In silico T-cell epitope-based vaccine designing against *Mycobacterium tuberculosis*

A DISSERTATION SUBMITTED TO BRAC UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE IN BIOTECHNOLOGY

Submitted by: Winifred Claire Mondol
Student ID: 14136009

Biotechnology Program
Department of Mathematics and Natural Sciences
BRAC University
66, Mohakhali, Dhaka-1212
Bangladesh
September 2018

Declaration

I, Winifred Claire Mondol declare that this thesis and the work entitled **"*In silico* T-cell epitope-based vaccine designing against *Mycobacterium tuberculosis*"** submitted to the Department of Mathematics and Natural Sciences (MNS), BRAC University in partial fulfillment of the requirements for the degree of Bachelor of Science in Biotechnology is a record of work carried out by me under the supervision of my supervisors.

I further declare that this thesis has been composed solely by me and it has not been submitted, in whole or in part, in any previous institution for a degree or diploma. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

Candidate

_____

Winifred Claire Mondol

Certified:

Shamira Tabrejee
Supervisor,
Lecturer (On Leave), Biotechnology Program
Department of Mathematics and Natural Sciences, BRAC University

Dr. Mahboob Hossain
Co-supervisor,
Professor, Department of Mathematics and Natural Sciences, BRAC University

_____ _____

ShamiraTabrejee Dr. Mahboob Hossain

Dedicated to my parents

Acknowledgements

The completion of my undergraduate dissertation would not have been possible without the constant assistance of Almighty in every phase of my life.

I am highly indebted to my **Parents** for all the love and support they have provided on every step of the way.

It is an honor for me to express my sincere gratitude and thank respected **A F M Yusuf Haider**, Ph.D., Professor and Chairperson of Department of Mathematics and Natural Sciences, BRAC University, for his active cooperation and encouragement. A special word of gratitude is due to Late **Professor Dr. A. A. Ziauddin Ahmad** (Former Chairperson, Department of Mathematics and Natural Sciences) for his constant support and guidance.

I am immensely grateful and would like to express my gratefulness to my supervisors, **Shamira Tabrejee**, Lecturer, Department of Mathematics and Natural Sciences, BRAC University and **Dr. Mahboob Hossain,** Professor, Department of Mathematics and Natural Sciences, BRAC University for their constant supervision, constructive criticism and enthusiastic encouragement throughout my research work. Without their support, it would not have been possible to complete the study. I would like to thank them for being very understanding and guiding me to complete my undergraduate dissertation.

I would also like to thank and express my heartiest gratitude to **Dr. Aparna Islam,** Professor, Department of Mathematics and Natural Sciences, BRAC University and **Romana Siddique**, Senior Lecturer, Department of Mathematics and Natural Sciences, BRAC University for their enormous support and guidance.

I would like to cordially thank my friend, **Sheikh Anushe** for being my strength and encouraging me. My special thanks to my friend **Nashrah Mustafa** for her support.

Last but not the least I would like to thank my husband, **Farhan Zahir** for his valuable support and cooperation.

Sincerely,
Winifred Claire Mondol
September 2018

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3 D	3 Dimensional
APC	Antigen Presenting Cell
BCG	Bacillus Calmette–Guérin
bp	Base pair
Da	Dalton
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
GMQE	Global Model Quality Estimation
HIV	Human Immunodeficiency Virus
HLA	human leukocyte antigen
IEDB	Immune Epitope Database
IFNγ	Interferon
IL	Interleukin
kcal/mol	male:female (M:F)
M. africanum	*Mycobacterium africanum*
M. bovis	*Mycobacterium bovis*
M. microti	*Mycobacterium microti*
M. tuberculosis, Mtb	*Mycobacterium tuberculosis*
MDR-TB	Multidrug-resistant Tuberculosis
MHC	Major histocompatibility complex
MSA	Multiple sequence alignment
NCBI	National Center for Biotechnology Information
NCBI	National Center for Biotechnology Information
NK T cells	Natural Killer
nm	Nanometer
PVS	Protein Variability Server
RNA	Ribonucleic acid
rRNA	Ribosomal Ribonucleic Acid
SA	Structural Alphabet
SMM	Stabilized matrix method
TB	Tuberculosis
TCR	T-cell receptor
tRNA	Transfer Ribonucleic Acid
WHO	World Health Organization

Abstract

Mycobacterium tuberculosis is an obligate pathogenic bacterial species in the family Mycobacteriaceae and the causative agent of tuberculosis. At present BCG, an attenuated strain of *Mycobacterium bovis* is used as a vaccine against tuberculosis. However, the overall success of BCG is arguable as it has some serious limitations. Some of these include BCG's inability to protect against TB in adults and also in immunosuppressed patients. Thus, it is necessary to develop vaccines that can replace BCG. In this study, various computational methods were employed to identify T-cell epitopes from the ESX-2 secretion-associated protein EspG2, which has the potential for vaccine development against *Mycobacterium tuberculosis*. After analyzing the immune parameters of ESX-2 secretion-associated protein EspG2 using various databases and bioinformatics tools which included IEBD, PEP-FOLD, PyRx, PyMol, etc. One T cell epitope was identified which may be used as epitope-based peptide vaccine. Five highly conserved, non- allergenic, non-cytotoxic putative T-cell epitopes were analyzed for their binding with the HLA-C 12*03 molecule. Amongst them one epitope was chosen which interacted with the maximum number of MHC alleles with satisfactory world population coverage. Docking simulation assay further revealed that SGQRRYQVL has significantly lower binding energy, which verifies that the binding cleft epitope interaction to HLA molecule will occur when it will be applied *in vivo.* Additional *in vivo* investigation can further provide concrete evidence that SGQRRYQVL be used as a peptide vaccine to effectively promote immunity against TB.

CHAPTER 01: INTRODUCTION

INTRODUCTION

1.1 Background of the study:

Today, tuberculosis (TB) is one of the top 10 causes of death worldwide and in 2016, 10.4 million people were affected by the disease and 1.7 million died. This disease is caused by the bacterium *Mycobacterium tuberculosis*, which can form lesions that lead to cell death in any tissue or organ. The lungs are most commonly affected. Patients suffer fever and loss of body weight, and without treatment, tuberculosis is often fatal. Multidrug-resistant TB (MDR-TB) remains a public health crisis and a health security threat. WHO estimates that there were 600 000 new cases with resistance to rifampicin – the most effective first-line drug, of which 490 000 had MDR-TB [1].

TB, known also as the white plague, has been around for ages. Early last century, expectations were that TB could be conquered by vaccination with the newly developed M. bovis BCG vaccine, isolated by and named after Calmette and Guérin in Lille, France [2]. The development of the first anti-tuberculous drugs during WWII by Selman Waksman, who discovered that streptomycin was bacteriostatic for Mtb [3] further, boosted the expectation. Initially, treatment with streptomycin appeared highly effective, but problems arose when drug resistance rapidly developed. The misconception that TB could be conquered by antibiotics and BCG vaccination led to complacency for several decades. This situation dramatically changed only in the early 1990s, when the World Health Organization (WHO) declared TB as a global emergency. From that time onwards TB scientists, who had been focusing much of their efforts on other areas of research and development due to a lack of interest and funding for TB, were able to reorient efforts and initiate significant activities in the study of TB [4]. In most cases, TB can be treated successfully with multidrug combinations of antibiotics (except for MDR-, XDR-, and TDR-TB), treating TB cases is clearly insufficient to interrupt disease transmission in highly endemic populations [5]. A bacterial disease that was thought to be under control has again become a serious global public health problem.

Currently, in some countries, BCG, an attenuated strain of *Mycobacterium bovis* that was developed between 1906 and 1919, is still used as a vaccine against tuberculosis. However, the overall efficacy of BCG is controversial, and the use of this vaccine has some serious limitations. BCG has failed to protect against TB in different parts of the world, especially in adults with pulmonary TB [6]. While the BCG vaccine can prevent disseminated disease and mortality in newborns and children, it cannot prevent chronic infection or protect against pulmonary tuberculosis in adults. Consequently, *M. tuberculosis* establishes a latent chronic infection that

reactivates when there are diminished immune responses, e.g., in aged people, in individuals with genetic immune defects, and in those whose medication reduces their immune responses, such as patients treated with antibodies against tumor necrosis factor alpha. Immunosuppression caused by HIV is now an extremely important factor in the reactivation of tuberculosis, and in the 15 million people co-infected by HIV and *M. tuberculosis*, it is the major cause of mortality in this population. About 2 billion people carry a latent *M. tuberculosis* infection, and approximately 10% progress to the active disease at some time. Additional limitations to the use of BCG as a vaccine are that individuals treated with BCG respond positively to a common tuberculosis diagnostic test, which makes it impossible to distinguish between individuals infected with *M. tuberculosis* and those inoculated with BCG cells [7]. To overcome these issues of the currently available vaccine, an alternative is essential. It is necessary to design a vaccine that is safer, more immunogenic and induces longer lasting protection.

An answer could be peptide vaccines which is an alternative approach to immunization that requires the identification of peptide epitopes on immunogens that trigger the immunogenic response and use the synthetic versions of those peptides while engineering the vaccine. Unlike traditional vaccines, peptide vaccines being entirely synthetic do not carry the risk of reversion or of incomplete inactivation and the epitopes could be selected to avoid components that give rise to unwanted side effects.

1.2 Objectives of study

The objectives of this present study are to employ various bioinformatics tools in order to-

a) Predict and identify the linear T-cell epitopes from the EspG2 protein and analyze their conservancy in all the available strains of *Mycobacterium tuberculosis*
b) Analyze MHCI and MHCII interaction with the selected linear T-cell epitopes
c) Analyze the world population coverage by the predicted T- cell epitopic peptides
d) Determine the allergenicity and toxicity of the predicted T-cell epitopes
e) Predict the 3D structures of conserved T-cell epitopes and selected MHC I allele
f) Perform the docking simulation assay of conserved T-cell epitopes with the selected MHC I allele

1.3 Literature review

1.3.1 *Mycobacterium tuberculosis*: an Introduction

Mycobacterium tuberculosis is an obligate pathogenic bacterial species in the family Mycobacteriaceae and the causative agent of tuberculosis. Mycobacteria are aerobic, non-spore forming non-motile, slightly curved or straight rods. Colony morphology varies among species. The characteristic features of the tubercle bacillus include its slow growth, dormancy, complex cell envelope with an unusual waxy coating on its cell, intracellular pathogenesis, and genetic homogeneity. The generation time of *M. tuberculosis*, in synthetic medium or infected animals, is usually around 24 hours and thus the growth of mycobacterial species is slower compared to other bacteria. This contributes to the chronic nature of the disease. In the state of dormancy the bacillus remains inactive within the infected tissue. As immunity declines, through ageing or immune suppression, the dormant bacteria reactivate, causing an outbreak of disease often many decades after the initial infection. The cell envelope of *M. tuberculosis*, a Gram-positive bacterium with a G + C-rich genome, contains an additional layer beyond the peptidoglycan that is exceptionally rich in unusual lipids, glycolipids and polysaccharides. Novel biosynthetic pathways generate cell-wall components such as mycolic acids, mycocerosic acid, phenolthiocerol, lipoarabinomannan and arabinogalactan [8].It is thought that the progenitor of the *M. tuberculosis* complex, comprising *M. tuberculosis*, *M. bovis*, *M. bovis* BCG, *M. africanum* and *M. microti*, arose from a soil bacterium and that the human bacillus may have been derived from the bovine form following the domestication of cattle. The complex lacks interstrain genetic diversity, and nucleotide changes are very rare [9].

Since its isolation in 1905, the H37Rv strain of *M. tuberculosis* has found extensive worldwide application in biomedical research because it has retained full virulence in animal models of tuberculosis, unlike some clinical isolates; it is also susceptible to drugs and amenable to genetic manipulation. An integrated map of the 4.4 mega base (Mb) circular chromosome of this slow-growing pathogen had been established previously and ordered libraries of cosmids and bacterial artificial chromosomes (BACs) are available [10,11].

Figure 1.1 Scanning electron micrograph of *Mycobacterium tuberculosis* bacteria, which causes TB. Credit; NIAID

1.3.2 Organization and sequence of the genome

The genome of *M. tuberculosis* was studied generally using the strain *M. tuberculosis* H37Rv. The contiguous genome sequence was obtained by a combined approach that involved the systematic sequence analysis of selected large-insert clones (cosmids and BACs) as well as random small-insert clones from a whole-genome shotgun library. This approach revealed culminated in a composite sequence of 4,411,529 base pairs (bp) (Figs 1.2), with a G + C content of 65.6%. The genome is rich in repetitive DNA, particularly insertion sequences, and in new multigene families and duplicated housekeeping genes. The G + C content is relatively constant throughout the genome. Several regions show higher than average G + C content (Fig. 1.2); these correspond to sequences belonging to a large gene family that includes the polymorphic G + C-rich sequences (PGRSs).

There are fifty genes coding for functional RNA molecules. Most of the insertion sequences in *M. tuberculosis* H37Rv appear to have inserted in intergenic or non-coding regions, often near tRNA genes. Many are clustered, suggesting the existence of insertional hot-spots that prevent genes from being inactivated. 3,924 open reading frames were identified in the genome, accounting for ~91% of the potential coding capacity. A few of these genes appear to have in-frame stop codons or frameshift mutations and may either use frameshifting during translation or correspond to pseudogenes. Consistent with the high G + C content of the genome, GTG initiation codons (35%) are used more frequently than in *Bacillus subtilis* (9%) and *E. coli* (14%), although ATG (61%) is the most common translational start [8].

Genes that code for lipid metabolism are a very important part of the bacterial genome, and 8% of the genome is involved in this activity. The different species of the *Mycobacterium*

tuberculosis complex show a 95-100% DNA relatedness based on studies of DNA homology, and the sequence of the 16S rRNA gene is exactly the same for all the species.

Plasmids in *M. tuberculosis* are important in transferring virulence because genes on the plasmids are more easily transferred than genes located on the chromosome. The cell envelope contains a polypeptide layer, a peptidoglycan layer, and free lipids. The *M. tuberculosis* cell wall contains three classes of mycolic acids: alpha-, keto- and methoxymycolates. The cell wall also contains lipid complexes including acyl glycolipids and other complexes such as free lipids and sulfolipids. There are porins in the membrane to facilitate transport. Beneath the cell wall, there are layers of arabinogalactan and peptidoglycan that lie just above the plasma membrane.
The *M. tuberculosis* genome encodes about 190 transcriptional regulators, including 13 sigma factors, 11 two-component system, and more than 140 transcription regulators. Several regulators have been found to respond to environmental distress.

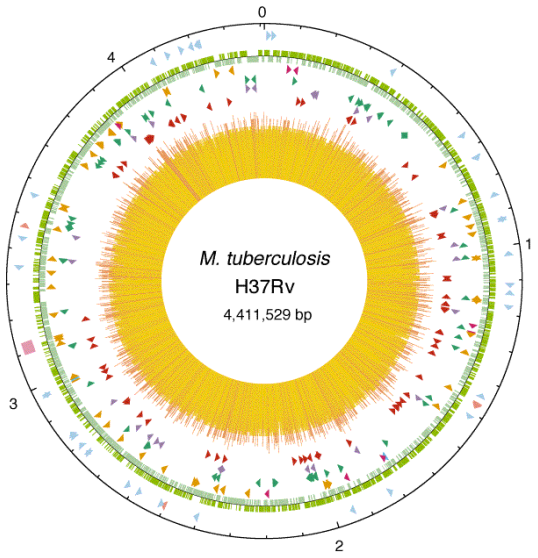


Figure 1.2 Circular map of the chromosome of *M. tuberculosis* H37Rv.The outer circle shows the scale in Mb, with 0 representing the origin of replication. The first ring from the exterior denotes the positions of stable RNA genes (tRNAs are blue, others are pink) and the direct repeat region (pink cube); the second ring inwards shows the coding sequence by strand (clockwise, dark green; anticlockwise, light green); the third ring depicts repetitive DNA (insertion sequences, orange; 13E12 REP family, dark pink; prophage, blue); the fourth ring shows the positions of the PPE family members (green); the fifth ring shows the PE family members (purple, excluding PGRS); and the sixth ring shows the positions of the PGRS sequences (dark red). The histogram (cent re) represents G + C content, with <65% G + C in yellow, and >65% G + C in red. The figure was generated with software from DNASTAR.

