CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Pan Genome Analysis of *Streptococcus agalactiae* for Drug Target Prioritization

by

Tayuba Ishaq

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

Every challenging work needs self-efforts as well as the guidance of elders. I dedicate this thesis to my parents whose affection makes me able to get such success and to my teachers whose encouragement has always been my source of inspiration.

CERTIFICATE OF APPROVAL

Pan Genome Analysis of *Streptococcus agalactiae* for Drug Target Prioritization

by

Tayuba Ishaq

(MBS203037)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Sobia Tabassum	IIU, Islamabad
(b)	Internal Examiner	Dr. Sania Riaz	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 Bioinformatics & Biosciences
December, 2022

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

Author's Declaration

I, **Tayuba Ishaq** hereby state that my MS thesis titled "**Pan Genome Analysis of *Streptococcus agalactiae* for Drug Target Prioritization** " 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.

(Tayuba Ishaq)

Registration No: MBS203037

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Pan Genome Analysis of *Streptococcus agalactiae* for Drug Target Prioritization** " is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully 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.

(Tayuba Ishaq)

Registration No: MBS203037

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 continuous 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 Bio-sciences, Dr. Syeda Marriam Bakhtiar for giving me the opportunity to pursue 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 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.

Thanks to all.

(Tayuba Ishaq)

Abstract

Current life styles, stress and toxic wastes have extremely increased the incidence of various infections in humans. Bacterial infections account for a major cause of deaths throughout the developing world. It is estimated that by the end of 2025, 226 million people worldwide will be affected with infectious diseases. Due to massive use of antibiotics bacteria gain resistance by biofilm formation against certain antibiotics. Multidrug resistance bacteria commonly use cell to cell communication which is called quorum sensing which leads to the regulation of numerous virulence factors in bacterial biofilm for developing antibiotic resistance. Bacteria may infect humans ,animals and the infections they cause are harder to treat than those caused by non-resistant bacteria. A recognized *Streptococcus agalactiae* is a pathogen known to cause infections in newborns, the elderly and immune compromised individuals. Bacterial infections such as urinary tract infection, intra amniotic infection, bones and joint infection, tissue infection, brain fever, pneumonia and postpartum infection especially with opportunistic bacteria such as *Streptococcus agalactiae* is hard to treat due to their high potential. The increasing global expansion of multi-resistant bacteria which cause diseases and cannot be treated with antibiotics or other anti-microbial drugs is particularly concerning. Because there is no effective drug against these bacteria to treat infections. In this study, pan genome approach was utilized to develop drug targets for 127 strains of *Streptococcus agalactiae*. From the total 572 core proteins, there were 355 non-host homologous proteins from which we selected 3 essential proteins by applying 2 thresholds that are identity greater than 35 and e-value=0.001. From these 3 proteins were selected as drug targets and docked against 9 anti-microbial compounds. After docking, against each protein one compound is selected based on the docking score and 3D structure of one docked compound is shown that either have highest docking score or number of residues interaction. In the Drug-targeting study, there are also some compounds that has shown highest docking score with more than one protein. Lead compound was selected on the basis of highest docking score and residues interations with protein named DNA binding

response regulator. These results can be further validated by in vitro analysis and can proceed for clinical trials.

Contents

List of Figures

List of Tables

Abbreviations

CPS:-	Capsular polysaccharide
CNS:-	Central Nervous System
DEG :-	Database of Essential Genes
EOS:-	Early Onset Sepsis
ECM:-	Extracellular matrix
ERRAT:-	Overall quality factor
IAI:-	Intra-Amniotic Infection (IAI)
KEGG:-	Kyoto Encyclopedia of Genes and genome
LOS:-	Late Onset Sepsis
MOE:-	Molecular Operating Environment
MW:-	Molecular Weight
MICs:-	Minimum Inhibitory Concentrations
NT:-	Non-toxin
NCBI :-	National Center for Biotechnology Information
PDB:-	Protein Data Bank
S.agalactiae:-	*Streptococcus agalactiae*
SBDD :-	Structure Based Drug Design
SPA:-	Spinal epidural abscess
SRRP :-	Serine-rich repeat proteins
UniProt:-	Universal Protein Resource
UTI :-	Urinary Tract Infection
VLBW:-	Very low birth weight
WHO:-	World Health Organization

Chapter 1

Introduction

Streptococci are one of the important genera found in the oral cavity including skin and mucosa [1]. *Streptococcus agalactiae* is Gram-positive cocci, which is among the significant reasons for cow-like mastitis, a disease with high financial impacts in agriculture. It is a normal flora in healthy adults but one of the most common opportunistic pathogens in immune-compromised adults and old people [2]. And also, can cause a hazardous infection that could even be fatal [3]. About 30 percent of bacteria are found in the digestive and reproductive tracts of children, females, old age individuals, and also in immune compromised people. *S.agalactiae* is a common pathogen for various clinical conditions including sepsis, brain fever, pneumonia, urinary tract infections, bone infection and the epidermis or tissue infections.

S. agalactiae has additionally been demonstrated to be present in breast milk which could be a threat to the health of neonates, especially preterm babies [4]. In late onset bacterial infection, i.e., after 4 months of birth, breast milk could be a source of exposure [5].

Meningitis is an extreme sign of invasive *Streptococcus agalactiae* infection. Microbial illness-causing sepsis and meningitis in newborns which could result in acute illness, disability for a long time, and death [5]. This bacterium is progressively connected with infection and enters or invade non-pregnant women and

men and elderly patients with malignant growth, causing blood sugar level high, renal dialysis, and other huge hidden sicknesses [6]. Every year, more than 21 million pregnant ladies worldwide are colonized with *Streptococcus agalactiae*, which involve approximately 18% of pregnancies [3]. In addition to human illness, *Streptococcus agalactiae* causes serious veterinary and agricultural problems, therefore it can affect fishes and ruminant animals' mammary glands.

The rate of pregnant women suffering from the colonization of *S. agalactiae* varies but it is a common practice all over the world to have screening for *S. agalactiae* between 35 and 37 weeks of pregnancy. it is also reported that the infection rate among pregnant women is 12.5 % in south Asia, 11% in eastern Asia, and varies between 11% to 34.7% worldwide [7]. As indicated by the part of the united nation that deals with major health issues around the world called (WHO), *S. aglactiae* causes 150,000 stillbirths and newborn children passing globally [8]. As the amniotic fluid can be the source of infection in neonates, especially the children born by vaginal birth. *S. agalactiae* must release a number of virulence factors in order to cause illness, including capsular polysaccharide, the alpha and beta antigens of the C protein, surface protein Rib, hyaluronate lyase, and C5a peptidase, which are encoded by the cps, BCA, bac, rib, and scpBgenes, respectively. It is biologically possible that direct face contact with anybody who is infected with this virus, such as coworkers, family members, or healthcare staff, might result in neonatal infection [2]. The major virulence component of S. agalactiae is the capsular polysaccharide, which is often used for strain classification [2].

Antigenic differentiation of 9 distinct serotypes of *S. agalactiae* (Ia, Ib, II-VIII) is possible because of differences in capsular polysaccharide structure. Serotypes Ia, II, III and V are recognized in 80-90 percent of all clinical isolates, as per studies from the United States and Europe. The division of serotypes seen then again, varies by ethnic background and geological region. Serotypes VI and VIII strains are s most ordinarily in Japanese colonization examinations [27]. The rise of new serotypes seems to be a non-stop process. Serotype V strains have been created as a novel and major serotype in the last 10 years reflecting continuous epidemiological changes and the need for proceeded with epidemiological observation [28], [9].

Although based on the capsular polysaccharide *S. agalactiae* has historically been divided into nine serotypes (Ia, Ib, II, III, IV, V, VI, VII, VIII) and in the the United States, a tenth serotype (IX) among non-pregnant women was found in 2007 [9]. Serotypes Ia, II, and III were found to be the most frequently occurring serotypes causing severe illness in V group(29% from 2005 to 2006) [11].Based on the capsular polysaccharide, the dominant serotypes that cause illness differ by invasive and colonizing isolates and vary regional-ly [10].

Serotyping of *S. agalactiae* is helpful to get the study of disease trans-mission, for contributing to existing serotype profiles in the area for rational, viable, and broad serotype inclusion *S. agalactiae* immunization advancement and monitoring of serotype replacement or capsular switching. But there is a shortage of serotype profile information on this microbe colonizing pregnant women and babies in Africa where studies have *S. agalactiae* have reported resistance to tetracycline, chloramphenicol, ciprofloxacin, erythro-mycin, ampicillin, vancomycin, and penicillin. As Penicillin is the most common medication of choice for the treatment of GBS diseases. Intermediate sensitivity to penicillin as well as reduced Minimum Inhibitory Concentrations (MICs) have also been reported. Alternative an-tibiotics including macrolides and lincosamide are available for penicillin hypersensitive individuals, while vancomycin may be avoided for penicillin allergic women who are at high risk of hypersensitivity [11].

Similarly, clindamycin, erythromycin, and levofloxacin are utilized as choices for Betalactam hypersensitive patients. An unfortunate increase in antibiotic resistance among GBS as well as a side effect of the excessive use of antibiotics has started to appear [3]. As a result, developing and implementing strategies to not only identify hosts and use antibiotics properly, but also to prevent invasive infections with medications is necessary to reduce the burden of *S. agalactiae* disease [12], [13]. Computational techniques have been developed to quickly find new targets in the post genomic period. This strategy has been quite successful in the cases of *Cyanobacterium Dipheria* and *Burkholeria pseudiomellei* [14], [15] along with comparative microbial and differential analyses of the genomes of *Mycobacterium tuberculosis* [16], *Helicobacter pylori* [17], *Salmonella typhi* [18],

Pseudomonas aeruginosa [19] and *N. gonorrhe* [20], [21]. In the drug discovery process, the identification of therapeutic targets is in earlier stages. When this process is completed, research on target identification and drugs will step into a new era [22], [23]. Host genome sequences on a genomic basis for any pathogens and availability of pathogen has facilitated the efficient identification of drug targets [24]. For identification of the drug targets which are properly characterized and functional analysis is validated, computational methods could be a way out to avoid a lengthy process of hit and trial in the conventional drug discovery process [25]. The pace of the ventures increases with the decrease in the expense that systems can now see entire microbes by *In- silico* methods. The primary disease molecular pathways have evolved in drug discovery processes, moving from classical ligand bbased drug discovery to structural and selective drug design. With cell biology principles and an overall understanding of the microbe as a whole, new opportunities for identifying computational drug targets will be opened [26].

1.1 Problem Statement

The main cause of morbidity and mortality in cystic fibrosis and immune compromised patients is an opportunistic pathogen called *S. agalactiae*. These bacteria form biofilm which helps them to resist multiple antibiotics. Therefore, its eradication has become more challenging. Increasing extent of pathogenic resistance to drugs has encouraged the search for new anti-virulence drugs. Our study will help will help toward the identification of novel drug that may lead to discovery of active treatment option against disease which are causing by *S.agalactiae* in humans.

1.2 Research Objectives

This study is defined with the aim to explore the essential non host homologous genes of *Streptococcus agalactiae* for the identification of drug targets.

To achieve the aim of the project is designed with the following objectives.

1. To identification of core genome of *Streptococcus agalactiae* by Pan genome analysis.

2. To identify antibacterial compounds of *Streptococcuss agalactiae* as potential inhibitors of target proteins.

3. To identify anti-bacterial drug candidate from host lead compound.

Chapter 2

Review of Literature

2.1 *Streptococcus agalactiae* and its Characteristics

Streptococcus, Lactococcus, and *Lactovum* are the three genera that make up the Streptococcaceae family; however, Streptococcus is now the most diverse within 79 species. These Gram-positive pathogens often appear in pairs or chains, are circular to ovoid in shape, need fermentative digestion, and form capsules in high numbers [27]. *Streptococcus* species (for example *S. agalactiae, S. pneumoniae* and *S. pyogenes* are divided from group A to V based on the carbohydrates found in their cell walls [28]. *Streptococci* are characterized based on biochemical reactions, cell wall pili, morphological variations, type of hemolysis on blood agar, associated protein, and polysaccharide capsule (specific for Group B streptococci). More than 85 capsules antigenic of *S. pneumoniae*, there have been proposed nine CPS (capsular polysaccharide) serotypes of *S. agalactiae* and 124 serotypes of *S. pyogenes*. Streptococci's cell wall is one of the most studied bacterial cell walls [27]. *S. agalactiae* or group *B. streptococci*, is a human pathogen that causes signifi-cant infections in human digestive and urogenital tracts like meningitis in babies and pregnant ladies; as of late, its pathogenic significance in older and immune-compromised patients, pneumonia and sepsis has been re-evaluated [28].

In addition, 60 Streptococcus species have been identified so far.

Few of these, like *S. agalactiae, S. equi, S. canis, S. pyogenes,* and *S. iniae* produce hemolytic elements and when developed on strong me-dia containing blood can be classified as beta-hemolytic [29]. *S. agalactiae* also contains a surface polysaccharide capsule that acts as a protective antigen and virulence factor. The genome of *S.agalactiae* is also prone to involve in recombination events [30].

2.2 Pathogenesis

2.2.1 Urinary Tract Infection(UTI)

During pregnancy the most well-known infection can occur is called urinary tract Infection. GBS causes cystitis, asymptomatic bacteriuria, and pyelonephritis when it enters the body through the vaginal canal. In pregnant and non-pregnant women, UTI caused by GBS is clinically indistinguishable from UTI and caused by other bacteria.

Up to 7% of pregnancies can be complicated by GBS bacteriuria, which is generally associated with a low bacterial count with 70% of cases being asymptomatic [31]. Asymptomatic bacteriuria acquired earlier in life might cause higher prevalence of symptomatic UTI in pregnancy. Around 10% of patients of acute pyelonephritis are caused by GBS, which occurs primarily during the second trimester.

Acute pyelonephritis presents a severe risk to the health of both the fetus and the mother. Septicemia, transient, low birth weight, early delivery, and fetal death are all possible prenatal problems. Maternal acute pyelonephritis is connected to transient renal failure, anemia, septicemia, thrombocytopenia, preeclampsia, pregnancy induced hypertension, and pulmonary insufficiency [30], [31]. Direct spinal cord compression, vascular compromise, and mechanical spine instability may increase the risk of a severe neurologic outcome in the spine surgical emergency known as spinal epidural abscess (SEA) [32], [33].

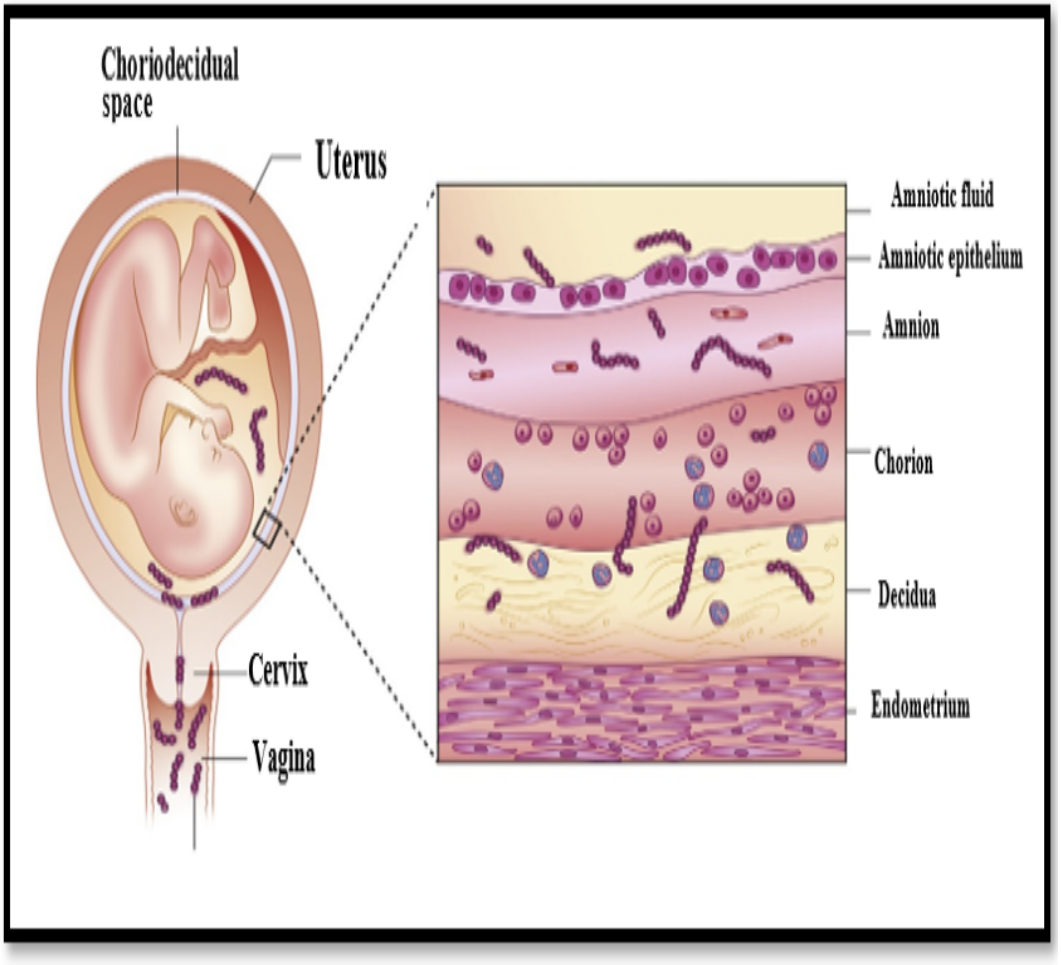

FIGURE 2.1: Ascending Group B *Streptococcus* (GBS) infection [34].

