

Immunoinformatics approaches towards developing vaccine/therapeutic agent against the Middle East Respiratory Syndrome Coronavirus by targeting nsp3 protein



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

Submitted by

Niaj Mohammad Tanvir

Student ID: 13326010

Microbiology Program

Department of Mathematics and Natural Sciences

BRAC University, Bangladesh

Submitted on

March 28, 2019

DECLARATION

I hereby solemnly declare that the research work embodying the results reported in this thesis entitled **“Immunoinformatics approaches towards developing vaccine/therapeutic agent against the Middle East Respiratory Syndrome Coronavirus by targeting nsp3 protein”** submitted by the undersigned has been carried out under the supervision of Dr. M. Mahboob Hossain, Professor, Department of Mathematics and Natural Sciences, BRAC University, Dhaka. It is further declared that the research work presented here is original and any part of this thesis has not been submitted to any other institution for any degree or diploma.

Author:

(Niaj Mohammad Tanvir)

Candidate

Certified:

(Dr. M. Mahboob Hossain)

Supervisor

Professor, Microbiology Program, Department of Mathematics and Natural Sciences

BRAC University, Dhaka

ACKNOWLEDGEMENTS

I want to convey my deepest appreciation and special thanks to Professor A.F.M Yusuf Haider, the Chairperson of the Department of Mathematics and Natural Sciences, late Professor A. A. Ziauddin Ahmed, former Chairperson of the Department of Mathematics and Natural Sciences and Professor Naiyyum Choudhury, former coordinator of the Biotechnology and Microbiology Program, for giving me their outstanding guidance and support during my study at BRAC University. I am grateful to all the faculty members and respective lab officers of the Department of Mathematics and Natural Sciences for their incredible support and valuable teachings throughout the period of my bachelor's studies. I would also thank the authority and management of BRAC University for continuously providing me with many opportunities and facilities.

I am truly indebted to my supervisor, Professor Dr. M. Mahboob Hossain, Department of Mathematics and Natural Sciences, BRAC University, for his constant supervision, constructive criticism, exemplary guidance, encouragement and finally for believing in me throughout the entire period of my research work. I am also indebted to Dr. Khademul Islam, Associate Professor, Department of Genetic Engineering and Biotechnology, Dhaka University for sparing his precious time to provide me with valuable insights regarding some technical aspects of this project.

I consider this thesis, my very first research work, as an important milestone in my career. I will strive to implement the knowledge I have gained in the best possible way and I will continue to work on the improvement of the idea that I carry.

Sincerely,

Niaj Mohammad Tanvir

Department of Mathematics and Natural Sciences, BRAC University

ABSTRACT

MERS coronavirus is an emerging virus which causes Middle East respiratory syndrome. MERS outbreak has been appearing in 27 different countries over a span of around six years with a mortality rate greater than 35% and claimed over 800 lives in the process. However, there is not any clinically approved vaccine or therapeutic agent available for treatment of MERS. Therefore, it is crucial to design an effective vaccine or therapeutic agents against MERS coronavirus. This study aimed to find vaccines/therapeutic agents against MERS coronavirus using immunoinformatics which could reduce both time and cost needed for laboratory analysis and vaccine development. Since nsp3 protein is an essential component of the replication/transcription complex of MERS coronavirus, the discovery of an nsp3 inhibitor will be a major leap towards developing an anti-viral agent that can interfere with MERS coronavirus replication. In the present study, two different strategies were explored. The first strategy was to design an epitope-based vaccine. For designing an epitope-based vaccine, nsp3 protein sequence was extracted from the NCBI database and then the sequence was put in T-cell and B-cell epitope prediction servers to generate a list of potential T-cell and B-cell epitope candidates. T-cell and B-cell epitope candidates are then screened using several software and tools. FAFETGLAY appeared to be the best T-cell epitope candidate. However, the only drawback was that FAFETGLAY was found as an allergen in allergenicity prediction tools. FVDWRSYNYAVSSAFWLF, LKFKEVCKTTTGIPEY and LKFKEVCKTTTGIPEYNF showed promise as epitope candidates for peptide-based vaccine design among the selected B-cell epitopes. The second strategy was focused on identifying effective flavonoids that can be used as nsp3 inhibiting therapeutic agents against MERS coronavirus. In this study, 18 flavonoids were selected as potential nsp3 inhibitor candidates and their anti-viral activities were assessed using molecular docking study. Molecular docking study revealed that among 18 flavonoids, apiin and naringin exhibited the most potent antiviral activity against MERS coronavirus nsp3 protein as they showed the best binding affinity of -10.1 kcal/mol which was higher than the binding affinity of ADP-ribose. In addition, apiin and naringin had the lowest K_i value of 0.0390649 μM . Furthermore, the molecular visualization of the docked complexes suggested that both apiin and naringin formed three or more hydrogen bonds ranging from moderate to weak. Therefore, apiin and naringin can be considered good candidates for further evaluation as potential anti-viral agents against MERS-CoV.

Table of Contents

SL	Content	Page
	Declaration	i
	Acknowledgements	ii
	Abstract	iii
	Table of contents	iv
	List of Tables	viii
	List of Abbreviations and Symbols	ix
1	Introduction & Literature Review	1
1.1	Genome Structure and Function of MERS-CoV	2
1.2	nsp3 protein of MERS-CoV	7
1.3	Major Histocompatibility complex (MHC)	8
1.4	MERS-CoV: Epidemiology	8
1.5	MERS-CoV: Pathogenesis, Pathology and Immunity	10
1.6	Vaccine Development and Treatment	11
1.7	Epitope-based Vaccine	12
1.8	Advantages of Epitope-based Vaccine	12
1.9	Flavonoids: Potential Anti-viral Agents against MERS-CoV	13
1.10	Aims and Objectives	14
2	Materials and Methods	15
2.1	Method Summary	16
2.1.1	In silico Analysis of Epitope-based Vaccine Candidates	18
2.1.2	Assessing Potential Anti-Viral Activity of Selected Flavonoids against MERS-CoV using Immunoinformatics	19
2.2	Databases used for obtaining data	20
2.2.1	NCBI	20
2.2.2	PDB	20
2.3	Software and tools used for analysis	20
2.3.1	UCSF chimera 1.13	20

SL	Content	Page
2.3.2	PyRx	20
2.3.3	VaxiJen 2.0	21
2.3.4	BepiPred 2.0	21
2.3.5	BCPPREDS	21
2.3.6	IEDB tools	21
2.3.6.1	Chou & Fasman Beta-Turn Prediction	22
2.3.6.2	Emini surface accessibility prediction tool:	22
2.2.6.3	Karplus and Schulz Flexibility prediction tool:	22
2.3.6.4	Kolaskar and Tongaonkar antigenicity prediction tool:	22
2.3.6.5	Parker Hydrophilicity prediction tool:	22
2.3.6.6	Epitope Conservancy Analysis	23
2.3.6.7	Proteasomal cleavage/TAP transport/MHC class I combined predictor	23
2.3.6.8	T-cell class I pMHC immunogenicity predictor	23
2.3.6.9	Peptide binding to MHC class II molecules predictor	23
2.3.6.10	Population Coverage	23
2.3.7	NetCTL 1.2 Server	23
2.3.8	NetMHC 4.0 Server	24
2.3.9	ToxinPred	24
2.3.10	AllergenFP v1.0	24
2.3.11	AllerTOP v2.0	24
2.3.12	PEP-FOLD 2.0	24
3	Results	25
3.1	Result Summary	26
3.2	Sequence retrieval	29
3.3	VaxiJen result for nsp3 Protein: Primary result of antigenicity	30
3.4	Prediction of T-cell epitopes	31
3.4.1	Prediction of CD8+ T-cell epitopes using NetCTL 1.2 Server	31
3.4.2	T-cell Epitope candidate screening using VaxiJen 2.0 and IEDB Immunogenicity Tool	31

SL	Content	Page
3.4.3	Prediction of peptide-MHC class I binding using Proteasomal cleavage/TAP transport/MHC class I combined predictor	34
3.4.4	Prediction of peptide-MHC class I binding using NetMHC 4.0 Server	35
3.4.5	Prediction of peptide-MHC class II binding using IEDB Peptide binding to MHC class II molecules predictor	37
3.4.6	Population Coverage Analysis	44
3.4.7	Conservancy and Toxicity Prediction	45
3.4.8	Allergenicity Prediction	46
3.4.9	Docking result and Analysis	46
3.5	Prediction of B-Cell epitopes	48
3.5.1	BCPREDS	48
3.5.2	BepiPred 2.0	55
3.5.3	Screening epitope candidates using VaxiJen 2.0	57
3.5.4	Overlapping Sequence Identification of B-cell epitopes and T-cell Epitopes	62
3.5.5	Conservancy Analysis	63
3.5.6	Checking Epitopes as Ideal Vaccine Candidates	63
3.5.6.1	Analysis of IFVDWRSYNYAVSS	63
3.5.6.2	Analysis of FVDWRSYNYAVS	65
3.5.6.3	Analysis of FVDWRSYNYAVSSAFW	65
3.5.6.4	Analysis of FVDWRSYNYAVSSAFWLF	65
3.5.6.5	Analysis of LKFKEVCKTTTGIPEYNF	66
3.5.6.6	Analysis of LKFKEVCKTTTGIPEY	66
3.5.6.7	Analysis of FKEVCKTTTGIPEYNFIIYD	67
3.5.6.8	Analysis of VCKTTTGIPEYN	67
3.5.7	Toxicity and Allergenicity Analysis	67
3.6	Assessing Potential Anti-Viral Activity of Selected Flavonoids against MERS-CoV	68
3.6.1	Ligand Selection	68

SL	Content	Page
3.6.2	Macromolecule selection	69
3.6.3	Macromolecule and Ligand preparation	70
3.6.4	Molecular Docking using PyRx	72
3.6.5	Measuring Ki Value	79
3.6.6	Docking Visualization Analysis using UCSF Chimera 1.13	80
4	Discussion	89
5	References	99

List of Tables

SL	Title	Page
1.1	Functions of MERS-CoV polypeptides	5
3.1	Analysis of predicted T-cell epitope candidates using VaxiJen 2.0	32
3.2	Predicted T-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0	33
3.3	Analysis of predicted T-cell epitope candidates using IEDB T-cell class I pMHC immunogenicity predictor	34
3.4	Predicted T-cell epitope candidates having a positive score in IEDB T-cell class I pMHC immunogenicity predictor	34
3.5	Prediction of peptide-MHC class I binding of selected T-cell epitope candidates using Proteasomal cleavage/TAP transport/MHC class I combined predictor	35
3.6	Prediction of peptide-MHC class I binding (%Rank<2.0) of selected T-cell epitope candidates using NetMHC 4.0 Server	36
3.7	Prediction of peptide-MHC class II binding (ic50<3000) using IEDB Peptide binding to MHC class II molecules predictor	37
3.8	Allergenicity prediction using AllergenFP v1.0 and AllerTOP v2.0	46
3.9	Binding affinity of best docking pose against HLA-A*01:01 allele	47
3.10	Predicted B-cell epitopes using BCPREDS	49
3.11	Predicted B-cell epitopes using BepiPred 2.0	55
3.12	Predicted B-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0	57
3.13	B-cell and T-cell Epitopes having overlapping Sequence	62
3.14	Toxicity and allergenicity prediction of selected B-cell epitopes	68
3.15	Selected flavonoids and their Pubchem ID	68
3.16	Macromolecule Minimization parameter	72
3.17	Binding affinity of re-docked ADP-ribose against MERS-CoV nsp3 macro domain	73
3.18	Binding affinity of flavonoid compounds against MERS-CoV nsp3 macro domain	74
3.19	Binding affinity of best docking pose against MERS-CoV nsp3 macro domain	79
3.20	Ki value calculation for each flavonoid compound	80
3.21	Intermolecular H bond between each compound with the nsp3 macro domain	80

List of Abbreviations and Symbols

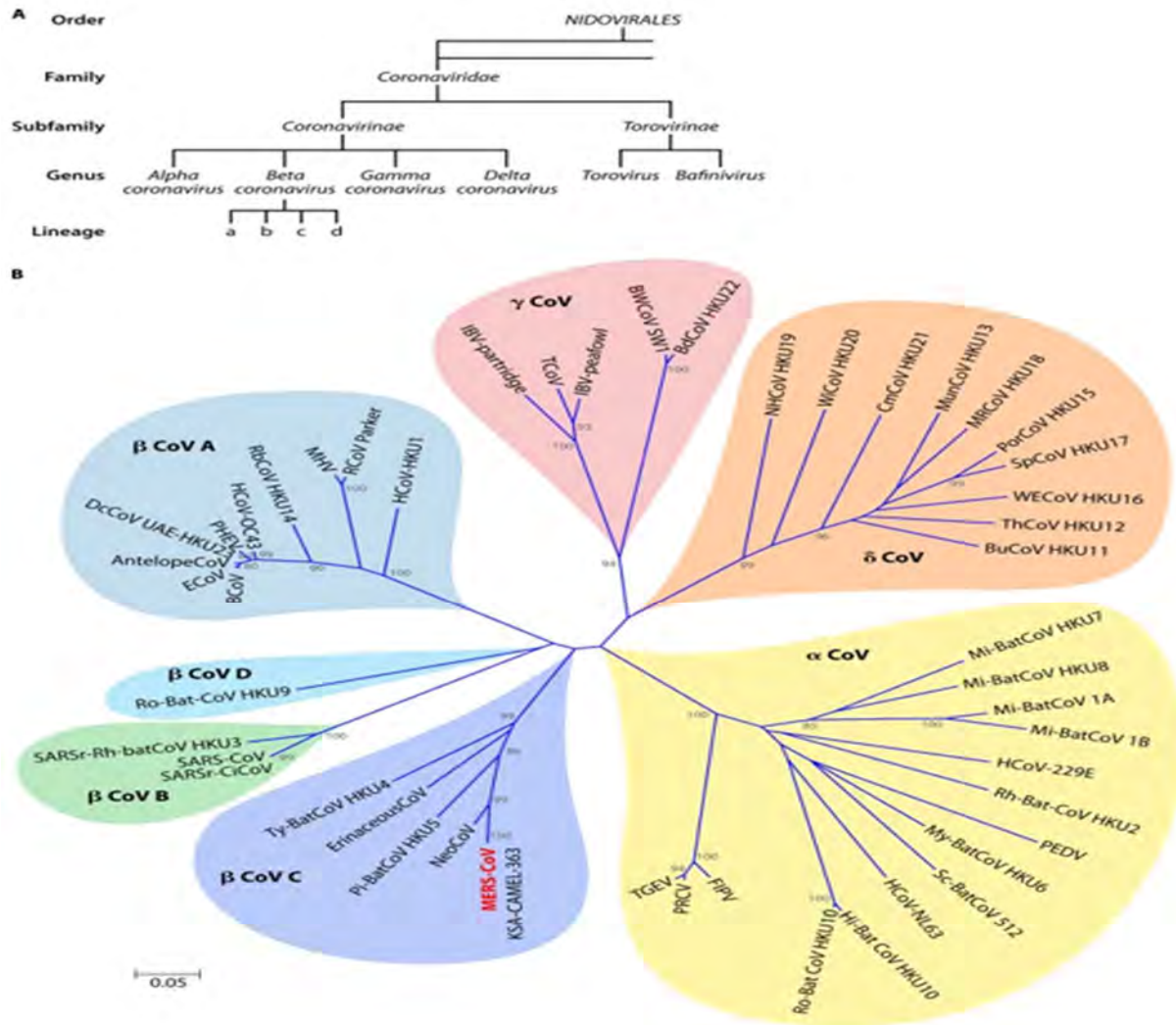
Short Form	Abbreviation/Meaning
MERS	Middle East Respiratory Syndrome
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
ORF	Open Reading Frame
IFN	Interferon
Nsp3	Non-Structural Protein 3
MHC	Major Histocompatibility Complex
HLA	Human Leukocyte Antigen
WHO	World Health Organization
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
NCBI	National Centre for Biotechnology Information
PDB	Protein Data Bank
RCSB	Research Collaboratory for Structural Bioinformatics
IEDB	Immune Epitope Database and Analysis Resources
SDF	Structure-Data File
ADP-ribose	Adenosine Diphosphate Ribose
IC50	Half Maximal Inhibitory Concentration
RMSD	Root-Mean-Square Deviation
H-bonds	Hydrogen Bonds

Introduction & Literature Review

The Middle East is a region centered on most of western Asia, Turkey and Egypt containing 18 countries. The history of the Middle East dates back to ancient times and even now it is one of the busiest politico-economic centers in the world. The Middle East exhibits many unique religious and cultural practices because many major religions which are still practiced today throughout the world originated in this region. The Middle East comprises a vast number of ethnic groups with an estimated population of over 411 million as of 2016. In addition to that, people living in this densely populated region have relied on camels for food and transportation for ages. These distinct regional features have provided favorable conditions for new emerging viruses such as Middle East respiratory syndrome coronavirus to appear. Middle East respiratory syndrome coronavirus (MERS-CoV) has originated from animal reservoirs and crossed interspecies barriers to infect humans and caused a severe outbreak of respiratory infection in the Middle East since 2012 and has spread to Europe, Africa, Asia, and North America. The disease caused by MERS-CoV is known as the Middle East respiratory syndrome (MERS). It was originally reported as a “SARS-like” infection because unlike other human CoVs infections which cause only mild upper respiratory tract infections such as the common cold, MERS causes lower respiratory tract infection which is often fatal. MERS-CoV is listed as Category C Priority Pathogen in NIAID’s pathogen priority list and this virus is considered to be a potential pandemic threat due to person-to-person transmission capability and lack of effective drugs. In addition, MERS epidemic has kept on appearing in different countries for several years with a mortality rate greater than 35% (Chafekar et al., 2018; Chan et al., 2015). As of November 30, 2018, the total number of laboratory-confirmed MERS-CoV cases reported globally to WHO is 2274 with 806 associated deaths.

1.1 Genome Structure and Function of MERS-CoV

MERS-CoV is a member of the family Coronaviridae which comprises enveloped single-stranded RNA viruses. The Coronaviridae family is divided into four genera based on phylogenetic clustering: alpha, beta, gamma, and delta coronaviruses. These genera are further subdivided into distinct lineages. MERS-CoV is a lineage C β -Coronavirus. It has a positive-sense single-stranded RNA (ssRNA) genome about 30-kb in size (Chafekar et al., 2018).

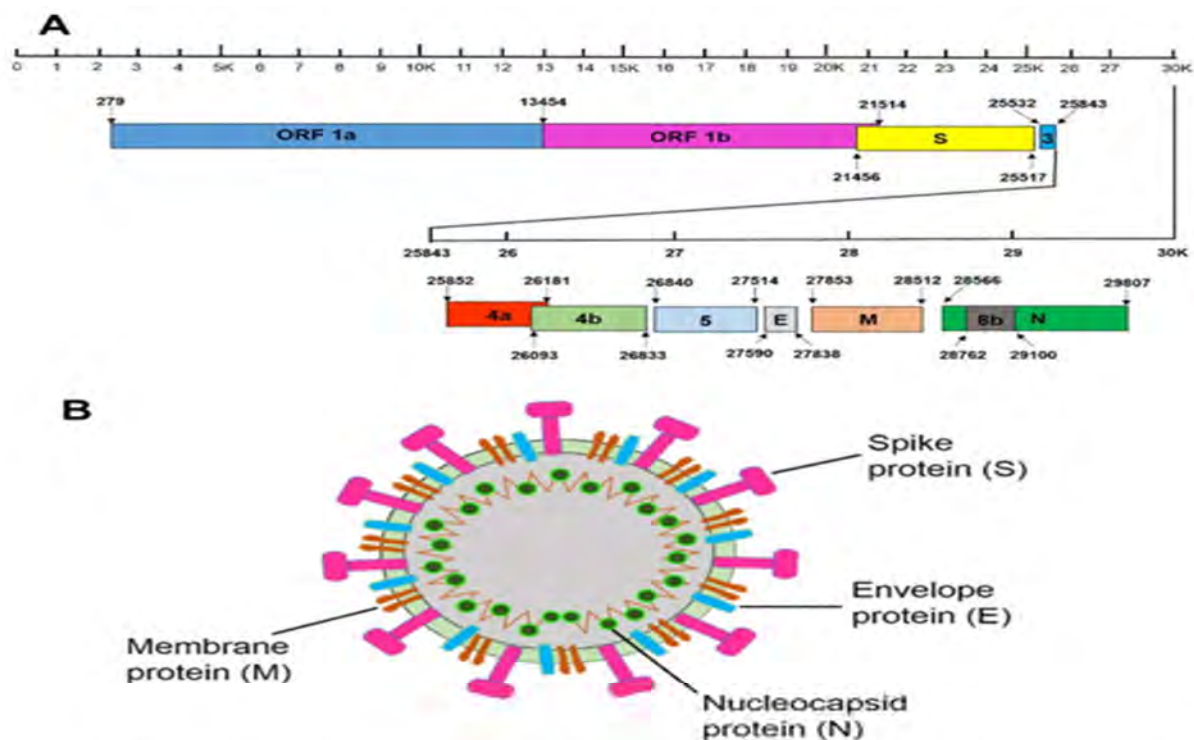


(A) Taxonomy of Coronaviridae according to the International Committee on Taxonomy of Viruses.
 (B) Phylogenetic tree of 50 coronaviruses with partial nucleotide sequences of RNA-dependent RNA polymerase constructed by the neighbor-joining method using MEGA 5.0.

Figure 1.1: Taxonomy of Coronaviridae (Chan et al., 2015)

182 full-length genomes or multiple concatenated genome fragments have been analyzed as of 2016. Among these genomes/genome fragments, 94 are from humans and 88 are from dromedary camels. MERS-CoV genomes share more than 99% sequence identity suggesting low mutation rate and low variance among the genomes. MERS-CoV genomes are roughly divided into two clades: clade A and clade B. Clade A contains only a few strains whereas clade B contains most strains. (Chafekar et al., 2018)

The MERS-CoV genome consists of 11 open reading frames (ORFs) (Boheemen et al., 2012). The first 5' two-thirds of the MERS-CoV genome encodes the replicase complex (ORF1a and ORF1b) whereas the remaining 3' one-third encodes the structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as five accessory proteins (ORF3, ORF4a, ORF4b, ORF5, and ORF8b). These accessory proteins are not required for genome replication but are likely involved in pathogenesis as recent studies by reverse genetics demonstrated that the absence of the genes encoding these proteins as a group may attenuate viral titers and these accessory proteins do not share homology with any known host or virus protein, apart from those of its closely related lineage C β CoVs (Chafekar et al., 2018; Zhang et al., 2014). The flanking regions of the MERS-CoV genome contain UTR regions (Chafekar et al., 2018).



(A) MERS-CoV genomic structure. Viral genes, including ORF1a, ORF1b, S, 3, 4a, 4b, 5, E, M, 8b and N (GenBank accession number: [JX869059](https://www.ncbi.nlm.nih.gov/nuccore/JX869059)) and their respective lengths, are indicated by rectangular boxes in the scheme.

(B) Schematic diagram of MERS-CoV structure. MERS-CoV contains four structural proteins. The S protein is a type I transmembrane glycoprotein displayed on the surface of viral membrane as an oligomer. The E protein is also a transmembrane protein which forms an ion channel on the viral surface. The N protein plays an important role in encapsidating the genomic RNA and interacting with the membrane M protein and other N molecules. E: Envelope; M: Membrane; N: Nucleocapsid; MERS-CoV: MERS-coronavirus; S: Spike.

Figure 1.2: Genomic and schematic diagram of MERS-CoV structure (Zhang et al., 2014)

The S protein of MERS-CoV is a heavily glycosylated type I membrane protein that is of paramount importance for attachment to the host receptor. MERS-CoV uses dipeptidyl peptidase 4 (DPP4), a multifunctional 766-amino-acid-long type II transmembrane glycoprotein present at the surface of many different cell types, as the receptor which mediates cell entry (Zhang et al., 2014).

Table 1.1: Functions of MERS-CoV polypeptides (Zhang et al., 2014)

Gene	Encoded polypeptides	Length (nucleotides)	Function
1	1a	13,176	Encodes viral proteases mainly
2	1b	8,061	Encodes RNA polymerase, helicase and ribonucleases mainly
3	S	4,062	Mediates receptor binding and membrane fusion
4	N	1,242	Associates with RNA genome and interacts with C-terminal domain of M protein
5	4b	741	Blocks host interferon production
6	5	675	Interferon antagonist with no effect on interferon beta promoter activation
7	M	660	Incorporates viral components into virions and interacts with the N protein in infected cells
8	8b	339	Not known
9	4a	330	A dsRNA-binding protein with a dsRNA-binding domain (residues 3 to 83) that potently antagonizes host interferon response via inhibition of interferon production (interferon beta promoter activity, IRF-3/7 and NF- κ B activation), ISRE promoter element signaling pathways, and/or suppression of PACT-induced activation of RIG-I and MDA5 in an RNA-dependent manner; not essential for virus replication in Vero A66 and Huh-7 cells
10	3	312	Not known
11	E	249	Ion channel activity

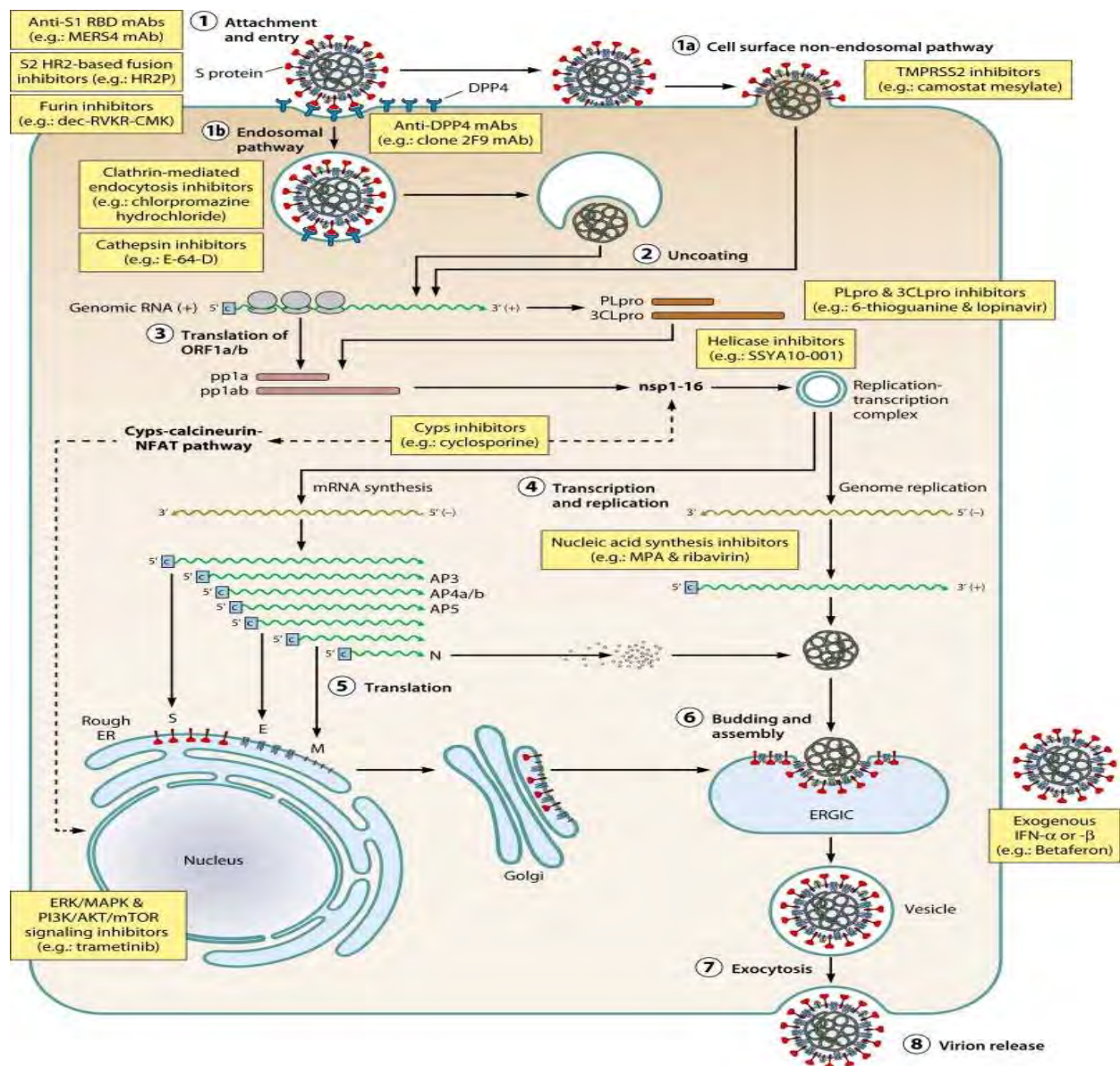


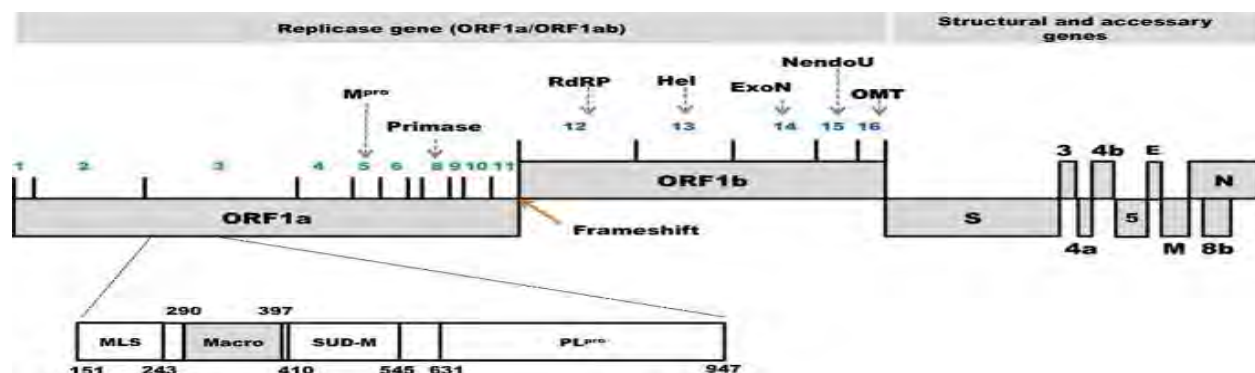
Figure 1.3: MERS-CoV replication strategy (Chan et al., 2015)

The interferons are a family of cytokine mediators that have anti-cancer, anti-proliferative, anti-viral and immunomodulatory functions (Taylor et al., 2013). Interferons alert the cellular immune system and activate immune cells whenever viral infection occurs. The most important immune response to any viral infection is the activation of the type I interferon-mediated innate immune response through the production of type I IFNs (IFN- α and IFN- β) (Chafekar et al., 2018). Viruses evade such host innate immunity by synthesizing IFN antagonist proteins which block one or more key signaling proteins in the IFN and NF- κ B pathways to ensure and enhance viral replication and

pathogenesis (Chafekar et al., 2018; Taylor et al., 2013). MERS-CoV uses these mechanisms to evade the innate immunity of host as well. For instance, MERS-CoV M, ORF4a, ORF4b, and ORF5 proteins are found to be strong IFN antagonists in a study (Yang et al., 2013). In addition to that, these accessory proteins can impede not only type I IFN induction but also NF- κ B signaling pathways (Niemeyer et al., 2013; Matthews et al., 2014). MERS-CoV ORF4a particularly acts as an IFN antagonist by inhibiting both the interferon production (IFN- β promoter activity, IRF-3/7, and NF- κ B activation) and the ISRE promoter element signaling pathways. Moreover, MERS-CoV ORF4b is an enzyme in the 2H-phosphoesterase (2H-PE) family with phosphodiesterase (PDE) activity which can prevent activation of RNase L. Furthermore, MERS-CoV replicase proteins also interfere with the innate immune response signaling pathways through different mechanisms (Chafekar et al., 2018).

1.2 nsp3 protein of MERS-CoV

The MERS-CoV genome comprises 16 non-structural proteins. In the midst of these non-structural proteins, the non-structural protein 3 (nsp3) is the largest protein encoded by the MERS-CoV genome. nsp3 has several domains. Among them, a macro domain is embedded in nsp3 which can bind to adenosine diphosphate ribose (ADP-ribose), an ester formed between the aldehydic carbon of ribose and the terminal phosphate of adenosine diphosphate (Cho et al., 2016).



Schematic diagram of the composition of structural and non-structural proteins (*NSPs*) in MERS-CoV genome. Functional domains of nsp3 are highlighted. NSPs encoded by ORF1a and ORF1b are numbered in green and blue, respectively.

