Analysis of Envelope Glycoprotein (E2) of Hepatitis C Virus (HCV) for Proposing Efficient Multi-Epitope Peptide Vaccine Based on Immuno-Informatics Approach

By

Adiba Rahman

18146072

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy (Hons.)

School of Pharmacy
Brac University
November 2022

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Adiba Rahman
ID: 18146072

Approval

The thesis titled "Analysis of Envelope Glycoprotein (E2) of Hepatitis C Virus (HCV) for Proposing Efficient Multi-Epitope Peptide Vaccine Based on Immuno-Informatics Approach" submitted by Adiba Rahman (18146072), of Spring, 2018 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervised By:

_____

Mohammad Kawsar Sharif Siam
Senior Lecturer
School of Pharmacy
BRAC University

Approved By:

Program Director:

_____

Professor Dr. Hasina Yasmin
Program Director and Assistant Dean
School of Pharmacy
BRAC University

Dean:

_____

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
BRAC University

Ethics Statement

The thesis was conducted in compliance with all ethical standards required. No unethical activities were part of this thesis. No human or animal trials were used in this research as well.

Abstract

Hepatitis C virus (HCV) that causes viral hepatitis is producing 1.5 million patients every year worldwide to boost chronically infected number of people to 58 million susceptible in developing liver carcinoma and cirrhosis. Hence, for designing a preventive vaccine, cost-friendly and convenient immuno-informatics approach can be exploited that constructs multi-epitope peptide vaccine. In this study, envelope glycoprotein (E2) sequences from genotype 1, 2, 3 and 4 were evaluated due to their prevalence. E2 protein was considered as the most suitable antigen to run multiple sequence alignment (MSA) and numerous immuno-informatics web servers were used to predict efficient cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL) and B cell epitope that resulted in 1 CTL, 1 HTL and 3 B cell epitopes to get included in the vaccine construct along with HSP60 adjuvant. Finally, the vaccine peptide was analyzed in different parameters to present it as a potential candidate against HCV.

Keywords: E2 protein; MSA; CTL; HTL; B cell ; HCV

Acknowledgement

This work would not be successful without the supervision and assistance of my supervisor, Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, Brac University, who gave me directions throughout my research work and guided me to upgrade continuously.

Adiba Rahman

Id: 18146072

School of Pharmacy

Brac University

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Viral hepatitis due to hepatitis C virus (HCV) is found among 1.5 million people nearly every year worldwide and this results in 58 million patients approximately suffering from chronic infection of HCV (WHO, 2022). HCV causes both acute and chronic infection but 80% of acute illness cases turn into chronic ones according to studies that ultimately develop liver cirrhosis and liver cancer in 10-20% and 1-5% population respectively (Di Bisceglie et al., 1991; Fattovich et al., 1997; Hutin et al., 2004; Kiyosawa et al., 1990; Seeff et al., 1992). Mainly, liver carcinoma and cirrhosis due to HCV infection were the reason for the death of 2,90000 patients in 2019 specified by the World Health Organization (WHO) (WHO, 2022). Hence, HCV-inducing liver diseases are considered as burdensome to well-being and wealth for their prevalence and contribution in occurring death and have become a global concern (Blach et al., 2017). According to the "WHO Global Health Sector Strategy (GHSS) on viral hepatitis" of 2016, 65% in terms of mortality and 90% in case of incidence have to be reduced by 2030, particularly for viral hepatitis caused by hepatitis B and C virus (WHO, 2021).

To reach this target of WHO, direct-acting antivirals (DAAs) are playing a noteworthy role by rehabilitating more than 95% of HCV patients (Coppola et al., 2016; Suwanthawornkul et al., 2015) but still, they are not sufficient alone to combat transmission of the virus as well as other associated challenges. Additionally, the low rate of diagnosis of viral infection that remains symptom-less (Chen et al., 2019) and higher drug expenses have diminished the succession rate of DAAs and also have limited the reach of treatment options for chronically infected patients which increases the necessity of preventive vaccine development against HCV (Bartenschlager et al., 2018). In contrast to drugs, vaccines have always been proved to

the game changer whenever barriers have been faced in fighting infectious diseases (WHO, 2017). The success of immense immunization through vaccine against smallpox or polio are the two landmark events of humankind since, without vaccines, 5 million death is estimated due to smallpox every year and 600000 paralytic cases or death would be encountered for poliovirus (Andrei, 2021) that emphasizes again the requirement of a vaccine that offers prophylaxis to eradicate HCV.

Despite genetic heterogenicity, numerous vaccine designing have been attempted to date and those have been undertaken in phase I or II clinical trials or preclinical trials (Yu & Chiang, 2010). Several causes have been highlighted that have affected the efficacy of some candidates negatively, for instance, shortage of epitopes that are exerted from virus, misfolded proteins addition in recombinant technology, poor immune (both humoral and cellular) simulation through DNA vaccines, and implementation of low potent adjuvants (Puig et al., 2006; Torresi et al., 2011). In such reality, epitope-based peptide vaccines under broad subunit categorization may bring new dimensions regarding prevention overcoming the abovementioned obstacles in proposing an efficient vaccine for HCV.

In general, subunit vaccines are developed using single or multiple antigenic proteins (Li et al., 2014) and this idea has been further extended towards developing peptide vaccines where specific epitopes of a target antigen or protein are used that can mediate in exerting the B and T cell immune pathway (Sesardic, 1993). The reason behind taking a single protein as the primary target is that it contains several epitopes all of which are not even essential and requires filtering too. Hence, multiepitope peptide vaccines offer immune stimulant molecules consisting of multiple antigenic epitopes that induce a response in the host body likely to natural pathogens without adding allergenicity caused by additional proteins or undesired epitope fragments (Li et al., 2014).

To achieve these advantages of a peptide vaccine, the most appropriate way will be utilizing

the insilico approach exploiting tools of immunoinformatics which is an interface of immunology and computer science. In this field, computational resources are employed to understand the information of immunology (Tomar & De, 2014) and thus it is economical and expedient since, a lower number of experiments will be required (Oli et al., 2020; Tomar & De, 2014). Thereby, many available servers and tools have been applied throughout this study for predicting CTL (Cytotoxic T Lymphocytes), HTL (Helper T Lymphocytes) and B cell epitopes of the achieved conserved sequence by multiple sequence alignment (MSA) and for further analysis of the constructed vaccine sequence.

Selection of a particular structural protein was the first requisite, and in this study E2 (envelope) protein has been selected. However, a clinical trial of a recent vaccine, designed based on intense cellular immunity using nonstructural proteins has not shown success in decreasing chronic rates of HCV (Page et al., 2021). Consequently, the humoral immune response along with cellular has the better potential to fight against chronicity (Bartenschlager et al., 2018). Now, HCV-specific neutralizing antibodies (NAb) target envelope glycoproteins(E1&E2) of HCV (Augestad et al., 2021), and stimulating this humoral immunity is vital in terms of viral eradication (Keck et al., 2019; Kinchen et al., 2019; Osburn et al., 2014; Pestka et al., 2007). Since E2 protein is influential in one significant kind of immune response that is antibody effectuated as well as showed better antigenicity, it can be noted as an indispensable antigen while developing epitope-based peptide vaccine and thus E2 was chosen as primary antigen.

Chapter 2

Review on Genome, Morphology, and Life Cycle of Hepatitis C Virus

2.1 Hepatitis C Virus: Genome and Morphology

HCV belongs to the Flaviviridae family consisting of single-stranded enveloped RNA viruses and the Hepacivirus genus (Khan, A. et al., 2014). It carries a positive-strand RNA and the genome is 9.6 kb long accommodating three portions which are a 5' noncoding region or NCR, 3' NCR, and an open reading frame or ORF. The 5' NCR contains an IRES or internal ribosome entry site and the ORF eventually encodes the two kinds of proteins named structural and nonstructural for the virus. There are three structural proteins of HCV- the core protein and two envelope glycoproteins denoted by E1 and E2. Together, these three proteins functioned for the formation of the virus particle. The remaining seven are nonstructural proteins formed inside the HCV organism, those are p7 viroporin, NS5A, and NS4B proteins, RNA-dependent RNA polymerase protein (NS5B), NS2 protease protein, NTPase or RNA helicase and lastly NS3-4A harboring complex protease (Moradpour & Penin, 2013).

Viral nucleocapsid protein is formed from the core protein which is translated at first from the ORF of HCV (Santolini et al., 1994). Envelope glycoproteins of HCV have a vital contribution in its entire lifecycle for instance viral entry, aggregation of particles that are contagious as well as endosomal membrane fusion (Zeisel et al., 2011). Both E1 and E2 proteins are classified as type I transmembrane proteins consisting of 192 and 363 amino acids respectively. They are made up of the N-terminal ectodomain and C-terminal transmembrane domain (TMD) which is comparatively shorter (Moradpour & Penin, 2013).

According to the E2 ectodomain model of structural organization, the ectodomain consists of three distinct domains-D1 which carries hypervariable region 1(HVR1) at the extension part of the N terminal, D2 which carries HVR2, and D3 which is connected to D1 through the

intergenotypic variable region or igVR and also linked to TMD by a flexible region on another side (Krey et al., 2010). Mutable genes encode E1 and E2 and as mentioned before numerous HVRs are found in E2 particularly, which elicit variation of around 80% across genotypes. Among the HVRs, HVR1 has been marked as an immunodominant one by exhibiting neutralizing antibodies specifically by type (Moradpour & Penin, 2013). On the other hand, HVR2, as well as igVR, are not acknowledged as contributors to inducing humoral immune responses rather they have important roles in maintaining structural solidarity, heterodimer activities, and in modification of receptor binding of E2 (McCaffrey et al., 2011).

2.2 Life Cycle of Hepatitis C virus

The entire life cycle of HCV can be explained in four steps – (1) HCV intrusion into the host cell, (2) Translation of genome followed by polyprotein processing, (3) Replication of genome, and finally (4) Association of particles and releasing from the host cell. Having entered the liver, HCV gets attached to hepatocytes at the very beginning before initiating a complicated multistage entrance procedure into the cells (Gerold & Pietschmann, 2014). Though several factors have contributed to the host cell entry process, four of them are inevitable for providing success in the whole mechanism; Scavenger receptor class B of type 1 (SCARB1), Tetraspanin (CD81), and Tight junction molecules named Claudin-1 (CLDN1) and Occludin (OCLN) are those four, (Evans et al., 2007; Liu et al., 2009; Pileri et al., 1998; Ploss et al., 2009; Scarselli et al., 2002). After that, HCV is grasped inside through endocytosis which is clathrin-mediated and its envelope amalgamates with the membrane of premature endosomes in a condition of low pH. In the end, the capsid of HCV dissociates into the cytosol of a host cell to liberate the RNA genome and thus the virus gets invaded ultimately (Gerold & Pietschmann, 2014).

Following the invasion of HCV, host cell ribosomes translate the RNA genome at ER in the

cytoplasm producing a polyprotein bound with a membrane. This polyprotein acts as a precursor to give rise to ten characterized proteins as mentioned in chapter 1.2, as a result of reactions with viral as well as host cell proteases. Additionally, NS5B enlarges the genome using an RNA (minus stranded) (Gerold & Pietschmann, 2014).

In the next step, a membranous web gets constructed which is the compartment of the cytosol to let replication occur and instigated by the virus itself (Egger et al., 2002; Gosert et al., 2003; Paul et al., 2013; Romero-Brey et al., 2012). Inside the ER extracted membranous web which is induced by host factors, numerous non-structural viral proteins (NS3, NS4A, NS5A, etc.) in association with host factors integrate the replication complex of HCV (Quinkert et al., 2005).

At the termination stage of the HCV life cycle, budding virions are synthesized from four essential constituents, three structural proteins of HCV and genome or RNA which are produced in the previous replication process. Firstly, capsids of virions are developed along with RNA insertion; there is the contiguity of this new formation with lipid droplets and then the capsids bud into ER of the host cell to extract the envelope and dimers of E1E2 that will be implanted in the lipid bilayer. Later on, these virions get released into the host bloodstream making a complex shape with lipoproteins and this is because of the adjacent state of virions with the very low density lipoprotein(VLDL) composition route (Gerold & Pietschmann, 2014). To summarize the whole cycle, it is a sophisticated procedure based on versatile cellular receptors and factors and the ultimate success of this will make the virus an infectious agent.

Chapter 3

Methodology

3.1 General Overview

The purpose of this study is to construct a multiepitope-based peptide vaccine that has the capacity of stimulating broad-spectrum immunity containing conserved epitopes among different strains of pathogen for which MSA is preferred (Li et al., 2014).

HCV is well known for its massive genetic diversity accompanied by a natural tendency to escape the defense of the immune system (Phelps et al., 2021). Hence, to fulfill this challenging project of designing a vaccine against HCV, the beginning study was conducted to identify the predominant genotypes and subtypes of HCV over the world. As far as acknowledged by studies, seven distinguished genotypes, and 87 subtypes of HCV (confirmed 67 and others are provisional) exist out of which the most widespread one is genotype 1(G1) accounts for 49.1% of global cases, and later on comes genotype 3(G3), 4(G4) & 2(G2) responsible for 17.9%, 16.8% and 11.0 % respectively. More specifically, subtypes 1a,1b,3a & 2a are found worldwide and cause HCV infection in huge numbers to be considered an “epidemic” in nature (Petruzziello et al., 2016).

