CHAPTER 3: TARGET SELECTION AND COMPUTATIONAL ASSISTED DRUG DESIGNING OF NOVEL INHIBITORS

3.1 Introduction

Amongst all, the genus Streptococcus, a Gram-positive cocci, facultative anaerobe, comprises a wide variety of commensal and pathogenic species. Notably, the oral streptococci, such as *Streptococcus mutans* inhabit the oral cavity and attach to the dental structures as dental plaque leading to the development of dental caries [83]. There are several incidences of occurrence of life threatening infective endocarditis involving biofilm formation on indigenous or prosthetic heart valves by *Streptococcus mutans* that enters occasionally in the human blood circulatory system from the gingival pocket during cariogenic dental infection [84].

In *S. mutans*, the persistent nature of biofilm due to a high production of extra-polymeric substances and its intrinsic ability to resist antibiotics are the main causes of the chronic nature of this infection [85]. The biofilm formation is postulated as one of the major virulence factors well known to be expressed and regulated by quorum sensing system prevalent in many Gram-positive and Gram-negative bacteria [86]. Quorum sensing refers to the approach that bacteria utilize to converse with the other bacteria of the same species as well as with bacteria from other species in a cell-density dependent manner via., autoinducers (in Gram-negative bacteria) and peptides (in Gram-positive bacteria) [87]. This system enables bacteria to respond to various growth conditions such as nutritional availability, population stress due to antibiotics in the form of co-ordinated group behaviour for successful survival in stress conditions [88]. It is postulated that bacterial cells may be continuously secreting quorum sensing signals on the course of their growth in a particular environment. Thus, as the
population density increases, the concentration of these signals will also increase, reaching the threshold value. As a consequence, the cell surface or intracellular receptors will activate the cell signalling pathway which encodes a particular gene expression in bacteria [89]. Apart from biofilm formation, various other physiological group behaviours have also been reported in many bacteria that are regulated through quorum sensing system. The spore formation in *Bacillus subtilis* [90], expression of virulence factor in *Staphylococcus aureus* [91], bioluminescence in *Vibrio fischeri* [92], production of antibiotics in *Lactococcus lactis* [91], and DNA competence development in *Streptococcus pneumoniae*, *Streptococcus gordonii*, and *S. mutans* [93]. Like many other Gram-positive and Gram-negative bacteria, biofilm formation is regulated by QS circuit in *S. mutans*. As stated earlier, QS pathway of *S. mutans*, the competence-stimulating peptide (CSP) acts as a signalling molecule and is cleaved from the precursor peptide ComC and matured peptide is exported out of the bacterial cell by ComA with the help of an accessory protein ComB. As a result, the accumulated CSP binds ComD (cell surface receptor) and consequently phosphorlylates ComE by its histidine kinase activity and thus, induces transcription of genes such as *comX*, *comR*, and *comS*, resulting in DNA competence, recombination, biofilm formation and bacteriocin production [12, 94]. Remarkably, the involvement of ComA is crucial for the initiation of the QS system of *S. mutans*. As described in the previous chapter, among all the three domains possessed by ComA, peptidase domain is believed to be involved in cleaving their cognate propeptides after the consensus Gly-Gly motif. The presence of a cysteine residue is critical in the sequence for the proteolytic activity of this family of ABC transporters [65]. Recently, the crystal structure of peptidase (PEP) domain involved in the QS pathway of Streptococcus was reported by Ishii *et al.*, [24] and its proteolytic activity substrate recognition mechanism was reported by Kotake *et
In the present study, we propose a computational approach to target and screen ComA against small compound library to quench the QS pathway in *S. mutans*. Since, homologues of the *com* genes are found in other *Streptococcus* sp. and are found to be involved in the biofilm formation such as *S. gordonii* or *S. pneumonia* and bacteriocin associated ABC transporters has solely been reported in prokaryotes till now, so ComA acts as potential broad spectrum therapeutic agent. Moreover, quenching the QS pathway will not lead to any survival stress on the bacteria, thus, the occurrence of resistance against the drugs targeting QS pathway may be a rare event.

