CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Identification of Therapeutic Targets Against *Streptococcus gordonii* by Subtractive Genomic Analysis

by

Sadaf Kiani

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

2020

I dedicate this thesis to my parents and my teachers.

CERTIFICATE OF APPROVAL

Identification of Therapeutic Targets Against *Streptococcus gordonii* by Subtractive Genomic Analysis

by

Sadaf Kiani

(MBS183014)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Mazhar Qayyum	PMASAAU, RWP
(b)	Internal Examiner	Dr. Shoukat Iqbal Malik	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
September, 2020

Dr. Sahar Fazal
Head
Dept. of Biosciences & Bioinformatics
September, 2020

Dr. Muhammad Abdul Qadir
Dean
Faculty of Health & Life Sciences
September, 2020

Author's Declaration

I, **Sadaf Kiani** hereby state that my MS thesis titled "**Identification of Therapeutic Targets Against *Streptococcus gordonii* by Subtractive Genomic Analysis**" 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.

(Sadaf Kiani)

Registration No: MBS183014

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Identification of Therapeutic Targets Against *Streptococcus gordonii* by Subtractive Genomic Analysis**" 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.

(Sadaf Kiani)

Registration No: MBS183014

Acknowledgements

Abstract

Streptococcus gordonii belongs to viridans group streptococci, it is Gram-positive cocci. *Streptococcus gordonii*, are primarily involved in development of biofilms on tooth surfaces also known as dental plaque can result in development of infectious endocarditis i.e. inflammation of inner lining of heart. As treatment of this disease is very expensive and different bacteria has shown resistance against various commonly used antibiotics it is now important to find the novel therapeutic targets and drugs to cure and treat this disease. In this study we predict modelome of 11 strains of *Streptococcus gordonii* by using MHOLine workflow. 1,255 core proteins were identified by using pan genomic approach, later subtractive proteomics was used to extract set of 20 essential proteins for bacteria. Considering human as host, 15 of these proteins (Glutamine–fructose-6-phosphate aminotransferase, UDP-3-O-acylglucosamine N-acyltransferase, Ribonuclease P protein component, tRNA N6-adenosine threonylcarbamoyltransferase, Pseudouridine synthase, Bifunctional protein, AMP nucleosidase, Probable DNA-directed RNA polymerase subunit delta, ATP-dependent DNA helicase, Peptidase S24-like protein, TetR family transcriptional regulator, Chromosomal replication initiator protein, UDP-N-acetylmuramoyl-tripeptide–D-alanyl-D-alanine ligase, HTH-type transcriptional regulator, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate–2,6-diaminopimelate ligase) were considered as essential and non-host homologs, and have been subjected to virtual screening using ZINC library. The selected ligand molecules indicated positive interactions. Comprehensive development of novel drugs and vaccine of *Streptococcus gordonii* putative proteins is shown from which some have already been reported and validated in other species.

Keywords: *Streptococcus gordonii*, Infectious endocarditis, Pan-genomics, Subtractive proteomics

Contents

List of Figures

List of Tables

Abbreviations

Blast	Basic Local Alignment Search Tool
DEG	Database of Essential Genes
Edgar	Efficient Database Framework for Comparative Genome Analyses using BLAST Score Ratios
Kegg	Kyoto Encyclopedia of Genes and Genomes
MOE	Molecular Operating Environment
MW	Molecular Weight
S.gordonii	*Streptococcus gordonii*
Uniprot	Universal Protein Resource
Vfdb	Virulence factors Database

Chapter 1

Introduction

1.1   Background

Bacteria are commonly considered as Germ or cause of diseases in human beings, animals and plants. But on the other hand, various bacterial species are present on human skin, nasal passage and most commonly in mouth and gut, these species reside on human body for their whole lifetime and never cause diseases rather sometimes are considered quite beneficial to maintenance of health. Among these normal flora, imparting health benefits, some bacterial species referred as opportunistic pathogens can cause or mediate in onset of diseases. For example, streptococci among the normal flora of nose, mouth, upper airways and female genital tract, are pathogenic in nature but have very low virulence [1]. Streptococci is heterogeneous groups of organisms with complex taxonomy and belongs to Viridans group of Gram-positive cocci [1]. This group is categorized into six main groups, which includes the Streptococcus mutans group, Streptococcus mitis group, Streptococcus anginosus group, Streptococcus salivarius group, Streptococcus bovis group and *Streptococcus sanguinis* group [2, 3]. The *Streptococcus sanguinis* group is further subdivided into three groups *Streptococcus sanguinis*, *Streptococcus parasanguinis*, *Streptococcus gordonii* [3] and this subdivision is represented in figure 1.1.
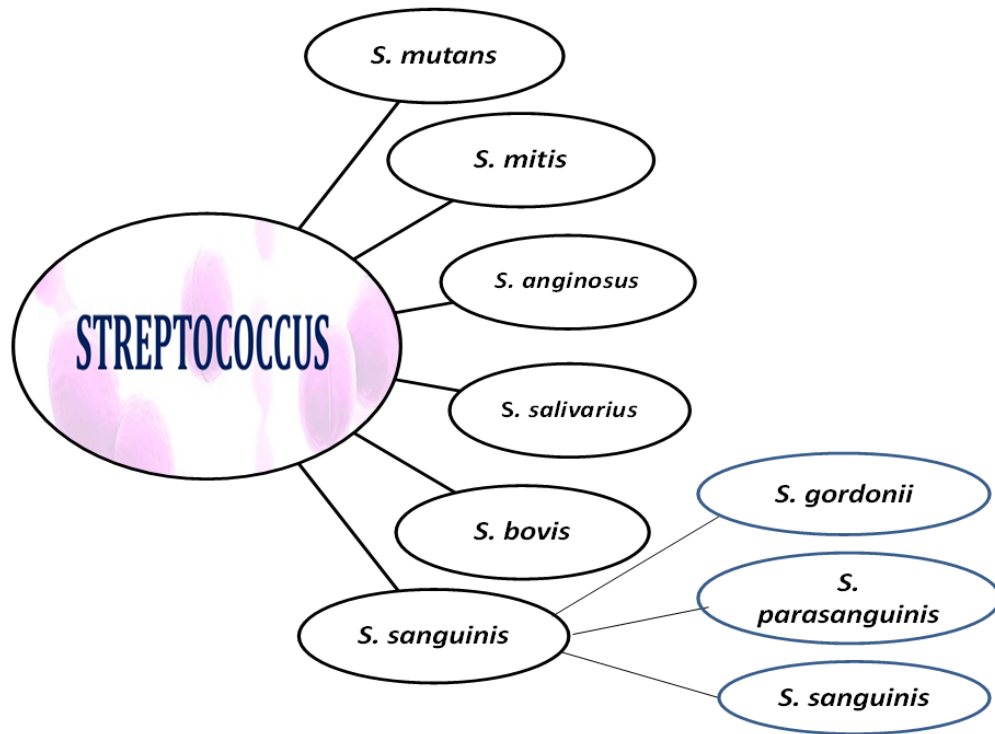
FIGURE 1.1: Subdivisions of Streptococcus group of Gram-Positive Bacteria

TABLE 1.1: Microbiology of Infectious Endocarditis [11].

Organism	Frequency (%)
Staphylococcus aureus	31
Viridans *Streptococcus*	17
Coagulase-negative staphylococci	11
Enterococci	11
Streptococcus bovis	7
Other streptococci	5
Fungi	2

Streptococcus gordonii, is a subdivision of *Streptococcus sanguinis*, and is a common inhabitant of mouth making biofilms on teeth. *Streptococcus gordonii* is basically a mesophilic, immotile cocci it grows in set of two or bead like chains [1] as depicted in figure 1.2. This bacterial specie got attention when it was found to cause infective endocarditis, where it was referred as that a dirty mouth can result into heart disease [4, 5], in addition to these *Streptococcus gordonii* was reported to be involved in onset of septic arthritis [6, 7] and spontaneous bacterial peritonitis [8, 9]. Although various other species are also reported to play their role in development of infective endocarditis, as summarized in Table 1.1.

FIGURE 1.2: Morphology of *Streptococcus gordonii* [10].

1.2   *Streptococcus gordonii* in Infective Endocarditis

Streptococcus gordonii, is a commensal (mutual benefit for host and bacteria) non-pathogenic bacteria among the normal flora of human oral microbiome. The specie is considered to play a significant role in initializing the formation of dental plaque or biofilm on teeth. It produces a surface where other bacteria can adhere. These dental plaques later cause dental cavities and other damages to teeth health. This bacterium can enter blood as a result of an injury or wound and then colonizes blood clotting agents such as platelet fibrin thrombi, and starts damaging endocardium and heat valves (Figure 1.3) Infective endocarditis could develop into a fatal disease, and this medical condition is often treated with antibacterial drugs [12]. The pathogenesis starts with the ability of *Streptococcus gordonii* to develop colonies very rapidly on the surfaces clean tooth. Human teeth are quite a reliable and promising site for biofilm formation as they are moist, warm, non-shedding
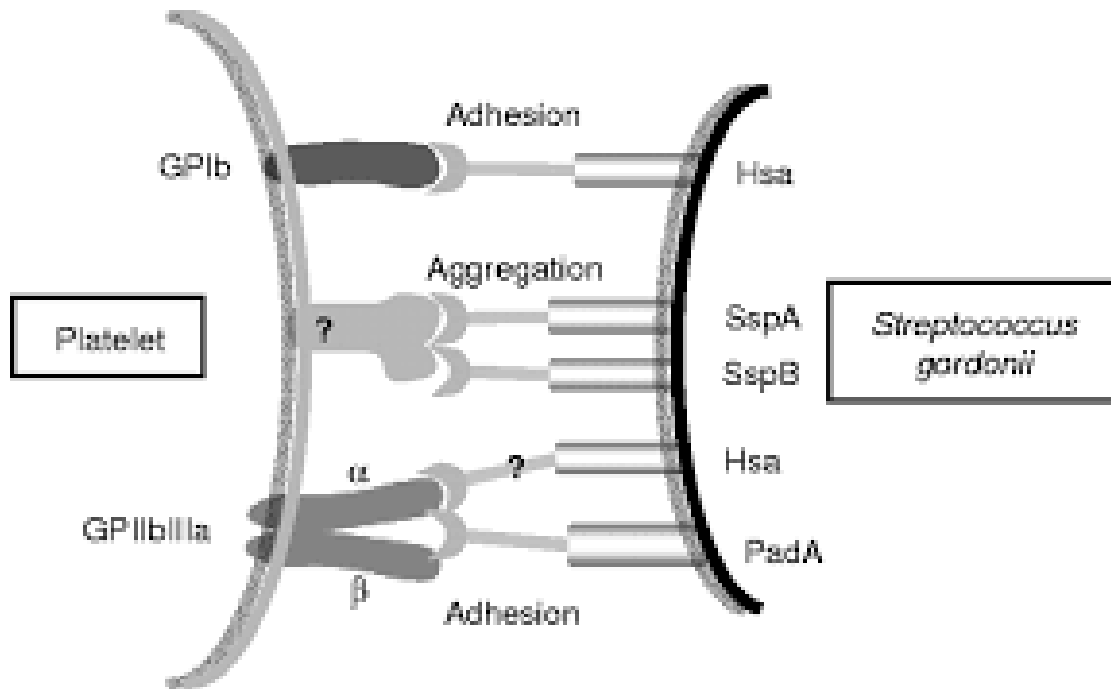
FIGURE 1.3: Interactions between *Streptococcus gordonii* and Human Platelets causing Thrombus Formation [13].

and mouth provides rich nutrition sources. This bacterial species itself is non damaging and just provides the adherence to other species which may be pathogenic. Although, the organism is not dangerous in the mouth area, but can cause acute endocarditis once reach blood stream as a result of trauma in oral region. Since *Streptococcus gordonii* is nonpathogenic but can adapt the pathogenicity of other secondary colonizers via mechanisms known as interspecies communication. The focus of research has been shifted to understand the mechanism of pathogenicity in opportunistic organisms such as streptococci which are a main cause of bacteremia inpatients whose immune system is compromised [14, 15]. Various strategies have been tried to control this species focusing especially on its ability to form biofilms on tooth surface and to decrease it is adhesion capabilities [16].

Streptococcus gordonii and *Streptococcus sanguinis* are among the initiators of biofilm formation and were also reported in bacterial spreads of infective endocarditis patients. Initially two species were considered one but later were identified as two separate species. Although these two species are considered safe but understanding the mechanism by which they can contribute to disease progression and how they could be controlled is need of time.

A fundamental interaction that causes pathogenesis of infective endocarditis is the one that bind bacteria directly to platelets (Figure 1.4). Due to *Streptococcus gordonii* Platelet binding by GspB emerges to be a main interaction in the pathogenesis of endocarditis. Endocarditis could be treated with antibiotic treatment and can cause mortality of host. There are few medication treatments for infective endocarditis but if the patient is allergic or resistant to these treatments than surgery is also recommended. Surgery is highly costly so we have to find novel therapeutic and drug targets to prevent the disease.

GPIbα	αIIbβ3 (direct)	Ganglioside GD3	IgG - FcγRIIA	Fibrinogen - αIIbβ3 (indirect)	?	Unknown receptors
S. gordonii (Hsa/GspB) *S. sanguinis* (SrpA)	*S. gordonii* (PadA)	*S. mitis* (PblA, PblB)	*S. sanguinis* *S. mitis* *S. mutans* *S. pyogenes* *S. agalactiae* *S. oralis*	*S. agalactiae* (FbsA) *S. mitis* (Lysin_{SM}) *S. pyogenes* (M protein)	*S. mitis* (sm-hPAF/ lectinolysin)	*S. sanguinis* (PAAP) *S. gordonii* (SspA/SspB) *S. mutans* (PAc,RGPs)

FIGURE 1.4: Platelet-Streptococcus interactions in Pathogenisis of Infective Endocarditis [17].

1.3 Therapeutic Target Identification

Advent of bioinformatics and Computational techniques such as comparative genomics, reverse vaccinology, Pan genomics and subtractive genomics have facilitated scientists to explore various genetic components and understand their significance in health and disease [18, 19]. Therapeutic target are basically nucleic acid or proteins whose function or activity can be altered by using external stimulation. Drug discovery is fundamentally laid on the exploration and consequent testing of drug candidates acting as therapeutic target. Advancement in genomics,

proteomics, and disease mechanisms has led to a develop concentration in finding novel targets and efficient exploration of existing targets. Knowledge of these targets is supportive for mechanism of action of drugs and for predicting features that helps to guide new drug design.

Methods such as subtractive and comparative microbial genomics and differential genome analysis [20] are also being apply for the identification of targets in a many human pathogens like *Mycobacterium tuberculosis* [21], *Burkholderia pseudomallei* [22], *Helicobacter pylori* [23], *Pseudomonas aeruginosa* [24], *Neisseria gonorrhea* [25] and *Salmonella typhi* [26]. The main principle of these methods is to hit upon targeted genes/proteins that are crucial and necessary for the organism that cause disease and are non-homologous to the host [14], so that drugs targeting these "pathogen-essential non-host homologs" can be applied with little off targets in the host. Several essential proteins for pathogen might have a little homology to the proteins of host. On the other hand, these proteins may still choose as potential molecular targets for structure-based specific inhibitor development. Important dissimilarity in the active sites or in other druggable pockets may present, so that the pathogenic protein possibly will be targeted [15, 16].

The procedure to identify the direct molecular target, for instance protein or nucleic acid, of a small molecule is known as target identification. In medical pharmacology, target identification is proposed intended to find the effective target of a pharmaceutical or other xenobiotic. In the early hours drug discovery begin with preliminary steps of target identification and progress to the afterward phases of development. Target identification and classification starts with identification of the function of a potential therapeutic target (gene/protein) and its aim in the disease. Identification of the target is subsequently characterization of the molecular mechanisms deal with the target. An excellent target must be effective, secure, meet clinical and business necessities and must be 'druggable'. Wet lab strategies of therapeutic target identification are time consuming, extensive and elaborate so to save the time we use insilico or computational approach for the identification of therapeutic target identification.

1.4   Subtractive Genomics

At this time the drugs used in treatment of diseases caused by pathogenic organisms demonstrate little to greater side effects in patients and there is wide distressing mount in the development of drug resistance strains.

Hence there is a requirement for identifying novel and effective drugs to fight with the diseases. A contemporary method known as "Subtractive genomics" is now extensively engaged to identify new and exact drug targets in organism that cause pathogenicity, as a footstep to recognizing new and potential drugs.

The process in which the host and the pathogen proteome are subtracted that's assists in giving information for a protein set that is essential to pathogen but are not present in the host. Subtractive genomics plays a significant role in identification of potential drug targets. These targets are the proteins which are considered necessary for the organisms to survive [2]. Subtractive genome study rely on cataloging the essential proteins for a pathogen as unique in order to make possible particular drug designing (Figure 1.5).

Subtractive genomics method useful in 3 types of cases: Multi-Drug resistant pathogenic organism, Pathogenic organisms with No effective medications available, Pathogenic organism with no virulence factor identified.

This study is designed to use in-silico approaches to find the genomic information of *Streptococcus gordnii* with its proteomics to identify potent drug targets. Proteomes of 11 complete genomes which were already present were comparing using core genome approach.

Core genome includes genes present in all strains of that specie then essential proteins of organism are filtered out from the core genome. Then subtractive genomic approach is use in which all essential proteins are further check for non-homologous to the human host. The putative targets that will identified might be used to design vaccines and propose novel drugs that can bind to the proposed target protein [27].
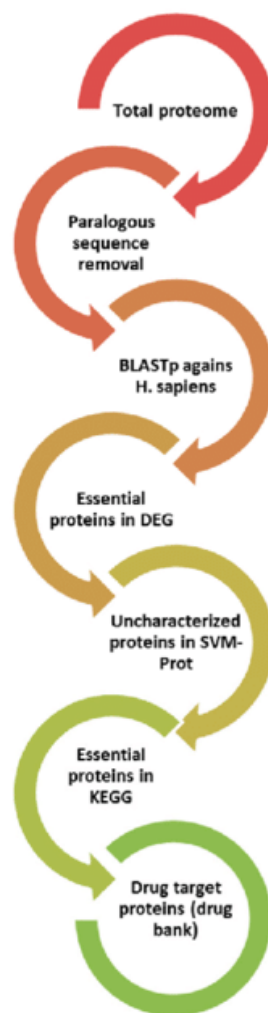
FIGURE 1.5: Flow chat of Subtractive Genomics [50].

1.5 Aims and Objectives

Streptococcus gordonii causes dental plaque and infective endocarditis which is disease of inner lining of heart. For many years this disease is being treated with antibodies but recent research on this disease have shown antibiotic resistance against the strains of *Streptococcus gordonii*. Bioinformatics, in general, contributes through prediction of therapeutic targets which ultimately reduce men efforts and cost of experimentation. So, in this study we will contribute towards drug development against endocarditis disease by predicting novel therapeutic targets and potent lead compound for inhibition of identified targets. The promising ligand molecule can be tested in experimental laboratory that can ultimately result in commercial product in future. In this study, we use an integrative insilico

method for the predictive proteome of *Streptococcus gordonii* species to correlate the information about genome with the identification of putative therapeutic targets based on their 3D structure.Theproteomes predicted from the 11 strains of *Streptococcus gordonii* were modeled by using the MHOLline workflow that is proposed by Hassanetal, 2014 [26]. There is a need of alternative novel targets and potent therapeutics to prevent the onset of the disease. The aim of the study is to identify new novel and potent therapeutics targets to prevent the onset of the infective endocarditis disease by using pan genomic approach.

