CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Analysis of Pangenome of *Streptococcus sanguinis* for Drug Targeting

by

Shumaila Amjad

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

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2022

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

CERTIFICATE OF APPROVAL

Analysis of Pangenome of *Streptococcus sanguinis* for Drug Targeting

by

Shumaila Amjad

(MBS203026)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Bashir Ahmed	IIUI, Islamabad
(b)	Internal Examiner	Dr. Erum Dilshad	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
November, 2022

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinformatics & Biosciences
November, 2022

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

Author's Declaration

I, **Shumaila Amjad** hereby state that my MS thesis titled "**Analysis of Pangenome of *Streptococcus sanguinis* for Drug Targeting.**" 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.

(Shumaila Amjad)

Registration No: MBS203026

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Analysis of Pangenome of *Streptococcus sanguinis* for Drug Targeting.** " 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.

(Shumaila Amjad)

Registration No: MBS203026

Acknowledgement

All thanks to Allah, the Most Gracious and Merciful Alhamdulillah, for granting me strength and blessings in completing my MS thesis. First and foremost, I'd want to thank Capital University for providing me with the chance to pursue my dream of advancing my education. I'd like to begin by thanking my supervisor, Dr. Syeda Marriam Bakhtiar, Assistant Professor, CUST, for her encouragement, support, and advice during the thesis completion. She was quite helpful and promptly responded to all of my questions. I'd want to express my heartfelt prayers and gratitude to her. I'd want to thank all of my respected instructors, including, Dr. Sania Riaz, Dr. Sohail Ahmad Jan, our honourable HoD, Dr. Sahar Fazal and all the faculty members of Biosciences at CUST university. Finally, I want to thank my parents and siblings for their support, cooperation and assistance during my years of education and the process of conducting research and writing this thesis successfully.

(Shumaila Amjad)

Abstract

Streptococcus sanguinis belongs to the streptococcus viridans group, a Gram-positive, facultative aerobic bacterium. It is prevalent in human mouths in dental plaque, which causes both dental cavities and bacterial endocarditis and has a 25% death rate. Furthermore, *S. sanguinis* is regarded as opportunistic pathogen and one of the most common causes of infective endocarditis. The goal of this study is to look at the involvement of *Streptococcus sanguinis* in virulence, pathogenicity, biofilm development, and infection mechanism. *S. sanguinis* are part of the human oral microbiota and have the capacity to create dental plaques, thus they might play a role in the development of IE. The pan genome method was used in this investigation to identify therapeutic targets for *Streptococcus sanguinis sk36* strain and design an anti-infection medicine because traditional drugs generate health hazards and sometimes it's not effective for human. Number of bioinformatics tools was used to retrieve essential proteins, find out metabolic pathway, defining the pockets, active sites and draggability parameters. In this investigation only two proteins were selected for drug targeting and docked against antibacterial compounds. Following docking, the top ten compounds against each protein are chosen based on binding affinity, and the 2D structure of one docked compound with the greatest binding affinity or number of residues interaction is revealed. Our findings aided in determining the sort of protein to be regarded as an effective therapeutic target. This study will help researchers create and identify more effective and targeted anti- *S. sanguinis* treatment drugs.

Contents

List of Figures

List of Tables

Abbreviations

ABE: Acute bacterial endocarditis

CADD: Computer aided drug designing.

COG: Cluster of orthologous group

CSP: Competance stimulating peptide.

EPS: Exopolysaccharides

GtFP: Glocosyltransferase protein.

HK: Histidine kinase.

RR: Response regulator

SAFJ: Septic arthritis of facet joint.

SBE: Sb acute bacterial endocarditis.

TCS: Two component system.

Chapter 1

Introduction

Streptococcus sanguinis, originally called *Streptococcus sanguis*, is a facultative, anaerobic and Gram-positive bacterium related to the group of *Viridans Streptococcus* . *S. sanguinis* is a common resident in the oral cavity of the healthy human, in the cavity it causes dental plaque, and alters the environment to develop them less friendly for other Streptococcus strains that cause cavitiy in human teeth, such as *Streptococcus mutans*. It's a Gram-positive, facultative anaerobe that doesn't produce spores. *S. sanguinis*, like other streptococci, present in chains or pairs of cocci, divides along a single axis [1].

S. sanguinis is a Gram-positive bacterium with a dense cell wall made up of peptidoglycan (also known as murein) and teichoic acid, murein is unique character for bacteria and gives them its form and stiffness [2]. Although some *S. sanguis* members exhibit an unusual "twitching motility" by using their polar fimbriae of type IV in a grappling connector fashion to translocate throughout surfaces.

The name "*sanguinis* (previously *sanguis*)" comes from a group of streptococcal strains obtained from individuals suffering from subacute bacterial endocarditis [3]. *S. sanguinis* is the most often detected species for causing native-valve infection among mitis group members, according to epidemiological investigations. As a result, its virulence potential has been recognized a causative link between infectious cardiovascular disease development and develop bacterial contact with the

immune system and human tissues. Due to its antagonistic connection with the cariogenic pathogen *Streptococcus mutans*, *Streptococcus sanguinis* is known as an important key colonizer of teeth and is generally regarded to be beneficial. *S. sanguinis* on the other hand can behave as an opportunistic pathogen, colonizing a damaged heart valve and inducing infective endocarditis if it enters the blood circulation [4].

Gram-positive pathogens' surface proteins are covalently connected to their thick cell walls and perform a role in a various biological activities, including identification of host matrix molecules, contact with human cells and immune system evasion. *S. sanguinis* surface proteins have been linked to the development of infective endocarditis. Those proteins usually have a cell wall sorting signal at the C-terminus, which is made up of a pentapeptide LPXTG motif followed by side chains which are hydrophobic in nature and a positively charged tail. These proteins are digested by the transpeptidase sortase A after being directed to the bacterial surface (SrtA). SrtA catalyzes a covalent bond between a free amino group in the developing cell wall and the carboxyl group of threonine in the LPXTG motif. In *Sanguinis* strain, thirty-three potential cell wall anchored proteins were discovered. However, the bulk of those proteins' biological functions is unknown. Initially, a salivary pellicle is produced on the enamel or cemental surface, which is quickly colonized by microorganisms, like bacteria. By adhering to the pellicle-coated surface, *Streptococcus sanguis* binds both specifically and non-specifically. Co-adherence also allows additional microorganisms to attach to the first plaque forms [4].

S. sanguinis SK36 contains the highest streptococcal genome, with 2,388,435 bp, which is 7 to 24 percent bigger than previous reported genomes (285). The amount of expected encoded proteins reflects the variation in genome size between different species. *S. sanguinis SK36* is expected to encode 2,274 proteins [5]. Carlsson and colleagues were among the first to characterize *S. sanguinis's* taxonomic and ecological characteristics in the oral cavity. Its GC content is 43.4 percent greater than that of other streptococci [6]. The act of single cell is to adhere on hard surface and fimbrie helps to adherence of *S. sanguinis* [7].

1.1 Drug Target

Drug discovery is the process of finding potential new medicinal substances by combining computational, translational, experimental, and clinical design approaches. Despite breakthroughs in biotechnology and our understanding of biological systems, drug development remains time-consuming, expensive, complex, and ineffective process with high attrition rate of novel therapeutics. Medication-targeting has the potential to increase drug delivery effectiveness, decrease adverse effects, and lower treatment costs significantly.

Drug design is the process of discovery of novel treatments based on a biological target's information. Drug design, at its earlier level, is creating molecules that are complementary in form and charge to the molecular target by which they interact and bind. Drugs are required for illness therapy. Many illnesses are continually threatening human life. As a result, optimal medications are in high demand. Computer-aided drug design (CADD) method are becoming increasingly important in drug development, and they are vital in identifying viable therapeutic candidates at a low cost. These computational methods are useful for reducing the use of animal models in pharmacological research, assisting in the rational design of novel and safe drug candidates, and repositioning marketed drugs, as well as assisting medicinal chemists and pharmacologists throughout the drug discovery process.

1.2 Problem Statement

Streptococcus sanguinis (S. sanguinis) is a common oral occupant commensal that, if it enters the circulation, can cause widespread human illness. Infectious endocarditis is the most serious disease. Initially, *Streptococcus sanguinis* makes a biofilm on the tooth surface in humans and damages oral health. Secondly, when it enters the bloodstream it damages the heart valves or breaks down the tissues of valves and makes them leaky valves which causes heart blockage. Without the

treatment, it could not cure. Some traditional methods, herbal medicines, and pants are available to reduce its effects. Interaction between a target protein and ligand needs to be studied for drug design.

1.3 Aims and Objectives

The major purpose of this research is to understand the genomic diversity of the *Streptococcus sanguinis* strain and identification of novel drugs against this pathogen. The study is designed with following objective in order to attain this aim.

1. To explore the pan-genome and essential genome of *Streptococcus sanguinis*.
2. To identify putative pathogenic variables in *Streptococcus sanguinis* and assessment as pharmacological target .
3. To investigate the efficacy of natural substances as a potential drug against *Streptococcus sanguinis*.

Chapter 2

Review of Literature

2.1 Characteristics of *Streptococcus sanguinis*

Streptococci are nonmotile, non-spore-forming, Gram-positive catalase-negative cocci that are present in colonies or in pair form. Gram-positive cultures may go extinct in older societies. Few Streptococci are obligatory (strict) anaerobes, whereas on the hand some are facultative anaerobes. The majority of people demand richer media (blood agar). A hyaluronic acid capsule is found on Group A *Streptococci* [8].

Members of the indigenous microbiota of the human oral dental cavity make up the diverse collection of oral bacteria that make up the *sanguinis (sanguis) streptococci*. While the link between *Streptococcus sanguinis* and bacterial endocarditis has been well documented, *S. sanguinis* is assumed to behave as benign, if not helpful, role in the oral cavity. However, little is known about S. sanguinis natural history or its association with other oral bacteria. We investigated the initial acquisition of *S. sanguinis* and described its colonization relative to tooth emergence and its proportions in plaque and saliva as a function of other biological processes, including subsequent colonization with mutans, as part of longitudinal study concerning the transmission and acquisition of oral pioneer bacteria within mother-infant [9]. Oral Mitis *Streptococci* are part of the commensal flora in the

human oral cavity and are also the first colonizers in the production of dental biofilms [10].

2.1.1 Taxonomy

Streptococcus sanguinis is the most common viridans type of *Streptococcus* found in patients with subacute bacterial endocarditis. It is difficult to determine whether any of the newly identified and re-identified species of viridans *Streptococci* are associated with human disease. Although the specific species were not well defined in this study, *S. mitis* was the most frequently detected and most likely resistant group of species [11].

2.1.2 Taxonomic Hierarchy of *Streptococcus sanguinis*

TABLE 2.1: Taxonomic Hierarchy of *Streptococcus sanguinis*.

Kingdom	**Bacteria**
Subkingdom	Posibacteria
Phylum	Firmicutes
Class	Bacilli
Order	Lactobacillales
Family	Streptococcaceae
Genus	*Streptococcus*
Species	*Streptococcus sanguinis.*

2.1.3 Definition and Importance

Streptococcus sanguinis is a gram-positive nonmotile spore-framing cocci that might be found in the solid human mouth this microorganism is generally found in dental plaque and it can colonize tooth pits in the oral depression sanguinis is a commensal microbes that are broadly spread it could be found on the surfaces

of teeth oral mucosa and human spit *Streptococcus sanguinis* is pervasive in both subgingival and supragingival plaque as a facultative anaerobic animal varieties its likewise considered normal found in the flow where it can taint heart valves and cause bacterial endocarditis a lethal cardiac condition that can prompt passing [13].

2.1.4 Discovery

Carlsson and colleagues were among the first to characterize *S. sanguinis's* taxonomic and ecological characteristics in the oral cavity. It was the Carlsson group that discovered that S. sanguinis colonizes babies following the eruption of teeth and that S. sanguinis colonization precedes that of mutans streptococci [14].

2.1.5 Genome

Laboratories at Virginia Commonwealth University determined the full genomic sequence of *S. sanguinis* in 2007. The circular DNA molecule in Streptococcus sanguinis comprises 2,388,435 base pairs, , which ranges from 177 to 590 kb larger than the commonly known streptococcal genomes.

Overall G+C concentration of the *S. sanguinis* genome is 43.4 percent, which is much higher than to the G+C contents of other streptococci strains. There are 2,274 predicted proteins in the genome, four rRNA operons and 61 tRNAs in the genome [8].

Horizontal gene transfer appears to have gained a 70-kb area containing pathways for vitamin B12 production and the breakdown of ethanolamine and propanediol.

The gene complement of *S. sanguinis* varies from the gene complements of other pathogenic and Nonpathogenic *Streptococci*, suggesting novel ideas for pathogenesis and virulence. *S. sanguinis*, in particular, has a surprising number of putative surface proteins [15]. The figure below 2.1 shows the Circular *S. sanguinis* SK36 genome map.

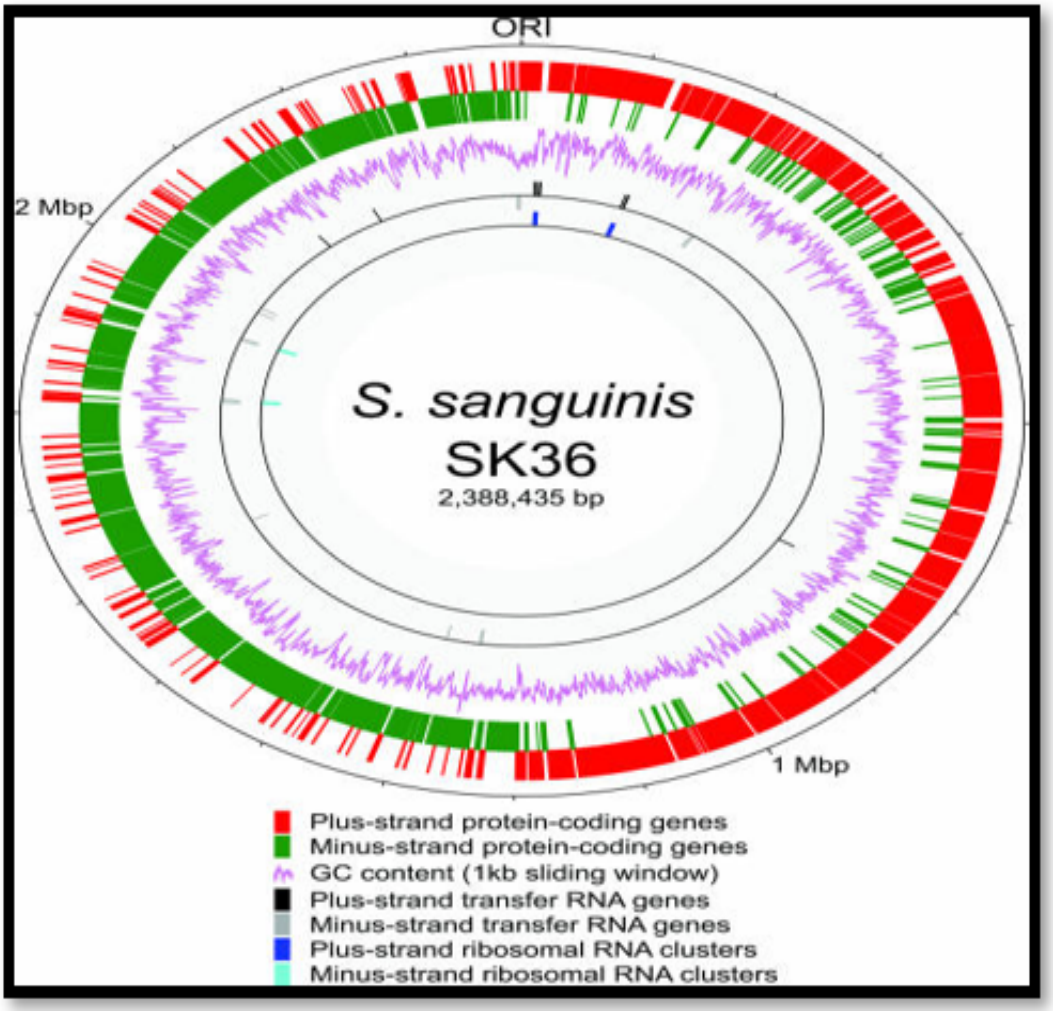

FIGURE 2.1: Circular *S. sanguinis* SK36 genome map. Starting from the outside, the circles show (i) the genome positions (in base pairs) starting from the origin of replication (ORI);(ii and iii) predicted coding regions on the two strands (different colors are used for clarity); (iv) G+C content (in 1-kb); (v and vi) rRNA clusters on the two strands; and (vii and viii) tRNA on the two strands.

2.1.6 The Relationship of *S. sanguinis* with Dental Caries

Dental caries is a chronic, contagious illness that causes demineralization of dental hard tissues. It is produced by the adhesion of specific microorganisms to tooth surfaces via the production of the biofilm known as dental plaque, followed by sugar metabolism into organic acids and, as a result, enamel disintegration. *Mutans Streptococci, Lactobacillus spp., Bifidobacterium spp,* and *Actinomyces spp* are among the bacteria that can induce an acidic environment in oral biofilms. *Candida albicans* also plays a significant role in caries pathogeneses [16]. *Becker*

et al. discovered that *S. sanguinis* was the only species identified that was significantly related with oral health when comparing caries-active and caries-free children by comparing colony counts of caries and caries-free samples. As sequencing technology have advanced, more 16S rRNA gene sequencing and metagenomic investigations have been conducted to identify bacterial species that are favourably or negatively associated to dentalcaries. *Streptococcus sanguinis* has been found to be commonly, though not always, related with oral health in these investigations [17].

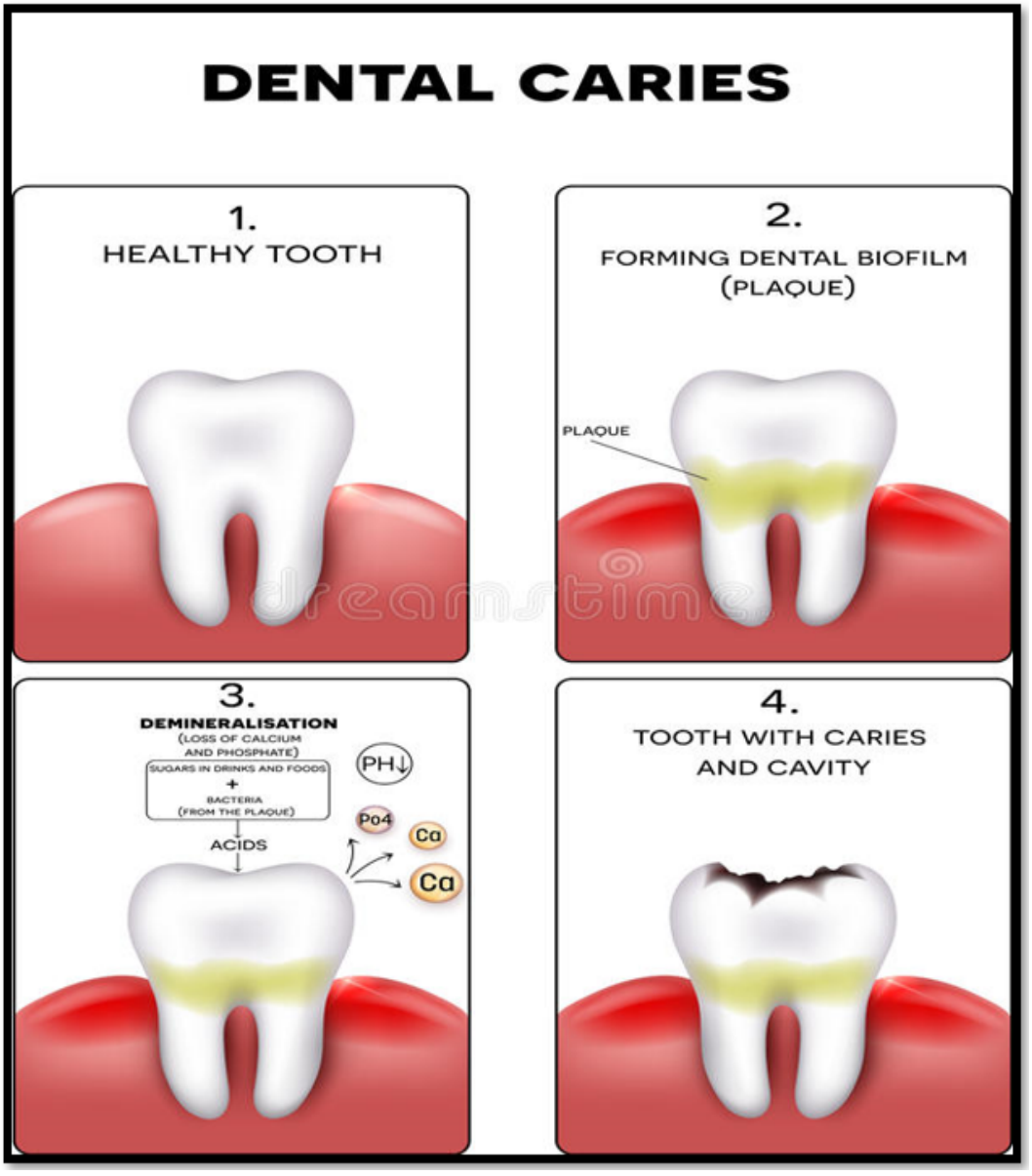

FIGURE 2.2: Shows formation of dental plaque (biofilm development), phosphate and calcium loss (demineralization) and eventually caries and cavities occur in tooth.

