CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Reverse Vaccinology Approach for Vaccine Development Against *Streptococcus agalactiae*

by

Kashaf Sajjad

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2022

I dedicate this thesis to all the great people came in my life specially my beloved Parents and my Supervisor who encouraged me to stand out in this world with nobility and motivated me to step ahead without any fear.

CERTIFICATE OF APPROVAL

Reverse Vaccinology Approach for Vaccine Development Against *Streptococcus agalactiae*

by

Kashaf Sajjad

(MBS203024)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Uzma Abdullah	AAU, Rawalpindi
(b)	Internal Examiner	Dr. Arshia Amin Butt	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
December, 2022

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. and Biosciences
December, 2022

Dr. Sahar Fazal
Dean
Faculty of Health and Life Sciences
December, 2022

Author's Declaration

I, **Kashaf Sajjad** hereby state that my MS thesis titled "**Reverse Vaccinology Approach for Vaccine Development Against *Streptococcus agalactiae***" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Kashaf Sajjad)

Registration No: MBS203024

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Reverse Vaccinology Approach for Vaccine Development Against *Streptococcus agalactiae***" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Kashaf Sajjad)
Registration No: MBS203024

Acknowledgement

I would like to thanks Allah Almighty, The Most Magnificent and Compassionate, indeed, all praises are due to Him and His Holy Prophet Muhammad (PBUH). They gave me the strength and aptitude to complete this target. I want to acknowledge the efforts of my thesis supervisor **Dr. Syeda Marriam Bakhtiar**, Department of Biosciences and Bioinformatics, whose encouragement, guidance and support helped us to complete our project. I want to give my sincere gratitude to my co-supervisor **Dr.Syed Babar Jamal Bacha**, National University of Medical Sciences who supported, motivated and guided me throughout my research journey and without whom it would have been difficult for me to complete this study. I want to say thank you to **Dr. Muhammad Faheem** for his continues help during my research work. I want to acknowledge Dean of Faculty of Health and Life Sciences, Dr. Sahar Fazal and head of Department of Bioinformatics and Biosciences, Dr. Syeda Marriam Bakhtiar, for giving me the opportunity to persue MS with thesis and complete my research within time. I owe a great deal of appreciation and gratitude to all the Faculty members, Dr. Shaukat Iqbal Malik, Dr. Erum Dilshad, Dr. Sohail Ahmed Jan and Dr. Arshia Amin Butt. A special thanks go to all my friends and seniors for their support, coordination and help from time to time. In the end, I am gratefully wanted to acknowledge my parents for their countless contributions, all their support without which I was unable to do anything. I am out of words to explain my gratitude towards my parents, siblings for their love, care, encouragement and prayers that enlightened my whole life.

(Kashaf Sajjad)

Abstract

Streptococcus agalactiae is a gram positive bacteria anaerobe, non-spore forming coccus belongs to streptococcus family (GBS). It is opportunistic and rounded bacteria present in the form of chains or colonies in the host's body. It mainly effects gastrointestinal and urogenital tract of host. It effects people of all age groups especially neonates. In this study, pan-genomic analysis and reverse vaccinology approach was used for designing vaccine against 127 strains of *Streptococcus agalactiae*. Out of 580 proteins, 335 proteins were non-host homologous proteins. From 335 proteins, 2 proteins were selected for epitope-based study on the basis of 2 thresholds i-e identity more than 25 and e-value = 0.003. 2 vaccines were designed manually on the basis of B and T-cell epitopes of these proteins by using linkers and adjuvant and docked against 2 TLRs i-e TLR 2 and TLR4 that play important role in human immune system. In epitope-based study, vaccine 2 with adjuvant shows higher interactions with TLRs and can be further validated by in vitro analysis and clinical trials.

Keywords: Reverse vaccinology, in-vitro analysis, non-homologous proteins

Contents

List of Figures

List of Tables

Abbreviations

Blast	Basic Local Alignment Search Tool
DEG	Database of Essential Genes
DIANNA	Diamine acid Neural Network Application
Edgar	Efficient Database framework for comparative Genome Analyses using BLAST score Ratios
ERRAT	Overall Quality Factor
GBS	Gram Positive Streptococcus
GRAVY	Grand Average of Hydrophaticity
HLA	Human Leukocyte Antigen
IEDB	The Immune Epitope Database
I-TASSER	Iterative Threading ASSEmbly Refinement
MHC	Major Histocompatibility Complex
MW	Molecular Weight
NA	Non-Allergen
NCBI	National Center for Biotechnology Information
NT	Non-Toxin
PDBePISA	Protein, Interfaces, Structures and Assemblies
PSIPRED	PSI-blast Based Secondary Structure Prediction
S. agalactiae	*Streptococcus agalactiae*
TLRs	Toll-like receptors
TMHMM	Transmembrane Helices; Hidden Markov Model
Uniprot	Universal Protein Resource

Chapter 1

Introduction

Streptococci is the most diversified group of pathogenic and opportunistic microorganism causing diseases from throat infection to life-threatening infections of the blood or other parts of body. People from all age groups from Neonates, toddlers to adults can be affected by it [1]. The most effective and successful treatment against infection diseases is antibiotics, but development of new treatment strategies and preventive measures including vaccine development is always focused. On the basis of hemolytic properties of Streptococci, Streptococci family has been divided into several groups such as Group A streptococcus, Group B streptococcus, Group C streptococcus and Group G streptococcus. In blood cells, Alpha-hemolytic members may cause oxidization of iron with hemoglobin protein which gives green color to the blood on agar. While complete breakdown of blood cells may cause by Beta-hemolytic members of Streptococci. *Streptococcus agalactiae* is a member of group B streptococci which is abbreviated as GBS which is rounded gram positive, hemolytic and facultative bacteria and forms chain when attack on host cells. It can be cultured in various growth media at 37 °C. It is an opportunistic pathogen of humans and effects immuno-compromised people such as pregnant females, adults, newborns and also can affects animals such as cows by causing a disease called bovine mastitis and Nile tilapia by causing meningitis [1]. It is present in gastrointestinal and genitourinary tract of the host in the forms of colonies. It is an asymptomatic commensal bacterium in adults but during infection it causes symptoms including fever, difficulty in breathing, irritability,

chills and cough [30]. The mortality rate is 15-20 % if it goes untreated. It is commonly treated through conventional antibiotics such as penicillin, clindamycin and erythromycin as first treatment strategies but for preventive measure there is no available vaccine.

The *Streptococcus agalactiae* enters and leaves the bodies of host naturally. It does not transmit from person to person through sexual activity or not even through water or food. The mode of transmission is unknown except in newborn babies where they get infected from mother positive with *Streptococcus agalactiae* especially during a vaginal delivery. It is also reported 10-30% of women are positive in their genital tracts as the pathogen is normally a resident of the gastrointestinal tract, increases the chances of further transmission of infection [2].

Streptococcus agalactiae colonized throat and rectum of newborns after 48 hours and remain colonized throughout childhood, but in adults the colonization generally shifts to genitourinary tracts. The transmission from infected animal to human i.e., zoonosis is rare. In general, it is estimate that 10-40% humans are colonized with *Streptococcus agalactiae* and remain asymptomatic.

Therefore, *Streptococcus agalactiae* is considered as vital neonatal opportunistic pathogen transmitted from mother. It colonizes pregnant females by forming chains or biofilms and causes infection of placenta, amniotic fluid, urinary tract and bloodstream which may result in prematurity or death of fetus before birth. It is estimated that only 1 out of 4 pregnant females are positive with *Streptococcus agalactiae* and transfer infective agent to the baby during delivery [5]. Normally, pregnant females may have a test for *Streptococcus agalactiae* bacteria when they are 36 through 37 weeks pregnant [3].

During delivery, antibiotics have been given to pregnant females via vein (IV) for preventing most early-onset infection caused by *Streptococcus agalactiae* in newborns [4]. Neonates may show symptoms of infection caused by *Streptococcus agalactiae* into two phases called EOS which is early-onset disease and second is late-onset disease. In early-onset disease, neonates show symptoms during first week of their birth while in late-onset disease, neonates may show symptoms from 1 week to 3 months after their birth. The common symptoms of *Streptococcus*

agalactiae in newborn babies are fever, difficulty in breathing and feeding, color of skin and lips become blue. It is estimated that up to 63% fetuses may have *Streptococcus agalactiae* infection if mothers are also infected with *Streptococcus agalactiae* and about 2% of the exposed newborns may develop an invasive disease, and the mortality rate is approximately 2% to 3% in child, and up to 20% in premature babies. *Streptococcus agalactiae* also causes sepsis (infection in blood), meningitis (infection of the tissues of brain and spinal cord). These babies are usually treated with the help of common antibiotics.

Streptococcus agalactiae is mostly harmless in adults, but it causes infection in those patients who are suffering from serious disease or have weak immune system. The patients may suffer from Infection of skin, blood, lungs, urinary tract, heart (endocarditis), brain, spinal cord and other major parts of body.

Streptococcus agalactiae also causes mastitis which is the infection of udder in cows and cattle which greatly effects the economy of dairy industry. In cows, the production of milk may decrease due to *Streptococcus agalactiae* which can cause a febrile disease which may be acute or chronic for them.

Streptococcus agalactiae also infects fishes, such as sepsis and hemorrhages (internal and external) [12]. Vaccination is used to cure the infection caused by *Streptococcus agalactiae* in aquaculture because it is most effective against pathogens. Scientists are trying to develop various kinds of vaccines against *Streptococcus agalactiae* and few of them have successfully developed vaccines. *Streptococcus agalactiae* is also effecting and causing infection in many animals such as dolphins, elephants, dogs, cats, camels and crocodile.

Various organizations are working to control the morbidity and mortality in newborns due to *Streptococcus agalactiae*, such as Centers for Disease Control and Prevention (CDC) are working to guide those females who has chances of delivering a *Streptococcus agalactiae* -colonized infant with intrapartum prophylaxis and as a result of it, up to 24% of women are given with antibiotics during labor and delivery. Mostly, *Streptococcus agalactiae* infection is treated with antibiotics such as penicillin or ampicillin. In case of penicillin allergic females, clindamycin and erythromycin are used instead of penicillin. Alternative treatment methods

should be used because antibiotics are major public health concern and alternative approach is vaccine which is more effective than antibiotics and cure the infection rapidly. The vaccine development is under process, but is not available yet [6].

The development of various vaccines against clinically vital diseases is one of the most important applications to medicine publicly health. While ancient or traditional vaccine preparation methods are targeted to reduce or inactivate the whole virus or partly purified virous macromolecule or protein. The intrinsic viral characteristics of such methods involve weak or null replication and matter hypervariability of pathogens in vitro have some disadvantages [7].

To reduce or overcome these kinds of problems, many approaches are developed and the epitope-based immunogen preparation is most effective and promising technique among them. This technique utilizes many peptides like peptides of beta cells, MHC advanced peptides of sophistication I and class II. Epitope based antigen preparation technique having many benefits such as peptides is simply created in vitro which will cut back the assembly costs which can change the immunogen production methods at a large-scale. It also overcomes the pathogen culturing problems as a result of it does not need in vitro infective agent growth for the amide expression happiness to microorganism proteins [8].

Some other benefits of this system include safety advantages concerning mutations and aspect effects of attenuated viruses. It may also use as immune agents to activate body substance and cell mediate immunologic response on an important domain of microorganism protein with well-defined synthesized peptides [9]. The central and most vital aim of immunogen preparation is to spot epitopes which might generate body substance and cell mediate responses. Due to the presence of each infectious agent and host genome sequence, it becomes easier to spot drug targets [10].

In this study, pan genomic approaches as well as detection of target site and immunogen based on amide epitope is intended to cure and target *Streptococcus agalactiae* bacteria can be used. *Streptococcus agalactiae* bacteria has 127 strains that contain 2626 genes together. Identification of potential targets for vaccine development are going to be done by subtractive genome techniques.

1.1 Aim and Objectives

The main aim of this research is to understand the genomic diversity of *Streptococcus agalactiae* and the identification of novel vaccine targets against this infective agent which will result in the exploration of vaccine targets. To attain this aim, research methodology is designed with following objectives:

1. To explore the pan-genome and essential genes of *Streptococcus agalactiae*.
2. To analyze the potential of prioritized virulent factor as immunogen or vaccine target.

Chapter 2

Literature Review

This study focuses on the review of the relevant literature.

2.1 *Streptococcus agalactiae*

Streptococcus agalactiae is an opportunistic and rounded pathogen causing various diseases in humans and animals by colonizing them in the form of chains. It is mostly found in gastrointestinal and urogenital tract of host [11].

2.1.1 Taxonomy

Streptococcus agalactiae is present in hot weather and also found in fresh water, marine and estuary fish species. *Streptococcus agalactiae* are the normally founded in the gastrointestinal tract and urogenital tract of the host and it has been isolated from about 35% of healthy adult women [26].

2.1.2 Phenotypic and Genotypic Characteristics of *Streptococcus agalactiae*

GC content of *Streptococcus agalactiae* is 35.70% and genome size is 2.2. The serotyping is commonly used phenotypic assessment for *Streptococcus agalactiae*.

TABLE 2.1: Taxonomy of *Streptococcus agalactiae* [36].

Level	**Taxonomy**
Kingdom	Bacteria
Sub-kingdom	Posibacteria
Phylum	Firmicutes
Class	Bacilli
Order	Lactobacillales
Family	Streptococcaceae
Genus	Streptococcus
Species	*Streptococcus agalactiae*

Other methods have been used for the examination of *Streptococcus agalactiae* proteins like polysaccharide capsule and also utilized for the evaluation of effect of proteins of host. These strategies help us to understand the infection caused by *Streptococcus agalactiae*, but they are not beneficial because they are not providing relevant complete information related to genome. For this reason, mostly scientists now use genotypic techniques in association with phenotypic methods [14].

The polysaccharide of cell wall is a virulence factor of *Streptococcus agalactiae* and it is divided into various serotypes on the basis of polysaccharide and these are; Ia, Ib, Ia/c, II, III, IV, V, VI, VII and VIII. All of them have ability to induce infections in host but the one type is present frequently in host which is serotype III. The serotype V causes infections in association with serotypes Ia, II, and III in hosts other than pregnant women which is approximately 29%. All serotypes are varying region to region and have invasive and colonizing isolates [13], [25].

2.2 Pathogenesis

It is a process in which disease develops in host. It includes microbial infection, inflammation, malignancy and tissue breakdown.

2.2.1 Pathogenesis in Pregnant Females and Neonates

Streptococcus agalactiae gets attach to a variety of host cells such as human cells including blood–brain barrier epithelium, placental membranes, respiratory tract,

and vaginal epithelium. Lipoteichoic acid and surface proteins which are components of cell wall of *Streptococcus agalactiae* mediated this adhesion action to epithelial cells of host.

Fibronectin, laminin and fibrinogen are components of extracellular matrix which are used by *Streptococcus agalactiae* for mucosal colonization in host cell. ScpB and FbsA which are adhesive and surface proteins, helps in mucosal colonization. *Streptococcus agalactiae* has ability to avoid immunological clearance with the help of polysaccharide capsules and also has ability to penetrate the host cell with help of various strategies. It can also cause infection in fetus by penetrating the placental membrane which may damage the placental layer and premature birth of a child. [18].

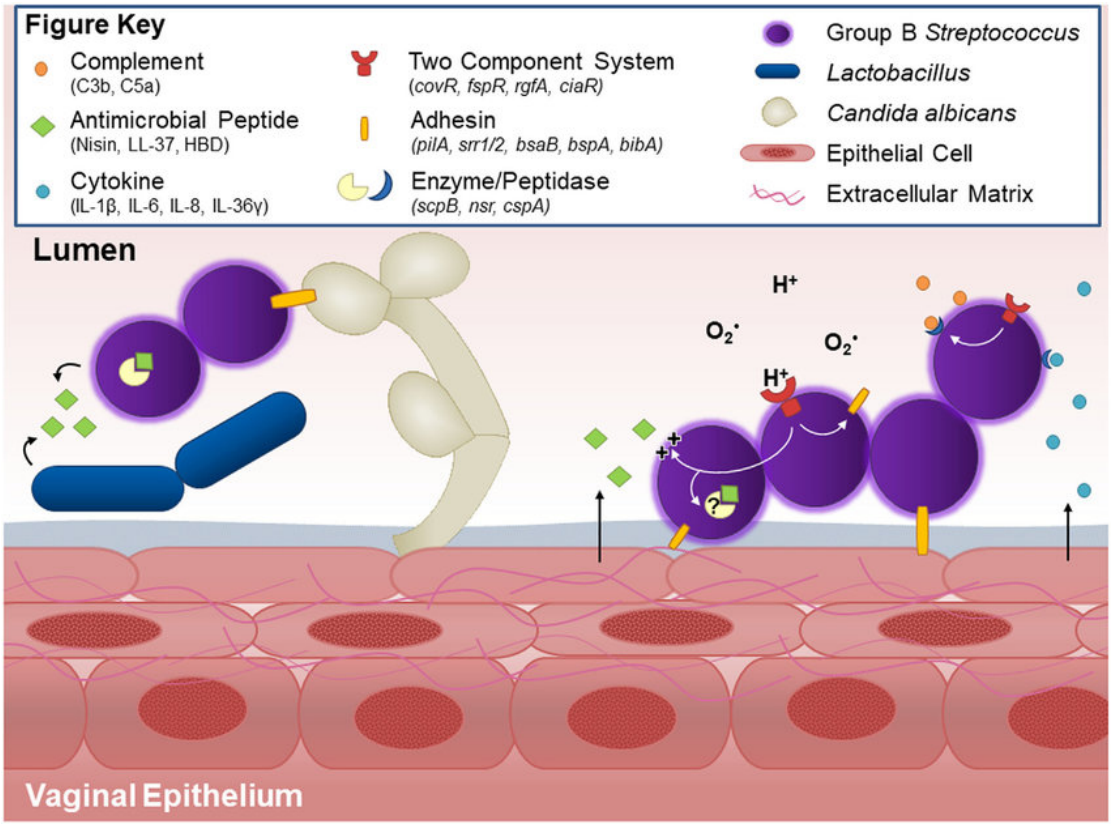

FIGURE 2.1: Pathogenesis of *Streptococcus agalactiae* in females [38].

Sepsis and meningitis are common infections caused by *Streptococcus agalactiae* in neonates, newborns. It does not harm the carrier or patient and even no treatment is required during pregnancy and delivery. Then during delivery, such type of patients must be treated with medicines such as antibiotics to lessen chances of infections due to *Streptococcus agalactiae* in newborns [17]. If a woman who

conceives second time had a *Streptococcus agalactiae* infection is at the risk of premature birth after 37 weeks of pregnancy.

During pregnancy of 35 to 37 weeks, rectal and vaginal swabs will be taken from mother to check whether she is positive for bacteria or not. If they are diagnosed positive against a particular bacterium then drugs such as antibiotics will be given to mother. Later in pregnancy, females with bacteriuria caused by *Streptococcus agalactiae* should not be screened for pathogen colonization in rectal or vagina, but considered a permanent GBS-colonized and treated with intrapartum chemoprophylaxis at the time of delivery. But *Streptococcus agalactiae* is not treated in non-pregnant women. Sometimes, when a woman has rectal or vaginal secretion, having *Streptococcus agalactiae* then antibiotic treatment is given [16].

The transmission rate from mother to neonates is 50%. The pathogen enters the body of pregnant females vertically in uterus then transfers from genital tract of mother to neonates during delivery. Only 1 to 2% effected neonates show symptoms of disease and diagnosed with Streptococcus colonization. Pneumonia with bacteremia is common, whereas meningitis is less likely [15].

Streptococcus agalactiae colonizes the genitourinary or gastrointestinal tract of host by adhering with vaginal epithelial cells, buccal epithelial cells, pulmonary epithelial cells and edothelial cells with the help of proteinaceous components of cell wall of bacteria and surface proteins. The first interaction of pathogen with host is mediated by proteinaceous component of cell wall which is weak interaction while second interaction is mediated by surface proteins which is firm and strong adhesion with eukaryotic cells. When pathogen colonizes the body, the immune system of host is activated and remove the infection via phagocytosis [38].

