ORIGINAL RESEARCH

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

Simrika Thapa[1,2]
Abdullah Zubaer[1,2]

[1]Swapnojaatra Bioresearch Laboratory, DataSoft Systems, Dhaka, Bangladesh, [2]Department 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

Correspondence: Simrika Thapa
Swapnojaatra Bioresearch Laboratory,
DataSoft Systems Bangladesh Limited,
73-d, New Airport Road, Manipuripara,
Dhaka-1215, Bangladesh
Tel +880 2 9110136
Fax +880 2 9114747
Email simrikathapa@gmail.com

131

almost completely avirulent.[7] 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.[8] 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).[9] For template selection, the amino acid sequence of CpsB of 6B with accession number YP_003878593 was subjected to the BLASTp[10] programs against a nonredundant structure database, Protein Data Bank (PDB).[11] The one with best sequence similarity was taken as the template, and aligned with a query using the ClustalW2 program.[12] 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),[13] Verify3D,[14] and ERRAT (v2.0)[15] software. Moreover, the model was rechecked in SuperPose.[16] The overall stereochemical property of the protein was assessed by Ramchandran plot analysis[17] that was provided by PROCHECK.

For virtual screening of ligands, the ZINC database was used.[18] The ZINC database was searched to retrieve the ligands, which were then prepared and energy minimized using AutoDock 4[19] (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 $20 \times 20 \times 20$ 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.[20,22] 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*.[7,23] 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.[8,24] 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.[25] The active sites of CpsB of serotype 4 are Arg139 and Arg209,[20] 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. 2WJD,[20] 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 model[26] and comparative models sharing >40% of sequence identity with the template are considered good quality models.[27,28] 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-

```
CpsB_4    GAMGMIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRRKGMFETPEEKIA  60
CpsB_6B   ------- MIDIHSHIVFDVDDGPKSREESKALLTEAYRQGVRTIVSTSHRRKGMFETPEEKIA  56
                  ********************************* . * . ********************************************

CpsB_4    ENFLQVREIAKEVASDLVIAYGAEIYYTPDVLDKLEKKRIPTLNDSRYALIEFSMNTPYR  120
CpsB_6B   ENFLQVREIAKEVASDLVIAYGAEIYYTPDVLDKLENNRIPTLNNSRYALIEFSMNTPYR  116
          ********************************************** . : ********* . *************************

CpsB_4    DIHSALSKILMLGITPVIAHIERYDALENNEKRVRELIDMGCYTQVNSSHVLKPKLFGER  180
CpsB_6B   DIHSALNKILMLGITPVIAHIERYDVLENNEKRVRELIDMGCYTQINSSHVLKSKLFGEP  176
          ********** : *************************** : *************************** . *********** : ********

CpsB_4    YKFMKKRAQYFLEQDLVHVIASDMHNLDGRPPHMAEAYDLVTQKYGEAKAQELFIDNPRK  240
CpsB_6B   YKFMKKRAQYFLERDLVHIIASDMHNVDGRPPHMAEAYDLVSQKYGEAKAQELFIDNPRK  236
          ********************* : ****** . ************* : *************************** . *************************

CpsB_4    IVMDQLI 247
CpsB_6B   IVMDQLI 243
          ***********
```

Figure 1 The sequence alignment of template and target, which are CpsB protein of *Streptococcus pneumoniae* 4 serotype and 6B serotype, respectively, using the ClustalW2[12] 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.

ment of the predicted model was performed using SuperPose, which showed the maximum superposition of model with the template. The SuperPose-engendered visual output is shown in Figure 3B. For model quality validation, the model was made available to SAVES, with the main chain parameters of: Ramachandran plot quality, peptide-bond planarity, side-chain parameters, main-chain bond length, and overall quality factor (ERRAT).[15] The stereochemical quality of the predicted models and accuracy of the protein model were evaluated after the refinement process using Ramachandran map calculations computed within PROCHECK.[13] The Ramachandran

