INSILICO PREDICTION OF IMMUNOGENIC PEPTIDE OF HEMAGGLUTININ PROTEIN IN H1N1 INFLUENZA VIRUS

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

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

By

BANPREET GUJRAL
Registration No. 300901004

Under the Supervision of
DR. MANOJ BARANWAL

DEPARTMENT OF BIOTECHNOLOGY & ENVIRONMENTAL SCIENCES
THAPAR UNIVERSITY, PATIALA
JULY, 2011

DEDICATED TO MY PARENTS…..

Certificate

This is to certify that the project entitled **"Insilico prediction of immunogenic peptide of Hemagglutinin protein in H1N1 Influenza virus"** being submitted by Banpreet Gujral, Roll No. 300901004 in partial fulfillment of the requirements for the award of degree of Master of Science 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
DBTES,
Thapar University, Patiala

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

Candidate's Declaration

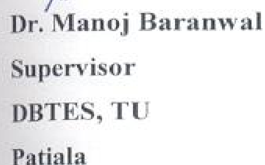
I hereby declare that the work which is being presented in the dissertation entitled **"Insilico prediction of immunogenic peptide of Hemagglutinin protein in H1N1 Influenza virus"** in partial fulfillment of the requirements for the award of the degree of Master in Science in Biotechnology, Department of Biotechnology and Environmental Sciences, Thapar University, Patiala is an authentic record of my own work during a period of six months from January 2011 to June 2011, 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.

Banpreet Gujral

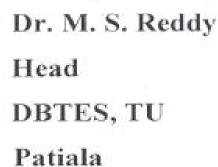
Place: Patiala

Date: 6-7-2011

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, TU
Patiala



Dr. M. S. Reddy
Head
DBTES, TU
Patiala

Acknowledgement

In pursuit of this academic endeavor, I feel that I have been singularly fortunate because inspiration, guidance, direction, co-operation, love and care- all come in my way in abundance and it seems almost an impossible task for me to acknowledge the same in adequate term.

My wholehearted indebtness goes to my erudite guide, **Dr. Manoj Baranwal**, Assistant Professor, Department of Biotechnology and Environmental Sciences, Thapar University, Patiala, for his support and patience. His invaluable assistance and precious guidance helped me in executing this arduous task from its conception to its completion.

I am sincerely thankful to **Dr. MS Reddy, Head,** Department of Biotechnology and Environmental Sciences for his immense concern throughout the project work.

I thank Ms. Neha Lohia and Mr. Raghavendra Amenedi, research scholars for their kind cooperation during the project work.

Life at Thapar University would be unforgettable for me throughtout my life because I was blessed to spend it with friends like karnika, Gunveen, Maitri, Nidhi. I thank them all for their great company.

I take pleasure to acknowledge the help given by Taran Chhabra who was always there when I needed his help.

Words fail me to express my thanks to my parents & my sister for their selfless sacrifice, encouragement and heart full blessings that continue to enlighten my life.

Above all I thank almighty God for blessing me with strength and wisdom to complete this project successfully.

Banpreet Gujral

CONTENTS

ABBREVIATIONS……………………………………………1

1. INTRODUCTION……………………………………….3

2. REVIEW OF LITERATURE……………………………5

2.1 Influenza virus…………………………………………5

2.2 H1N1…………………………………………………12

2.3 Presently available treatment against H1N1…………….13

2.4 Synthetic peptide vaccines………………………………15

2.5 Major Histocompatibility Complex……………………...16

2.6 Immunoinformatics………………………………………18

3. AIM OF THE STUDY……………………………………...19

4. MATERIALS AND METHODS……………………………20

4.1 Sequence Retrieval……………………………………….20

4.2 Conservancy Analysis……………………………………20

4.3 Epitope Prediction………………………………………..22

4.4 Epitope Selection…………………………………………30

4.5 Optimization of the protocol for isolation and culturing of PBMC……………………………………………………...30

5. RESULTS…………………………………………………33

6. DISCUSSIONS……………………………………………44

7. SUMMARY………………………………………………46

8. REFERENCES…………………………………………...48

ABBREVIATIONS

RNA	-	Ribonucleic acid
WHO	-	World Health Organization
CDC	-	Centre for Disease Control and Prevention
HLA	-	Human Leukocyte Antigen
MHC	-	Major Histocompatibility
HA	-	Hemagglutinin
NA	-	Neuraminidase
NP	-	Nucleoprotein
M1	-	Matrix protein 1
RNP	-	Ribonucleoprotein
NEP	-	Nuclear export protein
NS1	-	Non- Structural protein
IFN	-	Interferon
v RNA	-	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

1. INTRODUCTION

Influenza viruses are among the most common cause of human respiratory infections. Influenza virus is RNA virus which belongs to *Orthomyxoviridae* family. Influenza pandemics occur when a new influenza virus (typically influenza A virus) appears, against which the human population has little or no immunity. Outbreaks of the influenza infection are very common throughout the world which results in significant morbidity in the general population and increased mortality in high risk patients. The 1918 flu pandemic called as Spanish Influenza pandemic is the catastrophe against which all modern pandemics are measured. It was the first pandemic involving H1N1 influenza virus and recent pandemics of 2009 also belongs to the same. 1918 pandemic caused approximately 675,000 total deaths in the United States (Taubenberger JK *et al.*, 2008) and killed upto 50 million people worldwide (Taubenberger JK *et al.*, 2006) and it lasted from June 1918 to December 1920. Descendants of H1N1 circulated in humans until 1957 and then were replaced by an H2N2 subtype pandemic strain. H2N2 circulated until 1968 which was then replaced by H3N2 viruses. In 1977, H1N1 strains reappeared and since then both H3N2 and H1N1 have co-circulated in humans (Webster RG *et al*., 1992).

In April 2009, second outbreak of H1N1 influenza pandemic detected and first reported in the United States and Mexico (Christy BS *et al*., 2009) which subsequently spread to all parts of the world. An influenza pandemic was declared by the World Health Organization (WHO) on June 11th 2009. The pandemic was declared over on August 10th 2010. As of May 16th 2010, worldwide more than 214 countries and overseas territories or communities reported laboratory confirmed cases of pandemic influenza H1N1 2009, including over 18097 deaths (Pandemic H1N1 2009 - update 103). This virus appeared to be a novel strain of H1N1 which resulted when a previous triple reassortment of bird, swine, and human flu viruses combined with a Eurasian pig flu virus leading to the term ‘swine flu” to be used for this pandemic.

Antiviral drugs like Oseltamivir (Tamiflue) and Zanamivir (Relenza) are the choice for treatment but they had not shown effective results because majority of pandemic H1N1 are resistant to these drugs (Hajjar SA *et al.,* 2010). Recommendation of these drugs based upon recent viral surveillance and antiviral resistance data (CDC, 2010-2011). But an effective vaccine is always the best choice to control the spread of the influenza pandemic. Currently available vaccines

induce 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 provide a new strategy for prophylactic and therapeutic application of pathogen-specific immunity. Peptide-based vaccines are those in which small peptides containing epitopes derived from target proteins are used to provoke an immune reaction. These epitopes are common to the vast majority of influenza virus strains regardless of their antigenic drifts and shifts. 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.

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 HH *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 design epitope- based universal vaccines by selecting epitopes in conserved peptide sequences of Hemagglutinin (HA) viral protein that can act as candidates for vaccine design.

2. Review of Literature

2.1 Influenza virus

2.1.1 An introduction

Influenza is an acute infectious disease caused by a member of the Orthomyxovirus family. Influenza is caused by different subtypes in humans designated as A, B, C. Major outbreaks of influenza are associated with influenza virus type A or B. Infection with type B influenza is usually milder than type A. Type C virus is associated with minor symptoms. (www.pathmicro.med.sc.edu/mhunt/flu.html). These viruses are characterized by mutability and high frequency of genetic reassortment which leads to tremendous changes in the antigenic nature of the viral surface glycoproteins (Garfinkel MS *et al.*, 1994). Influenza viruses are single stranded RNA viruses surrounded by a lipid envelop with two virus encoded glycoproteins which play an important role in entry of the virus into the host cell through membrane fusion. 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).

Table 1: Various Characteristics of Influenza viruses.

Feature	Characteristic
Symmetry	Helical
Particle size	80-120nm (highly pleiomorphic)
Inner Ribonucleoprotein helix	9nm 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

2.1.2 Taxonomy

Influenza belongs to the genus Orthomyxovirus in the family of Orthomyxoviridae. Within this family are five genera: Influenza A, B and C, Thogotoviruses and Salmon Anemia Virus (http://www.umanitoba.ca/faculties/medicine/units/medical_microbiology/courses/MicroPath97.705/PDF/Influenza_report_mar2011.pdf) on the basis of their nucleocapsid (NP) and Matrix (M) protein antigens. Influenza type A viruses can infect humans, birds, pigs, horses, seals, whales, and other animals, but wild birds are the natural hosts for these viruses. Influenza B viruses are normally found only in humans. (www.pathmicro.med.sc.edu/mhunt/flu.html). The antigenic variance of the Hemagglutinin (HA) and the Neuraminidase (NA) antigens of influenza A viruses provide the foundation of their classification into subtypes. Three subtypes of type A influenza viruses, namely H1N1, H2N2 and H3N2, are able to transmit in humans (Parrish *et al.*, 2005; Taubenberger *et al*., 2007; Webster *et al.,* 1992). Till date, about 16 subtypes of HA and nine subtypes of NA have been identified in different combinations, from birds, animals and humans. Six have been isolated from humans at the molecular level (H1, H2, H3, H5, H7, H9), three of these have been involved in past pandemics (H1, H2, H3) (Gatherer D, 2009)**.** Among these three subtypes, H1N1 virus has the longest recorded history of human infection.