### 3.2 Materials and methods

#### 3.2.1 Ligand and protein preparation

The energy minimized 3D coordinates were generated for all ligands and ligand file was prepared for docking using Schrödinger LigPrep software [95]. An independent ligand dataset library Zinc database was used for this study [96]. Since none of the inhibitors have been reported for PEP-ComA, thus, the comparison studies were not carried out with known inhibitors. The X-ray crystal structure of the PEP domain of ComA (PDB ID: 3K8U) was retrieved from PDB database and processed for addition of polar hydrogen and Kollmann charges using Protein Preparation Wizard (PrepWizard) in Maestro (Schrödinger Suite) [97]. The prepared protein was treated to be completely rigid for all docking procedures to minimize the excessive computation cost. Since, our study is the first to exploit the PEP domain for computational analysis, so active sites were not known. Hence, a grid box encompassing the complete macromolecule was constructed and used for all docking runs in this study [98]. The hydrogen bonds were optimized and protein minimization was carried out using the standard protein preparation protocol. The possible options available for protein
minimization included hydrogen only or all-atom with a criterion based on the root-
mean-square deviation (RMSD) of the heavy atoms relative to their initial location.
Water molecules were retained through the H-bond optimization and minimization
stages, as without water molecules, the protein could collapse or the H-bond networks
required for ligand binding would be disrupted. Although, prior to the docking, all the
water molecules (>5Å) were removed. The resultant protein structure was used for
further docking studies with the prepared ligands.

3.2.2 Molecular docking studies

All docking calculations between macromolecule and ligand datasets were performed
with Glide (Grid based LIdgand Docking with Energetics) program in Schrödinger suite
[99]. All the datasets were run in three consecutive steps. Precisely, HTVS mode (High-
Throughput Virtual Screening), SP mode (Standard Precision), followed by XP mode
(eXtra Precision mode) using the default settings. All the prepared ligands were docked
with the target protein using the induced fit docking (IFD) protocol [100]. The extensive
sampling was carried out by SP mode which defines the interaction sites of ligand with
the protein molecule. Resultant G-scores were considered as the ranking criteria for the
selection of best docked ligands to the target protein.

3.2.3 De novo drug designing and lead optimization

The selected ligands were screened based on their respective glide scores, interaction
pattern at the site of interest in the protein and the functional group of the compounds.
The selected ligands were further run for clustering using Schrödinger Canvas module
[101]. The 3D- Pharmacophore and binary fingerprinting studies optimize the ligands
by ruling out the ligands with similar structures and their interaction with the protein.
Clustering of ligands was followed by optimization of lead molecules with Schrödinger
QikProp for toxicity analysis of the selected ligands [102]. The parameters in ADMET and TOPKAT (TOxicity Prediction by Komputer Assisted Technology) was used for the final validation of the selected compounds and deemed to be the potential leads.

3.3 Results

3.3.1 Docking studies, lead optimization and toxicity analysis:

In the present study, docking analysis revealed several hits that favourably occupied the active site of ComA. Each ligand was further evaluated and the compounds were modified to generate the next generation of molecules with improved binding affinity and less toxicity with the active site. For better activity of ComA inhibitors, ligands should optimally interact with at least one of the four crucial amino acid residues Q11, C17, H96, and D112. Lipinski’s rule of five predicts a high probability of success or failure of drug likeness for various selected molecules. Eighty one drug like ligands were subsequently guided by Lipinski’s rule of three and five, pharmacokinetics (ADME) and pharmacodynamics. Rule of five states that the drug likeness for molecules to comply with three or more of the following rules: Molecular mass less than 500 Dalton, high lipophilicity, less than five hydrogen bond donors, and less than 10 hydrogen bond acceptors.

