Virulent Strain of *Streptococcus pyogenes*: Unveiling Vaccine Candidates through *In-Silico* Exploration

Submitted By

Shanza Tariq

(56-FOC/MSBI/F22)

Supervised by

Ms. Tahira Noor

Lecturer

Department of Bioinformatics

Faculty of Computing and Information Technology

International Islamic University, Islamabad

2025

Virulent Strain of *Streptococcus pyogenes*: Unveiling Vaccine Candidates through *In-Silico* Exploration

MS Research Thesis

Submitted By

Shanza Tariq

(56-FOC/MSBI/F22)

Supervised by

Ms. Tahira Noor

Lecturer

Final year project report submitted to the Department of Bioinformatics as a part of course of studies of Master's degree in Bioinformatics of the International Islamic University, Islamabad.

Department of Bioinformatics

Faculty of Computing and Information Technology

International Islamic University, Islamabad

2025

Department of Bioinformatics

Faculty of Computing and Information Technology

International Islamic University Islamabad

Dated: 23-12-2025

FINAL APPROVAL

It is certificate that we have read the thesis "Virulent Strain of *Streptococcus pyogenes*: Unveiling Vaccine Candidates through *In-Silico* Exploration" submitted by **Ms. Shanza Tariq, 56-FOC/MSBI/F22.** It is our judgment that this project is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for the **Master Degree in Bioinformatics.**

COMMITTEE

External Examiner:
Dr. Abdul Rauf Siddiqi
Associate Professor
Department of Biological Sciences
COMSATS, Islamabad ______________________

Internal Examiner:
Dr. Mehrosh Khalid
Lecturer, Department of Bioinformatics
Faculty of Computing & Information Technology ______________________
International Islamic University Islamabad

Supervisor:
Ms. Tahira Noor
Lecturer, Department of Bioinformatics
Faculty of Computing & Information Technology ______________________
International Islamic University Islamabad

In-Charge:
Dr. Attiya Kanwal
Assistant Professor, Department of Bioinformatics
Faculty of Computing and Information Technology (FCIT), IIUI ______________________

Chairperson/Dean:
Prof. Dr. Hosam Hasan M Alhakami
Faculty of Computing and Information Technology
International Islamic University, Islamabad

____

DISSERTATION

A dissertation submitted to Department of Bioinformatics, Faculty of Computing & Information Technology, International Islamic University Islamabad as a partial fulfillment of the requirements for the award of the degree of Master in Bioinformatics (MSBI).

DEDICATION

I dedicated this research to ALLAH Almighty, thanking him for his guidance, strength, mental power, protection, and abilities, as well as for providing me with a healthy existence. This research is dedicated to my loving parents, who have always been a source of inspiration and courage when I wanted to give up, and who continue to provide ethical, spiritual, emotional, and financial support. I would also like to dedicate this research to my friends who give me suggestions and inspired me to complete the work. I want to pay special thanks to my supervisor Miss Tahira Noor who not only helped me throughout the course of this time to solve every query but also gave me moral support and guidance to fulfill this task.

DECLARATION

I hereby declare that the work present in the thesis "Virulent Strain of *Streptococcus pyogenes*: Unveiling Vaccine Candidates through *In-Silico* Exploration" is my own effort, except where otherwise acknowledged and that the thesis is my own composition. No part of the thesis has been previously presented for any other degree.

Date: ______________________________

Shanza Tariq
(56-FOC/MSBI/F22)

ACKNOWLEDGEMENTS

I would like to express gratitude to the Head of Department and to all faculty members of Bioinformatics for providing, with all the essential research resources. I am deeply grateful to my supervisor, Ms. Tahira Noor, for her continuous support, insightful guidance, patience, and valuable suggestions that greatly contributed to the success of this work. I am very grateful to my friends for their assistance in finishing this research. Finally, I want to acknowledge everyone who has helped me in this study, either directly or indirectly. I am very thankful to all the concerned people who cooperated with me in this regard.

PROJECT IN BRIEF

Project Title:	Virulent Strain of *Streptococcus pyogenes*: Unveiling Vaccine Candidates through *In-Silico* Exploration
Organization:	Department of Bioinformatics Faculty of computing and Information Technology International Islamic university H-10, Islamabad
Undertaken By:	Shanza Tariq 56-FOC/MSBI/F22
Supervised By:	Ms. Tahira Noor
Start Date:	2023
Completion Date:	2025
Objective:	This study aims to design MEBSV against *S. pyogenes* using *in-silico* approaches by identifying conserved, immunogenic antigens and predicting both humoral and cell-mediated immune responses with the inclusion of an appropriate adjuvant.
Tools & Technologies:	IEDB, I-TASSER, PROCHECK, C-ImmSim, Vector Builder, SnapGene, HDOCK
Documentation Tool:	MSWord, MS PowerPoint
Operation System:	Windows 10 Pro
System Used:	Haier core m3

Table of Contents

ACKNOWLEDGEMENTS i
PROJECT IN BRIEF ii
List of Abbreviations v
List of Figures vi
List of Tables vii
ABSTRACT viii
Chapter 1 1
Introduction 1
1 Introduction 2
1.1 ***Streptococcus Pyogenes*** 2
1.2 Epidemiological Background and Clinical Symptoms of *S. pyogenes* 2
1.2.1 Prevalence and Global Disease Burden of *S. pyogenes* 3
1.2.2 Mortality Rate of *S. pyogenes* 3
1.2.3 Current Treatments and Limitations 3
1.2.4 Transmission 4
1.2 Virulent proteins of *S. pyogenes* 6
1.3 Vaccine Design 8
1.4 Overview of Research Approach 10
1.5 Problem Statement 10
1.6 Motivation 11
1.7 Aim and Objectives 11
1.8 Research Questions 11
1.9 Proposed Solution 12
1.10 Scope and Limitations of the Study 12
Chapter 2 13
Literature Review 13
2.1 Introduction to Literature Review on *S. pyogenes* 14
2.2 Pathogenesis and molecular mimicry of *S. pyogenes* 14
2.3 Research Gap of *S. pyogenes*. 14
2.4 Critical Analysis of *S. pyogenes* 15
2.5 Limitations, Gaps and Future Directions in Current *S. pyogenes* Literature 15
2.6 Clinical Characteristics and advancements of *S. pyogenes* 16
Chapter 3 19
Materials and Methods 19
3.0 Methodology and Research Design 20
3.1 Vaccine Target Selection 22
3.2 B-cell and T-cell Epitope Prediction 23
3.3 Population Coverage Analysis 23
3.4 Construction of MEBSV 23

3.5 Structure Prediction, Refinement and Validation of MEBSV 24
3.6 Docking 24
3.6.1 Energy Minimization and Interaction Improvement 25
3.7 Immune Simulation 25
3.8 Expression analysis, Codon optimization and Molecular cloning 25
Chapter 4 27
Results 27
4.1 Vaccine Target Selection 28
4.2 B-cell and T-cell Epitope Prediction 32
4.3 Population Coverage Analysis 36
4.4 Construction of MEBSV 39
4.5 Structure Prediction, Refinement and Validation of MEBSV 41
4.6 Docking 47
4.6.1 Energy Minimization and Interaction Improvement 50
4.7 Immune Simulation 54
4.8 Codon optimization and Molecular cloning 58
Chapter 5 61
Discussion 61
5.1 Discussion 62
Chapter 6 65
Conclusion 65
6.1 Conclusion 66
Chapter 7 68
References 68

List of Abbreviations

Acronym	Abbreviation
BLAST	Basic local alignment search tool
UniProtKB	UniProt Knowledgebase
S. pyogenes	*Streptococcus pyogenes*
ARF	Acute rheumatic fever
RHD	Rheumatic heart disease
STSS	Streptococcal toxic shock syndrome
STSS	Streptococcal toxic shock syndrome
MEBSV	multiepitope-based subunit vaccine
iGAS	Invasive group A streptococcal disease
PSIPRED	PSI-blast based secondary structure prediction
RCSB PDB	RCSB Protein Data Bank
NCBI	National Center for Biotechnology Information
BLASTp	Protein Basic local alignment search tool
IEDB	Immune Epitope Database
GRAVY	Grand Average of Hydropathicity
MHC-II	Major histocompatibility complex class II
MHC-I	Major histocompatibility complex class I
HLA	Human leukocyte antigen
IC50	Half-maximal inhibitory concentration.
CTL	Cytotoxic T lymphocytes
HTL	Helper T lymphocyte
C-ImmSim	Cellular Immune Simulation
CTB	Cholera Toxin Subunit B

List of Figures

Figure 1.1	Streptococcal DNases role in the degradation of neutrophil extracellular traps……..	7
Figure 1.2	Different types of vaccine ……………………………………………………….	9
Figure 3.1	Illustrative Workflow of Multiepitope Vaccine Construct………………………….	21
Figure 4.1	Distribution of core-genome and accessory genome………………………………..	29
Figure 4.2	Transmembrane Topology analysis of SpnA family nuclease……………………..	31
Figure 4.3	B cell Epitope Prediction………………………………………………………	33
Figure 4.4	Population coverage calculation of MHC class combined world……………………	37
Figure 4.5	Population coverage calculation of MHC class combined Pakistan………………..	38
Figure 4.6	MEBSV Construct………………………………………………………………..	40
Figure 4.7	Secondary Structure Prediction…………………………………………………..	42
Figure 4.8	First model generated by I-TASSER……………………………………………..	43
Figure 4.9	Refined structure of vaccine construct generated using GalaxyRefine (model 1)…..	45
Figure 4.10	Ramachandran plot of improved MEBSV 3D structure………………………………….	45
Figure 4.11	Final vaccine model validated by Prosa-Web………………………………………	46
Figure 4.12	HDOCK result of docked molecule…………………………………………………	48
Figure 4.13	Best dock model is 1 so it is downloaded from HDOCK and its interaction and bond seen on discovery studio ……………………………………………………………	48
Figure 4.14	Energy minimization and the improvements in molecular interactions graphs……	53
Figure 4.15	In silico immune simulation using the C-ImmSim server…………………..........	57
Figure 4.16	*In-silico* Cloning of MEBSV (vector and the fragment along with the product)….	59
Figure 4.17	Vaccine candidate's codon-optimized gene sequence (shown in red) was cloned between NdeI and FspI restriction sites, within pET28a(+) vector, represented by a black circular structure………………………………………………………...	60

List of Tables

Table 1.1 Transmission mode of diseases associated with *S. pyogenes* ……….…………….…...... 5

Table 2.1 Summary of Key Studies Related to *S. pyogenes* Pathogenesis, Vaccine, and Therapeutic Development…………………………………………………….…...... 18

Table: 4.1 Subcellular location, Antigenicity score, blastp, Allergenicity and Solubility score of *S. pyogenes* ……………………………………………………………………. 30

Table 4.2 B cell Epitopes…………………….………………………………………............ 32

Table 4.3 MHC-I and MHC-II Selected Epitopes…………………………………………… 35

Table 4.4 Structure Information of top five models GalaxyRefine…………………………….. 44

Table 4.5 Docking score of top five models…………………………………………………. 47

Table 4.6 Predicted Hydrogen bonding and Salt Bridges of docked vaccine via PDBePISA.. 49

ABSTRACT

A wide variety of infections, both invasive and non-invasive, are brought on by *Streptococcus pyogenes*. These include cellulitis, impetigo, scarlet fever, pharyngitis, and even post-streptococcal glomerulonephritis. Even now, these infections continue to have a significant worldwide impact. However, there is currently no approved vaccine to prevent diseases associated with *S. pyogenes*. To develop a multi-epitope-based subunit vaccine that targets *S. pyogenes*, a computational tools were used to analyse proteins from *S. pyogenes*. This assisted in testing for toxicity, antigenicity, allergenicity, and B-cell and T-cell epitope predictions. In addition, physicochemical characteristics and population coverage was also examined. To determine the binding strength, the final vaccine design was docked against Toll-like receptor 2, or TLR2. For simulating the process of cloning and expressing this vaccine in the pET-28a vector, SnapGene software proved useful. A total of 422 amino acids were used in the vaccine construct. It included the adjuvant Cholera Toxin Subunit B, or CTB. It contained three predicted B-cell epitopes, seven helper T-cell epitopes, and seven cytotoxic T-cell epitopes. All these got linked together with suitable connectors. Computational analyses revealed the construct to be non-toxic, strongly antigenic, and non-allergenic. Moreover, it demonstrated favorable solubility, stability, and low potential to trigger autoimmune responses. The construct exhibited compatibility with TLR2 molecules, and in silico cloning indicated successful insertion into the pET-28a vector, suggesting potential expression feasibility. This study successfully designed MEBSV with desirable immunogenic and structural properties. The construct appears to be a viable future option for experimental validation against *S. pyogenes* infections.

Chapter 1

Introduction

1 Introduction

1.1 *Streptococcus Pyogenes*

S. pyogenes is a Gram-positive extracellular bacterium that often colonizes the human skin and nasopharynx [3]. It is considered a strict human pathogen and is classified as a Group A β-hemolytic streptococcus. The bacterial cells are coccoid in shape, non-motile, and do not form spores. Their diameter generally ranges between 0.5 and 1.0 µm, and they usually appear in chain-like clusters. To date, over 150 serotypes of *S. pyogenes* have been described. These types differ mainly in the antigenic composition of their M-protein. The organism possesses a circular genome of about 1.85 million base pairs. Most strains produce several toxins and virulence-related proteins, including M-protein, streptolysin O, streptolysin S, and pyrogenic exotoxins SpeA and SpeB [17]. This bacterium is facultatively anaerobic, meaning it can grow in both oxygen-rich and oxygen-poor environments. It also lacks catalase activity. *S. pyogenes* is responsible for a wide range of infections, from mild and non-invasive to severe invasive diseases, and may also lead to immune-mediated disorders [4].

1.2 Epidemiological Background and Clinical Symptoms of *S. pyogenes*

From an epidemiological standpoint, *S. pyogenes* has been classified into more than 220 emm types. These types are differentiated by sequence variations in the amino-terminal region of the M protein gene, which is expressed on the bacterial surface. Emm types show clear differences in their spread across various regions. These typically show up close to the mouth or nose. However, invasive cases vary. They consist of cellulitis and bacteremia. Necrotising fasciitis and puerperal sepsis are also included in this category. In these, the infection spreads deeply or via the blood. Results frequently become serious. The pain intensifies. Swelling occurs. A high fever develops. In certain circumstances, organs may fail. After the infection goes away, *S. pyogenes* may cause immunological problems. One is acute rheumatic fever. Another is post-streptococcal glomerulonephritis. These lead to arthritis. The heart is impacted by carditis. Involuntary movements are brought on by chorea. Swelling is a result of oedema. Blood in the urine is a sign of haematuria. The symptoms of all these infections are diverse. This demonstrates the pathogen's true adaptability. Over time, its evolution has been successful. Pathogenesis is a multi-step process [5].

1.2.1 Prevalence and Global Disease Burden of *S. pyogenes*

Worldwide, *S. pyogenes* infections continue to be fairly prevalent. Every year, they have a variety of effects on millions of people. One of the main problems caused by this bacterium is pharyngitis. According to experts, it causes about 288.6 million cases annually alone. That ranks it among the most common bacterial issues in the world. There are roughly 111 million cases of impetigo worldwide. This demonstrates the extent to which the bacteria causes various skin disorders. There are five to fifty-one cases of acute rheumatic fever for every 100,000 people. Rheumatic heart disease, its chronic form, affects almost 40 million people worldwide. Everywhere, this problem adds a significant burden to cardiovascular issues. Approximately 600,000 cases of bacteremia and other invasive GAS infections occur annually. They pose serious risks to medical resources in general and public health setups in particular. These infections significantly increase healthcare expenses. They also cause patients to have long-term problems. This has a greater impact in developing nations where access to testing equipment and antibiotics is still limited. Environmental, social-economic, and host-specific factors influence the spread and persistence of *S. pyogenes*. Together, those components improve the bacterium's ability to spread among local groups [5].

1.2.2 Mortality Rate of *S. pyogenes*

Despite the availability of effective antibiotics, *S. pyogenes* continues to cause a significant number of deaths. Its invasive disease types are primarily responsible for this. Each year, rheumatic heart disease alone results in nearly 300,000 fatalities. This results from persistent heart damage following untreated or recurring infections. Bacteremia is among the most deadly invasive diseases. Each year, it causes roughly 160,000 deaths worldwide. High death rates are also associated with necrotising fasciitis. It spreads quickly and frequently necessitates amputations or the cutting out of infected areas. In many instances, these infections become fatal due to delayed diagnosis. The prevalence of antibiotic resistance is beginning to increase. In addition, poorer communities lack access to quality healthcare. All of this maintains *S. pyogenes* as a major cause of infections that we could avoid that result in death. It has an impact on the entire world [5].

1.2.3 Current Treatments and Limitations

Antibiotics are the mainstay of treatment for *S. pyogenes* infections. The best options are beta-lactam antibiotics and penicillin. In most situations, they offer consistent efficacy and reliability. Oral antibiotics are frequently prescribed by doctors for less serious conditions like impetigo or

pharyngitis. Intravenous antibiotics are necessary for severe illnesses like streptococcal toxic shock syndrome, necrotising fasciitis, and bacteremia. Immunoglobulin therapy also plays a role in those circumstances. Occasionally, surgery is also required. Symptom relief is used to treat post-streptococcal issues such as acute rheumatic fever or post-streptococcal glomerulonephritis. Prophylactic use of secondary antibiotics aids in preventing recurrence of these problems. Nowadays, antibiotics are still widely accessible. However, treatment problems continue to arise. The number of delayed diagnoses is higher than it should be. Occasionally, patients do not complete their entire course of treatment. Particularly against macrolides, antibiotic-resistant strains are becoming more prevalent. There is currently no approved vaccine against *S. pyogenes*. In a sense, that makes managing the illness even more difficult. Even when bacteria are effectively eradicated, immune-related issues may arise in the future. This demonstrates the urgent need for improved preventative strategies and advancements in vaccines. Overall, *S. pyogenes* remains extremely susceptible to beta-lactam antibiotics. However, over time, resistance to other kinds, such as lincosamides and macrolides, has increased. Recurrent infections are frequently caused by those resistance patterns. Additionally, they lead to less than optimal treatment outcomes. Low-level beta-lactam resistance in some GAS strains has been mentioned in recent reports. In clinical settings, this is a major concern. It emphasises the significance of continuing observation and novel therapeutic approaches [5].

1.2.4 Transmission

Direct contact with respiratory droplets from infected individuals is the primary way that *S. pyogenes* is spread. When those droplets are released into the atmosphere, that is what occurs. The bacterium can travel too through objects or surfaces that carry contamination. This allows for indirect spread in community settings. Things like pharyngitis and pyoderma tend to start up when the germ enters via small cuts or breaks in the respiratory lining or skin. Sometimes contaminated food plays a role as well. It has led to infections in certain cases. Those outbreaks stay pretty much local. The key ways *S. pyogenes* passes on in infections show up summarized in **Table 1.1** [6].

Table 1.1: Transmission methods for diseases linked to *S. pyogenes* infection [6]

Category of Disease	**Specific Condition**	**Mode of Transmission**
Direct Infections	Pharyngitis (Strep throat)	Transmitted primarily via direct human contact, particularly through saliva or nasal secretions.
	Cellulitis	Spread through direct contact with skin lesions or respiratory droplets from an infected person.
	Impetigo	Acquired through close physical contact with an infected individual.
Toxin-Mediated Conditions	Scarlet fever	Spread mainly through saliva or nasal secretions via direct interpersonal transmission.
	Necrotizing fasciitis	Entry of the bacteria may occur through minor or penetrating trauma, enabling either bloodstream dissemination (bacteremia) or direct environmental access.
	Streptococcal toxic shock syndrome (STSS)	Occurs when bacteria invade through disrupted skin or mucous membranes, progressing to deeper tissues and systemic circulation.
Immune-Mediated Complications	Acute rheumatic fever	A delayed autoimmune response following an episode of streptococcal pharyngitis.
	Post-streptococcal glomerulonephritis	An immune complex-mediated consequence resulting from prior throat or skin infections caused by *S. pyogenes*.

1.2 Virulent proteins of *S. pyogenes*

S. pyogenes produces a wide range of surface-associated and secreted virulence factors that enable infection establishment and immune evasion. The bacterium relies on certain main surface components to get by in the host. These include the M protein, the hyaluronic acid capsule, and the S protein. Such structures help the bacterium stick to host tissues pretty effectively. They also allow it to dodge immune recognition and kick off the process of colonization.When it comes to secreted virulence molecules, several proteolytic enzymes stand out. SpyCEP and ScpA are good examples of those. Pyrogenic exotoxins join the mix too, like SpeA, SpeB, and SpeC. In their own way, haemolysins exacerbate the harm. Host cells are broken down by streptolysins O (also called SLO) and SLS (also called SLS). Over time, that causes actual tissue damage. Additional enzymes that are secreted include NADase, also known as NAD glycohydrolase. There are also several DNases involved. Tissue invasion is supported by antibody-degrading proteins like IdeS, EndoS, and Mac-2. In the process, they encourage immune escape. All of these virulence factors ultimately cooperate to perform important tasks. They deal with *S. pyogenes* adhesion, invasion, and persistence within the host [5]. In order to adapt and overcome the host's immune system, several pathogenic Streptococcus species release these extracellular DNases. Every strain of *S. pyogenes* that has been sequenced has at least one gene for producing the secreted DNase enzyme, according to genome-based studies. In all, eight DNase genes have been identified by researchers in GAS. Six of those, namely sda1, sda2, spd1, spd3, spd4, and sdn, bind to prophage elements. In fact, the other two, spnA and spdB, are located directly on the bacterial chromosome. Here, SpnA is particularly noticeable. It is the only DNase that is attached to the cell wall. This occurs because sortase can attach to the peptidoglycan layer thanks to the LPXTG motif. From a functional standpoint, these enzymes disassemble the backbone of DNA in neutrophil extracellular traps, or NETs. When bacteria become entangled in immune defence responses, this aids their escape. DNases significantly alter the virulence of GAS and the course of those infections, according to results from various infection models [5]. The degradation of NETs by streptococcal DNases is shown in **Figure 1.1**. All of this ultimately enables *S. pyogenes* to adhere more effectively and endure longer within the host [63].