For the identification of *Streptococcus agalactiae*, Real -time PCR technique is recommended if vaginal rectal swab has a *Streptococcus agalactiae* which is capable of decreasing the time and increasing sensitivity. This real time PCR technique can also be applicable for testing other specimens in certain samples such as amniotic fluid, neonatal screening swabs, blood, breast milk, urine, and serum. Universal 16S PCR is another technique which is used for the identification of *Streptococcus agalactiae* from blood, bone and joint infections. The targets genes for GBS-PCR

include the sip, cfb, scpB, and ptsI. The sip gene codes for the Sip immunogenic protein. The cfb gene codes for the Christie-Atkins-Munch-Petersen factor. The scpB gene codes for the C5a peptidase and the ptsI gene, which codes for phosphotransferase [17].

The neonates diagnosed with early-onset disease, are treated with antibiotics which is basic treatment for such type of infections. Once *Streptococcus agalactiae* is diagnosed in patient, then penicillin monotherapy is given to the patient which is antibiotic treatment.

Antibiotics are recommended for those infants whose age is 7 days and have 250,000 to 450,000 units/kg/day weight and also recommended for infants whose age is more than 7 days and have 450.000 to 500,000 units//kg/day weight. 10 to 14 days of antibiotic treatment is recommended for neonates to cope up with infection, whereas sophisticated infections are also treated for extended amount with antimicrobial medical care. Septic inflammatory disease or osteitis is treated for 3–4 weeks and a minimum of four weeks of medical care is usually recommended for carditis or ventriculitis [19].

2.2.2 Pathogenesis in Adults

Streptococcus agalactiae can also infect adults or non-pregnant individuals. It does not harm or cause any type of infection in healthy person but if individual is suffering from serious illness such diabetes, heart issues, cancer or has weak immune system then *Streptococcus agalactiae* actively causes the infection in such type of individuals and has high mortality rate.

Streptococcus agalactiae causes various infection in adults like skin infection, bacteremia, soft tissue infection, sepsis, meningitis and urinary tract infections. Normally, penicillin is used as a medicine against infection but ampicillin, gentamicin can also be used if someone is allergic to penicillin [20]. *Streptococcus agalactiae* also causes skin and soft tissue infection which includes infection of foot (decubitus ulcer), cellulitis and abscesses. The patients who are suffering from skin and soft tissue infection caused by *Streptococcus agalactiae* could have a Diabetes mellitus

[21]. One of the intense and critical infection is rumored in adults, youngsters and neonates that is named osteitis. *Eubacterium agalactiae* might get direct entry through skin tissues or might enter through injury or rapture skin like ulceration and causes infection in these areas.

Eubacterium agalactiae additionally causes a unwellness of joint called septic inflammatory disease and infect most joints however specially infect knee, ankle, and shoulder. Septic inflammatory disease might prepare prosthetic joints with the bulk of infections occurring a minimum of three months when corrective placement; some eubacterium agalactiae prosthetic joint infections need corrective removal additionally to antimicrobial medical aid for successful treatment [21].

Penicillin G is that the most suggested and first-line treatment for true bacteria agalactiae infection in adults. The period of treatment depends upon the condition of infection and clinical presentation. Normally, 10 days medical care is suggested for the treatment of the infections like skin and soft tissue infection, bacteriaemia, pneumonia and urinary tract infections. If a patient is suffering from severe infection then 10 days treatment is recommended while in case of osteitis, carditis and ventriculitis infections, 4 weeks treatment is recommended. In case of carditis infection, additional antibiotic drugs are recommended for the initial 2 weeks of medical care [19].

2.2.3 Bovine Mastitis in Cows

Streptococcus agalactiae could be extremely contagious obligate parasite of the bovine duct gland in cows. It usually causes inferior persistent variety of infection and doesn't have a high self-cure rate. It continues to be a serious reason behind sub-clinical redness in milcher and a supply of economic loss for the business and is vulnerable to treatment with a spread of antibiotics [23].

Streptococcus agalactiae has the flexibility to stick to the duct gland tissue of cows and therefore the specific micro-environment of the bovine bag is important for the expansion of the bacteria [22]. *Streptococcus agalactiae* might cause heat, pain and swelling of the bag furthermore as abnormal milk consisting of white to yellow clots

and flakes [29]. Antibiotic mixtures of cloxacillin, ampicillin, cephapirin, antibiotic drug, cephalexin, penethamate, antibiotic drug, penicillin, are used [24].

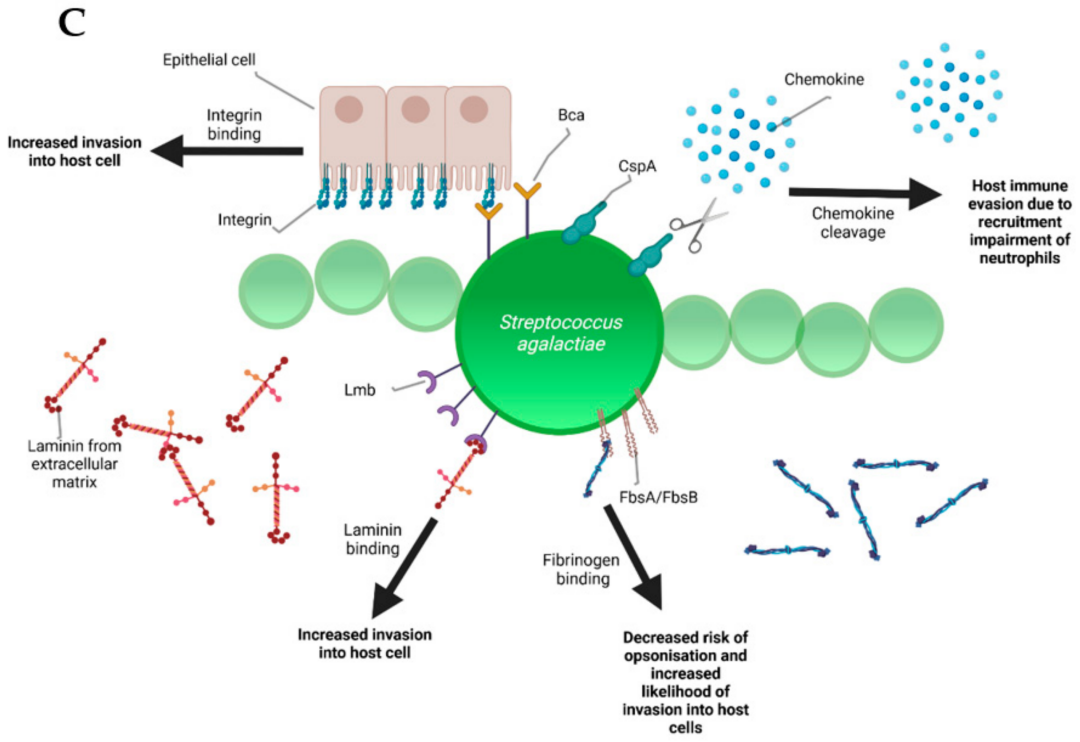

FIGURE 2.2: Interaction of GBS with cow cell [42].

2.3 Syndromes of *Streptococcus agalactiae*

There are various diseases caused by infection of *Streptococcus agalactia*. The most common diseases caused by *Streptococcus agalactiae* are given below;

2.3.1 Bacteremia/ Sepsis

Sepsis or bacteremia is a bloodstream infection and extensive hemodynamic changes which causes reduced cardiac functions, multi-organ failure and metabolic acidosis. It may also causes seeding of cardiac valves and endocarditis.

The bloodstream infection caused by *S. agalactiae* may spread bacterial infection to all other parts of body and may cause various serious diseases. It is commonly occuring in neonates and adults [38].

2.3.2 Meningitis

Meningitis is a common disorder of late-onset GBS infection in neonates but it occurs rare in adults. *Streptococcus agalactiae* causes this infection when it enters CSF of brain and causes an inflammatory cascade in subarachnoid space. It is treated through antibiotics [38].

2.3.3 Pneumonia

Pneumonia is a respiratory disorder caused by *Streptococcus agalactiae* and infects lower respiratory tract. It may reaches respiratory tract through blood or inspiration of amniotic fluid. It normally occurs in infants with early-onset occuring within seven days of life than with late-onset infection which occurs after 7 to 27 days of life. It may also causes neurological issues in patient and occurs in patient of cystic fibrosis [38].

2.3.4 Skin and Soft-Tissue Infections

Skin and soft-tissue infections caused by *Streptococcus agalactiae* are cellulitis,foot infections, abscesses and decubitus ulcers. It is commonly occuring in adults, neonates and children. *Streptococcus agalactiae* may infects the patient by attacking rupture skin and soft tissues or through blood. Septic arthritis is also observed in some patients due to *Streptococcus agalactiae* which may effects joints such as knee, shoulders and ankles [38].

2.4 Vaccine Development

A vaccine is a solution containing weak germs or pathogens which boosts the immune system against a particular disease. These weak pathogens have resemblance with disease causing agent and boost immune system to produce antibodies to kill the pathogen. The branch of science which deals with development and preparation of vaccine is known as vaccinology. After vaccination, when a pathogen

attacks the body of host for causing disease then immune system of host is ready to fight against that specific pathogen rapidly and aggressively. The protein coat of a pathogen is toxic in nature and when it enters the host body, it is recognized by immune system of the host due to its toxic protein coat which is virulent and has ability to cause disease in host cell and the immune system of host starts producing antibodies.

The host body also has ability to destroy infected cells of body by recognizing them before the multiplication of the pathogen. Sometimes vaccine does not work efficiently because of certain factors such as age or genetic disorders. Patients are vaccinated properly but due to these factors, their immune system is unable to produce antibodies. Mostly, vaccines work correctly if it is better than vaccine or strain of vaccine work efficiently against a disease [27].

The vaccine development has few steps. The first step is performed in laboratory to identify the foreign agent (antigen) by performing various tests. The antigens are toxic in nature and are composed of various toxic compounds. This step is called exploratory phase and it takes almost 2-3years. In next phase, the prepared vaccine is experimentally tested on various samples such as tissues, cells, and animals.

These experiments tell us the efficiency of vaccine, how it contributes to the functioning of immune system and it is safe for humans without harming them. It takes 1-2 years and it is called pre-clinical phase of vaccine development. In third phase, now vaccine is approved for utilization and it is safely applied to humans. This stage is called clinical stage which is further divided into three stages. In first stage of clinical phase, the efficacy, side effects and immune response will be tested by giving a dose to 20 – 80 patients.

This whole process will take 2 years. If vaccine shows positive response in stage I then it will be given to hundreds of patient's successfully in stage II and it takes 2-3 years. The method of administration of vaccine to patients could be oral, subcutaneous, intramuscular, intradermal, or intranasal. The stage III is just like a stage II but at third stage, vaccine will be given to thousands of patients in order to check efficacy and side effects which is the main objective of this stage

because it is not possible to check side effects among small group of patients and this procedure takes 5-10 years. In second last step, the vaccine will be approved by government on the basis of results of III trails performed in third step. Vaccine will only be approved if it is safe and secure for humans without harming them government will monitor its whole production process to ensure its effectiveness. In the last step, vaccine manufacture companies will start manufacturing vaccine at large scale and to test its reliability and efficiency and this stage is called pharmacovigilance [27] as summarized in figure 2.4.

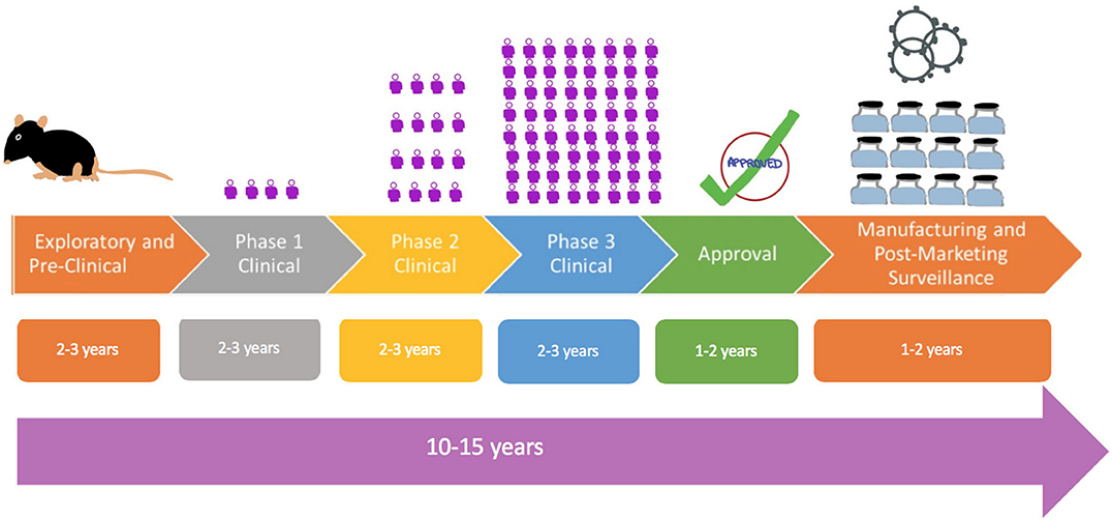

FIGURE 2.3: Vaccine Development Process [40].

2.4.1 Types of Vaccine

There are various types of vaccine which are described in table 2.2.

TABLE 2.2: Various types of vaccines [37].

Types of vaccine	**Function**
Attenuated	Attenuated are such types of vaccines which Boosted the immune responses in host for longer Period but these types are not safe for those whose Immune system is already weak or suering from seious disease. Hence these vaccines are called weak vaccines. These vaccines are mutant to virulent form

	and cause disease but itoccurs rarely such as viral diseases including measles and yellow fever
Inactivated	Some vaccines are present in inactivated form because they may have old dead virulent which is killed by heat or rays. Examples include IPV (polio vaccine).
Toxoid	These vaccines evolved from inactivated toxic substances that cause disease in host instead of killing pathogen. It is also given to animals. Examples include tetanus and diphtheria.
Subunit	Subunit vaccine is somehow different from other vaccines as it utilizes a segment of inactivated pathogens to boost the immune system instead of whole inactivated agents. For example; subunit vaccine against hepatitis B.
Conjugate	Conjugate vaccine can be manufactured on the bassis of toxic polysaccharide outer wall of pathogen which is poorly immunogenic, the defensive system. Only proteinic antigens recognize these toxic polysaccharides such as Haemophilus inuenza type B vaccine.
Heterotypic	The pathogens are being taken from other animals which may or may not cause any illness in organisms being treated and these are commonly called "Jennerian vaccines" because cowpox was used to treat a smallpox by developing a vaccine. Recently, BCG vaccine has been developed to immunize against tuberculosis by using Mycobacterium sp.
Viral vector	These vaccines use virus as a protective vector for injecting genes of pathogen in the host cell in order to generate specic antigens, for example surface proteins to provoke immune system to gene-

	rate a response.
Experimental	These vaccines provoke immune system by combbining antigen with dendritic cells in WBCs. Examples are vaccine against brain tumors and malignant melanoma.

2.5 Reverse Vaccinology Approach

The traditional approach for the development of the vaccine requires a live pathogen and it has been exploring by using various methods such as micro-immunological and biochemical techniques for identifying the parts which are essential for immunity. It is proven that it is beneficial in some cases, but it is time consuming as well as not successful to find a solution for those pathogens for which a vaccine is not available.

In order to solve this problem, there is an alternative method which is called reverse vaccinology in which genomic data of a pathogen is used to develop a vaccine in-silico by using various bioinformatic tools and this approach is less time consuming. This approach was used for vaccine development against serogroup B meningococcus. Now, this approach is applied for vaccine development against *Streptococcus agalactiae* [28].

The reverse vaccinology approach explores the proteins in pathogens by using various tools and design a vaccine with help of these proteins and these are composed of purified proteins. These vaccines are safe to use and has no side effect [31]. The main advantage of this vaccine is that they form correct folding of the proteins. About 90% of B cell epitopes are conformational and antibodies may bind with them. One or more than proteins can be used for vaccine development by reverse vaccinology because there is chance to develop a immunity against more than one strain of a pathogen. Reverse vaccinology approach is utilized to design a vaccine against infection caused by *Streptococcus agalactiae* with the help of various bioinformatic tools [32].

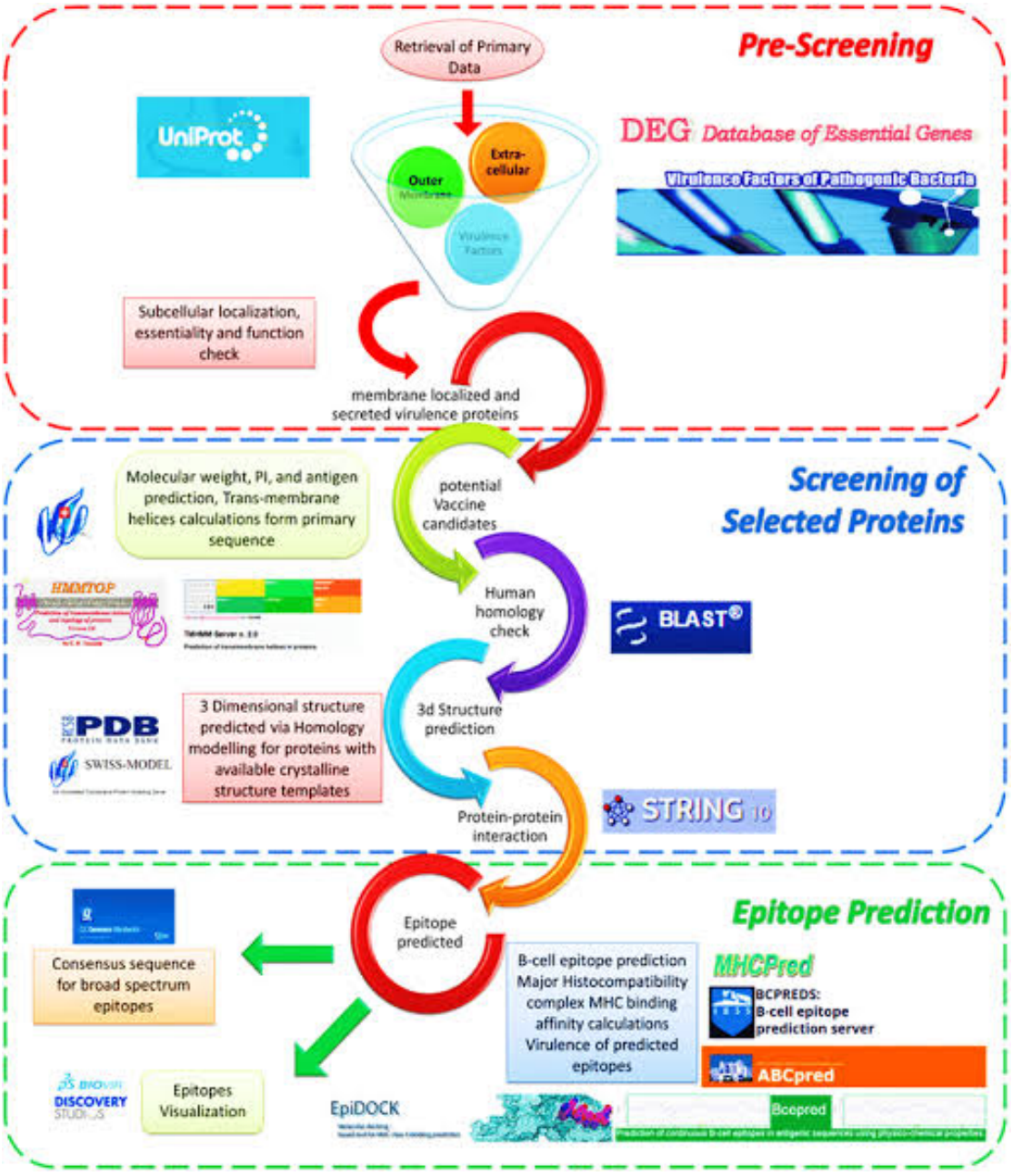

FIGURE 2.4: Reverse Vaccinology Screening Process [39].