2.1.3 Structure of Influenza Virus

Influenza A and B consist of 8 single-stranded, negative sense RNA (vRNA) segments encoding 11 proteins. The genome is encapsulated by a lipid bilayer, traversed by matrix protein 2 (M2) proton channels, these channels function during viral uncoating. The membrane is embedded with 400-500 spikes comprised of an 8:1 ratio of HA to NA, with HA constituting approximately 40% of the total mass of the viral particle. Underlying the envelope is a layer of matrix protein (M1) which provides structural integrity to the viral particle. Each genome segment is coated with multiple copies of nucleoprotein (NP) as well as with one copy of each of the three polymerase genes, PA, PB1 and PB2, together these four components make up the ribonucleoprotein complex (RNP). The RNPs interact with M1 protein early in infection and nuclear export protein (NEP) to facilitate packaging late in infection. The most recently identified protein, PB1-F2 is encoded by an alternative reading frame in the PB1 polymerase gene and following localization of a large fraction of this protein to the inner mitochondrial

membrane in influenza-A infected cells, it appears to function as a virulence factor through enhancement of virus-induced cell death in a cell-dependent manner, reliant on the loss of inner mitochondrial membrane potential. The non-structural protein (NS1) plays a large role in immune evasion through interference with production of cellular type I IFN3. (http://www.umanitoba.ca/faculties/medicine/units/medical_microbiology/courses/MicroPath97.705/PDF/Influenza_report_mar2011.pdf).

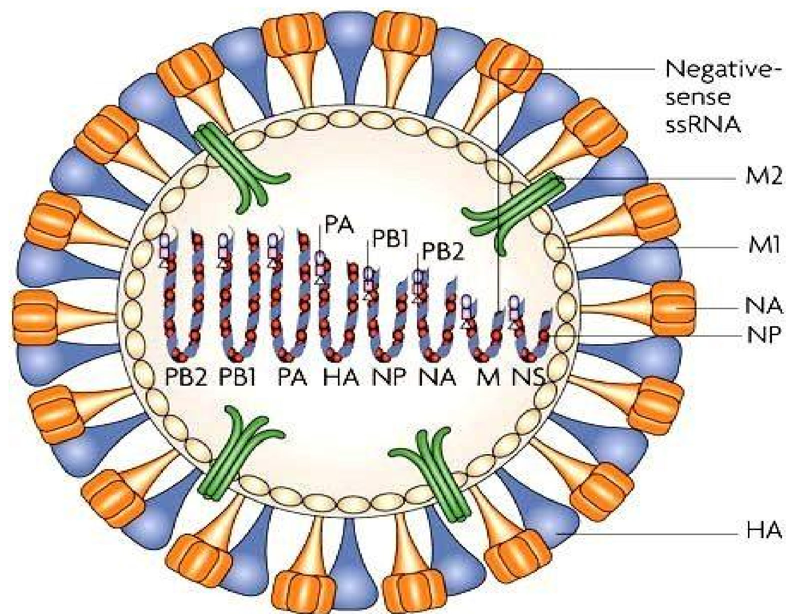


Figure 1: The structure of typical Influenza virus.

Table 2: Functions of various influenza virus proteins.

	Protein	Function
HA	Hemagglutinin	Envelope proteins, capable of being recognized by the immune system,
NA	Neuraminidase	Binds to plasma membrane
NP	Nucleoprotein	Forms a complex with the viral RNA genome and packages the RNA into a helical ribonucleoprotein core
M1	Matrix protein 1	Surrounds the ribonucleoprotein core
M2	Matrix protein 2	Forms an ion channel during lysosome fusion
PA, PB1, PB2	Polymerase proteins	Components of viral RNA transcriptase
NS1, NS2	Non-structural proteins	Believed to be involved in the control of nucleocytoplasmic transport

(Adapted from Michele S., *et al,* 1993)

2.1.4 Types of Influenza Virus

Type A and B influenza viruses consists of ribonucleoprotein core (consisting of 8 viral RNA segments and three proteins: PA, PB1, PB2), and the NEP/NS2 protein where as influenza C viruses consists of a ribonucleoprotein made up of viral RNA and four proteins. (http://www.virology.ws/2009/09/22/the-a-b-and-c-of-influenza-virus/). Comparative analysis of three types of influenza virus is shown in Table 3.

Table 3: A comparison between various Influenza virus types.

Features	Type A	Type B	Type C
Membrane proteins	HA, NA, M1	HA, NA, NB, BM2	HEF, CM2
RNA segments	8	8	7
Proton Channels	M2	BM2	CM2
Mutations	Antigenic shift and antigenic drift	Antigenic drift	-
Pandemic	Yes	No	No

2.1.5 Genome and replication of influenza virus

The Influenza A viral genome consists of eight, single negative-strand that can range between 890 and 2340 nucleotides long (Figure 2). Each RNA segment encodes one to two proteins.

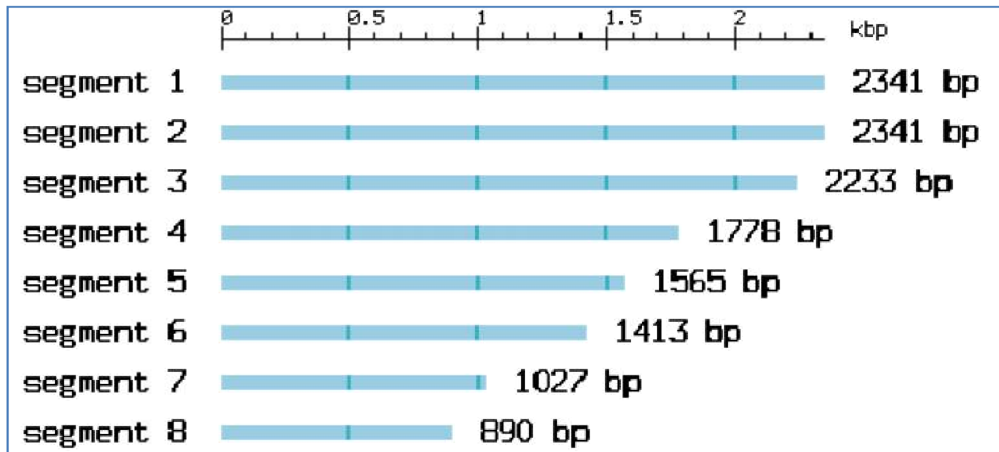


Figure 2: Genome fragments of Influenza virus.

(Adapted from Genome, NCBI)

The first step in replication, is the attachment of virion to the host cell membrane via HA and its entry into the cytoplasm by receptor-mediated endocytosis (Figure 3), thereby forming an endosome. A cellular trypsin-like enzyme cleaves HA into products HA1 and HA2. HA2 promotes fusion of the virus envelope and the endosome membranes. A minor virus envelope protein M2 acts as an ion channel thereby making the inside of the virion more acidic. As a result, the major envelope protein M1 dissociates from the nucleocapsid and vRNPs are translocated into the nucleus via interaction between NP and cellular transport machinery. In the nucleus, the viral polymerase complexes transcribe and replicate the vRNAs. Newly synthesized mRNAs migrate to cytoplasm where they are translated. Posttranslational processing of HA, NA, and M2 include transportation via Golgi apparatus to the cell membrane. NP, M1, NS1 (non-structural regulatory protein) and NEP (nuclear export protein), move to the nucleus, where bind freshly synthesized copies of vRNAs. The newly formed nucleocapsids migrate 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 the progeny virions.

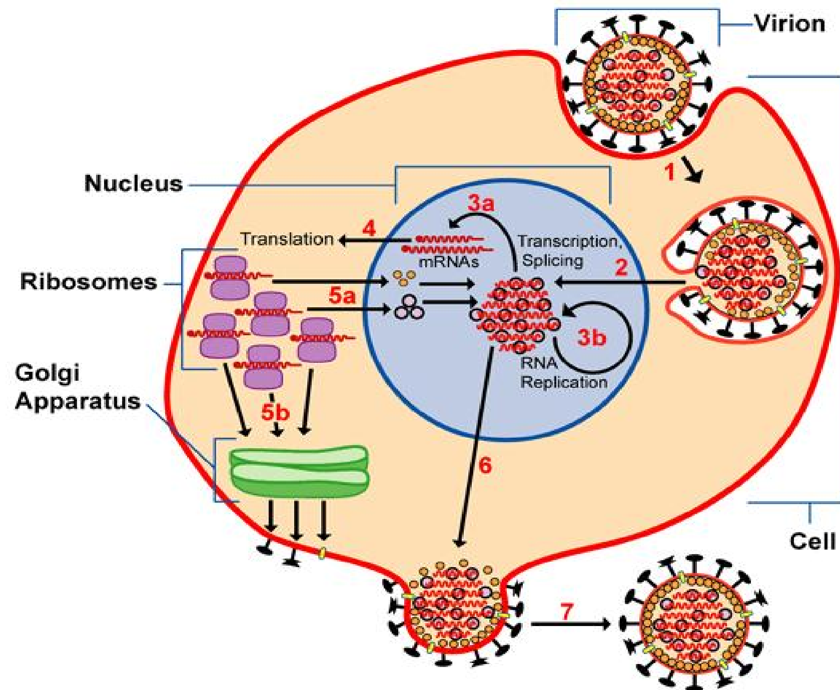


Figure 3: Scheme of Influenza A virus replication.

(Adapted from Genome, NCBI)

2.1.6 Role of Hemagglutinin (HA) and Neuraminidase (NA) protein

Hirst first discovered the presence of an enzyme activity on the surface of influenza viruses in 1942. After incubating red cells with virus, he observed the Hemagglutination reaction. The Hemagglutinin (Wharton *et al*., 1989), binds to red cells via sialic acid-containing glycoconjugates on the red cell surface. HA is a homotrimer composed of three identical, each HA subunit consists of two disulfide-linked polypeptides that are proteolytic fragments of the HA polypeptide encoded by one of the eight viral genomic RNA molecules subunits.

The primary sequence of HA contains 566 amino acid. During the maturation of the virus, the monomers are cleaved by a protease to generate two chains known as HA1 and HA2 which are joined through disulphide bridge. The HA1 forms a globular bulb at the top of the structure. This bulbous structure contains sialic acid receptor binding site. The amino acids that line the receptor binding pocket determines the specificity of the virus, that is, whether it can infect human, bird, or swine cells. This portion of the protein is the most exposed, it is also the site most targeted by immune responses and the most affected by mutation. The HA2 is primarily responsible for facilitating membrane fusion. The C-terminal end of the protein is embedded in the viral membrane. The N-terminal end, known as the fusion peptide consists of 10 hydrophobic amino acids. When the protein undergoes pH induced conformational changes, this fusion peptide becomes embedded in the host cell membrane. (http:// www. proteopedia. Org /wiki /index.php /influenza_hemagglutinin#structure).

