Peptides containing multiple T cell epitopes in Carcinoembryonic Antigen

Dissertation
Submitted in the partial fulfillment of the requirement for the award of the degree of

MASTER OF TECHNOLOGY
IN
BIOTECHNOLOGY

Submitted by
Dibyangana Raul
601404006

Under the guidance of
Dr. Manoj Baranwal
Assistant Professor

DEPARTMENT OF BIOTECHNOLOGY
THAPAR UNIVERSITY, PATIALA (PUNJAB)-147004

JULY 2016

CANDIDATE'S DECLARATION

I, hereby declare that the work presented in the thesis entitled **"Peptides containing multiple T cell epitopes in Carcinoembryonic antigen"** in the partial fulfilment of the requirement for the award of the degree of Master of Technology in Biotechnology, Department of Biotechnology, Thapar University, Patiala, is an authentic record of my work during the period of one year from August 2015 to July 2016, under the guidance of Dr. Manoj Baranwal, Assistant Professor, Thapar University, Patiala. I have not submitted the matter embodied in this thesis for the award of any other degree or diploma.

Dibyangana Raul

Place: Patiala

Dibyangana Raul

Date:

(601404006)

CERTIFICATE

This is to certify that the thesis entitled **"Peptides containing multiple T cell epitopes in Carcinoembryonic antigen"** submitted by Dibyangana Raul in partial fulfilment of the requirement for the award of Degree of Master of Technology in Biotechnology to Thapar University, Patiala, is a record of student's own work carried out by him. The report has not been submitted for the award of any other degree or certificate in this or any other University or Institute.

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

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

Dr. S S Bhatia
Dean
Academic Affairs
Thapar University
Patiala

ACKNOWLEDGEMENT

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

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

I express my sincere thanks to **Dr. Dinesh Goyal**, Professor and Head, Department of Biotechnology (DBT), Thapar University, Patiala, for providing the best laboratory facilities.

A special thanks to **Miss. Neha Lohia and Mr. Sahil Jain** research scholars in DBT, for her prodigious help and valuable suggestions. I was fortunate enough to avail the ungrudging help and assistance from all our laboratory staff, **Mr**. **Ram Newal, Mr**. **Lallan** and **Mr**.**Surinder**.

Last but not the least I am thankful to my wonderful friends for their encouragement. I thank my friends **Ankita, Tania and Suchandra** whenever required. I am forever indebted to my **parents** for their love and support.

Dibyangana Raul
(601404006)

CONTENTS

ABSTRACT

The present study involves the identification of potential peptides containing multiple epitopes from a tumor antigen named carcinoembryonic antigen (CEA). This protein has been reported to be existing as raised levels in different tumor patients while in normal adults its existence is negligible. An immunoinformatic drive was performed to recognize the possible epitopes. Six immunoinformatic softwares were utilized to predict $CD4^+$ (HLA class I) and $CD8^+$ (HLA class II) T cell epitopes. Finally five peptide fragments from the protein were obtained that comprised of overlapping T cell epitopes of both the HLA molecules. Population coverage of these peptides was analysed to show their potential to elicit an immune response by a wide population of the world and the result indicates that these peptides are expected to response in wide populations. The binding affinity of the peptides with the HLA class I and II alleles were evaluated through peptide docking tool. Molecular docking results show that the binding affinity of these peptides with different HLA molecules is within the range of natural peptides. One peptide fragment was synthesized and analysed for its immunogenic property through PBMC proliferation assay *in-vitro*. Thus the study contributes a few peptide targets in CEA which may be considered as candidates for cancer vaccine design.

LIST OF ABBREVIATIONS

APC	Antigen Presenting Cell
ANN	artificial neural networks
BGP	Biliary Glycoprotein
CEA	Carcinoembryonic antigen
CTL	Cytotoxic T lymphocyte
CYT	Cytoplasmic Domain
COL1	Collagen 1
FDA	Food and Drug Administration
gp100	Glycoprotein100
HLA	Human leukocyte
HER-2	Human Epidermal Growth Factor Receptor
HMMs	Hidden Markov Model
IFNγ	Interferon gamma
IL2	Interleukin 2
KRAS	Kirsten rat sarcoma viral oncogene homolog
MART1/Melan-A/MAGE1	melan-A/melanoma antigen recognized by T cells 1
MUC1	Mucin 1
MD	Molecular Dynamics
MTT	3-(4,5-Dimethylthiazol-2-Yl)-2,5-Diphenyltetrazolium Bromide
NY-ESO-1	New York-esophagus-1 antigen
NCA	Non cross reactive antigen
PBMCs	Peripheral blood mononuclear cells
PSA	Prostate-specific antigen
PSG	Pregnancy specific glycoprotein
QM	Quatitative Matrix
QSAR	Quantitative Structure Activity Relationship
SCC	squamous cell carcinoma
SMM	Stabilized matrix method
TAP	Transporter associated with Antigen Processing
TAAs	Tumor associated antigens
mg	milligram
KDa	Kilo Daltons
ml	Millilitre
µg	Microgram
ºC	Degree celcius
µl	Microlitre

LIST OF TABLES

LIST OF FIGURES

Chapter 1: INTRODUCTION

The International Agency for Research on Cancer (IARC) accounted 8.2 million cancer deaths and 14.1 million new cancer cases worldwide in 2012. They predicted 13 million cancer deaths worldwide by 2030.Cancer includes about 200 forms of malady that share two common characteristics viz. uncontrolled growth and the capability to breach tissues hence damaging them (CRI, 2003). Cancer treatment has evolved over the years with the advent of radiotherapy and chemotherapy which provided a solution but wasn't satisfactory. Researchers recognized the potential of the immune cells for therapeutic treatment against cancer. They also established the fact that the immune system normally is unable to fight against cancer because the tumors do not activate the immune system efficiently enough to produce either humoral or cell mediated response against them. Anti-tumor monoclonal antibodies, adoptive T cell transfer by ex-vivo numeral expansion and vaccination are the immunotherapeutic approaches that have paved a novel path in cancer treatment field (Gajewski et al., 2012).

Anti-cancer vaccination has been the most impressive and effective ways of treatment as it trains the immune system to specifically destroy the tumor cells. The only FDA approved therapeutic vaccine is Sipuleucel-T (Provenge^{TM}), It is a dendritic cell vaccine which is used against prostate cancer. In this approach, dendritic cells from patients are collected and re-injected back after labelling with the prostate acidic protein thereby inducing the T-cell response against cancer (Guo et al.,2013).The limitation to these kinds of vaccines is that they are specific to a person and are not effective against a larger population.This challenge is overcome with the utilisation of peptide vaccines which provide an opportunity to target a wide range of population through one single vaccine formulation. The ease in production, small size, chemical stability and biocompatility of the peptide certifies the peptide vaccine to be counted amongst the most desired cancer therapies (Patronov and Doytchinova,2012).With the discovery of Tumor associated antigens (TAAs), targeting cancerous cells through vaccination has been made possible. TAAs are specific cancerous cell markers which are recognized by the immune cells and induce them to kill the tumor. Several categories of tumor antigens have been described and a peptide database available at cancer immunity web site has reported 403 well defined tumor antigenic peptides till 2013 (Vigneron et al., 2013). TAAs such as tyrosinase, gp100, Carcinoembryonic antigen, MART1/Melan-A, NY-ESO-1, p53 and HER2/neu proteins have been reported to contribute

cytotoxic T cell epitopes that would lead to the destruction of cancerous cells (Kobayashi et al., 2002). Many of these peptides have been considered as potential targets for vaccine formulation and are reported to be studied at different stages of clinical trials.

Carcinoembryonic antigen (CEA) protein has emerged as a potential target for vaccination against cancer as links have been found between its rising levels with the increasing growth of various tumors like colorectal cancer, head and neck cancer, non small cell lung carcinomas, pancreatic carcinomas, breast cancer and gastrointestinal cancers (Campi et al.,2003). The CEA is a glycolipid rich transmembrane protein of molecular weight 180kd (Ojima et al., 2006). The use of a recombinant vaccinia-CEA vaccine in patients to generate human cytotoxic T lymphocytes(CTLs)against the CEA epitopes(Tsang et al.,1995),b)autologous dendritic cells bearing CEA peptide in malignant patients to develop CEA specific T cell response (Nair et al.,1999), c) anti idiotype antibody vaccine to generate immune response against patients having raised CEA protein(Foon et al.,1995) and many similar cases were reported over the years which validated the potential of CEA protein to contribute immunogenic epitopes.

An immunoinformatic approach which included epitope prediction softwares was carried out to identify promising peptide fragments from the whole CEA protein. These prediction softwares determined epitopes based on the protein sequence and their 3D structure. Among these *in-silico* determined peptides only the ones which were anticipated to promote both B cell and T cell mediated response are the ones believed to be most efficient and acceptable as candidate for vaccine design. Further *in-vitro* and *in-vivo* experimentation of these peptides in simulated environment and animal models would help it to proceed for clinical trials. Considering CEA protein as a favourable vaccine target as reported by many research studies, the present study focuses on finding of peptides containing multiple epitopes in this protein based on immunoinformatics approach. With the help of different *in-silico* tools, peptide candidates were perceived that bear the capability to activate immune responses. These determined peptides were further validated by analysing their binding efficiency different HLA class I and II based on molecular docking approach.

Chapter 2: REVIEW OF LITERATURE

Cancer is a global public health problem causing deaths worldwide. There are about 100 different types of cancer identified till date. Among the major ailments, the annual death count due to cancer is only next to cardiovascular diseases (CDC, 2016). According to the International Agency for Research on Cancer (IARC), in the year 2014, fourteen million cancer cases were reported worldwide, of which eight million cases were recorded from the part of the world which consisted of 82% of the world's population. However, it is expected that cancer shall be the leading cause of deaths in a few years (Siegel et al., 2015). Cancer is characterized by abnormal growth of cells and their ability to damage other tissues leading to the generation of tumor lumps, vascularization and metastasis. Over the years, advances have been made in the treatment of metastatic cancers using surgery, radiotherapy, chemotherapy and targeted therapies but significant amelioration in the survival rates is yet to be achieved. Thus, immunotherapy has emerged as an alternative approach in the field of cancer treatment (Gajewski et al., 2012).

After years of research, scientists have come to the conclusion that our immune system is the perfect weapon to fight cancer because of its remarkable specificity. Antigen recognition is the striking property of the immune system which makes it so specific and at the same time, it has attracted investigators to look at the immune-based therapy as a possible treatment for cancer. The tumor associated antigens that are exclusively present on the cancer cells paved the way to success, for the development of targeted vaccines and approaches mediated through T cells directed against tumors. Moreover the awareness about the relationship between the patient's immune response and the growing tumor brought forth abundant scope for this specific cancer immunotherapy. It has been studied that the immune cells do respond against the tumor cells but their action gets suppressed when the tumor cells outnumber them. Hence a primary aim of this therapy has been to elevate the count of tumor antigen-specific T cells. Immunotherapy is the novel field in cancer treatment that has made shown promising results in treating cancer. Cancer immunotherapy includes two crucial propositions viz. the administration of antigen-specific vaccines and adoptive transfer of anti-tumor T cells (http://www.cancerresearch.org).

2.1 Cancer vaccines

The efficacy of different types of vaccines against various cancer types has been studied. These vaccines may include peptides, proteins, recombinant viruses and microorganisms, killed tumor cells and peptide activated dendritic cells (Table 1) (Schlom, 2012).

Table1: Different vaccination approaches that have been used against cancers.

Vaccination approaches	Cancer type	References
1)Peptides/proteins		
Peptide	Melanoma	Schwartzentruber et al., 2011 and Disis et al., 2011
	Lung	Butts et al., 2005
Protein	Melanoma	Karbach et al., 2011
Antibody	Lymphoma	Schuster et al., 2011
Glycoprotein	Melanoma	Gilewski et al., 2007
2)Recombinant vectors		
Poxvirus	Prostrate	Moss et al., 1996
Saccharomyces cerevisiae (yeast)	Pancreatic	Wansley et al., 2008
Listeria	Pancreatic	Singh et al., 2006
adenoviruses	Carcinoma	MacDonald et al., 2000
3) Tumor Cells		
Autologous	Colon, melanoma	Hoover et al., 1993 Luiten et al., 2005
Autologous tumor cell fusion	Myeloma	Avigan et al., 2004
Allogeneic	Pancreatic	Lutz et al., 2011
4)Antigen presenting cells(APCs)		
APC-protein	Prostrate	Higano et al., 2009
Dendritic cell-peptide	Glioma	Banchereau et al., 2001

A cancer vaccine is required to induce antigen specific $CD4^+$ T helper cells that lead to the production of cytokines which enhances cytotoxic T lymphocyte (CTL) activity and antigen specific $CD8^+$T cells that which indulges in tumor lysis. Hence they are able to mount a strong attack against the cancerous cells. These features make immunotherapy the most reliable one in the field of cancer treatment. Till date, only one immunotherapeutic cancer

vaccine, Provenge® (Sipuleucel-T) for the treatment of prostate cancer has received FDA approval (Kantoff et al., 2010).

