Insights from *Streptococcus pneumoniae* glucose kinase structural model

Chaitanya Mulakayala[1], Babajan Nawaz Banaganapalli[1], C. M. Anuradha[2], Suresh Kumar Chitta[1*]

[1]DBT-Bioinforamatics Facility (BIF) and Department of Biochemistry, [2]Department of Biotechnology, Sri Krishnadevarya University College of Engineering and Technology, Sri Krishnadevaraya University, Anantapur-515 003, A.P. India; Suresh Kumar Chitta - E-mail: chitta34c@gmail.com; Phone: 91-8554-255644; Fax: 91-8554-255644; *Corresponding author

Abstract:

Streptococcus pneumonia is the common cause of sepsis and meningitis. Emergence of multiple antibiotic resistant strains in the community-acquired bacterium is catastrophic. Glucose kinase (GLK) is a regulatory enzyme capable of adding phosphate group to glucose in the first step of streptomycin biosynthesis. The activity of glucose kinase was regulated by the Carbon Catabolite Repression (CCR) system. Therefore, it is important to establish the structure-function relation of GLK in *S. pneumoniae.* However, a solved structure for *S. pneumoniae* GLK is not available at the protein data bank (PDB). Therefore, we created a model of GLK from *S. pnemoniae* using the X-ray structure of Glk from *E. faecalis* as template with MODELLER (a comparative modeling program). The model was validated using protein structure checking tools such as PROCHECK, WHAT IF and ProSA for reliability. The active site amino acid Asp114 in the template is retained in *S. pneumoniae* GLK model (Asp115). Solvent accessible surface area (ASA) analysis of the GLK model showed that known key residues playing important role in active site for ligand binding and metal ion binding are buried and hence not accessible to solvent. The information thus discussed provides insight to the molecular understanding of glucose kinase in *S. pneumoniae.*

Keywords: Glucose kinase, active site, model, homology, function

Background:

The continuing battles in the hospitals against drug-resistant bacteria highlight the need for new antibiotics. The world largest causes of mortality and morbidity of pneumonia is caused by *Streptococcus pneumoniae* (*S. pneumonia*). *S. pneumoniae* is the most common cause of fatal community-acquired pneumonia, middle ear infection and meningitis. The disease rate is especially high in young children, the elderly and immuno-compromised individuals with predisposing conditions such as asplenia or acquired immunodeficiency syndrome (AIDS) **[1, 2, 3].** More than 90 different pneumococcal serotypes have been identified by their difference in polysaccharide capsules. However, about 90% of clinical episodes of invasive pneumococcal infections in humans are caused by 23 pneumococcal serotypes **[4].** The mortality in the United Staes alone is about 40,000 per yea which is, higher than that for other bacterial diseases **[5, 6].** The efficacy of available vaccines is limited. The use of antibiotics results in either capsular type shifting or in the rapid appearance and spread of antibiotic resistance determinants **[7, 8, 9].** Thus prevention and treatment of the infection is a top priority for the scientific community **[10].** This requires a better understanding of the *S. pneumoniae* proteins using structural and functional data for target definition and validation. Here, we describe the modeled structure of glucose kinase (GLK) from *S. pneumonia* towards establishing its molecular function. GLK (E.C.2.7.1.2) is an important enzyme in the biosynthesis of streptomycin. Glk catalyzes the conversion of glucose to glucose-6-phosphate and is catabolically repressed by higher concentrations of glucose (carbon catabolite repression (CCR)). However, a 3-dimenisonal structure for *S. pneumonia* GLK is yet not available. Hence, we constructed the model structure for *S. pneumonia* GLK using known structural templates and describe its structural features to understand molecular function.