1.6   Objectives

So, the objectives of this study include:

1. To identify core genomes of all strains of *Streptococcus gordonii*

2. To identify Essential proteins for bacteria using subtractive genomic.

3. To prioritize our protein targets, identification of potent lead compound using protein-ligand interaction.

Chapter 2

Literature Review

Many bacteria are present on human skin, in nose area and most commonly in mouth area and gut in normal condition, sometimes lives throughout the life of human being and does not cause any disease but when they get chance or interact with other bacteria can cause diseases such type of bacteria are known as opportunistic bacteria. The streptococci are pathogenic in nature but have very low virulence which is normally present in the mouth area and upper airways, and the female genital tract [1]. *Viridans streptococci* including *Streptococcus gordonii*, is primary bacteria which is the main cause of development of biofilms on surface of tooth known as dental plaque and ultimately cause infectious endocarditis. In some studies it was revealed that *Streptococcus gordonii*is vital cause of infective endocarditis in immunocompetent patients who have undergone a dental procedure. *Streptococcus gordonii* is basically a mesophilic, immotile cocci it grows in couples or bead like chains and comprises of heterogeneous groups of organisms with vary complex taxonomy [1].

Infective endocarditis is a disease of inner lining of heart. This disease is more common in man as compare to women. This disease is very prevalent in western countries as well as in Asian countries .In western countries this disease is most prevalent in elder people but in Asia it is more prevalent to young people age ranges from 34 to 40 years. This difference is mainly due to rheumatic heart disease is more common in Asian countries. In Pakistan the person who is suffering from

rheumatic heart disease is likely to have infective endocarditis. The main issue in treating the disease is the antibiotic resistance so in this study pan-genomic approach will be used to find the new and common drug targets for all strains of *Streptococcus gordonii* to overcome antibiotic resistance.

2.1 *Streptococcus gordonii*

Streptococcus gordonii belongs to group of Gram-positive bacterium integrated along with several earlier colonizers of the periodontal environment.[28] The organism, associated to oral streptococci, has an elevated resemblance for molecules in the salivary pellicle (or coating) on surface of tooth. *Streptococcus gordonii* consequently can swiftly colonize on the hygienic surface of tooth, and *Streptococcus gordonii* along with associated organisms consist of a high percentage i.e. up to 70%, of the bacterial biofilm forms on hygienic surface of tooth. Usually risk-free in mouth area, *Streptococcus gordonii* can cause acute bacterial endocarditis upon gaining admittance systemically. *Streptococcus gordonii* also forms an accessory layer for other colonizers of tooth surface and can modulate the pathogenicity of these secondary colonizers via a mechanism known as interspecies communication.

DNA Data Bank of Japan, European Nucleotide Archive and GenBank in 2016 deposit the whole genome sequence of *Streptococcus gordonii* [30].

2.1.1 Significance

The Streptococci belongs to group of gram positive, immoveable cocci that develop in the foam of pairs or bead like strings. Bacteria contained by the genus consist of both type's i.e. pathogenic bacteria and non-pathogenic bacteria that live in the mouth area, skin and upper respiratory tract of host i.e. human being and it contain *Streptococcus gordonii* and *Streptococcus mutans* [31]. *Streptococcus gordonii* is member of the viridans group of Strepotococci that are fundamental
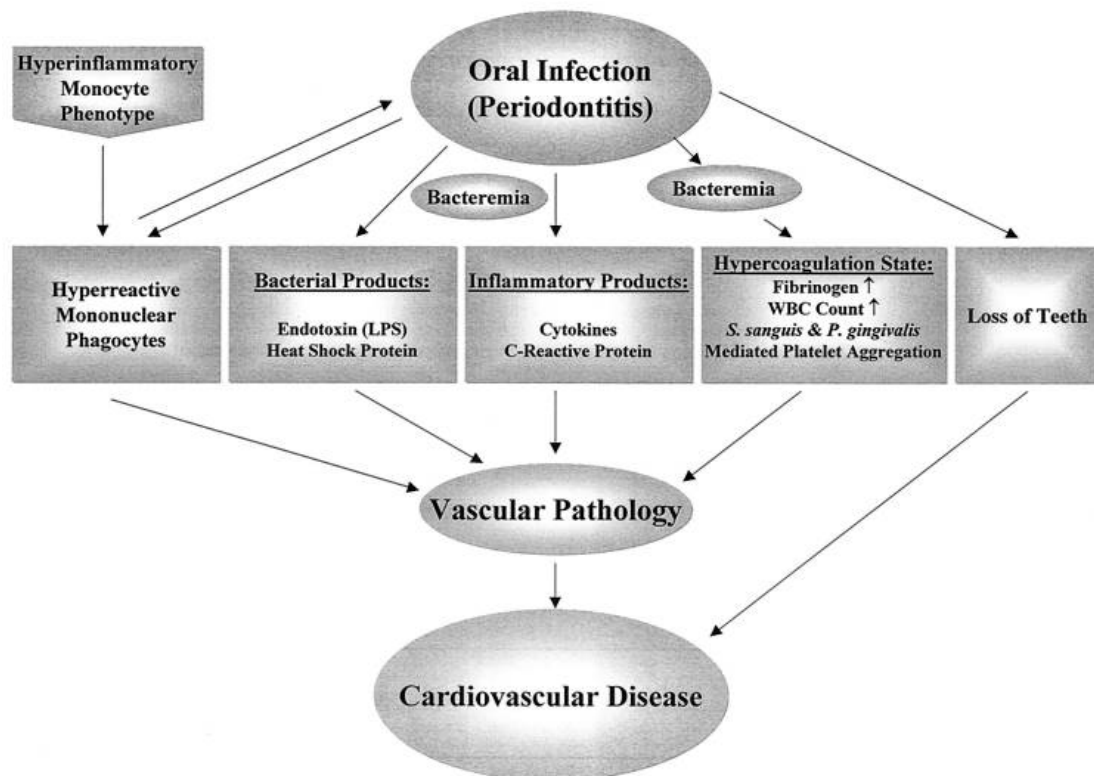
FIGURE 2.1: Proposed Mechanism linking Oral Infection to Cardiovascular Disease [29].

part of the human oral flora. These bacteria inhabit surface of tooth by making biofilms, also renowned as dental plaque. Ultimately dental plaque leads to periodontal infection and dental cavities both are the most widespread diseases in developed countries [32] (Figure 2.1). Mostly after the oral treatment *Streptococcus gordonii* is vital cause of bacterial endocarditis by entering the blood stream of host. *Streptococcus gordonii* colonizes platelet-fibrin thrombi that clot blood, in injured valves of the heart that's lead to damage function of the valves of heart. This disease can mainly treat by antibiotic therapy if not treated on time it might leads to death of host [12].

2.1.2 Genome Structure

Even though the full genome of *Streptococcus gordonii* has not so far been determined, but several appealing regions of chromosomal have been illustrated. Glucosyltransferase enzymes synthesize glucan polymers that are crucial for attachment

to the tooth.Parallel determinants to rgg have been illustrated in other associated bacteria present in mouth. Rgg-like determinants also present in other streptococci species which helps in regulationof different proteins with various functions. The data shows that rgg-like genes plays essential role as regulatory determinants and the function of rgg can be determined after the complete genomic sequencing of *Streptococcus gordonii* [36].

2.1.3   Structure of Cell and Metabolism

An individual oral cavity gives some degree of and unreliable source of diet for microorganisms live in the oral micro flora [33]. As an energy source oral *streptococci*, counting *Streptococcus gordonii*, depend on sugars that are producedprimarily from carbohydrates. A key constituent in the human diet i.e. Fructose, can be attained by means of glucosyltransferases and from fructans through fructanases. Oral streptococci depend chiefly upon the phosphotransferase system (PTS) to transport carbohydrates via phosphorylation and translocation by means of a membrane. [33].

The extracellular matrix of mammalscontains a lot of glycosaminoglycans containing recurring beta-linked dissaccharide units. Beta-linked disaccharides are released when glycosaminoglycans are degraded. These beta-glucoside sugar substrates, including "cellobiose, arbutin, salicin and esculin," are fermented by *Streptococus gordonii* to produce energy [32]. Numerous putative regulons include these genes that encodes for enzymes which are essential for the process of metabolism.

2.1.4   Ecology

For the biofilm expansion the teeth of human beings provides an ideal environment as it is damp and lukewarm place [32]. Biofilm development starts with pioneer organisms that adhere to the surface of tooth in the human mouth area during dental plaque development. One of these pioneer organisms, *Streptococcus gordonii* is main bacteria which start colonization and help out the more colonization of

different organisms by building up a biofilm on which microscopic organisms may stick [32]. In a modern research it was revealed that *Streptococcus gordonii* was main organism that contains necessary genes that assist the addition of drifting *P. gingivalis* cells into the initial stages of a functioning biofilm [32]. Ecology plaque hypothesis is depicted in figure 2.2 Primarily, Streptococcus gordonii starts
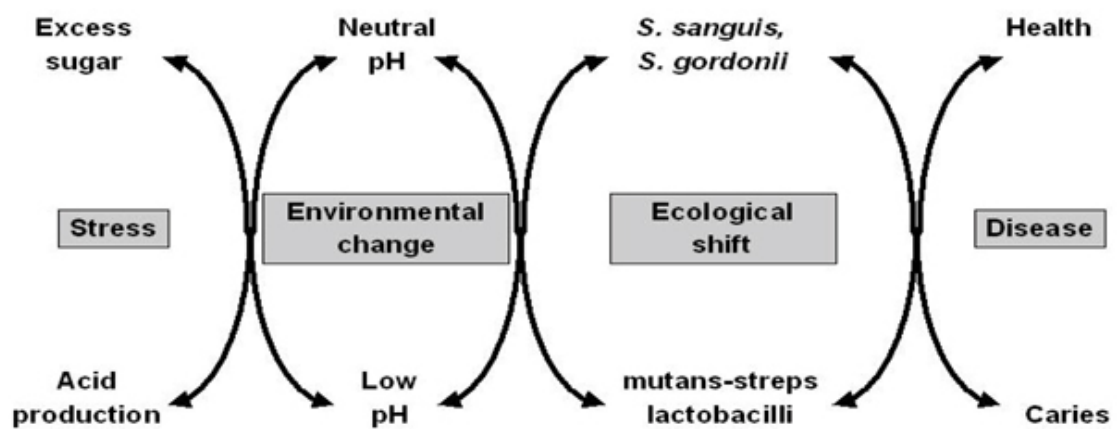
FIGURE 2.2: Ecological plaque hypothesis [34].

colonization by means of arrangement of a monospecies biofilm. The surface of human tooth is coated by pellicle containing lipids and proteins which also include salivary agglutinin glycoprotein (Figure 2.2). In *Streptococcus gordonii* and other colonizers the receptors for salivary agglutin glycoprotein and bind to the pellicle [32]. Cells of *Streptococcus gordonii* , bind to the tooth surface, then starts a signal transduction pathway, which is known as BrfAB, that regulates adhesive activity. The *Streptococcus gordonii* biofilm then play an important roleas a binding site for adhesion of some additional pathogenic organisms such as *Porphyromonas gingivalis* and the process is known as coaggregation. In coaggregation specific bacteria become internally connected by specific attachments [32].

2.1.5 Pathology

Despite the fact that *Streptococcus gordonii* is main cause of dental plaque and establish the colonies of other pathogenic microorganisms on the surface of tooth, it is not legitimately pathogenic in the mouth. Alternatively, if *Streptococcus gordonii* once enters the circulatory system through oral bleeding it can colonize

injured heart valves so that cause endocarditis in humans beings [35]. On the damaged heart valves and endocardium blood platelets, cell fragments that assist the clotting of blood, attach to fibrinogen, the inner lining of heart, and then makes platelet-fibrin thrombi. These platelet-fibrin thrombi can become colonized by *Streptococcus gordonii* which cause injury to the valves of heart and function of the heart [35]. Figure 2.3 shows Casual Model of Dental Diseases associated with Bacterial Endocarditis.
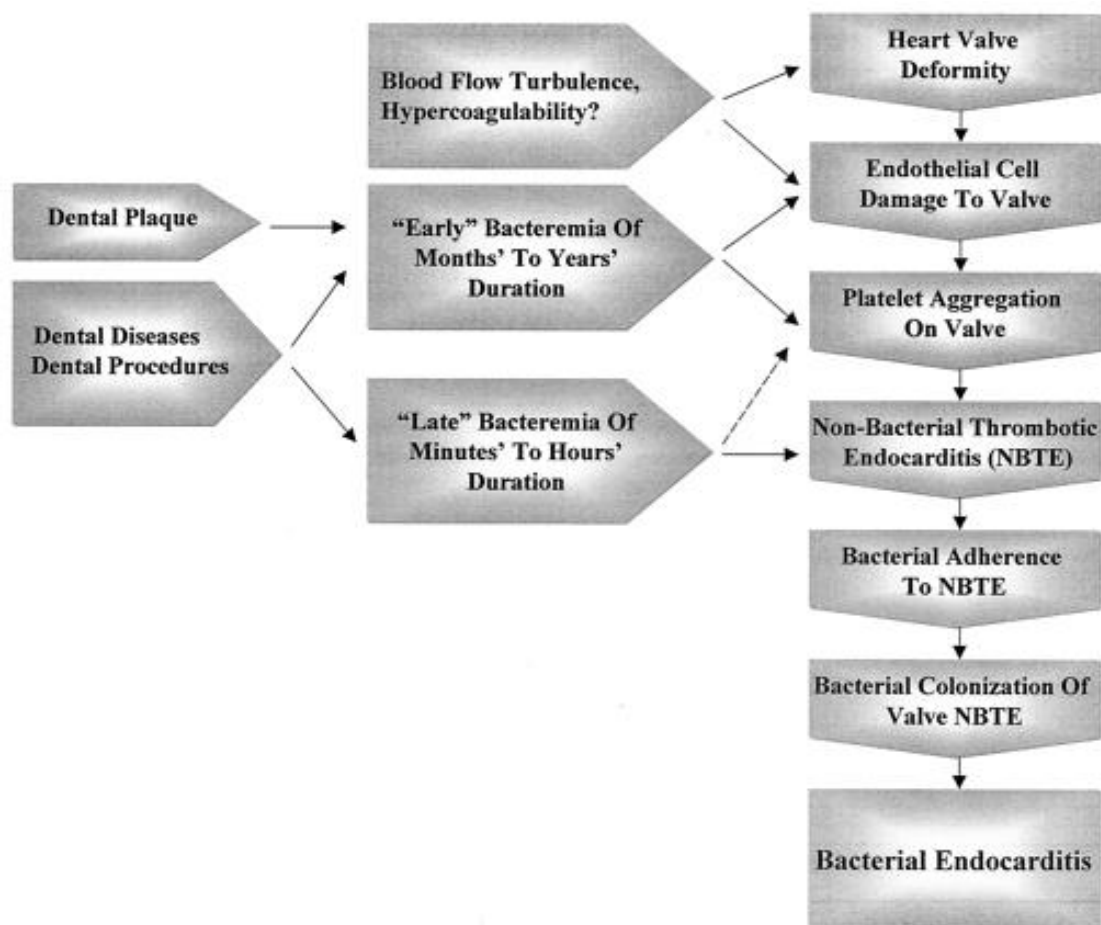
FIGURE 2.3: Casual Model of Dental Diseases Associated with Bacterial Endocarditis [36].

During bleeding of oral pit more than 700 species of bacteria may enter to the circulatory system regardless of all that streptococci are the most well-known and significant reason of infective endocarditis [32]. *Streptococcus sanguis*, *Streptococcus oralis*, and *Streptococcus gordonii* are the three most promising pathogens that lead to enodcarditis. It is inquiring that oral streptococci are competent in binding

to blood platelets particularly as the blood is not their normal habitation. Due to adaptation in one natural habitat during evolution it has allow *Streptococcus gordonii* and other oral streptococci to occupy an additional habitat [33].

Humans being with most probably having artificial heart valves develop bacterial endocarditis, heart disorders [12]. The chances of endocarditis will increased bydental procedure, urologic or gynecologic procedure and by use of intravenous medication. Various indications of bacterial endocarditis incorporates:Tiredness, Loss of hunger, Chilliness, Headaches, Joint inconvenience, The Treatment of this disease consists of intravenous anti-microbial treatment and occasionally antibiotics for a half a month [12].

2.2 Infective Endocarditis

Infective endocarditis is a disease of the endocardium. The endocardium is the coating of the inner layer of the chambers of the heart. This situation is commonly brought about bacteria that enter the circulatory system and infect the heart. Microbes might originate in the: mouth, skin, digestive organs, respiratory framwork, and urinary tract. At the point when this situation is brought about by bacteria, it is also termed as bacterial endocarditis. In unusual cases, it may be brought about by the growth of fungus or other microorganisms. On time medical treatmentis needed for infective endocarditis as it is a serious condition. If don't treated on time, the infection can injure your heart valves. This can cause issues like:Stroke, damage to different organs death.

People with healthy hearts mostly do not have this condition. Person with other heart problemsis at elevated risk. If someone is at elevated risk for infective endocarditis, may require taking antibiotics before any medical and dental treatment. Antibiotics facilitate to stop bacteria to enter bloodstream and cause infection. The number of people affected is about 5 per 100,000 per year. Rates, still, differ from region to region throughout the world [36]. Mostly males are affected as compare to females[38]. The risk of death among those infected is about 25% [37].

But if remain untreated it is nearly universally fatal [38]. Symptoms of this disease may differ from mam to man. In a few people, symptoms develop slowly, while others may develop symptoms suddenly. Consult to the physician if someone experiences any of the symptoms given below. People at elevated risk of endocarditis must take proper care.

The infection can cause changeless harm to the heart. If it is not fixed and treated on time, it may leads to death. We shall require having treatment in a hospital to prevent the infection from getting worst condition and causing complications. In the hospital, patient's fundamental symptoms will be observed. Patient will be treated with intravenous antibiotics (IV). When patient go to home, patient will have to use oral IV antibiotics for at least a month. At some stage in this time, patient must be keep visiting to physician. Proper blood tests after a regular period of time will show that the infection is going away.

If the heart valves of the patient have been damaged he may require surgery. Surgeon may suggest fixing the valve of heart or the valve of a patient may be changed by new valve that is made up of animal tissue or any artificial materials. Surgery might also be compulsory if the antibiotics do not work or if the infection is fungal. Infective endocarditis (IE) is a rare but distressing illness faced by both developing and developed countries

2.3   Prevalence of Infective Fndocarditis in Pakistan

Infective endocarditis is a vital cause of morbidity and mortality in Pakistan. In our country infective endocarditis occurs at early age. Many of the patients suffering from this disease have age less than 40. In 2004 Tariq et al. reported that median age of infective endocarditis was 24 years [40, 41] and in 2015 same group reported a shift in the median age which was 34 years. This gradual shift could be explained due to shift from communicable to non-communicable disease.

The non-communicable disease includes cardiovascular disease, diabetes, cancers and chronic airways diseases. Proper medical care is required in managing this disease [42]. these results are quite different from developed countries, as in developed countries the age of infective endocarditis is greater than 50 [43], in our country the infective endocarditis occurring at the low age is due to high frequency of rheumatic heart disease, multifactorial diseases and unrepaired congenital heart disease. The study have shown that male are more effected than female from infected endocarditis [44]. The ratio of infective endocarditis in men and women is 2:1 respectively [41].

The high prevalence of this disease in men is because men has more health check so they are exposed to nosocomial infections and intravenous drug usage is also very frequent in males [42]. Another reason for this was explained by Durente et al. That male majority of infective endocarditis decreases with the age and female harmones play a protective role against infective endocarditis [43].

That's why it is more common un men as compare to women. From the blood and tissue culture of infected patients the most frequently organism isolated is streptococcus group 36% in which about 14.2% are susceptible to penicillin [42]. The results show that in Pakistan the frequent group causing infective endocarditis is streptococcus group.