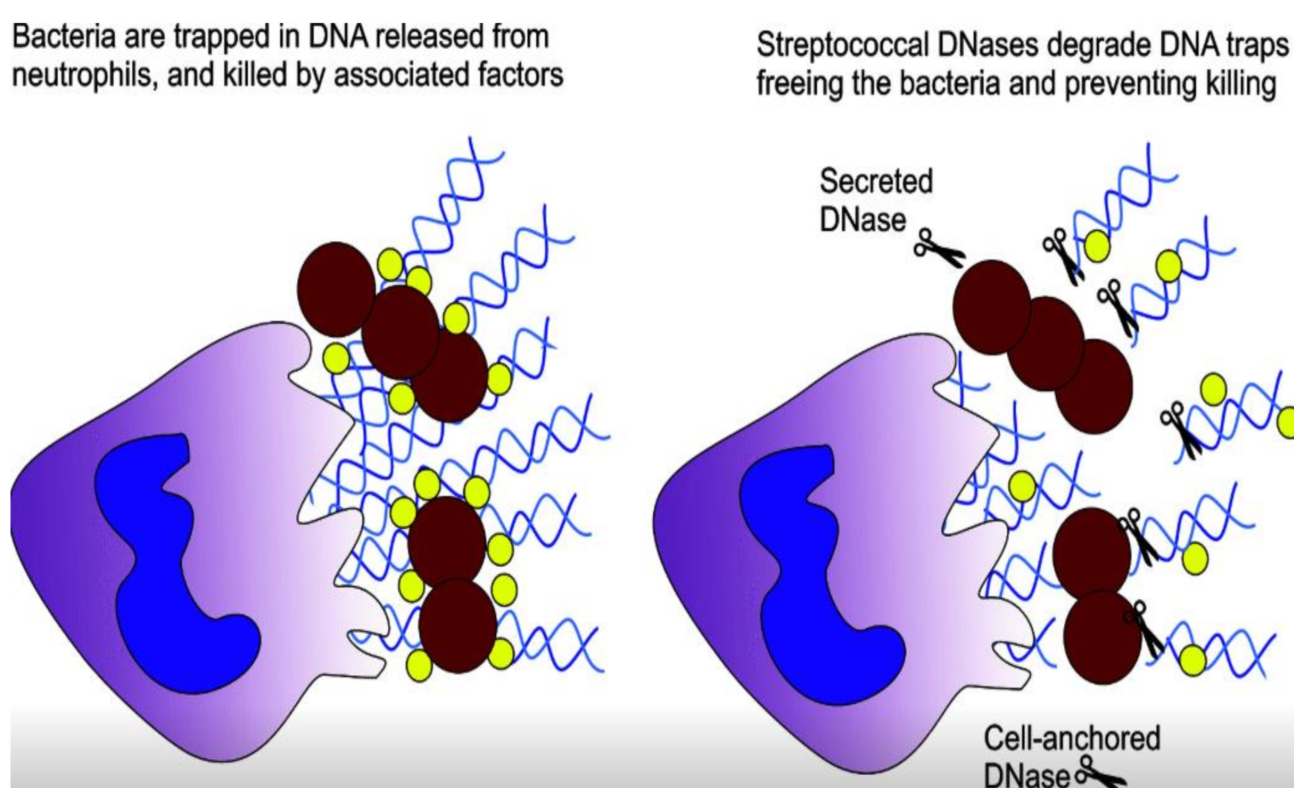


Figure 1.1: DNases are essential for dissolving neutrophil extracellular traps, or NETs. The blue-colored DNA filaments make up the majority of these NETs. Yellow circles indicate the histones and antimicrobial proteins that coat the filaments. Bacteria become entangled in these DNA networks during an infection. As seen on the left, the bacteria are subsequently killed by the bound antimicrobial molecules. DNases produced by Streptococcus are either secreted or adhered to its surface. These DNases work similarly to molecular scissors. They break down the NETs' DNA structure. The bacteria are able to escape immune destruction thanks to this type of enzymatic action. Consequently, it increases their overall virulence and survival chances, as seen on the right [63].

1.3 Vaccine Design

A vaccine functions as a sort of biological setup that safely activates the immune system. It aids the body in developing defences against future infections or diseases caused by a specific germ. The presence of antigens actually determines how well a vaccine works. These are small molecules that resemble pathogen components and cause the immune system to become alert. These antigens can be extracted directly from microorganisms or synthesised in a lab to mimic the structure and function of the real thing. Proteins are the primary components that trigger immunity in many vaccines. They are responsible for those beneficial reactions that guard against illness. However, polysaccharide-based antigens have also shown good efficacy. This is particularly evident in vaccines developed after the late 1980s that target bacterial infections such as meningitis and pneumonia caused by Streptococcus pneumoniae. Trials that verify actual health outcomes from the immune boost are how doctors determine whether a vaccine is effective. This includes preventing infections, reducing the severity of illnesses, and reducing hospitalisations. Finding immune correlates of protection is essential in this situation. They are evident physiological indicators of vaccine efficacy. They accelerate the development of new vaccines, their assessment and approval. Vaccines may be divided into two broad categories. One is live-attenuated and the other can be inactivated or non-live. The live-attenuated type uses a softer form of the pathogen that can multiply within the body and not cause any harm to it. Such a strategy generates a strong, long-term immune response. Conversely, inactivated vaccines utilize dead germs or cleaned up remnants of germs. Since nothing can grow or spread, they safely boost immunity. Recent advances in molecular biology and biotechnology have produced newer vaccine types. Consider DNA or RNA injections, viral vectors, and particles that resemble viruses. These configurations have improved the field of vaccination. They improve accuracy, safety, and the efficiency of eliciting immunity. Additionally, they enable developers to quickly adjust for new or resurfacing threats. Ultimately, vaccines stand out as important contributors to people's long-term health. They significantly reduce the number of illnesses and fatalities worldwide. The different vaccine types currently used in immunisation programs are depicted in **Figure 1.2** [7].

Type of vaccine		Licensed vaccines using this technology	First introduced
Live attenuated (weakened or inactivated)		Measles, mumps, rubella, yellow fever, influenza, oral polio, typhoid, Japanese encephalitis, rotavirus, BCG, varicella zoster	1798 (smallpox)
Killed whole organism		Whole-cell pertussis, polio, influenza, Japanese encephalitis, hepatitis A, rabies	1896 (typhoid)
Toxoid		Diphtheria, tetanus	1923 (diphtheria)
Subunit (purified protein, recombinant protein, polysaccharide, peptide)		Pertussis, influenza, hepatitis B, meningococcal, pneumococcal, typhoid, hepatitis A	1970 (anthrax)
Virus-like particle		Human papillomavirus	1986 (hepatitis B)
Outer membrane vesicle	Pathogen antigen; Gram-negative bacterial outer membrane	Group B meningococcal	1987 (group B meningococcal)
Protein–polysaccharide conjugate	Polysaccharide; Carrier protein	*Haemophilus influenzae* type B, pneumococcal, meningococcal, typhoid	1987 (*H. influenzae* type b)
Viral vectored	Viral vector; Pathogen gene; Viral vector genes	Ebola	2019 (Ebola)
Nucleic acid vaccine	DNA; RNA; Lipid coat	SARS-CoV-2	2020 (SARS-CoV-2)
Bacterial vectored	Pathogen gene; Bacterial vector	Experimental	–
Antigen-presenting cell	Pathogen antigen; MHC	Experimental	–

Figure 1.2: Summary of the main vaccine types, such as live-attenuated, inactivated, subunit, viral vector-based, nucleic acid (DNA and RNA), and virus-like particle (VLP) vaccines. Safety profiles, immune activation mechanism, and composition of these methods vary [7].

1.4 Overview of Research Approach

Recent advances in bioinformatics and computational biology have significantly improved the vaccine development process today. Instead of relying exclusively on outdated laboratory tests, these domains offer innovative methods for vaccine prediction based on epitopes. The amount of manual laboratory testing is decreased by these computer-based methods. They also speed up the entire process of preparing vaccines. Early detection of possible side effects is facilitated by computer-based tools. This reduces the overall cost of research. In the end, it makes vaccine design more effective. Due to enormous quantities of biological information, the search of new therapeutic targets has been advanced. This is possible with intricate computer systems. Subtractive genomics is one of the methods. It examines the pathogenic and host proteins. The objective is to eliminate proteins that complement each other. It also determines the pathogen-specific proteins. Those could be useful for therapy. The other field of bioinformatics that focuses on the use of computers to study the immune responses is known as immunoinformatics. Predictive programs are used to identify the locations of B-cell and T-cell epitopes, which speeds up the process of turning peptides into vaccines. They target bacteria, viruses, and parasites. Another noteworthy development is reverse vaccination, which uses genome data to analyse the composition of pathogen antigens and then predicts promising vaccine candidates. This approach has had a significant impact on vaccine research. It substitutes genome-based predictions for random testing. The combination of these computer techniques has altered the way that vaccines are discovered. They speed up and improve the accuracy of the process. Combining immune and genomic data also lowers costs [1]. Complete sets of proteomic and genomic data are now available in public databases. This has improved our use of bioinformatics in vaccine design. Researchers can identify antigenic regions in protein chains by using immunoinformatics analysis. Strong immune responses may be triggered by those components. These results lend support to the development of multi-epitope vaccines. Both T-cell and B-cell responses are triggered by these MEVs. They provide long-lasting protection. Additionally, they exhibit genuine promise in the treatment of specific infections [2].

1.5 Problem Statement

S. pyogenes targets particular hosts as a bacterial pathogen. There are numerous options for treating it. Currently, there is no approved vaccine available to prevent its infections. The importance of developing effective vaccines is emphasised by that missing component. Building a MEBSV

targeted at *S. pyogenes* is the goal of the current project. It makes use of computational methods and immunoinformatics tools. Reverse vaccinology finds promising antigen targets by using genomic information. This method provides a innovative production method of vaccines. Those vaccines could not be made by the use of obsolete methods [8]. The search of epitopes that induce robust responses of the immune system continues. It is also a facilitator of improved cross-reaction computer foretelling. It also expands our understanding of the interaction of pathogens and hosts. This is all very important. It encourages effective planning of new interventions. One of such treatments is effective vaccinations against *S. pyogenes* infections.

1.6 Motivation

A better insight into the response of the immune system towards infections caused by *S. pyogenes* should be achieved once this in silico analysis is completed successfully. It could also provide new areas of therapy and contribute to the development of more precise and effective approaches to enhance immune responses.

1.7 Aim and Objectives

The overall goal of the proposed research is to develop a possible vaccine against MEBSV which targets *S. pyogenes* specifically. Its design work is done using in silico techniques. The vaccine structure elicits certain B-cell and T-cell epitopes. It also incorporates appropriate adjuvant. This enhances immunogenicity. Comprehensively, it generates a balanced immune response. The objectives of this study are as follows:

- Identifying epitopes that can trigger humoral and cell-mediated immune responses.
- Identifying *S. pyogenes* antigens that are immunogenic and highly conserved. Those might make good targets for vaccines.

1.8 Research Questions

This study aims to address a few key research questions.

- It examines which conserved proteins on *S. pyogenes*' surface can elicit an immune response are most effective as targets for generating a MEBSV?

- In order to create an efficient MEBSV that prevents *S. pyogenes* infections, it also investigates which immunoinformatics techniques are employed for identifying and evaluating potent B-cell and T-cell epitopes?

1.9 Proposed Solution

Analysing the protein structures and genomic sequences of different *S. pyogenes* strains is necessary to predict putative antigens. Additionally, it allows us to evaluate their ability to trigger immunological responses. Different traits linked to immunity are identified by bioinformatics tools. These findings guide lab work and lead to better approaches to disease control. Reverse vaccination is useful in this situation. It predicts possible vaccine candidates by identifying antigens that elicit strong immune responses and are consistent across strains. This makes it possible to evaluate the body's possible defences against *S. pyogenes* infections using a computer.

1.10 Scope and Limitations of the Study

This study covers the precise prediction and selection of epitopes for MHC class I, MHC class II, and B-cells. This is aimed at building robust cell-based and antibody-based immunity. This was done by using a variety of bioinformatics databases and tools. We examined the fundamental physical and chemical characteristic features of the epitopes, their coverage of various populations, toxicity, antigenicity, and allergenicity. Dynamics simulations and molecular docking were used to improve the vaccine design. These tested its possibilities to bind with human immune receptors as well as its structural integrity. This, in general, develops a computer-based strategy in creating a vaccine against *S. pyogenes*. It also creates space to conduct more laboratory research and human testing.

This work has certain major shortcomings. Everything is based solely on computer predictions. The vaccine's safety and immunological effects are not supported by laboratory testing or animal research. The accuracy of the databases and algorithms determines the outcomes. The messiness of the interactions between hosts and pathogens may not be fully captured by those. Variations in *S. pyogenes* strains' genetic makeup also matter. Individuals in various groups have different immune responses. All of that could affect the vaccine's effectiveness. Thus, this creates a strong theoretical foundation. However, a number of practical experiments and clinical tests are still required to determine whether the vaccine is effective in real-world scenarios.

Chapter 2

Literature Review

2.1 Introduction to Literature Review on *S. pyogenes*

This literature review summarizes previous studies relevant to this research. It presents key findings and methods that help identify vaccine candidates against *S. pyogenes*. It is a Gram-positive, aerobic, non-motile bacterium. It usually lives in the human nasopharynx and on the skin. This bacterium is a strict human pathogen. It spreads through respiratory droplets, contact with infected skin, or nasal secretions. *S. pyogenes* remains a major health concern, especially in areas with limited healthcare resources [3], [10]. It is still sensitive to penicillin. However, treatment failures have been reported, especially in cases of GAS-induced pharyngitis. In addition, antibiotics do not always prevent post-infection complications. Therefore, developing a safe and effective vaccine is considered the most reliable strategy to control and prevent the wide range of diseases caused by this bacterium [6].

2.2 Pathogenesis and molecular mimicry of *S. pyogenes*

Developing a broad-spectrum vaccine faces several challenges. One major issue is the antigenic similarity between certain bacterial components and human tissues. This molecular mimicry may cause post-infection complications similar to autoimmune diseases. As a result, vaccine development requires careful design to avoid unwanted immune responses. These problems have held back the effectiveness of earlier vaccine efforts. Research now targeting *S. pyogenes* vaccines centers on two kinds of antigens. One set comes straight from the M protein. The other has no connection to it. The M protein stands out as a major factor in bacterial virulence right on the surface. It triggers a powerful immune response. Additionally, it enables the bacterium to adhere to host cells. It simultaneously avoids being ingested by phagocytes [10].

2.3 Research Gap of *S. pyogenes*.

Better computational tools are crucial to the development of vaccines against *S. pyogenes*. These tools facilitate accurate antigen detection, cross-reactivity testing, and determining whether an antigen can elicit a positive immune response. Standard animal models and thorough comparisons between studies are also very important. This kind of work begins to bridge the gap between what bioinformatics computers predict and what works in real-world clinical settings. Collaboration on these issues could accelerate the development of a vaccine against *S. pyogenes* [5], [11]. Typically, *in-silico* work adheres to reference genomes. These may not fully represent the genetic diversity found in human populations. Determining how host gene variations impact immune responses to

S. pyogenes appears to be essential. Solid vaccines and individualised treatments are supported by that knowledge. This flaw is addressed by bringing in diverse genomic data [6]. *S. pyogenes* is linked to numerous life-threatening infections. For now, antibiotics treat those infections. However, they are unable to provide enduring protection. Because the body's immune system isn't strong enough, infections keep recurring. A vaccination strategy is clearly the more intelligent one. It strengthens defences against specific pathogens [9].

2.4 Critical Analysis of *S. pyogenes*

To get around the host's immune system, *S. pyogenes* employs a number of strategies. It includes modifications to penicillin-binding proteins and molecular mimicry. These changes have reduced the efficacy of antibiotics. The issue of antibiotic resistance is becoming even worse. It emphasizes the necessity of alternative types of treatment. The significance of vaccinations against this pathogen is escalated through safe vaccinations [10]. It is important to know the mode of spread of the pathogen and its mechanism of disease. This kind of understanding forms the useful strategies of addressing and preventing the issues. *In-silico* procedures can now help researchers overcome *S. pyogenes* vaccine development challenges in the present day. Such methods accelerate the identification of appropriate vaccine targets. They increase the likelihood that formulas will be developed that actually work [6].

2.5 Limitations, Gaps and Future Directions in Current *S. pyogenes* Literature

S. pyogenes vaccine development is lagging behind that of many other important infections. A few aspirant candidates are currently seated in studies. However, the lineup as a whole is narrow and unvarnished. Finding epitopes that trigger a strong, balanced immune response without negative side effects is a significant challenge. Selecting those epitopes is aided by computational tools. They reduce the risk of immune damage while aiming for long-lasting protection. Developing vaccines that protect against different strains of *S. pyogenes* while minimising adverse effects is also difficult. Achieving broad efficacy across different strains while avoiding unwanted immune responses adds further complexity. In this context, *in-silico* tools are valuable for predicting antigenicity and cross-reactivity, supporting the safety and effectiveness of potential vaccine candidates [5], [11].

2.6 Clinical Characteristics and advancements of *S. pyogenes*

The M protein plays a major role in the disease-causing ability of *S. pyogenes* and is encoded by the *emm* gene. The hypervariable region of this gene is the basis for *emm* typing, which helps classify different *S. pyogenes* strains. More than 240 *emm* types and over 1000 subtypes have been identified. Their distribution often varies across geographic regions [5]. *S. pyogenes* was previously considered largely susceptible to antibiotics, with penicillins and macrolides commonly used for treatment. Over the past decade, however, drug-resistant strains have been reported worldwide, including in Europe and Asia. Clindamycin remains effective for severe infections, and fluoroquinolones have shown promising results. Although resistance to macrolides has increased in many regions, reports of erythromycin resistance in Pakistan are still relatively low [41]. The primary research on the pathogenicity of S. pyogenes is compiled in **Table 2.1**. It also covers vaccine research and treatment development. The main findings from all of this work are indicated in the table. It explains how experiments are conducted. It also demonstrates the final results of those efforts.

Table 2.1: Summary of Key Studies Related to *S. pyogenes* Pathogenesis, Vaccine, and Therapeutic Development

S. No	Objective	Methodology	Key Findings	References
1	To emphasize the urgent need to accelerate vaccine development efforts against *S. pyogenes* due to its global disease burden.	A review of eight vaccine candidates, focusing on preclinical proof-of-concept, safety, and immunogenicity data, along with insights from a Phase 1 clinical trial.	Eight candidates demonstrated safety and immunogenicity. The formation of the Strep A Vaccine Global Consortium (SAVAC) is a pivotal step in expediting vaccine development.	[5]
2	To discuss obstacles in designing a universal vaccine and review potential candidates.	A literature review on pathogenic mechanisms of *S. pyogenes*, vaccine development challenges, and ongoing research in candidate vaccines.	Understanding of adhesion, colonization, and immune evasion mechanisms has improved, facilitating informed vaccine design.	[6]
3	To highlight ongoing disease outbreaks and the importance of integrated public health and research approaches.	Focused on key virulence determinants in *S. pyogenes* and recent research advances in controlling invasive infections.	WHO has published a roadmap for *S. pyogenes* research, encouraging vaccine development aligned with global health priorities.	[10]
4	To underline the need for effective vaccines despite continued penicillin sensitivity.	Analysis of vaccine composition, particularly immunogens and adjuvants (mainly alum), and evaluation of immune	WHO advocates for vaccine development. However, strain diversity and antigenic variability remain major hurdles.	[11]

		response elicitation.		
5	To design a computationally predicted multi-epitope vaccine against S. aureus and *S. pyogenes* targeting cellulitis.	Bioinformatics-based selection and evaluation of membrane proteins, followed by epitope prediction and immunogenicity assessment.	Three epitopes were identified: one from S. aureus and two from *S. pyogenes* proteins (SPy and scpA), with potential for vaccine development.	[12]
6	To identify conserved and immunogenic antigens suitable for broad-spectrum vaccine design.	Eight surface proteins were examined for T-cell epitope prediction, structural analysis, and conservation across Streptococcus strains.	Seven potential epitopes derived from three different proteins were recognized as promising targets for future experimental validation	[13]
7	To compare virulence factors responsible for TSS in S. aureus and S. pyogenes.	Protein sequences for TSST-1, SpeA, and SpeC were retrieved and analyzed for structural, antigenic, and functional characteristics.	All three toxins displayed high antigenicity and hydrophilicity, highlighting their potential roles in TSS pathogenesis and vaccine targeting.	[14]
8	To explore novel therapeutic targets for drug-resistant *S. pyogenes* infections.	Core proteome analysis of four *S. pyogenes* strains, protein prioritization, and docking simulations with 1,000 phytochemicals.	Two proteins were proposed as novel drug targets, and multiple phytochemicals showed promise as potential inhibitors.	[15]

Chapter 3

Materials and Methods

3.0 Methodology and Research Design

The computational setup used to develop the multi-epitope vaccine known as MEBSV is covered in this chapter. *S. pyogenes* is its target. The entire concept is based on concepts from reverse vaccinology and immunoinformatics. It uses various *in-silico* techniques to identify antigens. It also oversees epitope mapping, structural checks, and immunological reaction simulations. The sequence of the target protein is first retrieved and analysed. Antigenic spots and conserved areas were then found. These, along with MHC class I and class II and B-cell epitopes, may trigger robust immune responses. This setup was intended to start humoral and cell-mediated immunity. Additionally, a population coverage was examined. This investigated the possibility of the chosen epitopes being recognised globally. The top epitopes were joined by the use of adjuvants and suitable linkers. This offered a MEBSV a reliable option. The physicochemical properties were assessed and tested in terms of allergenicity, toxicity, and overall antigenicity. All of that confirmed its safety and encouraging immunogenic qualities. Immune simulations and molecular docking were used. Those examined how stable the vaccine interactions were with the TLR2 receptor. It made sure of structural soundness and real effectiveness. In the end, they did *in-silico* cloning. Each step in this pipeline was selected to ensure scientific accuracy, reproducibility, and alignment with current immunoinformatics standards. This integrative *in-silico* approach provides a rational pathway for identifying novel antigenic vaccine candidates against *S. pyogenes*. A complete schematic of the workflow is shown in **Figure 3.1**.