Table 1.1 General classifications of M. tuberculosis genes [12]

Function	No. of genes	% of total	% of Total coding capacity
Lipid metabolism	225	5.7	9.3
Information pathways	207	5.2	6.1
Cell wall and cell processes	517	13.0	15.5
Stable RNAs	50	1.3	0.2
IS elements and bacteriophages	137	3.4	2.5
PE and PPE proteins	167	4.2	7.1
Intermediary metabolism and respiration	877	22.0	24.6
Regulatory proteins	188	4.7	4.0
Virulence, detoxification and adaptation	91	2.3	2.4
Conserved hypothetical function	911	22.9	18.4
Proteins of unknown function	607	15.3	9.9

1.3.3 Species

The *Mycobacterium tuberculosis* complex (MTC) consists of *Mycobacterium africanum, Mycobacterium bovis, Mycobacterium canettii, Mycobacterium microti, Mycobacterium tuberculosis* [9]. Gene Rv3889c of gene family ABC transporter family signature of *Mycobacterium tuberculosis* is the EspG2 gene which produces the protein ESX-2 secretion-associated protein G, EspG2, hence the protein of interest in this study [13].

1.3.4 Proteins

Mycobacterium tuberculosis has about 4000 genes which in turn express about 30000 proteins carrying out different functions for the bacteria. Amongst which are ESXC, ESX-2 secretion-associated protein EspG2, eccD2 secretion system protein, probable alanine and proline rich membrane-anchored mycosin mycP2, eccE2, eccA2, etc. with the presence of many transmembrane proteins. Two mysterious protein families, called Pro-Glu motif–containing (PE)

and Pro-Pro-Glu motif–containing (PPE) proteins, are highly expanded in Mtb and have been linked to virulence. ESAT-6 secretion system (ESX) secretion-associated protein G (EspG), a component of the secretion system translocates PE-PPE proteins to the bacterial cell surface. Epsg2 is the gene that expresses ESX-2 secretion-associated protein G, EspG2. It acts as a specific chaperone for cognate PE/PPE proteins. It is present in the cytoplasm. It plays an important role in preventing aggregation of PE/PPE dimers and interacts specifically with ESX-2-dependent PE/PPE proteins [14]. The gene is located at 4372800 bp which is shown in Fig 1.3. The protein has a length of 276 amino acids, the mass of 29,596 Da and an isoelectric point of 6.1372.

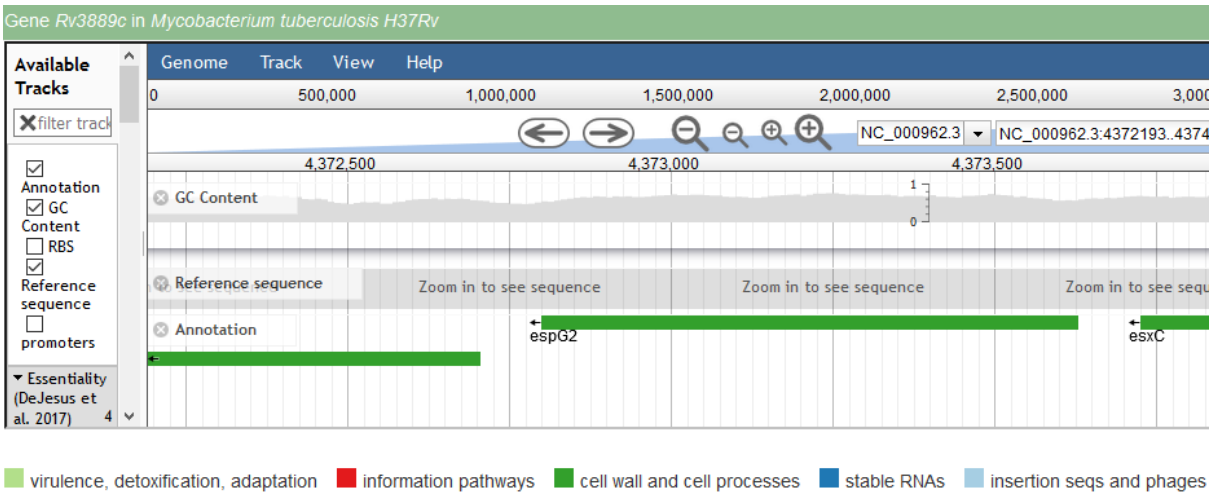


Figure 1.3 Location of the gene, EspG2 [15].

1.3.5 Pathogenesis

Infection can develop when a person breathes in tubercle bacilli from expelled droplets from an infected individual. The droplets reach the alveoli of the lungs where the bacilli can be deposited. Alveolar macrophages ingest the tubercle bacilli and destroy most of them. Some can multiply within the macrophage and be released when the macrophage dies. From there, the bacilli can spread to other regions of the body through the bloodstream. The areas in which TB is most likely to develop are the apex of the lung, the kidneys, the brain, bones and lymph nodes.

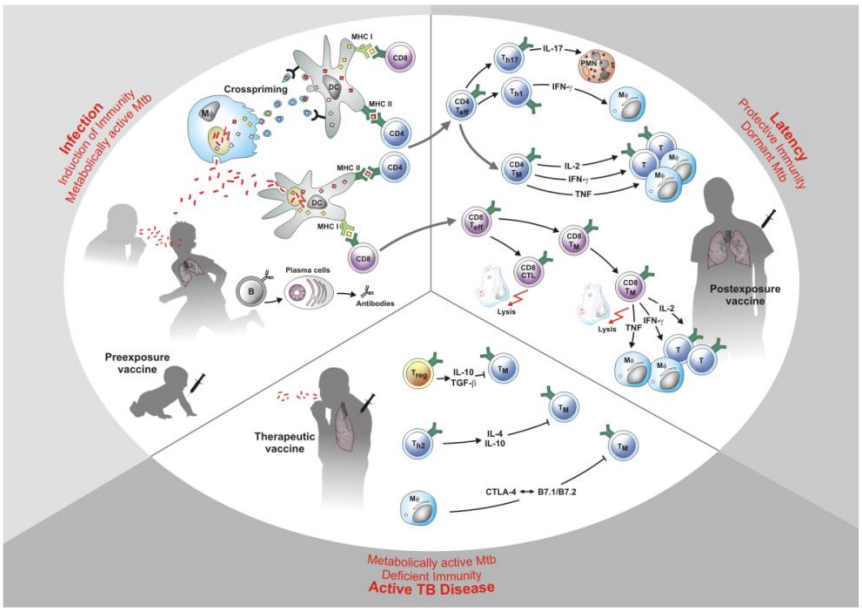


Figure 1.4 The three stages of tuberculosis.

Stage 1: Infection of *Mycobacterium tuberculosis* (*Mtb*) frequently occurs at a young age. Metabolically active *Mtb* are inhaled and subsequently, T-cells are stimulated which carry the major burden of acquired immunity. These include major histocompatibility complex class II (MHC II)-restricted CD4 T-cells and MHC I-restricted CD8 T-cells.

Stage 2: Acquired immunity comprising CD4 and CD8 T-cells contains *Mtb* in a dormant stage within solid granulomas. T-cells produce type I cytokines and cytolytic effector molecules. They become memory T-cells which concomitantly produce multiple cytokines. Individuals remain latently infected without clinical signs of active tuberculosis (TB).

Stage 3: Mechanisms leading to deficient immunity and disease reactivation are numerous and include production of suppressive cytokines including interleukin (IL)-10 and transforming growth factor-beta (TGFβ) by T helper 2 (Th2) cells and regulatory T(reg) cells as well as T-cell exhaustion mediated by inhibitory receptor- co-receptor interactions on antigen presenting cells (APCs) and T-cells. *Mtb* becomes metabolically active and granulomas become caseous. *Mtb* can be spread to other organs and to other individuals [16].

1.3.6 Symptoms and Diagnosis of TB

Common symptoms of active lung TB are a cough with sputum and blood at times, chest pains, weakness, weight loss, fever and night sweats. Many countries still rely on a long-used method called sputum smear microscopy to diagnose TB. Trained laboratory technicians look at sputum

samples under a microscope to see if TB bacteria are present. Microscopy detects only half the number of TB cases and cannot detect drug-resistance [17].

Diagnostic tests for TB disease include the following:

Rapid molecular tests– The only rapid test for diagnosis of TB currently recommended by the WHO is the XpertR MTB/RIF assay. It can provide results within 2 hours. The test has much better accuracy than sputum smear microscopy;

Sputum smear microscopy–this technique requires the examination of sputum samples using a microscope to determine the presence of bacteria. In the current case definitions recommended by WHO, one positive result is required for a diagnosis of smear-positive pulmonary TB;

Culture-based methods– The current reference standard and require more developed laboratory capacity and can take up to 12 weeks to provide results. There are also tests for TB that is resistant to first-line and second-line anti-TB drugs. They include Xpert MTB/ RIF; rapid line probe assays (LPAs); a rapid LPA; and sequencing technologies. Culture-based methods currently remain the reference standard for drug susceptibility testing.

1.3.7 Statistics

Globally in 2016, 6.6 million people with tuberculosis (TB) were notified to national TB programmes (NTPs) and reported to WHO. Of these, over 6.3 million had an incident episode (new or relapse) of TB. The global number of new and relapse TB cases per 100 000 population have both been increasing since 2013. In 2016, 41% of the 3.4 million new bacteriologically confirmed and previously treated TB cases notified globally were reported to have been tested for resistance to rifampicin, up from 31% in 2015. Coverage was 33% for new TB patients and 60% for previously treated TB patients. Globally, 153 119 cases of multidrug-resistant TB and rifampicin resistant TB (MDR/RR-TB) were notified in 2016, and 129 689 were enrolled in treatment. Globally in 2016, 57% of notified TB patients had a documented HIV test result, up from 55% in 2015 and a 19-fold increase since 2004. Globally, children (aged <15 years) accounted for 6.9% of the new and relapse cases that were notified in 2016.

Most of the global increase in notifications of new TB cases since 2013 is explained by a 37% increase in India 2013–2016. The global male:female (M:F) ratio for notifications was 1.7. Results from national TB prevalence surveys of adults show higher M:F ratios, indicating that notification data understate the share of the TB burden accounted for by men in some countries. Table 1.2 shows the Notifications of TB, HIV-positive TB and MDR/RR-TB cases, globally and for WHO regions in 2016 [17].

Table 1.2 Notifications of TB, HIV-positive TB and MDR/RR-TB cases, globally and for WHO regions, 2016

	TOTAL NOTIFIED	NEW AND RELAPSE*	PULMONARY NEW AND RELAPSE		EXTRAPULMONARY NEW AND RELAPSE (%)	HIV-POSITIVE NEW AND RELAPSE	MDR/RR-TB	XDR-TB
			NUMBER	OF WHICH BACTERIOLOGICALLY CONFIRMED (%)				
Africa	1 303 483	1 273 560	1 065 327	66%	16%	358 237	27 828	1 092
The Americas	233 793	221 008	186 940	77%	15%	20 528	3 715	112
Eastern Mediterranean	527 693	514 449	390 367	53%	24%	1 367	4 713	152
Europe	260 434	219 867	187 898	64%	15%	24 871	49 442	3 114
South-East Asia	2 898 482	2 707 879	2 291 793	61%	15%	60 245	46 269	2 926
Western Pacific	1 400 638	1 372 371	1 268 798	38%	8%	11 526	21 152	618
GLOBAL	6 624 523	6 309 134	5 391 123	57%	15%	476 774	153 119	8 014

1.3.8 Medication

TB is a treatable and curable disease. Consequently, treatment requires the use of multiple drugs for several months. With appropriate antibiotic treatment, TB can be cured in most people. Currently, there are 10 drugs approved by the U.S. Food and Drug Administration for the treatment of TB. Of the approved drugs, isoniazid (INH), rifampin (RIF), ethambutol (EMB), and pyrazinamide (PZA) are considered first-line anti-tuberculosis agents. These four drugs form the foundation of initial courses of therapy. Treatment of active, drug-susceptible TB disease usually combines the four different antimicrobial drugs mentioned above that are provided with information, supervision, and support to the patient by a health worker or trained volunteer over a course of 6 months, sometimes for as long as 12 months. Without such support, treatment adherence can be difficult and the disease can spread. The vast majority of TB cases can be cured when medicines are provided and taken properly [18]. However, many *M. tuberculosis* strains are resistant to one or more of the standard TB drugs, which complicates treatment greatly.

1.3.9 Vaccine

Current vaccines typically consist of either an inactivated or an attenuated pathogen. Usually, the pathogen is grown in culture, purified, and either inactivated or attenuated without losing its ability to induce an immune response that is effective against the virulent form of the pathogen.

Adaptation to growth in culture, the requirement for large-scale growth, production and yield, relevant safety issues, insufficient attenuation, the introduction of virulent organisms, attenuation

reversion, limited shelf life, and need for refrigeration pose some of the limitations associated with vaccine production. Significant advances have been made to develop vaccines by optimization of delivery systems, size of particulate vaccines, targeting of antigen-presenting dendritic cells, and the addition of components (T-cell epitopes) to improve vaccine efficacy. However, some problems must still be solved to successfully develop these vaccines. The vaccine developed should not elicit illness or death, and it must protect against illness resulting from exposure to the live pathogen. Protection against illness must be sustained for several years resulting in T- and B-cell-mediated immune memory. In the case of intracellular, it must induce a protective CD8+ CTL response. Vaccines should be cost-effective biologically stable with ease of administration and have few (or no) adverse side effects.

1.3.10 Types of vaccine

There are many approaches to designing vaccines against a microbe. The choices are usually based on fundamental information about the microbe, such as how it infects cells and how the immune system responds to it, as well as practical considerations, such as regions of the world where the vaccine would be used. Table 1.3 tabulates some of the types of vaccine strategies [7]. The following are some of the different types of vaccines:

- **Live, Attenuated Vaccines:**

Live, attenuated vaccines contain a version of the living microbe that has been weakened in the lab so it is unable to cause disease. Because a live, attenuated vaccine is the closest thing to a natural infection, these vaccines elicit strong cellular and antibody responses and often confer lifelong immunity with only one or two doses.

Despite the advantages of live, attenuated vaccines, there are some disadvantages. There is a possibility of the live attenuated microbe in the vaccine to revert to a virulent form and cause disease. Also, people who have damaged or weakened immune systems cannot safely receive live, attenuated vaccines. Another limitation is that live, attenuated vaccines usually need to be refrigerated to stay potent. If the vaccine needs to be shipped overseas and stored by healthcare workers in developing countries that lack widespread refrigeration, a live vaccine may not be the best choice. Live, attenuated vaccines are more difficult to create for bacteria. Bacteria have thousands of genes and thus are much harder to control.

- **Inactivated Vaccines**

Scientists produce inactivated vaccines by killing the disease-causing microbe with chemicals, heat, or radiation. Such vaccines are more stable and safer than live vaccines as the dead microbes cannot mutate back to their disease-causing state. Inactivated vaccines usually don't require refrigeration, and they can be easily stored and transported in a freeze-dried form. Most inactivated vaccines, however, stimulate a weaker immune system response than do live vaccines. So it would likely take several additional doses, or booster shots, to maintain a person's immunity. This could be a drawback in areas where people don't have regular access to health care and cannot get booster shots on time.

- **Toxoid Vaccines**

These vaccines are used when a bacterial toxin is the main cause of illness. Scientists have found that they can inactivate toxins by treating them with formalin, a solution of formaldehyde and sterilized water. When the immune system receives a vaccine containing a harmless toxoid, it learns how to fight off the natural toxin. The immune system produces antibodies that lock onto and block the toxin.

- **Conjugate Vaccines**

With a bacterium that possesses an outer coating of polysaccharides which disguise a bacterium's antigens so that the immature immune systems of infants and younger children cannot recognize or respond to them. Conjugate vaccines, a special type of subunit vaccine, get around this problem. When making a conjugate vaccine, scientists link antigens or toxoids from a microbe that an infant's immune system can recognize to the polysaccharides. The linkage helps the immature immune system react to polysaccharide coatings and defend against the disease-causing bacterium [18].

Table 1.3 Examples of different types of vaccines are provided, and the nature of the protective immune responses induced by these vaccines is summarized [7].

Type of vaccine	Example(s)	Form of protection
Subunit (antigen) vaccines	Tetanus toxoid, diphtheria toxoid	Antibody response
Conjugate (peptide) vaccines	*Haemophilus influenzae* infection	Th cell-dependent antibody response
DNA vaccines	Clinical trials ongoing for several infections	Antibody and cell-mediated immune responses
Live attenuated or killed bacteria	BCG, cholera	Antibody response
Live attenuated viruses	Polio, rabies	Antibody and cell-mediated immune responses
Vector vaccines (viruses, bacteria)	Clinical trials of HIV and vaccinia virus and tuberculosis bacteria	Antibody and cell-mediated immune responses

1.3.11 *M. bovis* BCG: The Current TB Vaccine

BCG is one of the most widely administered vaccines worldwide, having been given over 4 billion times. Individuals with genetic defects in key immune genes or infants with clinically active HIV infection are highly susceptible to developing disseminating BCG disease, posing a significant risk in HIV-burdened populations where TB is often highly endemic [19]. Despite the relative efficacy of BCG in infants, BCG fails to prevent pulmonary TB in adolescents. In addition to that since BCG is an attenuated live vaccine, it might revert back to its virulent state. BCG is insufficient for worldwide TB control. Thus, there is a strong need to develop vaccines that can either boost BCG's initial priming and protective effects or replace BCG by superior vaccines.