Figure 1.4: Genome organization of MERS-CoV (Cho et al., 2016)

1.3 Major Histocompatibility Complex (MHC)

The portion of antigen molecules which can be specifically bound by the antibody or antigenic receptor of lymphocytes is called an epitope. The size of an epitope is generally equivalent to 5-15 amino acids or 3-4 sugar residues or 6-8 nucleotide. T-cell epitopes are presented by the MHC molecules of antigen presenting cell. There are two types of classes: MHC-I and MHC-II. MHC class II is presented by specialized cell types such as B-cells, macrophages and dendritic cells whereas MHC class I is presented by all nucleated cell bodies. The MHC-I presents peptides of 8-11 amino acids whereas MHC-II present 11-25 amino acids. MHC molecules are the most polymorphic proteins which contain over 6000 classes listed in IMGT/HLA (Patronov et al., 2013). As determining all the peptide binding preferences of alleles in vitro is a very difficult technical challenge, in silico methods are used instead as a time-saving and cost-effective solution for primary virtual screening of potential peptide binding preferences of alleles. T-cell epitope prediction algorithms can screen and predict specific epitopes which can be later produced in vitro to check whether these epitopes are effective or not.

1.4 MERS-CoV: Epidemiology

Even though the first MERS case was reported from Jeddah in September 2012, retrospective studies identified the first case from an outbreak involving 13 persons in March/April 2012 in Zarqa, Jordan (Zumla et al., 2015). Since then, MERS cases have been reported in 27 countries. The largest series of outbreaks occurred in 2014 in Saudi Arabia, when around five hundred hospital-acquired cases appeared throughout the country within a few months. Increased awareness and meticulous focus on infection control measures finally put an end to these outbreaks. Later an outbreak of 186 MERS cases occurred in the Republic of Korea in 2015. This outbreak was the second largest worldwide and the largest reported outside the Middle East region. MERS outbreak in South Korea was halted after approximately 17,000 individuals were quarantined. Among them, 36 patients died from the infection (Fehr et al., 2016). According to the WHO website, the total number of laboratory-confirmed MERS-CoV cases reported globally to WHO is 2274 with case fatality rate around 35.4% as of November 30, 2018.

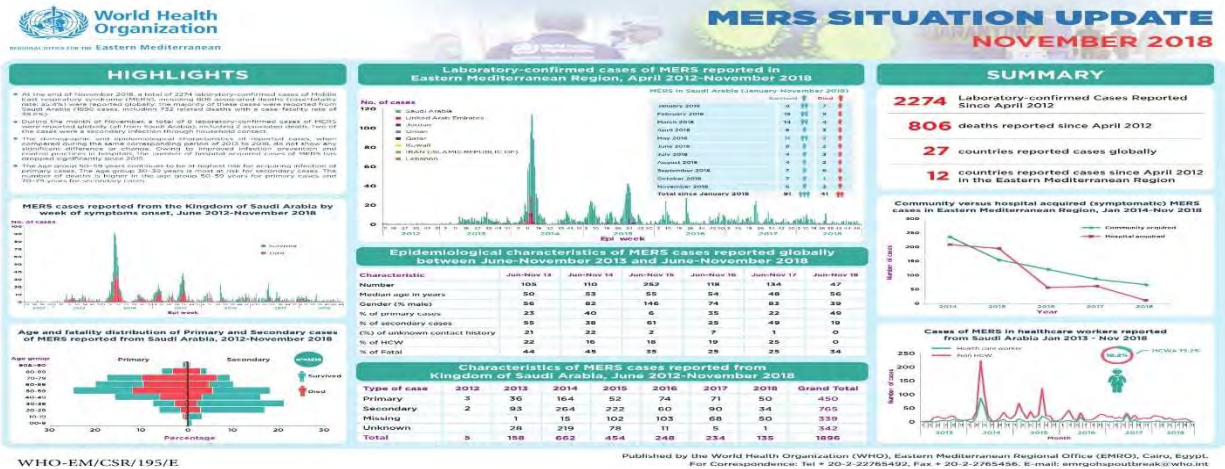


Figure 1.5: MERS situation update according to the World Health Organization (WHO)

Transmission from camels has been linked to human MERS-CoV infection, although very few MERS patients have a history of direct camel exposure. Less direct exposure such as consumption of unpasteurized camel milk may lead to such cases as MERS-CoV RNA has been detected in raw milk collected in the marketplace of Qatar. However, the source of infection in many patients remains undetermined. MERS-CoV may have originally transmitted from bats to camels and other

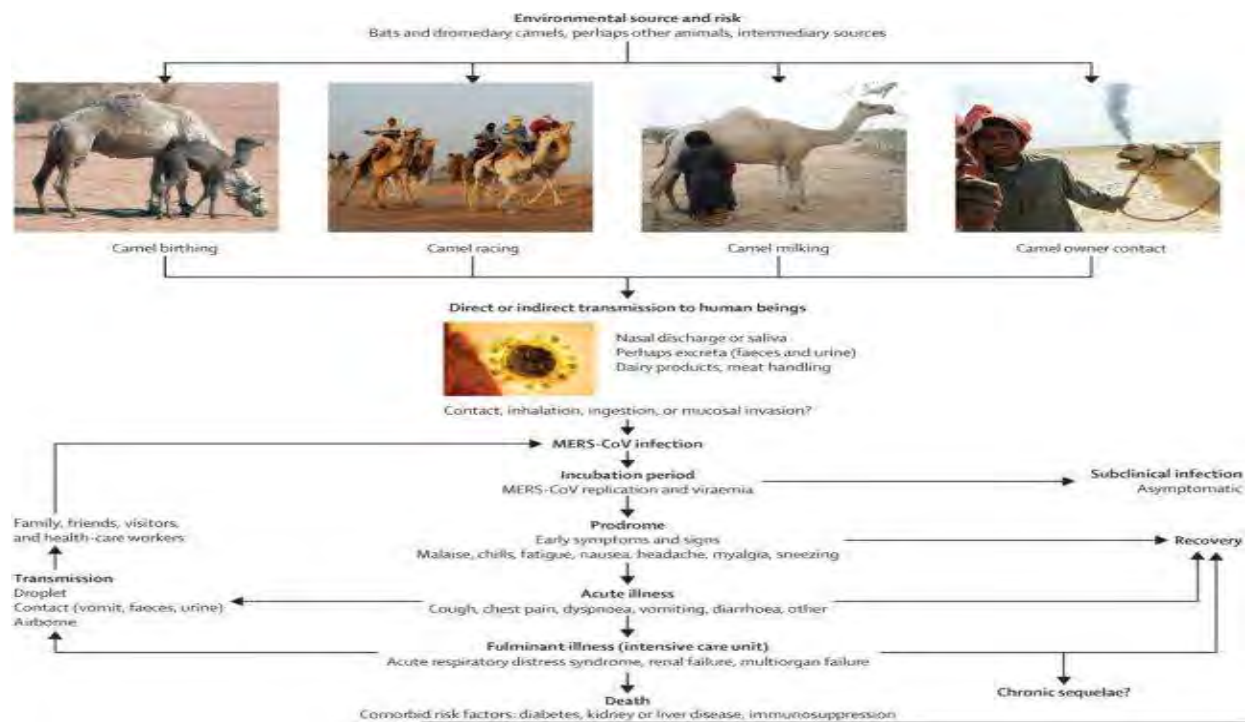


Figure 1.6: Ecology and transmission of MERS-CoV (Zumla et al., 2015)

intermediate hosts. After interspecies-transmission, MERS-CoV has circulated in camel populations in Africa and the Arabian Peninsula for at least 20 years. In 2012, MERS-CoV crossed interspecies barriers to infect human populations (Fehr et al., 2016).

Person-to-person transmission occurs through large droplets. However, aerosol or fomite transmission has not been ruled out (Fehr et al., 2016). Conclusive evidence of human-to-human transmission of MERS-CoV was first found in the United Kingdom where an adult male transmitted the virus to two of his family members (Milne-Price et al., 2014). MERS-CoV may persist in the environment for up to 24 hours, which may also contribute to human infection. Sequence analyses of virus isolated from patients revealed little evidence for directed mutation, lead to the belief that unlike SARS-CoV, MERS-CoV is not adapting to human populations (Fehr et al., 2016).

The fact that the average MERS patient is ~50 years old suggests that age is a risk factor for developing severe MERS. Elderly patients have a greater risk of dying from the disease, with a fatality rate of nearly 90% for patients over 80 compared to ~10% for those under the age of 20. Presence of underlying co-morbidities is another risk factor for developing severe MERS since approximately 75% of all documented cases occurred in patients with co-morbidities (Fehr et al., 2016).

1.5 MERS-CoV: Pathogenesis, Pathology and Immunity

MERS pathogenesis begins with the entry of the virus via the respiratory tract where the spike (S) protein interacts with its cellular receptor DPP4. DPP4 is expressed in the respiratory tract on type I and II pneumocytes, non-ciliated bronchial epithelial cells, endothelial cells, and some hematopoietic cells. Besides respiratory tract, DPP4 is also widely expressed on the epithelial cells of several other organs and tissues such as kidneys, intestine, liver, thymus and bone marrow. Lack of patient autopsy or surgical pathology samples from the Middle East or the Korean outbreak has limited studies of MERS-CoV pathogenesis (Fehr et al., 2016).

Severe MERS-CoV infection causes acute pneumonia which is often lethal. In addition, renal dysfunction/failure may occur as a result of either hypoxic damage or direct infection of the kidney as DPP4 is expressed at high levels in the kidney (Fehr et al., 2016).

Despite having a lack of knowledge on what constitutes a protective immune response in MERS patients who recover, it can be concluded that coordinated innate and adaptive immune responses are required for effective and long-lasting immunity. MERS-CoV induces attenuated innate immune responses with delayed pro-inflammatory cytokine induction in cell culture and *in vivo*, which may lead to a dysregulated immune response. Based on studies conducted on SARS-CoV infection, it can be expected that vaccines inducing antibody responses only may not provide long-lasting immunity against MERS-CoV despite being useful in the short term (Zumla et al., 2015).

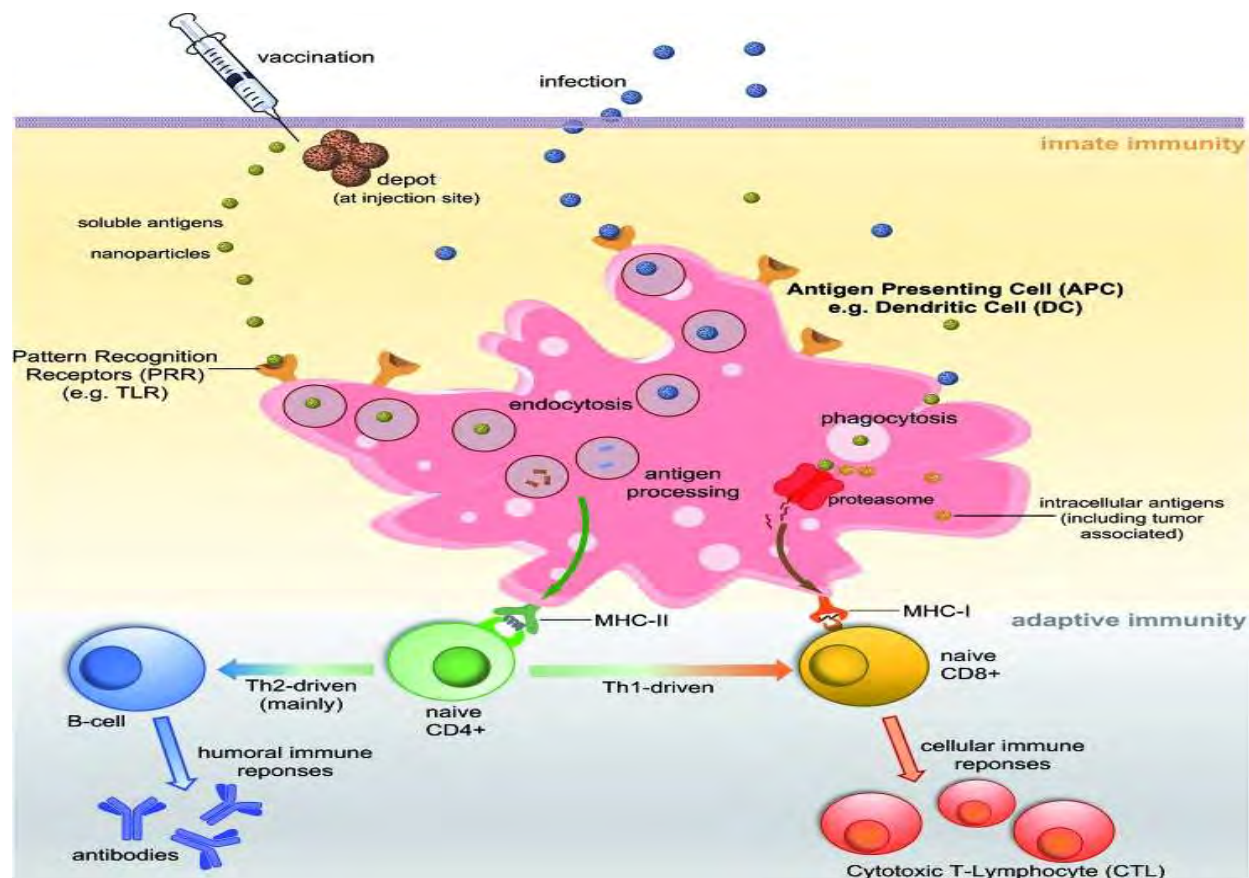


Figure 1.7: Schematic representation of major pathways of immune response (Skwarczynski et al., 2015)

1.6 Vaccine Development and Treatment

No MERS-CoV-specific vaccines are currently approved for use in humans. Even though several types of vaccines have been developed, none have been approved for clinical trials (Fehr et al.,

2016). As there is no specific approved therapeutic agent or vaccine available to treat or prevent MERS infection, supportive care is the main focus of treatment.

1.7 Epitope-based Vaccine

T-cell epitope-based vaccines have potential as therapeutic vaccines against viral infection as epitope-based vaccines with adjuvants can induce a strong immune response with high immunogenicity (Kametani et al., 2015). However, identifying a few epitopes among a mixture of above 10,000 MHC class I associated epitopes extracted from virus-infected cells is very difficult. Many software can correctly identify epitopes on a protein sequence using several databases. In this way, a probable epitope can be identified easily prior to vaccine design. Such computational approach not only speeds up the time but also lowers the cost needed for laboratory analysis as well as vaccine development.

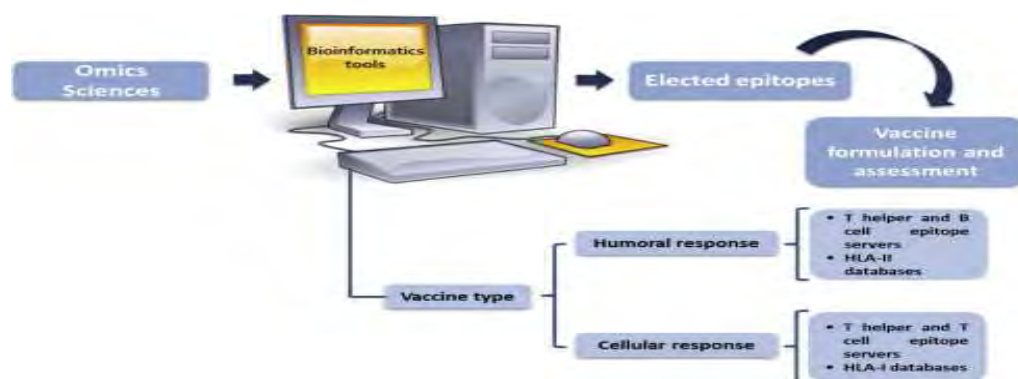


Figure 1.8: Schematic diagram of epitope-based vaccine formulation using immunoinformatics

1.8 Advantages of Epitope-based Vaccine

An epitope-based vaccine formulated with an adjuvant can induce a strong immune response with high immunogenicity. Several peptide vaccines are already in development and some of them are being clinically tested (Lambert et al., 2015). T-cell epitopes are usually peptide fragments whereas B-cell epitopes can be either protein, lipids or carbohydrates.

Peptides have become ideal candidates for vaccines because:

- Peptide-based vaccines are produced almost exclusively using chemical synthetic approaches.

- Production of peptides is simple, easily reproducible, fast and cost-effective.
- Chemical synthesis eliminates all the problems associated with the biological contamination of the antigens.
- These vaccines are typically water-soluble and stable. In addition, they can be freeze-dried.
- Peptides can be customized to target very specific objectives. The immune responses can be directed against naturally non-immunodominant epitopes. By the use of a multi-epitope approach, a single peptide-based vaccine can be designed to target several strains, different stages of the life cycle or even different pathogens.
- Peptide antigens are less likely to induce allergic or autoimmune responses due to the lack of redundant elements (Skwarczynski et al., 2015).

1.9 Flavonoids: Potential Anti-viral Agents against MERS-CoV

Flavonoids are part of the polyphenol class of phytonutrients. Polyphenols have traditionally been utilized in Chinese and Ayurvedic medicine. There are several groups of flavonoids having many subgroups, including flavones, flavonols, flavanones, flavanonols, flavanols or catechins, anthocyanins and chalcones (Panche et al., 2016). Each type of flavonoid carries its own distinct set of actions and beneficial properties.

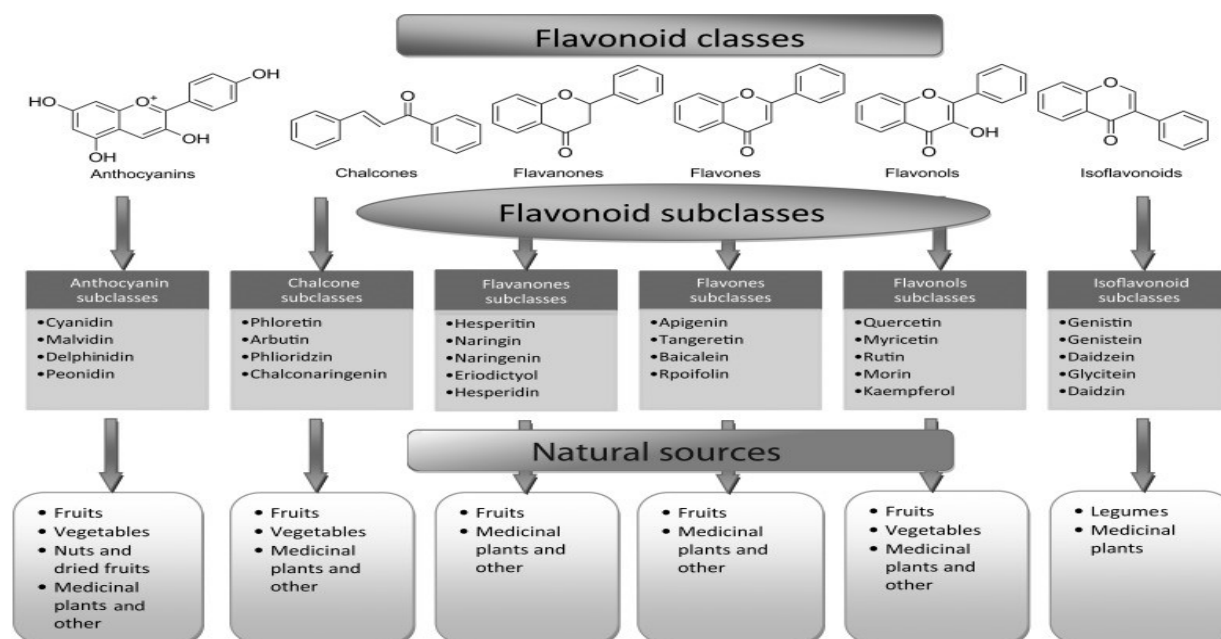


Figure 1.9: Flavonoid classes, subclasses and natural sources (Panche et al., 2016)

Plant-derived flavonoids are a large group of naturally occurring phenylchromones found in fruits, vegetables, tea, and wine. They have been shown to have a wide range of biological activities, including antiallergic, antibacterial, antidiabetic, antiinflammatory, antiviral, antiproliferative, antimutagenic, antithrombotic, anticarcinogenic, hepatoprotective, oestrogenic, insecticidal, and antioxidant activities (Orhan et al., 2010). Large studies have successfully shown that various types of flavonoids have significant antiviral activities against a wide range of viruses. Therefore, flavonoids may become potential anti-viral agents against MERS-CoV.

1.10 Aims and Objectives

- Studying about MERS-CoV and its current consequences.
- Exploring the possibilities of drug development against MERS-CoV using immunoinformatics.
- Identifying competent peptides for developing a vaccine against MERS-CoV.
- Exploring potential anti-viral effects of Flavonoids against MERS-CoV.

Materials and Methods

2.1 Method Summary

This study was divided into two sections. The first section was focused on designing an epitope-based vaccine whereas the second section was focused on identifying effective flavonoids in order to use them as nsp3 inhibiting therapeutic agents against MERS-CoV. For this study, only open-source immunoinformatics software and tools were used.

To design an epitope-based vaccine, nsp3 protein sequence was first extracted from the NCBI database and was checked for antigenicity using VaxiJen 2.0. For prediction of T-cell epitopes, the sequence of nsp3 protein was put in the NetCTL 1.2 server to identify probable T-cell epitopes in the target sequence. Antigenicity of selected epitopes was then evaluated using VaxiJen 2.0 followed by IEDB T-cell class I pMHC immunogenicity predictor. Then prediction of peptide-MHC class I binding was performed using both Proteasomal cleavage/TAP transport/MHC class I combined predictor and NetMHC 4.0 server. Afterward, the selected epitopes were used for the prediction of MHC-II alleles using IEDB Peptide binding to MHC class II molecules predictor. Then, population coverage of identified MHC-I-binding alleles with a high binding affinity of selected epitopes was analyzed using the IEDB Population Coverage Analysis tool. Then selected epitopes were checked for conservancy using IEDB conservancy analysis tool. Then, the toxicity of the epitope candidates was predicted using ToxinPred. Finally, allergenicity was anticipated using AllergenFP v1.0 and AllerTOP v2.0. For prediction of B-cell epitopes, nsp3 protein sequence was put in B-cell epitope predictor tools such as BCPREDS and BepiPred 2.0. and epitopes generated by these tools were screened using VaxiJen 2.0. Then overlapping B-cell and T-cell epitopes were identified and were selected for further evaluation. Conservancy analysis was performed using the IEDB conservancy analysis tool. Afterward, these epitopes were checked for the presence of beta-turn, surface accessibility, flexibility, antigenicity and hydrophilicity using several IEDB B-cell tools. Then, the toxicity of these epitope candidates was predicted using ToxinPred. After that, allergenicity was anticipated using AllergenFP v1.0 and AllerTOP v2.0. Finally, docking analysis was performed using common MHC allele as the macromolecule. For docking analysis, the 3D structures of selected epitope candidates were predicted using PEP-FOLD 2.0 server. A control ligand was also selected for comparing results. Then, energy minimization of macromolecules and ligands were carried out using UCSF Chimera 1.13. After the minimizing

process, PyRx was used for molecular docking. Finally, molecular visualization of the best docking poses was performed using UCSF Chimera 1.13

Molecular docking was used to identify effective flavonoids in order to use them as nsp3 inhibiting therapeutic agents against MERS-CoV. In this study, several flavonoids were selected as potential nsp3 inhibitor candidates. Each of these flavonoids was used as a ligand separately in molecular docking. MERS-CoV macro domain within nsp3 protein was selected as the macromolecule. To perform molecular docking, 3D structures of ligands were retrieved from PubChem in SDF format. After that, the three-dimensional crystal structure of the macro domain within nsp3 protein was retrieved from RCSB Protein Data Bank in PDB format. Then, undesired ligands along with water molecules were removed, and energy minimization was done using Dock Prep tool in UCSF Chimera 1.13. Ligand minimization was done by PyRx prior to docking. After the minimizing process, PyRx was used for molecular docking. The protein was placed in a grid box and ADP-ribose was first re-docked into the ADP-ribose binding site of nsp3, and the resulting interactions were later compared with those found by docking 18 flavonoids into the similar active site using the same grid box. The docking poses were ranked according to their docking scores. The conformation with the lowest binding affinity was selected as the best docking pose. Afterward, the K_i value was measured for each compound. Finally, the molecular visualization of the docked complexes was performed using UCSF Chimera 1.13, and the intermolecular H bonds between each compound with MERS-CoV nsp3 macro domain were listed along with their respective distances.

2.1.1 In silico Analysis of Epitope-based Vaccine Candidates

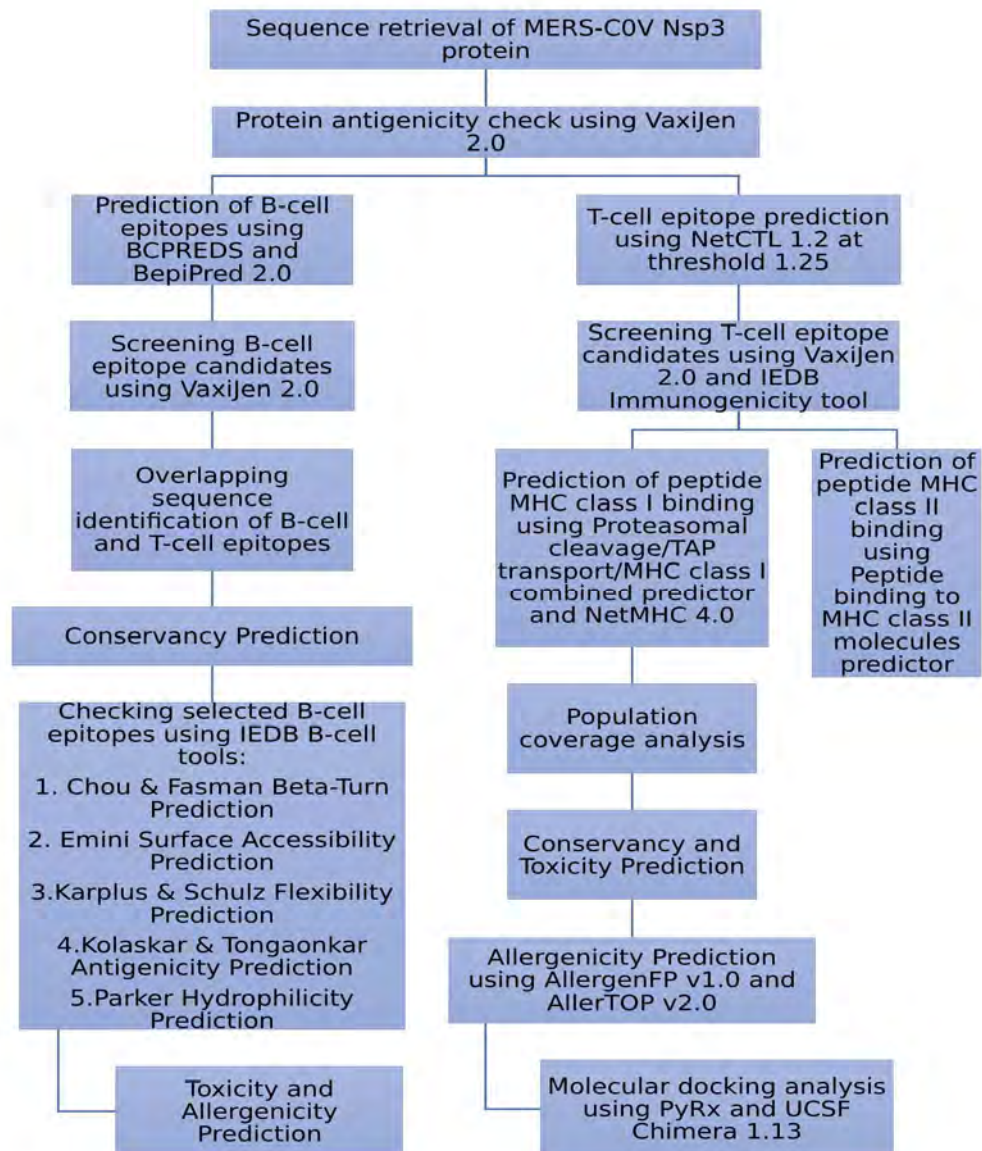


Figure 2.1 Summary of the methodology of in silico analysis of epitope-based vaccine candidates

2.1.2 Assessing Potential Anti-Viral Activity of Selected Flavonoids against MERS-CoV using Immunoinformatics

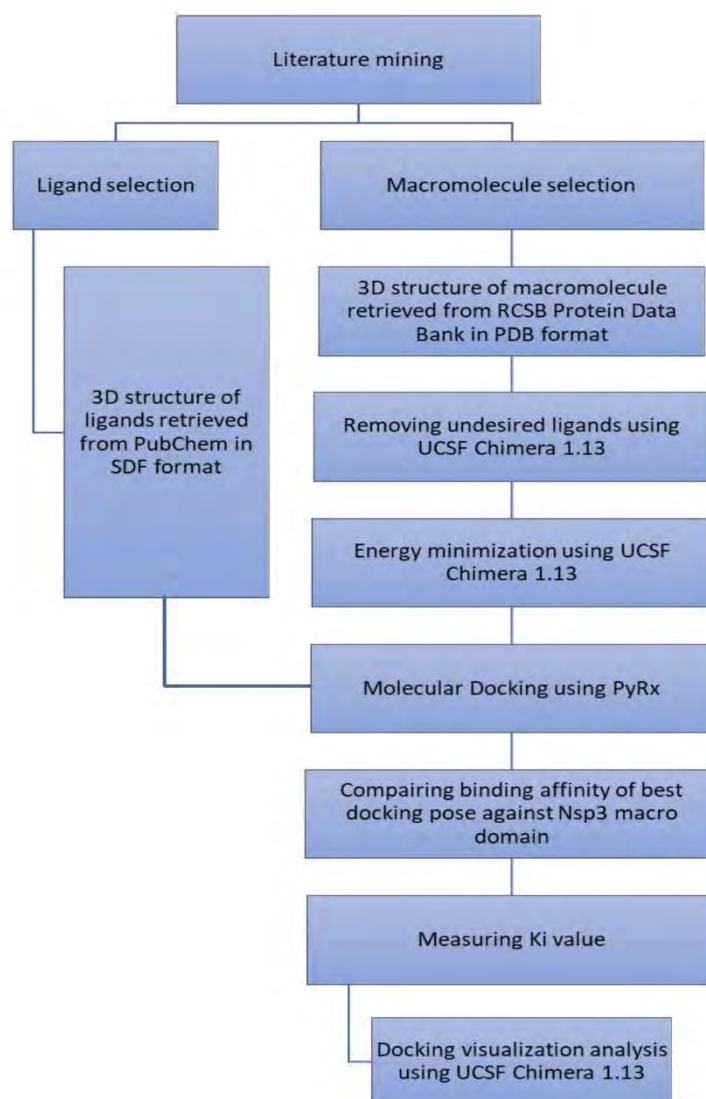


Figure 2.2: Summary of the methodology of assessing the potential anti-viral activity of selected flavonoids against MERS-CoV using immunoinformatics

2.2 Databases used for obtaining data:

2.2.1 NCBI

The National Center for Biotechnology Information (NCBI) was established in 1988 as a national resource for molecular biology information. NCBI houses several public databases relevant to biotechnology and biomedicine, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information.

URL link: <https://www.ncbi.nlm.nih.gov>

2.2.2 PDB

The Protein Data Bank (PDB) is a crystallographic database for the three-dimensional (3D) structural data of large biological molecules, such as proteins and nucleic acids. The data, typically obtained by X-ray crystallography or NMR spectroscopy and submitted by biologists and biochemists from around the world, are freely accessible on the Internet via websites of its member organizations (PDBe, PDBj and RCSB). The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB.

URL link: <https://www.rcsb.org>

2.3 Software and tools used for analysis:

2.3.1 UCSF chimera 1.13

Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). It is a highly extensible program for interactive visualization and analysis of molecular structures and related data. High-quality animations and images can be produced by this tool. It can be downloaded from the UCSF Chimera website (<http://www.cgl.ucsf.edu/chimera>).

2.3.2 PyRx

PyRx is a Virtual Screening software for Computational Drug Discovery that can be used to screen libraries of compounds against potential drug targets. PyRx includes docking wizard with an easy-to-use user interface which makes it a valuable tool for Computer-Aided Drug Design. PyRx also

includes chemical spreadsheet-like functionality and powerful visualization engine that are essential for structure-based drug design. It can be downloaded from the website (<https://pyrx.sourceforge.io>).

2.3.3 VaxiJen 2.0

VaxiJen is the first server for alignment-independent prediction of protective antigens of bacterial, viral and tumor origin. VaxiJen contains models derived by auto- and cross-covariance pre-processing of amino acids properties. The models showed remarkable stability, as tested by combinations of the positive set and five different negative sets. Thus, VaxiJen is a reliable and consistent tool for the prediction of protective antigens.

URL link: <http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>

2.3.4 BepiPred 2.0

BepiPred 2.0 is a web server that allows users to predict B-cell epitopes from antigen sequences. The server is based on a random forest algorithm trained on epitopes annotated from antibody-antigen protein structures.

URL link: <http://www.cbs.dtu.dk/services/BepiPred>

2.3.5 BCPREDS

BCPREDS is another web server used for predicting B-cell epitopes from sequences. It currently uses three prediction methods. It lets the user change various parameters such as sequence length and specify threshold as well. BCPREDS server allows users to choose the method for predicting B-cell epitopes among several developed prediction methods.

