DQVRESRNPGNAEIEDLIFLARSALILRGSVAHKSCLPA C	LIFLARSAL IFLARSALI FLARSALIL ILRGSVAHK LRGSVAHKS
H3N2 CS 10	SGYDFEKEGYSLVGIDPFKLLQNSQ	YSLVGIDPF LVGIDPFKL
H3N2 CS 11	YSLIRPNENPAHKSQLVWMACHSAAFEDLRLLSFIRGT KVSPRGKLSTRGVQIASNENMDNMGSSTLELRSGYWA IRTRSGGNTNQQRASAGQ	FEDLRLLSF LRLLSFIRG LLSFIRGTK LSFIRGTKV VQIASNENM YWAIRTRSG
H3N2 CS 12	SVQPTFSVQRNLPF	-
H3N2 CS 13	MAAFTGNTEGRTSDMRAEIIRMMEGAKPEEVSFRGRG VFELSDEKATPNIVPSFDMSNEGSYFFGDNAEEYD	IIRMMEGAK IRMMEGAKP MEGAKPEEV FFGDNAEEY

5.3.3 Common immunogenic peptide selection of NP from both Classes of MHC molecule for H3N2

Common immunogenic peptides were found by identifying and selecting the common region of both MHC class I and MHC class II immunogenic peptides If the common region between MHC

class I and MHC class II immunogenic peptides is less than 9 amino acids then it cannot act as an immunogenic peptide.

Seven common immunogenic peptides were found in NP of H3N2 and their length varied from 9-18 amino acid residues in length (Table 22).

TABLE 22: Combined Immunogenic peptide of NP for H3N2

Sr. No.	MHC Class I Immunogenic Peptide	MHC Class II Immunogenic Peptide	COMMON IMMUNOGENIC PEPTIDE
H3N2 CS 1	ASQGTKRSY	_	_
H3N2 CS 2	KMIDGIGRFYIQMCTELKL	FYIQMCTELKL	FYIQMCTELKL
H3N2 CS 3	EGRLIQNSLTIEKMVLSAF	_	_
H3N2 CS 4	DPKKTGGP	_	_
H3N2 CS 5	MRELVLYDK	VLYDKEEIRRIWRQAN	_
H3N2 CS 6	_	_	_
H3N2 CS 7 (i)	TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKGIGTM	YQRTRALVR	YQRTRALVR
H3N2 CS 7 (ii)		LMQGSTLPR	LMQGSTLPR
		IGTMVMELI	
H3N2 CS 8	KRGINDRNFWR	YERMCNILK	YERMCNILK
	AYERMCNILKGKFQTA		

H3N2 CS 9	NAEIEDLIFLARSALILRG SVAHK	LIFLARSALILRGSVAHKS	LIFLARSALILRG SVAHK
H3N2 CS 10	FEKEGYSLVGIDPFK	YSLVGIDPFKL	_
H3N2 CS 11 (i)	NPAHKSQLVWMACHSA AFEDLRLLSFIRGTKVSPR GKLSTRGVQIASNENMD NMGSSTLELRSGYWAIR	FEDLRLLSFIRGTKV	FEDLRLLSFIRGT KV
H3N2 CS 11 (ii)		VQIASNENM	VQIASNENM
		YWAIRTRSG	
H3N2 CS 12	FSVQRNLPF	_	_
H3N2 CS 13	GAKPEEVSFRGRGVFELS DEKATPNIVPSF	IIRMMEGAKPEEV	_
	MSNEGSYFF	FFGDNAEEY	

5.3.4 Verification of predicted common immunogenic peptide of NP (H3N2) from literature

Six epitopes were found out of seven common immunogenic peptides for H3N2 were found to be already reported as shown in table 23. Hence it is interesting to consider all these peptide for evaluation of immunogenic response in the *in vitro* approach. These peptides may be considered for target of vaccine design for H3N2 virus.

