Molecular-docking study of capsular regulatory protein in *Streptococcus pneumoniae* portends the novel approach to its treatment

Simrika Thapa1,2
Abdullah Zubaer1,2

1Swapnojaatra Bioresearch Laboratory, DataSoft Systems, Dhaka, Bangladesh, 2Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet, Bangladesh

Abstract: *Streptococcus pneumoniae* is the major pathogen that causes pneumonia in the pediatric population in Bangladesh and other developing countries. The capsule of this pathogen is the main virulent factor, the synthesis being governed by the capsular polysaccharide gene cluster that encodes for regulatory proteins such as CpsB. Studies have suggested CpsB to be a potential drug target in case of a high rate of vaccine failure as vaccines are intended to target only a few serotypes among 90 available ones. Thus the protein-targeted drug against CpsB is a good alternative to protect against pneumonia. In this study, homology modeling and validation of this protein was performed with certain bioinformatic tools which facilitate the execution of molecular docking against the ligands from the ZINC database. The two potential ligands bind substantially in the active site of this regulatory protein with considerable binding energy in the docking study, which means they could be considered good inhibitors. This study aims to understand the functional aspects of this and also to examine the development of a novel drug against pneumonia.

Keywords: CpsB, drug target, homology modeling, docking, *S. pneumoniae*

Introduction

Pneumonia remains a leading cause of mortality in the pediatric population in Bangladesh.1 The World Health Organization has estimated that it caused 1.4 million childhood deaths in the year 2008.2 Published comparisons of incidence between developing and developed countries indicate that 90%–95% of clinical pneumonia occurs in developing countries.1 “Pneumonia” is defined as the presence of cough or difficulty breathing, along with other clinical symptoms and signs. The pathogens that cause pneumonia are *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, and Gram negative bacteria such as *Escherichia coli* and *Klebsiella* spp.3 Of these, *S. pneumoniae* serotype 6B is the most prevalent in Bangladesh.3 It is a Gram positive bacteria, which frequently colonizes the nasopharynx, and invasive infection can develop in a variety of body compartments including blood, lungs, cerebrospinal fluid, and middle ear.4 Though availability of different drugs and vaccines, the increasing incidence of multiple-antimicrobial resistance and vaccine failure among *S. pneumoniae* isolates is becoming a problem throughout the world.5 Thus, the development of novel drugs is important.

The capsule has long been recognized as the major virulence factor of *S. pneumoniae*. Encapsulated strains were found to be at least 0.1 million times more virulent than strains lacking the capsule.6 All fresh isolates from patients with pneumococcal infection are encapsulated, and spontaneous nonencapsulated derivatives of such strains are
almost completely avirulent. Capsular polysaccharide (CPS) biosynthesis has been shown to be dependent upon a regulatory system with the involvement of capsular regulatory proteins, CpsB, CpsC, and CpsD. CpsB is the regulatory key, as it is required to dephosphorylate CpsD; in cpsB-deletion mutants, the proportion of CpsD that is phosphorylated increases dramatically, and there is a significant decrease in the amount of CPS that is produced. Thus, in this study, the structure-based virtual screening method is adopted to identify therapeutic inhibitors of CpsB as drug target in support of the need for protein-targeted drug discovery in the case of *S. pneumoniae*.

**Methods**

The modeling of the three-dimensional structure of the CpsB protein of 6B pneumococcal serotype as potential drug target of *S. pneumoniae* for protein-targeted drug discovery was performed using the homology modeling program MODELLER (v9.9; Andrej Sali, San Francisco, CA). For template selection, the amino acid sequence of CpsB of 6B with accession number YP_003878593 was subjected to the BLASTp program against a nonredundant structure database, Protein Data Bank (PDB). The one with best sequence similarity was taken as the template, and aligned with a query using the ClustalW2 program. Then the alignment file of amino acid sequences of CpsB of 6B and the template PDB file was submitted to MODELLER to generate the required model. The quality and validation of the model obtained was performed using the Strutural Analysis and Verification Server (SAVES), which provides PROCHECK (v3.5.4), Verify3D, and ERRAT (v2.0) software. Moreover, the model was rechecked in SuperPose. The overall stereochemical property of the protein was assessed by Ramachandran plot analysis that was provided by PROCHECK.

For virtual screening of ligands, the ZINC database was used. The ZINC database was searched to retrieve the ligands, which were then prepared and energy minimized using AutoDock 419 (Scripps Research Institute, La Jolla, CA) and virtually screened. The active sites were revealed on the basis of previous studies and using the Computed Atlas of Surface Topography of proteins (CASTp) database.20,21 The modeled CpsB of 6B serotype was further subjected to AutoDock 4 for macromolecule preparation of rigid and flexible residues. Using the software, polar hydrogen atoms were added to the modeled protein and its nonpolar hydrogen atoms were merged. All ligand bonds were set to be rotational. The grid box, with the dimension 20 × 20 × 20 points, was used around the entire active site to accommodate ligands to move freely. The ligands and the prepared molecule were fed to AutoDock 4 to perform the docking studies, which uses a Lamarckian genetic algorithm for about 50 iterations by placing the grid arbitrarily at the active site of the structure. The ligand-bound complexes were further analyzed for the binding affinity.