The presence of GBS vaginal colonization increases the risk of infection during pregnancy. *S.agalactiae* infection during pregnancy is induced by bacterial trafficking from the vagina which eventually leads to bacterial invasion of the placental membranes (chorion and amnion), the amniotic cavity and the baby. *S.agalactiae* generates number of of virulence factors that facilitate vaginal colonization, host cell adhesion and invasion as well as the activation or suppression of inflammatory responses [34].

2.2.2 Intra Amniotic Fluid

Intra-amniotic infection also known as IAI refers to infection layers followed by signs and symptoms in the mother or fetus, as well as a clinical state of placental disease. While the therapy based on clinical side effects, there are currently no

commonly accepted standards /388C maternal fever, fetal tachycardia (/160 beats/min), amniotic fluid that smells bad and uterine pain are often utilized criteria. IAI occurs as a result of *S.agalactiae*, which spreads from the vaginal area. The most well-known amniotic liquid segregates in women with IAI because of bacteria that are routinely present in the vaginal area. *S.agalactiae* was found in 15.4% of amniotic fluid patients, and it's one of the most often isolated species infect children delivered to IAI mothers [36].

2.2.3 Neonatal Sepsis

Based on the time of the infection and the presumed mechanism of transmission, neonatal sepsis is categorised into late onset disease. Neonatal sepsis stays a serious complication and remains a feared particularly extremely low birth weight (VLBW) preterm infant. Early onset sepsis (EOS) is characterized by initial seven days of life with some investigations limiting to EOS to disease happening in initial 72 hours that are brought about by maternal intrapartum transfer of invasive pahtogen. Late onset sepsis (LOS) is normally characterized as infection occurring after first week and is attributed to microorganisms obtained after birth [36]. Risk factors for neonatal sepsis contain destructiveness of infecting organism life form, maternal elements and neonatal host factors are also shown table 2.1.

TABLE 2.1: Risk factors for neonatal sepsis [36].

Sr No.	Source	Risk Factors
1	Early-onset neonatal sepsis (EOS)	Colonization of group B *streptococci* in the mother. UTI in number of pregnancies. It is also important in recovery after C- section and birth

TABLE 2.1: Risk factors for neonatal sepsis [36].

Sr No.	Source	Risk Factors
2	Late –onset neonatal sepsis (LOS)	Premature Rupture of maternal membrane. UTI in number of pregnancies. Prolonged use of antibiotics. Prematurity Decreased the passage of specific antibodies and mater immuno-globulins.

2.2.4 HIV Infection in Relation to Neonatal GBS Relation

Recto vaginal GBS carriage in women with HIV infection is a significant risk factor for GBS neonatal disease. HIV infection in women has been associated with a higher prevalence of vaginal candidiasis and sexually transmitted infections, and it is possible that HIV might have an impact on the microbial composition of the vagina, rectum, and colon18–22.

HIV infection of women is associated with recto vaginal GBS carriage, a major risk factor for GBS neonatal disease [37]. The aim of preventive techniques is to prevent and reduce the transmission of GBS to neonates by giving antibiotics to mothers who are GBS colonized during delivery and selectively giving antibiotics to neonates after delivery. No strategy can completely stop newborn GBS sepsis, even when it is strictly implemented.

The postpartum neonatal prophylaxis alone or in combination with intrapartum maternal prophylaxis reduces the risk of early-onset rates by 80% and 95% respectively [38]. The occurrence of *S. agalactiae* in the epaxial muscles of a wild bottlenose dolphin that had stranded itself was reported. Human *S. agalactiae* strains from fish, a dolphin and a frog revealed zoonotic and anthroponotic hazard

by causing severe illness in fish and affecting food security. Dangerous fish disease is another name for *S. agalactiae* [39].

2.2.5 Bones and Joint Infection

The involvement of bones of the foot are often connected with overlying ulcers and spreads from contiguous skin and soft tissue infection and vascular insufficiency. Vertebral osteomyelitis, normally in the lumbosacral region, is one more common type of GBS osteo articular infection; hematogenous seeding is the most probable system of disease, and vertebral destruction is insig-nificant. GBS septic joint inflammation is generally mono articular, most frequently including the knee, hip, or shoulder joints [40].

2.2.6 Postpartum Infections

Mastitis is a parenchymatous infection of the mammary organs, and puerperal mastitis caused by GBS can be either symptomatic or non-symptomatic [40], [41]. Intense puerperal mastitis affects 2.9 percent to 24 percent of pregnant women [42]. Another instance of *S.agalactiae* meningitis has been reported as an obstetric epidural an aesthetic complication [42]. Bacterial meningitis is usually caused by a hematogenous spread of bacteria. The pathogen then enters the subarachnoid space after crossing the blood-brain barrier. GBS shown that a high degree of bacteremia is a major determinant of meningeal invasion in an animal experimental model [43]. To employ the circulatory system to trigger meningitis, bacteria must avoid the host's defenses duplicate, and reach the threshold level bacteremia before attacking the meninges [43].

2.2.7 Presence of *Streptococcus agalactiae* in Animals

Staphylococcus aureus and *Streptococcus agalactiae* are the most common bacteria that cause mastitis. Milk is a basic food for people all over the world, but it

also causes a public health danger when polished off unpasteurized because it has a high bacterial concentration and is an ideal environment for bacterial growth. Included identification of bacteria in milk and noticed that long chain streptococci were mostly found in large numbers such as farmers can't be blamed for having saved the milk [43]. It is brought about by multi etiological microbes in that bacterial species are viewed as the main causative agent that leads to loss of milk production. The major microorganisms causing mastitis could be credited to *Staphylococcus aureus*, *Streptococcus agalactiae* among the bacterial species [44].

Camelids are a significant production resource in numerous regions of the world. In addition to low birthing rates, camel herds suffer from neonatal loss. Reproductive productivity is generally viewed as very low. Infectious life forms are responsible for the myriad of illnesses that directly or indirectly affect the regenerative progress of camelids. Uterine diseases are considered to be the most common reason for regenerative failure in camelid [45]. Pathogens that cause mastitis can be categorised as contagious and environmental [46], [46]. Contagious pathogens are adapted to survive inside their hosts and they primarily transfer from cow to cow during the milking process. Within a herd, contagious microorganisms have the potential to spread quickly and widely. Environmental infections, on the other hand, are able to live outside the host and are a normal part of the microflora of the cow's vicinity.

During the dry season or prior to a heifer's parturition and between milking is a period when environmental streptococci exposure occurs. Pathogen exposure is correlated with environmental pathogen abundance, which is regulated by factors including temperature and humidity. When the teat canal is opened after milking or when there has been injury, environmental pathogens enter the udder [47]. Between 30 and 50 percent of all cows normally suffer from bovine mastitis each year [48], [49]. Next to the financial losses due to lower milk yield and quality, the veterinary care, medication, and higher human costs, mastitis is a crucial issue of animal welfare and the primary factor in dairy cow culling. Symptoms of mastitis infected cows include swelling, heat and pain in the udder, abnormal milk appearance , higher body temperatures, lethargy, and anorexia [50]. According to the

degree of the inflammation, there are three types of bovine mastitis: clinical, subclinical, and chronic. Subclinical mastitis does not have the visible abnormalities as-sociated with clinical mastitis in the cow or milk. Only the milk production and somatic cell count are changed in this case. Subclinical mastitis is estimated to be 15–40 times occurs more frequently than clinical mastitis [51]. Since subclinical mastitis occurs more frequently and has the ability to decrease milk production while going unnoticed, it is economically more significant. Streptococcal mastitis normally lasts 12 days but it can last up to 300 days in chronic situations [52].

2.3 Serotypes of *Streptococcus agalactiae*

All strains of GBS isolated from people and can be classified based on CPS structure and serology different encapsulated GBS serotypes have been identified: Ia, Ib, II, III, IV, V, VI, VII, VIII, and IX. Ia, Ib, II, and III were similarly prevalent in normal vaginal carriage and beginning of onset sepsis (i.e., that developing at under 7 days old enough). Even so, type V is presently an important reason for *S.agalactiae* infection [53]. So, the strains of types VI and VIII have be-come dominant among Japanese women [54].

A pan- European governed in the range be-tween 2008 and 2010 showed that GBS serotypes Ia, III, and V together represented 88%, 96%, and 67% of strains isolated from youngsters with early onset illness, children with late onset illness, and vaginal rectal swabs of colonized pregnant women who delivered good health babies, respectively types Ia and III and showed that type V is an important serotype responsible for neonatal *S.agalactiae* infection in many regions of the world [55], [56].

2.4 Biofilm Formation

GBS's capacity to colonize and persist in different host niches is determined by its ability to adhere to cells and tissues. Biofilms are known as the arrangement

of sessile communities. Biofilms formed by bacteria are well-known virulence factors that contribute to illness persistence and chronicity. Bacteria in the human environment are commonly protected from the immune system by forming sessile colonies inside an extracellular polysaccharide matrix known as the biofilm.The type II a pili and bacterial capsule have been shown to play an important role in the production of biofilm in GBS. The host's ecological characteristics are important variables in the formation of bacterial biofilms. pH that is acidic appears to be ideal for GBS colonization as a regular vaginal inhabitant. GBS adhesion to vaginal epithelial cells was shown to be higher at low pH than at neutral pH in early studies. GBS isolates developed more biofilms in acidic pH environments than in neutral pH environments, and the strongest biofilm-forming GBS isolates were assigned to the ST-17 grouping type. When GBS was compared to strains, neonatal strains had more solid biofilm producers than colonizing strains [57]. GBS isolated from asymptomatic transporters has been shown to be a powerful biofilm producer [58].

Furthermore, human plasma was known to enhance the growth of GBS biofilms. Biofilm production protects bacteria from immune system detection and allows for long-term bacterial survival. Plasma and low pH are ecologically important factors for GBS increase bacterial biofilm formation by affecting the expression of bacterial surface-related structures such as capsule and pili which are both included [59].

2.4.1 Mode of Action

Group B streptococcus is a deadly pathogen that inhabits the gastrointestinal tract or vagina in healthy people. *S.agalactiae* as a pathobiont has the ability to expand from asymptomatic mucosal carriage to a bacterial infection that causes severe illnesses in any situation. The ability of GBS is to adhere with cells and extracellular matrix (ECM) is critical for colonization, persistence, motility, and targeting host barrier. GBS is an opportunistic pathogen that lives in the physiologic flora of the vagina and intestines [60]. Group B streptococcal fibronectin restricting protein A-(SfbA), streptococcal C5a peptidase (ScpB), fibrinogen binding proteins-(Fbs),

laminin-binding proteins (Lmb), and GBS immunogenic bacterial strains adhesion are functionally characterized adhesions that mediate GBS adherence and attack inside the host (Bi-bA).Biofilm formation, *S.agalactiae* colonization, persistence and central nervous system (CNS) attacks are considered to be boosted by surface-protruding structures involving a range of genes, such as pili [61]. Srr1

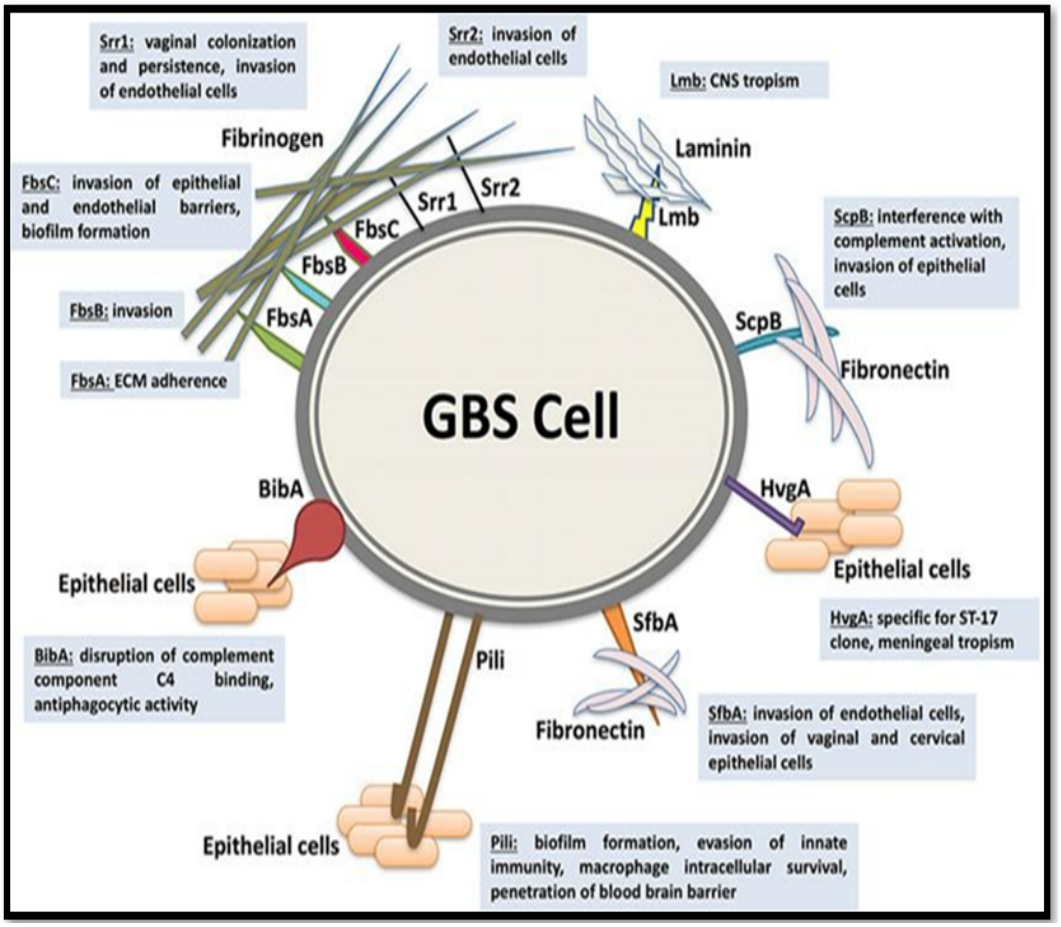

FIGURE 2.2: The interaction of S. agalactiae (GBS) with host cells is mediated through main adhesions. Ca-pacity to attach with cells and extracellular matrix proteins (ECM), which is a crucial stage in cell barrier breakdown, is a condition for GBS persistence, colonization, motility or penetration of host barriers [61].

and Srr2, the laminin binding protein (Lmb), the streptococcal C5a peptidase from group B (ScpB), the streptococcal fibronectin binding protein A (SfbA) and the GBS immunogenic bacterial adhesion BibA is the most well studied surface protein promoting the adherence of GBS.Furthermore, pili and other surface protruding features are known as crucial adhesions in facilitating GBS colonization, persistence, biofilm formation, and invasion of the central nervous system.While FbsB has been found to be essential for human cell penetration, FbsA was largely

proven to increase adherence. According to reports, Srr1 and Srr2 have a role in the invasion of micro vascular endothelial cells. Srr1 has also been demonstrated to promote vaginal colonization and persistence. Fbs has recently been shown to promote the penetration of epithelial and endothelial barriers and the formation of biofilms [61]. The LMB adhesion appears to play an important role in the bacterial tropism of the central nervous system (CNS). ScpB cleaves the neutrophil chemo attractant C5a to interrupt complement activation.

Additionally, it contributes to the invasion of human epithelial cells. Invasion of human brain micro vascular endothelial cells is facilitated by the SfbA adhesion. Cervical epithelial cells and SfbA can help in GBS colonization and niche formation in the vagina because it promotes GBS penetration to vaginal and epithelial cells. By attaching to the C4-binding protein and inhibiting the traditional complement pathway, BibA has been shown to allow GBS remain in bloodstream. It also shows anti-phagocytic effect against opsonophagocytic destruction of human neutrophils.The hyper virulent clone ST-17 is specific to HvgA. Promoting meningeal tropism in new borns was suggested. Pili have been found to have a major role in the invasion, translocation, biofilm formation, and colonization of epithelial cells in GBS. Additionally, it was shown that PI-1 pili were crucial in the evasion of the innate immune system. However, it has been shown that the PI-2b protein increases macrophage intracellular survival. Pilus 2b was also identified to be important for blood brain barrier penetration and infection [62].