2.2 Tumor associated antigen Peptides and their T-Cell Immunogenicity

Research works carried out in the last few decades ascertained the fact that peptide vaccines are the most effective considered to other vaccination methods. The sequencing of the human genome has led to the revelation of many novel proteins, some of which can act as a plausible target against tumor cells. Intracellular processing of proteins forms short peptides of mainly 8-10 and 13–20 amino acids which bind with class I and II HLA molecules respectively (human leukocyte antigen). HLA I present peptides on the surface of all nucleated cells while HLA II on antigen presenting cells. Studies revealed that divergent HLA alleles bind only those peptides that share preserved amino acid residues at particular positions (Falk et al., 1991). Crystallographic studies of the peptide and receptor binding sites revealed conformational similarity for both class I and II molecules. HLA class I molecules generally bind at the terminals of the bound peptide, leaving the middle region (Singh and Raghava , 2001). These interactions between the peptide binding groove and the amino and carboxyl terminal restricts the peptide to interact with only 8-10 amino acid residues. In case of class II molecules, the peptide binding groove is open at the terminals and the peptides can interact with a more variable length of 10-28amino acids (Singh and Raghava , 2001). The capability of the peptide to bind with the appropriate HLA allele is facilitated by the location and chemical character of the amino acids and this is declared as the peptide binding motif. The T lymphocytes can recognize the peptide-HLA complex with the help of the T cell receptors (TCRs) present on their surface. The T cell response generated by a tumor associated antigen (TAA) peptide is determined by its ability to bind HLA allele as well as the affinity of the peptide HLA complex affinity towards TCR. A number of TAAs and their peptides that are recognized by tumor-specific T cells have been identified through different methods. In 1991, first TAA was reported from melanoma antigen-1 (MAGE-1) protein (Van der et al., 1991). MAGE-1 protein nonamer peptide was diagnosed by human leukocyte antigen (HLA)-A1 restricted CD8+ T cells in the following year by Traversari et al., 1992. TAAs identified so far are generally expressed by melanoma while only a few of these identified TAAs are found to be expressed in other common tumors such as breast, lung, colon, pancreatic and gastric tumors express. In addition, only a few HLA alleles that are universally present throughout a population do spot the known TAA epitopes, limiting the number of epitopes for recognition

by T cells. Henceforth patients are HLA typed so that they can resort to the appropriate peptide as distinct polymorphic class I molecules express disparate peptides on them. Investigation of the patient tumor has been carried out in parallel to identify the antigens of interest. Therefore such studies require the consideration of both the above parameters. For a vaccine to generate an effective response, it is desired to activate both antibody dependant and cell mediated immunity. It has been found that some TAA peptides like Her-2/neu that are expressed on neoplastic cells, induce both antibody mediated and cell mediated immune response (Jager et al.1998). Some of the other well known TAAs that fall in this category include Carcinoembryonic antigen(CEA)(von Mehren et al.,2000), New York-esophagus-1 antigen (NY-ESO-1)(Caballero et al.,2000),MART-1, gp100, MELAN A(de Vries et al.,2001), prostate-specific antigen(PSA)(Gulley et al.,2010), MUC1, mutated antigen k-Ras and p53(Wolfel et al.,1995).

2.3 Carcinoembryonic antigen

Carcinoembryonic antigen (CEA) was initially isolated from a colon carcinoma specimen in 1965 by Gold and Freedman. Human CEA is a 180kDa glycoprotein commonly expressed in carcinomas of the colon, breast and lung (Averbach and Sugarbaker, 1995). Generally it is found at high levels in the fetal colon and low levels in normal adult. Also, elevated levels of CEA have been observed in individuals having tumor. CEA is also expressed by epithelial cells in non malignant diseases including pancreatitis, cirrhosis, hepatitis, bronchitis and renal failure (Jothy et.al., 1986). This fact makes the serum CEA determination method controversial for screening of cancer. But the discrepancies were removed with the discovery of various other proteins like Pregnancy specific protein and biliary glycoprotein which are a part of the CEA gene family but are not identical to CEA protein. CEA is considered to be one of the most important tumor markers for both diagnostic and prognostic matters.

2.3.1 CEA gene family

Till date 24 genes and pseudogenes of the human CEA gene family have been discovered. This gene family can be divided into 3 subgroups based on their nucleotide sequence similarity.

1. The CEA subgroup contains 12 members. Seven of these members are expressed while the others are pseudogenes (CGM8, CGM9, CGM10, CGM11 and CGM12).
2. The Pregnancy specific protein (PSG) subgroup containing 11members where all of them are expressed.
3. The third subgroup contains six members where all are pseudogenes (CGM13 to CGM18).

The CEA family is loacalized on the chromosome 19, with 1.8 Mb regions between CY2A and D19S15 (Thompson et al., 2001).The PSG and 3rd subgroup genes are located terminal to the CEA genes. These two clusters of genes (250kb and 850kb) are separated by a region of approx 700kb.The 3rd subgroup genes are distributed in between the PSG genes.

At the extreme end of the CEA protein chain, there is an amino terminal domain continued with three disulfide linked repeating domains of 178 amino acids which share extensive sequence similarity with immunoglobulin gene family depicting an evolutionary connection. The hydrophobic carboxyl terminal segment lies at the other end of the chain serving as the domain anchored within the membrane. Two Ig domains are seen here – an N-terminal domain consisting of 108 residues that is similar to the Ig variable domain and six domains similar to the Ig constant domain.

The CEA members are attached to the cell surface membrane via two kinds of attachment. BGP, CGM7 and CGM1 comprise of a trans-membrane domain that is hydrophobic followed by a cytoplasmic domain (CYT). While, CEA, CGM6, CGM2 &NCA are attached through a glycosylphosphatidylinositol moiety. The most important difference between CEA and PSG is that CEA remains attached to the transmembrane while the PSG molecules are secreted out of the cell and remains as a free molecule. A short hydrophilic tail present in the PSGs helps them to get exported secretory vesicles to the maternal circulation (Hammarstrom , 1999).

Two features that are observed in all expressed CEA genes are-

- They comprise of an amino domain with the absence of intra chain disulfide linkage.
- The molecules are heavily glycosylated on the asparagine residues. Carbohydrate composition can be upto 50% of the total mass.

The genes of this family contain both housekeeping and regulated genes.

2.3.2 Function of CEA family genes

Different proteins of CEA family play various roles. CEACAMs or CEA related adhesion molecules perform a crucial part in the cell to cell adhesion where it's necessary to transport cells through the functional organs during embryonic development (Pavlopoulou et al.,2014).This family also serves as receptors for pathogens of bacterial and viral origin such as *Haemophilus influenza*, *Nisseria gonorrhoea* etc. They bind to these proteins with help of their IgV like N-terminal domain (Bos et al., 2002).PSG protein secreted from the trophoblasts of fetus is expected to monitor the interactions between mother and fetus during pregnancy.

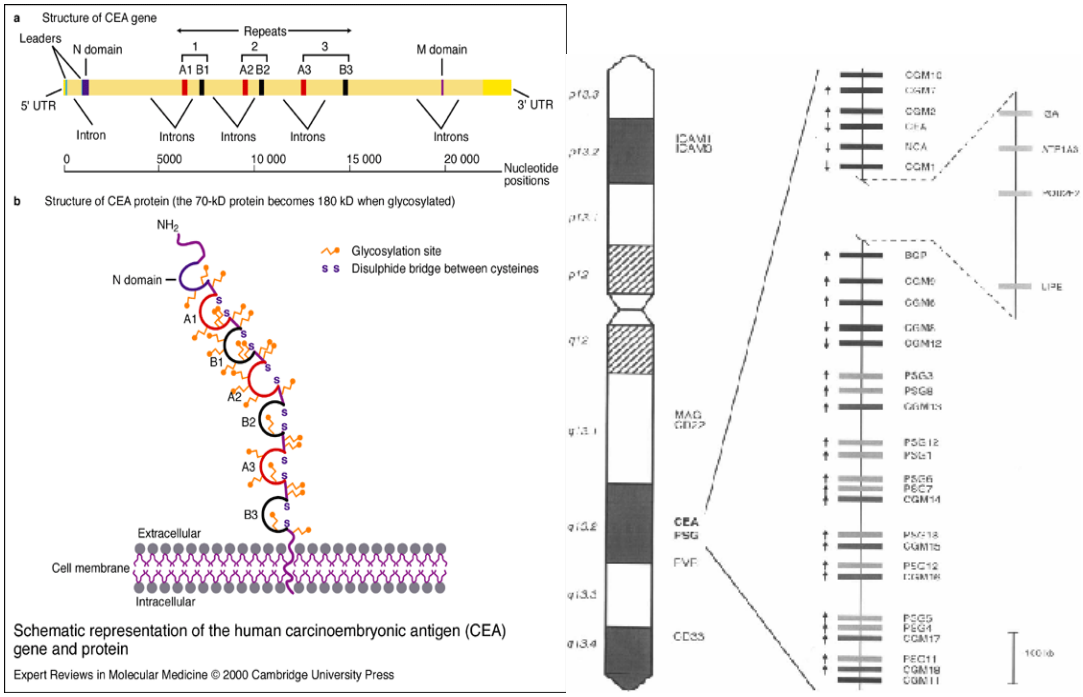


Figure 1 . The gene structure of the CEA gene family and the molecular structure depicting the three repeated domains and glycosylated sites of CEA protein. The chromosome idiogram pointed out the q13.2 region as the location of the CEA family. Locations of all the 29 individual members are indicated.(Hammarstrom., 1999)

2.3.3 CEA –a potential tumor marker

During the initial fetal developmental stage, CEA is found in the epithelial cells, colon cells, cells of mucous, neck and pyloric cells. CEA is an adhesion molecule which plays different functions in normal and neoplastic colon tissue. It helps to differentiate the localization of two tissues.The intercellular adhesion is interrupted due to the modified pattern of localization in tumor cells

leading to movement and unregulated growth of malignant cells,thus clearly indicating its purpose in the evolution of the metastatic disease. Ordonez et al.,(1998) showed that CEA inhibits cell death and established that it helps several proto oncogenes(as BCL2 and c-myc)in cellular transformation.CEA has been found to be immunogenic according to several reports.Its potential to generate a T cell response was firstly observed in individuals with colon cancer. It has been already reported that a few CEA epitopes have been recognized by $CD8^+$ T cells (Tsanget al., 1995 & Nukaya et al., 1999) and $CD4^+$ T cells (Kobyashi et al., 2002).The reasons that make CEA a potential target for immunotherapeutic purposes are its stability,its restricted expression in normal adult tissue while raised levels in positive tumors and lastly its immunogenicity. Other CEA like proteins viz.BGP (biliary glycoprotein) & NCA (non cross reactive antigen) are present in higher levels in the blood of a normal adult as compared to CEA and thus, are unlikely to be used as critical markers.

2.4 **Occurence of CEA in different tumors**

Many tumor markers have been identified till date but CEA is one of the few which has been characterized and traced in a vast number of tumors. It has also been considered as a standard for comparison of new serum markers.

2.4.1 Breast Cancer

CEA has been reported in breast tumors by many research groups including Harvey et al who determined the existence of CEA in considerable amount in metastatic liver of primary breast cancer patients. In a study, it was reported that of the 124 patients with primary breast cancer ,11% showed an elevated level of CEA.A study by the same group showed that in 45 cases of recurrences of breast cancer, 47% of the patients showed an increase in the CEA level(O'Dwyer et al.,2009).It has been proved by several researchers that any increase or decrease in the CEA level reflects the progression or regression of the disease. Thus CEA has been stated to be useful in the postsurgical follow up of breast cancer patients for diagnosis of reoccurrence and for monitoring of response to treatment (Nicoliniet al., 2006).

2.4.2 Colon Cancer

Patients at the initial stages of colon cancer have a low sensitivity for CEA but the sensitivity increases with the growing stages of the disease.Winawer et al., 2003 studied every stage of the disease and did the analysis for the sensitivity and specificity at each stage. The sensitivity was

accounted to be 36% and specificity to be 87% for a CEA greater than 2.5 ng/mL in patients at stage I and II of the disease. The sensitivity for stages III and IV of the disease having similar CEA levels was found to be 74% and 83%, respectively. Investigation shows that colorectal cancers that are well differentiated produce greater amount of CEA per gram of total protein than the poorly differentiated cancers. For instance, a recent report has shown average concentrations of CEA in colorectal tumors that may be categorized as well-differentiated, moderately differentiated, and poorly differentiated to account to 18.0, 5.5, and 2.2 mg/g of protein, respectively.

2.4.3 Head and Neck Cancer

The head and neck cancer (HNC) has been estimated to account for 4% of all the cancers occurring in the United States and many other parts of the world each year. A study showed that out of 69 cases of squamous cell carcinoma (SCC) of the head and neck, the majority responded to be positive for CEA when an immunohistochemical analysis of their tumor tissue was performed using COL-1,a CEA specific monoclonal antibody(Kass et al.,2002). Yanagawa T et al., 1986 checked the presence of CEA in 45 squamous cell neck and head cancers in which 19 showed a positive response for the presence of CEA. The levels of plasma CEA was determined to be upraised remarkably in the stage-IV patients only. It was also noted that three patients who had plasma CEA values greater than 5 ng/ml before the therapy exhibited a CEA level decrease to below 2.5 ng/ml with the cancer treatment leading to complete remission. These facts clearly denoted CEA to be a tumor marker for squamous cell head and neck cancer.

2.4.4 Lung Cancer

A report presented the data that 69% and 68% of CEA sensitivity and specificity respectively was diagnosed in lung cancer patients. Study was carried out to determine the CEA levels in lung cancer patients where the standard level of CEA was considered to be 3.2 ng/mL. This study tells that the number of patients having increased levels than the usual cut-off level were 453 (69%) in lung cancer (Zateska et al., 2010). The increase in the level of CEA is dependent on the stage at which the cancer exists. It has been reported that 60% of the patients in the stage I and II showed CEA levels greater than the standard quantity while 78% of the patients in the stage III and IV showed increased levels of CEA.

2.5 IMMUNOINFORMATICS

Immunological research has been immensely augmented by the accelerating growth of different techniques and applications in the field of bioinformatics. This drive has led the field of computation immunology towards prosperity along with the immunology dependant resources and software which help in interpreting the properties of the immune system (Brusic et al., 2005).This has brought forth the inception of a new field entitled as immunoinformatics.

This field accelerates the time involved in the process and reduces the cost needed for laboratory analysis by focusing mainly on the research and formulation of algorithms that potentially maps B- and T-cell epitopes. Discovery of new vaccines can thus be established by analyzing the sequences having potential binding sites with the support of these immunoinformatics tools and information. This method of predicting the prospective antigenic proteins or peptides from the whole genome of an organism is defined as 'reverse vaccinology'(Tomar et al.,2010).Identification of structural binding motifs, matrix driven methods, protein threading, docking techniques,homology modelling,QSAR analysis and design of various machine learning algorithms are the different approaches in this field. New and upgraded algorithms and tools make the process painless by enhancing the predictive performance unlike old days when these techniques deduced the sequence characteristics only. These prediction model development programs can be classified into a) sequence based methods that explore the amino acid sequence and b) structure based method that retrieve the information from the three-dimensional arrangement of the proteins.

2.5.1SEQUENCE BASED METHODS

The presentation of potential epitopes by HLA class I and II involves protein degradation, relocation of peptides to the endoplasmic reticulum (HLA-I), compartmental cleavage of lysosomes (HLA-II), binding of antigen to HLA I and HLA II binding with antigen, HLA haplotype specificity, and T cell receptor identification. Each of these steps are predicted via different algorithms along with combination algorithms which incorporate the individual predictions too. The essential steps comprised in these techniques are illustrated below.