2.6 Pan-Genome Analysis

Pan genome is the set of all genes present in the all strains of group having same ancestor in genetics and molecular biology. The pan-genome study is divided into three groups: (i) "core pan-genome" that contains genes found in all individuals, (ii) "shell pan-genome" that comprises genes found in two or more strains, and (ii) "cloud pan-genome" that only contains genes found in one strain. Cloud genome

is also called accessory genome which contains dispensable genes found in only a fraction of strains as well as strain-specific genes. The advancement in next generation sequencing techniques helps in the understanding of cellular localization, functional diversity at the meta-genomic level and microbial genetic repertoire [31]. In addition to this, whole-genome sequencing of bacterial pathogens helps in prioritizing the insert of researchers towards the pathogenicity by accurately measuring the genetic variations among the pathogenic groups [32]. SNPs is a single nucleotide polymorphisms, which is used to infer genetic variants among numerous genomes at the bench-top level and it is cost-effective and time-consuming. The whole-genome multi-locus sequence typing (MLST) technique can be used to accomplish this purpose [33].

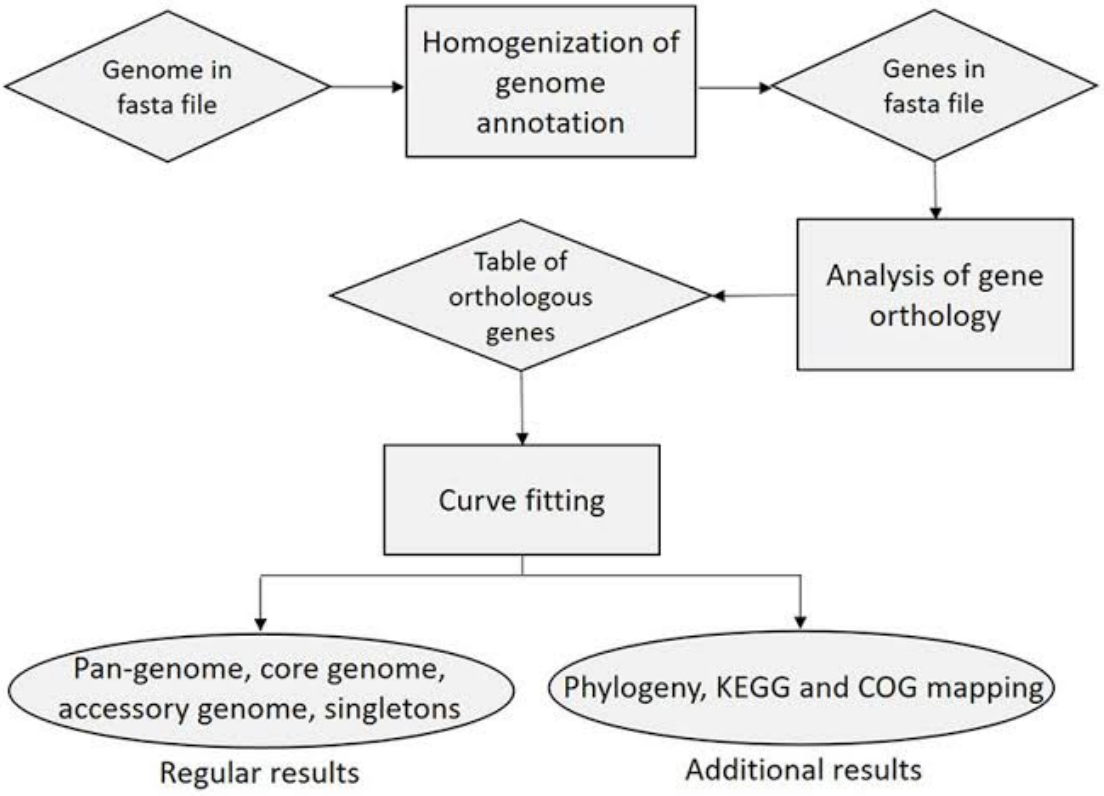

FIGURE 2.5: Main steps of Pan-genome Analysis [41].

To overcome the possible limitations of these reference-based methodologies, a comparative genomic method is applied. This method is based on sequence similarity search analysis, and it is causing a shift in worldwide attention in omics methods. These can be done by the whole-genome multi-locus sequence typing (MLST) approach. A comparative genomic approach is used to reduce and sort

out the limitations that are relevant to these approaches which are based on references. This approach depends upon the analysis of search of similar sequences and it shifts the global interest towards the omics strategies. The foundation of omics is laid by the availability of sequenced data at the public repositories and its free accesses and it also helps in the formation of the consecutive system biology principles [34].

The comparative microbial genomics strategy that is based on the sequence similarity helps in identifying the essential genetic content that is shared by all the pathogenic isolates with the help of statistical analysis. It also helps in finding the genes that encode virulence and novel functions as a variable genome [35].

2.7 B-cell Epitope Mapping

It is a potential method for finding microbes' key antigenic determinants, particularly those with discontinuous conformational properties. Epitope-based vaccinations have several advantages over traditional vaccinations, including being more precise, avoiding unwanted immune reactions, generating long-lasting protection, and being less expensive. The essential to antigen-antibody interactions is the humoral immune reaction to invading pathogen. A particular antigen (Ag) is recognized as the antigen determinant of B-cell epitopes by the particular antibody (Ab) in the distinct region. Surface available clusters of amino acids recognized as B cell epitopes by secreted antibodies. These epitopes of the B-cells may produce cell or humoral immune response [43]. Many surface antigens may be used in the epitopes after antibodies have been recognized. This method is not yet known about the mechanism [44]. By lacking the antigen reconfiguration in the Ag-Ab complex, the idea of classifying antigen in epitopes and non-epitopes is not capable of reliably reflecting the bioactivity [45]. Precise identification of the B-cell epitopes [45] is another immune diagnostic technique [46] which is the basis of the advancement of antibody therapeutics and vaccines which are based on peptides. It can be categorized as a nonlinear and continuous conformational structure, depending on the spatial structure. Amino acids are in closest interaction with

the discontinuous epitopes because of the three-dimensional conformation [46]. A minimal amino acid sequence is expected of the indigenous proteins for the proper folding of discontinuous epitopes. The range of the amino acid series is from 20 to 400. The bulk of the linear antigenic determinants found are thought to be the pieces of the conformational B-cell epitopes [47]. Analysis has shown that more than 70 percent of discontinuous epitopes consist of 1 to 5 linear parts. The length of these segments can be of amino acids varying from 1 to 6 [47]. The methods for determining the epitopes can be developed experimentally, hence can be separated as functional and structural studies approximately. The precise position of the laborious can be located precisely by the crystallography of X-rays which is a technique that takes time and is technically hard. This procedure therefore should not apply to all antigens [47]. The most popular approaches used for functional mapping of beta cells include proteolytic fragments screening extracted from antigen for the binding of antibody and measuring the reactivity of mutants with Ag-Ab. Site-driven or spontaneously mutating are these mutants [47].

Any of these techniques are inexpensive, fast and versatile in comparison to other techniques, and are used in the study of the epitope mapping [48]. An antigenic surface is homogeneously antigenic, according to which Rubinstein and his colleagues have established a null hypothesis. The epitopes were also described in the form of broad statistical analyses of Ag-Abco crystals deriving from protein databases to describe the physicochemical, structural and geometrical aspects. With this data, it can be inferred that the epitopes can be distinguished from the other surface of the antigen [47].

Another research was carried out by the Kringelum and colleagues who identify the smooth, extended, oval-shaped bundles which have an unknown secondary structure as the B-cell epitope [49].. The distinguishable characteristics of epitopes and non-epitopes are recognized using systematic laboratory studies and silicon analysis. The bulk of the epitopes are between 15 and 20 residues and are arranged into loops between 600 and 1000 to 1000 A2. The epitope's surface usability is the most common function. The epitopes series is complemented by Y, W, loaded polar amino acids. Relevant amino acid pairs are also available. The relationship between Ag-Ab and complex CDR loops involves epitope compression [47].

In recent years, the distinctions between residues between peptides and other residues are not taken into account [44]. Advances in epitope mapping technologies and bioinformatics are very important for immune technology creation and include uses of the numerical approach to disclose anticorps, B-cells, T-cells, and other allergen structures. The results will also be discussed [50] in the early prediction methods, the linear epitopes were identified by the propensity stage. Multiple techniques, like machine study methods like the Hidden Markov Model [51], supporting vector machines and the recurrent neural network [52], have been introduced to improve prediction efficiency. Despite these advances, there is only a small range of approaches used to forecast discontinuous epitopes. These approaches need to combine knowledge such as statistics for amino acids, sensitivity to surface areas and spatial data [53]. B-cell epitope discovery is very significant in advancement for testing of diagnostic therapeutics and vaccines [53]. Epitope imaging has been used in drug development [52]. Examinations are underway. Despite getting results in the mapping of B-cell epitope, this is also essential for observing antibodies against peptides because they cannot bind to native proteins because of the unstructured nature of peptides [53].

2.8 Virulence Factors of *Streptococcus agalactiae*

As all infections, *S. agalactiae* infection has also cope up with variety of various cell varieties like macrophages, animal tissue cells, and epithelium cells throughout the invasive illness method. To beat these defensive barriers and survive within the host, an organism should possess a spread of virulence factors. Such virulence factors not only permit invasion of the host tissue, leading to delicate to severe diseases but also shield the organism against the immune system of the host's body. *S. agalactiae* also has many such mechanisms or factors that shield and permit the organism to cause differing types of infections [54]. The capsular polysaccharide is the significant virulence issue of *S. agalactiae*. The capsular polysaccharide is formed of more than hundred continuation units of the monosaccharides sucrose,

glucose, N-acetylglucosamine, and N-acetylneuraminic acid (sialic acid). The primary perform of the *S. agalactiae* capsule is assumed to be the protection of the organism from phagocytosis by the host's system. The sialic acid part of the capsule inhibits the choice pathway of complement by preventing the deposition of active C3 complement on the surface of *S. agalactiae*. If C3 will deposit, the capsule promotes the conversion of C3b to iC3b on the microorganism surface, leading to the organism being immune to uptake and killing by neutrophils [54].

Lipoteichoic acids (LTA) area unit semipermeable membrane polymers containing alcohol phosphate or ribitol phosphate that's found in most gram-positive microorganism.Various functions are related to this compound, one amongst that is mediating adherence of Gram positive microorganism to organism cells. These have additionally been incontestable to bind to the cell membranes of erythrocytes and animal tissue cells.It is found that LTA binding to human craniate and embryonic animal tissue cells happens in a very ballroom dancing process; the initial step being hydrophobic interactions between host cells and *S. agalactiae*, followed by the interaction of the glycerolphosphate backbone with the embryonic organism cell surface.

As a result, lipoteichoic acid facilitates the binding of the microorganism to the cell surface of animal tissue cells in each adults and neonates. *S. agalactiae* shows a slim zone of beta hematolysis on sheep nutrient agar, and this is often used united of the primary composition options in distinguishing this organism within the clinical laboratory [54]. The β-hemolysin could be a pore-forming, non-immunogenic lysin that's active against a range of cell varieties. The beta-hemolysin made by *S. agalactiae* has been shown to push the induction of interleukin-8 (IL-8), a potent white corpuscle chemotactic agent. In addition to being a bunch signal to initiate innate immune responses to the organism, IL-8 mediate white corpuscle achievement can also contribute to harmful acute inflammatory processes seen in some cases of invasive B streptococci unwellness. The contused cells show membrane disruptions, cellular swelling, changes in organelles, and body substance and breastfeed dehydrogenase release [54]. Hyaluronate lyase could be a protein virulence of *S. agalactiae* that may enzymatically degrade hyaluronic acid, a predominant part of the extracellular matrix of animal animal tissue and nervous

system. The operation of hyaluronate lyase is to act as an enzyme, destroying the traditional animal tissue structure of the host and promoting microorganism dissemination. Additionally, each body fluid and placenta contain a high concentration of mucopolysaccharide that the accelerator might aid the organism in traversing this barrier to realize access to the craniate [54].

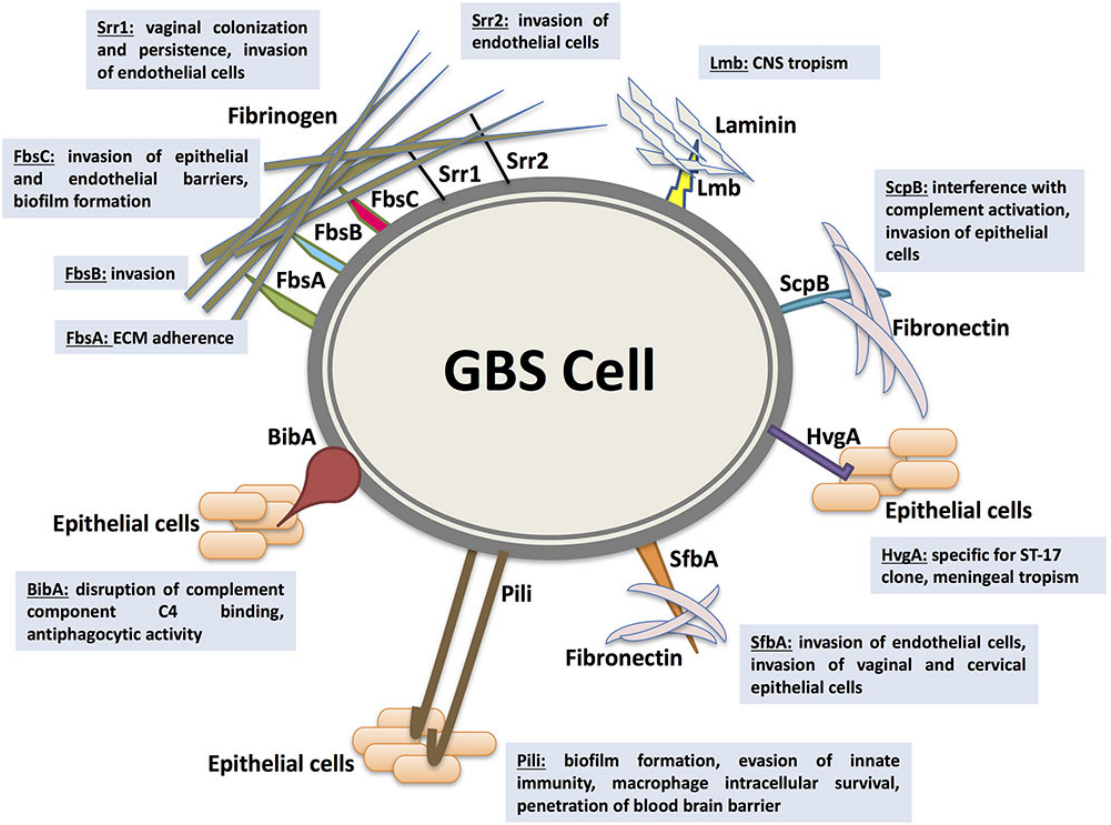

FIGURE 2.6: Virulence Factors of *S.agalactiae* [38].

Chapter 3

Materials and Methods

Multiple methodologies were used to find the potential vaccine targets. Vaccine developing techniques include, selection of genome, core genome identification, identification of non-host homologous proteins and target identification.

3.1 Selection of Genome

The complete genome of *Streptococcus agalactiae* was obtained from NCBI (http: //ncbi. nlm. nih. gov) [55]. 127 strains of *Streptococcus agalactiae* were obtained and used in this study.

3.2 Core Genome Selection

Core genome of *Streptococcus agalactiae* was identified by using EDGAR 3.0 (https://edgar3.computational.bio.uni-giessen.de/cgi-bin/edgar–login.cgi). EDGAR is designed to automatically perform genome comparisons in a high throughput approach. EDGAR provides novel analysis features and significantly simplifies the comparative analysis of related genomes. From 127 strains, one strain was selected as a reference strain on the basis of release date which was *Streptococcus–agalactiae–*2603VRAE009948. The selected reference strain was compared

with other 126 strains and genes that were common in all strains were used for further analysis [56].

3.3 Identification of Non-host Homologous Proteins

After retrieval of core genome, these genes/proteins were blast against human through BlastP to obtain non-host homologous protein (https:// blast. ncbi. nlm. nih. gov/ Blast.cgi). These non-homologous proteins were inserted into DEG Database (Database of essential genes) to obtain essential genes (http:// origin.tubic.org /deg/ public/ index. php/ blast/ bacteria) [56]. The two essential genes/ proteins were obtained on basis of thresholds i-e identity >25 and e-value 0.003.

3.4 Vaccine Target Identification

For the determination of potential therapeutics, multiple factors such as molecular weight, pathway analysis, protein localization etc were used. Multiple bioinformatic tools were applied on essential proteins. CELLO was used to determine the location of protein in the cell either surface or cytoplasmic (http:// cello. life. nctu. edu.tw/) [57].

Protparam tool was used to find out the molecular weight of essential proteins (https://web.expasy.org/protparam/) [58]. Uniport was used to find out the biological function and molecular function of essential genes (https://www.uniprot.org/) [59].

3.5 Epitope Based Vaccine Target Identification

Epitope-based immunogen/vaccine target identification involves a listing of steps that has knowledge/data retrieval and structural analysis, prediction of B-cell and

T-cell epitope, Imperative options identification of selected T cell epitopes, Epitope conservation analysis, multi-epitope immunogen/vaccine design and construction, Physio-chemical analysis of multi-epitope vaccines, multi-epitope vaccines 3D Structure prediction and Molecular Docking.

3.5.1 Data Retrieval and Structural Analysis

The physical properties of two surface proteins such as ABC transporter, permease protein and nrdI protein have been found by Protparam tool (https:// web. expasy. org/ protparam/) [58]. The physical properties include the half-life of the proteins, GRAVY (Grand Average of Hydrophathicity), molecular weight, amino acid atomic composition, and stability Index [60].

The secondary structure of both proteins has been analyzed by PSIPRED tool (http://bioinf.cs.ucl.ac.uk/psipred/) [61]. The transmembrane topology analysis of both the proteins was performed by TMHMM tool (http:// services. healthtech. dtu.dk /service .php? TMHMM-2.0) [61]. The presence of disulphite linkages in both proteins was identified by DIANNA (http:// clavius. bc. edu/ clotelab/ DiANNA/) [62]. Allergenicity of both the proteins was found by AllerTOP v2.0 (https://www.ddg-pharmfac .net/ AllerTOP/ [64]. and Antigenicity was found by Vaxijenv2.0 (http://www.ddg-pharmfac.net /vaxijen/ VaxiJen/ VaxiJen.html) [64].

3.5.2 Prediction of B-cell Epitope

The B cell epitopes of both proteins were predicted by BCPRED tool (http://ailab-projects1.ist.psu.edu:8080/bcpred/predict.html) [65]. Criteria for B-cell epitope prediction was set as 14 residue lengthy epitopes and 75% specificity. For checking the antigenicity of the resulted epitopes Vaxijen v2.0 was used and only antigenic epitopes used for further analysis [64]. B cell epitope recognition depends on multiple factors such as antigenicity, surface accessibility, flexibility, linear epitope prediction, hydrophobicity [66] other factors are; the Kolaskar and Tongaonkar antigenicity scales, Parker hydrophilicity prediction algorithms, Karplus and Schulz

flexibility prediction tool [67], Emini surface accessibility prediction tool, and Chou and Fasman beta-turn prediction algorithm[68].

These factors were determined by using IEDB analysis resources (http://tools. iedb. org/ bcell/). Discontinuous epitopes have been determined by using Disco-Tope server (http://tools. iedb. org/ discotope/) [69] because they have a greater number of explicit than linear epitopes. The parameters for this study were set at threshold -7.7, which is 75 percent specificity.

3.5.3 Prediction of T-cell Epitope

The MHC I and MHC II epitopes of both the proteins were identified by IEDB tool (http://tools.iedb.org/bcell/). The length of MHC I epitopes was 9 mer and the length of MHC II epitopes was 15 mer. Vaxigen v2.0 was used to find the antigenicity of these epitopes and only antigenic epitopes were used for further analysis [64].