2.4   Pan Genome Analysis

The word "pan-genome" reflects the entire number of non-repetitive genes that are available in a given dataset. It incorporates: Core genome, Accessory genome ,species-specific or strain-specific genes [45].

It provides the genomic diversity present between the strains of a distinct species [46]. Here are some tools mentioned in table 2.1 that are available for pan-genome analysis. In this study EDGAR tool will be used. It is the software which sustains a fast and easy overview of transformative connections between genomes of microbes

TABLE 2.1: List of Some Available Tool for Pan Genome Analysis Along with Their Functions.

Tool	URL	Function	Ref
EDGAR	edgar.computational. bio.uni-giessen.de	It performs homology analyses based on a particular cutoff that'sconsequently balanced to the query data.	[47]
PGAT (Prokaryotic Genome Analysis Tool)	nwrce.org/pgat	It performs comparison between different strains of the same species, to anticipate genetic differences.	[45]
PGAP – Pan-genome Analysis Pipeline	http:// pgap.sf.net	It could be a partitioned instrument utilized for the pan-genome examination, hereditary variety, advancement and work investigation of gene clusters.	[45]
PanGP: Quickly Analyzing Bacterial Pan-genome Profiles	http://PanGP.big.ac.cn	It could be a isolated apparatus that was created to carry out pan-genome examination for large-scale strains withan extremely low time cost.	[45]
GET_HOMOLOGUES	http://www.eead.csic.es/co mpbio/soft/gethoms.php	It was created to perform pan-genome and comparative -genomic examination of bacterial strains	[45]
Panseq – Pan-genome Sequence Analysis Program	http://76.70.11.198/panseq	This tool provides data of the core and accessory genome based on the sequence identity and segmentation length.	[45]
OrthoMCL	http://www.cbil.upenn.edu	This provides a scalable technique for constructing orthologous groups across multiple eukaryotic taxa, using a Markov Cluster algorithm to group (orthologs and paralogs).	[45]

and makes simpler the procedure of obtaining new biological knowledge into their differential gene content, which permits simple browsing of precompiled datasets [47].

2.5   Subtractive Genomic Analysis

Subtractive genomics is the mechanism by which sequences between the host and the pathogen proteome are subtracted, which helps to provide data for a collection of proteins that are essential for pathogen but not present in the host [48]. Essential genes are the gene that helps an organism to survive.

Removal of these genes causes cell death which indicates that these genes perform necessary biological function [49]. Table 2.2 shows the some data Databases and Tools Used for Subtractive Genomics. In this study essential genes are identified by DEG database (http://origin. tubic.org /deg/public /index.php).

The next step is to identify non-homologous proteins. Non-homologous proteins are identified by using Blastp tool (https:// blast .ncbi.nlm.nih.gov /Blast.cgi ?PAGE= Proteins) which helps us to filter out those essential proteins which are not present in host (human).

2.6   Drug Target Prioritization

Several factors are required for the drug target prioritization. Molecular weight of eseential and non-homologous targets will be identified by ProtParamtool for determining the Molecular function and biological process of target proteins Uniprot tool is used (https://www.uniprot.org/). For cellular localization of protein target CELLO online tool (http://cello.life.nctu.edu.tw/) will be used. For pathway analysis KEGG is used which will identify the role of the protein target in various cellular pathways. To check putative targets are involved in the pathogen virulence VFDB is used.

TABLE 2.2: List of Databases and Tools for Subtractive Genomics Along with their Functions

Database/tools	URL	Function	References
NCBI bacterial genomes Recourse of bacterial	http://www.ncbi.nlm.nih.gov/genomes/ genlist.cgi?taxid 5 2&type 5 0& name 5 Complete%20 Bacteria	Recourse of bacterial genomes	
GOLD: Genomes	http://www.genomesonline.org/	Recourse of genome projects	Bernal et al. [2001]
Swiss-port	Proteome database http://www.expasy.org/sprot/	Proteome database	Bairoch & Apweiler [1997]
Database of Essential Genes (DEG)	http://origin.tubic.org/deg/public/index.php	Database of Screening of essential genes	Zhang et al. [2004]
Kyoto Encyclopedia of Genes and Genomes	http://www.genome.jp/kegg/	Pathway comparison and subtraction	Kyoto Encyclopedia of Genes Ogata et al. [1999]
Genomic Target Database (GTD)	www.iioab.webs.com/GTD.htm	Genomic Target Database List of bacteria targets based on subtractive genomics	Barh et al. [2009]
Virulence Factors of Pathogenic Bacteria Database (VFDB)	http://www.mgc.ac.cn/VFs/main.htm	Resource of virulence factors of various medically significant bacterial pathogens	Chen et al. [2005]
CELLO	http://cello.life.nctu.edu.tw/	Subcellular localization prediction for bacteria and eukaryotes	Yu et al. [2004]
PSORTb	http://www.psort.org/psortb/	Subcellular localization prediction for gram-negative and gram-positive bacteria proteins	Gardy et al. [2005]
NCBI human BLAST	http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid 5 9606	Subtraction of non-human homologue genes	Altschul et al. [1990]

2.7 Catalytic Pocket Detection and Molecular Docking

For the catalytic pocket detection DoGSiteScorer is used. This tool is automated pocket detection and analysis tool which is used for calculation of draggability of protein cavities. This tool will return the pocket residue and draggability score which ranges from 0-1. The score nearer to 1 indicates extremely druggable protein cavity. The predicted cavities are possibly bind to ligands with high affinity [53]. For molecular docking MOE tool is used. It has three main steps retrieval of ligands, protein preparation and molecular docking.

2.7.1 Retrieval of Ligands

The inhibitors will be retrieved by literature review. The structures of these inhibitors will be built utilizing MOE-Builder tool (Emerson Electric Co, St. Louis, MO). The related 3D structures will be displayed and partial charges will be determined utilizing MOE. The energies of all the molecules that are identified will be minimized by means of the algorithm of energy minimization. The accompanying parameters will be utilized for energy minimization; gradient: 0.05, Force Field: MMFF94X, Chiral Constraint: Current Geometry. All the minimized molecules will be saved in the (mdb) file format. In the subsequent stage, the prepared ligands will be utilized as input files for MOE-Dock [27].

2.7.2 Protein Preparation

The protein molecule considered in our investigation will be retrieved from Protein Data Bank. Water molecules will be dispensed and the 3D protonation of the molecule was done by mean of MOE. The energy of the protein molecule will be minimized utilizing the energy minimization algorithm of MOE tool. The accompanying parameters will be used for energy minimization; gradient: 0.05,

force field: MMFF94X+dolvation, chiral constraint: current geometry. Energy minimization will be terminated when the root mean square gradient falls below 0.05. The minimized structure will be used as the template for docking [31].

2.7.3 Molecular Docking

The binding of the ligands into the binding pocket of protein molecule will be calculated by MOE-Dock implemented in MOE. In the process of Molecular mostly used the default parameters. After performing docking, we will examine the best poses for hydrogen bonding/$\pi - \pi$ interactions and root mean-square deviation (RMSD) calculation using MOE applications[30].

Chapter 3

Material and Methods

To identify therapeutic targets in Streptococcus gordonii methodology present in figure 3.1 was used the detail of each step are as follow:

3.1   Genome Selection

All the sequences of gene and protein of this bacterium are retrieved from NCBI (https://www.ncbi.nlm.nih.gov/genome/). Genome Assembly and Annotation report shows 53 strains of *Streptococcus gordonii* out of which 37 are contig, 5 are scaffold and 11 are complete. The eleven strains of *Streptococcus gordonii* are included in this study these strains have complete sequence.

3.2   Identification of Core Genomes

The word "pan-genome" reflects the entire number of non-repetitive genes that are available in a given dataset. The core genome are identified using EDGAR software [47]. In this, from all, one strain is selected as a reference strain which is used as a template strain and all other the strains are compared with the reference strains and then only those core genomes are selected which are common in all the

strains. The algorithm that it used was BLASTp with the standard scoring matrix BLOSUM62 and cut off value of E= 1x 10-5 [52].

3.3   Identification of Essential Genes

Essential genes are those genes which are vital for the species to survive or involve in growth. Subtractive genomics approach will be used to select conserved target which are essential to *Streptococcus gordonii.* The list of conserved proteins of *S. gordonii* which are retrieved from EDGAR software will be given to the database of essential genes: DEG (http://origin.tubic.org/deg/public/index.php).

The DEG contains tentativelyapproved evidence from bacteria and other organisms from currently available essential genomic elements for instance protein-coding genes and non-coding RNAs. Essential genes for a bacterium form a minimal genome, i.e. a series of functional modules with main roles in the emerging synthetic biology field. The default parameters will be selected, E-value=0.0001, bit score $\geq$100, scoring matrix BLOSUM62 and identity $\geq$25% [53].

3.4   Identification of Non-Homologous and Homologous Protein

Subtractive genomics is the mechanism by which sequences between the host and the pathogen proteome are subtracted, which helps to provide data of proteins that are important for pathogen but not present in the host [12]. The identification of non-homologous and homologous protein will be carried out using NCBI BLASTp, default parameter are; E-value=0.0001, bit score $\geq$100, scoring matrix BLOSUM62 and identity $\geq$25% [51].

Subtractive genomics plays a significant role in identification of potential drug targets. These targets are proteins which are considered necessary for the organisms to survive [2].

3.4.1 Drug Target Prioritization

There are numerous factors that can help in determining potential therapeutic targets such as molecular weight, molecular function, cellular localization, pathway analysis and virulence [56]. Molecular weight is determined by ProtParam tool (http://web.expasy.org/protparam/). Molecular functions and biological functions for every protein target is determined by Uniprot(https://www.uniprot.org/).

Cellular localization of pathogen isperformed by CELLO (http://cello.life.nctu.edu.tw/). For pathway analysis the KEGG web tool (https://www.genome.jp/kegg/) is used which will determine the function of protein targets in various cellular pathways. To identify virulence of protein targets VFDB is used which determines the pathogenic virulence of the protein targets.

3.4.2 Catalytic Pocket Detection

For the catalytic pocket detection DoGSiteScorer will be used. It is automated pocket detection and analysis tool which is used for calculation of draggability of protein cavities. This tool will return the pocket residue and draggability score which ranges from 0-1. The score closer to 1 indicates highly druggable protein cavity. The cavities that are predicted possibly bind ligands with high affinity [51].

3.4.3 Molecular Docking

For the protein-ligand binding the MOE (Molecular Operating Environment) software is used. The binding of the ligands into the binding pocket of protein molecule is calculated by MOE-Dock implemented in MOE. Molecular docking is performed with most of the default parameters. After performing docking, we analyzed the most excellent poses for hydrogen bonding/$\pi-\pi$interactions and root mean-square deviation (RMSD) calculation using MOE applications [30].
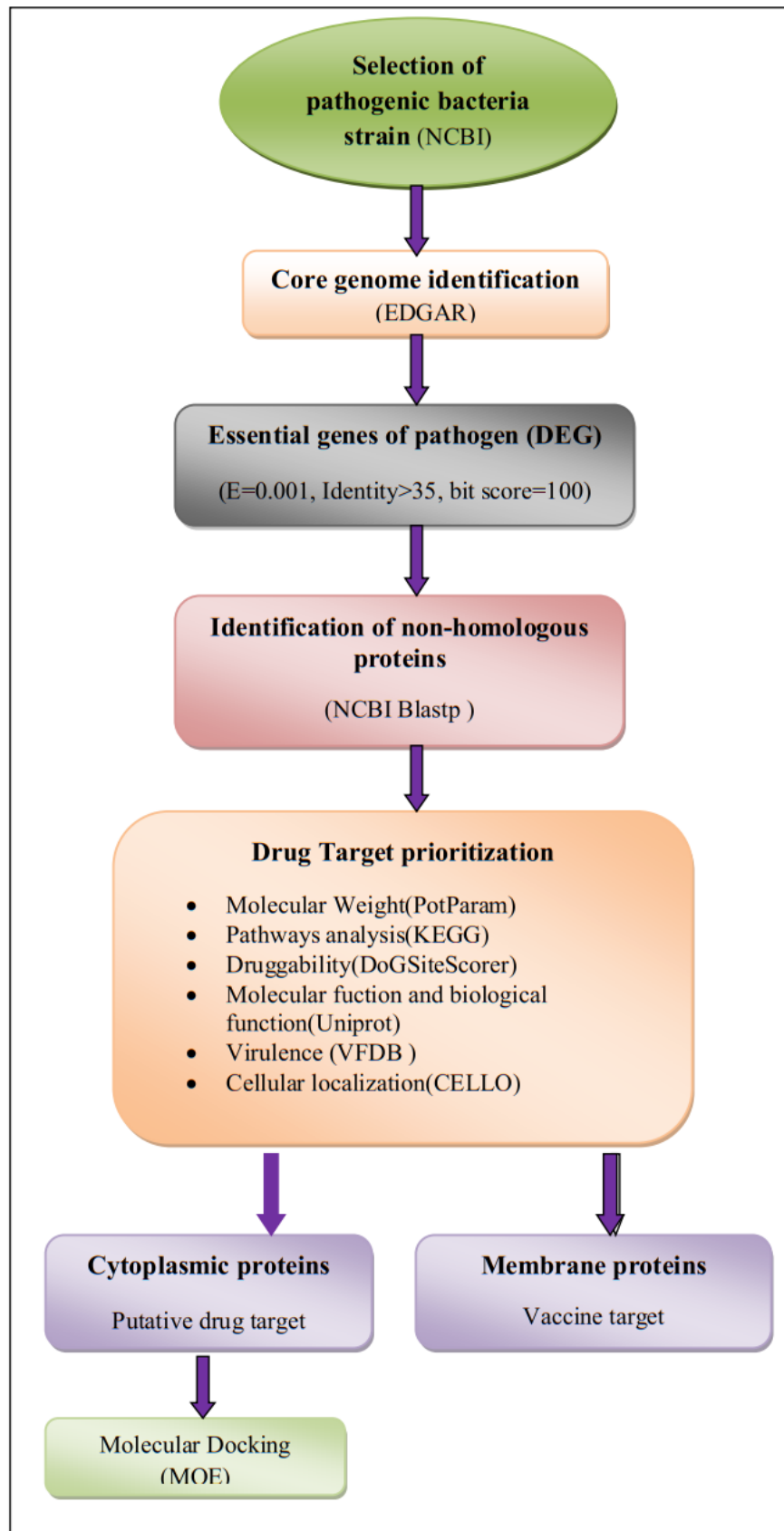
FIGURE 3.1: Methodological steps for Identification of Therapeutic Targets against *Streptococcus gordonii*

Chapter 4

Result and Analysis

4.1 Core Genome Identification of *Streptococcus gordonii*

Identification of core genome is based on two steps these are genome selection and identification of core genes by using an approach i.e. pan genome approach.

4.1.1 Genome Selection

The number of selected strains for this study is eleven. To get more accurate results we select only those stains which have complete genome. The proteins and genes of these complete genomes of eleven strains of *Streptococcus gordonii* were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/genome/). Genome selection is first step for identification of core genome. Genome Assembly and annotation report provide data of 53 strains of Streptococcus gordonii out of which 37 are contig, 5 are scaffold and 11 are complete. Table 4.1 shows the 11 strains of *streptococcus gordonii* having complete genome and names of all selected strains,genome size in Mb ,GC content, total number of genes present in strain and total number of proteins. The above mentioned information on genome statistics plays an important role in core genome identification.

28

TABLE 4.1: Strains of *Streptococcus gordonii* with Information on Genome Statistics

Sr. #	Organisms/ Name	Strain	Genome size (Mb)	GC %	Total genes	Total Protein
1	*S.gordonii str. Challis substr. CH1*	Challis substr. CH1	2.19666	40.50	2152	2014
2	*S. gordonii*	NCTC7865	2.18761	40.60	2150	2004
3	*S. gordonii*	KCOM 1506 (= ChDC B679)	2.28331	40.60	2273	2131
4	*S. gordonii*	NCTC9124	2.22391	40.40	2180	2012
5	*S. gordonii*	FDAARGOS_257	2.22271	40.40	2206	2060
6	*S. gordonii*	NCTC7868	2.19666	40.50	2149	2015
7	*S. gordonii*	FDAARGOS_683	2.19659	40.50	2133	2010
8	*S. gordonii*	IE35	2.1901	40.50	2144	2032
9	*S. gordonii*	NCTC10231	2.18555	40.60	2132	1993
10	*S. gordonii*	NCTC3165	2.1548	0.50	2156	2005
11	*S. gordonii*	FDAARGOS_371	2.12131	40.70	2102	1963

4.1.2 Core Genome Identification using Pan-genome Approach

EDGAR website was used to identify Pan Genomes. The EDGAR requires one strain as reference strain all other strains which were selected as complete genome from NCBI are compared with the reference strain. In this study the selected reference strain was *Challis substr.* CH1then rest of strains was compared with this reference strain. The result was in the foam of table which contain the locus tag and description of genes along with the Fasta file. Total number of genes identified in pan-genome is 2,835 genes, out of these 1,255 core genes were identify. Then this data is further used for the subtractive genome analysis.

4.1.3 Subtractive Genome Analysis from Core Genome

Subtractive genome analysis is a computational approach used to recognize the non-homologous proteins and these non-homologous are essential genes which are necessary for the survival of an organism.

4.1.4   Non-homologous Proteins Identification

NCBI-BLASTp was used to identify the non-homologous proteins. Default parameters were used against human genome to filter out the non-homologous proteins. Core genome of all 1,255 Protein sequences was subjected to Blastp, from which 643 were non-homologous proteins. Then these non-homologous proteins were used to identify essential proteins.

4.2   Essential Genes Identification

Essential genes are genes that are needed necessarily for the continued existence of pathogen and help to perform all important functions of life for instance entry into host, adherence, causing disease and determination within the host. The core 643 non-host homologous proteins were given to the Database of Essential Gene (DEG) for the recognition of essential proteins, and then final set of 20 proteins was attained shown in table 4.2.

TABLE 4.2: List of Non-Homologous Pathogen-Essential Proteins

Query_ID	DEG ID	Percent Identity	E_Value	Protein Name
SGO_ RS00230	DEG10480279	68.571	0.001	Glutamine-fructose- 6-phosphate aminotransferase
SGO_ RS00780	DEG10200254	56.41	0.001	UDP-3-O- acylglucosamine- N-acyltransferase
SGO_ RS00875	DEG10410143	52.941	0.001	Ribonuclease P protein component
SGO_ RS01020	DEG10070130	52.381	0.001	YSIRK_signal domain protein

Query_ID	DEG ID	Percent Identity	E_Value	Protein Name
SGO_RS01040	DEG10200056	52.239	0.001	Peptidase C51 domain -containing protein
SGO_RS01090	DEG10130190	50.898	0.001	tRNA N6-adenosine threonyl carbamoyltransferase
SGO_RS01400	DEG10320340	49.505	0.001	AraC family transcriptional regulator
SGO_RS01590	DEG10140111	47.368	0.001	Pseudouridine synthase
SGO_RS01610	DEG10430020	47.059	0.001	Glycosyl transferase
SGO_RS01705	DEG10120121	46.296	0.001	Bifunctional protein GlmU
SGO_RS01715	DEG10220213	45.833	0.001	AMP nucleosidase
SGO_RS02060	DEG10060014	44.828	0.001	Probable DNA-directed RNA polymerase subunit delta
SGO_RS02090	DEG10050245	44.444	0.001	ATP-dependent DNA helicase RecQ
SGO_RS02330	DEG10480104	42.424	0.001	Peptidase S24-like protein
SGO_RS02440	DEG10470298	41.463	0.001	TetR family transcriptional regulator
SGO_RS02520	DEG10480275	40.625	0.001	Chromosomal replication initiator protein DnaA

Query_ID	DEG ID	Percent Identity	E_Value	Protein Name
SGO_ RS02545	DEG10230225	40	0.001	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase
SGO_ RS02885	DEG10430422	37.757	0.001	Peptidoglycan D, D-transpeptidase FtsI
SGO_ RS03190	DEG10360030	37.395	0.001	HTH-type transcriptional regulator PrtR
SGO_ RS03305	DEG10320013	36.531	0.001	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate—2,6-diaminopimelate ligase

4.3 Drug Prioritization and Protein-Ligand Interactions

The study was further divide into two steps to identify the best therapeutics target. First step was drug prioritization it will provide information of target proteins about their druggability and protein-ligand interaction was determined the catalytic pockets and will perform molecular docking of target protein.