2.4.2 Transmission of Disease

Although it is unknown if some *S. agalactiae* strains may transfer from a stage of colonization to a stage of disease transmission ,there are number of risk factors have been discovered that may affect this process [63], [64]. Host factors include like low levels of type specific maternal antibodies during birth, a weakened immune system or poor immunological response, race or ethnicity, etc [65], [66]. In addition, other behavioural factors including sexual behavior in association with various bacterial characteristics including pathogenic capability e.g., adherence

capabilities, polysaccharide type, host evasion mechanisms and inoculum density contribute to *S.agalactiae* colonization and subsequent disease [67], [68], [69]. Pathogenesis such as the majority of infectious diseases is a complicated process whereas colonization leads to invasion and clinical symptoms in many individuals [70], [71].

2.5 Role of *Streptococcus agalactiae* Srr1 Protein

Streptococcus agalactiae Srr1 adhere to number of host cell types. *Streptococcus, Staphylococcus,* and *Lactobacillus* are Gram-positive genera that produce a group of huge, glycosylated serine-rich repeat proteins (SRRPs) [72]. SRRPs enhance adhesion to sialic acid, fibrinogen, keratin, and even unknown compounds on other bacteria in a strain or species dependent way.The adaptability of binding partners implies that serine rich repeated proteins(SRRPs) promote colonization of numerous and oral streptococci may remain in the mouth and produce infectious vegetations on injured heart valves in different habitats [73]. Fibrinogen and keratin are bound by the *S. agalactiae* SRRP (Serine rich repeat proteins), Srr1, which enables the organism to colonize various body sites [74].

The capacity of Srr1 to bind fibrinogen on the cell surface of human platelets and brain endothelial cells is a direct result of its ability, whereas adhesion to the vaginal and cervical epithelium is mediated by binding to both fibrinogen and keratin [74], [75]. In a mouse vaginal colonization model, *S. agalactiae* srr1 mutants are carried for shorter periods of time and at lower densities indicating that Srr1 encourages persistent vaginal colonization [75]. A mutation in the srr1 gene impairs *S. agalactiae* can bind to brain endothelial cells, laryngeal and lung epithelial cells, gut epithelial cells, vaginal and cervical cells, and platelets [75].

In the figure below 2.3 model of *S. agalactiae* and *S. pneumoniae*, RRP illustrates the role in colonization and illness and group of bacteria's which invades.

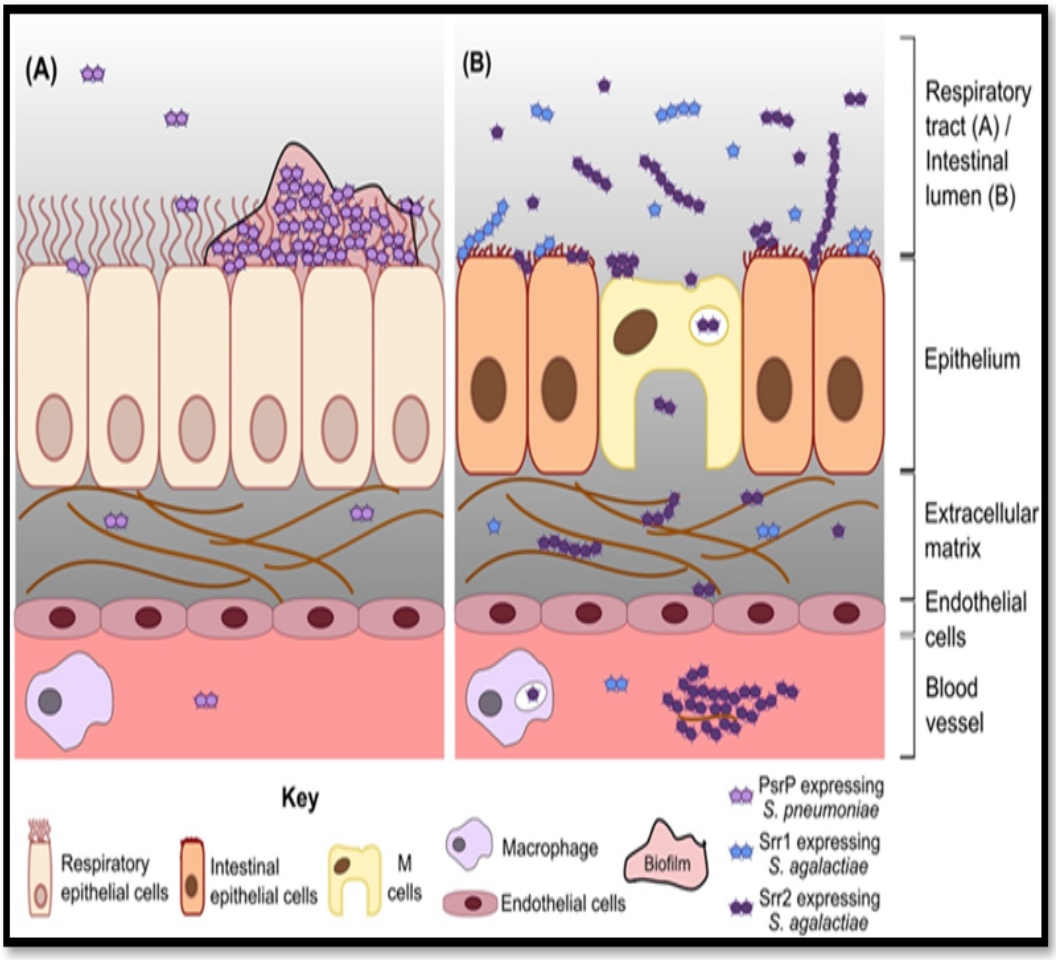

FIGURE 2.3: Model of *S. agalactiae* and *S. pneumoniae* RRP illustrates the role in coloniza-tion and illness.(A) A subgroup of bacteria invade deeper tissue by *Pneumococcal* PsrP-mediated respiratory system, biofilm formation and persistent colonization.(B) During coloniza-tion, *S. agalactiae*(blue and purple diplococci)that express Srr1 and Srr2 adhere to intes-tinal epithelial cells, respectively. In comparison to Srr1, Srr2 promotes invasion by transcytosis through M cells and mediating bacterial adhesion. Srr2 increases the survival of phagocytosed bacteria and forms large bacterial-plasma aggregates in the blood circulation to facilitate persis-tence. Immune cell migration may be used by internalised *S. agalactiae* to spread and cause other illnesses, such as meningitis [76].

2.6 Risk Factors

GBS carriers who are not pregnant is to identify risk factors which are associated with colonization. Age less than 20 years, the presence of an intrauterine device, the length of time since the last menstrual period, the use of tampons, the consumption of milk, infrequent hand washing, the use of yeast medications and African American ethnicity all appear to be common behaviours associated

with GBS colonization [77], [63]. The sexual habits are linked with colonization including fewer days since last sex ,increased frequency of sexual and having multiple partners. GBS virulence is complex and multifactorial. Several virulence determinants are involved in the adhesion to and invasion of host cells, as well as in the immune system evasion. Surface components, including a polysaccharide capsule and proteins, such as C(alpha), C(beta), Rib and the laminin binding protein (LMB), and a number enzymes (like the C5a peptidase) and toxins/cytolysins, are produced and have been associated with GBS virulence [78]. Numerous virulence factors expressed by GBS promote the colonization of the vagina, the adhesion and invasion of host cells and the activation or repression of inflammatory responses.

TABLE 2.2: Virulence factors of GBS

Virulence Factor	**Host Target**	**Function**	**Phenotype**	**References**
HylB	Hydraulic acid	Blocks TLR2/4 signaling	Mouse vaginal colonization and increase infection.	[79]
CovR/S	pH	Controls the expression of hemolytic pigment	Colonization in the mouse	[80], [81]
BsaB/FbsC	Fibrinogen and laminin	Adherence to vaginal epithelial cells.	Immortalized human cell line. Mouse vaginal colonization.	[82]
FspS/R	Fructose 6-phosphate	Vaginal persistence	Immortalized human cell line. Cytolysis	[83], [84]

TABLE 2.2: Virulence factors of GBS

Virulence Factor	**Host Target**	**Function**	**Phenotype**	**References**
Hemolytic Pigment	Amnion Epithelial cells Neutrophils	Cytolysis Mast cell degranulation Proptosis	Colonization of mouse vagina, human chorio amnion and placenta. Immortalized human cell line.	[85], [86], [87], [88], [89]
Srr	Fibrinogen	Adherence to vaginal	Mouse vaginal colonization.	[90], [91], [92]

2.6.1 Extracellular Virulence Factors of GBS

A large number of GBS virulence factors are essential elements of the bacterial surface. The GBS surface polysaccharide capsule is the most important of them which through its terminal sialic acid motif to prevent complement deposition and opsonophagocytic [93]. Other GBS extracellular virulence factors may be associated with the bacterial surface but can be eliminated in an active state by proteolysis or binding to host cell components. Some GBS extracellular virulence factors are secreted directly into the medium [94].

TABLE 2.3: (a)Extracellular virulence factors of GBS and their contribution to cause diseases.

Virulence Factor	**Genetic Basis**	**Biochemical Nature**
Beta-hemolysin	cylE	CylE protein (87 kD)
Hyaluronate lyase	scpB	HylB enzyme (110 kD)
C5a peptidase /invasion	pepB	ScpB protein (110 kD)

TABLE 2.3: (a)Extracellular virulence factors of GBS and their contribution to cause diseases.

Virulence Factor	**Genetic Basis**	**Biochemical Nature**
CAMP factor	Unknown	PepB peptidase (24 kD)
Oligo-peptidase	Unknown	PepB peptidase (70 kD)

TABLE 2.4: (b)Extracellular virulence factors of GBS and their contribution to cause diseases.

Virulence Factor	**Cellular Actions**	**Proposed contribution(s) to disease**	**Virulence Role**
Beta-hemolysin	forms pores in cell membranes induces apoptosis promotes cellular invasion triggers.	Penetration of epithelial barriers induction of sepsis syndrome, direct tissue injury.	Yes (Rabbits, Rats)
Hyaluronate lyase	cleaves chondroitin sulfate and hyaluronan	spread through host tissues.	No
C5a peptidase invasion	cleaves human C5a binds fibronectin.	Reduce opsonophagocytic host cell attachment and inhibit PMN recruitment.	Yes (Mice)
CAMP factor	CAMP (co-hemolysin) Binds to Fc portion IgG, IgM.	Impairment of antibody function and reaction tissue injury.	Yes (Rabbits)

TABLE 2.4: (b)Extracellular virulence factors of GBS and their contribution to cause diseases.

Virulence Factor	**Cellular Actions**	**Proposed contribution(s) to disease**	**Virulence Role**
Oligo-peptidase	cleaves bioactive peptides and perhaps collagen.	promotion of tissue invasion.	No

2.7 Available Drugs

For non-allergic patients having *S. pyogenes* and *S. agalactiae* infections, penicillin or one of its derivatives (e.g., amoxicillin and ampicillin) is the recommended antitoxin treatment [94]. Azithromycin and clarithromycin are recommended for hypersensitive persons, and azithromycin is recommended more commonly than penicillin. A combination of penicillin and clindamycin is suggested for severe *S. pyogenes* infections including toxic shock and necrotizing fasciitis [27].

S. pyogenes and *S. agalactiae* are not penicillin resistant, but they do develop resistance to clindamycin, antibiotics, vancomycin, and macrolides (e.g., erythromycin, azithromycin, and clarithromycin) over time [94].

2.8 Pan Genome Analysis

The advancements in next-generation sequencing technologies have transformed our knowledge of a cellular localization, functional variation at the basic genome level and more microbial hereditary repertoire [95], [96].

Also, the entire genome sequencing of bacterial pathogens helps in prioritizing the insert of researchers towards the pathogenicity by precisely estimating the hereditary varieties among the pathogenic groups [97]. The hereditary varieties among

the number of genomes at the bench high level are deduced by the time consuming and cost effective recognizable proof of the variable sites that are described as the SNPs. These can be done by the entire genome multilocus sequence typing (MLST) approach [98].

A comparative genomic approach is utilized to defeat the potential limitations that are con-nected with these reference based approaches. This approach depends on the sequence similarly search analysis [99]. The comparative microbial genomics system that depends on the arrangement closeness helps in distinguishing the essential hereditary content that is shared by all the pathogenic isolates with the assistance of statistical analysis. It also helps in observing the qualities that encode the novel abilities and virulence as a variable genome [100]. Both variables and core genome content of a life form is signified by the pan-genome [101]. The entire genetic repertoire of the isolates is addressed by the supra-genome. Phyllogenomic examination with help of the pan-genome helps in the determination of the genomic items in a group, for example, the core, variable, and pan genome of an organism along with the way of life can be allopatric or sympatric [102].

2.9 Plant Derived Antimicrobial Compounds

PubChem (www.pubchem.ncbi.nlm.nih.gov) is the largest database of widely accessible chemical information in the world. So, the chemical compounds that could be used as ligands were selected directly from PubChem database.

TABLE 2.5: Showing the plant-derived antimicrobial compounds for docking in drug designing.

Hyperenone A	**Hypecalin B**
Rutin	Sperimidine
Coumarin	Stigmasterol
Querecetin	Berberine
Kaempferol	Scopoletin

2.9.1 Molecular Docking in Drug Discovery

Molecular Docking is technique used to estimate the strength of a bond between a ligand and a target protein through a special scoring function and to determine the correct structure of the ligand within the target binding site. The 3D structure of the target proteins and the ligands is taken as the input for docking. It represents a frequently used approach in structure based drug design since it requires a 3D structure of a target protein. It can be used to determine the correct structure of the ligand within the target binding site, and to estimate the strength of the binding between the ligand and the target proteins through a specific scoring function [103]. It also helps in the recognition of new small molecular compounds, revealing the essential properties, such as high interaction between binding with target protein having reasonable absorption, distribution, metabolism and excretion, which help in the selection of lead compound for the target [104]. So, the docking process includes following compounds which are discussed below:

1. The docking process requires a protein 3D structure which is downloaded from pro-tein data bank (PDB).
2. Minimum size of molecules or compounds or virtual compounds that contain a database is required.
3. A computational framework is also needed to perform the docking and find the scoring process.

Protein and ligand docking is one of the key areas of molecular docking, which is obtain high popularity and appreciation due to its role in structure based drug designing [105].

Chapter 3

Research Methodology

This methodology includes selection of problem, genome selection, identification of core genome, non-host homologues protein identification, target identification, catalytic pocket detection and molecular docking.

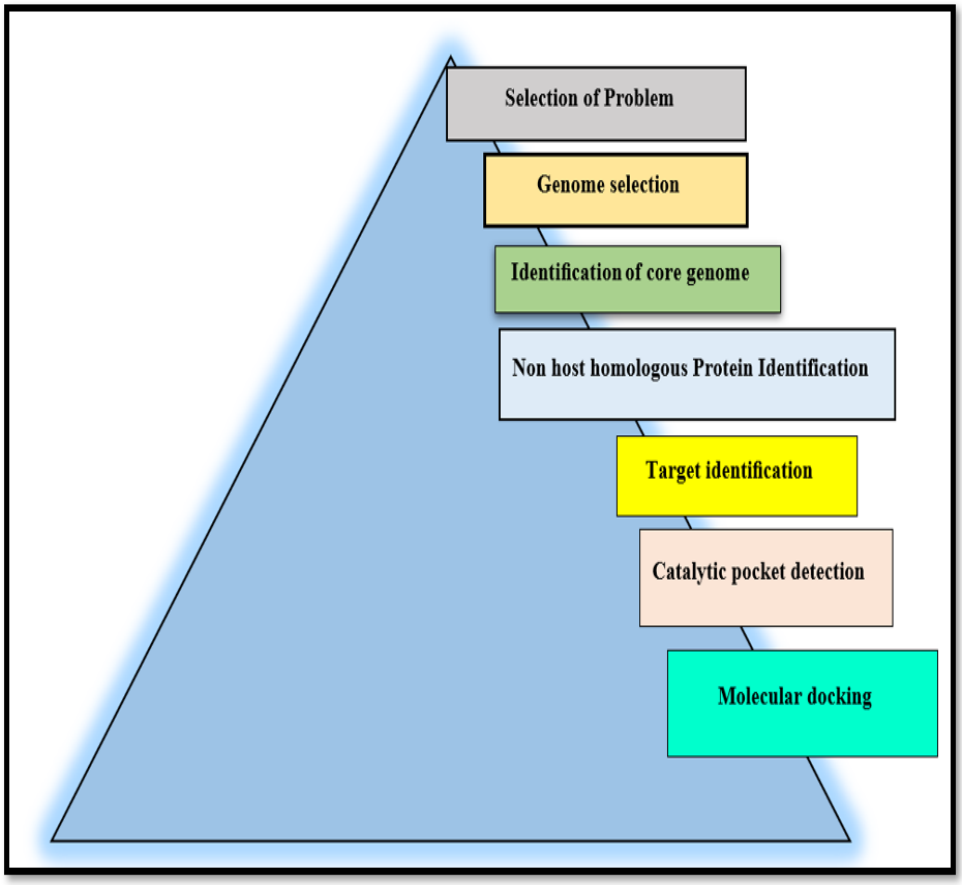

FIGURE 3.1: The flowchart of research methodology.