Integrating all these data, genotypes 1,2&4 were selected for this study since they are predominant, on the other hand genotype 3 was not considered for including in the MSA process due to its extensive protein sequence variation that requires separate inter-genotype MSA instead of combining with the other three genotypes to achieve an acceptable conserved sequence. To evaluate in a more specific manner, particular subtypes were kept in focus which were 1a,1b and 2a, based on HCV cases data from the entire world. All the other subtypes available like 4d, 4g, 4l, 4m, 4n, 4o, 4r and 4v in the used database were also in consideration along with selected strains under G2 as a source of E2 protein sequences at the

aim of the research. This step was important as it was unavoidable to ensure across genotype efficiency of the proposed vaccine. Targeting a single genotype or strain will provide protection only against that strain and will not be effective against the other strains because of carrying mutated epitopes by those strains. To sum up, strains of 12 genotypes and subtypes of HCV were picked to screen out the best antigenic E2 proteins which were undergone the MSA. Following MSA, the entire procedure has been illustrated in figure 1.

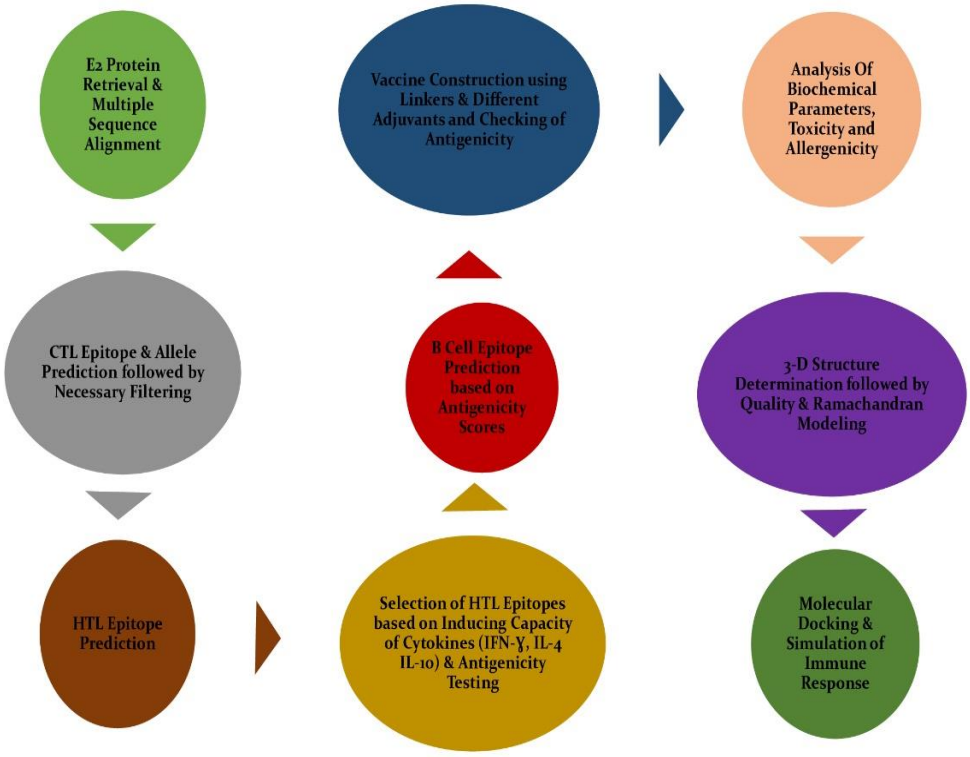

Figure 1: Schematic presentation of the whole procedure used for in-silico vaccine designing

3.2 Protein Sequence Retrieval

From the targeted 12 subtypes and genotypes, (1a,1b, 2, 2a, 4d, 4g, 4l, 4m, 4n, 4o, 4r, and 4v) 690 different E2 protein sequences were retrieved from ViPR database at https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjFlYrio_X3AhUq73MBHWLaAE8QFnoECAcQAQ&url=https%3A%2F%2Fwww.viprbrc.org%2F&usg=AOvVaw01beEA5V1fwiBWkTI-fNoa (Pickett et al., 2012) and

evaluated using Vaxijen v 2.0 server at https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiawLGFpPX3AhXW4XMBHYeNBaAQFnoECAQQAQ&url=http%3A%2F%2Fwww.ddg-pharmfac.net%2Fvaxijen%2F&usg=AOvVaw06bHtMWUBXXk_duNvLhcM5 (Doytchinova & Flower, 2007) where the threshold was kept as 0.5 and organism selected was "virus". The latter server results in 70–89 % prediction accuracy by applying the auto-cross covariance (ACC) method (Doytchinova & Flower, 2007). Among these 690 E2 protein sequences, some proteins were chosen based on antigenicity score since higher ones are essential to maintain an acceptable antigenic score while conserved sequences are made. During this preselection period, representation from all three genotypes was maintained so that result of MSA becomes validated to stand against diversified strains presenting the common epitopes.

3.3 Developing Conserved Sequences by Multiple Sequence Alignment and Testing of Antigenicity

In order to have conserved sequences, selected proteins mentioned were used in 3 different combinations and numbers. For doing multiple sequence alignment, the ClustalOmega tool at https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwic4961pfX3AhXf7HMBHTpoBsYQFnoECAYQAQ&url=https%3A%2F%2Fwww.genome.jp%2Ftools-bin%2Fclustalw&usg=AOvVaw2hnyIpytwhHAIUU8CzSLuf was used (Sievers & Higgins, 2018). This tool includes an algorithm named mBED that gets used for guide trees and computing distance matrix in terms of numerous motifs (Sievers & Higgins, 2018). Again, the Vaxijen v 2.0 server was utilized to check the antigenicity of the 3 sequences (Doytchinova & Flower, 2007). Additionally, the stability profile of the preferred sequence was evaluated using the ProtParam tool located at

https://web.expasy.org/protparam/ with a focus on comparing it with the final vaccine construct protein (Gasteiger et al., 2005).

3.4 Cytotoxic T Lymphocyte (CTL) Epitope Prediction and Evaluation

CTL epitopes are an important part of inducing the cellular immune response in a host body. Hence, CTL epitopes were predicted using NetCTL 1.2 server available at https://services.healthtech.dtu.dk/service.php?NetCTL-1.2 (Larsen et al., 2007). This server employs weight-matrix and artificial neural networks taking protein sequences in fasta format to identify CTL epitopes by calculating three attributes which are C-terminal cleavage, TAP transport efficiency, and MHC-I binding peptides (Larsen et al., 2007). Here, all the default parameters except one were kept the same like A1 supertype, 0.15 as weight on C terminal cleavage, 0.05 as weight on TAP transport efficiency, and 0.75 as epitope prediction threshold. The epitope sorting score was changed only to the combined score and the results were recorded accordingly.

For evaluation of these CTL epitopes, NetMHCpan 4.1 server found at http://www.cbs.dtu.dk/services/NetMHCpan-4.1 (Reynisson et al., 2020) was used to determine MHC-I binding alleles for strongly binding CTL epitopes. The web server provides results based on the Artificial Neural Network (ANN) which is built upon the combination of binding affinity (BA)and eluted ligand (EL) of mass spectroscopy of peptides (Reynisson et al., 2020). For this study, peptide length and species were chosen as 9-mer and HLA supertype representative respectively. All the alleles available at the server were taken for binding prediction and strong and weak binding thresholds were taken as given in the default which was 0.5 & 2 respectively and BA scores were also included to get on the result page.

Lastly, each CTL epitope screened out after the MHC-I binding assessment was tabulated

along with four properties – antigenicity, immunogenicity, allergenicity, and toxicity. Vaxijen v 2.0 (Doytchinova & Flower, 2007) with 0.5 threshold, IEDB MHC-I Immunogenicity tool at http://tools.iedb.org/immunogenicity/result/ (Calis et al., 2013), AllerTOP v 2.0 at https://www.ddg-pharmfac.net/AllerTOP/ (Dimitrov et al., 2013) and ToxinPred at https://webs.iiitd.edu.in/raghava/toxinpred/algo.php (Gupta et al., 2013) were used for the four mentioned properties respectively. For the current study, the ToxinPred server was used selecting SVM or Swiss-Prot based method and all other parameters as given in default settings.

3.5 Prediction and Evaluation of Helper T Lymphocyte (HTL) Epitope

HTL epitopes are at the center of adaptive immune system due to inducing role of cellular and humoral immune responses while HTL cells are bound to them. At first, NetMHCII pan 4.0 server at https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0 was used to predict all the MHC-II binding peptides (Reynisson et al., 2020). This server uses protein sequence in fasta format as input and carries similar methods as NetMHC pan 4.1 server for prediction purpose as mentioned in 3.4. In this study, peptide length and species were chosen 15 and HighQ-DRB respectively followed by strong and weak binding threshold as 2% and 10% as the lower the percentile rank, the higher the binding affinity. Sort by prediction score option was enabled too to get the scores in the output page.

Furthermore, each of the epitope that was screened out through NetMHCII pan 4.0 server, was assessed to achieve IFN-γ, IL-4 and IL-10 inducing profile. This assessment of sorting out effective HTL epitopes was done with the assistance of three servers named IFN epitope at https://webs.iiitd.edu.in/raghava/ifnepitope/help.php (Dhanda, Vir, et al., 2013), IL4pred at https://webs.iiitd.edu.in/raghava/il4pred/design.php (Dhanda, Gupta, et al., 2013) and IL-10 Pred available at https://webs.iiitd.edu.in/raghava/il10pred/ (Nagpal et al., 2017) to predict IFN-γ, IL-4 and IL-10 induction capacity respectively. While using IFN epitope server,

Motif and SVM hybrid and IFN-gamma versus Non IFN-gamma were enabled as the approach and model respectively for prediction. In terms of IL4pred and IL-10 Pred, thresholds were taken as given in default settings which was 0.2 and -0.3. To end up HTL epitope evaluation, antigenicity testing was also carried out similarly as CTL epitopes using the same server aiming to ensure desired efficacy of the selected epitopes.

3.6 Linear B Lymphocyte (LBL) Epitope Prediction and Evaluation

B cell epitopes are crucial for the induction of humoral immune response to produce specific antibody as well as memory cells. Prediction of LBL epitopes was done using B Cell Epitope Prediction Tool of IEDB Analysis Resource available at http://tools.iedb.org/main/bcell/ (Fleri et al., 2017) that uses BepiPred Linear Epitope Prediction 2.0 method. Finally, B cell epitopes were also assessed based on antigenic property again with Vaxijen v 2.0 (Doytchinova & Flower, 2007) server before inclusion in the vaccine construct.

3.7 Construction of Vaccine Sequence and Antigenicity Evaluation

Figure 2: Vaccine construction sequence using linkers

To construct the vaccine, a suitable adjuvant was used at the beginning as shown in figure 2 followed by CTL, HTL, and LBL epitopes which were merged by applying three different linkers EAAAK, GPGPG, and KK (Ahmad et al., 2021). EAAAK, GPGPG, and KK are the linkers implemented widely in in-silico vaccine design in order to add adjuvant and CTL epitopes, numerous HTL and LBL epitopes respectively. In this study, 5 different adjuvants which are TLR4 agonists like Beta(β)-Defensin (Ahmad et al., 2021), heat shock protein of 60 kDa (HSP60) (Lei et al., 2019), truncated Ov-Asp1 (Guo et al., 2015; Naz et al., 2020), Cholera toxin subunit B (CTB) (Mahmud et al., 2021; Rafi et al., 2022) and OmpA protein (Kathwate, 2022) were used as they will serve the purpose of enhancing immunogenicity (Ahmad et al., 2021). Once done with the sequence designing, again the antigenicity scores were derived from the Vaxijen v 2.0 server (Doytchinova & Flower, 2007) to validate the sequences primarily before proposing them as a potential candidate.

3.8 Biochemical Analysis of Vaccine Constructs

The physicochemical properties were examined for the 5 sequences through the ProtParam tool at https://web.expasy.org/protparam/ (Gasteiger et al., 2005) that provides information of various parameters. These properties aid in the primary and secondary analysis of structure (Nain et al., 2020).

3.9 Evaluation of Allergenicity and Toxicity of Vaccine Constructs

Ensuring safety for a vaccine candidate is inevitable which will prevent any kind of undesired effects while administered into the host body and to accomplish that allergenic and toxic profile determination are essential. AllergenOnline (Goodman et al., 2016) and T3DB (Wishart et al., 2015) are the two servers at http://www.allergenonline.org/ and http://www.t3db.ca/ respectively that were used to evaluate whether the prepared vaccine constructs exert any risk of allergenic cross-reactivity and toxic nature or act as a toxin.

AllergenOnline server includes criteria to categorize putative and bonafide allergens and proteins of unproven allergenicity (Goodman et al., 2016).T3DB contains around 136 toxic proteins of weight above 1500 Da based on which it provides results (Wishart et al., 2015). Both the servers used here use protein sequence in fasta format as input and all the default parameters remained the same for this study to get proper results.

3.10 Generation of 3D Model or Homology Modeling of the Vaccine Constructs

Prediction of the 3D model is beneficial for the current study through comparison with a model or template which is homologous to evaluate the tertiary structure of the constructed peptide vaccines. In view to fulfill this purpose, Phyre^2 v 2.0 web server was used which builds a 3D model incorporating remote and advanced methods of homology detection (Kelley et al., 2015). While using the server, protein sequences were given in plain format as input and normal mode for modeling was enabled.