**Results and discussion**

**Protein structure and active sites**

Capsule, which is synthesized by the CPS gene cluster, is the major factor for virulence in cases of *S. pneumoniae*. The regulation pathway of the amount of CPS synthesis, therefore, can be key for drug targeting, which involves the posttranslational tyrosine phospho-regulatory system, the essential components of which are the regulatory genes *cpsA–D* in the CPS cluster. Among these regulatory proteins, the CpsB protein was considered a potential drug target as this protein is conserved at amino-acid level among serotypes (shown in Figure 1) and is also important for capsule synthesis. The serotype 6B was chosen for structural analysis as it is a common invasive serotype in Asia. The active sites of CpsB of serotype 4 are Arg139 and Arg209, as suggested by CASTp result. Although the residues of CpsB of 6B are similar to that of serotype 4, the binding pocket analysis by CASTp excluded Arg139 and identified only Arg209 as the substrate binder and catalytic key player for CpsB of 6B, as shown in Figure 2. This difference led to the modeling of the CpsB of serotype 6B.

The model of CpsB protein of 6B serotype has been built to identify its inhibitors as a suitable drug for pneumonia treatment. The built model was validated for its reliability and usability by Ramachandran plot in PROCHECK. Moreover, certain bioinformatic tools were utilized to strengthen the confirmation of the validity of the model. 2WJD20 representing the crystal structure of CpsB of *S. pneumoniae* serotype 4, was found as a suitable template for generating a model with the similarity e-value of 1e-135 suggested by BLASTp with the template sharing the highest percentage of amino-acid residue identity with the query sequence (94%) (Figure 1). The sequence identity between target protein and template structure is commonly seen as a first indicator for the expected accuracy of a model26 and comparative models sharing >40% of sequence identity with the template are considered good quality models.27 This suggests that the generated model, with a 94% sequence homology between template and model, is a highly acceptable model. The three-dimensional structure of the metabolic protein CpsB of 6B strain was constructed with MODELLER, as indicated in Figure 3A. The assess-
Molecular-docking study of CpsB

The sequence alignment of template and target, which are CpsB protein of *Streptococcus pneumoniae* 4 serotype and 6B serotype, respectively, using the CLUSTALW program and showing homology of 94%. This is also revealed by BLASTp and the Swiss Model Repository, suggesting that CpsB of serotype 4 is suitable as a template for the modeling of CpsB of 6B.

**Notes:** *"* means that the residues in that column are identical in all sequences in the alignment; "*:"* means that conserved substitutions are observed.

The difference in active sites between the existing template model and generated target model predicted by the Computed Atlas of Surface Topography of proteins showing the different significant binding pockets (one in blue, the other in green). Below are the amino acid residues involved in configuration of respective binding pockets.

(A) The binding pockets of CpsB of serotype 4 involving Arg139 and Arg209 as active sites, which agrees with previous study. (B) The binding pockets of CpsB of 6B serotype involving only Arg209 as active sites to form pocket showing conformational difference between the CpsB of the two serotypes.
The homology modeling of CpsB of 6B done on the basis of the template 2WJD. (A) The resulting model generated from MODELLER on the basis of the template 2WJD. (B) The overlapped comparison of model with template is verified using the SuperPose program. The main chain is shown in green and template in red. (C) In the Ramachandran plot analysis, the residues were classified according to their regions in the quadrangle. The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. The Ramachandran plot of backbone conformational angles for each residue of the refined structure revealed that 91.4%, 7.7%, and 1.2% of the residues were located in the most favorable, additionally allowed, and disallowed regions, respectively. (D) The overall quality factor results in ERRAT show model quality of 92.34%. (E) The Verify3D plot predicted 89.34% of residues had good 3D-1D score. (F) The PROVE estimated RMS Z-score was $\approx 1$, which indicates the model is very good. 

**Figure 3** The homology modeling of CpsB of 6B done on the basis of the template 2WJD. (A) The resulting model generated from MODELLER on the basis of the template 2WJD. (B) The overlapped comparison of model with template is verified using the SuperPose program. The main chain is shown in green and template in red. (C) In the Ramachandran plot analysis, the residues were classified according to their regions in the quadrangle. The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. The Ramachandran plot of backbone conformational angles for each residue of the refined structure revealed that 91.4%, 7.7%, and 1.2% of the residues were located in the most favorable, additionally allowed, and disallowed regions, respectively. (D) The overall quality factor results in ERRAT show model quality of 92.34%. (E) The Verify3D plot predicted 89.34% of residues had good 3D-1D score. (F) The PROVE estimated RMS Z-score was $\approx 1$, which indicates the model is very good. 