2.2 Ecology

The interaction between *Streptococci: S. sanguinis* and *S. mutans* can be influenced by environmental variables such as cell density, nutritional availability, and pH. Chemicals generated by bacteria can be used to bring such interactions. *S. sanguinis* produces hydrogen peroxide, which inhibits the development of *S. mutans* when present. Hydrogen peroxide synthesis is shut off in a nutrient-rich environment, and the energy necessary for hydrogen peroxide generation is instead utilized for cell development. If there are insufficient nutrients or the pH of the environment is too low, hydrogen peroxide production is activated, allowing *S. mutans* to compete. As a result, the environment controls *S. sanguinis* acts as a tether for other oral microbes to colonize the tooth surface, produce dental plaque, and contribute to the development of caries and periodontal disease once bonded [18].

S. sanguinis may also prevent *Streptococcus mutans*, the principal species linked to dental caries, from colonizing the tooth, therefore its presence may be advantageous to oral health. *S. sanguinis* acts as a tether for other oral microbes to colonize the tooth surface, produce dental plaque, and contribute to the development of caries and periodontal disease once bonded.

S. sanguinis may also prevent *Streptococcus mutans*, the major species linked to dental caries, from colonizing the tooth, therefore its presence might be harmful to competition and coexistence between *S. sanguis* and *S. mutans* [19].

2.3 Mode of Action

By inactivating the *S. mutans* competence-stimulating peptide (CSP), a quorum-sensing signal that induces mutacin gene expression, *Streptococcus sanguinis* has been demonstrated to limit mutacin synthesis. *S. sanguinis* method may be comparable to that of *S. gordonii*, which generates a challisin-like protease that destroys *Streptococcus mutans* competence-stimulating peptide [20].

2.3.1 *Streptococcus sanguinis* in Early Childhood Caries

Dental decay is a biofilm-induced, mucus acidic demineralization of the teeth's enamel and dentin. Early childhood caries (ECC) serves particularly aggressive form of caries, beginning on smooth surfaces, advancing quickly, and having a negative long-term effect on the dentition. Dental caries develops as a result of a sucrose-induced ecological shift in the bacterial makeup toward cariogenic species [21]. *Streptococcus sanguinis* is a common tooth coloniser whose presence

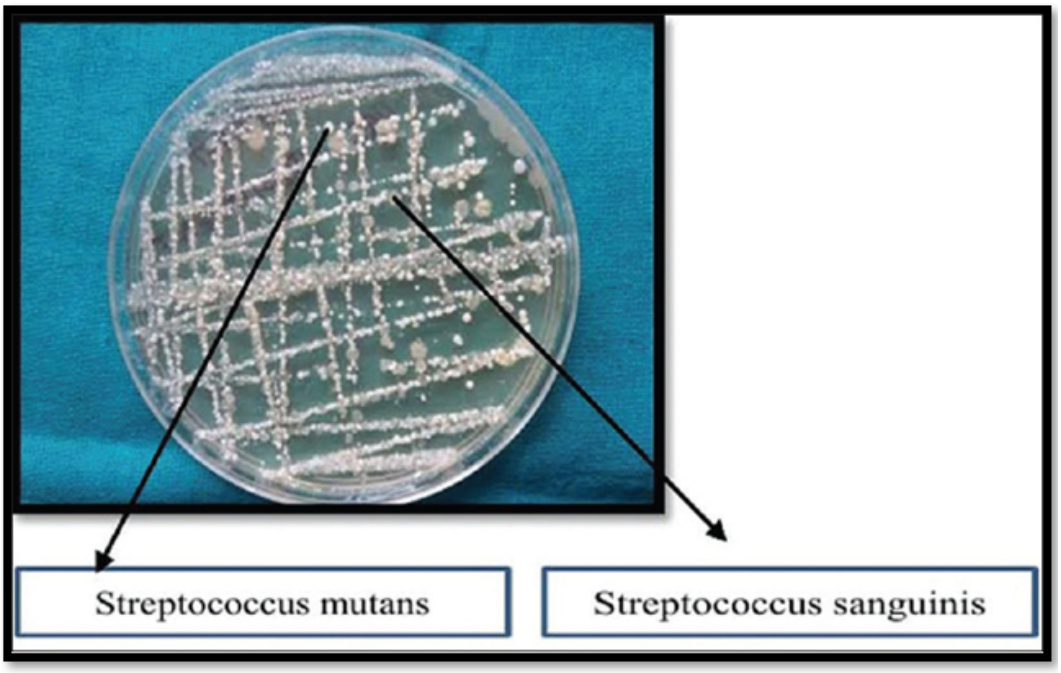

FIGURE 2.3: Shows *S. sanguinis* colonies were smooth or rough, hard and rubbery in uniformity, either grey, white, or colourless, and firmly attached strongly to the agar medium, making them difficult to eliminate with an inoculating loop. They can be distinguished from *S. mutans* colonies, which were irregular, rough, white, or yellow in colour, representing granular frosted glass, and easily removable from the agar medium.

is associated with healthy biofilms and the lack of dental cavities. *S. sanguinis* can compete for colonisation on tooth surfaces with *S. mutans*. There is a positive and inverse link between S. mutans and S. sanguinis and decaying teeth in the mouth. Because newborns and young children with ECC are more likely to acquire caries, preventative and treatment approaches are frequently required for them [22]. A research was conducted in 'The Jefferson County Health Department's Maternal and Infant Care programme in Birmingham, Alabama that *S. sanguinis* was found

in all 45 newborns sometime after their primary teeth emerged. By 8.0 months of age, 25% of the newborns had *S. sanguinis*, and 75percent of the total had *S. sanguinis* at the age of 11.4 months [23].

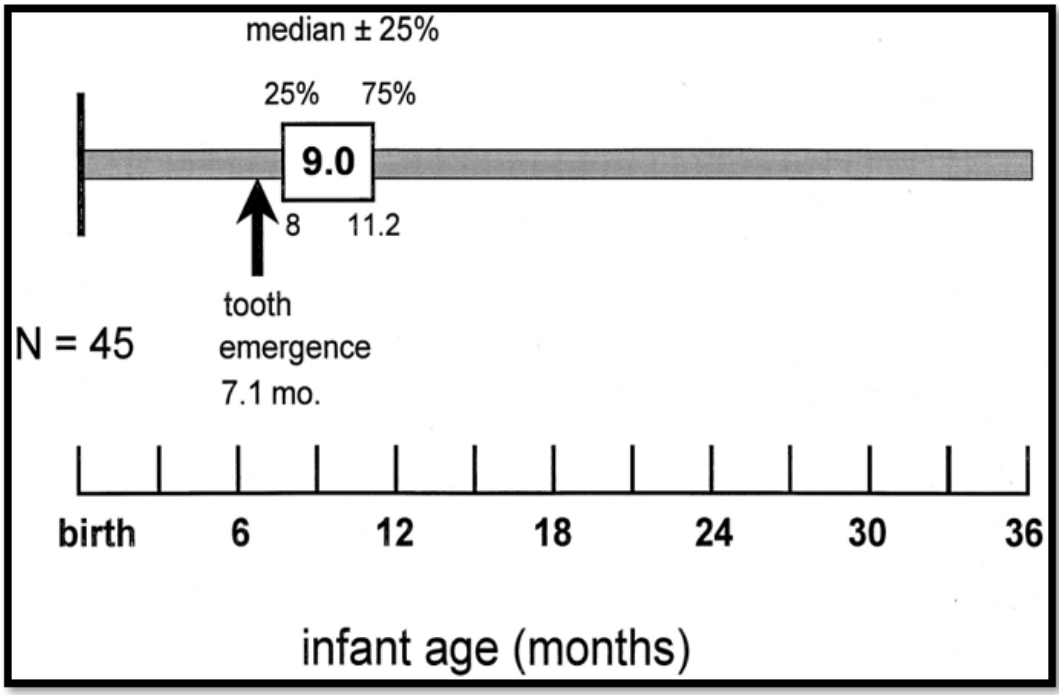

FIGURE 2.4: Shows the percentage of new born having the infection of *Streptococcus sanguinis*.

2.3.2 Biofilm Formation

The act of single cells adhering to a surface is the initial stage in biofilm development. In many bacteria species, fimbriae are involved in attachment to both animate and inanimate surfaces, as well as the production of biofilms. Fimbriae were shown to facilitate the adherence of *S. sanguinis* to saliva-coated hydroxyapatite (the primary material of the tooth surface) by *Fachon-Kalweit* et al. in 1985. Later, *Okahashi* et al. discovered three pilus (a class of fibrous protein) proteins in *S. sanguinis SK36*, named PilA, PilB, and PilC. A pilABC mutant was shown to be deficient in saliva-coated surface accumulation and biofilm formation. According to these researcher studies, PilB and PilC bind with human saliva in its whole. PilC is also attached to a variety of salivary components, one of which was discovered to be salivary [24].

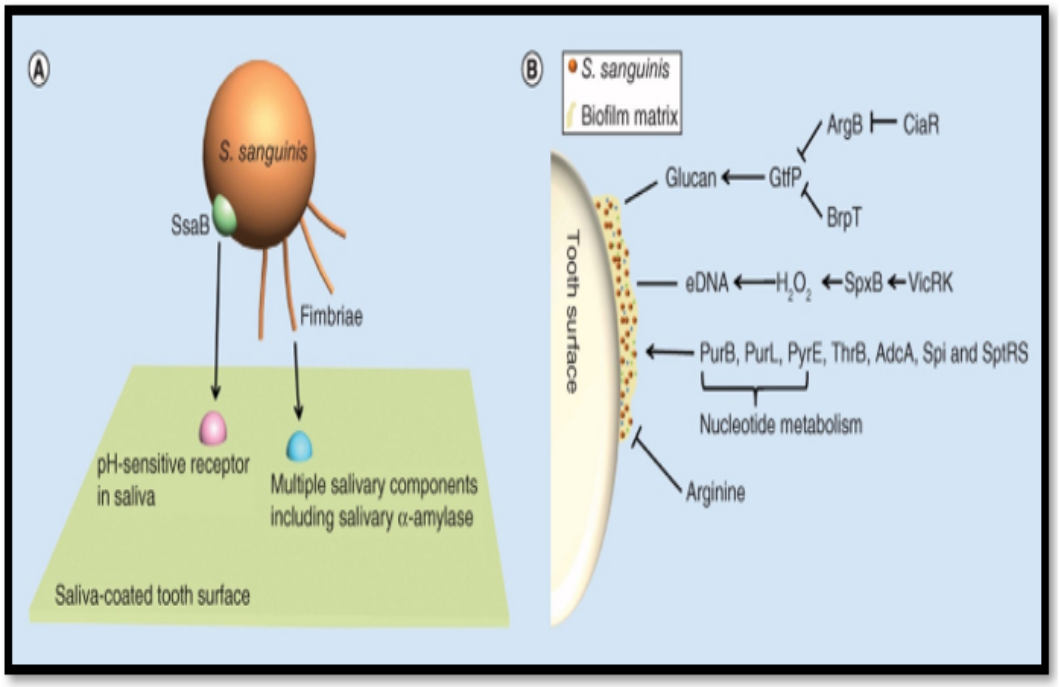

FIGURE 2.5: Biofilm formation factors in *Streptococcus sanguinis*. Figures A and B show the factors involve in biofilm formation

2.3.3 The Maturation of *S. sanguinis* Biofilm

Microorganisms make up around 10% of the dry mass in most biofilms, with the biofilm matrix accounting for the rest. Polysaccharides, proteins, nucleic acids, and lipids are the most essential components of the biofilm matrix, which mediate cell-cell and cell–surface adhesion to build cohesive, 3D polymeric networks [25]. Glucans are the predominant biofilm polymer generated in *S. sanguinis* in the presence of sucrose, and they are largely made up of 1,6-linked and 1,3-linked glucose. Glucans linked in the presence of sera, starch hydrolysates, and dextran had a different linkage than control glucans. The capacity of *S. sanguinis* to generate biofilms varies considerably depending on the growing media. Different medium components can change glucan levels and structure, impacting biofilm development, according to this research. Furthermore, even when cultivated in the same medium, the oxygen concentration impacts biofilm development, perhaps owing to alterations in metabolic pathways impacting glucan production [26].

The production of sticky glucans from sucrose is carried out by glucosyltransferases. *S. sanguinis SK36* has two gtf genes,gtfA, and gtfP, with gtfP being the

only one that has been demonstrated to synthesize glucan and enhance S. sanguinis adhesion to saliva-coated hydroxyapatite and biofilm formation. GtfP deletion inhibits the generation of both water-soluble and water-insoluble glucan and hence reduces biofilm formation. The CiaRH two-component system's response regulator CiaR and a transcriptional regulator BrpT have both been shown to affect the expression of the gtfP gene. The upregulation of gtfP transcription is reduced when ciaR is deleted and it is greatly conserved in *Streptocci* [27].

In several species, including *Pseudomonas aeruginosa* and *Staphylococcus aureus*, extracellular DNA (eDNA) is an important component for biofilm development. H_2O_2 generated by SpxB can trigger DNA release in *S. sanguinis*, which leads to cell aggregation. DNase I treatment reduces this behavior, indicating that eDNA enhances cell-cell interaction and may contribute to biofilm development [27]. A Vick knockout mutant reduces biofilm formation by inhibiting H_2O_2 generation and eDNA release. DNase I therapy significantly reduced biofilm development when biofilms were cultivated in brain heart infusion (BHI) medium with 1 percent sucrose for 2 or 4 hours in this research, but another investigation found that DNase I treatment had no impact when added after biofilms had formed [28]. In the figure

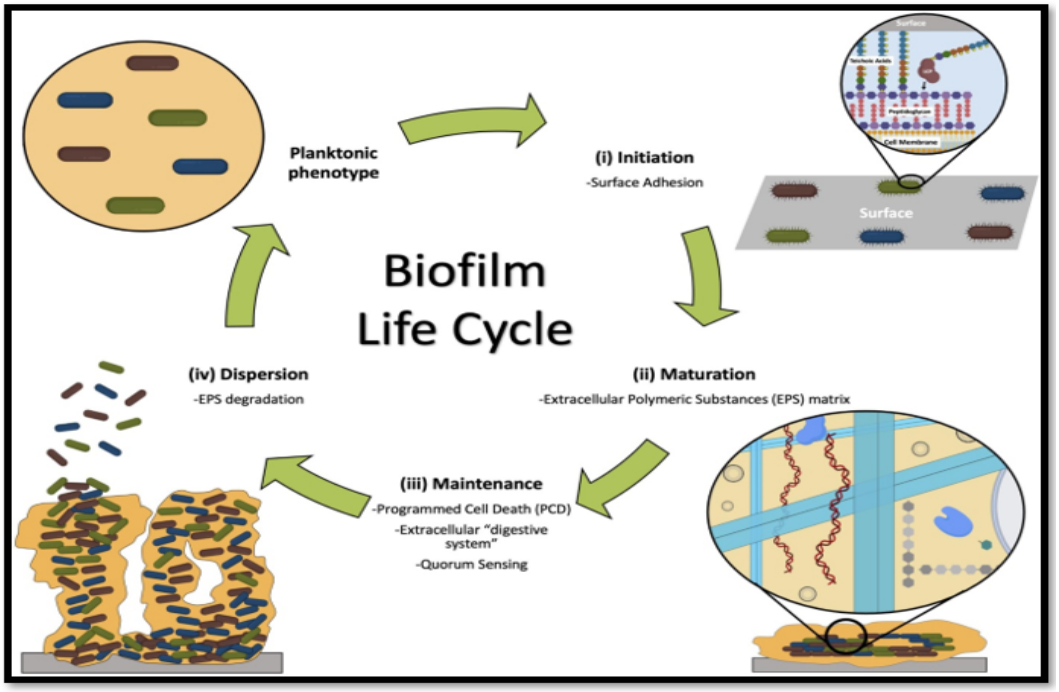

FIGURE 2.6: The fundamental mechanisms involved in biofilm growth are divided into three phases.

above 2.6 (i) When bacteria link to the surface, the process begins called initiation, (ii) maturation occur ; whenever microcolonies are formed, adhere on the surface, and become entrenched in the EPS matrix ; and (iii) When some bacteria in the population break out from the biofilm as well as reach the planktonic phenotype, this is referred to as dispersion [29].