TABLE 23: Predicted peptides verification (of H3N2) from Literature (Source: IEDB, Immune Epitope Database)

Sr. No.	Common Immunogenic Peptide	Reference Epitope	Epitope ID	References
H3N2 CS 2	FYIQMCTELKL	MIDGIGRFYIQMCTELKL	1598788	Laurel Yong-Hwa Lee; J Clin Invest 2008
H3N2 CS 7 (i)	YQRTRALVR	TYQRTRALVRTGMDPRM	1927779	Tom M Wilkinson; Nat Med 2012
H3N2 CS 8	YERMCNILK	RMCNILKGKFQTAAQRAM	1598798	Laurel Yong-Hwa Lee; Clin Invest 2008
H3N2 CS 9	LIFLARSALILRGSVAHK	ILRGSVAHK	16144	J T Voeten; Clin Exp Immunol 2001
H3N2 CS 11 (i)	FEDLRLLSFIRGTKV	VWMACHSAAFEDLRLLSF	1598805	Laurel Yong-Hwa Lee; J Clin Invest 2008
H3N2 CS 11 (ii)	VQIASNENM	GKLSTRGVQIASNENMDN	1598807	Laurel Yong-Hwa Lee; J Clin Invest 2008

5.4 Optimization of protocol for PBMC Proliferation Assay at preliminary steps (MTT Assay)

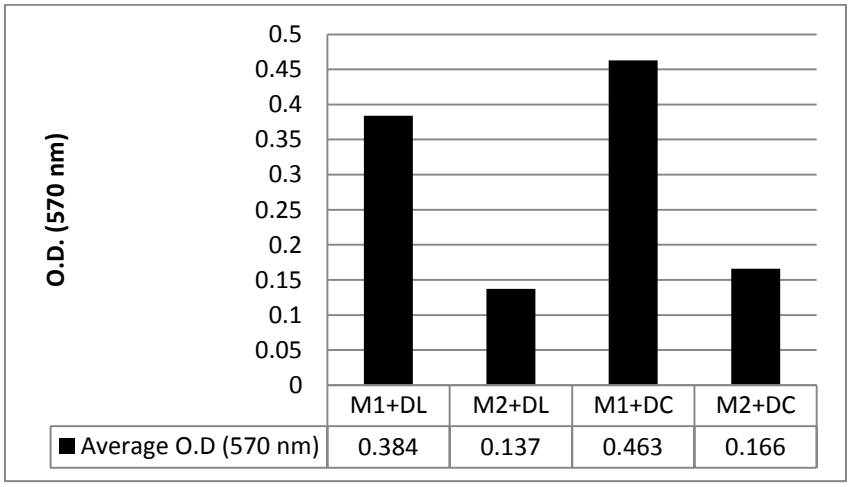
5.4.1 Optimization of protocol for Media and DMSO (solvent) to be used.

For preliminary optimization two types of DMEM medium were used, one with phenol red and other without phenol red. Two different types of DMSO were also used, one of laboratory grade and the other one which is used for cell freezing.

TABLE 23: Absorbance for Media and DMSO

Media	Average O.D (570 nm)
M_1+D_L	0.384
M_2+D_L	0.137
M_1+D_C	0.463
M_2+D_C	0.166

M_1- DMEM (with phenol red), M_2-DMEM (without phenol red)
D_L- Laboratory grade DMSO, D_C-Cell Freezing DMSO