URL link: <http://ailab.ist.psu.edu/bcpred/>

2.3.6 IEDB tools

IEDB is a collection of tools for predicting and analyzing of immune epitopes. It is as a companion site of the Immune Epitope Database (IEDB), a manually curated database of experimentally characterized immune epitopes.

URL link: <http://www.iedb.org/>

2.3.6.1 Chou & Fasman Beta-Turn Prediction

The Chou–Fasman method is an empirical technique that helps to predict secondary structures in proteins. This method originally was developed in the 1970s by Peter Y. Chou and Gerald D. Fasman. It relies on the relative frequencies of each amino acid in alpha helices, beta sheets, and turns based on known protein structures solved with X-ray crystallography.

2.3.6.2 Emini surface accessibility prediction tool:

The calculation is based on the surface accessibility scale on a product instead of an addition within the window. It is one of the parameters required to be an ideal epitope.

2.3.6.3 Karplus and Schulz Flexibility prediction tool:

In this method, flexibility scale based on the mobility of protein segments on the basis of the known temperature B factors of the α -carbons of 31 proteins of known structure was constructed. The calculation based on a flexibility scale is similar to classical calculation, except that the center is the first amino acid of the six amino acids window length, and there are three scales for describing flexibility instead of a single one.

2.3.6.4 Kolaskar and Tongaonkar antigenicity prediction tool:

It is the simplest method for the prediction of antigenic determinants. This method predicts antigenic epitopes of given sequence based on physicochemical properties of amino acid residues that frequently occur in experimentally determined antigenic epitopes. Previously reported data appreciated this method as it gives 75% experimental accuracy.

2.3.6.5 Parker Hydrophilicity prediction tool:

In this method, the hydrophilic scale based on peptide retention times during high-performance liquid chromatography (HPLC) on a reversed-phase column was constructed. A window of seven residues was used for analyzing the epitope region. The corresponding value of the scale was introduced for each of the seven residues and the arithmetical mean of the seven residues value was assigned to the fourth, $(i+3)$, residue in the segment.

2.3.6.6 Epitope Conservancy Analysis

This tool computes the degree of conservancy of an epitope within a given protein sequence set at a given identity level. Conservancy is defined as the fraction of protein sequences that contain the epitope, and Identity is the degree of correspondence (similarity) between two sequences.

2.3.6.7 Proteasomal cleavage/TAP transport/MHC class I combined predictor

This tool combines predictors of proteasomal processing, TAP transport, and MHC binding to produce an overall score for each peptide's intrinsic potential of being a T-cell epitope.

2.3.6.8 T-cell class I pMHC immunogenicity predictor

This tool uses amino acid properties as well as their position within the peptide to predict the immunogenicity of a class I peptide MHC (pMHC) complex.

2.3.6.9 Peptide binding to MHC class II molecules predictor

This tool employs different methods to predict MHC Class II epitopes, including a consensus approach which combines NN-align, SMM-align and Combinatorial library methods.

2.3.6.10 Population Coverage

This tool calculates the fraction of individuals predicted to respond to a given set of epitopes with known MHC restrictions. This calculation is made on the basis of HLA genotypic frequencies assuming non-linkage disequilibrium between HLA loci.

2.3.7 NetCTL 1.2 Server

NetCTL 1.2 server predicts CTL epitopes in protein sequences. The version 1.2 expands the MHC class I binding prediction to 12 MHC supertypes including the supertypes A26 and B39. The accuracy of the MHC class I peptide binding affinity is significantly improved compared to the earlier version. Also, the prediction of proteasomal cleavage has been improved and is now identical to the predictions obtained by the NetChop-3.0 server. The updated version has been trained on a set of 886 known MHC class I ligands. The method integrates prediction of peptide MHC class I binding, proteasomal C terminal cleavage and TAP transport efficiency.

URL link: <http://www.cbs.dtu.dk/services/NetCTL/>

2.3.8 NetMHC 4.0 Server

NetMHC 4.0 Server predicts peptide-MHC class I binding using artificial neural networks (ANNs) which generate rank of the predicted affinity(%Rank) by comparing to a set of 400.000 random natural peptides. This measure is not affected by the inherent bias of certain molecules towards higher or lower mean predicted affinities. The peptide will be identified as a weak binder if the % Rank is above the threshold of the strong binders (0.5% by default) but below the specified threshold for the weak binders (2% by default).

URL link: <http://www.cbs.dtu.dk/services/NetMHC/>

2.3.9 ToxinPred

ToxinPred is an in-silico method, which is developed to predict and design toxic/non-toxic peptides. The main dataset used in this method consists of 1805 toxic peptides (≤ 35 residues).

One of the major features of the server is that it also calculates various physicochemical properties. Peptide analogs can be displayed in sorting order based upon desired properties.

URL link: <http://crdd.osdd.net/raghava/toxinpred/>

2.3.10 AllergenFP v1.0

AllergenFP v1.0 is an alignment-free method to predict allergenicity of target peptide.

URL link: <http://www.ddg-pharmfac.net/AllergenFP/>

2.3.11 AllerTOP v2.0

AllerTOP v2.0 is a robust bioinformatics tool for in-silico allergenicity prediction.

URL link: <https://www.ddg-pharmfac.net/AllerTOP/>

2.3.12 PEP-FOLD 2.0

PEP-FOLD is a de novo approach aimed at predicting peptide structures from amino acid sequences.

URL link: <http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/>

Results

3.1 Result Summary

In the present study, nsp3 protein sequence was extracted from the NCBI database and was checked for antigenicity using VaxiJen 2.0. The result of VaxiJen 2.0 indicates that nsp3 protein is antigenic with a value of 0.4794 which is over the threshold for virus model (0.4). For prediction of T-cell epitopes, the sequence of nsp3 protein was put in NetCTL 1.2 server to identify probable T-cell epitopes in the target sequence, and 30 T-cell epitopes were selected which achieved threshold value of 1.25. Antigenicity of selected epitopes was then evaluated using VaxiJen 2.0 followed by IEDB T-cell class I pMHC immunogenicity predictor. Only six T-cell epitopes had achieved the threshold value of 0.4 in VaxiJen 2.0 and positive immunogenicity score in IEDB T-cell class I pMHC immunogenicity predictor. Then prediction of peptide-MHC class I binding was performed using both Proteasomal cleavage/TAP transport/MHC class I combined predictor and NetMHC 4.0 server. In Proteasomal cleavage/TAP transport/MHC class I combined predictor tool, FAFETGLAY showed the highest affinity and was recognized by 20 MHC-I alleles. FAFETGLAY showed the highest affinity in NetMHC 4.0 server as well and was recognized by 37 MHC-I alleles. Moreover, FAFETGLAY was recognized by 44 MHC-I alleles combined. Next, FVDWRSYNY had the second highest affinity and was recognized by 23 MHC-I alleles combined. After that, LLLAGTLHY was recognized by 17 MHC-I alleles combined. Next, KTTTGIPEY was recognized by 14 MHC-I alleles combined. Finally, both LSSVYHLYV and STDFIALIM showed the least affinity as LSSVYHLYV, along with LSSVYHLYV, was recognized by 11 MHC-I alleles combined. HLA-A*01:01 is the only MHC-I allele that had affinity with all of the selected T-cell epitopes. Afterward, the selected epitopes were used for the prediction of MHC-II alleles using the IEDB Peptide binding to MHC class II molecules predictor. Afterward, identified MHC-I-binding alleles with high binding affinity of six epitopes were considered to analyze population coverage using the IEDB Population Coverage Analysis tool. IEDB Population Coverage Analysis tool revealed that these epitopes and their HLA-alleles cover 98.55% of the world population cumulatively. The highest population coverage was found in the South Africa region (99.66%) while the lowest population coverage was found in Central America (9.07%). The cumulative population coverage in Southwest Asia was 96.40%. whereas the cumulative population coverage in East Asia was 96.86%. Then those six epitopes were checked for conservancy using IEDB conservancy analysis tool. Conservancy analysis revealed that all of them had the maximum identity (100%) for conservancy hit. In addition, all six of the selected

epitopes were found non-toxic in ToxinPred. Finally, allergenicity was anticipated using AllergenFP v1.0 and AllerTOP v2.0. STDFIALIM, LSSVYHLYV, FAFETGLAY, LLLAGTLHY, FVDWRSYNY were found as probable allergens in AllergenFP v1.0 while KTTTGIPEY was found non-allergenic. On the contrary, STDFIALM and LSSVYHLYV were found non-allergenic in AllerTOP v2.0. KTTTGIPEY was found non-allergenic in both AllergenFP v1.0 and AllerTOP v2.0. Then, molecular docking analysis was performed using HLA-A*01:01 as the macromolecule. The crystal structure of the HLA-A*01:01 protein molecule (PDB ID:4nqx) was retrieved from the RCSB Protein Data Bank in PDB format, and UCSF Chimera 1.13 was used to remove undesired ligands and molecules. The 3D structures of selected epitope candidates were predicted using PEP-FOLD 2.0 server. Besides these T-cell epitope candidates, NP44-S7N mutant peptide (CTELKLNDY) was also selected as a ligand to be used as a control. After the minimizing process, PyRx was used for molecular docking. HLA-A*01:01 protein was placed in a grid box measuring $52.8351 \text{ \AA} \times 68.2709 \text{ \AA} \times 61.7293 \text{ \AA}$ along the x, y and z axis, respectively, where the position of the center was X:-63.5001, Y:-17.1718, Z:7.5672. The docking procedure was performed using the instructed command prompts. The docking poses were ranked according to their docking scores. Since the docking result revealed two different binding sites, the conformation with the lowest binding affinity that used the same binding site as control (NP44-S7N mutant peptide) was selected as best docking pose in order to compare between sample and control for critical evaluation. Then, molecular visualization of the best docking poses was performed using UCSF Chimera 1.13. Only FAFETGLAY formed an intermolecular hydrogen bond. For prediction of B-cell epitopes, nsf3 protein sequence was put in BCPREDS and BepiPred 2.0. These tools generated a repertoire of probable B-cell epitope candidates which were screened using VaxiJen 2.0. After that, the number of epitopes candidates was reduced to 178. Then overlapping B-cell and T-cell epitopes were identified. Out of 178 B-cell epitopes, only eight had the sequence similarity with the selected T-cell epitopes. IFVDWRSYNYAVSS, FVDWRSYNYAVS, FVDWRSYNYAVSSAFW and FVDWRSYNYAVSSAFWLF had the sequence similarity with FVDWRSYNY whereas LKFKEVCKTTTGIPEYNF, LKFKEVCKTTTGIPEY, FKEVCKTTTGIPEYNFIIYD and VCKTTTGIPEYN had the sequence similarity with KTTTGIPEY. Conservancy analysis of these B-cell epitopes using IEDB conservancy analysis tool revealed that all of them had the maximum identity (100%) for conservancy hit. Afterward, these epitopes were checked for the presence of beta-turn, surface

accessibility, flexibility, antigenicity and hydrophilicity using several IEDB B-cell tools. Results of IEDB B-cell tools showed that among eight B-cell epitopes, LKFKEVCKTTTGIPEYNF had the presence of beta-turn, surface accessibility, flexibility, high antigenicity and hydrophilicity. On the contrary, epitope FVDWRSYNYAVSSAFWLF and LKFKEVCKTTTGIPEY had the presence of beta turns and they were found antigenic, flexible, hydrophilic but performed a bit poorly in Emini Surface Accessibility Prediction tool. All epitopes were found as non-toxic in ToxinPred. However, allergenicity prediction results were not conclusive for LKFKEVCKTTTGIPEYNF as it was found as a probable allergen in AllergenFP v1.0 but AllerTOP v2.0 identified LKFKEVCKTTTGIPEYNF as non-allergen. FVDWRSYNYAVSSAFWLF and LKFKEVCKTTTGIPEY were found as probable non-allergen in both AllergenFP v1.0 and AllerTOP v2.0.

In the present study, 18 flavonoids were selected as potential nsp3 inhibitor candidates to be used as ligands in molecular docking. MERS-CoV macro domain within nsp3 protein was selected as the macromolecule. 3D structures of ligands were retrieved from PubChem in SDF format whereas the three-dimensional crystal structure of the macro domain within nsp3 protein was retrieved from RCSB Protein Data Bank (PDB ID:5DUS) in PDB format. After the minimizing process, PyRx was used for molecular docking. The protein was placed in a grid box measuring $37.3660 \text{ \AA} \times 43.4316 \text{ \AA} \times 43.1478 \text{ \AA}$ along the x, y and z axis, respectively, where the position of the center was X:8.9843, Y:17.6095, Z:68.5928 and then the docking procedure was performed using the instructed command prompt. The docking poses were ranked according to their docking scores and then conformation with the lowest binding affinity was selected as the best docking pose. The best docking conformation of ADP-ribose showed a binding affinity of -8.7 kcal/mol. Only 12 flavonoids had binding affinities greater than -8.7 kcal/mol. Among them, apiin and naringin had the best binding affinity of -10.1 kcal/mol. Afterward, the K_i value was measured for each compound. The present study revealed that apiin, naringin, luteoloside, hesperidin and fisetin had significantly lower K_i values (less than 0.08 \mu M), with apiin and naringin having the lowest K_i value (0.0390649 \mu M). Finally, the molecular visualization of the docked complexes was performed revealed that both apiin and quercetin had five intermolecular hydrogen bonds (same as ADP-ribose) while naringin, luteoloside, fisetin, and pinostrobin had three intermolecular hydrogen bonds each. Others had less than three intermolecular hydrogen bonds each. However, kaempferol did not form any visible hydrogen bond at all.

3.2 Sequence retrieval

The protein sequence of MERS-CoV nsp3 protein was retrieved from NCBI database in fasta format.

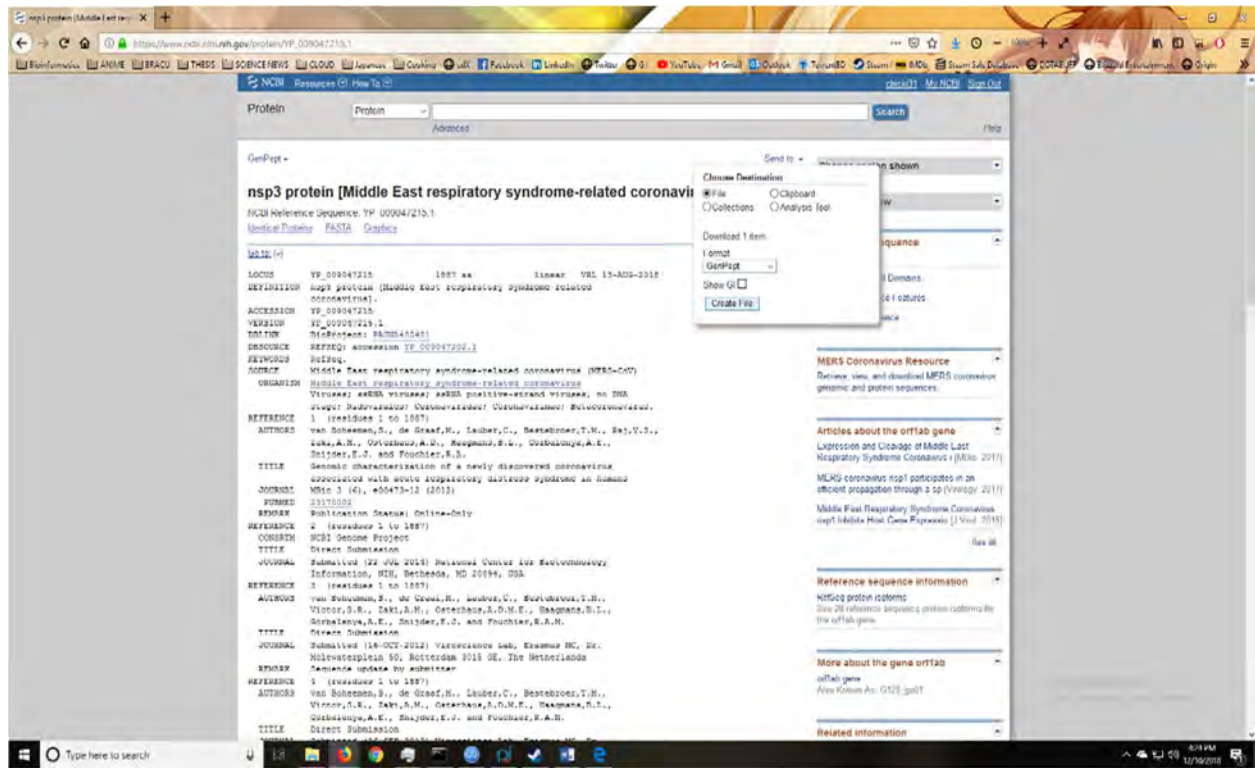


Figure 3.1: Protein sequence retrieval from NCBI database

Here the target sequence was MERS-CoV nsp3 protein:

>YP_009047215.1 nsp3 protein [Human betacoronavirus 2c EMC/2012]

```
APVKKVAFGGDQVHEVA AAVRSVTVEYNIHAVLDTLLASSSLRTFVVDKSLSIEEFADV  
KEQVSDLLVKLLRGMPIDFDLDDFIDAPCYCFNAEGDASWSSTMIFSLHPVECDEECSE  
VEASDLEEGESECISSETSTEQVDVSHETSDDEWAAAVDEAFPLDEAEDVTESVQEEAQP  
VEVPVEDIAQVVIADTLQETPVVPTDVEVPPQVVKLPAPQTIQPEVKEVAPVYEADTEQ  
TQNVTVKPKRLRKKRNVDP LSNFEHKVITECVTIVLGDAIQVAKCYGESVLVNAANTHL  
KHGGGIAGAINAASKGAVQKESDEYILAKGPLQVGDSVLLQGHSLAKNILHVVGPDAR  
AKQDVSLLSKCYKAMNAYPLVVTPLVSAGIFGVKPAVSFDYLIREAKTRVLVVVNSQD  
VYKSLTIVDIPQSLTFSYDGLRGAIRKAKDYGFTVFVCTDNSANTKVLRNKGVDYTKKF  
LTVDGVQYYCYTSKDTLDDILQQANKSVGIISMPLGYVSHGLDLMQAGSVVRRVNPY  
VCLLANKEQEAILMSEDEVKLNPSDFIKHVRTNGGYN SWHLVEGELLVQDLRLNKLH  
WSDQTICYKDSVFYVVKNSTAFFETLSACRAYLDSRTTQQLTIEVLVTVDGVNFRTVV  
LNNKNTYRSQ LGCVFENGADISDTIPDEKQNGHSLYLADNLTADETKALKELYGPVDPT  
FLHRFYSLKAAVHGWMVCDKVRSLKLS DNNCYLNAVIMTLDLLKDIKFVIPALQHA
```

FMKHKGGDSTDFIALIMAYGNCTFGAPDDASRLHTVLAKAELCCSARMVWREWCNV
 CGIKDVVLQGLKACCYVGVQTVEDLRARMTYVCQCGERHRQLVEHTTPWLLLSGTP
 NEKLVTSTAPDFVAFNVFQGIETAVGHYVHARLKGGILKFDSGTVSKTSDWKCKVTD
 VLFPGQKYSSDCNVVRYSLDGNFRTEVDPDLFAFYVKDGYFTSEPPVTYSPATILAGS
 VYTNSCLVSSDGQPGGDAISLSFNLLGFDSSKPVTKKYTSFLPKEDGDVLLAEFDITYD
 PIYKNGAMYKGPILWVNKASYDTNLNKFNRASLRQIFDVAPIELENKFTPLSVESTPVE
 PPTVDVVALQQEMTIVKCKGLNKPVKDNVSFVADDSGTPVVEYLSKEDLHTLYVDPK
 YQVIVLKDNLSSMLRLHTVESGDINVVAASGSLTRKVKLLFRASFYFKEFATRFTTATT
 AVGSCIKSVRHLGVTKGILTGCFSAKMLFMLPLAYFSDSKLGTTEVKVSALKTAGVV
 TGNVVKQCCTAAVDLSMDKLRVDWKSTLRLLMLCTTMVLLSSVYHLYVFNQVLSS
 DVMFEDAQGLKKFYKEVRAYLGISSACDGLASAYRANSFDVPTFCANRSAMCNWCLIS
 QDSITHYPALKMVQTHLSHYVLNIDWLWFAFETGLAYMLYTSAFNWLLLAGTLHYFFA
 QTSIFVDWRSYNYAVSSAFWLFTHIPMAGLVRMYNLLACLWLLRKFYQHVIINGCKDTA
 CLLCYKRNRLTRVEASTVVCGGKRTFYITANGGISFCRRHNWNCVDCDTAGVGNTFICE
 EVANDLTTALRRPINATDRSHYYVDSVTVKETVVQFNYYRRDGQPFYERFPLCAFTNLDK
 LKFKEVCKTTTGIPEYNFIYDSSDRGQESLARSACVYYSQVLCKSILLVDSSLVTSVGDS
 SEIATKMFDSFVNSFVSLYNVTRDKLEKLISTARDGVRRGDNFHSVLTTFIDAARGPAGV
 ESDVETNEIVDSVQYAHKHDIQITNESYNNYVPSYVKPDSVSTSDLGSLIDCNAASVNQI
 VLRNSNGACIWNAAYMKLSDALKRQIRIACRKCNLAFRLTTSKLRANDNILSVRFTAN
 KIVGG

3.3 VaxiJen result for nsp3 Protein: Primary result of antigenicity

When nsp3 protein sequence was run in VaxiJen 2.0, antigenicity was found 0.4794, which is above the threshold value of ≥ 0.4 . This indicates that nsp3 protein has good antigenic property.

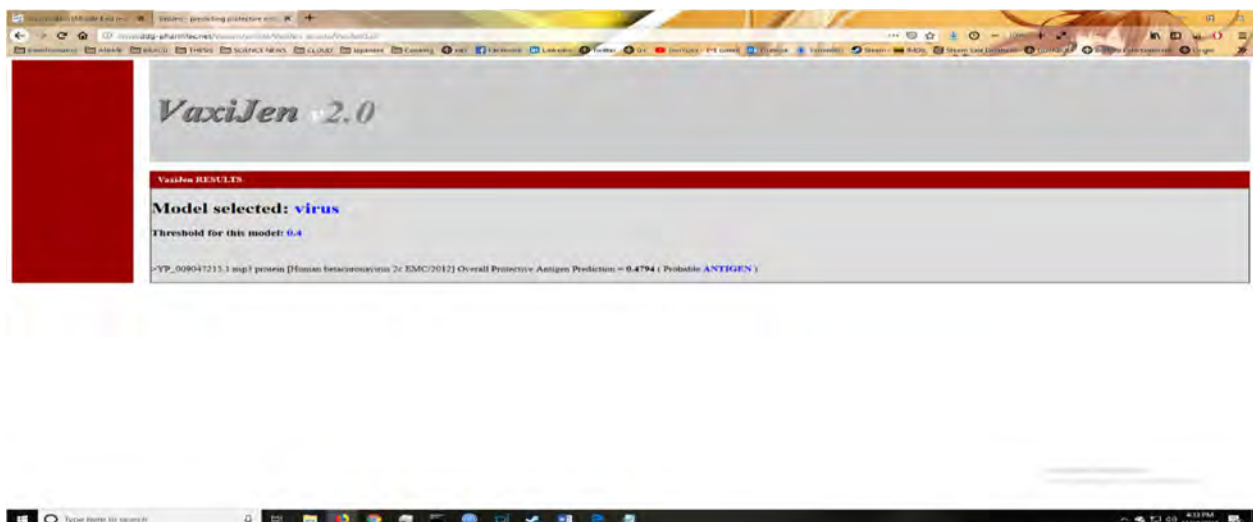


Figure 3.2: Analysis of nsp3 protein of MERS-CoV using VaxiJen 2.0 showing that the sequence is probably antigenic

3.4 Prediction of T-Cell epitopes

The sequence of nsp3 protein was put in T-Cell epitope predictor tools to identify probable T-cell epitopes in a target sequence. Selection of T-cell epitope candidates was done using NetCTL 1.2 Server.

3.4.1 Prediction of CD8+ T-cell epitopes using NetCTL 1.2 Server

The NetCTL 1.2 web server was used for the prediction of CD8+ T-cell epitopes, and for this purpose, the combined score for the epitope selection was considered. After putting the target sequence in the NetCTL 1.2 server, 30 epitopes in total were found which achieved the selected threshold value of 1.25.

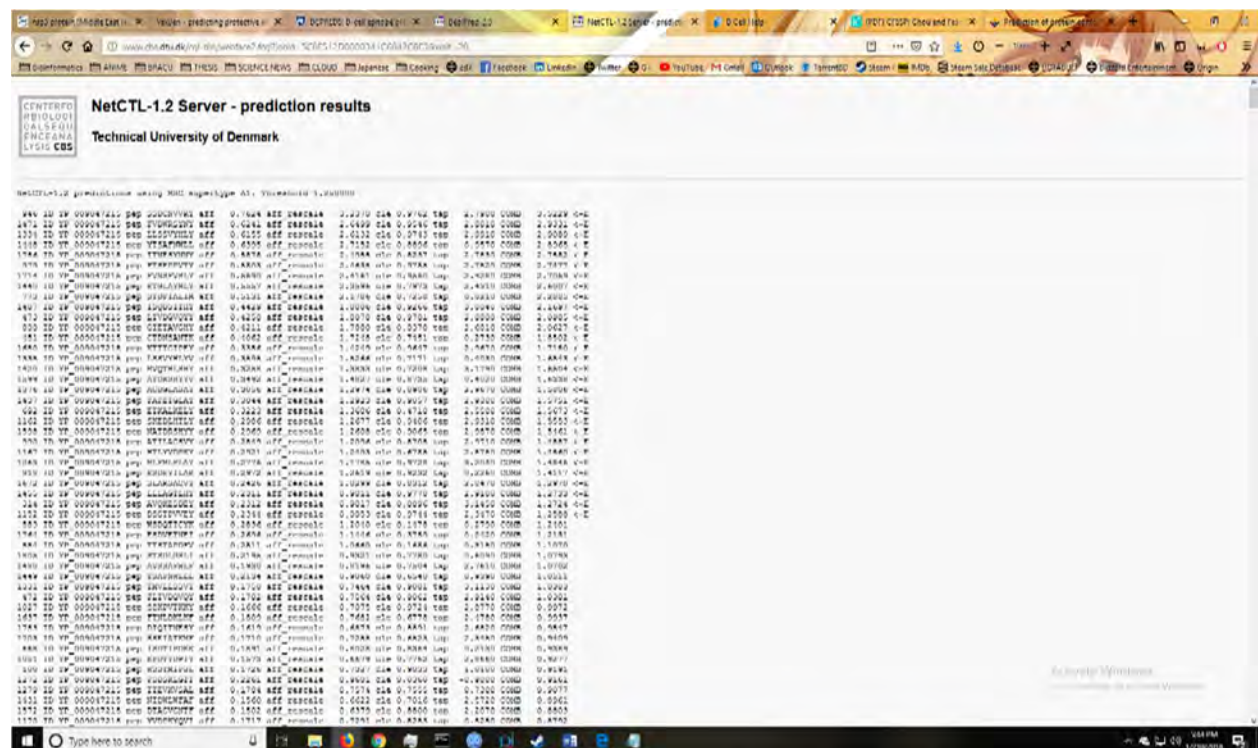


Figure 3.3: Prediction of CD8+ T-cell epitopes using NetCTL 1.2 Server

3.4.2 T-Epitope candidate screening using VaxiJen 2.0 and IEDB Immunogenicity Tool

After using the NetCTL 1.2 server and obtaining 30 T-cell epitope candidates, the predicted T-cell epitopes were further evaluated by the VaxiJen 2.0 server.

Table 3.1: Analysis of predicted T-cell epitope candidates using VaxiJen 2.0

Position	Epitope	NetCTL 1.2 Score	VaxiJen 2.0 Score
1471	FVDWRSYNY	2.9331	1.6645
1167	HTLYVDPKY	1.486	1.2941
1598	NATDRSHYY	1.5461	0.9564
1786	ITNESYNNY	2.7582	0.7731
1455	LLLAGTLHY	1.2733	0.7587
1437	FAFETGLAY	1.5751	0.7437
1599	ATDRSHYYV	1.6338	0.7165
1263	MLFMLPLAY	1.4846	0.6369
1650	KTTTGIPEY	1.718	0.4986
773	STDFIALIM	2.2885	0.4864
1440	ETGLAYMLY	2.6007	0.4838
1335	LSSVYHLYV	1.6543	0.4116
1420	MVQTHLSHY	1.6504	0.3976
899	GIETAVGHY	2.0627	0.2926
451	CTDNSANTK	1.8502	0.2589
1334	LLSSVYHLY	2.9089	0.2519
1448	YTSAFNWLL	2.8965	0.2423
979	FTSEPPVTY	2.7477	0.2059
1376	ACDGLASAY	1.5806	0.1838
314	AVQKESDEY	1.2724	0.1702
1162	SKEDLHTLY	1.5553	0.1463
692	ETKALKELY	1.5673	0.0669
1672	SLARSACVY	1.297	-0.0173
1152	DSGTPVVEY	1.2588	-0.0222
990	ATILAGSVY	1.4887	-0.0235
473	LTVDGVQYY	2.0985	-0.0371
1407	ISQDSITHY	2.1697	-0.196
946	SSDCNVVRY	3.5229	-0.2274
318	ESDEYILAK	1.4117	-0.2766
1714	FVNSFVSLY	2.7059	-0.4734

Among 30 T-cell epitope candidates, only 12 epitope candidates achieved a score over the desired threshold value 0.4.

Table 3.2: Predicted T-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0

Position	Epitope	NetCTL 1.2 Score	VaxiJen 2.0 Score
1471	FVDWRSYNY	2.9331	1.6645
1167	HTLYVDPKY	1.486	1.2941
1598	NATDRSHYY	1.5461	0.9564
1786	ITNESYNNY	2.7582	0.7731
1455	LLLAGTLHY	1.2733	0.7587
1437	FAFETGLAY	1.5751	0.7437
1599	ATDRSHYYV	1.6338	0.7165
1263	MLFMLPLAY	1.4846	0.6369
1650	KTTTGIPEY	1.718	0.4986
773	STDFIALIM	2.2885	0.4864
1440	ETGLAYMLY	2.6007	0.4838
1335	LSSVYHLYV	1.6543	0.4116

These 12 T-cell epitopes were further evaluated by the IEDB T-cell class I pMHC immunogenicity predictor.

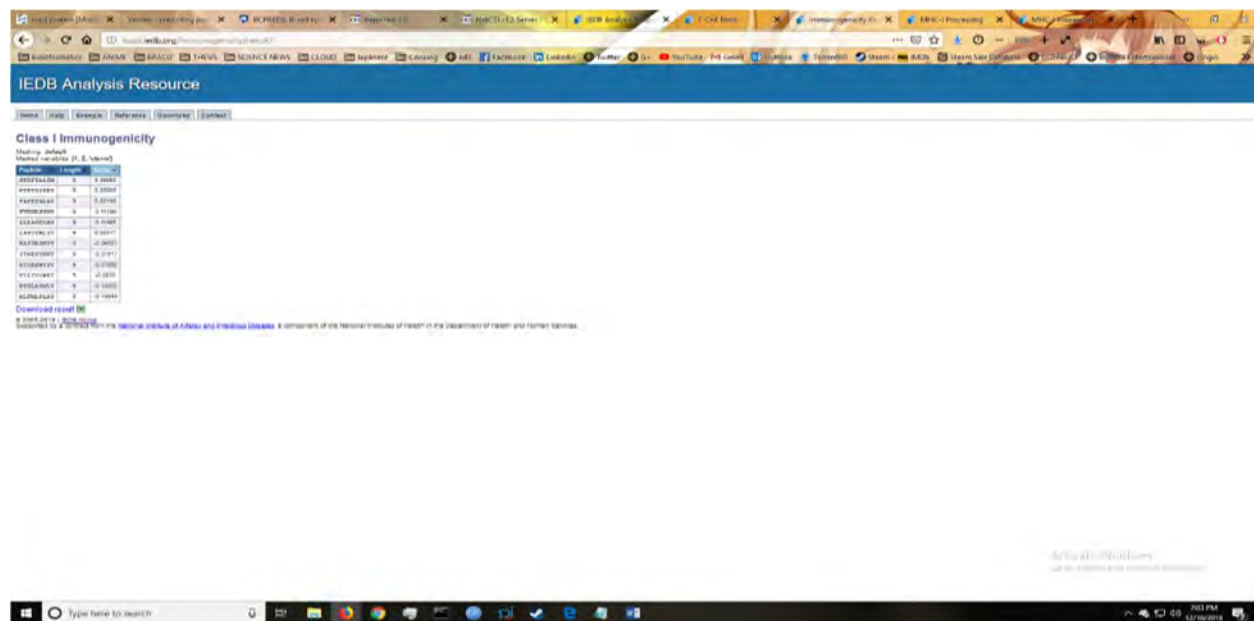


Figure 3.4: Analysis of selected epitope candidates using IEDB T-cell class I pMHC immunogenicity predictor

Table 3.3: Analysis of predicted T-cell epitope candidates using IEDB T-cell class I pMHC immunogenicity predictor

Position	Epitope	NetCTL 1.2 Score	VaxiJen 2.0 Score	Immunogenicity score
773	STDFIALIM	2.2885	0.4864	0.35983
1650	KTTTGIPEY	1.718	0.4986	0.25908
1437	FAFETGLAY	1.5751	0.7437	0.22195
1471	FVDWRSYNY	2.9331	1.6645	0.11786
1455	LLLAGTLHY	1.2733	0.7587	0.11485
1335	LSSVYHLYV	1.6543	0.4116	0.00317
1598	NATDRSHYY	1.5461	0.9564	-0.04527
1786	ITNESYNNY	2.7582	0.7731	-0.07517
1599	ATDRSHYYV	1.6338	0.7165	-0.07665
1167	HTLYVDPKY	1.486	1.2941	-0.0816
1440	ETGLAYMLY	2.6007	0.4838	-0.12022
1263	MLFMLPLAY	1.4846	0.6369	-0.14644

Among these 12 epitope candidates, only 6 epitopes had positive immunogenicity score.