After binding to sialic acid residues on the cell surface, and HA mediate virus entry into the cell by fusing together the virus envelope and cell membrane in a low pH-dependent process.

Neuraminidase (NA) glycoprotein is a homotetramer with a box shaped catalytic head and a slender tail (Varghese *et al.,* 1983). Each NA spike has four active sites (Kielian et *al.*, 1990). The function of neuraminidase was probably associated with the release of virus from host cells (Seto *et al.*, 1966). It was then found that antibody directed specifically against flu neuraminidase, and which abolished the activity of the enzyme for large substrates, did not prevent the infection of susceptible cells, but blocked the release of newly formed virus particles (Webster *et al.*, 1967). Neuraminidase is also known as sialidase because it breaks the linkages between sialic acid and cellular glycoproteins and glycolipids found in cell walls. Neuraminidase forms mushroom-like projections on the surface of the influenza virus.

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. Continuous and extensive antigenic variation shown by Hemagglutinin (HA) and Neuraminidase (NA) is the main reason of tremendous variability of influenza virus (Young *et al*., 1980).

Antigenic drift involves the slow and successive evolution of viral strains, due to frequent mutations (Both *et al*., 1983). In response to selection pressure to evade human Immunity it occurs on average every 2–8 years (Smith *et al*., 2004). This subtle process involves point mutations within antibody-binding sites in the HA protein, the NA protein, or both, which potentially occur each time the virus replicates (Finkenstadt *et al*., 2005).
Type A influenza also undergoes infrequent and sudden changes, called antigenic shift. Antigenic shift occurs when two different flu strains infect the same cell and exchange genetic material. The novel assortment of HA or NA proteins in a shifted virus creates a new influenza A subtype. Because people have little or no immunity to such a new subtype, their appearance tends to coincide with severe flu epidemics or pandemics.

2.2 H1N1

Swine influenza is an acute respiratory disease in pigs that can cause epidemics. Although pigs are the main host, the strains of swine origin influenza virus (S-OIV) can also be directly transmissible to humans, and reciprocally. H1N1 influenza viruses were first isolated from swine in 1930 (Shope, 1931). It has been shown to be antigenically highly similar to a recently reconstructed human 1918 A (H1N1) virus (Tumpey *et al.,* 2004*)* and likely share a common ancestor. From 1930 to the late 1990s, these “classical swine influenza” viruses circulated in swine and remained relatively antigenically stable. In April 2009, H1N1 influenza virus was isolated from humans in Mexico and the United States. This 2009 H1N1 virus contains a combination of gene segments.

Like all influenza viruses, swine flu viruses changes constantly. Pigs can be infected by avian influenza and human influenza viruses as well as swine influenza viruses. When influenza viruses from different species infect pigs, the viruses can reassort and new viruses that are a mix of swine, human and avian influenza viruses have emerged. At this time, four subtypes of influenza A virus are enzootic in pigs worldwide: H1N1, H1N2, H3N2, H3N1 (Osterhaus *et al.,* 2009). However, most of the recently isolated influenza viruses from pigs have been H1N1 viruses.

Symptoms of swine flu are similar to most influenza infections: fever (100F or greater), cough, nasal secretions, fatigue, and headache, with fatigue being reported in most infected individuals. Some may also get nausea, vomiting, and diarrhea.

2.3 Presently available treatment against H1N1

Influenza epidemics are responsible for massive disruption, and for a significant number of deaths, particularly in the elderly and the very young. At present, treatment of influenza is entirely symptomatic.

2.3.1 Chemotherapy

The HA and NA are important in the immune response against the virus; antibodies (proteins made by us to combat infection) against these spikes may protect against infection. The NA protein is the target of the antiviral drugs Relenza and Tamiflu. Also embedded in the lipid membrane is the M2 protein, which is the target of the antiviral adamantanes – amantadine and rimantadine.

1) **Oseltamivir** (Tamiflu) – It is a potent inhibitor of influenza neuraminidase. It can be given orally. Oseltamivir phosphate is an oral prodrug which undergoes hydrolysis by hepatic esterases to form active oseltamivir carboxylate. Oseltamivir carboxylate acts by selective inhibition of influenza A and B viral neuraminidase. A lipophilic side chain of the active drug binds to the virus enzyme, blocking its ability to cleave sialic acid residues on the surface of the infected cell and resulting in an inability to release progeny virions. (McNicholl *et al*., 2001). It is approved by the FDA for use as treatment for influenza A and B in persons 18 years or older
2) **Zanamivir** - Zanamivir was the first neuraminidase inhibitor available for clinical use and is effective against both influenza A and B (Monto *et al.,* 1999). Because of its poor bioavailability, Zanamivir must be administered by inhalation. Zanamivir had been shown to be effective and devoid of significant side effects in clinical trials. It is now approved by the FDA for use as treatment for influenza A and B in persons 12 years or older but not for prophylaxis. Because it cannot be given orally, it popularity has been eclipsed by oseltamivir.

3) **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 (Sugrue 1991). Amantidine is only effective against influenza A, and some naturally occurring strains of influenza A are resistant to it. The use of Amantidine is associated with the rapid emergence of drug-resistant variants (http://www.ukmi.nhs.uk/NewMaterial /Html /Docs/DrugDataNo59.pdf).
4) **Rimantadine** - For both the prevention and treatment of influenza A, Rimantadine has a comparable efficacy to Amantadine, but a lower potential for causing adverse effects (Stephenson I *et al.,* 2001, Jefferson T *et al*., 2004). 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 (Sugrue 1991). It is approved by the FDA for the treatment and prophylaxis of influenza A infection in children of one year or older.

2.3.2 Vaccines

Vaccination is an acquisition of protective immunity in advance by administration with viral antigenic glycoproteins, such as HA and NA. Therefore, the vaccination is widely considered to be the first line of defence for protecting populations in advance against influenza virus infection. Since HA antibody in the serum has the most consistent relation to the immunity against influenza virus, it is considered that the HA as transmembrane protein exhibits the principal immunogenicity in our bodies. Moreover, other transmembrane proteins, such as NA and M2, are also considered candidates for immunization.

Types of vaccine

a) **Whole virus vaccines**: Whole inactivated virus vaccines were the first influenza vaccines to be produced. The currently circulating strain of influenza is inoculated into embryonated eggs, harvested 2-3 days later and inactivated. This vaccine confers protection in 60-90% of population. However, the subsequent infecting virus may show

slow antigenic drift and the vaccine induced antibody will be less effective in conferring protection against the new strains.

b) **Split virus vaccines**: Because of the high incidence of reactions seen in vaccines given whole, inactivated virus vaccine, attempts have been made to produce a vaccine which is less antigenic but conserving immunogenicity. Split vaccines were prepared inactivated particles disrupted with detergents. Among split vaccine, the surface antigen vaccine contains predominantly purified HA and NA. These vaccines have been shown to induce fewer side effects in the vaccinees and just are immunogenic as whole virus vaccine. Whole virus vaccine should not be used in children.

c) **Live attenuated vaccines**: Immunization with live, attenuated influenza virus vaccines induce a solid immunity than do inactivated vaccines. Normal methods for attenuation, such as repeated passages and temperature adaptation require a long period to complete, and probably too long for the vaccine to become available for immunization against the current influenza strain. Currently available vaccines require annual revision of viral strains. Although research to develop live attenuated vaccines has been pursued for 20 years, basic problems remains particularly in the area of purification.

d) **Subunit virus vaccines**: Subunit vaccines have been prepared which contained only the HA and NA antigens. These are used in aqueous suspension or may be absorbed to carriers such as alhydrogel.

2.4 Synthetic peptide vaccines

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.

The immune response induced by an immunogen is not directed against the whole molecule as such, but against different parts of the molecule called antigenic determinants or epitopes. Epitopes are small active parts of the molecule that are recognized and bound by secreted or membrane-bound antibody or T-cell receptor. Proteins posses many epitope on their surface. Epitopes may be continuous (linear) or discontinuous (conformational/assembled). Assembled

epitopes are formed when amino acids which are separated in the primary sequence are brought close together in the tertiary structure due to folding of the polypeptide chain.

The design of a vaccine that guarantees antibody- mediated immunity to new influenza viruses is not currently feasible because the structural determinants of B-cell immunity are highly complex and there is no effective means for predicting the antibody epitope structure of target pathogens (Asif M. Khana *et al.,* 2006). To prepare effective vaccine, studies have shown that epitope-based vaccines should contain epitopes capable of generating immune response.

Identification of T-cell epitopes is not an easy task. A number of computational methods have been developed to identify T-cell epitopes. Epitope-based vaccines have number of advantages: they can be more potent, be controlled better, induce subdominant epitopes, and target multiple conserved epitopes in rapidly mutating pathogens like HIN1 (Ishioka *et al.,* 1999). One potential drawback of epitope-based vaccines is that the HLA genes are extremely polymorphic. Each of the corresponding molecules has a different specificity. If a vaccine needs to contain a unique peptide for each of these molecules, it will need to comprise hundreds of peptides. One way to counter this by similarities in peptide binding specificity between MHC molecules. Sets of molecules with similar specificities are called Supertype. Supertype-binding motifs and quantitative matrices have been incorporated into several computational prediction algorithms and it is now possible to identify, HLA-restricted T cell epitopes of protein sequences, allowing large-scale analysis of potential vaccine targets (Bian H *et al.,* 2004).