3.11 Validation of Tertiary Structure of Vaccine Constructs using Z-Score and Ramachandran Plotting

The structure validation step is significant for the current study since the implementation of a valid 3D structure will bring about reliable results whenever docking and immune simulation studies will be done. To obtain validation of the retrieved tertiary structures in PDB format, Z-score analysis, and Ramachandran plotting were done through ProSAweb located at https://prosa.services.came.sbg.ac.at/prosa.php (Wiederstein & Sippl, 2007) and Structure Assessment tool of SWISS-MODEL at https://swissmodel.expasy.org/assess (Wiltgen & Hospital, 2018) respectively. Both servers receive PDB files obtained after doing homology modeling previously. ProSaweb displays the quality score and questionable structural parts of the given input protein using the context of the already known protein structures and viewer

of 3D molecule respectively (Wiederstein & Sippl, 2007). On the other hand, the visualization plot (Ramachandran plot) produced by Swiss PDB Viewer shows the distribution of backbone dihedral angles (Φ and Ψ) together statistically and also the favored and forbidden regions in terms of energy for the two dihedral angles (Wiltgen & Hospital, 2018).

3.12 Docking between TLR4 Receptor and Vaccine Peptide

Molecular docking is one of the key steps to have an assumption either the proposed vaccine peptide or protein will bind with the target receptor or not that specifically initiates innate immune responses which ultimately induce the adaptive response pathway. In this study, TLR4(Toll-Like Receptor 4) was selected (PDB ID: 4G8A) as the target receptor (Ahmad et al., 2021) with which vaccine protein binding was assessed, so the best-constructed vaccine peptide identified after completion of the steps up to chapter 3.11 was the ligand. For docking analysis, the PATCHDOCK web server (Schneidman-Duhovny et al., 2005) was utilized where the receptor was chosen TLR4 (PDB ID: 4G8A) and the retrieved PDB file of the chosen vaccine peptide was given as ligand input keeping all other parameters same as provided by default. The server recognizes the best shape complementarity of a molecule by figuring out the docking transformations. The PatchDock method performs structure prediction of protein–protein and protein–small molecule complexes (Schneidman-Duhovny et al., 2005).

3.13 Simulation of Immune Responses by In-silico Method

Immune simulation in-silico is the final computational step that needs to be accomplished to achieve a detailed observation of the response pattern in the host body after administration of the vaccine peptide in terms of enhancement of various immune cells and also their duration of staying in higher concentration. For this purpose, the C-IMMSIM server was applied at

https://kraken.iac.rm.cnr.it/C-IMMSIM/ (Rapin et al., 2010) where 300 simulation step was taken to implement three injection doses at 28 days or 4-week interval (Nain et al., 2020) for which time steps were provided 1, 84 and 168 (assuming 8 hours equal to 1-time step that results in 3-time steps per day). The finalized vaccine construct sequence was used as input in fasta format. C-IMMSIM is an immune simulator with a dynamic ability to generate immune system reactions like the real responses encountered by the host body with the combined help of machine learning techniques and position-specific scoring matrix (PSSM) (Rapin et al., 2010).

3.14 Reflection on Methodology

To summarize the entire method applied in the present study, it needs to be mentioned that multiple sequence alignment (MSA) was the initial notable step which was an obvious targeting of a virus with a huge genetic variety that signifies the derivation of conserved sequence from diverse viral strains or subtypes. In addition, for building and analyzing a peptide vaccine sequence, all other measures including MSA were based on computational approaches entirely that include several immunoinformatic tools and web servers. However, all these actions intending to design or propose a vaccine candidate are not sufficient to develop a safe and efficacious vaccine product. Therefore, this study aims only to come up with a prospective vaccine design that can be further investigated and will give an edge in developing an effective vaccine.

Chapter 4

Results

4.1 Pre-Multiple Sequence Alignment (MSA) Protein Screening

After a thorough evaluation of 690 E2 protein sequences, 11 sequences were screened out possessing best antigenicity score, and belonging to all 3 genotypes (G1, G2 & G4). In Table 1, these 11 proteins' NCBI accession no. retrieved from https://www.ncbi.nlm.nih.gov/protein along with their antigenicity score keeping threshold as 0.5 have been provided.

Table 1:Screened out proteins with the best antigenic score

NCBI Accession Number	Genotype/Subtype	Antigenicity Score
ACJ37238.1	1a	0.5809
ACE82319.1	1a	0.5735
ACE63600.1	1a	0.5701
ACD13335.1	1b	0.5714
ACE63628.1	1b	0.5700
ABV46066.2	1b	0.5709
AGV23511.1	2	0.5701
AQW44229.1	2	0.5865
AQW44228.1	2	0.5794
AEJ86546.1	2a	0.5334
AFN53807.1	4m	0.5479

For screening E2 protein strains presented in Table 1, antigenicity scores of 0.57 or above for 1a,1b, and 2 proteins were chosen whereas the best scoring strain was taken in the case of 2a. Lastly, one more protein containing the highest score was added from the 8 subtypes assessed under genotype 4 to confirm its representation also.

4.2 Conserved Sequence Retrieval by Multiple Sequence Alignment (MSA)

Altogether, 3 conserved sequences were established presented in Table 2. In all 3 combinations, protein sequences from genotype 1,2& 4 were included.

Table 2:All the relevant information of the conserved sequences

Conserved Sequence No.	Aligned Protein Strains	Genotype/Subtype of Aligned Protein	Total number of Proteins Aligned	Total Number of Amino Acid	Antigenicity Score (Threshold 0.4)
1	DN14, HCV-1a/US/BID-V464/2006, HCV-1a/CH/BID-V252/2002, HCV-1b/US/BID-V154/2001, HCV-1b/US/BID-V149/2003, HCV-1b/US/BID-V151/2002, PTR1256, sP170026, sP147057, WYHCV286 & HCV-4m/GB/BID-G1657	1a,1b,2,2a & 4m	11	195	0.4220

2	HCV-1a/US/BID-V464/2006, HCV-1a/CH/BID-V252/2002, HCV-1b/US/BID-V154/2001, HCV-1b/US/BID-V149/2003, HCV-1b/US/BID-V151/2002, PTR1256, WYHCV286 & HCV-4m/GB/BID-G1657.	1a,1b,2,2a, & 4m	8	201	0.4368
3	DN14, HCV-1a/US/BID-V464/2006, HCV-1b/US/BID-V154/2001, sP170026, sP147057 & HCV-4m/GB/BID-G1657.	1a,1b,2 & 4m	6	208	0.4563

Among the 3 in Table 2, the last sequence with the best antigenic score was chosen for further epitopes prediction. A relatively lower antigenic score should not be a demerit point for the chosen 3rd sequence as post MSA 0.4 is an acceptable score (Chauhan et al., 2018). In an intention to compare the primary sequence stability with the designed vaccine sequence one, the stability of this 3rd sequence was also evaluated which gave a result of 40.36 as instability index that indicates the sequence as unstable.

Selected Conserved Sequence:

TGGFGQLNSWHINTLCNSLTGALYNGCRCFGWGRPYCWYPQCVPAVCGPVYCFTPS PVVGTTDGPYTWGNTDVLNTPPRGWFGCTWMNGFTKCGPPCCPTDCFRKHPTYCG SGPWTPCLVYYRLWHYPCTVNFKRMYVGGEHRLACNRGCLRDRLSPLLSTTWLPCF LPAL TGLHLHQNIVDVQYYGWEVLFLLLADARCCLWMLQEA

4.3 Identification of Cytotoxic T Lymphocyte(CTL) Epitopes

Taking the 3rd conserved sequence as the primary protein, initially, 5 CTL epitopes were predicted (Table 3). Among them, 3 showed strong binding results with one allele of the MHC-I complex (Table 4). Then, all the epitopes were evaluated based on four criteria, all of which were fulfilled by only 1 CTL epitope (Table 5).

Table 3: CTL epitopes with relevant scores

Serial no.	CTL epitopes	Combined Scores	Sensitivity	Specificity
1	CTVNFKRMY	1.5366	0.54	0.993
2	LVYYRLWHY	1.1265	0.70	0.985
3	HQNIVDVQY	0.9507	0.74	0.980
4	VVGTTDGPY	0.9360	0.74	0.980
5	CNSLTGALY	0.8485	0.74	0.980

Table 4: Strong binding CTL epitopes with corresponding scores

Serial no.	Binding allele	Peptide(epitopes)	Core	% Rank EL	% Rank BA
1	HLA-A*26:01	CTVNFKRMY	CTVNFKRMY	0.495	0.247
2	HLA-A*26:01	LVYYRLWHY	LVYYRLWHY	0.204	0.205
3	HLA-B*15:01	HQNIVDVQY	HQNIVDVQY	0.010	0.262

Table 5: CTL epitopes with different filtering parameters

Epitope	**Antigenicity**	Immunogenicity	Allergenicity	Toxicity
CTVNFKRMY	**0.7164**	-0.14103	Allergen	Non-Toxin
LVYYRLWHY	**0.2012**	0.24088	Allergen	Non-Toxin
HQNIVDVQY	1.4915	0.16006	Non-Allergen	Non-Toxin

In Table 3, the sensitivity and specificity values are the translated ones for different ranges of integrated scores, and these were taken from the server NetCTL-1.2. With the increasing number of combined scores, specificity also enhances for the epitopes (Reynisson et al., 2020). In table 4, percentile rank EL (Eluted ligands) and BA (Binding affinity) scores are provided for the epitopes and both the scores were below 0.5. The epitope in Table 5 highlighted by blue color was integrated into the final vaccine construct.

4.4 Prediction of Helper T Lymphocyte (HTL) Epitopes

In this stage, 9 distinct HTL epitopes were identified as strong binder (Table 6) which were further tested on the basis of IFN-γ, IL-4 & IL-10 inducing ability (Table 7). At the latter stage, only 1 epitope satisfied all three criteria which were why this single epitope were also assessed for antigenicity.

Table 6: Strong binding HTL epitopes with binding allele and relevant scores

Binding Allele	HTL Epitope	Core	Percentile Rank EL
DRB1_0101	TGGFGQLNSWHINTL	FGQLNSWHI	0.52
	GGFGQLNSWHINTLC		1.67
DRB1_0102	TGGFGQLNSWHINTL		1.61
DRB1_0103	TGGFGQLNSWHINTL		0.60
DRB1_0701	TGGFGQLNSWHINTL		1.35
DRB1_0901	TGGFGQLNSWHINTL		1.57
DRB1_1601	TGGFGQLNSWHINTL		0.53
DRB1_0405	CTVNFKRMYVGGEHR	FKRMYVGGE	1.13
DRB1_0405	TVNFKRMYVGGEHRL		1.47
DRB1_0801	CTVNFKRMYVGGEHR		1.43
DRB1_0701	GPVYCFTPSPVVGTT	YCFTPSPVV	0.74
	CGPVYCFTPSPVVGT		0.81

	VCGPVYCFTPSPVVG		1.19
DRB4_0101	GLHLHQNIVDVQYYG	LHQNIVDVQ	0.64
DRB4_0101	TGLHLHQNIVDVQYY		1.10
DRB4_0103	GLHLHQNIVDVQYYG		0.64
	TGLHLHQNIVDVQYY		1.10

Table 7: IFN- γ, IL-4, and IL-10 stimulating profile of HTL epitopes

Peptide Sequence	IFN- γ Induction	IL-4 Induction	IL-10 Induction
TGGFGQLNSWHINTL	NEGATIVE	Non IL4 inducer	IL10 inducer
GGFGQLNSWHINTLC	NEGATIVE	Non IL4 inducer	IL10 inducer
CTVNFKRMYVGGEHR	NEGATIVE	IL4 inducer	IL10 inducer
TVNFKRMYVGGEHRL	POSITIVE	IL4 inducer	IL10 inducer
GPVYCFTPSPVVGTT	NEGATIVE	Non IL4 inducer	IL10 non-inducer
CGPVYCFTPSPVVGT	NEGATIVE	Non IL4 inducer	IL10 non-inducer
VCGPVYCFTPSPVVG	NEGATIVE	Non IL4 inducer	IL10 non-inducer
GLHLHQNIVDVQYYG	POSITIVE	IL4 inducer	IL10 non-inducer
TGLHLHQNIVDVQYY	POSITIVE	IL4 inducer	IL10 non-inducer

In Table 6, percentile rank EL is presented. In Table 7, the single epitope passed all three criteria, has been highlighted. Lastly, this single HTL epitope had an antigenicity score of 0.7292 which is >0.5 and for this reason, this was selected to incorporate in the vaccine

construct.

4.5 Linear B Cell Epitope Identification and Antigenicity Evaluation

Here, 6 B cell epitopes were identified (Table 8) of length greater than or equal to 6-mer, and each of them was evaluated in terms of antigenicity where 3 epitopes scored > 0.5 (Table 8) which was why got integrated into the vaccine construct.

Table 8: Selected B cell epitopes as highlighted with blue color

No.	Start	End	Peptide Sequence	Length	Antigenicity
1	5	11	GQLNSWH	7	0.0877
2	37	43	CWYPQCV	7	1.1288
3	59	79	VGTTDGPYTWGNTDVLNTPPR	21	0.4598
4	90	95	FTKCGP	6	-0.5034
5	102	113	CFRKHPTYCGSG	12	0.6179
6	145	165	ACNRGCLRDRLSPLLSTTWLP	21	0.8454

For the conserved sequence, the Average residue score was 0.474, and Minimum and Maximum values were 0.607 and 0.233 respectively.