3.5.4 Identification of Imperative Features of Selected T-cell Epitopes

The chosen MHC I and MHC II epitopes on the basis of antigenicity have various important features such as s enzyme digestion, toxicity, hydrophobicity, hydrophilicity, charge, PI and molecular weight. The PI, hydrophobicity, hydrophilicity, charge and toxicity was determined by ToxinPred (http:// crdd. osdd. net/ raghava/ toxinpred/). ToxinPred is an in silico method, which is developed to predict and design toxic/non-toxic peptides. The main dataset used in this method consists of 1805 toxic peptides (<=35 residues) [70]. The molecular weight was found by Protparam. ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. Peptide cutter tool was used to find the non-digesting enzymes (https:// web. expasy. org/ peptide-cutter/) and allergenicity was predicted by AllergenFP 1.0 (https:// ddg-pharmfac. net/ AllergenFP/) [71]. Only non-toxic epitopes were used for further analysis.

3.6 Epitope Conservation Analysis

Multiple sequence alignment was performed by choosing distinctive sequence of multiple proteins of same species for both proteins. Conservation of all B cell, MHC class I and MHC class II epitopes were checked against all chosen proteins through IEDB conservation analysis tool (http://tools.iedb.org/bcell/) [71].

3.7 Multi-epitope Vaccine Design and Construction

Multiple epitope vaccine was designed by arranging the epitopes with the help of linkers. MHC Class-I epitopes were joined through GGGS flexible linker while MHC Class-II epitopes were joined through flexible GPGPG linker [72].

The second multi-epitope vaccine was joined by joining the B-defensin, an adjuvant, was added at the N terminus of multi-epitope with EAAAK linker [73].

3.7.1 Physio-chemical Analysis of Multi-epitope Vaccines

Physio-chemical properties of multiple-epitope vaccines were determined by using various tools. Antigenicity was found by Vaxigen V2.0. Allergenicity was predicted through online tool AllergenFP 1.0. Protein-Sol which is an online tool was used to check the solubility of protein (https://proteinsol.manchester.ac.uk/cgibin/ solubility/ sequenceprediction.php).

Protparam tool was used to find out the molecular weight, GRAVY, instability index, aliphatic index and PI of multiple epitope vaccines. Immune simulation for both the proteins were performed through C-IMMSIM tool (https://kraken.iac.rm.cnr.it/C-IMMSIM/). The C-ImmSim online server allows the user to define the antigen to be injected as a list of UniProt accession numbers, or PDB primary identifiers, or, as a multi-protein FASTA text. The haplotype is defined by drop-down menus. Other parameters are the simulation time and the volume. [74].

Population coverage in Pakistan for both MHC class I and MHC class II was predicted through IEDB tool (http://tools.iedb.org/ population/) [74].

3.8 Prediction and Validation of Multi-epitope Vaccine's 3D Structure

3D structures of both vaccines were predicted through I-TASSER (https:// zhang-group. org/ I-TASSER/. I- TASSER gave multiple models but only one top model for vaccine was selected based on high C score in PDB format. This model was refined by using GalaxyRefine tool (https:// galaxy. seoklab. org/ cgi-bin/ submit.cgi? type=REFINE) [74]. Ramachandran analysis of refined model of vaccine was performed by using Zlab tool (https:// zlab. umassmed. edu/ bu/ rama/ index.pl) [74]. The 3D structures of both vaccines were verified by ERRAT (https://saves.mbi.ucla.edu/) [74].

3.9 Molecular Docking

Toll-like receptors (TLRs) are proteins that play significant role in immune system. The predicted 3D structure was docked against TLR 2 (2Z7X) and TLR 4 (3FXI). 3-D structures of all the proteins were downloaded from the PDB RCSB (https://www.rcsb.org/) [75]. Docking was done through the Cluspro 2.0 (https://cluspro.bu.edu/publications.php) which is an online tool. PDB file for Vaccine is given as ligand for input while protein was used as a receptor and results are downloaded in PDB format. The best docked structures were selected [74]. The PDBePISA is an interactive tool which is used for interaction analysis of docked protein structure (https://www.ebi.ac.uk/pdbe/pisa/pistart.html).

3.10 Overview of Methodology

The overview of methodology acquired is represented in figure 3.1

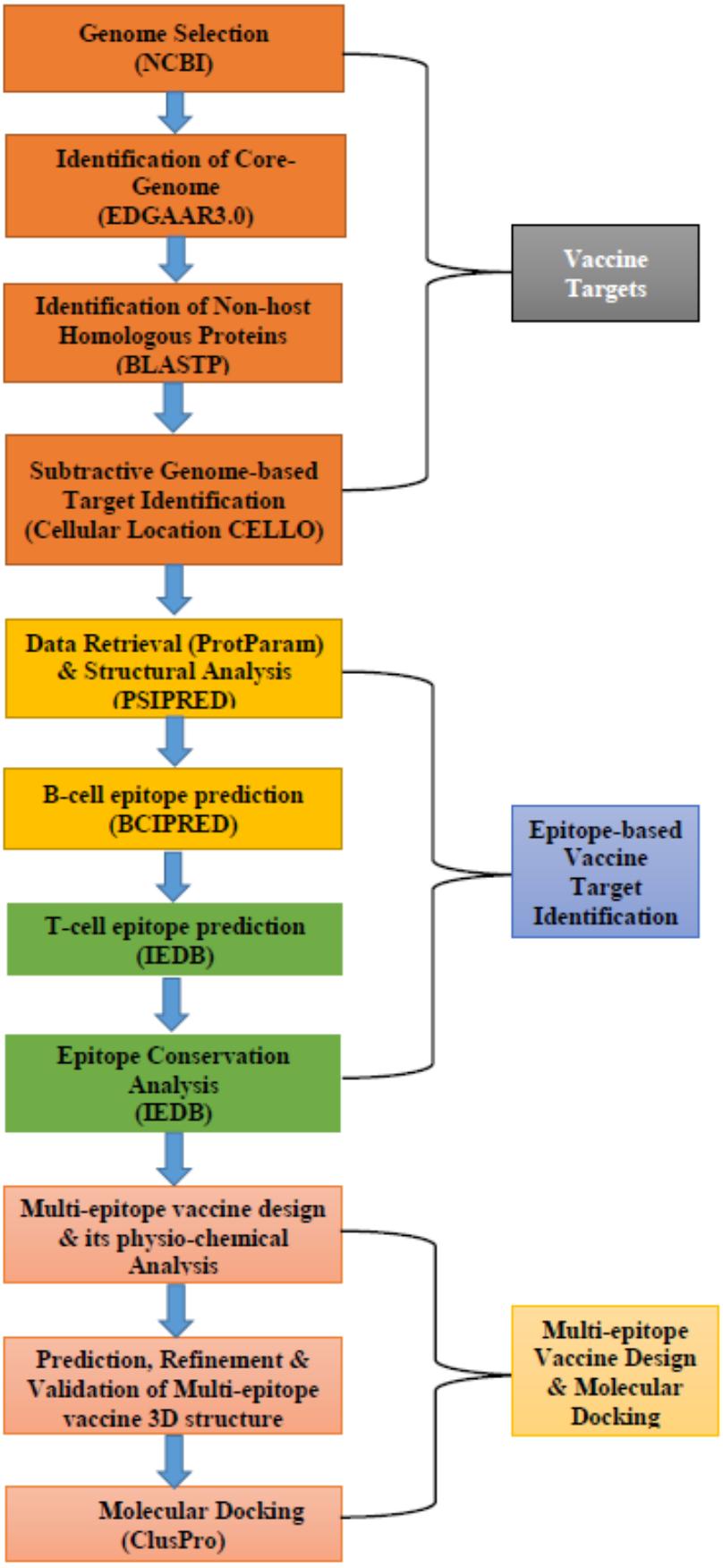

FIGURE 3.1: Methodology used for Epitope-based Study

Chapter 4

Results and Discussion

This chapter will explain the results that were obtained by following our methodological steps.

4.1 Epitope Based Vaccination Identification

4.1.1 Vaccine Targeting Analysis

From 127 strains of *Streptococcus agalactiae*, only one was selected as a reference strain named as Streptococcus–agalactiae–2603V–R–AE009948 and compared with other 126 strains to obtain common genes present in all strains. 580 core proteins were obtained which were common in all strains of *Streptococcus agalactiae*. Out of 580 core proteins, only 335 core proteins were non-host homologous proteins which were not present in humans and were only present in *Streptococcus agalactiae*.

From 335 non-host homologous proteins, 2 essential proteins were selected on the basis of cellular localization e.g extracellular protein and by applying 2 thresholds that are identity greater than 25 and e-value = 0.003. These 2 proteins named as ABC transporter permease protein, nrdI protein and these proteins are essential proteins and surface proteins. Substractive genome-based analysis was performed to find the pathways in which these proteins are involved, their molecular and

biological functions. Table 4.1 shows the molecular and biological characteristics of essential proteins.

TABLE 4.1: The molecular and biological characteristics of essential proteins.

Protein	**Uniprot ID**	**Gene name**	**Molecular function**	**Biological function**	**Cellular localization**
ABC transporter permease protein	A0A0 76Z8F2	yecS_3	Transmembrane Transporter activity	Amino acid transporter	membranous
nrdI protein	Q8E 1D8	nrdI-1	FMN binding	cellular protein modification process	extracellular

4.1.1.1 Structural Analysis

The physio-chemical properties of both proteins were computed by Protparam tool. ABC transporter permease protein has 230 amino acids, 24766.48kDa molecular weight. The nature of protein is negative as it has 6.72 theoretical isoelectric protein (pI) value. pI value more than 7 shows positive nature of protein while less than 7 shows negative nature of protein. The nrdI protein has 151 amino acids and 16799.96kDa. The nature pf protein is negative as it has 5.53 theoretical pI value. Table 4.2 shows all physio-chemical properties of 2 proteins including atomic composition, instability index, aliphatic index, GRAVY and half-life. According to instability index, both proteins are stable.

Vaxigen v2.0 shows that these proteins are probably antigen. The vaxijen value more 0.4 shows that protein is antigenic while below than 0.4 shows that protein is non-antigenic. ABC transporter permease protein has 0.4482 antigenic score while nrdI protein has 0.4882 antigenic score. According to AllerTOP results, both proteins are non-allergen.

DIANNA 1.1 shows that there is no disulphite bonds in ABC transporter permease protein while nrdI protein has 3 disulphite bonds. Table 4.3 shows the bonds for nrdI protein along with scores and distance. PSIPRED tool was used for the analysis of secondary structure of both proteins. Figure 4.1 and 4.2 shows the conformations of secondary structures of both proteins. The transmembrane topology of both proteins was checked by TMHMM tool. The TMHMM results shows that ABC transporter permease protein has 1 to 35, 93 to 95 and 192 to 200 amino acids were exposed to outside of the membrane, while 55 to 73, 119 to 168 and 224 to 230 amino acids were present inside the surface and 32 to 54, 74 to 92, 96 to 118, 169 to 191 and 201 to 223 amino acids were present in transmembrane helix. TMHMM results for nrdI protein shows that 1 to 151 amino acids were present outside the surface. Figure 4.3 and 4.4 shows the transmembrane topology of both proteins.

TABLE 4.2: Physio-chemical Properties of Both Proteins Computed Via Protparam

Parameters	**ABC transporter permease protein**	**nrdI protein**
Molecular weight	24766.48kDa	16799.96kDa
pI	6.72	5.53
Instability index	32.57	32.08
Amino acids	230	151
GRAVY	0.942	-0.215
Aliphatic index	136.87	88.41
Atomic composition	C 1165	C 746
	H 1847	H 1163
	N 281	N 203
	O 301	O 229
	S 5	S 5
Average half life	30 hours (mammalian reticulocytes, in vitro) Greater than 20 hours (yeast, in vivo).	30 hours (mammalian reticulocytes, in vitro) Greater than 20 hours (yeast, in vivo).

	Greater than 10 hours (E. coli, in vivo).	Greater than 10 hours (E. coli, in vivo).
Negatively	(Asp + Glu)	(Asp + Glu)
charged residue	13	14
Positively	(Arg + Lys)	(Arg + Lys)
charged residue	13	11

TABLE 4.3: Disulphite bonds along with bonds, scores and distance

Cysteine sequence position	**Distance**	**Bonds**	**score**
32 - 92	60	STWHNCQVSTI-NNVKNCIGIVG	0.01037
32 - 108	76	STWHNCQVSTI-FNNQYCLTAKQ	0.01037
92 - 108	16	NNVKNCIGIVG-FNNQYCLTAKQ	0.01185

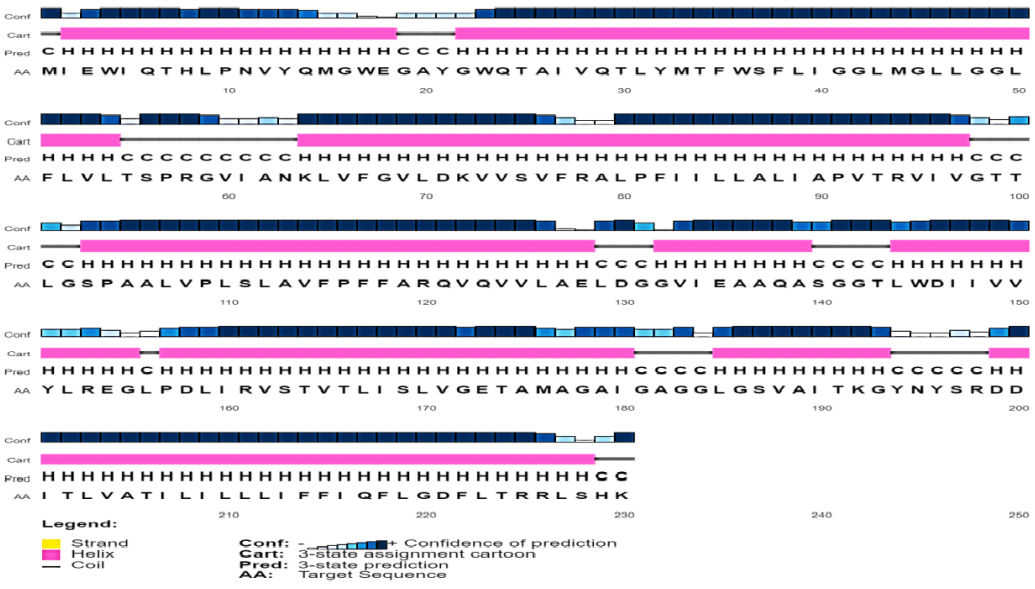

FIGURE 4.1: PSIPRED analysis of ABC transporter Permease Protein

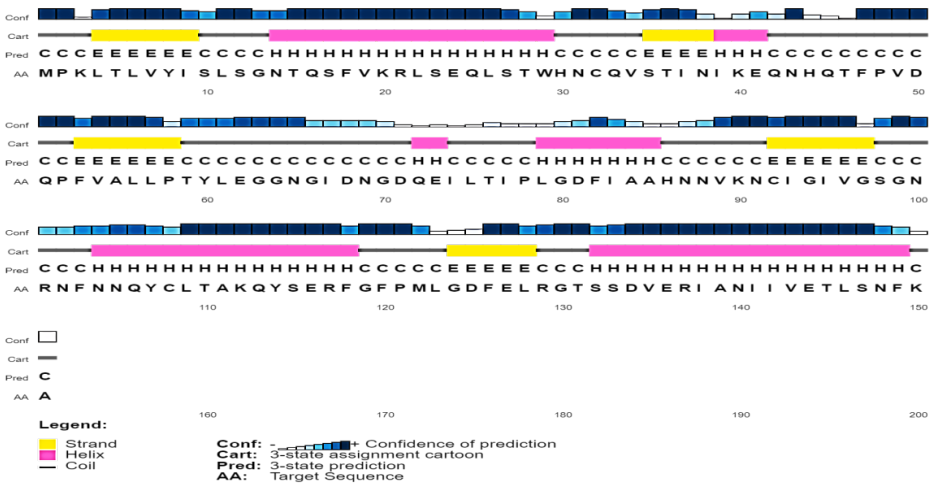

FIGURE 4.2: PSIPRED analysis of nrdI Protein

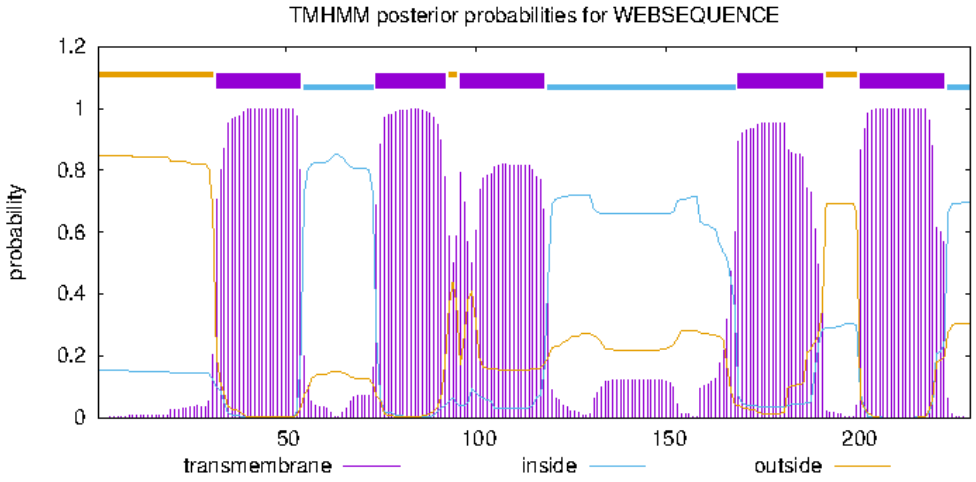

FIGURE 4.3: Transmembrane Topology analysis of ABC transporter Permease Protein

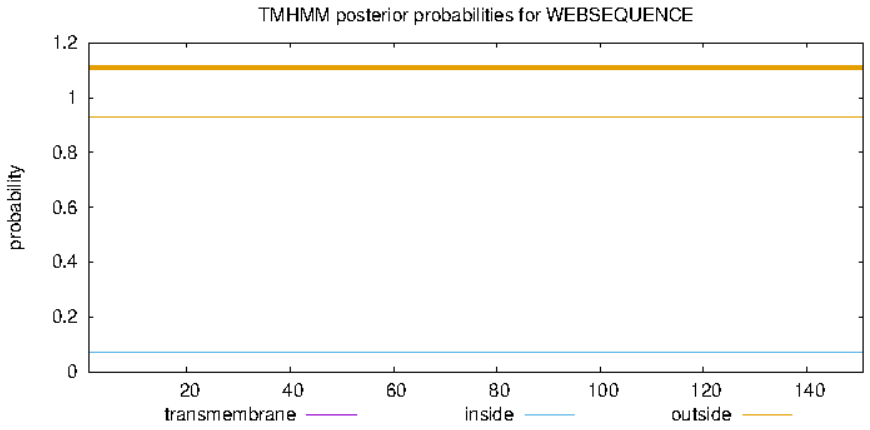

FIGURE 4.4: Transmembrane Topology analysis of nrdI Protein

4.2 B-Cell Epitope Prediction

There are different features of potential B-cell epitopes that directs them to recognize and then activate the defense responses. ABC transporter permease protein has only one B-cell epitope but it is non-antigenic. The nrdI protein has 5 B-cell epitopes but only 3 are antigenic (having high antigenicity score). These epitopes have 75% specificity with 14 residues (table 4.4). Bepipred linear epitope results for nrdI protein predicts 7 peptide epitopes at threshold 0.350 with average -0.042, minimum -0.004, maximum 1.742 values.

The first peptide GN starts from 13 position and ends at 14 position with length of 2 peptides. The second peptide Q starts from 16 position and also ends at 16 position because it has 1 peptide length. The third peptide EQNHQTFPVD starts from position 41 and ends at position 50 with 10 peptide length. The fourth peptide LEGGNGIDNGDQ starts from 61 position and ends at 72 position with 12 peptide length. The fifth peptide SGNRNF starts from position 68 and ends at position 103 with peptide length 6. The sixth peptide S starts at 115 position and also ends at same position 115 as it has 1 peptide length. The seventh peptide LRGTSSDVE starts at 128 position and ends at 136 position with 9 peptide length (Figure 4.5a).