2.5 Major Histocompatibility complex

The binding between peptide epitopes and major histocompatibility complex proteins (MHCs) is an important event in the cellular immune response. 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 (Wen Liu *et al.,* 2006). MHC complex is group of genes on a single chromosome that codes the MHC antigens. The HLA genes are the human versions of the MHC genes that are found in most vertebrates. It is a large genomic region that contains 128 genes (MHC Sequencing Consortium, 1999). The MHC complex resides in the short arm of chromosome 6 and overall size of the

MHC is approximately 3.5 million base pairs. It is the most gene-dense region of the mammalian genome and it plays an important role in the immune system and autoimmunity. The proteins encoded by the MHC are expressed on the surface of cells in all jawed vertebrates, and display both *self* antigens (peptide fragments from the cell itself) and *nonself* antigens (e.g., fragments of invading microorganisms) to a type of lymphocytes called a T cell that has the capacity to kill or coordinate the killing of pathogens in infected or malfunctioning cells. MHC antigens and their genes can be divided into three major classes: class I, class II and class III. Class I MHC molecules are found on all nucleated cells and present peptides to cytotoxic T cells. Class II MHC molecules are found on certain immune cells themselves, chiefly macrophages, B cells and dendritic cells, collectively known as professional antigen-presenting cells (APCs). These APCs specialize in the uptake of pathogens and subsequent processing into peptide fragments within phagosomes. The Class II MHC molecules on APCs present these fragments to helper T cells, which stimulate an immune reaction from other cells. The class I gene complex contains three loci A, B and C, each of which codes of α chain polypeptides. The class II gene complex also contains at least three loci, DP, DQ and DR; each of these loci codes for one α and a variable number of β chain polypeptides. Class III region is not actually a part of the HLA complex, but is located within the HLA region, because its components are either related to the functions of HLA antigens or are under similar control mechanisms to the HLA genes. Class III antigens are associated with proteins in serum and other body fluids (e.g.C4, C2, factor B, TNF) and have no role in graft rejection. MHC molecules are extremely polymorphic and over a thousand different human MHC (HLA) alleles are known. Therefore T cell epitopes which bind to multiple HLA molecules are advantageous for vaccine design as it provide much more extensive coverage of the population.

2.6 IMMUNOINFORMATICS

Defining epitope sequence specificity (including cleavage and transport signals and MHC binding) presents a tantalizing problem for biologists especially in the context of vaccine design. Therefore, the new development in immunoinformatics and other computational method 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. Epitope mapping tools gives a direct path from genome to vaccine candidate.

Immunoinformatics also known as computational immunology 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.

Immunoinformatics consists of 3 fields:

1) Hard facts, that can be verified by experience
2) Semi-soft facts, that consist of algorithms and parameters used to create the hard part
3) Soft facts, that can never be tested with hard facts

Immunoinformatics facilitates the study and research on astounding diversity of immune system components. It greatly simplifies the study of complex regulatory pathways and network-type interactions. In the current studies, immunoinformatics tool were employed to predict peptide containing epitope in HINI as target for vaccine design.

3. Aim of the study

Designing vaccines which can provide long-term protection against more than one subtype of influenza has become a hot topic in vaccine development. As the literature suggests (Carrat *et al.,* 2007; Adar *et a*l., 2009) the currently available vaccines against influenza are viral strain specific and, hence, their efficacy is limited when the circulating strain is not included in them. The main objective of the present study was to find the region of protein that can be used as vaccine candidate against H1N1.

The work strategy included the following steps:

a) Finding conserved peptide regions of Hemagglutinin protein, from all the available strains of H1NI that has been sequenced.
b) Epitope prediction in conserved peptide sequences using various computational tools which can act as targets for vaccine design and able to bind to multiple HLA molecules.
c) Optimization of the protocol for isolation and culturing of peripheral blood mononuclear cell (PBMC).

4. MATERIALS AND METHODS

4.1 Sequence Retrieval

The sequences of HA protein of H1N1 were retrieved from NCBI Influenza database (http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/select.cgi) from January 1918 to December 2010. 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".

Table 4: sequences of HA protein of H1N1 Influenza A virus from 1918 to 2010.

Protein	Total Sequences	Number of non-redundant sequences
HA	5721	2341

4.2 Conservancy analysis

Two bioinformatics tools were used to find out the conserved regions from all retrieved sequences of HA protein of H1N1 influenza virus, MUSCLE to align the sequences and AVANA to find out the conserved regions in the aligned sequences.

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

MUSCLE stands for **Mu**ltiple **S**equence **C**omparison by **L**og-**E**xpectation. It is one of the multiple sequence alignment tool provided by European Bioinformatics Institute. 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.

Limitation of MUSCLE tool is that it does not accept more than 500 sequences at one time. As the number of sequences obtained for HA protein was more than that, so it poses a limitation to the use of this tool. But this limitation can be solved by grouping the sequences according to their year of isolation.

Table 5: Sequences of HA protein of H1N1 influenza virus grouped into 7 groups.

1918-2005	303 sequences
2006-2008	481 sequences
January 2009 - June 2009	449 sequences
July 2009 -September 2009	451 sequences
October 2009-November 2009	448 sequences
1 December 2009- 31 December 2009	180 sequences
January 2010- December 2010	138 sequences

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 AM *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 HA protein 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 all the groups of HA 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

Peptide containing epitopes are immunogenic regions of protein instead of complete protein. Binding of a peptide to an MHC molecule is a prerequisite for activation of antigen-specific T-cells. With the development of bioinformatics tools, it is easy to predict T-cell epitope from primary protein sequences.

Different algorithms based programs were used for both $CD8^+$ T-cell binding peptide prediction (Class I) MHC and $CD4^+$ T-cell binding peptide prediction (Class II MHC).

4.3.1 $CD8^+$ T-cell binding peptide prediction

Five programs, NetCTL1.2 (Larsen MV *et al.,* 2005), BIMAS (Parker K *et al.*, 1994), Syfpeithi (Rammensee H *et al.,* 1999), IEDB Artificial Neural Network (ANN) (Nielsen *et al*., 2003) and IEDB Stabilized Matrix Method (SMM) (Peters B *et al.*, 2005), were used to predict the 9mer peptides because a length of nine amino acid represents the typical length of peptide that binds to class I HLA molecules.

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 effiency, and HLA class I affinity. The reliability of NetCTL is higher than other publicly available methods for $CD8^+$ Tcell epitope predictions. (Larsen MV *et al*., 2005). NetCTL uses the concept of MHC binding prediction.

The MHC peptide binding is predicted using neural networks. The proteasome cleavage (M. Nielsen *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 IC50 values 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 HA protein 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

MHC Supertypes

The identification of peptides that can bind to Major Histocompatibility Complex (MHC) molecules is important for anticipation of T-cell epitopes and for the design of epitope-based vaccines. Class I MHC molecules bind short peptides derived from the processing of proteins, and present them on the cell surface for T cell scrutiny. The peptide repertoire of a specific molecule is to a large extent determined by the MHC molecular structure accommodating the so called main anchor positions of the presented peptide. These receptors are extremely polymorphic, and much of the polymorphism influences the peptide-binding repertoire. However, despite this polymorphism, class I MHC molecules can be clustered into sets of molecules that bind largely overlapping peptide repertoires. Set of MHC molecules with similar specificities are called as Supertype. Analyzing the structure of the peptide-binding pockets of MHC molecules, cataloging motifs and performing MHC- peptide binding assays, similarities in peptide specificity between MHC molecules were identified.

Table 6: Classification of HL-A and HL-B Class 1 alleles into Supertypes.

SUPERTYPE	ALLELES
A1	A*0101,A*2601,A*2602,A*2603,A*2902,A*3001,A*3002,A*3003,A*3004, A*3201
A2	A*0201,A*0203,A*0204,A*0205,A*0206,A*0207,A*0202,A*0214,A*0217 A*6801,A*6901
A3	A*0301,A*1101,A*3001,A*3101,A*3301,A*3303,A*6601,A*6801,A*7401
A24	A*2301,A*2402,A*2902
B7	B*0702,B*0703,B*0705,B*3501,B*3503,B*4201,B*5101,B*5102,B*5103, B*1508,B*5301,B*5401,B*5501,B*5502,B*5601,B*6701,B*7801
B8	B*0801,B*0802
B27	B*1402,B*1503,B*1509,B*1510,B*1518,B*2702,B*2703,B*2704,B*2705 B*2706,B*2707,B*2709,B*4801,B*3801,B*3901,B*3902,B*3909
B44	B*4001,B*4002,B*4006,B*4402,B*4403,B*4501,B*1801,B*3701,B*7301
B58	B*1516,B*1517,B*5701,B*5702,B*5801,B*5802
B62	B*1501,B*1502,B*1510,B*1512,B*1513,B*4601,B*5201

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 highest NETCTL score are predicted as epitopes.

2. BIMAS (http://www.bimas.dcrt.nih.gov/molbio/hla_bind/)

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 Half life ($T_{1/2}$) taken: 5

3. Syfpeithi (http://www.syfpeithi.de/scripts/MHCserver.dII/EpitopePrediction.html)

The Syfpeithi database contains extensive information on MHC class I and class II anchor motifs and binding specificity, and includes more than 4,500 entries of MHC proteins and aligned sequences of their epitopes and natural ligands, with source proteins, organisms, and publication references for each peptide .The prediction is based on published motifs (pool sequencing, natural peptides) and takes into consideration the amino acids in the anchor and auxiliary anchor positions, as well as other frequent amino acids.

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: On the basis of score obtained in conserved region which contains no epitope.

4. IEDB- ANN (http://tools.immuneepitope.org/analyze/html/mhc_binding.html)

Neural network method to predict T-cell class I epitopes consists of a combination of sparse encoding, Blosum encoding and input derived from Hidden Markov Models. Combination of several neural networks derived using different sequence-encoding schemes has a performance superior to neural networks derived using a single sequence encoding scheme. This method has a high performance in comparison to other methods. Neural networks are ideally suited to integrate higher order correlations when predicting the binding affinity (Nielsen M *et al*., 2003).

A number of methods for predicting the binding of peptides to MHC molecules have been developed. In the simpler prediction tools it is assumed that the amino acids at each position along the peptide sequence contribute with a given binding energy, which can be independently added up to yield the overall binding energy of the peptide (Parker K *et al*., 1994; Meister *et al*.,1995; Stryhn *et al*.,1996). These predictions, however, fail to recognize correlated effects where the binding affinity of a given amino acid at one position is influenced by amino acids at other positions in the peptide. Two adjacent amino acids may, for example, compete for the space in a pocket in the MHC molecule. Artificial neural networks (ANN) are ideally suited to take such correlations into account.

This method gives the binding affinity as IC50 value.

- Lower the IC50 value, stronger would be the affinity.
- Peptides with IC50 values <50 nM have high affinity,
- Peptides with IC50 values <500 nM have intermediate affinity,
- Peptides with IC50 values <5000 nM have low affinity.

Most known epitopes have high or intermediate affinity. Some epitopes have low affinity, but no known T-cell epitope has an IC_{50} value greater than 5000.