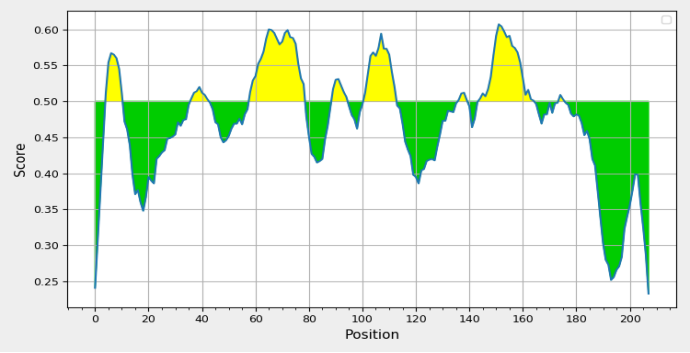

Figure 3: Score vs position graph obtained from the IEDB tool (Fleri et al., 2017).

According to the score vs position graph, (figure 3), yellow colored region denotes the peptides or LBLs scoring greater than 0.5, and green colored region is the representation of residues or particular amino acids those have a score below 0.5 and got discarded from being incorporated in the epitopes.

4.6 Designing and Screening of Final Vaccine Sequence

The final vaccine sequence was prepared by combining 1 CTL, 1 HTL, and 3 linear B cell (LBL) epitopes mentioned above sequentially using GPGPG and KK linkers. In the case of LBL epitopes, mapping was done following the higher to the lower order of antigenicity score (Nain et al., 2020). At the N-terminal, 5 distinct adjuvants were added with EAAAK linker to boost immunogenicity as well as efficiency of the proposed vaccine. Five separate adjuvants were incorporated to have 5 distinct vaccine constructs (1-5) in a view to draw a comparison in terms of 11 parameters as presented in Table 9.

Table 9: Comparison among 5 prepared vaccine constructs

Parameters	Vaccine 1 (Beta-Defensin)	Vaccine 2 (HSP60)	Vaccine 3(Truncated OV-ASP-1)	VACCINE 4(CTB)	Vaccine 5 (OmpA)
Antigenicity	0.6305	0.5510	0.5945	0.5873	0.7764
Physicochemical parameter (stability)	38.63	25.75	21.79	32.33	24.53
Physicochemical parameter (GRAVY value)	-0.682	-0.199	-0.701	-0.314	-0.295
Allergenicity	Identical with around 3 allergens	Identical with around 2 allergens	Identical with around 33 allergens	Identical with around 7 allergens	Identical with around 0 allergen
Toxicity	Non-toxin	Non-toxin	Non-toxin	Non-toxin	Non-toxin
3D modeling coverage & confidence	36% & 99.1%	81% % 100%	82% & 100 %	50% & 100%	47 % & 99.9%
Z-score	-4.75	-9.41	-4.49	-5.65	-2.53
Ramachandran plotting molprobity score	3.18	2.57	2.58	1.85	2.91

Clash score	80.86	44.16	63.71	22.57	52.66
Ramachandran favored region	74.42%	93.06%	95.60%	99.01%	81.57%
Ramachandran outliers	4.65%	2.31%	1.10%	0.00%	6.45%

From Table 9, it had been clear that all 5 candidates reached to satisfactory benchmark in case of antigenicity (>0.5), stability (<40), GRAVY value (negative), toxicity (non-toxin), and Z-score (score is plotted in the acceptable region in the context of known native structures). Vaccines 1 & 5 were discarded due to poor coverage of the 3D model (36 % & 47% with 99.9% confidence) along with unsatisfactory scores in Ramachandran favored and outlier regions which were lower and higher respectively for both vaccines. On the contrary, despite having good coverage of 82 % and 50% for vaccines 3 & 4 respectively and also Ramachandran favored regions of 95.6% and 99.01% which are up to the mark, being identical to 33 and 7 allergens were risky since any one of them could become allergenic in-vivo as potential numbers were excessive. Therefore, vaccine 2 with HSP60 adjuvant was finalized for proposing that had antigenicity score of 0.5510 and moderate scores in the other differentiator parameters discussed here.

Final Vaccine Sequence of 626 amino acids (when HSP60 is adjuvant):

MAKEIKFSDSARNLLFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSITKDGVSVA
KEIELSCPVANMGAQLVKEVASKTADAAGDGTTTATVLAYSIFKEGLRNITAGANPI
EVKRGMDKAAEAIINELKKASKKVGGKEEITQVATISANSDHNIGKLIADAMEKVGK
DGVITVEEAKGIEDELDVVEGMQFDRGYLSPYFVTNAEKMTAQLDNAYILLTDKKIS
SMKDILPLLEKTMKEGKPLLIIAEDIEGEALTTLVVNKLRGVLNIAAVKAPGFGDRRK

EMLKDIAILTGGQVISEELGLSLENAEVEFLGKAGRIVIDKDNTTIVDGKGHSHDVKD
RVAQIKTQIASTTSDYDKEKLQERLAKLSGGVAVIKVGAASEVEMKEKKDRVDDAL
SATKAAVEEGIVIGGGAALIRAAQKVHLNLHDDEKVGYEIIMRAIKAPLAQIAINAGY
DGGVVVNEVEKHEGHFGFNASNGKYVDMFKEGIIDPLKVERIALQNAVSVSSLLLTT
EATVHEIKEEKAAPAMPDMGGMGGMGGMGGMMEAAAKHQNIVDVQYGPGPGTV
NFKRMYVGGEHRLKKCWYPQCVKKACNRGCLRDRLSPLLSTTWLPKKCFRKHPTY
CGSG

4.7 Assessment of Biochemical Properties of the Preferred Vaccine Peptide

All the parameters computed are tabulated in Table 10.

Table 10: Physicochemical properties of proposed vaccine candidate

Biochemical Properties	Results
Total Number (amino acid)	626
Weight of Molecule	67.32768 kDa
Theoretical pI	7.18
Formula	$C_{2954}H_{4854}N_{822}O_{909}S_{28}$
Total atom number	9567
Estimated half-life	30 hours (mammalian reticulocytes, in vitro). >20 hours (yeast, in vivo). >10 hours (Escherichia coli, in vivo).
Instability index	25.75 (stable)
Aliphatic index	92.41
Grand average of hydropathicity (GRAVY)	-0.199

In Table 10, molecular weight is shown around 67.327 kDa which is acceptable compared to two other HCV peptide vaccine molecular weights, 29191.87g/mol (1 g/mol = 1kDa) and 69.2 kDa (Khalid & Ashfaq, 2020) (Ahmad et al., 2021). Likewise, theoretical pI (7.18) tends to basic category slightly. Half-life was achieved as 30 hours in vitro and greater than 20 & 10 hours in vivo. As in mammalian cells 30 hours $t_{1/2}$ is very high to ensure the long-lasting vaccine candidate activity in the human host because humans fall under the Mammalia class of animals too, the score was considered as a good one. The stability profile of the vaccine categorizes it as stable (below 40) whereas the primary target sequence was unstable, hence it was commendable as an unstable state turned out to be stable through specified epitopes and adjuvant integration. Finally, a higher aliphatic index value indicates that the vaccine candidate is thermostable, and a negative GRAVY value indicates the molecule is hydrophilic that will interact with water molecules which is satisfactory since hydrophobicity may induce toxicity in the host body.

4.8 Toxicity Evaluation of Preferred Vaccine Peptide

The preferred constructed peptide did not exert any toxic characteristics as shown in figure 18 (pointed with red box). The server contains data on toxins where any toxic substance input will exhibit results but input for this study returned no result.

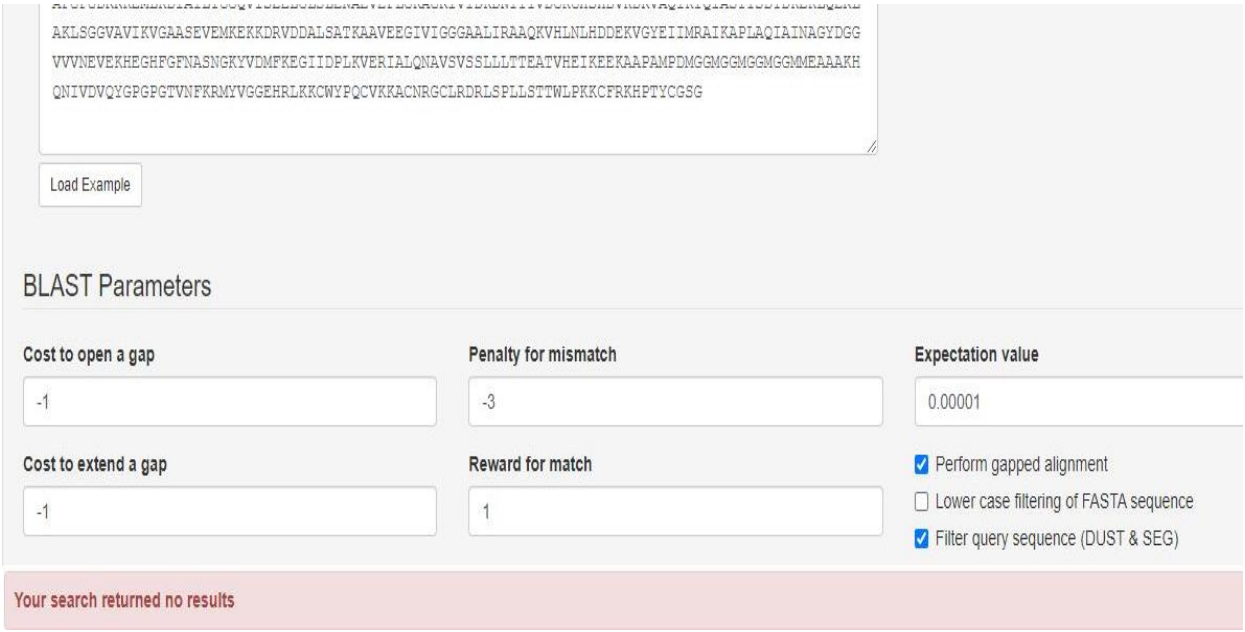

Figure 4:Output webpage from the T3DB server (Wishart et al., 2015) for Vaccine 2

4.9 Allergenicity Assessment of Preferred Vaccine Peptide

Two best-scoring known allergenic substances were found and presented in Table 11 along with the similarity percentage. They are identical by 50% to the vaccine molecule showing that it can be allergenic to fewer extent which is still moderate and not uncontrollable but must be kept under consideration. The reason for claiming the result as probably safe in the human body is an in vitro experimentation of IgE binding with a protein is not sufficient to prove that the same protein will be allergic in vivo. In vivo allergenicity depends on the involvement of multiple sites of IgE (Goodman et al., 2016).

Table 11: The two substances along with their similarity percentages

NCBI Accession No.	Name of the Substance	Similarity percentage (threshold 0.5)
AAG44478	Vacuolar serine protease [Penicillium oxalicum]	0.533
AAD25995	Alkaline serine protease Pen c2	0.529

4.10 Homology Modeling of the Vaccine Construct

The 3D model (figure 5) was achieved based on the single highest scoring template which covered 81% with 100% confidence (figure 6) with HSP60 adjuvant. A total of 507 residues were successfully included in building the model which is quite a competent score.

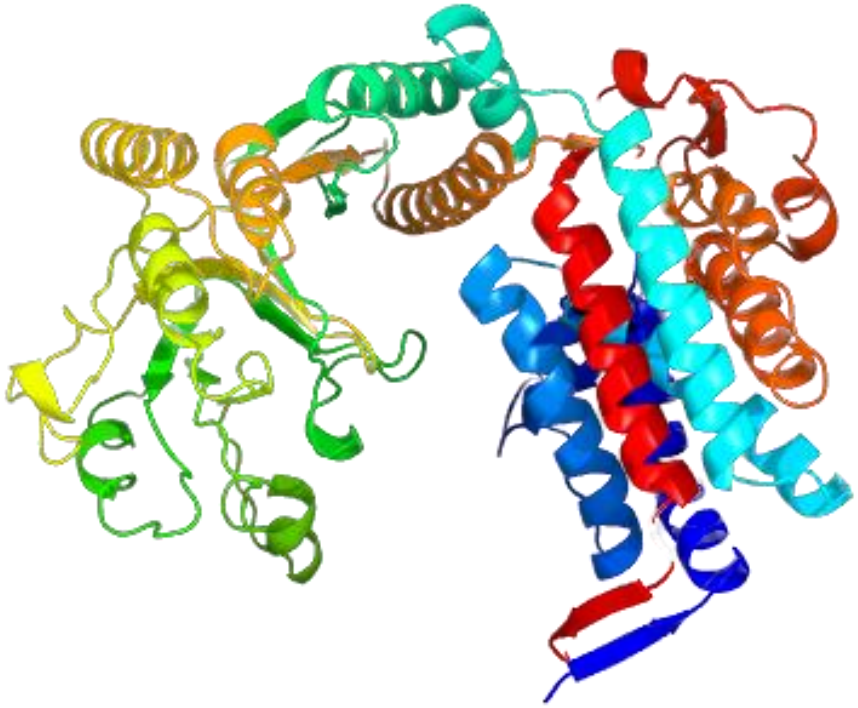

Figure 5:3D model of the finalized vaccine construct from Phyre2 server (Kelley et al., 2015).

