THE DEVELOPMENT OF PEPTIDE-BASED VACCINE FOR LEPTOSPIROSIS: *IN SILICO* APPROACH

SALWANA BINTI OTHMAN

INSTITUTE OF GRADUATE STUDIES

UNIVERSITY OF MALAYA

KUALA LUMPUR

2014

THE DEVELOPMENT OF PEPTIDE-BASED VACCINE FOR LEPTOSPIROSIS: *IN SILICO* APPROACH

SALWANA BINTI OTHMAN
(SGJ130005)

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF BIOINFORMATICS

INSTITUTE OF GRADUATE STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2014

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: SALWANA BINTI OTHMAN (I.C/Passport No: 870522-03-5382)

Registration/Matric No: SGJ130005

Name of Degree: MASTERS IN BIOINFORMATICS

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

THE DEVELOPMENT OF PEPTIDE-BASED VACCINE FOR LEPTOSPIROSIS: *IN SILICO* APPROACH

Field of Study: BIOINFORMATICS

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;

(2) This Work is original;

(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;

(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;

(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

Leptospirosis is a worldwide zoonosis infection caused by Gram-negative bacteria from the genus of *Leptospira*. Eight pathogenic species of *Leptospira* have been identified and there are more than 250 serovars that can cause leptospirosis to human and animals. Due to the indistinguishable symptoms of the disease, leptospirosis is often misdiagnosed, resulting to death if antibiotic treatment is not given immediately. Prophylactic vaccine can reduce the mortality rate resulting from the late treatment. However, the existing vaccines for leptospirosis are serovar-specific and can only provide limited protection to human. Therefore, a new and more effective vaccine is urgently needed. The vaccine can be developed in shorter time with reduced cost through rational design and *in silico* approach by maximizing the use of the abundant information on the sequences of leptospiral antigenic proteins and the growing number of bioinformatics tools. The objective of this research is to perform an *in silico* analysis to develop a vaccine from the peptide sequence of leptospiral outer membrane protein, LigB. The predicted peptide-based vaccine, which requires experimental validation in the future, is hoped to be effective against all pathogenic serovars of *Leptospira* and ultimately provide us immunity to leptospirosis.

ABSTRAK

Leptospirosis, yang merupakan jangkitan zoonosis di seluruh dunia, disebabkan oleh bakteria Gram-negatif daripada genus *Leptospira*. Lapan spesies *Leptospira* yang bersifat patogenik telah dikenal pasti dan terdapat lebih daripada 250 serovars yang boleh menyebabkan leptospirosis kepada manusia dan haiwan. Oleh sebab gejala penyakit tersebut sukar untuk dibezakan, kesilapan dalam diagnosis penyakit ini seringkali berlaku. Hal ini akan membawa maut, jika rawatan antibiotik tidak diberikan dengan segera. Vaksin yang merupakan agen pencegahan boleh mengurangkan kadar kematian yang disebabkan oleh leptospirosis yang lewat dikesan. Namun, vaksin yang sedia ada untuk leptospirosis hanya berkesan untuk satu serovar sahaja dan hanya dapat memberi perlindungan yang terhad untuk manusia. Oleh itu, vaksin baharu yang lebih berkesan amat diperlukan. Vaksin ini boleh direka dalam masa yang lebih singkat dengan kos yang rendah dengan menggunakan kaedah 'rational designing' dan '*in silico*' dengan memanfaatkan maklumat peptida dalam protein *Leptospira sp.* yang bersifat antigenik dan menggunakan bantuan perisian bioinformatik. Objektif kajian ini adalah untuk menjalankan analisis '*in silico*' untuk mencipta vaksin dari urutan peptida protein membran luar *Leptospira sp*. iaitu LigB. Vaksin baharu yang berasaskan peptida ini perlu diuji dengan eksperimen dan diharapkan berkesan terhadap semua serovar *Leptospira* yang patogenik dan juga dapat memberikan kita perlindungan daripada leptospirosis.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my deepest gratitude to my project supervisor, Dr Saharuddin bin Mohammad, for the patient guidance, encouragement, and advice throughout my time as his student. I would also like to thank all the lecturers and members of staff in Bioinformatics Department, especially Ms Sugunadevi Rajagopal, who has always promptly responded to my queries regarding the project. I am indebted to my fellow course mates, who have been the source of inspiration and motivation in the past 15 months. Finally, I would like to thank my family who has given me their full support and encouragement to complete this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

APC	Antigen Presenting Cell
HLA	Human Leukocyte Antigen
LigB	Leptospiral Immunoglobulin-like Protein B
MEGA	Molecular Evolutionary Genetics Analysis
MHC	Major Histocompatibility Complex
OMP	Outer Membrane Protein
PDB	Protein Data Bank
TMHMM	Transmembrane Hidden Markov Model

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Leptospirosis & *Leptospira sp.*

Leptospirosis is a worldwide zoonosis that is caused by pathogenic *Leptospira sp.* The bacterial disease is estimated to pose huge global burden on developed and developing countries with more than 500,000 severe cases per year around the world (Silva *et al,* 2007). The first leptospirosis case in Malaysia was documented in 1925 (Fletcher, 1928) involving rubber estate workers and rural inhabitants. Most of the cases of human leptospirosis worldwide have been attributed to rodents, which are believed to be the principal reservoir of leptospirosis (Miller e*t al,* 1991).

The pathogenic leptospires are harboured in the renal tubules of the wild and domesticated mammals, which serve as maintenance hosts, before the spirochaetal bacteria are shed in the animal urine into the environment. The bacteria can penetrate through abraded skin or mucosal membrane of human following accidental, occupational, or recreational exposure to the urine from the infected animals, or contaminated water sources (Faine *et al,* 1999). The patients who are infected with pathogenic *Leptospira sp* have shown diverse clinical manifestations ranging from asymptomatic infection, mild febrile symptoms to severe multi-organ failure. If immediate antibiotic treatment is not sought, mortality rate can increase from 5 to 40% (Farr, 1995).

There are at least 19 species of *Leptospira sp* that have been documented, with 8 species considered to be pathogenic while the rest are saprophytic. The eight pathogenic species, namely, *L. interrogans, L. borgpetersenii, L. santarosai, L. noguchii, L. weilli, L. kirscheneri, L. alexanderi* and *L. alstoni* are reported to be the main agents of leptospirosis (Evangelista & Coburn, 2010; Cinco, 2010). Phenotypic classification based on serology has revealed that there are more than 200 serovars of pathogenic leptospires (Faine *et al* 1999; Goncalves-de-Albuqurque *et al*., 2012).

A well-known *Leptospira, L. biflexa* is a free-living, saprophytic strain that causes no infection despite being genetically and morphologically similar to the pathogenic strain (Nascimiento *et al*, 2004). With the genome size of approximately 4.6 Mbp (Ricaldi *et al*, 2012, pathogenic *Leptospira* is indisputably well equipped with impressive arsenal of proteins at its disposal to facilitate host tropism as well as successfully evade host the immune response, ultimately guaranteeing its survival in the body of the host. Although the pathogenesis of leptospiral infection at molecular level remains enigmatic at the moment, several candidate virulence factors have been identified and implicated in the pathogenesis of leptospirosis (Evangelista and Coburn, 2010).

1.2 Human immune system and *Leptospira sp.*

Once a pathogenic *Leptospira sp.* invade the first line of defence in human body, the extracellular bacterium will be engulfed by professional antigen presenting cells (APC), for example dendritic cells and macrophages, and it will be trapped in an endocytic vesicle. The trapped bacterium is then digested by hydrolytic enzyme in the lysosome of APC into smaller peptide fragments. The particular immune cells continually synthesize

an important protein, major histocompatibility complex (MHC) Class II molecule, or human leukocyte antigen (HLA) molecule in human, and store the protein in a vesicle. Then, the vesicle containing the leptospiral peptides merges with a vesicle containing HLA. Only one of the pathogenic peptide fragments can implant itself into the binding groove of the Class II molecules to form a peptide-MHC Class II complex, before the complex is shuttled to the plasma membrane of APC to be presented to helper T cell, which will initiate a specific and appropriate immune response to the pathogenic bacteria (Weenink & Gautam, 1996).

The damaging effect of leptospirosis is partly due to the release of various inflammatory mediators and serum proteins that make up the host's immune response to the bacteria (Chirathaworn & Kongpan, 2014). Although leptospirosis can be treated with antibiotics, the symptoms of leptospirosis may vary and resemble those of other diseases, making leptospirosis prone for misdiagnosis. Hence, prophylactic vaccine for leptospirosis is required to build the immunity of an individual before the actual infection occurs. The existing vaccines for leptospirosis have been developed and shown disadvantages like suboptimal protection and failure to prevent the disease (Levett, 2001; Wang *et al,* 2007). There is an urgent need for an improved vaccine that can give broader protection and is effective against all pathogenic serovars of *Leptospira*.

1.3 *In silico* screening for vaccine

The adaptive human immune response, which is highly specific against a particular pathogen, is induced when the immune receptors from T cells are able to recognise an epitope, which is a molecular structure of peptides that is presented by MHC Class II

molecules. Only peptides interacting with Class II molecules above a certain affinity threshold are likely to be recognised by T cells and generate a specific immune response (Sette *et al,* 1994). Therefore, identification of the correct epitope is the key step toward the development of a new and effective vaccine. However, epitope identification and mapping are challenging areas of vaccine research due to the overwhelming variation in MHC Class II or HLA alleles in human and limitless permutations of peptide sequence in a single epitope. Thankfully, the epitope prediction can be done computationally and the *in silico* method has proven to significantly shrink the vast sample space and allow us to rationally designed a vaccine before the effectiveness of the vaccine can be further tested *in vitro* or *in vitro* (De Groot *et al*, 2009).

Prediction of which peptides can bind MHC molecules is commonly used to assist in the identification of T cell epitopes. Several tools of epitope prediction have been developed, nevertheless, large numbers of different MHC molecules, each associated with different predictive tools, tool generation and evaluation can be a very resource intensive task. These tools made full use of the available data sets obtained from laboratory experiments as training data and few mathematical algorithms to make predictions of the highly likely epitope to be presented by HLA molecules. To date, several methods have been widely used to predict class II epitopes such as simple binding motifs, quantitative matrices, Average Relative Binding (ARB) matrix (Bui *et al*, 2005), Hidden Markov Model (Lacerda *et al*, 2010) and artificial neural networks (Nielsen e*t al*, 2010).

1.4 Target protein for leptospirosis vaccine

Secreted and surface-exposed antigens are mostly antigenic and susceptible to antibody recognition and T-cell killing (Wang *et al,* 2001). Owing to this fact, outer-membrane proteins (OMPs) from pathogens have garnered huge interests due to their potential in becoming the primary vaccine targets (Amineni *et al*, 2010; Cafardi *et al* 2013). By far, *in silico* vaccine development has been successful in the design of vaccine for Hepatitis B (from HBsAg protein), Influenza (from HA and NA proteins) and Human Papilloma Virus (from capsid protein) (Delany *et al*, 2013).

To design a peptide-based vaccine for leptospirosis, it is important to choose an antigenic and immunogenic outer membrane protein (OMP) that is shared by all pathogenic serovars of *Leptospira* because the antigenic peptide is supposed to be able to elicit broader protection against all pathogenic serovars. It has been suggested that the outer membrane proteins from *Leptospira sp*. have the potential advantage of inducing comprehensive immunity besides being well-conserved and main virulence factors toward host tissues. Several leptospiral transmembrane OMPs have been identified, for example, OmpL1 (Haake e*t al,* 1999), LipL36 (Haake e*t al,* 2000), LipL32 (Haake e*t al,* 2000), LipL21 (Cullen e*t al,* 2005), LipL46 (Matsunaga e*t al, 2*003), LenA (Verma e*t al,* 2010), Loa22 (Matsunaga e*t al, 2*003) and Omp52 (Haake *et al,* 1999).

Extensive research has been done on the potential of the aforementioned proteins as the vaccine candidate for leptospirosis, however, the effectiveness of the proteins as peptide-based vaccine for leptospirosis remains inconclusive (Haake e*t al,* 1999; Guerreiro e*t al,* 2001). One of the recently discovered OMPs from *Leptospira sp*.,

namely LigB protein (Matsunaga e*t al,* 2003), was suggested to be an attractive target for leptospirosis vaccine (Raja & Natarajaseenivasani, 2014). *ligB* gene has been shown to be one of the genes that is upregulated under physiological osmolarity (Cinco, 2010). The protein has also been found in the sera of human patients with leptospirosis and it has been demonstrated to provide almost 100% cross-protection in mice (Koizumi *et al*, 2004). Therefore, the LigB protein is a suitable target protein for leptospirosis vaccine.