1.3.12 Peptide Based Vaccine

Instead of the entire microbe, subunit vaccines include only the antigens that best stimulate the immune system. In some cases, these vaccines use epitopes—the very specific parts of the antigen that antibodies or T cells recognize. They bind to a small peptide fragment of a protein, that is accessible to antibody binding act as an effective subunit vaccine and induce the production of neutralizing antibodies. They are immunologically recognized by an antibody. It can be anticipated that short peptides that mimic epitopes (antigenic determinants) are immunogenic and may be used as peptide vaccines.Since subunit vaccines contain only the essential antigens and not all the other molecules that make up the microbe, the chances of adverse reactions to the vaccine are lower. Firstly, it has to be effective; an epitope must consist of a short stretch of contiguous amino acids, which does not always occur naturally. Secondly, the peptide must be able to assume the same conformation as the epitope in the intact viral particle and a single epitope may not be sufficiently immunogenic. Conventional vaccines include unnecessary antigenic load that, not only contributes little to the protective immune response, but complicates the situation by inducing allergenic and/or reactogenic responses. Peptide vaccines are an attractive alternative strategy that relies on the usage of short peptide fragments to engineer the induction of highly targeted immune responses, consequently avoiding allergenic and/or reactogenic sequences. Conversely, peptide vaccines used in isolation are often weakly immunogenic and require particulate carriers for delivery and adjuvanting [7].

T cell epitopes are presented on the surface of an antigen-presenting cell, where they are bound to MHC molecules. In humans, professional antigen-presenting cells are specialized to present MHC class II peptides, whereas most nucleated somatic cells present MHC class I peptides. T cell epitopes presented by MHC class I molecules are typically peptides between 8 and 11 amino acids in length, whereas MHC class II molecules present longer peptides, 13-17 amino acids in length. Usage of databases containing known T cell epitopes or peptides including information of their respective MHC binding and affinity of binding in case of development of peptide vaccines, selecting the correct target MHC or HLA is very vital, as the vaccine candidate should bind to the majority of HLAs in the population.

CHAPTER 02: MATERIALS AND METHODS

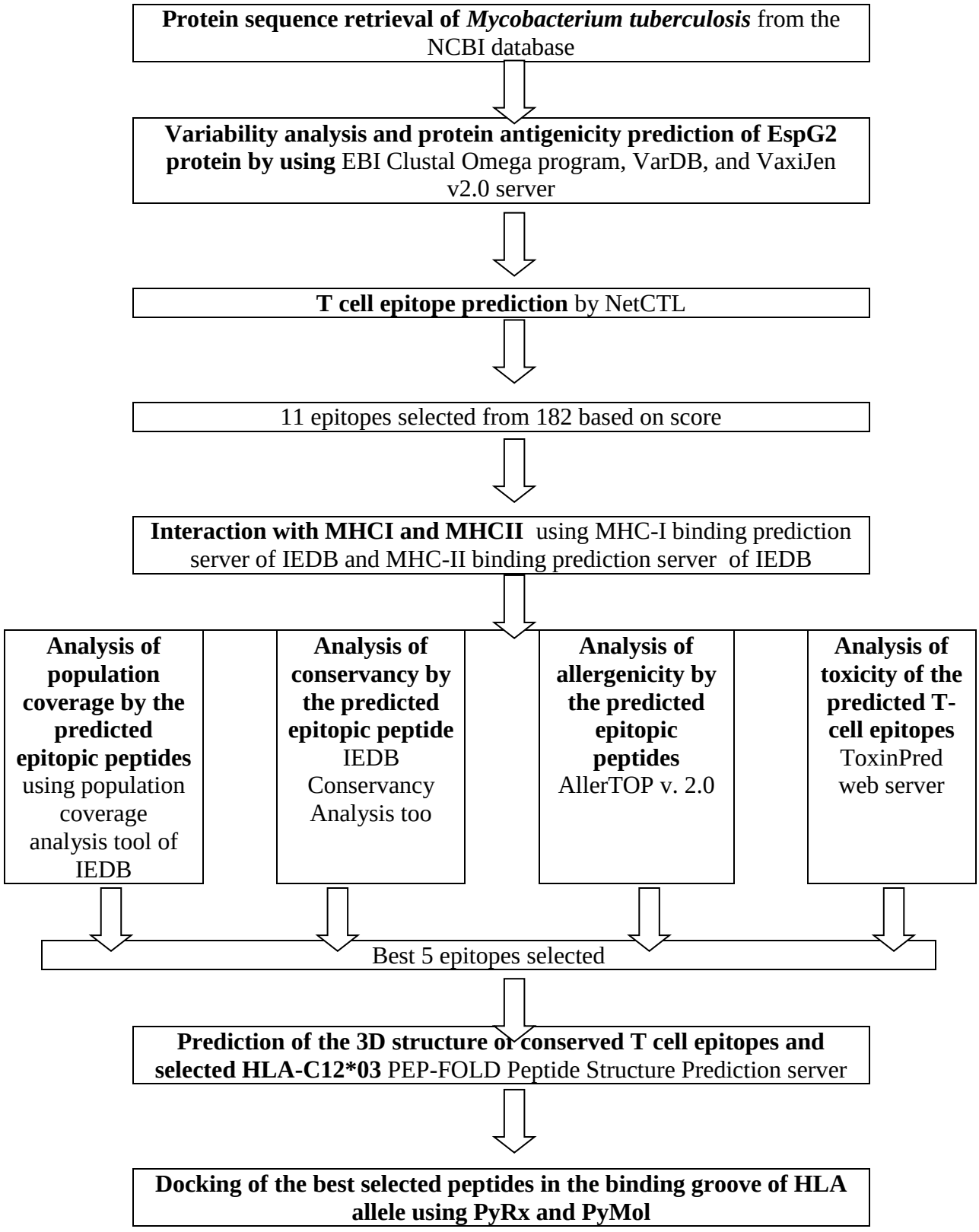


Figure 2.1 Flowchart displaying the protocols employed to predict T cell epitope

MATERIALS AND METHODS

2.1. Protein sequence retrieval

A total of 34 sequences of the EspG2 protein of *Mycobacterium tuberculosis* were retrieved from the NCBI database [20]. All the sequences were extracted from the database in the FASTA format. The dates of when they were isolated were also taken into account to cover the maximum outbreaks from the past. The length of each of the 34 sequences for *Mycobacterium tuberculosis* was 276 amino acids.

2.2. Variability analysis and protein antigenicity prediction of EspG2 protein

To analyze the level of conservation, all the retrieved sequences were aligned by using EBI-Clustal Omega program [21] and a multiple sequence alignment (MSA) was obtained. VarDB [22] is used which uses several information content metrics to calculate the degree of sequence variability within a multiple sequence alignment. In order to develop a peptide vaccine, it is essential to identify those proteins which display antigenic features. A reference EspG2 protein having accession number, ALB21149.1, was tested for its antigenicity. VaxiJen v2.0 server [23] was used to predict the antigenic property of the given sequence. VaxiJen makes the use of an approach independent of alignment, to predict the antigenicity of a given protein, which is entirely based on the physiochemical properties of amino acids.

2.3 T cell epitope prediction

T cell epitope was predicted by tools available in Immune Epitope Database (IEDB) [24] which provides a catalog of experimentally characterized T-cell epitopes, as well as data on Major Histocompatibility Complex (MHC) binding and MHC ligand elution experiments [24]. T cell epitopes were identified by NetCTL server [25] of IEDB. NetCTL brings out T cell epitopes by combining predictions of proteasomal cleavage, TAP transport efficiency, and MHC class I affinity. This integrated algorithm provides overall scores for epitope prediction. The threshold for epitope identification was set to 0.50 at which the sensitivity and specificity were 0.89 and 0.94 respectively. Weight on C terminal cleavage was set to 0.15 and weight on TAP transport efficiency was set to 0.05 for all the available supertypes. The epitopes above the threshold value were tabulated. All the peptides above the threshold value were predicted to be probable

epitopes. According to selected parameter settings, NetCTL server [25] identified potential T cell epitopes, but only 11 epitopes were chosen. The results generated from all the supertypes were analyzed and the epitopes were selected based on high combinatorial scores.

2.4 Interaction with MHCI and MHCII

MHC class I alleles that interact with the selected epitopes were determined by MHC-I binding prediction server [26] of IEDB. The nonamers were given as input in MHC class l binding prediction tool available in the Immune Epitope Database (IEDB) server. Prediction Method was chosen to be IEDB recommended and MHC source species was chosen to be human. Furthermore all the available MHC class I alleles were selected and the peptide lengths were set to 9 amino-acids prior to prediction. The peptides which interacted with the highest number of alleles were selected.

Similarly, the interaction was also determined for MHC-II using the IEDB MHC class II binding prediction tool [27]. The whole protein sequences were submitted since MHC class II can accommodate much longer peptides – possibly even whole proteins. The Stabilized Matrix Base Method (SMM) was used to calculate IC50 values of peptide binding to MHC II molecules. The peptides (containing 15 amino acid residues) that interacted with the highest number of alleles were again selected. The SMM-align method was employed to find out good binders and the cut-off value of IC50 was set at 100 nM. The overlapping epitopes between MHC I and MHC II binding predictions which interacted with the highest number of alleles were finally selected for further analysis.

2.5 Analysis of population coverage by the predicted epitopic peptides

In addition to predicting T cell epitopes, it is also vital to estimate population coverage is to become a worthy vaccine candidate. The predicted peptides should effectively cover the human population. To find out the population coverage of the individual epitopes, predicted epitopic sequences with the corresponding Class l HLA alleles were submitted to the population coverage analysis tool of IEDB [28] by maintaining the default analysis parameters and the population of the world was selected to get an overall idea about how well the predicted epitopes cover human population. This server employs the most comprehensive database allele frequencies.

2.6 Analysis of conservancy by the predicted epitopic peptides

The epitopes were subjected to analyze for comparing conservancy retrieved from different countries of the world using the IEDB Conservancy Analysis tool [29].

2.7 Analysis of allergenicity by the predicted epitopic peptides

Further, it is necessary to find out whether these proposed epitopes showed any sort of allergenicity. AllerTOP v. 2.0 [30] was used to determine the allergenicity of the selected T cell epitopes. AllerTOP is the first alignment-free server for *in silico* prediction of allergens based on the main physicochemical properties of proteins. In comparison to other servers for allergen prediction, AllerTOP outperforms them with 94% sensitivity.

2.8 Analysis of toxicity of the predicted T-cell epitopes

The convenience of manufacture, high specificity, and high penetration of peptides make the epitope-based peptide vaccine approach promising. However, the toxicity of peptides often obstructs the success of peptide-based therapy. An ideal epitope should have no or less toxicity, while having high antigenicity [31]. Therefore, ToxinPred web server was used [32] to determine the toxicity of the selected T-cell epitopes.

2.9 Prediction of the 3D structure of conserved T cell epitopes and selected HLA-C 12*03

For the docking simulation assay, 3D structures both of the peptides and allele are essential. The 3D structures of the selected peptides were generated using the PEP-FOLD Peptide Structure Prediction server [33-35]. PEP-FOLD uses a de novo approach to predict peptide structure from amino acids, ranging from 9 to 36 residues. This method is based on a Hidden Markov Model derived Structural Alphabet (SA) letter coupled the predicted series of SA letters to a greedy algorithm and a coarse-grained force field [35]. SA letters describe conformations of four consecutive residues. The best models provided by the server were chosen for the docking study. HLA-C 12*03 was found to interact with most of the predicted T cell epitopes. For this reason, the 3-D model of HLA-C 12*03 was generated using SWISS MODEL server [36]. The homology-modeling method includes the following four steps: (i) template selection; (ii) target

template alignment; (iii) model building; (iv) evaluation [37]. These steps can be iteratively repeated, until a satisfying model structure is obtained. The HLA-C 12*03 3D model was evaluated by PROCHECK software [38] and ProSA web tool [39]. Ramachandran plot [38], constructed using the PROCHECK software assesses the stereochemical quality of 3-D structure analyzing residue-by-residue geometry and overall structure geometry and ProSA tool provides a Z-score which is a measurement for the quality of the model [39].

2.10 Docking of the best selected peptides in the binding groove of HLA allele

Docking analysis was performed to investigate the interaction of the T cell epitopes with the corresponding MHC class I molecule, HLA-C 12*03. Computer-simulated ligand docking is a powerful technique for evaluating relative binding affinity of the ligand towards its receptor [40]. PyRx was employed for the docking purpose [41]. PyRx is a valuable tool for docking of protein to a ligand. PyRx is an open source software to perform virtual screening. It comprises a docking wizard with an easy-to-use user interface. It is a combination of many software such as AutoDockVina, AutoDock 4.2, Mayavi, Open Babel, etc. PyRx uses Vina and AutoDock 4.2 as the docking software. The docking generated PDBQT files which were further visualized in PyMOL[42] molecular Graphics system.

2.11 Validation of Workflow

Since the workflow used here included various computational tools developed by different platforms, so the validation of the workflow was required. The analysis for validating the strategy was repeated on the well-known West Nile Virus for the identification of CD8+ T-cell epitopes. This analysis supported the approach of prediction used in this study and gave reliability for considering the predicted peptides as potential candidate epitope. Six T-cell epitopes were identified through wet lab from the polyprotein precursor sequence of West Nile Virus [43]. The protein sequence of the polyprotein was retrieved from the NCBI database and fed into the workflow to check for whether the workflow can identify and qualify those as T-cell epitopes.

CHAPTER 03: RESULTS

RESULTS

3.1 EspG2 is conserved in most pathogenic *Mycobacterium tuberculosis* strains and is antigenic

The EspG2 protein sequence was of a length of 276 amino acids. To predict the degree of conservation, Multiple Sequence Alignment by Clustal Omega [21] and protein variability analysis were performed. From the Multiple Sequence Alignment, the EspG2 protein was found to be well conserved in all of the 34 strains except one. KBF74136.1 has Aspartic acid whereas all the other strains have Glycine in position 109. The absolute variability computed by VarDBServer [22] revealed 275 fully conserved nucleotides, which comprise more than 97% of the length of the EspG2 protein. Evaluation of EspG2 protein sequence having accession number ALB21149.1 by VaxiJen server [23] identified it as a probable antigen with a value of 0.4366. The threshold for the antigenicity of the bacterium was 0.4.

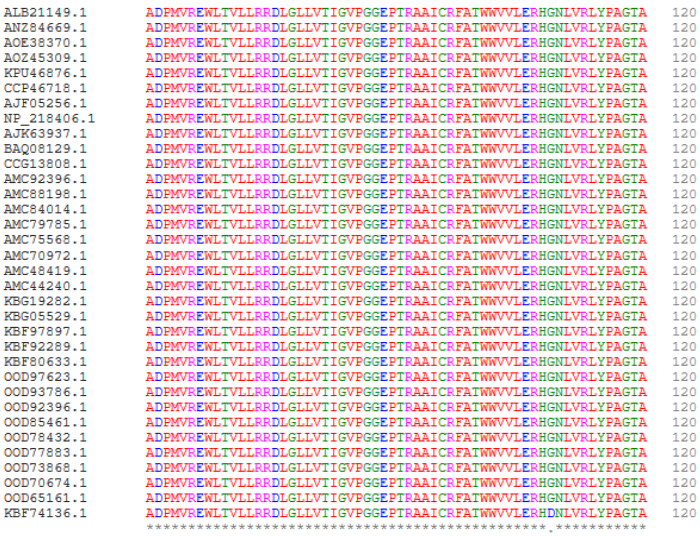


Figure 3.1 Multiple Sequence Alignment by Clustal Omega for the 34 strains of *Mycobacterium tuberculosis*

3.2 Prediction of T cell epitopes

The NetCTL server [25] predicts the different epitopes in Espg2 protein but only eleven most potential peptides were selected based on their high combinatorial score for further analysis. All the probable epitopes for each Supertype predicted are tabulated below. In the case of Supertype A1 seventeen probable epitopes are predicted (Table 3.1), for Supertype A2 eighteen (Table 3.2), A3 eleven (Table 3.3), A24 two (Table 3.4), A26 thirteen (Table 3.5), B7 nineteen (Table 3.6),

B8 twenty two (Table 3.7), B27 twenty one (Table 3.8), B39 seventeen (Table 3.9), B44 sixteen (Table 3.10), B58 eleven (Table 3.11) and B62 fifteen (Table 3.12). Amongst all these probable epitopes the ones with the highest overall score are selected. Epitope "MVREWLTVL" was predicted for both B7 and B8 so while shortlisting the best epitopes one additional epitope was considered. The summary Table 3.13 tabulates the best eleven epitopes in descending order according to their overall score along with their corresponding Supertype.