IC 50 threshold used: 100

5. IEDB_SMM (http://tools.immuneepitope.org/analyze/html/mhc_binding.html)

Stabilized Matrix Method (SMM) generates a model for the sequence specificity involving the recognition of short sequences of amino acid that binds to MHC molecules. The purpose of this model is two-fold. First, it can provide a summary of experimental results, allowing for a deeper understanding of the mechanisms involved in sequence recognition. Second, such models can be used to predict the experimental outcome for yet untested sequences. This method has been successfully applied to predicting peptide binding to MHC molecules, peptide transport by the transporter associated with antigen presentation (TAP) and proteasomal cleavage of protein sequences. Advantageous features of the package are:

(1) The output generated is easy to interpret.

(2) Input and output are both quantitative.

(3) Specific computational strategies to handle experimental noise are built in.

(4) The algorithm is designed to effectively handle bounded experimental data.

(5) Experimental data from randomized peptide libraries and conventional peptides can easily be combined.

IC 50 threshold used: 100

4.3.2 CD4$^+$ T-cell binding peptide prediction

Often the peptides interacting with Major Histocompatibility Complex (MHC) class-II molecules are longer (13–15 amino acids) but possess a core sequence of about 9mers, usually with three anchor residues, and their ends extend beyond the peptide-binding groove (Madden, 1995; Sinigaglia *et al.,*, 1995). Three methods, ProPred (Singh H, *et al*., 2001), MHC2PRED (Singh H *et al.,*), IEDB-SMM-align (Nielsen *et al*., 2007) were used to predict peptide which bind to classII MHC molecules.

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 *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, Raghava GP, 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 its binding to given MHC Molecule.

2. MHC2PRED (http://www.imtech.res.in/raghava.mhc2pred/)

The MHC2PRED is a support vector machine (SVM) method for prediction of promiscuous MHC class II binding peptides. MHC2PRED uses matrix optimization technique to detect the binding core of peptides. The overall accuracy of method is >78%, which is better than all the already existing methods in literature.

The binders and non-binders for all alleles have been obtained from MHCBN and JenPep database (Bhasin *et al.*, 2003; Blythe *et al.*, 2002). All the peptides having IC50 value less than 500nm has been considered as binders and peptides with IC50 value greater than 500nm are considered as non-binders. Peptides containing less than 9 amino acids have been deleted from the dataset. The binding core of 9 amino acids has been obtained from the binders of variable length without considering MHC binding motifs using Matrix Optimization Techniques (MOT) package (Singh H, *et al.*, unpublished).

Threshold used: 1%

3. 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. The predictive performance of the method

is validated on a large MHC class II benchmark data set covering 14 HLA-DR (human MHC) and three mouse H2-IA alleles. The predictive performance of the SMM-align method was demonstrated to be superior to that of the Gibbs sampler, TEPITOPE, SVRMHC, and MHCpred methods. 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 SMM-align method outperforms other state of the art MHC class II prediction methods. The method predicts quantitative peptide: MHC binding affinity values, making it ideally suited for rational epitope discovery. The method has been trained and evaluated on the, to our knowledge, largest benchmark data set publicly available and covers the nine HLA-DR supertypes suggested as well as three mouse H2-IA allele. Both the peptide benchmark data set and SMM-align prediction method (NetMHCII) are made publicly available.

IC 50 threshold used: 500

4.4 Epitope Selection

For selecting the epitopes that could be considered as putative epitopes following steps were used:

1. For all the programs used for each Class of MHC, all the peptides that gave significant score were listed together along with their score and the number of alleles for which they could act as epitopes.

2. 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 Class I and Class II were obtained.

3. Then the results of Class I and Class II were combined to obtain epitopes common for both MHC classes.

4.5 Optimization of the protocol for isolation and culturing of Peripheral Blood Mononuclear Cell (PBMC).

Table 7: List of Requirements

Requirements	Company
DMEM	HI-MEDIA
Penicillin	HI-MEDIA
Streptomycin	HI-MEDIA
Filter paper(0.22 micron)	Millipore
Sodium bicarbonate	Merck
200mM L- glutamine	Himedia
Foetal Bovine Serum	Himedia
EDTA(1.8mg per 1ml of blood)	Himedia

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

Procedure:-

1) 9.6 gms of the DMEM powder was suspended in 900 ml of tissue culture grade water with constant stirring until the medium was completely dissolved.
2) pH was adjusted to 4.0. The pH was adjusted to 0.2-0.3 unite below the desired pH since the pH tends to rise during filtration.
3) The volume was made to 1 litre with tissue culture grade water. Volumes of 7.5% sodium bicarbonate and 200mM L-glutamine solution were subtracted from the final volume. Sodium bicarbonate and L-glutamine was filter sterilized before use.
4) The medium was autoclaved.
5) The medium was removed promptly from the autoclave to avoid extended heating or evaporation and allowed to cool at room temperature.
6) 49.3 ml of sodium bicarbonate solution and 20ml of 200 mM L-glutamine solution was added to the final volume of the medium being prepared.
7) Medium was stored at 2-8° c till use.

*10% Heat inactivated serum (57^{0} C for 30 minutes) and filter sterilized antibiotics are added to media before culturing of cells.

4.5.2 Isolation of Peripheral Blood Mononuclear Cells

PBMCs were isolated from 5 milliliter of anticoagulant containing blood by Hisep LSM Ficoll plus density gradients. Blood samples were layered onto 5ml Hisep LSM Ficoll and centrifuged at 400 × g with the brake off, at room temperature for 20 minutes. PBMCs were removed from the density medium/plasma interface and were washed twice with PBS (400 × g with the brake off, at room temperature for 10 minutes).

The cells were counted and the number of viable cells was determined by trypan blue exclusion staining.

4.5.3 Preparing cell suspension

The cell suspension to be counted was well mixed by gentle agitation.100µl of cells was taken in eppendorf and 100 µl of tryphan blue was added to it and mixed gently.

4.5.4 Cell Counting

Cell suspension containing trphan blue were drawn out with th help of pipette on cleaned Heamocytometer. The fluid should run to the edges of the grooves.

The grid lines of the Haemocytometer were focused on using the 10X objective of the microscope. The cells were counted in all 4 sets of squares of Haemocytometer.

To obtain the count:-

The total count from 4 sets of squares of Haemocytometer= cells/4*10^4*dilution factor/ml

5. Results

5.1 Conserved regions of Hemagglutinin Protein in H1N1 virus

Conserved regions were obtained in different groups of HA protein using two tools: MUSCLE and AVANA, and the final conserved were calculated manually. Eleven peptide sequences ranging from 10-30 amino acids, in length were identified as final conserved peptides (Table 8 & 9).

Table 8: Conserved sequences of different groups of HA protein obtained using AVANA.

Conserved sequences 1918-2005	Conserved sequences 2006-2008	Conserved sequences Jan 2009-June 2009	Conserved sequences July2009-Sept2009	Conserved sequences Oct2009-Nov2009	Conserved sequences 1Dec2009-31Dec2009	Conserved sequences 1Jan2010-31Dec2010	Final Conserved sequences
17-59	1-51	21-51	1-219	1-390	1-390	1-390	21-51
61-69	53-98	75-85	221-567	392-566	392-566	392-566	-
75-85	100-110	114-136					75-85
104-136	112-157	234-243					114-136
186-199	159-202	342-384					342-361
212-221	211-225	395-418					363-384
242-268	227-269	425-435					437-456
271-287	271-283	437-456					458-467
295-326	292-398	458-469					471-489
328-361	434-572	471-491					495-512
363-418		495-512					531-546
420-467		531-546					548-566
469-489		548-572					-
491-566							-

Table 9: Final conserved Sequences from 1918 to 2010 of Hemagglutinin protein.

Serial no.	Sequence	Position
CS*1	CIGYHANNSTDTVDTVLEKNVTVTHSVNLLE	21-51
CS2	AGWILGNPECE	75-85
CS3	DYEELREQLSSVSSFERFEIFPK	114-136
CS4	PSIQSRGLFGAIAGFIEGGW	342-361
CS5	DGWYGYHHQNEQGSGYAADQK	363-384
CS6	DIWTYNAELLVLLENERTLD	437-456
CS7	HDSNVKNLYE	458-467
CS8	SQLKNNAKEIGNGCFEFY	471-489
CS9	CMESVKNGTYDYPKYSEE	495-512
CS10	YQILAIYSTVASSLVL	531-546
CS11	VSLGAISFWMCSNGSLQCR	548-566

*CS – Conserved Region

5.2 T cell epitopes from conserved peptides for class I MHC

Five different tools as NetCTL1.2, BIMAS, Syfpeithi, IEDB- ANN, IEDB-SMM were used to predict epitope which bind to Class I MHC molecules. Representative examples of each tool results are given in Table 10 to14. Threshold for each tool were taken based on literature and based on CS2 (Conserved region 2) in which no epitope was predicted. On comparing results from all five toos, 19 peptide sequence were selected as putative epitopes (Table 15). Out of 19 epitopes, 16 epitopes were predicted by all five tools and three of them by four tools. On the basis of overlapping epitopes, stretch of peptide were seleted in conserved region 3, conserved region 6, conserved region 10, conserved region 11 as putative target for vaccine design (Figure 4). In figure 4, the exact position of HA protein sequence were mentioned.

Table 10: Representative results of NetCTL.

SUPERTYPE	SEQUENCE	POSITION	SCORE T*=0.75
A1	STDTVDTVL	1:9-17	1.8043
	HQNEQGSGY	5:8-16	1.0369
	WTYNAELLV	6:3-11	0.7737
A2	TVLEKNVTV	1:15-23	0.9602
	WTYNAELLV	6:3-11	0.8772
	VLLENERTL	6:11-19	0.7722
B58	SSVSSFERF	3:10-18	1.3513
	VSSFERFEI	3:12-20	1.0277
	SSFERFEIF	3:13-21	0.9859
A24	IWTYNAELL	6:2-10	1.3519
	TYNAELLVL	6:4-12	1.1917

* Threshold for epitope identification: 0.75

Table 11: Representative results of Bimas.

SEQUENCE	POSITION	HLA molecule	$T^{*}(_{1/2})>5$
TVLEKNVTV	1:15	HLAA*0201	92.322
		HLAA*0205	36
		HLAB*5102	7.986
DYEELREQL	3:01	A24	432
		HLAB*3801	15.6
		HLAB*3901	9
VLLENERTL	6:11	A_0201	110.183
		A_0205	21.42
		A24	7.2
		B14	5
		B_2705	9

* Cut off value Dissociation time($T_{1/2}$) taken: 5

Table 12: Representative results of Syfpeithi.