4.3.1 Drug Prioritization

Several factors contribute to determine the potential therapeutic targets such as molecular weight, molecular function and cellular function, virulence and pathway analysis.

4.3.1.1 Molecular Weight

For drug targeting the determination of molecular weight (MW) is one of the important and basic step. Best molecular weight for targets are less than 100 kilo Dalton (kDa) the target who have MW less than 100 are considered as best Therapeutic targets [43]. Molecular weight of proteins was calculated by online tool Protparam.

This tool calculates the molecular weight in g/mol after converting the MW of proteins into kDa, molecular weight of all proteins was less than 100kDa. Molecular weight of all proteins is shown in table 4.4.

4.3.1.2 Subcellular Localization

The environment where proteins work is known as subcellular localization. Subcellular localization impacts the protein work by controlling access to and accessibility of a wide range of molecular interaction partner.

The data of localization of protein normally have a vital role in defining the cellular function of hypothetical and newly discovered proteins [50]. 15 proteins were cytoplasmic, 3 were membrane proteins and 2 were extracellular and periplasmic protein out of total 20 proteins. This division of proteins is shown in table 4.5.

4.3.1.3 Virulence of Target Proteins

Pathogens are responsible for causing several human diseases by invading the host immune mechanism, this is known as virulence.

When these non-homologous and essential proteins are declared as virulent, then these can be fundamental significance to reveal novel therapeutics targets.

In this study VFDB is used to find out the virulent proteins. All the targeted proteins are found to be virulent except Ribonuclease P protein component. A virulent target protein is shown in table 4.3.

4.3.1.4   Molecular and Biological Functions

Biological and molecular function plays a vital role in considerate of the protein. Biological function and molecular function are different from each other. Gene ontology shows that molecular function describes activities that take place at molecular level, e.g. transport or catalysis, whereas biological process is complex and extensive process and it is accomplished by multiple molecular activities [55]. In this study uniprot web is used to find out the molecular and biological function of the proteins. Biological and molecular function of all the target proteins is shown in table 4.5 and 4.6 respectively.

4.3.1.5   Pathway Analysis

Pathway analysis of the targeted proteins tells us about the function of these targets in various cellular pathways. Pathway analysis of genomic data signifies one potential track for computational inference of drug targets. KEGG database was used in this study to find out the pathways of the targeted proteins. These pathways of targeted proteins are shown in table 4.4.

4.3.2   Protein-Ligand Interaction

The basic step for the protein-ligand interaction is to detect the catalytic pockets of targeted proteins, this will show the binding sites of the protein for the binding to the corresponding ligand and then to perform molecular docking.

4.3.2.1   Catalytic Pocket Detection

DoGSiteScorer is used to find out the catalytic pockets. For all target proteins those pockets were selected and only those were selected whose dug ability score was above than 0.6. The draggability score greater than 0.6 is considered to be good but score greater than 0.8 is preferential. The score which is nearer to 1 shows extremely druggable protein cavity.

TABLE 4.3: Drug and Vaccine target prioritization parameters (Gene name, Protein name, Uniprot ID, Virulence Factor) and functional annotation of 20 essential non-host homologous putative targets

S.#	Gene Name	Protein Name	Uniprot ID	Virulence
1	glmS	Glutamine–fructose -6-phosphate aminotransferase [isomerizing]	V0V2E2	Yes
2	lpxD	UDP-3-O- acylglucosamine N-acyltransferase	B8GWR3	Yes
3	rnpA	Ribonuclease P protein component	Q6NC40	No
4	SPRM2 00_0325	YSIRK_signal domain protein	A0A5C 1BBB9	Yes
5	CC_0349	Peptidase C51 domain -containing protein	Q9AB82	Yes
6	tsaD	tRNA N6-adenosine threonyl carbamoyltransferase	Q6FCK9	Yes
7	gadX	AraC family transcriptional regulator	A0A0U 1ITI1	Yes
8	MYPU_2980	Pseudouridine synthase	Q98QR3	Yes
9	BGC29_04275	Glycosyl transferase	A0A1E3 M6M1	Yes
10	Glum	Bifunctional protein GlmU	A0Q565	Yes
11	PGN_1002	AMP nucleosidase	B2RJH6	Yes
12	rpoE	Probable DNA-directed RNA polymerase subunit delta	P47268	Yes
13	recQ	ATP-dependent DNA helicase RecQ	P71359	Yes
14	HMPREF 1604_02797	Peptidase S24-like protein	V0V7J8	Yes
15	BK761_05580	TetR family transcriptional regulator	A0A243 DXR5	Yes
16	dnaA	Chromosomal replication initiator protein DnaA	Q1R4N5	Yes
17	murF	UDP-N-acetylmuramoyl- tripeptide–D-alanyl- D-alanine ligase	Q8A1L7	Yes
18	ftsI	Peptidoglycan D,D -transpeptidase FtsI	D0CCM7	Yes
19	prtR	HTH-type transcriptional regulator PrtR	Q06553	Yes
20	murE	UDP-N-acetylmuramoyl -L-alanyl-4 D-glutamate–2,6- diaminopimelate ligase	A0A0H 3NH87	Yes

TABLE 4.4: Drug and Vaccine target prioritization parameters (Gene name, Protein name, KEGG Pathway and Molecular Weight) and functional annotation of 20 essential non-host homologous putative targets

S.#	Gene Name	KEGG Pathways	Molecular Weight (g/mol)
1	glmS	1.Alanine, aspartate and glutamate metabolism 2.Amino sugar and nucleotide sugar metabolism 3.Metabolic pathways 4.Biosynthesis of antibiotics	66866.28
2	lpxD	no hit	34329.93
3	rnpA	no hit	12879.05
4	SPRM200_0325	no hit	196144.2
5	CC_0349	no hit	24925.36
6	tsaD	no hit	36723.48
7	gadX	no hit	33325.26
8	MYPU_2980	no hits	34431.91
9	BGC29_04275	no hits	36093.57
10	Glum	1.Amino sugar and nucleotide sugar metabolism 2.Metabolic pathways 3.Biosynthesis of antibiotics	49637.92
11	PGN_1002	no hits	29031.2
12	rpoE	no hits	17057.05
13	recQ	Homologous recombination	70055.52
14	HMPREF1604_02797	no hits	28164.26
15	BK761_05580	no hits	23607.51
16	dnaA	Two-component system	52550.81
17	murF	1.Vancomycin resistance 2.Peptidoglycan biosynthesis 3.Lysine biosynthesis 4.Metabolic pathways	47729.11
18	ftsI	beta-Lactam resistance	64726.66
19	prtR	no hits	28111.87
20	murE	1. biosynthesis of Peptidoglycan 2.Metabolic pathways 3.Two-component system	53286.67

TABLE 4.5: Drug and Vaccine target prioritization parameters (Gene name, Subcellular localization and Molecular Function) and functional annotation of 20 essential non-host homologous putative targets

S.#	Gene Name	Subcellular localization	Molecular Function
1	glmS	Cytoplasmic	1. Carbohydrate derivative binding 2. glutamine-fructose-6-phosphate transaminase
2	lpxD	Cytoplasmic	N-acyltransferase activity
3	rnpA	Cytoplasmic	1. ribonuclease P activity 2.tRNA binding
4	SPRM200_0325	cellwall	1.carbohydrate binding 2.glycopeptide alpha-N-acetylgalactosaminidase activity
5	CC_0349	Extracellular	1.iron ion binding2.metalloendopeptidase activity
6	tsaD	Cytoplasmic	3.N(6)–L-threonylcarbamoyladeninesynthase activity
7	gadX	Membrane	1.DNA-binding transcription factor activity 2.sequence-specific DNA binding
8	MYPU_2980	Cytoplasmic	1.lyase activity 2.pseudouridine synthase activity 3.RNA binding
9	BGC29_04275	Membrane	transferase activity
10	Glum	Cytoplasmic	1.glucosamine-1-phosphate, N-acetyltransferase activity 2.magnesium ion binding 3.UDP-N-acetylglucosamine diphosphorylase activity
11	PGN_1002	Cytoplasmic	AMP nucleosidase activity
12	rpoE	Cytoplasmic	DNA-directed 5'-3' RNA polymerase activity
13	recQ	Cytoplasmic	1.3'-5' DNA helicase activity 2.ATP binding 3.DNA binding
14	HMPREF1604_02797	Cytoplasmic	DNA binding
15	BK761_05580	Cytoplasmic	DNA binding
16	dnaA	Cytoplasmic	1.ATP binding 2.DNA replication origin binding
17	murF	Cytoplasmic	1.ATP binding 2.UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase activity
18	ftsI	Membrane	1. Penicillin binding 2.eptidoglycan glycosyltransferase activity 3.serine-typeD-Ala-D-Alacarboxypeptidase activity
19	prtR	Cytoplasmic	DNA binding
20	murE	Cytoplasmic	1. ATP binding 2.magnesium ion binding 3.UDP-N –acetylmuramoylalanyl-D-glutamate-2,6 –diaminpimelate ligase activity

TABLE 4.6: Drug and Vaccine target prioritization parameters (Gene name, and Biological Function) and functional annotation of 20 essential non-host homologous putative targets

S.#	Gene Name	Biological Function
1	glmS	1.Carbohydrate derivative biosynthetic process 2.carbohydrate -metabolic process
2	lpxD	lipid A biosynthetic process
3	rnpA	tRNA 5'-leader removal
4	SPRM200_0325	
5	CC_0349	
6	tsaD	tRNA threonylcarbamoyladenosine modification
7	gadX	Transcription, Transcription regulation
8	MYPU_2980	
9	BGC29_04275	pseudouridine synthesis
10	Glum	1.cell morphogenesis 2.cell wall organization 3.lipid A biosynthetic process 4.lipopolysaccharide biosynthetic process 5.peptidoglycan biosynthetic process 6.regulation of cell shape 7.UDP-N-acetylglucosamine biosynthetic process
11	PGN_1002	1.AMP salvage 2.nucleoside metabolic process
12	rpoE	1.regulation of transcription, DNA-templated 2.transcription, DNA-templated
13	recQ	1.DNA duplex unwinding 2.DNA recombination 3.DNA repair 4.DNA replication 5.SOS response
14	HMPREF1604_02797	
15	BK761_05580	Transcription, Transcription regulation
16	dnaA	1.DNA replication initiation 2.regulation of DNA replication
17	murF	1.cell cycle 2.cell division 3.cell wall organization 4.peptidoglycan biosynthetic process
18	ftsI	1.cell wall organization 2.division septum assembly 3.FtsZ-dependent cytokinesis 4.peptidoglycan biosynthetic process 5.regulation of cell shape
19	prtR	1.bacteriocin biosynthetic process 2.negative regulation of secondary metabolite biosynthetic process
20	murE	1.cell cycle 2.cell division 3.cell wall organization 4.peptidoglycan biosynthetic process

4.3.3 Molecular Docking

Molecular docking has three main steps which includes: Selection of ligands, 3D structure prediction of targeted protein, Protein-ligand docking.

4.3.3.1 Ligands Selection

The ligand library was prepared by compounds from ZINC database (ZINC drug-like molecules). ZINC library has 1605 molecules, with Tanimoto cutoff level of 60%. 15 top positioned compounds for all targeted proteins were evaluated for figure complementarity and hydrogen bond interactions. This provides direction to the final selection of set of compounds with polypharmacology and polypharmacy characteristics for target proteins in *Streptococcus gordonii.*

4.3.3.2 3D Structure Prediction

As the structure of all these targeted proteins was not available at protein databank so the structure of these proteins was predicted by using the tool available for 3D structure prediction named SwissModel [46]. This tool is selected to find structures of targeted proteins because it is used worldwide, its performance is continuously evaluated and interface is user friendly. It has an easy access to modeling results, their visualization and its interpretation. To use this tool input all targeted protein sequences in fasta format then next step is Data Search, in this for the provided data it searches its evolutionary related protein structure against Swiss-Model Template Library (SMTL). This will use two databases on backend. One is BLAST which is quick and adequately precise. And the other is HHblits which add sensitivity to the remote homology structure [46]. The pair of proteins which have same structure and functions but lack easily detectable sequence similarity are known as remote homologs [45]. Than third step is selection of template. It shows all top ranked templates whose quality was estimated by Global Model Quality Estimate (GMQE). Those templates were selected whose

sequence similarity score was high. Final step is to build model upon the selected template it builds the 3D structure for the targeted proteins.

4.3.3.3   Validation of 3D Structures

Quality of 3D structures was validated by using RAMPAGE and ERRAT tool which are online tools available. Rampage stands for RNA Annotation and Mapping of Promoters for the Analysis of Gene Expression. Ramachandran plot was used by this tool and provide validation score for each target protein 3D structures. Score of RAMPAGE greater than 80 is considered as good score [47]. The input of this tool is .pdb file of 3D structure of targeted protein. So in this study the score of all targeted proteins are greater than 80.for more accuracy another tool ERRAT was used for validation. Quality factor greater than 37% is considered good [48]. The input of this tool is also .pdb file and quality factor for all predicted 3D structures were greater than 70 (Table 4.7).

TABLE 4.7: Validation Score of Models from Rampage and ERRAT

S.#	Protein Name	Rampage	ERRAT
1	Glutamine-fructose-6-phosphate aminotransferase	95.10%	94.382~
2	UDP-3-O-acylglucosamine-~N-acyltransferase	95.00%	81.7241
3	Ribonuclease P protein component	94.20%	94.8454
4	YSIRK_gnal domain protein	96.50%	95.1165
5	Peptidase C51 domain-containing protein	88.70%	82.6087
6	tRNA N6-adenosine ~threonylcarbamoyltransferase	94.50%	85.5219
7	AraC family transcriptionalregulator	98.00%	100
8	Pseudouridine synthase	92.10%	85.4545
9	Glycosyl transferase	91.40%	88.5135

10	Bifunctional protein	95.70%	87.8571
11	AMP nucleosidase	93.10%	95.2174~
12	Probable DNA-directed RNA polymerase subunit delta	86.70%	81.5789
13	ATP-dependent DNA helicase	95.70%	96.2
14	Peptidase S24-like protein	80.70%	70.8995
15	TetR family transcriptionalregulator	95.80%	99.4536
16	Chromosomal replication ~initiator protein	93.10%	88.0368
17	UDP-N-acetylmuramoyl-tripeptide-D–alanyl-D-alanine ligase	92.80%	86.0577
18	Peptidoglycan D,D-transpeptidase	89.50%	82.7236
19	HTH-type transcriptionalregulator	94.10%	100
20	UDP-N-acetylmuramoyl - L - alanyl - D- glutamate-2, 6–diaminopimelate ligase	96.30%	93.3619

4.3.3.4   Docking

MOE tool was used for the purpose of docking. Ligands were converted into .mdb file by minimizing molecular energy and by adding hydrogen having default parameters. Molecular docking was performed; the targeted proteins were opened in MOE. Than all solvent present in targeted protein i.e. ligands were deleted. In next step hydrogen was added, 3D protonation and energy minimization was done and save it in .mdb file. Than minimized structure of targeted protein was opened and docking was performed by selecting the ligands .mdb file here default parameters were used. Final step i.e. analysis was done in this step 1605 ligand of ZINC library were sort in ascending order of docking score. First 15 ligands were selected for each targeted proteins and there interaction was save as PNG. 3D figure representation for the best protein-ligand interaction of each targeted protein was drawn in chimera tool.

4.3.3.4.1 AMP Nucleosidase (PGN_1002)

AMPnucleosidase (AMN) enzyme is present in prokaryotesonly,in prokaryotes it have an important role in purine nucleoside salvage and intracellular AMP level control The activity of protein isinvigorated by ATP and smothered by phosphate. AMN is a common homohexamer, and every monomer is consisting of two domains:

    (1) Catalytic domain   (2) Putative regulatory domain.
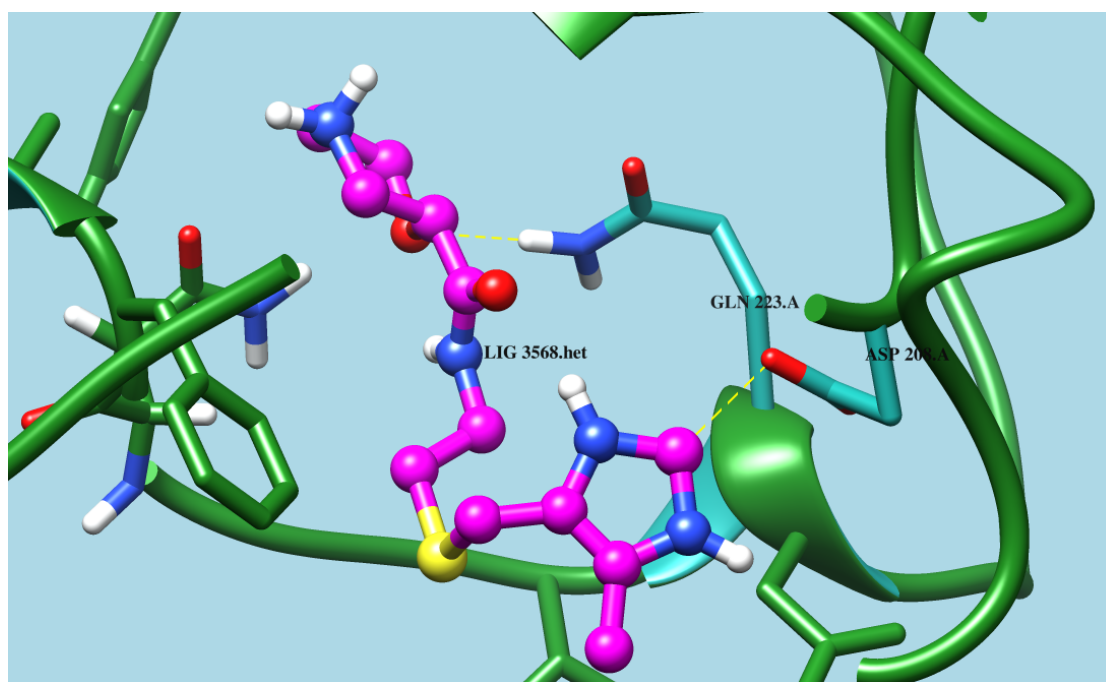
FIGURE 4.1: 3D Structure of Docking Analyses for most Druggable Protein Cavity of AMP nucleosidase with ZINC72153423

In general topology of the catalytic domain and a few features of the substrate binding site bear a resemblance to those of the nucleoside phosphorylases, revealthat AMN is a novel constituent of the family. It is a cytoplasmic protein which is essential for the survival of bacteria; this can be potential drug target after the wet lab research.

The compound name, MolDock score and interactive residues are shown in table 4.8 and 3D structural representation for the best protein-ligand interacton of this protein identified by MOE tool were drawn in chimera tool as shown in figure (figure 4.1).