2.5.1.1 Protein cleavage

Protein degradation is the initial step that depends on the proteosome complex. Cleavage prediction programmes are trained utilizing two sources of data. One involves the peptide

sequences developed from the *in vivo* degradation of particular proteins. Digestion assays carried out in the laboratory indicate that despite of the presence of favoured cleavage sites, the protein is available for cleavage at random sites resulting in formation of a pool of peptides which might be overlapping. The fact that only few proteins have been studied in this way and a huge set of data reviewing other existing proteins needs to be investigated brings out the limitation of this high-quality data. The other resource of data is the naturally presented peptides separated from HLA proteins that are assumed to be cleaved by the proteosome complex and thereby prove suitable to train predictors in union with the source protein sequences. Here, the limiting factor is that when a resulting peptide does not bind to HLA, protein cleavage sites are missed and the protein is not included in the peptide dataset.

2.5.1.2 Transporter associated with antigen processing

The next step is the transportation of the cleaved peptide sequences into the endoplasmic reticulum via TAP protein. Tapastin, calreticulin and ERp57 are the proteins that help in accomplishing this job (Abele et al.,2004).The deciphering of the motif of human TAP for peptide binding through combinatorial libraries has indicated its preference for particular amino acids at particular positions.

The final step in antigen presentation is the binding to HLA molecule. This is the most restraining step involved in antigen presentation. It is observed that less than 1% peptides bind a given HLA class I allele with energy sufficient to evoke a CTL response (Yewdell et al., 1999). HLA peptide binding has been predicted by various approaches and ways deduced for overall prediction of the entire presentation pathway.The arrangement of suitable amino acids at the particular peptide anchoring sites is characterized as a motif. With the help of a predefined motif library the amino acid sequence of a peptide is screened for motifs. Thus, HLA binding motifs for a particular peptide is identified by comparing the known binders and non binders. SYFPEITHI is a well known epitope prediction tool which has been developed based on the motif search approach (Rammensee et al., 1999 and De Groot et al., 2001). This method evaluates the immunogenicity of peptides by scoring them. For describing non linear data and supplying convenient method to sort relationships, researchers have come up with a new approach described as the artificial neural networks (ANNs). Utilizing the fundamentals of this model, the NetCTL server carries out the prediction of HLA class I binding peptides, C-terminal cleavage at the proteosome complex and transport efficiency associated with TAP transporter protein. This program has been

efficiently trained with the data of 67 HLA alleles and 55 HLA peptides.BIMAS is another peptide prediction tool that is based on the Quantitative matrix (QM) method (Parker et al., 1994). QM is a linear model with an advantage that it covers a vast range of potential peptides. BIMAS and SYFPEITHI servers are not only good performers in the prediction of known epitopes, but are quite impressive in search for novel epitopes.

2.5.2 STRUCTURE BASED METHODS

Rather than relying on binding data and sequence related information, these methods choose to exploit the structural information and further work in alliance with computational methods developed in structural biology field for prediction of potential good binders. Geometric and electrostatic adjustment of the receptor and ligand are an essential aspect for a stable complex formation proving to be crucial point for the antigenic peptide recognition by the HLA molecules. Residues convenient for the binding are deduced when peptides are aligned along sequences previously known to bind a given HLA molecule. The structure based methods generally involves the following steps illustrated below.

2.5.2.1 Peptide library screening for peptide docking

Screening of peptide library and ligand docking are the techniques commonly used in the field of drug design which are purposeful for bioinformatics. HLA class II models were also designed by analyzing the overall peptide affinity of a given amino acid (Davenport et al.,1995).The frequency in which the amino acid is present in a certain position was noted. Stryhn et al., (1996) studied the peptide library to analyse the specificities of peptide as HLA class I binders. The main purpose of simulation in docking is to search every possible translational, rotational and conformational proximity of a given pair of ligand and receptor pair and calculate their relative binding energy. The peptide structures are at first modelled through different bioinformatic tools (eg. PEPFOLD) and then are docked to the binding groove. Docking also helps in studying the synergy between the T cell receptor and the HLA ligand complex. An approach of docking which included the following steps (i) anchor residue docking; (ii) arrangement of the peptide backbone in the binding groove; and (iii) adjustment of the final position of the peptide and its side chains, showed improvement in accuracy. The prediction accuracy supremely depends on the structural information accessible for the receptor and correctly modelled structure of the peptide.

2.5.2.2 Threading algorithms

Threading algorithms that rely on information helps distinguish between the binding and non-binding peptides for a distinct HLA molecule. Inputs of each amino acid residue throughout the peptide that urges them to adhere the binding groove of HLA molecule are taken into account by these algorithms. These algorithms use threading to identify the peptides' anatomy by rearranging them over the known framework from X-ray study of HLA complexes. Combining the individual score of each amino acid at every single position the overall binding affinity is calculated (Altuvia et al., 1997). Higher affinity is designated by lower score of the peptide. Inspite of the close enough similarity between the standard and tested peptides the orientation of some side chains in different directions worsen the predictability thus emerging as a shortcoming for these algorithms (Patronov et al.,2013).

2.5.2.3 Binding energy and molecular dynamics

This method recognizes epitopes by calculating the Gibbs free energy while the complex formation between ligand and receptor and this is described as energy difference of the free and the bound peptide (Sezerman et al., 1997and Zhang et al.,2000).With help of scoring functions or molecular dynamics (MD) simulations comparison of the free energies of two peptides can be done which would lead to the identification of epitopes. This method is used for studying the binding of synthetic peptides (Scapozza. et al., 1995).A major flaw noted in this method is that due to huge amount of time and computational power required it cannot be accessed online.

2.6 *In-vitro* analysis of immunogenicity

Though *in-silico* approaches aids the peptide search process for vaccine design to a far extent but without undergoing the *in-vitro* and *in vivo* tests they completely cannot qualify as potential vaccine targets. Thus the *in silico* determined peptides are tested *in vitro* for their immunogenicity response on naive lymphocytes collected from healthy individuals. Limitations to the algorithm-based prediction programmes: (i) there may be epitopes which are tolerant to T cell response and access to them is restricted, (ii)amino acids in the neighbour of the functional region or outside the nonamer can influence the response which remains undetectable (iii)there is less knowledge about the stability of the peptide in the HLA pocket , (iv)whether the peptide is able to gather the T cell activation thresholds v)B cell response and lastly vi) post translational modifications can lead to misfolding of the protein that are not predicted by these sequence based algorithms. The software predicted peptides

are quite often found inefficient to mount a quality immune response and thus it is evident to check the immunogenic potential of these peptides. The main purpose of this *in vitro* analysis is to determine whether these peptide have the ability to induce immune response. This is generally done by collecting the peripheral blood mononuclear cells (PBMCs) from healthy donors, harvesting them and challenging them with target peptide stimulation. Within the body, when the epitopes of TAAs are expressed on the APCs in aid with the HLA class II molecules then CD4+ cells are activated and they release chemical messengers such as the cytokine,IL-2 and gamma interferon (IFN-g).The $CD4^{+}$T cells helps the immune system to build up an army of antibodies against the TAA epitopes. These messengers further helps in proliferation of the lymphocytes activating the $CD8^{+}$T cells. These cells tend to recognize the HLA class I antigen complex on the cancerous cells and secrete perforins and granzymes that lyse the cancerous cells. The T memory cells are generated which on further attack can immediately take action and degrade the tumors. In *in-vitro* culture, IL2 is provided externally to the lymphocytes to mimic the environment of the lymph node. The response is estimated in the form of cytokines produced by the cultured PBMCs which is measured through an ELISA test and the cell viability is estimated through MTT test.

Chapter 3: MATERIALS AND METHODS

3.1 Sequence Retrieval

The full length sequence of our target protein Carcinoembryonic antigen (CEA) was retrieved from the NCBI protein database bearing accession version AAA51967.1(Barnett et al., 1988).

3.2 Prediction of HLA Class I binding epitopes

HLA class I molecules bind to epitopes, expresses itself on the surface of nucleated cell. This HLA-peptide complex binds to the T cell receptor on the surface of $CD8^+$ T cells. In general, these epitopes are derived from endogenous intracellular proteins formed by proteasome-mediated cleavage of proteins in the cytosol. The HLA class I molecules bound peptides commonly have a length of nine amino acids and they contain specific amino acid residues that appear to be essential for binding to a particular HLA molecule. Three different epitope prediction tool were employed to predict HLA class I epitopes.

3.2.1 NetCTL 1.2 Server

The NetCTL is a prediction tool developed by the Center for Biological Sequence Analysis (CBS) at the Technical University of Denmark (Larsen et al., 2007). This tool determines the HLA peptide binding using the artificial neural network algorithm while the efficiency of transportation through TAP is predicted via weight matrix based method. This tool predicts the HLA class I binding epitopes for 12 different supertypes of HLA class I (HLA A and B). It has been trained for 886 known ligands of the HLA class I type. The method carries out the separate prediction of cleavage of the protein carboxyl terminal and efficiency of the TAP transporter molecule and finally binding of HLA class I with the peptide. At last integrating the scores from the three individual predictions as a weighted sum and taking a relative weight 1 for peptide HLA binding the combined score is displayed in the output result. The outcome appears in the form of IC50 values which is a log transformed value in nM units. The weight on the C terminal cleavage and TAP transport efficiency was set as 0.15 and 0.05 respectively in our experiment. The server after analysing the binding of more than 800 known ligands with HLA I have come up with a list of the sensitivity of the tool for the different threshold values of identifying the peptides were taken. Following this list we have taken the threshold value greater than 0.75 for which the sensitivity calculated is 0.80.

3.2.2. BIMAS

Bioinformatics and Molecular Analysis Section (BIMAS) is developed at the Centre of Information Technology of National Institute of Health (Parker et al., 1994).The tool has been trained for 152 peptides. Using this binding data and the position of these peptides on the epitope it predicts the potential epitopes. BIMAS predicts the binding of epitopes for 33 different HLAs. It mainly calculates the binding affinity in terms of half-time of dissociation of theβ2 microgobulin from class1 HLA molecules. Of the 152 peptides, 80 peptides that have been included have shown a half-life of binding with β2 microglobulin more than 5 min and are used for the prediction of epitopes. While the rest 72 peptides having half-life less than 5 min have been introduced in order to knock out the false prediction. The tool has been programmed to identify 8-mer, 9-mer or 10-mer peptides. The threshold value for epitope prediction was taken as $T_{1//2}$ equal to 50.

3.2.3 SYFPEITHI

This tool determines the HLA binding potential epitopes by evaluating every amino acid of the peptide (Rammensee et al., 1999). It dedicates an arbitrary value 1 to amino acids that are least preferred in the respective position while a value 15 for residues that optimal in terms of preference. Thus each amino acid residue is dedicated a score in between these two values. Residues which harm the binding capacity at a certain sequence position are given negative values. The frequency of an amino acid residue in natural ligands, T cell epitopes or binding peptides determines the allotment of values. Threshold value taken for epitope prediction was set to 20.

The epitopes predicted by all the three tools were taken for further consideration.

3.2 Prediction of HLA Class II binding epitopes

The CD4 restricted T cell epitopes are the ones which are presented on the antigen presenting cells (APCs) by HLA class II molecules. These HLA molecules induces cellular and humoral immune response through display of extra cellular peptides to T-helper cells through the release of chemicals like cytokines and inducing the growth of lymphocytes thereby activating the $CD8^{+}$T cells. Thus identification of these peptides HLA binding which can clarify the host pathogen interaction needs to be exposed. Various tools have been designed over the years to predict these $CD4^{+}$ T cell epitopes. We have identified these epitopes with the help of three such prediction tools.

3.3.1 NetMHCII 2.2 Server

This server has been been developed by the Center for Biological Sequence Analysis (CBS) at the Technical University of Denmark (Nielsen and Lund, 2009). This server predicts the binding of human HLA class II molecules (HLA-DQ, HLA-DP and HLA-DR) utilising the artificial neural network algorithm. 14 HLA-DR alleles with their 9 supertypes, 6 HLA-DQ and 6 HLA-DP are included within this tool for which the peptide bindings are predicted. The results are given in the form of IC50 values in nM units. The results also show that which method and HLA alleles were selected along with two threshold values, which declares the high binding and weak binding peptides. Peptides with IC50 value less than 50nM are defined as high binding peptides while an IC50 value less than 500nM defines it as weak binding peptide. Here we have considered only the strong binder peptides.

3.3.2 PROPRED

This tool has been developed by the Bioinformatics centre at IMTECH (Singh and Raghava, 2001). The tool carries on the prediction programme utilising the Quantitative matrices algorithm acquired from the literature published by Sturniolo et al., 1999. The literature provides a table where amino acid positions for the epitopes have been reported. Predictions are performed by the tool with reference to that table. The tool sets a percentage threshold parameter which defines the 'percentage of best scoring natural peptides'. This indicated that setting the threshold of 1% would return the peptides for any protein sequence which are ranked in the 1% best scoring natural peptides. The percentage threshold was set at 3% for the present work.

3.3.3 IEDB SMM Align

IEDB stands for Immune Epitope Database. IEDB SMM-Align is a tool of IEDB analysis resource which predicts HLA Class II binding epitopes. It follows SMM-QM (Stabilized Matrix Method-Quantitative Matrices) algorithm for prediction which is a modification of original Quantitative matrices (Nielsen et al., 2007). IEDB SMM-Align identifies the HLA class II binding motif in terms of a position specific weight matrix. The output of the SMM-Align method is IC50 binding affinity values, enabling direct readout of the peptide-HLA binding affinity. The lower the IC50 if the peptide the higher will be peptide-HLA binding affinity. Here peptides with an IC50 range less than 50nM are designated to be high affinity peptides while the one having scores less than 5000nM are said to low affinity higher affinity

is indicated through a low output score. Threshold for epitope identification was taken as IC50≤500.

Common epitopes predicted by all the three tools were selected.

3.4 BLAST Screening

In order to avoid any similarity of the peptides with any annotated human protein other than CEA BLASTp analysis was performed for epitopes predicted to bind HLA class I and II respectively (Altschul et al., 1990). The peptides showing similarity in 7 out of 9 amino acids without gap or mismatch were eliminated thus ruling out any possibility of development of autoimmune response against any human functional protein. The epitopes obtained after blastp analysis having overlapping sequences were merged together to generate single peptide fragment containing single or multiple $CD8^+$ and $CD4^+$ T cell epitope.