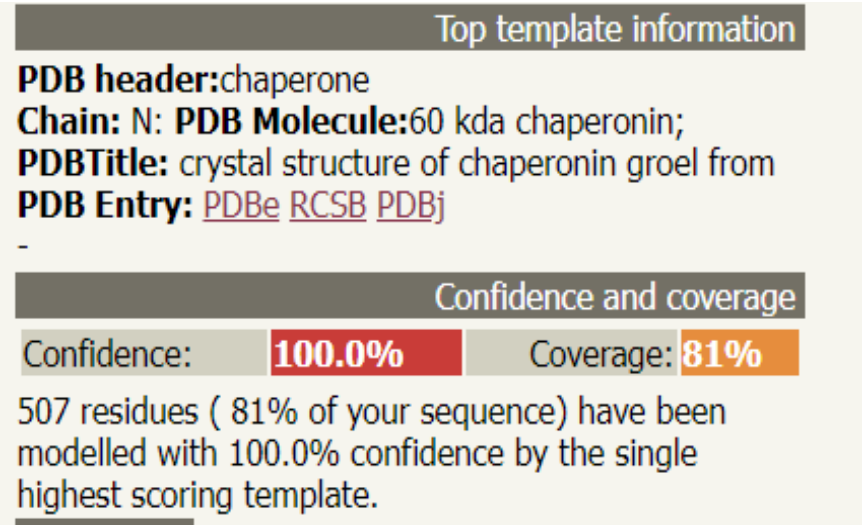

Figure 6:Homology modeling score for the finalized vaccine from the Phyre2 server (Kelley et al., 2015).

4.11 Tertiary Structure Validation by Z-score Analysis

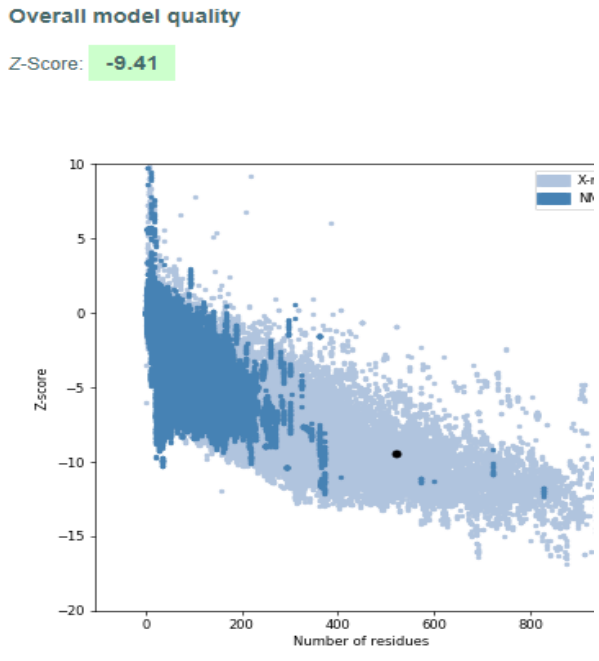

Figure 7:Overall model quality result of the finalized vaccine from ProsaWeb server (Wiederstein & Sippl, 2007).

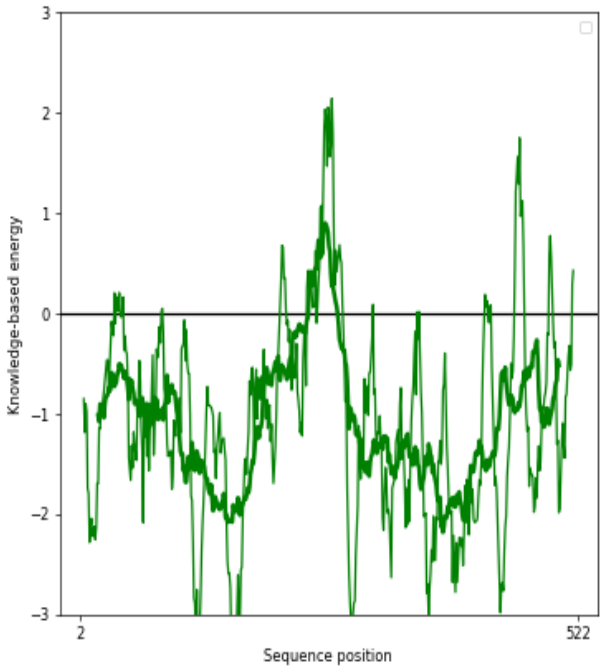

Figure 8:Knowledge based energy vs sequence position graph (Wiederstein & Sippl, 2007).

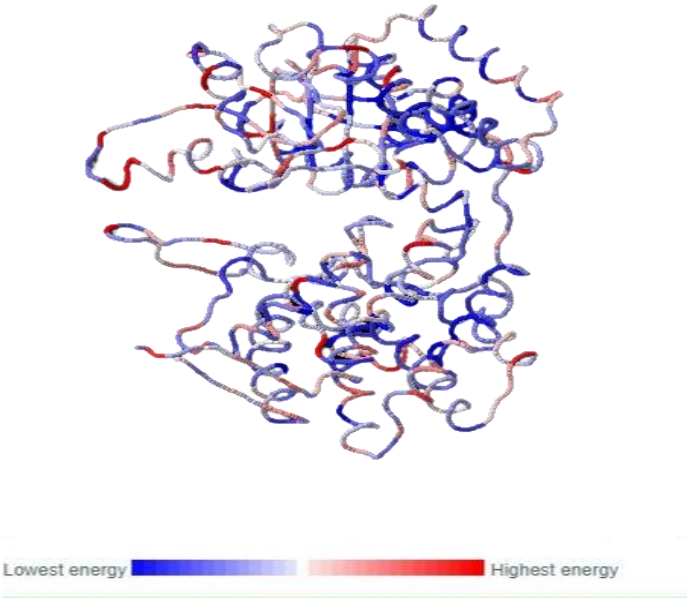

Figure 9: Schematic representation of lowest and highest energy residues (Wiederstein & Sippl, 2007).

In figure 7, the Z score was found -9.41 and the location of the score is close to the middle 9X Ray region) in the context of the native protein structure range. Again, from the knowledge-based energy vs sequence (figure 8) position graph, very few residues from both N and C terminals, and some more residues at the middle of the two terminals possess positive energy exhibiting the peaks, that take all the other or majority residues in negative energy region below the baseline. In figure 9 also, the consistency of the vaccine protein with typical proteins is evident with a greater blue colored region of the C^{α} model than the red colored region which represents the residues of highest energies. Therefore, the 3D structure prediction in chapter 4.10 is validated according to the analysis of the Z score (Wiederstein & Sippl, 2007).

4.12 Structure Validation through Ramachandran Plotting

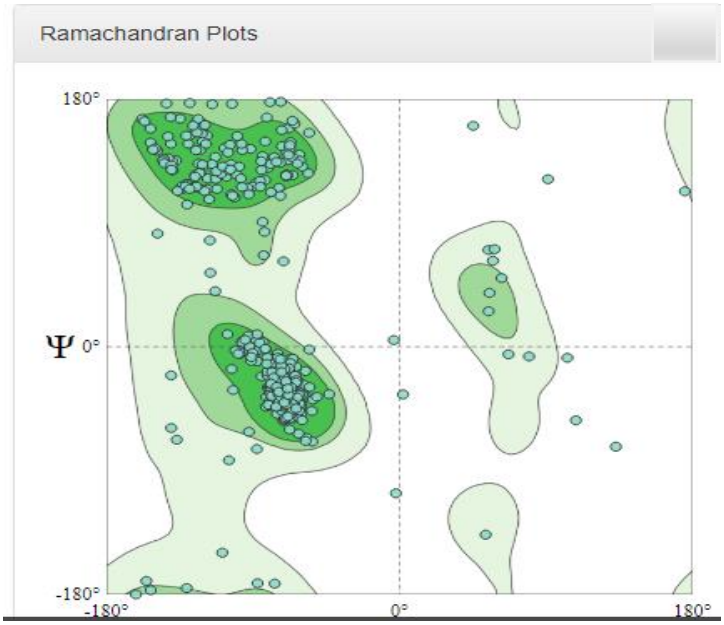

Figure 10:Ramachandran plotting of the proposed vaccine from SWISS PDB Plotter (Wiltgen & Hospital, 2018)

Table 12:Molprobity scores with involved amino acids

Parameter	Score	Involved Amino Acid
MolProbity Score	2.57	-
Clash Score	44.16	(218 TYR- 245 PRO), (431 ASP- 434 LYS), (129 GLU- 424 LYS), (432 ASP- 434 LYS), (23 HIS- 93 VAL), (431 ASP- 432 ASP), (226 ILE- 250 ALA), (281 GLY- 284 ARG), (473 PHE- 484 ASP), (144 THR- 162 ALA), (306 LEU- 307 SER), (434 LYS- 437 TYR), (224 LYS- 248 ILE), (359 TYR- 363 LYS), (430 HIS- 434 LYS), (101 GLU- 104 ARG), (430 HIS- 433 GLU), (222 THR- 248 ILE), (264 ASN- 265 LYS), (432 ASP- 433 GLU), (132

		LYS- 424 LYS), (432 ASP- 435 VAL), (125 ALA- 428 ASN), (337 HIS- 339 HIS), (360 ASP- 361 LYS), (241 LYS- 242 GLU), (429 LEU- 430 HIS), (3 LYS- 520 GLU), (125 ALA- 424 LYS), (465 GLU- 467 GLU), (20 ARG- 96 TYR), (422 ALA- 444 ILE), (178 GLU- 389 LYS), (285 LYS- 289 LYS), (202 TYR- 259 THR), (253 ILE- 256 GLU), (215 ASP- 321 ARG), (472 HIS- 484 ASP), (17 GLU- 21 GLN), (219 ILE- 317 GLY), (495 LYS- 499 ILE), (198 TYR- 198 TYR), (283 ARG- 359 TYR), (165 MET- 169 GLY)
Ramachandran Favored	93.06%	-
Ramachandran Outliers	2.31%	245 PRO, 254 GLU, 242 GLU, 265 LYS, 269 VAL, 249 ILE, 238 LYS, 432 ASP, 307 SER, 361 LYS, 433 GLU, 472 HIS
Rotamer Outliers	0.24%	229 MET
C-Beta Deviations	0	-
Bad Bonds	27 / 3954	426 HIS, 469 HIS, 430 HIS, 339 HIS, 23 HIS, 472 HIS, 155 HIS, 337 HIS, 519 HIS, 432 ASP, 465 GLU- 466 VAL, 253 ILE- 254 GLU, 278 PRO, 46 PRO, 112 PRO, 64 PRO, 447 PRO, 360 ASP- 361 LYS
Bad Angles	21 / 5321	(300 ILE- 301 SER), (237 GLU- 238 LYS), (257 ALA- 258 LEU), (465 GLU- 466 VAL), (431 ASP- 432 ASP), (247 LEU- 248 ILE), (353 ALA- 354 SER), (252 ASP- 253

		ILE), (360 ASP- 361 LYS), (244 LYS- 245 PRO), 430 HIS, 339 HIS, 426 HIS, 155 HIS, 472 HIS, 23 HIS, 519 HIS, 337 HIS, 469 HIS, (313 VAL- 314 GLU)

As visible in figure 10 and mentioned in Table 12, Ramachandran favored region was 93.06% and the outlier region was 2.31% (12 residues), both of which were sufficient for validation purposes. Therefore, within this stage, selection of vaccine 2 from Table 9 gets justified and accordingly the last two experiments were done only with this vaccine 2.

4.13 Molecular Docking of Vaccine and Receptor Protein

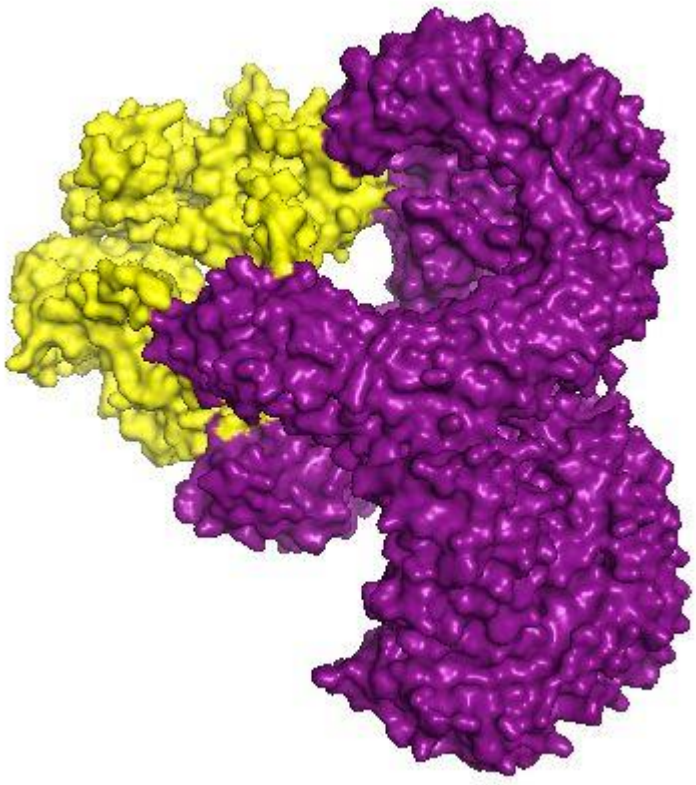

Figure 11: Ligand (vaccine molecule in yellow color) and receptor (TLR4 in purple color) docking

Here, figure 11 represents the binding pattern of the solution complex of vaccine and TLR4 which was primarily retrieved from PatchDock server before analysis.