B. Chou & Fasman Beta-Turn results predict the average value at 1.008, while the minimum at 0.696 and maximum at 1.390 at threshold 1.008 (Figure 4.5b). C. An Emini Surface Accessibility result predicts average value 1.000, while minimum 0.048 and maximum 6.195at threshold 1.000 with two peptides; first peptide IKEQNHQ starts from 39 position and ends at 45 position with 7 peptide length, while second peptide GNRNFN starts at 99 position and ends at position 104 with peptide length 6 and third peptide AKQYSE starts at position 111and ends at position 116 with peptide length 6 (Figure 4.5c).

Karplus & Schulz flexibility results predict average values as 1.003, minimum at 0.902 and maximum value lies at 1.117 at 1.003 threshold (Figure 4.5d). Kolaskar & Tongaonkar Antigenicity result predicts eight peptides; LTLVYISL starting from position 4 to 11 with 8 peptide length, FVKRLSEQ starting from position 18 to

25 with 8 peptide length, NCQVSTI starting from position 31 to 37 with 7 peptide length, TFPVDQPFVALLPTY starting from position 46 to 60 with 15 peptide length, ILTIPLGDFI starting from position 74 to 83 with 10 peptide length, KNCIGIV starting from position 90 to 96 with 7 peptide length, QYCLTAK starting from position 106 to 112 with 7 peptide length and NIIVETL starting from position 140 to 146 with 7 peptide length with average values 1.025 while the minimum at 0.849 and maximum at 1.174 at threshold 1.025 (Figure 4.5e).

E. Parker Hydrophilicity result predicts average value at 1.350, while the minimum at -3.571 and maximum at 5.686 at threshold 1.350 (Figure 4.5f). IEDB tool was used to predict the discontinuous B-cell peptides of both proteins with residue ID, residue name, contact number, propensity score and Discotope score. Table 4.5 and figure 4.6 is showing the discontinuous epitopes of ABC transporter permease protein and table 4.6 and figure 4.7 is showing the discontinuous epitopes of nrdI protein.

TABLE 4.4: B-cell epitopes present on the surface of nrdI protein predicted via BCPRED along with their starting position and antigenicity scores.

Sequence	**Position**	**Score**	**Antigenicity (Vaxigen 2.0)**
GSGNRNFNNQYCLT	97	0.98	1.0988
GGNGIDNGDQEILT	63	0.953	0.5655
GFPMLGDFELRGTS	119	0.91	0.9780

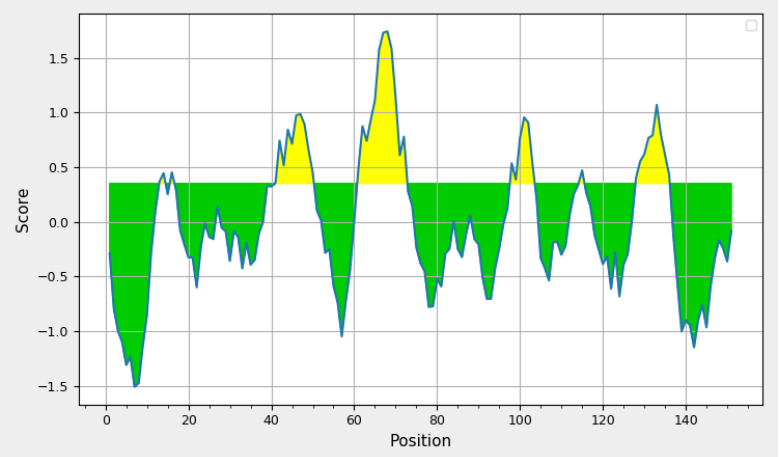

FIGURE 4.5A: A nrdI Protein Bepipred Linear Epitope prediction results

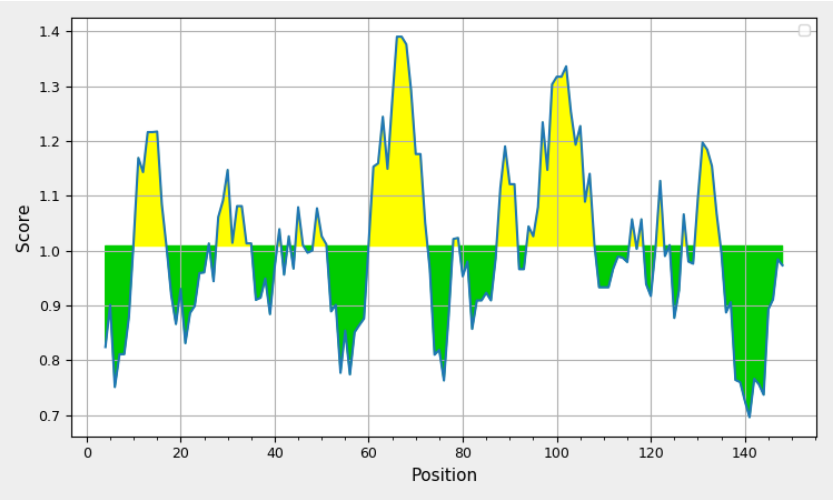

FIGURE 4.5B: A nrdI Protein beta turns analyses in structural polyprotein using Chou and Fasman beta-turn prediction

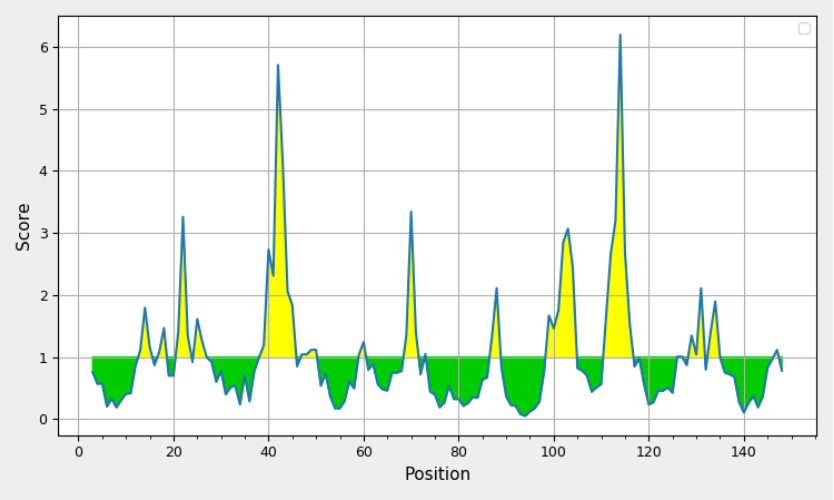

FIGURE 4.5C: A nrdI Protein surface accessibility analyses using Emini surface accessibility scale

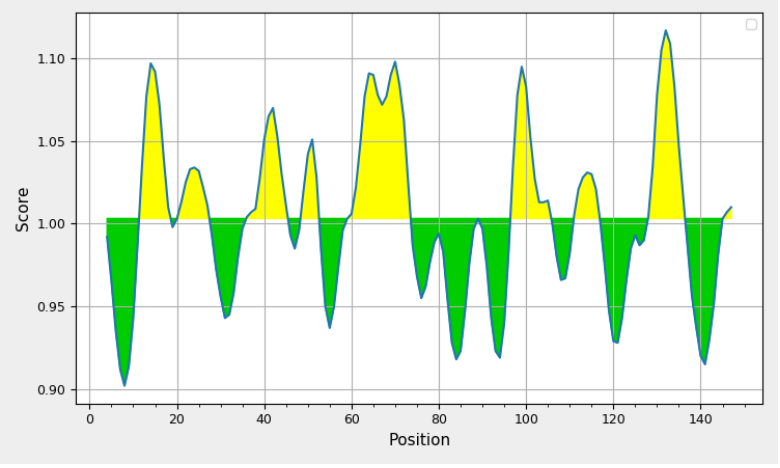

FIGURE 4.5D: A nrdI Protein flexibility analyses using Karplus and Schulz flexibility scale

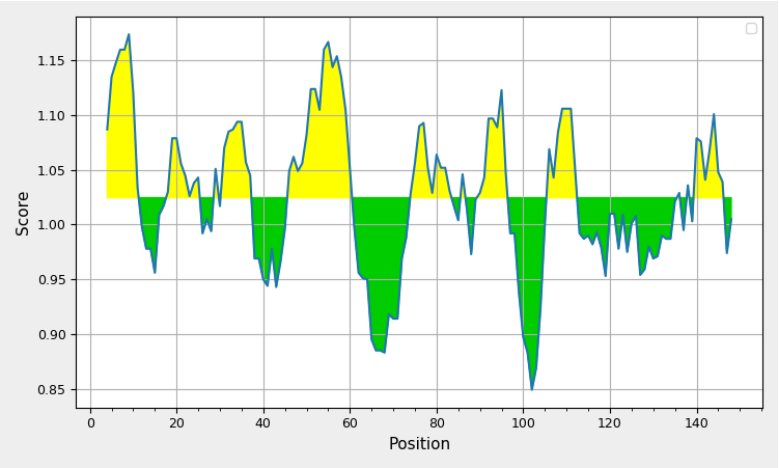

FIGURE 4.5E: A nrdI Protein prediction of antigenic determinants using Kolaskar and Tongaonkar antigenicity scale

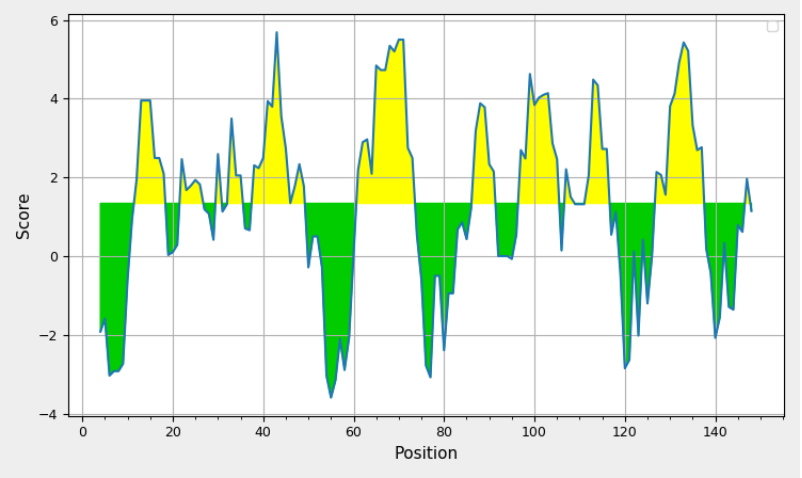

FIGURE 4.5F: A nrdI Protein hydrophilicity prediction using Parker hydrophilicity

TABLE 4.5: Discontinuous epitopes of ABC transporter permease protein through IEDB (DiscoTope)

Residue ID	**Residue name**	**Contact number**	**Propensity score**	**Discotope score**
194	TYR	13	-0.635	-7.135
195	ASN	10	-0.901	-4.099
196	TYR	14	-0.245	-7.245
197	SER	14	-0.044	-6.956
198	ARG	14	-0.267	-7.267
225	ARG	11	-1.716	-7.216

226	ARG	11	-1.716	-7.216
228	SER	12	-1.431	-7.431
229	HIS	8	-0.264	-4.264

TABLE 4.6: Discontinuous epitopes of nrdI protein through IEDB (DiscoTope)

Residue ID	**Residue name**	**Contact number**	**Propensity score**	**Discotope score**
20	LYS	12	-1.527	-7.527
28	THR	11	-2.05	-7.55
31	ASN	9	-1.79	-6.29
41	GLU	11	2.538	-2.962
42	GLN	12	2.956	-3.044
43	ASN	10	3.582	-1.418
44	HIS	11	2.522	-2.978
45	GLN	11	3.513	-1.987
46	THR	15	1.471	-6.029
47	PHE	12	0.13	-5.87
50	ASP	11	-1.397	-6.897
62	GLU	15	1.538	-5.962
63	GLY	17	2.788	-5.712
64	GLY	10	2.024	-2.976
65	ASN	10	2.641	-2.359
66	GLY	12	2.72	-3.28
67	ILE	11	2.803	-2.697
68	ASP	9	2.624	-1.876
69	ASN	12	2.592	-3.408
70	GLY	14	2.515	-4.485
72	GLN	16	1.397	-6.603
101	ARG	14	1.183	-5.817
102	ASN	9	1.811	-2.689
103	PHE	13	0.583	-5.917

104	ASN	10	1.202	-3.798
105	ASN	9	0.922	-3.578
133	SER	11	-1.533	-7.033
149	PHE	8	-3.064	-7.064

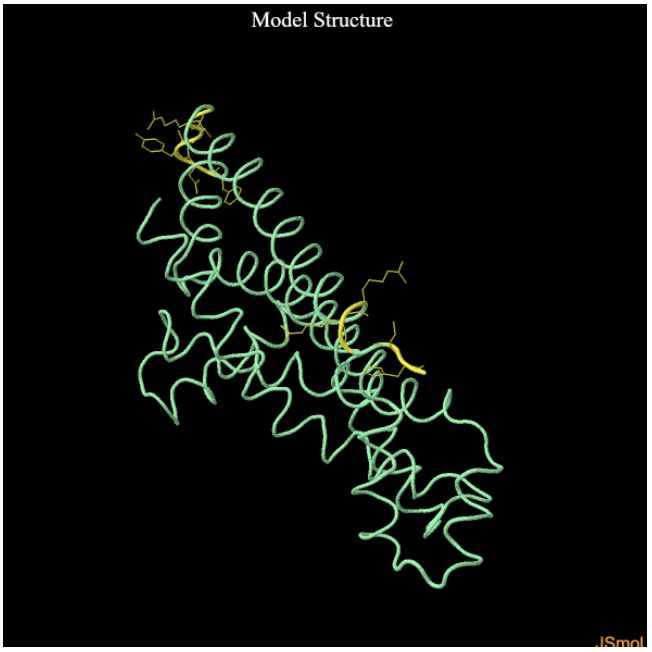

FIGURE 4.6: The site of discontinuous epitopes predicted through IEDB (DiscoTope) on 3D structure of ABC transporter permease protein

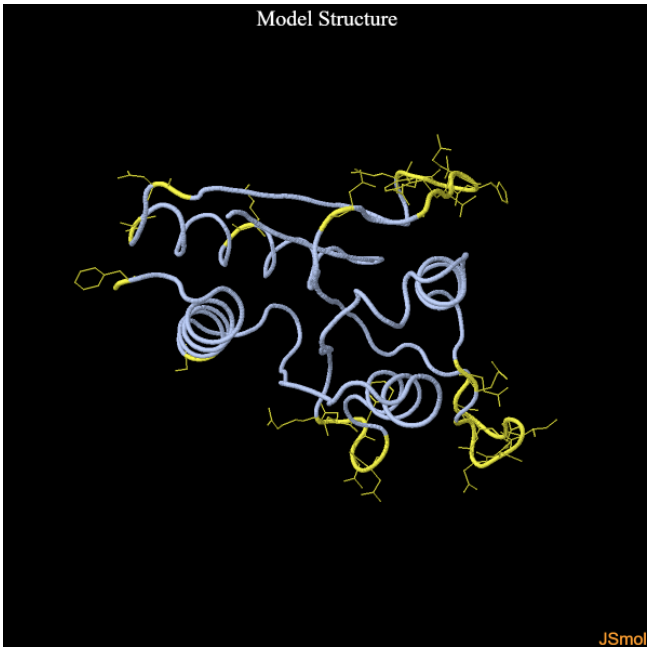

FIGURE 4.7: The site of discontinuous epitopes predicted through IEDB (DiscoTope) on 3D structure of nrdI protein

4.3 T-Cell Epitope Prediction

T-Cell epitopes were predicted by using an online tool IEDB. IEDB shows several MHC class-I with 9 mer peptide length and MHC class-II peptides with 15 mer peptide length for both proteins but on few peptides are selected for further analysis on the basis of high antigenicity.

For ABC transporter permease protein, 4 MHC class-I alleles and 4 MHC class-II alleles while for nrdI protein (Table 4.7 and 4.8), 3 MHC class-I alleles and 4 MHC class-II alleles are selected on the basis of high antigenicity (4.9 and 4.10). The antigenicity of all the peptides were checked through Vaxijen V2.0.

TABLE 4.7: MHC class-I allele binding peptides of ABC transporter permease protein predicted via Propred-1 with their antigenicity scores

Peptide sequence	**MHC class-I alleles**	**Vaxijen score**
LLLIFFIQF	HLA-A*02:06,HLA-A*32:01,	
	HLA-A*29:02,HLA-A*23:01,	
	HLA-A*02:01,HLA-A*24:02,	3.4763
	HLA-A*30:01,HLA-A*03:01,	
	HLA-A*01:01,HLA-A*68:01,	
	HLA-A*11:01,HLA-A*30:02,	
	HLA-A*68:02,HLA-A*26:01.	
	HLA-A*25:01	
ILLLIFFIQ	HLA-A*02:06,HLA-A*02:01,	
	HLA-A*29:02,HLA-A*30:02,	
	HLA-A*30:01,HLA-A*03:01,	2.8997
	HLA-A*32:01,HLA-A*11:01,	
	HLA-A*23:01,HLA-A*24:02,	
	HLA-A*68:02,HLA-A*68:01,	
	HLA-A*01:01,HLA-A*26:01,	
	HLA-A*25:01	
LLIFFIQFL	HLA-A*02:01,HLA-A*02:06,	
	HLA-A*68:02,HLA-A*30:01,	
	HLA-A*23:01,HLA-A*29:02,	2.7497

	HLA-A*03:01,HLA-A*26:01,	
	HLA-A*24:02,HLA-A*30:02,	
	HLA-A*32:01,HLA-A*68:01,	
	HLA-A*01:01,HLA-A*11:01,	
	HLA-A*25:01.	
LSLAVFPFF	HLA-A*23:01,HLA-A*02:06,	
	HLA-A*29:02,HLA-A*30:01,	
	HLA-A*24:02,HLA-A*30:02,	2.3066
	HLA-A*32:01,HLA-A*11:01,	
	HLA-A*02:01,HLA-A*68:01,	
	HLA-A*26:01,HLA-A*03:01,	
	HLA-A*01:01,HLA-A*68:02,	
	HLA-A*25:01	

Table 4.8: MHC class-II allele binding peptides of ABC transporter permease Protein predicted via IEDB with their antigenicity scores

Peptide sequence	**MHC class-II alleles**	**Vaxijen score (antigenicity)**
FRALPFIILL-ALIAP	HLA-DRB1*10:01,HLA-DRB1*16:02, HLA-DRB1*01:03,HLA-DRB1*08:01, HLA-DRB1*04:03,HLA-DRB1*04:02, HLA-DRB1*11:01,HLA-DRB1*04:04, HLA-DRB1*01:01,HLA-DRB1*13:01, HLA-DRB1*15:01,HLA-DRB1*07:01, HLA-DRB1*12:01,HLA-DRB1*04:01, HLA-DRB1*08:02,HLA-DRB1*09:01, HLA-DRB1*04:05,HLA-DRB1*13:02, HLA-DRB1*03:01	1.7007
VFRALPFIIL-LALIA	HLA-DRB1*10:01,HLA-DRB1*16:02, HLA-DRB1*07:01,HLA-DRB1*01:03, HLA-DRB1*15:01,HLA-DRB1*08:01, HLA-DRB1*11:01,HLA-DRB1*01:01,	1.4117