TABLE 4.8: Compound name, MolDock Score and Interactive residues for AMP nucleosidase from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC83246639	Phe_229_A,Gln_223_A.	-13.3335	-14.214
2	ZINC72447036	Phe_229_A	-12.7731	-19.571
3	ZINC71827951	Lys_94_A	-12.6645	-17.051
4	ZINC72354549	Val_225_A,Asp_208_A	-12.6202	-20.049
5	ZINC71772423	Phe_229_A	-12.6001	-19.225
6	ZINC78547466	Gln_223_A,Asn_228_A,Tyr_35_A	-12.3686	-14.877
7	ZINC71784817	Asp_208_A	-12.2603	-13.870
8	ZINC78870135	Asp_208_A	-12.2293	-63.451
9	ZINC76555281	Tyr_35_A	-12.1491	-16.163
10	ZINC75106924	Asn_228_A	-12.0858	-15.189
11	ZINC80316400	2/Lys_94_A	-11.9432	-12.586
12	ZINC83262599	No hit	-11.8772	-17.083
13	ZINC71783863	Gln_223_A	-11.8194	-12.617
14	ZINC82114814	Val_225_A,Asp_222_A	-11.7993	-16.761
15	ZINC72153423	Asp_208_A,Gln_223_A, 2/Asn_228_A,Phe_32_A	-11.7803	-60.722

4.3.3.4.2 Bifunctional Protein (GlmU)

The glmU gene plays a vital role to catalyze the development of UDP-N-acetyl glucosamine, a fundamental precursor for cell wall peptidoglycan and lipopolysaccharide biosyntheses [56]. GlmU protein is actually a bifunctional protein which moreover catalyzes acetylation of glucosamine-1-phosphate. It is essential protein
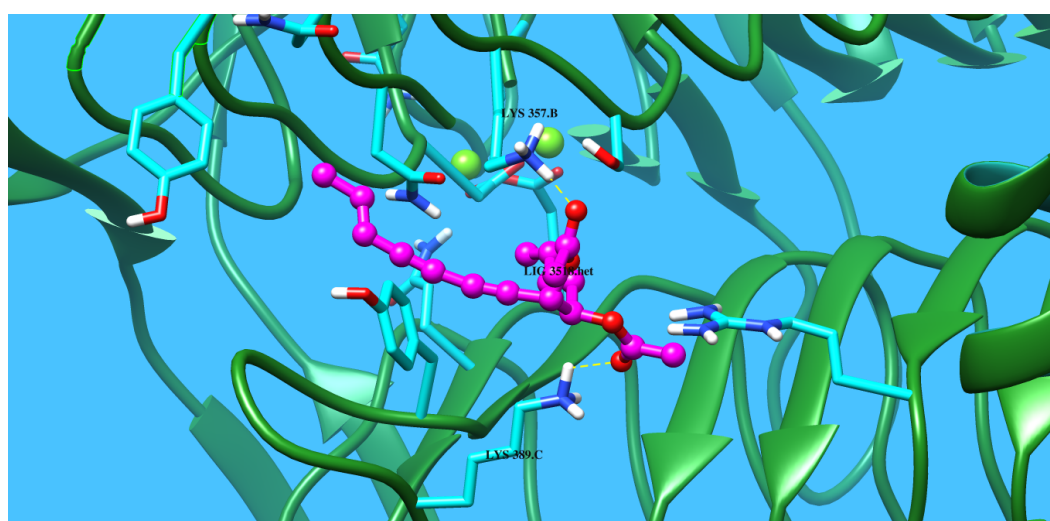
FIGURE 4.2: 3D Structure of Docking Analyses for most Druggable Protein Cavity of Bifurcational Protein with ZINC70503687

TABLE 4.9: Compound name, MolDock Score and Interactive residues for Bi-functional protein from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC71789643	Lys_389_B	-22.8298	-29.383
2	ZINC71404930	Lys_389_C	-20.9783	-21.836
3	ZINC71777127	Tyr_363_C	-19.2816	-37.031
4	ZINC83262599	Lys_389_C	-18.5302	-26.288
5	ZINC70503687	Lys_389_C,Lys_357_B	-18.1773	-41.343
6	ZINC82039595	Lys_389_C	-17.8068	-22.246
7	ZINC79036966	Arg_330_C	-17.8062	-17.160
8	ZINC78547526	Lys_389_C,Arg_330_C	-17.6660	-27.220
9	ZINC73736642	Tyr_363_C	-17.5962	-26.138
10	ZINC72187422	Lys_389_C,Arg_330_C	-17.5919	-19.049
11	ZINC82049692	Asn_383_C,Lys_348_C	-17.5387	-20.093
12	ZINC70632524	Lys_357_B,Lys_389_C	-17.5250	-13.521
13	ZINC71609301	Lys_389_C	-17.5180	-12.168
14	ZINC71782238	Arg_330_C	-17.3807	-20.835
15	ZINC71610591	Lys_389_C	-17.0571	-12.237

for the survival of bacteria and drug prioritization also confirm that it could act as drug target. The compound name,MolDock score and interactive residues are shown in table 4.9 and 3D structural representation for the best protein_ligand interacton of this protein identified by MOE tool were drawn in chimera tool as shown in figure (figure 4.2)

4.3.3.4.3 Chromosomal Replication Initiator Protein (DnaA)

In all organisms the start of DNA replication is a main occasion in the cell cycle. Replication initiation takes place at specific origin sequences that are predictable and processed by an oligomeric complex of the initiator protein DnaA in bacteria. The complex consists of three subcomplexes, and spatial arrangements of those over and above their components are vital for well-organized replication initiation. It is a cytoplasmic protein which is essential for the survival of bacteria, this can be potential drug target after the wet lab research. The compound name, MolDock score and interactive residues are shown in table 4.10 and 3D structural representation for the best protein-ligand interacton of this protein identified by MOE tool were drawn in chimera tool as shown in figure (figure 4.3)
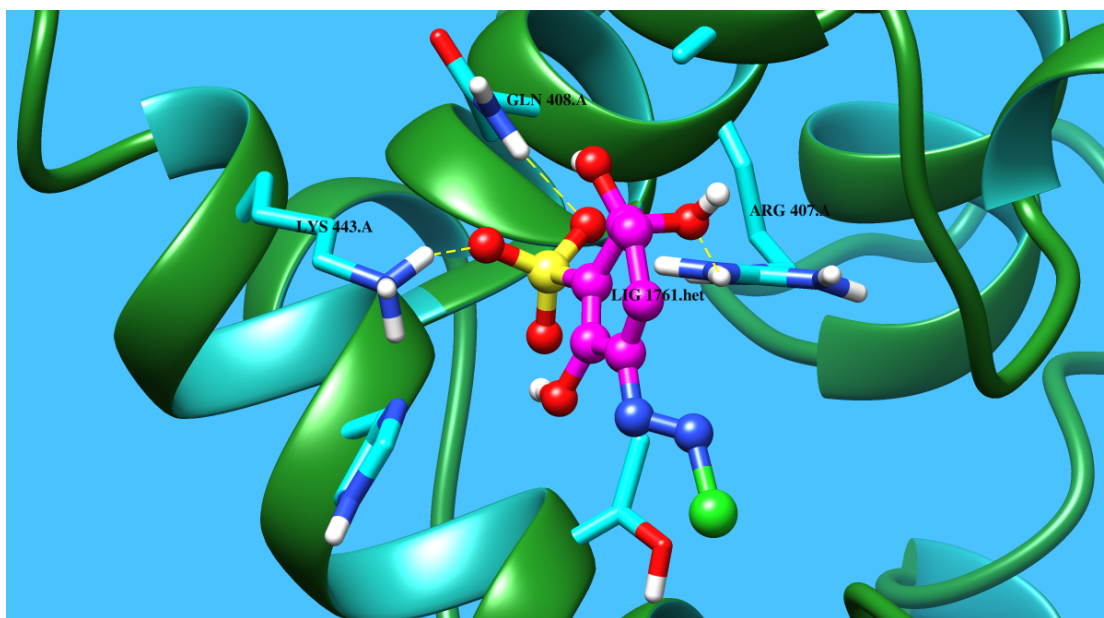
FIGURE 4.3: 3D Structure of Docking Analyses for most Druggable Protein Cavity of Chromosomal replication initiator protein with ZINC83442116

TABLE 4.10: Compound name, MolDock Score and Interactive residues for Chromosomal replication initiator protein from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC70503687	Thr_435,Asp_433, Thr-436,Gln_408, Arg_399,Arg_407	-11.9844	-52.042
2	ZINC72271115	Arg_399	-11.9760	-15.097
3	ZINC74723614	Lys_443,Arg_407	-11.7191	-20.346
4	ZINC71610591	2/Lys_443,Gln_408	-11.5903	-7.439
5	ZINC83312796	Thr_436, Arg_399,Arg_407	-11.1673	-21.306
6	ZINC71777127	Lys_443,Arg_407,Gln_408	-11.0657	-29.789
7	ZINC80053835	Lys_397,Thr_436	-10.9858	-22.135
8	ZINC71613886	No hit	-10.9660	-8.862
9	ZINC83442116	Lys_443,2/Arg_407,2/Gln_408	-10.9359	-12.800
10	ZINC71778026	Arg_432,Lys_397	-10.8917	-13.927
11	ZINC71781167	Lys_443,2/Arg_407,Gln_408	-10.8701	-27.019
12	ZINC71781091	2/Thr_436	-10.8696	-17.284
13	ZINC72338622	Lys_443,Arg_407	-10.8370	-16.870
14	ZINC71762339	Lys_443,Arg_407,Thr_436	-10.8120	-17.141
15	ZINC71777128	2/Lys_443,Arg_407	-10.7452	-28.377

4.3.3.4.4 UDP-3-O-acylglucosamine N-acyltransferase (lpxD)

Outer membrane of Gram-negative bacteria is mainly made up of a defensive, selectively permeable LPS.

TABLE 4.11: Compound name, MolDock Score and Interactive residues for UDP-3-O-acylglucosamine N-acyltransferase from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC71789643	Lys_79_,Ser_99_, Phe_13_,Ser_99_, Lys_79_	-14.8215	-23.508
2	ZINC71610591	Arg_08_	-13.8215	-6.901
3	ZINC71609301	Lys_79_	-13.8045	-20.434
4	ZINC71777127	Arg_08_,Arg_45_, Arg_63_	-12.9378	-21.482
5	ZINC82049692	Arg_63_,His_04_, Ala_09_,Gly_91_	-12.6946	-17.780
6	ZINC71783040	His_04_,Arg_63_	-12.6930	-4.519
7	ZINC80316400	Lys_79_,Lys_79_	-12.5641	-29.834
8	ZINC78976572	Ser_82_,Arg_63_	-12.5188	-15.795
9	ZINC78701498	Lys_79_,Lys_79_, Ser_99_	-12.3822	-13.645
10	ZINC83313306	Ser_82_,Arg_63_	-12.3730	-19.317
11	ZINC83406748	Asp_03_,Gly_93_, Arg_63_	-12.2312	-20.795
12	ZINC72271115	Ser_82_,Arg_63_	-12.1796	-16.614
13	ZINC77257146	Lys_79_,Lys_79_, Ala_97_	-12.1678	-23.648
14	ZINC72291787	Arg_63_,Glu_89_	-12.1589	-22.991
15	ZINC69741273	Ser_82_,Ala_90_, Arg_63_	-12.0914	-10.388

Biosynthesis of LPS depends on UDP-3-O-acyl-glucosamine N- acyltransferase (LpxD), which transfers 3-hydroxy-arachidic acid from acyl carrier protein, to the 2′ amine of UDP-3-O-myristoyl glucosamine in Chlamydia trachomatis. It is vital protein for the continued existence of bacteria and drug prioritization also confirm that it could act as drug target. The compound name, MolDock score and interactive residues are shown in table 4.11 and 3D structural representation for the best protein-ligand interacton of this protein identified by MOE tool were drawn in chimera tool as shown in figure (figure 4.4)
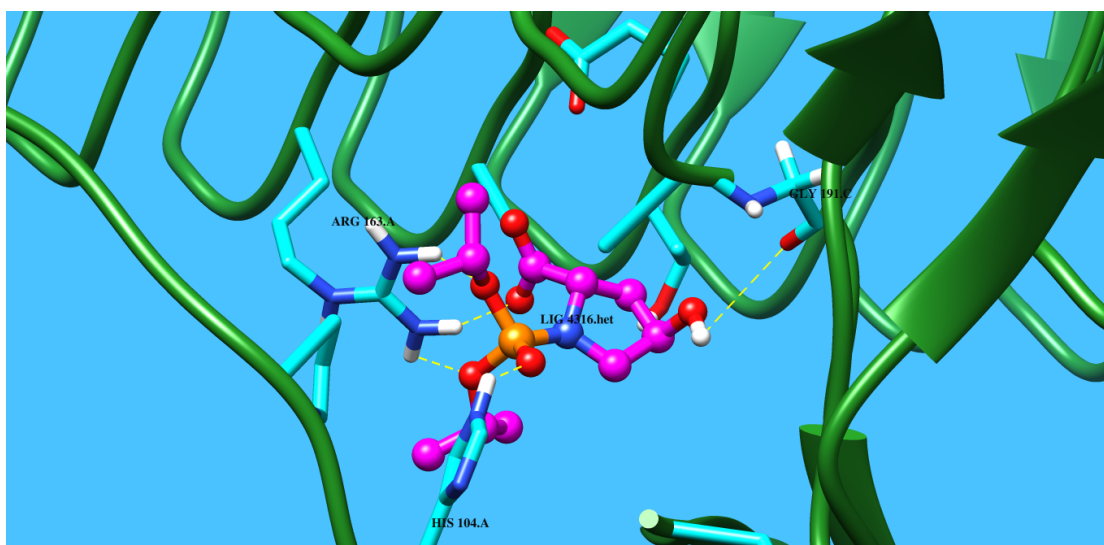
FIGURE 4.4: 3D Structure of Docking Analyses for most Druggable Protein Cavity of UDP-3-O-acylglucosamine N-acyltransferase proteinwith ZINC 82049692

4.3.3.4.5 UDP-N-acetylmuramoyl-L-alanyl-D-glutamate–2, 6-diaminopimelate Ligase (murE)

This is an enzyme present in cytoplasm, which catalyzes the addition of meso-diaminopimelic acid to nucleotide originator in the biosynthesis of cell-wall peptidoglycan of bacteria. The cell wall of bacteria composed of polymeric network of murein. This polymeric protects the cell wall from surplus water in its surrounding. This protein initiates the synthesis of murein by adding mesodiaminopimelic acid to UDP-N-acetylmuramoyl-l-alanyl-d-glutamate in cytoplasm. This protein play vital role in survival of bacteria the reason is that its cell wall synthesis could acts as a putative drug target. Table 4.12 illustrates a set of 15 potential ligands

according to their MolDock score and minimized energy value from the above mentioned ZINC library.

TABLE 4.12: Compound name, MolDock Score and Interactive residues for UDP-N-acetylmuramoyl-L-alanyl-D-glutamate–2,6-diaminopimelate ligase from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC72233583	Arg_0,Leu_,Ser_,Ala_,Gln_,His_, Gln_1, Asn_7, Thr_8,Arg_3,Ser_5	-17.5631	-85.785
2	ZINC71777792	3/Glu_9	-15.7656	-13.961
3	ZINC71780763	His_0,Arg_7, Lys_4,Asp_0, His_1,Lys_0,~Tyr_8	-15.7521	-22.630
4	ZINC70632524	Asp_0,Lys_4, Arg_7,Thr_8	-15.6213	-14.412
5	ZINC70632388	Lys_0,His_0,Thr_8,Thr_3	-15.5324	-16.037
6	ZINC78611957	Arg_7	-15.4265	-13.686
7	ZINC71618824	Lys_4,Asp_0, His_1,Tyr_8	-15.1914	-23.003
8	ZINC79482538	Tyr_1,3/Lys_0, Lys_5,2/Ser_6,His_7	-15.1322	-17.551
9	ZINC71827951	Arg_0,Arg_3,Thr_9, Asn_7,Lys_0	-14.6894	-14.002
10	ZINC71777127	2/Arg_7,Lys_4,His_0	-14.6312	-27.119
11	ZINC71610591	Lys_4,Arg_7	-14.5694	-14.478
12	ZINC70503687	Lys_4,Asp_0,2/His_0, Lys_0,Tyr_8	-14.5318	-39.500
13	ZINC72145573	Lys_4,Thr_5	-14.5214	-21.089
14	ZINC71404930	Arg_7,Thr_5,Lys_4	-14.4313	-21.566
15	ZINC71609301	Arg_7	-14.3715	-8.148
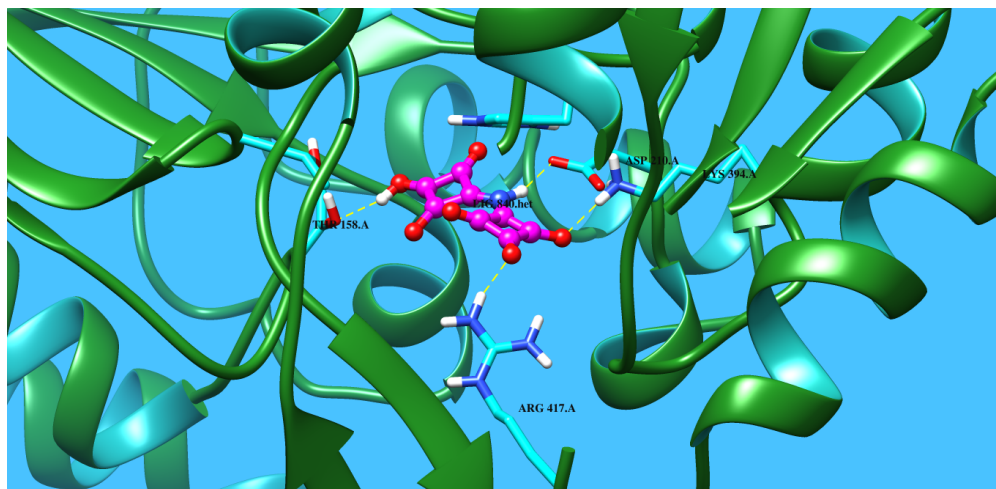
FIGURE 4.5: 3D Structure of Docking Analyses for most Druggable Protein Cavity of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate–2,6-diaminopimelate ligase withZINC70632524

4.3.3.4.6 UDP-N-acetylmuramoyl-tripeptide–D-alanyl-D-alanine Ligase (murF)

MurF is necessary to catalyze the ultimate step within the synthesis of the cytoplasmic precursor of the cell wall peptidoglycan of bacteria andis an attractive target for drug development such as antibacterial drugs. Evaluation with the complex crystal structures of MurD and its substrates, the topology of the N-terminal domain of MurF is unique, as its central and C-terminal domains exhibit similar mononucleotide and dinucleotide-binding folds, correspondingly. It is a cytoplas-
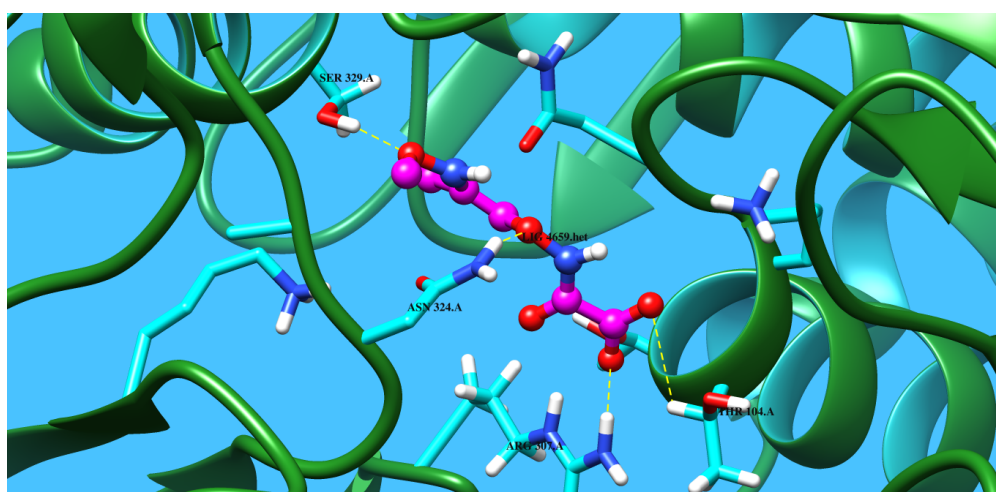
FIGURE 4.6: 3D Structure of Docking Analyses for most Druggable Protein Cavity of UDP-N-acetylmuramoyl-tripeptide–D-alanyl-D-alanine ligasewith ZINC71781357

mic protein which is essential for the survival of bacteria,this can be potential drug target after the wetlab research. Table 4.13 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.