SEQUENCE	POSITION	ALLELE	SCORE*
TVLEKNVTV	1:15-23	HLAA*0201	23
		HLAA*1101	15
		HLAB*1402	14
		HLAB*4901	12
		HLAB*6801	13
		HLAB*3901	14
		HLAB*5101	18
EQLSSVSSF	3:7-15	HLAA*2402	13
		HLAB*1510	18
		HLAB*2709	12

*Cut off value taken: >12(highest score given by CS2)

Table 13: Representative results of IEDB-ANN.

SEQUENCE	POSITION	ALLELE	IC50*
TVLEKNVTV	1:15-23	HLAA*0201	83.3
		HLAA*0206	17.4
		HLAA*0211	16.1
		HLAA*0212	34.8
		HLAA*0216	33.2
		HLAA*0219	49.6
		HLAA*0250	14.8
		HLAA*6802	31.7
		HLAA*6901	8.5
SSVSSFERF	3:10-18	HLAA*2602	23.4
		HLAB*1501	87.7
		HLAB*1503	25.3
		HLAB*1517	5.6
LSSVSSFER	3:9-17	HLAA*3101	18.5
		HLAA*6801	14.8
SSFERFEIF	3:13-21	HLAA*3201	80.3
		HLAB*1501	96.6

*IC 50 threshold: 100

Table 14: Representative results of IEDB –SMM.

SEQUENCE	POSITION	ALLELE	IC50
TVLEKNVTV	1:15-23	HLAA*0206	53.9
		HLAA*0211	45.8
		HLAA*0212	46.6
		HLAA*6901	26.2
		HLAA*0250	9.1
TVLEKNVTV	1:15-23	HLAA*0206	53.9
		HLAA*0211	45.8
DYEELREQL	3:1-9	HLAA*0250	15.5
QLSSVSSFE	3:8-16	HLAA*0202	40.7
		HLAA*0250	7
ELREQLSSV	3:4-12	HLAA*0203	43.1
		HLAA*0211	92.4
		HLAA*0250	36.3
LREQLSSVS	3:5-13	HLAA*0250	4.3
YEELREQLS	3:2-10	HLAA*0250	5.9

*IC 50 threshold: 100

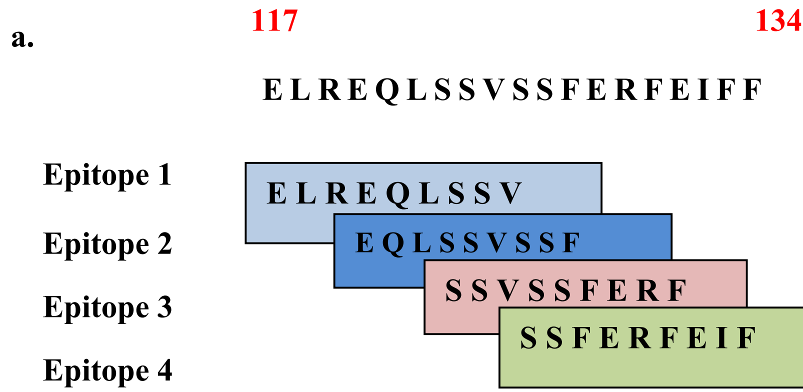
Table 15: Peptides (with position) selected as class I MHC epitopes by various tools.

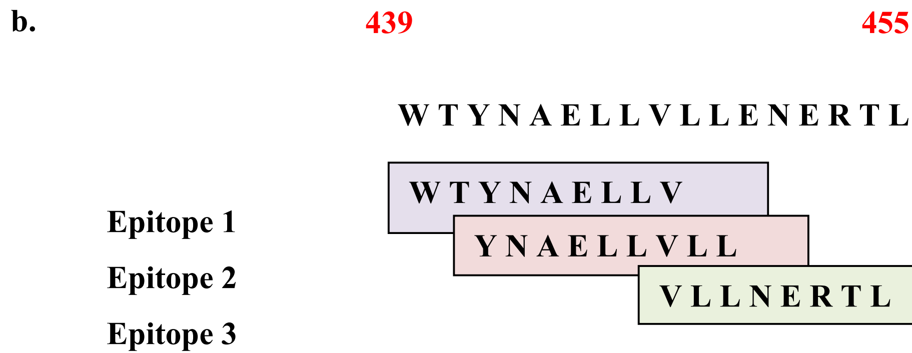
POSITION	IEDB:ANN		IEDB:SMM		NETCTL1.2		BIMAS		SYFPEITHI	
	N^a	R^b	N^a	R^b	S^d	R^c	N^a/S^d	R^c	N^a	R^c
35-43	9	8.5-83.3	4	9.1-53.9	1	0.9602	3	7.98-92.32	7	12-23
117-125	4	3.1-53.8	3	36.3-43.1	1	0.9989	1	6.6	2	12-25
120-128	4	2.3-49.1	2	1.7-59.1	3	0.84-1.28	5	5-80	4	12-18
123-131	4	5.6-87.7	3	10.3-49.1	1	1.35	3	5-52.8	4	12-15
126-134	4	9.9-96.6	3	7.3-62.3	2	0.98-1.13	5	5-96	4	14-15
348-356	4	23.3-99.3	3	69.1-89.9	3	0.86-1.31	4	15-67.5	7	12-22
353-361	-	-	4	45.7-97.8	1	1.42	1	90	1	12
370-378	3	18.6-45.4	4	2.8-89	4	0.81-1.47	3	20-88	1	21
439-447	6	27.5-60.2	4	16.2-61.2	2	0.87-1.03	5	7.26-14.22	5	12-18
441-449	1	60.1	3	13.2-51.6	1	1.60	7	5-20	13	12-23
447-455	5	3.3-18.8	6	2.3-74.7	1	0.77	-	-	13	12-27
498-506	3	11.7-18	4	13.3-88.7	3	0.92-1.47	3	6-12	3	13-21
531-539	1	11.3	2	15.2-20.6	1	0.87	1	20	2	13-15
532-540	7	9.8-57.5	2	21.5-93.6	1	1.09	3	6.6-53.07	2	13-22
533-541	4	9.3-97.4	5	2.3-34.3	1	0.7814	-	-	5	12-19
538-546	1	3.8	2	1.9-15	4	0.76-1.14	1	6	10	12-19
548-556	3	8.1-35.5	4	3.5-60.7	1	1.6542	1	96	2	13-15
549-557	6	13.7-61.5	3	6.3-61.2	1	1.0519	3	8-158.48	1	16
555-563	1	51.1	5	5.1-47.9	2	0.77-0.80	2	6	2	13-14

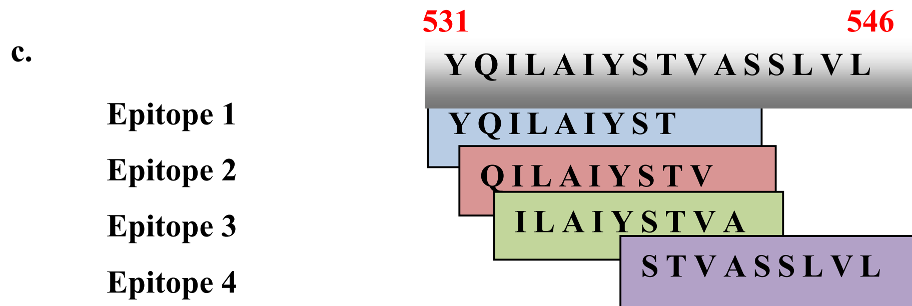
*Position wrt length of Hemagglutinin protein, a: - No. of Alleles

b: - IC50 Value (Range), Lower the value, Greater the Binding affinity

c: - Score(Range), Higher the value, Greater the Binding affinity, d:-Supertype







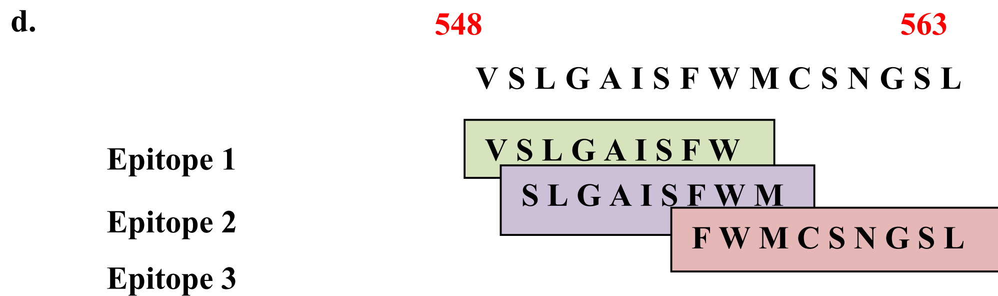


Figure 4(a to d): Stretch of peptide sequence of Hemagglutinin protein having overlapping predicted epitopes.

5.3 T cell epitope from conserved peptide for class II MHC

Three different tools as Propred, MHC2PRED and IEDB-SMM-align were used to predict epitope which bind to Class II MHC molecules. Representative examples of each tool results are given in table 16 to 18. Threshold for each tool were taken based on literature. On comparing results from all three tools, 14 peptide sequences were selected as epitopes (Table 19). Out of 14 epitopes, two epitopes were predicted by all three tools and 12 of them by two tools. On the basis of overlapping epitopes, stretch of peptide were seleted in conserved region 3, conserved region 6, conserved region 10, conserved region 11 as putative target for vaccine design (Fig 5). In fig 5, the exact position of HA protein sequence were mentioned.

Table 16:- Representative results of MHC2PRED.

SEQUENCE	POSITION	ALLELE	SCORE (1%)
DTVDTVLEK	1:11-19	HLADRB*0901	1.63
VLEKNVTVT	1:16-24	HLADRB*0901	1.512
STDTVDTVL	1:9-17	HLADRB*0901	1.498
VTHSVNLLE	1:23-31	HLADRB*0901	1.487
VLEKNVTVT	1:16-24	HLADRB1*1501	1.629
GWILGNPEC	2:2-10	HLADRB*0901	1.388
WILGNPECE	2:3-11	HLADRB*0901	1.189
AGWILGNPE	2:1-9	HLADRB*0901	1.091
FERFEIFPK	3:15-23	HLADRB1*0401	1.464

Table 17:- Representative results of propred.