LigB is one of Leptospiral Immunoglobulin-like proteins, which is a member of the bacterial immunoglobulin-like (Big) protein super family. The outer membrane protein is able to mediate interaction of bacteria with host cell, which leads to cell attachment and host invasion. The protein assists bacterial admission into the host cell by mediating the adherence of leptospires to the host's extracellular matrix, including fibronectin, fibrinogen, collagen, laminin (Lin *et al*, 2007) and kidney epithelial cells (Cinco, 2010). The LigB protein, which is exclusively found in pathogenic *Leptospira sp*, is predicted to have 12 tandem bacterial immunoglobulin-like repeated domains (Figure 1). The solution structure of the twelfth immunoglobulin-like (Ig-like) repeat domain from LigB (LigB-12) was identified to comprise mainly of β-strands that form a β-sandwich based on a Greek-key folding arrangement by using an NMR spectroscopy (Ptak e*t al,* 2014).

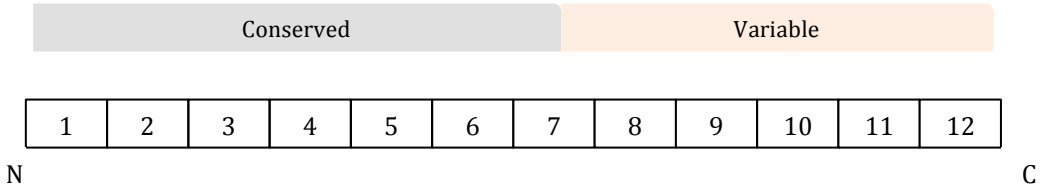


Figure 1: A schematic diagram showing the twelve domains within the LigB protein. The N-terminal 630 amino acids of LigB, which covers the first 6 ½ Ig-like domains, are highly conserved between LigB and its corresponding LigA protein. The remaining C-terminal domains are variable. The twelfth domain is connected to non-Ig-like domain. N: N-terminal, C: C-terminal (Ptak e*t al,* 2014)

1.5 Objective of the study

Having identified a relatively new potential antigenic OMP from pathogenic *Leptospira sp*., an *in silico* analysis was performed with the help of several mostly referenced bioinformatics tools to identify several candidates for peptide-based vaccine for leptospirosis using the LigB as the antigenic protein. The vaccine is hoped to evoke T cell response and give a broader immune protection to human against all pathogenic serovars of *Leptospira sp*. Firstly, a sequence-based analysis was done to identify the best peptide that has the motif of a T cell epitope, which is more likely to initiate immune response from T cells. Secondly, a structure-based analysis was done to computationally verify that the best peptide vaccine candidate could bind to a chosen HLA molecule. The predicted peptide sequence for the vaccine was established to allow an experimental validation in the future.

CHAPTER 2

METHODOLOGY

2.1 Selection of LigB protein sequences

The LigB sequence of *L. kirschneri* RM52 (AAP04736.1), deposited by Matsunaga *et al,* 2003 was retrieved from NCBI Protein database. The sequence of the protein was saved in the FASTA format. Using the LigB protein sequence, a BLASTP search was performed against UniProtKB database using the default parameters to search for homologous sequences. Only protein sequences that were named LigB and that originated from pathogenic *Leptospira sp*. were considered. The list was further trimmed by selecting only sequences that had at least 60% identity to the queried sequence and had a clear nomenclature i.e. the sequences have serovar and species clearly stated. The details of the sequences fulfilling the criteria are summarised in Chapter 3.

2.2 Multiple sequence alignment & divergence analysis

All the selected sequences of LigB, which were saved in a single FASTA file, were subjected to multiple sequence alignment using ClustalW algorithm implemented in MEGA v.6.06 using default parameters. The conserved residues were visualised and identified using Chimera software from the aligned sequences. Using MEGA v.6.06, a Neighbour Joining tree was constructed with 1000 bootstraps at uniform divergence rates with distance ‘p’ as the evolutionary model and with a data subset to use with gaps/ missing data treatment as complete deletion.

2.3 Prediction of the most antigenic protein

The most antigenic protein of all the selected LigB sequences was predicted using an online antigen prediction server, VaxiJen v.2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). The server is especially designed to predict the antigenicity of the protein based on the physicochemical properties of the amino acid (Doytchinova & Flower, 2007). The set of 32 LigB protein sequences was entered and a target organism was set as 'Bacteria'. The default threshold of 0.4 was used for the antigenicity prediction. The outputs given by the server were the antigenicity score and the status of the sequence i.e. whether the sequence is a probable antigenic or non-antigenic.

2.4 Prediction of the T cell epitopes

The sequence of the most antigenic protein was fed into NetMHC v.2.2 server (http://www.cbs.dtu.dk/services/NetMHCII/). The length of the peptide was set to be '15'. This step was performed to generate 15-mer peptides, which mimics the natural process of bacterial degradation in the endocytic vesicle of APC. The server is also able to predict the binding of the peptides to HLA-DR, HLA-DQ, HLA-DP, and mouse MHC class II alleles based on the artificial neural network method (Nielsen *et al,* 2007; Nielsen & Lund, 2009). Only human MHC class II alleles were used in the analysis, since the aim of the current study is to design a vaccine for human. All the 14 HLA-DR alleles, 5 HLA-DP alleles and 6 HLA-DQ alleles, which were used in the analysis, were summarised in Table 2.1.

Table 2.1: The HLA alleles used in the prediction of T cell epitope analysis.

HLA-DR	HLA-DRB1*0101
	HLA-DRB1*0301
	HLA-DRB1*0401
	HLA-DRB1*0404
	HLA-DRB1*0405
	HLA-DRB1*0701
	HLA-DRB1*0802
	HLA-DRB1*0901
	HLA-DRB1*1101
	HLA-DRB1*1302
	HLA-DRB1*1501
	HLA-DRB3*0101
	HLA-DRB4*0101
	HLA-DRB5*0101
HLA-DP	HLA-DPA1*0103-DPB1*0401 (HLA-DPA1*01-DPB1*0401)
	HLA-DPA1*0103-DPB1*0201
	HLA-DPA1*0201-DPB1*0101 HLA-DPA1*0201-DPB1*0501
	HLA-DPA1*0103-(HLA-DPB1*0301/DPB1*401)
	HLA-DPA1*0301-DPB1*0402
HLA-DQ	HLA-DQA1*0101-DQB1*0501
	HLA-DQA1*0102-DQB1*0602
	HLA-DQA1*0301-DQB1*0302
	HLA-DQA1*0401-DQB1*0402
	HLA-DQA1*0501-DQB1*0201
	HLA-DQA1*0501-DQB1*0301

The outputs given were the peptide sequences, affinity score (expressed as IC_{50} value in nM) and bind level (weak binder or strong binder). The default threshold of 500 nM was used to classify the peptides into binders and non-binders. Only high binding peptides (having an IC_{50} value below 50 nM) were filtered and selected. The strong binding peptides would be known as T cell epitopes as the peptides were predicted to have high affinity for the selected HLA molecules, hence more likely to be presented to T cells in the context of MHC Class II molecules. The strong binders were checked whether they covered all three types of HLA molecules, namely HLA-DP, HLA-DR, and HLA-DQ. Only strong binders that were found to be associated to all three HLA types were shortlisted for further analysis.

2.5 Mapping of T cell epitopes to conserved regions

All the shortlisted T cell epitopes from the most antigenic protein sequence were mapped to the conserved residues of the aligned LigB protein sequences. To qualify as a good peptide-based vaccine, the peptide should be a T cell epitope and it ideally needs to be as identical as possible to the conserved residues in the antigenic protein (Bui e*t al*, 2007). An epitope conservancy analysis was performed using epitope conservancy tool available in IEDB analysis resource and is available at (http://tools.immuneepitope.org/tools/conservancy/iedb_input). The conservancy level of each potential epitope was calculated by looking for identity in all the 32 LigB sequences and the server was able to compute the identity score of the epitope to the consensus residues.

The peptide/ T cell epitope with the highest percentage identity to the conserved residues was selected as the best vaccine candidate and would be used for docking analysis. The consensus sequence that corresponds to the T cell epitope was considered to be the best peptide-based vaccine candidate. The peptide sequence was then mapped to the whole 3D structure of LigB protein predicted by RaptorX server (http://raptorx.uchicago.edu/) and the location of the peptide was predicted whether it is located in the part of the protein that is exposed to the external environment by TMHMM v2.0 server (http://www.cbs.dtu.dk/services/TMHMM/). This step was taken to ensure that the part of the antigenic protein is exposed and accessible to the immune cells.

2.6 Rigid docking of peptide vaccine to HLA molecule

A rigid docking analysis was performed between the best candidate of peptide vaccine and a chosen HLA molecule. Prior to the docking, two PDB files (for HLA molecule and peptide vaccine) were prepared. The 3D structure of the peptide vaccine was predicted and designed using the PEP-FOLD Peptide Structure Prediction server at the RPBS mobile portal (Thévenet e*t al,* 2012). The best model provided by the server was validated by PROCHECK analysis. A crystal structure of HLA-DR1 molecule, which was to be docked with a peptide vaccine candidate, was retrieved from Protein Data Bank (PDB) server with the ID of 1KLU. The original peptide (staphylococcal enterotoxin C3 variant 3B2), which was complexed with the HLA molecule, was excluded using Chimera software, leaving only α- and β-chains.

For docking analysis, the two PDB files (the HLA-DR1 as receptor protein and the peptide vaccine as ligand protein) were uploaded onto GRAMM-X Protein-Protein Docking server v.1.2.0. The server uses Fast Fourier Transformation (FFT) methodology for the global search of the best rigid body conformations, followed by the refinement optimization in continuous coordinates and rescoring with several knowledge-based potential terms (Tovchigrechkov e*t al,* 2006). The analysis returned with a PDB file with 10 possible configurations of the peptide binding to the HLA molecule. The final and best conformation of the peptide binding to the HLA molecule was processed with Chimera. The molecular interaction between the peptide vaccine and the HLA molecule was predicted using LigPlot v.4.5.3 program (Wallace *et al,* 1995). The interaction between the residues of the peptide vaccine and the residues in the binding groove of HLA molecule was noted and evaluated.

CHAPTER 3

RESULTS

3.1 List of LigB sequences

A total of 32 sequences of LigB protein from different pathogenic serovars have been selected (Table 3.1). Analysis of antigenicity revealed that all the sequences were antigenic (met the criteria of default threshold level, ≥ 0.4, in VaxiJen v.2.0).

Table 3.1: The list of all LigB protein sequences from pathogenic serovars of *Leptospira* sp. that have been used in the analysis.

UniProt ID	Species	Serovar	Strain	Length (aa)	Source	Antigenicity score
Q04YJ5	borgpetersenii	Hardjo-bovis	L550	1912	Cattle	0.5854
Q04UY1	borgpetersenii	Hardjo-bovis	JB197	1896	Cattle	0.5846
M6F3M0	kirschneri	Bulgarica	Nikolaevo	1916	Human	0.5426
M6SBP2	interrogans	Copenhageni	HAI0188	1890	Human	0.5392
K6JSN9	interrogans	Icterohaemorrhagiae	Verdun LP	1890	Human	0.5391
Q72V39	interrogans	Copenhageni	Fiocruz L1-130	1890	Human	0.5352
M6R7K6	interrogans	Pomona	UT364	1889	Human	0.5315
M6HL92	interrogans	Zanoni	LT2156	1560	Human	0.5308
K6I198	interrogans	Canicola	Fiocruz LV133	1889	Human	0.5277
M6K1Q5	interrogans	Pyrogenes	L0374	1560	Human	0.5261
B5TXC6	interrogans	Pomona	pLPLIGB	1889	Pig	0.5257
C0J1Q6	interrogans	Kennewicki	PO-06-047	1889	Horse	0.5255
G7QKX4	interrogans	Lai	IPAV	1889	Human	0.5249
Q8EZS3	interrogans	Lai	56601	1889	Human	0.5249
M5UHA3	interrogans	Pomona	CSL10083	1844	Sea lion	0.5248
K6F9Q2	interrogans	Pomona	Pomona	1889	Human	0.5243
J7UHZ2	interrogans	Bulgarica	Mallika	1560	Human	0.5237
M6A6I9	interrogans	Pomona	CSL4002	1846	Sea lion	0.5224
M6PJ88	interrogans	Grippotyphosa	UI12764	1606	Human	0.5199
T0FPW9	noguchii	Panama	CZ214	1930	Opossum	0.5195
K8H670	kirschneri	Grippotyphosa	Moskva	1886	Human	0.5173
M6KST2	interrogans	Medanensis	L0448	1889	Human	0.517
M6NG55	interrogans	Pyrogenes	R168	1889	Human	0.5161
Q7X504	kirschneri	Grippotyphosa	RM52	1886	Human	0.516
K8JXB9	interrogans	Grippotyphosa	UI08368	1849	Human	0.5159
K8IAH3	kirschneri	Valbuzzi	200702274	1886	Human	0.5147
M6QEZ6	interrogans	Medanensis	UT053	1889	Human	0.5147
K8J0U3	interrogans	Bataviae	L1111	1889	Human	0.5137
M6I6K5	kirschneri	Bim	1051	1886	Dog	0.5126
K8IQP7	interrogans	Pyrogenes	2006006960	1487	Human	0.5123
M7A596	interrogans	Pyrogenes	2007001872	1889	Human	0.5095
M6F4P4	kirschneri	Bim	PUO1247	1842	Human	0.5062

3.2 Conserved residues of LigB proteins

Figure 3.1: The consensus residues in all the 32 aligned LigB sequences from pathogenic serovars of *Leptospira sp.* The grey bar illustrates the degree of amino acid conservations. The red letters written in capital represent highly conserved amino acids while the purple red letters written in capital represent moderately conserved residues. The small, black-coloured letters are residues with low conservation. The visualisation of the conserved regions was done with Chimera software.