plot provided by the performed SAVES analysis is shown in Figure 3C. The good-quality model is expected to have >90% in the most favored regions in terms of Ramachandran plot;[17] in the generated model it was found to be 92%, suggesting the model is highly reliable. The overall quality factor, suggested by ERRAT, was 92.34 (Figure 3D) – good high-resolution structures generally produce values around or higher than 90%.[15] In addition, the Verify3D plot predicted 89.34% of residues had 3D-1D score >0.2 (Figure 3E), revealing that the model was of good quality as negligible residues were found in the red area of the graph in Figure 3F
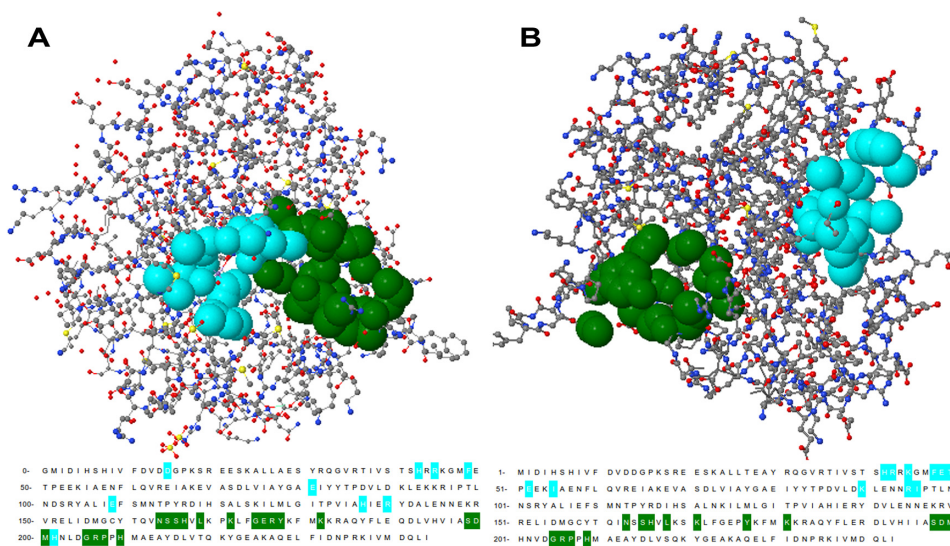