Present research was based on several computational approaches which are described below under headings sequentially:

3.1 Selection of Problem

In case of bacterial infections antibiotic treatment is recommended but it have seen now that certain bacteria become resistant to those antibiotics because of their massive use [106].

3.2 Genome Selection

The selection of strains was based on a review of the literature, which was obtained from NCBI and PubMed.127 strains of *Streptococcus agalactiae* were used for this study. The selec-tion was based on complete genome strains data for the accuracy in the results. The whole genome of *S.agalactiae* was searched and obtained from the NCBI (https://www.ncbi.nlm.nih.gov) [107].

3.3 Retrieval of Bacterial Sequences

The NCBI database is used to get genomic infor-mation (https://www.ncbi.nlm.nih.gov/genome). To be used for further analysis, data on both nucleotide and protein sequences are retrieved. A total of 127 strains are selected to be used in this study. To see the variations between strains, other variables like as genome sizes, G+C content, average gene count, coding DNA sequences (CDS).

3.4 Selection of Reference Strain

The availability of almost all complete genomes for *Streptococcus agalactiae* is useful in identifying the essential, accessory and unique genetic characteristics for

each strain. The core genome was identified using Edgar3.0 (https://edgar.computational.bio.unigies-sen.de/cgi-bin/edgar.cgi) .

Core genes were extracted from all 127 strains using Streptococcus agalactiae 2603V R AE009948 as a reference strain according to its release date. The genes found in all strains were chosen for further study. One strain, *S. agalactiae*, was chosen as the reference strain using Edgar 3.0. 2603V R AE009948 was a strain that was compared to all other strains [108].

3.5 Non-Host Homologous Protein Identification

Using the Blast P (https://blast.ncbi.nlm.nih.gov/Blast.cgiPAGE=Proteins), non-host ho-mologous genes or proteins were found after the retrieval of core genome. These non-homologous proteins were added to the DEG Database (http://tubic.tju.edu.cn/deg/) in order to identify essential proteins [108]. And show important proteins that were retrieved from the DEG Database. Applying criteria, i.e., identity e value=0.001, from the list of proteins 3 cytoplasmic proteins were chosen out for a drug targeting study based on cellular localization [109].

3.6 Target Identification

Multiple factors, including molecular weight pathway analysis and others were considered while determining possible therapeutics [110].

3.6.1 Molecular Weight

Protocols was used to identify essential genes such as molecular weight which is identified by Prot Param (https://web.expasy.org/protparam/) [111].

It calculates the molecular weight in g/mol of every protein when it was conserved in kiloDa that where units of weight.

3.6.2 Pathway Analysis

Kyoto Encyclopedia of Genes and genome (KEGG) (https://www.genome.jp/kegg/sess=ebfe2ad23e021e38540f798c803dd061) was used for the pathway analysis of an organism [112]. The TCA cycle, pentose phosphate, pentose and glucuronate interconversions, fructose and mannose metabolism, galactose metabolism, ascorbate and aldarate metabolism, and elonga-tion of fatty acid biosynthesis are all controlled by these three proteins.

3.6.3 Subcellular Localization

The location of cytoplasmic and membranous proteins were identified by using prediction system named CELLO (http://cello.life.nctu.edu.tw/) [113].

3.7 Catalytic Pocket Detection

The catalytic pocket of the essential proteins to associated to a certain drug score was identified using DoGSiteScorer (https://proteins.plus/) [114]. Dog Site Scorer is an automated tool for measuring the drug-ability of protein cavities and detecting pockets. PDB id or Protein's 3D model can be provided for analysis. From 0 to 1, the drug ability score can be given. A drug ability score of greater than 0.60 is taken into consideration, but it is favored if it is higher than 0.80 [114].

3.8 Molecular Docking

The list of ligands for antibacterial compounds was taken from the literature review and entered into the MOE software for molecular docking [115], [116]. By keeping all default parameters both the minimized and prepared ligands and proteins were subjected to docking. All of the ligand molecules docked into the target proteins'

binding sites. All of the ligands showed acceptable results. But we selected only one ligand that was effective against each ligand protein.

3.9 Antibiotic Resistance

One of the characteristics that a probiotic should ideally possess is antibiotic resistance. Probi-otic strains with intrinsic antibiotic resistance have the ability to increase their numbers in the gut after using antibiotics to treat infections. The worldwide problem, on the other hand, is bacterial species' resistance to antibiotics. Probiotic bacteria are screened for antibiotic resistance genes to ensure their compatibility to be used as probiotics so that they cannot transfer these resistance genes to other bacteria through horizonal gene transfer mechanisms. Comprehensive Antibiotic Resistance Database is a database used to check for signs of antibiotic resistance (https://card.mcmaster.ca/analyze/rgi). The database was used to identify that either particular strain harbors gene for resistance against various drugs as well compare and evaluate the differences [117].

Chapter 4

Results and Discussions

4.1 Sequence Retrieval of Protein

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

4.2 Genomic Data Collection

Streptococcus agalactiae is an opportunistic organism. For this study 127 strains of *Streptococcus agalactiae* were used shown in table (4.10).

4.3 Selection of Strains

For this project first step was to select an inclusion and exclusion criteria for selection of bacteri-al strains, all the strains with complete genomic sequence available along with a known source of isolation were selected. 127 bacterial strains were selected. Selected strains were verified using literature analysis and their genomic properties were analyzed. The whole genome sequences of all 127 strains were downloaded from NCBI database [117].

4.4 Selection of Core Genome

A nucleotide sequence assembly used as representative example of genes present in a particular bacterial species is referred reference genome. These reference genomes act as guide for annotation and as-sembly of new genomes.

4.5 Primary Sequence Retrieval

FASTA sequence of selected target proteins was retrieved through UniProt http://www.uniprot.org/. These proteins were selected on basis of their pathogenicity and virulence causing factors. The FASTA sequence of Thioredoxin family protein, DNA binding response regulator and UDP-N-D-glutamyl-2,6-D-alanyl ligase were downloaded from uniport under accession number, AAM99084.1 , AAN00807.1 and AAM99655.1

TABLE 4.1: Primary sequence of the targeted proteins.

Protein	**FASTA Sequence**
Thioredoxin family protein	>AAM99084.1thioredoxin family protein[Streptococcus agalactiae 2603V/R]MILPESYEEIAAYIDSTKKVVFF FTADWCPDCQFIYPVMPSIEKDFSDFVFVRVNRD DYIELQQWNIFGIPSFVVVENGQELGRLVNKNRK TKAEITKFLAEINYK
DNA binding response regulator	> AAN00807.1DNA-binding response regulator[Streptococcus agalactiae 2603V/R] MYRLLIVEDEHLIRKWLRY AIDYQSLNILVVGEAKDGKEGAQLIQEEQPDIVLSDI NMPIMTAFDMFEATKGQSYAKIILSGYADFPNAQSAI HYGVLEFLTKPLEKQALIDCLKTIMARIEEHKEKHLQ EHTELYLPLPQANDQVPEVIKDMLAWIHSHFHGKIVI SQLAHDLGYSESYLYTVTKKHLHITLSDYINQYRINQ AIQLMFREPDLMVYQIAEAVGIYDYRYFDRVFKKYL GQTVKAFKEEHIFKQMD

TABLE 4.1: Primary sequence of the targeted proteins.

Protein	**FASTA Sequence**
UDP-N-D-glutamyl-2, 6-D-alanyl ligase	>AAM99655.1UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate–D-alanyl-D-alanyliase[Streptococcusagalactiae2603V/R]M KLSLHEVAKVVGAKNQVSEFEDVPLGNIE FDSRNISEGDLFLPLKGARDGHEFIEMAFD NGAIATISEKEIEGHPYLLVSDALKAFQVLA QYYIEKMNVDVIAVTGSNGKTTTKDMIAAI LSTTYKTYKTQGNYNNEIGLPYTVLHMPED TEKIILEMGQDHLGDIHVLSEIAKPRIAVVTL IGEAHLEFFGSREKIAEGKMQITDGMSSDGI LIAPGDPIIDPYLPANQMTIRFGHDQELQVTE LKEEKHSLTFKTNALEHQLRIPVPGKYNATN AMVAAYVGKLLAVAEEDIVDALENLQLTRN RTEWKKSANGADILSDVYNANPTAMRLILE TFSAIPNNDGGKKIALLADMKELGEQSVDL HNQMIMSIRPDSIDTLICYGQDIEGLAQLAS QMFPIGKVYFFKKNQEVDQFDQLLAKVKD TLKEKDQILLKGSNSMNLSKIVDIL EVG

4.6 Identification of Non Host Homologous Protein

Through the use of the Blast P https://blast.ncbi.nlm.nih.gov/Blast.cgi\PAGE=Proteins tool,three non- homologous genes and proteins were identified as

shown in table 4.2.

TABLE 4.2: Shows essential proteins which were retrieve after DEG Database.

Sr.no	Protein
1	Thioredoxin family protein
2	DNA binding response regulator
3	UDP-N-D-glutamyl-2,6-D-alanyl ligase

4.7 Physicochemical Characterization of Target Proteins

ProtParam is a tool of Expasy which is used online for the prediction of different parameters including both physical and chemical properties of selected proteins. These several parameters calculate and estimate the following through ProtParam: molecular weight, composition of amino acid, theoretical value of protein index, atomic composition of protein, extinction coefficient, estimated half life of protein instability, aliphatic index and grand average of hydropathicity which was abbreviated as GRAVY. The physicochemical properties of the Thioredoxin family protein, DNA binding response regulator and UDP-N-D-glutamyl-2,6-D-alanyl ligase were shown in Table (4.3) respectively.

TABLE 4.3: Physicochemical Properties of Target Proteins

Target proteins	**Thioredoxin Family Protein**	**DNA response regulator**	**UDP-N-acetyl D-glutamyl -D-alanyl ligase**
MW	12573.44	30117.76	50399.66

TABLE 4.3: Physicochemical Properties of Target Proteins

Target proteins	**Thioredoxin Family Protein**	**DNA response regulator**	**UDP-N-acetyl D-glutamyl -D-alanyl ligase**
PI	4.78	5.73	4.92
NR	16	36	67
PR	12	27	44
Ext Co.1	18575	36330	24870
Ext CO.2	18450	36330	24870
Instability index	50.64	38.61	29.56
Aliphatic undex	87.38	100.58	98.84
GRAVY	-0.102	-0.235	-0.187

MW stands for molecular weight, for theoretical isoelectric point at which protein is neutral, without any charge), NR for total number of negatively charged residues (Asp + Glu), PR for total number of positively charged residues (Arg +Lys), Ext.Co1 for extinction coefficients when assuming all pairs of Cyst residues form cystines, Ext. Co2 for extinction coefficients when assuming all Cyst residues are reduced and GRAVY for grand average of hydropathicity. All these parameters which were selected for this research work were taken according to previous research work [118].

The calculated PI greater than 7 represents the basic nature of the protein while less than 7 shows acidic nature of protein. PI value of targeted proteins (Thioredoxin family protein , DNA response regulator and UDP-N-D-glutamyl-2,6-D-alanyl ligase) have shown less PI value so they are acidic basic in nature.

Extinction coefficient represents light absorption. Instability index if less than 40 show stability of the protein while greater than 40 indicates the instability of protein [119]. Instability index of DNA response regulator and UDP-N-D-glutamyl-2,6-D-alanyl ligase were less than 40. The aliphatic index represents the aliphatic content of a protein. The high value of the aliphatic index indicates the thermo stability of the protein. DNA binding response regulator has highest aliphatic index value among others. Molecular weight contains both positive and negative charged residues of protein. Low GRAVY shows better interaction with water molecules. Thioredoxin family protein showed lower GRAVY values among others.

4.8 Protein Structure Predictions

3D Structures of targeted proteins Thioredoxin family protein ,DNA response regulator and UDP-N-D-glutamyl-2,6-D-alanyl ligase were downloaded from RCSB PDB in PDB format. Protein Data Bank is a three dimensional database of complex molecules of living organisms like proteins and nucleic ac-ids [120].

Alphafold could be used if some structures were missing on PDB database. Alphafold `https://alphafold.com/` is also a protein structure database used for 3D structure prediction of pro-teins [121].

The 3D structures of Thioredoxin family protein ,DNA response regulator and UDP-N-D-glutamyl-2,6-D-alanyl ligase proteins were taken in PDB format under AAM99084.1, AAN00807.1 and AAM99655.1 accession numbers respectively. The protein structures were prepared in PyMOL by removing water molecules and extra ligands if existed. After the removal of ligands and other atoms the missing

polar hydrogens were added. This step was performed to get the stable conformation by preventing overlaps and saved the modified file in PDB.

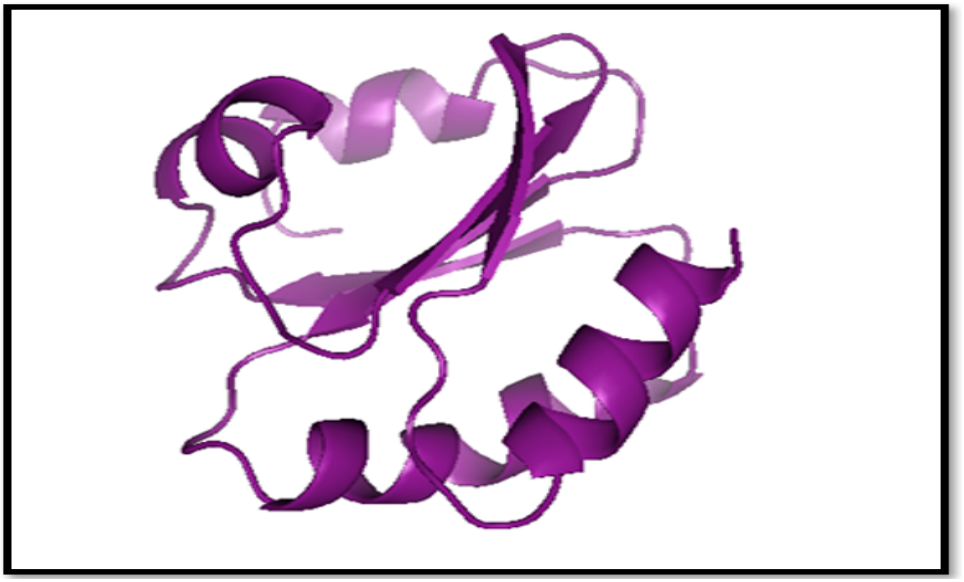

FIGURE 4.1: 3D structure of Thioredoxin family protein.

Above figure 4.1 represented the three dimensional structure of Thioredoxin family Protein. It is involved in redox signalling.

FIGURE 4.2: 3D structure of DNA binding response regulator

Above figure 4.2 represented the three-dimensional structure of DNA binding response regulator. It plays a key role in stress response and exerts either positive or negative regulation of genes.

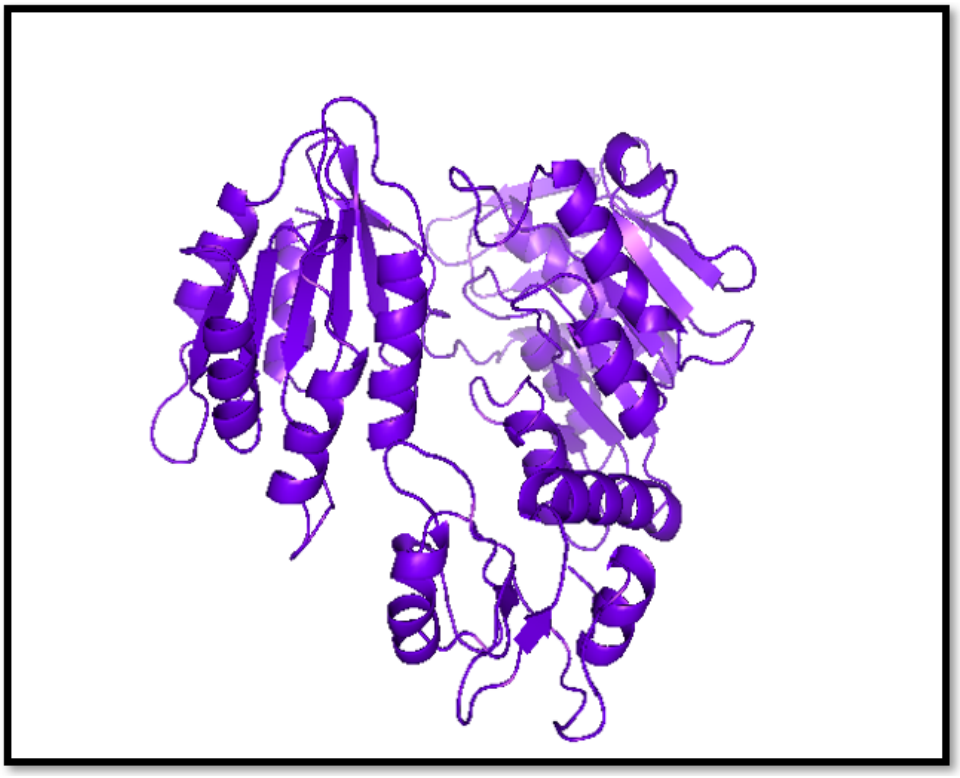

FIGURE 4.3: 3D structure of UDP-N-D-glutamyl-2,6-D-alanyl ligase (UMAL)

Above figure 4.3 represented the three-dimensional structure of UMAL.It catalyses the addition of mesodiaminopimelic acid in the biosynthesis of bacterial cell wall peptidoglycan.