**Abbreviation:** RMS, root mean square.

(a score in the red area indicates a poorer-quality model at this position). The protein volume evaluation (PROVE) analysis revealed root mean square (RMS) Z-scores =1 suggesting high model quality. For the generated model it was 1.40 (Figure 3F). The predicted structures conformed well to the stereochemistry indicating a reasonably good-quality model. The assessment of the predicted model is shown in Figure 3.

**Virtual screening and docking of ligands**

Protein–ligand interaction plays a significant role in structure-based drug designing. The representative molecules were taken from ZINC database and analyzed for their interactions with the target-modeled protein. The binding analysis indicates that these molecules can bind to the drug target efficiently and could be potential drugs for *S. pneumoniae*, suggesting that protein-based drug delivery could potentially resolve
Molecular-docking study ofcpsB

Table 1 The chemical compounds showing a good binding energy score in AutoDock 4.19 The first two potential molecules are similar and only differ in their structural conformation. The compound name, formula, and structure of each particular compound are shown according to their ZINC database18 identification number (ID).

<table>
<thead>
<tr>
<th>ZINC ID</th>
<th>Compound name</th>
<th>Molecular formula</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZINC40393531</td>
<td>3-(4,6-Diamino-1,3,5-triazin-2-ylamino)-propyl 9-fluoro-3-methyl-10-(4-methylpiperazin-1-yl)-7-oxo-3,7-dihydro-2H-[1,4] oxazino[2,3,4-ij] quinoline-6-carboxylate</td>
<td>C24H30FN9O4</td>
<td><img src="image1" alt="" /></td>
</tr>
<tr>
<td>ZINC40393532</td>
<td>3-(4,6-Diamino-1,3,5-triazin-2-ylamino)-propyl 9-fluoro-3-methyl-10-(4-methylpiperazin-1-yl)-7-oxo-3,7-dihydro-2H-[1,4] oxazino[2,3,4-ij] quinoline-6-carboxylate</td>
<td>C24H30FN9O4</td>
<td><img src="image2" alt="" /></td>
</tr>
</tbody>
</table>

Figure 4 Docking analysis of CpsB protein of 6B serotype of Streptococcus pneumoniae with suitable ligands selected by virtual screening from the ZINC database18 using AutoDock 419 (protein shown with molecular surface at left and with atoms only at right). (A) The best ligand (ZINC ID ZINC40393531) docked in the active site of the protein; with binding affinity of –8.37 kcal/mol. (B) The second-best ligand (ZINC ID ZINC40393532) docked in the active site of protein; with binding affinity of –6.25 kcal/mol.

Abbreviation: ID, identification number.
the problems of vaccine failure and antibiotic resistance commonly encountered with this organism. Overall, the best confirmation shows that the free energies of binding ($\Delta G_{\text{bind}}$ kcal/mol) for the two ligands were $-8.37$ kcal/mol and $-6.25$ kcal/mol. The selected ligands, according to their binding energy, were ZINC40393531 and ZINC40393532 respectively. The structures and description of these are shown in Table 1. The molecule chosen and the docked ligand-molecule are shown in Figure 4.

Free energy of binding ($\Delta G_{\text{bind}}$) was calculated from the sum of the intermolecular and the torsional free energies in AutoDock 4. The negative and low value of $\Delta G_{\text{bind}}$ indicated strong favorable bonds between protein and the ligand, indicating that the ligand was in its most favorable conformation. Molecular docking was applied to explore the binding mechanism and studies on the novel ligand against the CpsB protein showed that the free binding energy for the inhibitor was small, indicating that the ligand binds favorably to the binding site. The ligands were observed as the suitable inhibitor candidate, which may be considered as potential ligands for treatment of pneumonia, as this inhibition of CpsB would, in turn, inhibit the capsule synthesis of S. pneumoniae so that the organism would lose its disease-causing and novel serotype-evolving capabilities. The ligands thus developed are likely to inhibit infections by maximum serotypes since CpsB shares high sequence similarity among serotypes.

**Conclusion**

This study has focused on in silico drug design for S. pneumoniae based on its essential protein, CpsB, identified as a potential drug target. Further toxicological tests need to be conducted to validate the feasibility of the approaches used in this study in predicting endpoints and prioritizing chemicals for subsequent in vitro and in vivo screening.

Phenomenon of vaccine failure is a major problem in the case of pneumonia because different serotypes erupt unexpectedly. The current study has shown a novel solution for this problem by producing new drug against regulatory proteins like CpsB, inhibition of which will degrade the bacterial capsule and thus, the bacteria will fail to cause disease and frequent emergence.

**Acknowledgment**

The authors are indebted to Shakhnin Islam Mondal and Mohammad Lashkar for their valuable suggestions and supervision during this study. They would also like to thank Tabib Ahmed and Asraful Islam for their assistance in manuscript preparation.

**Disclosure**

The authors declare no conflicts of interest in this work.

**References**