Selected compounds were further analysed using the QikProp module in Schrodinger for toxicity analysis (ADMET). The properties considered for toxicity prediction were QPlogHERG, QPlogBB (Blood-brain barrier), QPlogKhsa (Human serum albumin) and human oral absorption. Among the selected 81 ligands, 7 ligands satisfied the Lipinski’s rule of five, and the parameters considered under ADMET properties as given in Table 3.1. Further, ligand number 5 (ComAI, ComA Inhibitor) was selected based on the ease of synthesis with the available resources.
Table 3.1: Final selected ligands after TOPKAT and ADMET properties as ComA inhibitors

<table>
<thead>
<tr>
<th>Ligand No.</th>
<th>Structure</th>
<th>Glide Score</th>
<th>Molecular weight</th>
<th>Binding site and other characteristics</th>
<th>clogP</th>
<th>H-Donors</th>
<th>H-Acceptors</th>
<th>Rotatable bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Structure 1" /></td>
<td>-3.915</td>
<td>336.84</td>
<td>Q11, R93, L94</td>
<td>2.780</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2" alt="Structure 2" /></td>
<td>-3.128</td>
<td>335.32</td>
<td>Q11</td>
<td>2.510</td>
<td>2</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3" alt="Structure 3" /></td>
<td>-3.433</td>
<td>307.34</td>
<td>H96, Q95, R93</td>
<td>2.220</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4" alt="Structure 4" /></td>
<td>-3.532</td>
<td>318.25</td>
<td>H96, R93, L94</td>
<td>2.580</td>
<td>1</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>
The selected ComAI was further used to generate derivatives with improved binding affinity by substituting various functional groups and were termed as ComAI\(^1\), ComAI\(^2\), ComAI\(^3\), and ComAI\(^4\). All the derivatives showed better binding affinity to the active site and improved Glide score as compared to the selected ComAI and no toxicity pattern was observed for the derivatives of ComAI as given in Table 3.2. The ComAI as well as the derived ligands of ComAI interacts and inhibits ComA protein at Q11 through hydrogen bonding along with various other interactions in protein cleft as
given in Figure 3.1 (A and B) and Figure 3.2 (A to D). Apart from the selected derivatives, one derivative (ComAI') with no interaction to the active site was also synthesized to compare and validate the computational data with *in vitro* data. Evidently, computational studies using docking tools exhibited excellent outcomes which show that the selected ligands have potential binding affinity to ComA.

Table 3.2: Schematic Representation of ADMET and TOPKAT properties of ComAI and its derivatives.

<table>
<thead>
<tr>
<th>Name</th>
<th>Structure</th>
<th>Glide Score/ Interaction Site</th>
<th>Toxicity prediction (ADMET and TOPKAT)</th>
<th>Human Oral Absorption</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rule of 3 and rule of 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QPlogHE</td>
<td>QPlogBB</td>
</tr>
<tr>
<td>ComAI</td>
<td><img src="image1" alt="Structure" /></td>
<td>-3.088 Q11</td>
<td><img src="image2" alt="Green" /> <img src="image3" alt="Green" /> <img src="image4" alt="Green" /> <img src="image5" alt="Green" /></td>
<td>100%</td>
</tr>
<tr>
<td>Derivatives of ComAI</td>
<td></td>
<td></td>
<td><img src="image2" alt="Green" /> <img src="image3" alt="Green" /> <img src="image4" alt="Green" /> <img src="image5" alt="Green" /></td>
<td>100%</td>
</tr>
<tr>
<td>ComAI¹</td>
<td><img src="image6" alt="Structure" /></td>
<td>-3.815 Q11</td>
<td><img src="image2" alt="Green" /> <img src="image3" alt="Green" /> <img src="image4" alt="Green" /> <img src="image5" alt="Green" /></td>
<td>100%</td>
</tr>
<tr>
<td>ComAI²</td>
<td><img src="image7" alt="Structure" /></td>
<td>-3.334 Q11, Q95</td>
<td><img src="image2" alt="Green" /> <img src="image3" alt="Green" /> <img src="image4" alt="Green" /> <img src="image5" alt="Green" /></td>
<td>100%</td>
</tr>
<tr>
<td>ComAI³</td>
<td><img src="image8" alt="Structure" /></td>
<td>-3.214 Q11, Q47, L94</td>
<td><img src="image2" alt="Green" /> <img src="image3" alt="Green" /> <img src="image4" alt="Green" /> <img src="image5" alt="Green" /></td>
<td>86.7%</td>
</tr>
</tbody>
</table>
Table 3.2 (Continued)