TABLE 4.13: Compound name, MolDock Score and Interactive residues for UDP-N-acetylmuramoyl-L-alanyl-D-glutamate–2,6-diaminopimelate ligase from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC78556566	Lys_8,Ser_8,Asn_3	-16.6791	-21.935
2	ZINC72194390	Lys_8,Thr_4	-13.8823	-28.881
3	ZINC72333100	Asn_6,Ser_8,Lys_8	-13.7611	-29.626
4	ZINC83442116	Lys_3,Asn_4, Arg-307,Thr_4	-13.6617	-19.529
5	ZINC73824605	No hit	-13.5822	-21.945
6	ZINC71951504	Lys_3	-13.4810	-16.213
7	ZINC71789643	Lys_3,Asn_4,Lys_8,Asp_1	-13.4496	-26.394
8	ZINC75163962	Thr_4	-13.3015	-27.270
9	ZINC71781357	Arg-307,Thr_4,Lys_3,Asn_4,Ser_9	-13.2128	-20.339
10	ZINC78547526	No hit	-13.1210	-25.325
11	ZINC71863887	Lys_8	-12.8958	-25.034
12	ZINC71177577	Asn_4	-12.8336	-22.438
13	ZINC83293606	Lys_8,Gly_2	-12.8308	-29.615
14	ZINC73825281	Asp_1,Ser_8,Ala_2	-12.7234	-4.895
15	ZINC72463312	Asn_6,Asp_1	-12.4934	-16.017

4.3.3.4.7 HTH-type Transcriptional Regulator (PrtR)

Helix-turn-helix (HTH) is a key structural motif able to bind to DNA. Every monomer includes two α helices that bind to the major groove of DNA connected by a short strand of amino acids. The HTH motif takes places in many proteins

TABLE 4.14: Compound name, MolDock Score and Interactive residues for HTH-type transcriptional regulator from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC71863768	Lys_3,Arg_15	-14.0707	-10.370
2	ZINC71777127	2/Lys_3,Val_67,Arg_15	-13.8433	-29.231
3	ZINC83442116	Arg_15,Ser_4	-13.3139	-17.542
4	ZINC83332969	Lys_3,Arg_15	-13.2665	-21.484
5	ZINC71787260	Lys_3,2/Arg_15	-13.2521	-10.998
6	ZINC71609301	Lys_3,Arg_15	-13.1373	-7.915
7	ZINC70632524	Lys_3,Ser_4	-12.6732	-11.859
8	ZINC83295231	2/Lys_3,Arg_15	-12.6568	-29.114
9	ZINC72372215	2/Lys_3,Arg_15	-12.5929	-27.290
10	ZINC83262599	No hit	-12.4528	-28.749
11	ZINC72343413	2/Arg_15	-12.3619	-33.209
12	ZINC71618824	2/Lys_3,Arg_15,Ser_4	-12.2688	-33.038
13	ZINC73630165	Lys_3	-12.2672	-18.166
14	ZINC71785005	Arg_15	-12.2487	-15.524
15	ZINC80316400	Lys_3,Arg_15	-12.2115	-25.639

that regulate gene expression. This will assume not to be mistaken for the helix-loop-helix motif [57]. The helix-turn-helix motif is a DNA-binding motif. It is essential protein for the survival of bacteria and drug prioritization also confirm that it could act as drug target. Table 4.14 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.
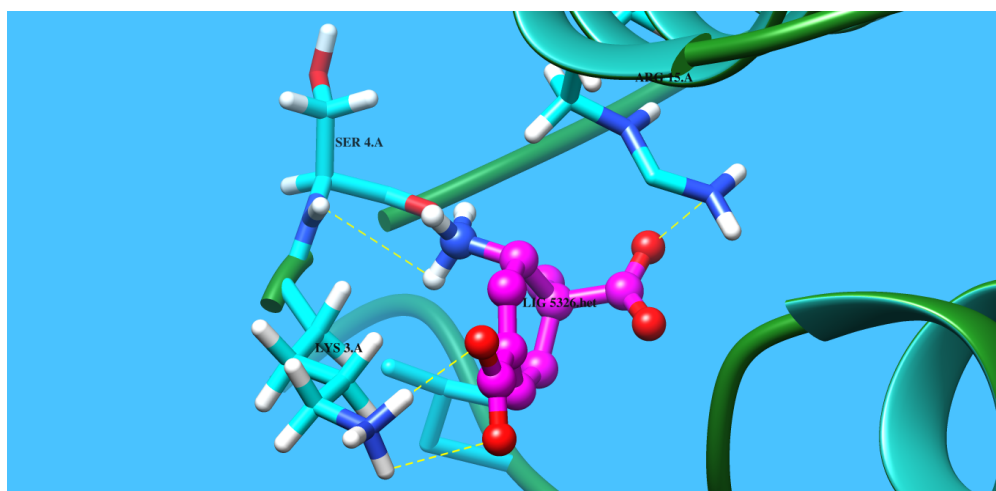
FIGURE 4.7: 3D Structure of Docking Analyses for most Druggable Protein Cavity of HTH-type transcriptional regulator with ZINC71618824

4.3.3.4.8 Probable DNA-directed RNA Polymerase Subunit Delta (rpoE)

TABLE 4.15: Compound name, MolDock Score and Interactive residues for Probable DNA-directed RNA polymerase subunit delta from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC71603491	Arg_3,Glu_0	-12.4968	-18.603
2	ZINC74941352	Glu_0	-12.2533	-15.459
3	ZINC83294593	Glu_0	-11.6314	-13.202
4	ZINC82163903	Glu_0,Glu_7	-11.5909	-21.346
5	ZINC71780841	Asp_7,Tyr_4,Tyr_3	-11.3396	-14.553
6	ZINC73822846	Asn_1,Arg_3	-11.1254	-10.826
7	ZINC72267068	Glu_0,Lys_3	-10.8458	-5.179
8	ZINC71967867	Lys_8,Glu_3	-10.8158	-17.461
9	ZINC72194481	Glu_0	-10.4065	-14.324
10	ZINC71776694	Tyr_4,Tyr_3	-10.1654	-15.233
11	ZINC72232841	No hit	-10.0472	-11.803
12	ZINC83250876	No hit	-9.9741	-7.017
13	ZINC70461088	Arg_3,Asp_9	-9.8596	-11.453
14	ZINC83439122	Tyr_3,Tyr_4	-9.8466	-15.506
15	ZINC72272542	Tyr_3	-9.7337	-18.518

The gene rpoE (RNA polymerase, extracytoplasmic E) predetermines the sigma factor sigma-24 (σ24, sigma E, or RpoE), a protein in *E. coli* and many species of bacteria. Based on thebacterial species, this gene might be known assigE. RpoE has been reported as an essential gene for exocytoplasmic stress response. Without rpoE *E. coli* mutants are not able to grow up at elevated temperatures (i.e.> 42 ℃) and illustrate growth deficiency at lower temperatures, although this might be due to compensatory mutations. In several bacterial species, for instance Clostridium botulinum, this sigma factor conceivably necessary for sporulation. The Hfq protein in *E. coli* transform the stress response activity of RpoE [6]. It is a cytoplasmic protein which is essential for the survival of bacteria,this can be potential drug target after the wetlab research. Table 4.15 illustrates a set of 15

potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.
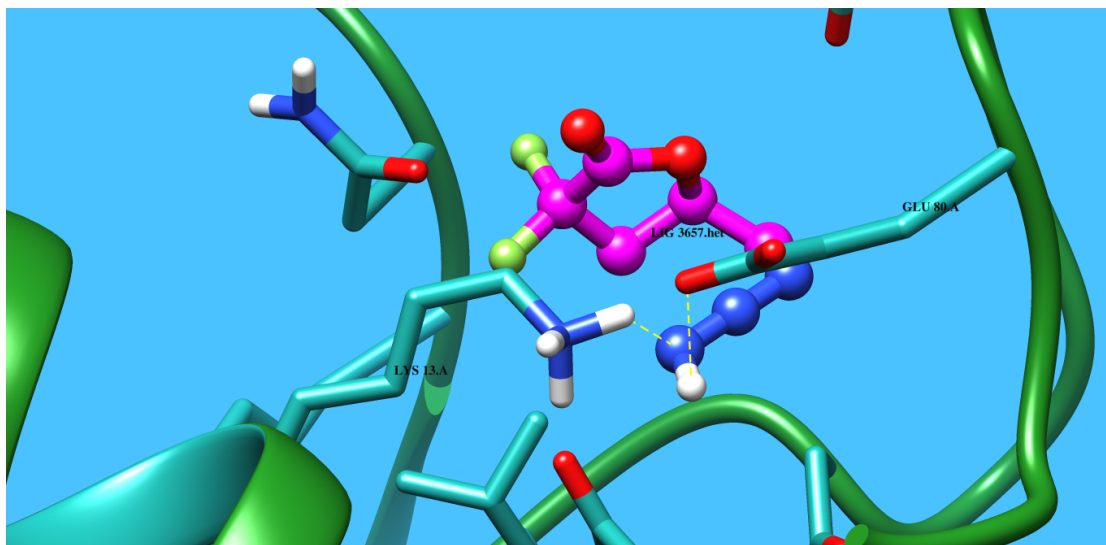
FIGURE 4.8: 3D Structure of Docking Analyses for most Druggable Protein Cavity of Probable DNA-directed RNA polymerase subunit delta with ZINC72267068

4.3.3.4.9 TetR Family Transcriptional Regulator (BK761_ 05580)

TetR family of regulators (TFRs) is generally correlated with antibiotic resistance and it helps to regulation of genes that encodes for the exporters of small-molecule.
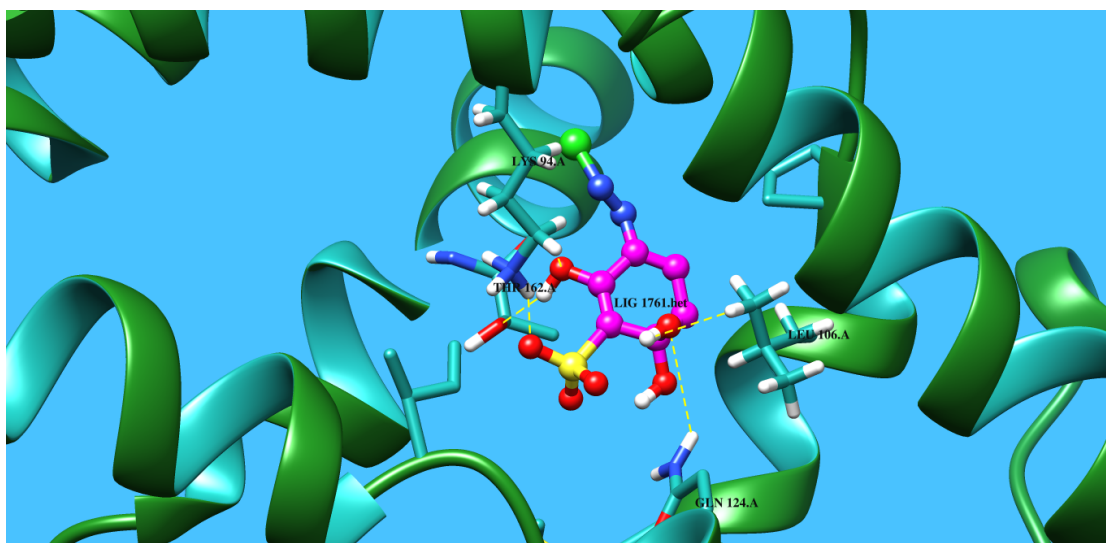
FIGURE 4.9: 3D Structure of Docking Analyses for most Druggable Protein Cavity of TetR family transcriptional regulator with ZINC83442116

TABLE 4.16: Compound name, MolDock Score and Interactive residues for TetR family transcriptional regulator from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC78611957	Lys_94	-14.2283	-11.564
2	ZINC70632388	Lys_94,Gln_124, Thr_162	-14.1354	-6.356
3	ZINC72145573	Gln_111	-14.0751	-9.421
4	ZINC78868921	Lys_94	-13.8527	-13.472
5	ZINC70503687	Thr_162,Asp_166	-13.5175	-29.403
6	ZINC83313306	Lys_94,Thr_162	-13.4713	-30.280
7	ZINC83442116	Gln_124,Leu_106, Thr_162,Lys_94	-13.2859	-18.230
8	ZINC82039595	Lys_94	-13.1124	-11.479
9	ZINC71781091	Lys_94,Thr_162	-13.0461	-18.023
10	ZINC71285802	Glu_69	-12.9859	-11.921
11	ZINC72233583	Lys_94	-12.8267	-11.215
12	ZINC72194390	Lys_94,Thr_162	-12.8232	-21.944
13	ZINC71186802	Gln_124,Lys_94, Thr_162	-12.8084	-20.718
14	ZINC70632524	Lys_127,Gln_124	-12.8031	-4.507
15	ZINC71973421	Leu_107,Lys_94, Asp_166	-12.7250	-14.120

On the other hand, TFRs plays a much wider role in controlling genes involved in metabolism, production of antibiotic, quorum sensing, and numerous other features of prokaryotic physiology. It is essential protein for the survival of bacteria and drug prioritization also confirm that it could act as drug target. Table 4.16 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.

4.3.3.4.10 ATP-Dependent DNAHelicase (RecQ)

The RecQ helicases plays a vital role in maintenance of chromosomes and their faults play a crucial role in many disorders and in human it cause cancer. The RecQ helicases belong to an significant family of highly conserved DNA helicases. RecQ is essential in prokaryotes for recombination of plasmid and DNA repair from Ultraviolet-light, free radicals, and alkylating agents. ATP-dependent DNA helicase protein also helps to reverse damage from replication errors.

Normally without RecQ proteins replication does not continue in eukaryotes, it also have important role in aging, silencing, recombination and DNA repair. It is a cytoplasmic protein which is essential for the survival of bacteria, this can be potential drug target after the wetlab research. The compound name, MolDock score and interactive residues are shown in table 4.17 and 3D structural representation for the best protein_ligand interacton of this protein identified by MOE tool were drawn in chimera tool as shown in figure (figure 4.10).

TABLE 4.17: Compound name, MolDock Score and Interactive residues for ATP-dependent DNA helicase from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC82087835	Asp_188	-11.9702	-40.180
2	ZINC72463312	Tyr_189,Asn _390,Gln_486	-11.6775	-6.108
3	ZINC76022877	Gln_193, 2/Ser_490	-11.3723	-14.770
4	ZINC79493313	Tyr_391	-11.3078	-17.908
5	ZINC71780157	2/Ser_490,Arg_197	-11.2574	-14.778
6	ZINC77969011	Ser_490	-11.2256	-8.337
7	ZINC72153423	Glu_372, Tyr_391	-11.1743	-44.055
8	ZINC76063543	Tyr_189,Asn_390	-10.8976	-21.397
9	ZINC71973421	Gln_487	-10.8918	-8.515
10	ZINC71780763	Tyr_189 ,Tyr_391	-10.8894	-3.922
11	ZINC71778026	2/Ser_490,Gln_193	-10.8796	-14.642
12	ZINC71778472	Ser_490,Gln_193	-10.7741	-18.675
13	ZINC83260916	Asp_188,Ala_190	-10.7485	-20.554
14	ZINC77291760	Gln_193	-10.7443	-23.516
15	ZINC72272542	Ser_490,Gln_193	-10.7442	-15.542
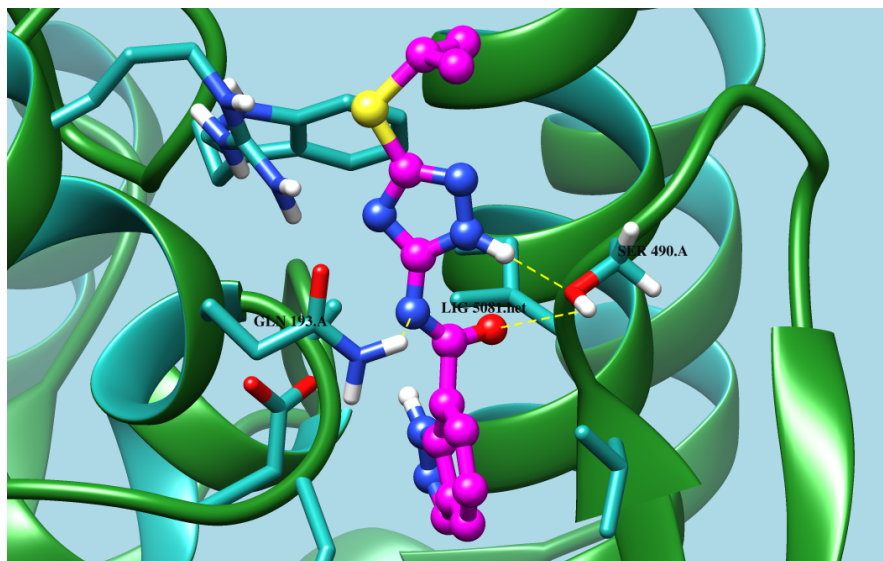
FIGURE 4.10: 3D Structure of Docking Analyses for most Druggable Protein
Cavity of ATP-dependent DNA helicase with ZINC76022877

4.3.3.4.11   Glutamine–fructose-6hosphate Aminotransferase (glmS)

Glucosamine-6-phosphate aminotransferase (GlmS) controls ammonia from glutamine at the glutaminase site to fructose 6-phosphate at the site of synthase. In *Escherichia coli* GlmS is consists of two C-terminal synthase domains that makes the dimer interface and two N-terminal glutaminase domains at its periphery. Addition of transposon Tn7 into the *E.coli* glmSwork astranscriptional terminator [58]. It is essential protein for the survival of bacteria and drug prioritization also confirm that it could act as drug target. Table 4.18 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.