SEQUENCE	POSITION	ALLELE	SCORE
IGYHANNST	1:2-9	DRB1_0401	2.1
		DRB1_0402	3.2
		DRB1_0421	2.8
		DRB1_0426	2.1
		DRB1_1501	3.6
		DRB1_1506	3.6
LREQLSSVS	3:5-13	DRB1_0401	2.3
		DRB1_0402	1.9
		DRB1_0404	3.2
		DRB1_0405	2.2
		DRB1_0408	2.2

Table 18:- Representative results of IEDB-SMM-align.

SEQUENCE	POSITION	ALLELE	IC50=500
LREQLSSVS	3:5-13	HLADRB1*0101	61
		HLADRB1*0101	61
		HLADRB1*0101	62
		HLADRB1*0101	173
		HLADRB1*0101	61
LVLLENERT	6:10-18	HLADRB1*0101	24
		HLADRB1*0101	27
ILAIYSTVA	10:3-11	HLADRB1*0101	4
		HLADRB1*1501	42
		HLADRB1*0401	43
		HLADRB1*0802	167
		HLADRB1*1501	89
		HLADRB5*0101	233

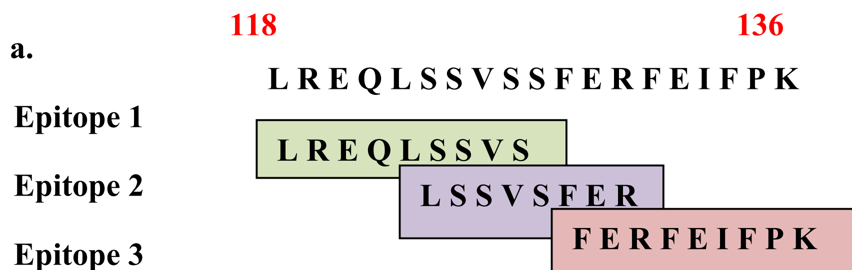
Table 19: Peptides (with position) selected as class II MHC epitopes by various tools.

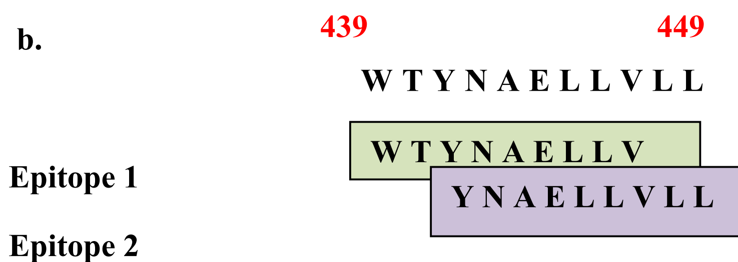
POSITION	IEDB:SMM-ALIGN		PROPRED		MHC2PRED	
	N[a]	R[b]	N[a]	R[c]	N[a]	R[c]
37-45	2	155-396	3	1.88-2.06	-	-
41-49	5	27-436	2	6.5	-	-
118-126	6	61-417	9	1.9-3.2	-	-
122-130	4	126-334	2	3.5	-	-
128-136	-	-	9	0.7-3	1	1.36
350-358	2	88-440	2	2.4	2	1.42-1.66
439-447	4	225-457	-	-	1	1.29
441-449	5	26-402	-	-	1	1.06
531-539	1	177	17	1-3.8	-	-
533-541	5	4-233	2	1.7	-	-
534-542	2	61-231	30	1.7-4.1	-	-
536-544	3	6-95	2	5.3	1	1.116
555-563	2	129-377	2	4.1	-	-
557-565	-	-	3	1.8	1	1.33

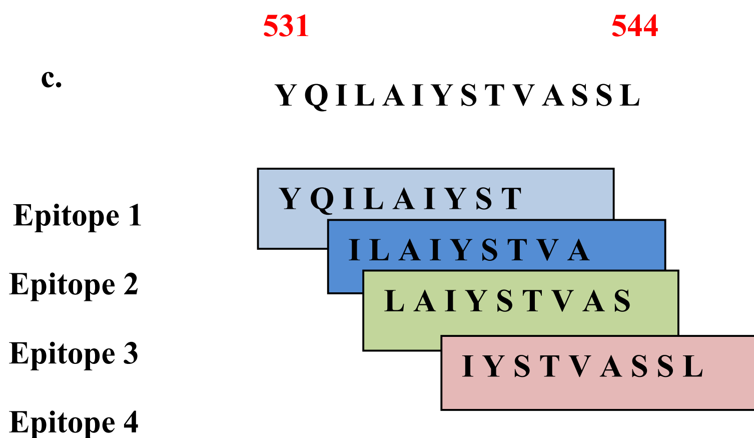
a:- No. of Alleles

b:- IC50 Value(Range), Lower the value, Greater the Binding affinity

c:- Score(Range), Higher the value, Greater the Binding affinity







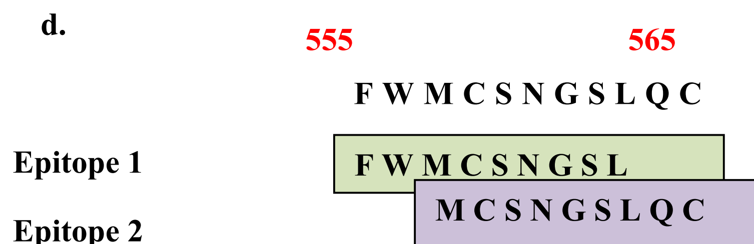


Figure 5(a to d): Stretch of peptide sequence of Hemagglutinin protein having overlapping predicted epitopes.

5.4 Common T cell epitope for class I and II MHC

Peptide sequences were predicted as epitopes and which are common in both class I and class II MHC molecules (Table 20). Two peptide stretch WTYNAELLVLLENERTL (Position 439-455) and YQILAIYSTVA (Position 531-541) that contain three and two epitopes shows positive results for both Class I and Class II MHC molecules.

Table 20: Epitopes which results for both MHC I and MHC II molecules.

SEQUENCE	POSITION	Class I	Class II
	Average length of HA protein 566 Amino acids	No. of tools that give positive result	No. of tools that give positive result
YNAELLVLL	441-449	5	2
VLLENERTL	447-455	4	2
WTYNAELLV	439-447	5	2
YQILAIYST	531-539	5	2
ILAIYSTVA	533-541	4	2
FWMCSNGSL	555-563	5	2

6. DISCUSSION

The rapid emergence of new viral pathogens, such as the recent avian and swine influenza strains, underscores the need for improved and expedited processes for developing and producing vaccines in response to such outbreaks. Timely responses to such outbreaks simply cannot be enacted by conventional approaches, thus, new immunoinformatics-based approach may have potential to solve the need of vaccines for these outbreaks. The new development in immunoinformatics and other computational methods have played an instrumental role in advancing peptide vaccine discovery, with promising results in melanoma, multiple sclerosis, malaria and anti-tumor vaccines (Tong JC *et al.,* 2006). Epitope specific T cell responses elicited by immunization with DNA or peptide and adoptive transfer of epitope specific T cell clones that could mediate protective immunity against various pathogens have been reported in murine experimental models.

We had taken HA protein, as subtyping of the influenza virus is done by the antigenic variations in the surface glycoprotein and HA is the major glycoprotein in influenza virus. Till date, 16 subtypes of HA protein have been recovered, out of which three are from humans.

In this study, 11 peptide sequences of 10-30 amino acid were found to be conserved (80%) in HA protein (566 amino acid) from 1918 to 2010 H1N1 strain which shows that HA is highly mutated as reported by Heiny *et al*., 2007. Many in-silico methods were available to predict T cell epitope, we had chosen tools which are widely used and reported in literature. Using BIMAS and SYFPEITHI prediction algorithm, a novel influenza matrix derived and HLA-A3- restricted 9-mer peptide RLEDVFAGK has been selected for immunogenic response (Trojan A, *et al*., 2003). In our study also we had used BIMAS and SYFPEITHI algorithm for class I MHC restricted T cell epitope.

We found 19 epitopes for class I MHC and 14 epitopes for class II MHC and five common epitopes for both class I and class I were predicted using different immunoinformatics tool. Four peptides sequences containing overlapping epitopes were finally selected in class I and II MHC as target for vaccine design. Three peptide sequences SLGAISFWM, STVASSLVL and VSLGAISFW which are predicted as epitopes for class I MHC molecules in our study are also shown as antigenic determinants by Immune Epitope Database.

It would be interesting to screen these peptides for immunogenic response in vitro. Our next step will be to synthesize these peptides and observe immunogenic response in human peripheral blood mononuclear cells. In one study, a number of $CD4^+$ T cell epitopes were predicted using epimatrix tool which are conserved between the 2008 and 2009 seasonal H1N1 vaccine strain and pandemic H1N1 (A/California/04/2009) hemagglutinin proteins and then biological evaluation were done using PBMCs from human donors not exposed to the pandemic virus (Schanen BC *et al.,* 2011). In this study, *insilico* results were found to be correlated with *invitro* result which shows that merging these two strategies to define potential epitopes have vast application in developing vaccines. Hence current insilico approach has significant potential in fulfilling the need of vaccine for current and future influenza attack.

7. SUMMARY

Influenza A viruses belongs 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.

Epitope-based vaccines offer an attractive strategy for improving protection against drift variants of seasonal influenza viruses and reduce the impact of future pandemic strains. This strategy requires the identification of T-cell epitopes. Experimental detection of T-cell epitope is expensive and time- consuming. Immunoinformatics algorithm has increased the pace of research from protein sequences to vaccine design.

In the course of our study, we have used various bioinformatics tools to determine peptide sequences in Hemagglutinin (HA) viral protein that act as epitope. Our approach was to find a stretch of immunogenic peptide from conserved peptide sequences as target for vaccine design which may be effective against current and future H1N1 influenza viruses as an alternative to current vaccines based on circulating virus strains.

The first step of our methodology was to retrieve various sequences of HA protein of H1N1 influenza virus from NCBI influenza database. The sequences were aligned using multiple sequence alignment tool MUSCLE. Conserved peptide sequences were find out from aligned sequences using AVANA. Then the conserved peptide sequences were submitted to different online tools for epitope prediction. We had used five programs (NETCTL 1.2, BIMAS, Syfpeithi, IEDB- ANN, and IEDB-SMM) to predict epitope recognized by class I HLA molecules. Three programs (ProPred, MHC2PRED, IEDB-SMM-align**)** were used to predict epitopes interacting with class HLA II molecules.