Figure 3.1, continued.

3.3 Phylogenetic tree

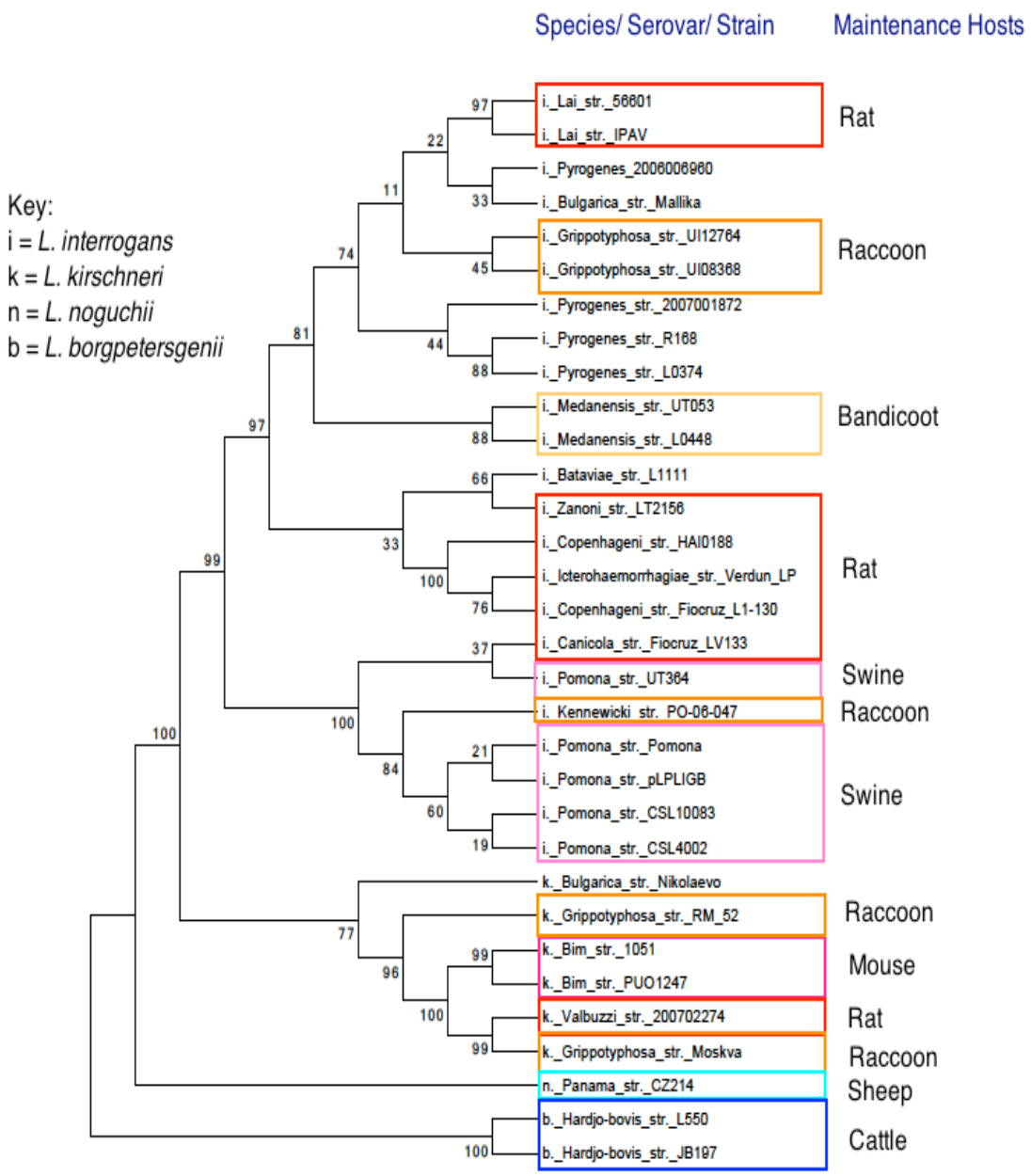


Figure 3.2: The phylogram illustrates, by bootstrap-based topology, the evolutionary relationship between all the LigB sequences from different pathogenic serovars of *Leptospira sp*. For the sake of brevity, the initial lowercase letters have been used to indicate the respective species, to which each serovar belongs: i. *L. interrogans*; k, *L. kirschneri*; n: *L. noguchii*, b: *L. borgpetersenii*. The bootstrap consensus values are indicated over each root. The phylogram, which was constructed by using the Neighbour-Joining algorithm, based on the LigBp protein sequences, demonstrates the clustering of serovars within the species.

3.4 Putative T cell epitopes

It has been predicted that there were 140 peptides associated with HLA-DP alleles, 456 peptides associated with HLA-DQ alleles, and 1526 peptides associated with HLA-DR alleles in the most antigenic protein sequence (Q04YJ5). Out of 2122 peptides/ T cell epitopes, only seven peptides were predicted to be able to interact with the three types of MHC class II alleles (Figure 3.3). These seven 15-mer peptides were considered to be the most potential epitopes for MHC class II alleles (Table 3.2).

>Q04YJ5 (L. borgpetersenii Hardjo-bovis str. L550)

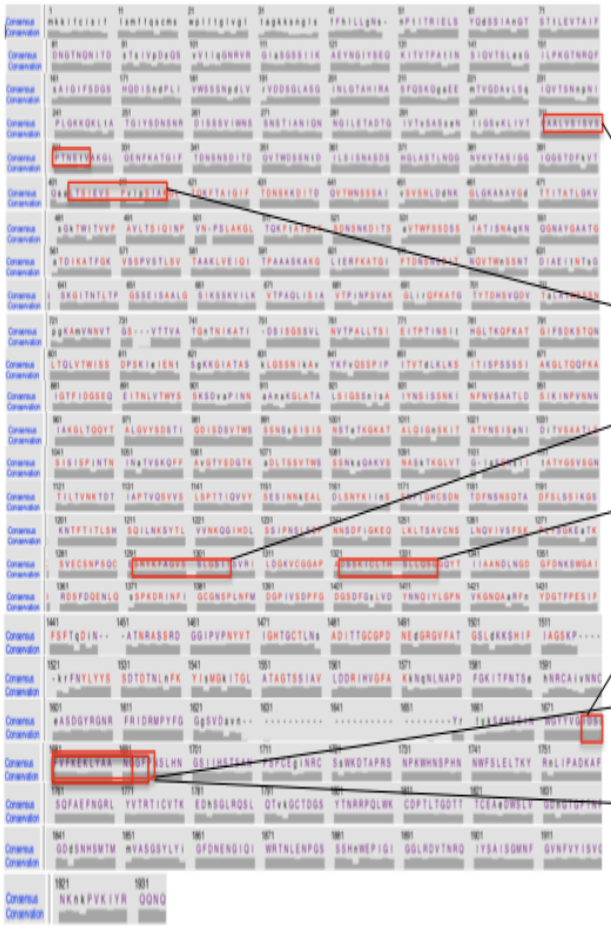
MKKIFSIFLSLFFQGCMVWPLVVGAVGLSTGNKGDGNNPFLFLLGIASDPVITRIELSA
QDSLIAKGTSTALQVTAIFDDGTNMDITDSTSIVSDSQAVVEVQGNRSRGISLGASTLQ
AEYKGLHSQLRITVTSATLDSIQVTSLDRDSLPKGLNRQFSAIGIFSDGSHQDLSDDPLT
IWSSSDSSLVRVNDSGLASGVNLGTAHIHASFGSKRGSATMTVGSATLSSIEVTPVNSN
LPLGKKQQLIATGIYSDNSNKDISSLVTWDILDNTIATIQPKGMVETVSTGSTTVLASF
GSSVGSTTLNVVSASLVSISVSSVNSSKAKGLKENFTATGIFSDDSNLDITNQVTWSSS
DTNILIISNSSGSHGSGSALNQGVVKVIASVGGIEGSMDFTVTQ**AALVSISVSPVLLSM**
PKGLTQQFKAMGIFTDNSKQDITSSVTWTSSSKALSVSNVSGREGMGQALAVGSATV
AATLGRTSGKTTIKVAPAILASIQISPMNTTALAKGLKKSFSARGIYSDNSSSDITSSVT
WFSSDPSVAVVGNVPSYKGEVRGEKIGTTNIKAALGNVSSPIVALSVTEAELVSIQVSP
YVSVTPKGITKNFKATGTFTDYSTQDITEQVTWKSSDTEIISIENAAGNKGLAHMLKQ
GDSYITATLGSISSSPELAMVASPTIVSVAVTPSNPSVVKGLTCQFKATATYTDNSTADI
TSMVSWSSSNSNKALVGNDILSGGLVTAVATGSANITARYENLSGSSAVNVTPATLTS
IEVTPVFPSVAKGLTEQFTATGIYSDKSTQDLTQVVTWISSDSSRVAIENTAGKKGLAL
ASTLGSSNIRATYNSIQSSPISMTVTEAKLVSITVSPEFTSKALGLTQQFKAKGIFTDGSE
RDITNLVTWFSSEPLVADAINADENRGLAVSHSIGSTEIHAYYDSVESNSVNFKVTSSE
LVSIEISPENGDLIKGLNQQYTAFGVYSDGSLQDISNSVTWYSSNTSSVSISNAVGSKG
KATALQVGTSKITATYRSVSGTTDLNVSAAILSSIVVSPTKPNVESTSKTKFFAVGMYS
DGTKKDLTSSVTWFSSSTNASVSNALKSKGLVVAGSGTGYSTITATYGSISGNTLLIVN
KYNKAVPTVKSVTSLSPTAIRIVYSEFVNNKEALKLSNYKVVDSSSLVGSCFDNADFK
KNSQTRDFSLKSISGAGDTFTITLSGSQNSGKVYTLIVNKPGIHDRSYAPQSLGCPNNA
DFVGKEQLKLTNAICNSVNRVIVTFSKPLYTGNDIAKSAECSNPSQCK**SRYKFAGVSV**
LGDVTSAKILNGKVCGGAPA**DPSKVCLTHTLLQSG**GHYTVMAANRLDGDGFDND
VWGAISDSSGQEALQPSPKDRANFVGCGGSPVNFADGPVISDPFGDGSSFASLTNYRY
QVYLGPNRKGNQAVRFRYDGTSPESVFFSFAKDAVGEQSSNTALSGISANYVTMGHA
GCTRNSADIVAGCGPDNEDGRGVFATGLLGGRPHMFMAGSKSLGGLDYLYYSSSAD
ADLDFKYIDMGTITGELTAGASSMAVFDNRVYVGFAKKNNRSNAPDFGKITFHTSDS
TRCVIGSNCDATDGQRGRRFRIDRMPYFGGESLDEEYKGRVGSTARLIRTLAAKNNG
NKPNQLGDSTINWGYYVG**IDSLFVFKGTLYAANGGF**PNSLHNGSIIRSTSANPAPCE
GKNRCSGWEDVAPRSNPKWHNSPDNNWFSLELVKERDLIPADKAFSQFAEFNGRMY
ATRTICVTSEDRSGLRKSLQTVKGCTDGSYTNRRPQLWKCDPTLSGNKSTCDSGDWS
VVGDDGTGITNFGNVFNHSITMLVANGSYLYVGFDNENGIQIWRTNLQNPGSSSSGW
EQVGGDGLGDVTNRRIYSGITVPKLSLNYVYVSTGGNNRPVKVYRQQNR

Figure 3.3: The sequence of the most antigenic protein (from serovar Hardjo-bovis str. L55) contains seven T cell epitopes, as predicted by NetMHCII v2.2. The T cell epitopes are bolded and coloured red. Three of the seven sequences of T cell epitopes overlap with each other.

Table 3.2: The list of all T cell epitopes found in the most antigenic LigB protein (UniProtKB ID Q04JY5) as predicted by NetMHCII v2.2 server. All the epitopes were associated with all the three types of HLA alleles namely, HLA-DP, HLA-DQ, and HLA-DR. The affinity score of the epitope to the corresponding alleles is in the bracket.

