In-Silico Based Multi Epitope Vaccine Construction against Salmonella Typhimurium: A Comparative Study

By

Zarmeen Yaqub
20146046

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy

School of Pharmacy
Brac University
September 2024

© 2024. Brac University
All rights reserved.

Declaration

It is hereby declared that

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

Student's Full Name and Signature:

____

Zarmeen Yaqub
Student ID: 20146046

Approval

The thesis titled "In-Silico Based Multi Epitope Vaccine Construction against Salmonella Typhimurium: A Comparative Study" submitted by Zarmeen Yaqub (ID: 20146046) of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy.

Supervised By:

Supervisor:

____

Mohammad Kawsar Sharif Siam

Senior Lecturer

School of Pharmacy, Brac University

Approved By:

Dean:

____

A F M Yusuf Haider,PhD

Acting Dean, School of Pharmacy

Professor, Department of Mathematics and Natural Sciences

Brac University

Ethics Statement

The thesis was completed following all the required ethical standards. There was no involvement of any human or animal trial.

Abstract

Salmonella Typhimurium is the major cause of foodborne illnesses globally, responsible for significant morbidity and mortality. Although, there are many efficacious antibiotics, development of vaccine against this pathogen is necessary for public health due to increasing antibiotic resistance. Furthermore, chronic infections with Salmonella Typhimurium have been implicated in the development of colon cancer in recent investigations. This study presents a comparative analysis of in silico based multi epitope vaccine construction against Salmonella typhimurium targeting 7 different proteins. Each protein was evaluated for vaccine antigenicity, physicochemical properties and structural stability using immune informative approach to identify the most suitable vaccine candidate. Various tools were used to identify the different epitopes of HTL, CTL B cell, which were connected using suitable linkers. Physicochemical properties was also check for the prepared vaccine through using Protparam. Molecular Docking was done with TLR- 4 for checking the residual interaction. Among the proteins investigated, lipoprotein emerged as the most promising candidate demonstrating superior vaccine antigenicity, favorable physical properties and structural stability. This finding highlights the potential of lipoprotein based vaccine in combating Salmonella typhimurium infection offering a pathway to improve preventive strategies. Additional research by collaborating both laboratory & biological system is required to confirm the reliability and potency of the vaccine.

Keywords: Salmonella Typhimurium; vaccine candidate; CTL; HTL; B-cell; Comparison

Acknowledgement

I would like to express my sincere gratitude to my supervisor Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, BRAC University for his continuous support, patience, motivation and valuable feedback. His guidance and support helped me all the time while performing this research.

Table of Contents

Declaration ...... **ii**

Approval ...... **iii**

Ethics Statement ...... **iv**

Abstract ...... **v**

Acknowledgement ...... **vi**

List of Tables ...... **ix**

List of Figures ...... **x**

List of Acronyms ...... **x**

Chapter 1 Introduction **1**

1.1 Genomic Structure of Salmonella Typhimurium 3

1.2 Lifecycle and Pathogenesis of Salmonella Typhimurium 4

Chapter 2 Methodology **5**

2.1 Choosing of Suitable Protein Sequence 5

2.2 Identifying Cytotoxic T Lymphocyte (CTL) Epitope 6

2.3 Analyzing CTL epitopes 6

2.4 Identifying Helper T Lymphocyte (HTL) Epitope 6

2.5 Analyzing HTL epitopes 7

2.6 Selection and Evaluation of B cell Epitopes 7

2.7 Construction of Vaccine incorporating Suitable Linkers 8

2.8 Evaluation of Vaccine Antigenicity, Allergenicity, Toxicity of Constructed Vaccine 8

2.9 Biochemical Analysis of Prepared Vaccine 8

2.10 Creating 3D Model of Constructed Vaccine 9

2.11 Generation and Analysis of Ramachandran Plots 9

2.12 Z- Score Analysis for 3D- Model Quality Assessment 9

2.13 Molecular Docking Analysis of the Vaccine 10

2.14 Immune Response Stimulation of the Vaccine 10

2.15 Key Aspects of Research Methodology 10

Chapter 3 Results 11

3.1 Protein Selection and Antigenicity Determination........ 11

3.2 Selection and Screening of CTL Epitopes 15

3.3 Selection and Screening of HTL Epitopes 21

3.4 Selection and Screening of B cell epitopes 23

3.5 Construction of Vaccine using Suitable Linkers with their Antigenicity, Allergenicity and Toxicity Profile Assessment........ 26

3.6 In-silico Biochemical Analysis of the Constructed Vaccines 28

3.7 Homology Modelling of the Prepared Vaccines 29

3.8 Z-score Determination........ 30

3.9 Ramachandran Plots Prediction 31

3.10 Molecular Docking with Toll like receptor 34

3.11 Immune Simulation using C-ImmSim 34

Chapter 4 Discussion 35

Chapter 5 Conclusion 37

References 38

List of Tables

Table 1:Protein Based on Antigenicity scores 15

Table 2: CTL epitopes of the 7 proteins with their respective combined scores 16

Table 3: Sequence number, Length Score_EL, Percentile rank of eluted ligand & binding affinity, binding affinity of the MHC I alleles particular to the epitopes using NetMHCpan 4.1 server 18

Table 4: Antigenicity, Allergenicity, Toxicity determination of the Selected CTL epitopes . 21

Table 5: HTL Epitopes for each protein (i-vii) with their Cytokine inducing Capability (IFN-γ, IL-4 and IL-10) 21

Table 6: Antigenicity, Allergenicity, toxicity profile assessment of the selected B-Cell epitopes 26

Table 7 : Vaccine Construction for the 7 proteins made by incorporating different epitopes and linkers 26

Table 8: Antigenicity, Allergenicity and toxicity profile assessment of the constructed vaccines 27

Table 9: Physico-chemical parameters of the constructed vaccines using Protparam Tool. ... 28

Table 10: Coverage and Confidence determination of the Constructed Vaccines using PHYRE-2 29

Table 11: Z-scores for the constructed vaccines (i-vii) 30

Table 12 : Ramachandran Plot Prediction for all the vaccines 31

Table 13: Summary table of the all the parameters assessed for the vaccine candidates 32

List of Figures

Figure 1: Step-by-Step Schematic overview of the entire process involved in constructing an In-Silico Vaccine 5

Figure 2: Vaccine construction using linkers 8

Figure 3: Determining antigenicity of the Protein no: 1 11

Figure 4: Determining antigenicity of the Protein no: 2 12

Figure 5: Determining antigenicity of the Protein no: 3 12

Figure 6 : Determining antigenicity of the Protein no: 4 13

Figure 7: Determining antigenicity of the Protein no: 5 13

Figure 8: Determining antigenicity of the Protein no: 6 14

Figure 9: Determining antigenicity of the Protein no: 7 14

Figure 10 (a-g): Position vs. score graph and B-Cell epitopes for all proteins 25

Figure 11: Protparam Results for Vaccine (vi) 29

Figure 12: Phyre-2 results for vaccine (ii) & (vi) 30

Figure 13: Overall model quality and local model quality of Vaccine (vi) 31

Figure 14: Ramachandran plot prediction of vaccine (vi) using SWISS-MODEL Tool. 32

Figure 15: Highest stable docked complex obtained via ClusPro 34

Figure 16: Molecular Docking performed between TLR-4 & vaccine (vi) using ClusPro 34

Figure 17: Immune simulation of vaccine (vi) illustrated graphically via C-ImmSim 35

List of Acronyms

CTL	Cytotoxic T- Lymphocyte
GRAVY	The Grand Average of Hydropathicity
HTL	Helper T-Lymphocyte
MHC	Major Histocompatibility complex

Chapter 1
Introduction

Non typhoidal Salmonellosis (NTS) which is transmitted via food known for causing diseases gastroenteritis, creates presence of bacteria in bloodstream and other sort of localized infections in the body. It poses serious risk for individuals who have weak immune system, including those with cancer, diabetes or receiving any sort of immunotherapy treatments. (Acheson & Hohmann, 2001). The spread of NTS through food is a serious public health problem across the whole world (Hoelzer et al., 2011). Known for causing disease worldwide NTS is responsible for causing more than ninety million cases of gastroenteritis leading to one hundred and fifty thousand deaths worldwide (Leung et al., 2013). Several research investigations confirmed the association between recurrent infections with NTS with an elevated risk of colonic malignancies. Findings from the studies revealed that even small, possibly unnoticed infections might be linked to acceleration of carcinogenesis in the colon due to prolonged Inflammation in the gut. The correlation observed between prior NTS infections along with the growth of carcinoma of the colon in individuals, most notably in those under the age of 60, implies that NTS might have been involved in the rising incidence of symptoms that promotes malignancies in the colon (Shanker & Sun, 2023). One of the strain of Salmonella that causes illness like acute gastroenteritis is Salmonella Typhimurium (Won & Lee, 2017). It is a gram negative bacteria which causes harmful diseases in the lumen of the intestine. Lipopolysaccharide (LPS) in the outer membrane of this pathogenic bacteria is the main reason for its toxicity which protects the bacteria from the environment (Slauch et al., 1995). Diarrhea, vomiting and abdominal pain are the common signs and symptoms that only appears after 12-24 hours of the organism entering the host and can last up to 2-5 days. The host's intestinal

mucosa is the first site of infection for S.Typhimurium, which then moves to the lymphoid organ and liver, causing disease in some organs of the body or might affect the whole body (Rosenberger et al., 2000). Serious complications may get triggered by Salmonella Typhimurium, notably among individuals who are in susceptible conditions as it leads to extended inflammation in the gut potentially facilitating carcinogenesis in the colon. This type of bacteria by the help of its virulence factors can initiate various immune response like type III Secretion System which can cause DNA damage & interference in the cell signaling pathways leading to development of malignancies and infections (Shanker & Sun, 2023). Effective Typhoid vaccines are available in the market. But there are some variation in the cellular structures among the Typhoidal and Non typhoidal Salmonella species which prevents this vaccine from working efficiently in case of NTS infections (Chinnasami et al., 2015). There are vast genetic variability in Salmonella species is creating a significant hindrance in the development of an effective vaccine targeting Salmonella (Lamichhane et al., 2024). There are also other reasons which add to the challenge of creating an effective Salmonella vaccine. These may include: improper handling of Vaccine leading to vaccine failure, Chances of risk as live attenuated vaccine can regain its virulence, when toxoid vaccines are given at high dose it can lead to toxoid tolerance, low immunogenicity of outer membrane vesicles which might result in a low immune responses. The progress in Salmonella vaccine development have been hindered by the combination of these factors. With these limitations in mind it is vital to highlight the necessity for continuous research in vaccine making process (Lamichhane et al., 2024). Recent studies show that antibiotic resistance is more prevalent in low income country like Bangladesh which has caused more difficulty for treating any disease effectively. This can happen due to improper use of antibiotics, low quality drugs, poor healthcare standard, unaffordability of expensive drugs etc. (Safain et al., 2020). Consequently it can be said

that vaccine is the best way which can prevent antimicrobial resistance. Because vaccinations particularly target the condition they are meant to treat, they are more unlikely to produce resistance than medications (Alghamdi, 2021). The standard vaccination design employs heat and pathogens that have been chemically degraded. Due to numerous failures in subsequent phases, this can take range around 15 to 20 years to complete (Rawat et al., 2023).In the recent decades, a new field called "Immunoinformatics "has emerged as a major contributor to vaccine design and development (Rawat et al., 2023). The traditional way of vaccine design takes 15-20 years for development process while immunoinformatics-based vaccine takes around 2-3 year for development which can screen multiple candidates at once (Rawat et al., 2023). By using immunoinformatics-based vaccine potential immunogenic epitopes are integrated using suitable linkers to create a vaccine sequence which could generate an immune response (Rawat et al., 2023). Other features such as half-life estimation, allergenicity toxicity, and antigenicity were considered to obtain better candidates for constructing multi epitope vaccine (Rawat et al., 2023).

In addition to addressing current health issues, a successful vaccination against Salmonella Typhimurium is required to reduce the potential danger of developing cancer in the years to come. Through this research we intend to present a comparative study of in silico multi epitope vaccine construction targeting Salmonella Typhimurium. Throughout the study 7 different proteins were examined. The antigenicity, physicochemical properties, structural stability and other relevant parameters were assessed to identify the most suitable vaccine candidate.