Table 3.4: Predicted T-cell epitope candidates having a positive score in IEDB T-cell class I pMHC immunogenicity predictor

Position	Epitope	NetCTL 2.0 Score	VaxiJen 2.0 Score	Immunogenicity score
773	STDFIALIM	2.2885	0.4864	0.35983
1650	KTTTGIPEY	1.718	0.4986	0.25908
1437	FAFETGLAY	1.5751	0.7437	0.22195
1471	FVDWRSYNY	2.9331	1.6645	0.11786
1455	LLLAGTLHY	1.2733	0.7587	0.11485
1335	LSSVYHLYV	1.6543	0.4116	0.00317

Therefore, these 6 epitope candidates were selected for further evaluation.

3.4.3 Prediction of peptide-MHC class I binding using Proteasomal cleavage/TAP transport/MHC class I combined predictor

For MHC-I processing, the analysis tool of the IEDB generates an overall score for each epitope's intrinsic potential of being a T-cell epitope-based on proteasomal processing, TAP transport, and

MHC-I-binding efficiency. Humans were selected as MHC source species, and the IEDB Recommended method was selected for the prediction of a distinct set of MHC HLA alleles for the humans. This tool gives an output result for HLA-binding affinity of the epitopes in the IC50 nM unit. A lower IC50 value indicates higher binding affinity of the epitopes with the MHC class I molecule. As a rough guideline, peptides with IC50 values <50 nM are considered high affinity, <500 nM intermediate affinity and <5000 nM low affinity. Most known epitopes have high or intermediate affinity. Some epitopes have low affinity, but no known T-cell epitope has an IC50 value greater than 5000. So, in this study, IC50 values less than 200 nM (IC50 < 200nM) were chosen for ensuring higher affinity.

Table 3.5: Prediction of peptide-MHC class I binding of selected T-cell epitope candidates using Proteasomal cleavage/TAP transport/MHC class I combined predictor

Epitope	Interacting MHC-I alleles with high affinity (ic50<200nM)	Total Number of Interacting MHC-I alleles
STDFIALIM	HLA-C*05:01,HLA-A*01:01,HLA-C*15:02,HLA-C*16:01	4
LSSVYHLYV	HLA-C*15:02,HLA-A*02:06,HLA-C*12:03,HLA-A*68:02	4
FAFETGLAY	HLA-B*35:01,HLA-C*03:02,HLA-B*15:25,HLA-C*12:03,HLA-B*15:02,HLA-C*16:01,HLA-C*12:02,HLA-A*29:02,HLA-B*15:01,HLA-C*03:03,HLA-C*14:02,HLA-C*02:02,HLA-C*02:09,HLA-B*53:01,HLA-B*46:01,HLA-C*08:01,HLA-B*18:01,HLA-A*26:01,HLA-A*30:02,HLA-A*68:01	20
LLLAGTLHY	HLA-A*29:02,HLA-B*15:25,HLA-B*15:01,HLA-B*15:02,HLA-A*03:01,HLA-A*30:02,HLA-C*03:02,HLA-B*35:01	8
FVDWRSYNY	HLA-A*01:01,HLA-A*29:02,HLA-B*35:01,HLA-C*16:01	4
KTTTGIPEY	HLA-A*30:02,HLA-B*58:01	2

3.4.4 Prediction of peptide-MHC class I binding using NetMHC 4.0 Server

For prediction of peptide-MHC class I binding using NetMHC 4.0 Server, only 9-mer peptide length was selected as most HLA molecules have a strong preference for binding 9mers. Default value 2.0 was used as a threshold for weak binders while 0.5 was used as a threshold for strong binders.

Table 3.6: Prediction of peptide-MHC class I binding (%Rank<2.0) of selected T-cell epitope candidates using NetMHC 4.0 Server

Epitope	Interacting MHC-I alleles with high affinity (%Rank<2.0)	Total Number of Interacting MHC-I alleles
STDFIALIM	HLA-A*01:01,HLA-A*26:03,HLA-A*32:01,HLA-A*69:01,HLA-A*80:01,HLA-B*15:17,HLA-B*39:01,HLA-C*05:01,HLA-C*08:02,HLA-C*15:02	10
LSSVYHLYV	HLA-A*02:05,HLA-A*01:01,HLA-A*68:23,HLA-A*69:01,HLA-B*15:17,HLA-B*58:01,HLA-C*05:01,HLA-C*15:02	8
FAFETGLAY	HLA-A*01:01,HLA-A*25:01,HLA-A*26:01,HLA-A*26:02,HLA-A*26:03,HLA-A*29:02,HLA-A*30:02,HLA-A*32:07,HLA-A*66:01,HLA-A*68:01,HLA-A*68:23,HLA-A*80:01,HLA-B*08:02,HLA-B*14:02,HLA-B*15:01,HLA-B*15:02,HLA-B*15:03,HLA-B*18:01,HLA-B*27:20,HLA-B*35:01,HLA-B*40:13,HLA-B*46:01,HLA-B*51:01,HLA-B*53:01,HLA-B*58:01,HLA-B*83:01,HLA-C*03:03,HLA-C*05:01,HLA-C*06:02,HLA-C*07:01,HLA-C*07:02,HLA-C*08:02,HLA-C*12:03,HLA-C*14:02,HLA-C*15:02,HLA-A*32:15,HLA-B*15:17	37
LLLAGTLHY	HLA-A*01:01,HLA-A*03:01,HLA-A*29:02,HLA-A*30:02,HLA-A*66:01,HLA-A*68:23,HLA-B*08:02,HLA-B*15:02,HLA-B*15:03,HLA-B*15:17,HLA-B*35:01,HLA-B*58:01,HLA-A*32:15,HLA-A*80:01,HLA-B*15:01	15
FVDWRSYNY	HLA-A*01:01,HLA-A*26:02,HLA-A*26:03,HLA-A*29:02,HLA-A*30:02,HLA-A*32:15,HLA-A*66:01,HLA-A*68:23,HLA-A*80:01,HLA-B*08:02,HLA-B*08:03,HLA-B*15:02,HLA-B*35:01,HLA-B*53:01,HLA-B*83:01,HLA-C*04:01,HLA-C*05:01,HLA-C*06:02,HLA-C*07:01,HLA-C*07:02,HLA-C*08:02,HLA-C*12:03	22
KTTTGIPEY	HLA-A*01:01,HLA-A*25:01,HLA-A*26:02,HLA-A*29:02,HLA-A*30:01,HLA-A*30:02,HLA-A*68:23,HLA-A*80:01,HLA-B*15:03,HLA-B*15:17,HLA-B*46:01,HLA-B*58:01,HLA-B*58:02,HLA-C*14:02	14

3.4.5 Prediction of peptide-MHC class II binding using IEDB Peptide binding to MHC class II molecules predictor

For the prediction of MHC-II alleles and their respective peptide or CD4+ T-cell epitope, identified MHC-I-binding alleles of 6 epitopes were considered. For peptide-MHC class II binding analysis, IEDB Peptide binding to MHC class II molecules predictor was used. In this study, IC₅₀ values less than 3000 nM (IC₅₀ < 3000nM) were chosen.

Table 3.7: Prediction of peptide-MHC class II binding (ic₅₀<3000) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic ₅₀ value
FAFETGLAY	HLA-DRB1*07:01	1433-1447	DWLWFAFETGLAYML	73
	HLA-DRB5*01:01	1434-1448	WLWFAFETGLAYMLY	123
	HLA-DRB5*01:01	1433-1447	DWLWFAFETGLAYML	125
	HLA-DRB1*01:01	1432-1446	IDWLWFAFETGLAYM	200
	HLA-DRB1*12:01	1433-1447	DWLWFAFETGLAYML	304
	HLA-DRB4*01:01	1437-1451	FAFETGLAYMLYTSA	330
	HLA-DRB5*01:01	1432-1446	IDWLWFAFETGLAYM	344
	HLA-DRB3*01:01	1435-1449	LWFAFETGLAYMLYT	552
	HLA-DRB3*01:01	1434-1448	WLWFAFETGLAYMLY	563
	HLA-DRB3*01:01	1433-1447	DWLWFAFETGLAYML	572
	HLA-DRB3*01:01	1432-1446	IDWLWFAFETGLAYM	757
	HLA-DRB1*11:01	1433-1447	DWLWFAFETGLAYML	825
	HLA-DRB1*11:01	1432-1446	IDWLWFAFETGLAYM	1290
	HLA-DRB1*12:01	1432-1446	IDWLWFAFETGLAYM	1592
	HLA-DRB1*04:04	1435-1449	LWFAFETGLAYMLYT	1739
	HLA-DRB1*03:01	1434-1448	WLWFAFETGLAYMLY	2982
	HLA-DPA1*01:03/DPB1*02:01	1434-1448	WLWFAFETGLAYMLY	70
	HLA-DPA1*01:03/DPB1*02:01	1433-1447	DWLWFAFETGLAYML	71
	HLA-DPA1*01:03/DPB1*02:01	1435-1449	LWFAFETGLAYMLYT	106
	HLA-DQA1*05:01/DQB1*03:01	1433-1447	DWLWFAFETGLAYML	170

Table 3.7: Prediction of peptide-MHC class II binding (ic50<3000) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic50 value
FAFETGLAY	HLA-DPA1*01:03/DPB1*02:01	1437-1451	FAFETGLAYMLYTSA	229
	HLA-DPA1*01/DPB1*04:01	1433-1447	DWLWFAFETGLAYML	242
	HLA-DPA1*01/DPB1*04:01	1434-1448	WLWFAFETGLAYMLY	242
	HLA-DPA1*02:01/DPB1*01:01	1435-1449	LWFAFETGLAYMLYT	257
	HLA-DPA1*01/DPB1*04:01	1435-1449	LWFAFETGLAYMLYT	282
	HLA-DPA1*01:03/DPB1*02:01	1436-1450	WFAFETGLAYMLYTS	302
	HLA-DPA1*03:01/DPB1*04:02	1433-1447	DWLWFAFETGLAYML	303
	HLA-DPA1*03:01/DPB1*04:02	1434-1448	WLWFAFETGLAYMLY	307
	HLA-DPA1*03:01/DPB1*04:02	1435-1449	LWFAFETGLAYMLYT	383
	HLA-DQA1*05:01/DQB1*03:01	1432-1446	IDWLWFAFETGLAYM	470
	HLA-DPA1*01/DPB1*04:01	1437-1451	FAFETGLAYMLYTSA	665
	HLA-DPA1*01/DPB1*04:01	1436-1450	WFAFETGLAYMLYTS	711
	HLA-DQA1*03:01/DQB1*03:02	1432-1446	IDWLWFAFETGLAYM	1426
	HLA-DQA1*05:01/DQB1*02:01	1435-1449	LWFAFETGLAYMLYT	1446
	HLA-DQA1*03:01/DQB1*03:02	1434-1448	WLWFAFETGLAYMLY	1608

Table 3.7: Prediction of peptide-MHC class II binding (ic50<3000) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic50 value
FAFETGLAY	HLA-DQA1*03:01/DQB1*03:02	1433-1447	DWLWFAFETGLAYML	1617
	HLA-DQA1*03:01/DQB1*03:02	1435-1449	LWFAFETGLAYMLYT	1773
FVDWRSYNY	HLA-DRB1*07:01	1471-1485	FVDWRSYNYAVSSAF	19
	HLA-DRB1*15:01	1468-1482	TSIFVDWRSYNYAVS	321
	HLA-DRB1*15:01	1469-1483	SIFVDWRSYNYAVSS	356
	HLA-DRB1*15:01	1467-1481	QTSIFVDWRSYNYAV	400
	HLA-DRB1*15:01	1466-1480	AQTSIFVDWRSYNYA	460
	HLA-DRB3*01:01	1471-1485	FVDWRSYNYAVSSAF	481
	HLA-DRB1*01:01	1468-1482	TSIFVDWRSYNYAVS	599
	HLA-DRB1*04:01	1468-1482	TSIFVDWRSYNYAVS	709
	HLA-DRB1*01:01	1467-1481	QTSIFVDWRSYNYAV	970
	HLA-DRB1*07:01	1467-1481	QTSIFVDWRSYNYAV	1015
	HLA-DRB1*01:01	1466-1480	AQTSIFVDWRSYNYA	1292
	HLA-DRB1*12:01	1466-1480	AQTSIFVDWRSYNYA	1309
	HLA-DRB1*12:01	1468-1482	TSIFVDWRSYNYAVS	1327
	HLA-DRB1*12:01	1467-1481	QTSIFVDWRSYNYAV	1390
	HLA-DRB1*12:01	1469-1483	SIFVDWRSYNYAVSS	1450
	HLA-DRB1*12:01	1471-1485	FVDWRSYNYAVSSAF	1507
	HLA-DRB1*04:01	1466-1480	AQTSIFVDWRSYNYA	1900
	HLA-DRB1*04:01	1467-1481	QTSIFVDWRSYNYAV	1974
	HLA-DRB1*04:04	1468-1482	TSIFVDWRSYNYAVS	1975
	HLA-DRB1*04:05	1468-1482	TSIFVDWRSYNYAVS	2132
	HLA-DRB1*12:01	1470-1484	IFVDWRSYNYAVSSA	2309
	HLA-DRB1*07:01	1466-1480	AQTSIFVDWRSYNYA	2366
	HLA-DRB5*01:01	1468-1482	TSIFVDWRSYNYAVS	2958
	HLA-DQA1*01:01/DQB1*05:01	1468-1482	TSIFVDWRSYNYAVS	1569
	HLA-DQA1*01:01/DQB1*05:01	1466-1480	AQTSIFVDWRSYNYA	1637

Table 3.7: Prediction of peptide-MHC class II binding (ic50<3000) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic50 value
FVDWRSYNY	HLA-DQA1*01:01/DQB1*05:01	1469-1483	SIFVDWRSYNYAVSS	1700
	HLA-DQA1*01:01/DQB1*05:01	1467-1481	QTSIFVDWRSYNYAV	2102
	HLA-DQA1*01:01/DQB1*05:01	1470-1484	IFVDWRSYNYAVSSA	2643
	HLA-DQA1*01:01/DQB1*05:01	1471-1485	FVDWRSYNYAVSSAF	2692
KTTTGIPEY	HLA-DQA1*04:01/DQB1*04:02	1650-1664	KTTTGIPEYNFIIYD	1577
	HLA-DQA1*05:01/DQB1*02:01	1648-1662	VCKTTTGIPEYNFII	2486
	HLA-DQA1*05:01/DQB1*02:01	1647-1661	EVCKTTTGIPEYNFI	2547
	HLA-DQA1*05:01/DQB1*02:01	1646-1660	KEVCKTTTGIPEYNF	2990
LLLAGTLHY	HLA-DRB1*04:04	1455-1469	LLLAGTLHYFFAQTS	51
	HLA-DRB1*01:01	1455-1469	LLLAGTLHYFFAQTS	87
	HLA-DRB1*12:01	1450-1464	SAFNWLLLAGTLHYF	137
	HLA-DRB1*12:01	1451-1465	AFNWLLLAGTLHYFF	147
	HLA-DRB1*12:01	1452-1466	FNWLLLAGTLHYFFA	151
	HLA-DRB1*12:01	1453-1467	NWLLLAGTLHYFFAQ	177
	HLA-DRB4*01:01	1455-1469	LLLAGTLHYFFAQTS	203
	HLA-DRB1*04:01	1450-1464	SAFNWLLLAGTLHYF	224
	HLA-DRB1*15:01	1450-1464	SAFNWLLLAGTLHYF	235
	HLA-DRB1*15:01	1452-1466	FNWLLLAGTLHYFFA	261
	HLA-DRB1*04:01	1451-1465	AFNWLLLAGTLHYFF	275
	HLA-DRB1*04:01	1452-1466	FNWLLLAGTLHYFFA	275
	HLA-DRB1*07:01	1453-1467	NWLLLAGTLHYFFAQ	275
	HLA-DRB1*12:01	1455-1469	LLLAGTLHYFFAQTS	342
	HLA-DRB1*04:01	1455-1469	LLLAGTLHYFFAQTS	356

Table 3.7: Prediction of peptide-MHC class II binding (ic50<3000) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic50 value
LLLAGTLHY	HLA-DRB1*15:01	1453-1467	NWLLLAGTLHYFFAQ	373
	HLA-DRB1*03:01	1452-1466	FNWLLLAGTLHYFFA	382
	HLA-DRB1*03:01	1451-1465	AFNWLLLAGTLHYFF	383
	HLA-DRB1*03:01	1453-1467	NWLLLAGTLHYFFAQ	391
	HLA-DRB1*03:01	1450-1464	SAFNWLLLAGTLHYF	394
	HLA-DRB1*04:01	1453-1467	NWLLLAGTLHYFFAQ	421
	HLA-DRB5*01:01	1453-1467	NWLLLAGTLHYFFAQ	444
	HLA-DRB1*09:01	1455-1469	LLLAGTLHYFFAQTS	775
	HLA-DRB1*04:01	1454-1468	WLLLAGTLHYFFAQT	800
	HLA-DRB1*13:02	1453-1467	NWLLLAGTLHYFFAQ	845
	HLA-DRB1*13:02	1452-1466	FNWLLLAGTLHYFFA	852
	HLA-DRB1*13:02	1451-1465	AFNWLLLAGTLHYFF	854
	HLA-DRB1*13:02	1450-1464	SAFNWLLLAGTLHYF	862
	HLA-DRB3*01:01	1450-1464	SAFNWLLLAGTLHYF	912
	HLA-DRB1*03:01	1455-1469	LLLAGTLHYFFAQTS	1135
	HLA-DRB1*03:01	1454-1468	WLLLAGTLHYFFAQT	1154
	HLA-DRB1*09:01	1453-1467	NWLLLAGTLHYFFAQ	1587
	HLA-DRB1*13:02	1454-1468	WLLLAGTLHYFFAQT	2153
	HLA-DRB1*13:02	1455-1469	LLLAGTLHYFFAQTS	2431
	HLA-DPA1*03:01/DPB1*04:02	1451-1465	AFNWLLLAGTLHYFF	558
	HLA-DPA1*03:01/DPB1*04:02	1452-1466	FNWLLLAGTLHYFFA	576
	HLA-DPA1*03:01/DPB1*04:02	1453-1467	NWLLLAGTLHYFFAQ	678
	HLA-DPA1*02:01/DPB1*01:01	1452-1466	FNWLLLAGTLHYFFA	697
	HLA-DPA1*02:01/DPB1*01:01	1453-1467	NWLLLAGTLHYFFAQ	697
	HLA-DPA1*02:01/DPB1*01:01	1451-1465	AFNWLLLAGTLHYFF	703

Table 3.7: Prediction of peptide-MHC class II binding (ic50<3000) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic50 value
LLLAGTLHY	HLA-DPA1*02:01/DPB1*01:01	1450-1464	SAFNWLLLAGTLHYF	783
	HLA-DPA1*03:01/DPB1*04:02	1455-1469	LLLAGTLHYFFAQTS	1033
	HLA-DPA1*03:01/DPB1*04:02	1454-1468	WLLLAGTLHYFFAQT	1187
LSSVYHLYV	HLA-DRB1*07:01	1330-1344	TTMVLLSSVYHLYVF	47
	HLA-DRB1*07:01	1331-1345	TMVLLSSVYHLYVFN	48
	HLA-DRB1*07:01	1332-1346	MVLLSSVYHLYVFNQ	61
	HLA-DRB1*07:01	1333-1347	VLLSSVYHLYVFNQV	65
	HLA-DRB1*15:01	1331-1345	TMVLLSSVYHLYVFN	113
	HLA-DRB1*07:01	1334-1348	LLSSVYHLYVFNQVL	137
	HLA-DRB1*01:01	1333-1347	VLLSSVYHLYVFNQV	138
	HLA-DRB1*07:01	1335-1349	LSSVYHLYVFNQVLS	145
	HLA-DRB1*15:01	1332-1346	MVLLSSVYHLYVFNQ	148
	HLA-DRB1*09:01	1333-1347	VLLSSVYHLYVFNQV	150
	HLA-DRB1*09:01	1330-1344	TTMVLLSSVYHLYVF	162
	HLA-DRB1*09:01	1331-1345	TMVLLSSVYHLYVFN	168
	HLA-DRB1*11:01	1330-1344	TTMVLLSSVYHLYVF	175
	HLA-DRB1*09:01	1332-1346	MVLLSSVYHLYVFNQ	175
	HLA-DRB1*04:05	1333-1347	VLLSSVYHLYVFNQV	197
	HLA-DRB1*15:01	1333-1347	VLLSSVYHLYVFNQV	208
	HLA-DRB1*04:05	1332-1346	MVLLSSVYHLYVFNQ	217
	HLA-DRB1*04:05	1330-1344	TTMVLLSSVYHLYVF	239
	HLA-DRB1*11:01	1331-1345	TMVLLSSVYHLYVFN	252
	HLA-DRB1*11:01	1332-1346	MVLLSSVYHLYVFNQ	252
	HLA-DRB1*04:05	1331-1345	TMVLLSSVYHLYVFN	289
	HLA-DRB1*11:01	1333-1347	VLLSSVYHLYVFNQV	298
	HLA-DRB1*09:01	1335-1349	LSSVYHLYVFNQVLS	322
	HLA-DRB1*09:01	1334-1348	LLSSVYHLYVFNQVL	329
	HLA-DRB1*15:01	1334-1348	LLSSVYHLYVFNQVL	342
	HLA-DRB4*01:01	1330-1344	TTMVLLSSVYHLYVF	361
	HLA-DRB4*01:01	1333-1347	VLLSSVYHLYVFNQV	387
	HLA-DRB4*01:01	1331-1345	TMVLLSSVYHLYVFN	412

Table 3.7: Prediction of peptide-MHC class II binding (ic50<3000) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic50 value
LSSVYHLYV	HLA-DRB1*11:01	1335-1349	LSSVYHLYVFNQVLS	524
	HLA-DRB5*01:01	1331-1345	TMVLLSSVYHLYVFN	538
	HLA-DRB5*01:01	1332-1346	MVLLSSVYHLYVFNQ	546
	HLA-DRB4*01:01	1332-1346	MVLLSSVYHLYVFNQ	571
	HLA-DRB1*04:04	1333-1347	VLLSSVYHLYVFNQV	628
	HLA-DRB1*11:01	1334-1348	LLSSVYHLYVFNQVL	677
	HLA-DRB5*01:01	1333-1347	VLLSSVYHLYVFNQV	899
	HLA-DRB5*01:01	1335-1349	LSSVYHLYVFNQVLS	1304
	HLA-DRB5*01:01	1334-1348	LLSSVYHLYVFNQVL	1305
	HLA-DRB1*08:02	1330-1344	TTMVLLSSVYHLYVF	1482
	HLA-DRB1*08:02	1331-1345	TMVLLSSVYHLYVFN	1820
	HLA-DRB1*08:02	1332-1346	MVLLSSVYHLYVFNQ	1926
	HLA-DRB1*08:02	1333-1347	VLLSSVYHLYVFNQV	2142
STDFIALIM	HLA-DRB1*01:01	770-784	GGDSTDFIALIMAYG	56
	HLA-DRB5*01:01	770-784	GGDSTDFIALIMAYG	312
	HLA-DRB1*09:01	770-784	GGDSTDFIALIMAYG	373
	HLA-DRB1*04:01	770-784	GGDSTDFIALIMAYG	780
	HLA-DRB1*08:02	770-784	GGDSTDFIALIMAYG	862
	HLA-DQA1*05:01/DQB1*03:01	769-783	KGGDSTDFIALIMAY	726
	HLA-DQA1*05:01/DQB1*03:01	768-782	HKGGDSTDFIALIMA	1373
	HLA-DQA1*03:01/DQB1*03:02	768-782	HKGGDSTDFIALIMA	1685
	HLA-DQA1*05:01/DQB1*02:01	769-783	KGGDSTDFIALIMAY	1867
	HLA-DQA1*01:01/DQB1*05:01	769-783	KGGDSTDFIALIMAY	2342
	HLA-DQA1*01:01/DQB1*05:01	768-782	HKGGDSTDFIALIMA	2365

Table 3.7: Prediction of peptide-MHC class II binding ($ic_{50} < 3000$) using IEDB Peptide binding to MHC class II molecules predictor

Core peptide	Allele	Position	Peptide	ic_{50} value
STDFIALIM	HLA-DQA1*01:01/DQB1*05:01	770-784	GGDSTDFIALIMAYG	2569
	HLA-DQA1*01:01/DQB1*05:01	771-785	GDSTDFIALIMAYGN	2882

3.4.6 Population Coverage Analysis

MHC HLA allele distribution differs among diverse geographic regions and ethnic groups around the world. Therefore, population coverage must be taken into consideration during the design of an effective vaccine. In this study, identified MHC-I-binding alleles of 6 epitopes were considered to analyze population coverage. For population coverage, the IEDB Population coverage tool was used. MHC-I binding alleles with high binding affinities that had been found using both the IEDB tool and the NetMHC 2.0 server were selected.

Population: World

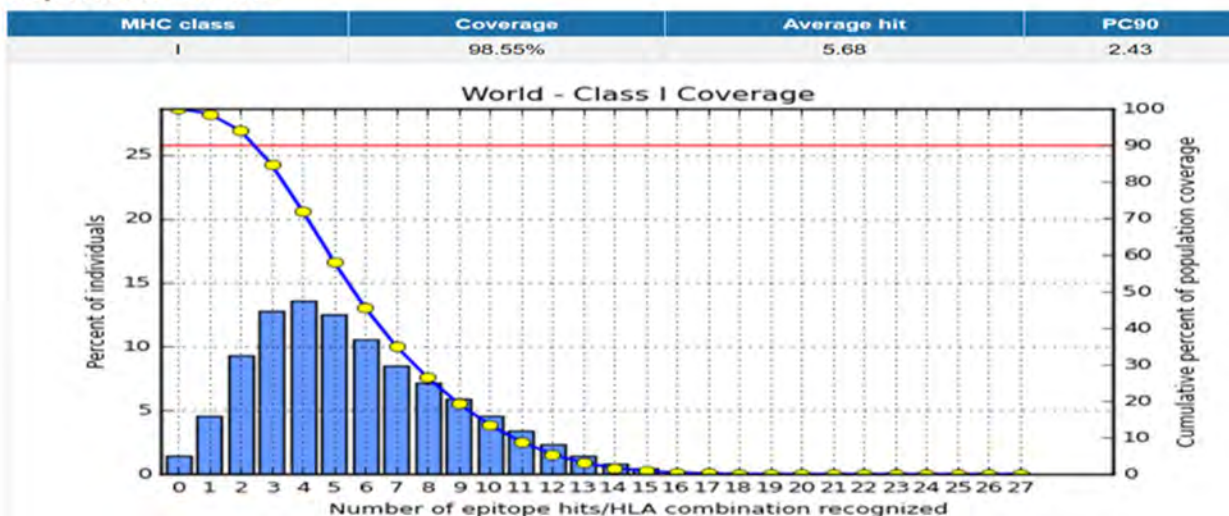


Figure 3.5: Population coverage of the epitopes and their respective HLA alleles showing the number of epitope hits against cumulative percent of population coverage

These epitopes and their HLA-alleles cover 98.55% of the world population combined.

population/area	Class I		
	coverage ^a	average_hit ^b	pc90 ^c
Central Africa	99.09%	6.06	2.82
Central America	9.07%	0.25	0.11
East Africa	98.73%	6.35	2.73
East Asia	96.86%	4.08	1.76
Europe	99.65%	6.49	3.13
North Africa	99.46%	6.68	3.17
North America	98.56%	5.77	2.46
Northeast Asia	94.95%	4.05	1.57
Oceania	87.56%	3.34	0.8
South Africa	99.66%	6.92	3.36
South America	91.44%	3.75	1.13
South Asia	97.48%	5.15	2.09
Southeast Asia	94.84%	3.84	1.5
Southwest Asia	96.4%	4.92	1.8
West Africa	99.41%	6.71	3.24
West Indies	87.22%	3.71	0.78
World	98.55%	5.68	2.43
Average	91.11	4.93	2.05
Standard deviation	20.87	1.67	0.95

^a projected population coverage
^b average number of epitope hits / HLA combinations recognized by the population
^c minimum number of epitope hits / HLA combinations recognized by 90% of the population

Figure 3.6: Population coverage of the epitopes and their respective HLA alleles

The highest population coverage was found in the South Africa region (99.66%) while the lowest population coverage was found in Central America (9.07%). The cumulative population coverage in Southwest Asia was 96.40%.


3.4.7 Conservancy and Toxicity Prediction

Conservancy of epitopes was checked using the IEDB conservancy analysis tool. All predicted epitope candidates had the maximum identity (100%) for conservancy hit.


IEDB Analysis Resource

[Home](#)[Help](#)[Example](#)[Reference](#)[Download](#)[Contact](#)

Epitope Conservancy Analysis Result

[Download result](#) 

Epitope #	Epitope name	Epitope sequence	Epitope length	Percent of protein sequence matches at identity <= 100%	Minimum identity	Maximum identity	View details
1	1	STDFIALIM	9	100.00% (1/1)	100.00%	100.00%	Go
2	2	LSSVYHLYV	9	100.00% (1/1)	100.00%	100.00%	Go
3	3	FAFEIGLAY	9	100.00% (1/1)	100.00%	100.00%	Go
4	4	LLLASTLHY	9	100.00% (1/1)	100.00%	100.00%	Go
5	5	FVDWRSYNY	9	100.00% (1/1)	100.00%	100.00%	Go
6	6	KITTIPIEY	9	100.00% (1/1)	100.00%	100.00%	Go

[Download result](#) 

© 2005-2019 | [IEDB Home](#)
Supported by a contract from the [National Institute of Allergy and Infectious Diseases](#), a component of the National Institutes of Health in the Department of Health and Human Services.

Figure 3.7: Epitope conservancy analysis showing all selected T-cell epitopes had 100% conservancy

A qualified vaccine should induce a specific immune response that targets only the virus rather than the host tissue. To ensure that cellular immunity induced by the selected epitopes would not damage host tissue, toxicity prediction was carried out using ToxinPred. All six of the selected epitopes were found as non-toxic.

3.4.8 Allergenicity Prediction

Allergenicity prediction was performed using AllergenFP v1.0 and AllerTOP v2.0.

Table 3.8: Allergenicity prediction using AllergenFP v1.0 and AllerTOP v2.0

Epitope	VaxiJen 2.0 Score	Immunogenicity	Toxicity	AllergenFP v1.0	AllerTOP v2.0
STDFIALIM	0.4864	0.35983	Non-Toxin	PROBABLE ALLERGEN	PROBABLE NON-ALLERGEN
LSSVYHLYV	0.4116	0.00317	Non-Toxin	PROBABLE ALLERGEN	PROBABLE NON-ALLERGEN
FAFETGLAY	0.7437	0.22195	Non-Toxin	PROBABLE ALLERGEN	PROBABLE ALLERGEN
LLLAGTLHY	0.7587	0.11485	Non-Toxin	PROBABLE ALLERGEN	PROBABLE ALLERGEN
FVDWRSYNY	1.6645	0.11786	Non-Toxin	PROBABLE ALLERGEN	PROBABLE ALLERGEN
KTTTGIPEY	0.4986	0.25908	Non-Toxin	PROBABLE NON-ALLERGEN	PROBABLE NON-ALLERGEN

3.4.9 Docking result and Analysis

HLA-A*01:01 was selected as macromolecule for docking purpose since it was found common in all six selected epitope candidates, and the 3D structure of HLA-A*01:01 allele was readily available in the PDB database. The crystal structure of the HLA-A*01:01 protein molecule (PDB ID:4nqx) was retrieved from the RCSB Protein Data Bank in PDB format. Since this predicted crystal structure of HLA-A*01:01 allele was in a complex form with Beta-2-microglobulin and NP44-S7N mutant peptide (CTELKLNDY), UCSF Chimera 1.13 was used to remove undesired

ligands and molecules. The 3D structures of selected epitope candidates were predicted using PEP-FOLD 2.0 server. Besides these T-cell epitope candidates, NP44-S7N mutant peptide (CTELKLNDY) was also selected for docking as ligand so that it can be used as control later. Then, ligands and macromolecule were prepared for docking using the same software. After the minimizing process, PyRx was used for molecular docking. HLA-A*01:01 protein was placed in a grid box measuring $52.8351 \text{ \AA} \times 68.2709 \text{ \AA} \times 61.7293 \text{ \AA}$ along the x, y and z axis, respectively, where the position of the center was X:-63.5001, Y:-17.1718, Z:7.5672. The docking procedure was performed using the instructed command prompts. The docking poses were ranked according to their docking scores. However, the docking result revealed two different binding sites.