Molecular Docking Algorithm Based on Shape Complementarity Principles

Solution No	Score	Area	ACE	Transformation	PDB file of the complex
1	18240	2669.00	441.57	-1.54 -0.01 1.97 41.01 19.81 22.00	result.1.pdb
2	17990	2399.10	444.86	0.91 -0.25 -1.93 -8.62 58.54 -145.55	result.2.pdb
3	16330	2345.50	22.05	-2.16 -0.04 1.76 -11.44 12.63 59.11	result.3.pdb
4	16288	2281.70	467.76	-0.02 -0.03 -0.26 -53.74 -17.41 -76.43	result.4.pdb
5	16198	2087.30	264.00	2.80 -0.27 2.21 -52.12 -35.47 9.37	result.5.pdb
6	16164	2422.30	-209.78	-2.01 0.60 -1.03 -12.50 83.79 46.43	result.6.pdb
7	16134	2418.70	207.82	0.13 -0.17 0.04 -35.78 -32.93 -77.31	result.7.pdb
8	15862	3895.20	323.57	3.10 0.56 -0.09 -21.66 92.48 -3.66	result.8.pdb
9	15794	2107.70	439.47	0.34 -0.24 0.13 -39.61 -27.02 -82.50	result.9.pdb
10	15776	2210.60	263.71	0.33 -0.10 0.19 -27.97 -29.86 -88.98	result.10.pdb
11	15740	2549.70	438.88	2.55 0.94 2.41 -60.68 -81.69 -17.29	result.11.pdb
12	15568	2317.40	-195.34	-2.39 -0.05 1.64 -32.38 15.33 51.95	result.12.pdb
13	15552	2774.70	83.14	-1.68 0.17 1.16 0.01 -53.44 91.97	result.13.pdb
14	15486	2474.10	314.19	3.07 0.10 -2.65 -3.16 -27.17 -31.73	result.14.pdb
15	15460	3261.70	401.83	1.01 -0.04 -1.92 -7.43 46.37 -154.44	result.15.pdb
16	15428	2035.20	489.94	-0.28 0.19 1.44 43.08 -33.04 -112.40	result.16.pdb
17	15370	2014.90	494.62	-1.70 0.45 2.90 38.17 41.84 -9.19	result.17.pdb
18	15304	2457.30	459.43	1.28 0.37 2.07 -52.14 -38.44 -144.34	result.18.pdb
19	15182	2215.10	461.79	-2.94 -0.86 -1.90 70.56 -57.33 79.50	result.19.pdb
20	15150	2573.30	222.42	-2.55 -0.15 2.29 -16.57 -8.19 4.61	result.20.pdb

Figure 12:PatchDock output page (Schneidman-Duhovny et al., 2005) of the 20 best scoring solutions in the PDB file along with four distinct results of corresponding solutions.

From figure 12, the results of the top scoring solution (no. 1) are 18240 and 2669.00 Å² as geometric score and interface size of receptor or area respectively. The area score expresses the strength of the binding of two proteins. Desolvation energy is achieved as 441.57 kJ/mole and transformation scores in the rigid state are -1.54, -0.01, 1.97, 41.01, 19.81, and 22.00 corresponding to the six dimensions of space for the sake of transformation.

4.14 Immune Simulation

From simulation studies, significant graphical images were found depicting immune cells number separately including all types and classes as well as antigen counting to enable the ideation that what can be the reactions of the host body in case of stimulating adaptive and innate immunity. In the C-IMMSIM server (Rapin et al., 2010), the prediction state for the preferred vaccine construct was normal with 119 elapsed time.

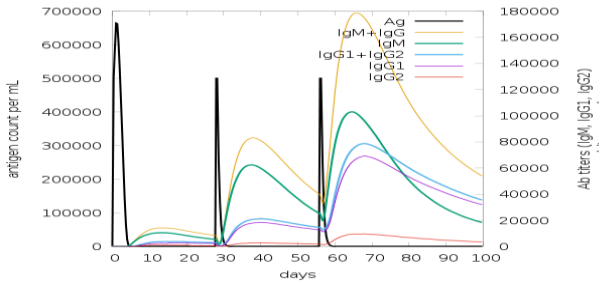

Figure 13:Cell count of Antigen (in per mL) and antibodies in numerous subtypes (in an arbitrary scale on the right of the y-axis) vs days graph

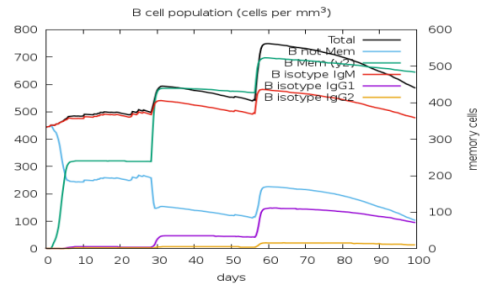

Figure 14:Graph of B cells per mm3 showing the total number, memory cell count, and subdivided immunoglobulins as the day progresses after vaccine administration

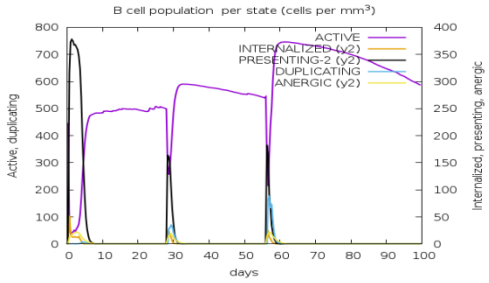

Figure 15:Graph of 5 entity states of B lymphocytes named active, duplicating, presenting, internalized, and anergic

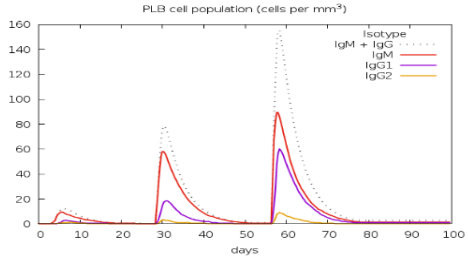

Figure 16:Plasma B cells (IgM, IgG1& IgG2) vs days graph following vaccine injection

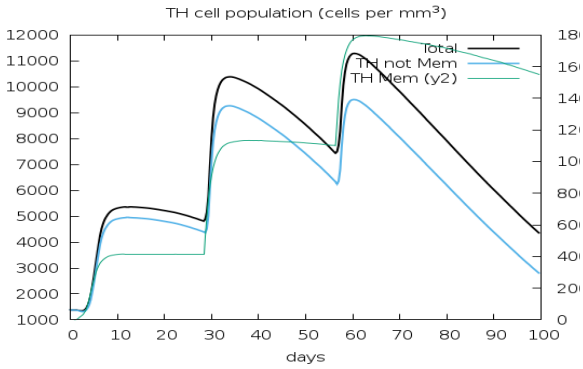

Figure 17:Graph of Helper T cell population (per mm3) in terms of total and memory cell count based on the count of days

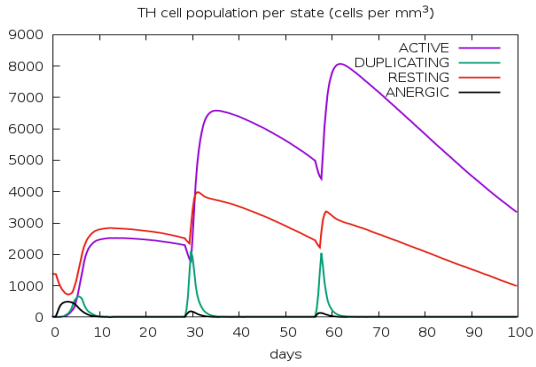

Figure 18:Graphical representation of activated, duplicating, resting, and anergic CD4+ T cells with days proceeding (post-vaccination)

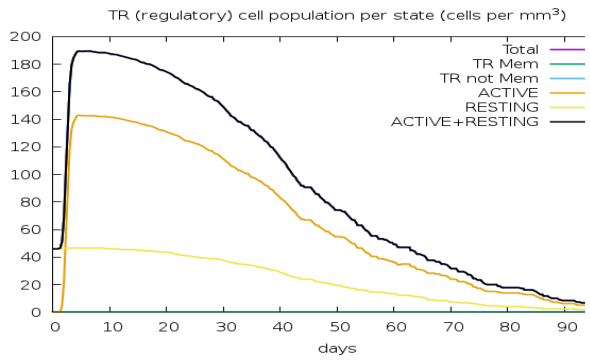

Figure 19:Graph presenting regulatory total, memory, active and resting helper T cell population after injecting vaccine

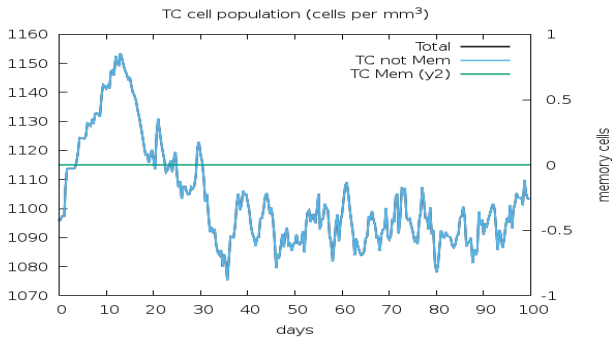

Figure 20:Graph of Cytotoxic T cell count (per mm3) in terms of total and memory cells post vaccine injection

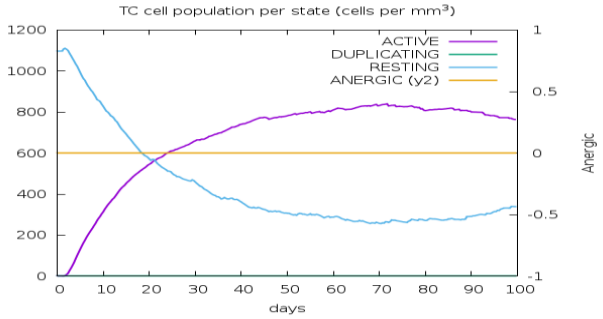

Figure 21:Active, duplicating, resting and anergic CD8+ cells population putting count of days on the x-axis

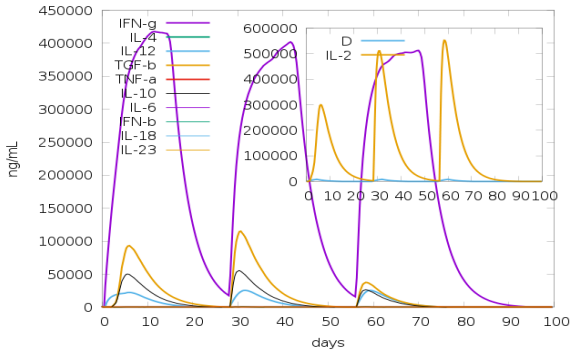

Figure 22:Graph plotting different interleukins

and cytokines in concentration (ng/mL) along with an additional inset box pointing danger signal (D).

Having analyzed all the graphs, it was quite visible that, humoral immune responses had induced in an amount that can be announced as up to the mark. The statement got justified since IgM, IgG1, and IgG2, all the observed antibodies gradually reached the highest

concentration number around day 70 which is close to the best possible score on the arbitrary scale (figure 13) and three clear peaks were easily distinguishable in between day 10 and 20, day 30 and 40 and day 60 and 70 following three vaccine doses which was successful consequences as desired. Moreover, antigen concentration was raised thrice too but got reduced in the latter two times which indicated activation of immune reactions (figure 13). Then, the number of B cells followed the same pattern in correspondence with immunoglobulins in figure 14, that enhanced confidence for expecting suitable humoral immunity in the host body. The number of memory B cells was the highest among all and IgG2 was the lowest but each of them increased with time whereas not memory B cells decreased hugely while the total cell population boosted the most around day 60. Again, active and duplicating B cells showed remarkable growth similarly while anergic cell number reduced and always remained lesser compared to the first two types mentioned (figure 15). Ultimately, plasma B cells were also kept aligned in the number of antibodies and B cells (figure 16) among which plasma cells are the bridges in the differentiation process and lead to releasing antibodies for fighting against the vaccine antigen specifically.

Next, T cell volume was monitored in silico which was not entirely appropriate as coveted, however, it was still at a level that can be useful for ensuring prevention for the host body. In the case of helper T cells, they elevated in three steps again on days 10, 30, and 60 in figure 17, where memory cells were more than not memory cells that fulfill the purpose of vaccination. In figure 18, though resting HTL cell proportion stayed above the duplicating cell proportion which can be defended by the exacerbated number of active HTL cells in total and the very poor number of anergic cells. Active regulatory CD4 cells were also higher in load than resting ones showing a top peak on day 5 approximately and even after continuous deduction they were in helpful size till day 56 when the last dose was given (figure 19). Figure 20, exhibited a good portion of not memory cells but memory cells were at level zero

that maybe create a ground of cytotoxic actions only from CTL cells. To end, active cytotoxic T cells were accelerated where resting cells maintained the complete opposite pathway and were lowest in number during highest load of active cells on day 70 (figure 21). Since anergic cell count was at zero, despite no amplification of duplicating cell number, required stimulation for CTL cells was confirmed instead of suboptimal induction.

Finally, IFN gamma, IL-4, IL-10, and IL-12 all were raised in concentration three times again around on day 10, between 30 and 50, and day between 60 and 80 where they all attained the highest peak after the very first injection. (figure 22). The inset box also indicates increasing of IL-2 following the same trend (figure 22) and thus the vaccine construct was assured as a suitable one for the proposition.