1.1 Genomic Structure of Salmonella Typhimurium

The complex genomic structure of Salmonella Typhimurium includes a circular chromosome that is roughly 4.8 megabases in size and has 4,500 genes on it. Salmonella Pathogenicity Islands (SPIs), or gene clusters encoding virulence factors, are some features unique to the genome. To be

more specific, SPI-1 and SPI-2 are necessary for the bacterium to enter host cells and avoid being recognized by the immune system. The SPIs are responsible for causing the disease and make the bacterium adaptable and pathogenic organism so that it can invade the host (McClelland et al., 2001). It also have multiple plasmids that are related in causing of this disease and are responsible for antibiotic resistance, toxin production etc. They encode genes which helps in Salmonella Typhimurium in spreading resistance traits which makes it more difficult to treat. Salmonella Typhimurium and E.coli have a core genome that is nearly identical (10% difference in DNA sequences), indicating that they have shared an ancestor around 100 million years ago (De Jong et al., 2012).

1.2 Lifecycle and Pathogenesis of Salmonella Typhimurium

After entering the body through the mouth, Salmonella Typhimurium interacts with the intestinal lining present in the large intestine. Consequently, its Type III Secretion System is activated which helps the bacteria in penetrating the cells and forming a vacuole (SCV) to enter inside the host's body. Once inside this vacuole, then it uses another secretion system called T3SS2 to multiply and evade the host's immune system (Galán, 2021).The pathogenesis of this disease starts with Salmonella Typhimurium causing inflammation in the intestine, which is essential for its survival and growth. The bacteria uses some of its proteins to cause inflammation without directly triggering the body's immune receptor. Due to this inflammation the existing normal gut bacteria in our body is disrupted. This inflammation gives bacteria an advantage by providing them with necessary nutrients and condition favoring the growth of Salmonella Typhimurium (Galán, 2021).

Chapter 2

Methodology

The sequence of steps that were maintained while designing the vaccine is outlined below:

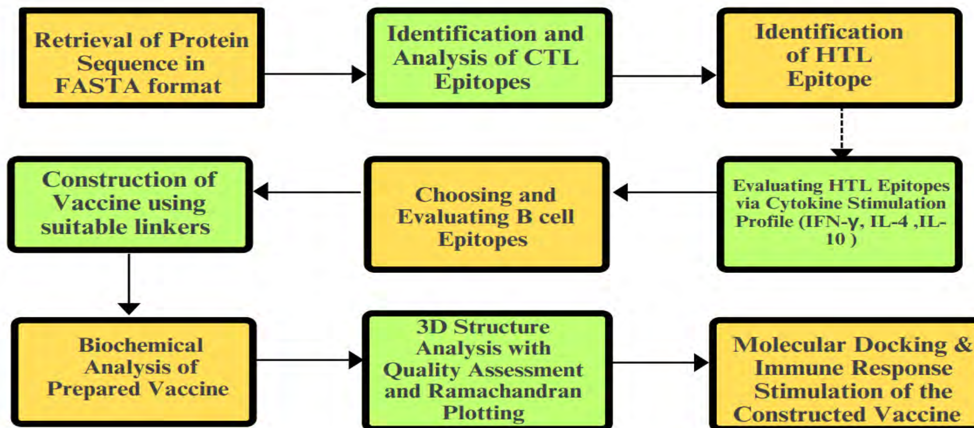


Figure 1: Step-by-Step Schematic overview of the entire process involved in constructing an In-Silico Vaccine

2.1 Choosing of Suitable Protein Sequence

Using the NCBI protein database, the suitable target protein sequences were selected which had the capacity for initiating an immunological reaction. The protein sequences were retrieved from NCBI and Uniprot (Bateman et al., 2023) in FASTA format successfully. The protein sequences with 400-600 Amino acids were selected for vaccine designing. The submitted protein sequences were then classified as either non-antigens (which could be targeted for a vaccine) or antigens (Doytchinova & Flower, 2007) employing the computer program Vaxijen v2.0. As stated by Doytchinova and Flower (2007) the target organism specified for the screening of protein sequences was "Bacteria," with an acceptable threshold of 0.5. According to Doytchinova and Flower (2007) it relies on an alignment-independent process to assist with selecting which proteins should be looked at greater detail during the vaccine design procedure.

2.2 Identifying Cytotoxic T Lymphocyte (CTL) Epitope

NetCTL 1.2 is a computational tool used for predicting CTL epitopes important for inducing immune response against the pathogen. It is a tool for CTL epitope prediction that offers a significant potential for enhancing vaccine design strategies. This tool integrates three parameters to pinpoint sequences that are probable to provoke an immune reaction. The protein sequences were inputted in FASTA format with 0.75 threshold applied for CTL epitope selection (Larsen et al., 2007).

2.3 Analyzing CTL epitopes

A computer-aided approach entitled NetMHCpan-4.1 is intended to forecast the binding strength between peptides and MHC class I molecules, which is vital for yielding a response from the immune system. This tool takes into account all HLA alleles and may assess peptides of any length, however it predominantly focusses on 9-mer peptides. In case of Strong Binding (SB) threshold of 0.5 was maintained and for Weak Binding threshold of 2 was maintained. Then Binding Affinity (BA) option and Sort by prediction score was selected in the server for the output page. After that, the output page generated in which the peptides were sorted by their binding affinity scores, highlighting that the peptides had met the specified thresholds. The peptides which had Strong Binding were chosen for further study. The CTL epitopes utilized for vaccine design have to be antigen and should not demonstrate any allergens & toxicity.

2.4 Identifying Helper T Lymphocyte (HTL) Epitope

The NetMHC-IIpan service that supports in HTL epitope forecasting leveraging artificial neural networks was used to recognize HTL epitopes. The protein sequence was inputted in FASTA format. 15 mer peptide was selected based on length, with a maximum of 20 alleles chosen per

submission. Repeating that process led to the picking of every allele. It was decided to keep the limit at 1% for strong binder alongside 5% for weak binder. After selecting the parameters to begin the estimation procedure, press on the "Submit" button. The results page was displayed on the server with alleles which fall under the threshold. Using the binding affinity scores and classifications, the potential HTL epitopes were predicted. Binding Affinity scores indicate the strength of the peptide -MHC binding. Through Classification peptides are categorized as Strong binder, Weak binder, non-binder based on % Rank thresholds. Strong binders (SB) are more likely to be effective HTL epitopes.

2.5 Analyzing HTL epitopes

There are many cytokines present in the body which helps in immune system activation. But for vaccine designing, the presence of three Cytokines is important: IFN-γ, IL-4, IL-10. The HTL epitopes taken from NetMHC-IIpan 4.0 were evaluated to check their ability to produce these three Cytokines. Firstly, MHC molecules required for antigen presentation resulting in generating immune system are activated by IFN-γ which is determined by IFN- epitope server. Secondly, for the growth and multiplication of B cells IL-4 is needed which can be determined by IL-4 pred. Lastly, for suppressing the inflammatory characteristics and controlling the homeostasis, IL-10 is needed which is checked using IL-10 pred server. Finally, the HTL epitopes which were IFN-γ positive, inducer of IL-4 & IL-10, non- allergen, non-toxin and antigenic were selected for vaccine designing.

2.6 Selection and Evaluation of B cell Epitopes

The B cell epitopes have been forecasted using an advanced computational technique called "Bepipred Linear Prediction 2.0" from the IEDB Resource Analysis server. It assists in determining the appropriate epitopes that can prompt the body’s immune defense by analyzing the

protein's most immunogenic regions. After the protein sequence was analyzed, a graph and a table showing the antigenicity scores were generated. The threshold for B cell epitope prediction was adjusted to 0.5 and epitopes falling within this range were selected. The selected epitopes' toxicity, antigenicity, and allergenicity were further investigated.

2.7 Construction of Vaccine incorporating Suitable Linkers

For constructing the in silico vaccine multi epitope vaccine various linkers were incorporated into the main protein sequence obtained from NCBI and Uniprot protein database .The primary protein sequence was joined to the CTL epitopes using EAAAK linker, the CTL epitope was coupled to the HTL epitope through GPGPG linker, and HTL and B cell epitopes were attached via KK linker.

Figure 2: Vaccine construction using linkers

2.8 Evaluation of Vaccine Antigenicity, Allergenicity, Toxicity of Constructed Vaccine

After designing the vaccine their antigenic potential was reviewed using Vaxijen v2.0. An online interactive allergen assessing portal named Allergen Online was used with the vaccine sequence entered in plain format (Goodman et al., 2016). To determine the toxicity, T3DB server was used to check for any sort of toxicity present in the vaccine (Wishart et al., 2015).

2.9 Biochemical Analysis of Prepared Vaccine

Physicochemical properties of constructed vaccine was investigated by Protparam server for checking its safety and efficacy. This server provides comprehensive information which includes

the molecular weight, theoretical PI, amino acid composition, GRAVY, Instability Index etc. (Garg et al., 2016).

2.10 Creating 3D Model of Constructed Vaccine

Phyre 2 was subsequently utilised to build 3D modelling of the vaccine using the vaccination sequence submitted in plain format. At first the email address was submitted to receive notification and results upon completion of the job. The vaccine sequence was pasted in the box for modeling. The results was sent to the provided email address which included the predicted 3D models, confidence score, coverage score, PDB file . The PDB file was accessed using PyMOL software to visualize the 3D vaccination model design. (Kelley et al., 2015).

2.11 Generation & Analysis of Ramachandran Plots

To validate 3D model of the protein and to confirm precision of the anticipated three-dimensional structure, we generated Ramachandran plots using Swiss Model Expasy. The PDB file that was obtained from Phyre 2 was uploaded here for structure assessment. Ramachandran plots display the phi and psi angles of the amino acid residues in the protein structure. By plotting these angles it would provide insights regarding the conformational feasibility of the protein model highlighting the regions that are highlighting regions that are likely to be in energetically favorable conformations (Schwede et al., 2003).

2.12 Z- Score Analysis for 3D- Model Quality Assessment

ProSa-web was utilised in order to ascertain the z-score of the designed vaccination (Wiederstein & Sippl, 2007). This server gained access to the vaccine's three-dimensional quality model and the model's energy distribution. It checks quality and energy distribution of the 3D model and provides

a Z score against number of residues graph & a graph of energy against position of sequence. Upon uploading PDB file output includes Z score and relevant graphs (Wiederstein & Sippl, 2007).

2.13 Molecular Docking Analysis of the Vaccine

By ClusPro, molecular docking was done to determine the extent to which the created vaccines bind to the particular Toll-like receptor. We sourced the PDB file for TLR-4 through RCSB which was assigned as the receptor molecule. The PDB file received through Phyre2 was obtained from mail and uploaded as receptor molecule. The result page featured 10 docked complexes where the highest energy scoring docked complex was obtained (Comeau et al., 2004).

2.14 Immune Response Stimulation of the Vaccine

In-silico simulations immune response pattern associated with vaccination sequence was carried out via C-ImmSim webserver. The simulation included three injections at four week intervals with all parameters set to default and time steps at 1, 84,168. Here each step represents 8 hours starting from first injection at time =0) (Rapin et al., 2010).

2.15 Key Aspects of Research Methodology

The significance behind employing this entire procedure was the development of vaccine using in-silico approach which involved screening, prediction and different analyses performed using online tools. Though this method is effective but it doesn't assess the vaccine’s efficacy and safety. That's why additional investigation is needed to determine whether this vaccine is effective against Salmonella Typhimurium.