	HLA-DRB1*12:01,HLA-DRB1*09:01,	
	HLA-DRB1*04:01,HLA-DRB1*04:03,	
	HLA-DRB1*04:02,HLA-DRB1*13:01,	
	HLA-DRB1*04:04,HLA-DRB1*08:02,	
	HLA-DRB1*13:02,HLA-DRB1*04:05,	
	HLA-DRB1*03:01	
SVFRALPFII-	HLA-DRB1*10:01,HLA-DRB1*16:02,	
LLALI	HLA-DRB1*12:01,HLA-DRB1*07:01,	1.2826
	HLA-DRB1*01:01,HLA-DRB1*11:01,	
	HLA-DRB1*15:01,HLA-DRB1*08:01,	
	HLA-DRB1*01:03,HLA-DRB1*13:01,	
	HLA-DRB1*09:01,HLA-DRB1*04:01,	
	HLA-DRB1*13:02,HLA-DRB1*04:04,	
	HLA-DRB1*04:02,HLA-DRB1*08:02,	
	HLA-DRB1*04:03,HLA-DRB1*04:05,	
	HLA-DRB1*03:01	
AIVQTLYMTF-	HLA-DRB1*16:02,HLA-DRB1*07:01,	
WSFLI	HLA-DRB1*12:01,HLA-DRB1*04:04,	1.3019
	HLA-DRB1*15:01,HLA-DRB1*04:05,	
	HLA-DRB1*09:01,HLA-DRB1*04:02,	
	HLA-DRB1*13:01,HLA-DRB1*10:01,	
	HLA-DRB1*04:01,HLA-DRB1*01:03,	
	HLA-DRB1*08:01,HLA-DRB1*04:03,	
	HLA-DRB1*01:01,HLA-DRB1*13:02,	
	HLA-DRB1*11:01,HLA-DRB1*03:01,	
	HLA-DRB1*08:02	

Table 4.9: MHC class-I allele binding peptides of nrdI Protein predicted via IEDB with their antigenicity scores

Peptide sequence	**MHC class-I alleles**	**Vaxijen score**
GDFELRGTS	HLA-A*30:01,HLA-A*30:02,	
	HLA-A*02:06,HLA-A*26:01,	1.662

	HLA-A*32:01,HLA-A*29:02,	
	HLA-A*25:01,HLA-A*01:01,	
	HLA-A*68:02,HLA-A*02:01,	
	HLA-A*03:01,HLA-A*11:01,	
	HLA-A*68:01,HLA-A*23:01,	
	HLA-A*24:02	
PMLGDFELR	HLA-A*68:01,HLA-A*29:02,	
	HLA-A*11:01,HLA-A*02:06,	1.5425
	HLA-A*03:01,HLA-A*30:02,	
	HLA-A*23:01,HLA-A*02:01,	
	HLA-A*01:01,HLA-A*26:01,	
	HLA-A*24:02,HLA-A*30:01,	
	HLA-A*25:01,HLA-A*32:01,	
	HLA-A*68:02	
ISLSGNTQS	HLA-A*30:01,HLA-A*02:06,	
	HLA-A*30:02,HLA-A*11:01,	1.36
	HLA-A*29:02,HLA-A*68:01,	
	HLA-A*01:01,HLA-A*32:01,	
	HLA-A*26:01,HLA-A*03:01,	
	HLA-A*23:01,HLA-A*02:01,	
	HLA-A*68:02,HLA-A*25:01,	
	HLA-A*24:02	

Table 4.10: MHC class-II allele binding peptides of nrdI Protein predicted via IEDB with their antigenicity scores

Peptide sequence	**MHC class-II alleles**	**Vaxijen score (antigenicity)**
FPMLGDFELR	HLA-DRB1*03:01,HLA-DRB1*04:02,	
GTSSD	HLA-DRB1*08:01,HLA-DRB1*04:04,	1.5338
	HLA-DRB1*16:02,HLA-DRB1*04:01,	
	HLA-DRB1*10:01,HLA-DRB1*08:02,	
	HLA-DRB1*01:03,HLA-DRB1*04:05,	

	HLA-DRB1*01:01,HLA-DRB1*11:01,	
	HLA-DRB1*13:01,HLA-DRB1*15:01,	
	HLA-DRB1*09:01,HLA-DRB1*04:03,	
	HLA-DRB1*12:01,HLA-DRB1*07:01,	
	HLA-DRB1*13:02	
PFVALLPTYL	HLA-DRB1*10:01,HLA-DRB1*09:01,	
EGGNG	HLA-DRB1*04:05,HLA-DRB1*08:02,	1.4533
	HLA-DRB1*12:01,HLA-DRB1*15:01,	
	HLA-DRB1*01:01,HLA-DRB1*16:02,	
	HLA-DRB1*04:04,HLA-DRB1*07:01,	
	HLA-DRB1*11:01,HLA-DRB1*04:01,	
	HLA-DRB1*04:02,HLA-DRB1*13:02,	
	HLA-DRB1*08:01,HLA-DRB1*13:01,	
	HLA-DRB1*04:03,HLA-DRB1*01:03,	
	HLA-DRB1*03:01	
RFGFPMLGDF	HLA-DRB1*03:01,HLA-DRB1*04:02,	
ELRGT	HLA-DRB1*16:02,HLA-DRB1*04:04,	1.1697
	HLA-DRB1*04:05,HLA-DRB1*10:01,	
	HLA-DRB1*01:01,HLA-DRB1*04:01,	
	HLA-DRB1*08:01,HLA-DRB1*15:01,	
	HLA-DRB1*09:01,HLA-DRB1*11:01,	
	HLA-DRB1*08:02,HLA-DRB1*12:01,	
	HLA-DRB1*01:03,HLA-DRB1*07:01,	
	HLA-DRB1*13:01,HLA-DRB1*04:03,	
	HLA-DRB1*13:02	
FVALLPTYLE	HLA-DRB1*10:01,HLA-DRB1*09:01,	
GGNGI	HLA-DRB1*12:01,HLA-DRB1*15:01,	1.0787
	HLA-DRB1*04:05,HLA-DRB1*07:01,	
	HLA-DRB1*01:01,HLA-DRB1*08:02,	
	HLA-DRB1*16:02,HLA-DRB1*04:02,	
	HLA-DRB1*08:01,HLA-DRB1*04:03,	
	HLA-DRB1*13:02,HLA-DRB1*04:01,	

	HLA-DRB1*13:01,HLA-DRB1*04:04, HLA-DRB1*11:01,HLA-DRB1*01:03, HLA-DRB1*03:01

4.3.1 Identification of Imperative Features of Selected T-cell Epitopes

Some of the important features of the selected T-cell epitopes of both proteins were checked to support our study such as molecular weight, toxicity, PI, allergenicity, hydrophobicity, charge and hydropathicity. Only non-toxin and non-allergen peptides are used for further study (Table 4.11, 4.12, 4.13 and 4.14). Another feature is peptide digesting enzymes which was predicted by Peptide cutter tool. Peptides are considered to be stable if digested by fewer enzymes and are more favorable vaccine targets while the peptides that are digested by several enzymes are considered to be non-stable.

TABLE 4.11: Peptides of ABC transporter permease Protein with non-digesting enzymes, mutation position, toxicity, allergenicity, hydrophobicity, hydrophilicity, charge, molecular weight and PI (MHC class-I alleles)

Peptide	**Allergenicity**	**Toxicity**	**Hydrophobicity**	**Hydrophilicity**	**Charge**	**pI**	**MW**
LLIFFIQFL	NA	NT	0.47	-1.81	0.00	5.88	115-3.47
LSLAVFPFF	NA	NT	0.37	-1.42	0.00	5.88	104-0.27
LLLIFFIQF	NA	NT	0.47	-1.81	0.00	5.88	115-3.47
ILLLIFFIQ	NA	NT	0.48	-1.73	0.00	5.88	111-9.46
Non-digesting enzymes	Arg-C proteinase, Asp-N endopeptidase, Asp-N endopeptidase, +N-terminal Glu, BNPS-Skatole,CNBr, Caspase 1-10, Clostripain, Enterokinase,Factor Xa,						

	Glutamyl endopeptidase, GranzymeB, Hydroxylamine, Iodosobenzoic acid, LysC, LysN, Proline-endopeptidase, Staphylococcal peptidase I, Thrombin, Tobacco etch virus protease.

Table 4.12: Peptides of ABC transporter permease Protein with non-digesting enzymes, toxicity, allergenicity, hydrophobicity, hydrophilicity, charge, molecular weight and PI (MHC class-II alleles)

Peptide	**Allergenicity**	**Toxicity**	**Hydrophobicity**	**Hydrophilicity**	**Charge**	**pI**	**MW**
SVFRALPFIILLALI	NA	NT	0.30	-1.12	1.00	10.11	1686.61
VFRALPFIILLALIA	NA	NT	0.33	-1.17	1.00	10.11	1670.16
FRALPFIILLALIAP	NA	NT	0.29	-1.07	1.00	10.11	1668.14
AIVQTLYMTFWSFLI	NA	NT	0.26	-1.43	0.00	5.88	1833.22
Non-digesting enzymes	Asp-N endopeptidase, Asp-N endopeptidase, +N-terminal Glu, BNPS-Skatole, LysC, LysN, Caspase 1-10, Clostripain, Enterokinase,Factor Xa, Glutamyl endopeptidase, Granzyme B, Hydroxylamine, Iodosobenzoic acid, Proline-endopeptidase, Staphylococcal peptidase I, Thrombin, Tobacco etch virus protease.						

Table 4.13: Peptides of nrdI Protein with non-digesting enzymes, toxicity, allergenicity, hydrophobicity, hydrophilicity, charge, molecular weight and PI (MHC class-I alleles)

Peptide	**Allergenicity**	**Toxicity**	**Hydrophobicity**	**Hydrophilicity**	**Charge**	**pI**	**MW**
PMLGDFELR	NA	NT	-0.12	0.18	-1.00	4.38	1077.26

GDFELRGTS	NA	NT	-0.23	0.51	-1.00	4.38	981.03
Non-digesting enzymes	BNPS-Skatole, Caspase 1-10, Enterokinase, Factor Xa, Glutamyl endopeptidase, Granzyme B, Hydroxylamine, Iodosobenzoic acid, LysC, LysN, Proline-endopeptidase, Staphylococcal peptidase I, Thrombin, Tobacco etch virus protease.						
ISLSGNTQS	NA	NT	-0.1	-0.3	0.00	5.88	905.96
Non-digesting enzymes	Arg-C proteinase, Asp-N endopeptidase, Asp-N endopeptidase, +N-terminal Glu, BNPS-Skatole, CNBr, Caspase 1-10, Clostripain, Enterokinase, Factor Xa, Glutamyl endopeptidase, Granzyme B, Hydroxylamine, Iodosobenzoic acid, LysC, LysN, Proline-endopeptidase, Staphylococcal pep tidase I, Thrombin, Tobacco etch virus protease, Trypsin.						

Table 4.14: Peptides of nrdI Protein with non-digesting enzymes, toxicity, allergenicity, hydrophobicity, hydrophilicity, charge, molecular weight and PI (MHC class-II alleles)

Peptide	**Allergenicity**	**Toxicity**	**Hydrophobicity**	**Hydrophilicity**	**Charge**	**pI**	**MW**
PFVALLPTYLEGGNG	NA	NT	0.13	-0.63	-1.00	4.00	1547.77
FVALLPTYLEGGNGI	NA	NT	0.18	-0.75	-1.00	4.00	1563.81
Non-digesting enzymes	Arg-C proteinase, Asp-N endopeptidase, Asp-N endopeptidase, +N-terminal Glu, BNPS-Skatole, CNBr, Caspase 1-10, Clostripain, Enterokinase, Factor Xa, Glutamyl endopeptidase, Granzyme B, Hydroxylamine, Iodosobenzoic acid, LysC, LysN, Proline-endopeptidase, Staphylococcal pep tidase I, Thrombin, Tobacco etch virus protease, Trypsin.						
FPMLGDFELRGTSSD	NA	NT	-0.12	0.15	-2.00	4.03	1671.84
RFGFPMLG-	NA	NT	-0.10	-0.05	0.00	6.42	1741-

DFELRGT		3.01
Non-digesting enzymes	BNPS-Skatole, Caspase 1-10,Enterokinase, Factor Xa, Glutamyl endopeptidase,Granzyme B, Hydroxylamine, Iodosobenzoic acid, LysC, LysN, Proline-endopeptidase, Staphylococcal peptidase I, Thrombin, Tobacco etch virus protease.	

4.4 Epitope Conservation Analysis

Two surface proteins like ABC transporter permease protein and nrdI protein were Blast against bacteria of same species and multiple sequences were obtained. After Blast, ABC transporter permease protein results in 2 proteins and nrdI protein also results in 2 proteins was subjected to multiple alignment.

The epitope conservancy analysis of B-cells, MHC class-I and MHC class-II alleles was performed by using IEDB conservancy tool (table 4.15 and 4.16). Figure 4.8 and 4.9 is showing multiple sequence alignment of ABC transporter protein and nrdI protein.

Table 4.15: Conservation of epitopes of ABC transporter permease Protein Via IEDB epitope conservancy analysis tool

Epitope sequence	**Epitope length**	**% age of sequence matches at identity equal or <100**	**Minimum identity**	**Maximum identity**
LLIFFIQFL	9	100.00% (1/1)	100%	100%
LSLAVFPFF	9	0.00% (0/1)	44.44%	44.44%
LLLIFFIQF	9	0.00% (0/1)	88.89%	88.89%
ILLLIFFIQ	9	0.00% (0/1)	77.78%	77.78%
SVFRALP-FIILLALI	15	0.00% (0/1)	33.33%	33.33%
VFRALPF-IILLALIA	15	0.00% (0/1)	33.33%	33.33%

FRALPFI-ILLALIAP	15	0.00% (0/1)	33.33%	33.33%
AIVQTLYM-TFWSFLI	15	0.00% (0/1)	26.67%	26.67%

TABLE 4.16: Conservation of epitopes of nrdI Protein via IEDB epitope conservancy analysis tool

Epitope sequence	**Epitope length**	**% age of sequence matches at identity equal or <100**	**Minimum identity**	**Maximum identity**
PMLGDFELR	9	0.00% (0/1)	55.56%	55.56%
ISLSGNTQS	9	0.00% (0/1)	55.56%	55.56%
GDFELRGTS	9	0.00% (0/1)	55.56%	55.56%
PFVALLPT-YLEGGNG	15	0.00% (0/1)	33.33%	33.33%
FVALLPTY-LEGGNGI	15	0.00% (0/1)	33.33%	33.33%
FPMLGDFE-LRGTSSD	15	0.00% (0/1)	40%	40%
RFGFPMLG-DFELRGT	15	0.00% (0/1)	40%	40%
GSGNRNF-NNQYCLT	14	0.00% (0/1)	21.43%	21.43%
GGNGIDN-GDQEILT	14	0.00% (0/1)	35.71%	35.71%
GFPMLGD-FELRGTS	14	0.00% (0/1)	42.86%	42.86%

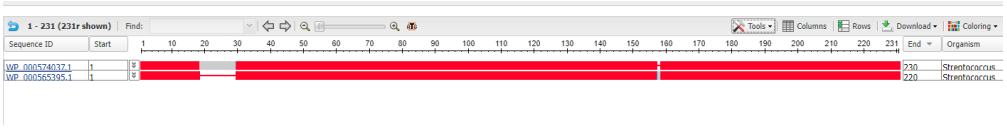

FIGURE 4.8: Multiple sequence alignment of ABC transporter permease Protein

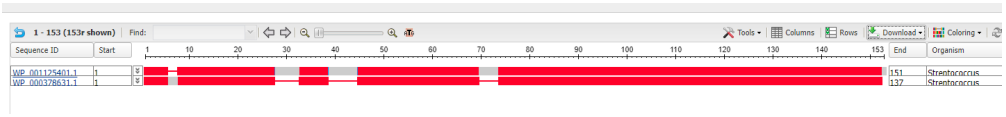

FIGURE 4.9: Multiple sequence alignment of nrdI protein

4.5 Multiple-epitope Vaccine Design and Construction

Two vaccines are designed by arranging the B-cell, MHC class-I and MHC class-II peptides with the help of linkers. Linkers improve the folding of peptides and stability of vaccine.

GGGS linker is used to join MHC class-I peptides while GPGPG linker is used to join MHC class-II peptides and KK linker is used to join B-cell peptides. B-defensin as an adjuvant and EAAAK linker is added at N-terminal to increase the immunogenicity of vaccine.

Figure 4.10 shows the graphical representation of vaccines.

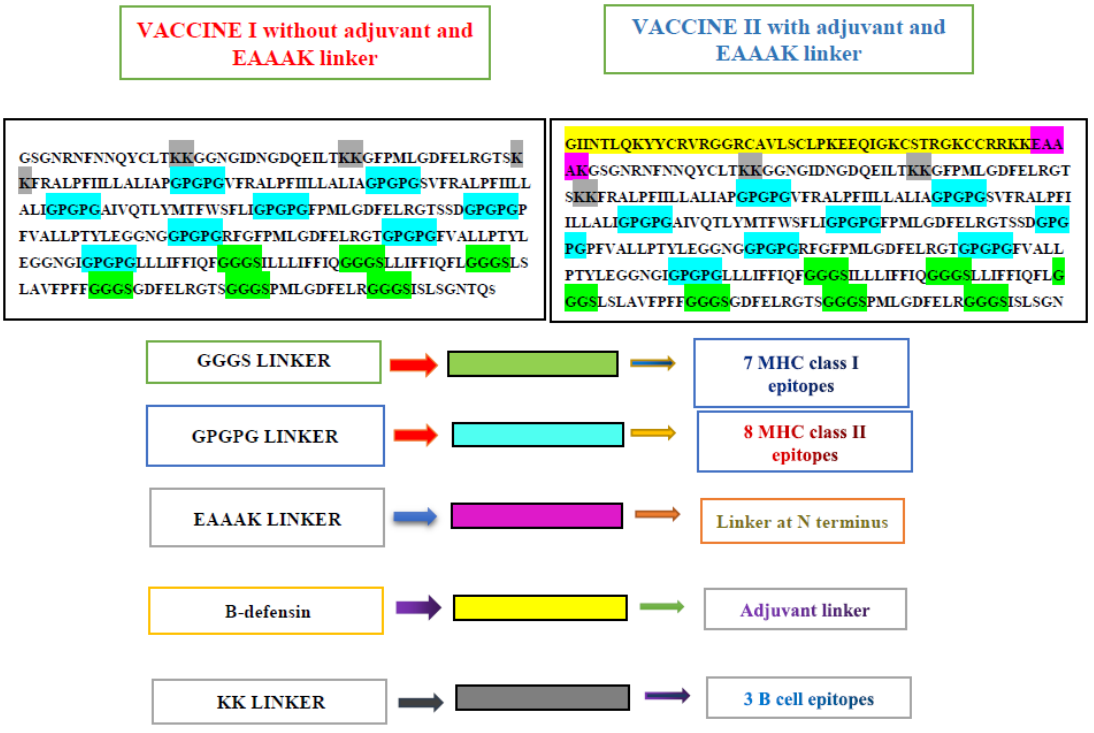

FIGURE 4.10: Graphical representation of VAC I and VAC II.

4.5.1 Physio-Chemical Properties of Multi-epitope Vaccines

Various physio-chemical properties of both vaccines were determined. Both vaccines are non-allergen determined by AllerTOP and are antigenic as vaccine 1 has 1.2382 Vaxijen score while vaccine 2 has 1.1671 Vaxijen score. The molecular weight of vaccine 1 is 30267.19KDa while vaccine 2 has 35724.71KDa. The PI value of vaccine 1 is 6.51 while vaccine 2 has 9.41. The instability index of vaccine 1 is 28.34 while vaccine 2 has 30.91 which means both vaccines are stable. The GRAVY of vaccine 1 is 0.437 while vaccine 2 has 0.291. The aliphatic index of vaccine 1 is 95.90 while vaccine 2 has 91.80. The protein solubility of vaccine 1 is 0.541 while vaccine 2 has 0.575. But their estimated half-life is same that is half-life is 30 hours in mammalian reticulocytes if tested in vitro while if it is tested in-vivo it is greater than 20 hours in yeast and greater than 10 hours in Escherichia coli if tested in vivo. Both vaccines have G (Gly) at N terminal. Table 4.17, 4.18 is showing physio-chemical properties of multi-epitope vaccines. The population coverage analysis was performed on all T and B-cell epitopes for the constructed vaccine. The population coverage for designed vaccine is 95.77 in Pakistan (Table 4.19). Figure 4.11 is showing the graphical representation of population coverage of vaccine in Pakistan. The simulation of immune response of designed vaccine was checked by C-IMMSIM. The immune responses of vaccine with and without adjuvant was predicted. The best results was shown by vaccine without adjuvant. The figure 4.12 and 4.13 is showing immune simulation response of vaccine with and without adjuvant. After injecting the vaccine in the body, it trigger the several immune responses such as IgM and IgC production, concentration of Cytokines, population level of B-cell, TH cell(Helper) and TC (Cytotoxic). The C-IMMSIM results show that designed vaccine trigger sufficient innate and adaptive immune response.