No.	Predicted T cell epitopes	Interacting MHC-II (HLA) alleles & affinity score (nM)
1	AALVSISVSPVLLSM	HLA-DPA10301-DPB10402 (44.6) HLA-DQA10102-DQB10602 (41.5) HLA-DQA10501-DQB10301 (44.9) HLA-DRB10101 (7.0) HLA-DRB10701 (17.8) HLA-DRB10901 (36.1) HLA-DRB40101 (47.9)
2	DPSKVCLTHTLLQSG	HLA-DPA10301-DPB110402 (39.9) HLA-DQA10102-DQB10602 (39.6) HLA-DRB10701 (31.9)
3	IDSLFVFKGTLYAAN	HLA-DPA10103-DPB10401 (13.5) HLA-DPA10103-DPB10201 (9.5) HLA-DPA10103-DPPB10301 (37.1) HLA-DQA10501-DQB10301 (19.2) HLA-DRB10101 (16.8)
4	LFVFKGTLYAANGGF	HLA-DPA10103-DPB10401 (18.4) HLA-DPA10103-DPB10201 (14.6) HLA-DQA10501-DQB10301 (24.9) HLA-DRB10101 (41.0)
5	LVSISVSPVLLSMPK	HLA-DPA10301-DPB10402 (41.1) HLA-DQA10102-DQB10602 (25.6) HLA-DRB10701 (30.9)
6	SLFVFKGTLYAANGG	HLA-DPA10103-DPB10401 (17.3) HLA-DPA10103-DPB10201 (13.2) HLA-DQA10501-DQB10301 (22.0) HLA-DRB10101 (21.8)
7	SRYKFAGVSVLGDVT	HLA-DPA10301-DPB10402 (48.4) HLA-DQA10501-DQB10301 (26.5) HLA-DRB10101 (25.7) HLA-DRB10405 (44.8) HLA-DRB10701 (19.0)

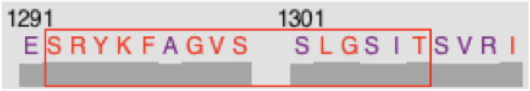
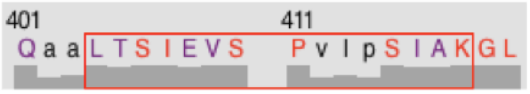
3.5 T cell epitopes in the conserved regions



Predicted epitopes	Start-end position
AALVSISVSPVLLSM	312-326
LVSISVSPVLLSMPK	404-418
SRYKFAGVSVLGDVT	1292-1326
DPSKVCLTHTLLQSG	1322-1336
IDSLFVFKGTLYAAN	1677-1691
SLFVFKGTLYAANGG	1679-1693
LFVFKGTLYAANGGF	1680-1694

Figure 3.4: All the T cell epitopes were mapped to the consensus residues in the LigB protein. All of the epitopes are in the highly or moderately conserved regions of LigB protein.

Table 3.3: The percentage identity of the T cell epitopes to the conserved region of LigB protein. The consensus amino acid sequences that correspond to the respective predicted epitope was boxed. The underlined residues are the residues in the epitope that vary from the consensus amino acid sequences.

No.	Predicted epitopes	Consensus amino acids / Degree of conservation	% identity of epitope to conserved region
1	DPSKVCLTHTLLQSG		80.00
2	SRYKFAGVSVLGDVT		73.33
3	AALVSISVSPVLLSM		53.33
4	LVSISVSPVLLSMPK		46.47
5	SLFVFKGTLYAANGG		33.3
6	LFVFKGTLYAANGGF		33.33
7	IDSLFVFKGTLYAAN		33.33

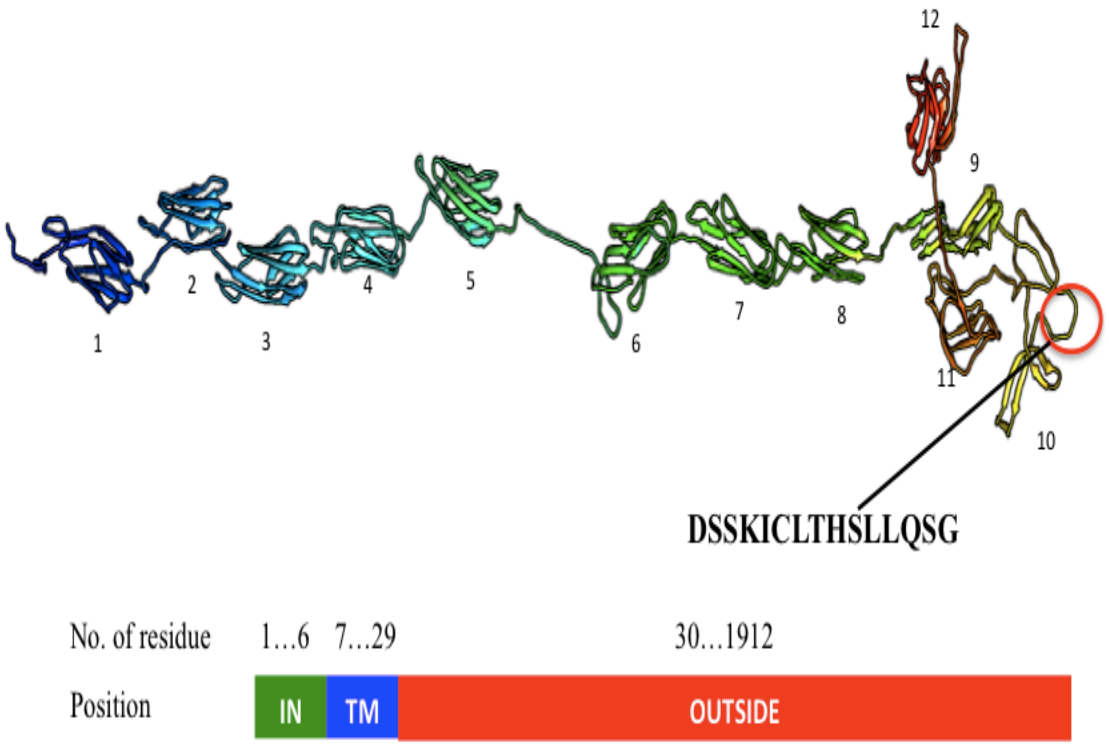


Figure 3.5: The target of the vaccine (in red circle) was predicted to be in the part of LigB that is exposed to the external environment and more likely to be seen and recognised by the immune cells. The 3D structure of the LigB protein was predicted using the most antigenic LigB sequence (Q04JY5) using RaptorX server. The location of the residues in the membrane was predicted by TMHMM v.2.0.

3.6 Peptide-HLA docking result

The stereo chemical quality of the 3D structure of the chosen peptide vaccine ("DSSKICLTHSLLQSG"), as shown in Figure 3.6, was assessed using PROCHECK tool. The Ramachandran plot generated from the analysis showed that >90% of residues are in the allowed region (Figure 3.7).

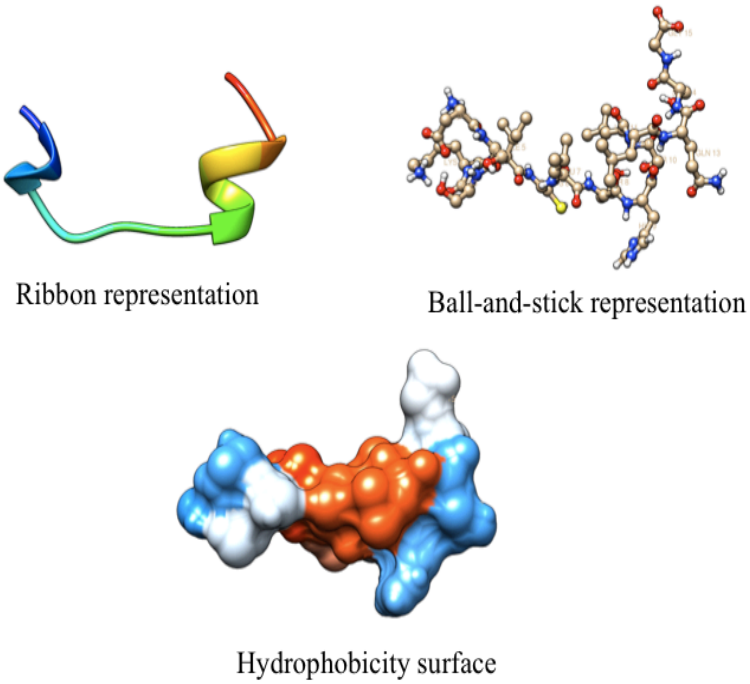


Figure 3.6: The predicted 3D structure of the vaccine candidate, modelled by PEP FOLD server. The ribbon representation suggests that the peptide vaccine has a core with coil structure while both ends contain helical structure. The hydrophobicity surface indicates that the peptide core is made of hydrophobic residues (red) while the ends have more hydrophilic residues (blue and white). The structures are visualised by Chimera.

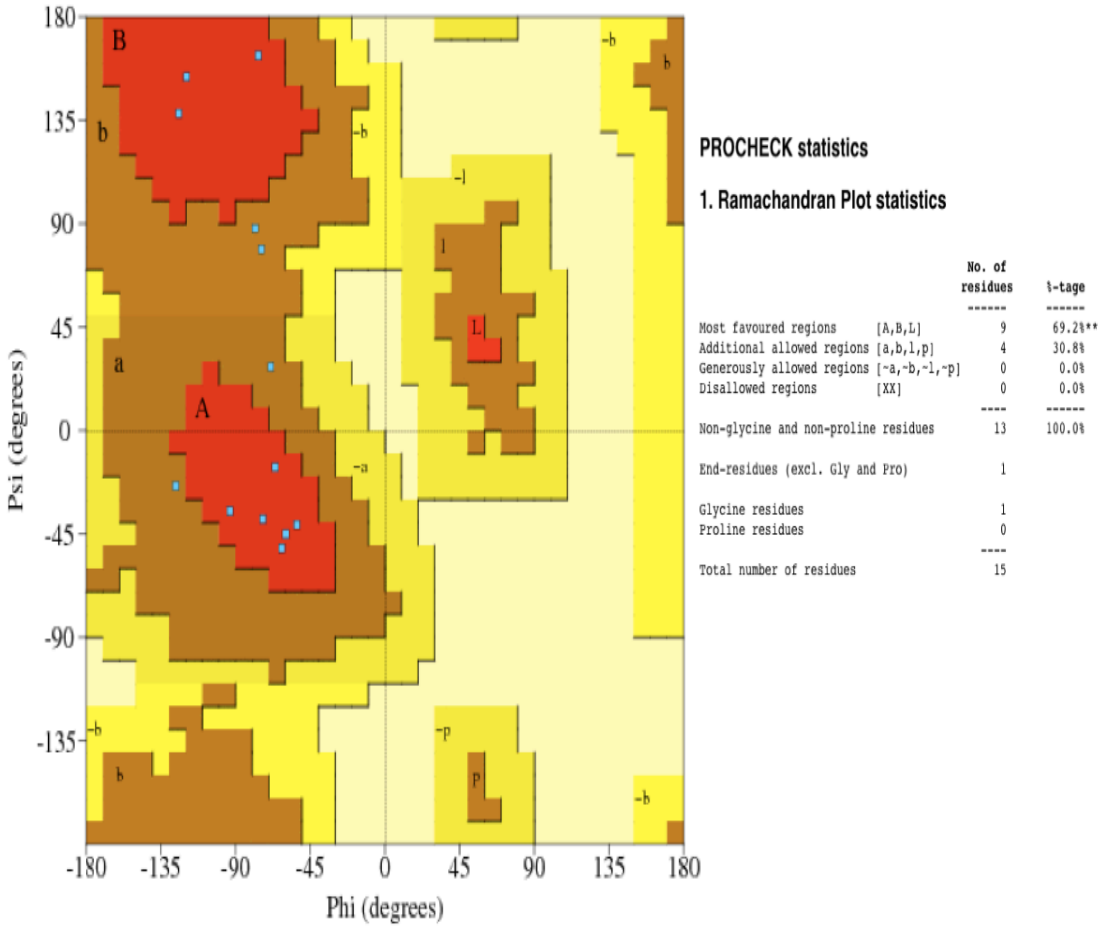
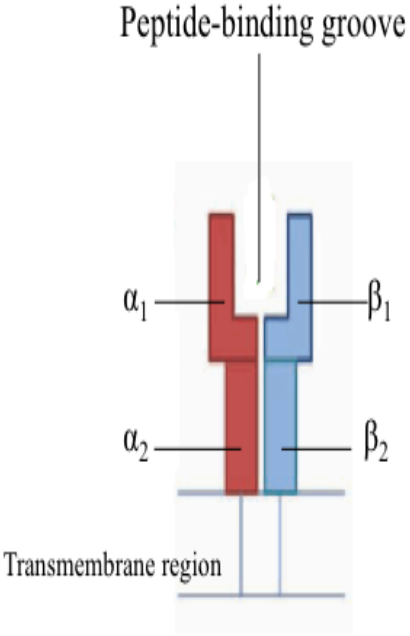


Figure 3.7: PROCHECK statistics was checked to validate the 3D structure. The Ramachandran plot suggests that none of the amino acids in the peptide vaccine is in the disallowed region. Therefore, the predicted structure was good for the *in silico* analysis.