2.4 Feature Contribute to Biofilm Formation

Cellular chain length is important feature that play a role of contribution in the biofilm production of *Streptococcus sanguinis* and developing endocarditis pathogenicity. The length of the chain, for example, might influence the cellular physical behaviour during the development of biofilm. The lengthier the cellular chains more easier it is for cells to unite with one another. Cellular chain length has been linked with invading pathogen in Streptococci species (like *Streptococcus pneumonia*) as well as phagocytosis susceptibility in Streptococcus mutans. Cell division geometry during binary fission determines the creation of cellular chains [30]. Cellular division occur between two axes that are perpendicular with one another leads to the production of triad of daughter cells, whereas division between three planes resulted in the formation of daughter cells. Cell division that is in irregular patterns that exist in various planes resulting in the form of clustered bundles or bunches of bacteria, in contrast to typical cell division patterns that occur accross single axis or plane results to the production of cellular chains. Cell division in streptococci typically proceeds in one plane, that leads to formation of chains of two or may be more cells. Because of streptococcal cellular chain elongation has been regulated by a variety of circumstances, this should be emphasised that streptococcal cell shape and chain development are very diverse and are known to change between strains [31]. In the above figure 2.7 For each COG function, the gene percentage was computed through dividing the number of shorter or longer chain mutants by the overall number of annotated genes present within the genome. Abbreviations for COG functions: Q stands for secondary metabolite production, catabolism and transport; P stands for metabolism and

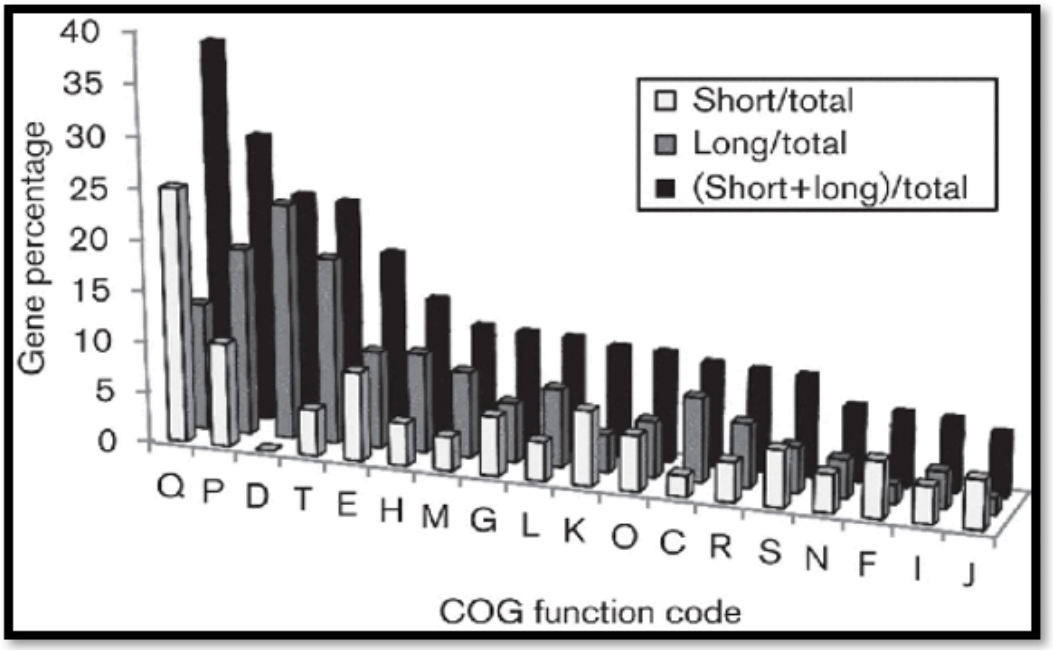

FIGURE 2.7: Shows following characteristics.Gene distribution by COG functions is propotional to the length of cellular chain .

the transportation of inorganic ion ; D stands for division of cell and chromosomal partitioning: T stands for the technique of signal transduction. E stands for the transportation of amino acid and metabolism; H stands for the metabolism of coenzyme ; M stands for cell envelope biosynthesis and, outer cell membrane; G stands for glucose metabolism and transport ; and L stands for the replication, repairing and recombination of DNA: letter K stands for the process of transcription; O stands for modification of post translation, chaperones as well as protein transformation; and C stands for the generation and conversion of energy; R stands to predict for general function only; S stands for uncertain process; N stands for the motility and excretion of cell; F stands for the metabolism and transportation of nucleotide; and letter I stands for metabolic process of lipid; J denoted for biogenesis, translation, and ribosomal architecture.

2.5 Genes Of Cellular Chain

Importantly, there are 12 genes linked with elongation of cellular chain and were implicated with COG annotations pathway for signal transduction. As a result,

we investigated the cellular morphology of mutants lacking TCS called two component system. This system often present in bacterial cell and are made up of a protein bounded by membrane termed as histidine kinase (HK) that detects alteration in the environment and autophosphorylates all its histidine residue. By transferring a phosphate toward an aspartate residue, the Histidine Kinase communicates with some of its matching cytosolic (RR) response regulator protein. The RR changes shape and causes a response via regulating gene transcription. We discovered that seven of the such 29 genes encoding 15 TCSs in the genome of S. sanguinis displayed statistical variances in morphological cellular chain length, included three HK – These are SSA 0205, SSSA 1972, SSA 1119, SSA1120, SSA 0517, SSA 0217 and SSA 1113, are among the RR. Longer chains were generated by all seven TCS gene mutations [32].

We have established a complete bank of gene deletion mutants. By using phase-contrast microscopy, they examined the alteration in morphology of number of gene omission mutants and discovered variety of genes associated with cellular chain expension [33].

2.5.1 Pilli Contribuite to Biofilm Formation

In this species, scientists discovered unusual structure called pili on cell surface which is long, thin and hair like extensions . In this study, they looked that pili involve in the development of biofilm in *S. sanguinis* [34]. In *S. sanguinis*, PilB is one of three minor pilin T4P, has a conventional N-terminal class III signal peptide, which is a distinguishing trait of type IV pilins. This sequence motif is made up with seven-residue leader peptide that is mostly made up of neutral and hydrophilic amino acids and concludes with the Gly [35].

2.5.2 Functions of Type IV Pili

DNA absorption, secretion, surface motility , eukaryotic cell adhesion, microcolony formation, and even electrical conductance and, in archaea, flagellar motility are

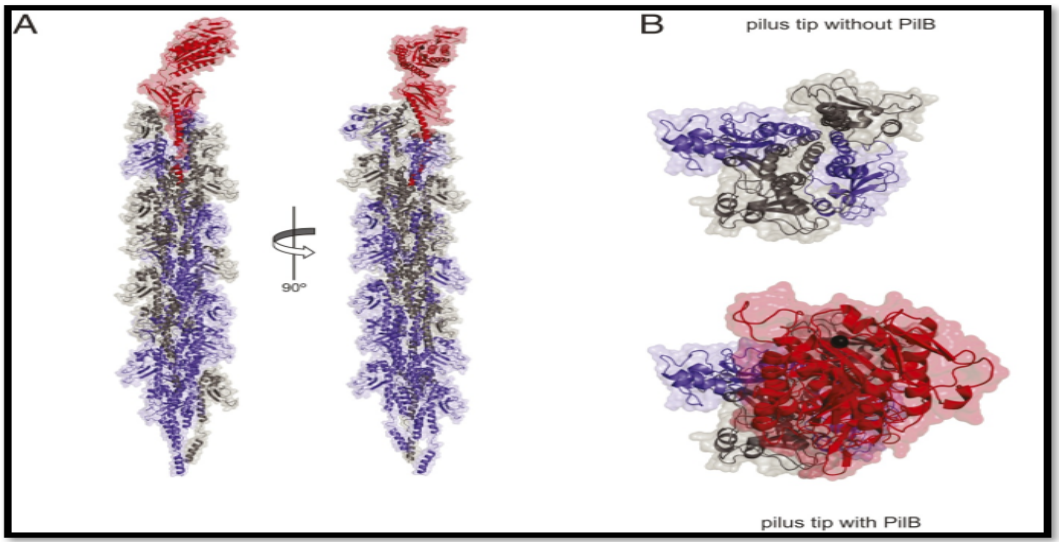

FIGURE 2.8: PilB 3D model of S. sanguinis T4P. (A) PilB shows (red) is packed into *S. sanguinis* T4P, right-handed helical heteropolymer comprising both main pilins, PilE1 (blue) and PilE2 (gray). (B) Whether or not the T4P tip is capped by PilB.

all activities of type IV pili. As a result, we anticipate that the roles of type IV pili in Gram-positive bacteria will be equally diverse. We may now classify known functions into two categories: adhesion and motility [36].

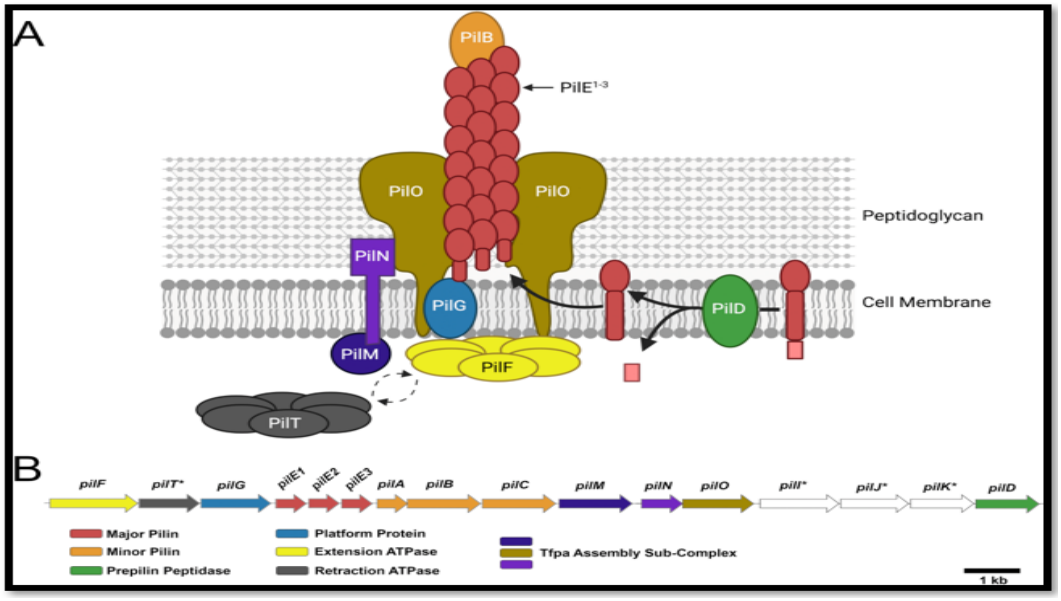

FIGURE 2.9: *S. sanguinis SK36* Type IV pili.

In the figure 2.9 (A) T4P model filament assembly of S. sanguinis based on structural modelling by Raynaud et al., with PilB confined at the filament's end. PilACIJK are really not displayed because their function or localisation is uncertain.

The image was generated using BioRender. (B) T4P biogenesis genes in monoderm bacteria, modified from Pelicic. Functions that are known or expected are noted underneath the operon. Genes of the same shade are projected will encode proteins with comparable functions, whereas white genes in other species have no known hmologs. An asterisk denotes genes that are not required for filament assembly but are required for functional motility.

Tfp have been extensively investigated because they are important virulence factors. These investigations have identified two types of Tfp. (i) Tfpa are made up of pilins with shortened leader peptides that share significant sequence conservation in 1N, (ii) Tfp are synthesised by a machinery composed of 15 conserved proteins (encoded by genes distributed throughout the genome), and (iii) promote a broader range of functions, including twitching motility and DNA uptake in naturally competent species. These extra features are a direct result of Tfpa's capacity to retract, which is driven by the PilT ATPase and generates enormous forces in the piconewton (pN) range [37]. The vast majority of these filaments were around 12 nm thick and wavy, although some exhibited a little more conventional Tfp shape, i.e. they were straight, thinner (approximately 6 nm wide), and occasionally formed tiny bundles [38].

2.5.3 Pathogenicity

Streptococcus sanguinis together with other oral streptococci, are thought to be responsible for 20% of IE cases globally. Such organisms are very widespread and frequently difficult to identify and treated in underdeveloped countries whereby rheumatic cardiovascular disease is a significant risk factor. Oral streptococci's potential to induce IE is linked to their ability to join and stimulate platelets, attach with hosts tissues and extracellular components, finally produce exopolysaccharides (EPS). During certain pathogenic factors are inherited across species, its also obvious that various species have distinct features which bestow varied pathogenic capabilities [39]. Type IV pili and dynamic pilar movement appear to be major contributions to certain species' pathogenicity [40].

2.6 Energy and Metabolism

S. sanguinis appears to be capable of adapting to a variety of carbohydrate sources .Over than 50 potential carbohydrate carriers were discovered, which includes phospho-transferase system enzymes specific for, mannose, glucose, fructose, cellobiose, lactose, glucoside, trehalose, mannose, galactitol, and maltose transport. All S. sanguinis proteins in genome were functionally classified and compared [41].

Although the citric acid cycle is incomplete, *Streptococcus sanguinis* possesses a well-developed enzyme system for energy generation. It enables the bacterium to ferment sugars and carbohydrates, as well as to produce gluconeogenesis from amino acids. The process also yields fructose-6-phosphate, an essential building ingredient for peptidoglycan production in the cell wall [42].

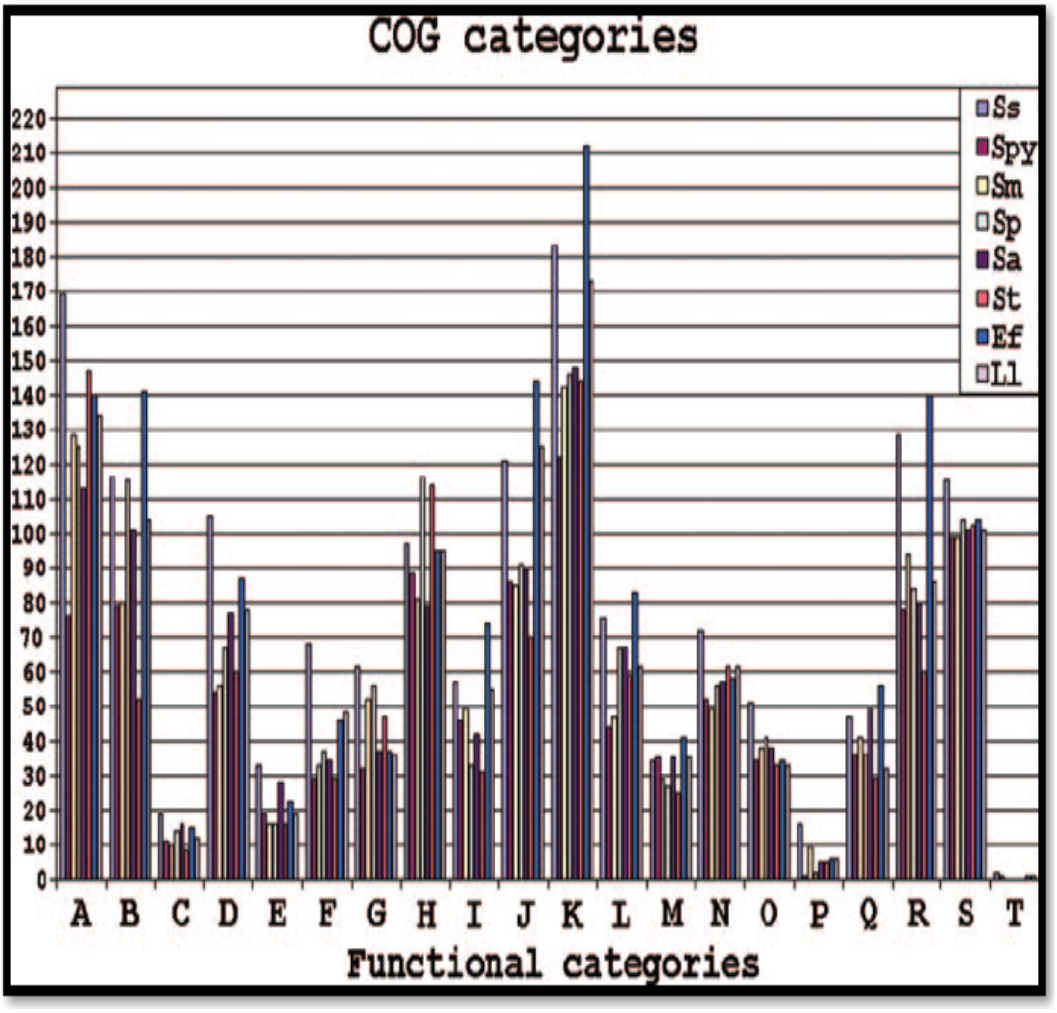

FIGURE 2.10: Shows the COG functional category.

2.7 Discovery Of Infective Endocarditis

William Osler in 1885 described endocarditis. Medical science advancements and microbiological studies have all provided a clear picture of the condition. The most prevalent risk factors for infective endocarditis include prior heart damage, current heart surgery, or poor dental hygiene [43]. The term infective endocarditis, coined by Thayer and later popularised by Lerner and Weinstein, is preferred to the previous term bacterial endocarditis since the condition may be caused by rickettsiae, chlamydiae, mycoplasmas, fungi, and maybe viruses [44]. Endocarditis begins when bacteria enter the circulation and travel to the heart, where they adhere to malfunctioning heart valves or damaged heart muscle. The majority of cases are caused by bacteria, although fungus or other microbes may also be to blame. Normally, dangerous germs that enter your bloodstream are destroyed by your immune system. Even if germs enter your heart, they may travel through without infecting you. Bacteria present in your mouth, oesophagus or as well as other regions of human body, such as your skin or gut, can, given the appropriate conditions, produce serious infections like endocarditis [45].

2.7.1 Types of Infective Endocarditis

Infective endocarditis defined as an infection of the heart walls that also affects the valves. It frequently affects the cardiac muscles. Infective endocarditis, or IE, is classified into two types. Acute IE (or acute bacterial endocarditis):- This type emerges suddenly and can be fatal within days.[46] (ABE) generally appears suddenly and advances quickly (within days). The source of infection or a entry point is frequently visible. ABE can harm normal valves in the presence of pathogenic bacteria and prolong bacterial transmission. S. aureus, group A hemolytic streptococci, gonococci and pneumococci are the most common pathogens [47].

Subacute or chronic IE (or subacute bacterial endocarditis) This type develops gradually over weeks to months [48]. (SBE), while aggressive, generally develops slowly and insidiously (within few weeks to months). Frequently there is

not obvious source of the infection nor entrance point. Streptococci (particularly viridans group anaerobic, microaerophilic, enterococci and nonenterococcal group D streptococci) are the most common cause of SBE, followed by *Streptococcus aureus, Gemella morbillorum, Staphylococcus epidermidis, Abiotrophia defective* (once called as *Streptococcus defectivus*), *Granulicatella species* and *Fastidious*, *Haemophilus species*, After asymptomatic bacteremia caused by periodontal, genitourinary tract and gastrointestinal infections, SBE frequently develops on defective valves [49].

TABLE 2.2: Shows the In acute and subacute infective endocarditis, culture consistency results from cardiac valves and dentigenous infected foci [50].

Endocarditis	**Acute**		**Subacute**		**Total**	
	(N)	**%**	**(N)**	**%**	**(N)**	**%**
Corresponding	8	100.0	22	43.2	30	50.5
Non- Corresponding	0	-	24	46.2	24	40.0
Negative	0	-	6	11.5	6	10.0
Total	8	100.0	52	100.0	60	100.0

2.7.2 Signs and Symptoms

The symptoms differ from individual to person. Some people get symptoms quickly, while others start showing symptoms gradually. People who are at greater risk of endocarditis should exercise extra caution.

1. Sweats or chills, particularly at night.
2. Rashes on the skin.
3. Tenderness, pain, redness, or swelling.
4. A wound or cut that will not heal.
5. Sore that is red, heated, or draining.
6. Sore throat, itchy throat, or swallowing pain.

7. Nasal congestion, sinus drainage, headaches, or pain along the upper cheekbones.

8. Continuously dry or wet cough that lasts more than two days [51].

2.7.3 Pathogenicity of IE

The typical heart valve endothelium is resistant to bacterial invasion when challenged intravascularly. Thus, the consequence of multiple independent factors is required for the prevalance of infective endocarditis; modification of the heart valve surface provide an appropriate place for bacterial adhesion and colonisation; pathogenic bacteria with an organism responsible for adhering with it and colonizing the tissue of heart valves; finally formation of the infected tissue or mass which we called it 'vegetation' via 'burying' the propagating pathogen inside a preventative matrix of blood components like platelets and fibrin [52]. Pathogens

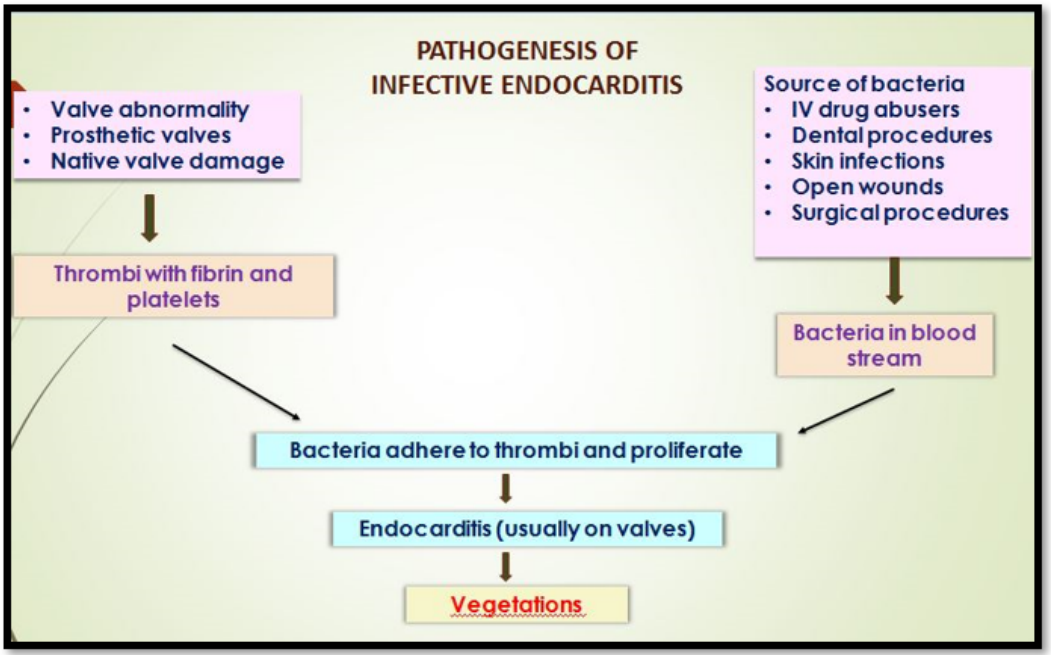

FIGURE 2.11: Shows how pathogenic bacteria develop infective endocarditis.

enter the bloodstream through several methods, including intravenous catheters, injection medication use, and dental sources. Pathogens attach to the surface of a faulty heart valve. The pathogen acquires intrinsic access to the valve endothelial cells. Infected vegetation develops by encasing the developing organism in a

serum-molecule-protective matrix.Vegetation particles have the ability to detach and scatter, resulting in emboli. These can cause mycotic aneurysms,ischemic stroke, and infarcts or abscesses in distant places.