Table 4.17: Physio-chemical properties of vaccine 1 and 2

Vaccine	**Antigenicity**	**Allergenicity**	**GRAVY**	**M.W**
Vaccine 1	1.2382	N.A	0.437	30267.19

Vaccine 2	1.1671	N.A	0.291	35724.71

TABLE 4.18: Physio-chemical properties of vaccine 1 and 2

Vaccine	**Instability index**	**Aliphatic index**	**PI**	**Protein solubility**
Vaccine 1	28.34	95.9	6.51	0.541
Vaccine 2	30.91	91.89	9.41	0.575

TABLE 4.19: Population coverage of designed vaccine in Pakistan

MHC class	**Coverage**	**Average hit**	**PC90**
Combined	95.77	11.33	7.16

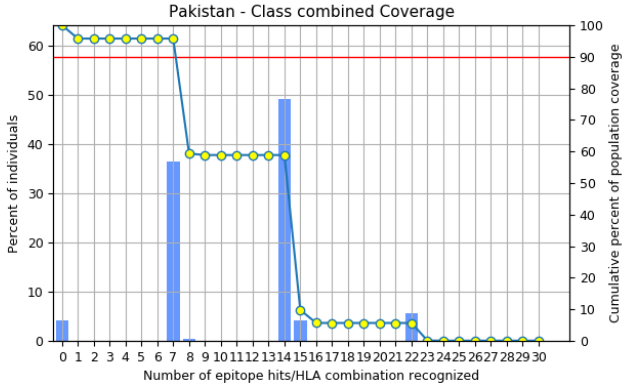

FIGURE 4.11: The Population Coverage of multi-epitope vaccine in Pakistan.

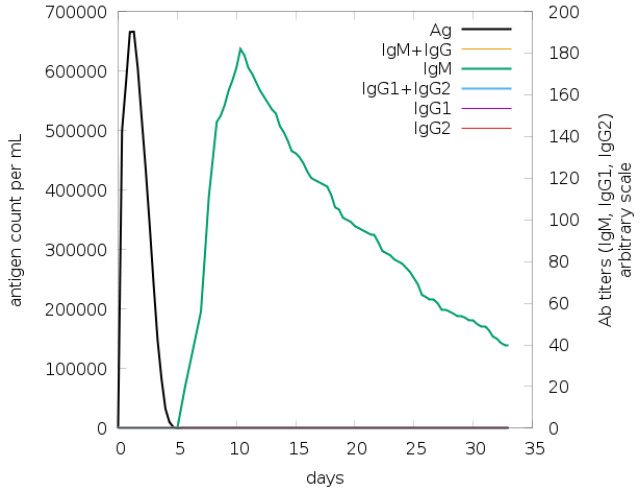

FIGURE 4.12: Immune Simulation of Multi-epitope Vaccine with adjuvant.

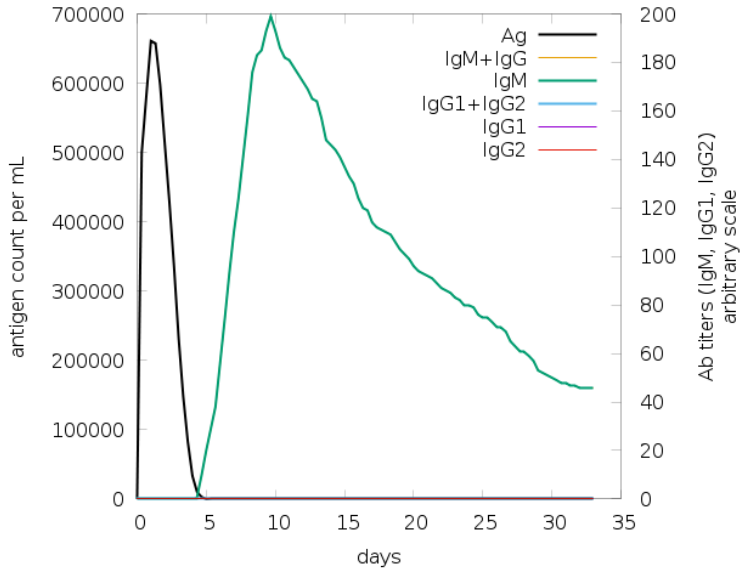

FIGURE 4.13: Immune Simulation of Multi-epitope Vaccine without adjuvant.

4.5.2 Prediction, Refinement and Validation of Multi-epitope Vaccine's 3D Structure

The 3D structure of multi-epitope vaccine was predicted by I-TASSER which gave top 5 models. On the basis of higher C-score (Confidence score), first model out of five models is selected. The selected model has C-score= -3.69, estimated TM-score = 0.31±0.10 and estimated RMSD = 15.7±3.3Å. The selected model was further refined by GalaxyRefine tool and then it is validated by Ramachandran analysis and ERRAT. ERAAT is a non-bounded atomic interactions "overall quality factor", with higher score signifying higher quality. ERAAT shows multi-epitope vaccine has 78.5714 before refinement (Figure 4.14a) and after refinement, it has 83.1395 accuracy (Figure 4.14b). Ramachandran analysis shows that 93.333% of all residues are in favored region accuracy of 3D structure of multi-epitope vaccine (Figure 4.15).

4.6 Molecular Docking

The selected refined model of vaccine 2 through GalaxyRefine is docked with TLR 2 and TLR 4 separately via ClusPro. ClusPro gave top 10 models for both TLRs

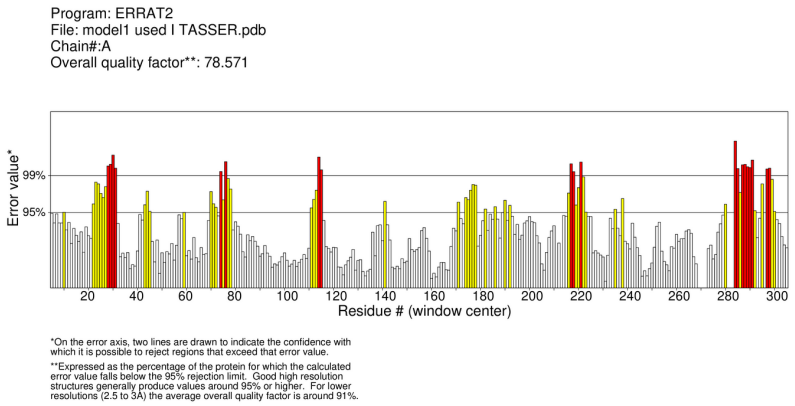

FIGURE 4.14A: ERAAT plot of Multi-Epitope Vaccine before refinement.

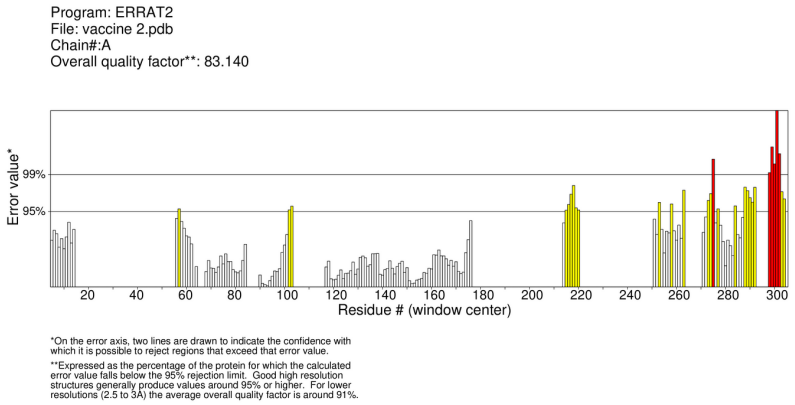

FIGURE 4.14B: ERAAT plot of Multi-Epitope Vaccine after refinement.

but top model was used for further analysis having lowest energy. The best docked structure was selected. The PDBePISA is used to determine the macro-molecular interfaces of docked structures. It analyses the protein-protein interaction and salt bridges, hydrogen bonding, covalent bonds and disulphite bonds. They show no disulphite and covalent bonding.

TABLE 4.20: Determination of hydrogen bonding of docked vaccine 2 with TLR2 via PDBePISA

Sr No	**Structure 1**	**Dist. (Å)**	**Structure 2**
1	A:TRY 376 [HH].	1.83	B:SER 309 [O].

2	A:LYS 347 [HZ3].	1.70	B:THR 361[OG1].
3	A:LYS 347 [HZ1].	1.78	B:THR 363[OG2].
4	A:ASN 345 [OD1].	1.67	B:LYS 385[HZ2].
5	A:GLU 369 [OE1].	1.81	B:LYS 385[HZ2].
6	A:GLU 374 [OE1].	2.02	B:ARG 337[HH21].
7	A:GLU 375 [OE2].	1.70	B:ARG 337[HH22].
8	A:GLU 375 [OE2].	2.11	B:ARG 337[HE].

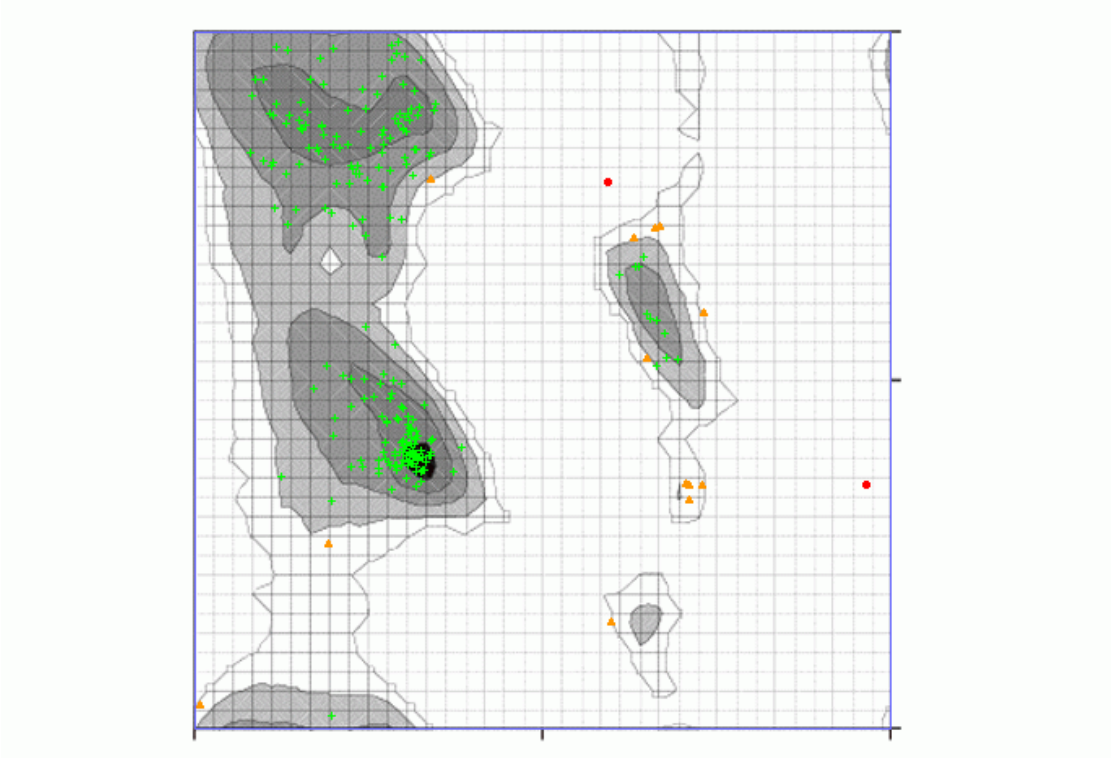

FIGURE 4.15: Ramachandran Analysis of Multi-Epitope Vaccine.

TABLE 4.21: Determination of Salt Bridges of docked vaccine 2 with TLR2 via PDBePISA

Sr No	**Structure 1**	**Dist. ($\AA$)**	**Structure 2**
1	A:GLU 369 [OE1].	2.67	B:LYS 385[NZ].
2	A:GLU 374 [OE1].	2.73	B:ARG 337[NH2].
3	A:GLU 375 [OE1].	3.27	B:ARG 337[NH2].
4	A:GLU 375 [OE2].	2.69	B:ARG 337[NH2].
5	A:GLU 375 [OE2].	2.92	B:ARG 337[NE].

TABLE 4.22: Determination of hydrogen bonding of docked vaccine 2 with TLR4 via PDBePISA

Sr No	**Structure 1**	**Dist. ($\AA$)**	**Structure 2**
1	A:ARG 289 [HH11].	2.49	C:SER 98 [OC].
2	A:ARG 289 [HH12].	1.77	C:ASP 99 [OD1].
3	A:ARG 234 [HH11].	2.08	C:ASP 100 [O].
4	A:ARG 234 [HH21].	1.91	C:ASP 100 [O].
5	A:ARG 264 [HH12].	2.11	C:ASP 101 [OD2].
6	A:ARG 289 [H].	1.82	C:ASP 101 [O].
7	A:ARG 289 [HE].	2.23	C:TYR 102 [OH].
8	A:ARG 289 [HH22].	1.79	C:TYR 102 [OH].
9	A:ARG 87 [HH12].	1.68	C:GLY 110 [O].
10	A:GLU 42 [OE2].	1.98	C:TYR 42 [HH].
11	A:GLU 42 [OE1].	1.83	C:ARG 68 [HH11].
12	A:GLU 42 [OE2].	1.92	C:ARG 68 [HH12].
13	A:ASN 265 [OD1].	1.96	C:SER 103 [H].
14	A: SER 183 [OG].	1.84	C:ARG 106 [HH22].
15	A:ASP 84 [OD2].	1.79	C:LYS 109 [HZ1].
16	A:SER 62 [OG].	1.72	C:LYS 109 [HZ2].
17	A:GLU 135 [OE1].	2.03	C:THR 112 [H].

TABLE 4.23: Determination of salt bridges of docked vaccine 2 with TLR4 via PDBePISA

Sr No	**Structure 1**	**Dist. ($\AA$)**	**Structure 2**
1	A:ARG 289 [NH1].	2.74	C:ASP 99 [OD1].
2	A:ARG 264 [NH1].	3.83	C:ASP 101 [OD1].
3	A:ARG 264 [NH1].	2.71	C:ASP 101 [OD2].
4	A:HIS 159 [ND1].	3.02	C:GLU 111 [0E1].
5	A:GLU 42[OE1].	2.67	C:ARG 68 [NH1].
6	A:GLU 42[OE2].	2.68	C:ARG 68 [NH1].
7	A:ASP 181 [OD2].	3.88	C:ARG 106 [NH2].

8	A:ASP 84 [OD1].	2.98	C:LYS 109[NZ].
9	A:ASP 84 [OD2].	2.98	C:LYS 109[NZ].
10	A:ASP 60 [OD1].	2.69	C:LYS 109[NZ].
11	A:ASP 60 [OD2].	2.76	C:LYS 109[NZ].

Vaccine 2 shows strong binding affinity and interactions with all proteins on the basis of interaction and PDBePISA analysis.

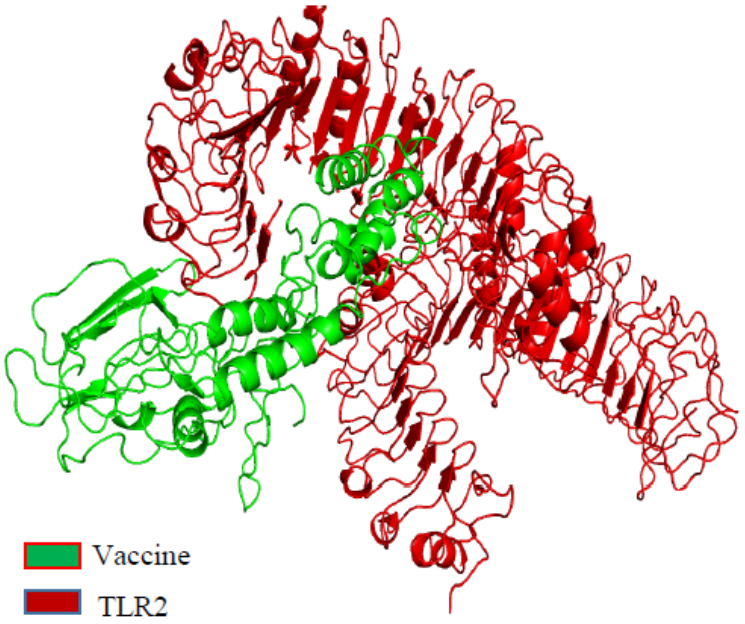

FIGURE 4.16: Protein-Protein interaction of Vac II with TLR2 via ClusPro

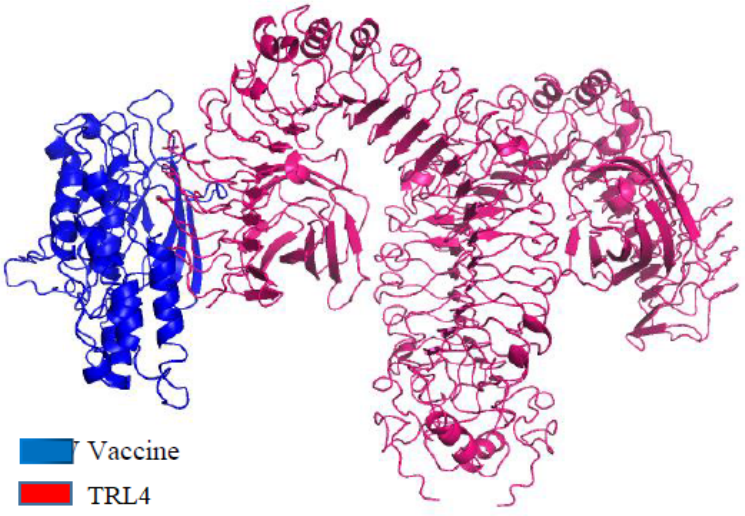

FIGURE 4.17: Protein-Protein interaction of Vac II with TLR4 via ClusPro

Chapter 5

Conclusion and Future Prospects

Streptococcus agalactiae is an anaerobe effecting people of all age groups especially pregnant females and neonates by causing infection in their body. As it is treated through various antibiotics which is not a permanent treatment and it has no specific symptoms so it is a major concern worldwide and nationwide to identify its permanent solution to stop its spread and treat it timely.

The purpose of this study to understand genomic diversity of *Streptococcus agalactiae* by using pan-genome approach and identify vaccine targets against *Streptococcus agalactiae*. 2 essential proteins were selected through core-genome analysis for epitope based study. On the basis of cellular localization, these two proteins were membranous or surface proteins. All the identified targets are playing vital role in the selected pathogen.

The first objective of this study was to explore the pan-genome and essential genes of *Streptococcus agalactiae* and for this purpose, 127 strains of *Streptococcus agalactiae* were analyzed through core-genome analysis approach.

The second objective of this study was to analysis the potential of prioritized virulent factors as immunogen or vaccine targets. For this purpose, 2 proteins were analyzed by epitope-based study. 3 B-cell epitopes, 15 T-cell epitopes (7 MHC class I epitopes and 8 MHC class II epitopes) of both proteins ABC transporter permease protein and nrdI protein were used for vaccine designing. 3 types of linkers were used to join B-cell and T-cell epitopes in vaccine designing. GGGS

linker was used to join MHC class I epitopes, GPGPG linker was used in MHC class II and KK linker was used in B-cell epitopes. The second vaccine has B-defensin, an adjuvant and linker EAAAK at the N-terminal. The 2 designed-vaccine were docked against TLRs but vaccine 2 with adjuvant and EAAAK linker showed best interactions with TLRs.

A reverse vaccinology approach was also used to find out the surface exposed peptides that are more efficient as it is cost-effective but less time-consuming approach. As there is no permanent treatment and FDA approved vaccine is available against *Streptococcus agalactiae* of 127 strains, epitope-based study is used to design vaccine which will enhance the immune response against all natural infections caused by *S. agalactiae*. A lot of efforts are put into developing epitope-based vaccines because it provides the opportunity to engineer the combinations of epitopes as per requirements. This will also help us to facilitate the required immune responses on the T-cell epitopes.