Methodology:
Tools and systems

The study was performed on an AMD 64 bits dual processor with Linux operating system. Sequence alignment was performed with ClustalX **[11]** and homology modeling was carried out using modeller (comparative modeling software). Energy minimization of the developed model was performed using Swiss PDB Viewer spdbv **[12].** Protein structure checks were conducted using the ADIT validation server **[13]**; WHAT IF web interface **[14]** and ProSA-web **[15]**. Figures were developed using Pymol 1.0 **[16].**
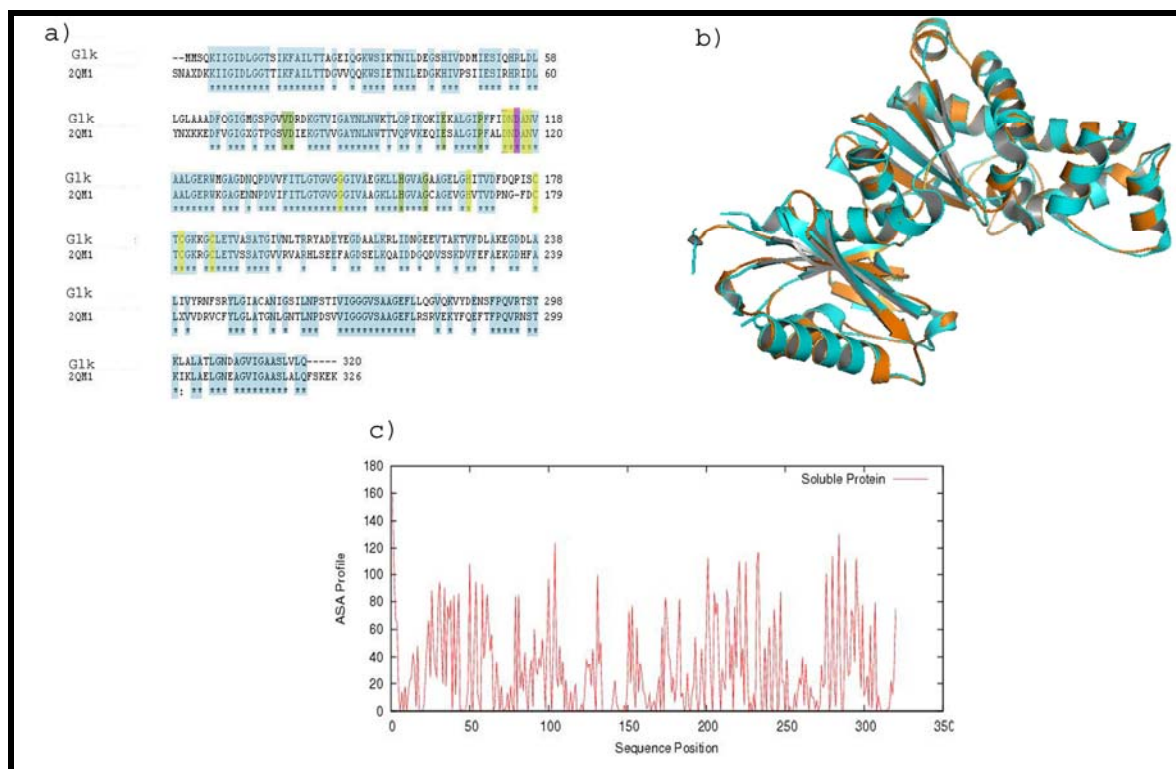
Figure 1: (a) Pairwise alignment of GLK from *S. pneumoniae* and the template PDB ID: 2QM1 from *E. feacalis*. Dash represents insertion and deletion: conserved residues which are not involved in active site, ligand binding and metal binding are shaded in blue. Conserved residues which are involved in active site are shaded in pink, conserved residues involved in ligand binding are shaded in green and conserved residues involved in metal binding are shaded in yellow. (b) Superposition of the predicted model of GLK from *S. pneumoniae* onto to the template PDB ID: 2QM1. (c) ASA analysis of the developed model of GLK from *S. pneumoniae*.

GLK sequence, template selection, sequence alignment and model building:
The protein sequence of *S. pneumoniae* GLK (Accession No: NP_345173) was downloaded from GenBank [17]. Subsequently, a protein-protein BLAST [18] search to PDB was conducted using the query GLK sequence for potential structural templates. The closest structural template (PDB ID: 2QM1) was GLK form *E. faecalis* with a 65% sequence identity to the query sequence. We then used the structure described in 2QM1 as a template for model building. We used ClustalX [11] to generate an alignment of the query and template. The GLK model from *S. pneumoniae* was superposed on to template 2QM1 and the Cα and back bone RMSD values using online server SUPERPOSE [19]. The model was further checked with WHAT IF [14], ProSA-Web [15] and Ramachandran plot at PROCHECK [13]. Accessible surface area prediction using ASAP was performed [20].

Discussion:
Streptococcus pneumoniae infction is the major cause of morbidity and mortality elderly and young children. GLK is the enzyme involved in streptomycin biosynthesis and converts glucose to glucose-6-phosphate. Hence, it is important to establish the structure-function relationship for *Streptococcus pneumoniae* GLK. We developed a structural model using 2QMI (PDB ID) as a template. The sequence similarity is 65% with the template and reliability

of the predicted model thus generated using MODELLER is high. Kinases are Zn^{2+} and Mg^{2+} dependent enzymes. The ions play an important role in the catalytic process of the enzyme by reducing the overall entropy by coordinating to substrates and water molecules. Here, we describe the consequence of model development towards establishing the specific function of *S. pneumoniae* GLK using predicted model.

Sequence homology
The protein sequence (320 residues long) for GLK (E.C.2.7.2.1, 320aa) from *S. pneumonaie* was retrieved from GenBank FASTA format. The basic local alignment search tool (BLAST) A search at PDB for *S. pneumonaie* GLK identified a template structure (PDB code: 2QM1 wih 2.02 Å resolution) *E. faecalis* with 65% identity (Figure 1a). Sequence alignment (Figure 1a) revealed that the active site amino acid residue Asp114 (represented in pink) in the template was conserved in *S. pneumonaie* Glk (Asp115). The ligand binding residues in the template Val76, Asp77, Glu102, Pro108, H155 and Gly159 (represented in green i) were also conserved both in target and template (Val77, Asp78, Glu103, Pro109, His 156 and Gly161. The metal ion interacting conserved residues are Asn113, Ala115, Asn116, Gly145, His166, Cys176, Cys178 and Cys183.