3.5 Population Coverage Analysis

An epitope can elicit an effective immune response only in individuals which express the HLA molecule that will bind that epitope specifically. HLA molecules are highly polymorphic and a diverse range of HLA alleles exists among the individuals of different races and geographical locations. The study for vaccine target identification aims to select peptides with multiple epitopes having different HLA binding specificities that allow the coverage of wide population. The subject of HLA polymorphism further adds complication as different HLA types are observed to be present in variable frequencies in individuals of different ethnicities and geographical zones. The tool is employed to calculate the percentage of individuals expected to respond to the predicted epitope with reference to HLA genotypic frequencies and HLA binding data (Bui et al. 2005). The Allele Frequency database (http://www.allelefrequencies.net/) provides the genotypic frequencies of the HLA allele. Population coverage was calculated for the individuals of 20 ethnicities and 16 different geographical areas.

3.6 Molecular Docking

Molecular Docking is a structure based technique to analyse the binding of HLA with various peptides. AutodockVina tool was used for docking HLA with the peptides (Trott et al., 2009).

3.6.1 Peptide structure generation

The structure of the epitopes from their amino acid sequence were predicted for CD8+ T cell epitopes and peptide containing multiple $CD4^+$ T cell epitopes employing peptide structure prediction server PEP-FOLD (Shen et al.,2014). The server is capable of running upto 100 simulations for an amino acid sequence and returns the 5 best models predicted. The tool accepts a sequence of maximum 36 amino acid residues. Structure prediction was done for eight CD8+ T cell epitopes and six peptides having $CD4^+$ T cell epitopes.

3.6.2 Native peptide separation from HLA molecule

The Protein Data Bank (PDB) is a depository of the three dimensional crystallographic structure of Proteins (Berman et al., 2000). Ten high resolution crystallographic structures each for HLA class I and HLA class II structures were downloaded from PDB. HLAs are generally available as HLA-peptide complex with their native peptides in the crystallized form. Using Discovery Studio V4.1, native ligand (peptide) was separated from HLA molecules. Also, during the separation process, water molecules and heteroatoms are removed from the crystal using various editing options. Hydrogen atoms are added to the separated native peptide and HLA molecules. The resultant HLA molecule and peptide is saved in PDB format.

3.6.3 AutoDockVina and AutoDock Tools

AutoDockVina have been developed in the Molecular Graphics Lab at The Scripps Research Institute. AutoDockVina treats the docking procedure as assuming global opimization of the scoring function, internal deduction of the grid maps and other implementation ways. It uses a particular type of structure format (PDBQT) to access maximum compatibility with auxiliary software. AutoDock Tools is included within the MGL Tools software package. It is used to obtain input files in the format compatible (PDBQT files) for AutoDockVina. It also helps in viewing the results and analyzes the output file. PDB file of receptors (HLA molecules) and ligands (predicted epitopes/peptides and the native HLA peptides) were opened in Auto Dock tools software and saved in PDBQT. Using the grid options widget in AutoDock tools software search space for docking is defined in the receptor. A configuration file labelled “conf.txt” is created, which contains the information regarding the dimensions

and location of grid (co-ordinates) along with name of receptor and ligand file. Following is an example of grid file.

Receptor = receptor.pdbqt

Ligand = ligand.pdbqt

Out = out.pdbqt

center_x = 72.065 (molecules specific obtained by grid box selection)

center_y = -40.833(molecules specific obtained by grid box selection)

center_z = -10.862(molecules specific obtained by grid box selection)

size_x = 60(molecules specific obtained by grid box selection)

size_y = 72(molecules specific obtained by grid box selection)

size_z = 92(molecules specific obtained by grid box selection)

Exhaustiveness = 16.

With the help of predefined commands, the Auto Dock Vina tool performs the docking. The predicted peptide ligands were docked to the receptors or the HLA class I and II molecules. Simultaneously, the native peptides were also re-docked to its HLA molecule to obtain its binding energy which was considered as a positive control for the peptides under investigation. The docking results were analysed via the Auto Dock Tools where the binding of the ligand to the user defined receptor groove can be visualized. The peptides that bind the molecule outside the groove were designated as non-binders.

3. 7 Assesment of immunogenic potential of the synthesized peptide

As *in silico* prediction of epitopes is just a screening step towards the identification of immunogenic epitopes. Epitopes are required to be tested experimentally to validate their immunogenicity (Wullner et al., 2010). CEA peptides was analysed for its ability to stimulate PBMCs proliferation and IFN-γ production to validate its immune response.

3.7.1 Peptide synthesis

One of the *in silico* identified peptides of CEA, (TYACFVSNLATGRNNSIVKS) each containing multiple $CD8^+$ and $CD4^+$ T cell epitopes was procured from GL Biochem (Shanghai) Ltd (China). Peptide was dissolved in sterile MilliQ water at a concentration of 1 mg/mL and stored at -20°C for future use.

3.7.2 Chemicals and Reagents

All the chemicals and reagents used throughout this work were purchased from the different companies as given in Table 2**.**All chemicals and reagents purchased were of cell culture grade (Table).

	Reagents/ Chemical/ Media	Company
1	ABTS (2,2'-Azinobis[3-ethylbenzothiazoline-6-sulfonicacid]-diammonium salt) substrate	ThermoFisher Scientific
2	Amphotericin B	Sigma Aldrich
3	Bovine serum albumin	Sigma Aldrich
4	Concanavalin A	Sigma Aldrich
5	Dimethyl Sulphoxide	Merck, Germany
6	Fetal bovin Serum	Gibco®Life Technologies
7	Glutamine	Himedia, India
8	HEPES buffer	Sigma Aldrich
9	Histopaque® -1077	Sigma Aldrich
10	Human IFN-γ Mini ABTS ELISA Development Kit	PeproTech, USA
11	MTT(3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide)	Sigma Aldrich
12	Penicillin Sodium	Himedia, India
13	Potassium Chloride (KCl)	Himedia, India
14	Potassium phosphate monobasic(KH_2PO_4)	Himedia, India
15	Recombinant human Interleukin-2	Sigma Aldrich
16	Rosewell Park Memorial Institute (RPMI)-1640 medium	Sigma Aldrich
17	Sodium Bicarbonate ($NaHCO_3$)	Himedia, India
18	Sodium Chloride (NaCl)	Himedia, India
19	Sodium phosphate dibasic (Na_2HPO_4)	Himedia, India
20	Streptomycin	Sigma Aldrich
21	Trehalose	Sigma Aldrich
22	Trypan blue	Himedia, India
23	Tween 20	Sigma Aldrich

3.7.3 Preparation of Buffers

a) **Phosphate buffer saline (1X):** Dissolved 8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4 and 0.24g KH_2PO_4 in 1 litre double distilled water and pH was adjusted to 7.4. Buffered was autoclaved.
b) **Blocking buffer for ELISA**: 1% BSA dissolved in 1X PBS
c) **Washing buffer for ELISA:** 0.05% of Tween20 in 1X PBS
d) **Sample Diluent**: 0.1% BSA and 0.05% Tween in 1X PBS

3.7.4 PBMC isolation and harvesting

Blood was drawn from healthy donor by trained technicians of Lifeline Blood bank Patiala and Nitin Hospital, Patiala. Peripheral Blood Mononuclear Cells (PBMCs) were isolated from the blood by density gradient centrifugation. In a sterile 15-mL conical centrifuge tube, 5 mL of whole blood was layered over 5 mL of Histopaque® -1077 and centrifuged at 400xg for 30 minutes at room temperature in swinging bucket rotor. This density based centrifugation technique fractionates blood into plasma, red blood cells (RBC) and peripheral blood mononuclear cells (PBMC). After centrifugation, upper plasma layer was carefully discarded and opaque interface (buffy coat) containing PBMCs was transferred in a sterile 15-mL conical centrifuge tube. Resuspended the cells in 10 mL of PBS (1X) and centrifuged at 250x g for 10 minutes. This washing step was done twice. Discarded the supernatant and the cell pellet was resuspended in 1 mL of complete RPMI-1640 (supplemented with 10% foetal bovine serum, 100 µg/mL streptomycin, 100 I.U./mL penicillin and 10 mM HEPES). Cells were counted and assessed for viability by Trypan Blue exclusion assay on Hemocytometer.

3.7.5 PBMC stimulation assay

In order to test the immunogenicity of peptide, isolated PBMCs were stimulated with the peptide. In a flat bottom 24 well cell culture plate, 2×10^6 cells were stimulated with the peptide (25µg/ml) in the presence of recombinant IL-2 (20 ng/mL) and complete RPMI-1640 media in total volume of 2 mL. Unstimulated cells served as control. The cultured plate was incubated at 5% CO_2 at 37°C in a humidified incubator. The cells were restimulated with peptide on 4^{th} and 7^{th} day by replacing 1 mL of complete media containing peptide and IL-2 in the same

concentration. The stimulated as well unstimulated cells were harvested on 10th day and recounted.

3.7.6 IFN γ detection by ELISA

Cells recovered from unstimulated wells were distributed in two set of triplicates; one set stimulated with 10 µg/ml of concanavalin A (Sigma Aldrich) served as positive control, whereas the other set having the un-stimulated cells served as negative control. Peptide stimulated cells distributed in triplicates were given final stimulus with 25 µg/mL peptide in the presence of IL-2 (20 ng/ml). The plate was incubated at 5% CO_2 at 37°C in a humidified incubator. Culture supernatant was collected after 48 h incubation (day 12). IFN-γ secretion in the culture supernatant was measured by sandwich ELISA (PeproTech).

Capture antibody was diluted with PBS to a concentration of 1µg/ml. 100µl of capture antibody was added to wells in 96 well plate ELISA plate.The plate was incubated at 4°C for overnight. The wells were aspirated to remove the liquid and washed 4times with 300µl of wash buffer per well. After each washing the plate is tapped well on tissue to ensure the removal of any residual liquid. 300µl block buffer was added to each well and incubated at room temperature for 1 hour. Discarded the blocking blocking buffer and repeated the washing steps. Human IFN-γ Standard from ELISA kit was diluted to two concentrations (3000 pg and 30 pg) in sample diluent. 100µl of standard and test sample (supernatant of peptide stimulated cells collected from the 96 well culture plate) were added to each ELISA plate in triplicate and incubated at room temperature for overnight. The plate wells were aspirated and washed 4 times. Detection antibody from was diluted in sample diluents to a concentration of 1µg/ml. 100µl of this was added to each well and incubated at room temperature for 2 and half hours. The plate was aspirated and washed 4 times. 5.5µl of avidin-HRP conjugate was diluted with diluent to a volume 11ml in the ratio 1:2000. Avidin-HRP Conjugate was added 100 µl per well and incubated for 30 minutes at room temperature. The plate was aspirated and washed 4times. 100µl of ABTS liquid substrate was added to each well and incubated at room temperature for about 15-20 minutes for colour development. The colour development is monitored using an ELISA plate reader at 405/630nm.

3.7.7 PBMCs proliferation assay

Cell were recovered and reseeded in the same way as discussed in the last section. The plate was incubated at 5% CO_2 at 37°C in a humidified incubator. On 16^{th} day, MTT assay was performed to estimate the proliferation of peptide stimulated cells as compared to the controls. 20µl of MTT (5mg/ml) was added to each well and incubated for 4hours. The media in the wells was aspirated carefully without disturbing the crystals formed at the bottom. 100µl of DMSO was added to each well to dissolve the crystal. Absorbance was measured at 575/630nm in multi-well plate reader.

Chapter 4: RESULTS

4.1 Predicted $CD8^+$ T cell epitopes

Nonameric peptides were predicted from SYFPEITHI, NetCTL and BIMAS prediction tools individually. The peptides that were predicted commonly by the three of the tools were finally accepted as potential $CD8^+$ T cell epitopes. Our motive is to identify the promising peptides and when three tools with different algorithms chooses a particular peptide as a epitope then its chance of being potential epitope is enhanced. Total 65 common peptides were selected by all the tools. Blast screening for all these 65 epitopes were performed to exclude the 30 peptides which showed sequence similarity of minimum 7 amino acid residues with other functional proteins present within our body. In the present study, some of the epitopes showed sequence similarity with biliary glycoprotein, Pregnancy specific glycoprotein, non specific cross-reactive antigen-3 protein, fibrous sheath-interacting protein 2 isoform X2 and phosphoprotein p65 proteins. Most of these proteins though are encoded by the CEA family genes but are different from the CEA protein. Of the 35 epitopes only seven epitopes were found to make overlapping peptide regions (Table 2 and 3).

Table 2 : CD8+T cell epitopes after blast screening

CD8+T CELL EPITOPES		
ARRSDSVIL	LYGPDTPII	TYYRPGVNL
DPVTLDVLY	NLPQHLFGY	VEDEDAVAL
ETQDATYLW	NRQIIGYVI	VLYGPDTPI
GQFRVYPEL	PEAQNTTYL	VYAEPPKPF
GRNNSIVKS	PEIQNTTYL	YECGIQNEL
HRWCIPWQR	QYSWRINGI	YLSGANLNL
HSDPVILNV	SARRSDSVI	YRPGVNLSL
HTQVLFIAK	SASGTSPGL	YSWFVNGTF
IMIGVLVGV	SDSVILNVL	YVCGIQNSV
IPQQHTQVL	SNDNRTLTL	LYGPDDPTI
KLTIESTPF	SSYLSGANL	TYACFVSNL
LYGPDAPT	TRNDTASYK	

Table 3: Overlapping peptide sequences containing CD8+ Tcell epitopes

Overlapping sequence	Position	Epitopes
QYSWRINGIPQQHTQVLFIAK	624-644	QYSWRINGI,IPQQHTQVL,HTQVLFIAK
SSYLSGANLNL	603-613	SSYLSGANL,YLSGANLNL
TYYRPGVNLSL	425-435	TYYRPGVNL,YRPGVNLSL

4.2 Predicted $CD4^+$ T cell epitopes

Identification of $CD4^+$ T cell epitopes epitopes is equally important as $CD8^+$ T cell epitopes as they combinedly generate a better immunogenic response against the antigens and henceforth gives more promising candidates to formulate vaccines. Seventy epitopes were commonly predicted by PROPRED, NetMHC2 and IEDB tools. After blast screening, 16 epitopes were found to have sequences matching with other proteins and were eliminated. Of these epitopes, 15 overlapping fragments were generated (Table 4 and 5).