T-cell epitopes which were predicted for multiple HLA molecules are advantageous for vaccine design to maximize population coverage.

Several epitopes were obtained for both Class I and Class II molecules. Some epitopes obtained were common that could act as epitopes for both classes. 19 epitopes for Class I and 14 epitopes for Class II and six epitopes for both classes were obtained for HA protein of H1N1 virus. Two peptide stretch WTYNAELLVLLENERTL (Position 439-455) and YQILAIYSTVA (Position 531-541) that contain three and two epitopes shows positive results for both Class I and Class II MHC molecules.

Our next approach will be to synthesize these immunogenic peptides and further validated for immunogenic response in human peripheral blood mononuclear cells.

8. REFERENCES

Adar Y, Singer Y, Levi R, Tzehoval, Perk S, Banet CN, Nagar S, Arnon R, Yedidia TB: A universal epitope-based influenza vaccine and its efficacy against H5N1, Vaccine 27 2099–2107, (2009).

Bhasin M, Singh H, Raghava GPS :MHCBN: a comprehensive database of MHC binding and non-binding peptides. Bioinformatics, 19(5): 665-666, (2003).

Blythe MJ, Doytchinova IA, Flower DR: JenPep: a database of quantitative functional peptide data for immunology. Bioinformatics. **18**:434–9, (2002).

Both GW, Sleigh MJ, Cox NJ, Kendal AP: Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: Multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J Virol; 48(1):52–60, (1983).

Bian H, Hammer J: Discovery of promiscuous HLA-II restricted Tcell epitopes with TEPITOPE Methods: 34:468-475, (2004).

Bui HH, Sidney J, Dinh K, SouthwoodS, NewmanMJ ,Sette A: Predicting population coverage of T-cell epitope-based diagnostic and vaccines. BMC Bioinformatics 7:153, (2006).

Carrat F, Flahault A :Influenza vaccine: The challenge of antigenic drift. Vaccine 25 6852–6862, (2007).

Christy BS, Robert GW, Richard JW: Diversity of Influenza viruses in swine and the emergence of a novel human pandemic influenza A (H1N1). Influenza other Respi Viruses 1; 3(5): 207-213, (2009).

Finkenstadt BF, Morton A, Rand DA: Modelling antigenic drift in weekly flu incidence. Statist Med; 24(22):3447–61. (2005).

Garfinkel MS, Katze MG: influenza virus control of protein synthesis. Science & Medicine. 64 -73, (1994).

Gatherer D: The 2009 H1N1 influenza outbreak in its historical context. Journal of Clinical Virology 45, 174–178, (2009).

Guan, P, Hattotuwagama C. K.; Doytchinova, I. A.; Flower, D. R. MHCPred 2.0: an updated quantitative T-cell epitope prediction server. Appl. Bioinf., *5* (1), 55–61, (2006).

Hammer J: New methods to predict MHC-binding sequences with in protein antigens. Curr Opin Immunol. **7**:263–9. (1995).

Heiny AT, Miotto O, Srinivasan KN, Khan AM, Zhang GL, Brusic V, Tan TW, August T: Evolutionarily Conserved Protein Sequences of Influenza A Viruses, Avian and Human, as Vaccine Targets. PLoS ONE 2(11), (2007).

Hajjar SA and McIntosh K: The first influenza pandemic of the 21st century. Ann. Saudi. Med 30(1), (2010).

Hirst GK: Adsorption of influenza haemagglutinins and virus by red blood cells. J Exp Med 76:195-209, (1942).

Ishioka GY, Fikes J, Hermanson G, Livingston B, Crimi C, Qin M, del Guercio MF, Oseroff C, Dahlberg C, Alexander J, Chesnut RW, Sette A: Utilization of MHC class I transgenic mice for development of minigene DNA vaccines encoding multiple HLA-restricted CTL epitopes. J Immunol 162:3915–3925, (1999).

Jefferson T, Deeks JJ, Demicheli V, Rivetti D, Rudin M: Amantadine and rimantadine for preventing and treating influenza A in adults. The Cochrane Database of Systematic Reviews , Issue 3,(2004).

Khan AM, Miotto O, Heiny AT, Salmond J, Srinivasand KN, Nascimento E, Marques ET, Brusica V, Tan TW, August JT: A systematic bioinformatics approach for selection of epitope based vaccine targets. Cell Immunol; 244(2): 141-147, (2006).

Khan AM, Miotto O, Nascimento EJM, Srinivasan KN, Heiny AT: Conservation and variability of Dengue virus proteins: implications for vaccine design. PLos NegaI Trop Dis 2 (8),(2008).

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

Kielian M, Jungerwirth S: Mechanisms of enveloped virus entry into cells. Mol Biol Med.7(1):17-31, (1990).

Larsen MV, Lundeguaard C, Lamberth K, Buus S, Brunak S, und O, Neilsen M. An integrative approach to CTL epitope prediction. A combined algorithm integrating MHC-I binding, TAP transport efficiency, and proteosomal cleavage predictions. European Journal of Immunology. 35(8): 2295-303, (2005).

Liu W, Meng X, Xu Q, Flower DR, Li T: Quantitative prediction of mouse class I peptide binding affinity using support vector machine regression (svr) models. BMC Bioinformatics, 7: 182 doi: 10.1186/1417-2105-7-182s, (2006).

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

Meister GE, Roberts CGP, Berzofsky JA, De Groot AS: Two novel T cell epitope prediction algorithms based on MHC binding motifs; comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences.Vaccine, **13**: 581–91,(1995).

McNicholl IR, McNicholl JJ: Neuraminidase inhibitors: zanamivir and oseltamivir. Ann Pharmaco ther; 35:57-70, (2001).

MHC Sequencing Consortium: Complete sequence and gene map of a human major histocompatiblity complex. Nature 401:921-923, (1999).

Nielsen M, Lundegaard C, Worning P, Lauemoller SL. Lamberth K, Buus S, Brunak S, Lund O. Reliable prediction of T- cell epitopes using neural networks with novel sequence representations. Protein Sci., 12:1007-17, (2003).

Nielsen M, Lundegaard C, Brunak S, Lund O, Kesmir C. the role of the proteasome in generating cytotoxic T cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. Immunogenetics., 57(1-2): 33-41.(2005).

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

Osterhaus A, Garau J: update on the H1N1 influenza A outbreak. 19th Eur Congr Clin Microbiol Infect Dis (ECCMID) Abst, (2009).

Parker KC, Bednarek MA, Coligan JE: Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunol.; 152: 163–75, (1994).

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

Peters B, Sette A: Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6;132,(2005).

Parrish CR, Kawaoka Y: 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, (2005).

Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S: SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50:213–219, (1999).

Rammensee HG, Bachmann J, Stefanovic S: MHC Ligands and Binding Motifs. Landes Bioscience, Georgetown,(1997).

Seto JT, Rott R: Functional significance of sialidase during influenza virus multiplication. Virology 30, 731–737, (1966).

Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, et al: Mapping the antigenic and genetic evolution of influenza virus. Science; 305:371–6, (2004).

Singh H, Raghava GP Propred: prediction of HLA-DR binding sites. BMC Bioinfomatics; 17(12):1236-7, (2001).

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

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

Stephenson I, Nicholson KG: Influenza: vaccination and treatment. Eur Respir J; 17: 1282–1293, (2001).

Schanen BC, De Groot AS, Moise L, Ardito M, McClaine E, Martin W, Wittman V, Warren WL, Drake DR: Coupling sensitive in vitro and in silico techniques to assess cross-reactive CD4+ T cells against the swine-origin H1N1 influenza virus. Vaccine in press 29 (17) 3299-3309, (2011).

Shope RE: Swine influenza. J. Exp. Med. 54, 373 (1931).

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

Varghese JN, Laver WG, PM Colman: Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution, Nature 303, 35-40, (1983).

Tong JC, Tan TW, Ranganathan S: Methods and protocols for prediction of immunogenic epitopes. Briefings in Bioinformatics 8 (2) 96 -108, (2006).

Trojan A, Urosevic M, Hummerjohann J, Giger R, Schanz U, Stahel RA: Immune reactivity against a novel HLA-A3-restricted influenza virus peptide identified by predictive algorithms and interferon-gamma quantitative PCR. Journal of Immunotherapy 26(1):41-6, (2003).

Taubenberger JK and Morens DM: The pathology of influenza virus infections. Annual Review of Pathology: Mechanisms of Disease 3:499-522, (2008).

Taubenberger JK and Morens DM: 1918 Influenza: The mother of all pandemics. Emerg. Infect. Dis. 12 15-22, (2006).

Taubenberger, JK, Hultin, JV, Morens, DM: Discovery and characterization of the 1918 pandemic influenza virus in historical context. Antivir. Ther. 12 (4 Pt B), 581–591. (2007).

Tumpey TM, Basler CF, Aguilar PV, Zeng H, Solórzano A, Swayne DE, Cox NJ, Katz JW, Taubenberger JK, Palese P ,Sastre AD: Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus,Science Vol.310(5745),77-80,(2005).

Webster RG, Bean WJ, Gorman OT, Chambers TM, Kwaoka Y: Evolution and ecology of influenza A viruses. Microbial. Rev. 56:152-79, (1992).

Webster RG Laver WG: Preparation and properties of antibody directed specifically against the neuraminidase of influenza virus. J. Immunol. 99, 49–55, (1967).

www.pathmicro.med.sc.edu/mhunt/flu.html

www.umanitoba.ca/faculties/medicine/units/medical_microbiology/courses/MicroPath97.705/PDF/Influenza_report_mar2011.pdf)

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

Wharton SA, Weis W, Skehel, JJ, Wiley, DC: Structure Function, and Antigenicity of the Hemagglutinin in The Influenza Viruses**.** New York, Plenum Press 153-174, (1989).

www. proteopedia. Org /wiki /index.php /influenza_hemagglutinin#structure.

Young JF, Elliott RM, Berkowitz EM, Palese P: Mechanisms of genetic variation in human influenza viruses. Annals of New York Acad Sci.