<table>
<thead>
<tr>
<th>ComAI&lt;sup&gt;4&lt;/sup&gt;</th>
<th><img src="image" alt="Chemical Structure" /></th>
<th>-3.342</th>
<th>Q11,Q95</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ComAI'</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>-0.485</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

- Toxicty free behaviour

Figure 3.1 (A) Binding of ComAI with PEP domain of ComA through Hydrogen bonding (yellow dotted line). (B) Schematic 2D representation of Ligand interaction pattern of ComAI (ComA Inhibitor) with PEP domain of ComA.
3.4 Discussion

Computer aided drug designing of the novel therapeutic ligands is an additional advantage in the designing of various drug leads which can be further experimentally validated to combat the ever increasing threat of Multidrug resistance in microbes. In the present study, we aimed at finding out novel therapeutic drug candidates for *S. mutans*, as biofilm inhibitors by using various bioinformatics modules available in Schrödinger suite. The fragment-based technique is based on local optimization of various sites in the protein of interest and provides a new insight in identifying new chemotypes and chemical scaffolds. The ComA protein was chosen from previously
published data of mutational studies, which specified that comA mutant strain was unable to process the CSP, an important factor in the process of biofilm development in S. mutans. Previously published data showed the active sites which are crucial for processing and maturation of CSP molecules are Q11, C17, H96, and D112. Since, all the four amino acids are arranged in such a way that it results in the formation of a cleft which could further favour to process the molecule to a mature peptide. Thus, blocking any one of the functional residues (Q11, C17, H96, and D112) will abolish the catalytic activity of the cleft to further process CSP as a mature peptide to signalize the QS pathway. During the initial evaluating steps in ligand docking process, the ligands that can bind and block the active site of the ComA protein were considered for further analysis. Multiple ligand conformations and orientations were generated, and the most appropriate 81 ligands that interact with the active site were selected.

In light of our present in silico studies, we have found that disubstituted ureas have better binding affinity to ComA, a bacteriocin associated ABC transporter. Moreover, all the derived compounds have been shown to bind at the active site of PEP domain of ComA. Thus, blocking this active site will render the cleaving property of PEP domain inactivated.

3.5 Conclusion

Various researchers worldwide have exploited mutational studies and showed that the mutation in one or more components of QS cascade can lead to downregulation of virulence genes [8]. Oral pathogenic streptococci such as S. mutans, are not only cariogenic but also cause several life-threatening systemic diseases such as infective endocarditis by forming a biofilm on heart valves. The CSP mediated QS system has been well established and reported to regulate the formation of biofilm as well as wide
variety of virulent expressions in *S. mutans*. The mutational studies by Ishii and group have shown that the mutations at the active site can lead to the impaired catalytic action of ComA and thus leading to the unprocessed CSP molecule which further hinders the bacterial communication. Considering this lead, we aimed at designing novel drug targets that can bind and block the catalytic activity of ComA and thus leading to impaired activity in processing of CSP into a mature signalling peptide. The use of computational tools in designing novel compounds serves as an initial step for drug discovery. With an existing structure of ComA, it appears appropriate to employ tools towards finding a lead among thousands of compounds for further studies. Based on above observations, the ComA specific inhibitor was designed to attenuate and quench the QS circuit in *S. mutans*. The *de novo* design and molecular docking were applied to find out the potential novel inhibitors and further studied for its binding to ComA protein. The ligand ComAI and its derivatives were observed as the suitable inhibitor candidates, and they can be taken forward as potential drug lead for further *in vitro* and *in vivo* studies. The ligand thus developed is likely to inhibit Streptococcal infections since ComA is highly conserved across the species of the same family. In contrast to prevailing strategy where the antimicrobial agents exerts a selection pressure on the pathogenic microbes and increase the adaptability to develop resistance, the present study reveals a rational design of selective inhibitor of ComA, as it is hypothesized to be an antibiofilm agent but not an antimicrobial agent. Further studies involve the synthesis and revelation of activity of small molecule inhibitors by validation through *in vitro* and *in vivo* studies of the selected ligand so that it can be taken forward for the future drug development process to show broad-spectrum activity against peptidase domains of the bacteriocin associated ATP-binding cassette transporters.