Table 4: $CD4^+$ T cell epitopes after blast screening

CD4+ specific T cell Epitopes			
IGVLVGVAL	YRPGVNLSL	FNVTRNDTA	IQNSVSANR
MIGVLVGVA	YYRPGVNLS	WVNNQSLPV	VCGIQNSVS
IMIGVLVGV	YTYYRPGVN	YLWWVNNQS	YVCGIQNSV
VGIMIGVLV	IQNELSVDH	FRVYPELPK	FNVTRNDAR
FVSNLATGR	LLSVTRNDV	IKSDLVNEE	YLWWVNGQS
YACFVSNLA	LSNDNRTLT	FYTLHVIKS	LYTCQANNS
FIAKITPNN	WVNNQSLPV	IQNDTGFYT	FVNGTFQQS
LFIAKITPN	YLWWVNNQS	LLIQNIIQN	WFVNGTFQQ
VLFIAKITP	IQNTTYLWW	IYPNASLLI	LNLSCHAAS
YLSGANLNL	FITSNNSNP	YSGREIIYP	VTRNDTASY
YGPDTPIIS	VTTITVYAE	YVIGTQQAT	WLIDGNIQQ
VTLDVLYGP	LNRTTVTTI	IIGYVIGTQ	LLTFWNPPT
FISNITEKN	IPNITVNNS	FNVAEGKEV	WQRLLLTAS
LFISNITEK	FIPNITVNN	LTIESTPFN	

Table 5: Overlapping peptide sequences containingCD4+ T cell epitopes

Overlapping sequence	**Position**	**Epitopes**
LTIESTPFNVAEGKEV	36-51	LTIESTPFN,FNVAEGKEV
IIGYVIGTQQAT	79-90	IIGYVIGTQ,YVIGTQQAT
IQNDTGFYTLHVIKSDLVNEE	113-133	IQNDTGFYT,FYTLHVIKS,IKSDLVNEE
YLWWVNNQSLPV	176-187	YLWWVNNQS,WVNNQSLPV
FNVTRNDTASY	203-213	FNVTRNDTA,VTRNDTASY
WFVNGTFQQS	271-280	WFVNGTFQQ,FVNGTFQQS
FIPNITVNNS	285-294	FIPNITVNN,IPNITVNNS
LNRTTVTTITVYAE	308-321	LNRTTVTTI,VTTITVYAE
IQNTTYLWWVNNQSLPV	349-365	IQNTTYLWW,YLWWVNNQS,WVNNQSLPV
LSNDNRTLTLLSVTRNDV	371-388	LSNDNRTLT,LLSVTRNDV
LFISNITEKN	461-473	LFISNITEK,FISNITEKN
VLFIAKITPNN	639-649	VLFIAKITP,LFIAKITPN,FIAKITPNN
YACFVSNLATGR	653-664	YACFVSNLA,FVSNLATGR
VGIMIGVLVGVAL	689-701	VGIMIGVLV,IMIGVLVGV,MIGVLVGVA, IGVLVGVAL

4.3 Determination of common $CD4^+$ and $CD8^+$T specific epitope containing peptide fragments

The sole purpose of this study is to identify multi-epitope peptide regions of the carcinoembryonic antigen protein that will elicit an effective immune response against cancerous cells and kill them. This peptide would be a more promising one when it can bring both $CD8^+$ and $CD4^+$ T cells into action. Thus, we tried to get peptide fragments that contained epitopes both $CD8^+$ and $CD4^+$ T cells. Five fragments were obtained containing overlapping CD4+ and CD8+ epitopes ranging from 11 to 20 amino acid residues (Table 6). These five peptides were enriched with eight each CD4+ and CD8+ epitopes. The peptide fragments were found to cover larger number of HLA class I and II alleles (Table 7).

Table 6: Peptides enriched with $CD8^+$ and $CD4^+$ T cell specific epitopes

Peptides containing $CD8^+$ and $CD4^+$ specific T-cell epitopes	**Position**	**$CD8^+$ specific T cell Epitopes**	**$CD4^+$ specific T-cell Epitopes**
VTLDVLYGPDTPII	585-598	LYGPDTPII VLYGPDTPI	VTLDVLYGP
SSYLSGANLNL	603-613	SSYLSGANL YLSGANLNL	YLSGANLNL
IPQQHTQVLFIAKITPNN	632-649	IPQQHTQVL	FIAKITPNN
TYACFVSNLATGRNNSIVKS	652-671	TYACFVSNL GRNNSIVKS	FVSNLATGR YACFVSNLA
VGIMIGVLVGVAL	689-701	IMIGVLVGV	IGVLVGVAL IMIGVLVGV VGIMIGVLV

Table 7: Number of HLA class I and II which are predicted to bind epitopes present in the selected peptides

Peptide Fragment	**Number of HLA Class I specific Alleles**	**Number of HLA Class II specific Alleles**
VTLDVLYGPDTPII	6	4
SSYLSGANLNL	10	11
IPQQHTQVLFIAKITPNN	6	8
TYACFVSNLATGRNNSIVKS	5	23
VGIMIGVLVGVAL	2	21

4.4<u>Population coverage analysis</u>

One objective of our recent study is also to review the expected response in the worldwide population when our predicted peptides are stimulated. HLA polymorphisms exist among population of different geographical regions and ethnicities and thus can affect the

immunogenic response in various ways. The IEDB population coverage tool helps determine the HLA frequencies in different populations thereby reveals the potential of the predicted peptides as whether they are capable to generate response in a particular population. Population coverage analysis was performed for 16 different geographical regions with 76.31%and 92.92% as the resulting average population coverage for class I and class II HLA alleles respectively (Fig 2a). The best respondent for class I were the European population with 92.63% coverage and for class II it was the North American population with 100% coverage (Fig 2a).. The least or negligible response was observed for the central American population for class I (0.80%) and south African population for class II(32.10%) alleles(Fig 2a).. Also the population coverage was carried out for 20 different ethnicities around the world which resulted in an average coverage of 76.43% and 93.66% for HLA class I and class II respectively (Fig 2b).. For class I the best and least response was shown by Amerindian (97.54%) and the Black (72.50%) respectively. While for class II the Caucasoid showed the best response with 99.96% and the Australian Aborigines responded the least with 66.34% coverage (Fig 2b)..

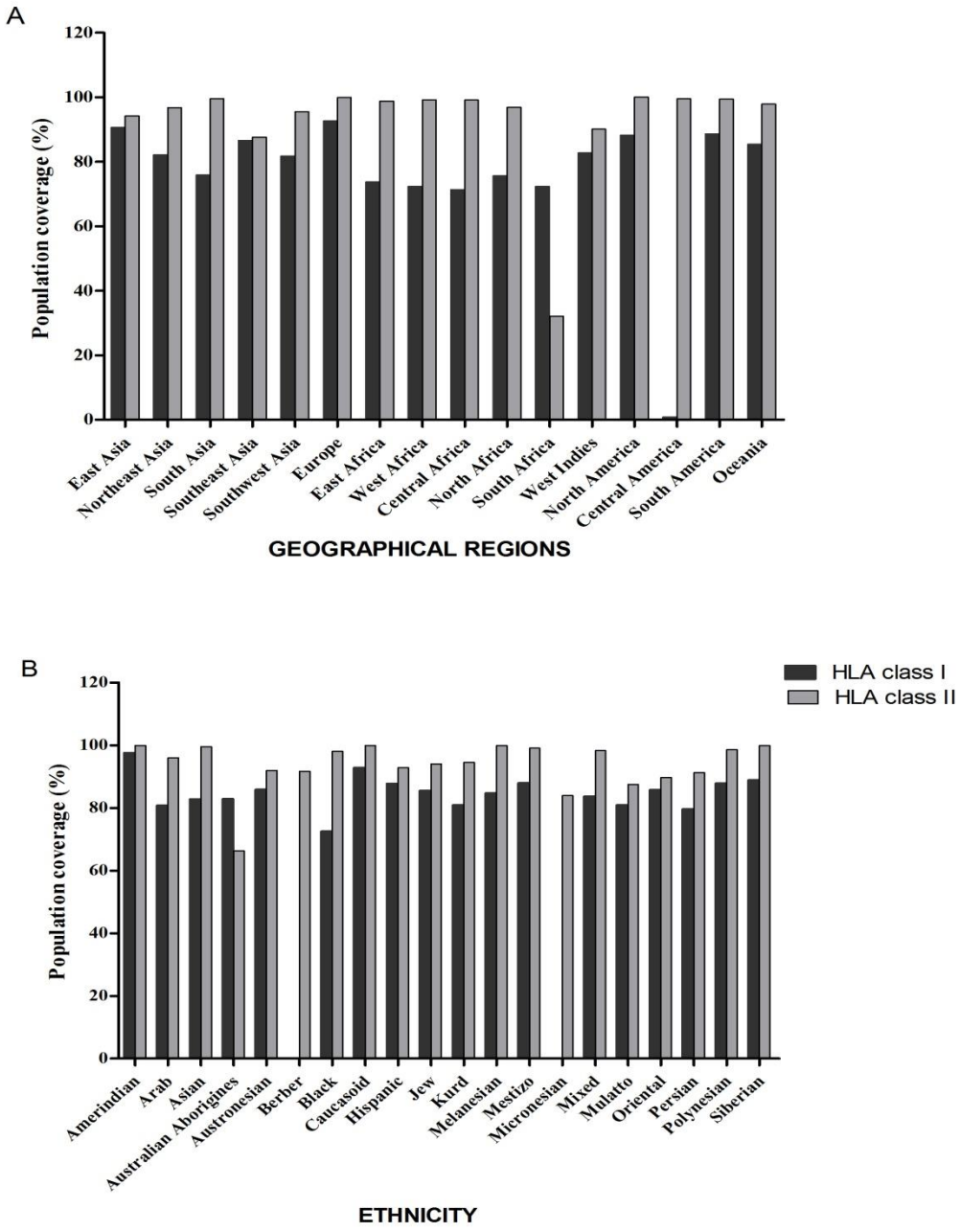
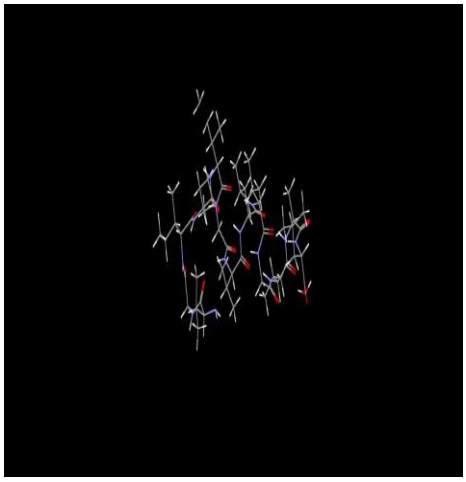


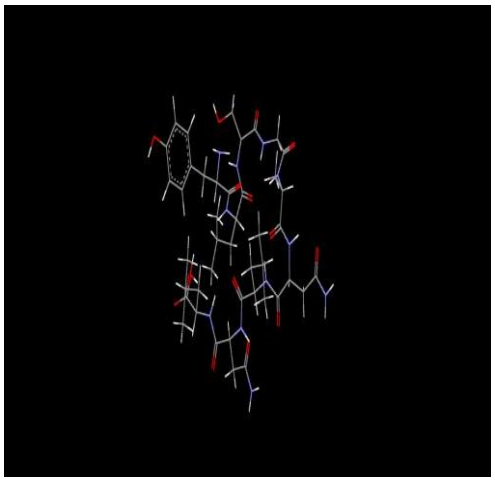
FIG 2 - Population coverage analysis of the predicted peptides specific for HLA class I and II showing the desired response in A)16 different geographical regions and b) 21 different ethinicities around the world as predefined in the IEDB database.

4.5 PEPFOLD generated peptide structure

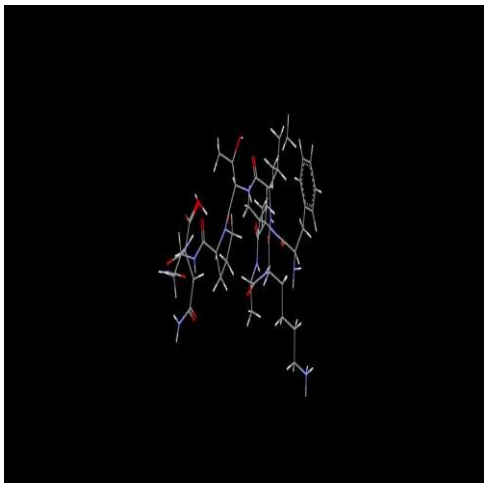
PEPFOLD server helped to predict the peptide structures from their amino acid sequences. It is based upon structure alphabet letters to describe the structure of the given sequence. It predicts the structure of four consecutive amino acids and then combines the series of structural alphabet letter to generate the structure (Thevenet *et al*., 2012).



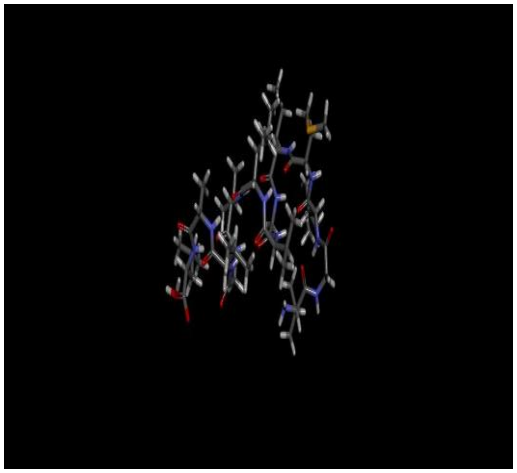
(1)YACFVSNLATGR (HLA I)



(2) YLSGANLNL (HLA I)



(3)FIAKITPNN (HLA II)



(4)VGIMIGVLVGVAL (HLA II)

FIG 3 : PEPFOLD predicted structures of the peptides

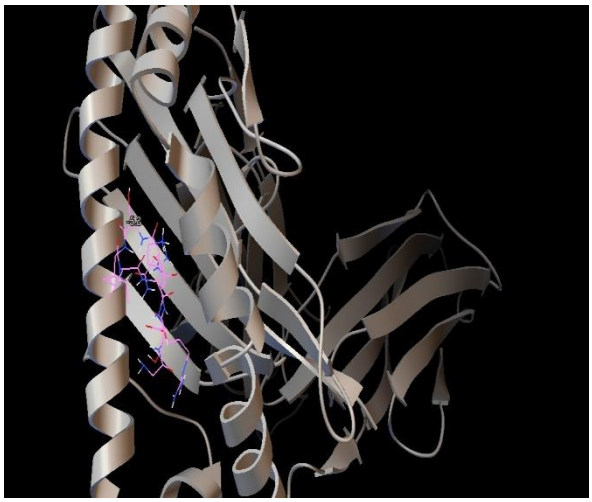
4.6 Docking of $CD8^+$ T cell epitopes

AutoDock Vina gives the binding efficiency of each epitope with HLA molecules in the form of a score. Binding score would tell us whether a given epitope binds efficiently with the

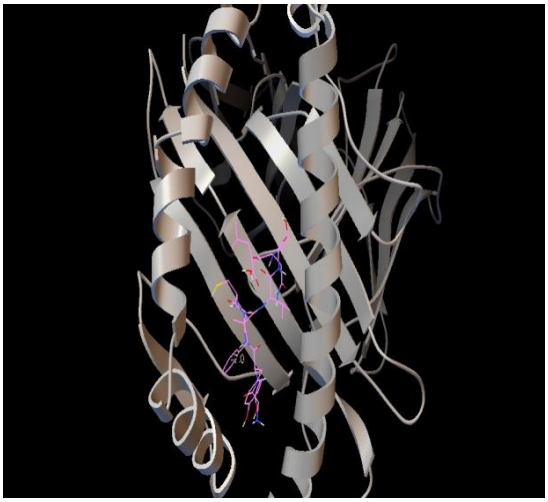
respective HLA groove or not. Often it is also seen that some peptides show good binding energy but bind at positions outside the groove. Thus after the run is complete the docking results should be analyzed in the AutoDock Tools to ensure binding of the peptide within the user defined grid. The epitopes that are found to bind outside the receptor groove have been defined as Non binders(NB) .We have performed the docking for 8 $CD8^+$ T cell epitopes with 10 HLA alleles that have been reported to be available widely in world population. YLSGANLNL and IPQQHTQVL were found to be binding outside the groove for three HLA alleles and two HLA alleles respectively (Table 8).One way ANOVA and Tukey's multiple comparison test was performed for the comparison of binding energies of these peptides with the native peptide considered to be standard. Except two peptides (GRNNSIVKS and IMIGVLVGV),the binding energies for CD^+ T cell peptides were found not to be significantly different from the respective native peptide which shows the good binding affinity with HLA molecules.