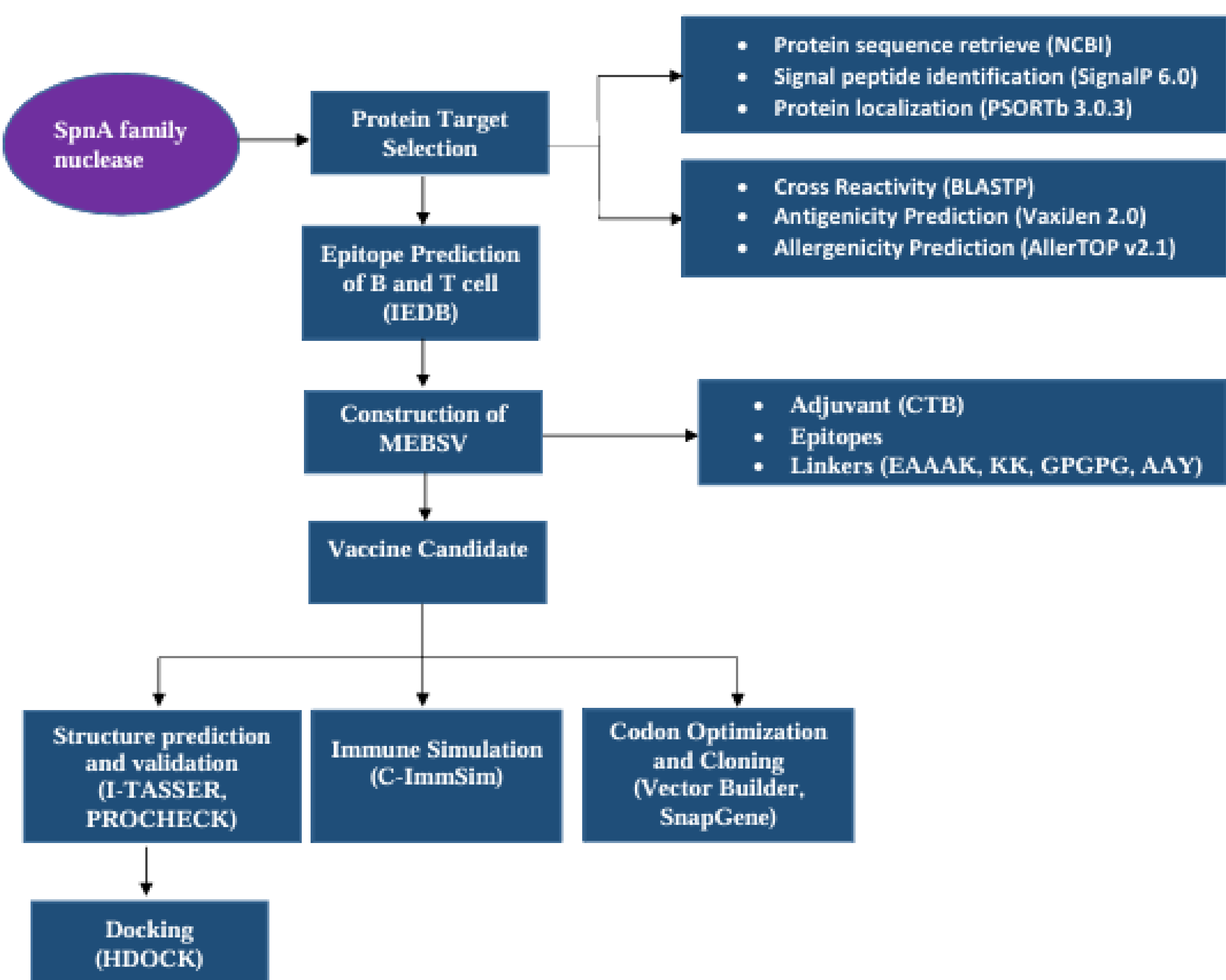


Figure 3.1: Illustrative Workflow of Multiepitope Vaccine Construct

3.1 Vaccine Target Selection

The complete genomic sequences of *S. pyogenes* strains were obtained from the NCBI database. Core genes were identified using the Pangenome Explorer tool. These core genes were then used for subsequent analyses [32, 3, 50]. Signal peptides have important characteristics that influence vaccine design. To identify proteins with signal peptides, analyses were performed using the SignalP 6.0 and TargetP 2.0 web servers [42], [43]. These tools can predict not only signal peptides but also mitochondrial, chloroplast, and thylakoid luminal transit peptides. This information helps in selecting suitable vaccine targets [2], [11]. Only proteins containing signal peptides were selected for further localization analysis using TMHMM 2.0 and PSORTb 3.0.3. These proteins were found to be associated with the outer membrane or to contain transmembrane domains. Subtractive genomics was then applied to identify pathogen-specific proteins by removing sequences homologous to human proteins. This approach helps identify proteins that are essential for the pathogen’s survival but do not share similarity with the host. Protein sequences in FASTA format were retrieved from NCBI and compared against the human proteome using BLASTp. Proteins that showed no real match to human sequences got labeled as non-homologous. Those turned out to be likely essential genes [11]. Epitopes matching human proteins too closely could dial down immune activation. In certain situations, they might even start autoimmune issues. BLASTp reduce those risks and increase vaccine safety. The selected epitopes are compared to the human proteome using this tool. It highlights any possible sequence matches. This encourages more thoughtful and intelligent epitope selection [2]. Protein or peptide antigenic strength predictions are handled by the VaxiJen web server [45]. For that assessment, the Auto Cross-Covariance transformation method is used. Accuracy ranges from roughly 70% to 90%. As a result, it is a good choice for determining antigenicity while developing vaccines. Using AllerTOP version 2.0, the selected protein's allergenicity was evaluated. That system sorts proteins into allergenic or non-allergenic categories. On top of that, the ProtParam server analyzed main physicochemical features. It focused on the protein-based vaccine options. These properties include chemical composition, molecular weight, theoretical isoelectric point (pI), aliphatic index, instability index, GRAVY (grand average of hydropathicity), and estimated half-life in both *E. coli* and mammalian cells [2], [44], [46]. An effective vaccine candidate should trigger a strong immune response in the host. Therefore, it must have essential properties such as high antigenicity, non-allergenicity, stability, non-toxicity, and good solubility [15].

3.2 B-cell and T-cell Epitope Prediction

B-cell epitope prediction was carried out using the IEDB database tool. The potential epitopes are shown graphically, indicating residue positions. This tool provides several reliable algorithms for predicting linear B-cell epitopes, helping to accurately identify antigenic regions within the protein sequence [2], [48]. T-cell epitope prediction was performed using the IEDB T-cell epitope prediction tool. The analysis included epitopes binding to both MHC class I and II molecules. Selection criteria focused on immunogenic potential, non-allergenicity, low IC50 values, and favorable physicochemical properties to ensure safety and efficacy. As these epitopes stimulate cytotoxic T lymphocytes (CTLs), which eliminate infected host cells, they are crucial for vaccine development. They improve cellular immune responses by activating both $CD4^+$ and $CD8^+$ T-cell subsets. MHC molecules present these epitopes on antigen-presenting cells (APCs), initiating adaptive immunity. Long-term protection also depends on the formation of memory T-cells after pathogen clearance. The IEDB platform allows prediction of binding affinities for MHC class I and II and supports allele-specific analyses, including HLA-DP, HLA-DQ, and HLA-DR [2], [6], [47].

3.3 Population Coverage Analysis

The prevalence of specific HLA alleles varies widely among human populations. Therefore, assessing epitope suitability for MHC binding requires knowledge of allele distribution in target populations or regions. Conserved CTL and HTL epitopes are often linked to broad-spectrum immune responses. The IEDB Population Coverage tool helps evaluate the selected epitopes by analyzing their binding to common HLA alleles across different ethnicities and geographic locations. This allows estimation of the potential effectiveness of a vaccine in diverse populations [2], [51].

3.4 Construction of MEBSV

Adjuvants often function as cytokine inducers or Toll-like receptor (TLR) agonists, enhancing immune activation. There are various adjuvants that are intended to elicit particular immune responses. The intended immunological effect determines their selection. When creating multi-epitope vaccines, linkers are essential. They preserve the correct shape of the epitope and promote proper folding. Aside from this, they maintain receptor accessibility. All of this results in better immune stimulation [38]. A variety of different types of linker were intentionally employed in the

formulation of the vaccine. This ensured the presentation of the epitopes in the best way. Separation of B-cell epitopes was performed by using KK linkers. They assisted with the strong antibody spotting. AAY linkers are connected to MHC class I epitopes and these are CTL types. GPGPG linkers dealt with MHC class II, or HTL epitopes. This had the effect of keeping their immune roles separate and providing processing. With the full sequence of the vaccine assembled, they were again investigating antigenicity and allergenicicity. If necessary, they would make changes to the layouts of the epitopes. That step aimed to strengthen the overall immune reaction. The final construct included an N-terminal adjuvant, various linkers (EAAAK, GPGPG, AAY, and KK), B-cell epitopes, MHC class I epitopes, MHC class II epitopes, and a C-terminal 6×His tag for purification. Cholera enterotoxin subunit B protein (UniProt ID: P01556) was selected as the adjuvant. Physicochemical properties of the construct were analyzed using the ProtParam tool. Since large amounts of recombinant protein are usually expressed in bacterial hosts, solubility is critical for production efficiency. Protein-Sol and SoluProt servers were employed to predict solubility in a prokaryotic expression system [52], [54]. Solubility affects not only protein yield but also downstream processing. Protein half-life is another important factor, particularly for candidates with therapeutic potential [2], [11].

3.5 Structure Prediction, Refinement and Validation of MEBSV

The PSIPRED server was used to predict the secondary structure of the vaccine construct, using PSI-BLAST-based algorithms. To generate the three-dimensional (3D) model of MEBSV, the vaccine sequence was submitted to the I-TASSER server with default settings. Structural refinement was then carried out using the GalaxyRefine server. The initial 3D model in PDB format was uploaded for refinement, which improved quality scores and produced a more stable and reliable vaccine structure [11], [56], [49], [57]. The refined 3D model was validated using a Ramachandran plot generated by the PROCHECK server. This plot evaluates the stereochemical quality of the protein model by analyzing the distribution of amino acid residues in favored, allowed, and disallowed regions [31], [59].

3.6 Docking

Molecular docking is a widely used computational method that predicts interactions between receptors and ligands. It estimates binding affinity, usually through scoring functions [2]. Toll-like receptor 2 (PDB ID: 2Z7X) was chosen as the receptor. Its 3D structure was obtained from the

RCSB Protein Data Bank [60]. The vaccine construct was used as the ligand. Molecular docking was performed using the HDOCK server to study the interaction between the vaccine protein and TLR2. Information on binding affinity and interaction patterns was obtained from this analysis [61].

3.6.1 Energy Minimization and Interaction Improvement

The vaccine receptor complex's structural flexibility and stability were investigated using molecular dynamics analysis on the iMODS server. It uses Normal Mode Analysis (NMA) to assess the protein complex's internal motions. The analysis reveals some important conclusions. Those consist of a covariance matrix and a deformability plot. It also covers an elastic network model, B-factor values, and eigenvalue data. These elements work together to reveal information about the complex's adaptability. Additionally, they show that it is generally stable [35], [62].

3.7 Immune Simulation

The C-ImmSim server was used to simulate the immune response resulting from the vaccine construct. They fed in the vaccination sequence under standard conditions. Its capacity to elicit an immune response, including the generation of antibodies, was predicted by this [33], [58]. C-ImmSim models adaptive immune responses through pathogen-host interactions. For that, position-specific scoring matrices are used. The tool provides details about the main immune components. Those cover activated B and T lymphocytes plus helper T cells. It includes plasma B-cell numbers after exposure to the antigen. This kind of simulation gives valuable details on the vaccines immunogenic strength. It helps in tweaking and assessing the construct [2].

3.8 Expression analysis, Codon optimization and Molecular cloning

Codon optimization helped boost gene expression levels in the host organism. This approach addresses variability in foreign gene expression caused by differences in mRNA codon usage [11]. VectorBuilder was employed for codon optimization [53]. *Escherichia coli* K12 strain was used as the host for the prokaryotic expression system [52]. In silico cloning is a computational method that simulates the insertion of DNA sequences into vectors using known sequence data. This approach allows researchers to predict experimental outcomes before performing laboratory procedures. It improves accuracy while reducing the time, cost, and resources needed for wet-lab experiments [39], [40]. Molecular cloning of the optimized vaccine sequence was performed using SnapGene software. The vaccine construct was inserted into the pET28a(+) plasmid vector, with

suitable restriction sites added at the N- and C-terminal ends to enable efficient cloning and expression in the host system [34], [55].

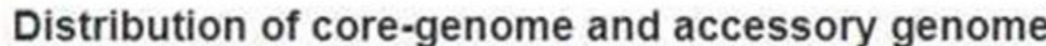
Chapter 4

Results

4.1 Vaccine Target Selection

NCBI database's genome assembly and annotation report reveals altogether, 2925 *S. pyogenes* strains. Out of 2925 strains, 270 exist as complete genomes. 1835 exist in contig form, and 797 in scaffold form (NCBI, accessed on 5 January 2024). This study contained strains with complete genome. The protein and genes of 270 complete genome were retrieved from NCBI. Pan genome analysis done through PanExplorer. The result shows 847 core genes, 868 Strain specific genes and 3287 Dispensable genes, depicted in **Figure 4.1**. Only core genes were selected for further analysis. Screening for signal peptides revealed that only fourteen proteins contained signal peptide sequences with accession numbers: WP_002992220.1, WP_011184773.1, WP_038431224.1, WP_012560804.1, WP_002990808.1, WP_002983307.1, WP_002992223.1, WP_002983723.1, WP_002984246.1, WP_011054207.1, WP_002992954.1, WP_038431426.1, WP_038431255.1 and WP_011054910.1. These proteins were selected for further subcellular localization analysis, with shortlisted candidates found to be either outer membrane-associated or containing transmembrane domains. Subtractive genomics was applied to eliminate proteins with similarity to human proteins. The FASTA sequences of the selected proteins were retrieved from the NCBI database and subjected to BLASTp analysis against the human proteome to filter out homologous proteins and retain only the non-homologous ones. **Table 4.1** presents the selected proteins along with their subcellular localization, antigenicity scores, BLASTp results, allergenicity, and solubility**.** Only proteins predicted to be soluble, antigenic, and non-allergenic was kept for further analyses.

The protein chosen was SpnA family nuclease (WP_038431255.1), with a subcellular location of cell wall, an antigenecity score of 0.4573, non-homologous, likely non-allergen, and soluble properties. TMHMM results for SpnA family nuclease shows that 1 to 883 amino acids were present outside the surface depicted in **Figure 4.2.** The protein contains 910 amino acids with molecular weight 99952.27, and predicted pI of 6.45. It has 111 negatively charged and 106 positively charged residues. The protein has a half-life of 30 hours in mammalian reticulocytes, 20 hours in yeast, and 10 hours in Escherichia coli. It has an instability value of 31.01, indicating stability. The protein had an aliphatic index of 91.74. Its Grand Average of Hydropathicity (GRAVY) value measured -0.339.



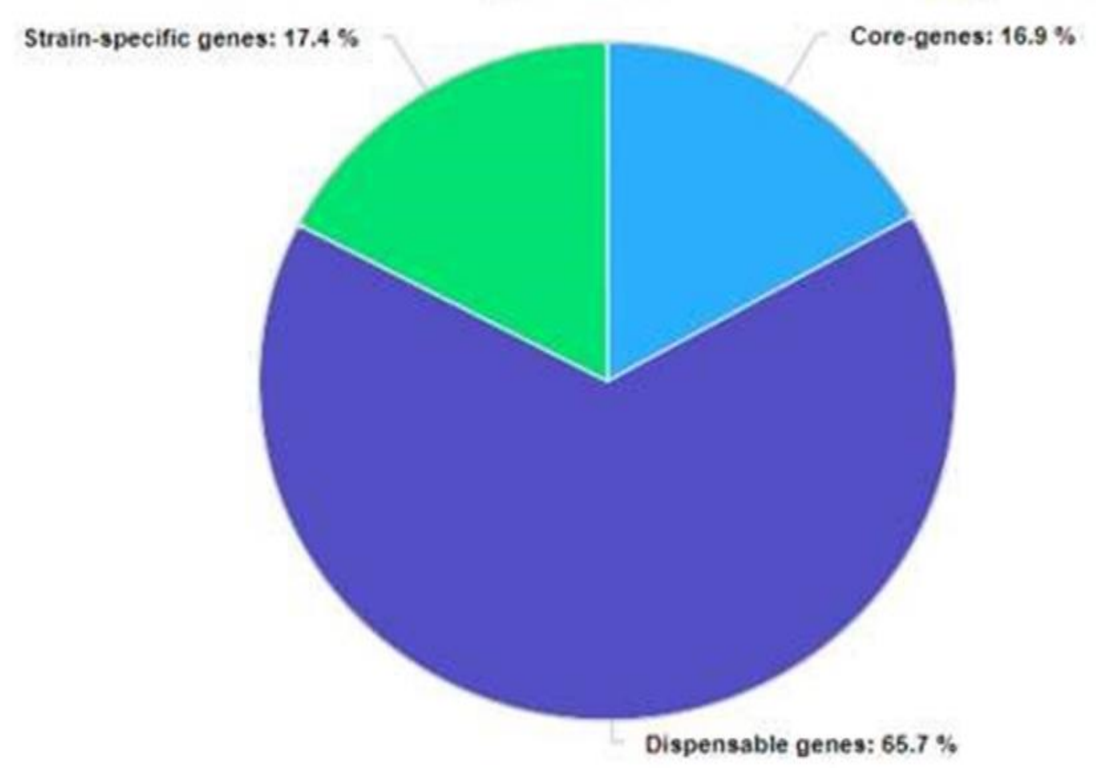


Figure 4.1: Distribution of core-genome and accessory genome (green colour shows strain specific gene, light blue colour shows core-genes and purple colour represents dispensable genes).

Table 4.1: Subcellular location, Antigenicity score, blastp, Allergenicity and Solubility score of *S. pyogenes*

Protein	**Subcellular Location**	**Antigenecity**	**blastp**	**Allergenicity**	**Solubility**
TlpA disulfide reductase family protein (WP_011184773.1)	Cell wall	0.5732	Non homologous	Probable allergen	Insoluble
glycoside hydrolase family 73 protein (WP_002983307.1)	Extracellular	0.5273	Non homologous	Probable Non-allergen	Insoluble
membrane protein insertase YidC (WP_011054207.1)	Cytoplasmic Membrane	0.6182	Non homologous	Probable Non-allergen	Insoluble
leucine-rich repeat domain-containing protein (WP_038431426.1)	Cell wall	0.5408	Non homologous	Probable Non-allergen	Insoluble
SpnA family nuclease (WP_038431255.1)	Cell wall	0.4573	Non homologous	Probable Non-allergen	Soluble

TMHMM result

```
# WEBSEQUENCE Length: 910
# WEBSEQUENCE Number of predicted TMHs:  1
# WEBSEQUENCE Exp number of AAs in TMHs: 21.69332
# WEBSEQUENCE Exp number, first 60 AAs:  1.00094
# WEBSEQUENCE Total prob of N-in:        0.05637
WEBSEQUENCE     TMHMM2.0        outside      1   883
WEBSEQUENCE     TMHMM2.0        TMhelix    884   903
WEBSEQUENCE     TMHMM2.0        inside     904   910
```

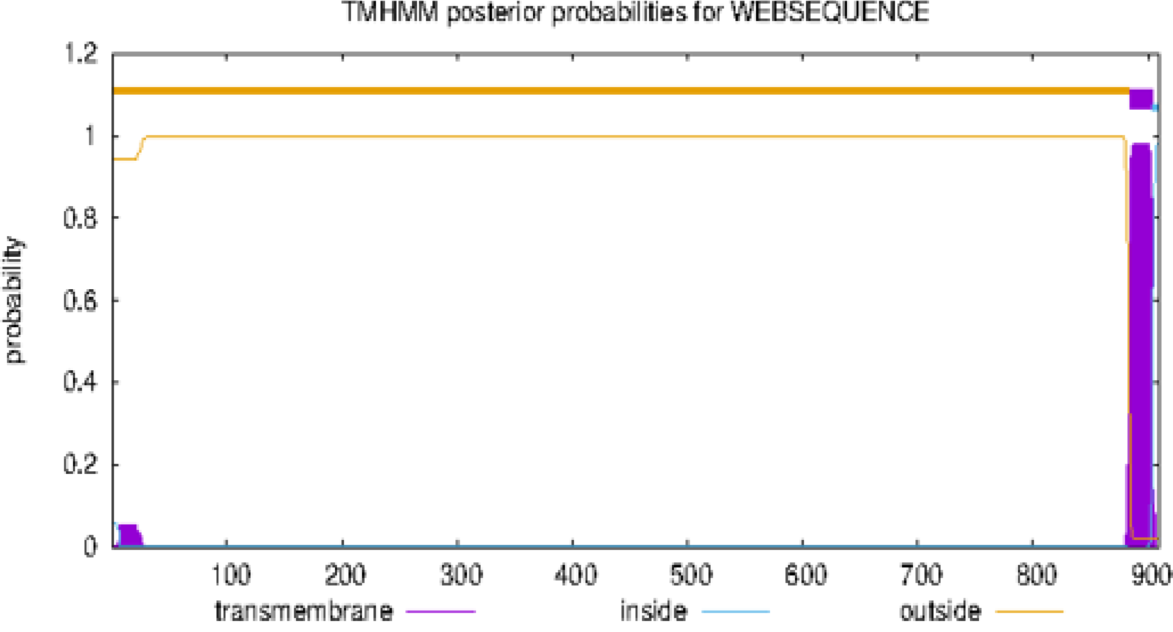


Figure 4.2: Transmembrane Topology analysis of SpnA family nuclease.