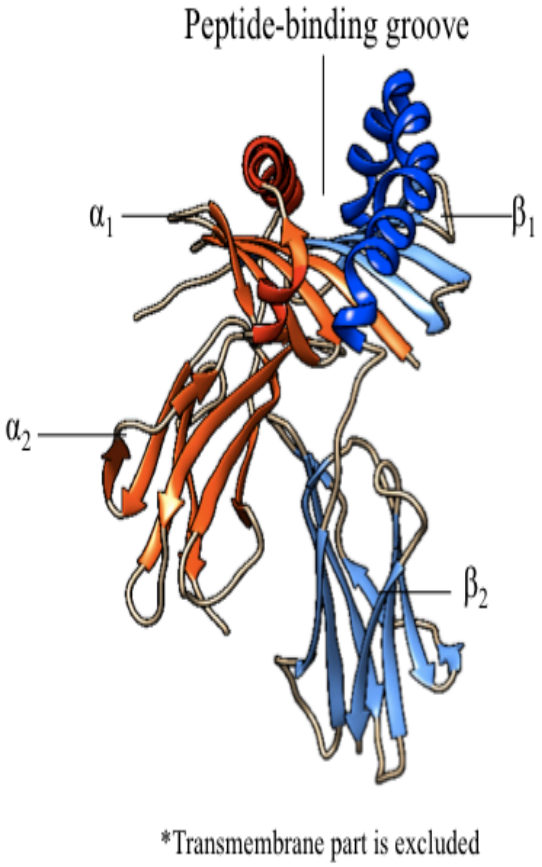


Figure 3.8a

Figure 3.8b

Figure 3.8a: The schematic diagram of general MHC Class II molecule in human. The molecule is made of two chains (α chain is coloured red while β chain is coloured blue). Each chain is folded into two domains. Figure 3.8b: The ribbon representation of the HLA-DR1 molecule. The 3D structure was taken from the crystallised HLA-DR1 molecule taken from PDB with the ID of 1KLU.

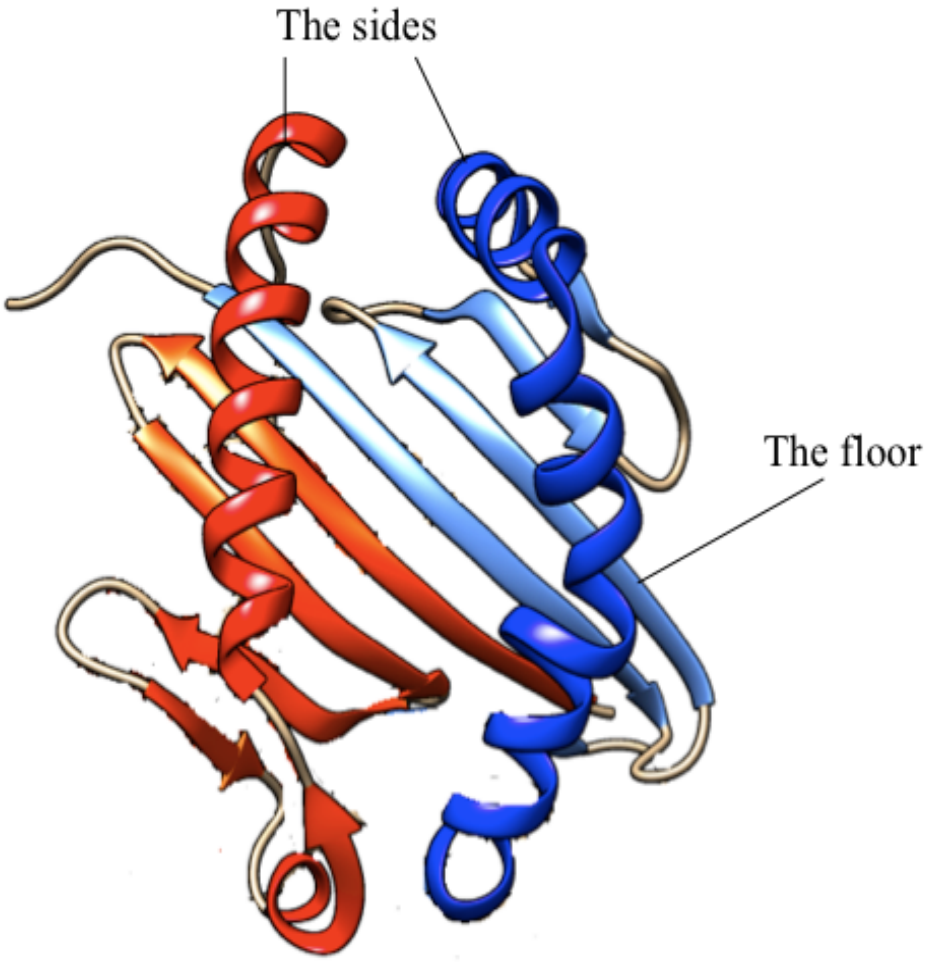


Figure 3.9: The ribbon representation of a vacant binding-groove of HLA-DR1 molecule. The sides of the groove are made of two α-helices (red) while the floor is made of eight antiparallel β-pleated sheets (blue). Both ends of the groove are open-ended.

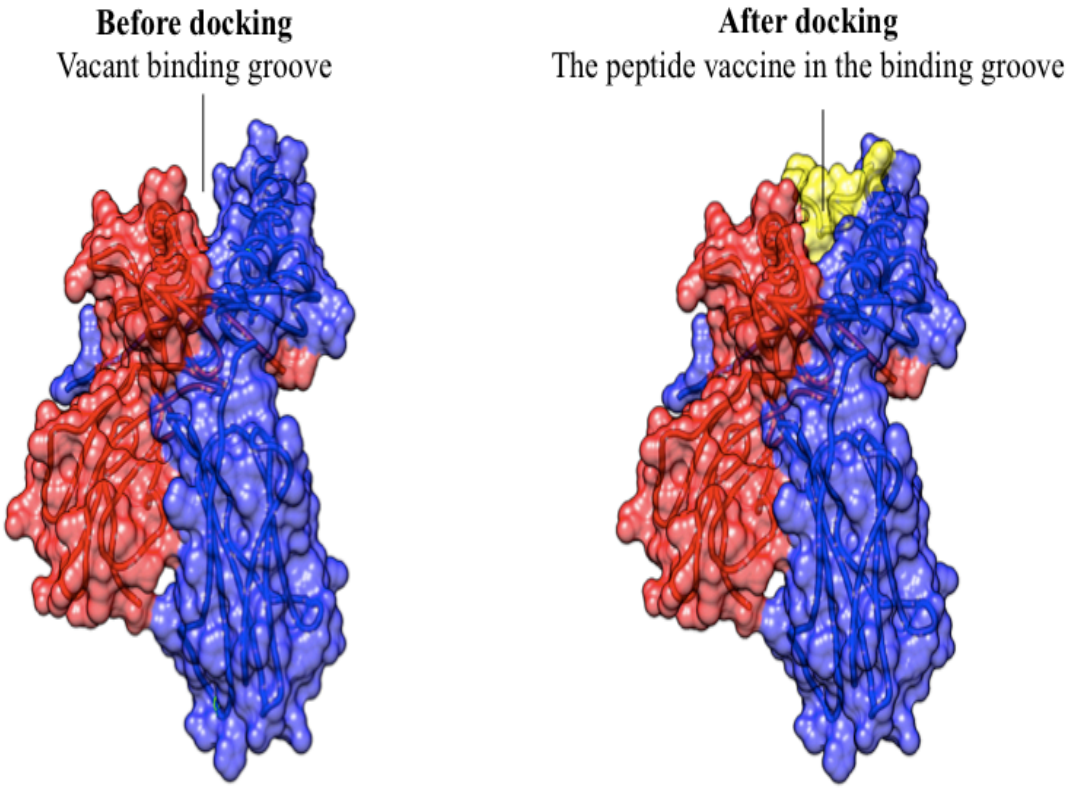


Figure 3.10: The HLA-DR1 molecule before and after the peptide vaccine was docked to it. The peptide vaccine (yellow) is predicted to occupy the binding groove of the HLA-DR1 molecule.

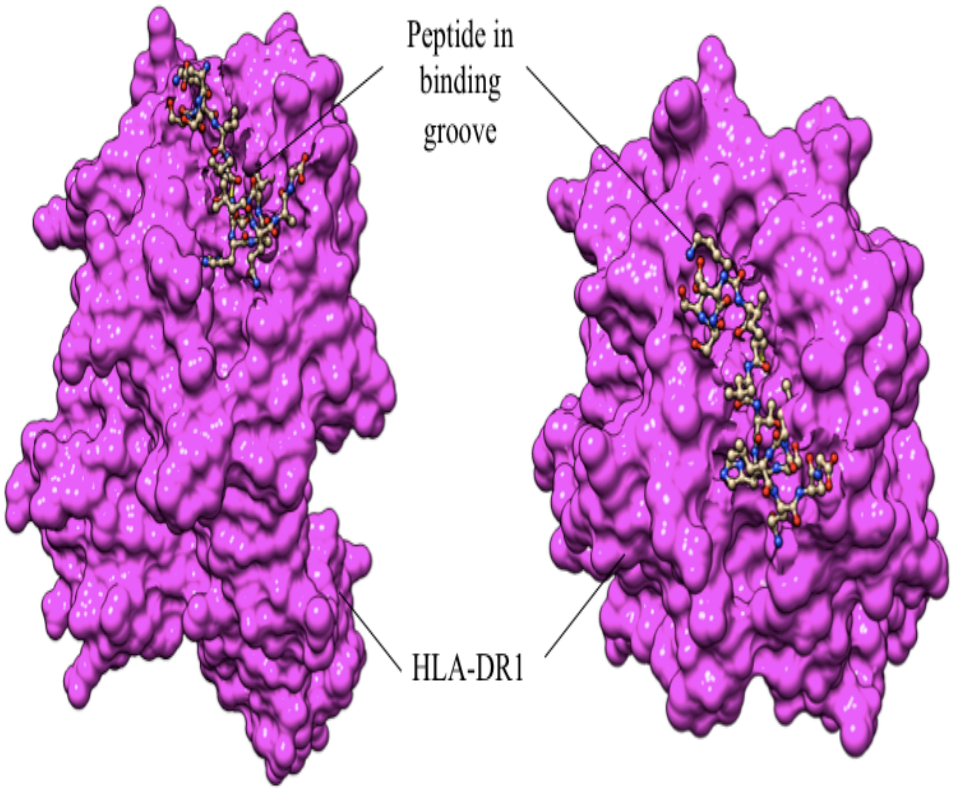


Figure 3.11: The peptide vaccine-HLA-DR1 complex from side view (left) and top view (right). The image on the right is probably the view of the peptide-HLA complex as 'viewed' by the immune cells.

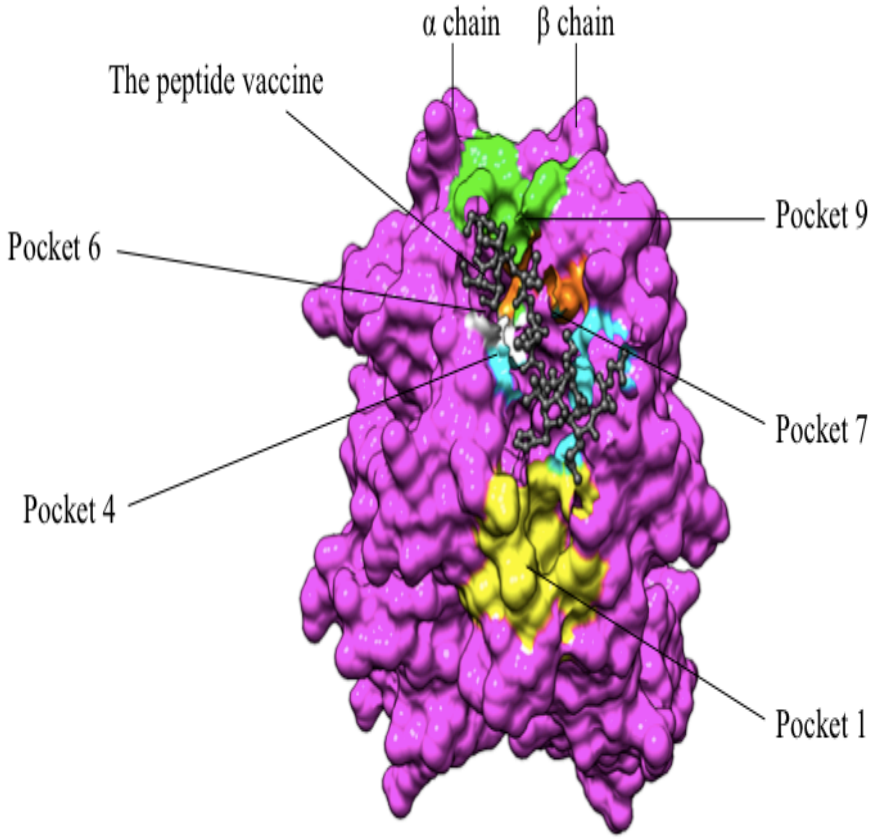


Figure 3.12: The peptide vaccine (grey) occupying the binding pockets of the HLA-DR1 molecule. Pocket 1 (yellow), Pocket 4 (cyan), Pocket 6 (white), Pocket 7 (orange), and Pocket 9 (green) are made of residues that interact with residues in the peptide vaccine. The pockets are coloured according to PPD (Program for Pocket Definition) Pocket Compositions (Androulakis e*t al,* 1997).