Chapter 3

Results

3.1 Protein Selection and Antigenicity Determination

7 proteins were selected and downloaded from the Uniprot and NCBI database. Afterwards, their antigenic potential was checked by Vaxijen. Antigenicity is an important aspect because it determines whether the protein has the capacity to show immune response. The protein sequences that are retrieved from Uniprot and NCBI are given below:

Protein 1

Protein Name: Porin

Antigenicity: 0.4005

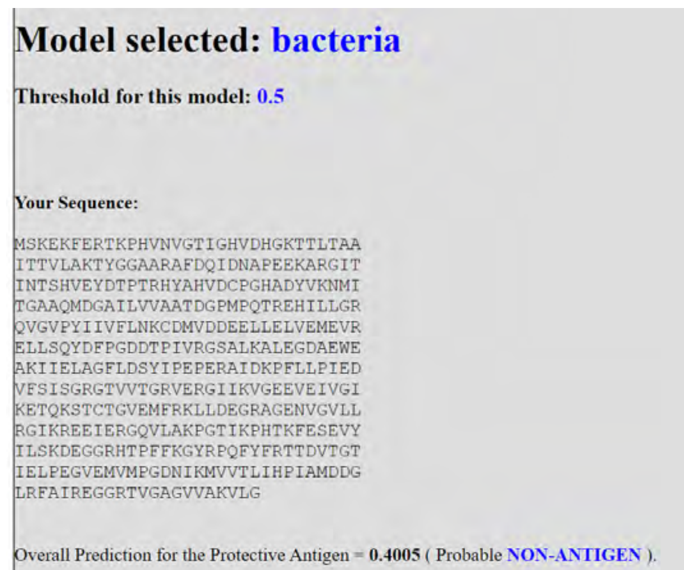


Figure 3: Determining antigenicity of the Protein no: 1 (Doytchinova & Flower, 2007)

Protein 2

Protein Name: Outer membrane protein

Antigenicity: 0.5911

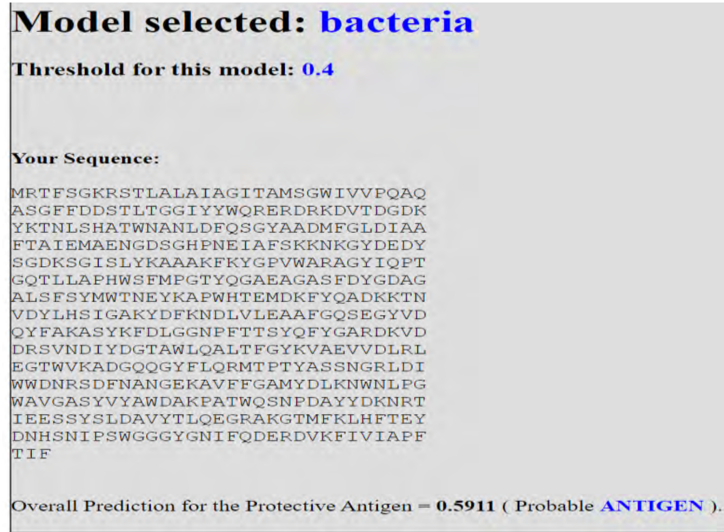


Figure 4: Determining antigenicity of the Protein no: 2 (Doytchinova & Flower, 2007)

Protein 3

Protein Name: Secreted Effector SptP

Antigenicity: 0.5192

Figure 5: Determining antigenicity of the Protein no: 3 (Doytchinova & Flower, 2007)

Protein 4

Protein Name: Fimbrial Protein

Antigenicity: 0.6286

Your Sequence:

```
MITKCDHRLMLLNTQRLKSVIKIISLLLLSF
VSFVTPSYGWSLYVGQYTKNEGGNVWGAVQD
FTKSYAPKNMLNPIAIHNGSGLREISVSNDA
TELDPGTGLSCNKDHDNHVLTLYHGYKNSGK
TYNGNMLWETNTPGMYMGIEVTSINFGGQYW
NEKFWVNWSNGETSGAKKFDLTNIMPKDGRF
ELCQRPTNWTYFGLGGIVLWVKYHLYIDDSF
NPAGATSLSVSLRKEQSYDYQFASSPAFSGS
HNVYYNIAPNQLTINYPTCQANAVTGDGVTN
STVPFGEQNAKDISANTINRKFSIKLSNCAY
VKELNVTLSSATVGGQDNTLLGNTLTGSEAA
GGIGVMIEGEKNPNSPADWTLLKPNISSSIY
SFTNSPDYENSEIGNAEQVMNFQATLKQDGN
KAITPGKFKATGKFTITYP
```

Overall Prediction for the Protective Antigen = **0.6286** (Probable ANTIGEN).

Figure 6 : Determining antigenicity of the Protein no: 4 (Doytchinova & Flower, 2007)

Protein 5

Protein Name: Scaffold Protein

Antigenicity: 0.5227

Model selected: bacteria

Threshold for this model: 0.5

Your Sequence:

```
MQRNLSHIISQATSAPLLLEPAYARVFFCAL
GRESGINSLHIPGNNESLDQSDMALVTGDFM
ATGKPQARFYQVVNGIAVLPVTGTLVHKLGG
MRPFSGMTGYDGVTARLQQAVSDPEVKGILL
DIDSPGGQAAGAFDCADMIYRMREQKPVWAL
ANETACSAAMLLAAACSHRLVTQTSRMGSIG
VVMAHTSYAEKLKQEGIDITLIYSGAHKADL
TPSQKLPESVYADYQQRMDEARKMFAEKVAR
YTGLSVDAVMATEAAVYDGQAIITTGLADGM
VNAADAIGVMAEAINSNKTGGTMPELSAADA
VTQENQRVMGILGCPEARGHEALAQMLAGQP
GMSVAQAKSILAAAAPADTTSTADRILALEE
AGGRETLAQTLAAMPEMTVEQARTILAASPI
AAATSLHDAVMALDEAKGREELAEKLAVMPG
MTTDQARDLLAAAPDKSGNAGLSMNNAFDAF
MQSHSPGPISGGKGHSNDTETTLLMSIPGTS
AT
```

Overall Prediction for the Protective Antigen = **0.5227** (Probable ANTIGEN)

Figure 7: Determining antigenicity of the Protein no: 5 (Doytchinova & Flower, 2007)

Protein 6

Protein Name: Lipoprotein

Antigenicity: 0.6122

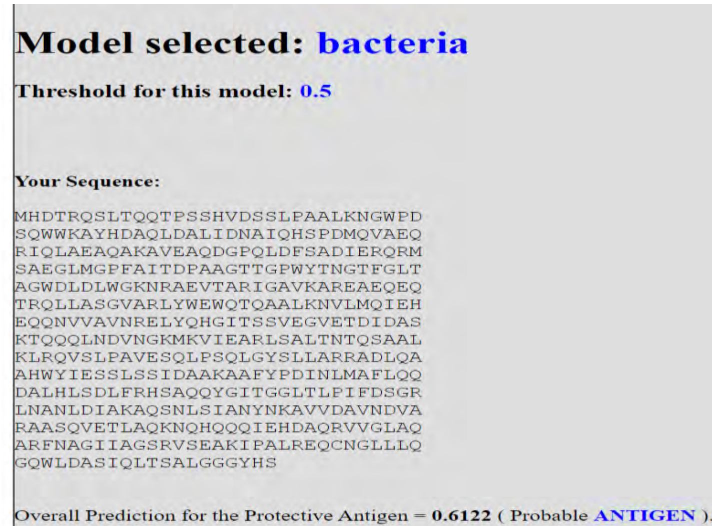


Figure 8: Determining antigenicity of the Protein no: 6 (Doytchinova & Flower, 2007)

Protein 7

Protein Name: Type III secretion system

Antigenicity: 0.5751

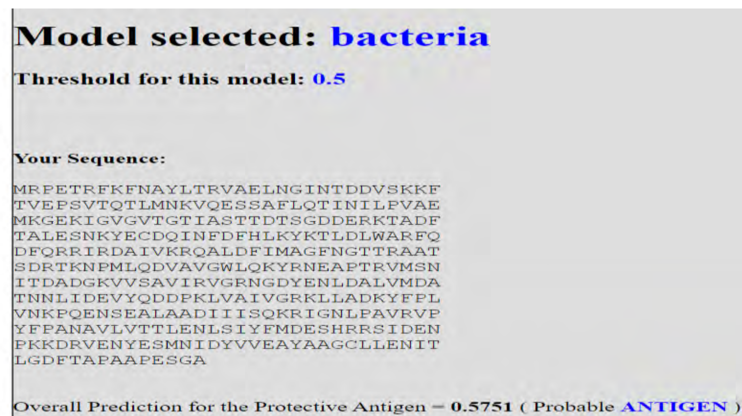


Figure 9: Determining antigenicity of the Protein no: 7 (Doytchinova & Flower, 2007)

Table 1: Protein Based on Antigenicity scores

Protein no.	Antigenicity score	Comments
Protein 1	0.4005	Low antigenicity, less preferred
Protein 2	0.5911	High antigenicity, strong potential
Protein 3	0.5199	High Antigenicity , Good Potential
Protein 4	0.6286	Highest Antigenicity ,Good Potential
Protein 5	0.5227	High Antigenicity, Good Potential
Protein 6	0.6122	Highest Antigenicity, Good potential
Protein 7	0.5751	High antigenicity, Good Potential

After evaluating the antigenicity scores it is observed that the Protein No-4 (Fimbrial protein) & 6 (Lipoprotein) showed the highest antigenicity. Protein no: 2, Protein no: 3, Protein no: 5, Protein no: 7 showed high antigenicity. Other proteins also showed good antigenicity scores but Protein no 4 & 6 stands out with the highest score.

3.2 Selection and Screening of CTL Epitopes

According to process explained by Larsen et al., for the selection of CTL (Cytotoxic Lymphocytes) epitopes, Net CTL1.2 was used keeping the threshold at 0.75. By using MHC-I supertype A1 the CTL epitopes were identified needed for making an effective vaccine. By the CTL epitopes selection process, a total of 10 Epitopes for Protein 1, 20 Epitopes for Protein 2, 12 epitopes of Protein no. 3, 11 epitopes of Protein no.4, 6 epitopes of Protein no.5, 8 epitopes of Protein no .6, 12 epitopes of Protein no. 7 with the potential to produce strong immune response which are shown below in Figure - in this study. The study also provides a table of combined scores associated with the CTL epitopes and Table 2 provides a complete compilation of these data.

Table 2: CTL epitopes of the 7 proteins with their respective combined scores

Protein no: 1		Protein no: 2		Protein no: 3		Protein no: 4	
CTL epitopes	Combined Scores	CTL epitopes	Combined Scores	CTL epitopes	Combined Scores	CTL epitopes	Combined Scores
HTKFESEVY	2.0415	QSEGYVDQY	2.9254	FLHALTEKY	1.4795	ATGKFTITY	2.7660
TINTSHVEY	1.8214	FSYMWTNEY	2.9109	GVSNDARLY	1.1931	FSGSHNVYY	2.7170
ITTVLAKTY	1.5928	STLTGGIYY	2.3271	STDQLEYLA	1.1034	YSFTNSPDY	2.0157
STCTGVEMF	1.1123	DSTLTGGIY	1.8994	MTNLQNLNK	0.9178	TSINFGGQY	1.9282
QMDGAILVV	1.0171	WLQALTFGY	1.4258	AIDQYNMQL	0.8957	FVSFVTPSY	1.4332
TTDVTGTIE	0.8957	YLHSIGAKY	1.3768	LSDGMPVPV	0.8848	HVLTLYHGY	1.1939
ELAGFLDSY	0.8615	SSYSLDAVY	1.3537	SSSDKHLPM	0.8206	CQRPTNWTY	1.0124
LTAAITTVL	0.8110	ITAMSGWIV	1.1708	LPSTDQLEY	0.8104	VLWVKYHLY	0.9734
EYDTPTRHY	0.7982	WQSNPDAYY	1.1565	SSDKHLPMI	0.7997	AVQDFTKSY	0.9261
ESEVYILSK	0.7830	FLQRMTPTY	1.1165	MLEDASQFV	0.7722	SIKLSNCAY	0.8702
-	-	NIPSWGGGY	0.9524	TSEDQMQAK	0.7689	VSLRKEQSY	0.7625
-	-	WAVGASYVY	0.9405	SQGEAIDQY	0.7605	-	-
-	-	FDDTTGKYY	0.8662	-	-	-	-
-	-	DAGALSFSY	0.8188	-	-	-	-
-	-	ATWNANLDF	0.8064	-	-	-	-

-	-	YVDQYF AKA	0.8043	-	-	-	-
-	-	AADMFG LDI	0.7942	-	-	-	-
-	-	SLDAVYT LQ	0.7862	-	-	-	-
-	-	KADGQQ GYF	0.7564	-	-	-	-
-	-	KAVFFGA MY	0.7505	-	-	-	-

Protein no: 5		Protein no: 6		Protein no: 7	
CTL epitopes	Combined Scores	CTL epitopes	Combined Scores	CTL epitopes	Combined Scores
VMATEAA VY	1.9205	LTSALGGGY	3.0778	FTALESNKY	3.3330
TTDQARDL L	1.4542	QSNLSIANY	2.1727	NIDYVVEAY	2.8807
AFDCADMI Y	1.2346	LASGVARLY	1.5005	TTLENLSIY	1.9582
STADRILAL	0.9742	VVAVNRELY	1.4742	NTDDVSKKF	1.4651
GVVMAHT SY	0.7789	LTAGWDLDL	0.9869	VMDATNNLI	1.1982
SLDQSDMA L	0.7742	ITDPAAGTT	0.8785	ETRFKFNAY	1.1937
-	-	SQLPSQLGY	0.7907	NYESMNIDY	1.1171
-	-	LTNTQSAAL	0.7845	LADKYFPLV	0.9336
-	-	-	-	TLENLSIYF	0.8400
-	-	-	-	TTDTSGDDE	0.8323
-	-	-	-	VTQTLMNKV	0.7563

With NetMHCpan 4.1, strong binding MHC-I alleles unique to CTL epitopes were found. The alleles that were collected shown in Table no -3 for further assessment which are the supertypes

of the alleles collected from Net CTL1.2 shown in Table no -2. Epitopes were categorized into two groups based on their binding strength: Strong Binder with score of 0.500 or lower and Weak Binder with scores of 2.000 or higher. This process helped us to identify and select the CTL epitopes that were most relevant for our Study.