4.9 Drug Targeting Analysis

Our methodology allowed us to identify three target proteins from the 127 S. agalactiae strains that can be used as drug design for all of these strains. RAMPAGE, which stands for RNA Annotation and Mapping of Promoters for the Analysis of Gene Expression, was used to validate the three dimensional (3D) structures of all of these proteins. ERRAT is a non-bonded atomic interaction overall quality factor with higherr scores signifying better quality. For Rampage,

a score more than 80 is considered to be favourable, but for ERRAT, a score greater than 37% is considered as favourable.

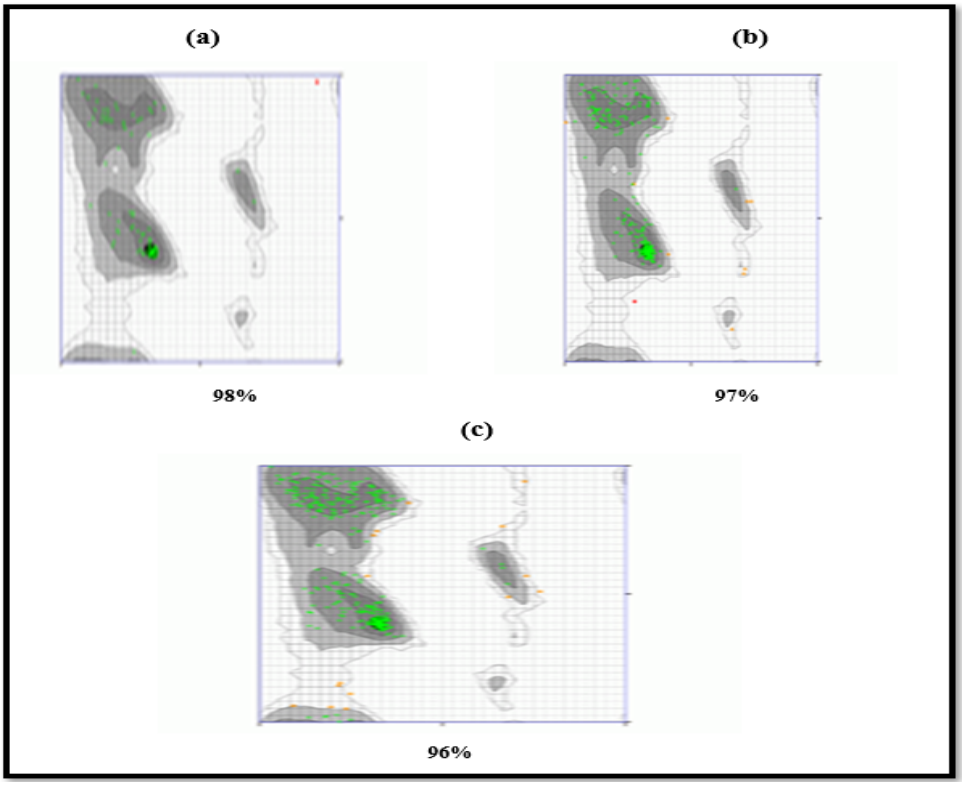

FIGURE 4.4: Shows the Ramachandran plot.

* Highly Preferred observations shown as GREEN Crosses.

TABLE 4.4: Validation scores of the selected proteins against Rampage and ERRAT value.

Protein name	**Rampage***	**ERRAT****
Thioredoxin family protein	98%	86.9565
DNA-binding response regulator	97%	87.395
UDP-N-acetyl muramoylalanyl-	96%	95. 9091

TABLE 4.4: Validation scores of the selected proteins against Rampage and ERRAT value.

Protein name	**Rampage***	**ERRAT****
D-glutamyl-2, 6-diaminopimelate--D-alanyl-D-alanyl ligase		

⋆Rampage: percentage above greater than or equal to 80 is considered to be a high quality model for drug targeting studies.

⋆⋆ERRAT:Range greater than or equal to 37% is acceptable for a high quality model.

4.10 Cellular Localization

CELLO tool was used for protein localization. It was online tool which was best for find localization.

TABLE 4.5: Protein localization for drug designing.

Sr.no	**Protein**	**Localization**
1	Thioredoxin family protein	Cytoplasmic
2	DNA-binding response regulator	Cytoplasmic

TABLE 4.5: Protein localization for drug designing.

Sr.no	Protein	Localization
3.	UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate--D-alanyl-D-alanyl ligase	Cytoplasmic

4.11 Catalytic Pocket Detection

Catalytic pocket detection was performed by the DogSiteScorer. For each selected protein, pockets with the highest drug score were chosen.11 anti bacterial com pounds were taken from the literature review and prepared by ChemBioDraw Ultra 11.

TABLE 4.6: Catalytic pocket detection for docking process

Thioredoxin Family protein residues	**DNA-binding response regulator residues**	**UDP-N-acetyl D-glutamyl D-alanyl ligase**
LYS_19_A	TRP_16_A	ASN_109_A
VAL_21_A	LEU_17_A	ASN_30
ILE_35_A	TYR_19_A	ARG_308_A
VAL_38_A	ALA_20_A	THR_309_A
MET_39_A	ILE_21_A	GLU_310_A

TABLE 4.6: Catalytic pocket detection for docking process

Thioredoxin	**DNA-binding**	**UDP-N-acetyl**
ILE_42_A	TYR_23_A	LYS_312
PHE_51_A	GLN_24_A	ILE_320_A
PRO_72_A	LEU_80_A	LEU_321_A
SER_73_A	LEU_101_A	SER_322_A
PHE_74_A	THR_102_A	ASP_323_A
VAL_75_A	LEU_105_A	VAL_324_A
VAL_76_A	GLU_106_A	TYR_325_A
LEU_83_A	LYS_107_A	ASN_326_A
GLY_84_A	GLN_108_A	ALA_327_A
ARG_85_A	ALA_109_A	ASN_328_A
LEU_86_A	LEU_110_A	PRO_329_A
VAL_87_A	ILE_111_A	ALA_331_
ASN_88_A	TYR_2_B	MET_332_
THR_93_A	LEU_12_B	MET_357_A

TABLE 4.6: Catalytic pocket detection for docking process

Thioredoxin	**DNA-binding**	**UDP-N-acetyl**
LYS_94_A	TRP_16_B	LYS_358_A
GLU_96_A	LEU_17_B	LEU_360_A
ILE_97_A	TYR_19_B	LEU_437_A
LYS_99_A	ALA_20_B	LYS_438_A
PHE_100_A	ILE_21_B	GLY_439_A
LEU_101_A	TYR_23_B	SER_440_A
GLU_103_A	GLN_24_B	ASN_441_A

4.12 Molecular Docking

MOE prepared these antibacterial compounds while using the energy minimization for these compounds as ligands. By using the Swiss Prot protein PDB structures were retrieved . MOE using 3D protonation and energy minimization. It proves that MOE is reliable for docking studies.

It can also be used to determine the correct structure of the ligand within the target binding site and to estimate the strength of the binding between the ligand and the target proteins through a specific scoring function.

It also helps in the recognition of new small molecular compounds, revealing the essential properties such as high interaction between binding with target protein which help in the selection of lead compound for the target [112].

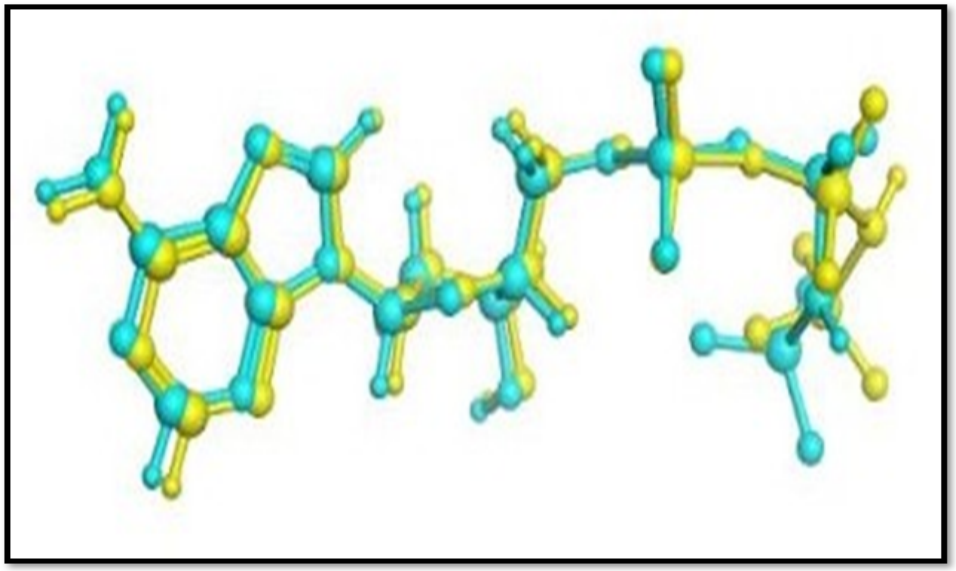

FIGURE 4.5: Docking of multiple ligands

The top nine compounds with high binding affinities were selected after each protein was docked against all of the prepared ligands.

4.13 Thioredoxin Family Protein

Thioredoxin is a cytoplasmic protein known to be present in all organisms. It plays a role in many important biological processes, including redox signaling. Molecular weight of Thioredoxin family protein is 12573.44. The docking results of Thioredoxin family protein are explained below in the table 4.7.

TABLE 4.7: Compound's name, binding affinity and Residue interaction of top 10 compounds with thioredox-in binding.

Compound	**Energy minimize (MM\GBVI)**	**S-Score**	**Residues Interaction⋆⋆**	**No. of Interaction⋆⋆⋆**
Sperimidine	-15.103	-7.9566	Arg91	2
Canpesterol	-8.710	-6.6651	Asn90 Asp45	1

TABLE 4.7: Compound's name, binding affinity and Residue interaction of top 10 compounds with thioredox-in binding.

Compound	**Energy minimize (MM\GBVI)**	**S-Score**	**Residues Interaction⋆⋆**	**No. of Interaction⋆⋆⋆**
Berberine	-10.473	-6.7166	Lys94	2
Coumarin	-10.530	-7.7234	Thr98 Thr93 Asp45	1
Quercitin	-13.12331	-9.3293	Thr94	3
Stigmasterol	-10.537	-6.5439	Lys94 Asn90	1
Rutin	-20.639	-11.0333	2\Lys94	2
Kaempferol	-18.906	-10.3699	Asp31	1
Hypercalin B	-16.308	-6.9508	Glu43	1

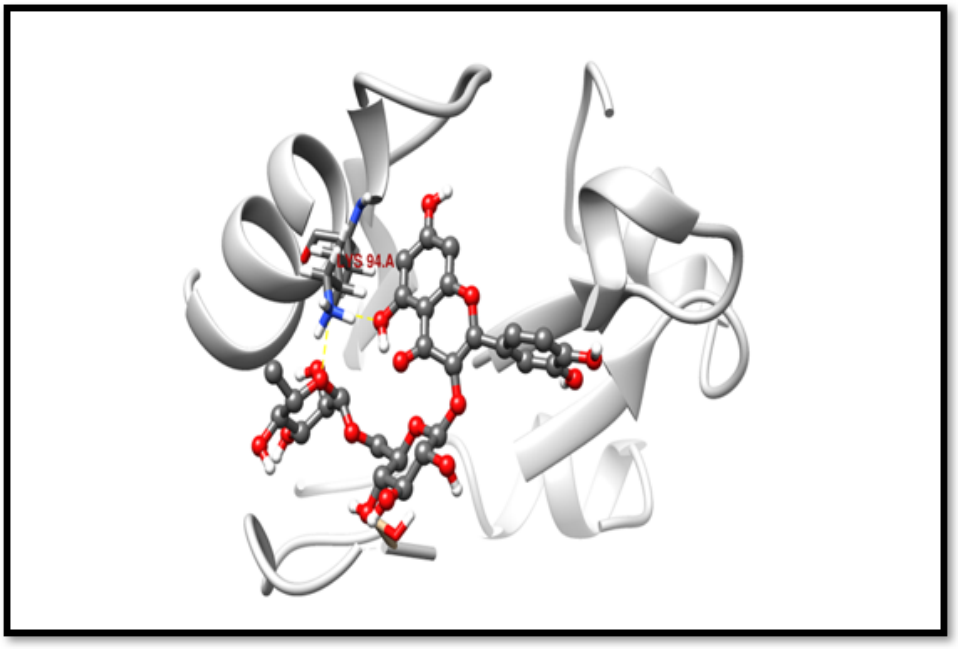

FIGURE 4.6: 3D interactionof Thioredoxin family protein

This structure (4.6) is extracted by using chimera. Rutin as a ligand was found to be most promise compound. In the case of Rutin,2 no of interactions were observed. We predicted that residues 2/lys of Thioredoxin family protein were interacted and docking score was predicted as -11.0333.

4.14 DNA Binding Response Regulator

It is a cytoplasmic protein. DNA binding response regulators are involved in various uptake processes, including nitrate/ nitrite. The molecular weight of DNA binding response regulator is 30117.76. The docking results of DNA binding response regulator protein are explained below 4.8.

TABLE 4.8: Compound's name, binding affinity and Residue interaction of top 10 compounds with DNA-binding response regulator.

Compound	**Energy minimize (MM\GBVI)**	**S-Score**	**Residues Interaction⋆⋆**	**No. of Interaction⋆⋆⋆**
Sperimidine	-9.733	-6.1197	ProA104	1
Campesterol	-10.953	-5.8724	GlnA108	2
Scopoletin	-13.134	-9.0037	AspA112 Thr93 LysA107	3
Berberine	-7.621	-6.6444	TyrB19 SrrA25	3

TABLE 4.8: Compound's name, binding affinity and Residue interaction of top 10 compounds with DNA-binding response regulator.

Compound	**Energy minimize (MM\GBVI)**	**S-Score**	**Residues Interaction⋆⋆**	**No. of Interaction⋆⋆⋆**
Coumarin	-7.629	-5.7264	LysA125	1
Hyperenone A	-14.604	-6.6166	TyrB19	1
Quercitin	-18.052	-8.8192	LysA107	1
Stigmasterol	-9.169	-5.1830	Gln108	1
Rutin	-13.127	-9.6658	SerA25 TyrB19	2
Kaempferol	-13.127	-7.0724	2\LysA107	1
Hypercalin B	-7.608	-76095	Glu43	1

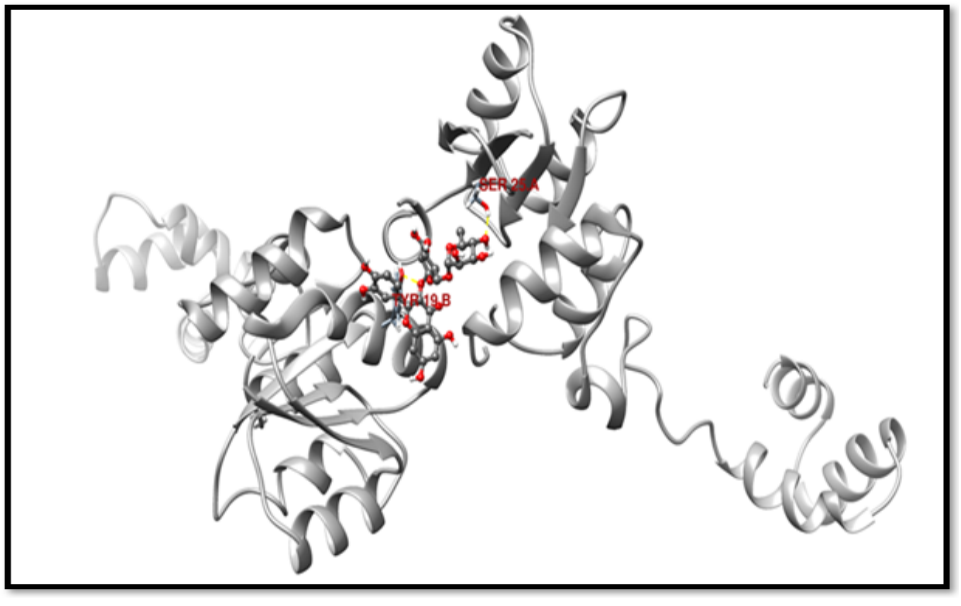

FIGURE 4.7: 3D interaction of DNA binding response regulator.