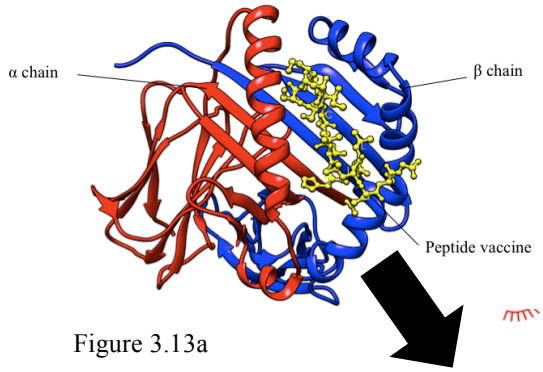


Figure 3.13a

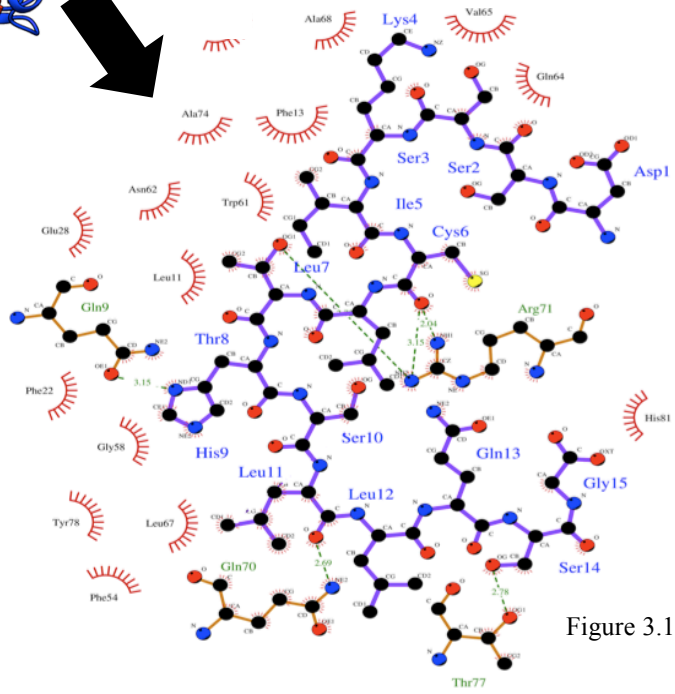


Figure 3.13b

Figure 3.13a shows the peptide vaccine (yellow) in the binding groove of the HLA-DR1 molecule. Figure 3.13b illustrates the molecular interactions between the residues in the peptide vaccine and the residues in the HLA-DR1 molecule. The bonds in the peptide vaccine (magenta); residues that make hydrophobic contacts (red-spoked arcs); hydrogen bonds (dashed green lines).

Table 3.4: The residues that are involved in the hydrogen bonding that holds the peptide-vaccine in the peptide-binding groove of HLA-DR1 molecule. The length between the atoms of the participating residues is predicted by LigPlot software.

Residues in HLA	**Residues in peptide**	**Length (Å)**
Gly-9	His-9	3.15
Gln-70	Leu-12	2.69
Arg-71	Leu-7	2.72
Arg-71	Cys-6	3.15
Arg-71	Cys-6	2.04
Thr-77	Ser-14	2.78

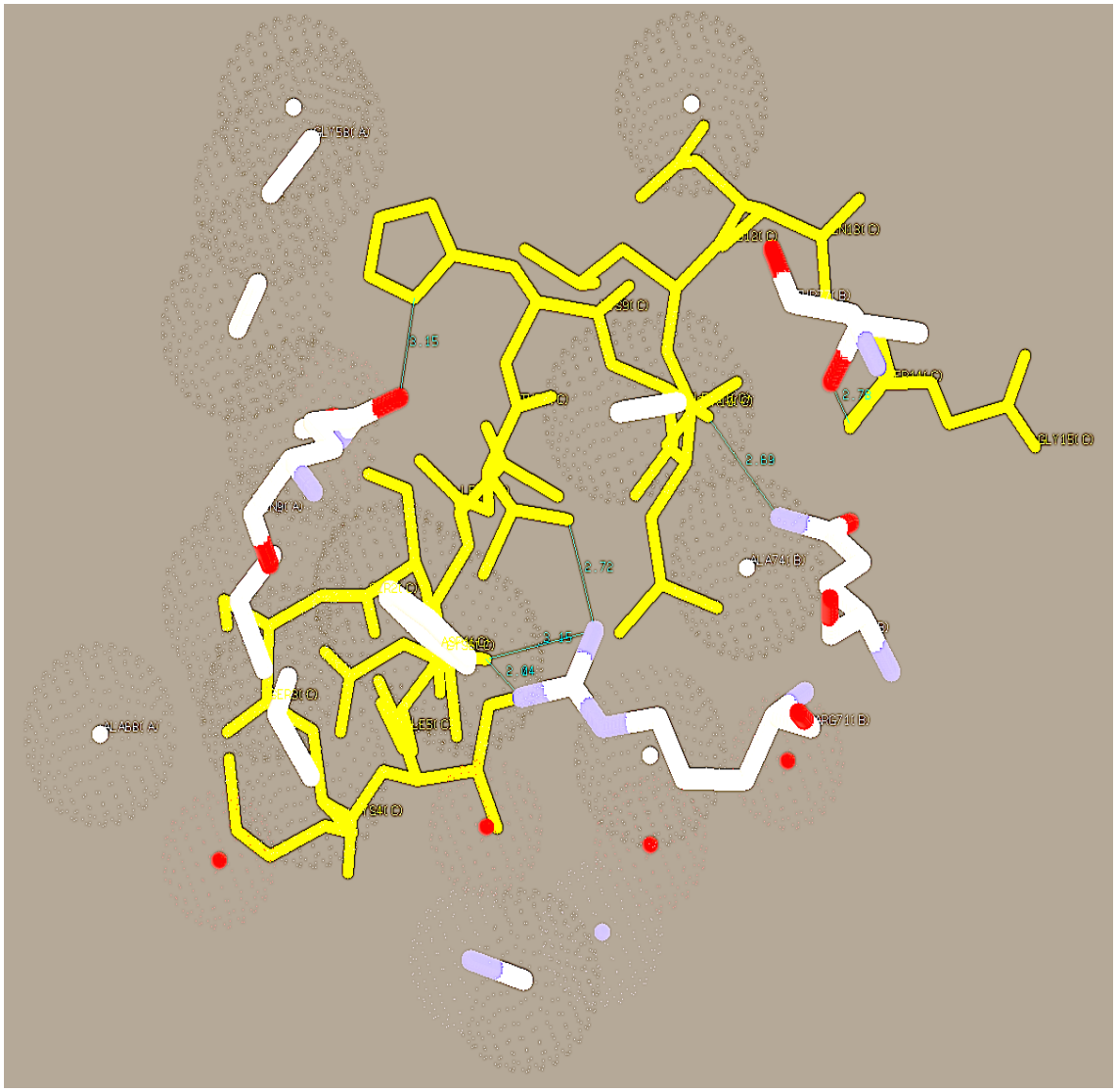


Figure 3.14: The 3D representation showing the interactions between the residues in the peptide vaccine and amino acids in the binding-pockets of HLA-DR1 molecule. Peptide vaccine (yellow), residues in the HLA molecule (grey sticks) making hydrogen bonds (cyan lines) residues in the HLA molecule (white sticks). The number on the cyan line is the distance (in Å) between the residues that form hydrogen bonds. The residues making hydrophobic interactions are surrounded by diffused white clouds.

CHAPTER 4

DISCUSSION

4.1 LigB protein as the target for vaccine

Leptospirosis, which is one the most important bacterial infections that has cost the lives of humans and livestock worldwide, requires an improved vaccine since the existing vaccines can only provide suboptimal protection. Although the disease can be easily treated with a prescription of antibiotics, the use of prophylactic vaccine to prevent the bacterial diseases is more appealing. The former could potentially lead to antibiotics resistance problem, which is harder to tackle once the pathogenic bacteria mutate and the existing antibiotics are no longer effective against the pathogens. Overexposure of the antibiotics to the bacteria will only give an advantage to the bacteria to be ahead of us in the antibiotics arms race. Furthermore, antibiotics are only effective if the disease is diagnosed and detected early. This is not always the case with leptospirosis because the symptoms of the disease are rather indistinguishable from those of other non-life-threatening diseases.

In the current study, an attempt to computationally design a peptide-based vaccine based on a recently discovered antigenic outer-membrane protein, LigB, was undertaken. Thirty-two LigB sequences were successfully retrieved from UniProtKB database. More than 80% of LigB sequences available in the database were partial sequences (having less than 300 amino acids) and had to be disregarded, leaving only complete or almost complete sequences (having length of greater than 1400 amino acids) that were amenable for the analysis. This stringent requirement was set to avoid

some consensus regions to be missing out from the analysis. As a result, only 17 unique pathogenic serovars (Lai, Pyrogenes, Bulgarica, Grippotyphosa, Pyrogenes, Medanensis, Bataviae, Zanoni, Copenhageni, Icterohaemorrhagiae, Canicola, Pomona, Kennewicki, Bim, Valbuzzi, Panama and Hardjo-bovis) were covered. Nevertheless, this number is adequate for the purpose of the study. Ideally, attempts to sequence the complete LigB proteins could be made for other obscured pathogenic serovars to give wider coverage of pathogenic *Leptospira* serovars for the future analysis and unveil more accurate representation of the conserved regions in LigB protein.

The immune cells need to be primed to recognise the part of the protein from the pathogen that is directly exposed to them. Therefore, the choice of LigB protein as the target for the vaccine is relevant. LigB protein is purported to be a transmembrane protein and it is only present in pathogenic species. Based on the analysis, the proposed peptide vaccine is supposed to activate the naïve helper T cells to induce Th2 response, which will mediate the activation and maintenance of the humoral immune response against the extracellular pathogenic *Leptospira sp.* Neutralising antibodies specific to the conserved region of the LigB protein will impede the function of LigB, which serves as a vital gateway to the host cell. This will effectively prevent the bacteria from invading the host cell. Most importantly, the protein has been predicted to contain several T cell epitopes, which means that the protein is likely to be presented to the helper T cell. The fact that the peptide vaccine is designed to be as identical as possible to the conserved region of the LigB proteins for all pathogenic serovars means that the vaccine could be effective against broader pathogenic serovars. Nevertheless, the effectiveness of the vaccine, along with its potential in causing allergenicity, needs to be confirmed *in vivo* and *in vitro* studies in the future.

The use of LigB sequence from *L. borgpetersenii* of serovar Hardjo-bovis and strain L550 in the prediction of T cell epitopes is apt because it has been predicted to be the most antigenic protein by VaxiJen, an antigenicity predictive server. Furthermore, serovar Harjo-bovis is suggested to be the most commonly implicated serovar in leptospirosis in dairy farmers, slaughterhouse workers, and animal-associated occupations (Black *et al,* 2001). Hence, an improved leptospirosis vaccine for these workers, who are at high risk of obtaining the infection from the infected animals, is urgently needed. The current vaccine against serovar Hardjo-bovis is only effective against that particular serovar. It is hoped that the improved vaccine will provide, not only protection against serovar Hardjo-bovis, but also against other pathogenic serovars.