Chapter 5

Discussion

The threat of hepatitis C virus (HCV) worldwide can be understood with the 1% infection rate (Blach et al., 2017) and in the USA particularly death number is already higher compared to deaths caused by human immunodeficiency virus or HIV (Ly et al., 2014). The situation becomes worst with its economic impact as it has turned out to be the principal disease that is increasing liver transplantation cases due to the impairment it causes to the liver in chronic form(Joshi et al., 2014; Mitchell & Gurakar, 2015). Thereby, the demand for a preventive vaccine for HCV has become very high which is leading scientists to propose a vaccine utilizing all dynamics possible to explore. The current study also aims to present a potential vaccine candidate exploiting in-silico approaches that have the capacity of digging into different dimensions to result in a comprehensive study and save valuable time initially. To prepare the vaccine, envelope glycoprotein (E2) of HCV was chosen considering its carried region which is dominant in terms of immunity induction that will be useful as a whole along with the humoral wing emphasizing (Moradpour & Penin, 2013).

For designing a peptide vaccine consisting of multiple epitopes, suitable prediction tools and servers were in use that ultimately gave 1 CTL, 1 HTL, and 3 B cell epitopes after all levels of screening. In the study, the reason for the extraction of less number of epitopes was conducting multiple sequence alignment (MSA) that ultimately ensures the efficiency of the vaccine peptide, especially for the pathogens which mutate frequently. For HCV, the extent of genetic diversity is massive against which an important weapon was MSA that provided a preserved protein sequence common across genotypes, and using that shortened sequence gave the epitopes only that would be expected in terms of invasion of all the genotypes went through MSA. This step was unavoidable in the context of proposing a vaccine against such a complex virus which makes the task of defeating it very challenging. In the case of a vaccine

proposition, efficiency and safety are the two prime attributes that must be complied and besides MSA, robust screening of CTL, HTL, and B cell epitopes was also very crucial. Combined with MSA, the filtering steps were the reason which was why fewer epitopes were predicted but it was possible to assure that all of them are safe and effective and the purpose of the peptide vaccine was also accomplished by excluding any unnecessary and harmful epitopes (Li et al., 2014). Moreover, the precision of genotype 1,2 and 4 and their subsequent subtypes for protein evaluation were done due to their prevalence.

Following the epitope predictions, the essential step was to identify the most suitable adjuvant which is a toll-like receptor 4 (TLR4) agonist as well as instigates the immune reactions to take place in an amplified manner. To fulfill these requirements 5 different adjuvants were investigated based on antigenicity, physicochemical properties, allergenicity, toxicity, homology modeling, and validation of the model using Z-score and Ramachandran plotting and molprobity outcomes. Among all, heat shock protein of 60kDa (HSP60) which is a chaperone protein showed a balanced result which was acceptable and thus the results were further analyzed in the latter subchapters under chapter 4. The allergenicity test gave two possibilities for it to become allergenic which needs to be evaluated in-vivo extensively before confirmation based on which necessary steps can be undertaken which is beyond the zone of computational approaches. On the other hand, docking and simulation results were well analyzed to propose this construct with HSP60 for future studies.

Chapter 6

Conclusion

The research was accomplished to propose an effective vaccine candidate against the hepatitis C virus for which till now, no vaccine is available or marketed. The candidate exhibited characteristics of being non-toxic, stable, hydrophilic, and possessing moderate molecular weight and adequate half-life as tested in mammals. Moreover, the predicted 3D model for the peptide had a high coverage compared to the template, hence the model was validated by Z-score and molprobity results of Ramachandran plotting and both of them showed satisfactory outcomes. Docking complex and immune simulation were also carried out which gave a strong ground for proposing the candidate. Additionally, future scopes must be discussed which are, carrying out population coverage and molecular dynamic simulation studies and also along with E2 protein, another structural protein like core protein can be added. Molecular dynamic simulation will give an additional edge to analyzing and improving the docking complex. Finally, core protein (Haller et al., 2007) can be helpful in terms of resulting in a compact innate and adaptive immune stimulation in the host body with a synergic effect being an added antigen. However, the boundary of this study was exploiting the vaccine peptide only from an immunoinformatic perspective, but to achieve more clarification on safety and efficacy, in-vitro and in-vivo experiments in a detailed manner are inevitable.

References

Ahmad, S., Shahid, F., Tahir Ul Qamar, M., Ur Rehman, H., Abbasi, S. W., Sajjad, W., Ismail, S., Alrumaihi, F., Allemailem, K. S., Almatroudi, A., & Saeed, H. F. U. (2021). Immuno-informatics analysis of pakistan-based hcv subtype-3a for chimeric polypeptide vaccine design. *Vaccines*, *9*(3), https://doi.org/10.3390/vaccines9030293

Andrei, G. (2021). Vaccines and Antivirals: Grand Challenges and Great Opportunities. *Frontiers in Virology*, *0*, 8. https://doi.org/10.3389/FVIRO.2021.666548

Attahullah, & Khan, Nasar & Wahab, Abdul & Tabassum, Anum & Umar, Muhammad & Basit, Abdul & Ahmad, Iqbal & Rahim, Kashif & Mahmood, Asif & Saleha, S. (2014). Molecular Biology of Hepatitis C Virus: An Overview. *Journal of Bio-Molecular Sciences, 2(2).* https://www.researchgate.net /publication /283298328_Molecular_Biology_of_Hepatitis_C_Virus_An_Overview

Augestad, E. H., Bukh, J., & Prentoe, J. (2021). Hepatitis C virus envelope protein dynamics and the link to hypervariable region 1. *Current Opinion in Virology*, *50*. https://doi.org/10.1016/J.COVIRO.2021.07.006

Bartenschlager, R., Baumert, T. F., Bukh, J., Houghton, M., Lemon, S. M., Lindenbach, B. D., Lohmann, V., Moradpour, D., Pietschmann, T., Rice, C. M., Thimme, R., & Wakita, T. (2018). Critical challenges and emerging opportunities in hepatitis C virus research in an era of potent antiviral therapy: Considerations for scientists and funding agencies. *Virus Research*, *248*. https://doi.org/10.1016/J.VIRUSRES.2018.02.016

Blach, S., Zeuzem, S., Manns, M., Altraif, I., Duberg, A. S., Muljono, D. H., Waked, I., Alavian, S. M., Lee, M. H., Negro, F., Abaalkhail, F., Abdou, A., Abdulla, M., Abou Rached, A., Aho, I., Akarca, U., Al Ghazzawi, I., Al Kaabi, S., Al Lawati, F., … Razavi, H. (2017). Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. *The Lancet. Gastroenterology & Hepatology*, *2*(3). https://doi.org/10.1016/S2468-1253(16)30181-9

Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., Keşmir, C., & Peters, B. (2013). Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLOS Computational Biology*, *9*(10). https://doi.org/10.1371/JOURNAL.PCBI.1003266

Chauhan, V., Singh, M. P., & Ratho, R. K. (2018). Identification of T cell and B cell epitopes against Indian HCV-genotype-3a for vaccine development- An in silico analysis. *Biologicals*, *53*(609). https://doi.org/10.1016/j.biologicals.2018.02.003

Chen, Q., Ayer, T., Bethea, E., Kanwal, F., Wang, X., Roberts, M., Zhuo, Y., Fagiuoli, S., Petersen, J., & Chhatwal, J. (2019). Changes in hepatitis C burden and treatment trends in Europe during the era of direct-acting antivirals: a modelling study. *BMJ Open*, *9*(6). https://doi.org/10.1136/BMJOPEN-2018-026726

Coppola, N., Minichini, C., Starace, M., Sagnelli, C., & Sagnelli, E. (2016). Clinical impact of the hepatitis C virus mutations in the era of directly acting antivirals. *Journal of Medical Virology*, *88*(10). https://doi.org/10.1002/JMV.24527

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*, *2013*. https://doi.org/10.1155/2013/263952

Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, *8*(1). https://doi.org/10.1186/1745-6150-8-30/TABLES/9

Di Bisceglie, A. M., Order, S. E., Klein, J. L., Waggoner, J. G., Sjogren, M. H., Kuo, G., Houghton, M., Choo, Q. L., & Hoofnagle, J. H. (1991). The role of chronic viral hepatitis in hepatocellular carcinoma in the United States - PubMed. *The American Journal of Gastroenterology, 86(3).* https://pubmed. ncbi. nlm.nih.gov/1847790/

Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). AllerTOP - a server for in silico prediction of allergens. *BMC Bioinformatics*, *14*(SUPPL6). https://doi.org/10.1186/1471-2105-14-S6-S4/FIGURES/4

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*(1). https://doi.org/10.1186/1471-2105-8-4/TABLES/2

Egger, D., Wölk, B., Gosert, R., Bianchi, L., Blum, H. E., Moradpour, D., & Bienz, K. (2002). Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *Journal of Virology*, *76*(12). https://doi.org/10.1128/JVI.76.12.5974-5984.2002

Evans, M. J., Von Hahn, T., Tscherne, D. M., Syder, A. J., Panis, M., Wölk, B., Hatziioannou, T., McKeating, J. A., Bieniasz, P. D., & Rice, C. M. (2007). Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature 2006 446:7137*, *446*(7137). https://doi.org/10.1038/nature05654

Fattovich, G., Giustina, G., Degos, F., Tremolada, F., Diodati, G., Almasio, P., Nevens, F., Solinas, A., Mura, D., Brouwer, J. T., Thomas, H., Njapoum, C., Casarin, C., Bonetti, P., Fuschi, P., Basho, J., Tocco, A., Bhalla, A., Galassini, R., … Realdi, G. (1997). Morbidity and mortality in compensated cirrhosis type C: a retrospective follow-up study of 384 patients. *Gastroenterology*, *112*(2). https://doi.org/10.1053/GAST.1997.V112.PM9024300

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Frontiers in Immunology*, *8*(MAR). https://doi.org/10.3389/FIMMU.2017.00278/BIBTEX

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*. https://doi.org/10.1385/1592598900

Gerold, G., & Pietschmann, T. (2014). The HCV Life Cycle: In vitro Tissue Culture Systems and Therapeutic Targets. *Digestive Diseases*, *32*(5). https://doi.org/10.1159/000360830

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., Ree, R. Van, Vieths, S., Baumert, J. L., Bohle, B., Lalithambika, S., Wise, J., & Taylor, S. L. (2016). AllergenOnline : A peer-reviewed , curated allergen cross-reactivity. *Molecular Nutrition Food Research, 2016*. https://doi.org/10.1002/mnfr.201500769

Gosert, R., Egger, D., Lohmann, V., Bartenschlager, R., Blum, H. E., Bienz, K., & Moradpour, D. (2003). Identification of the Hepatitis C Virus RNA Replication Complex in Huh-7 Cells Harboring Subgenomic Replicons. *Journal of Virology*, *77*(9). https://doi.org/10.1128/JVI.77.9.5487-5492.2003/ASSET/16A75B31-ACF1-468E-BFB6-9775BBFD3AB8/ASSETS/GRAPHIC/JV0932151005.JPEG

Guo, J., Yang, Y., Xiao, W., Sun, W., Yu, H., Du, L., Lustigman, S., Jiang, S., Kou, Z., & Zhou, Y. (2015). A truncated fragment of Ov-ASP-1 consisting of the core pathogenesis-related-1 (PR-1) domain maintains adjuvanticity as the full-length protein. *Vaccine*, *33*(16). https://doi.org/10.1016/J.VACCINE.2015.02.053

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In silico approach for predicting toxicity of peptides and proteins. *PloS One*, *8*(9). https://doi.org/10.1371/JOURNAL.PONE.0073957

Haller, A. A., Lauer, G. M., King, T. H., Kemmler, C., Fiolkoski, V., Lu, Y., Bellgrau, D., Rodell, T. C., Apelian, D., Franzusoff, A., & Duke, R. C. (2007). Whole recombinant yeast-based immunotherapy induces potent T cell responses targeting HCV NS3 and Core proteins. *Vaccine*, *25*(8). https://doi.org/10.1016/J.VACCINE.2006.10.035

Hutin, Y., Kitler, M. E., Dore, G. J., Perz, J. F., Armstrong, G. L., Dusheiko, G., Ishibashi, H., Grob, P., Kew, M., Marcellin, P., Seeff, L. B., Beutels, P., Nelson, C., Stein, C., Zurn, P., Clifford, G., Vranckx, R., Alberti, A., Hallaj, Z. S., … Lavanchy, D. (2004). Global burden of disease (GBD) for hepatitis C. *Journal of Clinical Pharmacology*, *44*(1). https://doi.org/10.1177/0091270003258669

Joshi, D., Pinzani, M., Carey, I., & Agarwal, K. (2014). Recurrent HCV after liver transplantation-mechanisms, assessment and therapy. *Nature Reviews. Gastroenterology & Hepatology*, *11*(12). https://doi.org/10.1038/NRGASTRO.2014.114

Kathwate, G. H. (2022). In silico Design and Characterization of Multi-epitopes Vaccine for SARS-CoV2 from Its Spike Protein. *International Journal of Peptide Research and Therapeutics*, *28*(1). https://doi.org/10.1007/S10989-021-10348-Z/FIGURES/7

Keck, Z. Y., Pierce, B. G., Lau, P., Lu, J., Wang, Y., Underwood, A., Bull, R. A., Prentoe, J., Velázquez-Moctezuma, R., Walker, M. R., Luciani, F., Guest, J. D., Fauvelle, C., Baumert, T. F., Bukh, J., Lloyd, A. R., & Foung, S. K. H. (2019). Broadly neutralizing antibodies from an individual that naturally cleared multiple hepatitis C virus infections uncover molecular determinants for E2 targeting and vaccine design. *PLoS Pathogens*, *15*(5). https://doi.org/10.1371/JOURNAL.PPAT.1007772