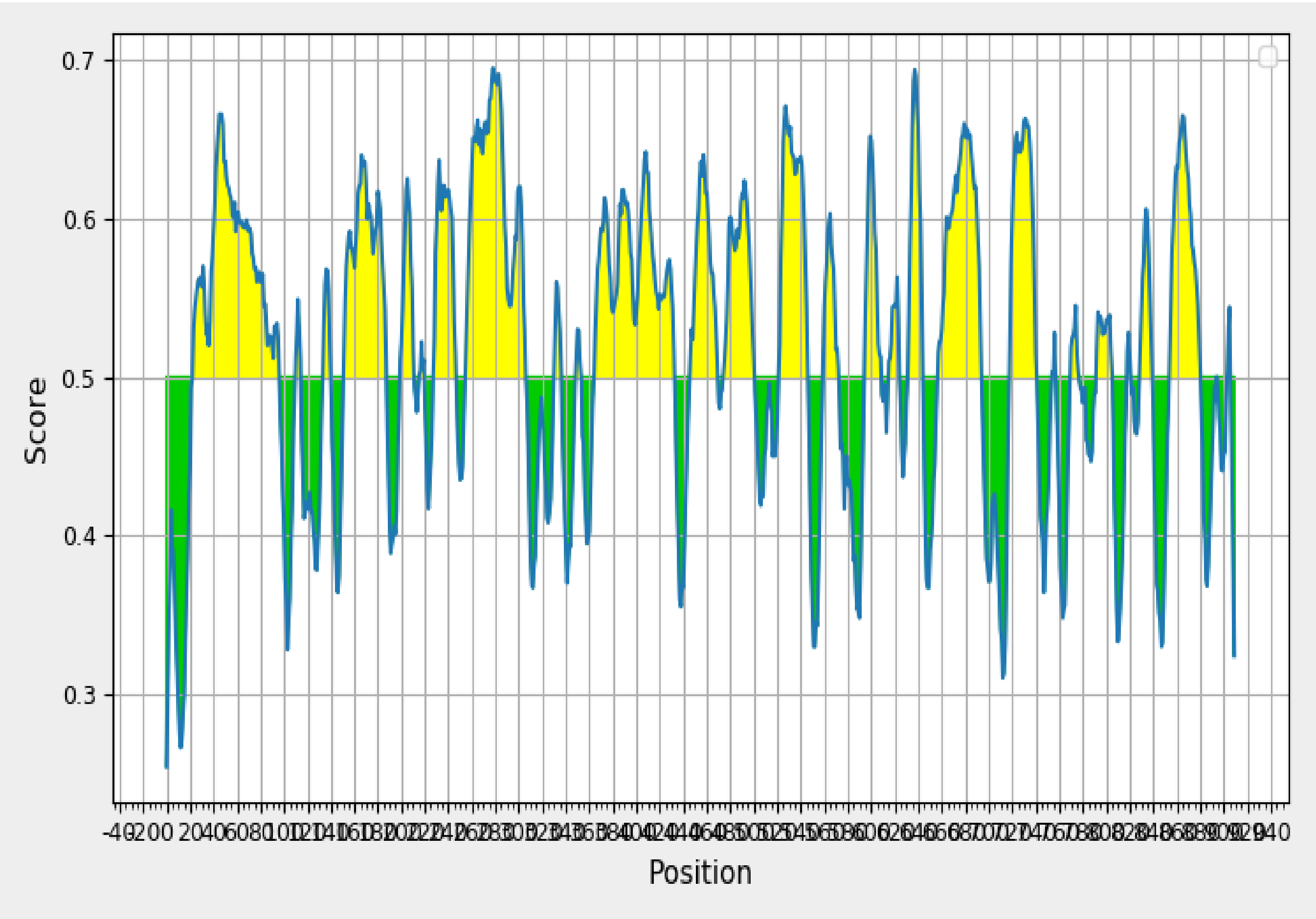
4.2 B-cell and T-cell Epitope Prediction

B cells are primarily involved in initiating the body's humoral immune response by producing antibodies. On the other hand, T cells are responsible for cell-mediated immunity**,** recognizing and eliminating infected or abnormal cells to protect the host from pathogens. B-cell epitope prediction of SpnA family nuclease (WP_038431255.1) was performed using the IEDB database, specifically utilizing Bepipred Linear Epitope Prediction 2.0 method. The predicted results is illustrated in **Figure 4.3** and **Table 4.2** presents the predicted B-cell epitopes The output data, exported in Excel format, was further analyzed to assess each epitope's antigenicity, allergenicity, and toxicity.

IEDB platform tools was used to predict epitopes on MHC class I and class II molecules. A good group of MHC alleles were selected to examine in detail. Since the epitopes with the lowest IC50 values were the most noticeable, they were first subjected to tests for toxicity, antigenicity, allergenicity, and solubility [20]. To facilitate delving into the specifics, we extracted the results into Excel files. **Table 4.3** shows the picked epitopes for MHC class I and II. Every one of those epitopes has to stay non-allergenic, non-toxic, and fully antigenic.

Table 4.2: B cell Epitopes

Epitope	**Position**	**Antigenicity**
SANPSSTKDEKVKR	560-573	2.0660
DNNGPTDDGTTDATQ	595-609	1.6684
ENDKAESSKQSVKAKKTSKGKLLPKTG	855-881	1.2316



Average: 0.520 **Minimum:** 0.254 **Maximum:** 0.695

Figure 4.3: B-cell epitope prediction results. Scores greater than 0.5 appear in yellow color, which means they indicate a B-cell epitope. Scores less than 0.5 appear in green color, which means they indicate it is not a B-cell epitope.

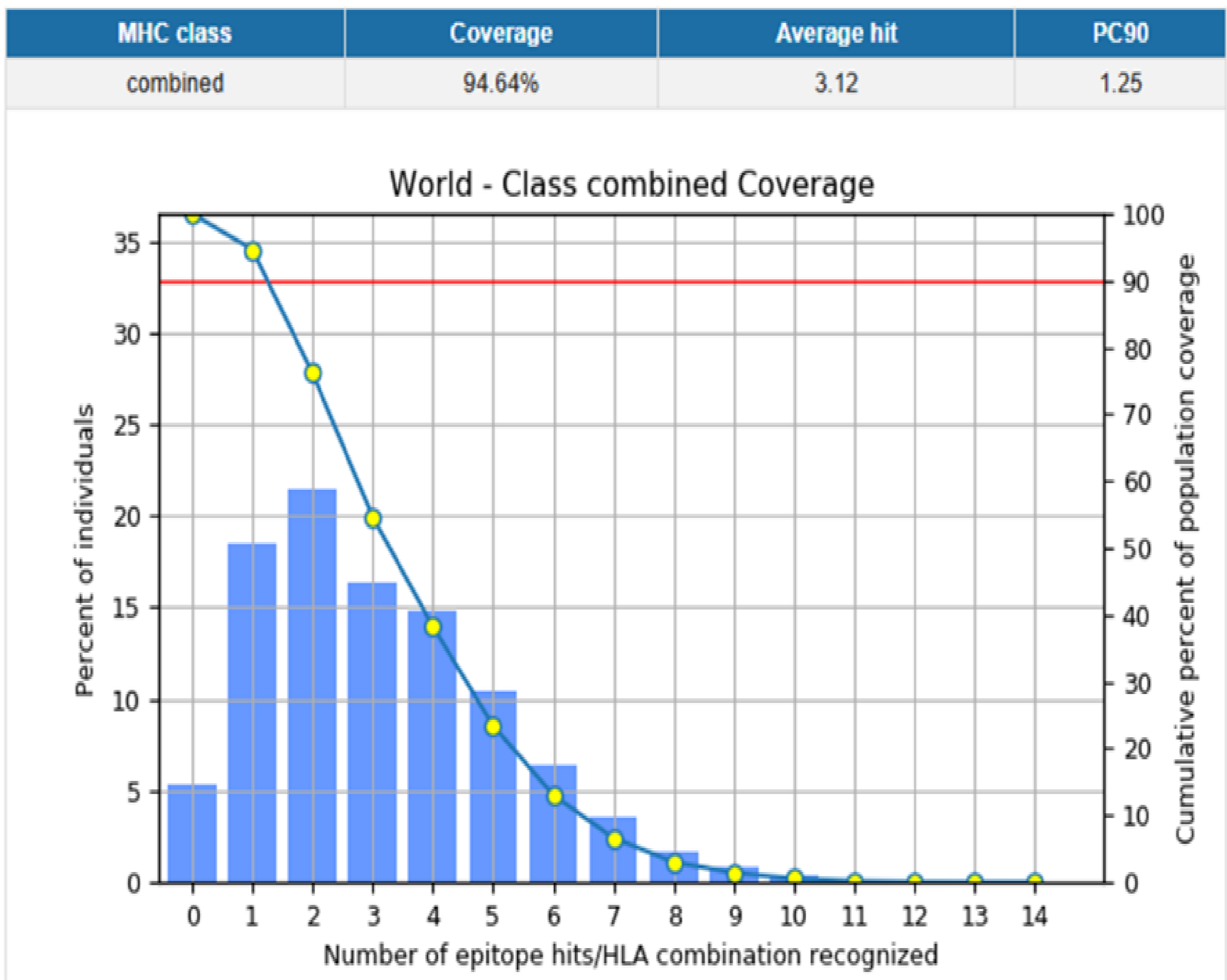
Table 4.3: MHC-I and MHC-II Selected Epitopes

Allele	**Peptide**	**ic50**	**Antigenicity**
MHC-I			
HLA-B*58:01, HLA-B*57:01, HLA-A*32:01, HLA-A*23:01	KSLAAEFIF	3.32, 8.7, 67.86, 193.76	0.8848
HLA-A*30:01	KAKKTSKGK	5.07	1.8439
HLA-A*02:03, HLA-A*02:01, HLA-A*02:06, HLA-A*68:02	NALYGRVQPV	8.39, 40.12, 176.28, 256.53	0.7484
HLA-A*30:01, HLA-A*11:01, HLA-A*03:01, HLA-A*31:01	KTSKGKLLPK	8.8, 20.1, 31.48, 170.34	0.6258
HLA-B*40:01	IEEGDMVNL	11.07	0.5454
HLA-B*07:02	KPKGGARDAL	13.25	1.1258
HLA-B*40:01	LETTTPSTL	14.09	0.7924
MHC-II			
HLA-DRB1*13:02, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB1*03:01, HLA-DRB1*12:01, HLA-DRB1*01:01	PANIIDNDGLRVFDP	3.9, 58.4, 203.7, 300.4, 371, 459.1	0.4576
HLA-DRB1*13:02, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB1*03:01, HLA-DRB1*12:01	ANIIDNDGLRVFDPE	4.3, 73.1, 220.3, 360.5, 383.9	0.4092
HLA-DRB3*02:02, HLA-DRB5*01:01, HLA-DQA1*05:01/DQB1*03:01, HLA-DRB1*13:02	IENFSANPSSTKDEK	4.4, 60.9, 139.5, 468.2	1.2975
HLA-DRB5*01:01, HLA-DPA1*02:01/DPB1*05:01, HLA-	LLVPILLLTKGKKES	6.6, 112.2, 15.7, 476.2, 221	0.6111

DRB1*11:01, HLA-DRB1*08:02, HLA-DRB1*01:01			
HLA-DRB3*02:02, HLA-DQA1*05:01/DQB1*03:01, HLA-DRB1*15:01, HLA-DRB1*04:01, HLA-DRB1*13:02, HLA-DRB1*04:05, HLA-DRB5*01:01, HLA-DRB1*01:01	SYNIENFSANPSSTK	9.9, 134.4, 211, 387.7, 258.6, 378.4, 232.5, 94.6	0.9641
HLA-DRB3*01:01, HLA-DRB1*03:01, HLA-DRB1*13:02	TSFLVRDDSGKSIVV	10.7, 124.9, 101.6	0.4796
HLA-DRB3*02:02, HLA-DRB5*01:01, HLA-DQA1*05:01/DQB1*03:01	ENFSANPSSTKDEKV	10.8, 91.5, 206.1	1.5309

4.3 Population Coverage Analysis

Researchers relied on the IEDB tool to carry out population coverage analysis. This approach evaluates the potential effectiveness of predicted epitopes in various human populations across the globe. Such an assessment supports the goal of broad immunological impact for the vaccine design. [20]. Coverage results for both MHC classes appear in **Figure 4.4.** They show worldwide coverage at 94.64 percent. Each person averages 3.12 epitope HLA hits. The PC90 value comes in at 1.25. In contrast, **Figure 4.5** represents the Pakistan-specific population coverage**,** showing a total coverage of 74.92%, an average of 1.09 epitope–HLA hits per person, and a PC90 value of 0.4**.**

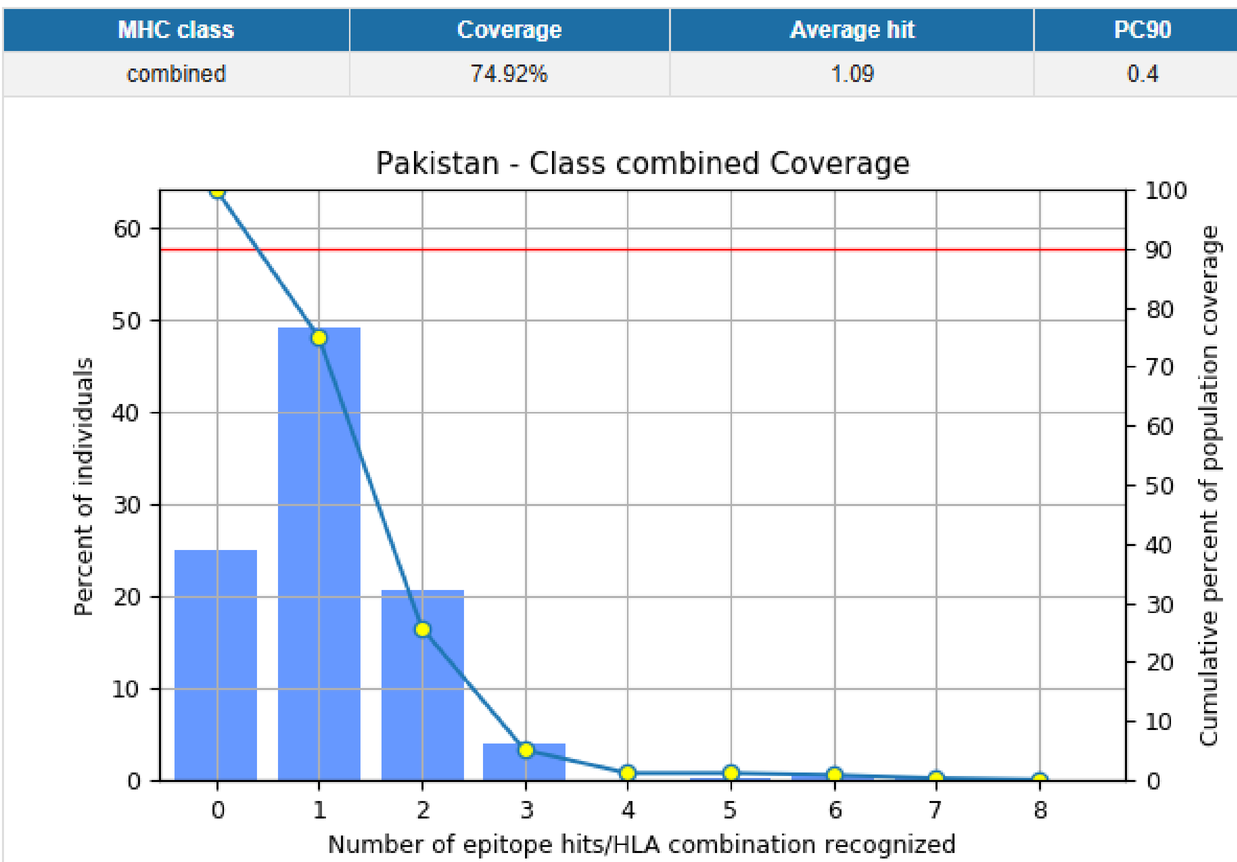


Population: World

MHC class	Coverage	Average hit	PC90
combined	94.64%	3.12	1.25

Figure 4.4: Population coverage calculation of MHC class combined world.

Population: Pakistan



MHC class	Coverage	Average hit	PC90
combined	74.92%	1.09	0.4

Figure 4.5: Population coverage calculation of MHC class combined Pakistan.

4.4 Construction of MEBSV

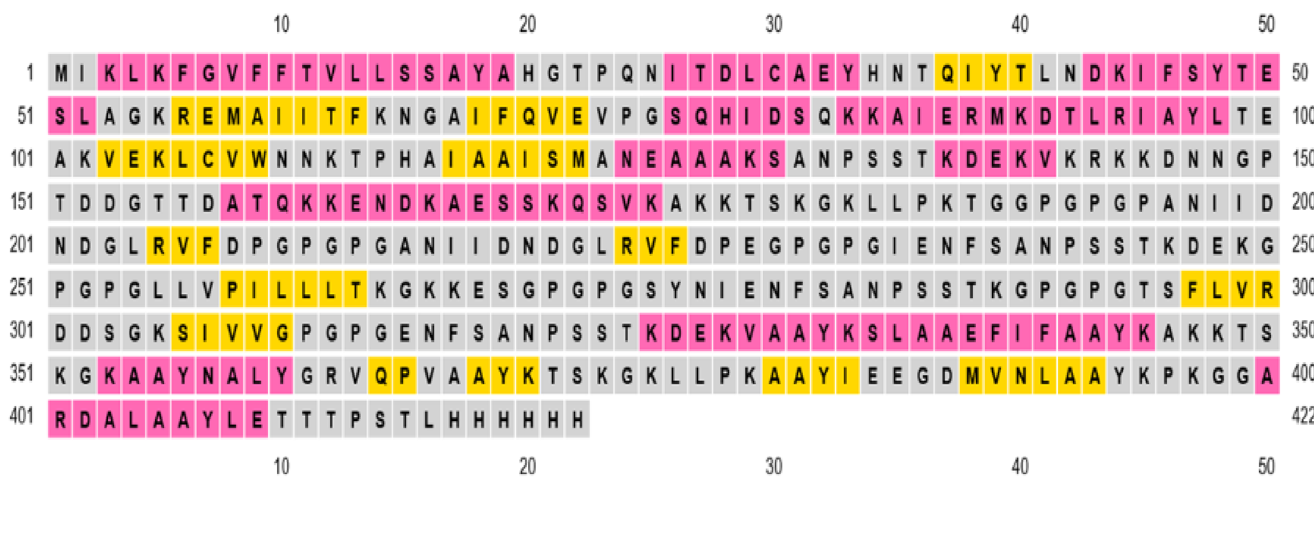
The MEBSV construct was designed to comprise several components in a defined order: an adjuvant at the N-terminal**,** connected via the EAAAK linker, followed by B-cell epitopes**, a** GPGPG linker**,** MHC class I epitopes, an AAY linker**,** MHC class II epitopes**,** and a 6× histidine tag at the C-terminal**.** The cholera toxin B subunit (CTB; UniProt ID: P01556) was chosen to serve as the adjuvant. **Figure 4.6** shows the MEBSV construct. Evaluation of MEBSV construct for antigenicity yielded a protective antigen score of 0.8469, categorizing it as a probable antigen. Subsequent predictions show that the construct was likely non-allergenic and non-toxic. Physicochemical properties depict that protein has 422 amino acids with molecular weight 44914.77. It contains a total of 44 negatively charged and 57 positively charged residues. The predicted half-life of the protein was approximately 30 hours in mammalian reticulocytes, more than 20 hours in yeast, and exceeds 10 hours in *Escherichia coli*. Also, the calculated instability index is 30.33, indicating protein stability. The protein has aliphatic index of 68.32, and GRAVY value of -0.577. Solubility was checked through tool Protein-Sol and SoluProt v1.0, and the predicted scaled solubility was 0.591 and 0.908.