3D zoomed image showed two interaction of SerA25 which is interacted with hydrogen group of rutin compound and TyrB19 is interacted with oxygen group of rutin compound.

4.15 UDP-N-acetyl-D-glutamylD-alanyl ligase

UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate–D-alanyl-D-alanyl ligases is a cytoplasmic protein and has 50399.66 kDa molecular weight. It is involved in the synthesis of a cell-wall peptide in bacteria and has 50399.66 molecular weight. The docking results of UDML are explained below 4.9.

TABLE 4.9: Shows compound's name, binding affinity and residue interaction of 10 compounds.

Compound	**Energy minimize (MM\GBVI)**	**S-Score**	**Residues Interaction⋆⋆**	**No. of Interaction⋆⋆⋆**
Sperimidine	-16.599	-7.2594	Thr309	1
Campesterol	-16.022	-9.2844	Asn134	1
Berberine	-9.954	-10.4248	Asn134	1
Coumarin	-4.939	-5.7318	Asn134	1
Hyperenone A	-8.184	-7.6417	Trp311	1
Quercitin	-15.439	-11.8090	Ser108 Gly185	2
Stigmasterol	-16.531	-9.1501	Gly185	1

TABLE 4.9: Shows compound's name, binding affinity and residue interaction of 10 compounds.

Compound	**Energy minimize (MM\GBVI)**	**S-Score**	**Residues Interaction⋆⋆**	**No. of Interaction⋆⋆⋆**
Rutin	-25.091	-15.0364	Gln161	2
Kaempferol	-10.079	-12.2667	Ser108	1
Hypercalin B	-28.481	-12.2500	Asn136	1

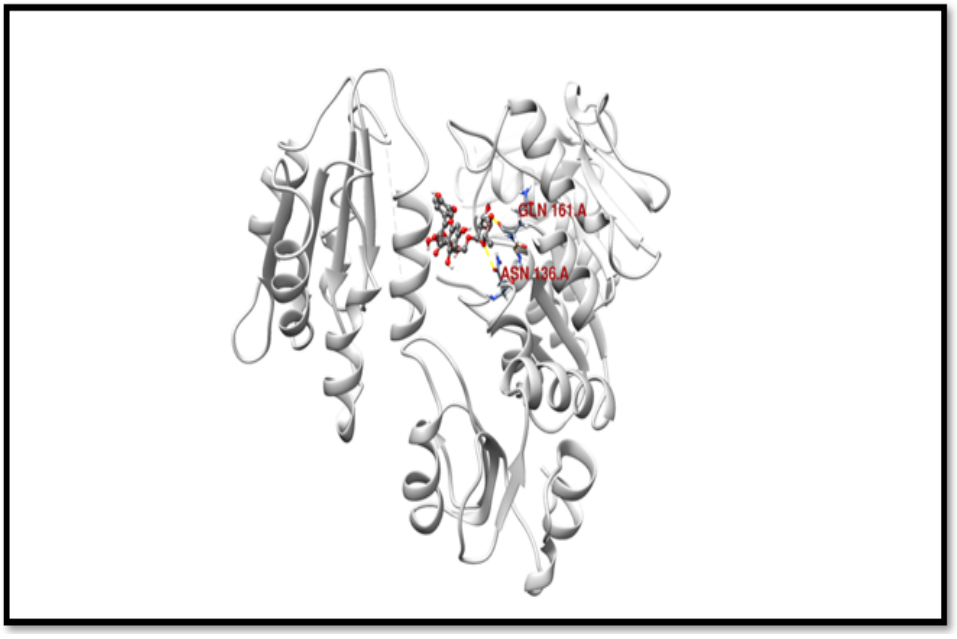

FIGURE 4.8: 3D interaction of UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-D-alanyl-D-alanyl ligase are interacting with rutin ligand.

3D zoomed image showed two interaction of Gln 161 which is interacted with hydrogen group of Rutin compound and Asn 136 is interacted with oxygen group of rutin compound. Rutin as a ligand was found to be most promising compound which is selected on the basis of binding score and number of interactions.

⋆The intensity of the binding relationship between a protein and compound (ligand).

⋆⋆Residues that are involved in the interaction.

⋆⋆⋆No of interactions formed between active site and ligand.

Previous studies have suggested that several antibiotic candidates have been explored for their anti-virulence effectiveness against bacteria. The research showed that these compounds inhibited the early stages of bacterial communication. Currently, there are hundreds of thousands of natural compounds that can be used for screening to find new therapeutic targets.

Hyperenone A, Rutin, Sperimidine, Coumarin, Stigmasterol, Querecetin Scopoletin, Hypercalin B , Kaempferol and Berberine compounds and three cytoplasmic protein namely Thioredoxin family binding protein, DNA binding response regulator and UDML showed in this study. Rutin showed docking score with these cytoplasmic proteins such as -11.0333 , -9.6658 and -15.0364.Rutin also showed 2 interactions with cytoplasmic proteins. Best lead compound is selected on the basis of number of interactions and docking score .Our analysis predicted that Rutin showed highest docking score and interactions with DNA binding response regulator as compared to other two proteins. However, these findings revealed Rutin had demonstrated itself as a promising potential anti-virulence agent against DNA binding response regulator proteins. Consequently, it might be an excellent candidate for drugs to treat bacterial infections.

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
NCTC13949	PRJEB6403	GCA_ 900638415	2.03288	35.6
SA111	PRJEB12926	GCA_900078265.1	2.27514	35.9
105B	PRJNA388485	GCA_003288035.1	2.27372	35.7

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
NCTC8184	PRJEB6403	GCA_900636375.1	2.26597	35.8
GBS1-NY	PRJNA24385	GCA_000831145.1	2.24371	35.5
GBS6	PRJNA244773	GCA_000831105.1	2.23148	35.8
HU-GS5823	PRJDB7410	GCA_003966545.1	2.23131	35.6
CUGBS5910	PRJNA293392	GCA_002197205.1	2.22768	35.8
NGBS 061	PRJNA246096	GCA_000730215.2	2.22121	35.5
Sag153	PRJNA521644	GCA_012222485.1	2.1745	35.8
NGBS357	PRJNA293858	GCA_001712835.1	2.17287	35.6
BM110	PRJEB 18603	GCA_900155855.1	2.17028	35.5
SG-M8	PRJNA293392	GCA_002197325.1	2.16712	35.6
GBS ST-1	PRJNA96923	GCA_001448985.1	2.16597	35.4

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
2603 V/R	PRJNA330	GCA_000007265.1	2.16027	35.6
874391	PRJNA395243	GCA_002289205.1	2.15394	35.5
BJ 01	PRJNA647240	GCA_013786965.1	2.14957	35.7
32790	PRJNA472222	GCA_006716245.1	2.1489	35.7
H002	PRJNA258310	GCA_001190885.1	2.14742	35.7
GBS11	PRJNA 556442	GCA_014218095.1	2.141	35.6
09mas 018883	PRJE1693	GCA_000427035.1	2.13869	35.5
GBS28	PRJNA556442	GCA_014218135.1	2.1374	35.7
NJ 1606	PRJNA430486	GCA_009930915.1	2.13644	35.7
FDAARGOS_512	PRJNA231221	GCA_003812805.1	2.13414	35.6
A909	PRJNA326	GCA_000012705.1	2.12784	35.6
C001	PRJNA252450	GCA_002214425.1	2.12137	35.6
SG-M1	PRJNA293392	GCA_001275545.2	2.11681	35.5
SG-M158	PRJNA293392	GCA_002197265.1	2.11681	35.5

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
SG- 50	PRJNA293392	GCA_002197245.1	2.11681	35.5
SG-M163	PRJNA293392	GCA_002197425.1	2.11681	35.5
SG-M- 29	PRJNA293392	GCA_002197285.1	2.11677	35.5
SGEHI2015-95	PRJNA293392	GCA_002812445.1	2.11677	35.5
ILR1005	PRJEB1694	GCA_000427075.1	2.10976	35.4
SG-M6	PRJNA293392	GCA_002197365.1	2.10602	35.6
1173	PRJNA627590	GCA_013000945.1	2.1053	35.7
GBS19	PRJNA556442	GCA_014218115.1	2.10056	35.5
Sag158	PRJNA376652	GCA_002025005.1	2.09688	35.7
NCTC13947	PRJEB6403	GCA_900638495.1	2.09519	35.6
PLG BS13	PRJNA473176	GCA_006874565.1	2.09503	35.5
CJB111	PRJNA6663970	GCA_015221735.2	2.09399	35.5
GBS- M002	PRJNA307137	GCA_001932715.1	2.09257	35.6
SS1	PRJNA274384	GCA_001026925.1	2.09207	35.5

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
GBS7	PRJNA556442	GCA_014218075.1	2.09043	35.5
FWL1402	PRJNA323692	GCA_001683515.1	2.09029	35.4
CU_G BS_08	PRJNA274685	GCA_001592385.1	2.08451	35.4
B507	PRJNA388485	GCA_003288055.1	2.08248	35.4
TFJ0901	PRJNA475097	GCA_003939065.1	2.08094	35.7
GBS30	PRJNA556442	GCA_014218155.1	2.07562	35.5
NGBS128	PRJNA293561	GCA_001552035.1	2.07912	35.7
SG- M4	PRJNA293392	GCA_002197385.1	2.07201	35.5
FDAAR GOS-669	PRJNA231221	GCA_008693505.1	2.06568	35.4
COH 1	PRJEB5232	GCA_000689235.1	2.06507	35.4
ZQ0910	PRJNA611468	GCA_011383065.1	2.06494	35.7
HN016	PRJNA258321	GCA_001190805.1	2.06472	35.7
NCTC11930	PRJEB6403	GCA_900474905.1	2.06327	35.4
GD201008-001	PRJNA169338	GCA_000299135.1	2.06311	35.6

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
NGBS572	PRJNA246096	GCA_000730255.1	2.06143	35.5
B509	PRJNA388485	GCA_003287995.	2.06064	35.5
S73	PRJNA479809	GCA_003319215.1	2.05992	35.4
SGEHI2015-25	PRJNA293392	GCA_002812465.1	2.05471	35.4
NCTC8187	PRJEB6403	GCA_900475355.1	2.05246	35.4
SGEHI2015-113	PRJNA293392	GCA_002812505.1	2.04385	35.4
515	PRJNA594846	GCA_012593885.1	2.03274	35.5
CU_ GBS98	PRJNA274685	GCA_001592425.1	1.029672	35.4
SGEHI2015-107	PRJNA293392	GCA_002812425.1	2.01681	35.4
CNCTC10/84	PRJNA229124	GCA_000782855.1	2.01384	35.4
GBS85147	PRJNA263907	GCA_001266635.1	1.99615	35.4
ILR I112	PRJEB1774	GCA_000427055.1	2.0292	35.3
SA95	PRJNA369821	GCA_002881375.1	1.8565	35.5
SA97	PRJNA369822	GCA_002881935.1	1.85641	35.5

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
SA132	PRJNA369827	GCA_002882035.1	1.852	35.5
SA85	PRJNA369820	GCA_002881355.1	1.84999	35.5
SA218	PRJNA369839	GCA_002881515.1	1.84998	35.5
SA102	PRJNA369826	GCA_002881395.1	1.84952	35.5
SA136	PRJNA369828	GCA_002881415.1	1.8491	35.5
SA75	PRJNA369817	GCA_002881865.1	1.84902	35.5
SA289	PRJNA369843	GCA_002881595.1	1.84899	35.5
SA256	PRJNA369842	GCA_002881575.1	1.84897	35.5
SA53	PRJNA369815	GCA_002881235.1	1.84897	35.5
SA245	PRJNA369841	GCA_002881555.1	1.84896	35.5
SA73	PRJNA369816	GCA_002881195.1	1.84884	35.5
SA191	PRJNA369832	GCA_002882125.1	1.84868	35.5
SA374	PRJNA369849	GCA_002881695.1	1.84225	35.5
SA375	PRJNA369850	GCA_002882745.1	1.84222	35.5

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
SA623	PRJNA369851	GCA_002882835.1	1.84211	35.5
SA341	PRJNA369846	GCA_002882555.1	1.84211	35.5
SA330	PRJNA369845	GCA_002882375.1	1.84208	35.5
SA627	PRJNA36985	GCA_002881775.1	1.84207	35.5
SA333	PRJNA369844	GCA_002882465.1	1.842	35.5
SA346	PRJNA369848	GCA_002882645.1	1.84198	35.5
SA343	PRJNA369847	GCA_002881615.1	1.84198	35.5
SA212	PRJNA369838	GCA_002882275.1	1.84196	35.5
SA20	PRJNA174852	GCA_000302475.3	1.84195	35.5
SA79	PRJNA369818	GCA_002881315.1	1.84195	35.5
SA5	PRJNA369810	GCA_002881255.1	1.84194	35.5
SA1	PRJNA369807	GCA_002881215.1	1.84194	35.5
SA9	PRJNA369809	GCA_002881275.1	1.84193	35.5
SA184	PRJNA369831	GCA_002881455.1	1.84189	35.5

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
SA16	PRJNA369811	GCA_002881295.1	1.84186	35.5
SA209	PRJNA369834	GCA_002881495.1	1.84186	35.5
SA201	PRJNA369835	GCA_002882205.1	1.84184	35.5
SA30	PRJNA369813	GCA_002881155.1	1.84173	35.5
SA195	PRJNA369833	GCA_002881475.1	1.84171	35.5
SA33	PRJNA369814	GCA_002881175.1	1.84163	35.5
SA159	PRJNA369829	GCA_002881435.1	1.84148	35.5
GX026	PRJNA254961	GCA_001190865.1	1.84065	35.5
SA81	PRJNA369819	GCA_002881335.1	1.84036	35.5
S25	PRJNA323652	GCA_001655175.1	1.83899	35.5
222	PRJEB643	GCA_000967445.1	1.83887	35.5
138P	PRJNA226756	GCA_000599965.1	1.8387	35.5
138Spar	PRJNA226756	GCA_000636115.1	1.83813	35.5
S13	PRJNA356737	GCA_001908255.1	1.83516	35.4

TABLE 4.10: Summarizes the details of selected strains of Streptococcus agalactiae for further analysis [117]

Strain name	**Bio project**	**Assembly**	**Size(Mb)**	**GC%**
QMA0271	PRJNA41753	GCA_003186145	1.80247	35.3
2012-845	PRJNA626549	GCA_016454805.1	1.53049	35.3

Chapter 5

Conclusions and Recommendations

Streptococcus agalactiae is responsible for a number of fatal illnesses that affect people all over the world, *S.agalactiae* is the reason of numerous fatal diseases that increases to find a solution that may effectively stop this dangerous infection. The purpose of this study was to understand the genomic diversity of *S.agalactiae* as well as to find novel drug against this pathogen. 3 essential proteins were selected through the core genome analysis that were used for this study. On the basis of cellular localization, these proteins were divided into cytoplasmic studies for drug targeting. All of the identified targets are playing an important role in the selected pathogen.

The first objective of this study was to identify of core genome *S.agalactiae* by pan genome analysis and for this purpose 127 strains of *S.agalactiae* were analyzed.

The second objective of this was to evaluate the potential of natural compounds against *S.agalactiae* as a drug. For this purpose, 10 natural anti-bacterial compounds were collected through literature against 3 essential genes that were found through the cellular localization after drug prioritization. This is done in the drug-targeting approach. For this objective 10 anti-bacterial compounds were selected against all the proteins based on residues interactions and binding affinities.

The third objective was to identify the antibacterial drug candidate from host lead compounds. One best lead Rutin compound which showed highest binding score and residues interactions with DNA binding response regulator for drug target against *S.agalactiae* that was selected can be further precede for clinical trials. The data given in this study require further experimental authentication for verification but we anticipate promising outcomes from this predicted drug targets against the deadly toxins of *S.agalactiae*.

Bibliography

[1] M. Yamaguchi, "Synergistic findings from microbiological and evolutional analyses of virulence factors among pathogenic streptococcal species," *J. Oral Biosci.*, vol. 60, no. 2, pp. 36– 40, 2018, doi: 10.1016/j.job.2018.02.004.

[2] M. Emaneini, B. khoramian, F. Jabalameli, S. Abani, H. Dabiri, and R. Beigverdi, "Comparison of virulence factors and capsular types of Streptococcus agalactiae isolated from human and bovine infections," *Microb. Pathog.*, vol. 91, pp. 1–4, 2016, doi: 10.1016/j.micpath.2015.11.016.