TABLE 4.18: Compound name, MolDock Score and Interactive residues for
Glutamine–fructose-6-phosphate aminotransferase from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC82039595	2/Ser_402,Thr_303	-14.0216	-43.505
2	ZINC79037048	2/Ser_402,Thr_303	-13.7116	-12.706
3	ZINC70632494	Ser_402	-13.6374	-12.402
4	ZINC70632058	Ser_402,Lys_486,Thr_353	-13.6263	-15.631
5	ZINC77291968	2/Ser_402, Thr_353,Ser_304	-13.4775	-4.611

6	ZINC68603627	Ser_402, Thr_353,Ser_304	-13.3565	-24.017
7	ZINC71784976	Ser_402, 2/ Thr_353,Ser_304, Ser_350,Thr_303,Lys_486	-13.2932	-14.609
8	ZINC71784254	Ser_350	-13.2688	-21.223
9	ZINC83262599	Ser_304,Gln_349, Ser_348 ,Glu_489	-12.9355	-24.523
10	ZINC83240075	Ser_402,Thr_303,Glu_489	-12.7459	-13.654
11	ZINC71782238	2/Ser_402, 2/Thr_353, Ser_304,Ser_350,Ser_348	-12.6694	-16.840
12	ZINC83330155	2/Ser_402, Thr_353,Ser_350	-12.6640	-23.802
13	ZINC77257146	2/Ser_402, Thr_303, Lys_486,Val_400	-12.5982	-17.838
14	ZINC72332550	Thr_303,Ser_350	-12.5801	-13.232
15	ZINC83324289	2/Ser_402, Thr_303,Lys_486, Val_400,Ala_603	-12.5482	-25.679
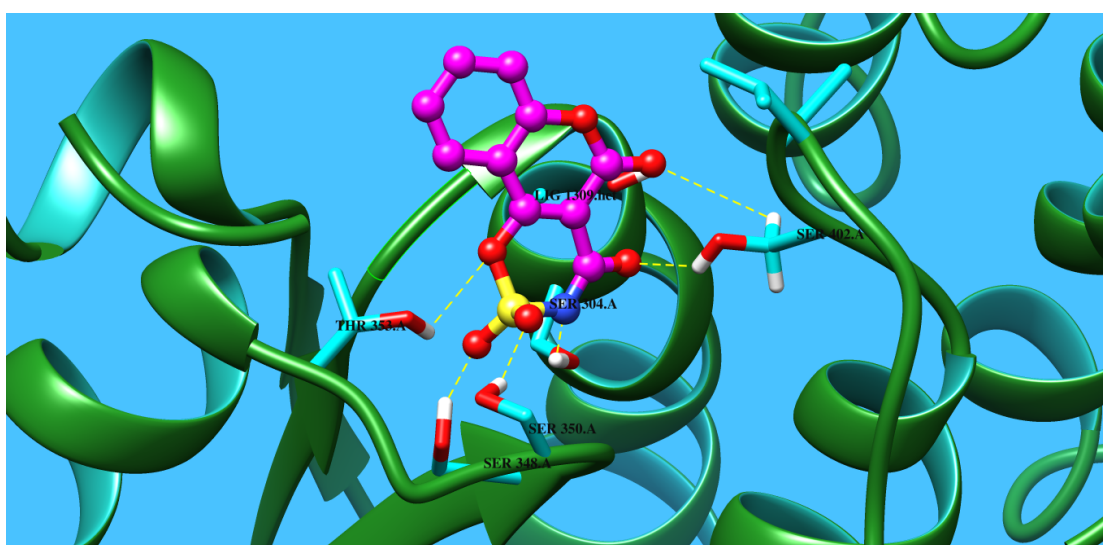
FIGURE 4.11: 3D Structure of Docking Analyses for most Druggable Protein Cavity of Glutamine–fructose-6-phosphate aminotransferase with ZINC 71782238

4.3.3.4.12 Ribonuclease P Protein Component (rnpA)

Ribonuclease P proteinis very old endonuclease that slice precursor tRNA and usually contains a catalytic RNA subunit (RPR) and one or more than one proteins (RPPs). It characterizes significant macromolecular complex and model system that is universally distributed in life. The gene that codes for the RNA subunit of ribonuclease P (RNase P) is necessary in all free-living organisms mean plays a crucial role for their survival. This had inferred form evolutionary tree that RNA subunit is very old molecule and is enzyme in nature. It is a cytoplasmic protein which is essential for the survival of bacteria,this can be potential drug target after the wetlab research.

TABLE 4.19: Compound name, MolDock Score and Interactive residues for Ribonuclease P protein component from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC71777128	Lys_65_D,Arg_61_D,Arg_3_D	-13.7073	-25.890
2	ZINC72145573	Asp_68_D,Glu_2_D	-13.6183	-17.224
3	ZINC71609301	Lys_65_D,Arg_58_D,Arg_3_D	-13.5622	-6.998
4	ZINC77323423	Lys_65_D,Arg_61_D, Arg_3_D ,Lya_65_D	-13.4705	-26.072
5	ZINC71603173	Asp_68_D, Glu_2_D	-13.1498	-92.963
6	ZINC70632388	Glu_2_D, Lys_65_D	-13.0197	-10.080
7	ZINC73825281	Glu_2_D	-12.6744	-13.724
8	ZINC71372392	Asp_68_D,Glu_2_D,Arg_69_D	-12.6161	-2.365
9	ZINC83324781	Arg_3_D, Asp_68_D	-12.5094	-74.330
10	ZINC71781167	Lys_65_D,Arg_61_D, Arg_3_D,Arg_58_D	-12.5044	-17.359
11	ZINC68604796	Asp_68_D,Glu_2_D	-12.2190	-17.589
12	ZINC72232764	Asp_68_D	-12.0268	-71.656
13	ZINC70632508	Lys_65_D,Glu_2_D	-11.8937	-11.861
14	ZINC71774240	Arg_69_D,Glu_2_D,Ile_37_D	-11.8699	-8.960
15	ZINC83312242	Arg_3_D,Glu_2_D	-11.8425	-28.299

Table 4.19 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.

FIGURE 4.12: 3D Structure of Docking Analyses for most Druggable Protein Cavity of Ribonuclease P protein component with ZINC 71781167

4.3.3.4.13 tRNA N6-adenosine Threonylcarbamoyl Transferase (tsaD)

Threonylcarbamoyladenosine (t(6) A) is a customized nucleoside generally conserved in tRNAs in 3 kingdoms system of classification. The newly discovered genes for t(6) A synthesis, counting tsaC and tsaD, are necessary in prokaryotes without which prokaryotes cannot survive,but not necessary in yeast. tsaC and tsaD was identified as antibacterial targets even prior to their functions were recognized. It was verify that t(6) A is necessary in *Escherichia coli* for survival and a review of genome-wide essentiality studies illustrates that genes for t(6) A synthesis are crucial in the majority prokaryotes. It is essential protein for the survival of bacteria and drug prioritization also confirm that it could act as drug target. Table 4.20 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.

4.3.3.4.14 Pseudouridine Synthase (MYPU_2980)
Pseudouridine synthases belongs to group of enzymes that plays a vital role in

TABLE 4.20: Compound name, MolDock Score and Interactive residues for tRNA N6-adenosine threonylcarbamoyltransferase from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC70503687	Asp_11_E,ZN-7:1	-13.2577	-74.138
2	ZINC83312254	Glu_12_E, ZN-7:1	-12.0951	-43.728
3	ZINC72264470	Asp_11_E	-11.6701	-41.364
4	ZINC71610591	ZN-7:1	-11.5735	-64.341
5	ZINC72145573	Asp_11_E	-11.4157	-39.292
6	ZINC72266002	No hit	-11.0264	-34.057
7	ZINC72187410	ZN-7:1	-10.9755	-298.362
8	ZINC72267263	Asp_298_E	-10.9198	-30.267
9	ZINC71777792	Asp_11_E	-10.8572	-3.093
10	ZINC71372392	Asp_11_E	-10.6414	-6.900
11	ZINC83324781	Glu_12_E,Asp_298_E, His_115_E,His_139_E, His_111_E	-10.6205	-85.498
12	ZINC74941352	Glu_12_E	-10.5529	-42.480
13	ZINC70632388	No hit	-10.5129	-21.005
14	ZINC78991044	Glu_12_E	-10.5048	-52.835
15	ZINC77323423	ZN-7:1	-10.4942	-34.465
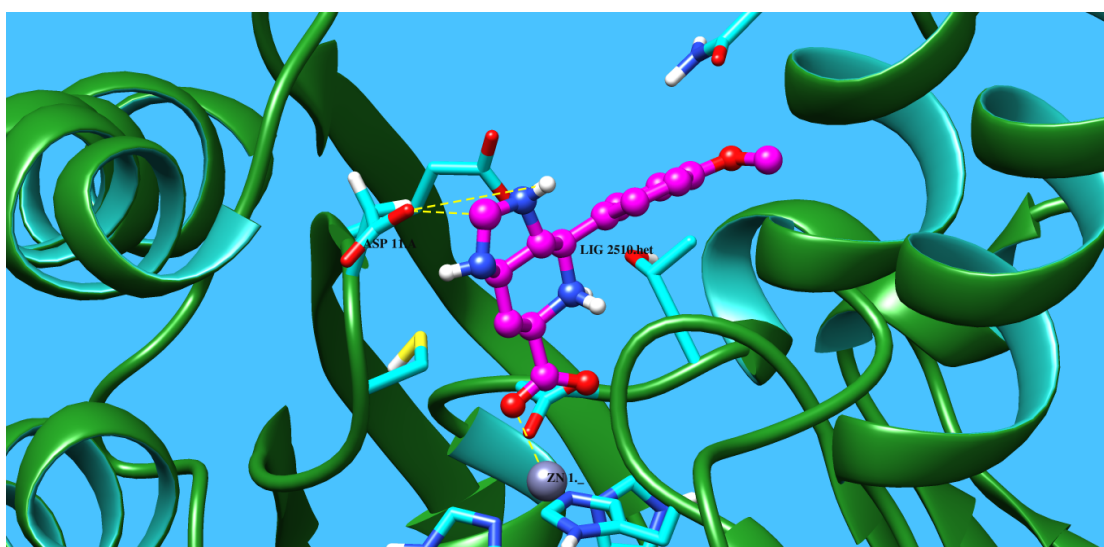
FIGURE 4.13: 3D Structure of Docking Analyses for most Druggable Protein Cavity of tRNA N6-adenosine threonylcarbamoyltransferase with ZINC 70503687

most abundant and phylogenetically conserved posttranscriptional modification of cellular RNAs, and appears to utilize both sequence and structural information to attain site specificity. Pseudouridine synthases catalyze an isomerization reaction of specific uridine residues contained by an RNA chain. Structure and Sequence

analyses have thus far demonstrated the existence of six synthase families that share a common core domain structure despite very low sequence identity.

TABLE 4.21: Compound name, MolDock Score and Interactive residues for Pseudouridine synthase from MOE

S.No	Ligand	Interactingresidues	Docking score	Energy (kcal/mol)
1	ZINC72360976	Tyr_65_	-12.7240	-16.166
2	ZINC71385043	Tyr_65_,His_234_,Arg_31_	-12.1696	-10.591
3	ZINC78903210	Tyr_65_	-12.1425	-12.038
4	ZINC83301304	Tyr_65_,His_34_, Arg_31_, Tyr_63_	-12.0788	-16.704
5	ZINC69533491	Tyr_65_,His_30_	-11.7710	-19.240
6	ZINC83429827	Tyr_65_,His_34_, Arg_31_, Asp_33-E	-11.7553	-16.153
7	ZINC72194195	Tyr_65_,His_30_, Arg_32_, Arg_63_	-11.7215	-19.744
8	ZINC78547526	Arg_31_,His_34_	-11.7139	-14.595
9	ZINC71777205	His_34_,Arg_31_,Gln_35_	-11.5940	-14.776
10	ZINC75629415	His_34_,Arg_32_,Tyr_65_	-11.5008	-16.669
11	ZINC71780763	Tyr_65_,Ile_36_	-11.4083	-21.586
12	ZINC71285802	Tyr_65_,Arg_31_	-11.3990	-9.802
13	ZINC71780174	Arg_31_,Arg_63-E,Tyr_65_	-11.3527	-14.881
14	ZINC79102849	Asp_33-E, His_30_,Tyr_65_	-11.2342	-31.298
15	ZINC78547466	Asp_33-E, Arg_31_, His_34_	-11.2211	-9.888

Synthases display exquisite specificity in pinpointing the site of pseudouridylation within their RNA substrates, and each enzyme achieves this by structural elaborations of the conserved catalytic domain. It is a cytoplasmic protein which is essential for the survival of bacteria,this can be potential drug target after the

wetlab research. Table 4.21 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.
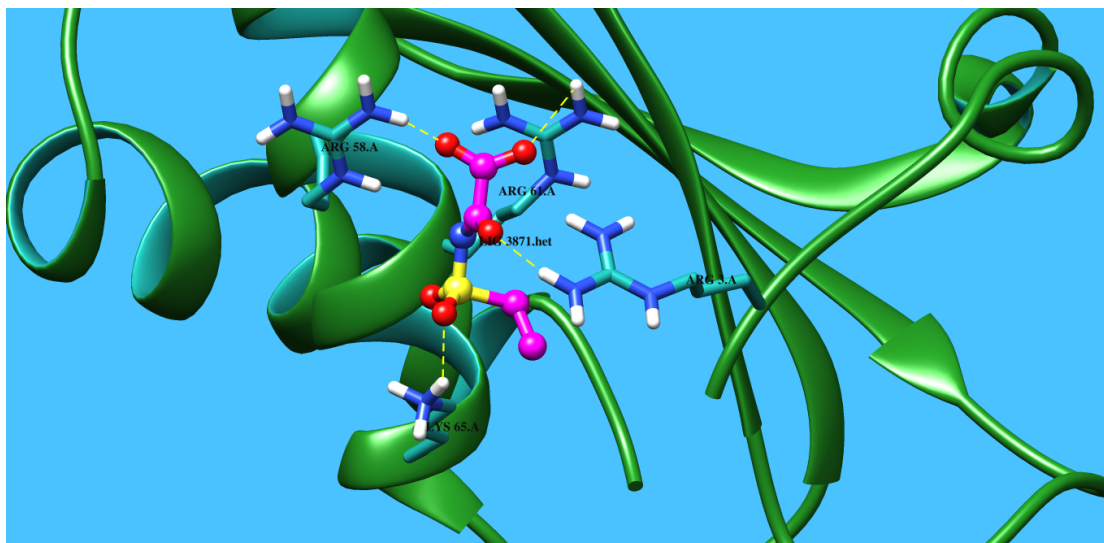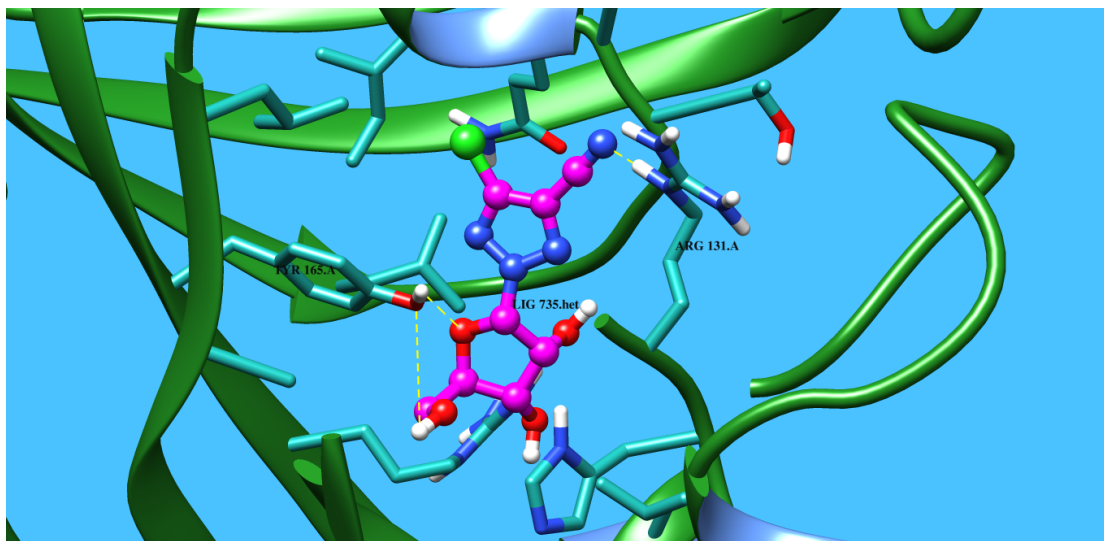

FIGURE 4.14: 3D Structure of Docking Analyses for most Druggable Protein Cavity of Pseudouridine synthase with ZINC83301304

4.3.3.4.15 Peptidase S24-like Protein (HMPREF 1604_ 02797)

Peptidase S24-like protein is cytoplasmic protein its molecular function is DNA binding. Peptidase is an enzyme it enhance the rate of proteolysis mean breakdown proteins into little polypeptides or single amino acids. This is typically done by breaking the peptide bonds within proteins by hydrolysis, a reaction where water breaks bonds.

Peptidase helps in several biological functions which includes digestion of proteins, protein catabolism [1][2] and cell signalling. It is essential protein for the survival of bacteria and drug prioritization also confirm that it could act as drug target. Table 4.22 illustrates a set of 15 potential ligands according to their MolDock score and minimized energy value from the above mentioned ZINC library.

Streptococcus gordonii, is a commensal (mutual benefit for host and bacteria) non-pathogenic bacteria among the normal flora of human oral microbiome. The specie is considered to play a significant role in initializing the formation of dental plaque or biofilm on teeth. It produces a surface where other bacteria can adhere. These

dental plaques later cause dental cavities and other damages to teeth health. This bacterium can enter blood as a result of an injury or wound and then colonizes blood clotting agents such as platelet fibrin thrombi, and starts damaging endocardium and heat valves.
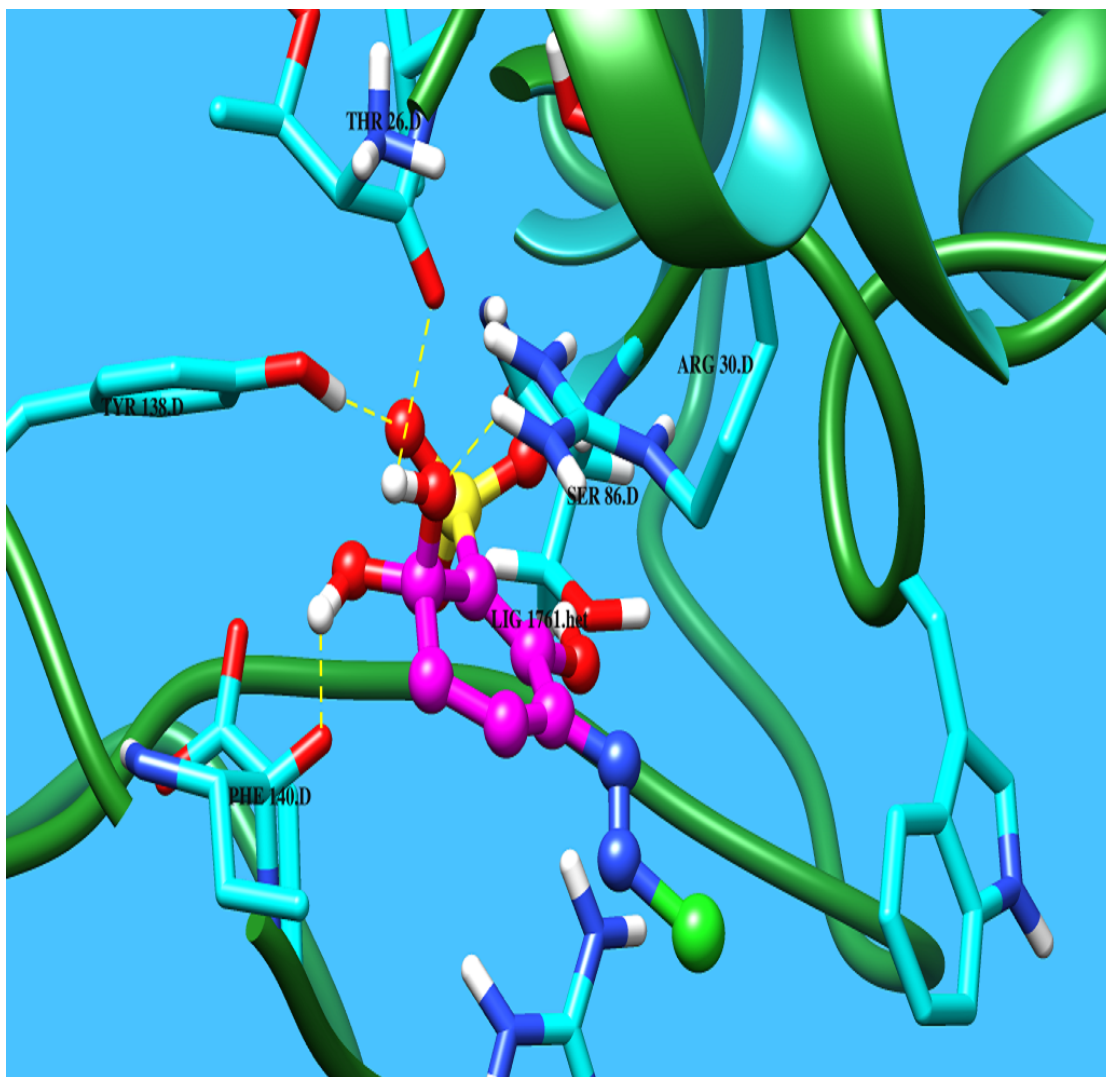
FIGURE 4.15: 3D Structure of Docking Analyses for most Druggable Protein Cavity Peptidase S24-like protein with ZINC83442116