MIKLKFGVFFTVLLSSAYAHGTPQNITDLCAEYHNTQIYTLNDKIFSYTESLAGKREMAIITFKNGAIFQVEVPG
SQHIDSQKKAIERMKDTLRIAYLTEAKVEKLCVWNNKTPHAIAAISMANEAAAKSANPSSTKDEKVKRKKDN
NGPTDDGTTDATQKKENDKAESSKQSVKAKKTSKGKLLPKTGGPGPGPANIIDNDGLRVFDPGPGPGANIID
NDGLRVFDPEGPGPGIENFSANPSSTKDEKGPGPGLLVPILLLTKGKKESGPGPGSYNIENFSANPSSTKGPGP
GTSFLVRDDSGKSIVVGPGPGENFSANPSSTKDEKVAAYKSLAAEFIFAAYKAKKTSKGKAAYNALYGRVQPV
AAYKTSKGKLLPKAAYIEEGDMVNLAAYKPKGGARDALAAYLETTTPSTLHHHHHH

Figure 4.6: MEBSV construct. Epitopes and linkers sequence were distinguished using different colours (Gray residues represent the CTB, pink residues indicate B-cell epitopes, dark blue residues correspond to MHC-I restricted (CTL) epitopes, orange residues denote MHC-II restricted (HTL) epitopes, black residues show linkers, and green residues mark the 6× histidine tag).

4.5 Structure Prediction, Refinement and Validation of MEBSV

Secondary structure was identified using PSIPRED program, which predicts the alpha, beta-sheet, and coil structures of MEBSV. 3-Dimensional structure model was created using I-TASSER webserver. **Figure 4.7** depicts secondary structure prediction, while **Figure 4.8** depicts I-TASSER's best model 1.



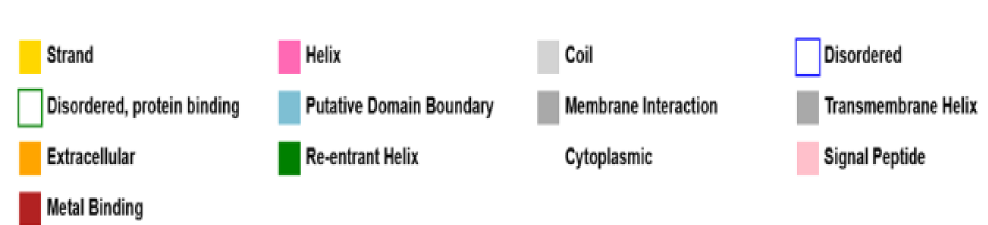


Figure 4.7: **Secondary Structure Prediction** (In the structural representation, yellow denotes beta strands, pink indicates alpha helices, and grey signifies coil regions. The protein structure comprises 26 coil segments (C), 14 beta-strand regions (E), and 11 alpha-helical regions (H)).

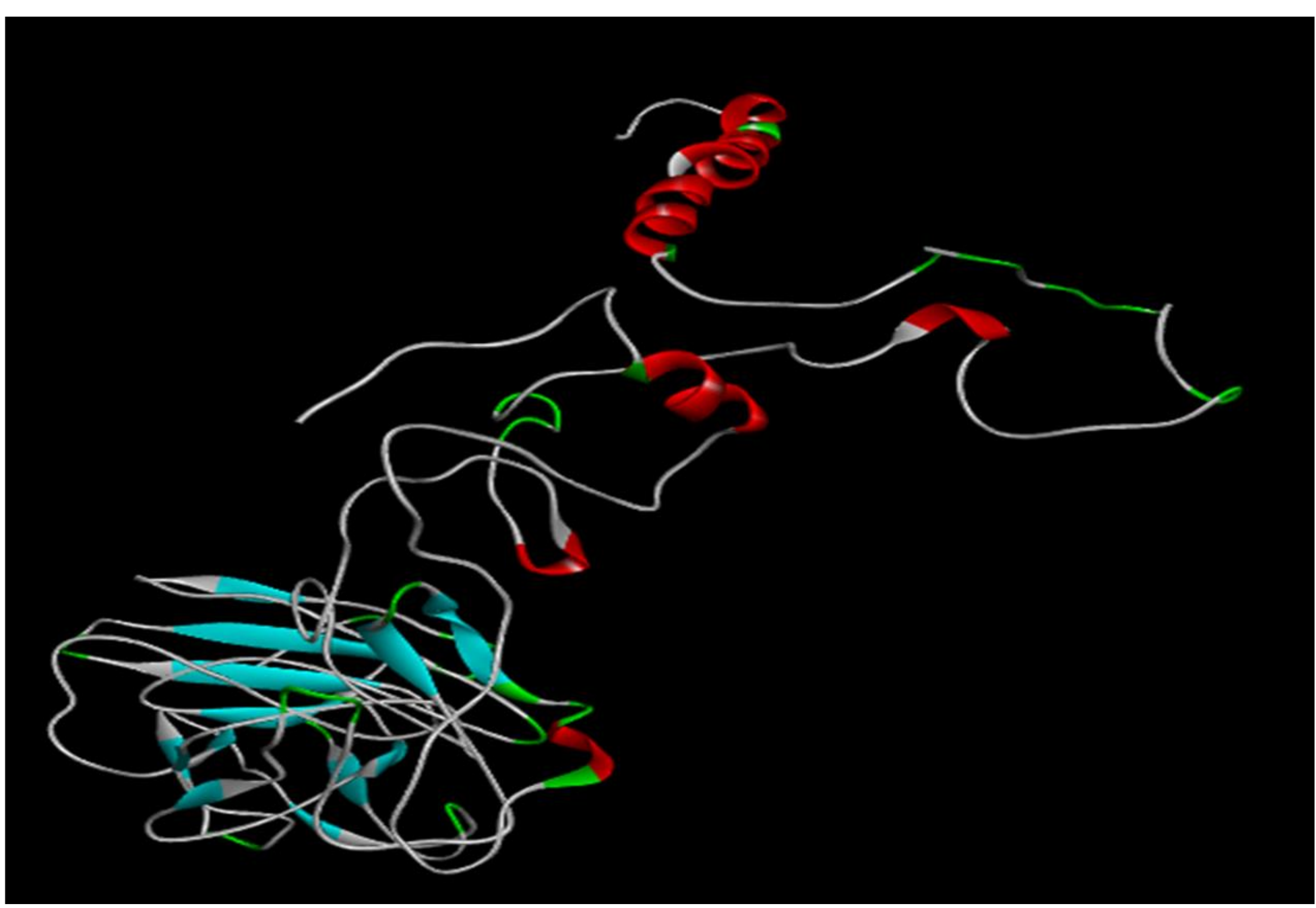


Figure 4.8: First model generated by I-TASSER was selected as it has best C value (-0.82) with TM value of 0.61±0.14 and RMSD value of 8.8±4.6Å.

The predicted structure from I-TASSER was further refined using GalaxyRefine, which improved the overall structural score. **Table 4.4** summarizes the structural details of the top five GalaxyRefine models. Among these, Model 1 was selected for further analysis. Ramachandran plot analysis of Model 1, performed using PROCHECK, indicated that 269 residues (77.1%) were located in the most favored regions, 65 residues (18.6%) in additional allowed regions, and 8 residues (2.3%) in generously allowed regions, while only 7 residues (2.0%) fell within disallowed regions. The refinement process boosted the share of residues in favored regions. It jumped from 53.0 percent to 77.1 percent. That really shows a solid gain in the protein's backbone geometry. **Figure 4.9** depicts the refined structure of the vaccine construct obtained through GalaxyRefine (Model 1). The corresponding Ramachandran plot is shown in **Figure 4.10**, illustrating the stereochemical quality of the refined structure. Finally, **Figure 4.11** presents the ProSA-web validation of the model, confirming its overall structural reliability.

Table 4.4: Structure Information of top five models GalaxyRefine

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.097	5.3	16.9	65.2
1	0.9159	0.492	2.286	14.9	0.9	87.6
2	0.9129	0.496	2.320	15.1	0.0	86.2
3	0.9064	0.503	2.335	15.9	0.6	86.4
4	0.9117	0.506	2.315	15.9	0.6	87.4
5	0.9153	0.498	2.326	15.7	0.3	86.7

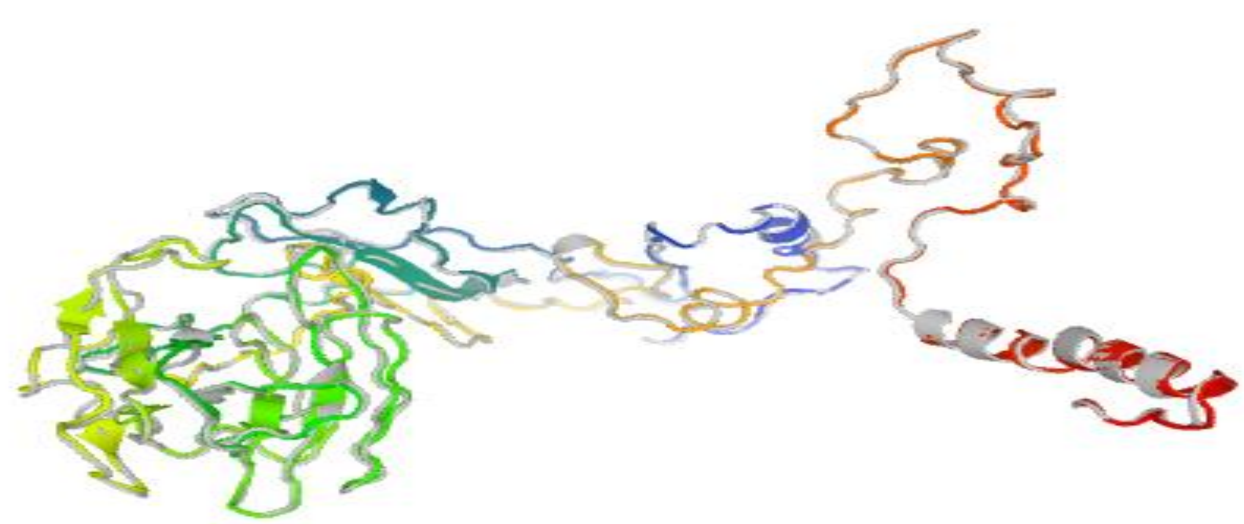


Figure 4.9: Refined structure of the vaccine construct generated using GalaxyRefine (model 1).

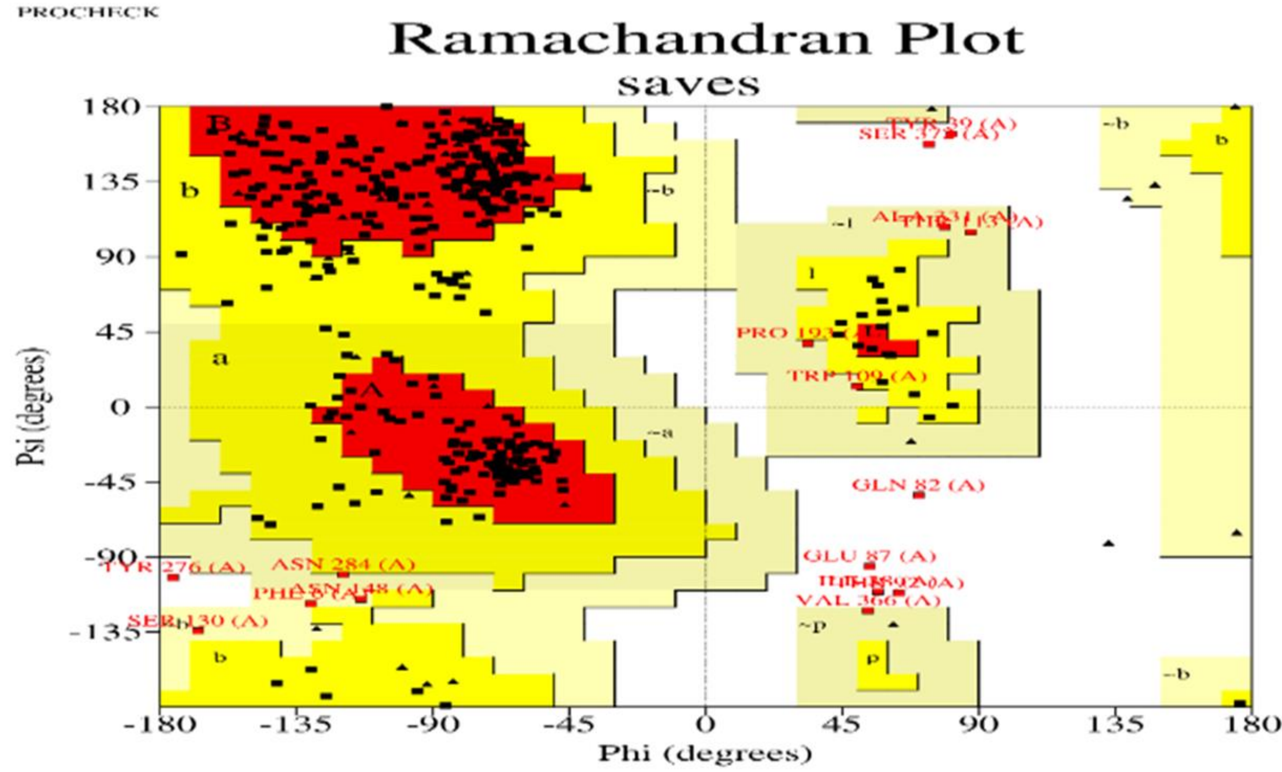


Figure 4.10: Ramachandran plot of improved MEBSV 3D structure (Red is most favourite region and yellow is less favourite region, as mostly balck dots in red region means our structure is right predicted and score increased).

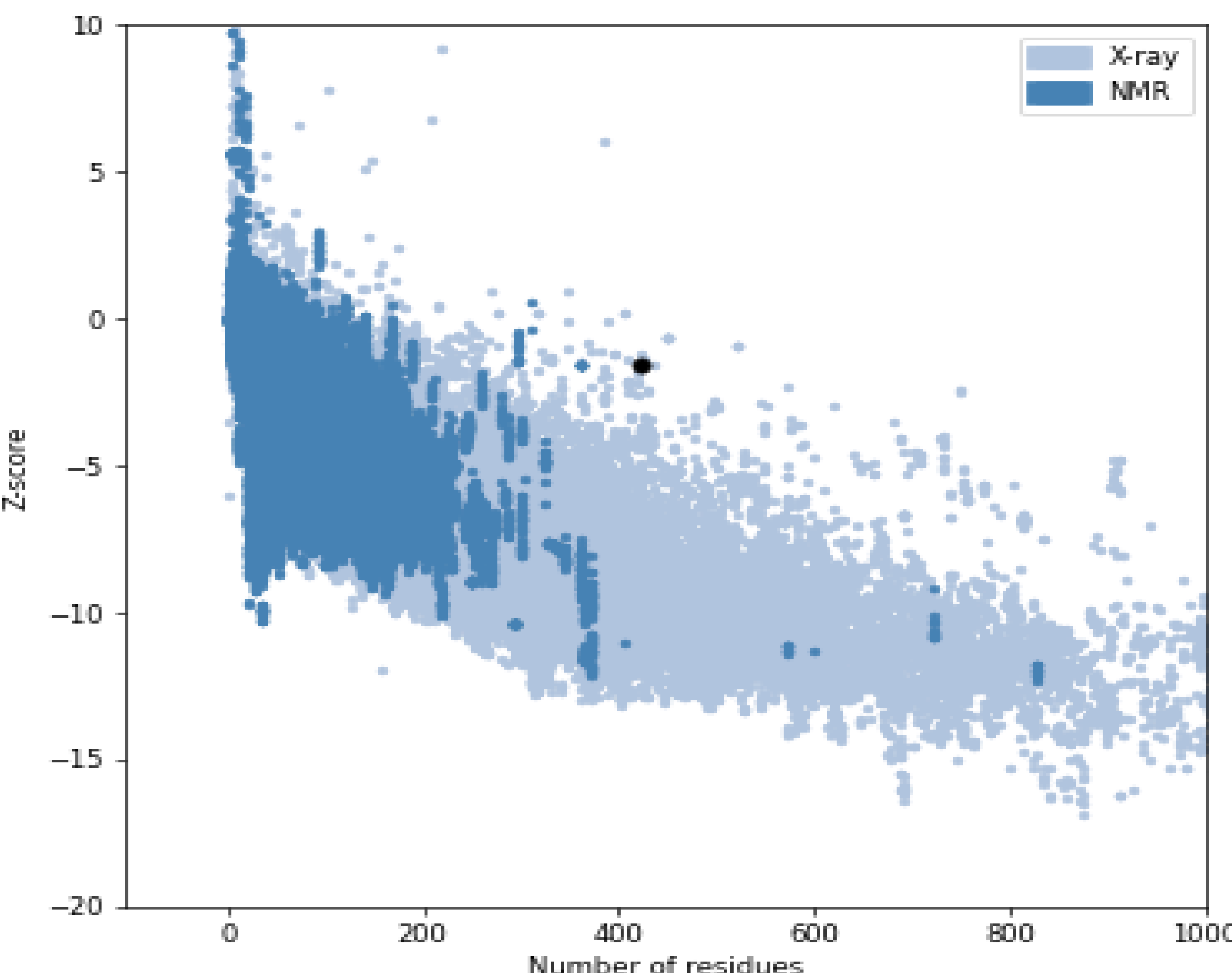


Figure 4.11: Final vaccine model validated by Prosa-Web.

4.6 Docking

TLR2 (PDB ID: 2Z7X) was selected as the receptor and downloaded from the RCSB Protein Data Bank, while the designed vaccine construct was used as the ligand. Molecular docking was performed using HDOCK to investigate potential interactions between the receptor and ligand. **Figure 4.12** illustrates the docking results obtained from HDOCK. The top-ranked docking model was chosen for detailed analysis, displaying a docking score of –331.22, a high confidence score of 0.9740, and a ligand RMSD of 145.63 Å. **Table 4.5** summarizes the docking scores of the top five models. The interactions and bonds of the best docking model (Model 1) are visualized in **Figure 4.13** using Discovery Studio. Additionally, **Table 4.6** lists the predicted hydrogen bonds and salt bridges of the docked complex, as determined by PDBePISA.

Table 4.5: Docking score of top five models

Rank	**1**	**2**	**3**	**4**	**5**
Docking score	-331.22	-324.05	-321.10	-315.84	-309.39
Confidence score	0.9740	0.9701	0.9684	0.9650	0.9604
Ligand rmsd (Å)	145.63	137.09	132.28	125.89	120.46

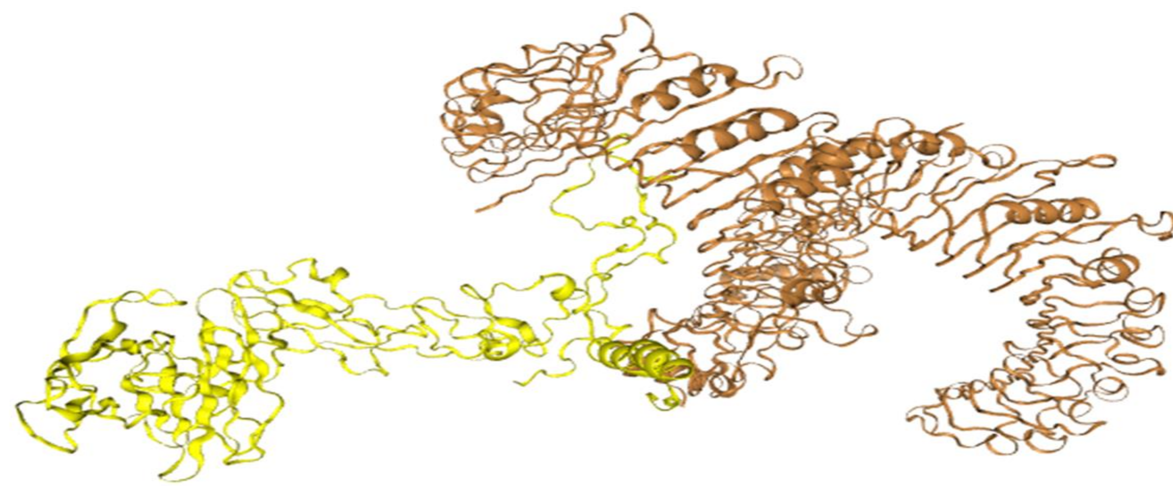


Figure 4.12: HDOCK result page depict the structure of docked molecule (receptor molecule is shown in brown colour and yellow colour shows ligand).