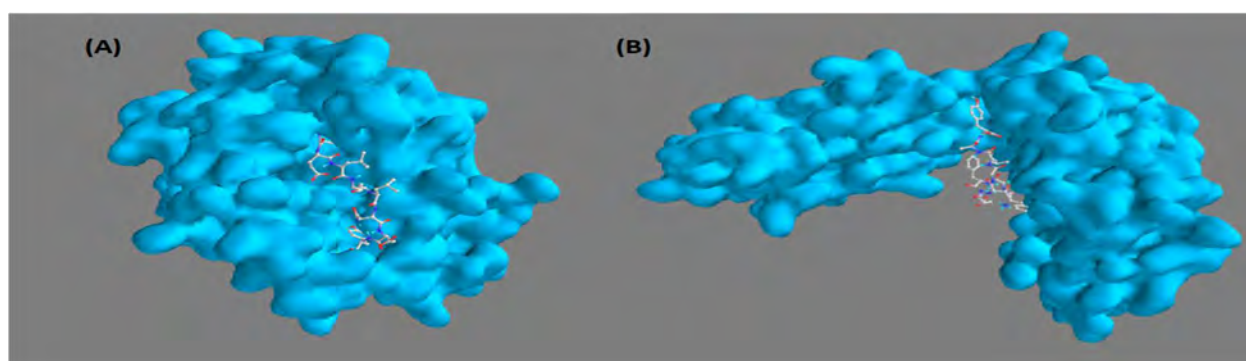


Figure 3.8: PyRx docking result revealed two different binding sites: (A) binding site of NP44-S7N mutant peptide (B) alternative binding site

Therefore, instead of choosing the conformation with lowest binding affinity as best docking pose, the conformation with the lowest binding affinity that used the same binding site as control (NP44-S7N mutant peptide) was selected as best docking pose in order to compare between sample and control for critical evaluation.

Table 3.9: Binding affinity of best docking pose against HLA-A*01:01 allele

Name	Binding Affinity (kcal/mol)
NP44-S7N mutant peptide	-8.7
LSSVYHLYV	-8
FVDWRSYNY	-7.7
LLLAGTLHY	-7.7
FAFETGLAY	-7.5
KTTTGIPEY	-7.5
STDFIALIM	-6

Then, molecular visualization of the best docking poses was performed using UCSF Chimera 1.13. Only FAFETGLAY formed visible intermolecular hydrogen bond similar to the control ligand (NP44-S7N mutant peptide).

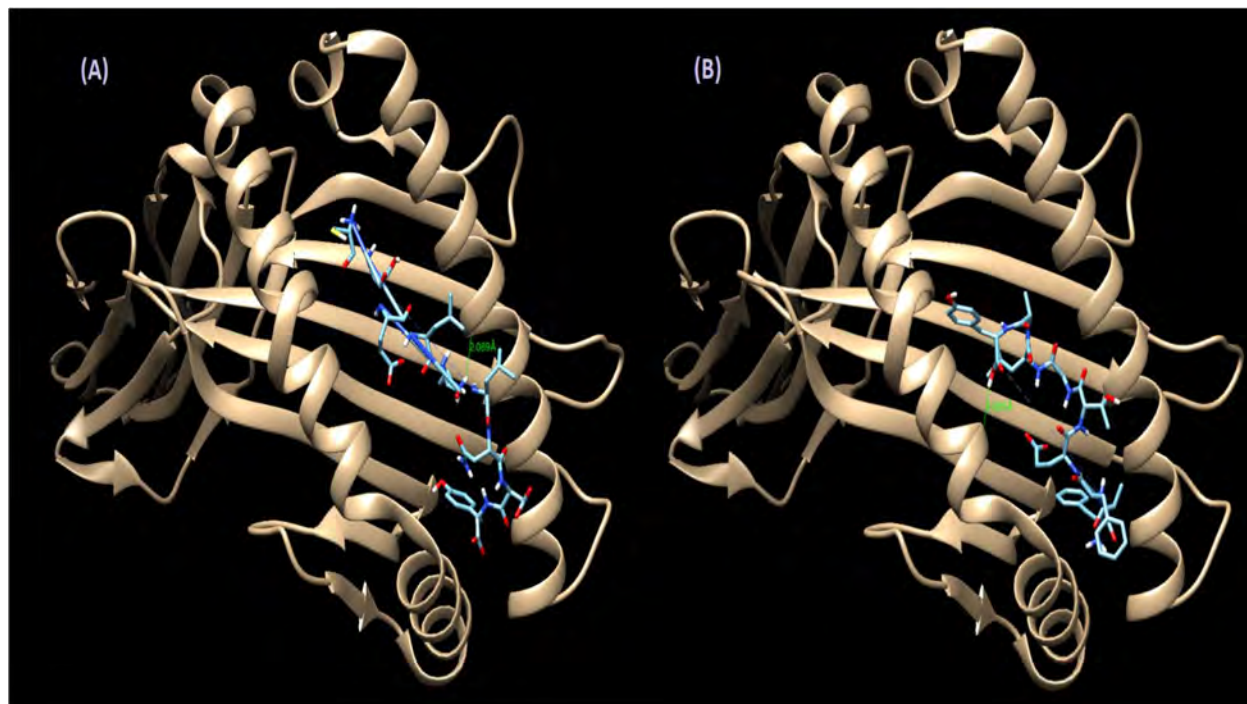


Figure 3.9: (A) Intermolecular bonding interaction between FAFETGLAY and HLA-A*01:01 allele with a binding affinity of -7.5 kcal/mol (B) Intermolecular bonding interaction between NP44-S7N mutant peptide and HLA-A*01:01 allele with a binding affinity of -8.7 kcal/mol

3.5 Prediction of B-Cell epitopes

After the antigenicity of nsp3 protein was confirmed, the sequence was put in B-cell epitope predictor tools to identify probable B-cell epitopes in a sequence. Selection of B-cell epitopes was done using BCPREDS and BepiPred 2.0.

3.5.1 BCPREDS

BCPREDS has 2 different modes for epitope prediction: fixed length epitope prediction and flexible length epitope prediction. For this case, the fixed length epitope prediction method was

used. Epitopes of 12 to 20 amino acid in length were desired for fixed length epitope prediction for selection. A specificity level of 75% was selected for all cases.

Table 3.10: Predicted B-cell epitopes using BCPREDS

Position	Epitope	Epitope length	Score
588	HWSDQTICYKDS	12	0.71
221	IQPEVKEVAPVY	12	0.94
925	VSKTSDWKCKVT	12	0.85
1471	FVDWRSYNYAVS	12	0.64
1169	LYVDPKYQVIVL	12	0.37
884	TTSTAPDFVAFN	12	0.7
436	GAIRKAKDYGFT	12	0.92
729	VRSLKLSDNNCY	12	0.3
667	DISDTIPDEKQN	12	0.76
1843	SDALKRQIRIAC	12	0.51
115	EECSEVEASDLE	12	0.54
348	HVVGPDARAKQD	12	0.99
648	LNNKNTYRSQLG	12	0.97
1299	KQCCTAAVDLSM	12	0.66
1677	ACVYYSQVLCKS	12	0.46
1192	HTVESGDINVVA	12	0.46
392	VSFDYLIREAKT	12	0.61
246	PKRLRKKRNVDP	12	0.71
522	VVRRVNVPPYVCL	12	1
1856	KCNLAFRLTTSK	12	0.75
134	SETSTEQVDVSH	12	0.98
904	VGHYVHARLKGG	12	0.85
1278	GTTEVKVSALKT	12	0.94
92	FNAEGDASWSST	12	0.84
1110	VESTPVEPPTVD	12	1
1250	TKGILTGCFSFA	12	0.73
168	VTESVQEEAQPV	12	1
535	ANKEQEAILMSE	12	0.37
295	THLKHGGGIAGA	12	1
1567	NCVDCDTAGVGN	12	0.92
194	TLQETPVVPDTV	12	0.93
324	LAKGPLQVGDSV	12	0.31
1546	GKRTFYITANGG	12	0.42
1601	DRSHYYVDSVTV	12	0.4
372	AYPLVVTPLVSA	12	0.78

Table 3.10: Predicted B-cell epitopes using BCPREDS

Position	Epitope	Epitope length	Score
481	YCYTSKDTLDDI	12	0.79
405	VLVVVNSQDVYK	12	1
627	TQQLTIEVLVTV	12	0.95
1697	VTSVGDSSEIAT	12	0.57
1648	VCKTTTGIPEYN	12	1
7	AFGGDQVHEVAA	12	0.42
73	GMPIPDFDLDDF	12	0.78
1793	NYVPSYVKPDSV	12	0.94
422	DIPQSLTFSYDG	12	0.82
1124	ALQQEMTIVKCK	12	0.75
807	LCCSARMVWREW	12	0.64
449	FVCTDNSANTKV	12	0.99
764	AFMKHKGGDSTD	12	0.56
1139	KPFVKDNVSFVA	12	0.72
1755	DAARGPAGVESD	12	0.98
1768	ETNEIVDSVQYA	12	0.94
559	HVRTNGGYN SWH	12	0.67
264	KVITECVTIVLG	12	0.38
1005	SSDGQPGGDAIS	12	1
1622	RRDGQPFYERFP	12	0.86
208	PPQVVKLPSAPQ	12	0.86
1226	ATRTFTATTAVG	12	0.84
950	NVVRYSLDGNFR	12	0.83
853	VCQCGGERHRQL	12	0.95
1661	IYDSSDRGQES	12	0.63
1409	QDSITHYPALKM	12	0.31
689	TADETKALKELY	12	0.78
1876	SVRFTANKIVGG	12	0.47
786	CTFGAPDDASRL	12	1
1737	RDGVRRGDNFHS	12	0.66
1058	IYKNGAMYKGKP	12	0.88
974	KDGKYFTSEPPV	12	0.95
1073	VNKASYDTNLNK	12	0.91
1531	RNRLTRVEASTV	12	0.47
1025	FDSSKPVTKKYT	12	0.33
1388	SFDVPTFCANRS	12	0.3
1722	YNVTRDKLEKLI	12	0.84
925	VSKTSDWKCKVTDV	14	0.75

Table 3.10: Predicted B-cell epitopes using BCPREDS

Position	Epitope	Epitope length	Score
587	LHWSDQTICYKDSV	14	0.79
1565	NWNCVDCDTAGVGN	14	0.81
647	VLNNKNTYRSQ LGC	14	0.88
1470	IFVDWRSYNYAVSS	14	0.78
1614	ETVVQFNYYRRDQGP	14	0.9
883	VTTSTAPDFVAFNV	14	0.9
1277	LGTTEVKVSALKTA	14	0.74
520	GSVVRRVNVVPYVCL	14	0.98
222	QPEVKEVAPVYEAD	14	0.98
366	CYKAMNAYPLVVTP	14	0.79
1192	HTVESGDINVVAAS	14	0.91
90	YCFNAEGDASWSST	14	0.84
105	IFSLHPVECDEECS	14	0.93
1644	KFKEVCKTTTGIPE	14	0.98
348	HVVGPDARAKQDVS	14	0.98
423	IPQSLTFSYDGLRG	14	0.86
1250	TKGILTGCFSFAKM	14	0.75
901	ETAVGHYVHARLKG	14	0.89
126	EEGESECISETSTE	14	0.81
199	PVVPDTVEVPPQVV	14	0.99
286	ESVLVNAANTHLKH	14	0.81
243	TVKPKRLRKKRNVD	14	0.97
1111	ESTPVEPPTVDVVA	14	1
9	GGDQVHEVA AVR SV	14	0.81
1768	ETNEIVDSVQYAHK	14	0.85
172	VQEEAQPV E PVED	14	1
695	ALKELYGPVDPTFL	14	0.86
1411	SITHYPALKMVQTH	14	0.8
1825	LRNSNGACIWNA AA	14	0.72
305	GAINAASKGAVQKE	14	0.9
1787	TNESYNNYVPSYVK	14	0.79
669	SDTIPDEKQNGHSL	14	0.8
1590	TTALRRPINATDRS	14	0.72
617	ACRAYLDSRTTQQL	14	0.71
851	TYVCQCGGERHRQL	14	0.86
1296	NVVKQCCTAAVDLS	14	0.92
764	AFMKHKGGDSTDFI	14	0.94
1004	VSSDGQPGGDAISL	14	1

Table 3.10: Predicted B-cell epitopes using BCPREDS

Position	Epitope	Epitope length	Score
399	REAKTRVLVVVNSQ	14	0.99
784	GNCTFGAPDDASRL	14	0.98
1035	YTYSFLPKEDGDVL	14	0.9
1750	LTTFIDAARGPAGV	14	0.99
720	GWKMVVCDKVRSLK	14	0.8
1735	TARDGVRRGDNFHS	14	0.83
1137	LNKPFVKDNVSFVA	14	0.76
974	KDGKYFTSEPPVTY	14	0.96
949	CNVVRYSLDGNFRT	14	0.9
556	FIKHVRTNGGYNW	14	0.8
1530	KRNRLTRVEASTVV	14	0.88
448	VFVCTDNSANTKVL	14	0.93
1071	LWVNKASYDTNLNK	14	0.88
1223	KEFATRTFTATTAV	14	0.94
1563	RHNWNCVDCDTAGVGN	16	0.93
646	VVLNNKNTYRSQLGCV	16	0.94
921	DSGTVSKTSDWKCKVT	16	0.91
1643	LKFKEVCKTTTGPIEY	16	1
119	EVEASDLEEGESECIS	16	1
1471	FVDWRSYNYAVSSAFW	16	0.94
1612	VKETVVQFNYYRRDGQP	16	0.98
663	FNGADISDTIPDEKQN	16	0.99
848	ARMTYVCQCGERHRQ	16	0.93
1106	TPLSVESTPVEPPTVD	16	1
519	AGSVVRRVNVVPYVCLL	16	1
88	PCYCFNAEGDASWSST	16	0.98
222	QPEVKEVAPVYEADTE	16	1
877	TPNEKLVTSTAPDFV	16	0.96
1192	HTVESGDINVVAASGS	16	0.92
288	VLVNAANTHLKHGGGI	16	0.98
142	DVSHETSDDEWAAAVD	16	0.95
245	KPKRLRKKRNVDPLSN	16	0.92
1790	SYNNYVPSYVKPDSVS	16	0.99
169	TESVQEEAQPEVPVE	16	1
1695	SLVTSVGDSSSEIATKM	16	0.91
399	REAKTRVLVVVNSQDV	16	1
205	VEVPPQVVKLPSAPQT	16	1
999	TNSCLVSSDGQPGGDA	16	1

Table 3.10: Predicted B-cell epitopes using BCPREDS

Position	Epitope	Epitope length	Score
1750	LTTFIDAARGPAGVES	16	0.99
1055	YDPIYKNGAMYKGKPI	16	0.96
948	DCNVVRYSLDGNFRTE	16	0.95
782	AYGNCTFGAPDDASRL	16	1
974	KDGKYFTSEPPVTYSP	16	1
1150	ADDSGTPVVEYLSKED	16	0.98
762	QHAFMKHKGGDSTDFI	16	0.97
42	RTFVVDKSLSIEEFAD	16	0.94
555	DFIKHVRTNGGYNWSH	16	0.96
1221	YFKEFATRTFTATTAV	16	0.97
447	TVFVCTDNSANTKVLR	16	0.98
1561	CRRHNWNCVDCDTAGVGN	18	0.81
921	DSGTVSKTSDWKCKVTDV	18	0.8
1643	LKFKEVCKTTTGIPEYNF	18	0.96
524	RRVNPYPVCLLANKEQEA	18	0.56
1610	VTVKETVVQFNYYRRDGQP	18	0.86
106	FSLHPVECDEECSEVEAS	18	0.94
1471	FVDWRSYNYAVSSAFWLF	18	0.62
1108	LSVESTPVEPPTVDVVAL	18	0.99
1272	FSDSKLGTTEVKVSALKT	18	0.72
1066	KGKPIWVWVKASYDTNLN	18	0.77
218	PQTIQPEVKEVAPVYEAD	18	1
243	TVKPKRLRKKRNVDPLSN	18	0.91
875	SGTPNEKLVTSTAPDFV	18	0.82
663	FNGADISDTIPDEKQNGH	18	0.65
591	DQTICYKDSVFYVVKNST	18	0.58
5	KVAFGGDQVHEVAAVRSV	18	0.85
1004	VSSDGQPGGDAISLSFNN	18	1
85	IDAPCYCFNAEGDASWSS	18	0.79
898	QGIETAVGHYVHARLKGG	18	0.6
365	KCYKAMNAYPLVVTPLVS	18	0.62
344	KNILHVVGPDARAKQDVS	18	0.91
1536	RVEASTVVC GGKRTFYIT	18	0.57
693	TKALKELYGPVDPTFLHR	18	0.61
197	ETPVVPDTVEVPPQVVKL	18	0.98
168	VTESVQEEAQPVEVPVED	18	1
809	CSARMVWREWCNVC GIKD	18	0.64
1292	VVTGNVVKQCCTAAVDLS	18	0.88

Table 3.10: Predicted B-cell epitopes using BCPREDS

Position	Epitope	Epitope length	Score
1787	TNESYNNYVPSYVKPDSV	18	0.84
443	DYGFTVFVCTDNSANTKV	18	0.83
298	KHGGGIAGAINAASKGAV	18	0.98
854	CQCGGERHRQLVEHTTPW	18	0.84
1220	FYFKEFATRTFTATTAVG	18	0.96
780	IMAYGNCTFGAPDDASRL	18	0.96
974	KDGKYFTSEPPVTYSPAT	18	0.94
399	REAKTRVLVVNSQDVYK	18	0.9
1025	FDSSKPVTKKYTYSF LPK	18	0.79
1756	AARGPAGVESDVETNEIV	18	0.98
1383	AYRANSFDVPTFCANRSA	18	0.73
1139	KPFVKDNVSFVADDSGTP	18	0.75
279	QVAKCYGESVLVNAANTH	18	0.58
551	NPSDFIKHVRTNGGYN	18	0.83
1558	ISFCRRHNWNCVDCDTAGVG	20	0.81
1104	KFTPLSVESTPVEPPTVDVV	20	0.98
919	KFDSGTVSKTSDWKCKVTDV	20	0.92
241	NVTVKPKRLRKKRNVDP LSN	20	0.99
522	VVRRVNPYPVCLLANKEQEA	20	0.86
405	VLVVVNSQDVYKSLTIVDIP	20	0.79
112	ECDEECSEVEASDLEEGESE	20	1
215	PSAPQTIQPEVKEVAPVYEA	20	1
1612	VKETVVQFNYYRRDGQPFYER	20	0.94
663	FNGADISDTIPDEKQNGHSL	20	0.75
8	FGGDQVHEVA AVRSVTVEYN	20	0.8
86	DAPCYCFNAEGDASWSSTMI	20	0.93
194	TLQETPVVPDTVEVPPQVVK	20	1
139	EQVDVSHETSDDEWAAAVDE	20	0.9
951	VVRYSLDGNFRTEVDPDLA	20	0.83
1645	FKEVCKTTTGIPEYNFIYD	20	1
439	RKAKDYGFTVFVCTDNSANT	20	0.82
1788	NESYNNYVPSYVKPDSVSTS	20	0.99
691	DETKALKELYGPVDPTFLHR	20	0.77
166	EDVTESVQEEAQPVEVPVED	20	1
473	LTVDGVQYYCYTSKDTLDDI	20	0.77
812	RMVWREWCNVC GIKDVVLQG	20	0.79
341	SLAKNILHVVGPDARAKQDV	20	0.79
304	AGAINAASKGAVQKESDEYI	20	0.95

Table 3.10: Predicted B-cell epitopes using BCPREDS

Position	Epitope	Epitope length	Score
1027	SSKPVTKKYTYSF LPKEDGD	20	0.83
1053	DTYDPIYKNGAMYKGK PILW	20	0.88
1142	VKDNVSFVADDSGTPVVEYL	20	0.92
974	KDGKYFTSEPPVTYSPATIL	20	1
1756	AARGPAGVESDVETNEIVDS	20	0.97
995	GSVYTNSCLVSSDGQPGGDA	20	1
780	IMAYGNCTFGAPDDASRL LH	20	0.99
1221	YFKEFATRTFTATTAVGSCI	20	0.9

3.5.2 BepiPred 2.0

Besides predicting epitopes, BepiPred 2.0 also predicts properties of the proteins like structure (Helix/ Coil/ Sheet), surface accessibility (Exposed/Buried) taking each residue of amino acid into account. Different threshold values can be set to identify the desired epitopes. This threshold value is a correlation between specificity and sensitivity. The more the value of specificity, the less the number of epitopes. Here, four different threshold values were taken for getting maximum numbers of epitope:0.5,0.55,0.6 and 0.65.

Table 3.11: Predicted B-cell epitopes using BepiPred 2.0

Position	Epitope	Threshold
296	HLKHGG	0.5
1559	SFCRRHNW	0.5
585	KLLHWS DQTICYKD	0.5
408	VVNSQD	0.5
1741	RRGDNFHS	0.5
607	STAFPFE	0.5
1786	ITNESYN	0.5
423	IPQSLTFSY	0.5
1274	DSKLGTT E VKVSALKT	0.5
621	YLDSRTTQQ	0.5
1594	RRPINATDRSH	0.5
1571	CDTAGV	0.5
1666	SDRGQES	0.5
1867	KLRANDNI	0.5
567	NSWHLVEGELLVQDLR	0.5

Table 3.11: Predicted B-cell epitopes using BepiPred 2.0

Position	Epitope	Threshold
844	EDLRARMTYVCQCGGE	0.5
512	HGLDLMQAGSVVRRVNV	0.5
918	LKFDSGTVSKTSD	0.5
537	KEQEAILMSEDEVKLNPS	0.5
1798	YVKPDSV	0.5
879	NEKLVTSTA	0.5
668	ISDTIPDEKQNGHSL	0.5
1425	LSHYVLNID	0.5
1006	SDGQPGGDA	0.5
8	FGGDQVHEV	0.5
1038	SFLPKEDGD	0.5
488	TLDDILQQANKSV	0.5
787	TFGAPDD	0.5
1380	LASAYRANSFDVPTFCA	0.5
650	NKNTYRSQ	0.5
1024	GFDSSKPVT	0.5
1755	DAARGPAGVESDVETN	0.5
1076	ASYDTNLNKFNRASL	0.5
1049	LAEFDTYDPIYKNGAMY	0.5
1220	FYFKEFATRTFTA	0.5
685	ADNLTADETKALKELYGPVD	0.5
321	EYILAKGPLQ	0.5
820	NVCGIKD	0.5
717	AVHGWKMV	0.5
1410	DSITHYP	0.5
277	AIQVAKC	0.5
1149	VADDSGTP	0.5
1723	NVTRDKLEK	0.5
1110	VESTPV	0.55
546	EDVKLNPS	0.55
1276	KLGTTE	0.55
671	TIPDEKQNG	0.55
1757	ARGPAGVES	0.55
1007	DGQPGGD	0.55
84	FIDAPCYCFNAEGDASWS	0.55
1384	YRANSFDVPT	0.55
880	EKLVTTS	0.55
490	DDILQQANK	0.55

Table 3.11: Predicted B-cell epitopes using BepiPred 2.0

Position	Epitope	Threshold
1354	FEDAQGLKK	0.55
1055	YDPIYKNGA	0.55
1078	YDTNLNKFNR	0.55
1222	FKEFATRTFT	0.55
170	ESVQEEAQP	0.6
202	PDTVEVPPQVVKLPS	0.6
451	CTDNSANTKVLNRK	0.6
1080	TNLNKF	0.6
238	QTQNVTVKPKRLRKKRN	0.65

3.5.3 Screening epitope candidates using VaxiJen 2.0

All predicted B-cell epitopes were further evaluated by the VaxiJen 2.0 server. Among them, 178 epitope candidates achieved a score over the desired threshold value 0.4.

Table 3.12: Predicted B-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0

Position	Epitope	Prediction Method	VaxiJen 2.0 Score
296	HLKHGG	BepiPred 2.0	1.838
1559	SFCRRHNW	BepiPred 2.0	1.6656
1558	ISFCRRHNWNCVDCDTAGVG	BCPREDS	1.5742
925	VSKTSDWKCKVTDV	BCPREDS	1.5713
585	KLLHWSQTICIKD	BepiPred 2.0	1.4921
587	LHWSQTICIKDSV	BCPREDS	1.4876
1561	CRRHNWNCVDCDTAGVGN	BCPREDS	1.4624
588	HWSQTICIKDS	BCPREDS	1.4528
1563	RHNWNCVDCDTAGVGN	BCPREDS	1.4061
221	IQPEVKEVAPVY	BCPREDS	1.3759
1110	VESTPV	BepiPred 2.0	1.3722
925	VSKTSDWKCKVT	BCPREDS	1.3515
921	DSGTVSKTSDWKCKVTDV	BCPREDS	1.3464
1471	FVDWRSYNYAVS	BCPREDS	1.3355
646	VVLNNKNTYRSQLGCV	BCPREDS	1.3326
1565	NWNCVDCDTAGVGN	BCPREDS	1.291
408	VVNSQD	BepiPred 2.0	1.2248
1643	LKFKEVCKTTTGIPEYNF	BCPREDS	1.2112

Table 3.12: Predicted B-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0

Position	Epitope	Prediction Method	VaxiJen 2.0 Score
1741	RRGDNFHS	BepiPred 2.0	1.2108
1104	KFTPLSVESTPVEPPTVDVV	BCPREDS	1.1992
1169	LYVDPKYQVIVL	BCPREDS	1.1904
921	DSGTVSKTSDWKCKVT	BCPREDS	1.1775
919	KFDSGTVSKTSDWKCKVTDV	BCPREDS	1.1712
647	VLNNKNTYRSQLGC	BCPREDS	1.1649
1643	LKFKEVCKTTTGIPEY	BCPREDS	1.1169
607	STAFPFE	BepiPred 2.0	1.1136
884	TTSTAPDFVAFN	BCPREDS	1.1112
119	EVEASDLEEGESECIS	BCPREDS	1.1092
1470	IFVDWRSYNYAVSS	BCPREDS	1.109
524	RRVNVVPYVCLLANKEQEA	BCPREDS	1.1034
1610	VTVKETVVQFNYRRDGQP	BCPREDS	1.0785
546	EDVKLNPSE	BepiPred 2.0	1.077
1786	ITNESYN	BepiPred 2.0	1.0758
436	GAIRKAKDYGFT	BCPREDS	1.0706
1614	ETVVQFNYRRDGQP	BCPREDS	1.061
423	IPQSLTFSY	BepiPred 2.0	1.0307
241	NVTVKPKRLRKKRNVDPLSN	BCPREDS	0.9976
883	VTTSTAPDFVAFNV	BCPREDS	0.9924
1274	DSKLGTTTEVKVSALKT	BepiPred 2.0	0.99
729	VRSLKLSDNNCY	BCPREDS	0.9883
1471	FVDWRSYNYAVSSAFW	BCPREDS	0.9881
667	DISDTIPDEKQN	BCPREDS	0.9616
1277	LGTTEVKVSALKTA	BCPREDS	0.9596
106	FSLHPVECDEECSEVEAS	BCPREDS	0.9582
1612	VKETVVQFNYRRDGQP	BCPREDS	0.9519
522	VVRRVNVVPYVCLLANKEQEA	BCPREDS	0.9459
621	YLDSTRTTQQ	BepiPred 2.0	0.9339
1276	KLGTE	BepiPred 2.0	0.923
1594	RRPINATDRSH	BepiPred 2.0	0.9174
520	GSVVRNVVPYVCL	BCPREDS	0.9166
222	QPEVKEVAPVYEAD	BCPREDS	0.906
663	FNGADISDTIPDEKQN	BCPREDS	0.9016
1471	FVDWRSYNYAVSSAFWLF	BCPREDS	0.8959
1843	SDALKRQIRIAC	BCPREDS	0.8891

Table 3.12: Predicted B-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0

Position	Epitope	Prediction Method	VaxiJen 2.0 Score
115	EECSEVEASDLE	BCPREDS	0.888
848	ARMTYVCQCGGERHRQ	BCPREDS	0.8753
1108	LSVESTPVEPPTVDVVAL	BCPREDS	0.8741
348	HVVGPDARAKQD	BCPREDS	0.8727
1272	FSDSKLGTTEVKVSALKT	BCPREDS	0.8707
1106	TPLSVESTPVEPPTVD	BCPREDS	0.8689
1066	KGKPILEWVNKASYDTNLN	BCPREDS	0.8639
519	AGSVVRRVNVVPYVCLL	BCPREDS	0.8564
1571	CDTAGV	BepiPred 2.0	0.8537
648	LNNKNTYRSQLG	BCPREDS	0.8514
1666	SDRGQES	BepiPred 2.0	0.847
366	CYKAMNAYPLVVTP	BCPREDS	0.8462
1867	KLRANDNI	BepiPred 2.0	0.8428
218	PQTIQPEVKEVAPVYEAD	BCPREDS	0.8382
1192	HTVESGDINVVAAS	BCPREDS	0.8362
1299	KQCCTAAVDLSM	BCPREDS	0.8312
1677	ACVYYSQVLCKS	BCPREDS	0.8299
1192	HTVESGDINVVA	BCPREDS	0.8284
90	YCFNAEGDASWSST	BCPREDS	0.8172
392	VSFDYLIREAKT	BCPREDS	0.8137
246	PKRLRKKRNVDP	BCPREDS	0.8105
405	VLVVVNSQDVYKSLTIVDIP	BCPREDS	0.8101
105	IFSLHPVECDEECS	BCPREDS	0.8016
522	VVRRVNVVPYVCL	BCPREDS	0.7994
112	ECDEECSEVEASDLEEGESE	BCPREDS	0.7964
243	TVKPKRLRKKRNVDP	BCPREDS	0.796
567	NSWHLVEGELLVQDLR	BepiPred 2.0	0.7843
1644	KFKEVCKTTTGIPE	BCPREDS	0.777
88	PCYCFNAEGDASWSST	BCPREDS	0.77
844	EDLRARMTYVCQCGGE	BepiPred 2.0	0.7675
1856	KCNLAFLRTTSK	BCPREDS	0.7654
512	HGLDLMQAGSVVRRVNV	BepiPred 2.0	0.7651
134	SETSTEQVDVSH	BCPREDS	0.7563
222	QPEVKEVAPVYEADTE	BCPREDS	0.754
875	SGTPNEKLVTSTAPDFV	BCPREDS	0.7486
215	PSAPQTIQPEVKEVAPVYEA	BCPREDS	0.7458

Table 3.12: Predicted B-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0

Position	Epitope	Prediction Method	VaxiJen 2.0 Score
1612	VKETVVQFNYRRDGQPFYER	BCPREDS	0.7425
904	VGHYVHARLKGG	BCPREDS	0.7377
238	QTQNVTVKPKRLRKKRN	BepiPred 2.0	0.736
348	HVVGPDARAKQDVS	BCPREDS	0.7308
1278	GTTEVKVSALKT	BCPREDS	0.7263
663	FNGADISDTIPDEKQNGH	BCPREDS	0.7247
918	LKFDSGTVSKTSD	BepiPred 2.0	0.72
92	FNAEGDASWSST	BCPREDS	0.7199
1110	VESTPVEPPTVD	BCPREDS	0.7039
1250	TKGILTGCFSFA	BCPREDS	0.7
168	VTESVQEEAQP	BCPREDS	0.6972
423	IPQSLTFSYDGLRG	BCPREDS	0.6899
537	KEQEAILMSEDKLNPS	BepiPred 2.0	0.6898
591	DQTICYKDSVFYVVKNST	BCPREDS	0.6884
535	ANKEQEAILMSE	BCPREDS	0.682
5	KVAFGGDQVHEVAAVRSV	BCPREDS	0.6727
295	THLKHGGGIAGA	BCPREDS	0.6614
663	FNGADISDTIPDEKQNGHSL	BCPREDS	0.6564
1567	NCVDCDTAGVGN	BCPREDS	0.6545
194	TLQETPVVPDTV	BCPREDS	0.6533
877	TPNEKLVTSTAPDFV	BCPREDS	0.6473
1250	TKGILTGCFSFAKM	BCPREDS	0.6444
170	ESVQEEAQP	BepiPred 2.0	0.6418
901	ETAVGHYVHARLKG	BCPREDS	0.6417
1192	HTVESGDINVVAASGS	BCPREDS	0.639
288	VLVNAANTHLKHGGGI	BCPREDS	0.6242
324	LAKGPLQVGDSV	BCPREDS	0.6192
8	FGGDQVHEVAAVRSVTVEYN	BCPREDS	0.6171
126	EEGESECISETSTE	BCPREDS	0.6161
199	PVVPDTVEVPPQVV	BCPREDS	0.6149
286	ESVLVNAANTHLKH	BCPREDS	0.6118
86	DAPCYCFNAEGDASWSSTMI	BCPREDS	0.6039
194	TLQETPVVPDTVEVPPQVVK	BCPREDS	0.5994
139	EQVDVSHETSDDEWAAVDE	BCPREDS	0.5916
1004	VSSDGQPGGDAISLSFNN	BCPREDS	0.5888
1546	GKRTFYITANGG	BCPREDS	0.5882