Considering human as host, 15 of proteins (Glutamine–fructose-6-phosphate aminotransferase, UDP-3-O-acylglucosamine N-acyltransferase, Ribonuclease P protein component, tRNA N6-adenosine threonylcarbamoyltransferase, Pseudouridine synthase, Bifunctional protein, AMP nucleosidase, Probable DNA-directed RNA polymerase subunit delta, ATP-dependent DNA helicase, Peptidase S24-like protein, TetR family transcriptional regulator, Chromosomal replication initiator protein,

TABLE 4.22: Compound name, MolDock Score and Interactive residues for Peptidase S24-like protein from MOE

S.No	Ligand	Interacting residues	Docking score	Energy (kcal/mol)
1	ZINC71618323	2/Arg_30_B,Ser_86_B,2/Arg_144_B	-14.4376	-17.082
2	ZINC72145573	Arg_30_B,Ser_86_B,Tyr_138_B	-13.9080	-15.808
3	ZINC77537571	Tyr_138_B,Lys_110_B	-13.8724	-10.346
4	ZINC82039595	Arg_30_B, Lys_110_B	-13.0064	-17.612
5	ZINC68732144	Lys_110_B	-12.9061	-10.550
6	ZINC72291787	Lys_110_B,Phe_129_B	-12.7718	-30.866
7	ZINC83442116	2/Tyr_138_B, 2/Arg_30_B,Thr_26_B,Phe_140_B, Thr_26_B	-12.6838	-18.400
8	ZINC83313306	2/Lys_110_B	-12.5315	-20.514
9	ZINC74723614	Lys_110_B	-12.4334	-19.849
10	ZINC77257146	Arg_156_B,Arg_127_B,Glu_137_B	-12.3925	-19.032
11	ZINC83312254	Tyr_138_B, Phe_140_B, Arg_144_B, 2/Ser_86_B	-12.3571	-20.810
12	ZINC80053835	Lys_110_B, Tyr_138_B,Asp_141_B	-12.3067	-0.842
13	ZINC71766299	2/Lys_110_B	-12.0701	-11.298
14	ZINC71771223	2/Lys_110_B	-12.0502	-11.006
15	ZINC71783234	Arg_30_B,Ser_86_B,2/Arg_144_B, Phe_140_B	-12.0324	-15.964

UDP-N- acetyl muramoyl-tripeptide–D-alanyl-D-alanine ligase, HTH-type transcriptional regulator, UDP-N-acetyl mu- ramoyl-L-alanyl-D- glutamate–2,6- di-aminopimelate ligase) were considered as essential and non-host homologs, and have been subjected to virtual screening using ZINC library. The selected ligand molecules indicated positive interactions and can be novel drugs target of *Streptococcus gordonii.*

Chapter 5

Conclusions and Recommendations

Several bacteria are present on human skin, in nose area and most commonly in mouth area and gut in normal condition, sometimes lives throughout the life of human being and does not cause any disease but when they get chance or interact with other bacteria can cause diseases such type of bacteria are known as opportunistic bacteria.

5.1 Conclusions

The streptococci are pathogenic in nature but have very low virulence. *Streptococcus gordonii*, are primarily involved in development of biofilms on tooth surfaces also known as dental plaque they are also involved in development of infectious endocarditis i.e. inflammation of inner lining of heart. As treatment of this disease is very expensive and different bacteria has shown resistance against various commonly used antibiotics it is now important to find the novel therapeutic targets and drugs to cure and treat this disease. By using bioinformatics approach, determine the core proteins of 11 strains of *Streptococcus gordonii* and exploitation of 3D structure information of these proteins give us prioritize putative drug

65

and vaccines targets. In this study 20 proteins are pathogenic essential, non-host homologous. Out of which 15 are cytoplasmic and their comprehensive structural evaluation between the host and the pathogen proteins propose that these proteins can act as a therapeutic drug. Considering human as host these are the 15 proteins (Glutamine–fructose-6-phosphate aminotransferase, UDP-3-O-acylglucosamine N-acyltransferase, Ribonuclease P protein component, tRNA N6-adenosine threonyl-carbamoyltransferase, Pseudouridine synthase, Bifunctional protein, AMP nucleosidase, Probable DNA-directed RNA polymerase subunit delta, ATP-dependent DNA helicase, Peptidase S24-like protein, TetR family transcriptional regulator, Chromosomal replication initiator protein, UDP-N-acetylmuramoyl-tripeptide–D-alanyl-D-alanine ligase, HTH-type transcriptional regulator, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate–2,6-diaminopimelate ligase) considered as essential , non-host homologs and shows best interaction with the ligands so act as drug targets.

5.2   Recommendations

Future perspective of this study is to validate all 15 proteins in wet lab as various drugs are determined in-silico and then validated by wet lab techniques are now a day's used as medicine.

Bibliography

[1]. Liao, C. Y., Su, K. J., Lin, C. H., Huang, S. F., Chin, H. K., Chang, C. W., & Yeh, Y. C. (2016). Plantar purpura as the initial presentation of viridians Streptococcal shock syndrome secondary to Streptococcus gordonii bacteremia. Canadian Journal of Infectious Diseases and Medical Microbiology, 2016(9463895), pp.5.

[2]. Facklam, R. (2002). What happened to the streptococci: overview of taxonomic and nomenclature changes. Clinical microbiology reviews, 15(4), pp. 613-630.

[3]. Doern, C. D., & Burnham, C. A. D. (2010). It's not easy being green: the viridans group streptococci, with a focus on pediatric clinical manifestations. Journal of clinical microbiology, 48(11), pp. 3829-3835.

[4]. Ikeda, A., Nakajima, T., Konishi, T., Matsuzaki, K., Sugano, A., Fumikura, Y., & Jikuya, T. (2016). Infective endocarditis of an aorto-right atrial fistula caused by asymptomatic rupture of a sinus of Valsalva aneurysm: a case report. Surgical case reports, 2(1), pp. 1-4.

[5]. Teixeira, P. G., Thompson, E., Wartman, S., & Woo, K. (2014). Infective endocarditis associated superior mesenteric artery pseudoaneurysm. Annals of vascular surgery, 28(6), pp. 1563-e1.

[6]. cyr Yombi, J., Belkhir, L., Jonckheere, S., Wilmes, D., Cornu, O., Vandercam, B., & Rodriguez-Villalobos, H. (2012). Streptococcus gordonii septic arthritis: two cases and review of literature. BMC Infectious Diseases, 12(1), pp. 1-5.

[7]. Fenelon, C., Galbraith, J. G., Dalton, D. M., & Masterson, E. (2017). Streptococcus gordonii—a rare cause of prosthetic joint infection in a total hip replacement. Journal of surgical case reports, 2017(1), pp. 235.

[8]. Collazos, J., Martínez, E., & Mayo, J. (1999). Spontaneous bacterial peritonitis caused by Streptococcus gordonii. Journal of clinical gastroenterology, 28(1), pp. 45-46.

[9]. Cheung, C. Y., Cheng, N. H., Chau, K. F., & Li, C. S. (2011). Streptococcus gordonii peritonitis in a patient on CAPD. Renal failure, 33(2), pp. 242-243.

[10]. https://microbewiki.kenyon.edu/index.php/File:Streptococcusgordonii.jpg

[11]. Murdoch, D. R., Corey, G. R., Hoen, B., Miró, J. M., Fowler, V. G., Bayer, A. S., & Chambers, S. T. (2009). Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century: the International Collaboration on Endocarditis–Prospective Cohort Study. Archives of internal medicine, 169(5), pp. 463-473.

[12]. Vickerman, M. M., Minick, P. E., & Mather, N. M. (2001). Characterization of the Streptococcus gordonii chromosomal region immediately downstream of the glucosyltransferase gene. Microbiology, 147(11), pp. 3061-3070.

[13]. Yajima, A., Y. Takahashi, and K. Konishi. 2005. Identification of platelet receptors for the Streptococcus gordonii DL1 sialic acid-binding adhesin. Microbiol. Immunol. 49, pp. 795-800.

[14]. Gonzalez-Barca, E., Fernandez-Sevilla, A., Carratala, J., Granena, A., & Gudiol, F. (1996). Prospective study of 288 episodes of bacteremia in neutropenic cancer patients in a single institution. European Journal of Clinical Microbiology and Infectious Diseases, 15(4), pp. 291-296.

[15]. Kern, W., Kurrle, E., & Schmeiser, T. (1990). Streptococcal bacteremia in adult patients with leukemia undergoing aggressive chemotherapy. A review of 55 cases. Infection, 18(3), pp. 138-145.

[16]. Ofek, I., Hasty, D. L., & Sharon, N. (2003). Anti-adhesion therapy of bacterial diseases: prospects and problems. FEMS Immunology & Medical Microbiology, 38(3), pp. 181-191.

[17]. Tilley, D., & Kerrigan, S. W. (2013). Platelet-Bacterial Interactions in the Pathogenesis of Infective Endocarditis—Part I: The Streptococcus. Recent Advances in Infective Endocarditis, pp. 13-33.

[18]. Chong, C. E., Lim, B. S., Nathan, S., & Mohamed, R. (2006). In silico analysis of Burkholderia pseudomallei genome sequence for potential drug targets. In silico biology, 6(4), pp. 341-346.

[19]. Dutta, A., Singh, S. K., Ghosh, P., Mukherjee, R., Mitter, S., & Bandyopadhyay, D. (2006). In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori. In silico biology, 6(1, 2), pp. 43-47.

[20]. Sakharkar, K. R., Sakharkar, M. K., & Chow, V. T. (2004). A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In silico biology, 4(3), pp. 355-360.

[21]. Barh, D., & Kumar, A. (2009). In silico identification of candidate drug and vaccine targets from various pathways in Neisseria gonorrhoeae. In silico biology, 9(4), pp. 225-231.

[22]. Rathi, B., Sarangi, A. N., & Trivedi, N. (2009). Genome subtraction for novel target definition in Salmonella typhi. Bioinformation, 4(4), 143.

[23]. Aronov, A. M., Verlinde, C. L., Hol, W. G., & Gelb, M. H. (1998). Selective tight binding inhibitors of trypanosomal glyceraldehyde-3-phosphate dehydrogenase via structure-based drug design. Journal of medicinal chemistry, 41(24), pp. 4790-4799.

[24]. Singh, S., Malik, B. K., & Sharma, D. K. (2008). Molecular modeling and docking analysis of Entamoeba histolytica glyceraldehyde-3 phosphate dehydrogenase, a potential target enzyme for anti-protozoal drug development. Chemical biology & drug design, 71(6), pp. 554-562.

[25]. Hassan, S. S., Tiwari, S., Guimarães, L. C., Jamal, S. B., Folador, E., Sharma, N. B., & Póvoa, F. D. (2014). Proteome scale comparative modeling for conserved drug and vaccine targets identification in Corynebacterium pseudotuberculosis. BMC genomics, 15(S7), pp. 3.

[26]. Jamal, S. B., Hassan, S. S., Tiwari, S., Viana, M. V., Benevides, L. D. J., Ullah, A., & Silva, A. (2017). An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae. PloS one, 12(10), pp. e0186401.

[27]. Yoneda, T., Hagino, H., Sugimoto, T., Ohta, H., Takahashi, S., Soen, S., & Urade, M. (2010). Bisphosphonate-related osteonecrosis of the jaw: position paper from the allied task force committee of Japanese society for bone and mineral research, Japan osteoporosis society, Japanese society of periodontology, Japanese society for oral and maxillofacial radiology, and Japanese society of oral and maxillofacial surgeons. Journal of bone and mineral metabolism, 28(4), pp. 365-383.

[28]. Herzberg M C, Meyer M W. Effects of oral flora on platelets: possible consequences in cardiovascular disease. J Periodontol. 1996;67: pp.1138–1142.

[29]. Salvà-Serra, F., Jakobsson, H. E., Thorell, K., Gonzales-Siles, L., Hallbäck, E. T., Jaén-Luchoro, D., & Bennasar, A. (2016). Draft genome sequence of Streptococcus gordonii type strain CCUG 33482T. Genome Announcements, 4(2), pp. 653-666.

[30]. Telford, J. L., Barocchi, M. A., Margarit, I., Rappuoli, R., & Grandi, G. (2006). Pili in gram-positive pathogens. Nature Reviews Microbiology, 4(7), pp. 509-519.

[31]. Kiliç, A. O., Tao, L., Zhang, Y., Lei, Y., Khammanivong, A., & Herzberg, M. C. (2004). Involvement of *Streptococcus gordonii* beta-glucoside metabolism systems in adhesion, biofilm formation, and in vivo gene expression. Journal of bacteriology, 186(13), pp. 4246-4253.

[32]. Plummer, C., & Douglas, C. W. I. (2006). Relationship between the ability of oral streptococci to interact with platelet glycoprotein Iba and with the salivary low-molecular-weight mucin, MG2. FEMS Immunology & Medical Microbiology, 48(3), pp. 390-399.

[33]. Marsh, P. D. (2003). Are dental diseases examples of ecological catastrophes?. Microbiology, 149(2), pp. 279-294.

[34]. Mosailova, N., Truong, J., Dietrich, T., & amp; Ashurst, J. (2019). Streptococcus gordonii: A Rare Cause of Infective Endocarditis. Case Reports in Infectious Diseases, 7127848, pp. 562-570.

[35]. Njuguna, B., Gardner, A., Karwa, R., & Delahaye, F. (2017). Infective endocarditis in low-and middle-income countries. Cardiology Clinics, 35(1), pp. 153-163.

[36]. Ambrosioni, J., Hernandez-Meneses, M., Téllez, A., Pericàs, J., Falces, C., Tolosana, J. M., & Moreno, A. (2017). The changing epidemiology of infective endocarditis in the twenty-first century. Current infectious disease reports, 19(5), pp. 21.

[37]. Dadon, Z., Cohen, A., Szterenlicht, Y. M., Assous, M. V., Barzilay, Y., Raveh-Brawer, D., & Munter, G. (2017). Spondylodiskitis and endocarditis due to Streptococcus gordonii. Annals of Clinical Microbiology and Antimicrobials, 16(1), pp. 68.

[38]. Li, X., Kolltveit, K. M., Tronstad, L., & Olsen, I. (2000). Systemic diseases caused by oral infection. Clinical microbiology reviews, 13(4), pp. 547-558.

[39]. Tariq, M., Alam, M., Munir, G., Khan, M. A., & Smego Jr, R. A. (2004). Infective endocarditis: a five-year experience at a tertiary care hospital in Pakistan. International journal of infectious diseases, 8(3), pp. 163-170.

[40]. Shahid, U., Sharif, H., Farooqi, J., Jamil, B., & Khan, E. (2018). Microbiological and clinical profile of infective endocarditis patients: an observational

study experience from tertiary care center Karachi Pakistan. Journal of cardiothoracic surgery, 13(1), pp. 94.

[41]. Netzer, R. O. M., Altwegg, S. C., Zollinger, E., Täuber, M., Carrel, T., & Seiler, C. (2002). Infective endocarditis: determinants of long term outcome. Heart, 88(1), pp. 61-66.

[42]. Durante-Mangoni, E., Bradley, S., Selton-Suty, C., Tripodi, M. F., Barsic, B., Bouza, E., & Koneçny, P. (2008). Current features of infective endocarditis in elderly patients: results of the International Collaboration on Endocarditis Prospective Cohort Study. Archives of internal medicine, 168(19), pp. 2095-2103.

[43]. Carlos Guimaraes, L., Benevides de Jesus, L., Vinicius Canario Viana, M., Silva, A., Thiago Juca Ramos, R., de Castro Soares, S., & Azevedo, V. (2015). Inside the pan-genome-methods and software overview. Current genomics, 16(4), pp. 245-252.

[44]. Marsh, P. D. (2003). Are dental diseases examples of ecological catastrophes?. Microbiology, 149(2), pp. 279-294.

[45]. Hinse, D., Vollmer, T., Rückert, C., Blom, J., Kalinowski, J., Knabbe, C., & Dreier, J. (2011). Complete genome and comparative analysis of Streptococcus gallolyticus subsp. gallolyticus, an emerging pathogen of infective endocarditis. BMC genomics, 12(1), pp. 400.

[46]. Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., & Goesmann, A. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic acids research, 44(W1), pp. W22-W28.

[47]. Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome research, 13(9), pp. 2178-2189.

[48]. Emmerth, M., Goebel, W., Miller, S. I., & Hueck, C. J. (1999). Genomic Subtraction Identifies Salmonella typhimurium Prophages, F-Related Plasmid

Sequences, and a Novel Fimbrial Operon, stf, Which Are Absent inSalmonella typhi. Journal of bacteriology, 181(18), pp. 5652-5661.

[49]. Zhang, Z., & Ren, Q. (2015). Why are essential genes essential?-The essentiality of Saccharomyces genes. Microbial cell, 2(8), pp. 280.

[50]. Uddin, R., Siraj, B., Rashid, M., Khan, A., Ahsan Halim, S., & Al-Harrasi, A. (2020). Genome Subtraction and Comparison for the Identification of Novel Drug Targets against Mycobacterium avium subsp. hominissuis. Pathogens, 9(5), pp. 368.

[51]. Wadood, A., Jamal, S. B., Riaz, M., & Mir, A. (2014). Computational analysis of benzofuran-2-carboxlic acids as potent Pim-1 kinase inhibitors. Pharmaceutical biology, 52(9), pp. 1170-1178.

[52]. Trost, E., Blom, J., de Castro Soares, S., Huang, I. H., Al-Dilaimi, A., Schröder, J., & Azevedo, V. (2012). Pangenomic study of Corynebacterium diphtheriae that provides insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. Journal of bacteriology, 194(12), pp. 3199-3215.

[53]. Luo, H., Lin, Y., Gao, F., Zhang, C. T., & Zhang, R. (2014). DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic acids research, 42(D1), pp. D574-D580.

[54]. Agüero, F., Al-Lazikani, B., Aslett, M., Berriman, M., Buckner, F. S., Campbell, R. K., & Crowther, G. J. (2008). Genomic-scale prioritization of drug targets: the TDR Targets database. Nature reviews Drug discovery, 7(11), pp. 900-907.

[55]. Brennan, R. G., & Matthews, B. W. (1989). The helix-turn-helix DNA binding motif. Journal of Biological Chemistry, 264(4), pp. 1903-1906.

[56]. Gay, N. J., Tybulewicz, V. L., & Walker, J. E. (1986). Insertion of transposon Tn7 into the Escherichia coli glm S transcriptional terminator. Biochemical Journal, 234(1), pp. 111-117.

[57]. Luscombe, N. M., Austin, S. E., Berman, H. M., & Thornton, J. M. (2000). An overview of the structures of protein-DNA complexes. Genome biology, 1(1), pp. 10-30.

[58]. Gay NJ, Tybulewicz VL, Walker JE. Insertion of transposon Tn7 into the Escherichia coli glmS transcriptional terminator. Biochem J. 1986;234(1), pp.111-117. doi:10.1042/bj2340111