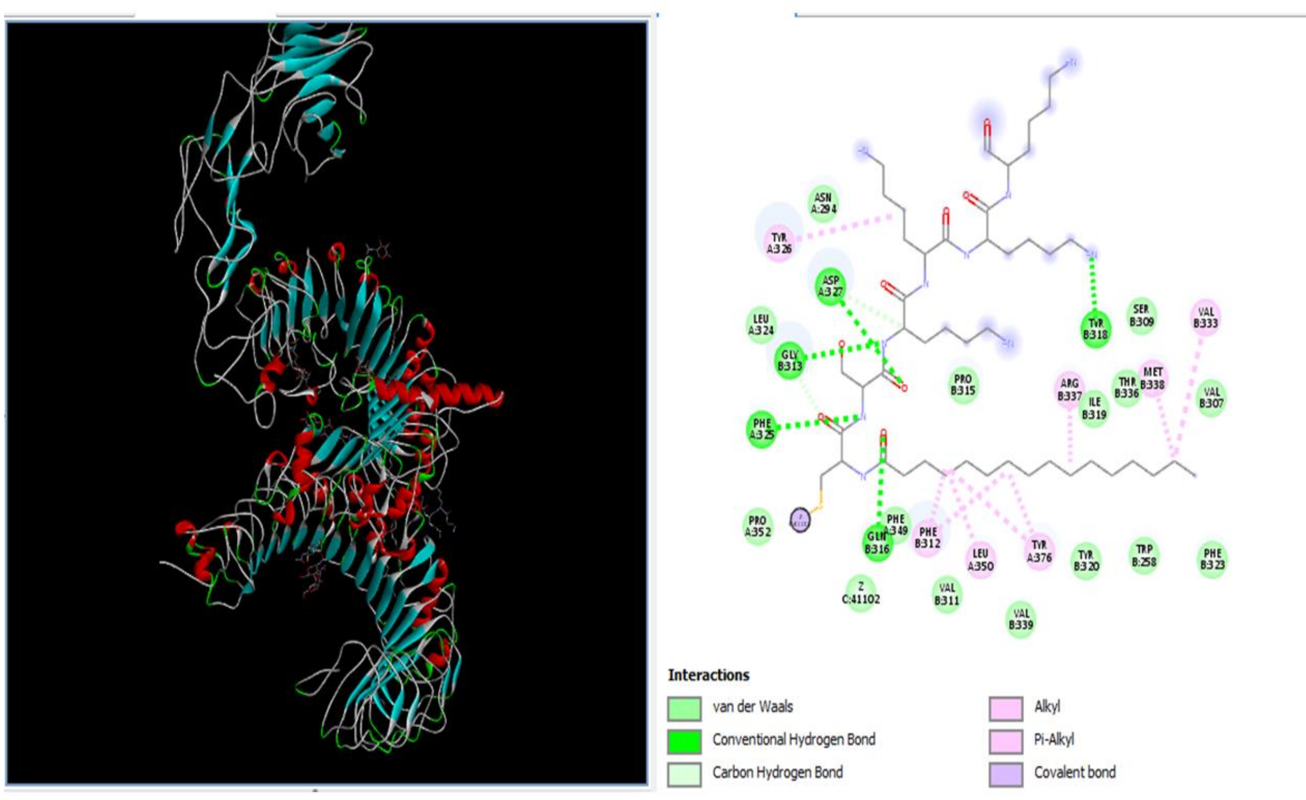


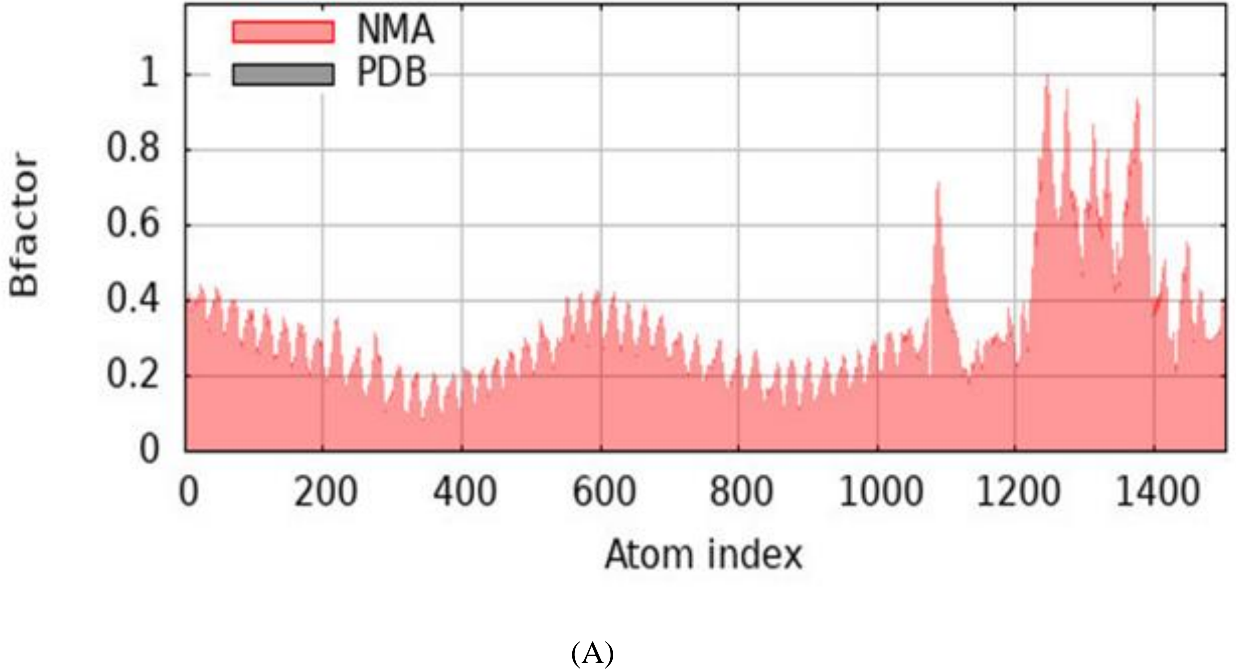
Figure 4.13: Best dock model was 1 so it was downloaded from HDOCK and its interaction and bond seen on discovery studio.

Table 4.6: Predicted Hydrogen bonding and Salt Bridges of docked vaccine via PDBePISA

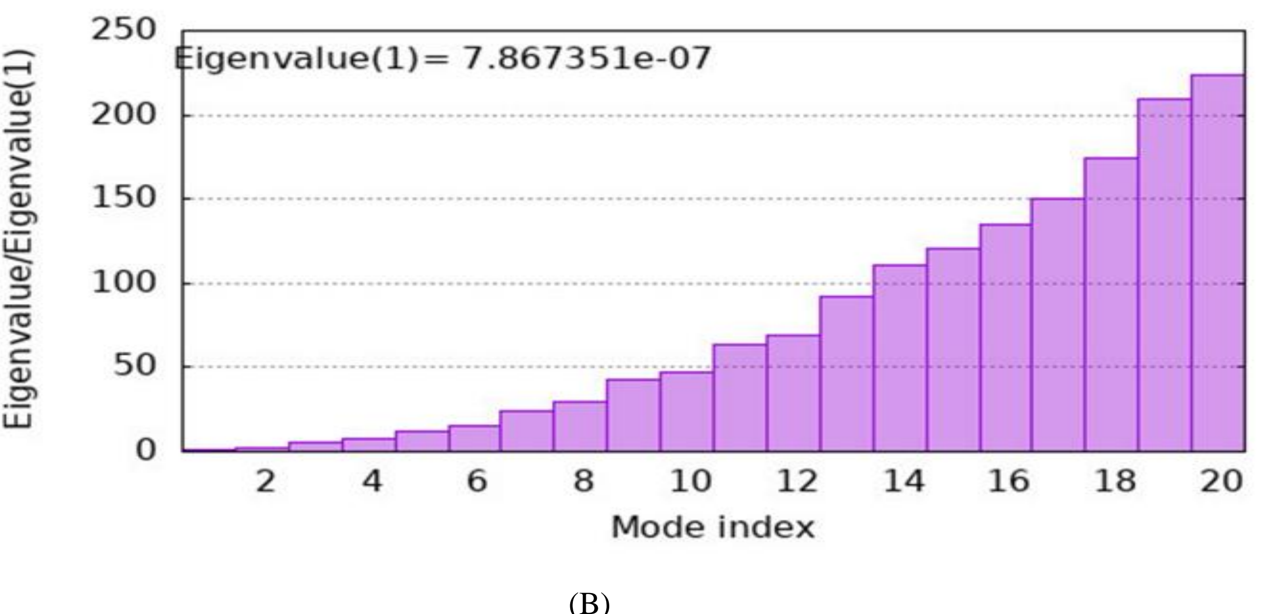
Interaction Type	Structure 1 (Chain A)	Bond Length [Å]	Structure 2 (Chain B)
Hydrogen Bonding	LYS 347 [NZ]	2.94	THR 361 [OG1]
	LYS 347 [NZ]	3.18	THR 363 [OG1]
	TYR 376 [OH]	2.78	SER 309 [O]
	GLN 396 [NE2]	3.41	SER 409 [OG]
	ASN 345 [OD1]	2.85	LYS 385 [NZ]
	GLU 369 [O]	3.31	LYS 385 [NZ]
	GLU 374 [OE1]	3.86	ARG 337 [NH2]
	GLU 375 [OE1]	2.89	ARG 337 [NH2]
	GLU 375 [OE2]	2.81	ARG 337 [NE]
Salt Bridges	GLU 374 [OE1]	3.86	ARG 337 [NH2]
	GLU 375 [OE1]	2.89	ARG 337 [NH2]
	GLU 375 [OE1]	3.35	ARG 337 [NE]
	GLU 375 [OE2]	3.72	ARG 337 [NH2]
	GLU 375 [OE2]	2.81	ARG 337 [NE]

4.6.1 Energy Minimization and Interaction Improvement

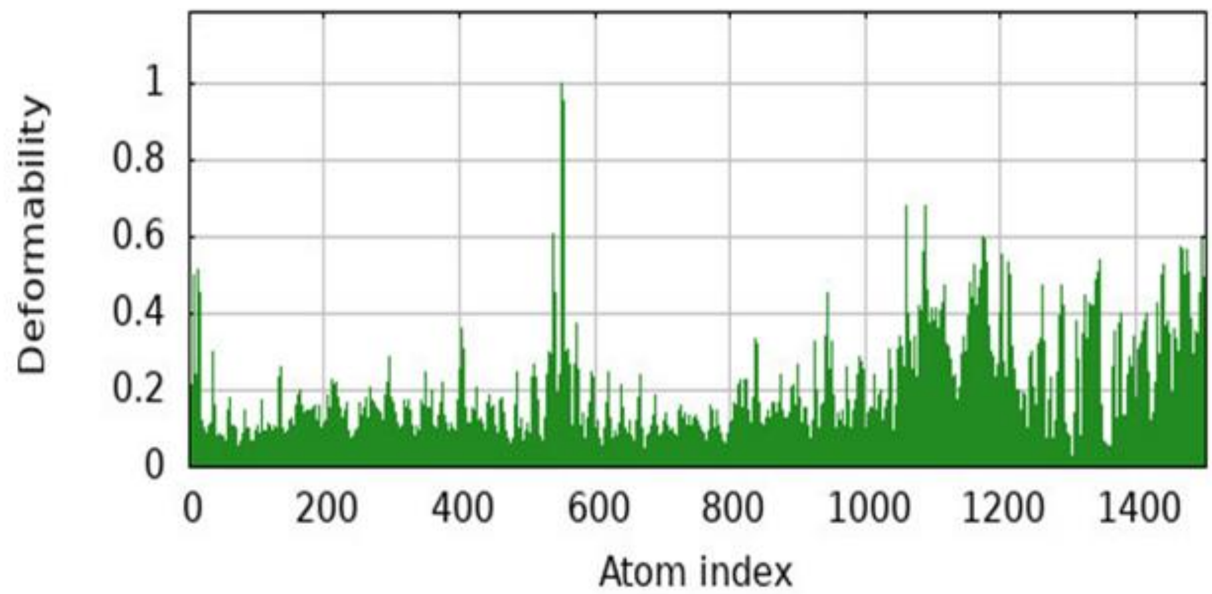
iMODS was employed for energy minimization and to enhance the interactions within the vaccine–receptor complex. The structural flexibility and stability of the complex were then examined using the iMODS server, which applies molecular dynamics-based normal mode analysis (NMA). This computational approach evaluates the internal motions of protein complexes to predict their conformational behavior. The analysis generated several outputs, including the deformability plot, covariance matrix, elastic network model, B-factor values, and eigenvalue, which together provide a comprehensive assessment of the dynamic stability and flexibility of the modeled complex [35], [62]. To bring in those optimised outcomes, the completed docking interaction model was uploaded. Graphs covering energy minimisation and the gains in molecular interactions shown in **Figure 4.14**.



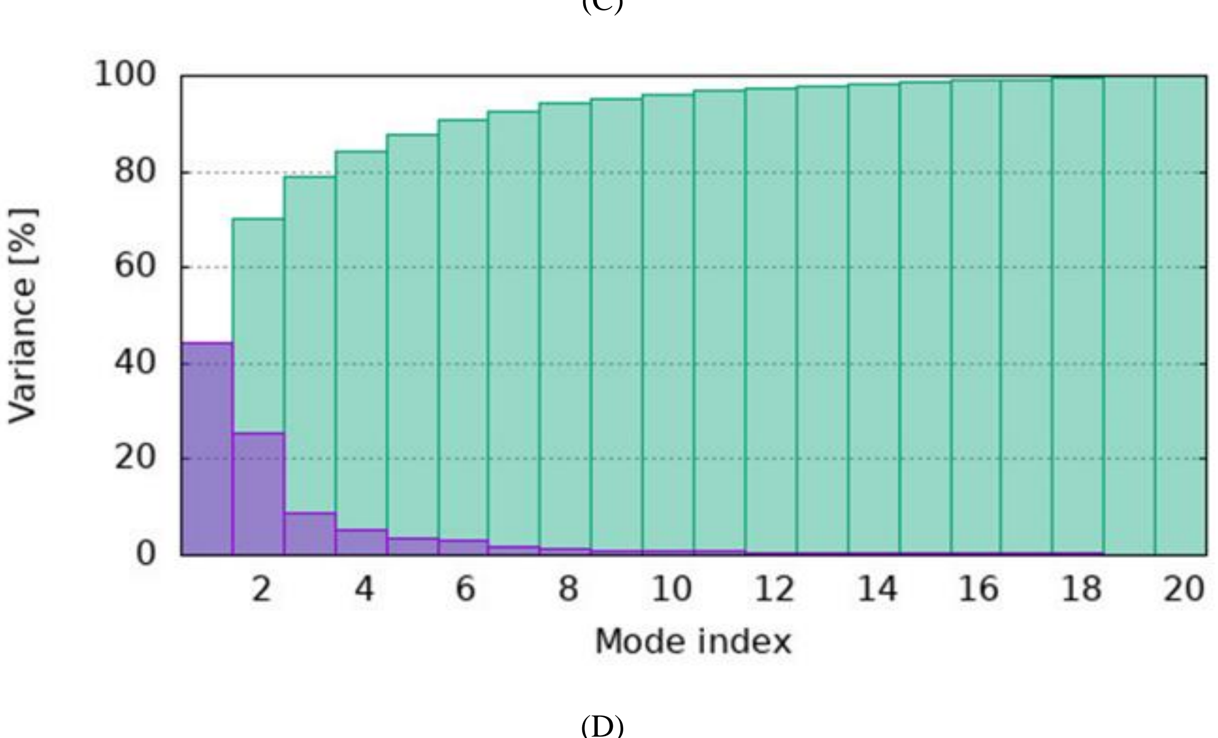
(A)



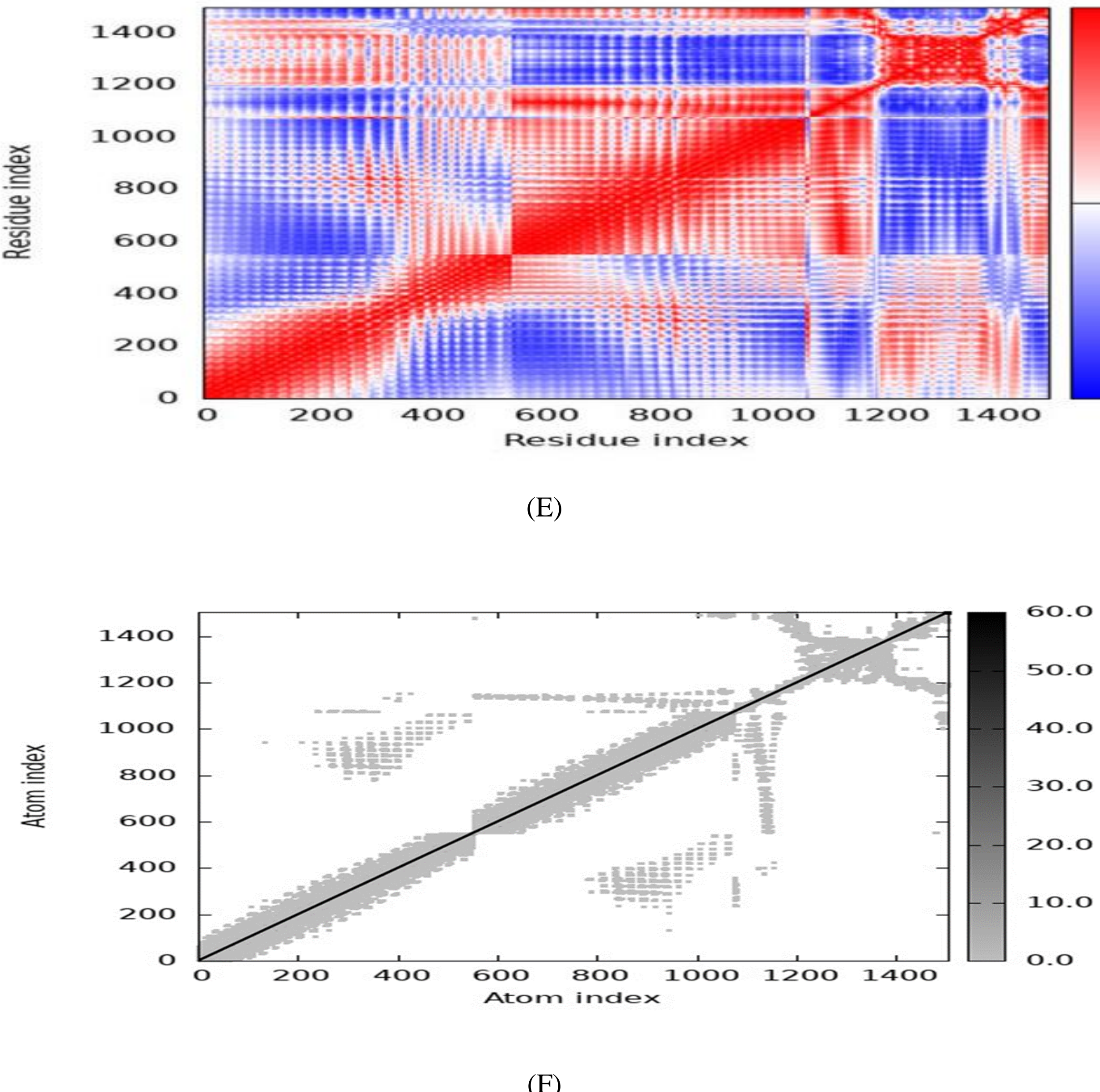
(B)



(C)



(D)



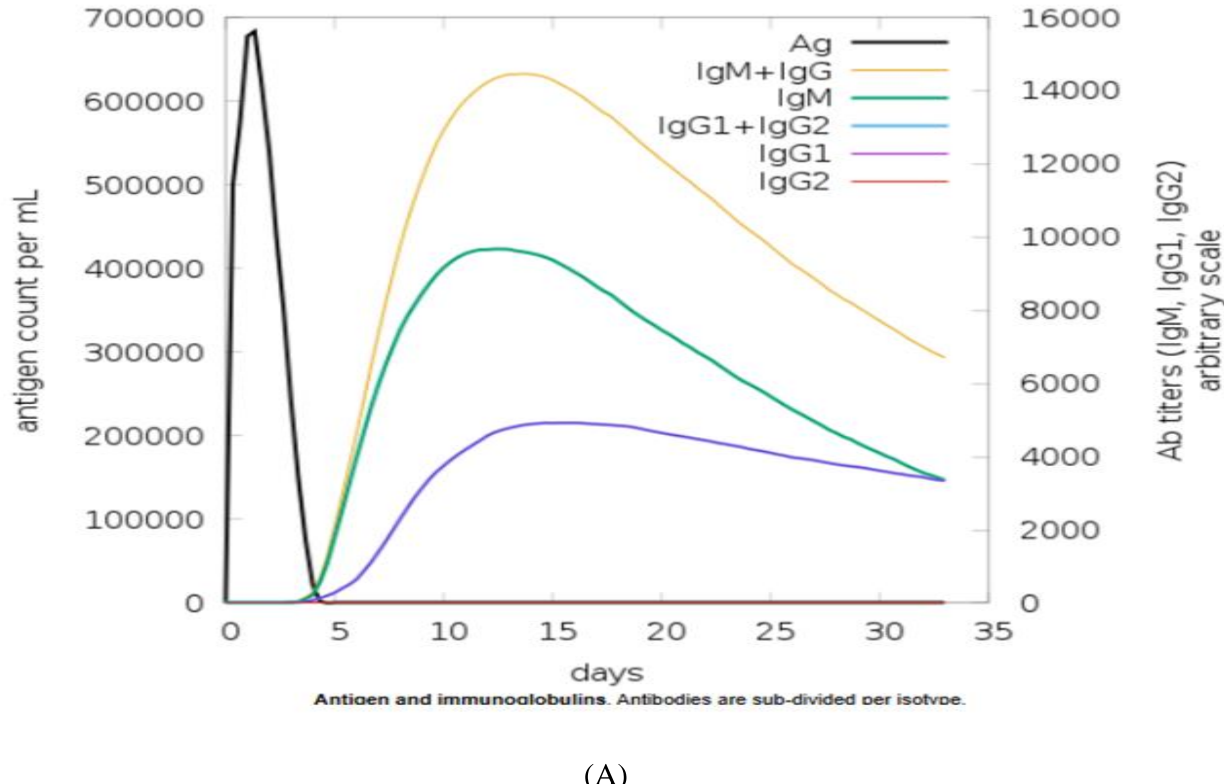
(E)

(F)