Figure 2 The difference in active sites between the existing template model and generated target model predicted by the Computed Atlas of Surface Topography of proteins[21] showing the different significant binding pockets (one in blue, the other in green). Below are the amino acid residues involved 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.[20] (**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.
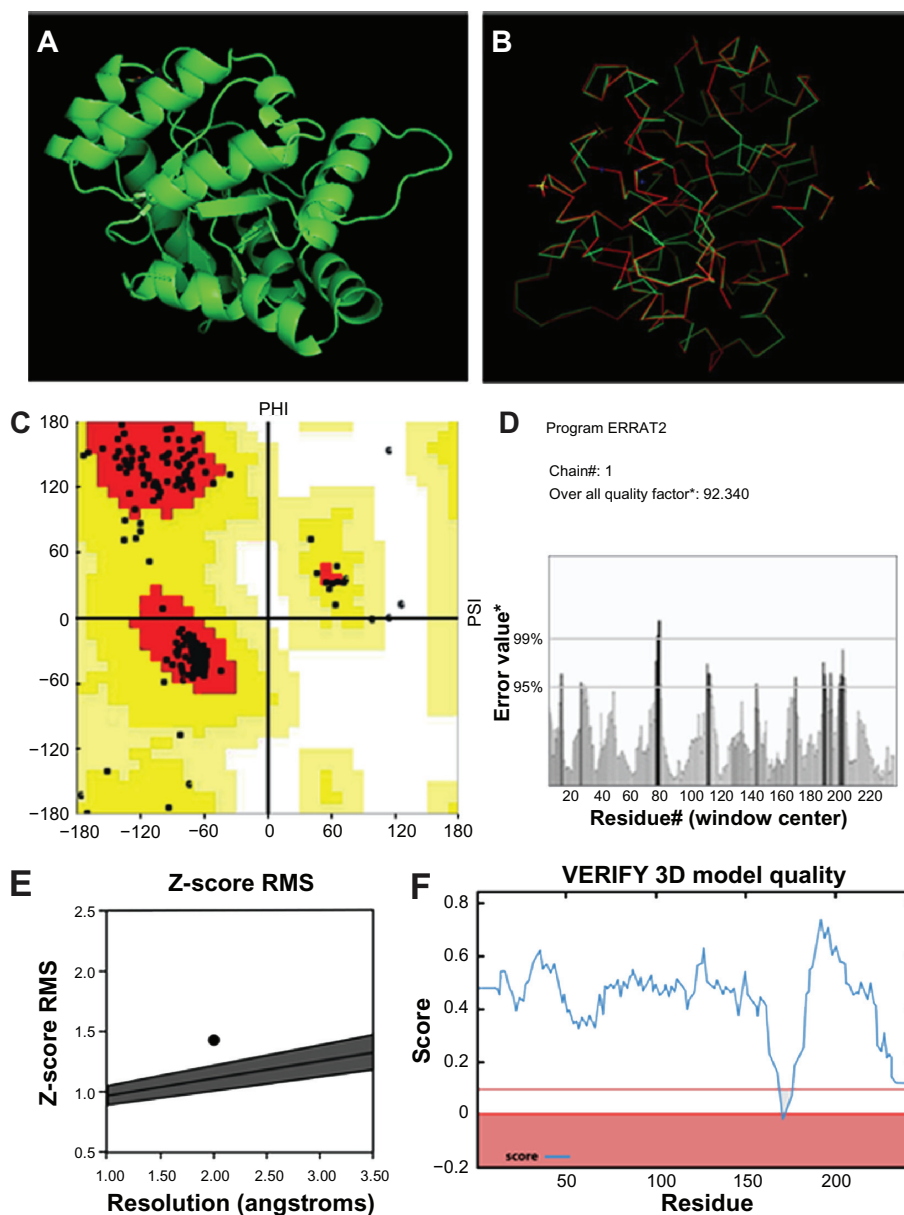
Figure 3 The homology modeling of CpsB of 6B done on the basis of the template 2WJD.[20] (**A**) The resulting model generated from MODELLER on the basis of the template 2WJD.[20] (**B**) The overlapped comparison of model with template is verified using the SuperPose[16] program. The main chain is shown in green and template in red. (**C**) In the Ramachandran plot analysis,[13] 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[15] show model quality of 92.34%. (**E**) The Verify3D[14] plot predicted 89.34% of residues had good 3D-1D score. (**F**) The PROVE estimated RMS Z-score was ≈1, which indicates the model is very good.
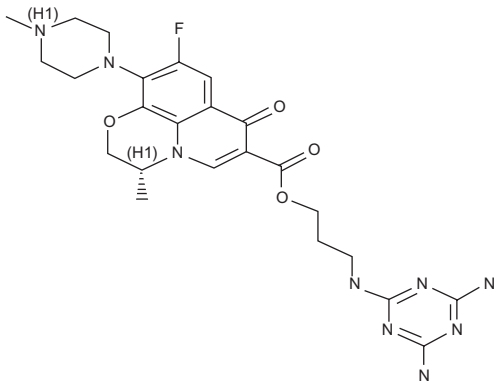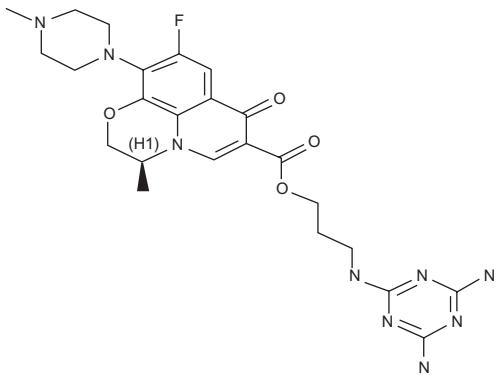**Abbreviation:** RMS, root mean square.