Table 3: Sequence number, Length Score_EL, Percentile rank of eluted ligand & binding affinity, binding affinity of the MHC I alleles particular to the epitopes using NetMHCpan 4.1 server

Protein no:1				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	HTKFESEVY	0.253	0.610	1500.32
HLA-A*01:01	TINTSHVEY	0.250	0.658	1661.01
HLA-A*02:01	QMDGAILVV	0.260	0.470	32.51
HLA-A*26:01	ELAGFLDSY	0.015	0.010	20.05
HLA-B*58:01	ITTVLAKTY	0.336	1.128	361.22
Protein no: 2				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	QSEGYVDQY	0.016	0.055	58.95
HLA-A*01:01	FDDTTGKYY	0.107	0.019	22.45
HLA-A*01:01	STLTGGIYY	0.096	0.197	299.14
HLA-A*01:01	YLHSIGAKY	0.307	0.659	1665.78
HLA-A*01:01	SSYSLDAVY	0.269	0.267	456.67
HLA-A*01:01	KFDDTTGKY	0.321	0.356	689.87
HLA-A*26:01	DSTLTGGIY	0.457	0.291	829.12
HLA-A*26:01	NIPSWGGGY	0.109	0.031	53.45
HLA-A*26:01	WAVGASYVY	0.474	0.234	611.27
HLA-A*26:01	DAGALSFSY	0.420	0.791	2611.32
HLA-B*15:01	WQSNPDAYY	0.161	0.136	25.74
Protein no:3				

Binding Allele	Sequence no.	% Rank EL	% Rank BA	Aff(nM)
HLA-A*01:01	FLHALTEKY	0.372	0.588	1419.53
HLA-A*01:01	STDQLEYLA	0.483	0.201	309.28
HLA-A*02:01	AIDQYNMQL	0.352	2.880	525.76
HLA-A*02:01	MLEDASQFV	0.199	0.209	15.15
HLA-A*02:01	MTNLQNLNK	0.277	0.292	74.04
HLA-B*15:01	SQGEAIDQY	0.079	0.867	176.06
Protein no: 4				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	FSGSHNVYY	0.046	0.023	26.13
HLA-A*01:01	YSFTNSPDY	0.184	0.158	217.81
HLA-A*01:01	TSINFGGQY	0.271	0.353	682.54
HLA-B*58:01	VSLRKEQSY	0.269	1.247	429.61
HLA-A*01:01	ATGKFTITY	0.126	0.204	313.46
HLA-A*01:01	FVSFVTPSY	0.288	0.242	392.39
HLA-B*15:01	CQRPTNWTY	0.028	0.107	21.74
HLA-B*15:01	AVQDFTKSY	0.057	1.865	163.92
HLA-B*15:01	SIKLSNCAY	0.149	0.185	34.25
Protein no: 5				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	TTDQARDLL	0.473	0.473	1026.34
HLA-A*02:01	SLDQSDMAL	0.236	1.259	126.85
HLA-A*26:01	STADRILAL	0.100	0.091	193.47
HLA-A*26:01	GVVMAHTSY	0.358	0.633	2017.98
HLA-B*15:01	TGRKARYIA	0.062	0.051	14.27
Protein no:6				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	QSNLSIANY	0.219	0.464	998.32

HLA-A*01:01	LASGVARLY	0.422	0.640	1598.49
HLA-B*15:01	SQLPSQLGY	0.009	0.310	56.14
Protein no:7				
Binding Allele	**Sequence no.**	**% Rank EL**	**% Rank BA**	**Aff(nM)**
HLA-A*01:01	FTALESNKY	0.028	0.029	30.29
HLA-A*01:01	NIDYVVEAY	0.080	0.105	130.67
HLA-A*01:01	TTLENLSIY	0.217	0.362	706.91
HLA-A*26:01	ETRFKFNAY	0.018	0.014	29.68

After confirming the strong binding affinity of the CTL epitopes by NetMHCpan 4.1, their antigenicity, allergenicity, toxicity profile were thoroughly screened using Vaxijen v.2.0 server, Allertop v.2.0 server, Toxinpred. This study used Vaxijen v.2.0 server to predict the immunogenicity of all identified epitopes resulting in the discovery of 3 antigenic epitopes for Protein no. 1, 8 antigenic epitopes for Protein no.2, 2 antigenic epitopes for Protein no.3, 4 antigenic epitopes for Protein no. 4, 1 antigenic epitopes for Protein no. 5, 1 antigenic epitope for Protein no. 6. Moreover, 3 antigenic epitopes for Protein no.1, 2 antigenic epitopes for Protein no.2, 1 antigenic epitopes for Protein no.3, 1 antigenic epitopes for Protein no.4 , 1 antigenic epitope for Protein for Protein 5-7 were found to be non-allergenic predicted by Allertop v.2.0 and non-toxin predicted by Toxinpred. Afterwards, CTL epitopes that satisfied every requirement were selected to be included in the creation of vaccines.

Table 4: Antigenicity, Allergenicity, Toxicity determination of the Selected CTL epitopes

No.	Selected CTL Epitopes	Antigenicity	Allergenicity	Toxicity
Protein no.1	HTKFESEVY	Antigen	Non-Allergen	Non-toxin
	TINTSHVEY	Antigen	Non-Allergen	Non-toxin
	QMDGAILVV	Antigen	Non-Allergen	Non-toxin
Protein no.2	FDDTTGKYY	Antigen	Non-Allergen	Non-toxin
	KFDDTTGKY	Antigen	Non-Allergen	Non-toxin
Protein no.3	SQGEAIDQY	Antigen	Non-Allergen	Non-toxin
Protein no.4	VSLRKEQSY	Antigen	Non-Allergen	Non-toxin
Protein no.5	TGRKARYIA	Antigen	Non-Allergen	Non-toxin
Protein no.6	QSNLSIANY	Antigen	Non-Allergen	Non-toxin
Protein no.7	NIDYVVEAYY	Antigen	Non-Allergen	Non-toxin

3.3 Selection and Screening of HTL Epitopes

The HTL epitopes that were Strong Binders were identified using the NetMHC-IIpan 4.0 server. After that, the HTL epitopes that were identified were identified were further evaluated based on their cytokine stimulating profile using IFN epitope, IL-4 pred, IL-10 pred server. The epitopes which maintained all this criteria were further assessed for antigenicity, allergenicity, toxicity using relevant bio-informatics tools.

Table 5: HTL Epitopes for each protein (i-vii) with their Cytokine inducing Capability

Protein no: 1			
HTL Epitopes	**IFN**	**IL-4 Prediction**	**IL-10 Prediction**
KPFLLPIEDVFSISG	Positive	IL-4 Inducer	IL-10 Inducer
GDNIKMVVTLIHPIA	Positive	IL-4 Inducer	IL-10 Inducer
IKMVVTLIHPIAMDD	Positive	IL-4 Inducer	IL-10 Inducer
EKARGITINTSHVEY	Positive	IL-4 Inducer	IL-10 Inducer
LIHPIAMDDGLRFAI	Positive	IL-4 Inducer	IL-10 Inducer
IHPIAMDDGLRFAIR	Positive	IL-4 Inducer	IL-10 Inducer
Protein no: 2			
HTL Epitopes	**IFN**	**IL-4 Prediction**	**IL-10 Prediction**
GNIFQDERDVKFIVI	Positive	IL-4 Inducer	IL-10 Inducer

TEMDKFYQADKKTNV	Positive	IL-4 Inducer	IL-10 Inducer
GYGNIFQDERDVKFI	Positive	IL-4 Inducer	IL-10 Inducer
ISLYKAAAKFKYGPV	Positive	IL-4 Inducer	IL-10 Inducer
GLDIAAFTAIEMAEN	Positive	IL-4 Inducer	IL-10 Inducer
Protein no: 3			
HTL Epitopes	**IFN**	**IL-4 Prediction**	**IL-10 Prediction**
TEVVQKHTENIRVQD	Positive	IL-4 Inducer	IL-10 Inducer
HSNLEQVRADFRDSR	Positive	IL-4 Inducer	IL-10 Inducer
Protein no: 4			
HTL Epitopes	**IFN**	**IL-4 Prediction**	**IL-10 Prediction**
EQSYDYQFASSPAFS	Positive	IL-4 Inducer	IL-10 Inducer
NSPADWTLLKPNISS	Positive	IL-4 Inducer	IL-10 Inducer
SPADWTLLKPNISSS	Positive	IL-4 Inducer	IL-10 Inducer
Protein no: 5			
HTL Epitopes	**IFN**	**IL-4 Prediction**	**IL-10 Prediction**
QAFWDIDYDSGERYR	Positive	IL-4 Inducer	IL-10 Inducer
WGDSQRYSVDYSNTA	Positive	IL-4 Inducer	IL-10 Inducer
SQRYSVDYSNTAWGS	Positive	IL-4 Inducer	IL-10 Inducer
PLEQYYLVQGGFKRT	Positive	IL-4 Inducer	IL-10 Inducer
LEQYYLVQGGFKRTD	Positive	IL-4 Inducer	IL-10 Inducer
Protein no: 6			
HTL Epitopes	**IFN**	**IL-4 Prediction**	**IL-10 Prediction**
EKTARIGAKKKRRAR	Positive	IL-4 Inducer	IL-10 Inducer
Protein no: 7			
HTL Epitopes	**IFN**	**IL-10 Prediction**	**IL-4 Prediction**
DIIISQKRIGNLPAV	Positive	IL-10 Inducer	IL-4 Inducer
RKTADFTALESNKYE	Positive	IL-10 Inducer	IL-4 Inducer

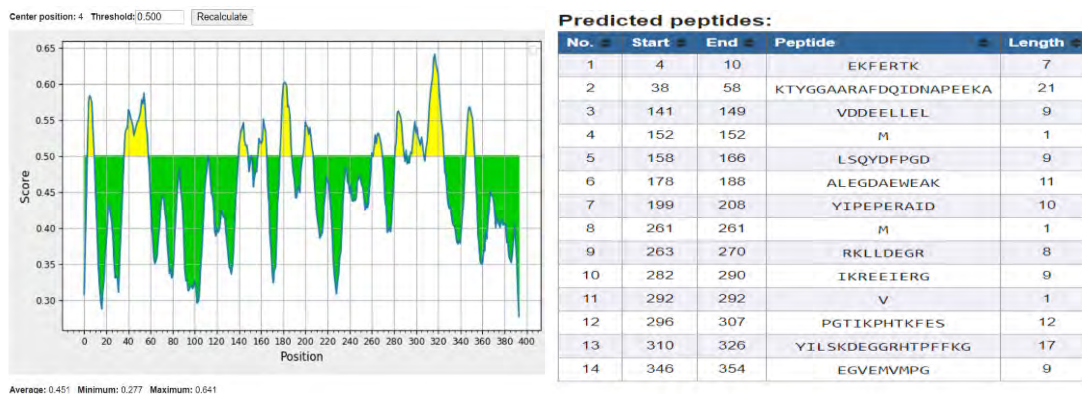
Table 06: Antigenicity, Allergenicity and Toxicity Assessment of the Selected HTL epitopes.

No.	Selected HTL Epitopes	Antigenicity	Allergenicity	Toxicity
Protein no.1	LIHPIAMDDGLRFAI	Antigen	Non-Allergen	Non-toxin
	IHPIAMDDGLRFAIR	Antigen	Non-Allergen	Non-toxin
Protein no.2	ISLYKAAAKFKYGPV	Antigen	Non-Allergen	Non-toxin
Protein no.3	TEVVQKHTENIRVQD	Antigen	Non-Allergen	Non-toxin
Protein no.4	NSPADWTLLKPNISS	Antigen	Non-Allergen	Non-toxin
Protein no.5	LEQYYLVQGGFKRTD	Antigen	Non-Allergen	Non-toxin
Protein no.6	EKTARIGAKKKRRAR	Antigen	Non-Allergen	Non-toxin
Protein no.7	RKTADFTALESNKYE	Antigen	Non-Allergen	Non-toxin

3.4 Selection and Screening of B cell epitopes

Using the IEDB Analysis Resource- B cell prediction tool, the B cell epitopes were identified. Among all the available methods present there, the Bepipred Linear Epitope prediction 2.0, the most advanced option was employed. A score vs. position graph was generated maintaining threshold at 0.5 where the green color portion represents non-desired B cell epitopes with scores below 0.5 and the yellow color part denotes the desired B cell epitopes with score above 0.5. For creating the vaccines the epitopes which demonstrated antigenic potential and showed no presence of any sort of allergens or toxins were desired which were obtained via various bioinformatics tools like: Allertop, Vaxijen, Toxinpred.