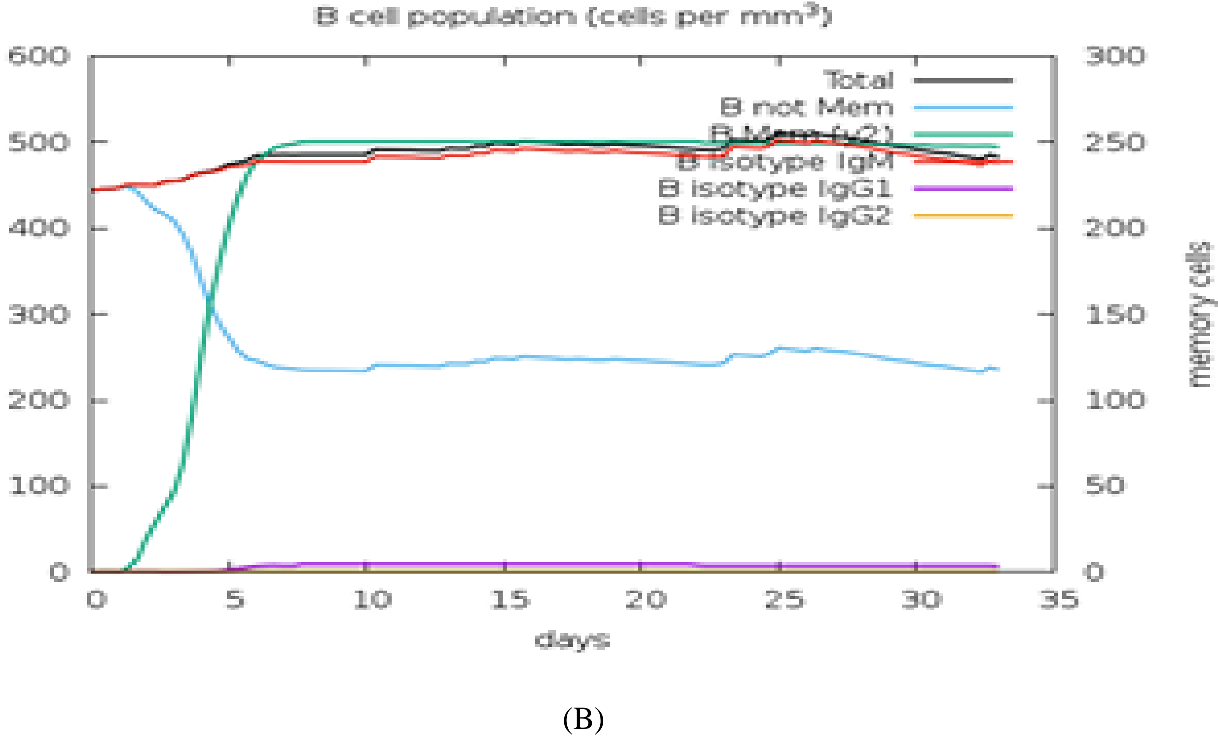
Figure 4.14: Mobility and the B-factor are covered in (A). (B) examines eigenvalues. Deformability is covered in (C). Variance is shown in (D). Individual differences are represented by the purple bars there. Cumulative variances are represented by green ones. A covariance matrix is denoted by (E). It highlights interactions between pairs of residues. Correlated motions in red are among them. White ones are uncorrelated. Blue indicates anti-correlated motions. An elastic network is shown in (F). Its dots are coloured according to their stiffness. Stiffer springs are indicated by darker greys. More flexibility is indicated by lighter hues.

4.7 Immune Simulation

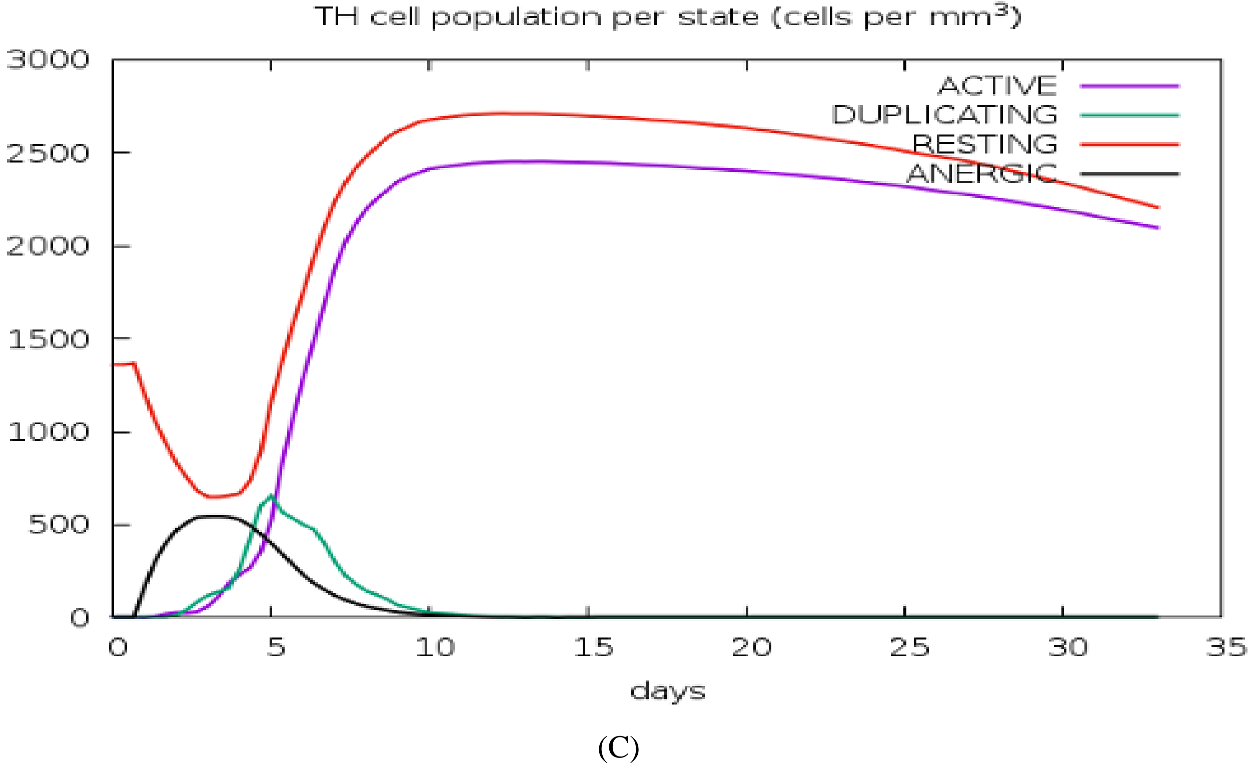
C-ImmSim simulates adaptive immune responses in order to function. Position-specific scoring matrices, or PSSMs, are used to model the interactions between a pathogen and its host. The tool provides thorough predictions for the primary immune components. These include helper T cells, plasma B-cell populations following antigen exposure, and activated B and T lymphocytes. When evaluating a vaccine construct's immunogenic potential, this type of simulation proves to be very helpful. It facilitates optimisation and identifies areas for further experimental investigation [2]. The server was used by researchers to simulate the host immune response to the vaccine construct that was created. Primary, secondary, and tertiary responses all had an impact on the immune responses that manifested. The total amounts of IgM and IgG antibodies clearly increased. The simulation also looked at the production of interleukins and various cytokines. All in all, these findings show the vaccine construct has solid antigenicity and immunogenicity. You can see the results laid out in **Figure 4.15**.



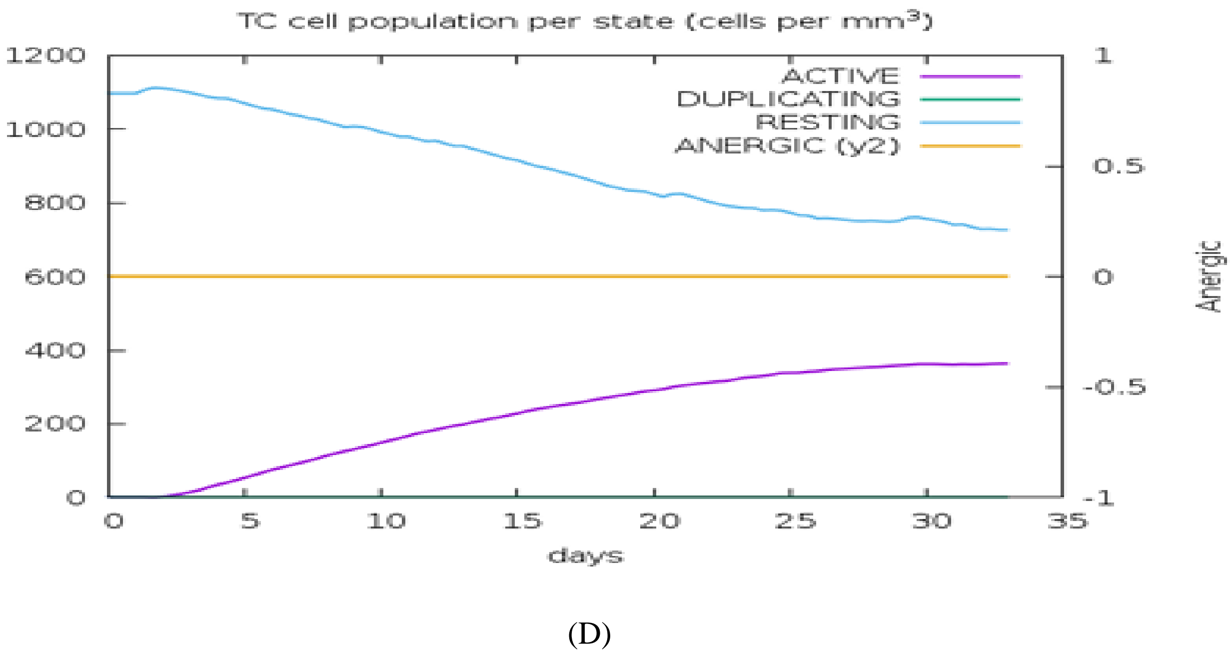
(A)



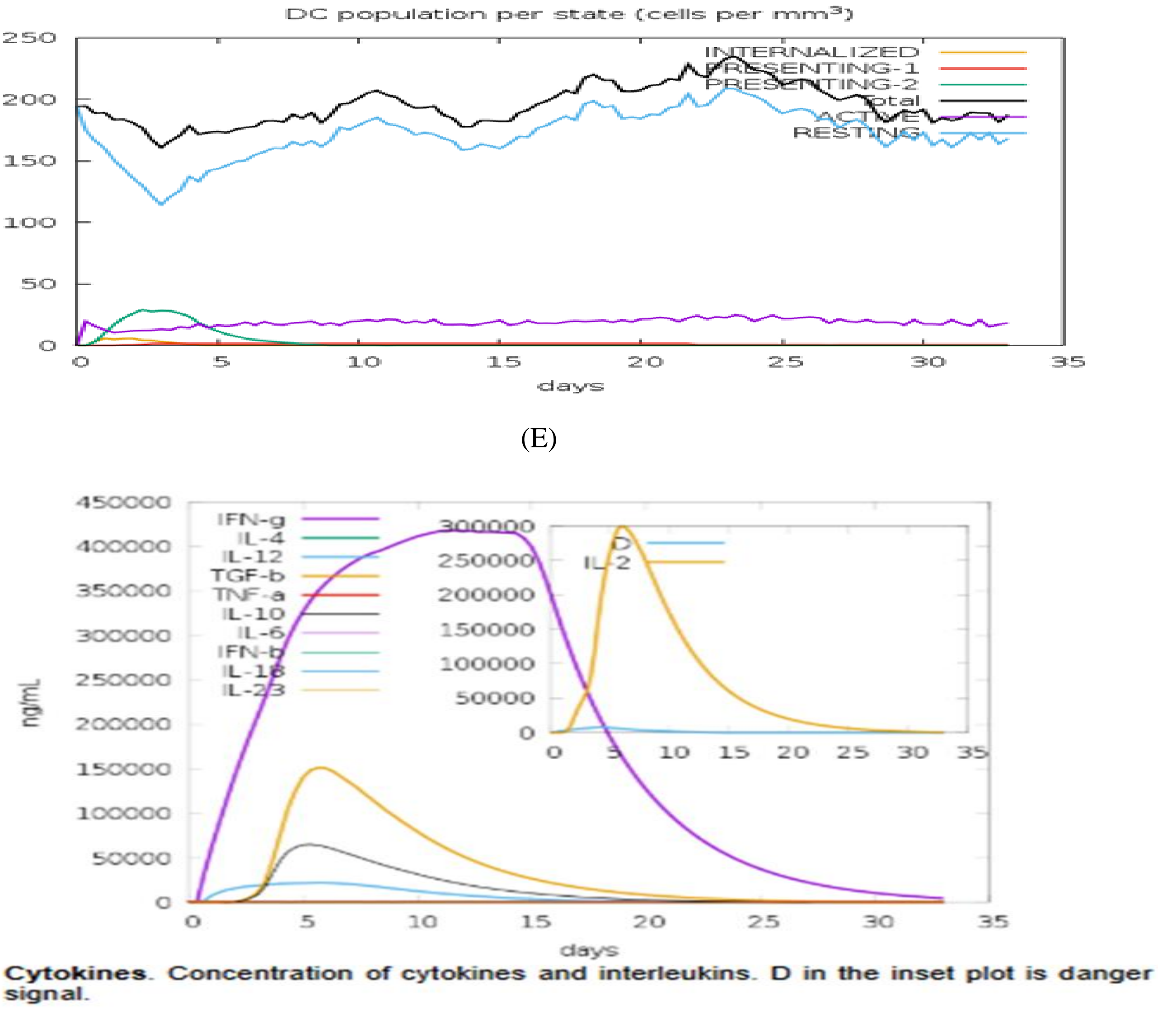
(B)



(C)



(D)



(F)

Figure 4.15: In silico immune simulation. They ran it on the C-ImmSim server to illustrate the immune response from the designed vaccine construct. (A) covers the generation of immune complexes and immunoglobulins right after vaccine administration. (B) presents the total B lymphocyte population following three successive injections. (C) depicts the expansion of CD4 plus T-helper cells. It includes active states, duplicating states, resting states, and anergic states. (D) represents the increase in cytotoxic CD8 plus T-cell numbers after vaccination. (E) highlights the proliferation of dendritic cells.Finally, (F) illustrates the stimulation and secretion patterns of key cytokines and interleukins induced by the vaccine.

4.8 Codon optimization and Molecular cloning

Codon optimization was performed to improve the efficiency of gene expression in the host organism. This approach helps overcome variability in the expression of foreign genes that arises from differences in mRNA codon usage [15]. Codon optimization was carried out using Vector Builder, adjusting the codon usage to match that of E. coli K12. This process generated a nucleotide sequence of 1,269 bases, with a Codon Adaptation Index (CAI) of 0.94 and a GC content of 53.51%. For cloning purposes, SnapGene software was used to insert the optimized sequence into the multiple cloning site (MCS) of an E. coli expression vector. The pET-28a(+) plasmid, which is 5,369 base pairs in length, served as the cloning vector. The codon-optimized sequence was positioned between the NdeI and FspI restriction sites, and overhangs were removed using SnapGene tools. The final recombinant construct measured 3,631 base pairs. **Figure 4.16** illustrates the vector, inserted fragment, and final product, while **Figure 4.17** presents the cloned vaccine construct.

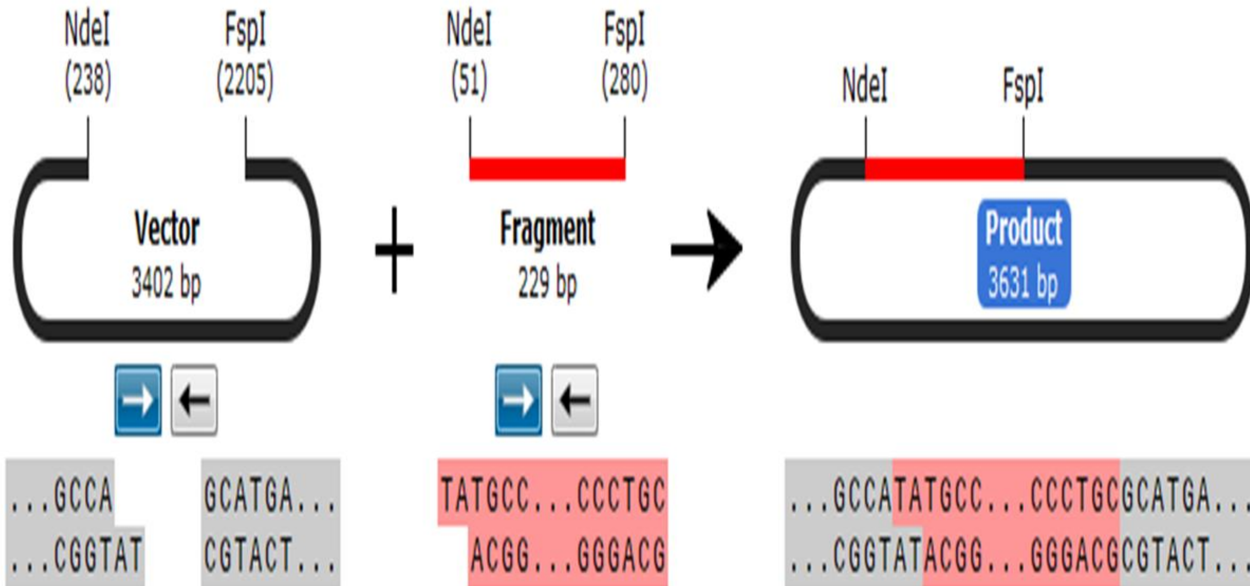


Figure 4.16: *In-silico* Cloning of MEBSV (vector and the fragment along with the product).

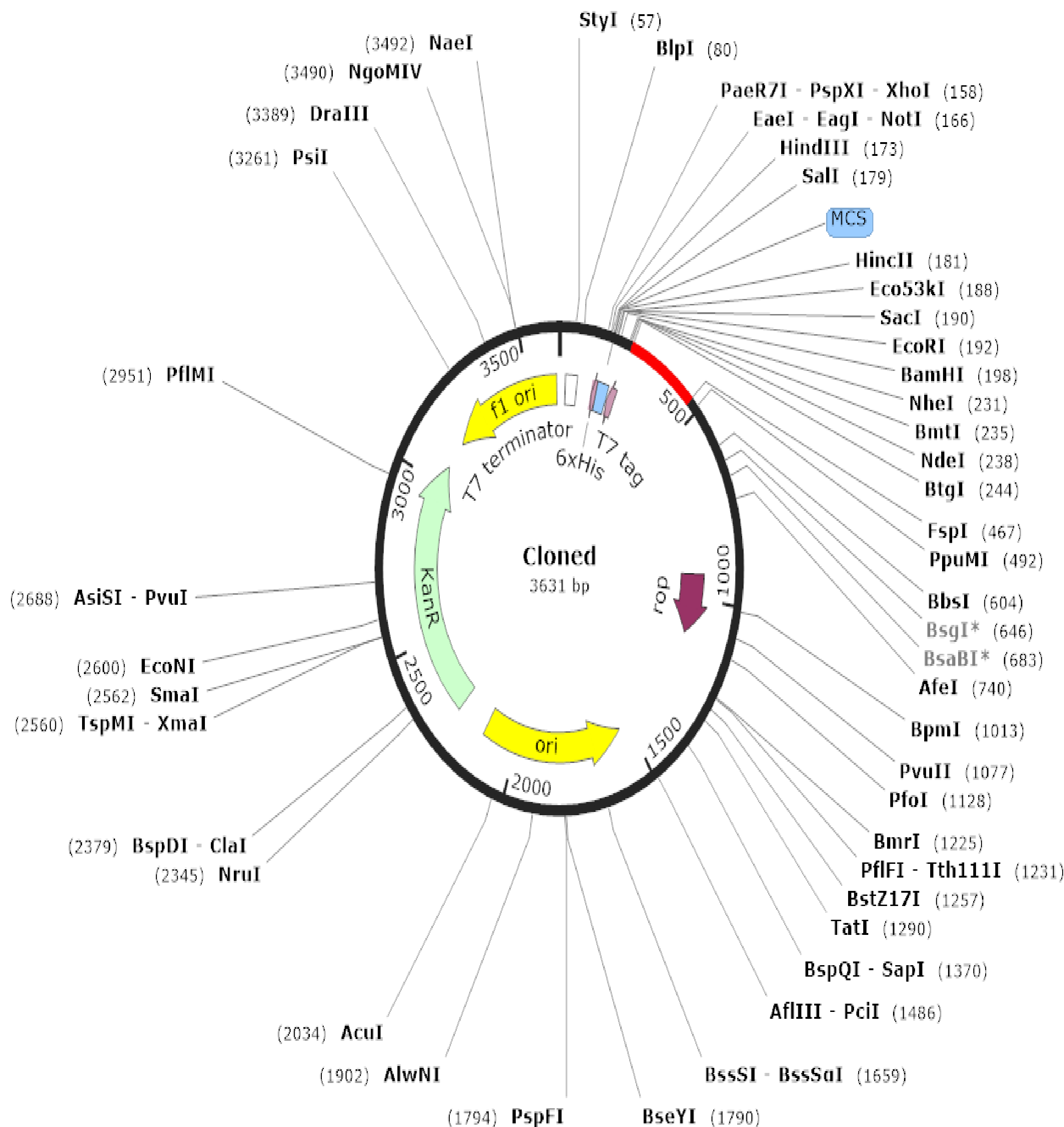


Figure 4.17: Vaccine candidate's codon-optimized gene sequence (shown in red) was cloned between NdeI and FspI restriction sites, within pET28a(+) vector, represented by a black circular structure.