Table 3.12: Predicted B-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0

Position	Epitope	Prediction Method	VaxiJen 2.0 Score
243	TVKPKRLRKKRNVD	BCPREDS	0.57
1111	ESTPVEPPTVDVVA	BCPREDS	0.5691
142	DVSHETSDDEWAAAVD	BCPREDS	0.5579
1601	DRSHYYVDSVTV	BCPREDS	0.5574
9	GGDQVHEVA AVRSV	BCPREDS	0.5553
671	TIPDEKQNG	BepiPred 2.0	0.5505
1768	ETNEIVDSVQYAHK	BCPREDS	0.5443
85	IDAPCYCFNAEGDASWSS	BCPREDS	0.5441
172	VQEEAQPVEVPVED	BCPREDS	0.5366
951	VVRYSLDGNFRTEVDPDLA	BCPREDS	0.5359
372	AYPLVVTPLVSA	BCPREDS	0.5353
898	QGIETAVGHYVHARLKGG	BCPREDS	0.5334
245	KPKRLRKKRNVDPLSN	BCPREDS	0.531
1757	ARGPAGVES	BepiPred 2.0	0.5293
365	KCYKAMNAYPLVVTPLV	BCPREDS	0.5235
344	KNILHVVGPDARAKQDVS	BCPREDS	0.5204
1790	SYNNYVPSYVKPDSVS	BCPREDS	0.5201
21	TGLAYMLYTSAFNWL	BepiPred 2.0	0.5128
695	ALKELYGPVDPTFL	BCPREDS	0.5113
1411	SITHYPALKMVQTH	BCPREDS	0.5101
1645	FKEVCKTTTGIPEYNFIID	BCPREDS	0.5087
1798	YVKPDSV	BepiPred 2.0	0.5082
481	YCYTSKDTLDDI	BCPREDS	0.5009
1536	RVEASTVVC GGKRTFYIT	BCPREDS	0.4996
439	RKAKDYGFTV FVCTDNSANT	BCPREDS	0.4994
169	TESVQEEAQPVEVPVE	BCPREDS	0.496
1007	DGQPGGD	BepiPred 2.0	0.4876
693	TKALKELYGPVDPTFLHR	BCPREDS	0.4858
197	ETPVVPDTVEVPPQVVKL	BCPREDS	0.484
84	FIDAPCYCFNAEGDASWS	BepiPred 2.0	0.4802
1825	LRNSNGACIWNAAA	BCPREDS	0.4769
168	VTESVQEEAQPVEVPVED	BCPREDS	0.4708
879	NEKLVTSTA	BepiPred 2.0	0.4699
405	VLVVVNSQDVYK	BCPREDS	0.462
809	CSARMVWREWCNVC GIKD	BCPREDS	0.4616
1788	NESYNNYVPSYVKPDSVSTS	BCPREDS	0.4599

Table 3.12: Predicted B-cell epitope candidates having threshold value over 0.4 in VaxiJen 2.0

Position	Epitope	Prediction Method	VaxiJen 2.0 Score
627	TQQLTIEVLVTV	BCPREDS	0.4468
1292	VVTGNVVKQCCTAAVDLS	BCPREDS	0.4454
1695	SLVTSVGDSSEIATKM	BCPREDS	0.4445
668	ISDTIPDEKQNGHSL	BepiPred 2.0	0.4383
1697	VTSVGDSSEIAT	BCPREDS	0.4374
691	DETKALKELYGPVDPTFLHR	BCPREDS	0.4355
166	EDVTESVQEEAQPVEVPVED	BCPREDS	0.433
1648	VCKTTTGIPEYN	BCPREDS	0.4272
1425	LSHYVLNID	BepiPred 2.0	0.4267
305	GAINAASKGAVQKE	BCPREDS	0.4204
7	AFGGDQVHEVAA	BCPREDS	0.4189
73	GMPIPDFDLDDF	BCPREDS	0.4168
1787	TNESYNNYVPSYVK	BCPREDS	0.4162
1793	NYVPSYVKPDSV	BCPREDS	0.4127
399	REAKTRVLVVVNSQDV	BCPREDS	0.407
1787	TNESYNNYVPSYVKPDSV	BCPREDS	0.4008

3.5.4 Overlapping Sequence Identification of B-cell epitopes and T-cell Epitopes

Eight B-cell epitopes shared common sequences with two of the selected T-cell epitopes.

Table 3.13: B-cell and T-cell Epitopes having overlapping Sequence

B-cell epitopes		T-cell epitopes	
Position	Epitope	Position	Epitope
1470	IFVDWRSYNYAVSS	1471	FVDWRSYNY
1471	FVDWRSYNYAVS		
1471	FVDWRSYNYAVSSAFW		
1471	FVDWRSYNYAVSSAFWLF		
1643	LKFKEVC KTTTGIPEY NF	1650	KTTTGIPEY
1643	LKFKEVC KTTTGIPEY		
1645	FKEVC KTTTGIPEY NFIIYD		
1648	VCK KTTTGIPEY N		

These eight B-cell epitopes were selected for further evaluation.


3.5.5 Conservancy Analysis

Conservancy of epitopes was checked using the IEDB conservancy analysis tool. All predicted epitope candidates had the maximum identity (100%) for conservancy hit.


IEDB Analysis Resource

[Home](#)[Help](#)[Example](#)[Reference](#)[Download](#)[Contact](#)

Epitope Conservancy Analysis Result

Download result 

Epitope #	Epitope name	Epitope sequence	Epitope length	Percent of protein sequence matches at identity <= 100%	Minimum identity	Maximum identity	View details
1	1	IFVDWRSYNYAVSS	14	100.00% (1/1)	100.00%	100.00%	Go
2	2	FVDWRSYNYAVS	12	100.00% (1/1)	100.00%	100.00%	Go
3	3	FVDWRSYNYAVSSAFW	16	100.00% (1/1)	100.00%	100.00%	Go
4	4	FVDWRSYNYAVSSAFWLF	18	100.00% (1/1)	100.00%	100.00%	Go
5	5	LKFKEVCKTIIGIPEYNF	18	100.00% (1/1)	100.00%	100.00%	Go
6	6	LKFKEVCKTIIGIPEY	16	100.00% (1/1)	100.00%	100.00%	Go
7	7	FKEVCKTIIGIPEYNFIID	20	100.00% (1/1)	100.00%	100.00%	Go
8	8	VCKTIIGIPEYN	12	100.00% (1/1)	100.00%	100.00%	Go

Download result 

© 2005-2019 | [IEDB Home](#)
Supported by a contract from the [National Institute of Allergy and Infectious Diseases](#), a component of the National Institutes of Health in the Department of Health and Human Services.

Figure 3.10: Epitope conservancy analysis showing all selected B-cell epitopes had 100% conservancy

3.5.6 Checking Epitopes as Ideal Vaccine Candidates

For being an ideal vaccine, an epitope should have properties like:

- Presence of beta-turn
- Hydrophobicity
- Surface accessibility
- Flexibility
- Antigenicity etc.

These properties of the selected 8 candidate epitopes were examined by IEDB B-cell tools.

3.5.6.1 Analysis of IFVDWRSYNYAVSS

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. The yellow peak indicates the residues that are above the threshold. 5 out of 8 peptide fragments were above the threshold.

Chou & Fasman Beta-Turn Prediction Results

Input Sequences

1 IFVDWRSYNY AVSS

Center position: 4 Window size: 7 Threshold: 1.089 Recalculate



Average: 1.089 Minimum: 0.910 Maximum: 1.234

Predicted residue scores:

Position	Residue	Start	End	Peptide	Score
7	S	4	10	DWRSYNY	1.234
6	R	3	9	VDWRSYN	1.143
10	Y	7	13	SYNYAVS	1.123
11	A	8	14	YNYAVSS	1.123
8	Y	5	11	WRSYNYA	1.12
9	N	6	12	RSYNYAV	1.054
5	W	2	8	FVDWRSY	1.006
4	D	1	7	IFVDWRS	0.91

Figure 3.11: Result of Chou & Fasman Beta Turn Prediction of IFVDWRSYNYAVSS

- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 3 out of 9 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 4 out of 7 peptide fragments were above the threshold.
- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 5 out of 8 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 5 out of 8 peptide fragments were above the threshold.

3.5.6.2 Analysis of FVDWRSYNYAVS

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. 4 out of 6 peptide fragments were above the threshold.
- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 3 out of 7 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 2 out of 5 peptide fragments were above the threshold.
- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 3 out of 6 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 4 out of 6 peptide fragments were above the threshold.

3.5.6.3 Analysis of FVDWRSYNYAVSSAFW

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. 5 out of 10 peptide fragments were above the threshold.
- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 4 out of 11 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 5 out of 9 peptide fragments were above the threshold.
- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 6 out of 10 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 6 out of 10 peptide fragments were above the threshold.

3.5.6.4 Analysis of FVDWRSYNYAVSSAFWLF

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. 7 out of 12 peptide fragments were above the threshold.
- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 4 out of 13 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 6 out of 11 peptide fragments were above the threshold.

- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 8 out of 12 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 7 out of 12 peptide fragments were above the threshold.

3.5.6.5 Analysis of LKFKEVCKTTTGIPEYNF

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. 7 out of 12 peptide fragments were above the threshold.
- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 7 out of 13 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 6 out of 11 peptide fragments were above the threshold.
- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 6 out of 12 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 7 out of 12 peptide fragments were above the threshold.

3.5.6.6 Analysis of LKFKEVCKTTTGIPEY

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. 5 out of 10 peptide fragments were above the threshold.
- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 5 out of 11 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 5 out of 9 peptide fragments were above the threshold.
- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 6 out of 10 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 6 out of 10 peptide fragments were above the threshold.

3.5.6.7 Analysis of FKEVCKTTTGIPEYNFIHYD

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. 7 out of 14 peptide fragments were above the threshold.
- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 7 out of 15 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 7 out of 13 peptide fragments were above the threshold.
- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 8 out of 14 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 9 out of 14 peptide fragments were above the threshold.

3.5.6.8 Analysis of VCKTTTGIPEYN

- The selected epitope was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool. 2 out of 6 peptide fragments were above the threshold.
- The selected epitope was checked for surface accessibility by Emini Surface Accessibility Prediction tool. 4 out of 7 peptide fragments were above the threshold.
- The selected epitope was checked for flexibility by Karplus and Schulz Flexibility Prediction tool. 3 out of 5 peptide fragments were above the threshold.
- The selected epitope was checked for antigenicity by Kolaskar & Tongaonkar Antigenicity Prediction tool. 3 out of 6 peptide fragments were above the threshold.
- The selected epitope was checked for hydrophilicity by Parker Hydrophilicity Prediction tool. 3 out of 6 peptide fragments were above the threshold.

3.5.7 Toxicity and Allergenicity Analysis

Toxicity prediction was carried out using ToxinPred. All eight of the selected epitopes were found as non-toxic. Afterward, allergenicity prediction was performed using AllergenFP v1.0 and AllerTOP v2.0.

Table 3.14: Toxicity and allergenicity prediction of selected B-cell epitopes

Epitope	Toxicity	AllergenFP v1.0	AllerTOP v2.0
IFVDWRSYNYAVSS	Non-Toxin	PROBABLE NON-ALLERGEN	PROBABLE ALLERGEN
FVDWRSYNYAVS	Non-Toxin	PROBABLE NON-ALLERGEN	PROBABLE NON-ALLERGEN
FVDWRSYNYAVSSAFW	Non-Toxin	PROBABLE ALLERGEN	PROBABLE ALLERGEN
FVDWRSYNYAVSSAFWLF	Non-Toxin	PROBABLE NON-ALLERGEN	PROBABLE NON-ALLERGEN
LKFKEVCKTTTGIPEYNF	Non-Toxin	PROBABLE ALLERGEN	PROBABLE NON-ALLERGEN
LKFKEVCKTTTGIPEY	Non-Toxin	PROBABLE NON-ALLERGEN	PROBABLE NON-ALLERGEN
FKEVCKTTTGIPEYNFIID	Non-Toxin	PROBABLE ALLERGEN	PROBABLE NON-ALLERGEN
VCKTTTGIPEYN	Non-Toxin	PROBABLE ALLERGEN	PROBABLE ALLERGEN

3.6 Assessing Potential Anti-Viral Activity of Selected Flavonoids against MERS-CoV

3.6.1 Ligand Selection

After browsing several articles in PubMed Central, 18 flavonoids which may have potential antiviral activities were selected as ligands along with ADP-ribose.

Table 3.15: Selected flavonoids and their Pubchem ID

Name	Pubchem ID
Apigenin	5280443
Apiin	5280746
Baicalein	5281605
Daidzein	5281708
Fisetin	5281614
Genistein	5280961
Glabranin	124049

Table 3.15: Selected flavonoids and their Pubchem ID

Name	Pubchem ID
Hesperetin	72281
Hesperidin	10621
Kaempferol	5280863
Luteolin	5280445
Luteoloside	5280637
Naringin	442428
Pinostrobin	73201
Quercetin	5280343
Ribavirin	37542
Rutin	5280805
Silymarin	7073228

3.6.2 Macromolecule selection

MERS-CoV macro domain within nsp3 protein was selected as macromolecule/receptor.

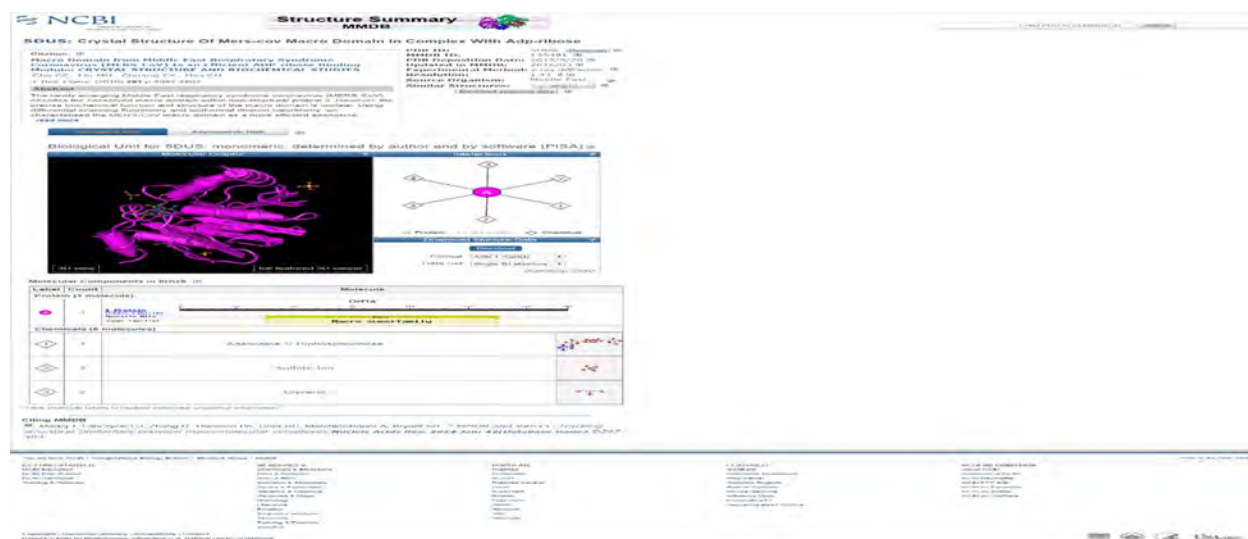


Figure 3.12: MERS-CoV nsp3 macro domain profile in NCBI

3.6.3 Macromolecule and Ligand preparation

Ligand structure retrieval

3D structures of ligands were retrieved from PubChem in SDF format.



Figure 3.13: 3D structure of selected flavonoids retrieval from PubChem

Macromolecule 3D structure retrieval

The three-dimensional crystal structure of the macro domain within nsp3 protein was retrieved from RCSB Protein Data Bank (PDB ID:5DUS) in PDB format.

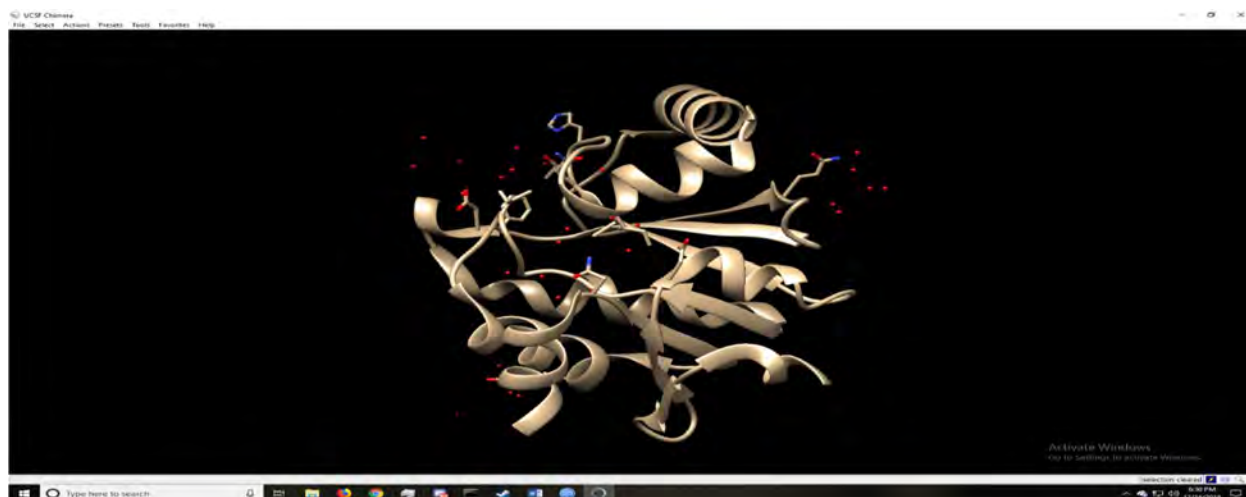


Figure 3.14: Macromolecule 3D structure retrieval from RCSB PDB

However, 3 ligands (ADP-ribose, Sulfate Ion and Glycerol) were already attached to the receptor in this crystal structure. Therefore, UCSF Chimera 1.13 was used to remove the undesired ligands.



(A) Crystal structure of MERS-CoV nsp3 macro domain complex with ADP-ribose, Sulfate Ion and Glycerol attached as ligands.



(B) ADP-ribose, Sulfate Ion and Glycerol were selected and removed

Figure 3.15: Undesired ligand removal using UCSF Chimera 1.13

In addition to that, the protein was cleaned by removing the water molecules.

Energy minimization

Energy minimization of target macromolecule was done using UCSF Chimera 1.13.

Table 3.16: Macromolecule Minimization parameter

Macromolecule Minimization parameter	
Force Field Method (Standard residues)	AMBER ff14SB
Force Field Method (other residues)	AM1-BCC
Steepest descent steps	100
Steepest descent step size	0.02
Conjugate gradient steps	10
Conjugate gradient step size	0.02
Update interval	10
Added hydrogen method	steric, also consider H-bonds
Protonation states for	Histidine

Ligand minimization was done by PyRx prior to docking.

3.6.4 Molecular Docking using PyRx

After the minimizing process, the protein was placed in a grid box measuring $37.3660 \text{ \AA} \times 43.4316 \text{ \AA} \times 43.1478 \text{ \AA}$ along the x, y and z axis, respectively, where the position of the center was X:8.9843, Y:17.6095, Z:68.5928.

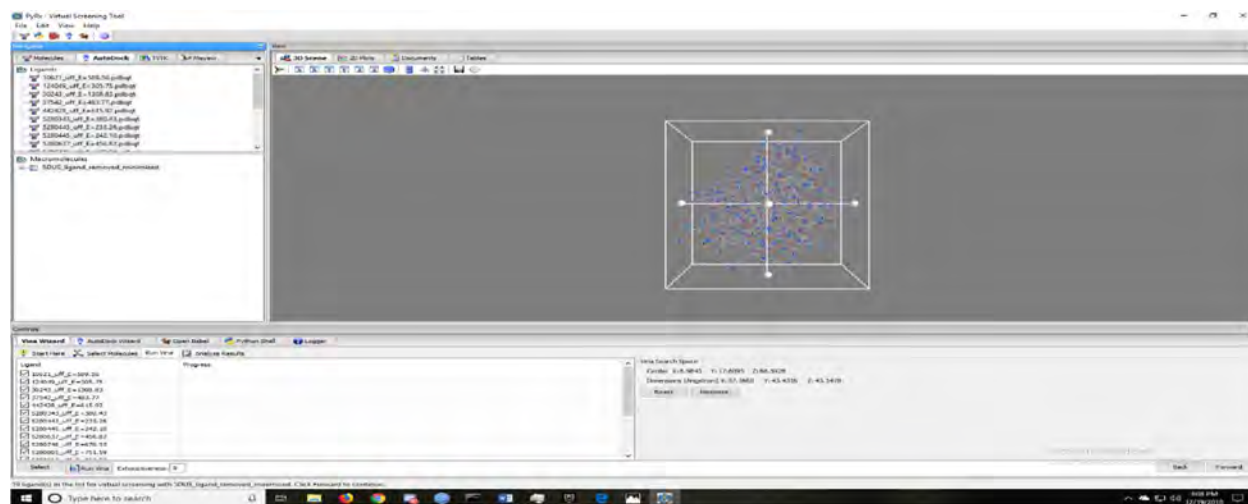


Figure 3.16: Grid box adjustment process before initiating the docking process in PyRx

ADP-ribose was first re-docked into the ADP-ribose binding site of nsp3 and the resulting interactions were compared with those found by docking 18 flavonoids into the similar active site using the same grid box. The docking procedure was performed using the instructed command

prompts. The docking results included the binding energy value given in kcal/mol, mode, RMSD upper bound (rmsd/ub) and RMSD lower bound (rmsd/lb).

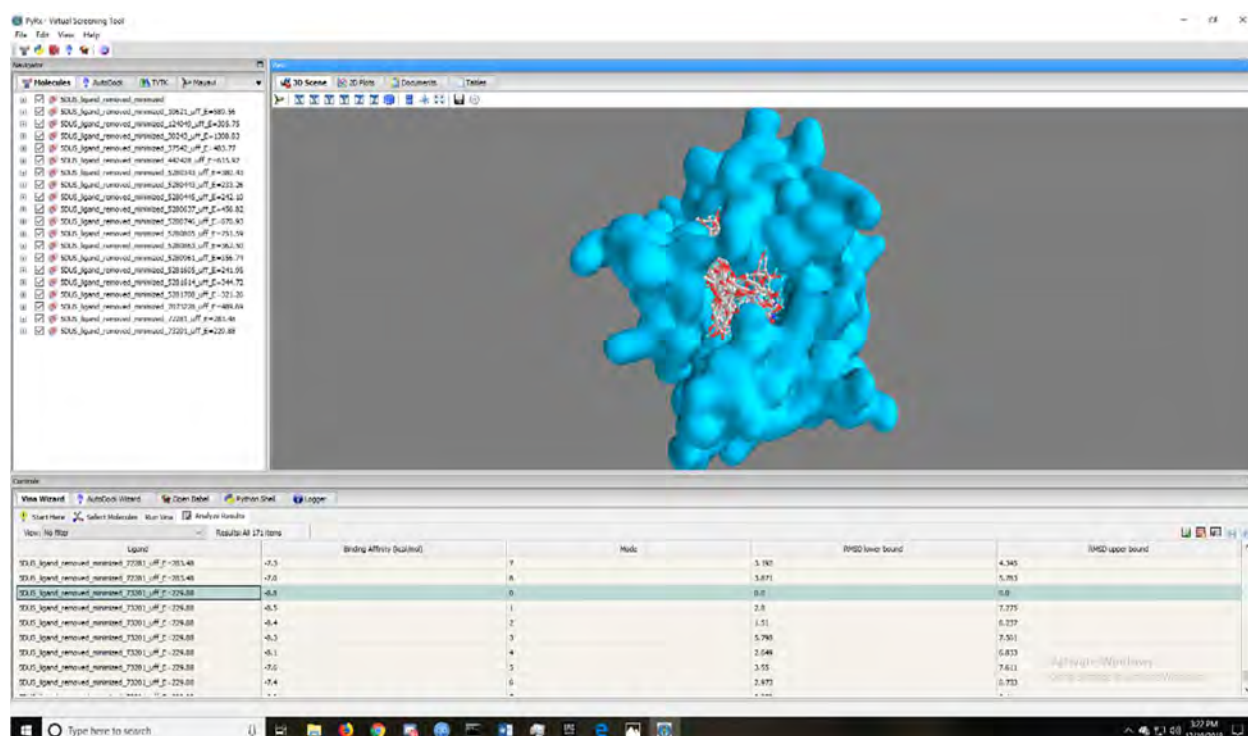


Figure 3.17: Docking result showing binding affinity of compounds against MERS-CoV nsp3 macro domain

Table 3.17: Binding affinity of re-docked ADP-ribose against MERS-CoV nsp3 macro domain

Ligand	Binding Affinity (kcal/mol)	rmsd/ub	rmsd/lb
ADP-ribose_uff_E=1308.83	-8.7	0	0
ADP-ribose_uff_E=1308.83	-8.6	4.443	2.804
ADP-ribose_uff_E=1308.83	-8	2.68	1.681
ADP-ribose_uff_E=1308.83	-7.9	4.732	2.899
ADP-ribose_uff_E=1308.83	-7.8	4.761	3.141
ADP-ribose_uff_E=1308.83	-7.6	4.112	2.914
ADP-ribose_uff_E=1308.83	-7.6	9.316	4.944
ADP-ribose_uff_E=1308.83	-7.5	5.626	3.775
ADP-ribose_uff_E=1308.83	-7.4	6.227	4.271

Table 3.18: Binding affinity of flavonoid compounds against MERS-CoV nsp3 macro domain

Ligand	Binding Affinity (kcal/mol)	rmsd/ub	rmsd/lb
Hesperidin_uff_E=589.56	-9.9	0	0
Hesperidin_uff_E=589.56	-9.1	2.861	1.836
Hesperidin_uff_E=589.56	-8.7	8.869	3.076
Hesperidin_uff_E=589.56	-8.6	2.899	1.898
Hesperidin_uff_E=589.56	-8.4	11.581	4.006
Hesperidin_uff_E=589.56	-8.2	2.498	1.643
Hesperidin_uff_E=589.56	-8.1	4.926	1.846
Hesperidin_uff_E=589.56	-8	4.324	2.16
Hesperidin_uff_E=589.56	-7.9	6.45	3.712
Glabranin_uff_E=305.75	-8.5	0	0
Glabranin_uff_E=305.75	-7.9	1.619	1.065
Glabranin_uff_E=305.75	-7.7	4.299	2.063
Glabranin_uff_E=305.75	-7.6	7.711	4.529
Glabranin_uff_E=305.75	-7.4	6.277	3.733
Glabranin_uff_E=305.75	-7	4.895	1.367
Glabranin_uff_E=305.75	-6.9	4.866	2.022
Glabranin_uff_E=305.75	-6.9	5.727	1.903
Glabranin_uff_E=305.75	-6.9	5.646	2.827
Ribavirin_uff_E=483.77	-6.6	0	0
Ribavirin_uff_E=483.77	-6.5	7.758	5.44
Ribavirin_uff_E=483.77	-6.4	2.456	1.21
Ribavirin_uff_E=483.77	-6.3	4.714	3.13
Ribavirin_uff_E=483.77	-6.3	5.757	2.822
Ribavirin_uff_E=483.77	-6.2	2.607	1.725
Ribavirin_uff_E=483.77	-6.1	6.149	3.619
Ribavirin_uff_E=483.77	-6	4.958	3.001
Ribavirin_uff_E=483.77	-5.9	3.337	2.071
Naringin_uff_E=615.92	-10.1	0	0
Naringin_uff_E=615.92	-9.6	3.833	2.29
Naringin_uff_E=615.92	-9.2	4.104	2.064
Naringin_uff_E=615.92	-8.6	5.312	3.227
Naringin_uff_E=615.92	-8.4	9.015	2.271
Naringin_uff_E=615.92	-7.8	9.028	2.261
Naringin_uff_E=615.92	-7.7	7.771	4.255

Table 3.18: Binding affinity of flavonoid compounds against MERS-CoV nsp3 macro domain

Ligand	Binding Affinity (kcal/mol)	rmsd/ub	rmsd/lb
Naringin_uff_E=615.92	-7.7	14.831	6.965
Naringin_uff_E=615.92	-7.5	16.169	7.064
Quercetin_uff_E=380.43	-9.1	0	0
Quercetin_uff_E=380.43	-8.8	1.79	1.296
Quercetin_uff_E=380.43	-8.8	7.138	2.24
Quercetin_uff_E=380.43	-8.3	7.097	2.064
Quercetin_uff_E=380.43	-8.3	2.912	2.389
Quercetin_uff_E=380.43	-8	4.136	2.357
Quercetin_uff_E=380.43	-7.9	7.433	1.935
Quercetin_uff_E=380.43	-7.6	7.105	3.407
Quercetin_uff_E=380.43	-7.6	6.847	4.75
Apigenin_uff_E=233.26	-8.7	0	0
Apigenin_uff_E=233.26	-8.4	6.929	2.334
Apigenin_uff_E=233.26	-8.3	4.831	2.692
Apigenin_uff_E=233.26	-8	6.697	2.518
Apigenin_uff_E=233.26	-7.8	4.572	2.91
Apigenin_uff_E=233.26	-7.8	3.263	1.66
Apigenin_uff_E=233.26	-7.8	5.959	3.871
Apigenin_uff_E=233.26	-7.6	7.229	3.287
Apigenin_uff_E=233.26	-7.6	6.277	3.665
Luteolin_uff_E=242.10	-8.6	0	0
Luteolin_uff_E=242.10	-8.6	7.382	1.54
Luteolin_uff_E=242.10	-8.5	4.122	2.385
Luteolin_uff_E=242.10	-8.5	7.384	2.712
Luteolin_uff_E=242.10	-8.3	4.546	2.475
Luteolin_uff_E=242.10	-8.2	2.934	1.322
Luteolin_uff_E=242.10	-7.8	6.991	2.746
Luteolin_uff_E=242.10	-7.3	6.296	4.072
Luteolin_uff_E=242.10	-7.1	6.259	3.624
Luteoloside_uff_E=456.82	-10	0	0
Luteoloside_uff_E=456.82	-9.7	1.186	0.689
Luteoloside_uff_E=456.82	-8.8	2.077	1.342
Luteoloside_uff_E=456.82	-8.7	9.23	1.944
Luteoloside_uff_E=456.82	-8.6	9.281	2.437

Table 3.18: Binding affinity of flavonoid compounds against MERS-CoV nsp3 macro domain

Ligand	Binding Affinity (kcal/mol)	rmsd/ub	rmsd/lb
Luteoloside_uff_E=456.82	-8	6.387	3.086
Luteoloside_uff_E=456.82	-8	6.81	3.259
Luteoloside_uff_E=456.82	-7.8	9.696	4.098
Luteoloside_uff_E=456.82	-7.7	6.458	4.184
Apiin_uff_E=670.93	-10.1	0	0
Apiin_uff_E=670.93	-9.6	2.765	1.667
Apiin_uff_E=670.93	-9	9.35	2.583
Apiin_uff_E=670.93	-8.7	4.059	1.946
Apiin_uff_E=670.93	-8.3	4.224	2.374
Apiin_uff_E=670.93	-8.2	3.636	2.706
Apiin_uff_E=670.93	-8.1	5.089	3.38
Apiin_uff_E=670.93	-7.8	9.245	2.749
Apiin_uff_E=670.93	-7.7	4.579	2.415
Rutin_uff_E=751.59	-8.8	0	0
Rutin_uff_E=751.59	-8.8	2.97	2.026
Rutin_uff_E=751.59	-8.8	1.816	1.102
Rutin_uff_E=751.59	-8.6	7.549	2.573
Rutin_uff_E=751.59	-8.5	5.075	1.402
Rutin_uff_E=751.59	-8.2	5.034	2.04
Rutin_uff_E=751.59	-8.1	2.635	1.557
Rutin_uff_E=751.59	-7.9	6.656	2.227
Rutin_uff_E=751.59	-7.9	5.805	1.929
Kaempferol_uff_E=362.50	-8.7	0	0
Kaempferol_uff_E=362.50	-8.5	6.608	2.282
Kaempferol_uff_E=362.50	-8	6.545	4.425
Kaempferol_uff_E=362.50	-7.9	5.205	3.395
Kaempferol_uff_E=362.50	-7.8	4.249	1.978
Kaempferol_uff_E=362.50	-7.7	6.273	1.466
Kaempferol_uff_E=362.50	-7.3	3.204	1.905
Kaempferol_uff_E=362.50	-6.7	9.035	6.09
Kaempferol_uff_E=362.50	-6.5	10.655	7.654
Genistein_uff_E=356.74	-8.3	0	0
Genistein_uff_E=356.74	-7.9	7.193	1.426
Genistein_uff_E=356.74	-7.8	6.928	1.816