(a score in the red area indicates a poorer-quality model at this position).[14] The protein volume evaluation (PROVE) analysis revealed root mean square (RMS) Z-scores ≈1 suggesting high model quality.[29] 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

Table I 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 database[18] identification number (ID).

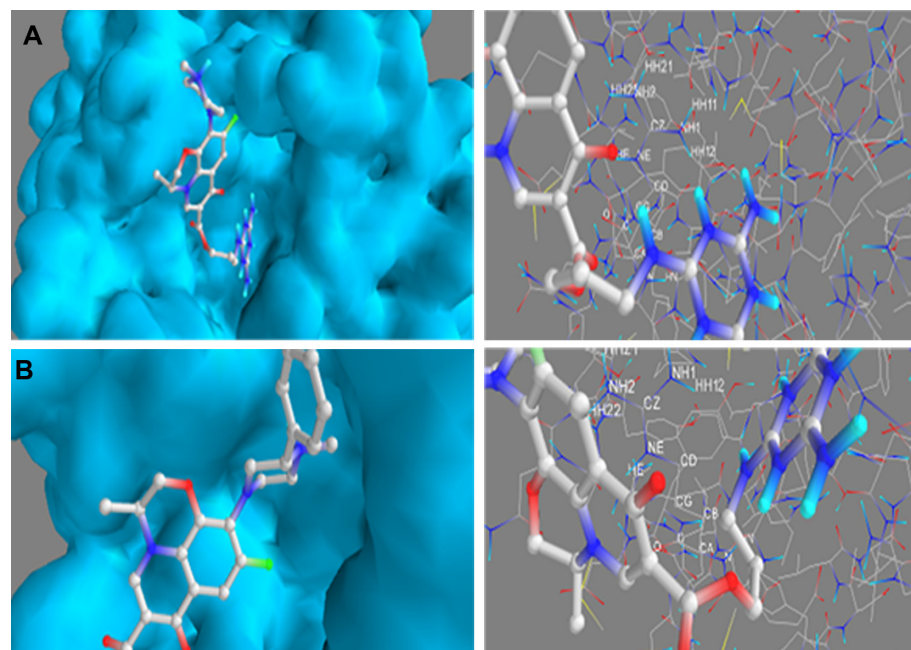| ZINC ID	Compound name	Molecular formula	Structure
ZINC40393531	3-(4,6-Diamino-1,3,5-triazin-2-ylamino)-propyl 9-fluoro-3-methyl-10-(4-methylpiperazin-1-yl)-7-oxo-3,7-dihydro-2H-[1,4]oxazino[2,3,4-ij] quinoline-6-carboxylate	$C_{24}H_{30}FN_9O_4$	
ZINC40393532	3-(4,6-Diamino-1,3,5-triazin-2-ylamino)-propyl 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	$C_{24}H_{30}FN_9O_4$	

Figure 4 Docking analysis of CpsB protein of 6B serotype of *Streptococcus pneumoniae* with suitable ligands selected by virtual screening from the ZINC database[18] using AutoDock 4[19] (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.[30,31] Overall, the best confirmation shows that the free energies of binding (ΔGbind 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 (ΔGbind) was calculated from the sum of the intermolecular and the torsional free energies in AutoDock 4. The negative and low value of ΔGbind indicated strong favorable bonds between protein and the ligand, indicating that the ligand was in its most favorable conformation.[34] 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.[8,33,34] 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.