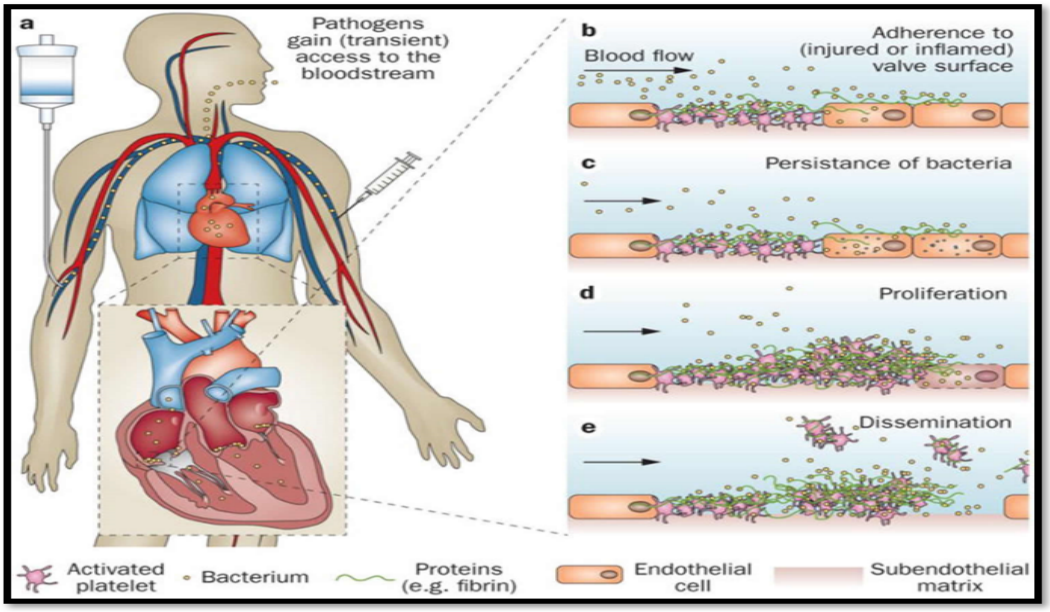

FIGURE 2.12: Shows that (a) Pathogens enter the circulation through a variety of routes, including intravenous catheters, injectable drug usage, and dental sources.(b) Pathogens cling to a defective cardiac valve surface.(c) Pathogen gains intracellular access to the valve endothelium. (d) Infected vegetation is formed by burying the growing organism within a serum-molecule-protective matrix. (e) Vegetation particles can separate and disperse, resulting in emboli. These can result in problems including ischemic stroke, mycotic aneurysms, and infarcts or abscesses in distant locations.

2.7.4 Epidemology

Infective endocarditis is more frequent in males than in women, and it gets worse with age. The average age of IE victim over the time has risen, from around 30 years before antibiotics in 1990s, around 60 years old. IE is much frequently linked to intracardiac prosthetic implants and microorganisms from the gastrointestinal (GIT) system in the elderly. In a major observational cohort research, IE more commonly affected the mitral valve alone (about 40% of people), next the aortic valve alone (36% of individual), and finally causing multivalvular illness . Except among injectable drug users, right-sided valves are hardly damaged. When it

comes to IE, the pulmonic valve is the least likely to be affected. IE is more frequent in males than in women, and it gets worse with age.

2.7.5 Aortic Valve and Mitral Valve

Because it causes turbulent blood flow, structural cardiovascular disease is a major factor in IE. About 75% of people who get IE have structural heart problems. In the past, the most common valvular abnormality in individuals with IE was rheumatic heart disease with mitral stenosis. Aortic valve disease (stenosis and regurgitation), mitral regurgitation, and congenital heart disease have recently been identified as the most common predisposing lesions. Mitral valve prolapse, especially with regurgitation, is a risk factor for IE [53].

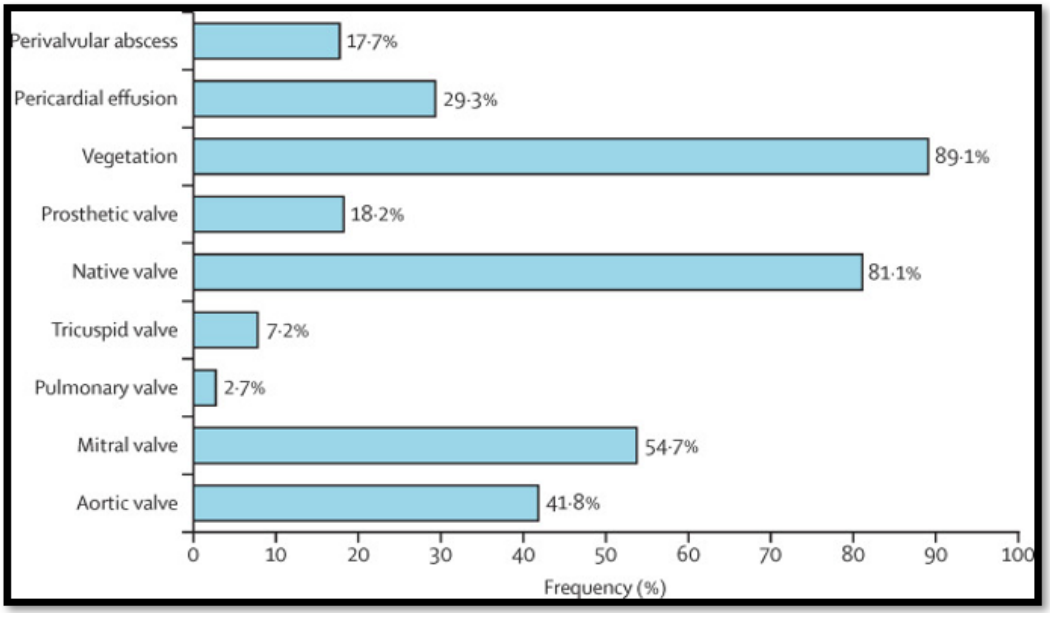

FIGURE 2.13: Shows the distribution of infective endocarditis. Highest Values showed by vegetation, native and mitral valve which are 89.%,81.1% and 54.7% respectively [53].

2.7.6 Localization of Vegetations in Patients

Infective endocarditis (IE) is an infection of the endothelial surface of heart. Despite modern advances, it remains a life-threatening disorder. Vegetation was most

common on the aortic valve (43%) and the mitral valve (36%).

TABLE 2.3: summarizes the percentage of vegetation at different localization.

Localization	**n**	**%**
Aortic valve	43	36.1
Mitral valve	36	30.3
Aorta+mitral valve	23	19.3
Tricuspid valve	7	5.9
Pulmonary valve	1	1.7
Endarteritis	2	1.7
Multivalvar(>2valve)	3	25
Pacemaker lead	2	1.7
Heart valve+endarteritis	2	1.7

2.7.7 Area of Infection at Aortic Valve

Streptococcus specie was found to be the most prevalent pathogen in research. In general, the aortic valve is more prone to develop abscess. Because of these abscesses spread, they affect the conductive tissues and cause various forms of AV block. Infective endocarditis causes a fresh onset AV block. *Streptococcus sanguinis* was also revealed as the causal organism, which is an uncommon pathogen to induce SAFJ. The observations, however, raised the possibility of the serious eccentric regurgitation of aortic valve in the presence of a bicuspid aortic valve. These growing conducting anomalies decreased the threshold for the further screening, finally assisting in the detection of perivalvular infection on cardiology CT [56].

2.7.8 Factors Contributing to the Valve's Vulnerability to Infection

The valve's vulnerability to infection is caused by tissue damage and inflammation. Mechanical injury or inflammatory processes of the cardiac endothelium, along

with an increased coagulation state and an unbalanced innate immune system, results in platelet-fibrin aggregation in a deposition-friendly environment. The accumulation of these aggregates renders the valve more susceptible to infections.

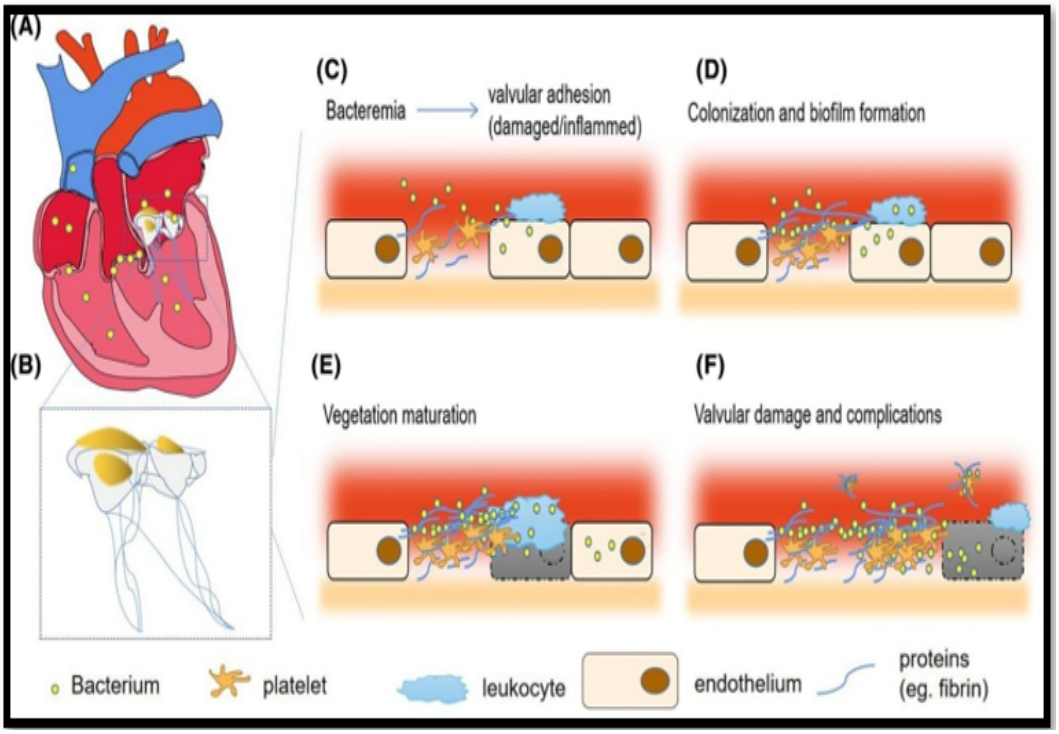

FIGURE 2.14: (A and B) Infectious endocarditis causes mitral valve vegetations. (C) Invasive infection is caused by bacteria in the bloodstream adhering to damaged or inflamed valves. (D) Bacterial endurance and growth result in valve colonisation and the production of biofilms . (E) The valvular lesion advances, followed by vegetation maturation. (F) Severe valvular injury and distributed vegetation particles cause clinical symptoms as well as a variety of complications [58] .

2.7.9 Pathophysiology

Bacteraemia generally causes the heart to get infected (delivers the organism to the valve). Non infective thrombi allow for adhesion and penetration of the organism. These can occur antecedent to bacteraemia and are caused by high pressure strains causing injury to the heart valves, resulting in fibrin and platelet aggregation. This is why high-pressure areas (e.g., the left side of the heart; structural/valve abnormalities) are more typically impacted than low-pressure locations (e.g. right

side of the heart or an isolated atrio-septal defect). These are far more susceptible to infection. Once it colonised, this is referred to as a vegetation [57].

2.8 Valve Colonization

Bacterial adherence to defective or inflamed valves is the very first step in valves colonisation and the subsequent creation and maturation of vegetation [58]. Most endocarditis infections must overcome the severe shear pressures present in the heart chambers by attaching and triggering platelets and utilising them as bridges. Proliferation and biofilm development are good survival strategies for harmful bacteria following relatively poor initial adhesion [59].

2.9 Complications

There are following some complications develop during infective endocardidtis.

2.9.1 Congestive Heart Failure

Congestive heart failure is the most prevalent complication of both NVE and PVE (prosthetic valve endocarditis) , resulting in more than half of all IE cases. It is most typically linked with aortic valve disease and is usually caused by valvular malfunction rather than cardiac failure [60].

2.9.2 Intracardiac Abscess

Intracardiac sores or abscesses develop in 10 to 40% of IE cases and are especially frequent in PVE and aortic valve dysfunction. Aortic valve abscess can spread into the conduction system, resulting heart block . TEE is typically used for diagnosis, however it offers a sensitivity of around 50% when compared to intraoperative assessment. Surgical treatment is frequently required to cure an abscess [61].

2.9.3 Embolism

About 20–50% of IE patients suffers from Embolic events. The central nervous system is the most prevalent location of embolism, then by the other important organism like spleen, lungs, kidneys, and liver. The risk of embolism is greatest soon after diagnosis, and it diminishes after antibiotic medication is started. Vegetations larger than 10 millimetres in diameter, as well as vegetation cover the anterior mitral valve leaflet, are more prone to embolize [62].

2.9.3.1 Neurological Complication

The presence of atrial fibrillation significantly increased the risk of neurological complications, which can be explained by the reason that atrial fibrillation is a potential risk for cerebral embolism in and of itself, so the progression of atrial fibrillation with IE raises the probability of cerebral embolism even more.

The presence of more pathogenic organisms in acute IE explains why neurological disorders occurred nearly equally in both infective endocarditis with delayed therapy and acute IE with early aggressive therapy. Other neurological problems reported in studies include seizures and headache [63].

2.10 Prevention

Regular dental examination and therapy are indicated prior to having surgery or replace heart's valves or treat congenital heart issues. Measures in reducing the wellbeing bacteremia are also advocated in order to reduce the increased prevalence of iatrogenic bacteremia and consequent endocarditis.

Endocarditis prevention after prosthetic device implantation is evolving to accommodate the rise in endocarditis caused by enterococci. Instead of a cephalosporin, amoxicillin/clavulanic acid is frequently used for prophylaxis. Oral (dental) and epidermal (cutaneous) cleanliness are recommended for the entire population [64].

2.11 Diagnosis

Following are some tests that can help doctors to determine whether or not the individual has infective endocarditis: Blood tests to look for germs; an infected person may need many blood samples taken at different periods. An urination test, which examines the patient's urine for blood and infection associated with endocarditis. An echocardiogram test which records the locations and movement of the cardiac walls or internal organs such as the valves [65].

2.12 Traditional Methods and Drugs

Streptococcus sanguis is responsible for 50% of all viridans *streptococci*. Because of the rise in penicillin and cephalosporin resistance, treatment has grown more difficult. The researchers wanted to see how effective various antibiotics were as monotherapy and in combination with gentamicin in a rabbit model of infective endocarditis. The results were compared to those of a control group and a group that received traditional penicillin-gentamicin therapy [66]. Treatment for bacteria developing in biofilms is difficult due to enhanced antibiotic resistance and protection from the host's immune system. The discovery of drugs that interact with biofilm formation may boost the efficacy of antimicrobials and enable host defences access, which may then resolve the illness. Innovative techniques to biofilm prevention target particular treatment targets [67].

Some plant extracts are also used for herbal medicine known to be anti-inflammatory, anti-fungal, and anti-allergenic and also have some biological properties.

For example Ginger (Zingiber officinale), Neem (Azadirachta indica),Tulsi (Ocimum sanctum), Aloevera (Aloe barbadensis Miller), Garlic (Allium sativum), Banana plant (Musa cavendishii), Orange fruit (Citrus sinensis), Sun flower (Arnica Montana), Cloves (Syzygium aromaticum) etc. These plants are used for relieving toothache and works as traditional healer and tooth whitening. Compared to synthesized compounds, plant extracts have the disadvantage of being time-consuming

and requiring specialized equipment for the separation and characterization of active chemicals. Other difficulties in the separation of active ingredients include the instability of the starting material, uncertainty in the isolation of the active ingredients, and the cost of extraction [68].

2.13 Pangenome Analysis

The accumulation of biological information and developments in computing technologies have permitted the application of pan-genome analysis, which attempts to improve understanding of genotype-phenotype relationships in a given set of species. Pan-genome approach has been found to be an excellent method for deeper understanding of pathogenic bacterial clad since this aids in the development of diverse and personalised therapy methods based on biological differences and commonalities [69]. The pan-genome is split into three sections: the core pangenome, which includes genes present in all individuals, the shell pangenome, which includes genes present in two or many strains, and the "cloud pangenome," which only includes genes identified in one strain.

Pangenome concept was created by *Tettelin et al* [70]. The pangenome is made up of all the genes identified in a specific microbial species' sequenced genomes, and it can vary if new genomes are processed and added into the study. A genomic lineage's pangenome compensates for intralineage gene content variations. The pangenome emerges as a result of gene duplication, gained and loss dynamics, and the genome's interaction with mobile elements influenced by selection and drift [71].

A pangenome research also allowed us to identify a strain's resistome, non-virulence genes, and mobilome (to identify selfish genes). By examining SNPs in the core genome, it is often feasible to calculate the age of clones. Furthermore, similar to the pangenome notion, the panmetabolism may be presented, providing a broad yet thorough overview of all shared metabolisms and/or variances between strains of interest [72]. This study designed to find the therapeutic agents

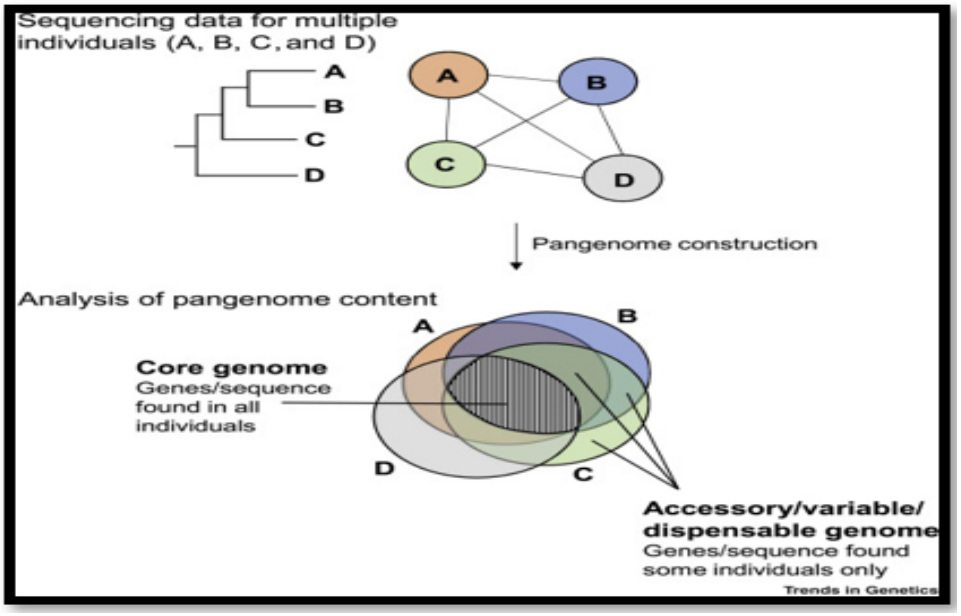

FIGURE 2.15: Shows analysis of pangenome.

against *S.sanguinis* through computer aided softwares and insilico approach. The development of therapeutic approaches that efficiently reduce bacterial adhesion, colonisation, and biofilm formation can lowering the chances of infectious disease known as infective endocarditis.