4.2 Future direction of the study

An individual inherits MHC or HLA genes in a Mendelian fashion from each parent as an HLA haplotype. The highly polymorphic nature of HLA molecules is attributed to the HLA loci, which are known to be the most polymorphic loci known in higher vertebrates. The polymorphism of HLA gives rise to the peptide-binding specificity, which is dictated by a limited number of amino acid residues in the peptide-binding pockets (Choo e*t al, 2*007). To verify that the peptide vaccine can, at least, bind to the HLA molecule *in silico,* a rigid docking could be conducted between the designed peptide vaccine and an HLA molecule.

In the current study, the structure-based analysis revealed that the predicted peptide could accommodate the binding-groove of an HLA molecule. Several forces such as hydrogen bonds and hydrophobic interactions were predicted to be the forces that are responsible to stabilise the peptide in the groove to be presented to helper T

cells. According to the rigid docking analysis, hydrophobic interactions are the forces that stabilise the peptide vaccine in the binding groove. The additional forces that are predicted to be present are hydrogen bonding, which involved several residues in the peptide vaccine and the residues in the HLA-DR1 molecule. Arg-71 residue in the β-chain (Pocket 4) of the HLA molecule is the residue that is majorly involved in the hydrogen bonding to hold the peptide in the binding pocket. A non-synonymous mutation that results in the change of the Arg-71 residue could potentially alter the interaction of the whole peptide with the HLA molecule, subsequently reduce the probability of the peptide to form a complex with an HLA molecule and ultimately diminish the effectiveness of the designed vaccine.

The distribution and frequency of HLA alleles can vary greatly among different ethnic groups. This factor, along with the polymorphism in the HLA structure itself, complicate the *in silico* development of the vaccine. However, the two factors need to be considered to determine how well the peptide vaccine will fare as an effective vaccine in individuals with diverse genetic background. In the future, the variability of HLA molecules resulting from different HLA haplotypes needs to be taken into account. For example, instead of docking the designed peptide vaccine to a single type of HLA molecule, a predicted 3D structure of HLA molecule, which represents all HLA-DR, HLA-DQ, and HLA-DP molecules, can be used. The predicted structure, which can be generated by homology modelling, will give a better and meaningful prediction. However, the accuracy of the structure relies on the template model used and the availability of the HLA sequences in the database. The future study should also consider the variability of the residues in the binding pockets of the HLA sequences among individuals in different ethnics. Since leptospirosis susceptibility was reported to have an association with a certain HLA haplotype (Lingappa *et al,* 2004), a closer look

at the molecular interaction of leptospiral peptide with the HLA molecule from susceptible individual and non-susceptible individual can be done to design an improved peptide vaccine that can give protection to all individuals, regardless of their HLA haplotype.

As of 2008, a total of 986 alleles for Class II molecules (HLA-DR, HLA-DP, HLA-DQ) was recorded and deposited in IMGT/ HLA Database release 2.22 (Shiina *et al,* 2009). This number is predicted to grow in years to come. Hence, the developers of the T cell epitope predictive tools need to revise and update their servers regularly. The current study was able to cover only 2.5% of the HLA alleles reported in the current IMGT/ HLA database. The author is aware of the limitations of *in silico* studies because all of the training data sets used in the T cell epitope prediction tool have been derived from laboratory experiments. The reliability of *in silico* methods relies heavily on the quantity and most importantly the quality of the experimental data. Although more HLA alleles are discovered and added into the IMGT/ HLA database every year, the results of the experiments that establish the association of peptides with the newly found alleles, are underreported and cannot be included in the training data. Immunologists who work in the laboratory should be encouraged to share their results with the bioinformaticians community to enable bioinformaticians to design a more accurate mathematical algorithm or improve the existing predictive tools. The concerted effort between immunologists and bioinformaticians will benefit both communities as well as humankind with the fast discovery of much more improved vaccine to eradicate or at least reduce the mortality rate caused by global infectious disease such as leptospirosis.

CHAPTER 5

SUMMARY

The present study has shown that an integrated computational approach could be used to predict an improved peptide-based vaccine candidate against bacterial pathogen, *Leptospira sp,* which is the causative agent of leptospirosis. Unlike traditional method, rational design and *in silico* studies can vastly minimise unnecessary trial and error repeats, thus saving time and cut costs for the vaccine development. Several publically available bioinformatics tools have been utilised in the research. The sequence-based analysis and structure-based analysis revealed that the epitope with the sequence of 'DSSKICLTHSLLQSG' could be the best peptide-based vaccine candidate. The peptide is predicted to be likely to be presented by MHC Class II in human and more likely to induce an immune response. The target of the vaccine is present in all pathogenic serovars of *Leptospira* and is predicted to be in the conserved region of LigB protein that is exposed to the external environment, thus, more likely to be recognised by T helper cells. Having known the sequence of the potential peptide vaccine, the vaccine can be synthesized using organic chemistry for experimental validation.

REFERENCES

Androulakis I.P., Nayak, N.N., Ierapetritous, M.G., Monos, D.S. & Floudas, C.A. (1997) A predictive method for the evaluation of peptide binding in Pocket 1 of HLA-DRB1 via global minimization of energy interactions. *PROTEINS: Structure, Function, and Genetics,* 29: 87-102.

Altschul S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res. 25:3389-3402.

Amineni, U., Pradhan, D., & Marisetty, H. (2010). In silico identification of common putative drug targets in Leptospira interrogans. *J Chem Biol, 3*(4), 165-173. doi: 10.1007/s12154-010-0039-1

Berzovsky, J.A. (1995) Designing peptide vaccines to broaden recognition and enhance potency. *Ann NY Acad Sci,* 754, 161-168.

Black, P.F., Corney, B.G., Smythe, L.D., Dohnt, M.F., Norris, M.A. & Symomds, M.L. (2001) Prevalence of antibodies to Leptospira serovars in beef cattle in central Queensland. *Aust Vet J*, 79(5): 344-348.

Bui, H. H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics, 8*, 361. doi: 10.1186/1471-2105-8-361

Bui, H. H., Sidney, J., Peters, B., Sathiamurthy, M., Sinichi, A., Purton, K. A., Mothe, B. R., Chisari, F. V., Watkins, D. I. & Sette, A. (2005). Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics, 57*(5), 304-314. doi: 10.1007/s00251-005-0798-y

Cafardi, V., Telford, J. L., & Serruto, D. (2013). Bacterial Genomes and Vaccine Design. 13-37. doi: 10.1007/978-1-4614-5070-2_2

Cerqueira, G. M., McBride, A. J., Picardeau, M., Ribeiro, S. G., Moreira, A. N., Morel, V., Reis, M. G., Ko, A. I. & Dellagostin, O. A. (2009). Distribution of the leptospiral immunoglobulin-like (lig) genes in pathogenic Leptospira species and application of ligB to typing leptospiral isolates. *J Med Microbiol, 58*(Pt 9), 1173-1181. doi: 10.1099/jmm.0.009175-0

Choo S.Y. (2007) The HLA system: Genetics, immunology, clinical testing, and clinical implications. *Yonsei Medical Journal,* 48(1): 11-23.

Cinco, M. (2010) New insights into the pathogenicity of leptospires: evasion of host defences. *New Microbiologica*, 33, 283-292.

Chirathaworn, C. & Kongpan S. (2013) Immune responses to Leptospira infection: roles as biomarkers for disease severity. *Braz J Infect*, 18(1): 77-81.

Delany, I., Rappuoli, R., & Seib, K. L. (2013). Vaccines, reverse vaccinology, and bacterial pathogenesis. *Cold Spring Harb Perspect Med, 3*(5), a012476. doi: 10.1101/cshperspect.a012476

De Groot, A. S., Moise, L., McMurry, J. A., & Martin, W. (2009). Epitope-Based Immunome-Derived Vaccines: A Strategy for Improved Design and Safety. 39-69. doi: 10.1007/978-0-387-79208-8_3

Dimitrov, I., Flower, D.R. & Doytchinova, I (2013) AllerTOP – a server for in silico prediction of allergens. *BMC Bioinformatics,* 14(6): 1471-2105.

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics, 8*, 4. doi: 10.1186/1471-2105-8-4

Evangelista, K. V., & Coburn, J. (2010). Leptospira as an emerging pathogen: a review of its biology, pathogenesis and host immune responses. *Future Microbiol, 5*(9), 1413-1425. doi: 10.2217/fmb.10.102

Faine, S., Adler, B., Bolin, C. & Perolat, P. (1999) Appendix 2 species and serovar list. In: Leptospira and leptospirosis, 2nd edn. Medisci, Melbourne, Australia, pp 138-139.

Farr, R.W. (1995) Leptospirosis. *Clin Infect Dis,* 21:1-6.

Fletcher, W. (1928) Recent work on leptospirosis, tustsugamushi disese and tropical typhus in the Federal Malay States. *Transaction of the Royal Society of Tropical Medicine and Hygiene,* 21: 265-288.

Goncalves-de-Albuquerque, C. F., Burth, P., Silva, A. R., Younes-Ibrahim, M., Castro-Faria-Neto, H. C., & Castro-Faria, M. V. (2012). Leptospira and inflammation. *Mediators Inflamm, 2012*, 317950. doi: 10.1155/2012/317950

Guerreiro, H., Croda, J., Flannery, B., Mazel, M. & Matsunaga J. (2001) Leptospiral proteins recognized during the humoral immune response to leptospirosis in humans. *Infect Immun,* 69: 4958-4968.

Haake D.A., Chao G., Zuerner R.L. & Barnett D. (2000) The leptospiral major outer membrane protein LipL32 is a lipoprotein expressed during mammalian infection. *Infect Immun,* 73: 4853-4863.

Haake, D.A., Mazel, M.K, McCoy, A.M., Milward, F. & Chao, G. (1999) Leptospiral outer membrane proteins OmpL1 and LipL41 exhibit synergistic immunoprotection. *Infect Immun,* 67: 6572-6582.

Holland, C. J., Cole, D. K., & Godkin, A. (2013). Re-Directing CD4(+) T Cell Responses with the Flanking Residues of MHC Class II-Bound Peptides: The Core is Not Enough. *Front Immunol, 4*, 172. doi: 10.3389/fimmu.2013.00172

Janeway, C.A., Travers, P., Walport, M. & Capra, J.D. (1999) "The recognition of antigen," in *Immunology: The Immune System in Health and Disease*, pp.79-194.

Koizumi, N. & Watanabe, H. (2004) Leptospiral immunoglobulin-like proteins elicit protective immunity. *Vaccine*, 22, 1545-1552.

Lacerda, M., Scheffler, K., & Seoighe, C. (2010). Epitope Discovery with Phylogenetic Hidden Markov Models. *Molecular Biology and Evolution, 27*(5), 1212-1220. doi: 10.1093/molbev/msq008

Laskowski, R.A., Macarthur, M.W., Moss, D.S. & Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26: 283-291.

Levett, P. N. (2001). Leptospirosis. *Clin Microbiol Rev, 14*(2), 296-326. doi: 10.1128/CMR.14.2.296-326.2001

Lin, Y. & Change, Y.F. (2007) A domain of the Leptospira LigB contributes to high affinity binding of fibronectin. *Biochem Biophys Res Commun.* 362, 443-440.

Lingappa, J., Kuffner, T., Tappero, J., Whitworth, W., Mize, A., Kaiser, R., & McNicholl, J. (2004). HLA-DQ6 and ingestion of contaminated water: possible gene-environment interaction in an outbreak of Leptospirosis. *Genes Immun, 5*(3), 197-202. doi: 10.1038/sj.gene.6364058

Maupetit, J., Derreumaux, P. & Tufféry, P. (2010) A fast and accurate method for large-scale de novo peptide structure prediction. *J Comput Chem*, 31-726-38.

Maupetit, J., Derreumaux, P. & Tufféry, P. (2009) PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res*, 37 (Web Server issue): W498-503, doi:10.1093/nar/gkp323

Matsunaga, J., Barocchi, M.A., Croda, J., Young, T.A., Sanchez, Y., Siqueira, I., Bolin, C.A., Reis, M.G., Riley, L.W., Haake, D.A. & Ko, A.I. (2003) Pathogenic Leptospira species express surface-exposed proteins belonging to the bacterial immunoglobulin superfamily. *Mol Microbiol*, 49(4): 929-945.

Mbow, M.L., De Gregorio, E, Valiente, N.M. & Rappuoli, R. (2010) New adjuvants for human vaccines. *Curr Opin Immunol,* 22: 411-416.

McBride, A. J., Cerqueira, G. M., Suchard, M. A., Moreira, A. N., Zuerner, R. L., Reis, M. G., Haake, D. A., Ko, A. I. & Dellagostin, O. A. (2009). Genetic diversity of the Leptospiral immunoglobulin-like (Lig) genes in pathogenic Leptospira spp. *Infect Genet Evol, 9*(2), 196-205. doi: 10.1016/j.meegid.2008.10.012

Miller, D.A., Wilson, M.A. & Beran, G.W. (1991) Relationshops between prevalence of Leptospira interrogans in cattle, and regional, climatic, and seasonal factors. A*merican Journal of Veterinary Research,* 52: 1766-1768.