[3] A. Dobrut and M. Brzychczy-Włoch, "Immunogenic Proteins of Group B Streptococ-cus— Potential Antigens in Immunodiagnostic Assay for GBS Detection," Pathogens, vol. 11, no. 1, 2022, doi: 10.3390/pathogens11010043.

[4] L. Thomas and L. Cook, "Two-component signal transduction systems in the human pathogen streptococcus agalactiae," *Infect. Imun.*, vol. 88, no. 7, 2020, doi: 10.1128/IAI.00931-19.

[5] K. Roloff, G. Stepanyan, and G. Valenzuela, "Prevalence of oropharyngeal group b streptococcus colonization in mothers, family, and health care providers," *PLoS One*, vol. 13, no. 9, pp. 1–8, 2018, doi: 10.1371/journal.pone.0204617.

[6] T. Motallebirad, H. Fazeli, A. Ghahiri, D. Shokri, S. Jalalifar, and S. Moghim, "Preva-lence , population structure , distribution of serotypes , pilus islands and resistance genes among erythromycin- resistant colonizing

and invasive Streptococcus agalactiae isolates recovered from pregnant and non-pregnant women in Isfahan , Iran," pp. 1–11, 2021.

[7] M. Gizachew, M. Tiruneh, F. Moges, and B. Tessema, "Streptococcus agalactiae ma-ternal colonization, antibiotic resistance and serotype profiles in Africa: A meta-analysis," *Ann. Clin. Microbiol. Antimicrob.*, vol. 18, no. 1, pp. 1–14, 2019, doi: 10.1186/s12941-019-0313-1.

[8] Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, 1982, doi: 10.1016/00222836(82)90515-0.

[9] N. Brimil, E. Barthell, U. Heindrichs, M. Kuhn, R. Lütticken, and B. Spellerberg, "Ep-idemiology of Streptococcus agalactiae colonization in Germany," *Int. J. Med.* Microbiol., vol. 296, no. 1, pp. 39–44, 2006, doi: 10.1016/j.ijmm.2005.11.001.

[10] V. N. Raabe and A. L. Shane, " Group B Streptococcus (Streptococcus agalactiae) ," Microbiol. Spectr., vol. 7, no. 2, pp. 1–13, 2019, doi: 10.1128/microbiolspec.gpp3-0007-2018.

[11] J. Y. Bolukaoto et al., "Antibiotic resistance of Streptococcus agalactiae isolated from pregnant women in Garankuwa, South Africa," BMC Res. Notes, vol. 8, no. 1, pp. 6–12, 2015, doi: 10.1186/s13104-015-1328-0.

[12] D. Barh, K. Gupta, N. Jain, G. Khatri, N. Le´on-Sicairos, A. CanizalezRoman, S. Tiwari, A. Verma, S. Rahangdale, S. Shah Hassan et al., "Conserved host–pathogen ppis globally conserved inter-species bacterial ppisbased conserved host-pathogen interactome derived novel target in c. pseudotuberculosis, c. diphtheriae, m. tuberculosis, c. ulcerans, y. pestis, and e.coli tar-geted by piper betel compounds," *Integrative Biology*, vol. 5, no. 3,pp. 495–509, 2013.

[13] W. Ruankham, P. Winyangkul, and N. Bunchu, "In silico analysis of potential diagnostictargets from burkholderia pseudomal-lei," *Asian Pacific J. Trop. Dis.*, vol. 6, no. 10, pp. 778–782, 2016.

[14] Lee, S. H., Kang, J. S., Min, J. S., Yoon, K. S., Strycharz, J. P., Johnson, R., ... Clark, J. M., "metabolic enzymes in pseudomonas aeruginosa," *Mol. Bio*l., vol. 19, no. 5, pp. 599–615, 2010.

[15] D. B. Thompson, K. Crandall, S. V. Harding, S. J. Smither, G. B. Kitto,R. W. Titball, and K. A. Brown, "In silico analysis of potential diagnostic-targets from burkholderia pseudomal-lei," *Transactions of the Royal Societyof Tropical Medicine and Hygiene*, vol. 102, no. Supplement 1, pp. S61–S65,2008.

[16] Amir, S. M. Faizan, Asif, A. Asad, M. S. Anjali, A. Arvind, K. Neelesh,K. Hirdesh, and K.Sanjay,"Potential targets for mycobacterium tuberculosis h37rv through comparative ge-nome analysis," *Bioinformation*, vol. 4,no. 6, p. 245, 2009.

[17] S. M. Asif, A. Asad, A. Faizan, M. S. Anjali, A. Arvind, K. Neelesh, K. Hirdesh, and K. Sanjay, "Dataset of potential targets for mycobacterium tuberculosis h37rv through compara-tive genome analysis," *Bioinformation*, vol. 4,no. 6, p. 245, 2009.

[18] B. Rathi, A. N. Sarangi, and N. Trivedi, "Genome subtraction for novel target definition in salmonella typhi," *Bioinformation*, vol. 4, no. 4, p. 143,2009.

[19] K. R. Sakharkar, M. K. Sakharkar, and V. T. Chow, "Biocomputational-strategies for microbial drug target identification," in New antibiotic targets.*Springer*, 2008, pp. 1–9.

[20] D. Barh and A. Kumar, "In silico identification of candidate drug and vaccine targets from various pathways in neisseria gonorrhoeae," *In silico biology*,vol. 9, no. 4, pp. 225–231, 2009.

[21] J. N. Chan, C. Nislow, and A. Emili, "Recent advances and method development for drug target identification," Trends in pharmacological sciences,vol. 31, no. 2, pp. 82–88, 2010.

[22] D. Bumann, T. F. Meyer, and P. R. Jungblut, "Proteome analysis of the common human pathogen helicobacter pylori," PROTEOMICS: International Edition, vol. 1, no. 4, pp. 473–479, 2001.

[23] A. N. Sarangi, R. Aggarwal, Q. Rahman, and N. Trivedi, "Subtractive genomics ap-proach for in silico identification and characterization of novel drug targets in neisseria meningiti-des serogroup b," Journal of Computer Science& Systems Biology, vol. 2, no. 5, pp. 255–258, 2009.

[24] T. Kabelitz, E. Aubry, K. van Vorst, T. Amon, and M. Fulde, "The role of strepto-coccus spp. In bovine mastitis," Microorganisms, vol. 9, no. 7, pp. 1–21, 2021, doi: 10.3390/microorganisms9071497

[25] N. Chaturvedi and S. Nayak, "Future of generic drugs and india's interest,"Available at SSRN 1762356, 2011.

[26] K. Mdluli and M. Spigelman, "Novel targets for tuberculosis drug discovery,"Current opinion in pharmacology, vol. 6, no. 5, pp. 459–467, 2006.

[27] S. Abachi, S. Lee, and H. P. V. Rupasinghe, Molecular mechanisms of inhibition of streptococcus species by phytochemicals, vol. 21, no. 2. 2016. doi: 10.3390/molecules21020215.

[28] A. H. Nobbs, R. J. Lamont, and H. F. Jenkinson, " Streptococcus Adherence and Col-onization ," Microbiol. Mol. Biol. Rev., vol. 73, no. 3, pp. 407–450, 2009, doi: 10.1128/mmbr.00014-09.

[29] A. Fischer et al., "Camel Streptococcus agalactiae populations are associated with specific disease complexes and acquired the tetracycline resistance gene tetM via a Tn916-like el-ement," Vet. Res., vol. 44, no. 1, 2013, doi: 10.1186/1297-9716-44-86.

[30] R. S. W. Tsang, "A narrative review of the molecular epidemiology and laboratory surveillance of vaccine preventable bacterial meningitis agents: Streptococcus pneumoniae, neis-seria meningitidis, haemophilus influenzae

and streptococcus agalactiae," Microorganisms, vol. 9, no. 2, pp. 1–19, 2021, doi: 10.3390/microorganisms9020449.

[31] D. S. McKenna, S. Matson, and I. Northern, "Maternal group B streptococcal (GBS) genital tract colonization at term in women who have asymptomatic GBS bacteriuria," Infect. Dis. Obstet. Gynecol., vol. 11, no. 4, pp. 203–207, 2003, doi: 10.1080/10647440300025522.

[32] C. Bureta et al., "Lumbar spine epidural abscess and facet joint septic arthritis due to Streptococcus agalactiae: A case report," BMC Surg., vol. 18, no. 1, pp. 2–7, 2018, doi: 10.1186/s12893-018-0350-2.

[33] S. Y. Chung, C. H. Chen, and W. L. Yu, "Spinal epidural abscess caused by group B Streptococcus in a diabetic woman presenting with febrile low back pain," Jpn. J. Infect. Dis., vol. 58, no. 3, pp. 177–179, 2005.

[34] J. Vornhagen, K. M. Adams Waldorf, and L. Rajagopal, "Perinatal Group B Strepto-coccal Infections: Virulence Factors, Immunity, and Prevention Strategies," Trends Microbiol., vol. 25, no. 11, pp. 919–931, 2017, doi: 10.1016/j.tim.2017.05.013.

[35] A. E. Muller, P. M. Oostvogel, E. A. P. Steegers, and P. J. Dörr, "Morbidity related to maternal group B streptococcal infections," Acta Obstet. Gynecol. Scand., vol. 85, no. 9, pp. 1027–1037, 2006, doi: 10.1080/00016340600780508.

[36] M. Haenni, A. Lupo, and J.-Y. Madec, " Antimicrobial Resistance in Streptococcus spp ," Microbiol. Spectr., vol. 6, no. 2, 2018, doi: 10.1128/microbiolspec.arba-0008-2017.

[37] P. Cools et al., "Role of HIV exposure and infection in relation to neonatal GBS dis-ease and rectovaginal GBS carriage: A systematic review and meta-Analysis," Sci. Rep., vol. 7, no. 1, pp. 1–13, 2017, doi: 10.1038/s41598-017-13218-1.

[38] B. S. Apgar, G. Greenberg, and G. Yen, "Prevention of group B streptococcal dis-ease in the newborn," Am. Fam. Physician, vol. 71, no. 5, pp. 903–910, 2005.

[39] D. Numberger, U. Siebert, M. Fulde, and P. Valentin-Weigand, "Streptococcal in-fections in Marine mammals," Microorganisms, vol. 9, no. 2, pp. 1–25, 2021, doi: 10.3390/microorganisms9020350.

[40] V. Kubín, H. Mraštíková, M. Paulová, J. Motlová, and J. Franěk, "Group B strepto-cocci in the milk of lactating mothers," Zentralblatt fur Bakteriol. Mikrobiol. und Hyg. - Abt. 1 Orig. A, vol. 265, no. 1–2, pp. 210–217, 1987, doi: 10.1016/S0176-6724(87)80168-2.

[41] N. Chopin, A. Bonnet, and J. Gabet, "MCningite h streptocoque B apr & s anesth & ie pkridurale obstktricale CCcitC corticale au tours de la toxkmie gravidique . Amaurose transi-toire par baisse de pression de perfusion cCr6brale," pp. 195–196, 1998.

[42] P. L. Ruegg, "A 100-Year Review: Mastitis detection, management, and prevention," J. Dairy Sci., vol. 100, no. 12, pp. 10381–10397, 2017, doi: 10.3168/jds.2017-13023.

[43] T. Kabelitz, E. Aubry, K. van Vorst, T. Amon, and M. Fulde, "The role of strepto-coccus spp. In bovine mastitis," Microorganisms, vol. 9, no. 7, pp. 1–21, 2021, doi: 10.3390/microorganisms9071497

[44] P. Krishnamoorthy, K. P. Suresh, K. S. Jayamma, B. R. Shome, S. S. Patil, and R. G. Amachawadi, "An understanding of the global status of major bacterial pathogens of milk con-cerning bovine mastitis: A systematic review and meta-analysis (scientometrics)," Pathogens, vol. 10, no. 5, 2021, doi: 10.3390/pathogens10050545.

[45] A. Tibary, C. Fite, A. Anouassi, and A. Sghiri, "Infectious causes of reproductive loss in camelids," Theriogenology, vol. 66, no. 3 SPEC. ISS., pp. 633–647, 2006, doi: 10.1016/j.theriogenology.2006.04.008.

[46] T. Kabelitz, E. Aubry, K. van Vorst, T. Amon, and M. Fulde, "The role of strepto-coccus spp. In bovine mastitis," Microorganisms, vol. 9, no. 7, pp. 1–21, 2021, doi: 10.3390/microorganisms9071497 .

[47] P. L. Ruegg, "A 100-Year Review: Mastitis detection, management, and prevention," J. Dairy Sci., vol. 100, no. 12, pp. 10381–10397, 2017, doi: 10.3168/jds.2017-13023.

[48] D. Numberger, U. Siebert, M. Fulde, and P. Valentin-Weigand, "Streptococcal infec-tions in Marine mammals," Microorganisms, vol. 9, no. 2, pp. 1–25, 2021, doi: 10.3390/microorganisms9020350.

[49] Harris, E., Videa, E., Sandoval, E., Cuadra, R., Rocha, J., Idiaquez, W., ... & Balmaseda, A., "Clinical, epidemiologic, and virologic features of dengue in the 1998 epidemic in Nicaragua.," *The American Journal of Tropical Medicine and Hygiene*, vol. 63, no. 1, pp. 5–11, Jul. 2000, doi: 10.4269/ajtmh.2000.63.5.

[50] A. Canini, D. Alesiani, G. D'Arcangelo, and P. Tagliatesta, "Gas chromatography–mass spectrometry analysis of phenolic compounds from Carica papaya L. leaf," *Journal of Food Composition and Analysis*, vol. 20, no. 7, pp. 584–590, Nov. 2007, doi: 10.1016/j.jfca.2007.03.009.

[51] F. Stanzione, I. Giangreco, and J. C. Cole, Use of molecular docking computational tools in drug discovery, 1st ed., vol. 60. Elsevier B.V., 2021. doi: 10.1016/bs.pmch.2021.01.004.

[52] P. Martin, H. W. Barkema, L. F. Brito, S. G. Narayana, and F. Miglior, "Symposium review: Novel strategies to genetically improve mastitis resistance in dairy cattle," J. Dairy Sci., vol. 101, no. 3, pp. 2724–2736, 2018, doi: 10.3168/jds.2017-13554.

[53] P. L. Ruegg, "A 100-Year Review: Mastitis detection, management, and prevention," J. Dairy Sci., vol. 100, no. 12, pp. 10381–10397, 2017, doi: 10.3168/jds.2017-13023.

[54] M. Kaur, N. C. Talniya, S. Sahrawat, A. Kumar, and E. E. Stashenko, "Ethnomedicinal Uses, Phytochemistry and Pharmacology of Carica papaya Plant: A Compendious Review," *Mini-Reviews in Organic Chemistry*, vol. 16, no. 5, pp. 463–480, Jun. 2019, doi: 10.2174/1570193x15666180816110733.

[55] C. S. Lachenauer et al., "Serotypes VI and VIII predominate among group B strepto-cocci isolated from pregnant Japanese women," J. Infect. Dis., vol. 179, no. 4, pp. 1030–1033, 1999, doi: 10.1086/314666.

[56] K. Le Doare et al., "Risk factors for Group B Streptococcus colonisation and disease in Gambian women and their infants," J. Infect., vol. 72, no. 3, pp. 283–294, 2016, doi: 10.1016/j.jinf.2015.12.014. [42]L. C. Paoletti and D. L. Kasper, "Surface structures of group B streptococcus important in human immunity," Gram-Positive Pathog., pp. 204–227, 2019, doi: 10.1128/9781683670131.ch13.

[57] D. Latousakis, D. A. MacKenzie, A. Telatin, and N. Juge, "Serine-rich repeat pro-teins from gut microbes," Gut Microbes, vol. 11, no. 1, pp. 102–117, 2020, doi: 10.1080/19490976.2019.1602428

[58] H. S. Seo, Y. Q. Xiong, and P. M. Sullam, "Role of the Serine-Rich Surface Glyco-protein Srr1 of Streptococcus agalactiae in the Pathogenesis of Infective Endocarditis," PLoS One, vol. 8, no. 5, pp. 1–6, 2013, doi: 10.1371/journal.pone.0064204.

[59] T. R. Sheen, A. Jimenez, N. Y. Wang, A. Banerjee, N. M. van Sorge, and K. S. Doran, "Serine-rich repeat proteins and pili promote Streptococcus agalactiae colonization of the vaginal tract," J. Bacteriol., vol. 193, no. 24, pp. 6834–6842, 2011, doi: 10.1128/JB.00094-11.

[60] M. Y. Mistou, S. Dramsi, S. Brega, C. Poyart, and P. Trieu-Cuot, "Molecular dis-section of the secA2 locus of group B streptococcus reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence," J. Bacteriol., vol. 191, no. 13, pp. 4195–4206, 2009, doi: 10.1128/JB.01673-08.

