IMMUNOGENIC PEPTIDE PREDICTION OF MATRIX PROTEIN 1 (M1) IN H1N1 AND H3N2 STRAINS OF INFLUENZA VIRUS A

A Thesis submitted in partial fulfillment of the requirements

for the award of the degree of

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

Submitted By

Rameez Hasan
Registration No. 301001020

Under the Supervision of
DR. MANOJ BARANWAL

DEPARTMENT OF BIOTECHNOLOGY & ENVIRONMENTAL SCIENCES
THAPAR UNIVERSITY, PATIALA

JULY, 2012

CERTIFICATE

This is to certify that the thesis entitled "**Immunogenic peptide prediction of Matrix protein 1 (M1) in H1N1 and H3N2 strains of Influenza Virus A**" being submitted by Rameez Hasan, Roll No. 301001020 in partial fulfillment of the requirements for the award of degree of Master of Science in Biotechnology, Department of Biotechnology and Environmental Sciences, Thapar University, Patiala, is a bonafide work carried out under my supervision and guidance. The thesis has not been submitted for award of any other degree or certificate in this or any other university.

Dr. Manoj Baranwal
Supervisor
DBTES
Thapar University, Patiala

Dr. M. S. Reddy
Head
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 thesis entitled "**Immunogenic peptide prediction of Matrix protein 1 (M1) in H1N1 and H3N2 strains of Influenza Virus A**" in partial fulfillment of the requirements for the award of degree of Master of 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 thesis has not been submitted for award of any other degree or certificate in this or any other university.

Place: Patiala

Date: 16/7/2012

Rameez Hasan

Roll No. 301001020

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
DBTES
Thapar University, Patiala

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

ACKNOWLEDGEMENT

I thank the Almighty whose blessings have enabled me to accomplish my dissertation work successfully. It is my pride privilege to express deep sense of gratitude and indebtedness to **Dr. Manoj Baranwal**, Assistant Professor, Department of Biotechnology and Environmental Sciences, Thapar University for his valuable advice, splendid supervision and constant patience throughout this work. His constant encouragement and confidence-imbibing attitude has always been a moral support for me throughout the project work.

I wish to express my deep sense of humbleness to **Dr. M. S. Reddy**, Head of the Department of Biotechnology and Environmental Sciences, Thapar University, Patiala (Punjab) for his support.

I express my esteem and profound sense of gratitude to research scholar Ms. Neha Lohia for her help and cooperation during project work. My sincere thanks are due to Laboratory assistants of DBTES, Mr. Babban and Mr. Iqbal for their time to time help.

I feel lacunae of words to express my most heartfelt and cordial thanks to all my friends who have always been a source of inspiration for me, stood by my side at the toughest times.

The whole credit of my achievements goes to my parents who were always there for me. It was their unshakable faith in me, which helped me to proceed further. I have always banked upon their moral, emotional, un-stinted support and unparalleled gestures. Knowing your support and love is behind me has made this possible. I am grateful to have continued enthusiasm in my pursuit of happiness in my career.

Place: Patiala **Rameez Hasan**

ABSTRACT

H1N1 and H3N2 strains of Influenza A viruses belong to one of the best studied viruses; however no effective prevention against influenza has been developed. Current influenza virus vaccines protect mostly against one particular strain thus regular immunization with updated formulations is necessary against the virus. Hence great challenge in the field of influenza virus research is to design universal vaccine. MUSCLE and AVANA tools were used to find out most conserved peptide sequences of M1. Immunoinformatics tools were used for prediction of immunogenic peptides of Matrix Protein 1 (M1) in H1N1 and H3N2 strains of influenza virus A. Putative epitopes for M1 were predicted from conserved peptide sequences of M1. Three tools NetCTL 1.2, BIMAS and Syfpeithi were used to predict the Class I putative epitopes while three tools, ProPred, IEDB-SMM-align and NetMHCII 2.2 were used to predict the Class II putative epitopes. Immunogenic peptides were identified and selected manually by overlapping putative epitopes predicted from online tools individually for both MHC classes. Finally sequences of predicted peptides for both MHC classes were looked for common region which was selected as common immunogenic peptide. Two common immunogenic peptides were found for M1 in H1N1: (i) MEWLKTRPILS (43-53) and (ii) IRHENRMVLASTTAKAM (173-189) while four common immunogenic peptides were found for M1 in H3N2: (i) VKLYRKLKR (97-105), (ii) FHGAKEIALSYSAGALASC (109-127), (iii) LIYNRMGAVTTEVAFGLVCA (130-149) and (iv) VLASTTAKA (180-188). These predicted peptides are promising candidates to be used as target for vaccine design. For establishing immunogenicity of predicted peptides T cell proliferation assay is need to be performed and for that we have optimized the protocol for PBMC proliferation assay (MTT Assay).

ABBREVIATIONS

RNA	-	Ribonucleic acid
WHO	-	World Health Organization
CDC	-	Centre for Disease Control and Prevention
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
NS	-	Non- Structural protein
IFN	-	Interferon
vRNA	-	Negative sense RNA
PA, PB	-	Polymerase protein A, B
HEF	-	Hemagglutinin-esterase-fusion
S-OIV	-	Swine origin influenza virus
FDA	-	Food and Drug Administration
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
TAP	-	Transporter Associated With Antigen Processing
SVM	-	Support Vector Machine
MOT	-	Matrix Optimization Techniques
PFR	-	Peptide Flanking Residue
DMEM	-	Dulbecco's Modified Eagle Medium
PBS	-	Phosphate Buffer Saline
EDTA	-	Ethylenediaminetetraaceticacid

CONTENTS

1. INTRODUCTION

Influenza viruses are negative sense, single-strand RNA viruses. Influenza viruses belong to the family of Orthomyxoviridae (Group V). Influenza is a major cause of sickness and death around the world and is one of the most important infectious diseases confronting the world today.

Human influenza virus can evolve rapidly by one of two mechanisms: (i) antigenic shift (genetic reassortment between a human and a non-human virus in a non-human host), and (ii) antigenic drift (accumulation of mutations that facilitate evasion of the host immune response). New influenza viruses constantly emerge from the environment, emanating from such disparate sources as migratory waterfowl, swine, domestic poultry, and sea mammals.

Between major epidemics, mutations produce minor but cumulative antigenic variants in a wide number of animal reservoirs. Major new viral subtypes have appeared in 1918, 1957, 1977, 1968 and 2009, in each case resulting in a pandemic (Bao Y. *et al*., 2008). The term "pandemic" literally means a worldwide epidemic, but for influenza there is also the assumption that a major subtype involving new hemagglutinin (HA) and/or neuraminidase (NA) antigens has appeared. These shifts are much more dramatic than the normal slow genetic drift of other subtypes. As a result, there is no opportunity for people to develop immunity to the new virus thus allowing the virus to spread faster.

Influenza virus infection can be effectively treated with anti-influenza antiviral agents (Ong A.K. *et al*., 2007) as it was during the rapid emergence of the pandemic H1N1 influenza virus in 2009 when no vaccine was available. At present, two classes of antiviral drugs are approved for influenza therapy: M2 ion channel blockers (oral amantadine and its derivative rimantadine) and neuraminidase inhibitors (oral oseltamivir and inhaled zanamivir). These agents either prevent viral uncoating inside the cell (M2-blockers) or prevent the release of progeny virions from infected cells (NA inhibitors) (Moscona A., 2008). The development of resistance is a major obstacle to the usefulness of both classes of anti-influenza agents, which could provide a growth advantage to variants carrying drug-resistance mutations, more likely under drug selective pressure.

An effective vaccine is always the best choice to control the spread of the influenza pandemic. 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 viruses. The influenza virus that affects humans mutates easily and results into new antigenic variants, hence it requires the inclusion of such variants in existing vaccine to ensure effective immunization of the population. In spite of the availability of antiviral and inactivated trivalent vaccines, which are effective for most recipients, Influenza remains a serious respiratory disease.

Therefore is a call for development of a vaccine, which would be protective against different virus strains and would not need to be updated every year.

Peptide based vaccines in which small peptides derived from target proteins, are used to provoke an immune reaction have attracted considerable attention recently as a potential means both of treating infectious diseases and promoting the destruction of cancerous cells by a patient's own immune system. With the availability of large sequence databases and supercomputers, rapid processing of large numbers of peptides is possible. Computer aided design of peptide based vaccines has emerged as a promising approach for screening of billions of possible immuneactive peptides. It helps to find those peptides which are likely to provoke an immune response to a particular cell type. The vaccine, activating both the humoral and cellular arms of the immune response, induces long- lasting protection against many strains of the influenza virus. Consequently, it is expected to protect against future strains as well.

Vaccine informatics is an emerging research area that focuses on development and applications of bioinformatics methods that can be used to facilitate area of vaccine research. Many immunoinformatics algorithms and resources have been developed to predict T and B cell epitopes and to study immunological aspects. Systematic transcriptomics and proteomics gene expression analyses facilitate rational vaccine design and identification of gene responses that are correlates of protection in vivo. Mathematical simulations have been used to model host-pathogen interactions and improve vaccine production and vaccination protocols. Computational methods have also been used for development of immunization registries or immunization information systems, assessment of vaccine safety and efficacy, and immunization modeling.

The immune response to attenuated intact viruses and subunit vaccines is to a very large degree dependent on T-cell recognition of peptide epitopes bound to MHC. Thus targeting antigens that contain many $CD4^+$ T helper epitopes may lead to the selection of good B-cell antigens as well as immunogens for effective CD8 responses— this is because $CD4^+$ T helper cells are critically important to the development of memory B-cell (antibody) and memory CTL (cytotoxic T-cell) responses, in addition to being active against pathogens on their own. T helper cells have been called the "conductors of the immune system orchestra" (Ahlers J.D *et al*., 2001). CTLs generally play a role in the containment of viral and bacterial infection (Plotnicky H. *et al*., 2003), and the prevalence of CTLs usually correlates with the rate of pathogen clearance.

Important requirement for this approach is the identification and selection of T-cell epitopes that act as vaccine targets. T cells can recognize antigen only when it is presented by a group of specialized proteins known as HLAs or MHCs, class I and class II on the surface of antigen presenting cells (Lin H.H. *et al*., 2008). Since the experimental methods to detect epitopes are expensive and time consuming so various computational tools can be employed to facilitate the process of epitope detection by reducing this experimental effort (Korber B. *et al.*, 2006).

In this study, our approach is to identify immunogenic peptides containing T cell epitopes in conserved peptide sequences of Matrix protein 1(M1) viral protein that can act as a candidate for vaccine design.

2. REVIEW OF LITERATURE

2.1 Influenza Virus

2.1.1 An Introduction

The viruses that lead to influenza are members of the Orthomyxoviridae family (Group V), which consists of enveloped viruses with a single, segmented, negative-strand ribonucleic acid (RNA) genome. According to the antigenic difference between their nucleoproteins (NP) and matrix proteins (M), influenza viruses can be classified into three types namely A, B, and C. Although these viruses possess similar proteins, each virus encodes proteins by different methods. The unique genetics of influenza A has led to additional divisions into subtypes on the basis of the hemagglutinin (HA) and neuraminidase (NA) antigens present in a given virus (Bull World Health Organ, 1980). Influenza A differs from influenza B and C in its ability to infect multiple species, including avian and mammalian species. Humans, horses, and swine have been infected most frequently. Influenza B infects only humans. In addition, the HA and NA proteins derived from influenza A viruses have greater sequence variability than influenza B in their amino acids (Lamb R.A. and Krug R.M., 2001). Error prone RNA dependent RNA polymerases and segmented genome influenza viruses 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. *et al.*, 2010).