Table 3.1 Predicted epitopes for Supertype A1.

Number	Epitope	Position	Total score
1	TVDADELLH	150-158	0.7964
2	SVTSGQRRY	233-241	0.7942
3	RLPAGDEWY	262-270	0.7578
4	PAGDEWYSY	264-272	0.7434
5	LTTTVDGLW	2-10	0.7362
6	IVDTAAGRI	221-229	0.7250
7	TTTVDGLWV	3-11	0.6423
8	DTAERALRH	35-43	0.6197
9	VLERHGNLV	104-112	0.6018
10	VTVDADELL	149-157	0.5855
11	RLDVDQLQM	174-182	0.5774
12	TTVDGLWVL	4-12	0.5632
13	DADELLHAV	152-160	0.5569
14	ATWWVVLER	99-107	0.5412
15	TVDGLWVLQ	5-13	0.5126
16	RHGNLVRLY	107-115	0.5064
17	RDAGTLRSY	167-175	0.5046

Table 3.2 Predicted epitopes for Supertype A2

Number	Epitope	Position	Total score
1	YLLSQRLDV	169-177	1.0857
2	MLTTTVDGL	1-9	0.9998
3	TLVALQAGV	196-204	0.9910
4	WVLQAVTGV	10-18	0.9815
5	VLLRRDLGL	71-79	0.9560
6	GLRPLLPRL	26-34	0.9212
7	TTVDGLWVL	4-12	0.8607
8	LVGDSTVAT	213-221	0.7864
9	GQVERLCGV	132-140	0.7428
10	MVREWLTVL	64-72	0.7388
11	RILVGDSTV	211-219	0.7268
12	QLQMVTMAA	179-187	0.6500
13	RLDVDQLQM	174-182	0.6342
14	VTVDADELL	149-157	0.6330
15	ILVGDSTVA	212-220	0.6129
16	LVVGQVERL	129-137	0.5897
17	RLDTAERAL	33-41	0.5223
18	SAHATLVAL	192-200	0.5091

Table 3.3 Predicted epitopes for Supertype A3

Number	Epitope	Position	Total score
1	ATWWVVLER	99-107	1.0341
2	LQAGVGPEK	200-208	0.9239
3	GVAEAAPLR	139-147	0.7501
4	SVTSGQRRY	233-241	0.7477
5	AVRDAGTLR	159-167	0.6320
6	RLPAGDEWY	262-270	0.6249
7	RLYPAGTAS	113-121	0.5839
8	TLRSYLLSQ	165-173	0.5621
9	REWLTVLLR	66-74	0.5405
10	RSAHATLVA	191-199	0.5366
11	RSYLLSQRL	167-175	0.5243

Table 3.4 Predicted epitopes for Supertype A24

Number	Epitope	Position	Total score
1	RFATWWVVL	97-105	1.3758
2	EWLTVLLRR	67-75	0.7108

Table 3.5 Predicted epitopes for Supertype A26

Number	Epitope	Position	Total score
1	TTVDGLWVL	4-12	1.5456
2	SVTSGQRRY	233-241	1.2567
3	DTAERALRH	35-43	0.8891
4	VTVDADELL	149-157	0.7359
5	DTAAGRICV	223-231	0.6573
6	RLPAGDEWY	262-270	0.6481
7	LVVGQVERL	129-137	0.6442
8	DIGGAVQRL	251-259	0.6277
9	WVLQAVTGV	10-18	0.6060
10	MVREWLTVL	64-72	0.5581
11	RDAGTLRSY	161-169	0.5343
12	TLVALQAGV	196-204	0.5241
13	PAGDEWYSY	264-272	0.5062

Table 3.6 Predicted epitopes for Supertype B7

Number	Epitope	Position	Total score
1	CPELGLRPL	22-30	1.4771
2	MVREWLTVL	64-72	1.4495
3	GPEKSARIL	205-213	1.3803
4	SAHATLVAL	192-200	1.1456
5	RPLLPRLDT	28-36	1.0071
6	LVGDSTVAI	213-221	0.8666
7	RAAICRFAT	92-100	0.8538
8	HPVAAELMA	43-51	0.8066
9	LPRLDTAER	31-39	0.8034
10	MAADPTRSA	185-193	0.6867
11	HAVRDAGTL	158-166	0.6597
12	AAPLRPVTV	143-151	0.6335
13	RALRHPVAA	39-47	0.6238
14	AVQRLIRRL	255-263	0.6153
15	RLDTAERAL	33-41	0.5664
16	VVLERHGNL	103-111	0.5521
17	DPTRSAHAT	188-196	0.5460
18	AERALRHPV	37-45	0.5356
19	VPGGEPTRA	85-93	0.5010

Table 3.7 Predicted epitopes for Supertype B8

Number	Epitope	Position	Total score
1	SGQRRYQVL	236-244	1.7979
2	MVREWLTVL	64-72	1.4997
3	VLLRRDLGL	71-79	1.3231
4	SAHATLVAL	192-200	0.8974
5	RFATWWVVL	97-105	0.8332
6	EWYSYRRVV	268-276	0.7967
7	VVLERHGNL	103-111	0.7824
8	VDQLQMVTM	177-185	0.7192
9	QLQMVTMAA	179-187	0.7172
10	RALRHPVAA	39-47	0.7143
11	LLRRDLGLL	72-80	0.7069
12	RAAICRFAT	92-100	0.6672
13	LRHPVAAEL	41-49	0.6530
14	PTRSAHATL	189-197	0.6522
15	RSYLLSQRL	167-175	0.5803
16	ERALRHPVA	38-46	0.5645
17	VREWLTVLL	65-73	0.5627
18	RLDTAERAL	33-41	0.5343
19	CPELGLRPL	22-30	0.5329
20	AGTLRSYLL	163-171	0.5076
21	HAVRDAGTL	158-166	0.5066
22	AVQRLIRRL	255-263	0.5020

Table 3.8 Predicted epitopes for Supertype B27

Number	Epitope	Position	Total score
1	RRYQVLSPG	239-247	1.2866
2	QRLIRRLPA	257-265	1.1717
3	ERHGNLVRL	106-114	1.0275
4	VREWLTVYLL	65-73	0.9833
5	REWLTVLLR	66-74	0.9333
6	LRHPVAAEL	41-49	0.8879
7	RRLPAGDEW	261-269	0.8439
8	QRRYQVLSP	238-246	0.8040
9	LRSYLLSQR	166-174	0.7657
10	YQVLSGSR	241-249	0.7595
11	RRDLGLLVT	74-82	0.7413
12	RDAGTLRSY	161-169	0.7278
13	RHGNLVRLY	107-115	0.7013
14	LQAGVGPEK	200-208	0.6938
15	SRSDIGGAV	248-256	0.6759
16	RSYLLSQRL	167-175	0.6665
17	CRFATWWVV	96-104	0.6579
18	LRRDLGLLV	73-81	0.6574
19	GRICVESVT	227-235	0.6126
20	QRLDVDQLQ	173-181	0.5867
21	TRSAHATLV	190-198	0.5513

Table 3.9 Predicted epitopes for Supertype B39

Number	Epitope	Position	Total score
1	RFATWWVVL	97-105	1.2838
2	VREWLTVLL	65-73	1.1769
3	TTVDGLWVL	4-12	1.0691
4	ERHGNLVRL	106-114	1.0644
5	LRHPVAAEL	41-49	0.8490
6	HAVRDAGTL	158-166	0.7411
7	SAHATLVAL	192-200	0.7260
8	SGQRRYQVL	236-244	0.7029
9	RLDTAERAL	33-41	0.5724
10	GPEKSARIL	205-213	0.5620
11	RSYLLSQRL	167-175	0.5605
12	SQRLDVDQL	172-180	0.5266
13	EQTCPELGL	19-27	0.5234
14	GVEQTCEL	17-25	0.5222
15	SDEAGAGEL	121-129	0.5179
16	ARLMAVGAL	47-55	0.5166
17	MVREWLTVL	64-72	0.5088

Table 3.10 Predicted epitopes for Supertype B44

Number	Epitope	Position	Total score
1	AELMAVGAL	47-55	1.6524
2	PELGLRPLL	23-31	1.5043
3	AERALRHPV	37-45	1.2997
4	AEAAPLRPV	141-149	0.9248
5	DEAGAGELV	122-130	0.8869
6	PEKSARILV	206-214	0.8248
7	SDEAGAGEL	121-129	0.7860
8	DEWYSYRRV	267-275	0.7825
9	RDLGLLVTI	75-83	0.6977
10	EQTCPELGL	19-27	0.6713
11	ADPMVREWL	61-69	0.6235
12	SQRLDVDQL	172-180	0.5553
13	RSYLLSQRL	167-175	0.5335
14	REWLTVLLR	66-74	0.5217
15	GQVERLCGV	132-140	0.5158
16	LERHGNLVR	105-113	0.5025

Table 3.11 Predicted epitopes for Supertype B58

Number	**Epitope**	**Position**	**Total score**
1	LTTTVDGLW	2-10	1.5383
2	NADPMVREW	60-68	1.2459
3	AAICRFATW	93-101	1.1467
4	RSYLLSQRL	167-175	1.0610
5	RRLPAGDEW	261-269	0.9625
6	RSAHATLVA	191-199	0.7679
7	TTVDGLWVL	4-12	0.7280
8	RLDVDQLQM	174-182	0.7208
9	AICRFATWW	94-102	0.6345
10	VTVDADELL	149-157	0.5894
11	HAVRDAGTL	158-166	0.5347

Table 3.12 Predicted epitopes for Supertype B62

Number	**Epitope**	**Position**	**Total score**
1	MVREWLTVL	64-72	1.0789
2	RLAGDEWY	262-270	1.0750
3	RDAGTLRSY	161-169	0.9512
4	SVTSGQRRY	233-241	0.8329
5	RFATWWVVL	97-105	0.7475
6	SQRLDVDQL	172-180	0.7152
7	SAHATLVAL	192-200	0.6990
8	LQAVTGVEQ	12-20	0.6523
9	DQAGNADPM	56-64	0.6335
10	RSYLLSQRL	167-175	0.6326
11	LQAGVPEK	200-208	0.6251
12	VLLRRDLGL	71-79	0.5541
13	RLYPAGTAS	113-121	0.5449
14	VVLERHGNL	103-111	0.5315
15	PAGDEWYSY	264-272	0.5084

Table 3.13 Epitopes with the highest scores

Number	Epitope	Position	Score	Supertype
1	SGQRRYQVL	236-244	1.7979	B8
2	AELMAVGAL	47-55	1.6524	B44
3	TTVDGLWVL	4-12	1.5456	A26
4	LTTTVDGLW	2-10	1.5383	B58
5	PELGLRPLL	23-31	1.5043	B44
6	MVREWLTVL	64-72	1.4997	B8
7	CPELGLRPL	22-30	1.4771	B7
8	MVREWLTVL	64-72	1.4495	B7
9	GPEKSARIL	205-213	1.3803	B7
10	RFATWWVVL	97-105	1.3758	A24
11	VLLRRDLGL	71-79	1.3231	B8

3.3 MHC class I and class II epitope identification

A good epitope should also interact with multiple MHC alleles. By employing MHC-I binding prediction tool [26], those MHC-I alleles for which the epitopes showed higher affinity (IC50 < 100nm) was determined. The total score of each epitope – HLA interaction was considered and higher the score meant higher the processing efficiency. This method retrieved the interacting MHC alleles along with the IC50 values, rank, and score for each of the previously selected epitopes (Table 3.14). Similarly MHC-II binding prediction tool [27] was used to further predict allele interaction with the eleven epitopes (Table 3.15). Among the total peptides, only those peptides which interacted with the maximum number of MHC class I alleles were found to be the core sequences of 15mer MHC class II alleles, were selected for further analysis. Amongst the eleven epitopes, MVREWLTVL was predicted twice so there are in fact just ten candidate epitopes. Table 3.16 summarizes both MHC I and MHC II allele interaction for the ten candidate epitope.

Table 3.14 Predicted T-cell epitopes along with their interacting MHC-I alleles

Epitope	Allele	Ic50	Rank	Score
SGQRRYQVL	HLA-C*03:03	29.35013	21	
	HLA-B*15:02	32.97843	1.6	
	HLA-C*12:03	48.32924	72	
	HLA-C*07:02	71.97307	2.2	
	HLA-C*14:02	120.3982	27	
	HLA-B*14:02	165.2152	0.3	
	HLA-B*08:01	187.7371	0.7	1.7979
AELMAVGAL	HLA-C*03:03	11.31384	5.7	
	HLA-B*40:01	25.11308	0.2	
	HLA-B*15:02	35.17467	1.8	
	HLA-B*40:02	49.18809	0.3	
	HLA-C*12:03	152.1283	100	1.6524
TTVDGLWVL	HLA-C*03:03	21.1646	15	
	HLA-C*15:02	44.60874	2.6	
	HLA-A*02:06	50.06682	1.7	
	HLA-B*15:02	53.60805	3.7	
	HLA-C*12:03	65.49528	85	
	HLA-A*68:02	126.0637	1.6	1.5456
LTTTVDGLW	HLA-B*58:01	10.19295	0.2	
	HLA-B*57:01	37.05271	0.2	
	HLA-C*12:03	43.37206	67	
	HLA-C*03:03	79.179	47	1.5383
PELGLRPLL	HLA-C*03:03	28.74816	21	
	HLA-B*15:02	105.9815	8.3	
	HLA-C*12:03	159.6651	100	1.5043
MVREWLTVL	HLA-B*15:02	41.99813	2.7	
	HLA-C*12:03	62.98106	83	
	HLA-A*02:06	141.4328	4.4	
	HLA-B*07:02	167.4519	0.9	
	HLA-C*15:02	189.4175	12	
	HLA-C*07:01	194.6122	9.2	
	HLA-C*06:02	196.2502	2.7	1.4997
CPELGLRPL	HLA-C*03:03	36.44265	26	
	HLA-B*15:02	53.60805	3.7	
	HLA-C*12:03	61.12375	82	
	HLA-B*07:02	75.14325	0.5	1.4771
MVREWLTVL	HLA-B*15:02	41.99813	2.7	
	HLA-C*12:03	62.98106	83	
	HLA-A*02:06	141.4328	4.4	
	HLA-B*07:02	167.4519	0.9	
	HLA-C*15:02	189.4175	12	
	HLA-C*07:01	194.6122	9.2	
	HLA-C*06:02	196.2502	2.7	1.4495
GPEKSARIL	HLA-C*03:03	51.71422	35	
	HLA-C*12:03	78.02073	92	
	HLA-B*15:02	80.58051	6.2	1.3803
FATWWVVL	HLA-C*03:03	30.31168	22	
	HLA-C*12:03	113.8178	98	
	HLA-B*15:02	144.9539	12	1.3758
VLLRRDLGL	HLA-C*03:03	51.2401	35	
	HLA-B*15:02	73.32127	5.7	
	HLA-C*14:02	86.82004	18	
	HLA-B*08:01	131.3863	0.5	
	HLA-C*07:02	191.942	12	
	HLA-C*12:03	195.5285	100	1.3231

Table 3.15 Predicted T-cell epitopes along with their interacting MHC-II alleles

Epitopes	allele	start	end	core_peptide	peptide	ic50	rank
SGQRRYQVL	HLA-DRB5*01:01	235	249	TSGQRRYQV	TSGQRRYQVLSPGSR	23	0.07
	HLA-DRB5*01:01	236	250	YQVLSPGSR	SGQRRYQVLSPGSRS	23	0.07
	HLA-DRB1*01:01	236	250	YQVLSPGSR	SGQRRYQVLSPGSRS	53	10.36
	HLA-DRB1*01:01	235	249	RYQVLSPGS	TSGQRRYQVLSPGSR	55	10.7
AELMAVGAL	HLA-DQA1*05:01/DQB1*03:01	46	60	AVGALDQAG	AAELMAVGALDQAGN	79	17.48
	HLA-DQA1*05:01/DQB1*03:01	47	61	AVGALDQAG	AELMAVGALDQAGNA	79	17.48
	HLA-DQA1*05:01/DQB1*03:01	45	59	VAAELMAVG	VAAELMAVGALDQAG	83	17.78
	HLA-DQA1*05:01/DQB1*03:01	41	55	PVAAELMAV	LRHPVAAELMAVGAL	84	17.85
	HLA-DQA1*05:01/DQB1*03:01	42	56	PVAAELMAV	RHPVAAELMAVGALD	84	17.85
	HLA-DRB1*04:04	46	60	LMAVGALDQ	AAELMAVGALDQAGN	97	0.83
TTVDGLWVL	HLA-DRB1*01:01	4	18	LWVLQAVTG	TTVDGLWVLQAVTGV	51	10
	HLA-DRB1*04:04	4	18	LWVLQAVTG	TTVDGLWVLQAVTGV	94	0.8
	HLA-DRB1*04:04	3	17	VDGLWVLQA	TTTVDGLWVLQAVTG	97	0.83
LTTTVDGLW	N/A	N/A	N/A	N/A	N/A	N/A	N/A
PELGLRPLL	HLA-DRB1*01:01	22	36	LRPLLPRLD	CPELGLRPLLPRLDT	88	15.99
	HLA-DRB1*01:01	23	37	LRPLLPRLD	PELGLRPLLPRLDTA	89	16.14
	HLA-DRB1*01:01	21	35	ELGLRPLLP	TCPELGLRPLLPRLD	90	16.28
MVREWLTVL	HLA-DRB1*11:01	64	78	LTVLLRRDL	MVREWLTVLLRRDLG	15	0.01
	HLA-DRB1*11:01	63	77	PMVREWLTV	PMVREWLTVLLRRDL	16	0.01
CPELGLRPL	HLA-DRB1*01:01	22	36	LRPLLPRLD	CPELGLRPLLPRLDT	88	15.99
	HLA-DRB1*01:01	21	35	ELGLRPLLP	TCPELGLRPLLPRLD	90	16.28
MVREWLTVL	HLA-DRB1*11:01	64	78	LTVLLRRDL	MVREWLTVLLRRDLG	15	0.01
	HLA-DRB1*11:01	63	77	PMVREWLTV	PMVREWLTVLLRRDL	16	0.01
GPEKSARIL	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RFATWWVVL	HLA-DRB1*11:01	97	111	WVVLERHGN	RFATWWVVLERHGNL	69	0.04
	HLA-DRB1*11:01	96	110	WWVVLERHG	CRFATWWVVLERHGN	74	0.05