2.14 Scope

Computational analyses aided in determining the sort of protein that should be regarded an efficient therapeutic target. Our research will aid researchers create and identify highly effective and specialized anti-*S. sanguinis* treatment drugs. Advanced drug design is based on receptor-ligand combinations, in which the drug interacts inside the binding pocket of the receptor and changes its activity.

The potential and limits of current medication targeting initiatives are reviewed in terms of the condition that targeting techniques must eventually be judged by their efficacy in achieving selectivity of pharmacological activity, therefore minimising or eliminating non-mechanism-related adverse effects. Local drug distribution, differential metabolism, carriers and vehicles, complex molecular recognition, site-specific activation, and molecular specificity are all examined by computer aided

drug designing technique. Each strategy is briefly evaluated for its viability as a viable choice for human health treatment against life threatening diseases. The computational method for drug development has aided in the rapid translation of biological data into useful knowledge therapy treatments against illnesses. The drug targeting approach has the ability to significantly boost drug delivery effectiveness, decrease adverse effects, and lessen treatment costs.

Chapter 3

Research Methodology

3.1 Drug Targeting Techniques

A drug's therapeutic response is determined by the interaction of the drug molecule with the cell membrane and the biological processes that occur at receptor sites in a concentration-dependent way. We will investigate several approaches for drug targeting. The methodological steps are as follows: A drug's therapeutic response is determined by the interaction of the drug molecule with the cell membrane and the biological processes that occur at receptor sites in a concentration-dependent way. For drug targeting, we will find out different methodologies. The fundamental goal of CADD is to screen, optimise, and assess the compound's activity against the target. It represents a multidisciplinary strategy for both academic institutions and big pharmaceutical firms seeking improved efficacy with no negative effects.

In this project we used insilico approach to determine the pathogenicity and drug designing against bacteria. Computational strategies are utilized in pc-aided drug design to find,create and take a look at medicines and other physiologically active compounds. The ligand-primarily based (CADD) method examines ligands identified to have interaction with target of interest.

The fundamental goal of CADD is to screen, optimise, and assess the compound's activity against the target.

FIGURE 3.1: Possible main features of drug targeting.

It represents a multidisciplinary strategy for both academic institutions and big pharmaceutical firms seeking improved efficacy with no negative effects. For drug targeting, we will find out different methodological steps.

3.2 Genome Selection

The whole genome of *Streptococcus sanguinis* was searched and retrieved by National Centre of Biotechnology Information commonly called (NCBI), from Genebank. NCBI establishes in November 4, 1988. It is a part of National Library Of Medicine at National Institute of Health. The web page is (https://www.ncbi.nlm.nih.gov) [73].

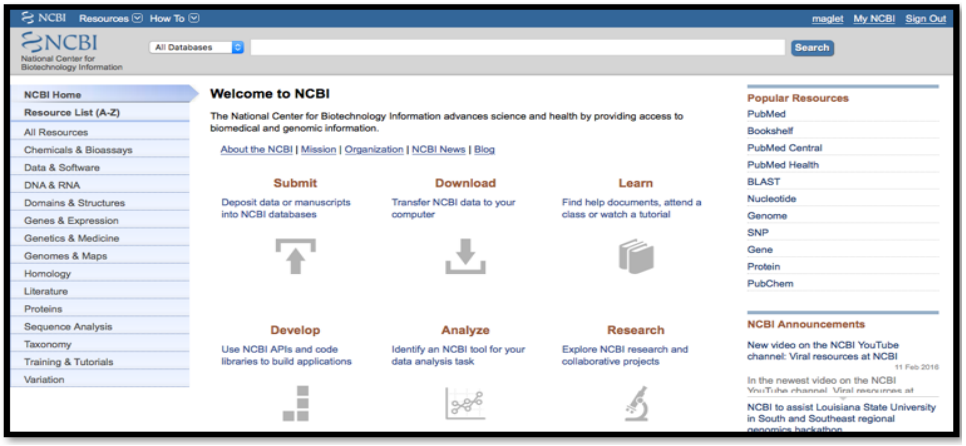

FIGURE 3.2: Home page of NCBI.

It is a public database its aim is to develop various software tools and to collect general information regarding gene.

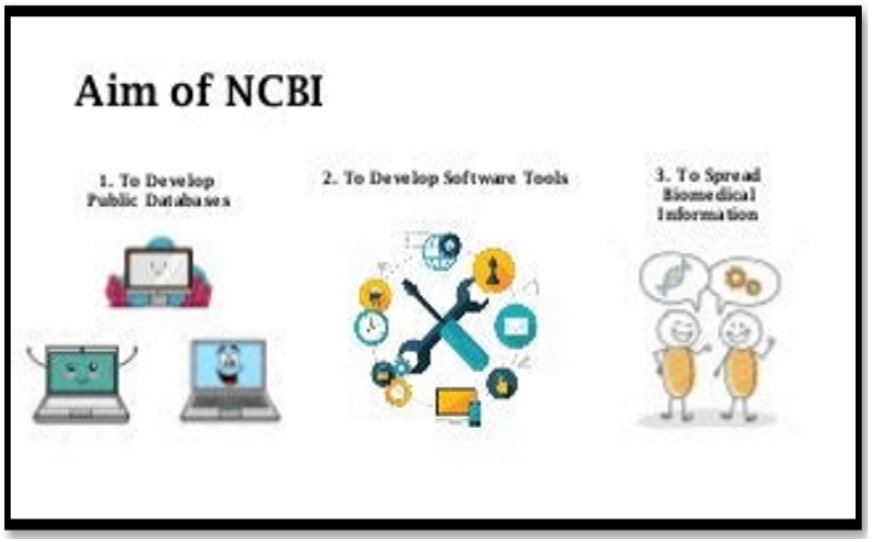

FIGURE 3.3: Indicates aim of NCBI

3.3 Identification of Core Genome

The core genome is just a collection of homologous genes present in all genomes of an investigated data. In 1984 EDGAR (Electronic Data Gathering Analysis and Retrieval) tool was created, it is electronically accessible database providing a more efficient and less costly method whereby the researcher investigate homologous gene by using the tool named EDGAR 3.0 through this web address (https://edgar.computational.bio.uni-giessen.de/cgi-bin/edgar.cgi) [74].

EDGAR provides for the rapid detection of differential gene content, such as the pangenome, core genome, or singleton genes. This tool also includes variety of assessment and visualisations such as Venn diagrams, phylogenetic trees, synteny plots, and Amino Acid Identity (AAI) and Average Nucleotide Identity (ANI) matrices. In previous several years, the average amount of genomes analysed in the EDGAR research has increased by approximately two orders of magnitude [75].

EDGAR introduces new analytic capabilities and greatly facilitates the comparison of related genomes. The programme allows for a fast assessment of phylogenetic

links and accelerates the method of gaining innovative biological insights about the differences in genetic makeup between related genomes [76].

FIGURE 3.4: Home page of EDGAR

3.4 Selection of Strain

Streptococcus sanguinis was used to extract core genes from all the 8 completed strains, while 25 strains are scaffold 68 are contigs out of total 101 strains. A reference strain (Sk36) determined with its official release. Number of genes discovered throughout all strains being chosen for further investigation. *Streptococcus sanguinis (Sk36)* was chosen as the reference strain in Edgar 3.0, and it was compared to the other 7 strains.

3.5 Quality Assessment

Patric Databases (https://www.patricbrc.org/) and Check M were used to examine the quality of Streptococcus sanguinis strains. All strains were human,

with a full genome, and they were of high quality.

3.6 Sequence Similarity

David J. Lipman and William R Pearson created the FASTA tool, which is useed for searching sequence similarity against protein databases. It works best for nucleotide searches.

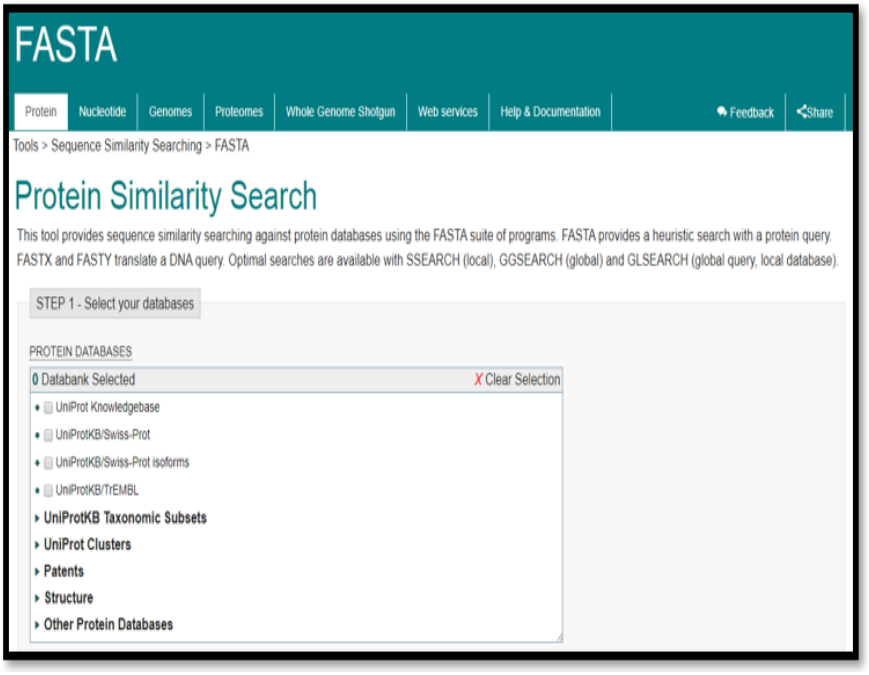

FIGURE 3.5: Home page of FASTA

FASTA is a format based on text for encoding amino acid sequences, with single-letter code for representing nucleotides or amino acids. FASTA searches a related sequence database for matches of comparable database sequences utilising local matching on a specified nucleotide or amino acid sequence.

3.7 Non-Homologous Protein Identification

BLAST is (basic local alignment search tool), which compares pairs of sequences to look for areas of local BLAST searches for areas of local resemblance between

nucleotide or protein sequences [77].

BLAST searches for areas of local resemblance between nucleotide or protein sequences. The software compares nucleotide or protein sequences to database sequences and computes the statistical validity of the matches.

3.8 UniProt

UniProt databases are made up of three types of databases. UniProt Archive or (UniParc): By keeping the whole collection of publicly accessible protein sequence data, provides a reliable, comprehensive, and non-redundant sequence collection.

FIGURE 3.6: Home page of uniprot

The UniProt Knowledgebase or (UniProt): is a central database of protein sequences that contains accurate, consistent, and extensive sequence and functional annotation. The UniProt NREF databases or (UniRef): offer datasets from the UniProt knowledgebase that are not duplicated, allowing for comprehensive or full coverage of sequence space at many resolutions. This tool is high quality, easily accessable and comprehensive. The URL (`www.uniprot.org/downloads`) [78].

3.9 Target Identification

Multiple criteria, such as molecular weight, route analysis, pathogenicity, and so on, were employed to determine possible therapies [79]. A good target should be effective, safe, satisfy clinical and commercial criteria, and be "druggable. "ProtParam (https://web.expasy.org/protparam/) was used to identify the molecular weight of the important genes and also calculate molecular weight, pI, amino acid composition, aliphatic Index and GRAVY, Charge and hydrophobicity, extinction co-efficient and protein concentration from inputed A280 [80].

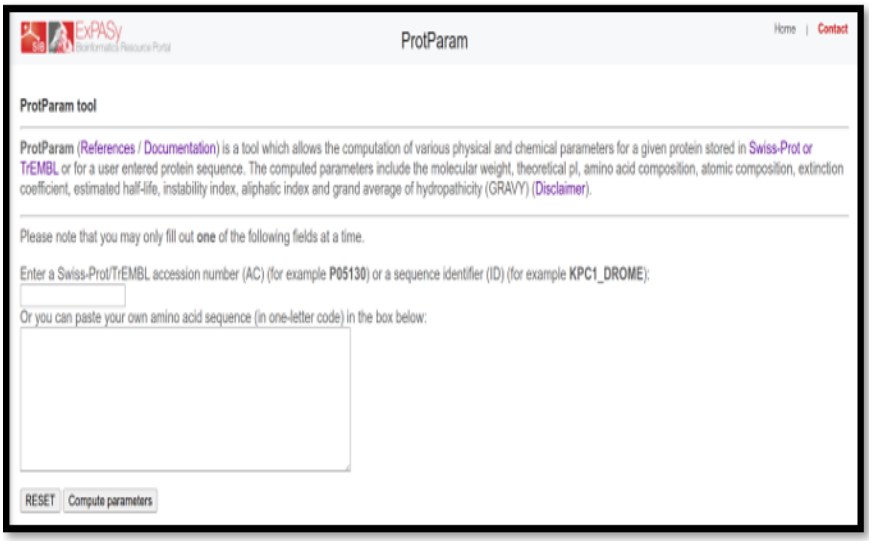

FIGURE 3.7: Home page of Protparam

3.10 Localization of Protein

CELLO:Subcellular localization predictive system (http://cello.life.nctu.edu.tw/) is publicly available web based tool, used to determine cytoplasmic and membrane protein [81].

CELLO attempts to predict at least single subcellular location based on the species of the protein. The aggregate of cellular compartment, molecular function, and biological process is shown as a pie chart illustrating various functional annotations for the searched protein.

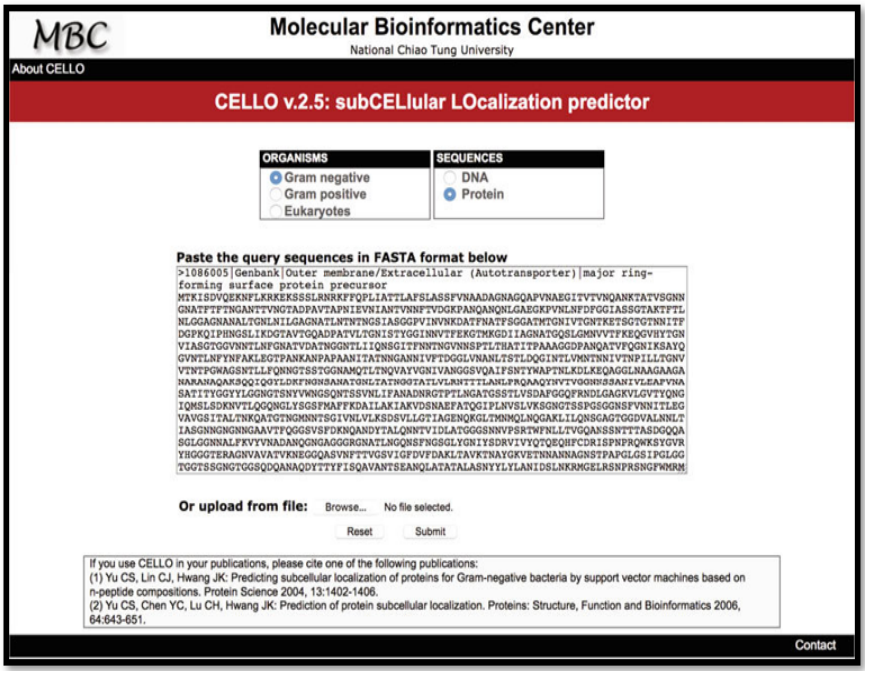

FIGURE 3.8: Home page of CELLO

3.11 Catalytic Pocket Detection

Many drug development attempts fail because the underlying target is eventually shown to be unsolvable. Protein structure elucidation progress has now paved the way for automated structure-based target evaluation. DoGSiteScorer is a recently created automated tool that combines pocket prediction, identification, and druggability assessment. It is currently accessible via a web server. The free web server DoGSiteScorer is available for academic use at(http://dogsite.zbh.uni-hamburg.de) [82].

FIGURE 3.9: Home page of CELLO

The DoGSiteScorer web server offers a simple interface for predicting pockets and subpockets of a target protein. In addition, important parameters defining the pocket and druggability estimates are provided.

3.12 Molecular Docking

Molecular docking is the technique of docking tiny molecules into macromolecular structures in order to score complementary responses at binding sites. It is a thriving research field with significant usage in structure-based drug design, lead optimization, biochemical pathway design, and drug design as the most appealing tools [83]. There are five docking programs available such as, FlexX, Surflex, Glide, GOLD, and DOCK, as well as nine scoring algorithms, were examined for their capacity to rank-order possible lead compounds for a comprehensive virtual screening. In this study we used Moe a computational software for drug discovery that combines visualisation, modelling, simulation, and methodology development into a single package. Biologists, medicinal chemists, and computational chemists employ MOE scientific applications in pharmacological, biotechnology, and academic research. [84].

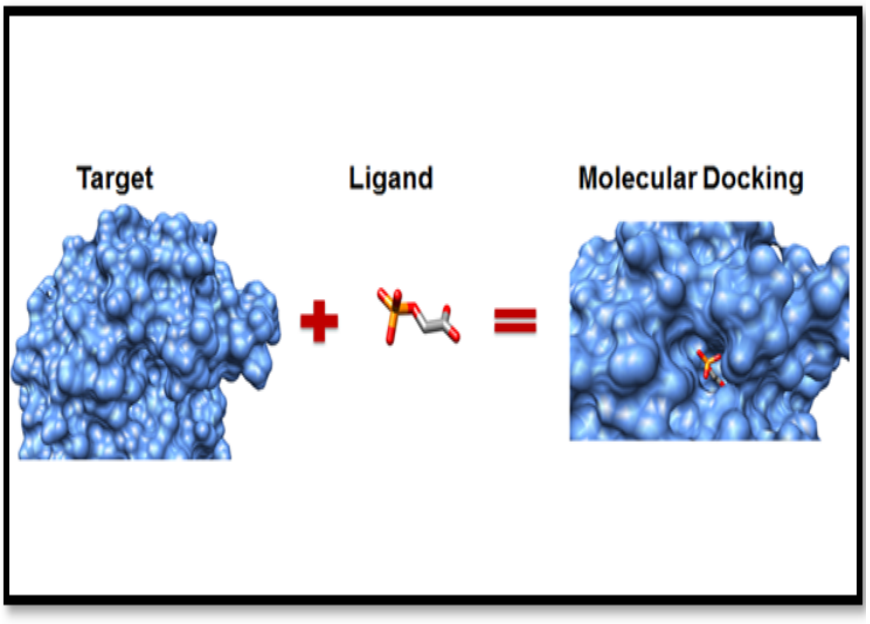

FIGURE 3.10: Home page of CELLO

Docking is used to find out the binding sites of protein, attempt to predict the structure of inter molecular complex formed between two molecules. It also play role in new drug designing.