Homology modeling and structure refinement

The homology model of *S. pneumonaie* GLK was built using MODELLER. We then incorporated the Mg^{2+} ion and Zn^{2+} ion from the template structure into the modeled structure. The atomic charges of the amino acid side chains involved in Mg^{2+} ion and Zn^{2+} ion coordination system is similar to those of the template. Structural refinement through energy minimization model was performed using spdbv.

Model validation

The constructed model of Glk from *S. pneumoniae* was examined for validation using different criteria. The RMSD analysis of the developed model was evaluated by means of deviation from its template using SUPERPOSE (Figure 1b). The Cα RMSD and the backbone RMSD deviations for the model and the template crystal structure are 0.36 Å and 0.42 Å respectively. The stereo chemical quality of the predicted model was evaluated using PROCHECK in ADIT. The Ramachandran plot of phi/psi distribution in the model is developed using PROCHECK for checking non-GLY residues at the disallowed regions. Standard bond lengths and bond angles of the model were determined using WHAT IF. The analysis revealed RMS Z-scores for bond lengths and bond angles to be 0.885 and 0.926, respectively. The values are almost equal to 1 suggesting high model quality. ProSA-Web analysis of the model revealed a Z-score value of -9.49 and it is in the range of native conformations of the template.

ASA analysis

Accessible surface area (ASA) analysis of the predicted GLK model showed the active site amino acid Asp115 with zero ASA value (Figure 1c) and is buried inside the catalytic cleft. Some of the ligand binding residues and metal ion binding residues were found to have high ASA values (Pro109, His156, Gly161, Cys179, Cys181, Cys186) and some others were found to have low ASA values (Val77, Asp78, Glu103, Asn114, Ala116, Asn117, Gly146, His168). Residues with high ASA values are on the surface of the cleft and those with low values are buried inside.

Conclusion

The structure of *S. pneumoniae* GLK is important for establishing its molecular function. However, a three dimensional structure is not available as yet at PDB. We developed a homology model for *S. pneumoniae* GLK using MODELLER. The model was further analyzed for residue solvent accessibility in establishing its molecular function. Solvent accessible surface area (ASA) analysis of the GLK model showed that known key residues playing important role in active site for ligand binding and metal ion binding are buried and not accessible to solvent. The analysis highlights the importance of solvent exposed catalytic residues in molecular function.

References

[1] B. M. Gray et al., *J. Infect. Dis.,* 142: 923-933 (1980) [PMID: 7462701]

[2] R. B. Johnston Jr, *Rev. Infect., Dis.* 13: S509-S517 (1991) [PMID: 1862280]

[3] D. M. Musher *Clin. Inect. Dis.,* 14:801-807 (1991) [PMID: 1576274]

[4] S. Neralla & K.C. Meyer, *Drugs Aging.,* 21:851-864 (2004) [PMID: 15493950]

[5] J. E. Alexander et al., *Infect. Immun.,* 62:5683-5688 (1994) [PMID: 7960154]

[6] D. R. Gwatkin, *Lancet* 2: 699-701 (1985) [PMID: 2863681]

[7] R. Novak et al., *Nature.,* 399:590-593 (1999) [PMID: 10376600].

[8] R. Dagan, *Curr. Opin. Otolaryngol. Head Neck Surg.*, 12: 488-494 (2004) [PMID: 15548905]

[9] A.W. McCormick et al., *Nat. Med.,* 9: 424-430 (2003) [PMID: 12627227]

[10] J. Cohen, *Science,* 265: 1371-1373 (1994) [PMID: 8073270]

[11] J.D. Thompson et al., *Nucl. Acids. Res.,* 22: 4673-4680 (1994) [PMID: 7984417]

[12] N. Guex & M.C. Peitsch, *Electrophoresis,* 18:2714-2723 (1997) [PMID: 9504803]

[13] R. A. Laskowski et al., *J. Appl. Cryst.,* 26:283-291, (1993) [PMID: 10089410]

[14] G. Vriend, *J. Mol. Graph.,* 8:52-56 (1990) [PMID: 2268628]

[15] M.J. Sippl, *Proteins,* 17:355-362 (1993) [PMID: 8108378]

[16] http://pymol.source-forge.net

[17] http:// www.ncbi.nlm.nih.gov/

[18] S.F.Altshul et al., J.Mol.Biol., 215:403-410 (1990) [PMID: 2231712]

[19] http://wishart.biology.ualberta.ca/SuperPose/

[20] http://ccb.imb.uq.edu.au/ASAP