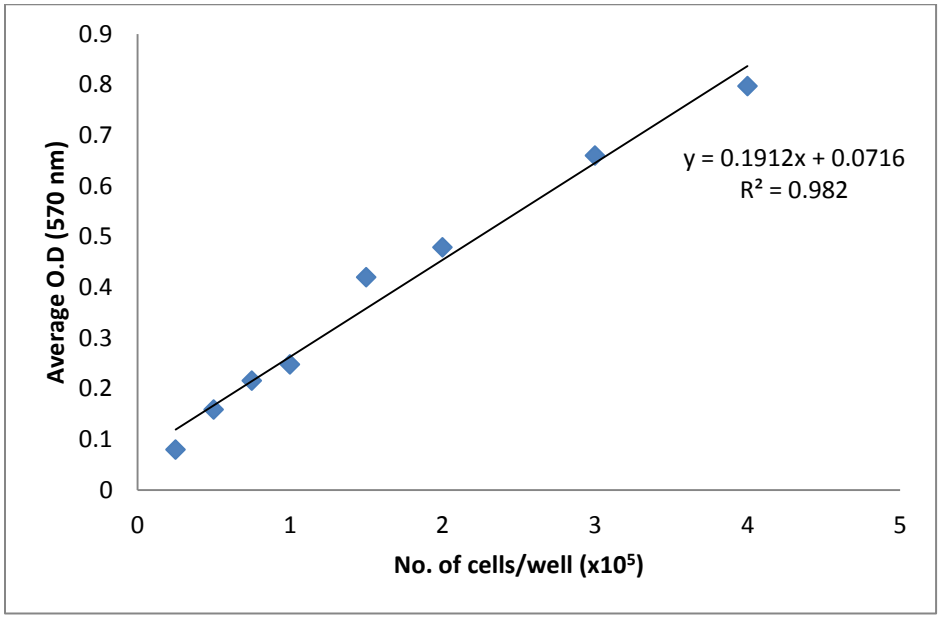
It was observed that DMEM with phenol red and cell freezing DMSO interferes in measurement of absorbance. It is inferred that phenol red interferes with the absorbance in the medium (with phenol red) thus subsequent assay was performed by using DMEM medium without phenol red. It was also observed that cell freezing DMSO interferes with absorbance as it is not in a pure solvent form and contains media components and fetal bovine serum.

5.4.2 Optimization of protocol for MTT cell proliferation assay

For optimization of protocol for MTT cell proliferation assay a range of cells/ well was taken from $0.25x10^5$ to $4x10^5$. In MTT assay O.D. is proportional to the formation of formazan crystals from MTT which depends on the number of cells. So as the cell number increases formation of formazan increases and so is the O.D.

TABLE 24: Absorbance for MTT cell proliferation assay

No. of cells/well	Average O.D (570 nm)
$4x10^5$	0.797
$3x10^5$	0.660
$2x10^5$	0.479
$1.5x10^5$	0.420
$1x10^5$	0.248
$0.75x10^5$	0.216
$0.5x10^5$	0.159
$0.25x10^5$	0.080



The graph shows linear relationship between number of cells /well and O.D. at 570 nm, In MTT assay, O.D. is proportional to the formation of formazan crystals from MTT which depends on the number of live cells. Hence it shows that MTT assay is working well.

CONCLUSION

We found that that nucleoprotein sequence show a high level of conservancy in case of both H1N1 and H3N2. Four immunogenic peptide of H1N1 NP: (i) EIRASVG RMIGGIGRFYIQM CTELKL (26 amino acid), (ii) TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAGAA (36 amino acid), (iii) AYERMCNILKGKFQTA (16 amino acid), (iv) FLARSALILRGSVAHKSC LPACVY (24 amino acid) and immunogenic peptide regions of H3N2 NP: (i) KMIDGIGR FYIQMCTELKL (19 amino acid), (ii) TYQRTRALVRTGMDPRMCSLMQGSTLPRRSGAAG AAVKGIGTMVMELI (48 amino acid), (iii) NAEIEDLIFLARSALILRGSVAHKS (25 amino acid), (iv) NPAHKSQLVWMACHSAAFEDLRLLSFIRGTKVSPRGKLSTRGVQIASN ENMDNMGSSTLELRSGYWAIRTRSG (72 amino acid), (v) IIRMMEGAKPEEVSFRGRGVF ELSDEKATPNIVPSF(36 amino acid) were found to be interesting as long stretch of immunogenic peptides were found ranging in length from 16 to 72 amino acid residues. These peptides covers large number of epitopes for both class I and class II MHC, as well as epitopes which are common for both MHC class I and class II. So it is inferred that these immunogenic peptides can be used as potential targets for epitope based vaccines.

These peptides can be used for further study to assess the affinity of immunogenic peptide to MHC molecule by structural analysis and molecular modeling. Further study can be carried out to assess the potential of these peptides for immunogenic response in the PBMC by T-cell proliferation assay (MTT assay) and cytokine production assay.