The data given in this study require further experimental authentication for verification but we anticipate promising outcomes from this predicted peptide-epitopes against the *Streptococcus agalactiae*.

Bibliography

[1]. S. Delfani, M. Bahmani, R. Mohammadrezaei-Khorramabadi, and M. Rafieian-Kopaei, "Phytotherapy in Streptococcus agalactiae: An Overview of the Medicinal Plants Effective against Streptococcus agalactiae," Journal of Clinical and Diagnostic Research: JCDR, vol. 11, no. 6, p. DE01, Jun. 2017, doi: 10.7860/JCDR/2017/25530.9988.

[2]. N. Al-Sweih, M. Hammoud, M. Al-Shimmiri, M. Jamal, L. Neil, and V. Rotimi, "Serotype distribution and mother-to-baby transmission rate of Streptococcus agalactiae among expectant mothers in Kuwait," Archives of Gynecology and Obstetrics, vol. 272, no. 2, pp. 131–135, 2005, doi: 10.1007/s00404-004-0705.

[3]. J. Vornhagen, K. M. Adams Waldorf, and L. Rajagopal, "Perinatal Group B Streptococcal Infections: Virulence Factors, Immunity, and Prevention Strategies," Trends in Microbiology, vol. 25, no. 11, pp. 919–931, Nov. 2017, doi: 10.1016/J.TIM.2017.05.013.

[4]. G. H. Rosen et al., "Group B Streptococcus and the Vaginal Microbiota," Journal of Infectious Diseases, vol. 216, no. 6, pp. 744–751, Sep. 2017, doi: 10.1093/INFDIS/JIX395

[5]. S. Borges, J. Silva, and P. Teixeira, "Survival and biofilm formation by Group B streptococci in simulated vaginal fluid at different pHs," Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, vol. 101, no. 3, pp. 677–682, Mar. 2012, doi: 10.1007/S10482-011-9666-Y.

[6]. S. D. Manning, "Molecular epidemology of Streptococcus agalactiae (Group B Streptococcus)," 2003.

[7]. S. A. Plotkin, "Vaccines: the fourth century," Clinical and Vaccine Immunology, vol. 16, no. 12, pp. 1709–1719, 2009.

[8]. M. Castelli, F. Cappelletti, R. A. Diotti, G. Sautto, E. Criscuolo, M. Dal Per-aro, and N. Clementi,"Peptide-based vaccinology: experimental and computational approaches to target hypervariable viruses through the fine characterization of protective epitopes recognized by monoclonal antibodies and the identification of t-cell-activating peptides," Clinical and Developmental Immunology, vol. 2013, 2013.

[9]. T. Ben-Yedidia and R. Arnon, "Epitope-based vaccine against influenza," Expert review of vaccines, vol. 6, no. 6, pp. 939–948, 2007.

[10]. B. Pulendran and R. Ahmed, "Immunological mechanisms of vaccination," Nature immunology, vol. 12, no. 6, pp. 509–517, 2011.

[11]. M. N. W. Butter and C. E. de Moor, "Streptococcus agalactiae as a cause of meningitis in the newborn, and of bacteraemia in adults," Antonie van Leeuwenhoek 1967 33:1, vol. 33, no. 1, pp. 439–450, Dec. 1967, doi: 10.1007/BF02045596.

[12]. C. M. Delannoy et al., "Human Streptococcus agalactiae strains in aquatic mammals and fish," BMC Microbiology 2013 13:1, vol. 13, no. 1, pp. 1–9, Feb. 2013, doi: 10.1186/1471-2180-13-41.

[13]. V. G. Dutra et al., "Streptococcus agalactiae in Brazil: Serotype distribution, virulence determinants and antimicrobial susceptibility," BMC Infectious Diseases, vol. 14, no. 1, pp. 1–9, Jun. 2014, doi: 10.1186/1471-2334-14-323/2.

[14]. C. S. do Nascimento, N. F. B. dos Santos, R. C. C. Ferreira, and C. R. Taddei, "Streptococcus agalactiae in pregnant women in Brazil: prevalence, serotypes, and antibiotic resistance," Brazilian Journal of Microbiology, vol. 50, no. 4, p. 943, Oct. 2019, doi: 10.1007/S42770-019-00129-8.

[15]. F. J. Bobadilla, M. G. Novosak, I. J. Cortese, O. D. Delgado, and M. E. Laczeski, "Prevalence, serotypes and virulence genes of Streptococcus agalactiae isolated from pregnant women with 35–37 weeks of gestation," BMC

Infectious Diseases, vol. 21, no. 1, pp. 1–11, Dec. 2021, doi: 10.1186/S12879-020-05603-5/3.

[16]. K. S. Doran and V. Nizet, "Molecular pathogenesis of neonatal group B streptococcal infection: No longer in its infancy," Molecular Microbiology, vol. 54, no. 1, pp. 23–31, Oct. 2004, doi: 10.1111/J.1365-2958.2004.04266.X.

[17]. K. L. Atkins, A. Shanks, C. A. Parvin, W. M. Dunne, and G. Gross, "Evaluation of polymerase chain reaction for group b streptococcus detection using an improved culture method," Obstetrics and Gynecology, vol. 108, no. 3, pp. 488–491, 2006, doi: 10.1097/01.AOG.0000228961.42272.31.

[18]. C. Marconi, T. T. Rocchetti, V. L. M. Rall, L. R. de Carvalho, V. T. M. Borges, and M. G. da Silva, "Detection of streptococcus agalactiae colonization in pregnant women by using combined swab cultures: Cross-sectional prevalence study," Sao Paulo Medical Journal, vol. 128, no. 2, pp. 60–62, 2010, doi: 10.1590/S1516-31802010000200003.

[19]. J. Y. Bolukaoto et al., "Antibiotic resistance of Streptococcus agalactiae isolated from pregnant women in Garankuwa, South Africa," BMC Research Notes, vol. 8, no. 1, pp. 1–7, Aug. 2015, doi: 10.1186/S13104-015-1328-0/FIGURES/1.

[20]. L. A. Jackson et al., "Risk factors for group B streptococcal disease in adults," Ann Intern Med, vol. 123, no. 6, pp. 415–420, Sep. 1995, doi: 10.7326/0003-4819-123-6-199509150-00003.

[21]. C. Ruppen, J. Notter, C. Strahm, B. Sonderegger, and P. Sendi, "Osteoarticular and skin and soft-tissue infections caused by Streptococcus agalactiae in elderly patients are frequently associated with bacteremia," Diagnostic Microbiology and Infectious Disease, vol. 90, no. 1, pp. 55–57, Jan. 2018, doi: 10.1016/2017.09.008.

[22]. U. B. S. Sørensen, I. C. Klaas, J. Boes, and M. Farre, "The distribution of clones of Streptococcus agalactiae (group B streptococci) among herdspersons and dairy cows demonstrates lack of host specificity for some lineages,"

Veterinary Microbiology, vol. 235, pp. 71–79, Aug. 2019, doi: 10.1016/2019.06.008.

[23]. Y. Yang et al., "Molecular Characterization of Streptococcus agalactiae Isolated from Bovine Mastitis in Eastern China," Plos One, vol. 8, no. 7, p. e67755, Jul. 2013, doi: 10.1371/0067755.

[24]. J. Gao et al., "Antibiotic resistance of Streptococcus agalactiae from cows with mastitis," The Veterinary Journal, vol. 194, no. 3, pp. 423–424, Dec. 2012, doi: 10.1016/J.TVJL.2012.04.020.

[25]. R. C. Lancefield, "A serological differentiation of human and other groups of hemolytic streptococci," Journal of Experimental Medicine, vol. 57, no. 4, pp. 571–594, Apr. 1933, doi: 10.1084/JEM.57.4.571.

[26]. V. N. Raabe and A. L. Shane, "Group B Streptococcus (Streptococcus agalactiae)," Microbiol Spectr, vol. 7, no. 2, Apr. 2019, doi: 10.1128/ GPP3-0007 -2018.

[27]. A. K. Dutta, "Vaccine Against Covid-19 Disease – Present Status of Development," Indian Journal of Pediatrics, vol. 87, no. 10, pp. 810–816, Oct. 2020, doi: 10.1007/S12098-020-03475-W/1.

[28]. R. Rappuoli, "Reverse vaccinology, a genome-based approach to vaccine development," Vaccine, vol. 19, no. 17–19, pp. 2688–2691, Mar. 2001, doi: 10.1016/S0264-410X (00)00554-5.

[29]. "Mastitis in Cows: Causes, Symptoms, Prevention and Treatment, Cargill India." https://www.cargill.co.in/en/mastitis-in-cows-causes, symptoms, prevention and treatment (accessed Apr. 13, 2022).

[30]. "Group B Strep - Affects 25% of Health Pregnant Women, familydoctor.org." https://familydoctor.org/condition/group-b-strep-infection/ (accessed Apr. 13, 2022).

[31]. R. Sriram, M. Shoffff, G. Booton, P. Fuerst, and G. S. Visvesvara, "Survival of acanthamoeba cysts after desiccation for more than 20 years," Journal of clinical microbiology, vol. 46, no. 12, pp. 4045–4048, 2008.

[32]. M. Adams, J. Kelley, J. Gocayne, M. P. Dubnick, and X. MH, "H. merril, cr wu, a. olde, b. moreno, rf kerlavage, ar mccombie, wr and venter. jc 1991," Complementary DNA sequencing: Expressed sequence tags and human genome project. Science, vol. 252, pp. 1651–1656.

[33]. M. M. Gutacker, B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser, "Single-nucleotide polymorphism–based population genetic analysis of mycobacterium tuberculosis strains from 4 geographic sites," The Journal of infectious diseases, vol. 193, no. 1, pp. 121–128, 2006.

[34]. D. Barh, M. S. Khan, and E. Davies, PlantOmics: the omics of plant science. Springer, 2015.

[35]. L. Zhang, D. Xiao, B. Pang, Q. Zhang, H. Zhou, L. Zhang, J.Zhang, and B.Kan, "The core proteome and pan proteome of salmonella paratyphi a epidemic strains," PloS one, vol. 9, no. 2, p. e89197, 2014.

[36]. "Streptococcus agalactiae." https://www.uniprot.org/taxonomy/1311 (accessed Apr. 13, 2022).

[37]. Vaccine Types HHS.gov." https://www.hhs.gov/ immunization/ basics/ types/ index.html (accessed Apr. 13, 2022).

[38]. K. A. Patras and V. Nizet, "Group B Streptococcal maternal colonization and neonatal disease: Molecular mechanisms and preventative approaches," Frontiers in Pediatrics, vol. 6, p. 27, Feb. 2018, doi: 10.3389/FPED.2018.00027.

[39]. M. I. Rashid, S. Rehman, A. Ali, and S. Andleeb, "Fishing for vaccines against Vibrio cholerae using in silico pan-proteomic reverse vaccinology approach," PeerJ, vol. 2019, no. 6, 2019, doi: 10.7717/PEERJ.6223/FIG-1.

[40]. O. Sharma, A. A. Sultan, H. Ding, and C. R. Triggle, "A Review of the Progress and Challenges of Developing a Vaccine for COVID-19," Frontiers in Immunology, vol. 11, p. 2413, Oct. 2020, doi: 10.3389/FIMMU.2020.585354.

[41]. "Flow diagram representing the main steps in a pan-genomic analysis.. Download Scientific Diagram." https://www.researchgate. net/ figure/ Flow-diagram -representing -the- main- steps- in-a-pan- genomic-analysis-Each-process–fig3343520319 (accessed Apr. 13, 2022).

[42]. T. Kabelitz, E. Aubry, K. van Vorst, T. Amon, and M. Fulde, "The Role of Streptococcus spp. in Bovine Mastitis," Microorganisms 2021, Vol. 9, Page 1497, vol. 9, no. 7, p. 1497, Jul. 2021, doi: 10.3390/9071497.

[43]. M. R. Popoffff, "Clostridial pore-forming toxins: powerful virulence factors," Anaerobe, vol. 30, pp. 220–238, 2014.

[44]. H. Yoon, C. Ansong, J. E. McDermott, M. Gritsenko, R. D. Smith, F. Heffron, and J. N. Adkins, "Systems analysis of multiple regulator perturbations allows discovery of virulence factors in salmonella," BMC systems biology, vol. 5, no. 1, pp. 1–16, 2011.

[45]. L. Zhao, L. Wong, and J. Li, "Antibody-specified b-cell epitope prediction in line with the principle of context-awareness," IEEE/ACM transactions on computational biology and bioinformatics, vol. 8, no. 6, pp. 1483–1494, 2011.

[46]. B. Poulain, J. D. Wadsworth, E. A. Maisey, C. C. Shone, J. Melling, L. Tauc, and J. O. Dolly, "Inhibition of transmitter release by botulinum neurotoxin a: contribution of various fragments to the intoxication process," European journal of biochemistry, vol. 185, no. 1, pp. 197–203, 1989.

[47]. L. Potocnakova, M. Bhide, and L. B. Pulzova, "An introduction to b-cell epitope mapping and in silico epitope prediction," Journal of immunology research, vol. 2016, 2016.

[48]. J. Kubrycht, K. Sigler, and P. Sou˘cek, "Virtual interactomics of proteins from biochemical standpoint," Molecular biology international, vol. 2012, 2012.Bibliography 86

[49]. J. V. Kringelum, M. Nielsen, S. B. Padkjær, and O. Lund, "Structural analysis of b-cell epitopes in antibody: protein complexes," Molecular immunology, vol. 53, no. 1-2, pp. 24–34, 2013.

[50]. E.-M. Yasser and V. Honavar, "Recent advances in b-cell epitope prediction methods," Immunome research, vol. 6, no. 2, pp. 1–9, 2010.

[51]. J. D. Black and J. O. Dolly, "Interaction of 125i-labeled botulinum neurotoxins with nerve terminals. ii. autoradiographic evidence for its uptake into motor nerves by acceptor-mediated endocytosis." The Journal of cell biology, vol. 103, no. 2, pp. 535–544, 1986.

[52]. M. Law, "Editorial overview: Preventive and therapeutic vaccines," Current opinion in virology, vol. 11, p. viii, 2015.

[53]. Z. Chen, J. Li, and L. Wei, "A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue," Artificial Intelligence in Medicine, vol. 41, no. 2, pp. 161– 175, 2007.

[54]. S. Shabayek and B. Spellerberg, "Group B streptococcal colonization, molecular characteristics, and epidemiology," Frontiers in Microbiology, vol. 9, no. MAR, p. 437, Mar. 2018, doi: 10.3389/FMICB.2018.00437/.

[55] N. A. Qureshi, S. M. Bakhtiar, M. Faheem, M. Shah, A. Bari, H. M. Mahmood, M. Sohaib, R. A. Mothana, R. Ullah, and S. B. Jamal, "Genomebased drug target identification in human pathogen streptococcus gallolyticus," Frontiers in Genetics, vol. 12, p. 303, 2021.

[56] A. M. Altenhoff, N. ˇSkunca, N. Glover, C.-M. Train, A. Sueki, I. Pili ˇzota, K. Gori, B. Tomiczek, S. M¨uller, H. Redestig et al., "The oma orthology database in 2015: function predictions, better plant support, synteny view and other improvements," Nucleic acids research, vol. 43, no. D1, pp. D240–D249, 2015.

[57] M. S. Scott, S. J. Calafell, D. Y. Thomas, and M. T. Hallett, "Refining protein subcellular localization," PLoS computational biology, vol. 1, no. 6, p. e66, 2005.

[58] S. I. Mondal, S. Ferdous, N. A. Jewel, A. Akter, Z. Mahmud, M. M. Islam, T. Afrin, and N. Karim, "Identification of potential drug targets by subtractive

genome analysis of escherichia coli o157: H7: an in silico approach,"Advances and applications in bioinformatics and chemistry: AABC, vol. 8,p. 49, 2015.

[59] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "Kegg," New perspectives, vol. 855, 2008.

[60] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch et al., "Protein identification and analysis tools on the expasy server," The proteomics protocols handbook, pp. 571–607, 2005.

[61] D. W. Buchan, F. Minneci, T. C. Nugent, K. Bryson, and D. T. Jones, "Scalable web services for the psipred protein analysis workbench," Nucleic acids research, vol. 41, no. W1, pp. W349–W357, 2013.

[62] F. Ferr'e and P. Clote, "Dianna 1.1: an extension of the dianna web server for ternary cysteine classification," Nucleic acids research, vol. 34, no. suppl 2, pp. W182–W185, 2006.

[63] I. A. Doytchinova and D. R. Flower, "Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines," BMC bioinformatics, vol. 8, no. 1, pp. 1–7, 2007.

[64] C. N. Magnan, A. Randall, and P. Baldi, "Solpro: accurate sequence based prediction of protein solubility," Bioinformatics, vol. 25, no. 17, pp. 2200–2207, 2009.

[65] Y. EL-Manzalawy, D. Dobbs, and V. Honavar, "Predicting linear b-cell epitopes using string kernels," Journal of Molecular Recognition: An Interdisciplinary Journal, vol. 21, no. 4, pp. 243–255, 2008.

[66] J. M. S´anchez-Calvo, G. R. Barbero, G. Guerrero-V´asquez, A. G. Dur´an, M. Mac´ıas, M. A. Rodr´ıguez-Iglesias, J. M. Molinillo, and F. A. Mac´ıas, Bibliography 89 "Synthesis, antibacterial and antifungal activities of naphthoquinone derivatives: A structure–activity relationship study," Medicinal Chemistry Research, vol. 25, no. 6, pp. 1274–1285, 2016.

[67] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "Bepipred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes," Nucleic acids research, vol. 45, no. W1, pp. W24–W29, 2017.

[68] P. Koehl and M. Levitt, "Structure-based conformational preferences of amino acids," Proceedings of the National Academy of Sciences, vol. 96, no. 22, pp. 12 524–12 529, 1999.

[69] P. Sun, H. Ju, Z. Liu, Q. Ning, J. Zhang, X. Zhao, Y. Huang, Z. Ma, and Y. Li, "Bioinformatics resources and tools for conformational b-cell epitope prediction," Computational and mathematical methods in medicine, vol. 2013, 2013.

[70] E. A. Johnson and M. Bradshaw, "Clostridium botulinum and its neurotoxins: a metabolic and cellular perspective," Toxicon, vol. 39, no. 11, pp. 1703–1722, 2001

[71] Alqahtani, "Epitope-based peptide vaccine design and target site depiction against middle east respiratory syndrome coronavirus: an immuneinformatics study," Journal of translational medicine, vol. 17, no. 1, pp. 1–14, 2019.

[72] M. Munia, S. Mahmud, M. Mohasin, and K. M. K. Kibria, "In silico design of an epitope-based vaccine against choline binding protein A of Streptococcus pneumoniae," Informatics in Medicine Unlocked, vol. 23, p. 100546, Jan. 2021, doi: 10.1016/J.IMU.2021.100546.

[73] H. B. Kolla, C. Tirumalasetty, K. Sreerama, and V. S. Ayyagari, "An immunoinformatics approach for the design of a multi-epitope vaccine targeting super antigen TSST-1 of Staphylococcus aureus," Journal of Genetic Engineering & Biotechnology, vol. 19, no. 1, Dec. 2021, doi: 10.1186/S43141-021-00160-Z.

[74] M. Sana, A. Javed, S. Babar Jamal, M. Junaid, and M. Faheem, "Development of multivalent vaccine targeting M segment of Crimean Congo Hemorrhagic Fever Virus (CCHFV) using immunoinformatic approaches," Saudi Journal of Biological

[75] S. Ebrahimi, H. Mohabatkar, and M. Behbahani, "Predicting Promiscuous T Cell Epitopes for Designing a Vaccine Against Streptococcus pyogenes," Applied Biochemistry and Biotechnology, vol. 187, no. 1, pp. 90–100, Jan. 2019, doi: 10.1007/S12010-018-2804-5/6.