[61] L. C. Paoletti and D. L. Kasper, "Surface structures of group B streptococcus im-portant in human immunity," Gram-Positive Pathog., pp. 204–227, 2019, doi: 10.1128/9781683670131.ch13

[62] S. Shabayek and B. Spellerberg, "Group B streptococcal colonization, molecular characteristics, and epidemiology," Front. Microbiol., vol. 9, no. MAR, pp. 1–14, 2018, doi: 10.3389/fmicb.2018.00437

[63] Z. Yang, "[Frontiers in Bioscience 8, d440-450, January 1, 2003] MOLECULAR EP-IDEMIOLOGY OF TUBERCULOSIS Zhenhua Yang," pp. 34–44, 2003.

[64] C. J. Baker, D. L. Kasper, and I. B. Tager, "Quantitative determination of antibody to capsular polysaccharide in infection with type III strains of group B Streptococcus," J. Clin. In-vest., vol. 59, no. 5, pp. 810–818, 1977, doi: 10.1172/JCI108703.

[65] V. A. Oxelius, V. Linden, K. Kvist Christensen, and P. Christensen, "Deficiency of IgG subclasses in mothers of infants with group B streptococcal septicemia," Int. Arch. Allergy Appl. Immunol., vol. 72, no. 3, pp. 249–252, 1983, doi: 10.1159/000234876.

[66] A. K. Rundgren, K. K. Christensen, and P. Christensen, "Increased frequency of high serum IgM among mothers of infants with neonatal group-B streptococcal septicemia1," Int. Arch. Allergy Immunol., vol. 77, no. 3, pp. 372–373, 1985, doi: 10.1159/000233847.

[67] K. K. Christensen, P. Christensen, G. Faxelius, J. Henrichsen, and F. K. Pedersen, "Immune response to pneumococcal vaccine in mothers to infants with group B streptococcal sep-ticemia: Evidence for a divergent IgG/IgM ratio1," Int. Arch. Allergy Immunol., vol. 76, no. 4, pp. 369–372, 1985, doi: 10.1159/000233723.

[68] L. A. Meyn, D. M. Moore, S. L. Hillier, and M. A. Krohn, "Association of sexual ac-tivity with colonization and vaginal acquisition of group B Streptococcus in nonpregnant women," Am. J. Epidemiol., vol. 155, no. 10, pp. 949–957, 2002, doi: 10.1093/aje/155.10.949.

[69] G. Teti, F. Tomasello, M. S. Chiofalo, G. Orefici, and P. Mastroeni, "Adherence of group B streptococci to adult and neonatal epithelial cells mediated by lipoteichoic acid," Infect. Immun., vol. 55, no. 12, pp. 3057–3064, 1987, doi: 10.1128/iai.55.12.3057-3064.1987.

[70] B. Spellerberg et al., "Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human laminin," Infect. Immun., vol. 67, no. 2, pp. 871–878, 1999, doi: 10.1128/iai.67.2.871-878.1999.

[71] C. Beckmann, J. D. Waggoner, T. O. Harris, G. S. Tamura, and C. E. Rubens, "Iden-tification of novel adhesins from group B streptococci by use of phage display reveals that C5a peptidase mediates fibronectin binding," Infect. Immun., vol. 70, no. 6, pp. 2869–2876, 2002, doi: 10.1128/IAI.70.6.2869-2876.2002.

[72] L. C. Madoff, S. Hori, J. L. Michel, C. J. Baker, and D. L. Kasper, "Phenotypic di-versity in the alpha C protein of group B streptococci," Infect. Immun., vol. 59, no. 8, pp. 2638–2644, 1991, doi: 10.1128/iai.59.8.2638-2644.1991.

[73] D. Latousakis, D. A. MacKenzie, A. Telatin, and N. Juge, "Serine-rich repeat pro-teins from gut microbes," Gut Microbes, vol. 11, no. 1, pp. 102–117, 2020, doi: 10.1080/19490976.2019.1602428.

[74] H. Yumoto et al., "The pathogenic factors from oral streptococci for systemic dis-eases," Int. J. Mol. Sci., vol. 20, no. 18, 2019, doi: 10.3390/ijms20184571

[75] H. S. Seo, Y. Q. Xiong, and P. M. Sullam, "Role of the Serine-Rich Surface Glyco-protein Srr1 of Streptococcus agalactiae in the Pathogenesis of Infective Endocarditis," PLoS One, vol. 8, no. 5, pp. 1–6, 2013, doi: 10.1371/journal.pone.0064204.

[76] T. R. Sheen, A. Jimenez, N. Y. Wang, A. Banerjee, N. M. van Sorge, and K. S. Doran, "Serine-rich repeat proteins and pili promote Streptococcus agalactiae colonization of the vaginal tract," J. Bacteriol., vol. 193, no. 24, pp. 6834–6842, 2011, doi: 10.1128/JB.00094-11.

[77] M. Y. Mistou, S. Dramsi, S. Brega, C. Poyart, and P. Trieu-Cuot, "Molecular dissec-tion of the secA2 locus of group B streptococcus reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence," J. Bacteriol., vol. 191, no. 13, pp. 4195–4206, 2009, doi: 10.1128/JB.01673-08.

[78] S. E. Park, S. Jiang, and M. R. Wessels, "CsrRS and environmental pH regulate group B Streptococcus adherence to human epithelial cells and extracellular matrix," Infect. Im-mun., vol. 80, no. 11, pp. 3975–3984, 2012, doi: 10.1128/IAI.00699-12.

[79] V. G. Dutra et al., "Streptococcus agalactiae in Brazil: Serotype distribution, viru-lence determinants and antimicrobial susceptibility," BMC Infect. Dis., vol. 14, no. 1, pp. 1–9, 2014, doi: 10.1186/1471-2334-14-323.

[80] S. L. Kolar et al., "Group B Streptococcus Evades Host Immunity by Degrading Hya-luronan," Cell Host Microbe, vol. 18, no. 6, pp. 694–704, 2015, doi: 10.1016/j.chom.2015.11.001.

[81] K. A. Patras et al., "Group BStreptococcusCovR regulation modulates host immune signalling pathways to promote vaginal colonization," Cell. Microbiol., vol. 15, no. 7, pp. 1154–1167, 2013, doi: 10.1111/cmi.12105.

[82] S. E. Park, S. Jiang, and M. R. Wessels, "CsrRS and environmental pH regulate group B Streptococcus adherence to human epithelial cells and extracellular matrix," Infect. Im-mun., vol. 80, no. 11, pp. 3975–3984, 2012, doi: 10.1128/IAI.00699-12.

[83] M. Buscetta et al., "FbsC, a novel fibrinogen-binding protein, promotes Streptococ-cus agalactiae-host cell interactions," J. Biol. Chem., vol. 289, no. 30, pp. 21003–21015, 2014, doi: 10.1074/jbc.M114.553073.
[83]

[84] S. Jiang and M. R. Wessels, "BsaB, a novel adherence factor of group B Strep-tococcus," Infect. Immun., vol. 82, no. 3, pp. 1007–1016, 2014, doi: 10.1128/IAI.01014-13.

[85] C. Faralla et al., "Analysis of two-component systems in group B streptococcus shows that RgfAC and the novel FspSR modulate virulence and bacterial fitness," MBio, vol. 5, no. 3, 2014, doi: 10.1128/mBio.00870-14.

[86] E. Boldenow et al., "Group B Streptococcus circumvents neutrophils and neutrophil extracellular traps during amniotic cavity invasion and preterm labor," Sci. Immunol., vol. 1, no. 4, 2016, doi: 10.1126/sciimmunol.aah4576.

[87] C. Whidbey et al., "A streptococcal lipid toxin induces membrane permeabilization and pyroptosis leading to fetal injury," EMBO Mol. Med., vol. 7, no. 4, pp. 488–505, 2015, doi: 10.15252/emmm.201404883.

[88] C. Whidbey et al., "A hemolytic pigment of Group B Streptococcus allows bacterial penetration of human placenta," J. Exp. Med., vol. 210, no. 6, pp. 1265–1281, 2013, doi: 10.1084/jem.20122753.

[89] A. J. Carey et al., "Infection and Cellular Defense Dynamics in a Novel 17beta-Estradiol Murine Model of Chronic Human Group B Streptococcus Genital Tract Colonization Reveal a Role for Hemolysin in Persistence and Neutrophil Accumulation," J. Immunol., vol. 192, no. 4, pp. 1718–1731, 2014, doi: 10.4049/jimmunol.1202811.

[90] A. Costa et al., "Activation of the NLRP3 Inflammasome by Group B Streptococci," *J. Immunol.*, vol. 188, no. 4, pp. 1953–1960, 2012, doi: 10.4049/jimmunol.1102543.

[91] M. Y. Mistou, S. Dramsi, S. Brega, C. Poyart, and P. Trieu-Cuot, "Molecular dissec-tion of the secA2 locus of group B streptococcus reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence," J. Bacteriol., vol. 191, no. 13, pp. 4195–4206, 2009, doi: 10.1128/JB.01673-08.

[92] T. R. Sheen, A. Jimenez, N. Y. Wang, A. Banerjee, N. M. van Sorge, and K. S. Doran, "Serine-rich repeat proteins and pili promote Streptococcus agalactiae colonization of the vaginal tract," J. Bacteriol., vol. 193, no. 24, pp. 6834–6842, 2011, doi: 10.1128/JB.00094-11.

[93] P. Date, "UC San Diego," 2004.

[94] M. Y. Mistou, S. Dramsi, S. Brega, C. Poyart, and P. Trieu-Cuot, "Molecular dissec-tion of the secA2 locus of group B streptococcus reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence," J. Bacteriol., vol. 191, no. 13, pp. 4195–4206, 2009, doi: 10.1128/JB.01673-08.

[95] N. D'Urzo et al., "Acidic pH strongly enhances in vitro biofilm formation by a sub-set of hypervirulent ST-17 Streptococcus agalactiae strains," Appl. Environ. Microbiol., vol. 80, no. 7, pp. 2176–2185, 2014, doi: 10.1128/AEM.03627-13.

[96] R. E. Parker, C. Laut, J. A. Gaddy, R. N. Zadoks, H. D. Davies, and S. D. Manning, "Association between genotypic diversity and biofilm production in group B Streptococcus," BMC Microbiol., vol. 16, no. 1, pp. 1–12, 2016, doi: 10.1186/s12866-016-0704-9.

[97] F. Di Xia, A. Mallet, E. Caliot, C. Gao, P. Trieu-Cuot, and S. Dramsi, "Capsular pol-ysaccharide of Group B Streptococcus mediates biofilm formation in the presence of human plas-ma," Microbes Infect., vol. 17, no. 1, pp. 71–76, 2015, doi: 10.1016/j.micinf.2014.10.007.

[98] T. R. Sheen, A. Jimenez, N. Y. Wang, A. Banerjee, N. M. van Sorge, and K. S. Doran, "Serine-rich repeat proteins and pili promote Streptococcus agalactiae colonization of the vaginal tract," J. Bacteriol., vol. 193, no. 24, pp. 6834–6842, 2011, doi: 10.1128/JB.00094-11.

[99] S. Landwehr-Kenzel and P. Henneke, "Interaction of Streptococcus agalactiae and cellular innate immunity in colonization and disease," Front. Immunol., vol. 5, no. OCT, pp. 1–11, 2014, doi: 10.3389/fimmu.2014.00519.

[100] M. Lahsoune, H. Boutayeb, K. Zerouali, H. Belabbes, and N. El Mdaghri, "Préva-lence et état de sensibilité aux antibiotiques d'Acinetobacter baumannii dans un CHU marocain," Med. Mal. Infect., vol. 37, no. 12, pp. 828–831, 2007, doi: 10.1016/j.medmal.2007.05.006.

[101] O. Olsvik et al., "Use of automated sequencing of polymerase chain reactiongener-atedamplicons to identify three types of cholera toxin subunit B in Vibrio cholerae O1strains," *J. Clin. Microbiol.*, vol. 31, no. 1, pp. 22–25, 1993, doi: 10.1128/jcm.31.1.2225.1993.

[102] M. D. Adams et al., "Complementary DNA sequencing: Expressed sequence tags and human genome project," Science (80-.)., vol. 252, no. 5013, pp. 1651–1656, 1991,doi: 10.1126/science.2047873.

[103] A. A. K. Khalil, W. S. Park, H. J. Kim, K. M. Akter, and M. J. Ahn, "Anti-helicobacter pylori compounds from polygonum cuspidatum," Nat. Prod. Sci., vol. 22, no. 3, pp. 220–224, 1182016, doi: 10.20307/nps.2016.22.3.220.

[104] E. M. A. Hons, L. Ac, D. H. Med, D. Chm, and K. Vall, "Naphthoquinone Phenols Mitochondria and tumor metabolic flexibility mechanisms and therapeutic perspective," 2018.

[105] A. Kacmaz et al., "Synthesis and antiproliferative evaluation of some 1,4-naphthoquinone derivatives against human cervical cancer cells," Open Chem., vol. 17, no. 1, pp. 337–345, 2019, doi: 10.1515/chem-2019-0030.

[106] E. P. Abraham, "The Antibiotics," Compr. Biochem., vol. 11, no. 4, pp. 181–224, 1963, doi: 10.1016/B978-1-4831-9711-1.50022-3.

[107] M. M. Gutacker et al., "Single-Nucleotide Polymorphism–Based Population Genet-icAnalysis of Mycobacterium tuberculosis Strains from 4 Geographic Sites," *J. In-fect.Dis.*, vol. 193, no. 1, pp. 121–128, Jan. 2006, doi: 10.1086/498574.
[108]

[108] M. W. Peck, Biology and Genomic Analysis of Clostridium botulinum, vol. 55, no. 09.Elsevier.

[109] L. Zhang et al., "The Core Proteome and Pan Proteome of Salmonella Paratyphi AEpidemic Strains," *PLoS One*, vol. 9, no. 2, p. e89197, Feb. 2014, doi:10.1371/journal.pone.0089197.

[110] L. Zhang et al., "The Core Proteome and Pan Proteome of Salmonella Paratyphi AEpidemic Strains," *PLoS One*, vol. 9, no. 2, p. e89197, Feb. 2014, doi:10.1371/journal.pone.0089197.

[111] L. Rouli, V. Merhej, P. E. Fournier, and D. Raoult, "The bacterial pangenome as a newtool for analysing pathogenic bacteria," *New Microbes New Infect.*, vol. 7, pp. 72–85,Sep. 2015, doi: 10.1016/j.nmni.2015.06.005.

[112] J. W. Sahl et al., "Evolution of a Pathogen: A Comparative Genomics AnalysisI-dentifies a Genetic Pathway to Pathogenesis in Acinetobacter," *PLoS One*, vol. 8, no.1, p. e54287, Jan. 2013, doi: 10.1371/journal.pone.0054287.

[113] D. Savoia, "Plant-derived antimicrobial compounds: alternatives to antibiotics," pp. 979– 990, 2012

[114] 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.

[115] A. M. Altenhoff, N. Skunca, N. Glover, C.M. Train, A. Sueki, I. Pilizota,K. Gori, B. Tomiczek, S. Muller, H. Redestig et al., "Theoma orthology database in 2015: function predictions, better plant support, synteny viewand other improvements," *Nucleic acids research*, vol. 43, no. D1, pp. D240–D249, 2015.

[116] H. Luo, Y. Lin, F. Gao, C.-T. Zhang, and R. Zhang, "Deg 10, an update ofthe data-base of essential genes that includes both protein-coding genes andnoncoding genomic elements," *Nucleic acids research*, vol. 42, no. D1, pp.D574–D580, 2014.

[117] M. M. Gutacker et al., "Single-Nucleotide Polymorphism–Based Population Genet-icAnalysis of Mycobacterium tuberculosis Strains from 4 Geographic Sites," *J. In-fect.Dis.*, vol. 193, no. 1, pp. 121–128, Jan. 2006, doi: 10.1086/498574.

[118] A. Jaiswal, A. Chhabra, U. Malhotra, S. Kohli, and V. Rani, "Comparative analysis of human matrix metalloproteinases: Emerging therapeutic targets in diseases," *Bioinformation*, vol. 6, pp. 23-30, 2011

[119] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic char-acter of a protein," *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, 1982, doi: 10.1016/00222836(82)90515-0.

[120] A. Altunkaya et al., "OUP accepted manuscript," *Nucleic Acids Res.*, vol. 45, no. October 2016, pp. 271–281, 2016, doi: 10.1093/nar/gkw1000.

[121] Protein complex prediction with AlphaFold-Multimer. Evans R, O'Neill M, Pritzel A, et al. bioRxiv 2021. doi: https://doi.org/10.1101/2021.10.04.463034

[122] L. RA and S. MB, "LigPlot+: multiple ligand-protein interaction diagrams for drug discov ery," *J. Chem.* Inf. Model., vol. 51, pp. 2778–2786, 2011.