SUMMARY

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

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

Conventional influenza virus vaccines protect against one particular strain of influenza and thus it is not effective against novel Influenza viruses which emerged as a result of antigenic variation. So there is a requirement for a broad range vaccine. Epitope based vaccines imparts an efficient strategy for protection against antigenic variations of influenza viruses. Immune response in epitope based vaccine is directed against a specific stretch of protein sequence called epitopes. So if we find out the epitope conserved in different strains of influenza virus then it can be used as a vaccine target which will be effective against a broad range of influenza strains. Consequently, it will also protect against future (novel) strains as well. Nucleoprotein is considered to be a good candidate as it is a internal protein and it shows high level of conservancy.

Our approach was to find a stretch of immunogenic peptide from conserved peptide sequences of nucleoprotein of H1N1 and H3N2 strains of Influenza Virus using immunoinformatics tools. 17conserved regions were found in NP of H1N1. In these conserved regions of H1N1 16 MHC class I specific and 10 MHC class II specific immunogenic peptides were found. In these immunogenic peptides 6 immunogenic peptides were common for both classes of MHC. While in H3N2 13 conserved regions were found. In these conserved regions of H3N2 14 MHC class I specific and 13 MHC class II specific immunogenic peptides were found. In these immunogenic peptides 7 immunogenic peptides were common for both classes of MHC.

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

REFERENCES

Arnon R., (2006) “A novel approach to vaccine design: epitope-based vaccines” published in BTi

Betakova T., Nermut M.V., Hay A.J. (1996) “The NB protein is an integral component of the membrane of influenza B virus.” J Gen Virol, 77:21689-94.

Chen J. and Deng Y.M. (2009) “Influenza virus antigenic variation, host antibody production and new approach to control epidemics” Virology Journal 2009

Dimitrov I., Garnev P., Flower D.R., and Doytchinova I. (2010) “MHC Class II Binding Prediction—A Little Help from a Friend” Journal of Biomedicine and Biotechnology,Volume, Article ID 705821, 8 pages,

Edgar R.C. (2004) “MUSCLE: multiple sequence alignment with high accuracy and high throughput” Nucleic Acids Research, 2004, Vol. 32, No. 5,

Epstein S.L., Kong W.P., Misplon J.A., Lo C.Y., Tumpey T.M., et al. (2005) “Protection against multiple influenza A subtypes by vaccination with highly conserved nucleoprotein.” Vaccine 23: 5404–5410.

Huang Q., Sivaramakrishna R.P., Ludwig K., Korte T., Bottcher C., Herrmann A. (2003) “Early steps of the conformational change of influenza virus hemagglutinin to a fusion active state: stability and energetics of the hemagglutinin.” Biochim Biophys Acta. ; 1614(1):3–13

Ishioka G.Y., Fikes J., Hermanson G., Livingston B., Crimi C., Qin M., del Guercio M.F., Oseroff C., Dahlberg C., Alexander J., Chesnut R.W. and Sette A. (1999) “Utilization of MHC class I transgenic mice for development of minigene DNA vaccines encoding multiple HLA-restricted CTL epitopes” J. Immunol. 162, 3915-3925.

JIN XIONG, 2006 “Essential Bioinformatics” Cambridge University Press, ISBN-13 978-0-521-60082-8

Khan A.M., Miotto O., Heiny A.T., Salmond J., Srinivasand K.N., Nascimento E., Marques E.T., Brusica V., Tan T.W., August J.T. (2006) "A systematic bioinformatics approach for selection of epitope based vaccine targets." Cell Immunol; 244(2): 141-147

Kindt T.J., Osborne B.A., Goldsby R.A., Freeman W.H. *et al.*, Kuby Immunology 6th Edition, ISBN 13: 978-1-4292-0211-4

Larsen M.V., Lundegaard C., Lamberth K., Buus S., Lund O., Nielsen M. (31-10-2007) "Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction" , BMC Bioinformatics, 8:424

Laver W.G., Bischofberger N., Webster R.G. (1999) "Disarming the flu viruses." Sci Am1999;280:78-87

Li M.L., Rao P., Krug R.M.. (2001) "The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits." EMBO J.; 20(8):2078–2086.