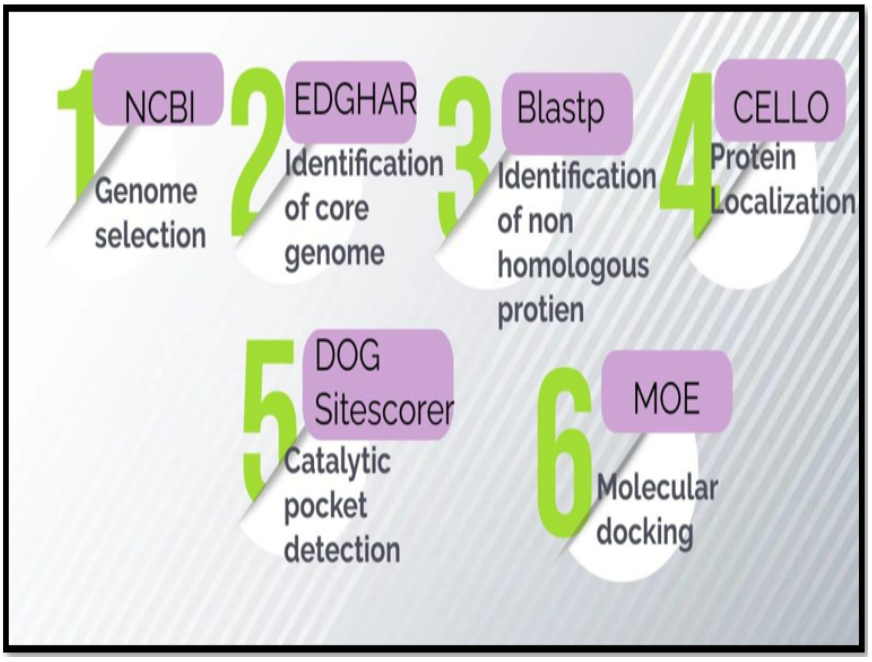

FIGURE 3.11: Summarizes the methodological steps to analyze the drug targeting approach.

These are very essential steps used in drug targeting analysis, contributing to new and reliable results and accelerating the drug discovery process. Each step provides sufficient evidence of biological action at disease-relevant targets, as well as satisfactory, safety and drug-like properties

Chapter 4

Results and Discussions

Streptococcus sanguinis is a member of gram positive, non-motile, human oral microbiota play key role in biofilm development. When it gain entrance into heart valves it causes bacterial endocarditis, leading towards life threatening of human being. It is not always harmful because it is commensal bacteria. It enters through injury in mouth, during tooth extraction and through surgical wound. The thesis design attempt to analyse pangenome of strain for drug targeting.

4.1 Selection of Strain

The initial stage in this study was to choose an inclusion and exclusion criterion for bacterial strain selection, with whole genomic sequence and G.C content accessible. At NCBI there are about 102 genomes are available in which some are cotings, some are present in scaffold level and few are completed as mentioed in table 4.1.

Table 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
SK405	SAMN 00253311	PRJNA53175	2.29829	43.3
SK408	SAMN 00253319	PRJNA53177	2.39857	42.8

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
SK353	SAMN 00253310	PRJNA53171	2.30108	43.4
VMC66	SAMN 00253296	PRJNA53191	2.31945	43.3
CC94A	SAMN 01162043	PRJNA71581	2.50704	42.9
SK1057	SAMN 00253315	PRJNA53183	2.34878	43.3
SK72	SAMN 00253313	PRJNA53159	2.37601	43.3
SK340	SAMN 00253970	PRJNA53169	2.40873	43
SK115	SAMN 00253314	PRJNA53161	2.34673	43.4
SK160	SAMN 00253316	PRJNA53165	2.33387	43.3
SK150	SAMN 02415615	PRJNA53163	2.32433	43.2
SK1058	SAMN 00253321	PRJNA53185	2.35283	43.2
ATCC	29667	SAMN00262642	PRJNA64743	2.42935
SK49	SAMN 00253323	PRJNA53157	2.2794	43
SK1087	SAMN 00253322	PRJNA53189	2.31502	43.2
SK1059	SAMN 00253969	PRJNA53187	2.42827	42.9
SK355	SAMN 00253324	PRJNA53173	2.37043	43.4

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
CGMH010	SAMN 11812007	PRJNA544156	2.34949	42.8
BCA8	SAMN 14598581	PRJNA625332	2.21824	43.2
SRR9217452 -mag-bin.9	SAMEA 8395238	PRJEB43277	2.26586	43.1
ERR2764863 -bin.2-meta WRAP-v1.1 -MAG	SAMEA 9943374	PRJEB41354	2.03643	43.6
JCVI-7 -bin.18	SAMN 14571057	PRJNA624185	2.14905	43.1
1112-SSAN	SAMN 03197066	PRJNA267549	3.31855	42.2
SRR2034637 -bin.1-meta WRAP-v1. 1-MAG	SAMEA 9921446	PRJEB26096	2.07514	43.3
DRR046102 -mag-bin.4	SAMEA 8391447	PRJEB42319	1.4973	44
SRR15235659 -bin.14-meta WRAP-v1. 1-MAG	SAMEA 10220496	PRJEB47694	1.34911	44.1
JCVI-49-bin.6	SAMN 14571056	PRJNA624185	1.86059	43.8

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
ERR3827271 -bin.6-meta WRAP-v1. 1-MAG	SAMEA 12514440	PRJEB49987	1.27049	43.6
JCVI-15 -bin.5	SAMN 14571055	PRJNA624185	1.79259	43.2
MA17	SAMN 20399954	PRJNA749742	2.40183	43
BCC37	SAMN 09631761	PRJNA480251	2.38689	43.1
KO71	SAMN 20399952	PRJNA749742	2.34899	43.4
S28	SAMN 12139124	PRJNA551103	2.36775	43.1
PJM8	SAMN 03480699	PRJNA280681	2.3679	43.2
S24	SAMN 26876568	PRJNA817585	2.3289	43.1
FSS4	SAMN 03480635	PRJNA280669	2.31267	43.2
10556	SAMN 20399946	PRJNA749742	2.28032	43.3
BCA3	SAMN 09631783	PRJNA480251	2.33834	43.3
BCC23	SAMN 09631774	PRJNA480251	2.38233	43.4
BCC54	SAMN 09631765	PRJNA480251	2.47088	42.9

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
MB451	SAMN 03480686	PRJNA280678	2.45276	42.9
KO29	SAMN 20399951	PRJNA749742	2.4093	43.3
SY17	SAMN 20399969	PRJNA749742	2.33834	43.4
ST499	SAMN 20399962	PRJNA749742	2.36188	43.2
BCC24	SAMN 09631775	PRJNA480251	2.31336	43.1
BCA13	SAMN 09631784	PRJNA480251	2.35932	43.3
FSS9	SAMN 03480643	PRJNA280671	2.42926	43.1
SK164	SAMN 20399957	PRJNA749742	2.34542	43.4
KO16	SAMN 20399949	PRJNA749742	2.42464	43.3
BCC28	SAMN 09631777	PRJNA480251	2.38096	43.1
BCC53	SAMN 09631753	PRJNA480251	2.47194	42.9
BCC46	SAMN 09631763	PRJNA480251	2.32262	43.4
KO19	SAMN 20399950	PRJNA749742	2.43162	43.3
BCC20	SAMN 09631773	PRJNA480251	2.37548	43.1

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
BCC03	SAMN 09631768	PRJNA480251	2.36813	43.1
KLC03	SAMN 09631779	PRJNA480251	2.36222	43.1
BCC04	SAMN 09631769	PRJNA480251	2.39259	42.9
BCA9	SAMN 09631712	PRJNA480251	2.30389	42.9
SY10	SAMN 20399967	PRJNA749742	2.46857	43.3
BCC16	SAMN 09631771	PRJNA480251	2.43373	43
BCC25	SAMN 09631776	PRJNA480251	2.29976	43.4
BCC18	SAMN 09631772	PRJNA480251	2.33724	43.2
OH953	SAMN 10478698	PRJNA506243	2.47692	43
KLC10	SAMN 09631782	PRJNA480251	2.34563	43.3
216-SSAN	SAMN 03197408	PRJNA267549	2.35743	43.1
711-SSAN	SAMN 03197916	PRJNA267549	2.33603	43.1
SW01	SAMN 20399965	PRJNA749742	2.41282	43.1
BCC39	SAMN 09631762	PRJNA480251	2.41613	43.2

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
BCC31	SAMN 09631778	PRJNA480251	2.29217	43.4
BCC14	SAMN 09631770	PRJNA480251	2.41836	43.1
BCC64	SAMN 09631767	PRJNA480251	2.36308	43.2
SY16	SAMN 20399968	PRJNA749742	2.43996	43.4
KLC08	SAMN 09631781	PRJNA480251	2.30582	43.6
OH0843	SAMN 13850658	PRJNA192961	2.45471	43
SK678	SAMN 00253312	PRJNA53179	2.29872	43.3
SK330	SAMN 00253320	PRJNA53167	2.41567	43
SK1056	SAMN 00253325	PRJNA53181	2.37669	43.1
SK36	SAMN 02604299	PRJNA13942	2.38843	43.4
NCTC	7863	SAMN03480625	PRJNA280651	2.30261
SK1	SAMN 00253318	PRJNA53155	2.28048	43.3
2908	SAMEA 3207422	PRJEB7884	2.30861	43.3
KO9	SAMN 20399953	PRJNA749742	2.31205	43.4
CR29	SAMN 20399948	PRJNA749742	2.48631	43

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
KLC04	SAMN 09631780	PRJNA480251	2.36619	43.1
STn373	SAMN 20399964	PRJNA749742	2.41021	43.2
ST02	SAMN 20399960	PRJNA749742	2.40461	43.2
STn342	SAMN 20399963	PRJNA749742	2.62747	43
BR3	SAMN 20399947	PRJNA749742	2.35511	43.1
S138	SAMN 14209003	PRJNA608797	2.35075	43.2
ST01	SAMN 20399959	PRJNA749742	2.40815	43.2
ST249	SAMN 20399961	PRJNA749742	2.39497	43.2
BCC61	SAMN 09631766	PRJNA480251	2.37952	43.1
MA4-6	SAMN 20399955	PRJNA749742	2.3947	43
SK162	SAMN 20399956	PRJNA749742	2.3469	43.2
SW02	SAMN 20399966	PRJNA749742	2.54622	42.9
SK37	SAMN 20399958	PRJNA749742	2.3732	43.4
NCTC7863	SAMEA 3672886	PRJEB6403	2.29693	43.3

TABLE 4.1: The table comprises 102 genomes of streptococcus sanguinis strain

Strain	**BioSample**	**BioProject**	**Size(Mb)**	**GC%**
NCTC11085	SAMEA 3632064	PRJEB6403	2.44953	43
NCTC	10904	SAMEA3307890	PRJEB6403	2.36022
FDAA RGOS -770	SAMN 11056485	PRJNA231221	2.35885	43.3
CGMH058	SAMN1157	PRJNA541204	2.34257	43
NCTC 11086	SAMEA364	PRJEB6403	2.33139	43.4

We chose eight complete bacterial strains with full genome annonated sequences. Table 4.2 summarises the details of the chosen strain of Streptococcus sanguinis for further investigation. These strains were validated by literature searches, and the genomic sequences of all eight strains were obtained from the NCBI database. By comparing gene content in publically accessible completed genomes and assemblies, these genes may be discovered in a sample set-independent way.

TABLE 4.2: Comprises all the complete strainswith its size, G.C content,and year of discovery.

Strain	**Size (mb)**	**G.C Content**	**Level**	**Year**
NCTC7863	2.45	43.3	Complete	2018
NCTC11085	2.3	43	Complete	2018
NCTC 10904	2.36	43.3	Complete	2018
FDAARGOS770	2.36	43.3	Complete	2020
CGMH058	2.34	43	Complete	2020
NCTC11086	2.33	43.4	Complete	2018
CGMH010	2.35	42.8	Complete	2020
SK36	2.39	43.4	Complete	2007

4.2 Selection of Reference Genome

Streptococcus sanguinis is a common inhabitant of human mouth and reported about 101 strains at NCBI, while only 8 strains are present in complete sequence with whole genome annotation.

TABLE 4.3: Summarizes the genomic characteristics of selected genome.

Genomic Feature	**Patric**	**Refseq**
Pseudogenes	2	4
Misc RNA	4	1
rRNA	8	12
tRNA	61	61
CDS	2263	2270

4.3 Sequence Similarity

In FASTA, the line, which comes before the nucleotide sequence, must start with a symbol (>) known as carat and be followed by one unique SeqID (sequence identifier). Using FASTA, 5 genes from S. sanguinis SK36 were obtained out from National Center for Biotechnology Information.

4.4 Non-Homologous Protein Identification

The obtained 5 critical proteins were tested against Homo sapiens using BlastP on the website of National Center for Biotechnology Information (NCBI). Proteins

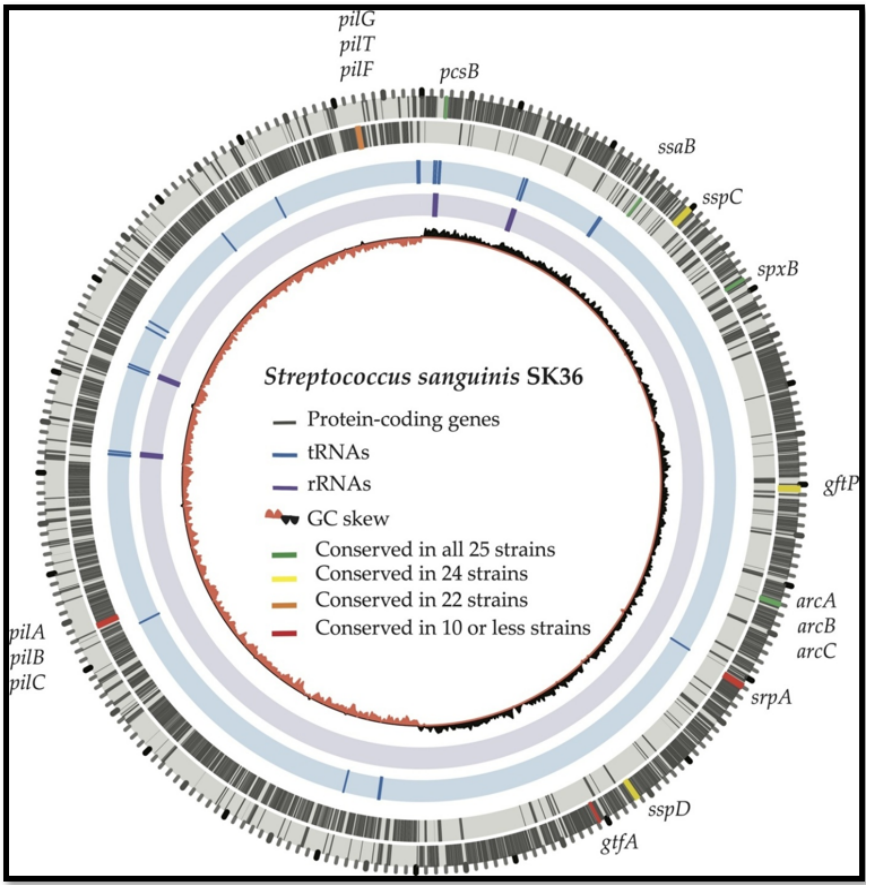

FIGURE 4.1: The diagram shows the circular view of geneome of Streptococcus sanguinis with its conserved regions

having dissimilarities to human nonhomologous essential proteins of *S.sanguinis* screened to differentiate human nonhomologous essential proteins of *S.sanguinis*. Out of 5 only 2 proteins having dissimilarities with the human nonhomologous essential protein.

4.5 Target Identification

Subcellular localization of *S. sanguinis* unique metabolic proteins (essential non-human homologues) to identify the proteins on the surface membrane that are or may be accessible proteins. An approach like this may be useful in identifying potential treatment targets. Following are the 2 proteins which is computed by ProtParam tool, it gives the information about molecular weight, amino acid composition, theoretical pI, overall quality factor etc. This will help us to find out the

chemical as well as physical value of under consideration protein. As a result, the ProtParam tool was used to compute protein properties.

TABLE 4.4: Protparam results shows the overall quality factor of amino acid count,Mw and PI Value of redox regulated ATPase protein.

Redox Regulated ATPase	
Number of amino acids	363
Molecular weight	39650.95
Theoretical pI	4.54
Overall quality factor	89.226

TABLE 4.5: Protparam results shows the overall quality factor of amino acid count,Mw and PI Value of aminoglycoside 3 phosphotransferase protein.

Aminoglycoside 3 Phosphotransferase	
Number of amino acids	258
Molecular weight	29500.42
Theoretical pI	5.43
Overall quality factor	90.8297

An ERRATE value of high-quality model should have a range greater than or equal to 37%, in this study the E- value is 89.226 and 90.8297 of the selected protein which is considered to be the ideal value for drug targeting analysis.

4.6 Rampage Analysis

RAMPAGE is the abbreviation of 'RNA Annotation and Mapping of Promoters for the Analysis of Gene Expression' serves to examine the stereochemical characteristics of a protein structure, whether it is experimentally available or has been

predicted. Various criteria have been developed to measure the stereochemical quality of protein structures both locally and globally. A number of parameters can be calculated using the coordinates of a particular structure. The psi, phi, and chi 1 torsion angles and hydrogen bond energy. The parameters used in this evaluation are ambiguous and can be easily calculated from structural coordinates. A suite of applications is currently being developed to quickly probe specific structures and display unexpected stereochemistry and errors.distributions are global parameters.

Errate is the overall quality factor for non-bonded atomic connections, where higher values indicating higher quality. The typical recognised limit or range for a high-quality model usually greater than 50. The ERRAT server projected the overall quality factor equals to 80 for the present 3D image. The given below graphs shows (97.819%) and (97.768%) respectively which is highly favourable.

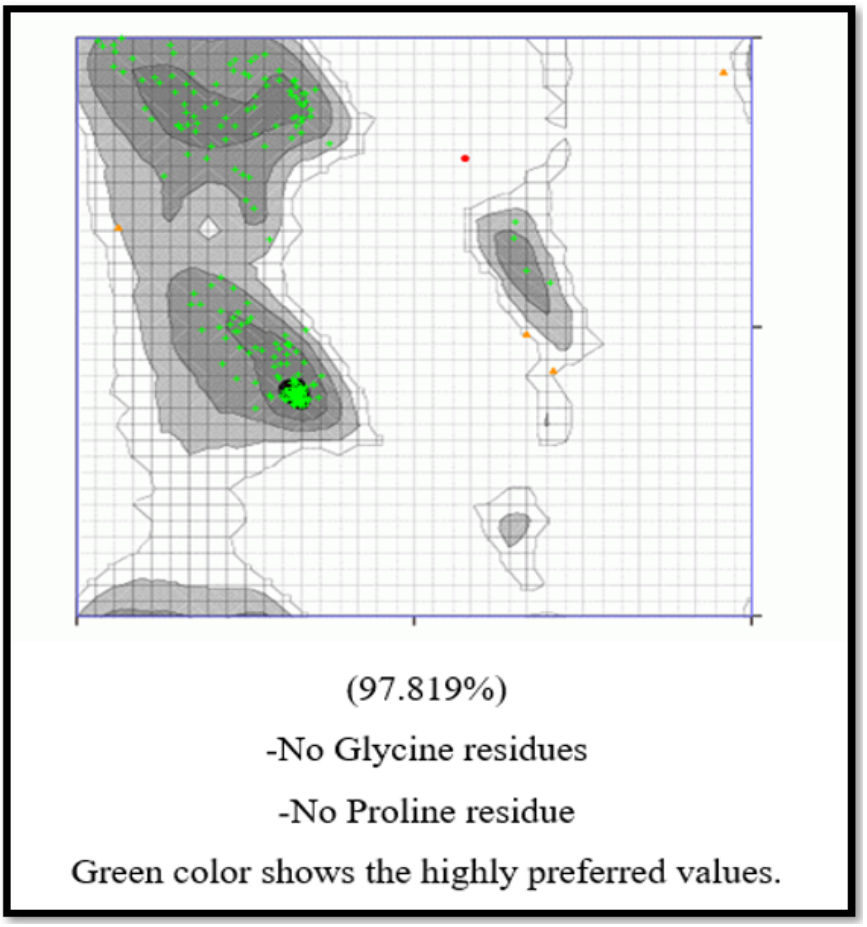

FIGURE 4.2: (a) The above Shown graphs of protein with coloured area and percentage.