Characteristics of Influenza viruses

Feature	Properties
Virion	Spherical, pleiomorphic
Particle size	80-120 nm
Composition	RNA(1%), Protein(73%), Lipid(20%), Carbohydrate (6%)
Inner Ribonucleoprotein helix	9 nm in diameter
RNA in nucleocapsid	RNase sensitive
Fusion of virus with cell	Host Endosomal membrane
Transcription of viral RNA	Nucleus of host cell
Genetic reassortment	Frequent
Rate of antigenic change	High

Adapted from Chapter 39. Orthomyxoviruses (Influenza Viruses), Section IV. Virology, Jawetz, Melnick, & Adelberg's Medical Microbiology, 25e

2.1.2 Taxonomy

Influenza viruses are negative sense, single-strand RNA viruses. Influenza viruses belong to the family of Orthomyxoviridae which comprises influenza A, B, C, Isavirus and Thogotovirus, based on differences in their nucleoprotein (NP) and matrix protein (M) (Cheung T.K.W. and Poon L.L.M., 2007). Influenza A viruses are enveloped, pleomorphic, and contain genomes of 8 single-stranded negative-sense segments of RNA. Influenza viruses have three key structural proteins: hemagglutinin (HA), neuraminidase (NA) and Matrix 2 (M2). Both HA and NA are surface glycoproteins diverse enough that their serological recognition gives rise to the traditional classification into different subtypes. At present, 17 subtypes of HA (H1-H17) (Rivailler P. *et al*., 2012) and 9 subtypes of NA (N1-N9) have been identified. Three subtypes of type A influenza viruses, namely H1N1, H2N2 and H3N2, are able to transmit in humans (Parrish C.R. and Kawaoka Y., 2005; Taubenberger J.K. *et al*., 2007).

2.1.3 Types of Influenza Virus

Based on antigenic features of M1 and NP proteins

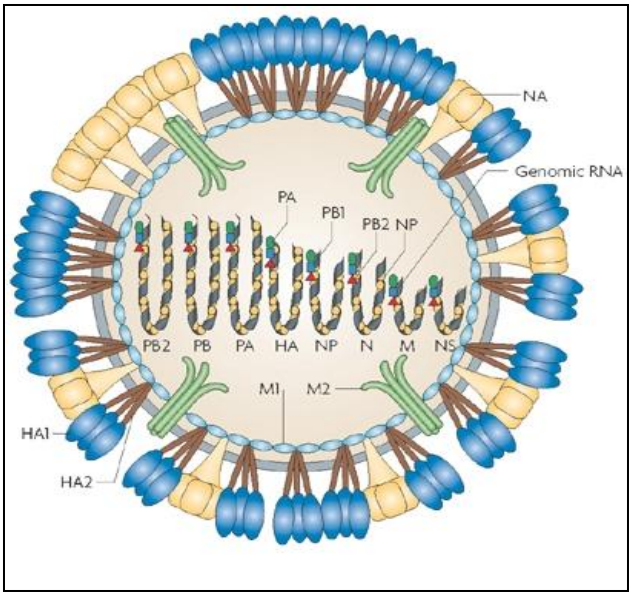
Influenza Type A: It infects humans, birds, pigs, horses and other animals; its natural hosts are wild birds; they are classified by subtype (based on HA and NA proteins) and strains.

Influenza Type B: They are usually found in humans only; they are classified by strain (not subtype); they are associated with less severe epidemics than A, but never with pandemics.

Influenza Type C: They cause mild illness in humans, but don't cause epidemics or pandemics.

2.1.4 Structure of Influenza Virus

Its genome consists of 8 single stranded negative-sense segments of RNA and the virus particle contains a lipid envelope which is acquired from the host cell membrane during the viral budding process (Cheung T.K.W. and Poon L.L.M., 2007; Webster R.G. *et al*., 1992; Lamb R.A. and Krug R.M., 2001). Three proteins: hemagglutinin (HA), neuraminidase (NA) and Matrix (M2) are included in its envelope, two of which, HA and NA, are surface glycoproteins. The M1 protein forms a layer to separate the ribonucleoprotein (RNP) from the viral membrane and interacts with the viral RNA, vRNA (Negative sense RNA), and the RNP protein components during the assembly and disassembly of the influenza A virus.



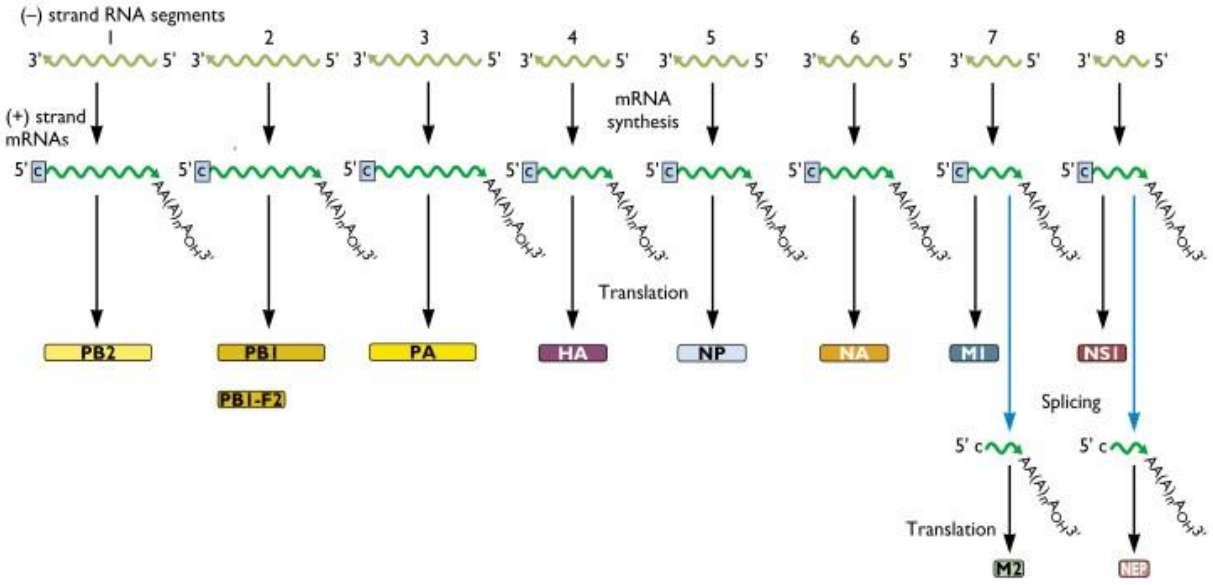
Adapted From Nature Reviews Microbiology 6, 143-155 (February 2008)

Functions of various influenza virus proteins

	Protein	**Function**
PA, PB 1, PB2	Polymerase Proteins	Virus replication
HA	Hemagglutinin	Viral attachment, antigenic determinant
NP	Nucleoprotein	RNA coating, nuclear targeting, RNA transcription
NA	Neuraminidase	Antigenic determinant, viral release from host cell
M1	Matrix Protein 1	Stability, viral assembly and disassembly, RNP trafficking
M2	Matrix Protein 2	Viral uncoating
NS	Non-structural Proteins	Regulation of virus life cycle, mRNA transcription and localization of viral RNP

Adapted from Lamb R.A. and Krug R.M., 2001

2.1.5 Genome and life cycle of influenza virus



Adapted from http://www.virology.ws/2009/05/01/influenza-virus-rna-genome/

The life cycle of the influenza A virus is initiated by the binding of HA to the sialic acid receptor on the host cell surface. The sialic acid linkage to galactose by α2-3 or α2-6 is the main factor that determines host specificity (Hay A.J., 1998). The virus enters the host cell by clathrin-coated receptor mediated endocytosis and the vesicle harboring the whole virus then fuses with endosomes. The acidic environment of the endosome triggers a conformational change of HA, allowing the fusion of HA with the vesicular membrane. The proton influx mediated by the M2 ion channel induces dissociation of the M1 from the vRNP, resulting in uncoating of the virus. The viral RNPs are translocated to the host cell nucleus where transcription and replication occur.

vRNAs serve as templates for both cRNA and mRNA. PB2 polymerase recognizes and binds to the 5'end of host mRNA and cleaves it 10-15 nucleotides away from the cap structure. The viral polymerase complex uses the cleaved cap structure as primers for viral mRNA transcription (Plotch S.J. *et al.*, 1981). Viral mRNA that is capped and polyadenylated is then exported to the cytoplasm. The mRNAs encoding viral membrane proteins such as HA, NA, and M2 are translated by ribosomes in the rough endoplasmic reticulum.

Subsequently, these proteins pass to the secretory pathway at the trans-Golgi for glycosylation, then move to the host cell surface and are incorporated into the host cell membrane (Hay A.J., 1998). The mRNAs encoding the other proteins, PB2, PB1, PA, NP, NS1, and NS2 are translated by free ribosomes in the cytosol. PB2, PB1, PA, and NP are then transported back into the nucleus and initiate complementary RNA (cRNA) transcription using vRNAs as templates. As a full length copy of the vRNA, cRNA serves as a template for the synthesis of progeny virus genomes and newly synthesized progeny vRNA forms vRNP binding proteins. M1 proteins, which contain nuclear localization domains, then bind to vRNP and promote the nuclear export of vRNPs. NS2 protein is thought to be involved in enhancing the function of M1 protein.

vRNPs are transported into the cell membrane where the envelope proteins are located and are composed of viral particles. When newly assembled virus particles bud off from the host cell membrane, NA cleaves the sialic acid receptors on the host cell and allows virus particles to leave the host cell (Hay A.J., 1998).