Muh, H. C., Tong, J. C., & Tammi, M. T. (2009). AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. *PLoS One, 4*(6), e5861. doi: 10.1371/journal.pone.0005861

Nardin, G.A., Oliveira, G.A., Calvo-Calle, J.M. & Nussenzweig, R.S. (1995) The use of multiple antigen peptides in the analysis and induction protective immune responses against infectious diseases. *Adv. Immunol,* 60, 105-149.

Nascimento, A. L. T. O., Ko, A. I., Martins, E. A. L., Monteiro-Vitorello, C. B., Ho, P. L., Haake, D. A., Verjovski-Almeida, S., Hartskeerl, R. A., Marques, M. V., Oliveira, M. C., Menck, C. F. M., Leite, L. C. C., Carrer, H., Coutinho, L. L., Degrave, W. M., Dellagostin, O. A., El-Dorry, H., Ferro, E. S., Ferro, M. I. T., Furlan, L. R., Gamberini, M., Giglioti, E. A., Goes-Neto, A., Goldman, G. H., Goldman, M. H. S., Harakava, R., Jeronimo, S. M. B., Junqueira-de-Azevedo, I. L. M., Kimura, E. T., Kuramae, E. E., Lemos, E. G. M., Lemos, M. V. F., Marino, C. L., Nunes, L. R., de Oliveira, R. C., Pereira, G. G., Reis, M. S., Schriefer, A., Siqueira, W. J., Sommer, P., Tsai, S. M., Simpson, A. J. G., Ferro, J. A., Camargo, L. E. A., Kitajima, J. P., Setubal, J. C. & Van Sluys, M. A. (2004). Comparative Genomics of Two Leptospira interrogans Serovars Reveals Novel Insights into Physiology and Pathogenesis. *Journal of Bacteriology, 186*(7), 2164-2172. doi: 10.1128/jb.186.7.2164-2172.2004

Nielsen, M., Lundegaard, C., & Lund, O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics, 8*, 238. doi: 10.1186/1471-2105-8-238

Nielsen, M., & Lund, O. (2009). NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics, 10*, 296. doi: 10.1186/1471-2105-10-296

Nielsen, M., Lund, O., Buus, S., & Lundegaard, C. (2010). MHC class II epitope predictive algorithms. *Immunology, 130*(3), 319-328. doi: 10.1111/j.1365-2567.2010.03268.x

Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biol, 3*(1), 120139. doi: 10.1098/rsob.120139

Ptak, C. P., Hsieh, C. L., Lin, Y. P., Maltsev, A. S., Raman, R., Sharma, Y., Oswald, R. E. & Chang, Y. F. (2014). NMR solution structure of the terminal immunoglobulin-like domain from the leptospira host-interacting outer membrane protein, LigB. *Biochemistry, 53*(32), 5249-5260. doi: 10.1021/bi500669u

Raja, V., & Natarajaseenivasan, K. (2013). Pathogenic, diagnostic and vaccine potential of leptospiral outer membrane proteins (OMPs). *Crit Rev Microbiol*. doi: 10.3109/1040841X.2013.787387

Rajkumar Soni, A. M. K., & Archana Tiwari, A. G. (2013). Reverse Vaccinology: Basics and Applications. *Journal of Vaccines & Vaccination, 04*(06). doi: 10.4172/2157-7560.1000194

Ricaldi, J. N., Fouts, D. E., Selengut, J. D., Harkins, D. M., Patra, K. P., Moreno, A., Lehmann, J. S., Purushe, J., Sanka, R., Torres, M., Webster, N. J., Vinetz, J. M. & Matthias, M. A. (2012). Whole genome analysis of Leptospira licerasiae provides insight into leptospiral evolution and pathogenicity. *PLoS Negl Trop Dis, 6*(10), e1853. doi: 10.1371/journal.pntd.0001853

Sakib, M.S., Islam, M.R., Mahbub Hasan, A.K.M. & Nurun Nabi, A.H.M. (2014) Prediction of epitope-based peptides for the utility of vaccine development from fusion and glycoprotein of Nipah virus using in silico approach. *Advances in Bioinformatics,* Volume 2014, Article ID 402492, http://dx.doi.org/10.1155/2014/402492

Shiina, T., Hosomichi, K., Inoko, H. & Kulski J.K. (2009) The HLA genomic loci map: expression, interaction, diversity, and disease. *Journal of Human Genetics,* 54: 15-39.

Silva, E.F., Medeiros, M.A., McBride, A.J.A, Matsunaga, J, Esteves, G.S., Ramos, J.G.R., Santos, C.S., Croda, J., Homma, A., Dellagostin, O.A., Haake, D.A., Reis, M.G. & Ko, A.I. (2007) LigA confers protective immunity against lethal infection in the hamster model of leptospirosis. *Vaccine.* 25(33): 6277-6286.

Sette, A., Vitiello, A., Reherman, B., Fowler, P., Nayersina, R., Kast, W.M., Melief, C.J., Oseroff, C., Yuan, L. & Ruppert, J. (1994) The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol,* 153: 5586-5592.

The UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. 42: D191-D198.

Thévenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P., & Tufféry, P. (2012) PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res*, 40, W288-293.

Tovchigrechko, A., & Vakser, I. A. (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res, 34*(Web Server issue), W310-314. doi: 10.1093/nar/gkl206

Verma, A., Brissette, C.A., Bowman, A.A., Shah, S.T. & Zipfel, P.F. (2010) Leptospiral endostatin-like protein A is a bacterial cell surface receptor for human plasminogen. *Infect Immun,* 78: 2053-2059.

Wang, L., Harcourt, B.H. & Yu, M. (2001) "Molecular biology of Hendra and Nipah viruses," vol.3, no.4, pp.279-287.

Wang, Z., Jin, L., & Wegrzyn, A. (2007). Leptospirosis vaccines. *Microb Cell Fact, 6*, 39. doi: 10.1186/1475-2859-6-39

Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 8(2): 127-134. doi:10.1093/protein/8.2.127

Weenink, S.M. & Gautam, A.M. (1996) Antigen presentation by MHC class II molecules. *Immmunology and Cell Biology,* 75: 69-81.

APPENDIX

```
# Peptide length 15
        NetMHCII version 2.2.

Strong binder threshold  50.00. Weak binder threshold 500.00.
```

Allele	pos	peptide	core	1-log50k(aff)	affinity(nM)	%Rank	Identity	Bind Level
HLA-DPA10103-DPB10401	1	IFSIFLSLFFQGCMV	FLSLFFQGC	0.8702	4.1	0.30	Q04YJ5	SB
HLA-DPA10103-DPB10401	2	KIFSIFLSLFFQGCM	FLSLFFQGC	0.8650	4.3	0.30	Q04YJ5	SB
HLA-DPA10103-DPB10401	3	FSIFLSLFFQGCMVW	FLSLFFQGC	0.8628	4.4	0.30	Q04YJ5	SB
HLA-DPA10103-DPB10401	4	SIFLSLFFQGCMVWP	FLSLFFQGC	0.8624	4.4	0.30	Q04YJ5	SB
HLA-DPA10103-DPB10401	5	IFLSLFFQGCMVWPL	FLSLFFQGC	0.8603	4.5	0.30	Q04YJ5	SB
HLA-DPA10103-DPB10401	6	KKIFSIFLSLFFQGC	FLSLFFQGC	0.8543	4.8	0.30	Q04YJ5	SB
HLA-DPA10103-DPB10401	7	FLSLFFQGCMVWPLV	FQGCMVWPL	0.8267	6.5	0.50	Q04YJ5	SB
HLA-DPA10103-DPB10401	8	LSLFFQGCMVWPLVV	FQGCMVWPL	0.7728	11.7	1.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	9	IDSLFVFKGTLYAAN	FKGTLYAAN	0.7594	13.5	1.10	Q04YJ5	SB
HLA-DPA10103-DPB10401	10	SLFFQGCMVWPLVVG	FQGCMVWPL	0.7498	15.0	1.25	Q04YJ5	SB
HLA-DPA10103-DPB10401	11	DSLFVFKGTLYAANG	FKGTLYAAN	0.7484	15.2	1.25	Q04YJ5	SB
HLA-DPA10103-DPB10401	12	DSSSLVGSCFDNADF	LVGSCFDNA	0.7431	16.1	1.50	Q04YJ5	SB
HLA-DPA10103-DPB10401	13	SSSLVGSCFDNADFK	LVGSCFDNA	0.7376	17.1	1.50	Q04YJ5	SB
HLA-DPA10103-DPB10401	14	SLFVFKGTLYAANGG	FKGTLYAAN	0.7363	17.3	1.50	Q04YJ5	SB
HLA-DPA10103-DPB10401	15	LFVFKGTLYAANGGF	FKGTLYAAN	0.7307	18.4	1.50	Q04YJ5	SB
HLA-DPA10103-DPB10401	16	LFFQGCMVWPLVVGA	FQGCMVWPL	0.7297	18.6	1.50	Q04YJ5	SB
HLA-DPA10103-DPB10401	17	VVDSSSLVGSCFDNA	LVGSCFDNA	0.7294	18.7	1.50	Q04YJ5	SB
HLA-DPA10103-DPB10401	18	VDSSSLVGSCFDNAD	LVGSCFDNA	0.7261	19.4	2.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	19	SSLVGSCFDNADFKK	LVGSCFDNA	0.7147	21.9	2.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	20	FVFKGTLYAANGGFP	FKGTLYAAN	0.6997	25.8	2.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	21	FFQGCMVWPLVVGAV	FQGCMVWPL	0.6831	30.8	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	22	SLVGSCFDNADFKKN	LVGSCFDNA	0.6729	34.4	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	23	TAIRIVYSEFVNNKE	IVYSEFVNN	0.6719	34.8	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	24	PTAIRIVYSEFVNNK	IVYSEFVNN	0.6697	35.6	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	25	MKKIFSIFLSLFFQG	IFSIFLSLF	0.6660	37.1	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	26	AIRIVYSEFVNNKEA	IVYSEFVNN	0.6643	37.8	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	27	SPTAIRIVYSEFVNN	IVYSEFVNN	0.6580	40.5	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	28	VFKGTLYAANGGFPN	FKGTLYAAN	0.6413	48.5	4.00	Q04YJ5	SB
HLA-DPA10103-DPB10401	29	IRIVYSEFVNNKEAL	IVYSEFVNN	0.6372	50.7	4.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	30	RIVYSEFVNNKEALK	IVYSEFVNN	0.6006	75.3	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	31	TSKTKFFAVGMYSDG	FAVGMYSDG	0.5855	88.6	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	32	STSKTKFFAVGMYSD	FAVGMYSDX	0.5794	94.7	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	33	LVTWFSSEPLVADAI	FSSEPLVAD	0.5781	96.1	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	34	SKTKFFAVGMYSDGT	FAVGMYSDG	0.5691	105.9	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	35	AELVSIQVSPYVSVT	IQVSPYVSV	0.5664	109.1	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	36	ELVSIQVSPYVSVTP	IQVSPYVSV	0.5650	110.7	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	37	KTKFFAVGMYSDGTK	FAVGMYSDG	0.5649	110.7	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	38	LIPADKAFSQFAEFN	KAFSQFAEF	0.5633	112.8	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	39	NLVTWFSSEPLVADA	FSSEPLVAD	0.5590	118.2	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	40	DLIPADKAFSQFAEF	KAFSQFAEF	0.5559	122.1	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	41	TNLVTWFSSEPLVAD	FSSEPLVAD	0.5545	123.9	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	42	EAELVSIQVSPYVSV	IQVSPYVSV	0.5545	124.0	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	43	LVSIQVSPYVSVTPK	IQVSPYVSV	0.5541	124.5	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	44	VTWFSSEPLVADAIN	FSSEPLVAD	0.5528	126.2	8.00	Q04YJ5	WB
HLA-DPA10103-DPB10401	45	TKFFAVGMYSDGTKK	FAVGMYSDG	0.5516	128.0	8.00	Q04YJ5	WB

A snapshot of output given by NetMHCII v.2.2 server. The 15-mer peptides/ T cell epitopes (column 3), the corresponding HLA-DP alleles (column 1) and the affinity score (column 6) were used in the analysis. Only short binders were considered in the study. Similar outputs were also generated for HLA-DQ alleles and HLA-DR alleles.

```
awk '{print $1, $3, $6}' epitopes_dp.txt > strong_binders_dp.txt
awk '{print $1, $3, $6}' epitopes_dq.txt > strong_binders_dq.txt
awk '{print $1, $3, $6}' epitopes_dr.txt > strong_binders_dr.txt
paste strong_binders_dp.txt strong_binders_dq.txt strong_binders_dr.txt > all_epitopes.txt
```

Few Shell scripts used to process the raw outputs generated by NetMHCII v.2.2