Table 3.16 Shortlisted T-cell overlapping epitopes between MHC I and MHC II binding predictions

Epitope	MHC I allele	No. of MHC I allele	peptide	MHCII	MHC II allele	No. of MHC II allele
SGQRRYQVL	HLA-C*03:03, HLA-B*15:02, HLA-C*12:03, HLA-C*07:02, HLA-C*14:02, HLA-B*14:02, HLA-B*08:01	7	TSGQRRYQVLSPGSR SGQRRYQVLSPGSRS SGQRRYQVLSPGSRS TSGQRRYQVLSPGSR	HLA-DRB5*01:01 HLA-DRB5*01:01 HLA-DRB1*01:01 HLA-DRB1*01:01	HLA-DRB5*01:01 HLA-DRB1*01:01	2
AELMAVGAL	HLA-C*03:03, HLA-B*40:01, HLA-B*15:02, HLA-B*40:02, HLA-C*12:03	5	AAELMAVGALDQAGN AELMAVGALDQAGNA VAAELMAVGALDQAG LRHPVAAELMAVGAL RHPVAAELMAVGALD AAELMAVGALDQAGN	HLA-DQA1*05:01/DQB1*03:01 HLA-DQA1*05:01/DQB1*03:01 HLA-DQA1*05:01/DQB1*03:01 HLA-DQA1*05:01/DQB1*03:01 HLA-DQA1*05:01/DQB1*03:01 HLA-DRB1*04:04	HLA-DQA1*05:01/ DQB1*03:01 HLA-DRB1*04:04	2
TTVDGLWVL	HLA-C*03:03, HLA-C*15:02, HLA-A*02:06, HLA-B*15:02, HLA-C*12:03, HLA-A*68:02	6	TTVDGLWVLQAVTGV TTVDGLWVLQAVTGV TTTVDGLWVLQAVTG	HLA-DRB1*01:01 HLA-DRB1*04:04 HLA-DRB1*04:04	HLA-DRB1*01:01 HLA-DRB1*04:04	2
PELGLRPLL	HLA-C*03:03, HLA-B*15:02, HLA-C*12:03	3	CPELGLRPLLPRLDT PELGLRPLLPRLDTA TCPELGLRPLLPRLD	HLA-DRB1*01:01 HLA-DRB1*01:01 HLA-DRB1*01:01	HLA-DRB1*01:01	1
MVREWLTVL	HLA-B*15:02, HLA-C*12:03, HLA-A*02:06, HLA-B*07:02, HLA-C*15:02, HLA-C*07:01	6	MVREWLTVLLRRDLG PMVREWLTVLLRRDL	HLA-DRB1*11:01 HLA-DRB1*11:01	HLA-DRB1*11:01	1
CPELGLRPL	HLA-C*03:03, HLA-B*15:02, HLA-C*12:03, HLA-B*07:02	4	CPELGLRPLLPRLDT TCPELGLRPLLPRLD	HLA-DRB1*01:01 HLA-DRB1*01:01	HLA-DRB1*01:01	1
RFATWWVVL	HLA-C*03:03, HLA-C*12:03, HLA-B*15:02	3	RFATWWVVLERHGNL CRFATWWVVLERHGN	HLA-DRB1*11:01 HLA-DRB1*11:01	HLA-DRB1*11:01	1
VLLRRDLGL	HLA-C*03:03, HLA-B*15:02, HLA-C*14:02, HLA-B*08:01, HLA-C*07:02	5	REWLTVLLRRDLGLL VREWLTVLLRRDLGL EWLTVLLRRDLGLLV VLLRRDLGLLVTIGV WLTVLLRRDLGLLVT LTVLLRRDLGLLVTI EWLTVLLRRDLGLLV LTVLLRRDLGLLVTI TVLLRRDLGLLVTIG WLTVLLRRDLGLLVT LTVLLRRDLGLLVTI WLTVLLRRDLGLLVT EWLTVLLRRDLGLLV TVLLRRDLGLLVTIG VLLRRDLGLLVTIGV LTVLLRRDLGLLVTI WLTVLLRRDLGLLVT EWLTVLLRRDLGLLV TVLLRRDLGLLVTIG VLLRRDLGLLVTIGV VLLRRDLGLLVTIGV	HLA-DRB1*11:01 HLA-DRB1*11:01 HLA-DRB1*11:01 HLA-DRB1*01:01 HLA-DRB1*11:01 HLA-DRB1*11:01 HLA-DRB1*01:01 HLA-DRB1*01:01 HLA-DRB1*01:01 HLA-DRB1*01:01 HLA-DRB1*03:01 HLA-DRB1*03:01 HLA-DRB1*03:01 HLA-DRB1*03:01 HLA-DRB1*03:01 HLA-DRB1*13:02 HLA-DRB1*13:02 HLA-DRB1*13:02 HLA-DRB1*13:02 HLA-DRB1*13:02 HLA-DRB1*07:01	HLA-DRB1*11:01 HLA-DRB1*01:01 HLA-DRB1*03:01 HLA-DRB1*13:02 HLA-DRB1*07:01	5

3.4 Allergenicity analysis

AllerTOP v. 2.0 [30] was used to predict the allergenicity of the ten T-cell epitopes. All except two of the epitopes were found to be non-allergens. The non-allergens are SGQRRYQVL, AELMAVGAL, TTVDGLWVL, PELGLRPLL, MVREWLTVL, CPELGLRPL, GPEKSARIL and RFATWWVVL. The allergens were LTTTVDGLW and VLLRRDLGL and were hence excluded from further consideration.

3.5 Toxicity assessment

The toxicity of the selected ten T-cell epitopes was analyzed by the ToxinPred server [32] and the information was tabulated in Table 3.17. All of the selected T-cell epitopes were found to be non-toxic to cell proving their potential as candidate vaccines.

Table 3.17 Toxicity of candidate epitopes

Epitope	SVM score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Molecular weight
SGQRRYQVL	-1.02	Non-toxin	-0.43	-1.17	0.12	2	1106.38
AELMAVGAL	-1.07	Non-toxin	0.24	1.69	-0.54	-1	874.19
TTVDGLWVL	-0.76	Non-toxin	0.18	1.09	-0.87	-1	1003.30
LTTTVDGLW	-0.78	Non-toxin	0.10	0.54	-0.74	-1	1005.27
PELGLRPLL	-1.07	Non-toxin	-0.03	0.40	-0.13	0	1007.38
MVREWLTVL	-0.68	Non-toxin	0.02	0.92	-0.63	0	1146.55
CPELGLRPL	-0.92	Non-toxin	-0.08	0.26	-0.04	0	997.35
GPEKSARIL	-1.44	Non-toxin	-0.24	-0.51	0.58	1	970.26
RFATWWVVL	-1.09	Non-toxin	0.14	1.09	-1.33	1	1177.54
VLLRRDLGL	-1.44	Non-toxin	-0.16	0.72	0.03	1	1054.44

3.6 Epitope conservancy

The conservancy analyses revealed that all the nonamers were 100% conserved among all the 34 (Table 3.18)

.

Table 3.18 Conservancy analysis of the candidate epitopes

Epitope sequence	Percent of protein sequence matches at identity <= 100%	Minimum identity	Maximum identity
SGQRRYQVL	100.00% (34/34)	100.00%	100.00%
AELMAVGAL	100.00% (34/34)	100.00%	100.00%
TTVDGLWVL	100.00% (34/34)	100.00%	100.00%
LTTTVDGLW	100.00% (34/34)	100.00%	100.00%
PELGLRPLL	100.00% (34/34)	100.00%	100.00%
MVREWLTVL	100.00% (34/34)	100.00%	100.00%
CPELGLRPL	100.00% (34/34)	100.00%	100.00%
GPEKSARIL	100.00% (34/34)	100.00%	100.00%
RFATWWVVL	100.00% (34/34)	100.00%	100.00%
VLLRRDLGL	100.00% (34/34)	100.00%	100.00%

3.7 Population Coverage

Over a thousand different human MHC alleles are known and different HLA types are expressed at different frequencies in different ethnicities. Population coverage by the probable epitopes varied between 20 % – 65 % when both MHC-I and MHC-II alleles were considered. The epitopes which interacted with MHC alleles with large world population coverage were selected as the most probable epitopes for vaccine design. The population coverage of the following epitopes SGQRRYQVL, AELMAVGAL, TTVDGLWVL, PELGLRPLL, MVREWLTVL, CPELGLRPL, RFATWWVVL are 54.97%, 32.4%, 38.92%, 29.23%, 49.64%, 38.39% and 28.44% respectively. Top five candidate epitopes, SGQRRYQVL, MVREWLTV, TTVDGLWVL, CPELGLRPL, and AELMAVGAL were chosen after the exclusion of the epitopes which showed allergenicity and interacted with the least number of MHC allele. Table 3.19 shows all the information gathered at this stage of the study.

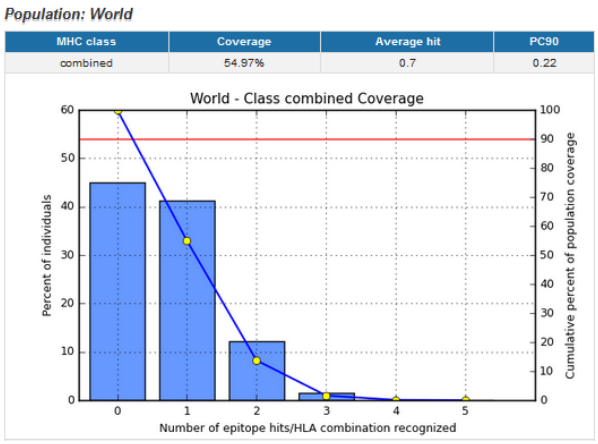


Figure 3.2 A Population coverage analysis for SGQRRYQVL based on the HLA interaction.

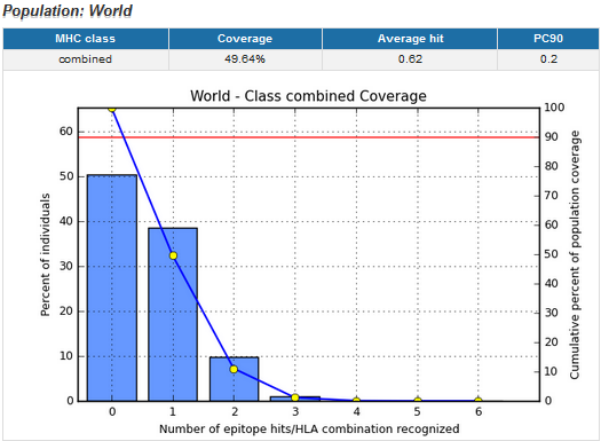


Figure 3.2 B Population coverage analysis for MVREWLTV based on the HLA interaction.

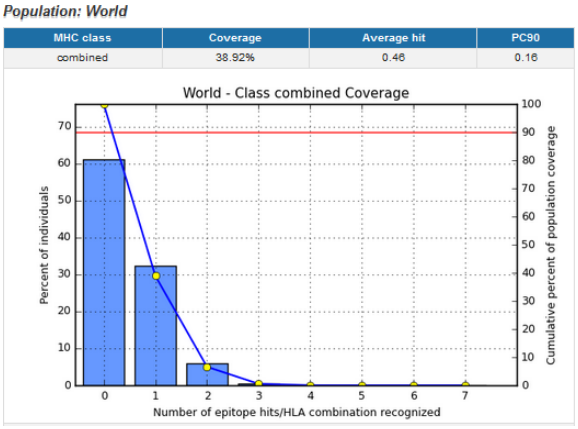


Figure 3.2 C Population coverage analysis for TTVDGLWVL based on the HLA interaction.

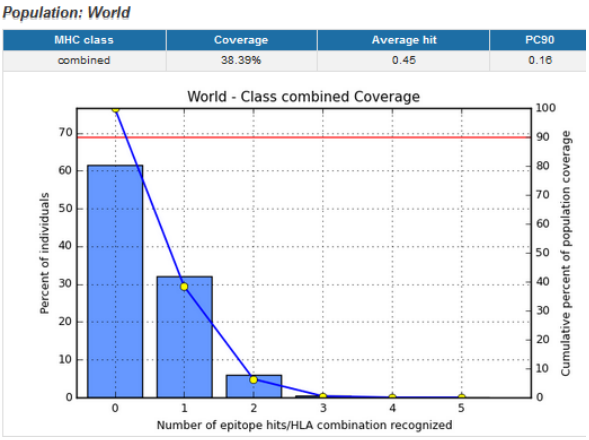


Figure 3.2 D Population coverage analysis for CPELGLRPL based on the HLA interaction.

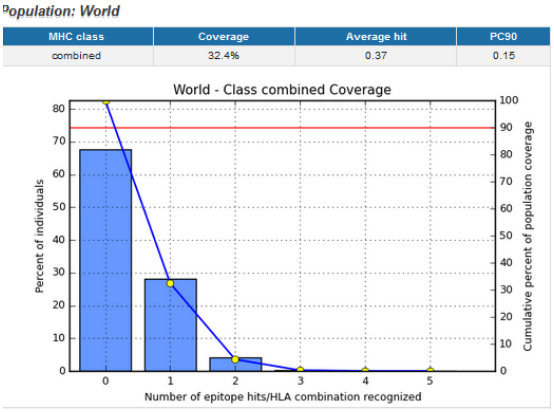


Figure 3.2 E Population coverage analysis for AELMAVGAL based on the HLA interaction.

Figure 3.2 Population coverage of the combined prediction for both of the MHC I and MHC II alleles. Here, the number 1 bar for all the analyses represents out-predicted epitope. Notes: in the graphs, the line (-o-) represents the cumulative percentage of population coverage of the epitopes; the bars represent the population coverage for each epitope.

Table 3.19 Top five epitopes

Epitopes	Population coverage	MHCI	No. of alleles	MHCII	No. of alleles
SGQRRYQVL	54.97	HLA-C*03:03, HLA-B*15:02, HLA-C*12:03, HLA-C*07:02, HLA-C*14:02, HLA-B*14:02, HLA-B*08:01	7	HLA-DRB5*01:01 HLA-DRB1*01:01	2
MVREWLTVL	49.64	HLA-B*15:02, HLA-C*12:03, HLA-A*02:06, HLA-B*07:02, HLA-C*15:02, HLA-C*07:01	6	HLA-DRB1*11:01	1
TTVDGLWVL	38.92	HLA-C*03:03, HLA-C*15:02, HLA-A*02:06, HLA-B*15:02, HLA-C*12:03, HLA-A*68:02	6	HLA-DRB1*01:01 HLA-DRB1*04:04	2
CPELGLRPL	38.39	HLA-C*03:03, HLA-B*15:02, HLA-C*12:03, HLA-B*07:02	4	HLA-DRB1*01:01	1
AELMAVGAL	32.4	HLA-C*03:03, HLA-B*40:01, HLA-B*15:02, HLA-B*40:02, HLA-C*12:03	5	HLA-DQA1*05:01/DQB1*03:01, HLA-DRB1*04:04	2

3.8 3D structures of the predicted epitope peptides and HLA-C 12*03 allele were predicted and validated

For the molecular docking simulation assay, the 3-D structure of all the epitopes and MHC molecule are required. PEP-FOLD Peptide Structure Prediction server [33,34,35] was used for generating 3-D structure of five T cell epitopes SGQRRYQVL,MVREWLTVL, TTVDGLWVL, CPELGLRPL and AELMAVGAL (Figure 3.3- 3.7) As all the epitopes were predicted to interact with HLA-C 12*03 allele, the 3-D structure of HLA-C 12*03 allele (Figure 3.7) was generated by homology modeling [36]. Numerous models were generated by SWISS MODEL server [37]. Based on the GQME and QMEAN4 scores of 0.74 and 1.02 respectively, the best model was selected for which the template was 5w69.3.A. The target and template sequences showed 98.19%

sequence identity. The model was further validated using the Ramachandran plot and Z-score. Ramachandran plot generated by Procheck (Figure 3.8) software showed that 98% residues were in the favorable region. To check whether the input structure is within the range of scores typically found for a native protein of similar size, ProSAz-score was calculated [39]. The Z-score which indicates the overall model quality was -9.27 (Figure 3.9). This score established the quality of the generated model.