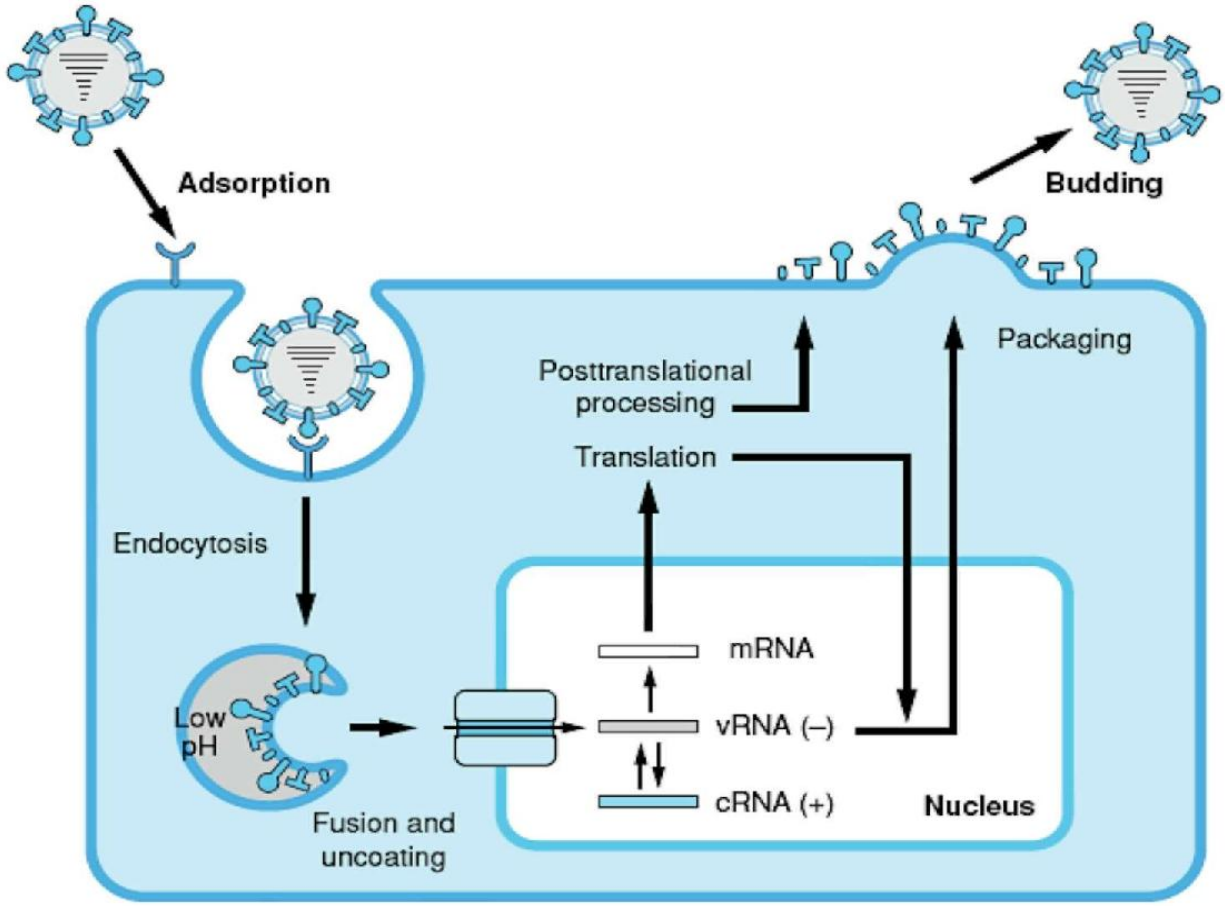


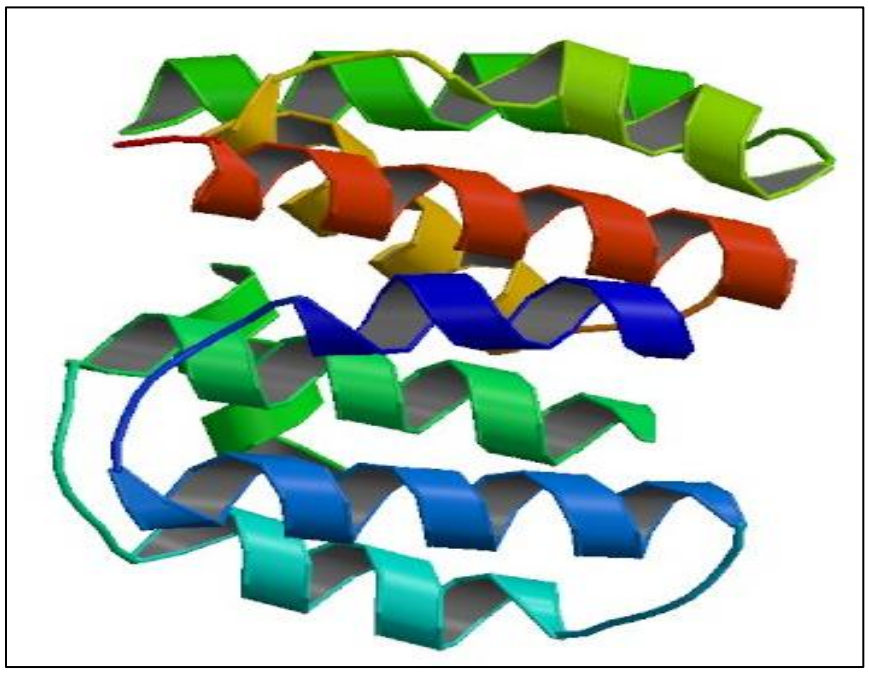
Illustration of the influenza virus replication cycle

Adapted from Fields Virology; Palese and Shaw, 2007

2.1.6 Role of Matrix Protein 1 (M1)

Matrix protein 1 (M1) plays critical roles in virus replication, virus entry and uncoating to assembly and budding of the virus particle. M1 binds to ribonucleocapsids (RNPs) in nucleus and seems to inhibit viral transcription. Interaction of viral NEP with M1-RNP is thought to promote nuclear export of the complex, which is targeted to the virion assembly site at the apical plasma membrane in polarized epithelial cells. Interactions with NA and HA may bring M1, a non-raft-associated protein, into lipid rafts. It forms a continuous shell on the inner side of the lipid bilayer in virion, where it binds to RNP. During virus entry into cell, the M2 ion channel acidifies the internal virion core, inducing M1 dissociation from the RNP. M1-free RNPs are transported to the nucleus, where viral transcription and replication can take place (Huang X. *et al*., 2001).

M1 determines the virion's shape to be spherical or filamentous. Filamentous virions are thought to be important to infect neighboring cells, and spherical virions are more suited to spread through aerosol between host's organisms. Thus M1 plays an important role in infectivity by determining shape of the virion (Huang X. *et al*., 2001).

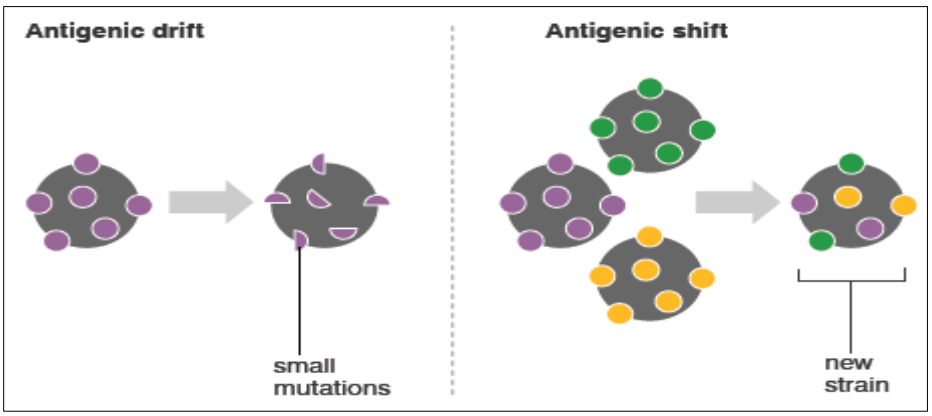


Crystal structure of the Matrix protein 1 (M1) from influenza A virus (A/California/04/2009 H1N1) PDB ID: 3MD2

2.1.7 Antigenic Drift and Antigenic Shift of the Influenza Virus

Influenza virus has a remarkable ability in escaping host defense mechanisms by altering its antigenic character. Due to the error-prone activity of the viral RdRp (RNA dependent RNA ploymerase), a high mutation rate of $\geq 5 \times 10^5$ nucleotide changes per nucleotide and replication cycle, thus approaching almost one nucleotide exchange per genome per replication, is observed among the influenza viruses (Drake J.W., 1993). In case selective pressures (such as neutralising antibodies, suboptimal receptor binding or chemical antivirals) are acting during viral replication on a host or population scale, mutants with corresponding selective advantages (e.g. escape from neutralisation, reshaped receptor-binding units) may be singled out and become the dominant variant within the viral quasispecies (group of viruses related by a similar mutation or mutations, competing within a highly mutagenic environment) in that host or population. Antigenic determinants of the membrane glycoproteins HA and NA are affected by mechanisms driven by immunity, such a (gradual) process is referred to as antigenic drift (Ferguson N.M. *et al.*, 2003).

Antigenic shift, in contrast, denotes a sudden and profound change in antigenic determinants, i.e. a switch of H and/or N subtypes, within a single replication cycle. This occurs in a cell which is simultaneously infected by two or more influenza A viruses of different subtypes. The distribution of replicated viral genomic segments into budding virus progeny occurs independently from the subtype origin of each segment, and due to this replication competent progeny carrying genetic information of different parental viruses (so-called reassortants) may spring up (Webster R.G. *et al.*, 2004). While the pandemic human influenza viruses of 1957 (H2N2) and 1968 (H3N2) clearly arose through reassortment between human and avian viruses, the influenza virus causing the Spanish flu in 1918 appears to be entirely derived from an avian source (Belshe R.B., 2005).



2.2 Influenza Treatment

2.2.1 Antiviral Drugs

2.2.1.1 Neuraminidase Inhibitors

1) Oseltamivir:- Oseltamivir is a potent and selective inhibitor of the neuraminidase enzyme of the influenza viruses A and B. The neuraminidase enzyme is responsible for cleaving sialic acid residues on newly formed virions and plays an essential role in the release and spread of progeny virions. When exposed to oseltamivir, the influenza virions aggregate on the surface of the host cell, thereby limiting the extent of infection within the mucosal secretions and reducing viral infectivity (McNicholl I.R. and McNicholl J.J., 2001).

2) Zanamivir:- Zanamivir is an orally inhaled powder currently approved in 19 countries for the treatment of, and in two for the prophylaxis of influenza A and B. Zanamivir is a competitive inhibitor of the neuraminidase glycoprotein, which is essential in the infective cycle of influenza viruses. It closely mimics sialic acid, the natural substrate of the neuraminidase (Varghese J.N. *et al.,* 1992).

2.2.1.2 M2 Ion Channel Inhibitors

1) Amantadine:- Amantadine inhibits the replication of influenza A viruses by interfering with the uncoating of the virus inside the cell. It is an M2 inhibitor which blocks the ion channel formed by the M2 protein that spans the viral membrane (Hay A.J. *et al.,* 1985; Sugrue R.J. and Hay A.J., 1991; Bui M. *et al*., 1996). The influenza virus enters its host cell by receptor-mediated endocytosis. Thereafter, acidification of the endocytotic vesicles is required for the dissociation of the M1 protein from the ribonucleoprotein complexes.

2) Rimantadine:- Rimantadine is an M2 ion channel inhibitor which specifically inhibits the replication of influenza A viruses by interfering with the uncoating process of the virus. M2 inhibitors block the ion channel formed by the M2 protein that spans the viral membrane (Hay A.J. *et al.,* 1985; Sugrue R.J. and Hay A.J., 1991; Bui M. *et al*., 1996). The influenza virus enters its host cell by receptor-mediated endocytosis. Thereafter, acidification of the endocytotic vesicles is required for the dissociation of the M1 protein from the ribonucleoprotein complexes.

2.2.2 Vaccines

Vaccines are apathogenic entities that cause the immune system to respond in such a way, that when it encounters the specific pathogen represented by the vaccine, it is able to recognize vaccine and mount a protective immune response, even though the body may not have encountered that particular pathogen before.

Types of Influenza Vaccine

Killed vaccines

Killed virus vaccines can be divided into whole virus vaccines, and split or subunit vaccines.

Whole virus vaccines were the first to be developed. The influenza virus was grown in the allantoic sac of embryonated hen's eggs, subsequently purified and concentrated using red blood cells, and finally, inactivated using formaldehyde or β-propiolactone. Later, this method of purification and concentration was replaced with centrifuge purification, and then by density gradient centrifugation, where virus particles of a specific density precipitate at a certain level in a solution of increasing density. Subsequently, filter-membrane purification was added to the methods available for purification/concentration (Hilleman M.R., 2002; Potter C.W., 2005).

Split vaccines for influenza are produced in the same way as whole virus vaccines, but virus particles are disrupted using detergents.

Subunit vaccines consist of purified HA and NA proteins, with the other viral components removed. Split and subunit vaccines cause fewer local reactions than whole virus vaccines, and a single dose produces adequate antibody levels in a population exposed to similar viruses (Couch R.B. *et al.*, 1997; Hilleman M.R., 2002; Potter C.W., 2005). However, this might not be sufficient if a novel pandemic influenza virus emerges, and it is believed that two doses will be required.

Inactivated influenza virus vaccines are generally administered intramuscularly, although intradermal (Belshe R.B. *et al.*, 2004; Cooper C.L. *et al.*, 2004; Kenney R.T. *et al.*, 2004) and intranasal (mucosal) routes are being investigated (Langley J.M. *et al.*, 2005).

Live vaccines

Cold-adapted live attenuated influenza virus (CAIV) vaccines, for intranasal administration, have been available in the USA since July 2003, and in the former Soviet Union, live attenuated influenza vaccines have been in use for several years. The vaccine consists of a master attenuated virus into which the HA and NA genes have been inserted. The master viruses used are A/Ann Arbor/6/60 (H2N2) and B/Ann Arbor/1/66 (Hoffman E. *et al.*, 2005; Palese P. *et al.*, 1997). The vaccine master virus is cold-adapted. In other words, it has been adapted to grow ideally at 25°C, which means that at normal human body temperature, it is attenuated. The adaptation process has been shown to have caused stable mutations in the three polymerase genes of the virus, namely PA, PB1, and PB2 (Hilleman M.R., 2002).

The advantages of a live virus vaccine applied to the nasal mucosa are the development of local neutralising immunity, the development of a cell-mediated immune response, and a cross-reactive and longer lasting immune response (Couch R.B. *et al.*, 1997).

Major concern in the CAIV vaccine, is its use in immunocompromised patients and the possible interference between viral strains present in the vaccine which might result in decreased effectiveness. They may damage mucosal surfaces, while far less than with wild type virulent influenza viruses, may lead to susceptibility to secondary infections. The greater concern for the future is the possibility of genetic reversion where the mutations causing attenuation change back to their wild-type state and reassortment with wild-type influenza viruses, resulting in a new strain. However, studies done to test these vaccines have not detected problems so far (Youngner J.S. *et al.*, 1994).

Vaccines and technology in development

It is hoped that cell culture, using Madin-Darby Canine Kidney (MDCK) or Vero (African green monkey kidney) cells approved for human vaccine production, may eventually replace the use of hen’s eggs, resulting in a greater production capacity, and a less labor-intensive culturing process. However, setting up such a facility takes time and is costly, and most vaccine producers are only now beginning this process.