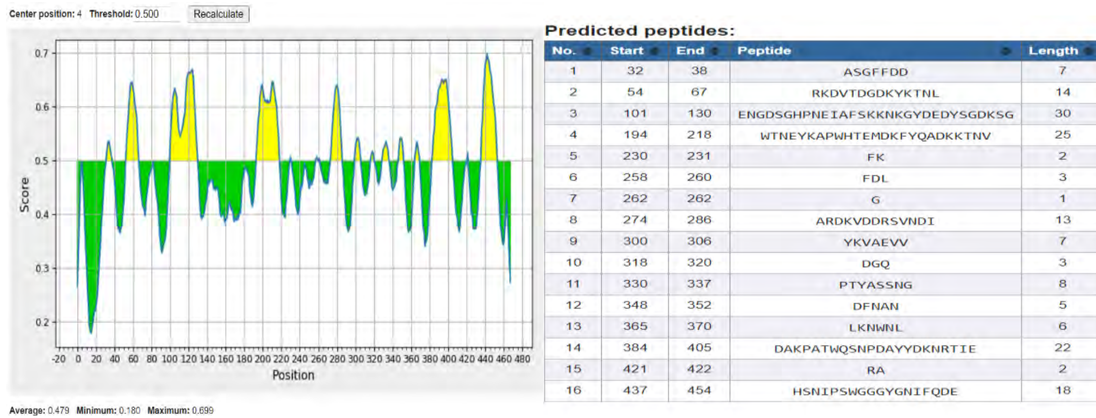
PROTEIN 1



No.	Start	End	Peptide	Length
1	4	10	EKFERTK	7
2	38	58	KTYGGAARAFDQIDNAPEEKA	21
3	141	149	VDDEELLEL	9
4	152	152	M	1
5	158	166	LSQYDFPGD	9
6	178	188	ALEGDAEWEAK	11
7	199	208	YIPEPERAID	10
8	261	261	M	1
9	263	270	RKLLDEGR	8
10	282	290	IKREEIERG	9
11	292	292	V	1
12	296	307	PGTIKPHTKFES	12
13	310	326	YILSKDEGGRHTPFFKG	17
14	346	354	EGVEMVMPG	9

Figure 10 (a)

PROTEIN 2

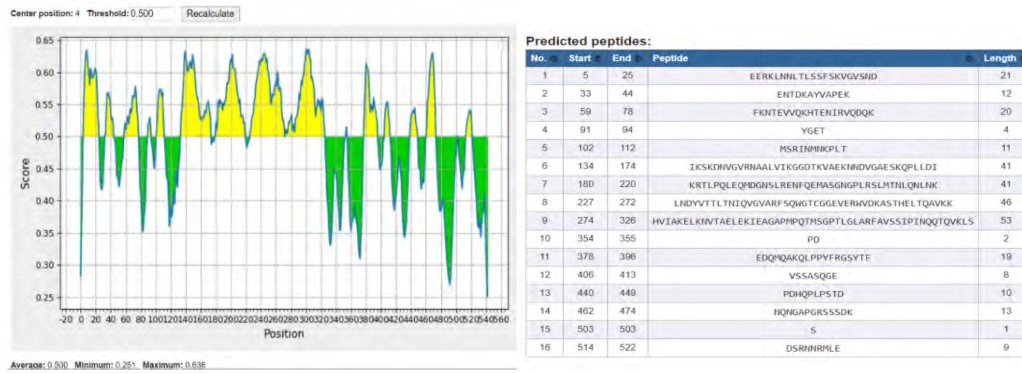


Predicted peptides:

No.	Start	End	Peptide	Length
1	32	38	ASGFFDD	7
2	54	67	RKDVTDGDKYKTNL	14
3	101	130	ENGDSGHPNEIAFSKKNKGYDEDYSGDKSG	30
4	194	218	WTNEYKAPWHTEMDKFYQADKKTNV	25
5	230	231	FK	2
6	258	260	FDL	3
7	262	262	G	1
8	274	286	ARDKVDDRSVNDI	13
9	300	306	YKVAEVV	7
10	318	320	DGQ	3
11	330	337	PTYASSNG	8
12	348	352	DFNAN	5
13	365	370	LKNWNL	6
14	384	405	DAKPATWQSNPDAYYDKNRTIE	22
15	421	422	RA	2
16	437	454	HSNIPSWGGGYGNIFQDE	18

Figure 10 (b)

PROTEIN 3

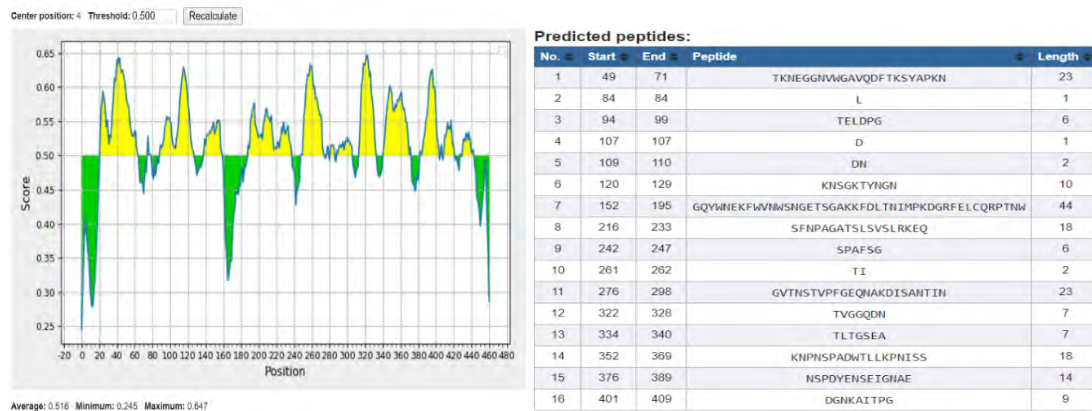


Predicted peptides:

No.	Start	End	Peptide	Length
1	5	25	EERKLNNLTLSSFSKVGVSND	21
2	33	44	ENTDKAYVAPEK	12
3	59	78	FKNTEVVQKHTENIRVQDQK	20
4	91	94	YGET	4
5	102	112	MSRINMNKPLT	11
6	134	174	IKSKDNVGVRNAALVIKGGDTKVAEKNNDVGAESKQPLLDI	41
7	180	220	KRTLPQLEQMDGNSLRENFQEMASGNGPLRSLMTNLQNLNK	41
8	227	272	LNDYVTTLTNIQVGVARFSQWGTCGGEVERWVDKASTHELTQAVKK	46
9	274	326	HVIAKELKNVTAELEKIEAGAPMPQTMSGPTLGLARFAVSSIPINQQTQVKLS	53
10	354	355	PD	2
11	378	396	EDQMQAKQLPPYFRGSYTF	19
12	406	413	VSSASQGE	8
13	440	449	PDHQPLPSTD	10
14	462	474	NQNGAPGRSSSDK	13
15	503	503	S	1
16	514	522	DSRNNRMLE	9

Figure 10 (c)

PROTEIN 4

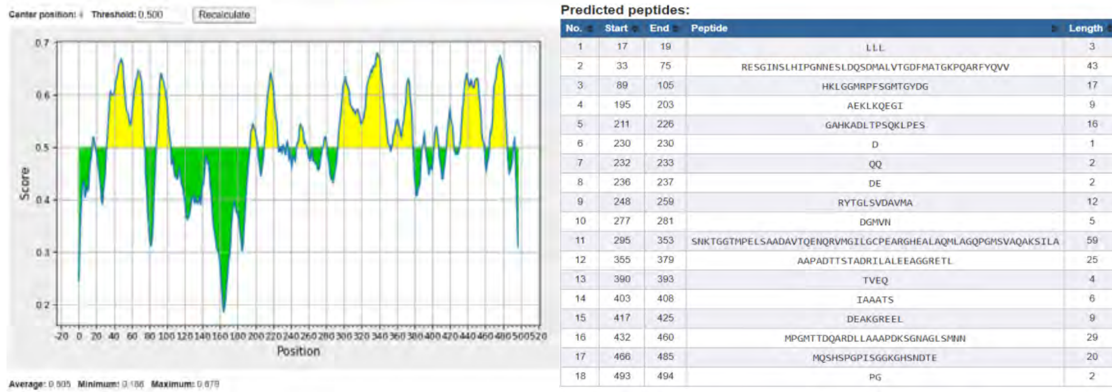


Predicted peptides:

No.	Start	End	Peptide	Length
1	49	71	TKNEGGNVWGAVQDFTKSYAPKN	23
2	84	84	L	1
3	94	99	TELDPG	6
4	107	107	D	1
5	109	110	DN	2
6	120	129	KNSGKTYNGN	10
7	152	195	GQYWNEKFWVNWSNGETSGAKKFDLTNIMPKDGRFELCQRPTNW	44
8	216	233	SFNPAGATSLSVSLRKEQ	18
9	242	247	SPAFSG	6
10	261	262	TI	2
11	276	298	GVTNSTVPFGEQNAKDISANTIN	23
12	322	328	TVGGQDN	7
13	334	340	TLTGSEA	7
14	352	369	KNPNSPADWTLLKPNISS	18
15	376	389	NSPDYENSEIGNAE	14
16	401	409	DGNKAITPG	9

Figure 10 (d)

PROTEIN 5

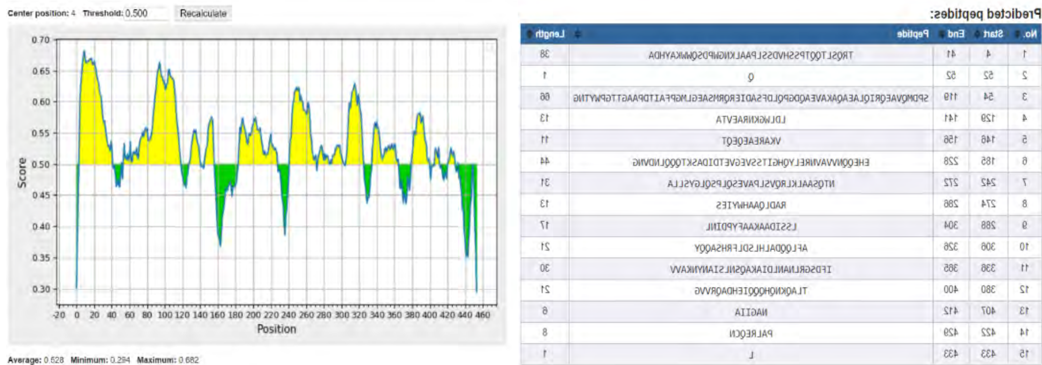


Predicted peptides:

No.	Start	End	Peptide	Length
1	17	19	LLL	3
2	33	75	RESGINSLHIPGNNESLDQSDMALVTGDFMATGKPQARFYQVV	43
3	89	105	HKLGGMRPFSGMTGYDG	17
4	195	203	AEKLKQEGI	9
5	211	226	GAHKADLTPSQKLPES	16
6	230	230	D	1
7	232	233	QQ	2
8	236	237	DE	2
9	248	259	RYTGLSVDAVMA	12
10	277	281	DGMVN	5
11	295	353	SNKTGGTMPELSAADAVTQENQRVMGILGCPEARGHEALAQMLAGQPGMSVAQAKSILA	59
12	355	379	AAPADTTSTADRILALEEAGGRETL	25
13	390	393	TVEQ	4
14	403	408	IAAATS	6
15	417	425	DEAKGREEL	9
16	432	460	MPGMTTDQARDLLAAAPDKSGNAGLSMNN	29
17	466	485	MQSHSPGPISGGKGHSNDTE	20
18	493	494	PG	2

Figure 10 (e)