The 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

Disclosure

References

1. Brooks WA, Breiman FR, Goswami D, et al. Invasive pneumococcal disease burden and implications for vaccine policy in urban Bangladesh. *Am J Trop Med Hyg*. 2007;77(5):795–801.
2. World Health Organization (WHO). Causes of child mortality for the year 2008. Geneva: WHO; 2008. Available from: http://www.who.int/gho/child_health/mortality/mortality_causes_text/en/. Accessed April 04, 2011.
3. Saha SK, Rikitomi N, Biswas D, et al. Serotypes of Streptococcus pneumoniae causing invasive childhood infections in Bangladesh, 1992 to 1995. *J Clin Microbiol*. 1997;35(3):785–787.
4. Moschioni M, Pansegrau W, Barocchi MA. Adhesion determinants of the Streptococcus species. *Microb Biotechnol*. 2010;3(4):370–388.
5. Aspa J, Rajas O, Rodríguez de Castro F, et al. Drug-Resistant pneumococcal pneumonia: clinical relevance and related factors. *Clin Infect Dis*. 2004;38(6):787–798.
6. Watson DA, Musher DM. Interruption of capsule production in Streptococcus pneumoniae serotype 3 by insertion of transposon Tn916. *Infect Immun*. 1990;58(9):3135–3138.
7. Jiang S, Wang L, Reeves PR. Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene clusters. *Infect Immun*. 2001;69(3):1244–1255.
8. Bender MH, Cartee RT, Yother J. Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in Streptococcus pneumoniae. *J Bacteriol*. 2003;185(20):6057–6066.
9. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modelling by MODELLER. *Proteins*. 1995;23(3):318–326.
10. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–3402.
11. Bourne PE, Addess KJ, Bluhm WF, et al. The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res*. 2004;32: D223–D225.
12. Larkin M, Blackshields G, Brown N, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–2948.
13. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK – a program to check the stereochemical quality of protein structures. *J App Cryst*. 1993;26:283–291.
14. Luethy R, Bowie J, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992;356:83–85.
15. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2(9):1511–1519.
16. Maiti R, Van Domselaar GH, Zhang H, Wishart DS. SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res*. 2004;32:W590–W594.
17. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol*. 1963;7:95–99.
18. Irwin JJ, Shoichet BK. ZINC – a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 2005;45(1): 177–182.
19. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDock-Tools4: automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785–2791.
20. Hagelueken G, Huang H, Mainprize IL, Whitfield C, Naismith JH. Crystal structures of Wzb of Escherichia coli and CpsB of Streptococcus pneumoniae, representatives of two families of tyrosine phosphatases that regulate capsule assembly. *J Mol Biol*. 2009;392:678–688. PDB ID: 2WJD.

21. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res*. 2006;34:W116–W118.

22. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*. 2003;32:335–373.

23. Alonsodevelasco E, Verheul AFM, Verhoef J, Snippe H. Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines. *Microbiol Rev*. 1995;59(4):591–603.

24. Morona JK, Morona R, Paton JC. Attachment of capsular polysaccharide to the cell wall of Streptococcus pneumoniae type 2 is required for invasive disease. *Proc Natl Acad Sci*. 2006;103(22):8505–8510.

25. Song JH, Jung SI, Ko KS, Kim NY, Son JS, Chang HH. High prevalence of antimicrobial resistant among clinical Streptococcus pneumoniae isolates in Asia (an ANSORP Study). *Antimicrob Agent Chemother*. 2004;48:2101–2107.

26. Arnold K, Kiefer F, Kopp J, et al. The protein model portal. *J Struct Funct Genomics*. 2009;10(1):1–8.

27. Hillisch A, Pineda LF, Hilgenfeld R. Utility of homology models in the drug discovery process. *Drug Discov Today*. 2004;9(15): 659–669.

28. Peitsch MC. About the use of protein models. *BioinformaticsI*. 2002;18: 934–938.

29. Mulakayala C, Banaganapalli BN, Anuradha CM, Chitta SK. Insights from Streptococcus pneumoniae glucose kinase structural model. *Bioinformation*. 2009;3(7):308–310.

30. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog*. 2007;3(11):e168.

31. Lynch JP 3rd, Zhanel GG. Streptococcus pneumoniae: epidemiology and risk factors, evolution of antimicrobial resistance, and impact of vaccines. *Curr Opin Pulm Med*. 2010;16(3):217–225.

32. Jitendra S, Vinay R. Structure based drug designing of a novel antiflaviviral inhibitor for nonstructural 3 protein. *Bioinformation*. 2011;6(2): 57–60.

33. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis*. 2003;187(9):1424–1432.

34. Spratt BG, Greenwood BM. Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet*. 2000;356(9237): 1210–1211.