Reverse genetics allows for specific manipulation of the influenza genome, exchanging genome segments for those desired (Palese P. *et al.*, 1997). Based on this method, several plasmid-based methods for constructing new viruses for vaccines have been developed, but are not yet in use commercially (Neumann G. *et al.*, 2005). A number of plasmids, small circular pieces of DNA, containing the genes and promoter regions of the influenza virus, are transfected into cells, which are then capable of producing the viral genome segments and proteins to form a new viral particle. If this method could be used on a larger scale, it may simplify and speed up the development of new vaccines. instead of the cumbersome task, for the live attenuated vaccines, of allowing reassortment in eggs, and then searching for the correct reassortment (6 genes from the vaccine master strain, and HA and NA from the selected strain for the new vaccine), the vaccine producers could simply insert the HA and NA genes into a plasmid.

DNA vaccines have been tested for a variety of viral and bacterial pathogens. The principle upon which the vaccine works is inoculation with DNA, which is taken up by antigen presenting cells, allowing them to produce viral proteins in their cytosol. These are then detected by the immune system, resulting in both a humoral and cellular immune response (Hilleman M.R., 2002).

Synthetic Peptide Vaccines

Vaccines to conserved proteins have been considered, and among the candidates are the M1 (Masanori T. *et. al.*, 2008), M2 and the NP proteins. It is hoped that, by producing immunity to conserved proteins, i.e. proteins that do not undergo antigenic change like HA and NA do, a vaccine can be produced that does not need to be replaced each year. This is also on the WHO's agenda for a pandemic vaccine (Couch R. *et al.*, 2005). Such vaccines have been shown to be effective in laboratory animals, but data are not available for human studies. Generic HA-based vaccines, aimed at conserved areas in the protein, are also being considered (Palese P., 2002).

Synthesizing peptides for use as vaccines requires identification of those epitopes in the protein antigen that stimulate protective immunity. Both B and T cell epitopes must be included in the peptide so that arms of the immune system humoral and cell-mediated are stimulated.

2.3 Major Histocompatibility Complex

MHC molecules enable T lymphocytes to recognize epitopes of antigens and discriminate self from nonself. Unlike B cell receptors on B lymphocytes that are able to directly bind epitopes on antigens, the T cell receptors (TCRs) of T lymphocytes can only recognize epitopes after they are bound to MHC molecules. Without these, there would be no presentation of internal or external antigens to the T cells and other aspects of the immune response cannot occur. Accurate prediction of the binding between short peptides and the MHC molecules has long been a principal challenge for synthetic epitope based vaccine. Therefore it is very important to understand the concept of MHC (Liu W. *et al.*, 2006).

MHC I molecules are designed to enable the body to recognize infected cells and tumor cells and destroy them with cytotoxic T lymphocytes or CTLs. CTLs are effector defense cells derived from naive T8 lymphocytes ($CD8^+$).

MHC I molecules are (i) expressed in all nucleated cells in the body; (ii) possess a deep groove that can bind peptides, typically 8-11 amino acids long, from endogenous antigens and (iii) present peptide to naive T8 lymphocytes and cytotoxic T lymphocytes that have a complementary shaped TCR.

MHC II molecules are designed to enable T4 lymphocytes ($CD4^+$) to recognize epitopes of exogenous antigens and discriminate self from non self.

MHC II molecules are (i) expressed in antigen presenting cells (APCs), such as dendritic cells, macrophages, and B lymphocytes; (ii) possess a deep groove that can bind peptides, typically 10-30 amino acids long but with an optimum length of 12-16 amino acids, from exogenous antigens and (iii) present peptide to naive T4 lymphocytes or effector T4 lymphocytes that have a complementary shaped TCR.

The expression of MHC molecules is increased by cytokines produced during both innate immune responses and adaptive immune responses. Cytokines such as interferon α, interferon β, interferon γ, tumor necrosis factor increase the expression of MHC I molecules, while interferon gamma is the main cytokine to increase the expression of MHC II molecules.

2.4 Immunoinformatics

Immunoinformatics, or computational immunology, is an emerging area that provides fundamental methodologies in the study of immunomics, that is, immune-related genomics and proteomics. The integration of immunoinformatics with systems biology approaches may lead to a better understanding of immune-related diseases at various systems levels. Such methods can contribute to translational studies that bring scientific discoveries of the immune system into better clinical practice (Yan Q., 2010). Immunoinformatics is an emerging specialization of bioinformatics that focuses upon the structure, function and interactions of the molecules involved in immunity (Evans M.C., 2008). One of the key goals of immunoinformatics is the development of computer aided vaccine design (CAVD) (Flower D.R. and Doytchinova I.A., 2002).

Immunoinformatics and other computational methods may reduce the time and number of wet laboratory experiments to identify the peptide for vaccine targets. Computational immunology methods dramatically reduce the time and effort involved in screening potential epitopes. Its 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.

In the current study, immunoinformatics tools were employed to predict immunogenic peptides of HINI and H3N2 as target for vaccine design.

3. AIM OF THE STUDY

Current influenza virus vaccines protect mostly against homologous virus strains; thus, regular immunization with updated vaccine formulations is necessary to guard against the virus' hallmark remodeling of regions that mediate neutralization. Development of a broadly protective influenza vaccine would mark a significant advance in human infectious diseases research (Wang T.T. *et al*., 2010).

The virus-specific cytotoxic T lymphocyte (CTL) induction is an important target for the development of a broadly protective human influenza vaccine, since most CTL epitopes are found on internal viral proteins and relatively conserved (Ichihashi T. *et al*., 2011).

In this study, the possibility of developing a strain/subtype-independent human influenza vaccine was explored by taking an immunoinformatics approach to predict immunogenic regions of matrix protein that can be used as vaccine candidate.

Work plan included the following objectives:

(i) Finding conserved peptide regions of Matrix protein 1, from all the available strains of H1NI and H3N2 that has been sequenced.

(ii) Prediction of immunogenic peptides containing overlapping epitopes in conserved peptide sequences which are able to bind to multiple HLA molecules using various computational tools and can act as targets for vaccine design.

(iii) Isolation of peripheral blood mononuclear cells (PBMC) and culturing of PBMC for cell proliferation assay (MTT assay).

4. MATERIALS AND METHODS

4.1 Sequence Retrieval

The sequences of Matrix protein 1 of H1N1 and H3N2 were retrieved from NCBI Influenza database (http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/select.cgi) from January 1918 to December 2011. Full length sequences were taken and the identical sequences were collapsed using the option available in the database search engine. The sequences were downloaded in fasta format and opened with WordPad and then these sequences were transferred to a word file.

Some sequences have sometimes an invalid letter code "J" which does not represent any amino acid, and thus these sequences need to be corrected by replacing the "J" with "X".

Sequences of Matrix protein 1 of H1N1 and H3N2 strains of Influenza A virus

Matrix Protein 1	Total Sequences	Non Redundant Sequences
H1N1	5902	285
H3N2	4230	236

4.2 Conservancy Analysis

Two bioinformatics tools were used to find out the conserved regions from all retrieved sequences of Matrix protein 1 of H1N1 and H3N2 strains of influenza virus; MUSCLE to align the sequences and AVANA to find out the conserved regions in the aligned sequences.

4.2.1 MUSCLE (http://www.ebi.ac.uk/Tools/muscle/index.html)

MUSCLE stands for Multiple Sequence Comparison by Log-Expectation. It is one of the multiple sequence alignment tool provided by European Bioinformatics Institute (Edgar R.C., 2004). MUSCLE is claimed to achieve both better average accuracy and better speed than Clustal W2 or T-Coffee.

Elements of the algorithm include fast distance estimation, progressive alignment and refinement using tree-dependent restricted partitioning. MUSCLE uses two distance measures for a pair of sequences: a Kmer distance (for an unaligned pair) and the kimura distance (for an aligned pair).

Multiple alignments of protein sequences are important in many applications, including phylogenetic tree estimation, secondary structure prediction and critical residue identification. Many multiple sequence alignment (MSA) algorithms have been proposed. Two attributes of MSA programs are of primary importance to the user: biological accuracy and computational complexity (i.e., time and memory requirements). MUSCLE, provides significant improvements in both accuracy and speed.

4.2.2 AVANA

Antigen Variability Analyzer tool (AVANA) was subsequently used to extract alignments of several subsets of the collected sequences, based on annotation values, such as viral subtype, host, and year of isolation. This tool finds conserved regions based on information entropy analysis (Khan A.M. *et al.*, 2008). It also compares alignments using mutual information, identifying the mutations that characterize specific sequence sets.

Assuming that each sequence represents an independent isolate, the information entropy methodology was used to measure the variability of M1 of Influenza A virus in the context of overlapping nine-amino acid peptides spanning the length of each influenza protein. The rationale of this selection was the length of peptides that are bound by HLA molecule for presentation to T cell receptor, typically from 8-20 amino acids, with nine amino acids being the predominant length class I peptides and the core of class II peptides.

Applying Shannon's formula, the nonamer peptide entropy H(x) at any given position x in the alignment is computed by

$$H(x) = -\sum_{i=1}^{n(x)} p(i, x) \log_2 p(i, x)$$

where p(i, x) is the probability of a particular nonamer peptide i being centered at position x. The entropy value increases with n(x), the total number of peptides observed at position x; it is also sensitive to the relative frequency of the peptides; such that it decreases when one peptide is clearly dominant (i.e. the position is conserved). Sites which are highly conserved have lower entropy because entropy is degree of randomness.

The threshold for conservation was fixed to 80% and based on this threshold, the conserved region were obtained using AVANA for M1 protein. The next step is to find regions that are conserved in all the groups of a protein which was done manually. These sequences were then further analyzed to predict epitopes.

4.3 Epitope Prediction

Reliable predictions of Cytotoxic T lymphocyte (CTL) epitopes are essential for rational vaccine design. Reverse immunogenetic approaches attempt to optimize the selection of candidate epitopes and can minimize the experimental effort needed to identify epitopes.

Different algorithms based programs were used to predict T cell epitopes for both; Class I MHC ($CD8^+$) and Class II MHC ($CD4^+$).

4.3.1 Class I MHC ($CD8^+$ T-cell Binding Epitope Predictions)

Three programs, NetCTL 1.2 (Larsen M.V. *et al*., 2005), BIMAS (Parker K. *et al*., 1994) and Syfpeithi (Rammensee H. *et al*., 1999) were used to predict the nonamers because a length of nine amino acid represents the typical length of peptide that binds to Class I HLA molecules.

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

NetCTL 1.2 is a tool designed for predicting human Cytotoxic T Lymphocytes (CTL) epitopes in any given protein. In this method, each nonameric peptide in a protein is assigned a score based on a combination of predictions of proteosomal cleavage, transported associated with antigen processing (TAP) transport efficiency, and HLA class I affinity. The reliability of NetCTL is higher than other publicly available methods for $CD8^+$ T cell epitope predictions (Larsen M.V. *et al*., 2005). NetCTL uses the concept of MHC binding prediction.

The MHC peptide binding is predicted using neural networks. The proteasome cleavage (Nielsen M. *et al*., 2005) event is predicted using the version of the NetChop neural networks trained on C terminals of known CTL epitopes as described for the NetChop-3.0 server. The TAP transport efficiency is predicted using the weight matrix based method (Peters B. *et al*., 2003). The output from the neural network predicting MHC/peptide binding is a log transformed value related to the IC_{50} in nM units.

The scores from the three individual prediction methods are integrated as a weighted sum with a relative weight on peptide/MHC binding of 1.

For Matrix Protein 1(M1) following threshold values were used:

(i) Weight on C terminal cleavage= 0.15

(ii) Weight on TAP transport efficiency= 0.05

(iii) Threshold for epitope identification= 0.75

NetCTL 1.2 uses the concept of MHC supertypes for MHC binding prediction. For each of the 12 HLA Class I supertypes, nonameric peptides with score greater than threshold are predicted as epitopes.

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

Bioinformatics and Molecular Analysis Section of the National Institute of Health (Parker K. *et al*., 1994), is the most popular prediction algorithm of peptide-MHC interaction on World Wide Web. The BIMAS tool ranks potential peptides on the basis of predicted half-time of disassociation from HLA class I molecules, which in turn is based on coefficient tables deduced from the published literature. Higher the binding affinity of a peptide to the MHC, the higher the likelihood that this peptide represents an epitope.