Figure 3.3 A Cartoon model showed in rainbow colors to show the different amino acids of SGQRRYQVL

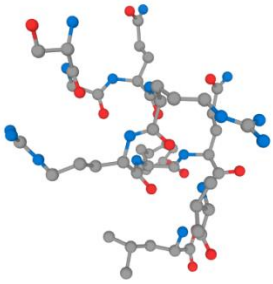


Figure 3.3 B Balls and sticks model with uniform color to show different atoms of SGQRRYQVL

Figure 3.3 3D model of SGQRRYQVL

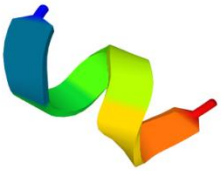


Figure 3.4 A Cartoon model showed in rainbow colors to show the different amino acids of MVREWLTVL

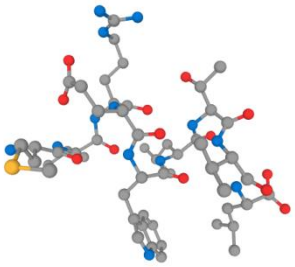


Figure 3.4 B Balls and sticks model with uniform color to show different atoms of MVREWLTVL

Figure 3.4 3D model of MVREWLTVL

Figure 3.5 A Cartoon model showed in rainbow colors to show the different amino acids of TTVDGLWVL

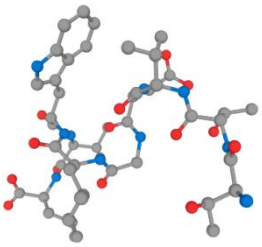


Figure 3.5 B Balls and sticks model with uniform color to show different atoms of TTVDGLWVL

Figure 3.5 3D model of TTVDGLWVL

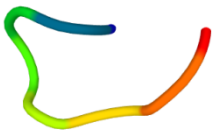


Figure 3.6 A Cartoon model showed in rainbow colors to show the different amino acids of CPELGLRPL

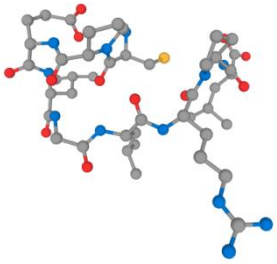


Figure 3.6 B Balls and sticks model with uniform color to show different atoms of CPELGLRPL

Figure 3.6 3D model of CPELGLRPL

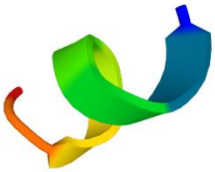


Figure 3.7 A Cartoon model showed in rainbow colors to show the different amino acids of AELMAVGAL

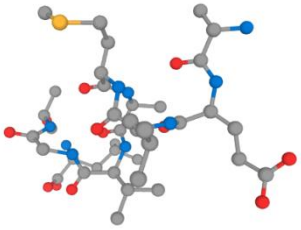


Figure 3.7 B Balls and sticks model with uniform color to show different atoms of AELMAVGAL

Figure 3.7 3D model of AELMAVGAL

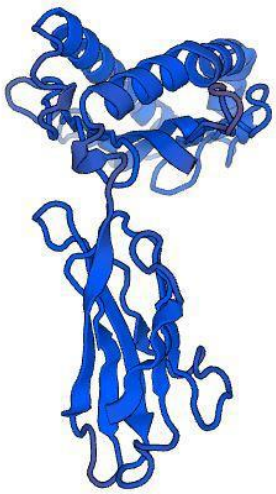
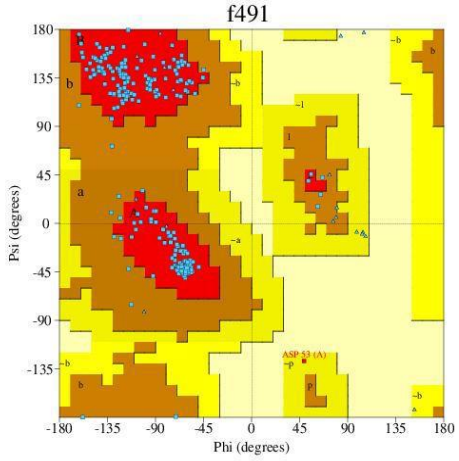


Figure 3.8 Predicted 3D structure of the HLA-C 12*03 allele



2. G-Factors

Parameter	Score	Average Score
Dihedral angles:-		
Phi-psi distribution	-0.07	
Chi1-chi2 distribution	0.01	
Chi1 only	-0.11	
Chi3 & chi4	0.43	
Omega	-0.57*	
		-0.14
Main-chain covalent forces:-		
Main-chain bond lengths	0.29	
Main-chain bond angles	-0.06	
		0.09
OVERALL AVERAGE		-0.04

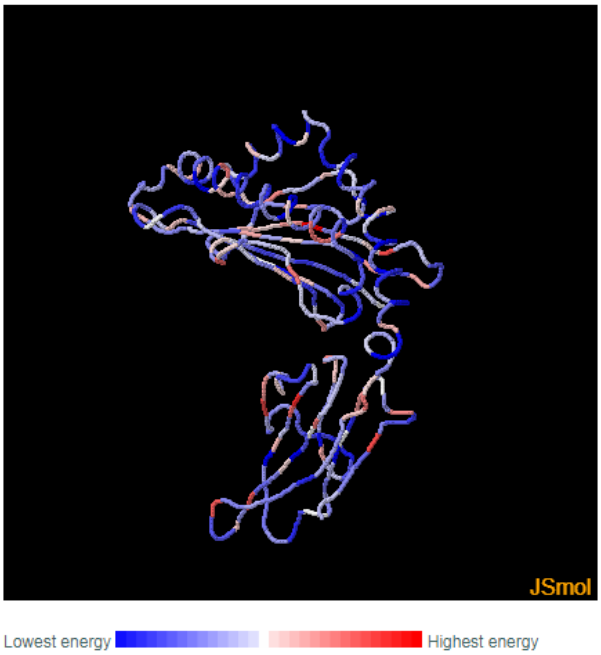
G-factors provide a measure of how **unusual**, or out-of-the-ordinary, a property is

Values below -0.5* - unusual
Values below **-1.0**** - highly unusual

1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	221	93.2%
Additional allowed regions	[a,b,l,p]	15	6.3%
Generously allowed regions	[~a,~b,~l,~p]	1	0.4%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		237	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		20	
Proline residues		14	
Total number of residues		273	

Figure 3.9 Ramachandran plot of HLA-C 12*03 along with statistics showing residues in the most favorable and disallowed regions and the G-factor for the model



3.10 A 3D image of HLA-C 12*03 showing different energy regions

Evaluation of residues

```
Residue [A  53 :ASP] (  49.16,-127.73) in Allowed region
Residue [A 138 :ASP] (-161.83, 109.67) in Allowed region
Residue [A 147 :TYR] (-112.44, -75.36) in Allowed region
Residue [A 186 :GLY] (-100.64, -82.33) in Allowed region
Number of residues in favoured region    (~98.0% expected)  :  267 ( 98.5%)
Number of residues in allowed region     ( ~2.0% expected)  :    4 (  1.5%)
Number of residues in outlier region                        :    0 (  0.0%)
```

Figure 3.10 B Evaluation of residues

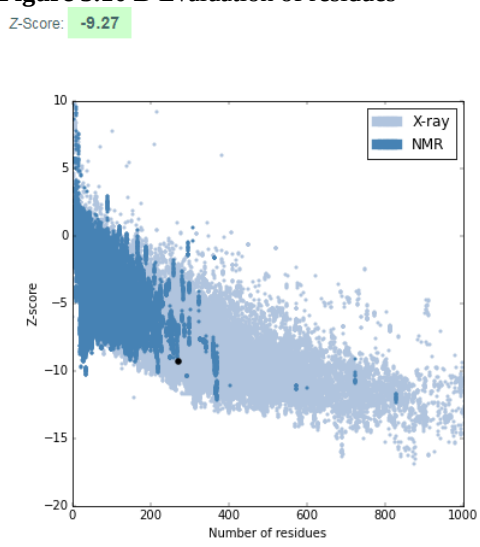


Figure 3.10 C Z-score for quality of the 3D structure HLA-C 12*03

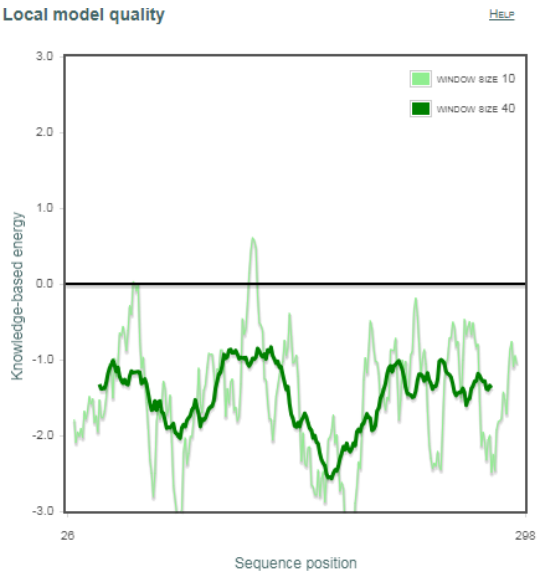


Figure 3.10 D Local model quality

Figure 3.10 Results of HLA-C 12*03 obtained from ProSA

3.9 Docking

At first binding models for all the top five T cell epitopes with the HLA-C 12*03 were generated using the AutoDOCKVina tool in PyRx [41]. Free energy of binding was estimated by AutodockVina tool according to the following equation:

Free energy of binding= Intermolenergy+Internalenergy+Torsional energy-Unbound energy

In docking analysis, intermolecular forces included Van der Waals forces, Hydrogen bonds, solvation and electrostatic energy. Epitope "SGQRRYQVL" bound to the binding groove of HLA-B*12:03 with the binding energy -11.5 kcal/mol and epitope "MVREWLTVL" with the binding energy -10.8kcal/mol, epitope "TTVDGLWVL" with a binding energy of -10.3 kcal/mol, epitope "CPELGLRPL" with a binding energy of -10.6 kcal/mol and epitope "AELMAVGAL" with a binding energy of -9.7kcal/mol. On the other hand, after setting same parameters, the binding energy of the control peptide KVITFIDLto the binding grooves of class I MHC allele-H2KB was estimated to be -10.2 kcal/mol(Figure 3.16). As lower binding energy favors the formation of stable interaction, we can expect that the candidate epitope will interact with MHC-I molecules *in vivo* readily. Binding models of the best probable epitopes to its specific HLA molecules were observed using Pymol [42] which are illustrated in Figure 3.11, Figure 3.12, Figure 3.13, Figure 3.14 and Figure 3.15.

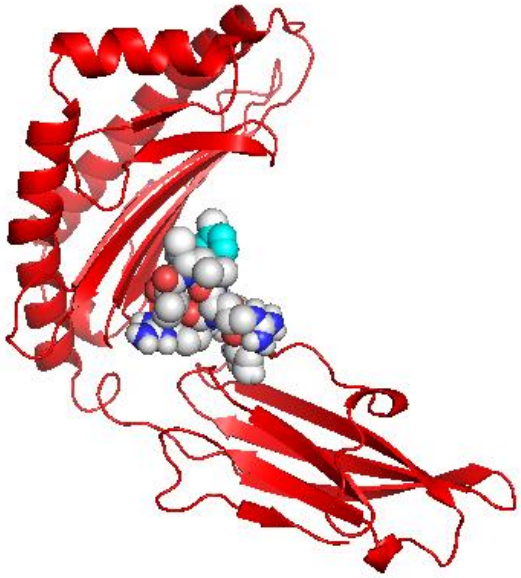


Figure 3.11 A Representing the cartoon view.

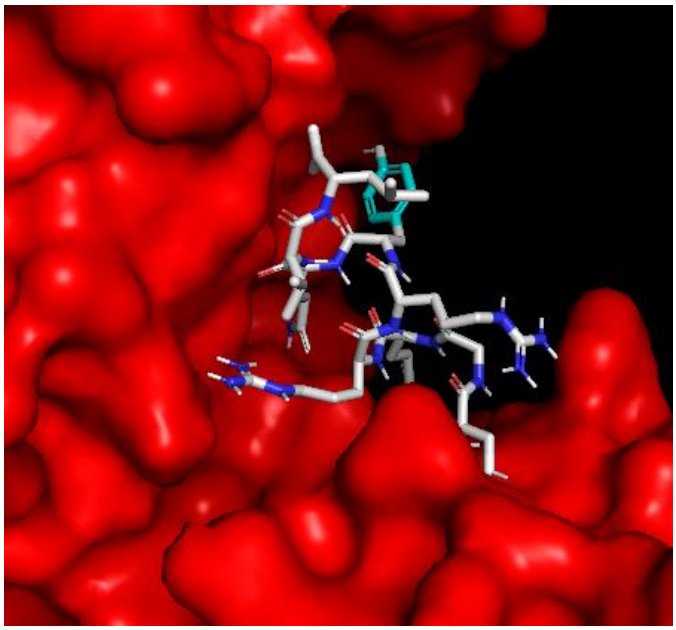
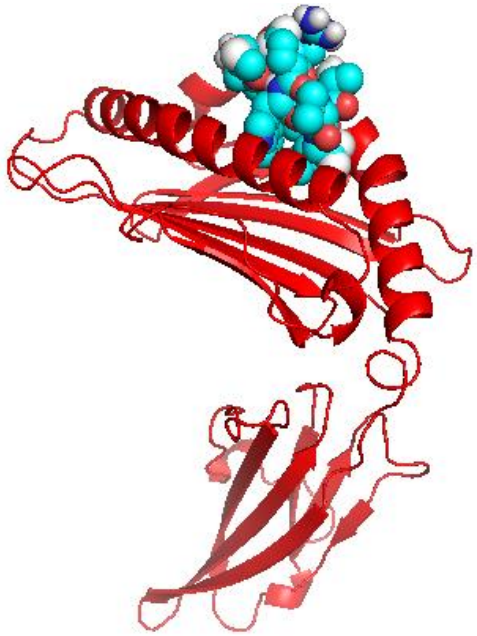


Figure 3.11 B Illustration of interaction

Figure 3.11 Docking simulation assay of SGQRRYQVL to the HLA-C 12*03 allele



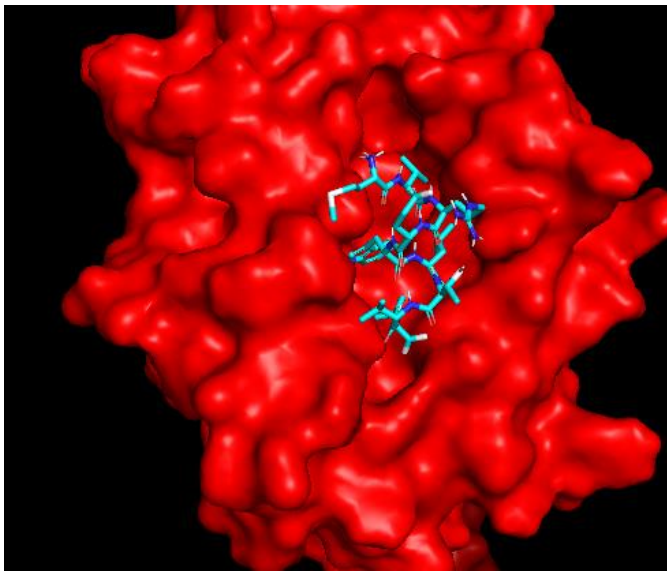
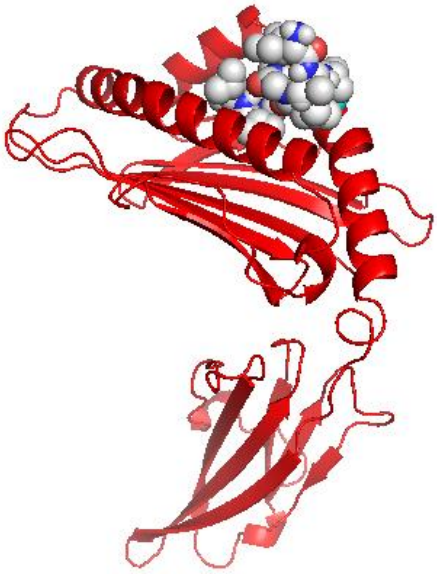