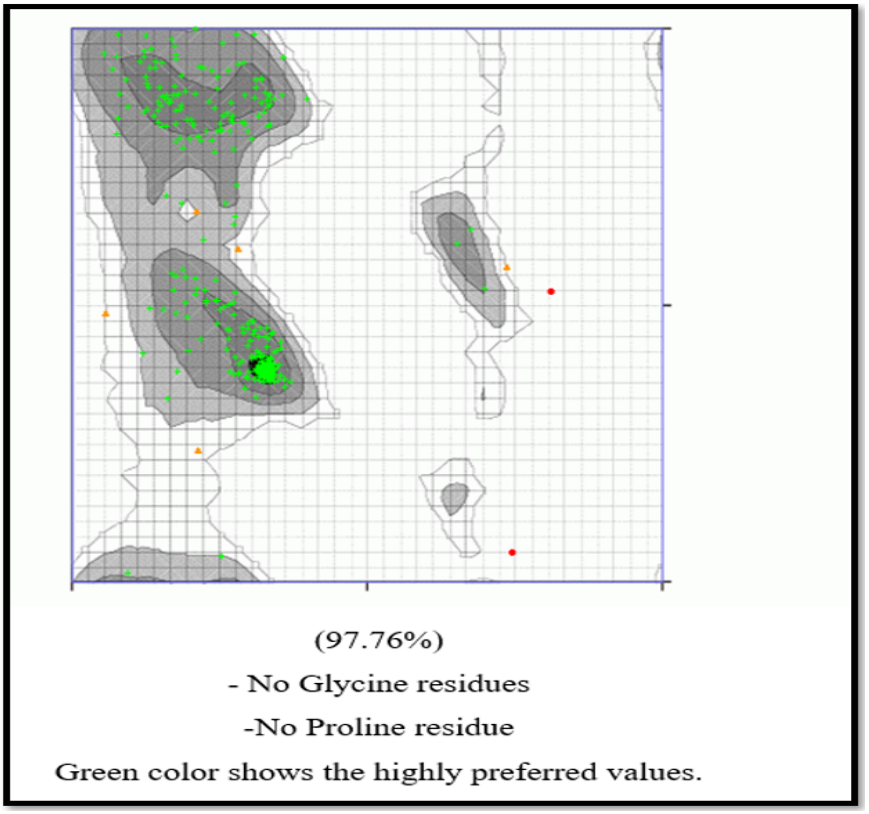

FIGURE 4.3: (b) The above Shown graphs of two protein with coloured area and percentage.

Rampage: a model with a percentage more than or equal to 80 is considered high-quality for drug targeting research. In this research the Rampage value of above proteins are 97.819 And 97.76%, these values are ideal and were taken for further analysis. We isolated 5 non-host homologous proteins from a total of 1570 proteins. Two important proteins were chosen by using two criteria which have identity more than 35 and e-value = 0.001.

4.7 Localization of Protein

The two proteins' prediction accuracies were obtained for both cytoplasmic and periplasmic proteins, extracellular and both(inner-membrane, outer-membrane) etc. The comparison of predictive performances of different approaches in the prediction of subcellular localization for eukaryotic sequences. CELLO attempts to predict at least single subcellular location based on the species of the protein.

The aggregate of cellular compartment, molecular function, and biological process is shown as a pie chart illustrating various functional annotations for the searched protein

TABLE 4.6: Summarises the localization and reliability of protein.

Redox Regulate ATPase	**Localization**	**Reliability**
Amino Acid Comp.	Cytoplasmic	0.979
N-peptide Comp.	Cytoplasmic	0.998
Partitiones Seq Comp.	Cytoplasmic	0.935
Physio-Chemical Comp.	Cytoplasmic	0.482
Neighbouring Seq. Comp.	Cytoplasmic	0.995

4.8 Cello Prediction at Cytoplasmic, Membrane, Extracellular and at Cell Wall Level

During computational studies we identified 2 different proteins which are cytoplasmic and easy to targeted.

TABLE 4.7: Table shows protein location and reliability of protein that is 4.389.

Localization	**Reliability**
Cytoplasmic	4.389 *
Membrane	0.476
Extracellular	0.124
Cell Wall	0.011

The table (4.7) explains that this protein is cytoplasmic and its reliability is high with a value of 4.389 and ideal for drug targeting.

TABLE 4.8: Summarises the localization and reliability of protein that is 4.389.

Aminoglycoside 3'-phosph-transferase	**Localization**	**Reliability**
Amino Acid Comp.	Cytoplasmic	0.643
N-peptide Comp.	Cytoplasmic	0.656
Partitiones Seq Comp.	Cytoplasmic	0.957
Physio-Chemical Comp.	Cytoplasmic	0.928
Neighbouring Seq.Comp.	Cytoplasmic	0.924

4.9 Cello Prediction at Cytoplasmic, Membrane, Extracellular and at Cell Wall Level

Subcellular localization predictive system is publicly available web based tool, used to determine cytoplasmic and membrane protein.

TABLE 4.9: Table shows protein location and reliability that is 4.107

Localization	**Reliability**
1)Cytoplasmic	4.107*
2)Extracellular	0.622
3)Membrane	0.245
4) CellWall	0.026

The table (4.9) explains that this protein is cytoplasmic and its reliability is high with a value of 4.107 and ideal for drug targeting.

4.10 Identification of Molecular and Biological Function

As indicated in the table, the molecular and biological functions of proteins, as well as their gene and protein, were collected by using the Uniprot tool.

TABLE 4.10: Aminoglycoside 3 phosphotransferase genome.

Aminoglycoside 3 phosphotransferase.
Gene name=FOC72-01690
Active site=182
Binding site=187
Length=258AA

TABLE 4.11: Table conclude the biological and molecular function of protein.

Molecular function	**Biological Function**
Metal ion binding.	Antibiotic resistance
Phosphotransferase activity, alcohol group as acceptor. ATP binding.	Phosphorylation

4.11 Catalytic Pocket Detection

DoGSiteScorer is a web-based application that may give both qualitative and quantitative data for druggability evaluation. A pocket's druggability can be predicted automatically by using this technique based on its size, shape, and chemical

properties. Furthermore, DoGSiteScorer may be used to evaluate the components, amino acid composition as well as functional groups, found on the selected pocket. The larger the value, the more likely the pocket is to be druggable.

The idea of molecular recognition patterns may be conserved throughout the binding pockets of proteins with similar capabilities inspired to create a new way to recognize protein relationships. Thera are about 27 residue detected from one protein while 24 residue from the other protein. The method entails automatically detecting and extracting potential binding sites from protein molecules.

4.12 Molecular Docking

Molecular docking is an in silico method for simulating the orientation and binding affinities (poses) of a ligand near the active center of a target macromolecule while trying to assess intermolecular forces such as hydrophobicity, ionic bonds, hydrogen bonds, Van der Waals forces and electrostatic interactions. Furthermore, this methodology provides interaction energy (docking scores), interaction types, and amino acid residues results in the preparation of the ligand-receptor complex. The scores are used as a guide to rank the ligand's most stable poses. As a result, the lower the S score, creating the stronger and more stable interaction with the chosen target.

4.13 Significance of Molecular Docking

Molecular docking facilitates rapid screening of databases of approved pharmaceuticals, natural products, or previously synthesized compounds against one or more biological targets of interest. This is a powerful and powerful method for in silico research. It is becoming increasingly important for rational dosing design. Docking is a computational process for finding suitable ligands that fit both energetically and geometrically into the binding site of a protein.

4.14 Aminoglycoside 3 Phosphotransferase

It is a cytoplasmic protein with amino acid composition 0.643. The aminoglycoside phosphotransferase (APH) family, which chemically modifies and inactivates aminoglycoside antibiotics such as streptomycin and kanamycin via ATP-mediated phosphate transfer. Its length is 268 and mass in Da is 29,986 and are widely present in resistant bacteria.

TABLE 4.12: Table shows the docking result of the observed protein.

Compound	**S-score**	**Binding energy (MM\GBVI)**	**Residues**	**No of interaction**
Sperimidne	-8.8154	-22.784	Glu 255 Tyr 215	3
Campesterol	-10.3668	-16.538	Ile 251 Tyr146	1
Berberine	-10.5669	-12.595	His127	2
Coumarin	-6.9817	-11.835	GLn135 Asn139	1
Quercitin	-8.8809	-14.777	Asn139 PHE 257	2
Scopoletin	-9.8777	-13.321	Tyr132 Gly181	2
Stigmasterol	-8.7781	-11.329	Ala252	1

TABLE 4.12: Table shows the docking result of the observed protein.

Compound	**S-score**	**Binding energy (MM\GBVI)**	**Residues**	**No of interaction**
Rutin	-14.0171	-25.036	Gln53 Gly181 Tyr132	4
Kaempferol	-12.8931	-22.932	Gln53 Asn139 Phe254 Phe257	4

Docking can be used to show the strength of a compound's binding interaction with a protein.The number of residues engaged in the interaction.The number of interactions made between the active site and the ligand.The table 4.6 showing the compound with the binding affinities and interactions with the Aminoglycoside 3 phosphotransferase .

4.15 Sperimidne

The compound sperimidne is selected on the basis of S- score with -8.8154 and having 3 interaction. Its molecular weight is 145.25g/mol. It is used in cardiovascular diseases and cancer treatment. Chemical formula of sperimidne is $C_7H_{19}N_3$.

4.15.1 Source

Sperimidne are found in specific mushrooms and some lentils, green peas and dry soy beans.

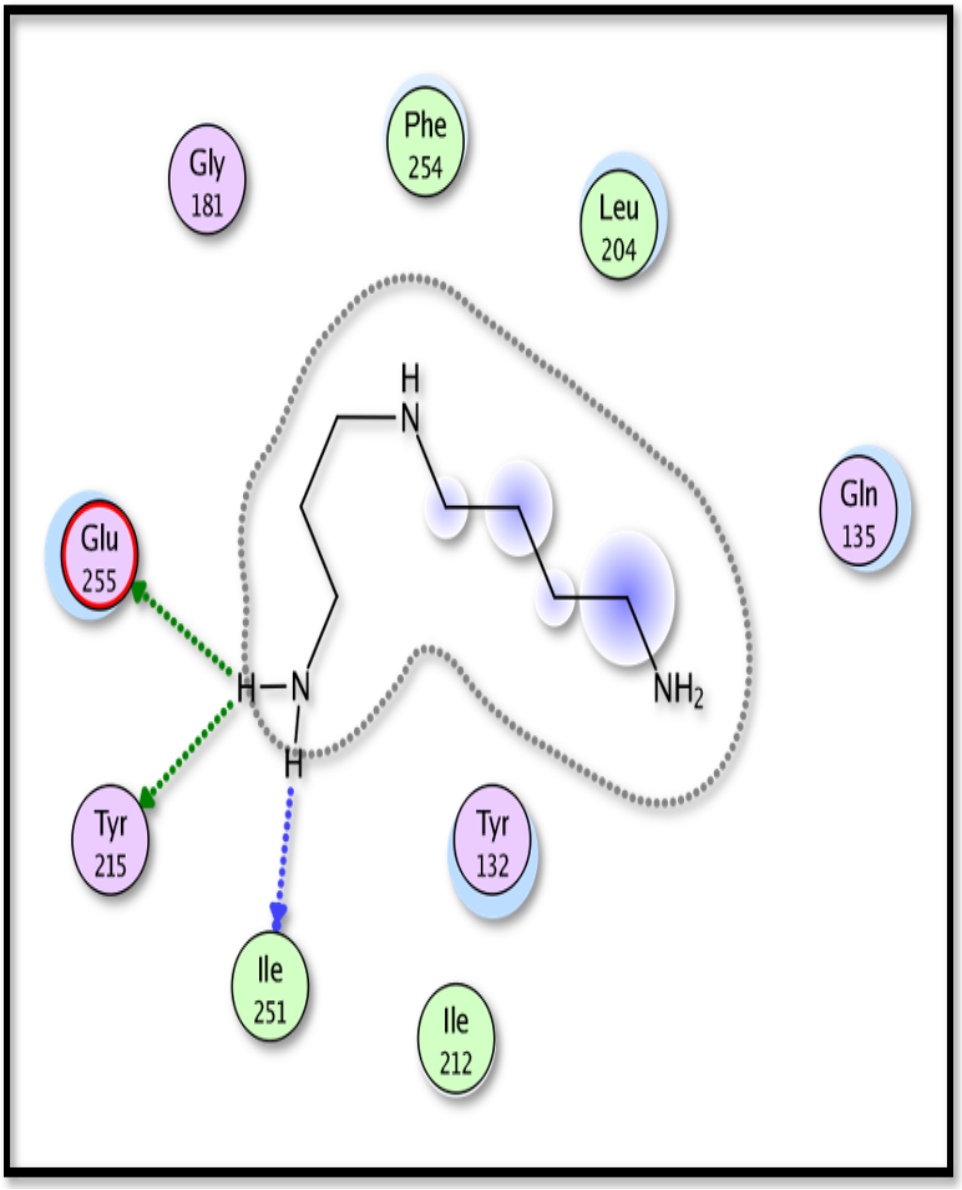

FIGURE 4.4: The structure of compound sperimidne shows the 3 interactions in 2D view.

4.16 Redox Regulated ATPase

The mitochondrial ATP synthase protein complex has a molecular weight of around 550 kDa. Furthermore, structural elucidation has opened up new avenues for the creation of innovative ATP synthase- directed drugs with potentially therapeutic benefits.

To far, over 250 both natural and synthetic inhibitors have also been classified, including data on their inhibitory locations and mechanisms of action.

TABLE 4.13: The table shows the docking result of selected protein.

Compound	**S-score and mseq**	**Binding Energy (MM/GBVI)**	**Residues**	**No of interactions**
Sperimidne	-8.6373 (1)	-12.368	Arg24 Glu 30	2
Berberine	-12.4962 (3)	-13.062	2/Arg24	2
Coumarin	-8.3116 (4)	-15.832	Arg24	1
Quercitin	-11.7852 (5)	-19.319	Arg24 Leu 242	2
Scopoletin	-9.7543 (6)	-11.341	Arg24	2
Rutin	-11.5280 (8)	-20.303	Asn 31 Pro 33	3
Kaempferol	-8.0087 (8)	-19.235	Arg 24 Arg 24 Glu247	2

Docking can be used to show the strength of a compound's binding interaction with a protein.The number of residues engaged in the interaction. The number of interactions made between the active site and the ligand. The table 4.13 showing

the compound with the binding affinities and interactions with Redox regulated ATPase.

4.17 Kaempferol

This compound is selected on the basis of S- score, having 2 interactions. It is a plant metabolite and widely distributed in edible plants with a molecular weight 286.24g/mol. It act as an antioxidant and emollient. It plays role as an antibacterial agent and also used in cancer treatment such as pancreatic and colorectal cancer. It also used in cardiovascular disrases. Chemical formula of Kaemferol is $C_{15}H_{10}O_6$.

4.17.1 Source

It is found in lotus and *Ardisia sanguinolenta*, beans, broccoli, spinach etc.

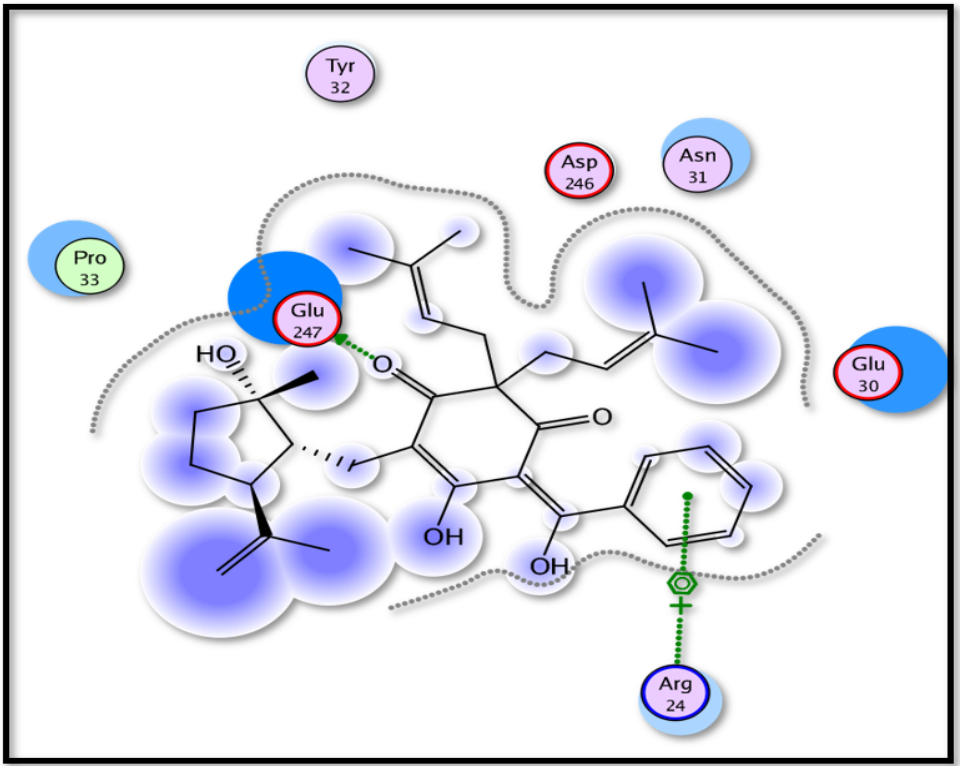

FIGURE 4.5: The structure of kaempferol shows the 2 interactions in 2D view.

Chapter 5

Conclusions and Recommendations

Streptococcus sanguinis belongs to the streptococcus viridans group, a Gram-positive, facultative aerobic bacterium. It is prevalent in human mouths in dental plaque, which causes both dental cavities and bacterial endocarditis and has a 25% death rate. Furthermore, *S. sanguinis* is regarded as opportunistic pathogen and one of the most common causes of infective endocarditis. The goal of this study is to look at the involvement of *Streptococcus sanguinis* in virulence, pathogenicity, biofilm development, and infection mechanism. *S. sanguinis* are part of the human oral microbiota and have the capacity to create dental plaques, thus they might play a role in the development of IE. The pan genome method was used in this investigation to identify therapeutic targets for *Streptococcus sanguinis sk36* strain and design an anti-infection medicine because traditional drugs generate health hazards and sometimes it's not effective for human. Number of bioinformatics tools was used to retrieve essential proteins, find out metabolic pathway, defining the pockets, active sites and draggability parameters. The goal of this research was to discover potential therapeutic targets to find alternative antibiotic therapies. To provide more information, we looked for the active sites of the retrieved two proteins and also examined their interactions with other proteins. The reason behind the selection of *S. sanguinis strain SK36* is that its genome

has been sequenced, and so that this bacterium is particularly receptive to genetic modification. Antibiotics are commonly used to treat dental plaque and common human occupants of the mouth, which are thought to be the most prevalent cause of infective endocarditis. Based on these findings, medications can be developed to more precisely target and eradicate *S. sanguinis* while causing no damage to the patient. To avoid any cross-reactivity of the identified drug with the human host, BlastP was performed alongside the 5 required proteins, and then the two nonhuman homologous proteins were retrieved. Molecular modelling and target analysis assisted in identifying the best feasible active sites for drug design. Because a subtractive genetic technique was used in this work, the host will not be harmed by this pathogen-specific treatment.

The treatment against this infection is totally baesd on natural compound derived from plants, these compound retrieved from the PubChem, with its resources and molecular weight.Hence these compound play essential role in drug designing. Computational analysis against these new targets will aid in the development of promising anti-*S. sanguinis* therapeutics. *S. sanguinis* is a significant source of future research. In comparison to cariogenic S. mutans, our genome study revealed key genetic data for researching the aetiology of caries. The genetics and metabolism of this essential bacteria have been elucidated, enabling the development of new therapeutic and preventative methods.

This research concentrates on numerous parts of the genome *S. sanguinis* in order to identify key nonhomologous metabolic membrane-bound cytoplasmic proteins that might be used as therapeutic targets. There is always need for betterment in every aspect that's why further work is required to elucidate the life threatening infections, and replacement of antibiotics which is harmful for human. With replacement of antibiotics, make new drugs for fighting against diseases in short time and provide better strategy against bacterial infection.