Cut off value for Predicted Half life ($T_{1/2}$) taken: 5

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

Syfpeithi provides scores based on the presence of certain amino acids in certain positions along the MHC-binding groove. The allocation of values is based on the frequency of the respective amino acids in natural ligands, T-cell epitopes or binding peptides (Rammensee H. *et al*., 1997). According to Syfpeithi, the top 2% of predicted peptide should contain the naturally presented epitopes in 80% of predictions.

Cut off value taken (Score): 13

4.3.2 Class II MHC ($CD4^+$ T-cell Binding Epitope Predictions)

Often the peptides interacting with Major Histocompatibility Complex (MHC) class-II molecules are longer (13–15 amino acids) but possess a core sequence of nonamers, usually with three anchor residues, and their ends extend beyond the peptide-binding groove (Sinigaglia F. *et al*., 1999). Three methods, ProPred (Singh H. and Raghava G.P., 2001), IEDB-SMM-align (Nielsen M. *et al*., 2007) and NetMHCII 2.2 (Nielsen M. and Lund O., 2009) were used to predict peptide which bind to Class II MHC molecules.

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

ProPred is a graphical web tool for predicting MHC class II binding regions in antigenic protein sequences. The tool implement quantitative matrix based prediction algorithm, employing amino-acid/position coefficient table deduced from literature (Sturniolo T. *et al*., 1999). The predicted binders can be visualized either as peaks in graphical interface or as colored residues in HTML interface. This server might be a useful tool in locating the promiscuous binding regions that can bind to 51 HLA-DR alleles (Singh H. and Raghava G.P., 2001).

Threshold used: 3%

Threshold means the percentage of best scoring natural peptides. A lower threshold corresponds to a higher stringency levels i.e. to a lower rate of false positives and to a higher rate of false negatives. Propred also gives scores to predicted peptides. The higher the score of any peptide frame the greater is the probability of it's binding to given MHC molecule.

4.3.2.2 IEDB-SMM-align (http://tools.immuneepitope.org/analyze/html/mhc_binding.html)

The MHC class II binding groove is open at both ends making the correct alignment of a peptide in the binding groove a crucial part of identifying the core of an MHC class II binding motif (Nielsen M. *et al*., 2007). A novel stabilization matrix alignment method, SMM-align allows for direct prediction of peptide: MHC binding affinities. Cross validation between peptide data set obtained from different sources demonstrated that direct incorporation of peptide length potentially results in over-fitting of the binding prediction method. Focusing on amino terminal peptide flanking residues (PFR), a consistent gain in predictive performance was demonstrated by favoring binding registers with a minimum PFR length of two amino acids.

The method predicts quantitative peptide: MHC binding affinity values, making it ideally suited for rational epitope discovery. The method has been trained and evaluated, and is largest benchmark data set publicly available and covers the nine HLA-DR supertypes. Both the peptide benchmark data set and SMM-align prediction method (NetMHCII) are made publicly available.

IC_{50} threshold used: 500 nM

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

NetMHCII 2.2 server predicts binding of peptides to HLA-DR, HLA-DQ, and HLA-DP Class II alleles using artificial neural networks (Nielsen M. and Lund O., 2009). Predictions can be obtained for 14 HLA-DR alleles covering the nine HLA-DR supertypes, six HLA-DQ, and six HLA-DP alleles (Nielsen M. *et al.,* 2007). 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.

4.4 Epitope Selection

For selection of putative epitopes following steps were followed:

(i) For all the programs used for both of MHC classes, all the nonamers that gave significant score (i.e greater or equal to cutoff/threshold value) were listed together along with their score and the number of alleles for which they could act as epitopes.

(ii) Then for each class, the results of all programs used were combined to find epitopes that gave significant result for all the programs used. This way putative epitopes for both classes were obtained.

(iii) Then the results of both classes were combined to obtain epitopes common for both MHC classes.

4.5 Isolation of Peripheral Blood Mononuclear Cell (PBMC) and optimization of the protocol for cell proliferation assay (MTT assay)

Requirements

Media Preparation- DMEM Media (Himedia) (of two types with Phenol Red and without Phenol Red), 7.5% Sodium bicarbonate (Himedia) solution, 200mM L-glutamine (Himedia) solution, 10% Fetal Bovine Serum (FBS) (Himedia), 100 Unit/ml Penicillin (Himedia), and 100 µg/ml Streptomycin (Himedia)

PBMC Isolation- Hisep LSM (Himedia), Blood (From adult healthy human being), EDTA coated Vacutainer system (Becton Dickinson) for drawing blood, and Phosphate Buffer Saline (PBS) pH 7.4

Cell Counting and Viability checking- 0.4% Trypan Blue (Himedia), Hemocytometer (Rohem)

MTT ASSAY- 5mg/ml MTT Reagent (3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide) (Himedia), and Dimethyl sulfoxide (DMSO) (SRL)

4.5.1 Preparation of Dulbecco's Modified Eagle Medium (DMEM)

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 and 20 ml of 200mM L-glutamine solution were added for 1 litre of medium and stirred until dissolved. pH was adjusted 0.2-0.3 pH units below the desired pH (7.4) using 1N HCl or 1N NaOH since the pH tends to rise during filtration. The final volume was made up to 1 litre with double distilled water. The medium was immediately sterilized by filtering through a sterile membrane filter with porosity of 0.22 µ, using syringe filter assembly. Liquid medium was stored at 2-8°C and in dark till use. Media was supplemented with 10% heat inactivated fetal bovine serum (57°C for 30 minutes) and filter sterilized penicillin (100 Unit/ml) and streptomycin (100 µg/ml) antibiotics before culturing of cells.

4.5.2 Isolation of Peripheral Blood Mononuclear Cells

Blood was drawn from a healthy human being with the use of vacutainer system. Blood was diluted in 1:1 ratio with PBS. Now blood sample was layered carefully drop wise through side of the falcon tube over equal volume of Hisep and it was centrifuged at 600×g for 30 minutes 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 minutes at 25^{o}C. Supernatant was discarded and pellet was again washed with 5 ml of PBS and centrifuged at 500×g for 10 minutes at 25^{o}C. Supernatant was discarded and pellet of PBMC was suspended in cell culture medium supplemented with FBS, Penicillin, and Streptomycin.

4.5.3 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 and damaged cells and stains them blue while live cells remain unstained. 30 µl 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. Haemocytometer was focused on using the 10X objective of the microscope and cells were counted in all four sets of squares of haemocytometer using 40X objective of the microscope.

Cell count and percent cell viability were calculated using following formula

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

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

4.5.4 Optimization of protocol for PBMC Proliferation Assay at preliminary steps (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 tissue culture plate (microtiter plate) (Exiva Sichem) containing 100 µl of cell culture medium supplemented with FBS, Penicillin, and Streptomycin. Plate was incubated in CO_2 incubator (Eppendorf) maintained at 37^oC and 5% CO_2 concentration for 24 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 microplate reader (Thermo Scientific).

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

Layout of microtiter plate used for media and solvent optimization

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

Layout of microtiter plate used for Cell no. optimization

	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												
F												
G												
H												

	No of Cells/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	$1.25x10^5$

P-PBS, M- DMEM(without phenol red)

5. RESULTS AND DISCUSSION

5.1 Most Conserved regions (MCRs) of Matrix Protein 1 (M1) in H1N1 and H3N2

Conserved regions were obtained for M1 using two tools: MUSCLE and AVANA; sequences were first aligned using MUSCLE and conserved regions were obtained for both strains with AVANA and the final conserved regions were found manually.

Eight conserved regions of M1 were found in H1N1 having conservancy in the range of 80.07%-97.43% and length of conserved regions vary from 12-70 amino acid residues (Table1). In M1 of H3N2, five conserved regions were found having conservancy in the range of 80.00%-80.19% and length of their conserved regions vary from 10-94 amino acid residues (Table 2). The data obtained complies with the fact that M1 is more conserved in comparison to highly variable surface proteins such as HA and NA since it is an internal protein (Ferguson N.M. *et al.*, 2003; Webster R.G. *et al.*, 2004).

Table 1: Final Most Conserved Regions of M1 in H1N1

Sr. No.	Sequence	Position	Conservancy (%)
MCR 1	MSLLTEVETYVLSI	1-14	≥95.60%
MCR 2	PSGPLKAEIAQRLE	16-29	≥95.59%
MCR 3	VFAGKNTDLEALMEWLKTRPILSPLTKGILGFVF TLTVPSERGLQRRRFVQNALNGNGDPNNMDRA VKLY	31-100	≥80.07%
MCR 4	KLKREITFHGAKE	102-114	≥96.34%
MCR 5	GALASCMGLIYNRMG	122-136	≥97.43%
MCR 6	CATCEQIADSQH	148-159	≥96.31%
MCR 7	TTTNPLIRHENRMVLASTTAKAMEQMAGSSEQA AEAMEVA	167-206	≥81.65%
MCR 8	DLLENLQAYQKRMGVQMQRFK	232-252	≥85.02%

Table 2: Final Most Conserved Regions of M1 in H3N2

Sr No.	Sequence	Position	Conservancy (%)
MCR 1	MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAG KNTDLEALMEWLKTRPILSPLTKGILGFVFTLTVP SERGLQRRRFVQNALNGNGDPNNMD	1-94	≥80.00%
MCR 2	AVKLYRKLKREITFHGAKEIALSYSAGALASCM GLIYNRMGAVTTEVAFGLVCATCEQIADSQHRS HRQMVATTNPLI	96-173	≥80.09%
MCR 3	HENRMVLASTTAKAMEQMAGSSEQAAEAMEIA SQARQMVQAMR	175-217	≥80.09%
MCR 4	GTHPSSSTGL	220-229	≥80.19%
MCR 5	DDLLENLQTYQKRMGVQMQRFK	231-252	≥80.09%

5.2 Immunogenic peptide prediction from overlapping putative epitopes of M1 for H1N1

Six different programs were used to predict putative epitopes for both MHC classes. Three programs: NetCTL 1.2, BIMAS and Syfpeithi were used to predict the putative MHC Class I ($CD8^+$) epitopes. Three programs: ProPred, IEDB-SMM-align and NetMHCII 2.2 were used to predict the putative Class II MHC ($CD4^+$) epitopes. All the nonamers that gave significant score (i.e greater or equal to cutoff/threshold value) and gave significant result for all three programs used were selected as putative epitopes. Immunogenic peptides were identified and selected manually by overlapping putative epitopes predicted from online tools individually for both MHC classes.

5.2.1 Immunogenic peptide prediction of M1 for MHC Class I ($CD4^+$) of H1N1

Nine Class I specific immunogenic peptides were predicted in seven conserved regions out of eight conserved regions of M1 in H1N1. Three out of nine immunogenic peptides were predicted in MCR 3 (31-100) of M1 in H1N1, while only one immunogenic peptide was predicted in each of the remaining six conserved regions. The lengths of predicted immunogenic peptides vary from 9-24 amino acid residues (Table 4). A representative example to identify and select Class I immunogenic peptide from predicted Class I MHC putative epitopes is shown in Table 3.