Figure 3.12 A Representing the cartoon view. **Figure 3.12 B** Illustration of interaction

Figure 3.12 Docking simulation assay of MVREWLTVL to the HLA-C 12*03 allele



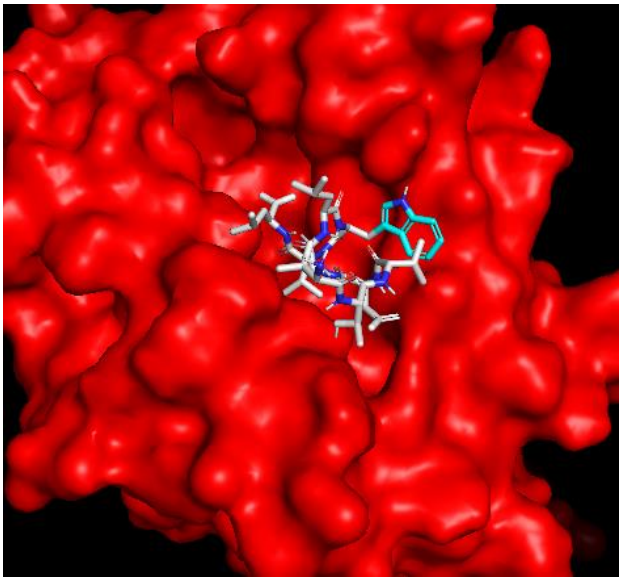


Figure 3.13 A Representing the cartoon view. **Figure 3.13 B** Illustration of interaction

Figure 3.13 Docking simulation assay of TTVDGLWVL to the HLA-C 12*03 allele

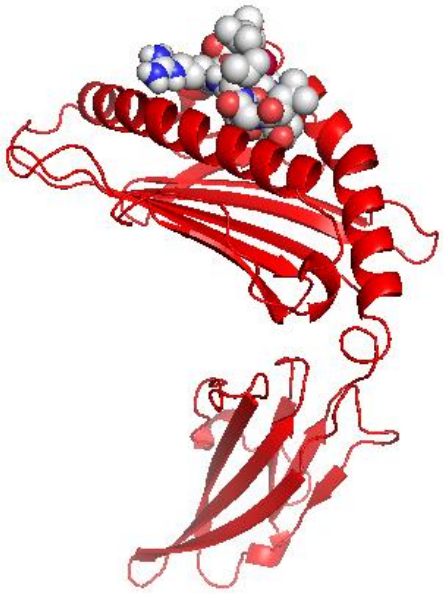


Figure 3.14 A Representing the cartoon view.

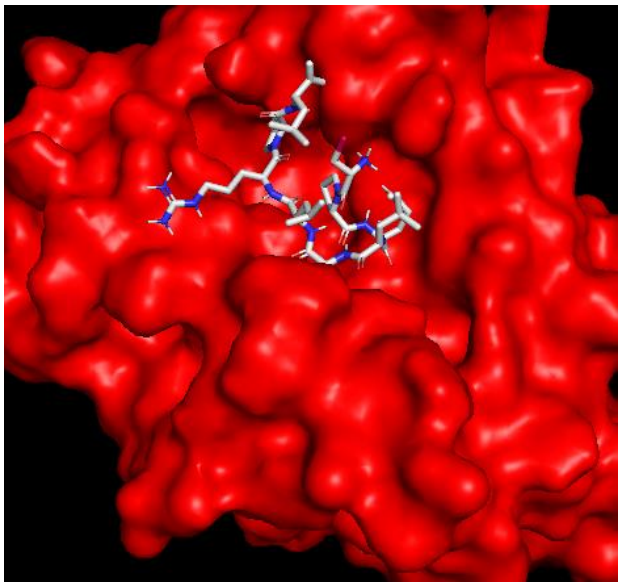


Figure 3.14 B Illustration of interaction

Figure 3.14 Docking simulation assay of CPELGLRPL to the HLA-C 12*03 allele

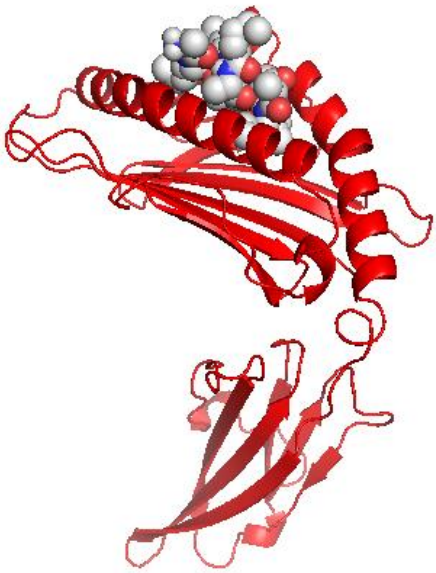


Figure 3.15 A Representing the cartoon view.

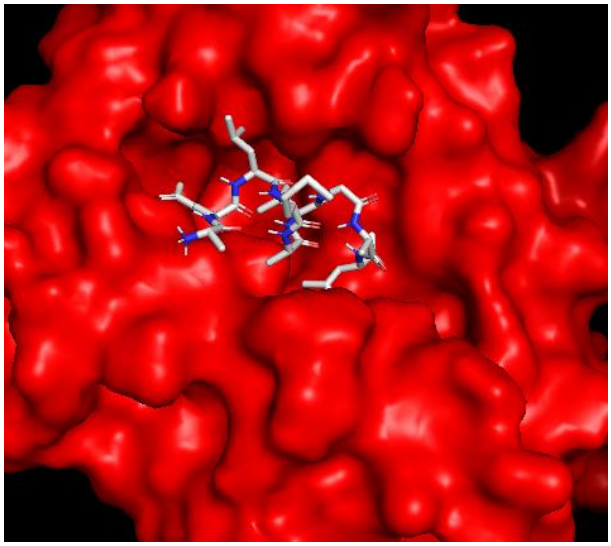


Figure 3.15 B Illustration of interaction

Figure 3.15 Docking simulation assay of AELMAVGAL to the HLA-C 12*03 allele

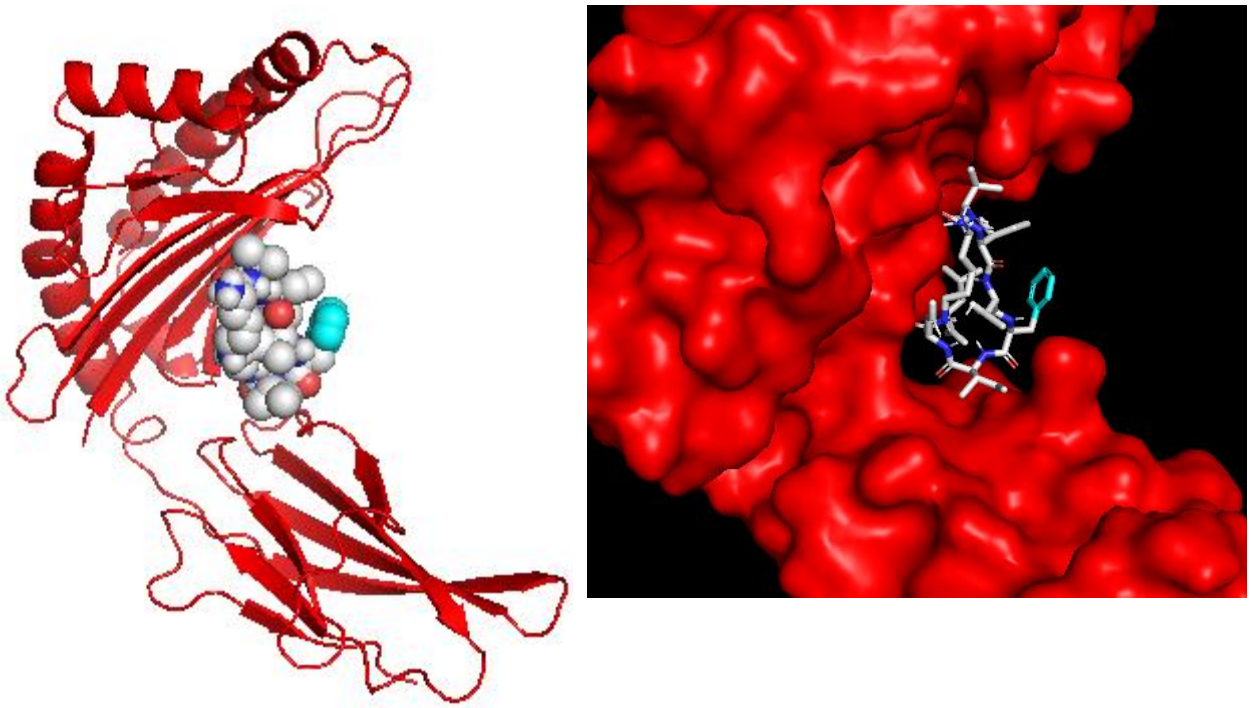


Figure 3.16 A Representing the cartoon view. **Figure 3.16 B** Illustration of interaction

Figure 3.16 Docking simulation assay of the control peptide KVITFIDL to the H2KB allele

3.9 Designed workflow synchronizes with experimental results

The workflow, designed to predict epitopes, used here was validated by positive controls. Six linear T-cell epitopes, SVGGVFTSV, RLDDDGNFQL, YTMDGEYRL, SLFGQRIEV, SLTSINVQA, and ATWAENIQV were mapped in the polyprotein sequence of West Nile virus NY99 [45]. When the sequence of the polyprotein was fed into the workflow, it identified all T-cell epitopes except RLDDDGNFQL.

CHAPTER 04: DISCUSSION

DISCUSSION

Tuberculosis (TB) is one of the top 10 causes of death worldwide at present. This disease is caused by the bacterium *Mycobacterium tuberculosis*, which can form lesions that lead to cell death in any tissue or organ. *M. tuberculosis* establishes a latent chronic infection that reactivates when there are diminished immune responses. For example in aged people, individuals with genetic immune defects, medically immune-suppressed patients and patients treated with antibodies against tumor necrosis factor alpha. At present BCG, an attenuated strain of *Mycobacterium bovis* is used as a vaccine against tuberculosis. However, the overall effectiveness of BCG is debatable as it has some serious limitations. BCG has failed to protect against TB in adults with pulmonary TB [6]. BCG is unable to prevent chronic infection or protect against pulmonary tuberculosis in adults. Immunosuppression caused by HIV is now a particularly significant factor in the reactivation of tuberculosis. In addition to that, individuals treated with BCG respond positively to a common tuberculosis diagnostic test, which makes it difficult to distinguish between individuals infected with *M. tuberculosis* and those inoculated with BCG cells [7].

Vaccines made from whole cells of organisms or parts of them have many drawbacks. The safety of this form of vaccination is one of the major concerns as it may cause autoimmune or strong allergic responses. Attenuation or inactivation of such vaccines might not be effective as the pathogen may revert back to its virulent state. Subunit vaccines being more controllable than whole cell vaccines are still not perfectly safe, and cause side effects and production difficulties similar to whole pathogen strategies which include manufacturing difficulties and poor or undesired immune response [46]. To overcome these issues of the currently available vaccine, an alternative is vital. It is necessary to design a vaccine that is safer, more immunogenic and induces longer lasting protection.

Therefore, only using synthetic epitope-based peptide vaccines which consist of the minimal antigenic epitopes to induce the desired immune response appears to be the most sensible and safest technique to develop vaccines. Peptide-based vaccines can serve as excellent alternatives to traditional vaccination approaches because of the ease with which chemical modifications can be introduced. Peptide vaccines avoid the inclusion of lipopolysaccharides, lipids, and toxins which possess high reactogenicity to the host [47] Epitopes lack infectious potential and chemical stability which eliminates the risk of reactivation of the pathogen to its virulent state. Besides they are able to provide long term protection against the pathogen which makes them suitable vaccine candidates. In addition to that they are comparatively easier to construct and

produce [48]. Computational determination of conserved epitope for the manufacture of peptide vaccines is not only less time consuming, but is also very economic [44]. They have the potential to not only render protection against the disease but also behave as the therapeutic tool to treat them.

An epitope, also known as antigenic determinant, is that part of an antigen which is recognized by the immune system, specifically by antibodies. The epitope is the specific part of the antigen to which an antibody binds [49]. Fully synthetic epitope-based peptide vaccines are the potential future of vaccine development. The development of this category of the vaccine of peptide-based immunogens is already taking place [46].

Though sequences from pathogens provide several potential vaccine candidates, it can be assumed that only one in 100 to 200 peptides bind to a particular MHC in real [50]. Therefore, a good computational resolving method could significantly narrow down the number of peptides that have to be manufactured and tested. With the creation of various bioinformatics tools and the availability of massive sequence data, epitope-based peptide vaccine design against highly conserved antigenic protein has become very convenient [51, 52]. The ground behind the epitope-based vaccine is the chemical synthesis of identified T-cell epitopes, which are immunodominant and can trigger specific immune responses [40]. Vaccine development is a long, complex process involving a combination of public and private involvement. With the advancement of sequence based technology and information about the proteomics and genomics of the different microorganisms, vaccine development has been facilitated.

Vaccine development for *Mycobacterium tuberculosis* is based on screening of multiple epitopes having most antigenic properties that direct the immune system to protect human beings from the infection. The purpose of this study was to screen new and highly potential immunogenic epitopes for T-cell because vaccine against T-cell epitopes is much more promising as it evokes a long lasting immune response, and also the antigen can easily escape the memory response of antibody due to antigenic drift. In this study peptide base vaccine has been designed using various bioinformatics tools and vaccinomics approach [53]. This *in silico* approach has already been used to address the development of new vaccines for combating with diseases like Multiple sclerosis [54], Dengue [55], Malaria [56], Influenza [57] and Tumor [58].

Recently vaccine based on T-cell has been encouraged as the host can generate a strong immune response by CD8+Tcell against the infected cell. *In silico* studies has recently received experimental validation, where a multi-epitope cluster secretary protein of *Mycobacterium tuberculosis* (Ag85B) which bind to HLA class-I and class-II molecules. Later their prediction has been experimentally validated *in vitro* [59].

There are many criteria that need to be fulfilled by a vaccine candidate epitope. Vaccine design must take into account both chemical modifications to antigens and heterogeneous length of these important targets of immunity, as well as the challenge of the diversity in human immunogenetics [47].

Prediction of T cell epitope can be done easily from the abundant algorithms those freely available. In this study, IEDB analysis tool [24] which is possibly the most wide-ranging database offering several T cell epitope-related analysis and prediction tools as well as provides both intrinsic biochemical and extrinsic context dependent information about them has been employed.

Firstly, the identification of T-cell epitopes is important and 182 epitopes were identified using the NetCTL server [25]. From those only best eleven epitopes were chosen depending on their total score. MVREWLTVL appeared twice among the eleven candidate epitopes, once for SupertypeB7 and then for Supertype B8. The other epitopes were SGQRRYQVL, AELMAVGAL, TTVDGLWVL, LTTTVDGLW, PELGLRPLL, CPELGLRPL, GPEKSARIL, RFATWWVVL and VLLRRDLGL.

T-cells recognize their particular antigen in terms of major histocompatibility complex (MHC) molecules and the genes which code for these molecules are highly variable within an outbred population of a particular species. Therefore, the identification of T-cell epitopes is crucial to the MHC alleles available in a particular species, an event called MHC restriction [60]. To identify the binding of MHC alleles to each of the selected epitopes, MHC-I prediction server [26] and MHC-II prediction server [27] of IEDB was employed. In this case, those peptides which showed a higher affinity, IC50 <200nm and IC50 <100nm for MHC-I and MHC-II alleles respectively were selected because the immunogenic property of each predicted T cell epitope was characterized by its IC50 value, which indicates the peptide's binding affinity to HLA molecules and the number of corresponding HLA alleles. Peptides with less IC50 values show good inhibition. The overlapping T-cell epitopes between MHC I and MHC II binding predictions which interacted with highest number of alleles were finally selected for further analysis. Furthermore, selected T cell epitopes were subjected to the IEDB conservancy analysis tool and were found to be 100% conserved.

Toxicity of peptides often obstructs the success of peptide-based therapy. An ideal epitope should have no or less toxicity while having high antigenicity [31]. ToxinPred web server was used [32] to discover that all the candidate epitopes were non-cytotoxic.