4.7 Docking of $CD4^+$ T cell epitopes

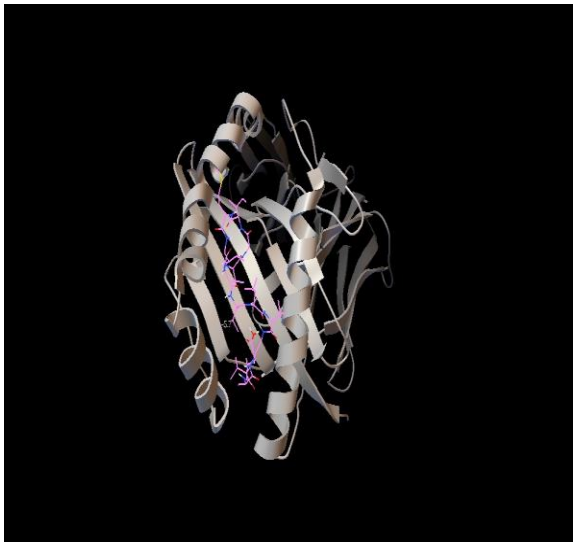
Similarly CD4+$^+$ T cell epitopes were docked to the widely known human HLA molecules. The only difference here was that large fragments were considered as class II HLA binds larger peptides. From the $CD4^+$ T cell epitopes which are part of selected peptides, we had obtained three overlapping peptide fragments. We carried out the docking for these three peptides and the rest three nonamers with 10 diversely identified HLA class II allele. VGIMIGVLVGVAL is the one peptide that bounded out of the defined groove for two of the HLA molecules (Table 9). One way ANOVA and Tukey's multiple comparison test was implemented to compare the binding affinities of the predicted $CD4^+$ T cell specific peptides with the corresponding native peptides that were already bound to the HLA molecule. Almost all the class II predicted molecules were calculated to have non-significant difference with the native peptide except one (YTYYRPGVNLSL) (Fig5).



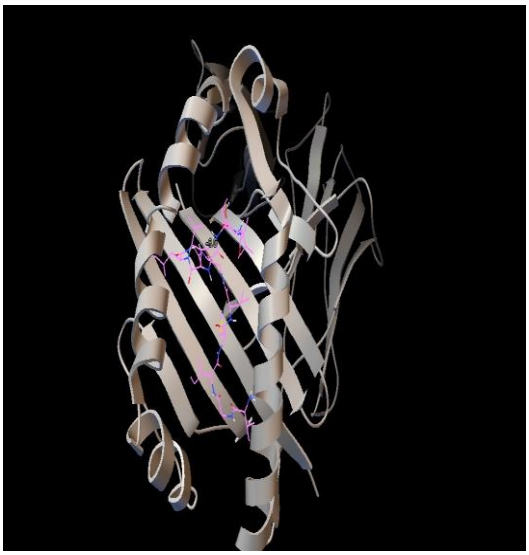
(1)GRNNSIVKS bound to HLA A24



(2) TYACFVSNL bound to HLA B57



(3)YACFVSNLATGR bound to HLA DQ1



(4) YACFVSNLATGR bound to HLADR3

FIG 4: Docking of the constructed peptide to HLA Class I and II

Table 8: Binding Energies of T cell epitopes with Class I HLA alleles

HLA alleles	PDB ID	Native peptide	GRNNSIVKS	IMIGVLVGV	IPQQHTQVL	LYGPDTPII	SSYLSGANL	TYACFVSNL	VLYGPDTPI	YLSGANLNL
HLA B8	3SPV	-9.1	-7.5	-7.9	-8.5	-8.7	-8.3	(NB)	-9	(NB)
HLA B62	1XR9	-9.7	-6.6	-6.9	-8.1	-9	-8	-9	-8.3	-8.6
HLA B5702	2BVP	-8.7	-7.8	-6.8	-6.4	-7.9	-7.5	-7.8	(NB)	(NB)
HLA B7	3VCL	-8.7	-7.9	-7.2	-6.1	-8.2	-8.2	-8	-8.2	-7.8
HLA B27	1K5N	-7.5	-6.7	-6.1	(NB)	-8	(NB)	-7	-7.9	(NB)
HLA B44	1N2R	-6.7	-5	-5.1	-6.1	-6.3	-7	-7	-6.4	-7.4
HLA A1	3BO8	-7.6	-5.8	-5.7	(N.B)	-7.1	-6.9	-6.6	-6.6	-7.1
HLA A2	3MRE	-9.3	-7.7	-7.9	-7.2	-8.3	-6.7	-7.9	-7.8	-7.5
HLA A24	3WL9	-8.2	-6.6	-6.8	-6.6	-6.2	-7.3	-8.1	-7.9	-6.7
HLA A3	3RL1	-6.6	(N.B)	-6.6	(N.B)	(N.B)	(N.B)	(N.B)	(N.B)	(N.B)

*N.B-non binder

Table 9 : Binding Energies of T cell epitopes with Class II HLA alleles

HLA alleles	Native peptide	VGIMIGVLVGVAL	YACFVSNLATGR	YTYYRPGVNLSL	VTLDVLYGP	YLSGANLNL	FIAKITPNN
HLA DR3	-6	-5.2	-6	-7	-6.8	-7.2	-6.6
HLA DR4	-7	-6.5	-6.2	-7	-8.1	-8.3	-8.4
HLA DR1	-5.7	-6.8	-6.3	-8.2	-6.8	-7.1	-6.9
HLADQ1	-6.9	-6.2	-5.7	-8	-7.5	-7.5	-8.1
HLA DQ2	-8.7	(N.B)	-6.8	-8.7	-8.2	-8.2	-7.5
HLADP2	-7.1	-6.7	-6	-8.3	-6.9	-7.5	-7.6
HLA Q0602	-6.6	-6.2	-6.1	-8.1	-7.6	-7.2	-7.2
HLA DR52C	-6.4	-5.7	-6.6	-6.9	-6.4	-8.6	-6.9
HLA DR2	-7.6	-6.3	-5.9	-8	-8.3	-7.5	-7.3
HLA DQ8	-7.1	(N.B)	(N.B)	-8	-8.9	-6	-7.8

*N.B-non binder

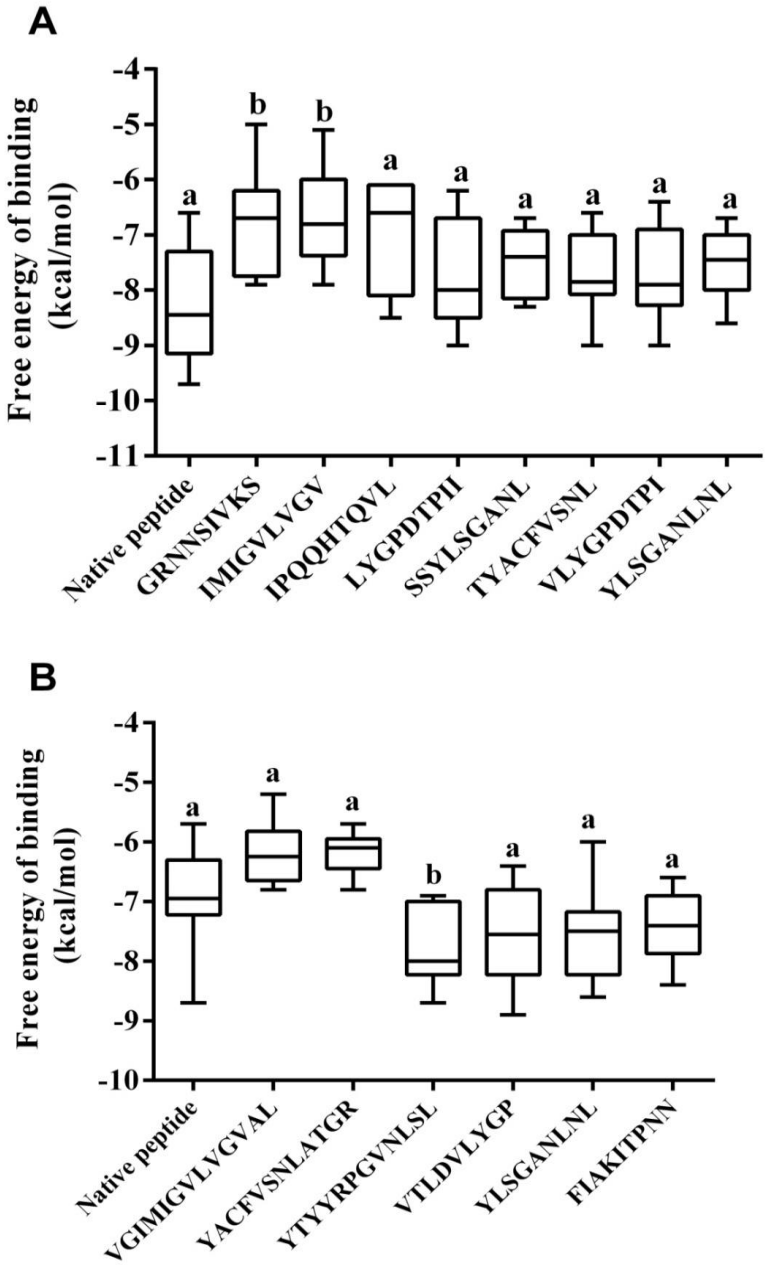


FIG 5 - Binding energies of the predicted epitopes/peptides and native peptides with A) HLA class I and B) HLA class II molecules obtained after docking .

4.8 Peptide specific interferon-γ

ELISA was performed to determine the cytokine production by peptide stimulated cells. After 2 days of stimulation of the cells by peptide (TYACFVSNLATGRNNSIVKS), cytokine production by these cells were analyzed. The positive control which were cells stimulated with mitogen Concanavalin A showed a sharp rise in the graph depicting the ability of these cells to generate a maximum response in cytokine production. While the peptide stimulated cells showed response greater than the unstimulated cells. This clearly depicts that the peptide has the ability to activate the immune cells and thereby cytokines are produced.

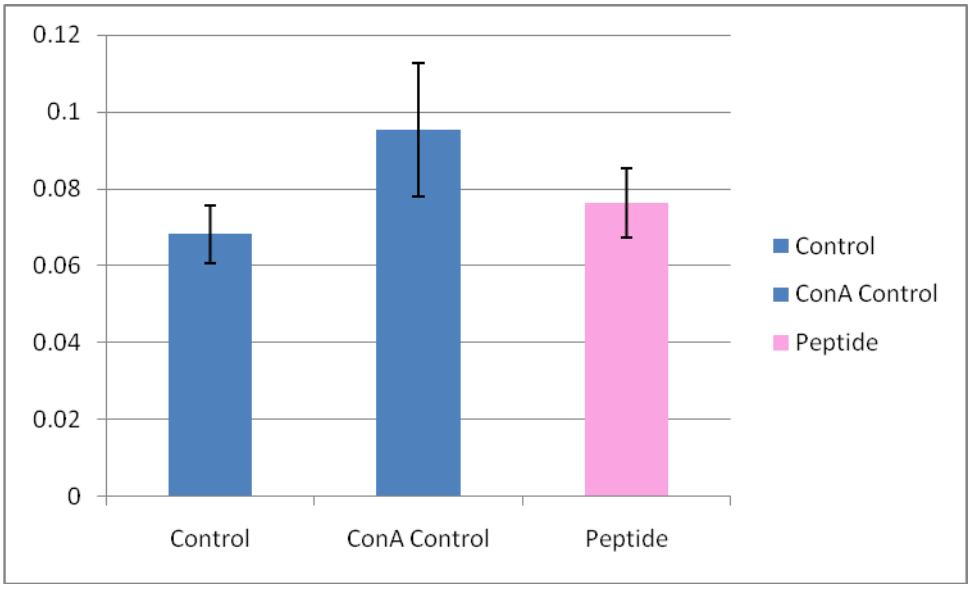


Fig 6 - Interferon-γ production in peptide stimulated peripheral blood mononuclear cells

4.9 Peptide specific T cell proliferation

The lymphocyte cell proliferation is assessed through MTT assay. The synthesized peptide (TYACFVSNLATGRNNSIVKS) was observed to stimulate PBMCs for proliferation.

In a 6 day culture the peptide stimulated PBMCs showed more growth than the control cells which were unstimulated (cells in media and IL2). The positive control containing the mitogen, ConcanavalinA (Con A) showed a sharp rise in the cell concentration. Immunogenic potential of the peptide is validated through its capability to stimulate the growth of PBMCs which is distinctly noticed in the figure2.