PROTEIN 6

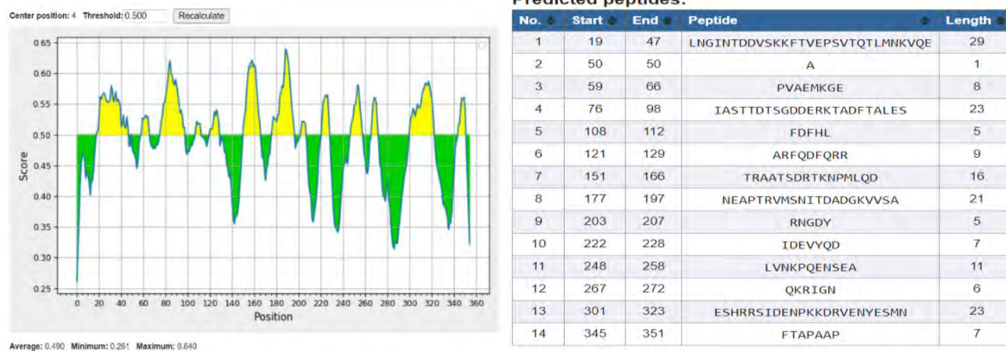


Predicted peptides:

No.	Start	End	Peptide	Length
1	4	41	TRQSLTQQTPSSHVDSSLPAALKNGWPDSQWWKAYHDA	38
2	52	52	Q	1
3	54	119	SPDMQVAEQRIQLAEAQAKAVEAQDGPQLDFSADIERQRMSAEGLMGPFAITDPAAAGTTGPWYTNG	66
4	129	141	LDLWGKNRAEVTA	13
5	146	156	VKAREAEQEQT	11
6	185	228	EHEQQNVVAVNRELYQHGITSSVEGVETDIDASKTQQQLNDVNG	44
7	242	272	NTQSAALKLRQVSLPAVESQLPSQLGYSLLA	31
8	274	286	RADLQAAHWYIES	13
9	288	304	LSSIDAAKAAFYPDINL	17
10	306	326	AFLQQDALHLSDLFRHSAQQY	21
11	336	365	IFDSGRLNANLDIAKAQSNLSIANYNKAVV	30
12	380	400	TLAQKNQHQQQIEHDAQRVVG	21
13	407	412	NAGIIA	6
14	422	429	PALREQCN	8
15	433	433	L	1

Figure 10 (f)

PROTEIN 7



Predicted peptides:

No.	Start	End	Peptide	Length
1	19	47	LNGINTDDVSKKFTVEPSVTQTLMNKVQE	29
2	50	50	A	1
3	59	66	PVAEMKGE	8
4	76	98	IASTTDTSGDDERKTADFTALES	23
5	108	112	FDFHL	5
6	121	129	ARFQDFQRR	9
7	151	166	TRAATSDRTKNPMLQD	16
8	177	197	NEAPTRVMSNITDADGKVVSA	21
9	203	207	RNGDY	5
10	222	228	IDEVYQD	7
11	248	258	LVNKPQENSEA	11
12	267	272	QKRIGN	6
13	301	323	ESHRRSIDENPKKDRVENYESMN	23
14	345	351	FTAPAAP	7

Figure 10 (g)

Figure 10 (a-g): Position vs. score graph and B-Cell epitopes for all proteins (Sun et al., 2013)

By analyzing the IEDB Analysis Resource- with Bepipred Linear Epitope Prediction 2.0 technique resulting in the identification of 11 peptides for Protein no. 1, 8 peptides for Protein no.2, 9 peptides for Protein no.3, 7 peptides for Protein no.4, 7 peptides for Protein no.5, 7 peptides for Protein no.6, 8 peptides for Protein no.7 as B cell epitopes. Considering that they have fewer possibilities to stimulate an immunological response, epitopes with incredibly short lengths were removed from the initial evaluation.

Table 6: Antigenicity, Allergenicity, toxicity profile assessment of selected B-Cell epitopes

No.	Selected B Cell Epitopes	Antigenicity	Allergenicity	Toxicity
Protein no.1	EKFERTK	Antigen	Non-Allergen	Non-toxin
Protein no.2	RKDVTDGDKYKTNL	Antigen	Non-Allergen	Non-toxin
Protein no.3	VSSASQGE	Antigen	Non-Allergen	Non-toxin
	DSRNNRMLE			
Protein no. 4	KNPNSPADWTLLKPNISS	Antigen	Non-Allergen	Non-toxin
Protein no.5	DEAKGREEL	Antigen	Non-Allergen	Non-toxin
Protein no.6	RADLQAAHWYIES	Antigen	Non-Allergen	Non-toxin
Protein no.7	PVAEMKGE	Antigen	Non-Allergen	Non-toxin

3.5 Construction of Vaccine using Suitable Linkers with their Antigenicity, Allergenicity and Toxicity Profile Assessment

The immunogenic Epitopes obtained from various bioinformatics tools were connected together by Linkers. To connect the epitopes, EAAAK (CTL epitope), GPGPG (HTL epitopes), KK (B-cell epitope) linkers were used. A total of 7 vaccines were prepared using the linkers from the 7 protein sequences.

Table 7 : Vaccine Construction for the 7 proteins made by incorporating different epitopes and linkers

	Construction of Vaccine
Vaccine 1 (Protein 1)	MSKEKFERTKPHVNVGTIGHVDHGKTTLTAAITTVLAKTYGGAARAFDQIDNAPEEKARGITINTSHVEYDTPTRHYAHVDCPGHADYVKN MITGAAQMDGAILVVAATDGPMPQTREHILLGRQVGVPYIIVFLNKCDMVDDEELLELVEMEVRELLSQYDFPGDDTPIVRGSALKALEGD AEWEAKIIELAGFLDSYIPEPERAIDKPFLLPIEDVFSISGRGTVVTGRVERGIIKVGEEVEIVGIKETQKSTCTGVEMFRKLLDEGRAGENVG DNIKMVVTLIHPIAIKMVVTLIHPIAMDDGVLLRGIKREEIERGQVLAKPGTIKPHTKFESEVYILSKDEGGRHTPFFKGYRPQFYFRTTDVT GTIELPEGVEMVMPGDNIKMVVTLIHPIAMDDGLRFAIREGGRTVGAGVVAKVLGEAAAKQMDGAILVVEAAAKTINTSHVEYGPGPGLIH PIAMDDGLRFAIGPGPGIHPIAMDDGLRFAIRKKEKFERTK
Vaccine 2 (Protein 2)	MRTFSGKRSTLALAIAGITAMSGWIVVPQAQASGFFDDSTLTGGIYYWQRERDRKDVTDGDKYKTNLSHATWNANLDFQSGYAADMFGLD IAAFTAIEMAENGDSGHPNEIAFSKKNKGYDEDYSGDKSGISEAAKTASVVKMSYLYKAAAKFKYGPVWARAGYIQPTGQTLLAPHWSFMP GTYQGAEAGASFDYGDAGALSFSYMWTNEYKAPWHTEMDKFYQADKKTNVDYLHSIGAKYDFKNDLVLEAAFGQSEGYVDQYFAKASY KFDLGGNPFTTSYQFYGARDKVDDRSVNDIYDGTAWLQALTFGYKVAEVVDLRLEGTWVKADGQQGYFLQRMTPTYASSNGRLDIWWDN RSDFNANGEKAVFFGAMYDLKNWNLPGWAVGASYVVYAWDAKPATWQSNPDAYYDKNRTIEESSYSLDAVYTLQEGRAKGTMFKLHFTEYD NHSNIPSWGGGYGNIFQDERDVKFIVIAPFTIFEAAAKFDDTTGKYYGPGPGISLYKAAAKFKYGPVKKRKDVTDGDKYKTNL
Vaccine 3 (Protein 3)	MLKYEERKLNNLTLSSFSKVGVSNDARLYIAKENTDKAYVAPEKFSSKVLTWLGKMPLFKNTEVVQKHTENIRVQDQKILQTFLHALTEKY GETAVNDALLMSRINMNKPLTQRLAVQITECVKAADEGFINLIKSKDNVGVRNAALVIKGGDTKVAEKNNDVGAESKQPLLDIALKGLKRTL PQGPGPGDSRGNYRYTDKDLVKLEQMDGNSLRENFQEMASGNGPLRSLMTNLQNLNKIPEAKQLNDYVTTLTNIQVGVARFSQWGTCGG EVERWVDKASTHELTQAVKKIHVIAKELKNVTAELEKIEAGAPMPQTMSGPTLGLARFAVSSIPINQQTQVKLSDGMPVPVNTLTFDGKPVA LAGSYPKNTPDALEAHMKMLLEKECSCLVVLTSEDQMQAKQLPPYFRGSYTFGEVHTNSQKVSSASQGEAIDQYNMQLSCGEKRYTIPVL HVKNWPDHQPLPSTDQLEYLADRVKNSNQNGAPGRSSSDKHLPMIHCLGGVGRTGTMAAALVLKDNPHSNLEQVRADFRDSRNNRMLED AQFVQLKAMQAQLLMTTASEAAAKMTNLQNLNKEAAAKMLEDASQFVGPGPGTEVVQKHTENIRVQDKKDSRNNRMLEKKVSSASQGE
Vaccine 4 (Protein 4)	MITKCDHRLMLLNTQRLKSVIKIISLLLLSFVSFVTPSYGWSLYVGQYTKNEGGNVWGAVQDFTKSYAPKNMLNPIAIHNGSGLREISVSND ATELDPGTGLSCNKDHDNHVLTLYHGYKNSGKTYNGNMLWETNTPGMYMGIEVTSINFGGQYWNEKFWVNWSNGETSGAKKFDLTNIM PKDGRFELCQRPTNWTYFGLGGIVLWVKYHLYIDDSFNPAGATSLSVSLRKEQSYDYQFASSPAFSGSHNVYYNIAPNQLTINYPTCQANAVT GDGVTNSTVPFGEQNAKDISANTINRKFSIKLSNCAYVKELNVTLSSATVGGQDNTLLGNTLTGSEAAGGIGVMIEGEKNPNSPADWTLLKP NISSSIYSFTNSPDYENSEIGNAEQVMNFQATLKQDGNKAITPGKFKATGKFTITYPEAAAKVSLRKEQSYGPGPGNSPADWTLLKPNISSKK KNPNSPADWTLLKPNISS
Vaccine 5 (Protein 5)	MQRNLSHIISQATSAPLLLEPAYARVFFCALGRESGINSLHIPGNNESLDQSDMALVTGDFMATGKPQARFYQVVNGIAVLPVTGTLVHKLGG MRPFSGMTGYDGVTARLQQAVSDPEVKGILLDIDSPGGQAAGAFDCADMIYRMREQKPVWALANETACSAAMLLAAACSHRLVTQTSRM GSIGVVMAHTSYAEKLKQEGIDITLIYSGAHKADLTPSQKLPESVYADYQQRMDEARKMFAEKVARYTGLSVDAVMATEAAVYDGQAIITTG LADGMVNAADAIGVMAEAINSNKTGGTMPELSAADAVTQENQRVMGILGCPEARGHEALAQMLAGQPGMSVAQAKSILAAAAPADTTSTA DRILALEEAGGRETLAQTLAAMPEMTVEQARTILAASPIAAATSLHDAVMALDEAKGREELAEKLAVMPGMTTDQARDLLAAAPDKSGNA GLSMNNAFDAFMQSHSPGPISGGKGHSNDTETTLLMSIPGTSATEAAAKTGRKARYIAGPGPGLIHPIAMDDGLRFAIGPGPGIHP IAMDDGLRFAIRKKDEAKGREEL
Vaccine 6 (Protein 6)	MHDTRQSLTQQTPSSHVDSSLPAALKNGWPDSQWWKAYHDAQLDALIDNAIQHSPDMQVAEQRIQLAEAQAKAVEAQDGPQLDFSADIER QRMSAEGLMGPFAITDPAAGTTGPWYTNGTFGLTAGWDLDLWGKNRAEVTARIGAVKAREAEQEQTRQLLASGVARLYWEWQTQAALK NVLMQIEHEQQNVVAVNRELYQHGITSSVEGVETDIDASKTQQQLNDVNGKMKVIEARLSALTNTQSAALKLRQVSLPAVESQLPSQLGYSL LARRADLQAAHWYIESSLSSIDAAKAAFYPDINLMAFLQQDALHLSDLFRHSAQQYGITGGLTLPIFDSGRLNANLDIAKAQSNLSIANYNKA VVDAVNDVARAASQVETLAQKNQHQQQIEHDAQRVVGLAQARFNAGIIAGSRVSEAKIPALREQCNGLLLQGQWLDASIQLTSALGGGYHS EAAAKQSNLSIANYGPGPGEVTARIGAVKAREAEKKRADLQAAHWYIES
Vaccine 7 (Protein 7)	MRPETRFKFNAYLTRVAELNGINTDDVSKKFTVEPSVTQTLMNKVQESSAFLQTINILPVAEMKGEKIGVGVTGTIASTTDTSGDDERKTADF TALESNKYECDQINFDFHLKYKTLDLWARFQDFQRRIRDAIVKRQALDFIMAGFNGTTRAATSDRTKNPMLQDVAVGWLQKYRNEAPTRV MSNITDADGKVVSAVIRVGRNGDYENLDALVMDATNNLIDEVYQDDPKLVAIVGRKLLADKYFPLVNKPQENSEALAADIIISQKRIGNLPAV RVPYFPANAVLVTTLENLSIYFMDESHRRSIDENPKKDRVENYESMNIDYVVEAYAAGCLLENITLGDFTAPAAPESGAEAAAKNIDYVVEAY GPGPGRKTADFTALESNKYEGPGPGDIIISQKRIGNLPAVKKPVAEMKGE

After preparation of the constructed vaccine, all of them were assessed for their antigenicity, allergenicity and Toxicity to observe whether they demonstrate immunogenic properties. The antigenicity of vaccine candidates verified using Vaxijen, allergenicity by Allergen online server & toxicity by T3DB server.