Table 3: Representation of immunogenic peptide selection of M1 for H1N1 Class I MHC

Most Conserved Region (PEPTIDE)		CLASS I					
		NetCTL(0.75)		BIMAS($t_{1/2}$=5)		SYFPEITHI(13)	
	NONAMER	No. Of Alleles	Score Range	No. Of Alleles	Score Range	No. Of Alleles	Score Range
MSLLTEVETYVLSI (1-14)	SLLTEVETY	3	0.8721-1.1593	3	15-44	6	13-23
	LLTEVETYV	1	1.2859	2	9.0-2666.276	3	13-23
	LTEVETYVL	2	0.9687-1.3517	4	6.0-10.0	3	13-16
PSGPLKAEIAQRLE (16-29)	LKAEIAQRL	1	0.7971	2	6.0-24.0	5	13-23
VFAGKNTDLEALMEWLKTRPILSPLTKGILGFVFTLT VPSERGLQRRRFVQNALNGNGDPNNMDRAVKLY (31-100)	MEWLKTRPI	1	0.8181	4	8-13.5	3	14-21
	EWLKTRPIL	1	1.0344	2	6-7.5	7	13-18
	KTRPILSPL	1	1.1491	4	13.44-40	9	13-24
	RPILSPLTK	1	0.9115	2	6.0-30	5	13-26
	ILSPLTKGI	1	0.764	3	5.72-17.736	5	13-25
	GILGFVFTL	1	1.2937	6	6-550.927	16	13-30
	ERGLQRRRF	1	0.8599	3	15-30	5	15-23
	RRRFVQNAL	3	1.2185-1.4922	1	6000	8	13-28
	RRFVQNALN	1	1.1797	1	3000	2	16-19
	NNMDRAVKL	1	0.7766	7	6.0-30	11	13-19
	NMDRAVKLY	1	2.4229	2	6-7.5	3	14-27
KLKREITFHGAKE (102-114)	KLKREITFH	1	1.0483	1	9	5	14-22
GALASCMGLIYNRMG (122-136)	ALASCMGLI	1	0.9204	2	5-9.0	3	13-22
	LASCMGLIY	3	1.0083-2.7186	1	5	2	14-19

Table 4: Putative Epitopes and Immunogenic peptide Of M1 for H1N1 Class I MHC

Immunogenic Peptide Class I (In Colour)	Putative Epitopes
MSLLTEVETYVLSI (1-14)	SLLTEVETY, LLTEVETYV, LTEVETYVL
PSGPLKAEIAQRLE (16-29)	LKAEIAQRL
VFAGKNTDLEALMEWLKTRPILSPLTKGI LGFVFTLTVPSERGLQRRRFVQNALNGNG DPNNMDRAVKLY (31-100)	MEWLKTRPI, EWLKTRPIL, KTRPILSPL, RPILSPLTK, ILSPLTKGI, GILGFVFTL, ERGLQRRRF, RRRFVQNAL,RRFVQNALN NNMDRAVKL, NMDRAVKLY
KLKREITFHGAKE (102-114)	KLKREITFH
GALASCMGLIYNRMG (122-136)	ALASCMGLI, LASCMGLIY
CATCEQIADSQH (148-159)	-
TTTNPLIRHENRMVLASTTAKAMEQMAG SSEQAAEAMEVA (167-206)	IRHENRMVL, MVLASTTAK, VLASTTAKA, LASTTAKAM
DLLENLQAYQKRMGVQMQRFK (232-252)	DLLENLQAY, QAYQKRMGV, YQKRMGVQM , KRMGVQMQR , RMGVQMQRF

5.2.2 Immunogenic peptide prediction of M1 for Class II MHC ($CD4^+$) in H1N1

Three Class II immunogenic peptides were predicted in only two of the eight conserved regions of M1 in H1N1. The lengths of predicted immunogenic peptides vary from 11-18 amino acid residues (Table 6). A representative example to identify and select Class II immunogenic peptide from predicted Class II MHC putative epitopes is shown in Table 5.

Table 5: Representation of immunogenic peptide selection of M1 for H1N1 Class II MHC

Most Conserved Region (PEPTIDE)		CLASS II					
		PROPRED(3%)		IEDB (Smm Alin.) <500		NET MHC II <500	
	NONAMER	No. Of Alleles	% Score Range	No. Of Alleles	Score Range	No. Of Alleles	Score Range
VFAGKNTDLEALMEWLKTRPILSPLTKGILGF VFTLTVPSERGLQRRRFVQNALNGNGDPNN MDRAVKLY (31-100)	MEWLKTRPI	32	3.49-35.23	9	42-408	1	117.6
	WLKTRPILS	34	3.18-34.69	10	40-455	2	44.1-162.9
	ILGFVFTLT	6	3.41-33.67	9	53-472	1	362.2
	LGFVFTLTV	48	5-45.86	8	33-452	1	135.9
	FVFTLTVPS	51	15.12-74.22	10	34-439	3	27.3-127.3
	FTLTVPSER	8	5.81-38.78	4	51-477	2	44.3-177.4
TTTNPLIRHENRMVLASTTAKAMEQMAGSS EQAAEAMEVA (167-206)	LIRHENRMV	25	5.68-55.21	8	137-494	1	87.2
	VLASTTAKA	31	6.73-31.4	6	90-437		161.9
	LASTTAKAM	4	9.18-26.72	4	81-292	1	227.3

Table 6: Putative Epitopes and Immunogenic peptide Of M1 for H1N1 Class II MHC

Immunogenic Peptide Class II (In Colour)	Putative Epitopes
MSLLTEVETYVLSI (1-14)	-
PSGPLKAEIAQRLE (16-29)	-
VFAGKNTDLEALMEWLKTRPILSPLTKGILGFVFTLTVPSERGLQRRRFVQNALNGNG DPNNMDRAVKLY (31-100)	MEWLKTRPI, WLKTRPILS, ILGFVFTLT, LGFVFTLTV, FVFTLTVPS,FTLTVPSER
KLKREITFHGAKE (102-114)	-
GALASCMGLIYNRMG (122-136)	-
CATCEQIADSQH (148-159)	-
TTTNPLIRHENRMVLASTTAKAMEQMAG SSEQAAEAMEVA (167-206)	LIRHENRMV, VLASTTAKA, LASTTAKAM
DLLENLQAYQKRMGVQMQRFK (232-252)	-

5.2.3 Common Immunogenic Peptide (from both MHC Classes) selection of M1 in H1N1

Finally sequences of predicted peptides for both MHC classes were looked for common region which was selected as common immunogenic peptide.

Two common immunogenic peptides were found out from both MHC classes for M1 in H1N1. The lengths of two common immunogenic peptides are 11 and 17 amino acid residues respectively (Table 7). If the common region between immunogenic peptides of both classes is shorter than 9 amino acid residues; it is neglected and not selected as common immunogenic peptide, since minimum length of immunogenic peptide is 9 amino acid residues (nonamer).

Table 7: Common Immunogenic Peptide of M1 for H1N1

Class I Immunogenic Peptide	Class II Immunogenic Peptide	Common Immunogenic Peptide
SLLTEVETYVL (2-12)	-	-
LKAEIAQRL (20-28)	-	-
(i) MEWLKTRPILSPLTKGILGFVFTL (43-66) (ii) ERGLQRRRFVQNALN (71-85) (iii) NNMDRAVKLY (91-100)	(i) MEWLKTRPILS (43-53) (ii) ILGFVFTLTVPSER (59-72)	MEWLKTRPILS (43-53)
KLKREITFH (102-110)	-	-
ALASCMGLIY (123-132)	-	-
IRHENRMVLASTTAKAM (173-189)	LIRHENRMVLASTTAKAM (172-189)	IRHENRMVLASTTAKAM (173-189)
DLLENLQAYQKRMGVQMQRF (232-251)	-	-

5.2.4 Verification of selected common immunogenic peptides of M1 for H1N1 from Literature (Source:- IEDB, Immune Epitope Database)

For verification of predicted common peptides, IEDB was searched for predicted immunogenic peptides and their results were recorded and tabulated as follows. Both of the common immunogenic peptides for H1N1 were found to be already reported specific to single HLA allele (Table 8), but in our study they are found to be reactive against multiple alleles. Thus they are promising candidates as target for peptide based vaccine design against H1N1.

Table 8: Verification of selected common immunogenic peptides of M1 for H1N1

Predicted Common Immunogenic Peptide	Epitope ID	Reference	Epitope
MEWLKTRPILS	1765977	Jenny Aurielle B Babon; Hum Immunol 2009	MEWLKTRPILS PLTKGI
IRHENRMVLASTTAKA	1542382	Junbao Yang 2009	TNPLIRHENRM VLASTTAKA

5.3 Immunogenic peptide prediction from overlapping putative epitopes of M1 for H3N2

Putative epitopes and immunogenic peptides of M1 for H3N2 were predicted in same manner as described in section 5.2.

5.3.1 Immunogenic peptide prediction for MHC Class I ($CD8^+$) of H3N2

Nine Class I specific immunogenic peptides were predicted in five conserved regions of M1 in H3N2. The lengths of predicted immunogenic peptides vary from 9-54 amino acid residues (Table 10). A representative example to identify and select Class I immunogenic peptide from predicted Class I MHC putative epitopes is shown in Table 9.

Table 9: Representation of immunogenic peptide selection of M1 for H3N2 Class I MHC

Most Conserved Region (PEPTIDE)		CLASS I					
		NetCTL(0.75)		BIMAS($t_{1/2}$=5)		SYFPEITHI(13)	
	NONAMER	No. Of Alleles	Score Range	No. Of Alleles	Score Range	No. Of Alleles	Score Range
MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGKNTDLEALM EWLKTRPILSPLTKGILGFVFTLTVPSERGLQRRRFVQNALNG NGDPNNMD (1-94)	SLLTEVETY	3	0.8721-1.1593	3	15-44	6	13-23
	LLTEVETYV	1	1.2859	2	9-2666.276	3	13-23
	LTEVETYVL	2	0.9687-1.3517	4	6.0-10	3	13-16
	VETYVLSIV	1	0.9695	2	8.8-9.0	1	13
	SIVPSGPLK	1	1.2079	1	9	5	13-25
	LKAEIAQRL	1	0.7971	2	6.0-24	5	13-23
	EIAQRLEDV	1	0.9013	1	18	4	14-21
	IAQRLEDVF	2	0.7542-0.8989	2	5.0-45	3	13-15
	RLEDVFAGK	1	1.1556	3	36-90	5	13-29
	MEWLKTRPI	1	0.8181	4	8-13.5	3	14-21
	EWLKTRPIL	1	1.0344	2	6.0-7.5	7	13-18
	KTRPILSPL	1	1.1491	4	13.44-40	9	13-24
	RPILSPLTK	1	0.9115	2	6.0-30	5	13-26
	ILSPLTKGI	1	0.764	3	5.72-17.736	5	13-25
	GILGFVFTL	1	1.2937	6	6-550.927	16	13-30
	ERGLQRRRF	1	0.8599	2	15-30	5	15-23
	RRRFVQNAL	3	1.2185-1.4922	1	6000	8	13-28
	RRFVQNALN	1	1.1797	1	3000	2	16-19
AVKLYRKLKREITFHGAKEIALSYSAGALASCMGLIYNRMGA VTTEVAFGLVCATCEQIADSQHRSHRQMVATTNPLI (96-173)	AVKLYRKLK	1	0.9121	1	120	3	15-26
	LYRKLKREI	2	0.8031-0.9301	2	6.921-66	4	13-20
	YRKLKREIT	1	1.2551	2	5-200	2	13-21
	RKLKREITF	2	0.87-1.4268	2	5-7.5	7	14-24
	KLKREITFH	1	1.0483	1	9	5	14-22

Table 10: Putative Epitopes and Immunogenic peptide of M1 for H3N2 Class I MHC

Immunogenic Peptide Class I (In Colour)	Putative Epitopes
MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGKNTDLEALMEWLKTRPILSPLTKGILGFVFTLTVPSERGLQRRRFVQNALNGNGDPNNMD (1-94)	SLLTEVETY, LLTEVETYV, LTEVETYVL, VETYVLSIV, SIVPSGPLK, LKAEIAQRL, EIAQRLEDV, IAQRLEDVF, RLEDVFAGK, MEWLKTRPI, EWLKTRPIL, KTRPILSPL, RPILSPLTK, ILSPLTKGI, GILGFVFTL, ERGLQRRRF, RRRFVQNAL, RRFVQNALN
AVKLYRKLKREITFHGAKEIALSYSAGALASCMGLIYNRMGAVTTEVAFGLVCATCEQIADSQHRSHRQMVATTNPLI (96-173)	AVKLYRKLK, LYRKLKREI, YRKLKREIT, RKLKREITF, KLKREITFH, FHGAKEIAL, GAKEIALSY, ALSYSAGAL, ALASCMGLI, LASCMGLIY, LIYNRMGAV,NRMGAVTTE, RMGAVTTEV, GAVTTEVAF, VTTEVAFGL,TEVAFGLVC, EVAFGLVCA, QMVATTNPL, MVATTNPLI
HENRMVLASTTAKAMEQMAGSSEQAAEAMEIASQARQMVQAMR (175-217)	MVLASTTAK,VLASTTAKA, LASTTAKAM, IASQARQMV, QARQMVQAM,ARQMVQAMR
GTHPSSSTGL (220-229)	THPSSSTGL
DDLLENLQTYQKRMGVQMQRFK (231-252)	DLLENLQTY, YQKRMGVQM, KRMGVQMQR, RMGVQMQRF

5.3.2 Immunogenic peptide prediction for MHC Class II ($CD4^+$) of H3N2

Five Class II immunogenic peptides were predicted in three of the five conserved regions of M1 in H3N2. Lengths of predicted immunogenic peptides vary from 9-21 amino acid residues (Table 12). A representative example to identify and select Class II immunogenic peptide from predicted Class II MHC putative epitopes is shown in Table 11.