Bibliography

[1] B. Zhu, L. C. Macleod, T. Kitten, and P. Xu, "Streptococcus sanguinis biofilm formation & interaction with oral pathogens," *Future Microbiol.*, vol. 13, pp. 915–932,

[2] C. Morita et al., "Cell wall-anchored nuclease of Streptococcus sanguinis contributes to escape from neutrophil extracellular trap-mediated bactericidal activity," *PLoS One*, vol. 9, no. 8, p. e103125, 2014.

[3] P. Xu et al., "Genome of the opportunistic pathogen Streptococcus sanguinis," *J. Bacteriol.*, vol. 189, no. 8, pp. 3166–3175, 2007

[4] A. M. Martini et al., "Association of novel streptococcus sanguinis virulence factors with pathogenesis in a native valve infective endocarditis model," *Front. Microbiol.*, vol. 11, p. 10, 2020.

[5] V. S. Lauren Senty Turner B, Ed., "A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University," in Director: TODD KITTEN, PH.D. ASSOCIATE PROFESSOR, PHILIPS INSTITUTE, SCHOOL OF DENTISTRY, 2002.

[6] P. Xu et al., "Genome of the opportunistic pathogen Streptococcus sanguinis," *J. Bacteriol.*, vol. 189, no. 8, pp. 3166–3175, 2007.

[7] F. L. Macrina and P. H. Wood, Eds., *Genetic Transformation of Streptococcus sanguis (Challis) with Cryptic Plasmids from Streptococcus ferus Authors.* Vol.28 no(3),p. 692–699. doi: 10.1128/iai.28.3.692-699.1980.

[8] J. Kreth, R. A. Giacaman, R. Raghavan, and J. Merritt, "The road less traveled - defining molecular commensalism withStreptococcus sanguinis," *Mol. Oral Microbiol.*, vol. 32, no. 3, pp. 181–196, 2017.doi: 101111/omi.12170.

[9] P. W. Caufield, A. P. Dasanayake, Y. Li, Y. Pan, J. Hsu, and J. M. Hardin, "Natural history of Streptococcus sanguinis in the oral cavity of infants: evidence for a discrete window of infectivity," *Infect. Immun.*, vol. 68, no. 7, pp. 4018–4023,

[10] W. J. Loesche, J. Rowan, L. H. Straffon, and P. J. Loos, "Association of Streptococcus mutans with Human Dental Decay," *Infect. Immun.*, vol. 11, no. 6, pp. 1252–1260, 1975.2000.doi;10.1128/iai.68.7.4018-4023.2000.

[11] C. W. Douglas, J. Heath, K. K. Hampton, and F. E. Preston, "Identity of viridans streptococci isolated from cases of infective endocarditis," *J. Med. Microbiol.*, vol. 39, no. 3, pp. 179–182, 1993.

[12] C. L. Schoch et al., "NCBI Taxonomy: a comprehensive update on curation, resources and tools," *Database (Oxford)*, vol. 2020, 2020.

[13] P. E. Kolenbrander and J. London, "Adhere today, here tomorrow:oral bacterial adherence," *J. Bacteriol*, vol. 175, pp. 3247–3252, 1993.

[14] J. Carlsson, H. Grahnén, G. Jonsson, and S. Wikner, "Establishment of Streptococcus sanguis in the mouths of infants," *Arch. Oral Biol.*, vol. 15, no. 12, pp. 1143–1148, 1970.

[15] P. Xu et al., "Genome of the opportunistic pathogen Streptococcus sanguinis," *J. E. M Bacteriol.*, vol. 189, no. 8, pp. 3166–3175, 2007. .

[16] J. van Houte, "Role of micro-organisms in caries etiology," *J. Dent. Res.*, vol. 73, no. 3, pp. 672–681, 1994

[17] Bik et al., "Bacterial diversity in the oral cavity of 10 healthy individuals," *ISME J.*, vol. 4, no. 8, pp. 962–974, 2010

[18] J. Kreth, J. Merritt, W. Shi, and F. Qi, "Competition and coexistence between Streptococcus mutans and Streptococcus sanguinis in the dental biofilm," *J. Bacteriol.*, vol. 187, no. 21, pp. 7193–7203, 2005.

[19] W. J. Loesche, J. Rowan, L. H. Straffon, and P. J. Loos, "Association of Streptococcus mutans with Human Dental Decay," *Infect. Immun.*, vol. 11, no. 6, pp. 1252–1260, 1975.

[20] B.Y. Wang and H. K. Kuramitsu, "Interactions between Oral Bacteria: Inhibition of Streptococcus mutans Bacteriocin Production by Streptococcus gordonii," *Applied and Environmental Microbiology*, vol. 71, no. 1, pp. 354–362, Jan. 2005, doi: 10.1128/aem.71.1.354-362.2005

[21] B. Ghassemia, "Long-Term Treatment of Early Childhood Caries (ECC)," *Interventions in Pediatric Dentistry Open Access Journal*, vol. 4, no. 2, Jun. 2020, doi: 10.32474/ipdoaj.2020.04.000181.

[22] R. A. Giacaman, S. Torres, Y. Gómez, C. Muñoz-Sandoval, and J. Kreth, "Correlation of Streptococcus mutans and Streptococcus sanguinis colonization and ex vivo hydrogen peroxide production in carious lesion-free and high caries adults," *Archives of Oral Biology*, vol. 60, no. 1, pp. 154–159, Jan. 2015, doi: 10.1016/j.archoralbio.2014.09.007.

[23] P. W. Caufield, A. P. Dasanayake, Y. Li, Y. Pan, J. Hsu, and J. M. Hardin, "Natural history of Streptococcus sanguinis in the oral cavity of infants: Evidence for a discrete window of infectivity," *Infect. Immun.*, vol. 68, no. 7, pp. 4018–4023, 2000.

[24] B. Zhu, L. C. Macleod, T. Kitten, and P. Xu, "Streptococcus sanguinis biofilm formation & interaction with oral pathogens," *Future Microbiol.*, vol. 13, pp. 915–932, 2018.

[25] H.-C. Flemming, T. R. Neu, and D. J. Wozniak, "The EPS matrix: the 'house of biofilm cells,'" *J. Bacteriol.*, vol. 189, no. 22, pp. 7945–7947, 2007.

[26] L. K. Kopec, A. M. Vacca Smith, D. Wunder, L. Ng-Evans, and W. H. Bowen, "Influence of antibody on the structure of glucans," *Caries Res.*, vol. 36, no. 2, pp. 108–115, 2002.

[27] P. Xu et al., "Genome of the Opportunistic Pathogen Streptococcus sanguinis," *Journal of Bacteriology*, vol. 189, no. 8, pp. 3166–3175, Feb. 2007, doi: 10.1128/jb.01808-06.

[28] K. Fujishima, M. Kawada-Matsuo, Y. Oogai, M. Tokuda, M. Torii, and H. Komatsuzawa, "dpr and sod in Streptococcus mutans are involved in coexistence with S. sanguinis, and PerR is associated with resistance to H2O2," *Appl. Environ. Microbiol.*, vol. 79, no. 5, pp. 1436–1443, 2013.

[29] C. Alves-Barroco, J. Paquete-Ferreira, T. Santos-Silva, and A. R. Fernandes, "Singularities of Pyogenic Streptococcal Biofilms – From Formation to Health Implication," *Frontiers in Microbiology*, vol. 11, Dec. 2020, doi: 10.3389/fmicb.2020.584947.

[30] K. Evans, V. Stone, L. Chen, X. Ge, and P. Xu, "Systematic study of genes influencing cellular chain length in Streptococcus sanguinis," *Microbiology*, vol. 160, no. Pt 2, pp. 307–315, 2014

[31] M. L. Higgins, H. M. Pooley, and G. D. Shockman, "Site of initiation of cellular autolysis in Streptococcus faecalis as seen by electron microscopy," *J. Bacteriol.*, vol. 103, no. 2, pp. 504–512, 1970.

[32] E. A. Thibodeau and C. M. Ford, "Chain formation and de-chaining in Streptococcus sobrinus SL-1," *Oral Microbiol. Immunol.*, vol. 6, no. 5, pp. 313–315, 1991.

[33] Evans, V. Stone, L. Chen, X. Ge, and P. Xu, "Systematic study of genes influencing cellular chain length in Streptococcus sanguinis," *Microbiology*, vol. 160, no. 2, pp. 307–315, Feb. 2014, doi: 10.1099/mic.0.071688-0.

[34] N. Okahashi et al., "Pili of oral Streptococcus sanguinis bind to salivary amylase and promote the biofilm formation," *Microb. Pathog.*, vol. 50, no. 3–4, pp. 148–154, 2011.

[35] J.-L. Berry et al., "Global biochemical and structural analysis of the type IV pilus from the Gram-positive bacterium Streptococcus sanguinis," *J. Biol. Chem.*, vol. 294, no. 17, pp. 6796–6808, 2019.

[36] K. H. Piepenbrink and E. J. Sundberg, "Motility and adhesion through type IV pili in Gram-positive bacteria," *Biochemical Society transactions*, vol. 44, no. 6, pp. 1659–1666, Dec. 2016, doi: 10.1042/BST20160221.

[37] Y.Y. M. Chen et al., "Molecular and functional analysis of the type IV pilus gene cluster in Streptococcus sanguinis SK36," *Appl. Environ. Microbiol.*, vol. 85, no. 6, 2019.

[38] Gurung et al., "Functional analysis of an unusual type IV pilus in the Gram-positive Streptococcus sanguinis," *Molecular Microbiology*, vol. 99, no. 2, pp. 380–392, Oct. 2015, doi: 10.1111/mmi.13237.

[39] T. E. Paterick, T. J. Paterick, R. A. Nishimura, and J. M. Steckelberg, "Complexity and subtlety of infective endocarditis," *Mayo Clin. Proc.*, vol. 82, no. 5, pp. 615–621, 2007. .

[40] L. L. Burrows, "Pseudomonas aeruginosa twitching motility: type IV pili in action," *Annual Review of Microbiology*, vol. 66, pp. 493–520, 2012.

[41] R. E. Mendes, L. M. Deshpande, J. Kim, D. S. Myers, J. E. Ross, and R. N. Jones, "Streptococcus sanguinis isolate displaying a phenotype with cross-resistance to several rRNA-targeting agents," *J. Clin. Microbiol.*, vol. 51, no. 8, pp. 2728–2731, 2013.

[42] Xu, J. Alves, T. kitten., "Genome of the Opportunistic Pathogen Streptococcus sanguinis," *Journal of Bacteriology*, vol. 189, no. 8, pp. 3166–3175, Feb. 2007, doi: 10.1128/jb.01808-06.

[43] F. Martínez, F. Martín-Luengo, A. García, and M. Valdés, "Treatment with various antibiotics of experimental endocarditis caused by penicillin-resistant Streptococcus sanguis," *Eur. Heart J.*, vol. 16, no. 5, pp. 687–691, 1995.

[44] F. A. G. Vallejo, "Epidemiology of Infective Endocarditis," in Contemporary Challenges in Endocarditis, InTech, 2016.

[45] J. J. Federspiel, S. C. Stearns, A. F. Peppercorn, V. H. Chu, and V. G. Fowler, "Increasing US rates of endocarditis with Staphylococcus aureus: 1999-2008," *Archives of Internal Medicine*, vol. 172, no. 4, pp. 363–365, Feb. 2012, doi: 10.1001/archinternmed.2011.1027.

[46] W. Tsang, "Recent advances in understanding and managing mitral valve disease," *F1000Research*, vol. 8, p. 1686, Sep. 2019, doi: 10.12688/f1000research.16066.1.

[47] I. E. B. Guy, P. Armstrong, W. Md, H. District, and W. Board, Auckland Last full review/revision. 2022.

[48] W. Tsang, "Recent advances in understanding and managing mitral valve disease," *F1000Research*, vol. 8, p. 1686, Sep. 2019, doi: 10.12688/f1000research.16066.1.

[49] I. E. B. Guy, P. Armstrong, W. Md, H. District, and W. Board, Auckland Last full review/revision. 2022.

[50] B. Wisniewska-Spychala et al., "Dentigenous infectious foci – a risk factor of infective endocarditis," *Medical Science Monitor*, vol. 18, no. 2, pp. CR93–CR104, 2012, doi: 10.12659/msm.882464.

[51] R. Mocchegiani and M. Nataloni, "Complications of infective endocarditis," Cardiovasc. Hematol. Disord. *Drug Targets*, vol. 9, no. 4, pp. 240–248, 2009.

[52] K. Werdan et al., "Mechanisms of infective endocarditis: pathogen-host interaction and risk states," *Nat. Rev. Cardiol.*, vol. 11, no. 1, pp. 35–50, 2014.

[53] J. R. McDonald, "Acute infective endocarditis," *Infect. Dis. Clin. North Am.*, vol. 23, no. 3, pp. 643–664, 2009

[54] J Freers, H Mayanja-Kizza, JL Ziegler, M Rutakingirwa 'Echocardiographic diagnosis of heart disease in Uganda,'Trop Doct, pp. 125-128, 26. 1996

[55] M. D. Mehmet Akif Öztürk et al., Infective endocarditis and complications; a single center experience Enfektif endokardit ve komplikasyonları; tek merkez deneyimi Olcay Özveren. İstanbul.

[56] P. Kovuri, S. Senthil Kumaran, and T. Chatterjee, "Streptococcus sanguinis Endocarditis of Bicuspid Aortic Valve Presenting as Septic Arthritis of Lumbar Facet Joint," Cureus, Apr. 2022, doi: 10.7759/cureus.24189.

[57] M. Yilmaz, "Right atrial thrombus following closure of an atrial septal defect," *Heart*, vol. 89, no. 7, pp. 726–726, Jul. 2003, doi: 10.1136/heart.89.7.726.

[58] Wang, J. G. Gaca, and V. H. Chu, "Management Considerations in Infective Endocarditis," *JAMA*, vol. 320, no. 1, p. 72, Jul. 2018, doi: 10.1001/jama.2018.7596.

[59] Y.-A. Que et al., "Fibrinogen and fibronectin binding cooperate for valve infection and invasion in Staphylococcus aureus experimental endocarditis," *Journal of Experimental Medicine*, vol. 201, no. 10, pp. 1627–1635, May 2005, doi: 10.1084/jem.20050125.

[60] M. R. Moon, E. B. Stinson, and D. C. Miller, "Surgical treatment of endocarditis," *Prog. Cardiovasc. Dis.*, vol. 40, no. 3, pp. 239–264, 1997

[61] W. G. Daniel et al., "Improvement in the diagnosis of abscesses associated with endocarditis by transesophageal echocardiography," *N. Engl. J. Med.*, vol. 324, no. 12, pp. 795–800, 1991

[62] E. Mylonakis and S. B. Calderwood, "Infective endocarditis in adults," *The New England Journal of Medicine*, vol. 345, no. 18, pp. 1318–1330, Nov. 2001, doi: 10.1056/NEJMra010082

[63] S. A. S. Khan, B. A. Yaqub, and S. M. Al Deeb, "Neurological Complications of Infective Endocarditis," *Annals of Saudi Medicine*, vol. 16, no. 3, pp. 254–256, May 1996, doi: 10.5144/0256-4947.1996.254.

[64] G. Habib et al., "2015 ESC Guidelines for the management of infective endocarditis," *European Heart Journal*, vol. 36, no. 44, pp. 3075–3128, Aug. 2015, doi: 10.1093/eurheartj/ehv319.

[65] B. D. Prendergast, "Diagnosis of infective endocarditis," *BMJ*, vol. 325, no. 7369, pp. 845–846, Oct. 2002, doi: 10.1136/bmj.325.7369.845.

[66] I. Jain, P. Jain, D. Bisht, A. Sharma, B. Srivastava, and N. Gupta, "Use of traditional Indian plants in the inhibition of caries-causing bacteria–Streptococcus mutans," *Braz. Dent. J.*, vol. 26, no. 2, pp. 110–115, 2015.

[67] P. Yadav et al., "Deciphering Streptococcal Biofilms," *Microorganisms*, vol. 8, no. 11, p. 1835, Nov. 2020, doi: 10.3390/microorganisms8111835.

[68] H. Elgamily, R. Safy, and R. Makharita, "Influence of medicinal plant extracts on the growth of oral pathogens Streptococcus mutans and Lactobacillus acidophilus: An in-vitro study," *Open Access Maced. J. Med. Sci.*, vol. 7, no. 14, pp. 2328–2334, 2019.

[69] Y. Kim, C. Gu, H. U. Kim, and S. Y. Lee, "Current status of pan-genome analysis for pathogenic bacteria," *Current Opinion in Biotechnology*, vol. 63, pp. 54–62, Jun. 2020, doi: 10.1016/j.copbio.2019.12.001.

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

[71] M. A. Brockhurst, E. Harrison, J. P. J. Hall, T. Richards, A. McNally, and C. MacLean, "The Ecology and Evolution of Pangenomes," *Current Biology*, vol. 29, no. 20, pp. R1094–R1103, Oct. 2019, doi: 10.1016/j.cub.2019.08.012.

[72] H. C. den Bakker et al., "Comparative genomics of the bacterial genus Listeria: Genome evolution is characterized by limited gene acquisition

and limited gene loss," *BMC Genomics*, vol. 11, no. 1, p. 688, 2010, doi: 10.1186/1471-2164-11-688.

[73] M. Suchitra, A. Ajjarapu, P. Tiwari, D. B. Wasudeo Ramteke, and S. Singh, "Chapter 15 - Ligand-based drug designing," in Bioinformatics, D. B. Singh and R. K. Pathak, Eds. Academic Press, 2022, pp233–252

[74] N. A. Qureshi, S. M. Bakhtiar, M. Faheem, M. Shah, A. Bari, H. M. Mahmood, M. Sohaib, R. A. Mothana, R. Ullah, and S. B. Jamal, "Genome based drug target identification in human pathogen streptococcus gallolyticus," *Frontiers in Genetics*, vol. 12, p. 303, 2021. 233–252.

[75] R Ing provides Novel Evolutionary Insights for Differentiation Raffael C," *BMC Genomics*, vol. 19, 2018.

[76] J. Blom et al., "EDGAR: A software framework for the comparative analysis of prokaryotic genomes," *BMC Bioinformatics*, vol. 10, no. 1, May 2009, doi: 10.1186/1471-2105-10-154.

[77] Altschul, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, 1990.

[78] R. Apweiler et al., "UniProt: The Universal Protein knowledgebase," *Nucleic Acids Res.*, vol. 32, no. Database issue, pp. D115-9, 2004.

[79] F. Ag¨uero, B. Al-lazikani, M. Aslett, M. Berriman, S. Frederick, R. Campbell, S. Carmona, and I. Carruthers, "Edith a w.; chen f; et al," *Database*, vol. 7, no. 11, pp. 900–907, 2008.

[80] S. I. Mondal, S. Ferdous, N. A. Jewel, A. Akter, Z. Mahmud, M. M. Islam, T. Afrin, and N. Karim, "Identification of potential drug targets by subtractive genome analysis of escherichia coli o157: H7: an in silico approach," Advances and applications in bioinformatics and chemistry: *AABC*, vol. 8, p. 49, 2015.

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

[82] M. Kanehisa, "Toward pathway engineering: a new database of genetic and molecular pathways," Science & Technology Japan, no. 59, pp. 34–38, 1996

[83] A. Volkamer, D. Kuhn, F. Rippmann, and M. Rarey, "Dogsitescorer: a web server for automatic binding site prediction, analysis and druggability assessment," *Bioinformatics*, vol. 28, no. 15, pp. 2074–2075, 2012.

[84] N. Tomar, V. Singh, S. Marla, R. Chandra, R. Kumar, and A. Kumar, "Molecular docking studies with rabies virus glycoprotein to design viral therapeutics," *Indian journal of pharmaceutical sciences*, vol. 72, no. 4, p. 486, 2010