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols 2015 10:6*, *10*(6). https://doi.org/10.1038/nprot.2015.053

Khalid, H., & Ashfaq, U. A. (2020). Exploring HCV genome to construct multi-epitope based subunit vaccine to battle HCV infection : Immunoinformatics based approach. *Journal of Biomedical Informatics*,

108(June). https://doi.org/10.1016/j.jbi.2020.103498

Kinchen, V. J., Massaccesi, G., Flyak, A. I., Mankowski, M. C., Colbert, M. D., Osburn, W. O., Ray, S. C., Cox, A. L., Crowe, J. E., & Bailey, J. R. (2019). Plasma deconvolution identifies broadly neutralizing antibodies associated with hepatitis C virus clearance. *The Journal of Clinical Investigation*, *129*(11). https://doi.org/10.1172/JCI130720

Kiyosawa, K., Sodeyama, T., Tanaka, E., Gibo, Y., Yoshizawa, K., Nakano, Y., Furuta, S., Akahane, Y., Nishioka, K., Purcell, R. H., & Alter, H. J. (1990). Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. *Hepatology (Baltimore, Md.)*, *12*(4 Pt 1). https://doi.org/10.1002/HEP.1840120409

Krey, T., D'Alayer, J., Kikuti, C. M., Saulnier, A., Damier-Piolle, L., Petitpas, I., Johansson, D. X., Tawar, R. G., Baron, B., Robert, B., England, P., Persson, M. A. A., Martin, A., & Rey, F. A. (2010). The Disulfide Bonds in Glycoprotein E2 of Hepatitis C Virus Reveal the Tertiary Organization of the Molecule. *PLOS Pathogens*, *6*(2). https://doi.org/10.1371/JOURNAL.PPAT.1000762

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*, 424. https://doi.org/10.1186/1471-2105-8-424

Lei, Y., Zhao, F., Shao, J., Li, Y., Li, S., Chang, H., & Zhang, Y. (2019). Application of built-in adjuvants for epitope-based vaccines. *PeerJ*, *2019*(1). https://doi.org/10.7717/PEERJ.6185/FIG-7

Li, W., Joshi, M. D., Singhania, S., Ramsey, K. H., & Murthy, A. K. (2014). Peptide Vaccine: Progress and Challenges. *Vaccines 2014*, *2*(3). https://doi.org/10.3390/VACCINES2030515

Liu, S., Yang, W., Shen, L., Turner, J. R., Coyne, C. B., & Wang, T. (2009). Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. *Journal of Virology*, *83*(4). https://doi.org/10.1128/JVI.01888-08

Ly, K. N., Xing, J., Klevens, R. M., Jiles, R. B., & Holmberg, S. D. (2014). Causes of death and characteristics of decedents with viral hepatitis, United States, 2010. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *58*(1). https://doi.org/10.1093/CID/CIT642

Mahmud, S., Rafi, M. O., Paul, G. K., Promi, M. M., Shimu, M. S. S., Biswas, S., Emran, T. Bin, Dhama, K., Alyami, S. A., Moni, M. A., & Saleh, M. A. (2021). Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach. *Scientific Reports 2021*, *11*(1). https://doi.org/10.1038/s41598-021-92176-1

McCaffrey, K., Gouklani, H., Boo, I., Poumbourios, P., & Drummer, H. E. (2011). The variable regions of hepatitis C virus glycoprotein E2 have an essential structural role in glycoprotein assembly and virion infectivity. *The Journal of General Virology*, *92*(Pt 1). https://doi.org/10.1099/VIR.0.026385-0

Mitchell, O., & Gurakar, A. (2015). Management of Hepatitis C Post-liver Transplantation: a Comprehensive Review. *Journal of Clinical and Translational Hepatology*, *3*(2). https://doi.org/10.14218/JCTH.2015.00005

Moradpour, D., & Penin, F. (2013). Hepatitis C virus proteins: from structure to function. *Current Topics in Microbiology and Immunology*, *369*. https://doi.org/10.1007/978-3-642-27340-7_5

Nagpal, G., Usmani, S. S., Dhanda, S. K., Kaur, H., Singh, S., Sharma, M., & Raghava, G. P. S. (2017). Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific Reports*, *7*. https://doi.org/10.1038/SREP42851

Nain, Z., Minnatul, M., Kumer, M., & Kumar, U. (2020). Structural basis and designing of peptide vaccine using PE-PGRS family protein of Mycobacterium ulcerans — An integrated vaccinomics approach. *Molecular Immunology*, *120*(February). https://doi.org/10.1016/j.molimm.2020.02.009

Naz, A., Shahid, F., Butt, T. T., Awan, F. M., Ali, A., & Malik, A. (2020). Designing Multi-Epitope Vaccines to Combat Emerging Coronavirus Disease 2019 (COVID-19) by Employing Immuno-Informatics Approach. *Frontiers in Immunology*, *11*(July). https://doi.org/10.3389/fimmu.2020.01663

Oli, A. N., Obialor, W. O., Ifeanyichukwu, M. O., Odimegwu, D. C., Okoyeh, J. N., Emechebe, G. O., Adejumo, S. A., & Ibeanu, G. C. (2020). Immunoinformatics and Vaccine Development: An Overview . *ImmunoTargets and Therapy*, *9*. https://doi.org/10.2147/ITT.S241064

Osburn, W. O., Snider, A. E., Wells, B. L., Latanich, R., Bailey, J. R., Thomas, D. L., Cox, A. L., & Ray, S. C. (2014). Clearance of hepatitis C infection is associated with the early appearance of broad neutralizing

antibody responses. *Hepatology (Baltimore, Md.)*, *59*(6). https://doi.org/10.1002/HEP.27013

Page, K., Melia, M. T., Veenhuis, R. T., Winter, M., Rousseau, K. E., Massaccesi, G., Osburn, W. O., Forman, M., Thomas, E., Thornton, K., Wagner, K., Vassilev, V., Lin, L., Lum, P. J., Giudice, L. C., Stein, E., Asher, A., Chang, S., Gorman, R., … Cox, A. L. (2021). Randomized Trial of a Vaccine Regimen to Prevent Chronic HCV Infection. *The New England Journal of Medicine*, *384*(6). https://doi.org/10.1056/NEJMOA2023345

Paul, D., Hoppe, S., Saher, G., Krijnse-Locker, J., & Bartenschlager, R. (2013). Morphological and biochemical characterization of the membranous hepatitis C virus replication compartment. *Journal of Virology*, *87*(19). https://doi.org/10.1128/JVI.01370-13

Pestka, J. M., Zeisel, M. B., Bläser, E., Schürmann, P., Bartosch, B., Cosset, F. L., Patel, A. H., Meisel, H., Baumert, J., Viazov, S., Rispeter, K., Blum, H. E., Roggendorf, M., & Baumert, T. F. (2007). Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(14). https://doi.org/10.1073/PNAS.0607026104

Petruzziello, A., Marigliano, S., Loquercio, G., Cozzolino, A., & Cacciapuoti, C. (2016). Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. *World Journal of Gastroenterology*, *22*(34). https://doi.org/10.3748/WJG.V22.I34.7824

Phelps, C. C., Walker, C. M., & Honegger, J. R. (2021). Where to Next? Research Directions after the First Hepatitis C Vaccine Efficacy Trial. *Viruses 2021*, *13*(7). 1351. https://doi.org/10.3390/V13071351

Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., Zhou, L., Larson, C. N., Dietrich, J., Klem, E. B., & Scheuermann, R. H. (2012). ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research*, *40*(D1). https://doi.org/10.1093/NAR/GKR859

Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G., & Abrignani, S. (1998). Binding of Hepatitis C Virus to CD81. *Science*, *282*(5390). https://doi.org/10.1126/SCIENCE.282.5390.938

Ploss, A., Evans, M. J., Gaysinskaya, V. A., Panis, M., You, H., De Jong, Y. P., & Rice, C. M. (2009). Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature*, *457*(7231). https://doi.org/10.1038/NATURE07684

Puig, M., Mihalik, K., Tilton, J. C., Williams, O., Merchlinsky, M., Connors, M., Feinstone, S. M., & Major, M. E. (2006). CD4+ immune escape and subsequent T-cell failure following chimpanzee immunization against hepatitis C virus. *Hepatology (Baltimore, Md.)*, *44*(3). https://doi.org/10.1002/HEP.21319

Quinkert, D., Bartenschlager, R., & Lohmann, V. (2005). Quantitative analysis of the hepatitis C virus replication complex. *Journal of Virology*, *79*(21). https://doi.org/10.1128/JVI.79.21.13594-13605.2005

Rafi, M. O., Al-Khafaji, K., Sarker, M. T., Taskin-Tok, T., Rana, A. S., & Rahman, M. S. (2022). Design of a multi-epitope vaccine against SARS-CoV-2: immunoinformatic and computational methods. *RSC Advances*, *12*(7). https://doi.org/10.1039/D1RA06532G

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System. *PLoS ONE*, *5*(4). https://doi.org/10.1371/JOURNAL.PONE.0009862

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*(W1). https://doi.org/10.1093/NAR/GKAA379

Romero-Brey, I., Merz, A., Chiramel, A., Lee, J. Y., Chlanda, P., Haselman, U., Santarella-Mellwig, R., Habermann, A., Hoppe, S., Kallis, S., Walther, P., Antony, C., Krijnse-Locker, J., & Bartenschlager, R. (2012). Three-Dimensional Architecture and Biogenesis of Membrane Structures Associated with Hepatitis C Virus Replication. *PLOS Pathogens*, *8*(12). https://doi.org/10.1371/JOURNAL.PPAT.1003056

Santolini, E., Migliaccio, G., & Monica, N. La. (1994). Biosynthesis and biochemical properties of the hepatitis C virus core protein. *Journal of Virology*, *68*(6). https://doi.org/10.1128/JVI.68.6.3631-3641.1994

Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., & Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate

receptor for the hepatitis C virus. *The EMBO Journal*, *21*(19). https://doi.org/10.1093/EMBOJ/CDF529

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*(suppl_2). https://doi.org/10.1093/NAR/GKI481

Seeff, L. B., Buskell-Bales, Z., Wright, E. C., Durako, S. J., Alter, H. J., Iber, F. L., Hollinger, F. B., Gitnick, G., Knodell, R. G., Perrillo, R. P., Stevens, C. E., & Hollingsworth, C. G. (1992). Long-term mortality after transfusion-associated non-A, non-B hepatitis. The National Heart, Lung, and Blood Institute Study Group. *The New England Journal of Medicine*, *327*(27). https://doi.org/10.1056/NEJM199212313272703

Sesardic, D. (1993). Synthetic peptide vaccines. *Journal of Medical Microbiology*, *39*(4). https://doi.org/10.1099/00222615-39-4-241

Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science : A Publication of the Protein Society*, *27*(1). https://doi.org/10.1002/PRO.3290

Suwanthawornkul, T., Anothaisintawee, T., Sobhonslidsuk, A., Thakkinstian, A., & Teerawattananon, Y. (2015). Efficacy of Second Generation Direct-Acting Antiviral Agents for Treatment Naïve Hepatitis C Genotype 1: A Systematic Review and Network Meta-Analysis. *PloS One*, *10*(12). https://doi.org/10.1371/JOURNAL.PONE.0145953

Tomar, N., & De, R. K. (2014). Immunoinformatics: a brief review. *Methods in Molecular Biology (Clifton, N.J.)*, *1184*. https://doi.org/10.1007/978-1-4939-1115-8_3

Torresi, J., Johnson, D., & Wedemeyer, H. (2011). Progress in the development of preventive and therapeutic vaccines for hepatitis C virus. *Journal of Hepatology*, *54*(6). https://doi.org/10.1016/J.JHEP.2010.09.040

WHO. (2017). *Vaccination and trust: How concerns arise and the role of communication in mitigating crises*. https://www.who.int/publications/i/item/vaccination-and-trust

WHO. (2021). *WHO releases first-ever global guidance for country validation of viral hepatitis B and C elimination.* https://www.who.int/news/item/25-06-2021-who-releases-first-ever-global-guidance-for-country-validation-of-viral-hepatitis-b-and-c-elimination

WHO. (2022). *Hepatitis C, fact sheet*. https://www.who.int/news-room/fact-sheets/detail/hepatitis-c

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*(SUPPL.2). https://doi.org/10.1093/nar/gkm290

Wiltgen, M., & Hospital, G. G. (2018). Algorithms for Structure Comparison and Analysis : Homology Modelling of Proteins. *Encyclopedia of Bioinformatics and Computational Biology*, 1. https://doi.org/10.1016/B978-0-12-809633-8.20484-6

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S. A., & Rappaport, S. M. (2015). T3DB: the toxic exposome database. *Nucleic Acids Research*, *43*(D1). https://doi.org/10.1093/NAR/GKU1004

Yu, C. I., & Chiang, B. L. (2010). A new insight into hepatitis C vaccine development. *Journal of Biomedicine & Biotechnology*, *2010*. https://doi.org/10.1155/2010/548280

Zeisel, M. B., Fofana, I., Fafi-Kremer, S., & Baumert, T. F. (2011). Hepatitis C virus entry into hepatocytes: Molecular mechanisms and targets for antiviral therapies. *Journal of Hepatology*, *54*(3). https://doi.org/10.1016/J.JHEP.2010.10.014