Chapter 5

Discussion

5.1 Discussion

The MEBSV came together through some pretty advanced computational methods. Those methods helped spot and pick out the top immunogenic epitopes from proteins in *S. pyogenes* strains. Things like the rising global load of infections from *S. pyogenes*, plus no approved vaccine out there yet, really push the need for safe and solid ways to prevent these issues. The design of vaccines has been significantly altered by immunoinformatics. It enables researchers to accurately predict and verify epitopes. The screening of both T and B cell epitopes carried out through this pipeline is more efficient, cost-effective and returns faster results as compared to conducting traditional laboratory studies. This also confirmed that the final antigen will provide a broad and robust immune response from the end-user along with maintaining the integrity of the safe and stable antigen. B cell epitopes (or epitopes that bind to antibodies produced by the human host) are critical areas on an antigen by which the humoral immunity initiates [21]. The vaccine can create antibodies that will destroy the infection should they be accurate. Helper T lymphocytes or HTLs are involved in many aspects of adaptive immunity. They aid B cells to create antibodies. They also promote the cytotoxic T lymphocytes to kill the infected cells [22]. It is important that these epitopes are brought in to total protection. We decided to use HTL and CTL epitopes for our study depending on their antigenicity and ability to cover different populations. The vaccine successfully stimulates a number of immune pathways in this manner. To increase the immune punch and overall response we added an adjuvant right at the N-terminal end of vaccine [18], [23].

The cholera toxin B subunit or CTB was chosen for this purpose. It is notable for its disease-free and nontoxic properties and its ability to improve systemic and mucosal immunity. The presence of this is commonly seen during live vaccination tests. Its excellent record as safe booster has also been demonstrated by the fact that Canadian and European health authorities have approved [24]. But the structural integrity of the vaccine was maintained by careful construction. The adjuvant and first B-cell-epitope are linked by an EAAAK linker. That one keeps domains from interfering excessively and provides rigidity. Then B-cell epitopes are chained up using KK linkers. HTL epitopes attach themselves to GPGPG linkers to attach to the final B cell location. The end of HTL is joined to CTL epitopes by AAY linkers. There are 422 amino acids in all. No coincidence in the selection of those linkers. There was sound science to that. EAAAK types are routinely used in computer-based vaccine types. They maintain the separation of domains and help in fold and steadiness [25]. The linkers KK, GPGPG and AAY, on the other hand, remain hydrophilic. This

enhances cross-domain solubility, flexibility & folding. The GPGPG one helps in finding out the HTL epitopes. Furthermore, it makes them more susceptible to the immune system cells. AAY correctly regulates antigen processing. It prevents the formation of the junk epitopes at joints which could interfere with immune targeting [16], [19]. So this vaccination programme should be carried out on a large population. Based on predictions across MHC alleles, however, the estimated global reach is approximately 94.64 percent. Stability and high immunogenicity were observed during its tests, related to physics. However, the molecular weight was maintained below 110 kDa. This makes it easier to express and clean up in setups with E. coli. Less than 40 was the index of instability. As a result, the protein exhibits structure well [26]. Increase in GRAVY score was also observed to decrease. This would indicate that it is usually hydrophilic in nature. Due to these traits it is able to adapt well in watery environments [27]. Solubility projections supported that. The vaccine should proceed to dissolve and prove its function under the normal physiological conditions. The Aliphatic Index was greater than 50. That indicates that it must be highly resistant to heat. It is important for maintenance of quality of the vaccine during transit and storage [28]. Solvent access predictions were greater than 55 percent. As a result, epitopes probably remain on the surface and wait for immune system capture [29]. Strong binding between the vaccine and that receptor were shown by TLR2 docking runs. The scores appeared to be good, and interaction patterns remained the same. All this suggests that the construct is indeed able to utilize TLR2 to stimulate innate immunity. Immediately after this, it leads to adaptive responses. These strong relationships indicate that the vaccine is a good immunogen. It is able to support antibody as well as cell-based immunity.

The codons were optimized to obtain smooth expression in E.coli. The sequence was then transferred into the pET-28a(+) vector of the computer [30]. The CAI and GC content fit the sweet spots. This suggests that there are no issues with translation or production that takes place within the bacteria. At this stage, the links between design and the actual work in the lab takes place. It demonstrates the fact that the product can be made effectively. Despite these strong computational and structural findings, the work has limitations. So everything was done using in silico tools. They are strong despite lacking the intricacy of real biology. For this reason, in vivo animal tests and in vitro testing are important. These will reveal the antigenicity, safety and immunological response that was anticipated by us. In addition, they will try to determine if that develops real protection and works in living systems. In conclusion, this work offers a comprehensive

computational roadmap for a reliable, highly immunogenic vaccine that gives widespread defence against *S. pyogenes*. Adjuvants, important epitopes and tuned linkers are combined to target the major physical and immune targets. The results show how immunoinformatics expedites procedures and lowers lab expenses. To perfect protection, the next steps should include animal testing, additional docking with immune players, and lab checks. When combined, this suggests that the construct is a promising candidate for the *S. pyogenes* vaccine. It advances the field of computational vaccine research.

Chapter 6

Conclusion

6.1 Conclusion

The MEBSV that we designed in this work stands out as a promising option. It is logically constructed to aid in preventing *S. pyogenes* infections. At the moment, no approved vaccine is available for this pathogen. We relied on immunoinformatics along with computational methods to pick out the most conserved epitopes. Those epitopes also proved highly immunogenic. After that, we integrated them into a single overall construct. Three B-cell epitopes, seven CTL epitopes, and seven HTL epitopes are all present in that construct. These are all linked to different HLA alleles. We anticipate that both humoral and cell-mediated immunity will be triggered by this arrangement. It should also provide coverage for a large population.

The vaccine's antigenic qualities were good. Its physicochemical characteristics were also appropriate. It seems to be non-toxic and non-allergic. Furthermore, there is minimal chance that it will result in autoimmunity. Its low resemblance to human proteins is the reason for this. Strong stability and solubility were confirmed by our physicochemical tests. Additionally, the molecular properties appear favourable. For expression and formulation to be effective, these factors are crucial. The GRAVY value was negative. Hydrophilicity is indicated by that. In the meantime, thermal stability is indicated by the aliphatic index.

TLR2 was tightly bound, according to molecular docking studies. This implies that the vaccine can successfully stimulate innate immune pathways. We optimised codons as well. After that, we cloned it into the pET-28a(+) vector in silico. All of this demonstrates that it is effective for high expression in E. coli. This facilitates future testing and production scaling. This study demonstrates the practical advantages of computational vaccination. It expedites the process of developing new vaccines. Conventional laboratory techniques are more expensive and time-consuming. Conversely, these in silico techniques will allow us to quickly find strong antigenic components. These findings give some preliminary evidence of immunogenicity of MEBSV. Also, they facilitate its structural potential against *S. pyogenes*.

Future perspectives:

- Nevertheless, most of the predictions of the study are computed.
- In vitro and in vivo testing is required to continue with it.
- Experiments are required in order to confirm immunological activation and epitope recognition. Protective effects need to be tested on model organisms.
- It could be enhanced by testing different adjuvants. Also beneficial would be stability by

molecular dynamics. In any case, it would be good to conduct additional testing of the immune coverage among individual groups.

To sum up, the MEBSV presented, satisfies all the factors of a potential vaccine candidate. It establishes a strong ground on which further research can be conducted. When we continue to test this construct and carry out the corresponding adjustments, it may become a viable tool of controlling *S. pyogenes* infections without an unjust cost. In general, it enhances the process of computational vaccine design.

Chapter 7

References

1. Aslam, S., Ashfaq, U. A., Zia, T., Aslam, N., Alrumaihi, F., Shahid, F., ... & Qasim, M. (2022). Proteome based mapping and reverse vaccinology techniques to contrive multiepitope based subunit vaccine (MEBSV) against *Streptococcus pyogenes*. Infection, Genetics and Evolution, 100, 105259.
2. Mortazavi, B., Molaei, A., & Fard, N. A. (2024). Multi-epitope vaccines, from design to expression; an in silico approach. Human Immunology, 110804.
3. Stevens, D. L., & Bryant, A. E. (2016). *Severe Group A streptococcal infections*. In J. J. Ferretti, D. L. Stevens, & V. A. Fischetti (Eds.), *Streptococcus pyogenes: Basic biology to clinical manifestations* [Internet]. University of Oklahoma Health Sciences Center. https://www.ncbi.nlm.nih.gov/books/NBK333425/
4. ks/NBK333425/Chaleshtori, Z. A., Rastegari, A. A., Nayeri, H., & Doosti, A. (2023). Use of immunoinformatics and the simulation approach to identify Helicobacter pylori epitopes to design a multi-epitope subunit vaccine for B-and T-cells. BMC biotechnology, 23(1), a 42.
5. Brouwer, S., Rivera-Hernandez, T., Curren, B. F., Harbison-Price, N., De Oliveira, D. M., Jespersen, M. G., ... & Walker, M. J. (2023). Pathogenesis, epidemiology and control of Group A Streptococcus infection. Nature Reviews Microbiology, 1-17.
6. Avire, N. J., Whiley, H., & Ross, K. (2021). A review of *Streptococcus pyogenes*: public health risk factors, prevention and control. Pathogens, 10(2), 248.
7. Pollard, A. J., & Bijker, E. M. (2021). A guide to vaccinology: from basic principles to new developments. Nature Reviews Immunology, 21(2), 83-100.
8. Wang, J., Ma, C., Li, M., Gao, X., Wu, H., Dong, W., & Wei, L. (2023). *Streptococcus Pyogenes*: Pathogenesis and the Current Status of Vaccines. Vaccines, 11(9), 1510.
9. Walkinshaw, D. R., Wright, M. E., Mullin, A. E., Excler, J. L., Kim, J. H., & Steer, A. C.(2023). The Streptococcus pyogenes vaccine landscape. npj Vaccines, 8(1), 16.
10. Doern, C. D. (2023). Group A Streptococcus (Streptococcus pyogenes): the Most Interesting Pathogen in the World. Clinical Microbiology Newsletter, 45(5), 39-46
11. Wonggo, D., & Wahjudi, M. (2023). The Construction of A Multi-epitope Vaccine Against Klebsiella pneumoniae Using in silico Approach. Molecular and Cellular Biomedical Sciences (MCBS), 7(2), 90-97.
12. Velusamy, S., Husain, S. A. M., Khan, J. M., Ahmad, A., & Muthusamy, J. (2023). Vaccine

candidates for cellulitis from Staphylococcus aureus and *Streptococcus pyogenes*–In silico approach. Journal of King Saud University-Science, 35(9), 102917.

13. Ebrahimi, S., Mohabatkar, H., & Behbahani, M. (2019). Predicting promiscuous T cell epitopes for designing a vaccine against *Streptococcus pyogenes*. Applied biochemistry and biotechnology, 187, 90-100.
14. Maglinao, R. B., Jamito, J. M. V., Kalaw, L. M., Lim, J. M. M., Manimtim, G. A. L., Marcelo, A. L. G., ... & Cano, E. A. A. (2024). In silico Comparative Analysis of the Pathogenicity and Molecular Characteristics of Toxic Shock Syndrome Caused by Staphylococcus aureus and *Streptococcus pyogenes*. Asian Journal of Biological and Life Sciences, 13(2), 473.
15. Rehman, A., Wang, X., Ahmad, S., Shahid, F., Aslam, S., Ashfaq, U. A., ... & Abd_Allah, E. F. (2021). In silico core proteomics and molecular docking approaches for the identification of novel inhibitors against *streptococcus pyogenes*. International Journal of Environmental Research and Public Health, 18(21), 11355.
16. Ayyagari VS, Venkateswarulu TC, Abraham Peele K, Srirama K. Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. J Biomol Struct Dyn. 2022; 40(7): 2963-77. 22.
17. Public Health Agency of Canada. (2021, August). *Pathogen safety data sheets – Infectious substances: Streptococcus pyogenes.* Government of Canada. https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/streptococcus-pyogenes.html
18. Albekairi, T. H., Alshammari, A., Alharbi, M., Alshammary, A. F., Tahir ul Qamar, M., Ullah, A., ... & Ahmad, S. (2022). Designing of a novel multi-antigenic epitope-based vaccine against E. hormaechei: an intergraded reverse vaccinology and immunoinformatics approach. Vaccines, 10(5), 665.
19. Oluwagbemi OO, Oladipo EK, Dairo EO, Ayeni AE, Irewolede BA, Jimah EM, et al. Computational construction of a glycoprotein multi-epitope subunit vaccine candidate for old and new South African SARS-CoV-2 virus strains. Inform Med Unlocked. 2022; 28: 100845. doi: 10.1016/j.imu.2022.10084
20. Yousaf, M., Ullah, A., Sarosh, N., Abbasi, S. W., Ismail, S., Bibi, S., ... & Bin Emran, T.

(2022). Design of multi-epitope vaccine for staphylococcus saprophyticus: Pan-genome and reverse vaccinology approach. Vaccines, 10(8), 1192.

21. Van Regenmortel MHV. What is a b-cell epitope? Methods Mol Biol. 2009; 524: 3-20.
22. Kalita P, Padhi AK, Zhang KYJ, Tripathi T. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. Microb Pathog. 2020; 145: 104236. doi: 10.1016/j.micpath.2020.104236.
23. Stewart E, Triccas JA, Petrovsky N. Adjuvant strategies for more effective tuberculosis vaccine immunity. Microorganisms. 2019; 7(8): 255. doi: 10.3390/microorganisms7080255.
24. Stratmann T. Cholera toxin subunit B as adjuvant--An accelerator in protective immunity and a break in autoimmunity. Vaccines. 2015; 3(3): 579-96
25. Rajendran Krishnamoorthy H, Karuppasamy R. Designing a novel SOX9 based multi-epitope vaccine to combat metastatic triplenegative breast cancer using immunoinformatics approach. Mol Divers. 2022; 2022: 1-14. doi: 10.1007/s11030-022-10539-w.
26. Gul S, Ahmad S, Ullah A, Ismail S, Khurram M, Tahir Ul Qamar M, et al. Designing a recombinant vaccine against Providencia rettgeri using immunoinformatics approach. Vaccines. 2022; 10(2): 189. doi: 10.3390/vaccines10020189.
27. Girija ASS, Shoba G, Priyadharsini JV. Accessing the T-cell and B-cell immuno-dominant peptides from A. baumannii biofilm associated protein (bap) as vaccine candidates: A computational Approach. Int J Pept Res Ther. 2021; 27(1): 37-45.
28. Dar HA, Ismail S, Waheed Y, Ahmad S, Jamil Z, Aziz H, Hetta HF, Muhammad K. Designing a multi-epitope vaccine against Mycobacteroides abscessus by pangenome-reverse vaccinology. Sci Rep. 2021; 11(1): 11197. doi: 10.1038/s41598-021-90868-2.
29. Salaikumaran MR, Kasamuthu PS, Aathmanathan VS, Burra VLSP. An in silico approach to study the role of epitope order in the multiepitope-based peptide (MEBP) vaccine design. Sci Rep. 2022; 12(1): 12584. doi: 10.1038/s41598-022-16445-3.
30. Pandey RK, Bhatt TK, Prajapati VK. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. Sci Rep. 2018; 8(1): 1125. doi: 10.1038/s41598-018-19456-1
31. Naorem, R. S., Pangabam, B. D., Bora, S. S., Fekete, C., & Teli, A. B. (2024). Immunoinformatics Design of a Multiepitope Vaccine (MEV) Targeting Streptococcus mutans: A Novel Computational Approach. Pathogens, 13(10), 916.

32. Dereeper, A., Summo, M., & Meyer, D. F. (2022). PanExplorer: a web-based tool for exploratory analysis and visualization of bacterial pan-genomes. Bioinformatics, 38(18), 4412-4414.
33. Naveed, M., Tehreem, S., Arshad, S., Bukhari, S. A., Shabbir, M. A., Essa, R., ... & Khan, I. (2021). Design of a novel multiple epitope-based vaccine: An immunoinformatics approach to combat SARS-CoV-2 strains. Journal of infection and public health, 14(7), 938-946.
34. Yazdani, Z., Rafiei, A., Ghoreyshi, M., & Abediankenari, S. (2024). In silico analysis of a candidate multi-epitope peptide vaccine against human brucellosis. Molecular Biotechnology, 66(4), 769-783.
35. Yılmaz Çolak, Ç. (2021). Computational design of a multi-epitope vaccine against clostridium chauvoei: An immunoinformatics approach. International Journal of Peptide Research and Therapeutics, 27(4), 2639-2649.
36. Ghandadi, M. (2022). An immunoinformatic strategy to develop new mycobacterium tuberculosis multi-epitope vaccine. International Journal of Peptide Research and Therapeutics, 28(3), 99.
37. National Center for Biotechnology Information (NCBI), n.d. NCBI. https://www.ncbi.nlm.nih.gov/
38. Dey, J., Mahapatra, S. R., Raj, T. K., Kaur, T., Jain, P., Tiwari, A., ... & Suar, M. (2022). Designing a novel multi-epitope vaccine to evoke a robust immune response against pathogenic multidrug-resistant Enterococcus faecium bacterium. *Gut Pathogens*, *14*(1), 21.
39. Humayun, F., Cai, Y., Khan, A., Farhan, S. A., Khan, F., Rana, U. I., ... & Shamas, S. (2022). Structure-guided design of multi-epitopes vaccine against variants of concern (VOCs) of SARS-CoV-2 and validation through In silico cloning and immune simulations. *Computers in Biology and Medicine*, *140*, 105122.
40. Gomes, L. G. R., Rodrigues, T. C. V., Jaiswal, A. K., Santos, R. G., Kato, R. B., Barh, D., ... & Tiwari, S. (2022). In silico designed multi-epitope immunogen "Tpme-VAC/LGCM-2022" may induce both cellular and humoral immunity against Treponema pallidum infection. *Vaccines*, *10*(7), 1019.
41. Khan, R. M. A., Anwar, S., & Pirzada, Z. A. (2020). *Streptococcus pyogenes* strains associated with invasive and non-invasive infections present possible links with *emm* types and superantigens. *Iranian journal of basic medical sciences*, *23*(1), 133–139.

https://doi.org/10.22038/IJBMS.2019.38635.9164

42. Technical University of Denmark, n.d. SignalP-6.0 server. https://services.healthtech.dtu.dk/services/SignalP-6.0/
43. Technical University of Denmark, n.d. TargetP-2.0 server. https://services.healthtech.dtu.dk/services/TargetP-2.0/
44. .Dimitrov I, Bangov I, Flower DR, Doytchinova I, n.d. AllerTOP v.2.0: Allergenicity Prediction Server. https://www.ddg-pharmfac.net/AllerTOP/method.html
45. Doytchinova IA, Flower DR, n.d. VaxiJen v2.0: Antigenicity Prediction Server. https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html
46. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A, n.d. ProtParam tool – ExPASy. https://web.expasy.org/protparam/
47. Immune Epitope Database and Analysis Resource, n.d. T cell epitope prediction tools. IEDB. http://tools.iedb.org/main/tcell/
48. IEDB Analysis Resource, n.d. BepiPred Linear B Cell Epitope Prediction Tool. http://tools.iedb.org/bcell/
49. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y, n.d. I-TASSER: Protein structure and function prediction. https://zhanggroup.org/I-TASSER/
50. South Green Bioinformatics Platform, n.d. Xanthomonas oryzae genome project. PanExplorer.https://panexplorer.southgreen.fr/cgi-bin/doc.cgi?project=Xanthomonas_oryzae
51. Immune Epitope Database and Analysis Resource, n.d. Population coverage tool. IEDB. http://tools.iedb.org/population/
52. Hebditch M, Carballo-Amador M.A., Charonis S, Curtis R, Warwicker J Protein-Sol: a web tool for predicting protein solubility from sequence.Bioinformatics (2017)
53. VectorBuilder.(n.d.). Codon Optimization Tool. https://www.vectorbuilder.com/tool/codon-optimization.html
54. Hon J, Marusiak M, Martinek T, Kunka A, Zendulka J, Bednar D, Damborsky J, n.d. SoluProt: Soluble protein expression prediction. Loschmidt Laboratories, Masaryk University. https://loschmidt.chemi.muni.cz/soluprot/
55. SnapGene, n.d. SnapGene® software (from Dotmatics). https://www.snapgene.com/

56. Buchan DWA, Jones DT, n.d. PSIPRED: Protein structure prediction server. http://bioinf.cs.ucl.ac.uk/psipred/
57. Heo L, Park H, Seok C, n.d. GalaxyRefine: Protein structure refinement. https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE
58. Rapin N, Lund O, Bernaschi M, Castiglione F, n.d. C-ImmSim: Immune simulation software. http://kraken.iac.rm.cnr.it/C-IMMSIM/
59. UCLA MBI, n.d. SAVES v6.0: Structural analysis and verification server. https://saves.mbi.ucla.edu/
60. RCSB Protein Data Bank, n.d. RCSB PDB: Protein Data Bank. https://www.rcsb.org/
61. Yan Y, Zhang D, Zhou P, Li B, Huang S, n.d. HDOCK: Protein–protein docking tool. http://hdock.phys.hust.edu.cn/data/6804a98c84d19/
62. López-Blanco, J. R., & Chacón, P. (n.d.). *iMODS: Normal mode analysis server*. https://imods.iqf.csic.es/
63. Remmington, A., & Turner, C. E. (2018). The DNases of pathogenic Lancefield streptococci. *Microbiology*, *164*(3), 242-250.