Table 3.18: Binding affinity of flavonoid compounds against MERS-CoV nsp3 macro domain

Ligand	Binding Affinity (kcal/mol)	rmsd/ub	rmsd/lb
Genistein_uff_E=356.74	-7	7.046	4.524
Genistein_uff_E=356.74	-6.8	7.126	4.525
Genistein_uff_E=356.74	-6.4	23.429	21.444
Genistein_uff_E=356.74	-6.1	1.657	1.516
Genistein_uff_E=356.74	-6	23.497	21.463
Genistein_uff_E=356.74	-6	8.929	5.966
Baicalein_uff_E=241.95	-9	0	0
Baicalein_uff_E=241.95	-8.4	3.828	1.853
Baicalein_uff_E=241.95	-8.3	6.474	2.295
Baicalein_uff_E=241.95	-8.3	2.559	1.275
Baicalein_uff_E=241.95	-8	2.114	1.817
Baicalein_uff_E=241.95	-8	6.725	2.391
Baicalein_uff_E=241.95	-8	6.067	4.252
Baicalein_uff_E=241.95	-7.8	6.443	2.878
Baicalein_uff_E=241.95	-7.2	4.616	2.698
Fisetin_uff_E=344.72	-9.7	0	0
Fisetin_uff_E=344.72	-9.6	1.648	1.096
Fisetin_uff_E=344.72	-9	2.535	1.594
Fisetin_uff_E=344.72	-8.7	7.275	2.501
Fisetin_uff_E=344.72	-8.5	2.524	1.568
Fisetin_uff_E=344.72	-7.9	7.453	2.891
Fisetin_uff_E=344.72	-7.9	7.501	4.048
Fisetin_uff_E=344.72	-7.8	7.172	4.397
Fisetin_uff_E=344.72	-7.7	7.421	2.909
Daidzein_uff_E=321.26	-8.3	0	0
Daidzein_uff_E=321.26	-8.3	7.629	1.927
Daidzein_uff_E=321.26	-8.1	3.618	2.147
Daidzein_uff_E=321.26	-8.1	2.455	1.268
Daidzein_uff_E=321.26	-7.8	6.985	1.579
Daidzein_uff_E=321.26	-7.8	6.663	1.254
Daidzein_uff_E=321.26	-7.7	8.027	2.672
Daidzein_uff_E=321.26	-7.3	6.855	2.677
Daidzein_uff_E=321.26	-7	6.203	4.82
Silymarin_uff_E=489.69	-8.5	0	0

Table 3.18: Binding affinity of flavonoid compounds against MERS-CoV nsp3 macro domain

Ligand	Binding Affinity (kcal/mol)	rmsd/ub	rmsd/lb
Silymarin_uff_E=489.69	-8.4	9.247	6.256
Silymarin_uff_E=489.69	-8.2	9.587	6.077
Silymarin_uff_E=489.69	-8	5.857	4.179
Silymarin_uff_E=489.69	-7.9	9.875	2.706
Silymarin_uff_E=489.69	-7.8	9.553	5.91
Silymarin_uff_E=489.69	-7.7	8.958	5.392
Silymarin_uff_E=489.69	-7.7	9.753	5.839
Silymarin_uff_E=489.69	-7.5	5.139	3.637
Hesperetin_uff_E=283.48	-8.9	0	0
Hesperetin_uff_E=283.48	-8.7	2.424	1.758
Hesperetin_uff_E=283.48	-8.2	7.13	1.887
Hesperetin_uff_E=283.48	-7.9	7.852	1.677
Hesperetin_uff_E=283.48	-7.6	7.284	4.59
Hesperetin_uff_E=283.48	-7.6	7.336	2.799
Hesperetin_uff_E=283.48	-7.5	2.93	2.096
Hesperetin_uff_E=283.48	-7.3	4.345	3.192
Hesperetin_uff_E=283.48	-7	5.783	3.871
Pinostrobin_uff_E=229.88	-8.8	0	0
Pinostrobin_uff_E=229.88	-8.5	7.275	2.8
Pinostrobin_uff_E=229.88	-8.4	6.237	1.51
Pinostrobin_uff_E=229.88	-8.3	7.561	5.798
Pinostrobin_uff_E=229.88	-8.1	6.833	2.649
Pinostrobin_uff_E=229.88	-7.6	7.611	3.55
Pinostrobin_uff_E=229.88	-7.4	6.733	2.973
Pinostrobin_uff_E=229.88	-6.9	6.4	3.879
Pinostrobin_uff_E=229.88	-6.6	4.485	3.139

The docking poses were ranked according to their docking scores. The conformation with the lowest binding affinity was selected as the best docking pose for further analysis.

Table 3.19: Binding affinity of best docking pose against MERS-CoV nsp3 macro domain

Name	Pubchem ID	Binding Affinity (kcal/mol)
Apiin	5280746	-10.1
Naringin	442428	-10.1
Luteoloside	5280637	-10
Hesperidin	10621	-9.9
Fisetin	5281614	-9.7
Quercetin	5280343	-9.1
Baicalein	5281605	-9
Hesperetin	72281	-8.9
Pinostrobin	73201	-8.8
Rutin	5280805	-8.8
Apigenin	5280443	-8.7
Kaempferol	5280863	-8.7
ADP-ribose	30243	-8.7
Luteolin	5280445	-8.6
Glabranin	124049	-8.5
Silymarin	7073228	-8.5
Daidzein	5281708	-8.3
Genistein	5280961	-8.3
Ribavirin	37542	-6.6

The best docking conformation of ADP-ribose showed a binding affinity of -8.7 kcal/mol (highlighted in grey in table 3.19). Only 12 flavonoids had a binding affinity greater than -8.7 kcal/mol.

3.6.5 Measuring K_i Value

K_i was calculated by the equation: $K_i = \exp [(\Delta G \cdot 1000)/(R \cdot T)]$, where ΔG is docking energy (binding affinity), R (gas constant) is $1.9859 \text{ cal K}^{-1} \text{ mol}^{-1}$ and T (temperature) is 298.15 K .

Table 3.20: Ki value calculation for each flavonoid compound

Name	Pubchem ID	Binding Affinity (kcal/mol)	Ki value (μM)
Apiin	5280746	-10.1	0.0390649
Naringin	442428	-10.1	0.0390649
Luteoloside	5280637	-10	0.0462525
Hesperidin	10621	-9.9	0.0547626
Fisetin	5281614	-9.7	0.0767681
Quercetin	5280343	-9.1	0.211482
Baicalein	5281605	-9	0.250393
Hesperetin	72281	-8.9	0.296463
Pinostrobin	73201	-8.8	0.351009
Rutin	5280805	-8.8	0.351009
Apigenin	5280443	-8.7	0.415592
Kaempferol	5280863	-8.7	0.415592
Luteolin	5280445	-8.6	0.492058
Glabranin	124049	-8.5	0.582592
Silymarin	7073228	-8.5	0.582592
Daidzein	5281708	-8.3	0.816699
Genistein	5280961	-8.3	0.816699
Ribavirin	37542	-6.6	14.4208

3.6.6 Docking Visualization Analysis using UCSF Chimera 1.13

The molecular visualization of the docked complexes was performed using UCSF Chimera 1.13.

Table 3.21: Intermolecular H bond between each compound with the nsp3 macro domain

Name	Pubchem ID	Binding Affinity (kcal/mol)	H-Bond Interacting residue	Distance (Å)
Apiin	5280746	-10.1	#0 ALA 21.A H-#1 UNK 1.N O 1.896Å	1.896
			#0 ALA 21.A H-#1 UNK 1.N O 2.512Å	2.512
			#0 GLY 128.A H-#1 UNK 1.N O 2.077Å	2.077
			#0 GLY 44.A H-#1 UNK 1.N O 2.154Å	2.154
			#1 UNK 1.N H-#0 LYS 42.A O 2.631Å	2.631
Naringin	442428	-10.1	#0 ALA 21.A H-#1 UNK 1.N O 1.808Å	1.808
			#0 GLY 128.A H-#1 UNK 1.N O 2.105Å	2.105
			#0 GLY 44.A H-#1 UNK 1.N O 2.521Å	2.521

Table 3.21: Intermolecular H bond between each compound with the nsp3 macro domain

Name	Pubchem ID	Binding Affinity (kcal/mol)	H-Bond Interacting residue	Distance (Å)
Luteoloside	5280637	-10	#0 GLY 128.A H-#1 UNK 1.N O 2.188Å	2.188
			#0 GLY 44.A H-#1 UNK 1.N O 2.031Å	2.031
			#1 UNK 1.N H-#0 VAL 152.A O 2.811Å	2.811
Hesperidin	10621	-9.9	#0 ALA 21.A H-#1 UNK 1.N O 2.104Å	2.104
			#1 UNK 1.N H-#0 GLY 44.A O 2.245Å	2.245
Fisetin	5281614	-9.7	#0 LEU 124.A H-#1 UNK 1.N O 2.356Å	2.356
			#0 PHE 130.A H-#1 UNK 1.N O 2.284Å	2.284
			#1 UNK 1.N H-#0 GLY 44.A O 2.563Å	2.563
Quercetin	5280343	-9.1	#0 GLY 128.A H-#1 UNK 1.N O 2.131Å	2.131
			#0 ILE 47.A H-#1 UNK 1.N O 2.368Å	2.368
			#0 LEU 124.A H-#1 UNK 1.N O 2.147Å	2.147
			#1 UNK 1.N H-#0 GLY 44.A O 2.353Å	2.353
			#1 UNK 1.N H-#0 LEU 124.A O 2.268Å	2.268
Baicalein	5281605	-9	#1 UNK 1.N H-#0 VAL 152.A O 2.403Å	2.403
Hesperetin	72281	-8.9	#0 GLY 128.A H-#1 UNK 1.N O 2.211Å	2.211
			#0 ILE 47.A H-#1 UNK 1.N O 2.083Å	2.083
Pinostrobin	73201	-8.8	#0 GLY 128.A H-#1 UNK 1.N O 2.200Å	2.2
			#0 ILE 47.A H-#1 UNK 1.N O 2.081Å	2.081
			#0 PHE 130.A H-#1 UNK 1.N O 2.385Å	2.385
Rutin	5280805	-8.8	#0 GLY 128.A H-#1 UNK 1.N O 2.256Å	2.256
			#0 ILE 47.A H-#1 UNK 1.N O 1.995Å	1.995
Apigenin	5280443	-8.7	#0 GLY 128.A H-#1 UNK 1.N O 2.523Å	2.523
			#0 GLY 44.A H-#1 UNK 1.N O 2.377Å	2.377
Kaempferol	5280863	-8.7	no hydrogen bond	invalid
ADP-ribose	30243	-8.7	#0 ILE 129.A H-#1 UNK 1.N O 1.862Å	1.862
			#0 ILE 47.A H-#1 UNK 1.N O 2.001Å	2.001
			#0 PHE 130.A H-#1 UNK 1.N O 2.212Å	2.212
			#1 UNK 1.N H-#0 ALA 36.A O 2.275Å	2.275
			#1 UNK 1.N HN-#0 VAL 152.A O 2.486Å	2.486

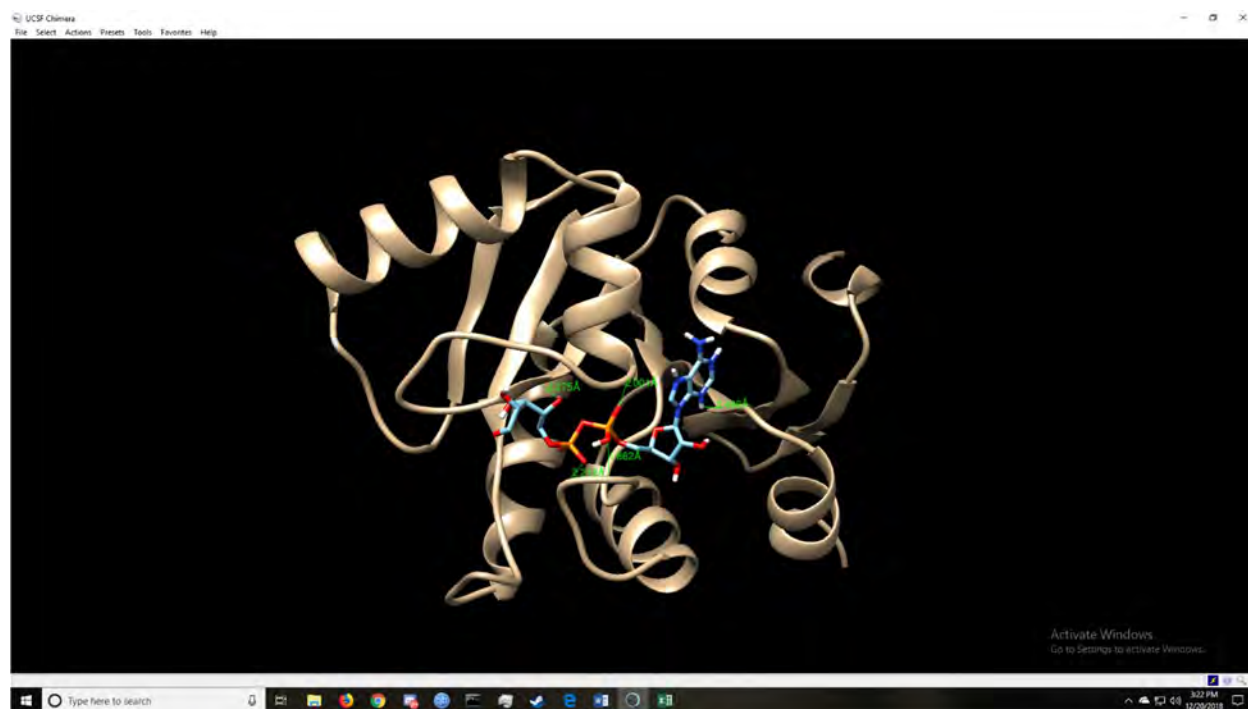


Figure 3.18: Intermolecular bonding interaction between ADP-ribose and the nsp3 residue



Figure 3.19: Intermolecular bonding interaction between Apiin (PubChem ID-5280746) and nsp3 residue with a binding affinity of – 10.1 kcal/mol, which is the best compared to other ligands except for Naringin

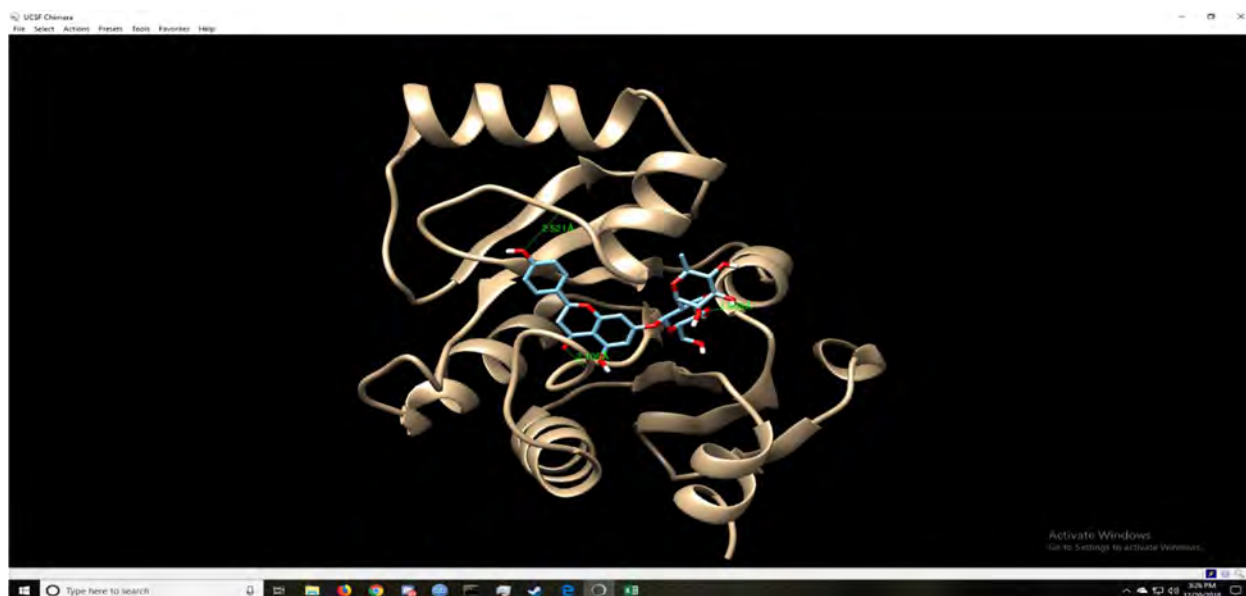


Figure 3.20: Intermolecular bonding interaction between Naringin (PubChem ID-442428) and nsp3 residue with a binding affinity of -10.1 kcal/mol, which is the best compared to other ligands except for Apiin

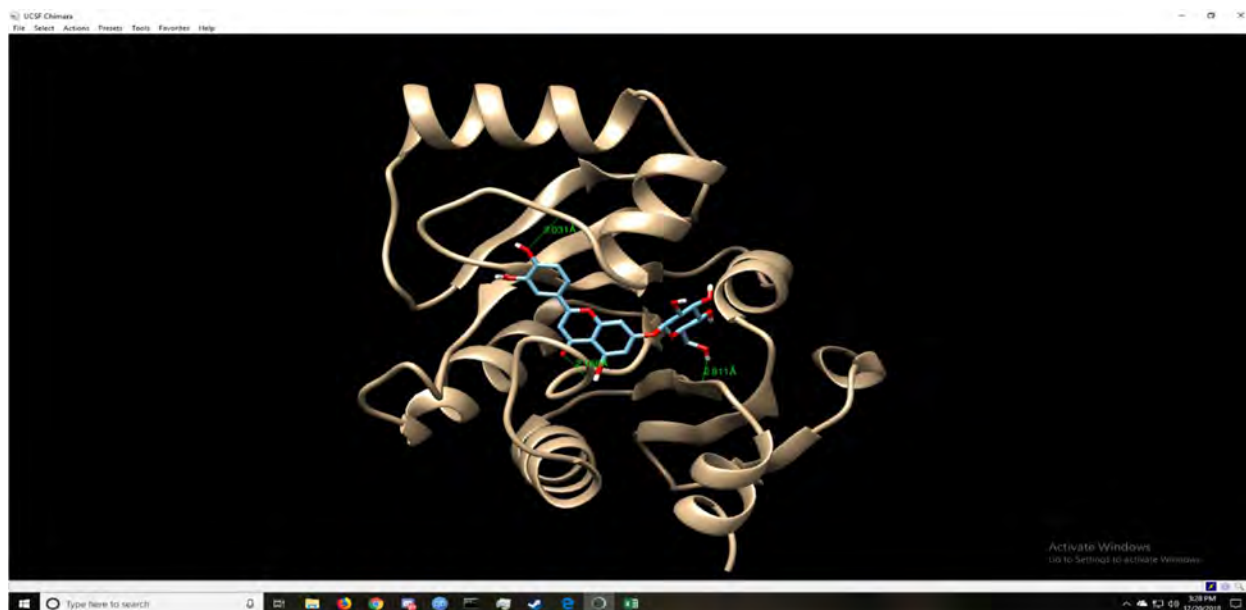


Figure 3.21: Intermolecular bonding interaction between Luteoloside (PubChem ID-5280637) and nsp3 residue



Figure 3.22: Intermolecular bonding interaction between Hesperidin (PubChem ID-10621) and nsp3 residue

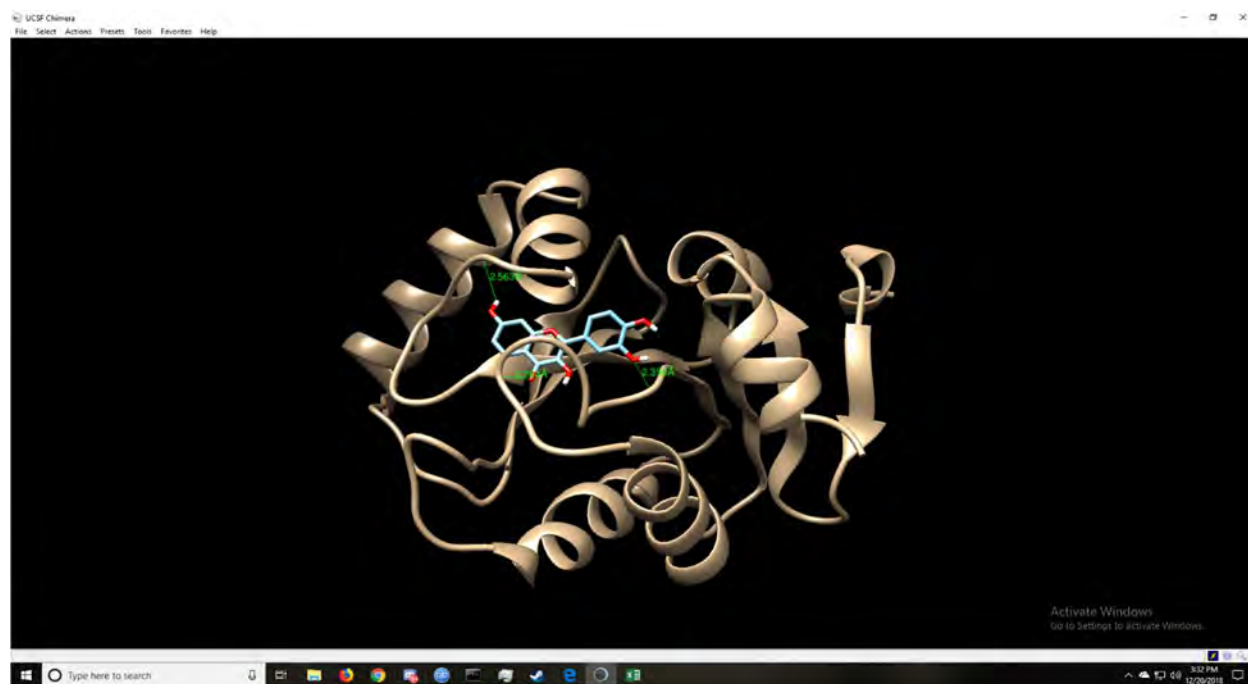


Figure 3.23: Intermolecular bonding interaction between Fisetin (PubChem ID-5281614) and nsp3 residue

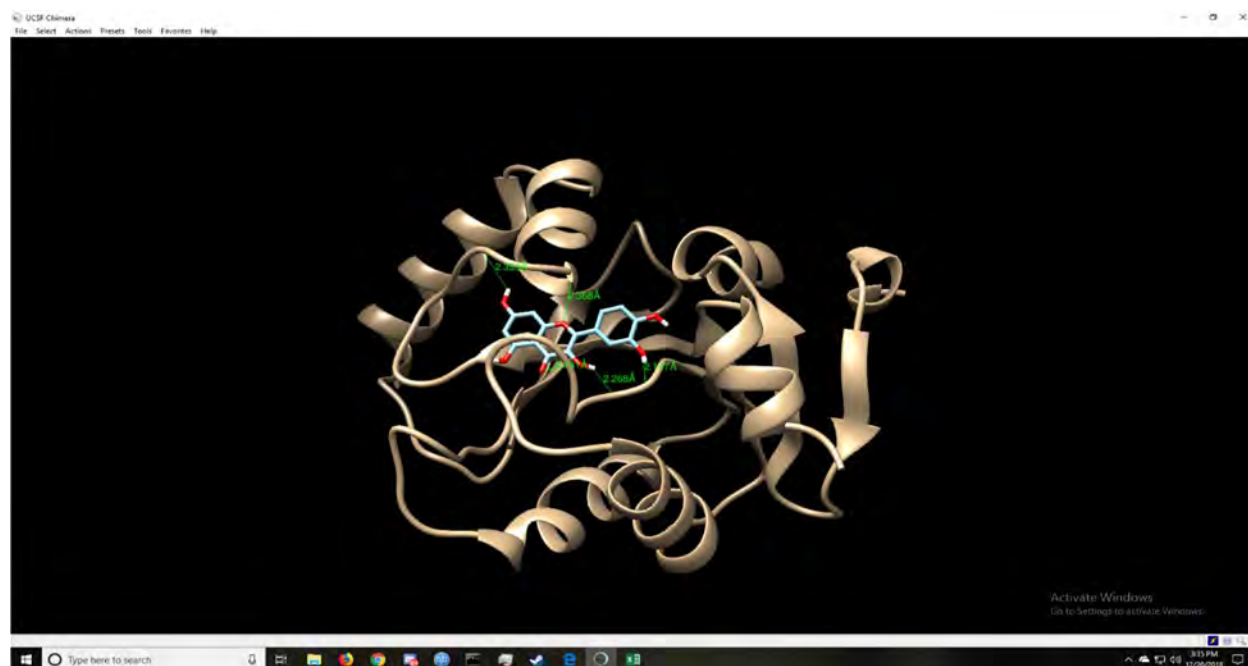


Figure 3.24: Intermolecular bonding interaction between Quercetin (PubChem ID-5280343) and nsp3 residue

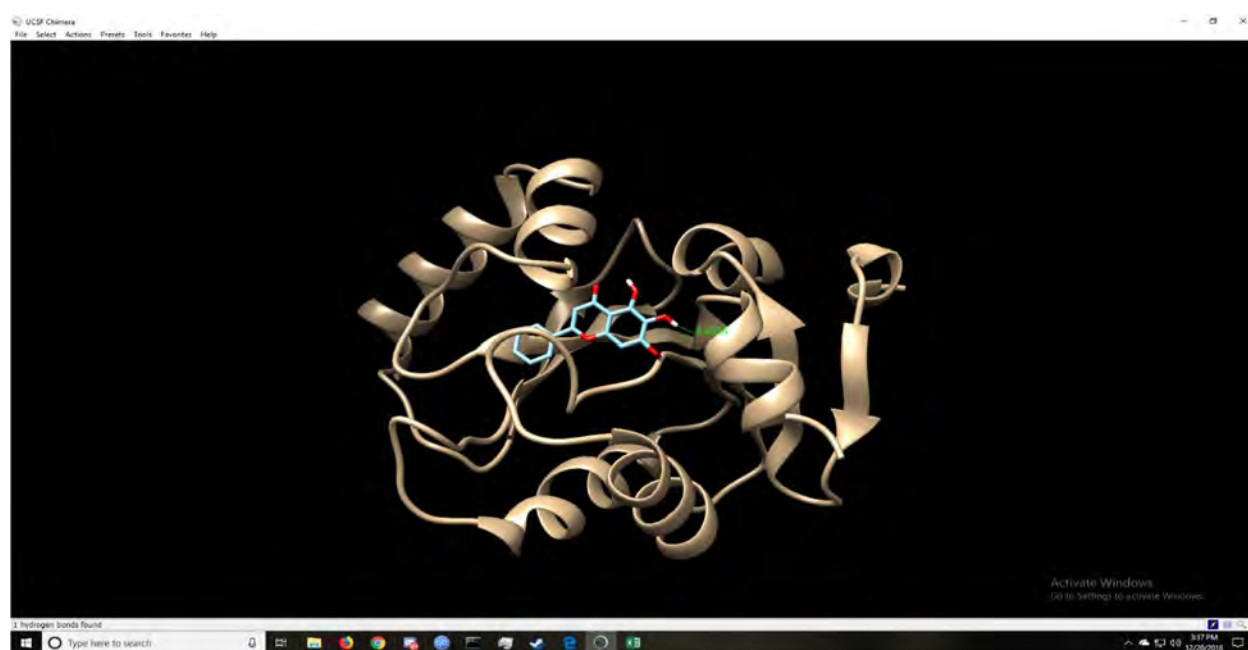


Figure 3.25: Intermolecular bonding interaction between Baicalein (PubChem ID-5281605) and nsp3 residue

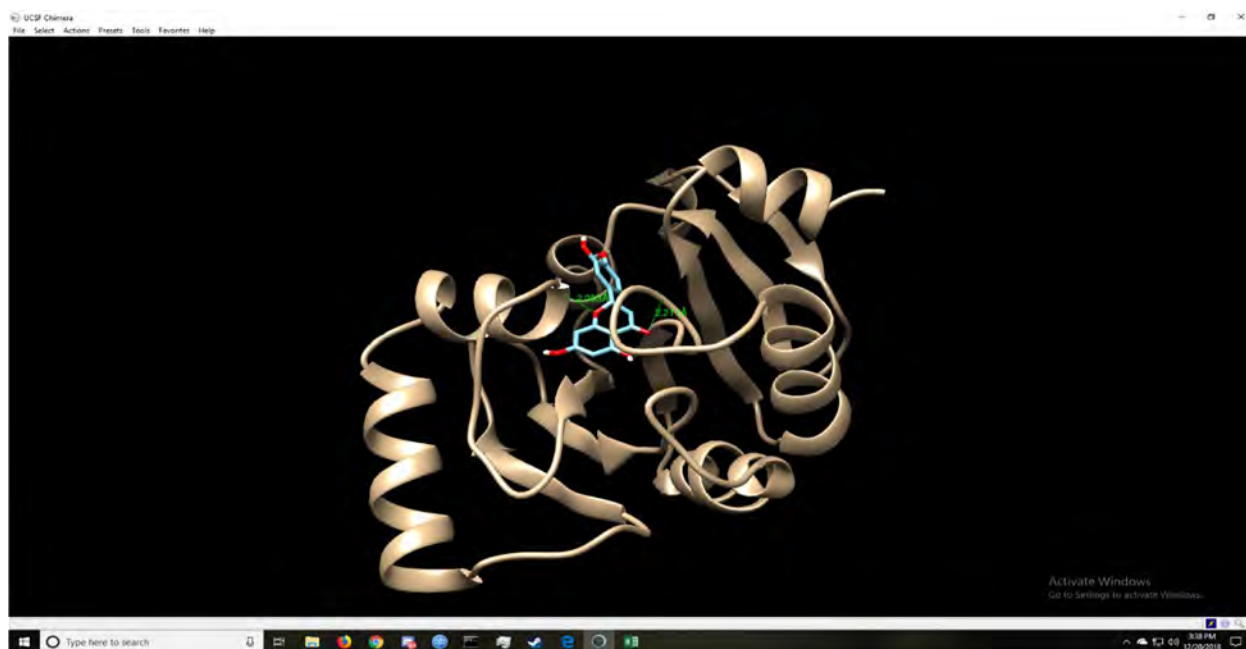


Figure 3.26: Intermolecular bonding interaction between Hesperetin (PubChem ID-72281) and nsp3 residue



Figure 3.27: Intermolecular bonding interaction between Pinostrobin (PubChem ID-73201) and nsp3 residue



Figure 3.28: Intermolecular bonding interaction between Rutin (PubChem ID-5280805) and nsp3 residue

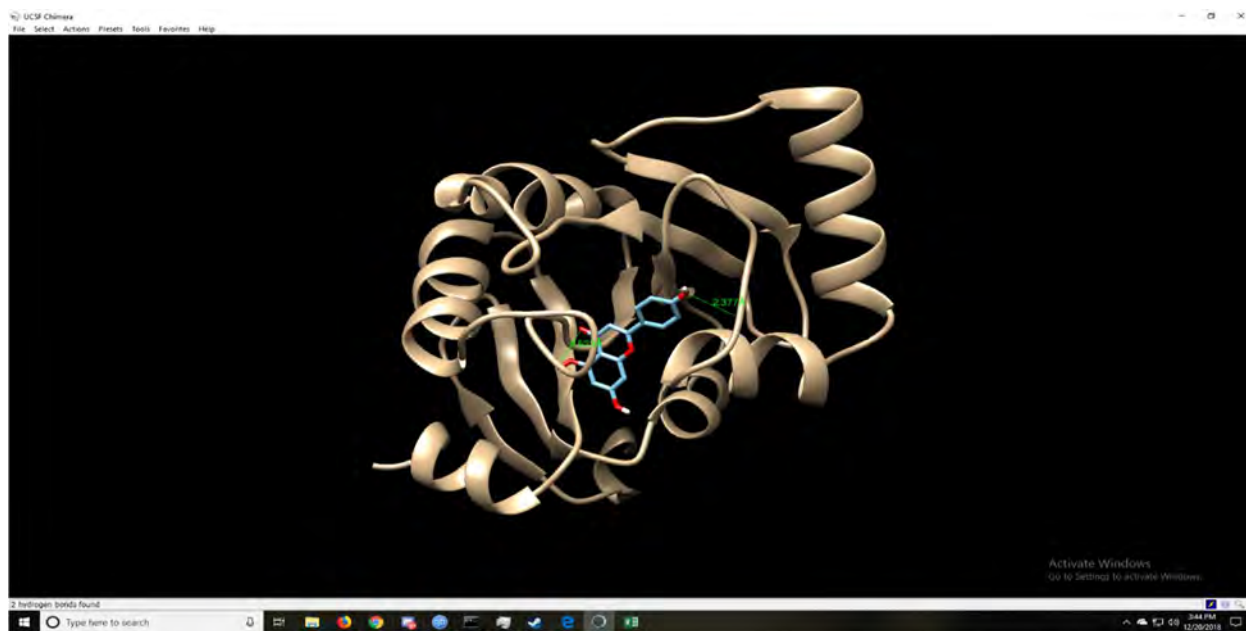


Figure 3.29: Intermolecular bonding interaction between Apigenin (PubChem ID-5280443) and nsp3 residue

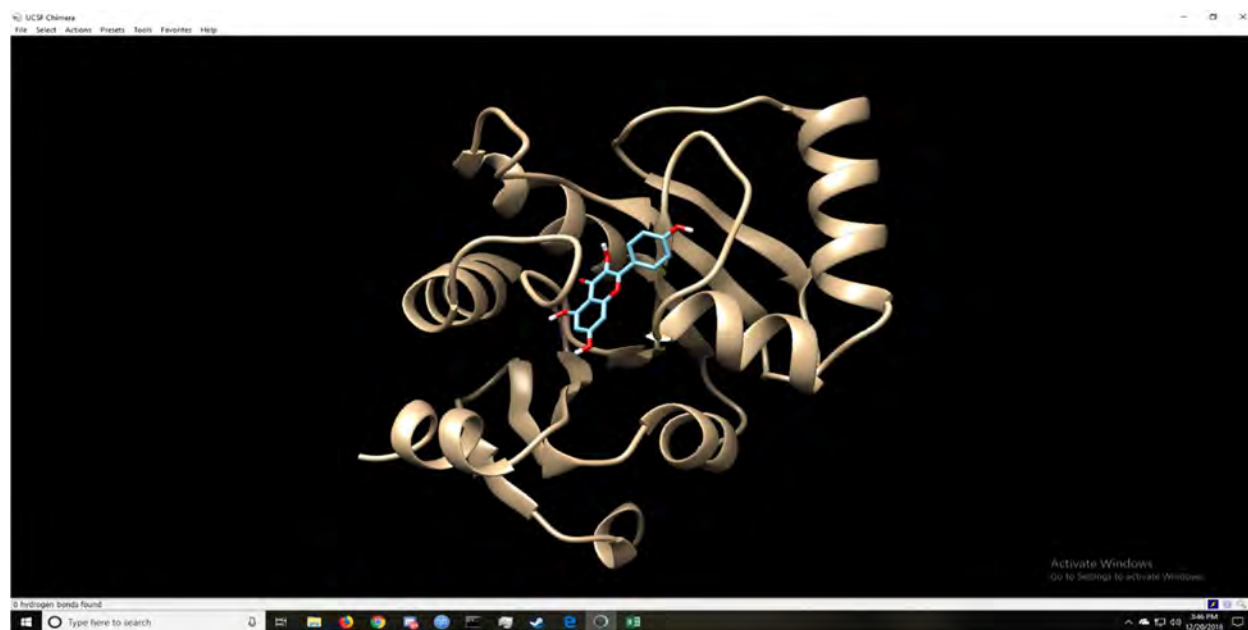


Figure 3.30: No visible intermolecular bonding interaction between Kaempferol (PubChem ID-5280863) and the nsp3 residue

Discussion

MERS-CoV is an emerging virus originated in the Middle East where this virus crossed the interspecies barrier and infected humans. MERS-CoV is an enveloped, positive-sense, single-stranded RNA viruses with a genome size of about 30-kb. Out of 16 non-structural proteins encoded by MERS-CoV, nsp3 is the largest protein. nsp3 has a macro domain embedded in which can bind to ADP-ribose. More importantly, nsp3 is an essential component of the replication/transcription complex. Therefore, the discovery of an nsp3 inhibitor can be a major leap towards developing a vaccine/therapeutic agent that can interfere with MERS-CoV replication.