Table 8: Assessment of the constructed vaccines

No.	Antigenicity	Presence of Allergens	Presence of Toxins
Vaccine-(i)	0.4675	Number of sequence with hits: 0	Your search returned no results
Vaccine-(ii)	0.6216	Number of sequence with hits: 0	Your search returned no results
Vaccine-(iii)	0.5871	Number of sequence with hits: 0	Your search returned no results
Vaccine-(iv)	0.6408	Number of sequence with hits: 0	Your search returned no results
Vaccine-(v)	0.5821	Number of sequence with hits: 0	Your search returned no results
Vaccine-(vi)	0.6733	Number of sequence with hits: 0	Your search returned no results
Vaccine-(vii)	0.6094	Number of sequence with hits: 0	Your search returned no results

3.6 Biochemical Analysis of the Constructed Vaccines

To assess the biochemical characteristics of the vaccines that have been developed, an online server called Protparam was used. This server provides information whether the vaccine is eligible or not by giving results of physicochemical parameters like: molecular weight, molecular formula, instability index, GRAVY ,amino acid number, molecular weight, therapeutic pI.

Table 9: Physico-chemical parameters of the constructed vaccines using Protparam Tool.

Parameters	Vaccine (i)	Vaccine (ii)	Vaccine (iii)	Vaccine (iv)	Vaccine (v)	Vaccine (vi)	Vaccine (vii)
Molecular weight	54755.08	62047.84	69675.40	52187.55	58896.10	54811.23	44462.20
Theoretical PI	5.53	5.78	8.86	8.34	5.27	8.25	5.06
Instability Index	28.10	27.09	34.06	29.78	34.26	38.74	34.68
Aliphatic Index	93.39	55.87	80.06	72.33	83.41	87.30	81.68
GRAVY(Grand Average of Hydropathicity)	-0.123	-0.578	-0.530	-0.433	-0.063	-0.375	-0.408

Observing the table we can observe that, GRAVY scores were negative, signifying that the vaccine is hydrophilic. GRAVY mainly indicates whether the vaccine is hydrophobic or hydrophilic in nature. The Hydrophilic nature is often required for vaccine development otherwise it might lead to toxicity and poor absorption in the body. The Instability Index was less than 40 and the therapeutic pI was also in acceptable range.

Number of amino acids: 503

Molecular weight: 54811.23

Theoretical pI: 8.25

```
Atomic composition:
Carbon      C         2398
Hydrogen    H         3794
Nitrogen    N          702
Oxygen      O          756
Sulfur      S            8

Formula: C2398H3794N702O756S8
Total number of atoms: 7658

Extinction coefficients:
Extinction coefficients are in units of  M-1 cm-1, at 280 nm measured in water.
Ext. coefficient    78380
Abs 0.1% (=1 g/l)   1.430, assuming all pairs of Cys residues form cystines


Ext. coefficient    78380
Abs 0.1% (=1 g/l)   1.430, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
                            >20 hours (yeast, in vivo).
                            >10 hours (Escherichia coli, in vivo).


Instability index:
The instability index (II) is computed to be 38.74
This classifies the protein as stable.

Aliphatic index: 87.30
Grand average of hydropathicity (GRAVY):-0.375
```

Figure 11: Protparam Results for Vaccine (vi) (Garg et al., 2016)

3.7 Homology Modelling of the Prepared Vaccines

For generating the 3D model for the constructed vaccine Phyre 2, a web based bioinformatics tool was used which can effectively predict protein structure, function, can build 3D models, predict binding site of the ligands. The 7 sequences were submitted one after another which gave detailed results on the confidence and coverage of the vaccine sequence within 30mins to 2hrs time period shown in the table below.

Table 10: Coverage and Confidence determination of the Constructed Vaccines using PHYRE-2

Parameters	Vaccine (i)	Vaccine (ii)	Vaccine (iii)	Vaccine (iv)	Vaccine (v)	Vaccine (vi)	Vaccine (vii)
Coverage	74%	81%	54%	70%	61%	82%	78%
Confidence	100%	100%	100%	99.9%	100%	100%	100%

From the table we can see, that Vaccine no. (vi) and (ii) showed coverage around 82% and rest of the other vaccine candidates showed results less than this (less than 80%) which was not desirable for this study. The confidence and Coverage for vaccine (ii) & (vi) is shown below:

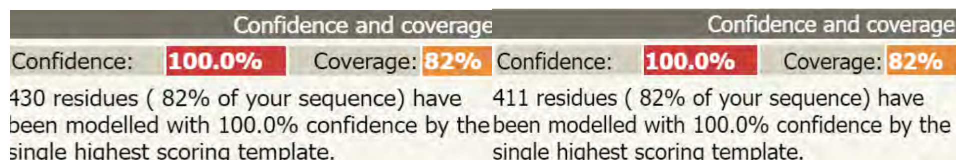


Figure 12: Phyre-2 results for vaccine (ii) & (vi) (Kelley et al., 2015)

3.8 Z-score Determination

ProSa web facilated to calculate the vaccine models' Z score, which represents their overall quality. It uses interactive web application tool to display quality scores and energy plots. It not only analyzes vaccine sequences but also refines and validates the vaccine sequences obtained from the X-ray region and NMR spectroscopy .The PDB (Protein Data Bank) file was obtained from Phyre 2 server. The local quality graph (knowledge- based energy vs sequence position of the model) shows some residues rise which exhibits both negative (residues rise below baseline) and positive energy (residues rise above baseline).

Table 11: Z-scores for the constructed vaccines (i-vii)

	Vaccine-(i)	Vaccine (ii)	Vaccine (iii)	Vaccine (iv)	Vaccine (v)	Vaccine (vi)	Vaccine (vii)
Z-scores	-4.56	-2.33	-5.12	-4.66	-4.65	-5.56	-4.88

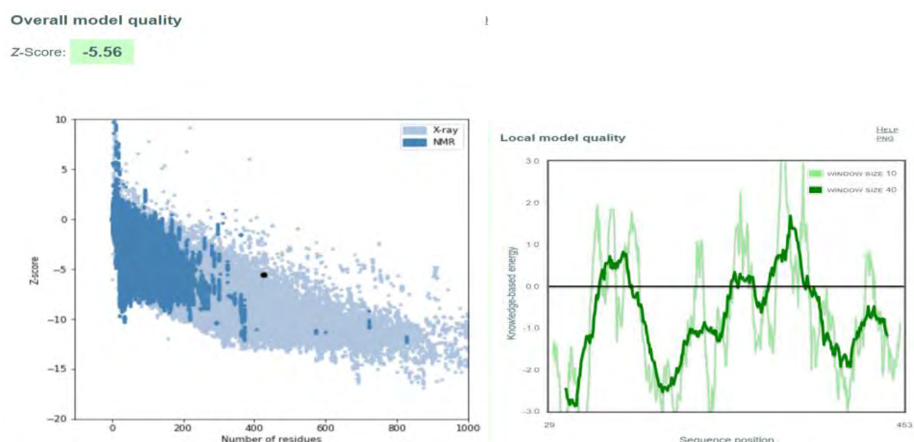


Figure 13: Overall model quality and local model quality of Vaccine (vi) (Wiederstein & Sippl, 2007)

3.9 Ramachandran Plots Prediction

The Ramachandran Plots Prediction was assessed through Swiss Model Expasy by inputting the PDB file obtained from Phyre 2 server. The Ramachandran Favored region and Ramachandran outlier for the 7 vaccine candidates are shown below in tabular form.

Table 12 : Ramachandran Plot Prediction for all the vaccines

Parameters	Vaccine (i)	Vaccine (ii)	Vaccine (iii)	Vaccine (iv)	Vaccine (v)	Vaccine (vi)	Vaccine (vii)
Ramachandran Favored	85.98%	93.51%	87.98%	86.49%	85.47%	95.51%	90.99%
Ramachandran Outlier	1.65%	1.57%	2.46%	6.03%	6.98%	1.89%	1.16%
Rotamer outliers	0.00%	0.00%	0.00%	0.00%	0.38%	0.00%	0.00%
MolProbity	2.23	2.65	2.77	3.17	3.24	2.52	2.53

From the table we can see that, the highest one is vaccine no 6 where the Ramachandran Favored region is 95.51% and Ramachandran Outlier region is 1.89%. So, the result is acceptable as Ramachandran Favored region is greater than 90% and Ramachandran outlier less than 2%.

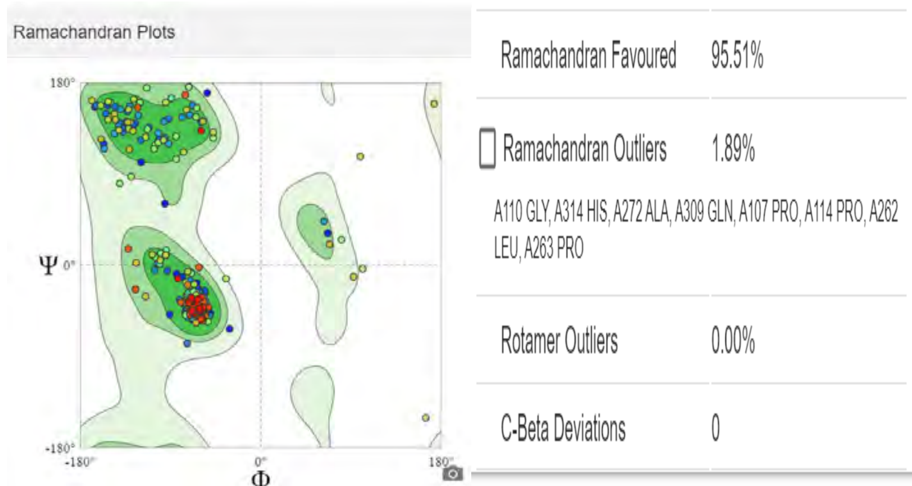


Figure 14: Ramachandran plot prediction of vaccine (vi) using SWISS-MODEL Tool. (Schwede et al., 2003)

Table 13: Summary table of the all the parameters assessed for the vaccine candidates

Parameters	Server	Vaccine (i)	Vaccine (ii)	Vaccine (iii)	Vaccine (iv)	Vaccine (v)	Vaccine (vi) FINAL VACCINE	Vaccine (vii)
Antigenicity of vaccine	**Vaxijen V.2.0**	0.4675	0.6216	0.5871	0.6408	0.5821	0.6733	0.6094
Allergenicity	**Allergen Online**	None	None	None	None	None	None	None
Toxicity	**T3DB**	None	None	None	None	None	None	None
Theoretical Isoelectric point (pI)	**Protparam**	5.53	5.78	8.86	8.34	5.27	8.25	5.06
Instability Index		28.10	27.09	34.06	29.78	34.26	38.74	34.68
GRAVY		-0.123	-0.578	-0.530	-0.433	-0.063	-0.375	-0.408
Confidence	**PHYRE-2**	100%	100%	100%	99.9%	100%	100%	100%
Coverage		74%	82%	54%	70%	61%	82%	78%

Ramachandran Favored	Ramachandran SWISS MODEL	85.98%	93.51%	87.98%	86.49%	85.47%	95.51%	90.99%
Ramachandran Outlier		1.65%	1.57%	2.46%	6.03%	6.98%	1.89%	1.16%

Based on the Summary table, The Vaccine (vi) meets every prerequisite. Additionally, the values of the physicochemical parameters were also between acceptable range: the therapeutic pI value was within the range of 7-9, the instability index (38.74) was lower than 40, and the GRAVY value was negative (-0.375). This suggests that the vaccine is stable and has a hydrophilic feature. The vaccine's antigenicity also changed, going from 0.6122 to 0.6733, suggesting that it may cause to initiate an immune reaction. It didn't demonstrate any toxicity or allergenicity. Z-score of the was observed plotted in acceptable range and the Ramachandran Favored region is greater than 90% (95.51%) and Ramachandran outlier is less than 2% (1.89%). Vaccine (ii) also had showed desirable results but vaccine (vi) appears to have slightly better characteristics for vaccine designing compared to other proteins. The main aim of this research was assess each vaccine candidate individually for their Antigenicity, physicochemical properties and structural stability and then compare them to determine the best vaccine candidate. Vaccine (vi) made from protein (vi) Lipoprotein has higher vaccine antigenicity score, higher percentage of residues in Favored regions of Ramachandran plots which suggest that it may be more promising candidate for further investigation compared to other proteins . Vaccine (ii) is stable too like protein (vi) in terms of Z-score but it can be outweighed by the structural stability and high antigenicity scores of Vaccine (vi). For further analysis , we will progress with vaccine (vi) and investigate it's capacity to produce immune response and check it's binding affinity .To validate this molecular docking and immune simulation was done to observe its potential for vaccine development.