In vaccine development, allergenicity is a fundamental problem. Nowadays, most vaccines stimulate an allergic reaction as an immune response, through induction of type-2 T helper T (Th2) cells and immunoglobulin E (IgE)[61]. Amongst all the candidate epitopes LTTTVDGLW and VLLRRDLGL were identified as allergens and were excluded for further analysis.

Moreover, it is also essential to estimate population coverage to become a worthy vaccine candidate. The predicted peptides should effectively cover the human population. The population coverage analysis tool of IEDB [28] predicted population of individuals with MHC alleles interacting with the epitopes SGQRRYQVL, AELMAVGAL, TTVDGLWVL, PELGLRPLL, MVREWLTVL, CPELGLRPL, RFATWWVVL to be 54.97%, 32.4%, 38.92%, 29.23%, 49.64%, 38.39% and 28.44% respectively. Then after considering the entire prior described methods and values, five best epitopes were selected for further research. The five most potent epitope candidates were SGQRRYQVL, MVREWLTV, TTVDGLWVL, CPELGLRPL, and AELMAVGAL.

Molecular docking studies are essential to provide sufficient evidence that whether the target is binding to the specific region or not. Ligand docking simulated on a computer is a fast and powerful technique to evaluate the relative binding affinity of the ligand towards its receptor. The validating results obtained from the epitope prediction workflow, which includes the availability for binding, hence validating if the predicted epitope is suitable to induce an immune response. PEPFOLD Peptide Structure Prediction server [33, 34, 35] was used for generating the 3-D structure of all five T cell epitopes SGQRRYQVL, MVREWLTVL, TTVDGLWVL, CPELGLRPL, and AELMAVGAL. As all the epitopes were predicted to interact with HLA-C 12*03 allele, the 3-D structure of the allele was generated by homology modeling in the SWISS MODEL server [36]. All the five epitopes were subjected to computer-simulated molecular docking by AutoDockVina in PyRx [41] to investigate the intermolecular interactions between the epitopes and the HLA-C 12*03 allele. From the outcomes of docking simulation assay, it was found that the binding energy to the binding groove of HLA-B*12:03 for epitope SGQRRYQVL was -11.5 kcal/mol, epitope MVREWLTVL was -10.8kcal/mol, epitope TTVDGLWVL was -10.3 kcal/mol, epitope CPELGLRPL was -10.6 kcal/mol and epitope AELMAVGAL was -9.7kcal/mol. Alternatively, after employing the same parameters, the binding energy of the control peptide KVITFIDL to the binding grooves of class I MHC allele- H2KB was estimated to be -10.2 kcal/mol (Figure 3.15). Lower binding energy favors the formation of stable interaction. Therefore it verifies the binding cleft epitope interaction to HLA molecule when it will be applied *in vivo.*

Taking all other previous data into consideration SGQRRYQVL is the most suitable epitope. It interacts with the maximum number of MHC-I (seven alleles) and MHC-II (two alleles) amongst all the other epitopes predicted in this study. In addition to that, the results of similarity analysis and epitope prediction showed that the prior selected epitope is a non-allergen, non-cytotoxic, conserved and also shows satisfactory population coverage. In SGQRRYQVL "S" stands for serine, "G" stands for glycine, "Q" stands for glutamine, "R" stands for arginine, "Y" stands for tyrosine, "V" stands for valine and "L" stands for leucine. All the predictions were maintained with high specificity, which can be defined by stringent threshold values. It can be suggested that the proposed epitopes would be able to trigger an efficacious immune response as a peptide vaccine *in vivo*. Thus this implies that the predicted candidate epitope can be considered as a broad-spectrum vaccine if developed.

Stability and antigenicity of the peptide vaccine can be further improved by attaching these epitopes to adjuvants because the peptides are very susceptible to enzymatic degradation. Adjuvants act as immune stimulants. Some adjuvants licensed for human use are Alum, MF59, AS03, and virosomes [62]. To support the concerns regarding actual immunogenicity, stability and efficacy it is necessary to also take into consideration delivery method of these epitopes inside human bodies, both *in vitro* and *in vivo* experiments [44].

CONCLUSION

Bioinformatics has emerged as a promising field for predicting epitopes. In this study, an attempt was made to design epitope based vaccines against *Mycobacterium tuberculosis* which elicit T-cell immunity. The mechanism used implies if the bacteria try to attach to the host cell the peptide vaccine, well recognize it and present this information to a broad spectrum of protector cells such as T-cells. The epitope can mimic antigen presentation and thus help the antibody formation inside the host. Bioinformatics was used to filter out an epitope that has the greatest potential to make happen. Several bioinformatics tools were employed to establish the qualities of the epitope, SGQRRYQVL to become an effective vaccine. Thus this analysis finally infers that SGQRRYQVL can help in promoting the immunity against TB. In this way, bioinformatics approach saves both expenditure and time required to screen a large number of epitopes as compared to experimental techniques. In addition to that, it guides the experimental work with high confidence to find the desired results. The results of this study provide computational data for the identification and screening of epitopes and may be used for the development of epitope vaccines with enhanced safety and efficacy.

References

1. Tuberculosis (TB) [Internet]. World Health Organization. 2018 Feb 16. Available from: http://www.who.int/news-room/fact-sheets/detail/tuberculosis

2. Calmette A, Guérin C, Boquet A, Négre L. La vaccination preventive contre la tuberculose par le “BCG". Masson et cie. 1927.

3. Waksman SA. The conquest of tuberculosis. Berkeley: University of California Press; 1964.

4. Kaufmann S, Parida S. Changing funding patterns in tuberculosis. Nat Med [Internet]. 2007; 13: 299–303. Available from: http://dx.doi.org/10.1038/nm0307

5. Floyd K, Glaziou P, Ulekar M. Global tuberculosis control: epidemiology, strategy, financing. Geneva, Switzerland: World Health Organization; 2009. 314.

6. Crampin AC, Glynn JR, Fine PE. What has Karonga taught us? Tuberculosis studied over three decades. Int J Tuberc Lung Dis. 2009; 13:153–164.

7. Glick BR, Delovitch TL, Patten CL. Medical Biotechnology. Washington DC: ASM Press; 2014. 632-663.

8. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T, Churcher, C., Harris, D. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature [Internet]. 1998 Jun 11; 393: 537-544. Available from: http://dx.doi.org/10.1038/31159

9. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl Acad. Sci. USA. 1997; 94 (18): 9869-9874.

10. Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C. Use of a *Mycobacterium tuberculosis* H37Rv Bacterial Artificial Chromosome Library for Genome Mapping, Sequencing, and Comparative Genomics. Infect Immun. 1998; 66(5): 2221-2229.

11. Philipp WJ, Poulet S, Eiglmeier K, Pascopella L, Balasubramanian V, Heym B, et al. An integrated map of the genome of the tubercle bacillus, Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium leprae. PNAS [Internet]. 1996 Apr 2; 93(7): 3132-3137; Available from: https://doi.org/10.1073/pnas.93.7.3132

12. Smith I. *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. Clin Microbiol Rev [Internet]. 2003; 16(3): 463-496. Available from: doi:10.1128/CMR.16.3.463-496.2003.

13. Gey van Pittius NC, Gamieldien J, Hide W, Brown GD, Siezen RJ, Beyers AD. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. Genome Biol. 2001; 2(10): research0044.1-research0044.18.

14. EspG2 - ESX-2 secretion-associated protein EspG2 - Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) - espG2 gene & protein [Internet]. Uniprot.org. 2018. Available from: https://www.uniprot.org/uniprot/P9WJC9.

15. Mycobrowser [Internet]. Ecole polytechnique fédérale de Lausanne. 2018 [cited 3 July 2018]. Available from: https://mycobrowser.epfl.ch/genes/Rv3889c.

16. Ottenhoff TH, Kaufmann SH. Vaccines against Tuberculosis: Where Are We and Where Do We Need to Go? PLoS Pathog. 2012; 8(5): e1002607. Avavilable from: https://doi.org/10.1371/journal.ppat.1002607.

17. World Health Organization. Global tuberculosis report 2017. Geneva. WHO; 2017. 262.

18. Vaccine Types. NIH: National Institute of Allergy and Infectious Diseases [Internet]. 2018 [cited 3 July 2018]. Available from: https://www.niaid.nih.gov/research/vaccine-types

19. Hesseling AC, Marais BJ, Gie RP, Schaaf HS, Fine PE, et al. The risk of disseminated Bacille Calmette-Guerin (BCG) disease in HIV-infected children. 2007; Vaccine 25: 14–18.

20. Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J. and Sayers, E. GenBank. Nucleic Acids Res. 2009; 37: D26-D31.

21. Sievers F, Wilm A, Dineen D, Gibson T, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2014; 7(1): 539-539.

22. Hayes CN, Diez D, Joannin N, Honda W, Kanehisa M, Wahlgren M, et al. VarDB: a pathogen-specific sequence database of protein families involved in antigenic variation. Bioinformatics England. 2008; 24: 2564-5.

23. Doytchinova I, Flower D. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007; 8(1): 4.

24. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. Nucleic Acid Res [Internet]. 2014 Oct 9; pii: gku938. Available from: http://www.iedb.org/

25. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics [Internet]. 2007 Oct 31; 8: 424. Available from: http://www.cbs.dtu.dk/services/NetCTL/

26. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics [Internet]. 2005; 6(1):132. Available from: DOI:10.1186/1471-2105-6-132

27. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. Frontiers in Immunology [Internet]. 2017; 8: 276. Available from: DOI: 10.3389/fimmu.2017.00278

28. Bui H, Sidney J, Dinh K, Southwood S, Newman M J, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics [Internet]. 2006; 7(1): 153. Available from: DOI: 10.1186/1471-2105-7-153.

29. Bui H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics [Internet]. 2007; 8(1): 361. Available from: DOI: 10.1186/1471-2105-8-361

30. Dimitrov I, Flower DR, Doytchinova I. AllerTOP - a server for in silicoprediction of allergens. BMC Bioinformatics [Internet]. 2013;14:S4. DOI:10.1186/1471-2105-14-S6-S4.

31. Vlieghe P, Lisowski V, Martinez J, Khrestchatisky M. Synthetic therapeutic peptides: Science and market. Drug Discov Today [Internet]. 2010 Jan; 15 (1-2): 40-56. doi: 10.1016/j.drudis.2009.10.009.

32. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. In Silico Approach for Predicting Toxicity of Peptides and Proteins. PLoS ONE [Internet]. 2013; 8(9). Available from: DOI: 10.1371/journal.pone.0073957

33. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res [Internet]. 2016 Jul 8; 44(W1):W449-54. Available from: DOI: 10.1093/nar/gkw329.

34. Shen Y, Maupetit J, Derreumaux P, Tufféry P.Improved PEP-FOLD approach for peptide and miniprotein structure prediction. J Chem Theory Comput [Internet]. 2014; 10(10): 4745-4758. Available from: DOI: 10.1021/ct500592m

35. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res [Internet]. 2012; 40: W288-293. Available from: DOI: 10.1093/nar/gks419.

36. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research [Internet]. 2014; 42(W1). Available from: DOI: 10.1093/nar/gku340

37. Schwede T. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. 2003; 31(13): 3381-3385.

38. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993; 26(2): 283-291.

39. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acid Res [Internet]. 2007; 35. Available from: DOI: 10.1093/nar/gkm290.

40. Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open Biol [Internet]. 2013; 3(1), 120139-120139. Available from: DOI: 10.1098/rsob.120139.

41. Dallakyan S, Olson AJ. Small-Molecule Library Screening by Docking with PyRx. Methods Mol Biol [Internet]. 2015; 1263: 243-250. Available from: DOI: 10.1007/978-1-4939-2269-7_19.

42. Seeliger D, Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. J Comput Aided Mol Des [Internet]. 2010 May; 24(5): 417-22. Available from: DOI: 10.1007/s10822-010-9352-6

43. Mcmurtrey CP, Lelic A, Piazza P, Chakrabarti AK, Yablonsky EJ, Wahl A, et al. Epitope discovery in West Nile virus infection: Identification and immune recognition of viral epitopes. Proc Natl Acad Sci U S A. 2008; 105(8): 2981-2986.

44. Parvege MM, Rahman M, Nibir YM, Hossain MS. Two highly similar LAEDDTNAQKT and LTDKIGTEI epitopes in G glycoprotein may be useful for effective epitope based vaccine design against pathogenic Henipavirus. Comput Biol Chem [Internet]. 2016; 61: 270-280. Available from: DOI: 10.1016/j.compbiolchem.2016.03.001.

45. Mcmurtrey CP, Lelic A, Piazza P, Chakrabarti AK, Yablonsky EJ, Wahl A, et al. Epitope discovery in West Nile virus infection: Identification and immune recognition of viral epitopes. Proc Natl Acad Sci U S A [Internet]. 2008; 105(8): 2981-2986. Available from: DOI: 10.1073/pnas.0711874105.

46. Skwarczynski M, Toth I. Peptide-based synthetic vaccines. Chem Sci [Internet]. 2016; 7(2): 842-854. Available from: DOI: 10.1039/c5sc03892h.

47. Purcell AW, Mccluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov [Internet]. 2007; 6(5): 404-414. Available from: http://dx.doi.org/10.1038/nrd2224

48. Huang J, Honda W. CED: A conformational epitope database. BMC Immunology.2006; 7(1): 7. Available from: https://doi.org/10.1186/1471-2172-7-7.

49. Li W, Joshi MD, Singhania S, Ramsey KH, Murthy AK. Peptide Vaccine: Progress and Challenges. Vaccines [Internet]. 2014; 2(3): 515-536. Available from: DOI: 10.3390/vaccines2030515.

50. Yewdell JW, Bennink JR. Immunodominance In Major Histocompatibility Complex Class I–Restricted T Lymphocyte Responses. Annu Rev Immunol [Internet]. 1999; 17(1), 51-88. Available from: DOI: 10.1146/annurev.immunol.17.1.51.

51. Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. PLoS Comput Biol [Internet]. 2006; 2(6): e71 Available from: https://doi.org/10.1371/journal.pcbi.0020071.

52. De Groot AS, Bosma A, Chinai N, Frost J, Jesdale BM, et al. From genometo vaccine: in silico predictions, ex vivo verification. Vaccine. 2001; 19(31): 4385–4395.

53. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. Pharmacogenomics [Internet]. 2009; 10(5):837–52. Available from: DOI: 10.2217/pgs.09.25.

54. Bourdette DN, Edmonds E, Smith C, Bowen JD, Guttmann CRG, Nagy ZP, et al. A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. Mult Scler [Internet]. 2005; 11(5): 552–61. Available from: DOI: 10.1191/1352458505ms1225oa.

55. Tambunan USF. In silico analysis of envelope dengue virus-2 envelope dengue Virus-3 Protein as the backbone of dengue virus tetravalent vaccine by using homology modeling method. OnLine Journal for Biological Sciences [Internet]. 2009; 9(1): 6–16. Available from: DOI : 10.3844/ojbsci.2009.6.16

56. Lopez JA, Weilenman C, Audran R, Roggero MA, Bonelo A, Tiercy JM, et al. A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immuneresponse in humans. Implications for vaccination strategies. Eur J Immunol [Internet]. 2001; 31(7): 1989–98. Available from: DOI: 10.1002/1521-4141(200107)31:7<1989::AID-IMMU1989>3.0.CO;2-M.

57. Shahsavandi S, Ebrahimi MM, Sadeghi K, Mahravani H. Design of a heterosubtypic epitope-based peptide vaccine fused with hemokinin-1 against influenza viruses. Virol Sin [Internet]. 2015; 30(3):200–7. Available from: https://doi.org/10.1007/s12250-014-3504-0

58. Knutson KL, Schiffman K, Disis ML. Immunization with a HER-2/neu helper peptidevaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. J Clin Invest [Internet]. 2001; 107(4) :477–84 . Available from: DOI: 10.1172/JCI11752

59. Khan MK, Zaman S, Chakraborty S, Chakravorty R, Alam MM, Bhuiyan TR, et al. In silico predicted mycobacterial epitope elicits in vitro T-cell responses. Mol Immunol [Internet]. 2014 Sep; 61(1):16-22. Available from: DOI: 10.1016/j.molimm.2014.04.009.

60. Gerner W, Hammer SE, Wiesmuller K, Saalmuller A. Identification of Major Histocompatibility Complex Restriction and Anchor Residues of Foot-and-Mouth Disease Virus-Derived Bovine T-Cell Epitopes. J Virol [Internet]. 2009; 83(9), 4039-4050. Available from: DOI: 10.1128/JVI.01534-08.

61. Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R. From ZikV genome to vaccine: *in silico* approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. Immunology [Internet]. 2016; 149(4): 386-399. Available from: DOI: 10.1111/imm.12656.

62. Lee S, Nguyen MT. Recent Advances of Vaccine Adjuvants for Infectious Diseases. Immune Netw [Internet]. 2015; 15(2), 51-57. Available from: DOI: 10.4110/in.2015.15.2.51.