Table 11: Representation of immunogenic peptide selection of M1 for H3N2 Class II MHC

Most Conserved Region (PEPTIDE)		CLASS II					
		PROPRED(3%)		IEDB (Smm Alin.) <500		NET MHC II <500	
	NONAMER	No. Of Alleles	% Score Range	No. Of Alleles	Score Range	No. Of Alleles	Score Range
MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGKNTDLEALME WLKTRPILSPLTKGILGFVFTLTVPSERGLQRRRFVQNALNGNG DPNNMD (1-94)	FVFTLTVPS	51	15.12-74.22	10	34-439	1	99.8
	FTLTVPSER	6	5.81-38.78	4	51-477	1	44.3
AVKLYRKLKREITFHGAKEIALSYSAGALASCMGLIYNRMGAVT TEVAFGLVCATCEQIADSQHRSHRQMVATTNPLI (96-173)	VKLYRKLKR	38	5.88-53.75	3	208-499	1	238.2
	FHGAKEIAL	6	4.49-37.07	7	54-333	1	99.1
	IALSYSAGA	16	3.19-20.93	8	37-453	1	78.3
	LSYSAGALA	4	5.68-5.68	6	40-446	1	89.3
	YSAGALASC	1	4.4	2	24-167	1	51.4
	LIYNRMGAV	3	11.22-21.43	5	55-486	3	46.8-372.6
	MGAVTTEVA	15	4.82-25	2	206-206	1	251.4
	VAFGLVCAT	9	3.3-21.05	1	98	1	173.8
HENRMVLASTTAKAMEQMAGSSEQAAEAMEIASQARQMVQ AMR (175-217)	VLASTTAKA	33	6.73-31.4	6	90-437	1	161.9

Table 12: Putative Epitopes and Immunogenic peptide Of M1 for H3N2 Class II MHC

Immunogenic Peptide Class II (In Colour)	Putative Epitopes
MSLLTEVETYVLSIVPSGPLKAEIAQRLED VFAGKNTDLEALMEWLKTRPILSPLTKGI LGFVFTLTVPSERGLQRRRFVQNALNGNG DPNNMD (1-94)	FVFTLTVPS, FTLTVPSER
AVKLYRKLKREITFHGAKEIALSYSAGAL ASCMGLIYNRMGAVTTEVAFGLVCATCE QIADSQHRSHRQMVATTNPLI (96-173)	VKLYRKLKR, **FHGAKEIAL, IALSYSAGA, LSYSAGALA, YSAGALASC, LIYNRMGAV, MGAVTTEVA,VAFGLVCAT**
HENRMVLASTTAKAMEQMAGSSEQAAEA MEIASQARQMVQAMR (175-217)	VLASTTAKA
GTHPSSSTGL (220-229)	-
DDLLENLQTYQKRMGVQMQRFK (231-252)	-

5.3.3 Common Immunogenic Peptide (from both MHC Classes) selection of M1 for H3N2

Finally sequences of predicted peptides for both MHC classes were looked for common region which was selected as common immunogenic peptide.

Four common immunogenic peptides were found out from both MHC classes for M1 in H3N2. The lengths of common immunogenic peptides vary from 9-20 amino acid residues (Table 13). If the common region between immunogenic peptides of both classes is shorter than 9 amino acid residues; it is neglected and not selected as common immunogenic peptide, since minimum length of immunogenic peptide is 9 amino acid residues (nonamer).

Table 13: Common Immunogenic peptide of M1 for H3N2

Class I Immunogenic Peptide	Class II Immunogenic Peptide	Common Immunogenic Peptide
(i) SLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGK (2-35) (ii) MEWLKTRPILSPLTKGILGFVFTL (43-66) (iii) ERGLQRRRFVQNALN (71-85)	FVFTLTVPSER (62-72)	-
(i) AVKLYRKLKREITFHGAKEIALSYSAGALASCMGLIYNRMGAVTTEVAFGLVCA (96-149) (ii) QMVATTNPLI (164-173)	(i) VKLYRKLKR (97-105) (ii) FHGAKEIALSYSAGALASC (109-127) (iii) LIYNRMGAVTTEVAFGLVCAT (130-150)	(i) VKLYRKLKR (97-105) (ii) FHGAKEIALSYSAGALASC (109-127) (iii) LIYNRMGAVTTEVAFGLVCA (130-149)
(i) MVLASTTAKAM (179-189) (ii) IASQARQMVQAMR (205-217)	VLASTTAKA (180-188)	VLASTTAKA (180-188)
THPSSSTGL (221-229)	-	-
DDLLENLQTYQKRMGVQMQRF (231-251)	-	-

5.3.4 Verification of selected common immunogenic peptides of M1 from Literature (Source:- IEDB, Immune Epitope Database)

For verification of predicted common peptides, IEDB was searched for predicted immunogenic peptides and their results were recorded and tabulated as follows. All four of the common immunogenic peptides for H3N2 were found to be already reported specific to single HLA allele (Table 14), but in our study they are found to be reactive against multiple alleles. Thus they are promising candidates as target for peptide based vaccine design against H3N2.

Table 14: Verification of selected common immunogenic peptides of M1 for H3N2

Predicted Common Immunogenic Peptide	Epitope ID	Reference	Epitope
(i) VKLYRKLKR (ii) FHGAKEIALSYSAGALASC (iii) LIYNRMGAVTTEVAFGLVCA	1598763	Laurel Yong-Hwa Lee; J Clin Invest 2008	KAVKLYRKL KREITFHGA
	1598766	Laurel Yong-Hwa Lee; J Clin Invest 2008	HGAKEIALSY SAGALA
	1598768	Laurel Yong-Hwa Lee; J Clin Invest 2008	ALASCMGLI YNRMGAV
	1598769	Laurel Yong-Hwa Lee; J Clin Invest 2008	AVTTEVAFG LVCATCEQI
VLASTTAKA	1598772	Laurel Yong-Hwa Lee; J Clin Invest 2008	IKHENRMVL ASTTAKAM

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

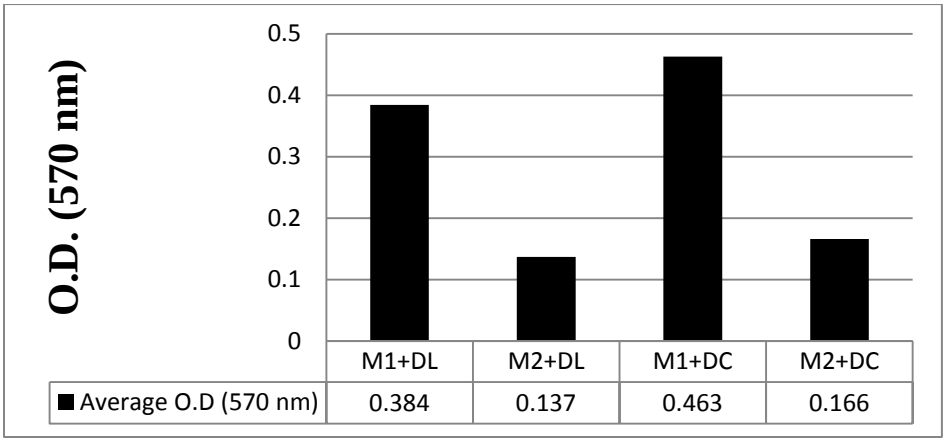
5.4.1 Optimization of protocol for Media and DMSO (solvent) to be used.

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

Table 15: Absorbance of microtiter plate (having different combinations of Media and DMSO) at 570 nm

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



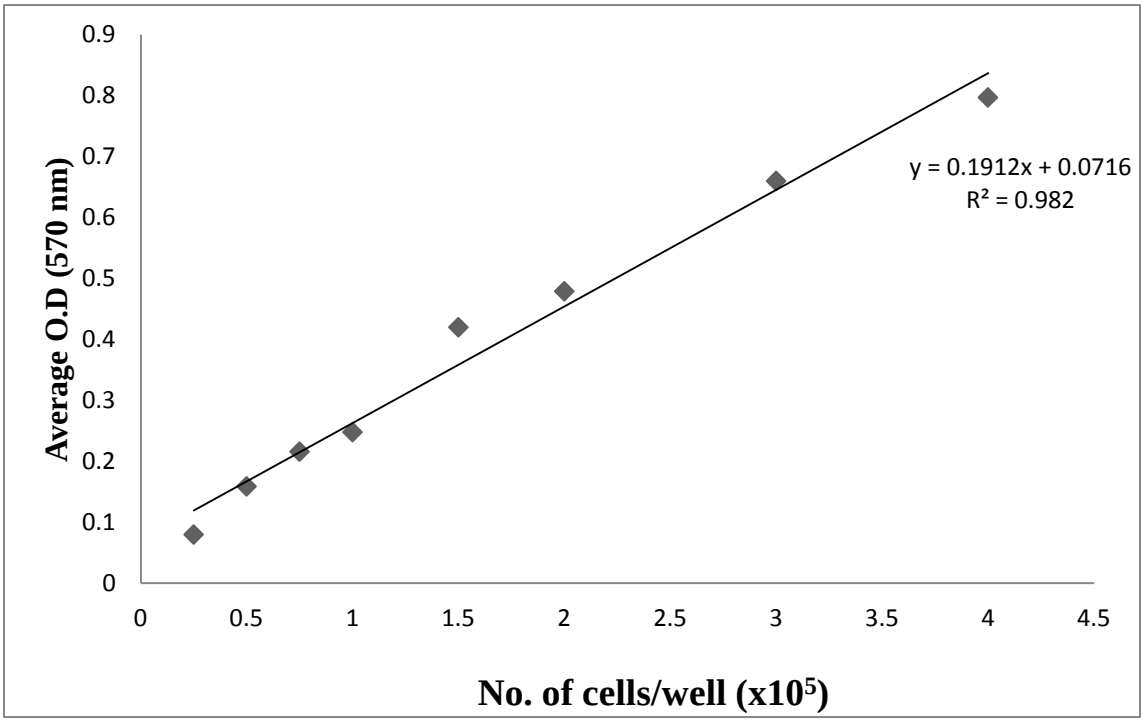
It was observed that DMEM medium (with phenol red) and DMSO (cell freezing grade), interferes with absorbance thus subsequent assay was performed by using DMEM medium without phenol red and DMSO (laboratory grade). It is due to fact that phenol red being a dye interferes with absorbance; while DMSO (cell freezing grade) interferes with absorbance as it is not in absolute solvent form; and contains DMEM medium and fetal bovine serum due to which absorbance tends to increase.

5.4.2 Optimization of protocol for PBMC Cell Proliferation Assay (MTT Assay)

For optimization of protocol for PBMC Cell Proliferation assay, a range of cells/well was used from 0.25×10^5 to 4.0×10^5. In MTT assay, O.D. is proportional to the formation of formazan crystals from MTT which depends on the number of live cells. The graph shows linear relationship between number of cells /well and O.D. at 570 nm from which it is inferred that MTT assay, is working well.

Table 16: Absorbance of microtiter plate for MTT assay at 570 nm

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



From the results obtained it was inferred that all of the different cell concentrations gave significant result (i.e linear relationship) for measurement of absorbance. As the objective is proliferation of cells, thus initial cell concentration to be used in the assay before stimulation should be ~1×10^5 cells /well because absorbance will increase as the cells proliferate.

6. CONCLUSION

It was concluded that H1N1 and H3N2 strains of influenza of influenza virus A show a high level of conservancy of Matrix protein 1 (M1). Two immunogenic peptides (i) MEWLKTRPILSPLTKGILGFVFTL (43-66) and (ii) IRHENRMVLASTTAKAM of M1 in H1N1 while two immunogenic peptides (i) MVLASTTAKAM (179-189) and (ii) AVKLYRKLKREITFHGAKEIALSYSAGALASCMGLIYNRMGAVTTEVAFGLVCA (96-149) of M1 in H3N2 were found to be most promising candidates as targets for vaccine design as they are long and cover large number of epitopes for both MHC classes.

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 studies can be carried out to assess the potential of these peptides for immunogenic response in the PBMC of healthy individuals by T-cell proliferation assay.