In the present study, the aim was to find vaccines/therapeutic agents against MERS-CoV using immunoinformatics since there is no approved MERS-CoV-specific vaccine/therapeutic agent available for treatment (Fehr et al., 2016). The main purposes of using immunoinformatics are:

- Saving time required to develop vaccines/therapeutic agents
- Reducing the cost needed for laboratory analysis and vaccine development

For this study, only open-source immunoinformatics software and tools were used.

This study was divided into two sections. The first section was focused on designing an epitope-based vaccine whereas the second section was focused on identifying effective flavonoids in order to use them as nsp3 inhibiting therapeutic agents against MERS-CoV.

To design an epitope-based vaccine, T-cell or B-cell epitopes are screened using several software and tools. In the present study, nsp3 protein sequence was extracted from the NCBI database and was checked for antigenicity using VaxiJen 2.0. The result of VaxiJen 2.0 indicates that nsp3 protein is antigenic with a value of 0.4794 which is over the threshold for virus model (0.4). Finding conserved regions using multiple sequence alignment was avoided since MERS-CoV genomes share more than 99% sequence identity indicating low mutation rate and low variance among the genomes (Chafekar et al., 2018).

For prediction of T-cell epitopes, the sequence of nsp3 protein was put in the NetCTL 1.2 server to identify probable T-cell epitopes in the target sequence. Thirty T-cell epitopes were selected which achieved threshold value of 1.25 (Table 3.1). In the NetCTL 1.2 server, achieving a score of 1.25 means having a specificity value of 0.993 and a sensitivity value of 0.54 (Larsen et al., 2007). Antigenicity of selected epitopes was then evaluated using VaxiJen 2.0 followed by IEDB

T-cell class I pMHC immunogenicity predictor. Only six T-cell epitopes had achieved the threshold value of 0.4 in VaxiJen 2.0 and positive immunogenicity score in IEDB T-cell class I pMHC immunogenicity predictor (Table 3.4). In a similar in-silico study on Oropouche Virus, the minimum requirement for screening epitope was a score of 0.0 instead of the threshold value of 0.4 in VaxiJen 2.0 (Adhikari et al., 2018). In the present study, however, the minimum requirement for screening T-cell epitope candidates was a score of 0.4 which is the threshold value for any virus model to ensure a better quality of epitope candidates while sacrificing quantity. Then prediction of peptide-MHC class I binding was performed using both Proteasomal cleavage/TAP transport/MHC class I combined predictor and NetMHC 4.0 server. For prediction of peptide-MHC class I binding, only 9-mer peptide length was selected as most HLA molecules have a strong preference for binding 9-mer peptides. Proteasomal cleavage/TAP transport/MHC class I combined predictor gives an output result for HLA-binding affinity of the epitopes in the IC₅₀ nM unit. A lower IC₅₀ value indicates higher binding affinity of the epitopes with the MHC class I molecule. According to the website, peptides with IC₅₀ values <50 nM are considered high affinity, <500 nM intermediate affinity and <5000 nM low affinity. Most known epitopes have high or intermediate affinity, but no known T-cell epitope has an IC₅₀ value greater than 5000. Therefore, IC₅₀ values less than 200 nM (IC₅₀ < 200nM) were chosen for ensuring higher affinity similar to the in-silico study on Oropouche Virus. On the other hand, percentile rank <2.0 was used as the threshold in NetMHC 4.0 server. According to the website, the peptide will be identified as a strong binder if the percentile rank is below the specified threshold for the strong binders, by default 0.5%. The peptide will be identified as a weak binder if the percentile rank is above the threshold of the strong binders but below the specified threshold for the weak binders, by default 2%. Therefore, percentile rank <2.0 was used as the threshold to include both strong and weak binders. The selected T-cell epitopes were found to be recognized by the significant MHC class-I molecule such as HLA-A, HLA-B, and HLA-C. In Proteasomal cleavage/TAP transport/MHC class I combined predictor tool, FAFETGLAY showed the highest affinity and was recognized by 20 MHC-I alleles (HLA-B*35:01,HLA-C*03:02,HLA-B*15:25,HLA-C*12:03,HLA-B*15:02,HLA-C*16:01,HLA-C*12:02,HLA-A*29:02,HLA-B*15:01,HLA-C*03:03,HLA-C*14:02,HLA-C*02:02,HLA-C*02:09,HLA-B*53:01,HLA-B*46:01,HLA-C*08:01,HLA-B*18:01,HLA-A*26:01,HLA-A*30:02,HLA-A*68:01). FAFETGLAY showed the highest affinity in NetMHC 4.0 server as well and was recognized by 37 MHC-I alleles (HLA-

A*01:01,HLA-A*25:01,HLA-A*26:01,HLA-A*26:02,HLA-A*26:03,HLA-A*29:02,HLA-A*30:02,HLA-A*32:07,HLA-A*66:01,HLA-A*68:01,HLA-A*68:23,HLA-A*80:01,HLA-B*08:02,HLA-B*14:02,HLA-B*15:01,HLA-B*15:02,HLA-B*15:03,HLA-B*18:01,HLA-B*27:20,HLA-B*35:01,HLA-B*40:13,HLA-B*46:01,HLA-B*51:01,HLA-B*53:01,HLA-B*58:01,HLA-B*83:01,HLA-C*03:03,HLA-C*05:01,HLA-C*06:02,HLA-C*07:01,HLA-C*07:02,HLA-C*08:02,HLA-C*12:03,HLA-C*14:02,HLA-C*15:02,HLA-A*32:15,HLA-B*15:17). In addition, FAFETGLAY was recognized by 44 MHC-I alleles combined (HLA-A*01:01,HLA-A*25:01,HLA-A*26:01,HLA-A*26:02,HLA-A*26:03,HLA-A*29:02,HLA-A*30:02,HLA-A*32:07,HLA-A*66:01,HLA-A*68:01,HLA-A*68:23,HLA-A*80:01,HLA-B*08:02,HLA-B*14:02,HLA-B*15:01,HLA-B*15:02,HLA-B*15:03,HLA-B*15:25,HLA-B*18:01,HLA-B*27:20,HLA-B*35:01,HLA-B*40:13,HLA-B*46:01,HLA-B*51:01,HLA-B*53:01,HLA-B*58:01,HLA-B*83:01,HLA-C*02:02,HLA-C*02:09,HLA-C*03:02,HLA-C*03:03,HLA-C*05:01,HLA-C*06:02,HLA-C*07:01,HLA-C*07:02,HLA-C*08:01,HLA-C*08:02,HLA-C*12:02,HLA-C*12:03,HLA-C*14:02,HLA-C*15:02,HLA-C*16:01,HLA-A*32:15,HLA-B*15:17). Next, FVDWRSYNY had the second highest affinity and was recognized by 23 MHC-I alleles combined (HLA-A*01:01,HLA-A*26:02,HLA-A*26:03,HLA-A*29:02,HLA-A*30:02,HLA-A*32:15,HLA-A*66:01,HLA-A*68:23,HLA-A*80:01,HLA-B*08:02,HLA-B*08:03,HLA-B*15:02,HLA-B*35:01,HLA-B*53:01,HLA-B*83:01,HLA-C*04:01,HLA-C*05:01,HLA-C*06:02,HLA-C*07:01,HLA-C*07:02,HLA-C*08:02,HLA-C*12:03,HLA-C*16:01). Then, LLLAGTLHY was recognized by 17 MHC-I alleles combined (HLA-A*01:01,HLA-A*03:01,HLA-A*29:02,HLA-A*30:02,HLA-A*66:01,HLA-A*68:23,HLA-B*08:02,HLA-B*15:01,HLA-B*15:02,HLA-B*15:03,HLA-B*15:17,HLA-B*15:25,HLA-B*35:01,HLA-B*58:01,HLA-C*03:02,HLA-A*32:15,HLA-A*80:01). Next, KTTTGIPEY was recognized by 14 MHC-I alleles combined (HLA-A*01:01,HLA-A*25:01,HLA-A*26:02,HLA-A*29:02,HLA-A*30:01,HLA-A*30:02,HLA-A*68:23,HLA-A*80:01,HLA-B*15:03,HLA-B*15:17,HLA-B*46:01,HLA-B*58:01,HLA-B*58:02,HLA-C*14:02). Finally, both LSSVYHLYV and STDFIALIM showed the least affinity as LSSVYHLYV was recognized by 11 MHC-I alleles combined (HLA-A*01:01,HLA-A*02:05,HLA-A*02:06,HLA-A*68:02,HLA-A*68:23,HLA-A*69:01,HLA-B*15:17,HLA-B*58:01,HLA-C*05:01,HLA-C*12:03,HLA-C*15:02), along with STDFIALIM (HLA-A*01:01,HLA-A*26:03,HLA-A*32:01,HLA-A*69:01,HLA-A*80:01,HLA-B*15:17,HLA-

B*39:01,HLA-C*05:01,HLA-C*08:02,HLA-C*15:02,HLA-C*16:01). HLA-A*01:01 is the only MHC-I allele that had an affinity with all of the selected T-cell epitopes. Afterward, the selected epitopes were used for the prediction of MHC-II alleles and their respective peptide or CD4+ T-cell epitope using IEDB Peptide binding to MHC class II molecules predictor while selecting IC50 < 3000nM as threshold since CD4+ T-cell epitopes play an important role in eliciting protective immune responses during peptide-based vaccination (Oyarzún et al., 2013). FAFETGLAY was recognized by 37 MHC-II alleles. As MHC HLA allele distribution differs among diverse geographic regions and ethnic groups around the world, population coverage must be taken into consideration during the design of an effective vaccine. In this study, identified MHC-I-binding alleles with high binding affinity of six epitopes were considered to analyze population coverage using the IEDB Population Coverage Analysis tool. The IEDB Population Coverage Analysis tool revealed that these epitopes and their HLA-alleles cover 98.55% of the world population cumulatively (Figure 3.5). The highest population coverage was found in the South Africa region (99.66%) while the lowest population coverage was found in Central America (9.07%). MERS was first found in the Middle East, and several outbreaks have been recorded in this region. Therefore, the population coverage prediction in the Middle East is essential for vaccine design. In the IEDB Population Coverage Analysis tool, Middle East is listed as Southwest Asia. The cumulative population coverage in Southwest Asia was 96.40%. In addition to that, the cumulative population coverage in East Asia where the Republic of Korea was located was 96.86%. These results indicate that these epitopes are promising vaccine candidates. Then those six epitopes were checked for conservancy using IEDB conservancy analysis tool. Conservancy analysis revealed that all of them are 100% conserved as they had the maximum identity (100%) for conservancy hit. In addition, to ensure that the epitope vaccines will not harm host cells, the toxicity of the epitope candidates was predicted using ToxinPred. All six of the selected epitopes were non-toxic. Finally, allergenicity was anticipated using AllergenFP v1.0 and AllerTOP v2.0 as many vaccines stimulate an allergenic reaction in the human body which can create several problems and may even hinder vaccine development (Oany et al., 2014). STDFIALIM, LSSVYHLYV, FAFETGLAY, LLLAGTLHY, FVDWRSYNY were found as probable allergens in AllergenFP v1.0 while KTTTGIPEY was found non-allergenic. On the contrary, STDFIALM and LSSVYHLYV were found non-allergenic in AllerTOP v2.0. KTTTGIPEY was found non-allergenic in both AllergenFP v1.0 and AllerTOP v2.0. The 3D structures of selected epitope

candidates were predicted using PEP-FOLD 2.0 server. NP44-S7N mutant peptide (CTELKLNDY) was also selected as control ligand. HLA-A*01:01 was selected as the macromolecule. After the minimizing process, HLA-A*01:01 protein was placed in a grid box measuring $52.8351 \text{ \AA} \times 68.2709 \text{ \AA} \times 61.7293 \text{ \AA}$ along the x, y and z axis, respectively, where the position of the center was X:-63.5001, Y:-17.1718, Z:7.5672. The docking procedure was performed using the instructed command prompts. The docking poses were ranked according to their docking scores. However, the docking result revealed two different binding sites. Therefore, instead of choosing the conformation with lowest binding affinity as best docking pose, the conformation with the lowest binding affinity that used the same binding site as control (NP44-S7N mutant peptide) was selected as best docking pose in order to compare between sample and control for critical evaluation. NP44-S7N mutant peptide had a binding affinity of -8.7kcal/mol. Among these six epitopes, the binding affinity of LSSVYHLYV was the best of all (-8.0kcal/mol). However, Only FAFETGLAY formed visible intermolecular hydrogen bond similar to the control ligand (NP44-S7N mutant peptide). FAFETGLAY had a binding affinity of -7.5 kcal/mol which was relatively high, and the distance of hydrogen bond was 2.089 \AA which was quite similar to control ligand (2.095 \AA).

For prediction of B-cell epitopes, nsp3 protein sequence was put in B-cell epitope predictor tools such as BCPREDS and BepiPred 2.0. These tools generated a repertoire of probable B-cell epitope candidates. These epitopes were screened using VaxiJen 2.0. After that, the number of epitopes candidates was reduced to 178. Then overlapping B-cell and T-cell epitopes were identified similar to a study focused on emerging Rift Valley fever virus (Adhikari et al., 2017). Out of 178 B-cell epitopes, only eight had the sequence similarity with the selected T-cell epitopes (Table 3.12). IFVDWRSYNYAVSS, FVDWRSYNYAVS, FVDWRSYNYAVSSAFW and FVDWRSYNYAVSSAFWLF had the sequence similarity with FVDWRSYNY whereas LKFKEVCKTTTGIPEYNF, LKFKEVCKTTTGIPEY, FKEVCKTTTGIPEYNFIIYD and VCKTTTGIPEYN had the sequence similarity with KTTTGIPEY. These eight B-cell epitopes were selected for further evaluation. Conservancy analysis of these B-cell epitopes using IEDB conservancy analysis tool revealed that all of them had the maximum identity (100%) for conservancy hit. Afterward, these epitopes were checked for the presence of beta-turn, surface accessibility, flexibility, antigenicity and hydrophilicity using several IEDB B-cell tools. Results of IEDB B-cell tools showed that among eight B-cell epitopes, LKFKEVCKTTTGIPEYNF was

the most promising B-cell epitope as it had the presence of beta-turn, surface accessibility, flexibility, high antigenicity and hydrophilicity. LKFKEVCKTTTGIPEYNF was checked for the presence of beta-turn by Chou & Fasman Beta Turn Prediction tool, and it was revealed that 7 out of 12 peptide fragments were above the threshold. In Emini Surface Accessibility Prediction tool, 7 out of 13 peptide fragments were above the threshold and found to be surface accessible. 6 out of 11 peptide fragments were found flexible in Karplus and Schulz Flexibility Prediction tool. Kolaskar & Tongaonkar Antigenicity Prediction tool indicated that 6 out of 12 peptide fragments were antigenic. Lastly, hydrophilicity was checked by Parker Hydrophilicity Prediction tool, and it was revealed that 7 out of 12 peptide fragments were above the threshold. LKFKEVCKTTTGIPEYNF, along with other epitopes, was found as non-toxic in ToxinPred. However, allergenicity prediction results were not conclusive as LKFKEVCKTTTGIPEYNF was found as a probable allergen in AllergenFP v1.0 but AllerTOP v2.0 identified LKFKEVCKTTTGIPEYNF as non-allergen. The analysis of antigenicity, conservancy, surface accessibility, flexibility, hydrophilicity, toxicity and allergenicity of B-cell epitopes revealed that the epitope FVDWRSYNYAVSSAFWLF and LKFKEVCKTTTGIPEY could be the most potential B-cell epitope candidate for peptide-based vaccine design among the selected eight B-cell epitopes because these epitopes are antigenic, 100% conserved, flexible, hydrophilic non-toxic and non-allergenic to the human. In addition to that, they performed well in Chou & Fasman Beta Turn Prediction tool. These epitopes, however, performed a bit poorly in the Emini Surface Accessibility Prediction tool. FVDWRSYNYAVSSAFWLF had only 4 out of 13 peptide fragments above the threshold in the Emini Surface Accessibility Prediction tool. Similarly, LKFKEVCKTTTGIPEY had 5 out of 11 peptide fragments above the threshold in the Emini Surface Accessibility Prediction tool.

Molecular docking is a study of how two or more molecular structures fit together and interact. Molecular docking, especially protein-ligand docking, has become a very popular bioinformatics-based drug designing tool in the medical industry and within academic communities for predicting preferred binding orientations or poses of a ligand to a macromolecule. In the present study, molecular docking was used to check whether flavonoids could be used as a therapeutic agent against MERS-CoV or not. In this study, 18 flavonoids were selected as potential nsp3 inhibitor candidates (Table 3.15). Each of these flavonoids was used as ligand separately in molecular docking. As for macromolecule, MERS-CoV macro domain within nsp3 protein was selected

which is known to be an efficient ADP-ribose binding module (Cho et al., 2016). First, 3D structures of ligands were retrieved from PubChem in SDF format which was PyRx compatible. Then, the three-dimensional crystal structure of the macro domain within nsp3 protein was retrieved from RCSB Protein Data Bank (PDB ID:5DUS) in PDB format. Undesired ligands such as ADP-ribose, Sulfate Ion and Glycerol were removed, and then minimization process was done using Dock Prep tool in UCSF Chimera 1.13. This process removed water molecules as well. Ligand minimization was done by PyRx prior to docking. After the minimizing process, PyRx was used for molecular docking. PyRx was chosen for molecular docking instead of AutoDock Vina due to its perceived simplicity. The protein was placed in a grid box measuring $37.3660 \text{ \AA} \times 43.4316 \text{ \AA} \times 43.1478 \text{ \AA}$ along the x, y and z axis, respectively, where the position of the center was X:8.9843, Y:17.6095, Z:68.5928. ADP-ribose was first re-docked into the ADP-ribose binding site of nsp3, and the resulting interactions were later compared with those found by docking 18 flavonoids into the similar active site using the same grid box. This type of comparative study is more reliable than binding affinity alone. The docking results included the binding energy value given in kcal/mol, mode, RMSD upper bound (rmsd/ub) and RMSD lower bound (rmsd/lb). The docking poses were ranked according to their docking scores. The conformation with the lowest binding affinity was selected as the best docking pose (Table 3.19). The best docking conformation of ADP-ribose showed a binding affinity of -8.7 kcal/mol. Only 12 flavonoids (Apiin, Naringin, Luteoloside, Hesperidin, Fisetin, Quercetin, Baicalein, Hesperetin, Pinostrobin, Rutin, Apigenin and Kaempferol) had binding affinity greater than -8.7 kcal/mol. Among them, apiin and naringin had the best binding affinity of -10.1 kcal/mol. Afterward, the K_i value was measured for each compound (Table 3.20). According to a similar study on potential anti-chikungunya activity of baicalin, naringenin and quercetagenin, drugs with a K_i value $<1 \text{ mM}$ are normally considered to be effective (Seyedi et al., 2016). The present study revealed that apiin, naringin, luteoloside, hesperidin and fisetin had significantly lower K_i values (less than $0.08 \text{ }\mu\text{M}$), with apiin and naringin having the lowest K_i value ($0.0390649 \text{ }\mu\text{M}$). Finally, the molecular visualization of the docked complexes was performed using UCSF Chimera 1.13 and the intermolecular H bonds between each compound with MERS-CoV nsp3 macro domain were listed along with their respective distances (Table 3.21). According to the previously mentioned study, a review by Szatyłowicz classified the energy borders setting for strong, moderate and weak H-bonds where 1.2–1.5 is considered strong, >1.5 –2.2 is considered moderate and >2.2 is considered

weak (Seyedi et al., 2016). In the present study, three intermolecular hydrogen bonds between apiin and nsp3 fell under the moderate bond class while two other intermolecular hydrogen bonds were considered as weak. Similarly, two intermolecular hydrogen bonds between naringin and nsp3 fell under the moderate bond class while one intermolecular hydrogen bond was considered as weak. Other than apiin, no flavonoid had more than two hydrogen bonds categorized as moderate bond. On the other hand, kaempferol did not form any visible hydrogen bond at all (Figure 3.30).

The key findings of the present study were:

- STDFIALIM, LSSVYHLYV, FAFETGLAY, LLAGTLHY, FVDWRSYNY, KTTTGIPEY are 100% conserved, non-toxic and their HLA-alleles cover 98.55% of the world population cumulatively. FAFETGLAY was recognized by 44 MHC-I alleles (cumulatively) and 37 MHC-II alleles. In addition, FAFETGLAY is the only epitope candidate that formed a visible intermolecular hydrogen bond with MHC HLA-A*01:01 allele in the molecular docking simulation study. However, both AllergenFP v1.0 and AllerTOP v2.0 suggested that FAFETGLAY was a probable allergen. On the other hand, KTTTGIPEY was recognized by 14 MHC-I alleles cumulatively and was found non-allergenic in both AllergenFP v1.0 and AllerTOP v2.0. However, molecular docking simulation revealed that KTTTGIPEY did not form any hydrogen bond with MHC HLA-A*01:01 allele. Since FAFETGLAY performed the best in the molecular docking simulation study, more studies are required to confirm whether FAFETGLAY really induces allergenic reactions or not.
- FVDWRSYNYAVSSAFWLF and LKFKEVCKTTTGIPEY could be the most potential B-cell epitope candidates for peptide-based vaccine design among the selected eight B-cell epitopes. However, LKFKEVCKTTTGIPEYNF would be the best B-cell epitope candidate if any conclusive evidence ensuring that LKFKEVCKTTTGIPEYNF will not induce any allergenic reaction is found in future as LKFKEVCKTTTGIPEYNF performed better than FVDWRSYNYAVSSAFWLF and LKFKEVCKTTTGIPEY in IEDB B-cell tools.

- Apiin and naringin exhibited the most potent antiviral activity against MERS-CoV nsp3 with a binding affinity of -10.1 kcal/mol and can be considered good candidates for further evaluation as potential antiviral agents against MERS-CoV.

These epitope candidates can be used to develop a multi-epitope vaccine against MERS-CoV. In addition to that, apiin and naringin can be used in the combination therapy along with other antiviral agents which may further increase the efficacy of the multi-epitope vaccine. However, more studies are required to develop an effective vaccine/therapeutic agent against MERS-CoV. Using commercial-grade bioinformatics tools may increase the quality of the study greatly. Furthermore, other factors such as the route of delivery, use of compatible adjuvants etc. should be considered while designing vaccine/therapeutic agent. Finally, since *in silico* results often deviate from the original outcome, *in vitro* experiment should be conducted to check whether a designed anti-viral agent is really effective or not.

References

References:

- Adhikari, U. K., & Rahman, M. M. (2017). Overlapping CD8+ and CD4+ T-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach. *Infection, Genetics and Evolution*, 56:75–91. doi:10.1016/j.meegid.2017.10.022
- Adhikari, U. K., Tayebi, M., & Rahman, M. M. (2018). Immunoinformatics Approach for Epitope-Based Peptide Vaccine Design and Active Site Prediction against Polyprotein of Emerging Oropouche Virus. *Journal of immunology research*, 2018, 6718083. doi:10.1155/2018/6718083
- Andreatta, M., & Nielsen, M. (2015). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*, 32(4), 511-517.
- Chafekar, A., & Fielding, B. C. (2018). MERS-CoV: Understanding the Latest Human Coronavirus Threat. *Viruses*, 10(2), 93. doi:10.3390/v10020093
- Chan, J. F., Lau, S. K., To, K. K., Cheng, V. C., Woo, P. C., & Yuen, K. Y. (2015). Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clinical microbiology reviews*, 28(2), 465-522.
- Cho, C. C., Lin, M. H., Chuang, C. Y., & Hsu, C. H. (2016). Macro Domain from Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Is an Efficient ADP-ribose Binding Module: CRYSTAL STRUCTURE AND BIOCHEMICAL STUDIES. *The Journal of biological chemistry*, 291(10), 4894-902.
- Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. In *Chemical Biology* (pp. 243-250). Humana Press, New York, NY.
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, 8(1), 4.
- El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2008). Predicting linear B-cell epitopes using string kernels. *Journal of molecular recognition : JMR*, 21(4), 243-55.
- Fehr, A. R., Channappanavar, R., & Perlman, S. (2016). Middle East Respiratory Syndrome: Emergence of a Pathogenic Human Coronavirus. *Annual review of medicine*, 68, 387-399.
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G. P., & Open Source Drug Discovery Consortium. (2013). In silico approach for predicting toxicity of peptides and proteins. *PloS one*, 8(9), e73957.
- Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic acids research*, 45(W1), W24-W29.
- Kametani, Y., Miyamoto, A., Tsuda, B., & Tokuda, Y. (2015). B Cell Epitope-Based Vaccination Therapy. *Antibodies*, 4(3), 225–239. doi:10.3390/antib4030225
- Lambert, D. M., & Fowler, C. J. (2005). Therapeutic Applications, 48(16), 1077–1088.
- Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC bioinformatics*, 8(1), 424.
- Matthews, K. L., Coleman, C. M., van der Meer, Y., Snijder, E. J., & Frieman, M. B. (2014). The ORF4b-encoded accessory proteins of Middle East respiratory syndrome coronavirus and two related bat coronaviruses localize to the nucleus and inhibit innate immune signalling. *The Journal of general virology*, 95(Pt 4), 874-82.
- Milne-Price, S., Miazgowicz, K. L., & Munster, V. J. (2014). The emergence of the Middle East respiratory syndrome coronavirus. *Pathogens and disease*, 71(2), 121-36.
- Nielsen, M., Lundegaard, C., Wornig, P., Lauemøller, S. L., Lamberth, K., Buus, S., ... & Lund, O. (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science*, 12(5), 1007-1017.

- Niemeyer, D., Zillinger, T., Muth, D., Zielecki, F., Horvath, G., Suliman, T., Barchet, W., Weber, F., Drosten, C., ... Müller, M. A. (2013). Middle East respiratory syndrome coronavirus accessory protein 4a is a type I interferon antagonist. *Journal of virology*, 87(22), 12489-95.
- Oany, A. R., Emran, A. A., & Jyoti, T. P. (2014). Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach. *Drug design, development and therapy*, 8, 1139-49. doi:10.2147/DDDT.S67861
- Orhan, DD., Özçelik, B., Özgen, S., Ergun, F. (2010). Antibacterial, antifungal, and antiviral activities of some flavonoids. *Microbial Research*, 165(6):496-504. doi:10.1016/j.micres.2009.09.002
- Oyarzún, P., Ellis, J. J., Bodén, M., & Kobe, B. (2013). PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. *BMC bioinformatics*, 14, 52. doi:10.1186/1471-2105-14-52
- Panche, A. N., Diwan, A. D., & Chandra, S. R. (2016). Flavonoids: an overview. *Journal of nutritional science*, 5, e47. doi:10.1017/jns.2016.41
- Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open biology*, 3(1), 120139. doi:10.1098/rsob.120139
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), 1605-1612.
- Seyedi, S. S., Shukri, M., Hassandarvish, P., Oo, A., Shankar, E. M., Abubakar, S., & Zandi, K. (2016). Computational Approach Towards Exploring Potential Anti-Chikungunya Activity of Selected Flavonoids. *Scientific reports*, 6, 24027. doi:10.1038/srep24027
- Shen, Y., Maupetit, J., Derreumaux, P., & Tufféry, P. (2014). Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *Journal of chemical theory and computation*, 10(10), 4745-4758.
- Skwarczynski, M., & Toth, I. (2015). Peptide-based synthetic vaccines. *Chemical science*, 7(2), 842-854.
- Taylor, K. E., & Mossman, K. L. (2013). Recent advances in understanding viral evasion of type I interferon. *Immunology*, 138(3), 190-7.
- Thévenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P., & Tufféry, P. (2012). PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic acids research*, 40(Web Server issue), W288-93.
- Van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T. M., Raj, V. S., Zaki, A. M., Osterhaus, A. D., Haagmans, B. L., Gorbalenya, A. E., Snijder, E. J., ... Fouchier, R. A. (2012). Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio*, 3(6), e00473-12. doi:10.1128/mBio.00473-12
- Yang, Y., Zhang, L., Geng, H., Deng, Y., Huang, B., Guo, Y., Zhao, Z., ... Tan, W. (2013). The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. *Protein & cell*, 4(12), 951-61.
- Zhang, N., Jiang, S., & Du, L. (2014). Current advancements and potential strategies in the development of MERS-CoV vaccines. *Expert review of vaccines*, 13(6), 761-74.
- Zumla, A., Hui, D. S., & Perlman, S. (2015). Middle East respiratory syndrome. *Lancet (London, England)*, 386(9997), 995-1007.