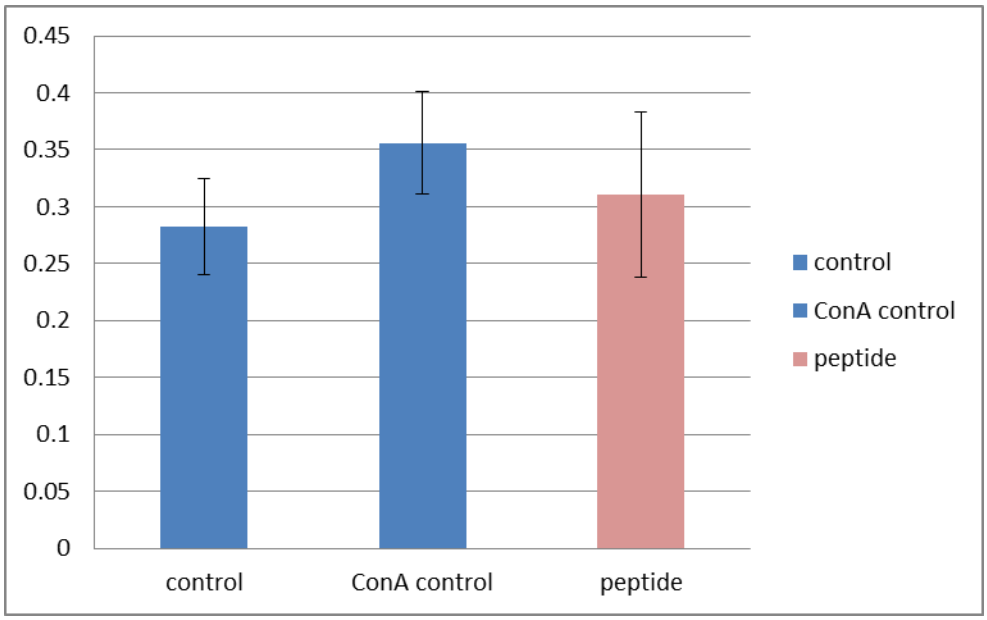


Fig 7- Proliferation assay of peripheral blood mononuclear cells (PBMCs) against peptide (TYACFVSNLATGRNNSIVKS) using MTT assay

Chapter 5: DISCUSSIONS

The use of bioinformatic tools for the purpose of various immunological studies brought a whole new revolution in the field of vaccine design. Computational approaches based on various algorithms made the long tiring experimental work not only painless but also saved a huge amount of resources and money. Vaccine design against cancer has gained some progress in the last few decades with declaration of some facts like the presence of tumor marker antigens and their capability of being immunogenic. Exploiting these facts, intensive work has been carried out for the identification of tumor markers or tumor associated antigens (TAAs) as they are addressed. Many possible vaccination approaches have been discovered using the whole protein or peptide of these TAAs, peptide loaded dendritic cells, vaccinia virus vector conjugated with TAAs. Studies have been reported showing progress of these different vaccination approaches as many of them are being evaluated at various clinical phase trials. At the same time *in silico* studies has also shown the significant results *in vitro* as well as *in vivo*, identification of HER/neu specific epitope predicted by SYPEITHI has shown immunogenic response *in vitro* and *in vivo* (Gritzapis et al., 2010). *In-silico* studies and its validation through *in-vitro* test of the carcinoembryonic antigen epitope at position 355-367 also considered to be a naturally processed epitope have been noted(Campi et al.,2003).There are other reports which promises the immunogenic response of *in silico* predicted peptide already confirmed in *in vitro* and *in vivo* tests. But majority of these studies have used one (Gritzapis *et al*., 2010) or two (Yuan *et al*., 2012) tool for the epitope prediction. This generates a sort of loop hole in the approach as many epitopes predicted by a single computational algorithm led to the generation of a number of false positives and false negatives (Lohia et al., 2015). Thus researchers have conceptualized a consensus approach (comparing and combining results of epitope prediction from different algorithms) which have been ratified *in-vitro* and *in-vivo* (Seyed et al., 2011). The present study includes the application of six different algorithms to predict the CEA epitopes making the process more impactful. Choosing the epitopes commonly predicted by all the tools confirms the potential of the epitope and then creating overlapping fragments from these epitopes help it to target a wide range of HLA alleles covering maximum possible population of the world. Simultaneously, the epitopes have been chosen to activate both $CD4^+$ and $CD8^+$T cell driven responses thus aiding the vaccine candidate to develop an overall immune response.CEA is reported to be tumor associated antigen as it is found to be increased in different cancers.

The unique fact about the protein being present in negligible amount in normal individuals makes it a more convincible target for vaccine design as the chances of developing any autoimmunity is somewhat ruled out. Further to ensure this development of any chance of tolerance by the epitope, BLAST screening has been performed to eliminate such epitopes. After the BLAST screening, five final peptide fragments were obtained which were enriched with both $CD4^+$ and $CD8^+$ T cells epitopes.

A population coverage study of the HLA alleles considered in our study has been executed to assure that the peptides can generate a strong immune response against the tumor cells. These tumor surface peptides are programmed to bind the HLA molecules and present them to immune T cells for development of immunogenic responses thus making the presence of the HLA molecules evident in the system of maximum possible world population to make the study a success. The population study revealed that these peptide fragments covered an average >84%population among various geographical regions and ethnicities. These results affirmed that the peptides along with covering multiple T cell epitopes are also reliable for spawning immunogenic response in individuals from different populations.

Molecular docking has extensively benefited the field of drug designing and has recently gained huge impact in the field of vaccine design. Through this tool, the structural binding of the peptide to the HLA molecule can be investigated. With eight nonameric epitopes $CD8^+$ T cells and six $CD4^+$T cell peptides, the docking study was carried out. The binding affinities were interpreted for all the peptides specific for CD8+ and CD4+ cells and comparison was done with the standard which is the native peptide. Statistical comparison was also accomplished between the native assumed to be the standard and the predicted peptides. Statistical deductions showed almost all the peptides to have nonsignificant difference with the standard peptide for both class I and II molecules. Only two of the eight identified $CD8^+$ T specific peptides and one from the $CD4^+$ T cell peptides showed some significant difference with the binding energy of the native peptide. This kind of result may be clarified by remembering the fact that the peptides were not predicted for all HLA interactions. Hence, these results bring forth the fact that peptides will be strongly presented by HLA molecules present on antigen presenting cells to induce T cell response. Similar kind of studies has been reported for nucleoprotein of H5N1 virus where they compare the binding energy with predicted epitope and bound peptides for only class I HLA molecules (Hou *et al* 2012). Peptide HLA docking studies has been also reported for other tumor antigen like melanoma antigen E (Akiyama *et al* 2012).

In-silico tools can only accelerate the process of vaccine design by reducing the cost, time and resources to recognize the actual deserving targets on which the *in-vitro* and *in-vivo* experimentation should be invested. Prediction tools only support the study by making it somewhat precise through narrowing the epitope selection process from few thousands to within fewer peptides. Henceforth, prediction of peptides alone cannot determine the immunogenicity of the peptides. Only after assessing a peptide *in-vitro* in simulated environment and then in animal models can it be announced as a vaccine target that can be tested for clinical trials. Thus one of the final five peptides (TYACFVSNLATGRNNSIVKS) comprising of the most number of epitopes amongst the others was synthesized for *in-vitro* analysis. Lymphocyte proliferation assay was carried out with mononuclear cells of healthy individuals which were stimulated with the synthesized peptide. The readings obtained clearly displayed that the proliferation of the peptide stimulated cells were higher than the unstimulated cells. But it is difficult to interpret its immunogenic potential as present study is limited to one sample. But single data result has given direction to carry out more experiments to confirm its immunogenicity.

With the evidence of providing epitopes which are unique to the immune system and deprived of any self tolerance risk, the CEA epitopes also have the capability to induce a strong overall immune response. Thus from the current study we can conclude that peptides derived from the carcinoembryonic antigen protein can act as a potential target for vaccine development.

Chapter 6: SUMMARY

Vaccination is an efficient approach that has gained ample importance in cancer treatment. Various vaccination approaches have been reported which displayed efficient results in the fight against cancer. Carcinoembryonic antigen (CEA) is a well studied tumor associated antigen whose presence in increased concentration has been recorded in various types of cancer. In the current study an effort was made to identify the potential epitopes of CEA which could generate an effective immune response when introduced in the form of vaccine within the body. An immunoinformatic approach was selected to determine the epitopes from the whole protein sequence which was downloaded through NCBI protein database. The study includes the use of six different *in-silico* tools to obtain $CD4^+$ and $CD8^+$T cell epitopes. Blast screening of these epitopes helped us to eliminate those epitopes which had sequence similiarity with the functional proteins of normal individual. Finally eightCD8+ and eightCD4+ nonameric epitopes were obtained. These epitopes were merged to obtain five overlapping peptide fragments containing both $CD4^+$ and $CD8^+$T cell epitopes. Structure of eight $CD8^+$T cell epitopes and six $CD4^+$T cell peptides were generated using PEPFOLD software. Docking of these peptides to the widely known HLA molecules was accomplished to ensure the binding potential of the predicted peptides with HLA molecules. Docking of these eight $CD8^+$T cell epitopes and six $CD4^+$T cell peptides were performed with 10 HLA class I and 10 HLA class II molecules respectively. Almost all the binding energies of both $CD4^+$ and $CD8^+$T cell epitopes/peptides were found non significant in comparison to the native peptide of the respective HLA molecule. In-vitro cytokine production and lymphocyte proliferation analysis through MTT assay was carried out for the synthesized peptide (TYACFVSNLATGRNNSIVKS). Although, the experiment could be carried out with one sample, hence it is difficult to confirm its immunogenic potential. But prelimarily data indicates positive immune response of peptide specific PBMCs. Present study selected five peptides containing multiple epitopes which may have the potential to elicit immune response and thus can be good cancer vaccine candidate.

REFERENCES

(1) Cancer and the Immune system: The vital connection,2003, publication by the Cancer Research Institute

(2) Guo, J. Y., Xia, B., & White, E. (2013). Autophagy-mediated tumor promotion. *Cell*, *155*(6), 1216-1219.

(3) Patronov, A., Dimitrov, I., Flower, D. R., & Doytchinova, I. (2012). Peptide binding to HLA-DP proteins at pH 5.0 and pH 7.0: a quantitative molecular docking study. *BMC structural biology*, *12*(1), 1.

(4) Vigneron, N., Stroobant, V., Van den Eynde, B. J., & van der Bruggen, P. (2013). Database of T cell-defined human tumor antigens: the 2013 update.*Cancer Immunity Archive*, *13*(3), 15.

(5) Kobayashi, H., Omiya, R., Ruiz, M., Huarte, E., Sarobe, P., Lasarte, J. J. & Celis, E. (2002). Identification of an antigenic epitope for helper T lymphocytes from carcinoembryonic antigen. *Clinical cancer research*, *8*(10), 3219-3225.

(6) http://www.cancerresearch.org

(7) Campi, G., Crosti, M., Consogno, G., Facchinetti, V., Conti-Fine, B. M., Longhi, R., ... & Protti, M. P. (2003). CD4+ T cells from healthy subjects and colon cancer patients recognize a carcinoembryonic antigen-specific immunodominant epitope. *Cancer research*, *63*(23), 8481-8486.

(8) Ojima, T., Iwahashi, M., Nakamura, M., Matsuda, K., Nakamori, M., Ueda, K., ... & Yamaue, H. (2007). Successful cancer vaccine therapy for carcinoembryonic antigen (CEA)-expressing colon cancer using genetically modified dendritic cells that express CEA and T helper-type 1 cytokines in CEA transgenic mice. *International journal of cancer*, *120*(3), 585-593.

(9) Tsang, K. Y., Zaremba, S., Nieroda, C. A., Zhu, M. Z., Hamilton, J. M., & Schlom, J. (1995). Generation of human cytotoxic T cells specific for human carcinoembryonic antigen epitopes from patients immunized with recombinant vaccinia-CEA vaccine. *Journal of the National Cancer Institute*,*87*(13), 982-990.

(10) Nair, S. K., Hull, S., Coleman, D., Gilboa, E., Lyerly, H. K., & Morse, M. A. (1999). Induction of carcinoembryonic antigen (cea)-specific cytotoxic t-lymphocyte responses In vitro using autologous dendritic cells loaded with cea peptide or cea rna in patients with metastatic malignancies expressing cea. *International journal of cancer*, *82*(1), 121-124.

(11) Foon, K. A., Chakraborty, M., John, W. J., Sherratt, A., Köhler, H., & Bhattacharya-Chatterjee, M. (1995). Immune response to the carcinoembryonic antigen in patients treated with an anti-idiotype antibody vaccine. *Journal of Clinical Investigation*, *96*(1), 334.

(12) CDC, 2016: http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

(13) Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: a cancer journal for clinicians*, *65*(1), 5-29.

(14) Schlom, J. (2012). Therapeutic cancer vaccines: current status and moving forward. *Journal of the National Cancer Institute*, *104*(8), 599-613.

(15) Schwartzentruber, D. J., Lawson, D. H., Richards, J. M., Conry, R. M., Miller, D. M., Treisman, J., ... & Kendra, K. L. (2011). gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *New England Journal of Medicine*, *364*(22), 2119-2127.

(16) Disis, M. L. (2011). Immunologic biomarkers as correlates of clinical response to cancer immunotherapy. *Cancer immunology, immunotherapy*,*60*(3), 433-442.

(17) Butts, C., Murray, N., Maksymiuk, A., Goss, G., Marshall, E., Soulières, D., ... & Davis, M. (2005). Randomized phase IIB trial of BLP25 liposome vaccine in stage IIIB and IV non–small-cell lung cancer. *Journal of Clinical Oncology*, *23*(27), 6674-6681.

(18) Karbach, J., Neumann, A., Atmaca, A., Wahle, C., Brand, K., von Boehmer, L., ... & Jäger, E. (2011). Efficient in vivo priming by vaccination with recombinant NY-ESO-1 protein and CpG in antigen naive prostate cancer patients. *Clinical Cancer Research*, *17*(4), 861-870.

(19) Schuster, S. J., Neelapu, S. S., Gause, B. L., Janik, J. E., Muggia, F. M., Gockerman, J. P., ... & McGaughey, D. S. (2011). Vaccination with patient-specific tumor-derived antigen in first remission improves disease-free survival in follicular lymphoma. *Journal of Clinical Oncology*, *29*(20), 2787-2794.