Vaccines designed in this way could be protective against different virus strains. Vaccines to conserved proteins have been considered, and among the candidates are the M1 (Masanori T. *et. al.,* 2008), M2 and the NP proteins. It is hoped that, by provoking immunity to conserved proteins, i.e. proteins that do not undergo antigenic change like HA and NA do, a vaccine can be produced that does not need to be replaced each year and are expected to protect against future strains as well.

7. SUMMARY

H1N1 and H3N2 strains of Influenza A viruses belong to one of the best studied viruses; however no effective prevention against influenza has been developed. Current influenza virus vaccines protect mostly against one particular strain thus regular immunization with updated formulations is necessary against the virus. Hence great challenge in the field of influenza virus research is to design universal vaccine.

The development of improved vaccines and vaccination strategies is clearly needed to enhance immune responsiveness following vaccination and to achieve high levels of protection in all risk groups. Vaccines to conserved proteins have been considered, which vaccine can be produced and does not need to be replaced each year. This is also on the WHO's agenda for a pandemic vaccine.

In the course of our study, we have used various immunoinformatics tools to determine immunogenic peptide of Matrix Protein (M1) from conserved peptide sequences as target for vaccine design. Eight conserved regions of M1 were found in H1N1 while in H3N2 five conserved regions are found. Putative epitopes for M1 were predicted from conserved peptide sequences of M1. Immunogenic peptides were identified and selected manually by overlapping putative epitopes predicted from online tools individually for both MHC classes. Nine Class I specific immunogenic peptides of M1 were predicted in H1N1 and H3N2 each. Three Class II specific immunogenic peptides were predicted in H1N1 while in H3N2 five were predicted. Finally sequences of predicted peptides for both MHC classes were looked for common region which was selected as common immunogenic peptide. Two common immunogenic peptides (from both MHC classes) for M1 in H1N1 were found while in H3N2 four were found.

For establishing immunogenicity of predicted peptides T cell proliferation assay is need to be performed and for that we have optimized the protocol for PBMC proliferation assay (MTT Assay).

Our next approach will be to synthesize these immunogenic peptides so they can be further validated for immunogenic response *In Vitro* and *In Vivo*.

8. REFERENCES

Ahlers J.D., Belyakov I.M., Thomas E.K., and Berzofsky J.A. (2001). High-affinity T helper epitope induces complementary helper and APC polarization, increased CTL, and protection against viral infection. J. Clinical Investigation, 108(11), 1677–1685.

Bao Y., Bolotov P., Dernovoy D., Kiryutin B., Zaslavsky L., Tatusova T., Ostell J., and Lipman D. (2008). The Influenza Virus Resource at the National Center for Biotechnology Information. J. Virol., 82(2), 596-601.

Belshe R.B. (2005). The origins of pandemic influenza lessons from the 1918 virus. N Engl J. Med, 353, 2209-11.

Belshe R.B., Newman F.K., Cannon J., *et al*. (2004). Serum antibody responses after intradermal vaccination against influenza. N Engl J. Med, 351, 2286-94.

Bui M., Whittaker G., and Helenius A. (1996). Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. J. Virol., 70(12), 8391.

Cassetti M.C., Couch R., Wood J., Pervikov Y. (2005). Report of meeting on the development of influenza vaccines with broad spectrum and long lasting immune responses. Vaccine, 2005, 23, 1529-33.

Cheung T.K.W. and Poon L.L.M. (2007). Biology of Influenza A Virus. Annals of the New York Academy of Sciences, 1102, 1–25.

Cooper C.L., Davis H., Cameron D.W. Influenza vaccination with 1/10th the full dose. (2004). N Engl J., 351, 2339-40.

Couch R.B., Keitel W.A., Cate T.R. (1997). Improvement of inactivated influenza virus vaccines. J Infect Dis., 176.

Drake J. W. (1993). Rates of spontaneous mutation among RNA viruses. Proc Natl Acad Sci, 90, 4171-5.

Edgar R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5).

Evans M.C. (2008). Recent advances in immunoinformatics: application of *in silico* tools to drug development. Curr Opin Drug Discov Devel., 11(2), 233-41.

Ferguson N.M., Galvani A.P., Bush R.M. (2003). Ecological and immunological determinants of influenza evolution. Nature, 422, 428-33.

Flower D.R. and Doytchinova I.A. (2002). Immunoinformatics and the prediction of immunogenicity. Appl Bioinformatics., 1(4), 167-76.

Govorkova E.A., Rehg J.E., Krauss S., Yen H.L., Guan Y., Peiris M., Nguyen T.M., Hanh T.H., Puthavathana P., Long H.T., Buranathai C., Lim W., Webster R.G., Hoffmann E. (2005). Lethality to ferrets of H5N1 influenza viruses isolated from humans and poultry. J. Virol, 79, 2191-2198.

Hay A.J. (1998). The virus genome and its replication. Oxford: Blackwell Science Ltd.

Hay A.J., Wolstenholme A.J., Skehel J.J., and Smith M.H. (1985). The molecular basis of the specific anti-influenza action of amantadine. J. EMBO, 4(11), 3021–3024.

Hilleman M.R. (2002). Realities and enigmas of human viral influenza:pathogenesis, epidemiology and control. Vaccine, 20, 3068-87.

Hoffmann E., Mahmood K., Chen Z., *et al*. (2005) Multiple gene segments control the temperature sensitivity and attenuation phenotypes of ca B/Ann Arbor/1/66. J. Virol., 79:11, 014-21.

Huang X., Liu T., Muller J., Levandowski R.A., Ye Z. (2001). Effect of influenza virus matrix protein and viral RNA on ribonucleoprotein formation and nuclear export. J. Virology, 287, 405-416.

Ichihashi T., Yoshida R., Sugimoto C., Takada A., Kajino K. (2011). Cross-Protective Peptide Vaccine against Influenza A Viruses Developed in HLA-A*2402 Human Immunity Model. PLoS ONE, 6(9).

Kenney R.T., Frech S.A., Muenz L.R., Villar C.P., Glenn G.M. (2004). Dose sparing with intradermal injection of influenza vaccine. N Engl J. Med., 004:351, 2295-301.

Khan A.M., Miotto O., Heiny A.T., Salmon J., Srinivasan K.N., *et al*. (2006). A systematic bioinformatics approach for selection of epitope based vaccine targets. Cell Immunol, 244,141–147.

Khan A.M., Miotto O., Nascimento E.J.M., Srinivasan K.N., Heiny A.T., *et al*. (2008). Conservation and Variability of Dengue Virus Proteins: Implications for Vaccine Design. PLoS Negl Trop Dis, 2(8), 272.

Korber B., LaBute M., Yusim K. (2006). Immunoinformatics comes of age. PLos Comput Biol, 2(6).

Lamb R.A., Krug R.M. Orthomyxoviridae: The viruses and their replication, in Knipe D.M., Howley P.M., Griffin D.E., *et al* (eds): Fields' Virology. Philadelphia, Lippincott Williams & Wilkins, 2001, vol 1, ed 4, 1487-1532.

Langley J.M., Halperin S.A., McNeil S., *et al*. (2005). Safety and immunogenicity of a trivalent™ inactivated influenza vaccine, given nasally to healthy adults. Vaccine.

Larsen M.V., Lundegaard C., Lamberth K., Buus S., Brunak S., Lund O., and Nielsen M. (2005). An integrative approach to CTL epitope prediction. A combined algorithm integrating MHC binding, TAP transport efficiency, and proteasome cleavage predictions. European Journal of Immunology, 35(8), 2295-303.

Lin H.H., Ray S., Tongchusak S., Reinherz E.L., Brusic V. (2008). Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunology, 9-8.

Liu W, Meng X, Xu Q, Flower DR, Li T. (2006). Quantitative prediction of mouse class I peptide binding affinity using support vector machine regression (svr) models. BMC Bioinformatics, 7, 182.

Masanori T., John C., Anita M., *et al*. (2008). Influenza a virus matrix protein 1 specific human $CD8^+$ T cell response induced in trivalent inactivated vaccine recipients. J. Virology, 82 (18), 9283-9287.

McNicholl I.R. and McNicholl J.J. (2001). Neuraminidase inhibitors: zanamivir and oseltamivir Ann Pharmacother , 35, 57-70.

Moscona A. (2008). Medical management of influenza infection. Annu. Rev. Med., 59, 397-413.

Neumann G., Fujii K., Kino Y., Kawaoka Y. (2005). An improved reverse genetics system for influenza A virus generation and its implications for vaccine production. Proc Natl Acad Sci, 102, 16825-9.

Nielsen M. and Lund O. NN-align. (2009). A neural network-based alignment algorithm for MHC class II peptide binding prediction.

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

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

Ong A.K. and Hayden F.G. (2007). John F. Enders lecture 2006: Antivirals for influenza. J. Infect. Dis., 196, 181-190.

Palese P. and Garcia S.A. (2002). New directions in vaccine research. J. Clin Invest., 109, 1517-8.

Palese P., Zavala F., Muster T., Nussenzweig R.S., Garcia S.A. (1997). Development of novel influenza virus vaccines and vectors. J. Infect Dis., 176.

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.

Parrish C.R. and Kawaoka Y. (2005). The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. Annu. Rev. Microbiol., 59, 553-586.

Peters B., Bulik S., Tampe R., Endert P., Holzhutter H.G. (2003). Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. J. Immunol., 171, 1741-49.

Plotch S.J., Bouloy M., Ulmanen I., Krug R.M. (1981). A unique cap (m7GpppXm) dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription, Cell. 23(3), 847-58.

Plotnicky H., Cyblat-Chanal D., Aubry J. P. *et al*. (2003). The immunodominant influenza matrix T cell epitope recognized in human induces influenza protection in HLA-A2/K transgenic mice. J. Virology, 309(2), 320-329.

Potter C.W. (2005). 'Influenza' in Principles and Practice of Clinical Virology. 5th ed. John Wiley & Sons, 271-297.

Rammensee H.G., Bachmann J., Emmerich N.N., Bachor O.A., Stevanovic S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics, 50, 213-219.

Rammensee H.G., Friede T., and Stevanoviic S. (1995). MHC ligands and peptide motifs: first listing. Immunogenetics, 41, 178-228.

Shannon C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423 and 623-656.

Singh H. and Raghava G.P.S. (2001). ProPred: Prediction of HLA-DR binding sites. Bioinformatics, 17(12), 1236-37.

Stanekova Z. and Vareckova E. (2010). Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. Virology Journal, 7, 351.

Sturniolo T., Bono E., Ding J., Raddrizzani L., Tuereci O., Sahin U., Braxenthaler M., Gallazzi F., Protti M.P., Sinigaglia F., Hammer J. (1999). Generation of tissue specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nature Biotechnol., 17, 555-561.

Sugrue R.J., Hay A.J. (1991). Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel.Virology, 180(2), 617-24.

Taubenberger J.K., Hultin J.V., Morens D.M. (2007). Discovery and characterization of the 1918 pandemic influenza virus in historical context. Antivir. Ther., 12(4), 581-591.

Tong S., Li Y., Rivailler P., *et al*. (2012). A distinct lineage of influenza A virus from bats. PNAS.

Varghese J.N., McKimm B.J.L., Caldwell J.B., Kortt A.A., Colman P.M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins, 14(3), 327-32.

Wang T.T., Tan G.S., Hai R., Pica N., Ngai L., Ekiert D.C., Wilson I.A., Sastre A.G, Moran T.M., and Palese P. (2010) . Vaccination with a synthetic peptide from the influenza virus hemagglutinin provides protection against distinct viral subtypes. PNAS.

Webster R.G., Bean W.J., Gorman O.T., Chambers T.M., Kwaoka Y. (1992). Evolution and ecology of influenza A viruses. Microbial. Rev., 56:1, 52-79.

WHO. (1980). A revision of the system of nomenclature for influenza viruses: A WHO memorandum. Bull World Health Organ, 58, 585-591.

Yan Q. (2010). Immunoinformatics and systems biology methods for personalized medicine. Methods Mol Biol., 662, 203-20.

Youngner J.S., Treanor J.J., Betts R.F., Whitaker D.P. (1994). Effect of simultaneous administration of cold-adapted and wild-type influenza A viruses on experimental wild-type influenza infection in humans. J Clin Microbiol., 32, 750-4.