3.10 Molecular Docking with Toll like receptor

By using ClusPro, molecular docking was conducted where Toll like receptor-4 was selected for ligand binding. The TLR-4 PDB file was uploaded in the receptor option and the PDB file obtained from PHYRE 2 was uploaded in the ligand folder. This resulted in 10 docked complexes out of which the first docked complex was the most stable docked complex. The score of highest docked complex was (-1307.8). The best binding Cluster was selected which had both high cluster population and reduced energy. “Cluster 0” was chosen for its greatest cluster membership (83) and lowest energy ratings (-1307.8) for the vaccine candidate (vii) and TLR-4 receptor.

Cluster	Members	Representative	Weighted Score
0	83	Center	-1105.7
		Lowest Energy	-1307.8
1	82	Center	-1068.9
		Lowest Energy	-1375.2
2	63	Center	-1113.3
		Lowest Energy	-1205.4
3	59	Center	-1105.1
		Lowest Energy	-1304.8
4	56	Center	-1199.7
		Lowest Energy	-1265.8
5	56	Center	-1157.5
		Lowest Energy	-1174.0

Figure 15: Highest stable docked complex obtained via ClusPro (Comeau et al., 2004)

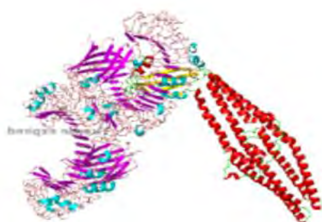
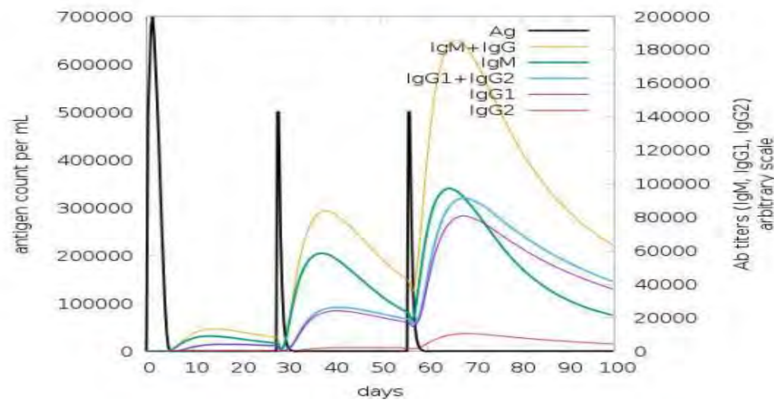


Figure 16: Molecular Docking performed between TLR-4 & vaccine (vi) using ClusPro (Comeau et al., 2004)

3.11 Immune Simulation using C-ImmSim

We used C-ImmSim to review immune response simulation of vaccine (vi). After inputting the sequence, the server showed normal termination and graphical representation shown below in Fig-18. Following vaccination, the antigen level rises as demonstrated in graph (a). Nevertheless, it also demonstrated that IgG and IgM peaked after 28 days, and antibody levels increased significantly after 60 days, indicating that the vaccine was successful to provoke an immune reaction & lymphocyte generates both antibodies as well as memory cells. These memory cells helps in storing the memory so that when the Salmonella pathogen again attacks the body's immune system, it can make a faster immune response.

Figure 17: Immune simulation of vaccine (vi) illustrated graphically via C-ImmSim (Rapin et al., 2010)



Chapter 4

Discussion

Globally, Salmonella Typhimurium is a leading cause of foodborne illnesses that causes gastrointestinal problems like: Gastroenteritis which may end in development of chronic diseases like colon cancer, gallbladder cancer etc. Development of an effective vaccine is necessary for public health due to rising incidence of antibiotic resistant strains. The In-Silico method used here

provides a new way of accelerating the development of vaccines possibly cutting the timeframe from years to only a few months.

By targeting 7 proteins involved in virulence and survival we seek to identify the most effective vaccine candidate against Salmonella Typhimurium which provides a unique and systematic comparison. For each target protein, we identified and evaluated the epitope sequences. After that, each protein were analyzed for their vaccine potential focusing on their antigenicity, allergenicity, toxicity, physicochemical properties, structural stability. Then, a comparative analysis was done to determine the optimal vaccine candidate. The comparative study focused on identifying the protein with most favorable characteristic for vaccine development. The comparative study revealed that Lipoprotein showed the highest antigenicity among other proteins. Lipoprotein showed superior coverage of epitope sequences, better Z- scores and high percentage of residues in favored regions of Ramachandran plots compared to other proteins. Physicochemical properties are important for vaccine's stability, solubility which in turn affects its efficacy and shelf life. Lipoprotein favorable scores in all of these parameters shows that it can be a potential vaccine candidate. Molecular Docking was also done with Toll-like receptors which provided insights into potential immunogenicity and safety of constructed Vaccine. At last, the vaccine made from protein – (vi) was further checked for its capacity to produce antibody (IgG, IgM Antibody & CD4, CD8 cells) in the body after administration through C-Immsim server. It generated some graphical representations of inducing immune response in body after giving the vaccine doses which re-validates the antibody producing capacity of the constructed vaccine-(vi). Lipoprotein is also less commonly used target but its characteristics make it a good option for further investigation and its novel nature might address the gaps in the current vaccine strategies for Salmonella Typhimurium. Bacterial Lipoprotein have also been used in some vaccine for disease caused by other pathogens.

Despite the results, it is essential to acknowledge the limitations of this study. Following the in silico method, biological system and laboratory based studies are required to conform the vaccine's reliability and effectiveness and to mitigate from any adverse effects later on.

Chapter 5
Conclusion

To conclude, this research mainly explored the potential of vaccine incorporating various multiple antigenic epitopes against Salmonella Typhimurium by concentrating seven different proteins. By using in-silico approaches, firstly epitope sequence for each target protein was selected and evaluated. After that, each protein were analyzed for their vaccine potential focusing on their antigenicity, allergenicity, toxicity, physicochemical properties, structural stability. Lastly, a comparative analysis was preformed among all of them to determine the most optimal vaccine target. Out of all the vaccine candidates, Lipoprotein emerged as the most promising candidate for vaccine design for superior antigenicity, physicochemical properties, and structural stability. In this way, it provides a new strategy for combating this pathogen due to increase cases of antibiotic resistance worldwide. However, further investigation including human trials and laboratory based study is essential for the development of safe and effective vaccine.

References

Alghamdi, S. (2021). The role of vaccines in combating antimicrobial resistance (AMR) bacteria. *Saudi Journal of Biological Sciences*, *28*.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., & Yeh, L. S. L. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *46*.

Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004). ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Research*, *32*.

de Jong, H. K., Parry, C. M., van der Poll, T., & Wiersinga, W. J. (2012). Host–Pathogen Interaction in Invasive Salmonellosis. *PLoS Pathogens*, *8*.

Dhanda, S. K., Gupta, S., Vir, P., & Raghava, G. P. (2013). Prediction of IL4 inducing peptides. *Clinical & Developmental Immunology*, *2013*.

Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2 - A server for in silico prediction of allergens. *Journal of Molecular Modeling*, *20*.

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*.

Garg, V. K., Avashthi, H., Tiwari, A., Jain, P. A., Ramkete, P. W. R., Kayastha, A. M., & Singh, V. K. (2016). MFPPI – Multi FASTA ProtParam Interface. *Bioinformation*, *12*.

Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., Baumert, J. L., Bohle, B., Lalithambika, S., Wise, J., & Taylor, S. L. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular Nutrition & Food Research*, *60*.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLOS ONE*, *8*.

Hoelzer, K., Switt, A. I. M., & Wiedmann, M. (2011). Animal contact as a source of human non-typhoidal salmonellosis. *Veterinary Research*, *42*.

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

Kupani, M., Pandey, R. K., Vashisht, S., Singh, S., Prajapati, V. K., & Mehrotra, S. (2023). Prediction of an immunogenic peptide ensemble and multi-subunit vaccine for Visceral leishmaniasis using bioinformatics approaches. *Heliyon*, *9*.

Lamichhane, B., Mawad, A. M. M., Saleh, M., Kelley, W. G., Harrington, P. J., Lovestad, C. W., Amezcua, J., Sarhan, M. M., El Zowalaty, M. E., Ramadan, H., Morgan, M., & Helmy, Y. A. (2024c). Salmonellosis: An Overview of Epidemiology, Pathogenesis, and Innovative Approaches to Mitigate the Antimicrobial Resistant Infections. In *Antibiotics*. Multidisciplinary Digital Publishing Institute (MDPI).

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*, *8*.

Leung, D. T., Das, S. K., Malek, M. A., Ahmed, D., Khanam, F., Qadri, F., Faruque, A. S. G., & Ryan, E. T. (2013). Non-Typhoidal Salmonella Gastroenteritis at a Diarrheal Hospital in Dhaka, Bangladesh, 1996–2011. *The American Journal of Tropical Medicine and Hygiene*, *88*.

Majowicz, S. E., Musto, J., Scallan, E., Angulo, F. J., Kirk, M., O'Brien, S. J., Jones, T. F., Fazil, A., & Hoekstra, R. M. (2010). The global burden of nontyphoidal salmonella gastroenteritis. *Clinical Infectious Diseases*, *50*.

McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., … Wilson, R. K. (2001). Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature 2001*.

Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational Immunology Meets Bioinformatics: The Use of Prediction Tools for Molecular Binding in the Simulation of the Immune System. *PLOS ONE*, *5*.

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

Rosenberger, C. M., Scott, M. G., Gold, M. R., Hancock, R. E. W., & Finlay, B. B. (2000). Salmonella typhimurium Infection and Lipopolysaccharide Stimulation Induce Similar Changes in Macrophage Gene Expression. *The Journal of Immunology*, *164*.

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, *31*.

Singh, O., Hsu, W. L., & Su, E. C. Y. (2021). ILeukin10Pred: A Computational Approach for Predicting IL-10-Inducing Immunosuppressive Peptides Using Combinations of Amino Acid Global Features. *Biology*, *11*.

Slauch, J. M., Mahan, M. J., Michetti, P., Neutra, M. R., & Mekalanos, J. J. (1995). Acetylation (O-factor 5) affects the structural and immunological properties of Salmonella typhimurium lipopolysaccharide O antigen. *Infection and Immunity*, *63*.

Sun, P., Ju, H., Liu, Z., Ning, Q., Zhang, J., Zhao, X., Huang, Y., Ma, Z., & Li, Y. (2013). Bioinformatics resources and tools for conformational B-cell epitope prediction. *Computational and Mathematical Methods in Medicine*, *2013*.

van Elsland, D. M., Duijster, J. W., Zhang, J., Stévenin, V., Zhang, Y., Zha, L., Xia, Y., Franz, E., Sun, J., Mughini-Gras, L., & Neefjes, J. (2022). Repetitive non-typhoidal Salmonella exposure is an environmental risk factor for colon cancer and tumor growth. *Cell Reports Medicine*, *3*.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*.

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

Won, G., & Lee, J. H. (2017). Salmonella Typhimurium, the major causative agent of foodborne illness inactivated by a phage lysis system provides effective protection against lethal challenge by induction of robust cell-mediated immune responses and activation of dendritic cells. *Veterinary Research*, *48*.