Londo D. R., Davis A. R. and Nayak, D. P. (1983) "Complete nucleotide sequence of the nucleoprotein gene of influenza B virus". Journal of Virology 47,642-648.

Nicole S.M., Bresee J.S., Shay D.K., Uyeki T.M., Cox N.J., Strikas R.A. (2006). "Prevention and Control of Influenza: Recommendations of the Advisory Committee on Immunization Practices (ACIP)". Morbidity and Mortality Weekly Report. Centers for Disease Control and Prevention.

Nielsen M. and Lund O.(2009) "NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction" BMC Bioinformatics, 10:296

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

Olsen L.R., Zhang G.L., Keskin D.B., Reinherz E.L., and Brusic V. (2011) "Conservation analysis of dengue virus T-cell epitope-based vaccine candidates using peptide block entropy"

Parker K.C., Bednarek M.A. and Coligan J.E. (1994) “Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains” J. Immunol., 152, 163-175.

Patterson K.D. and Pyle G.F. (1991) “The geography and mortality of the 1918 influenza pandemic” Bull Hist Med. 4-21

Pinto L.H., Lamb R.A. (2006) “The M2 proton channels of influenza A and B viruses.” J Biol Chem.; 281(14):8997–9000

Portela A., Digard P. (2002) “The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication.” J Gen Virol. ;83(Pt 4):723-34.,

Rammensee, Bachmann, Stevanovic (1997) “MHC ligands and peptide motifs” Landes Bioscience

Samji T. (2009) “Influenza A: Understanding the Viral Life Cycle”, Yale J Biol Med. 2009 December; 82(4): 153–159. PMCID: PMC2794490

Shannon C.E. (1948) “a mathematical theory of communication” Bell System Technical Journal

Shu L. L., Bean W. J. and Webster R. G. (1993) “Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990.” Journal of Virology 67, 2723-2729.

Singh H. and Raghava G. P. S. (2001) “ProPred: prediction of HLA-DR binding site” BIOINFORMATICS APPLICATIONS NOTE, Vol. 17 no. 12, Pages 1236–1237

Stanekova Z. and Vareckova E. (2010) “Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development” Virol J.; 7: 351

Townsend A.R.M., Rothbard J., Gotch F.M., Bahadur G., Wraith D., and Mcklichael A.J. (1986) “The Epitopes of Influenza Nucleoprotein Recognized by Cytotoxic T Lymphocytes Can Be Defined with Short Synthetic Peptides” Cell, Vol. 44, 959-968.

Vladimir T., Hossein K., Raul R. (2009). "Geographic Dependence, Surveillance, and Origins of the 2009 Influenza A (H1N1) Virus". New England Journal of Medicine361 (2): 115–119

van der Burg S.H., Visseren M.J., Brandt R.M., Kast W.M. and Melief C.J. (1996) “Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability” J. Immunol. 156, 3308-3314.

Viret C. and Janeway C.A. Jr. (1999) “MHC and T cell development” Rev. Immunogenet. 1, 91-104.

WHO,Weekly epidemiological record, 2002, 77th YEAR ,No. 28, 229–240

www.merckmanuals.com/home/infections/viral_infections/influenza.html

www.ncbi.nlm.nih.gov/genomes/FLU

www.virology.ws/2009/09/22/the-a-b-and-c-of-influenza-virus/

www.who.int/topics/influenza/en/

Yongqun He, Rappuoli R., De Grootand A.S., Chen R.T. (2010) “Emerging Vaccine Informatics”, Hindawi Publishing Corporation,Journal of Biomedicine and Biotechnology

Zhao B., Sakharkar K.R., Lim C.S., Kangueane P., Sakharkar M.K. (2007) “MHC-Peptide binding prediction for epitope based vaccine design” International Journal of Integrated Biology, IJIB, Vol. 1, No. 2, 127-140