(20) Gilewski, T. A., Ragupathi, G., Dickler, M., Powell, S., Bhuta, S., Panageas, K., ... & Houghton, A. N. (2007). Immunization of high-risk breast cancer patients with clustered sTn-KLH conjugate plus the immunologic adjuvant QS-21. *Clinical Cancer Research*, *13*(10), 2977-2985.

(21) Moss, B. (1996). Genetically engineered poxviruses for recombinant gene expression, vaccination, and safety. *Proceedings of the National Academy of Sciences*, *93*(21), 11341-11348.

(22) Wansley, E. K., Chakraborty, M., Hance, K. W., Bernstein, M. B., Boehm, A. L., Guo, Z., ... & Hodge, J. W. (2008). Vaccination with a recombinant Saccharomyces cerevisiae expressing a tumor antigen breaks immune tolerance and elicits therapeutic antitumor responses. *Clinical Cancer Research*, *14*(13), 4316-4325.

(23) Singh, R., & Paterson, Y. (2006). Listeria monocytogenes as a vector for tumor-associated antigens for cancer immunotherapy. *Expert review of vaccines*, *5*(4), 541-552.

(24) MacDonald, G. H., & Johnston, R. E. (2000). Role of dendritic cell targeting in Venezuelan equine encephalitis virus pathogenesis. *Journal of virology*,*74*(2), 914-922.

(25) Hoover, H. C., Brandhorst, J. S., Peters, L. C., Surdyke, M. G., Takeshita, Y., Madariaga, J., ... & Hanna, M. G. (1993). Adjuvant active specific immunotherapy for human colorectal cancer: 6.5-year median follow-up of a phase III prospectively randomized trial. *Journal of Clinical Oncology*, *11*(3), 390-399.

(26) Luiten, R. M., Kueter, E. W., Mooi, W., Gallee, M. P., Rankin, E. M., Gerritsen, W. R., ... & Sein, J. (2005). Immunogenicity, including vitiligo, and feasibility of vaccination with autologous GM-CSF–transduced tumor cells in metastatic melanoma patients. *Journal of clinical oncology*, *23*(35), 8978-8991.

(27) Avigan, D., Vasir, B., Gong, J., Borges, V., Wu, Z., Uhl, L., ... & Giallambardo, N. (2004). Fusion cell vaccination of patients with metastatic breast and renal cancer induces immunological and clinical responses.*Clinical Cancer Research*, *10*(14), 4699-4708.

(28) Lutz, E., Yeo, C. J., Lillemoe, K. D., Biedrzycki, B., Kobrin, B., Herman, J., ... & Onners, B. (2011). A lethally irradiated allogeneic granulocyte-macrophage colony stimulating factor-secreting tumor vaccine for pancreatic adenocarcinoma: a phase II trial of safety, efficacy, and immune activation.*Annals of surgery*, *253*(2), 328.

(29) Higano, C. S., Schellhammer, P. F., Small, E. J., Burch, P. A., Nemunaitis, J., Yuh, L., ... & Frohlich, M. W. (2009). Integrated data from 2 randomized, double-blind, placebo-controlled, phase 3 trials of active cellular immunotherapy with sipuleucel-T in advanced prostate cancer. *Cancer*,*115*(16), 3670-3679.

(30) Banchereau, J., & Steinman, R. M. (1998). Dendritic cells and the control of immunity. *Nature*, *392*(6673), 245-252.

(31) Kantoff, P. W., Higano, C. S., Shore, N. D., Berger, E. R., Small, E. J., Penson, D. F., ... & Xu, Y. (2010). Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *New England Journal of Medicine*,*363*(5), 411-422.

(32) Falk, K., Rötzschke, O., Grahovac, B., Schendel, D., Stevanović, S., Jung, G., & Rammensee, H. G. (1993). Peptide motifs of HLA-B35 and-B37 molecules. *Immunogenetics*, *38*(2), 161-162.

(33) Bruggen, P. V. D., Szikora, J. P., Boël, P., Wildmann, C., Somville, M., Sensi, M., & Boon, T. (1994). Autologous cytolytic T lymphocytes recognize a MAGE-1 nonapeptide on melanomas expressing HLA-Cw* 1601. *European journal of immunology*, *24*(9), 2134-2140.

(34) Traversari, C., van der Bruggen, P., Luescher, I. F., Lurquin, C., Chomez, P., Van Pel, A., ... & Boon, T. (1992). A nonapeptide encoded by human gene MAGE-1 is recognized on HLA-A1 by cytolytic T lymphocytes directed against tumor antigen MZ2-E. *The Journal of experimental medicine*, *176*(5), 1453-1457.

(35) Gajewski, T. F. (2012). Cancer immunotherapy. *Molecular oncology*, *6*(2), 242-250.

(36) Jäger, E., Chen, Y. T., Drijfhout, J. W., Karbach, J., Ringhoffer, M., Jäger, D., ... & Old, L. J. (1998). Simultaneous humoral and cellular immune response against cancer–testis antigen NY-ESO-1: Definition of human histocompatibility leukocyte antigen (HLA)-A2–binding peptide epitopes. *The Journal of experimental medicine*, *187*(2), 265-270.

(37) von Mehren, M., Arlen, P., Tsang, K. Y., Rogatko, A., Meropol, N., Cooper, H. S., ... & Weiner, L. M. (2000). Pilot study of a dual gene recombinant avipox vaccine containing both carcinoembryonic antigen (CEA) and B7. 1 transgenes in patients with recurrent CEA-expressing adenocarcinomas.*Clinical cancer research*, *6*(6), 2219-2228.

(38) Caballero, O. L., & Chen, Y. T. (2009). Cancer/testis (CT) antigens: potential targets for immunotherapy. *Cancer science*, *100*(11), 2014-2021.

(39) de Vries, T. J., Smeets, M., de Graaf, R., Hou-Jensen, K., Bröcker, E. B., Renard, N., ... & Ruiter, D. J. (2001). Expression of gp100, MART-1, tyrosinase, and S100 in paraffin-embedded primary melanomas and locoregional, lymph node, and visceral metastases: implications for diagnosis and immunotherapy. A study conducted by the EORTC Melanoma Cooperative Group. *The Journal of pathology*, *193*(1), 13-20.

(40) Gulley, J. L., Arlen, P. M., Madan, R. A., Tsang, K. Y., Pazdur, M. P., Skarupa, L., ... & Cereda, V. (2010). Immunologic and prognostic factors associated with overall survival employing a poxviral-based PSA vaccine in metastatic castrate-resistant prostate cancer. *Cancer Immunology, Immunotherapy*, *59*(5), 663-674.

(41) Wolfel, T., Hauer, M., Schneider, J., & Serrano, M. (1995). A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science*, *269*(5228), 1281.

(42) Gold, J. M., Freedman, S. O., & Gold, P. H. (1972). Human anti-CEA antibodies detected by radioimmunoelectrophoresis. *Nature*, *239*(89), 60-62.

(43) Jothy, S., Brazinsky, S. A., Chin-A-Loy, M., Haggarty, A., Krantz, M. J., Cheung, M., & Fuks, A. (1986). Characterization of monoclonal antibodies to carcinoembryonic antigen with increased tumor specificity. *Laboratory investigation; a journal of technical methods and pathology*, *54*(1), 108-117.

(44) Thompson, J. A., Grunert, F., & Zimmermann, W. (1991). Carcinoembryonic antigen gene family: molecular biology and clinical perspectives. *Journal of clinical laboratory analysis*, *5*(5), 344-366.

(45) Hammarström, S. (1999, April). The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. In *Seminars in cancer biology* (Vol. 9, No. 2, pp. 67-81). Academic Press.

(46) Pavlopoulou, A., & Scorilas, A. (2014). A comprehensive phylogenetic and structural analysis of the carcinoembryonic antigen (CEA) gene family.*Genome biology and evolution*, *6*(6), 1314-1326.

(47) Bos, M. P., Kao, D., Hogan, D. M., Grant, C. C., & Belland, R. J. (2002). Carcinoembryonic antigen family receptor recognition by gonococcal Opa proteins requires distinct combinations of hypervariable Opa protein domains. *Infection and immunity*, *70*(4), 1715-1723.

(48) Perez-Ordonez, B., & Rosai, J. (1998, May). Follicular dendritic cell tumor: review of the entity. In *Seminars in diagnostic pathology* (Vol. 15, No. 2, pp. 144-154).

(49) Tsang, K. Y., Zaremba, S., Nieroda, C. A., Zhu, M. Z., Hamilton, J. M., & Schlom, J. (1995). Generation of human cytotoxic T cells specific for human carcinoembryonic antigen epitopes from patients immunized with recombinant vaccinia-CEA vaccine. *Journal of the National Cancer Institute*,*87*(13), 982-990.

(50) Nukaya, I., Yasumoto, M., Iwasaki, T., Ideno, M., Sette, A., Celis, E., ... & Kato, I. (1999). Identification of HLA-A24 epitope peptides of carcinoembryonic antigen which induce tumor-reactive cytotoxic T lymphocyte. *International journal of cancer*, *80*(1), 92-97.

(51) Kobayashi, H., Omiya, R., Ruiz, M., Huarte, E., Sarobe, P., Lasarte, J. J., ... & Celis, E. (2002). Identification of an antigenic epitope for helper T lymphocytes from carcinoembryonic antigen. *Clinical cancer research*, *8*(10), 3219-3225.

(52) Sætrom, P., Biesinger, J., Li, S. M., Smith, D., Thomas, L. F., Majzoub, K., ... & Weitzel, J. (2009). A risk variant in an miR-125b binding site in BMPR1B is associated with breast cancer pathogenesis. *Cancer research*,*69*(18), 7459-7465.

(53) Nicolini, A., Carpi, A., & Rossi, G. (2006). Cytokines in breast cancer.*Cytokine & growth factor reviews*, *17*(5), 325-337.

(54) Winawer, S., Fletcher, R., Rex, D., Bond, J., Burt, R., Ferrucci, J., ... & Kirk, L. (2003). Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. *Gastroenterology*, *124*(2), 544-560.

(55) Kass, E. S., Greiner, J. W., Kantor, J. A., Tsang, K. Y., Guadagni, F., Chen, Z., ... & Van Waes, C. (2002). Carcinoembryonic antigen as a target for specific antitumor immunotherapy of head and neck cancer. *Cancer research*, *62*(17), 5049-5057.

(56) Brusic, V., & Petrovsky, N. (2005). Immunoinformatics and its relevance to understanding human immune disease. *Expert review of clinical immunology*, *1*(1), 145-157.

(57) Tomar, N., & De, R. K. (2010). Immunoinformatics: an integrated scenario.*Immunology*, *131*(2), 153-168

(58) Abele, R., & Tampé, R. (2004). The ABCs of immunology: structure and function of TAP, the transporter associated with antigen processing.*Physiology*, *19*(4), 216-224.

(59) Yewdell, J. W., & Bennink, J. R. (1999). Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses 1.*Annual review of immunology*, *17*(1), 51-88.

(60) Rammensee, H. G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A., & Stevanović, S. (1999). SYFPEITHI: database for HLA ligands and peptide motifs. *Immunogenetics*, *50*(3-4), 213-219.

(61) De Groot, A. S., Bosma, A., Chinai, N., Frost, J., Jesdale, B. M., Gonzalez, M. A., ... & Saint-Aubin, C. (2001). From genome to vaccine: in silico predictions, ex vivo verification. *Vaccine*, *19*(31), 4385-4395.

(62) Parker, K. C., Bednarek, M. A., & Coligan, J. E. (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *The Journal of Immunology*, *152*(1), 163-175.

(63) Davenport, M. P., Shon, I. A. H., & Hill, A. V. (1995). An empirical method for the prediction of T-cell epitopes. *Immunogenetics*, *42*(5), 392-397.

(64) Altuvia, Y., Sette, A., Sidney, J., Southwood, S., & Margalit, H. (1997). A structure-based algorithm to predict potential binding peptides to HLA molecules with hydrophobic binding pockets. *Human immunology*, *58*(1), 1-11.

(65) Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open biology*, *3*(1), 120139.

(66) Sezerman, O., & Best, G. (1997). Accurate alignment preserves polarization.*Laser focus world*, *33*(12), S27-S30.

(67) Zhang, G. L., Srinivasan, K. N., Veeramani, A., August, J. T., & Brusic, V. (2005). PREDBALB/c: a system for the prediction of peptide binding to H2d molecules, a haplotype of the BALB/c mouse. *Nucleic acids research*,*33*(suppl 2), W180-W183.

(68) Scapozza, L. E. O. N. A. R. D. O., Rognan, D., Folkers, G., & Daser, A. (1995). Molecular dynamics and structure-based drug design for predicting non-natural nonapeptide binding to a class I HLA protein. *Acta Crystallographica Section D: Biological Crystallography*, *51*(4), 541-549.

(69) Barnett, T., Goebel, S. J., Nothdurft, M. A., & Elting, J. J. (1988). Carcinoembryonic antigen family: characterization of cDNAs coding for NCA and CEA and suggestion of nonrandom sequence variation in their conserved loop-domains. *Genomics*, *3*(1), 59-66.

(70) Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*, *8*(1), 1.

(71) Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., & Nielsen, M. (2008). Net HLA-3.0: accurate web accessible predictions of human, mouse and monkey HLA class I affinities for peptides of length 8–11.*Nucleic acids research*, *36*(suppl 2), W509-W512.

(72) Singh, H., & Raghava, G. P. S. (2001). ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, *17*(12), 1236-1237.

(73) Peters, B., & Sette, A. (2005). Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC bioinformatics*, *6*(1), 1

(74) Toseland, C. P., Clayton, D. J., McSparron, H., Hemsley, S. L., Blythe, M. J., Paine, K., ... & Flower, D. R. (2005). AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome research*, *1*(1), 1

(75) Peters, B., Sidney, J., Bourne, P., Bui, H. H., Buus, S., Doh, G., ... & Nemazee, D. (2005). The immune epitope database and analysis resource: from vision to blueprint. *PLoS biol*, *3*(3), e91.

(76) Shen, Y., Maupetit, J., Derreumaux, P., & Tufféry, P. (2014). Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *Journal of chemical theory and computation*, *10*(10), 4745-4758.

(77) Thévenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P., & Tufféry, P. (2012). PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic acids research*,*40*(W1), W288-W293.

(78) 3silico analysis of six known Leishmania major antigens and in vitro evaluation of specific epitopes eliciting HLA-A2 restricted CD8 T cell response. *PLoS Negl Trop Dis*, *5*(9), e1295.