In silico Protein Structural Modeling and Active binding site Evaluation of *Streptococcus pneumoniae*

M.Balakrishnan[1], R.C.Srivastava[2] and M.Ramachandran[3]

Structure function relation of glucose kinese in *Streptococcus pneumoniae.* However, a solved structure for *Streptococcus pneumoniae* glucose kinese is not available at the protein data bank. Glucose kinase 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 system. So, we created a model of glucose kinese from *Streptococcus pnemoniae* using the X-ray crystallography structure of glucose kinese enzymes from *Enterobacteria faecalis* as template with Molsoft ICM v3.5 software. The model was validated using protein structure checking tools such as PROCHECK, WHAT IF: for reliability. The active site amino acid "Asp114" in the template is retained in *S. pneumoniae* Glucose kinese model "Asp115". Solvent accessible surface area analysis of the glucose kinese 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 *Streptococcus pneumoniae in* glucose kinase.

Streptococcus pneumoniae is the most common cause of bacterial meningitis in adults and children, and is one of the top two isolates found in ear infection, otitis media [1]. *S. pneumoniae* is normally found in the nasopharynx of 5-10% of healthy adults, and 20-40% of healthy children [21]. pneumonia is more common in the very young and the very old. Emergence of multiple antibiotic resistant strains in the community-acquired bacterium is catastrophic. The disease rate is especially high in young children, the elderly and immuno-compromised individuals with predisposing conditions such as asplenia [1, 2, 3]. Hydrogen peroxide causes damage to host cells can cause apoptosis in neuronal cells during meningitis and has bactericidal

[1]Scientist, Bioinformatics Centre, Central Agricultural Research Institute, Port Blair-744 101, A&N Islands, India, [2]Director, Central Agricultural Research Institute, Port Blair, [3]RA, Central Agricultural Research Institute, Port Blair, A&N Islands.

effects against competing bacteria *Haemophilus influenzae, Neisseria meningitidis* and *Staphylococcus aureus* [21,22]. 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 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 *Streptococcus pneumoniae* proteins using structural and functional data for target definition and validation. Here, we describe the modeled structure of glucose kinese from *Streptococcus pneumonia* towards establishing its molecular function. Glucose kinese - E.C.2.7.1.2., is an important enzyme in the biosynthesis of streptomycin. Glucose kinese catalyzes the conversion of glucose to glucose-6- phosphate and is catabolically repressed by higher concentrations of glucose. However, a 3D structure for *Streptococcus pneumonia* glucose kinese is yet not available. Hence, we constructed the model structure for *Streptococcus pneumonia* Glucose kinese using known structural templates and describe its structural features to understand molecular function.

RESULTS

Proteins Sequence Similarity and Amino acid Compositions

Consequence of model development towards establishing the specific function of *Streptococcus pneumoniae* GLK using predicted model. Compositions of proteins GLK – Molecular weight is 33552.48 Daltons (Protein sequence length 320amino acids) , 2QM1- Molecular weight is 34481.09 Daltons (sequence length 326 amino acids) Ala =10.0 - 8.3%, Cys=1.3 - 1.5%, Asp = 6.9 - 6.1%, Glu = 4.7 - 6.7%, Phe = 3.1 - 4.6%, Gly= 13.4 - 12.9%, His=1.3-1.8%, Ile = 9.1- 7.1%, Lys = 5.3-6.1%, Leu = 9.7-7.4%, Met = 1.9-0.9%, Asn = 4.1-4.5%, Pro=2.2-2.5%, Gln=3.8-2.8%, Arg = 2.8-3.1%, Ser = 4.7-5.5%, Thr=5.9-5.8%, Val=7.2-9.8%, Trp=0.9-0.9% and Tyr= 1.9-1.2% (Fig. 1). *Streptococcus pneumoniae* infection is the major cause of morbidity and mortality elderly and young children. GLK (Glucose kinase) 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 65.615% with the template and reliability of the predicted model thus generated using Molsoft ICM v3.5 software[26] 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.

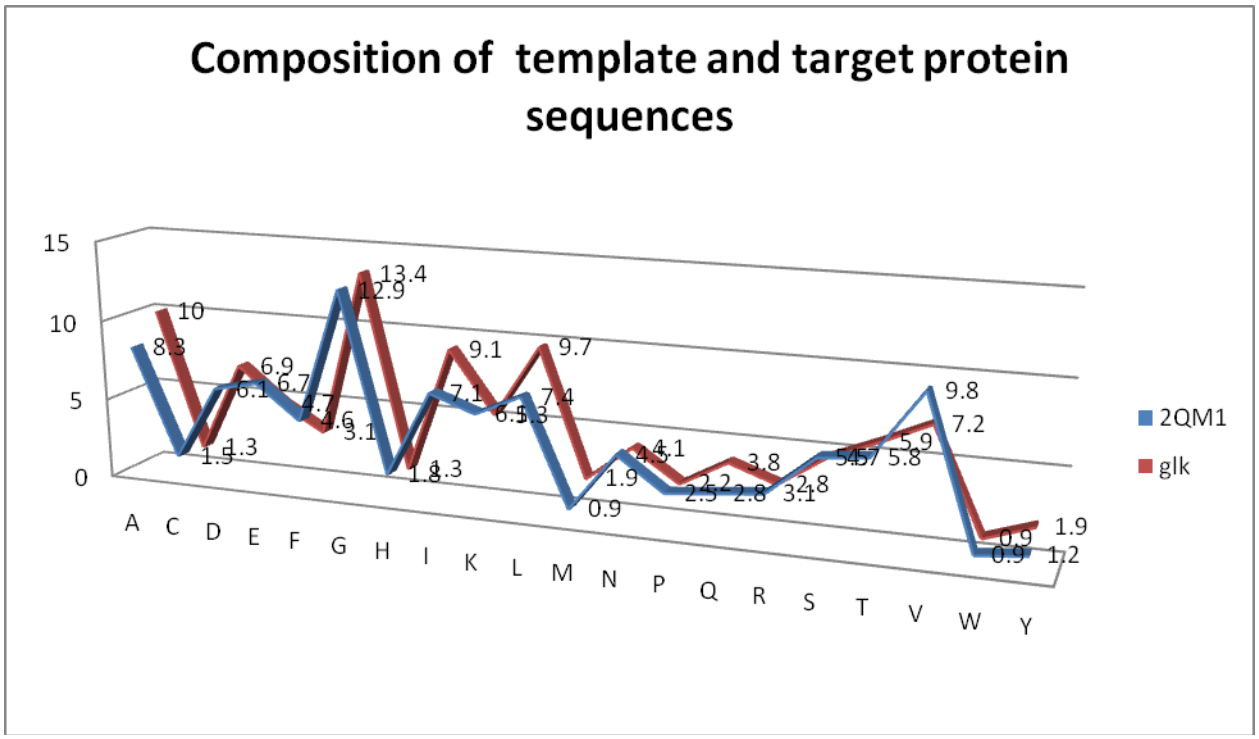


Figure 1 | Composition of *Streptococcus pneumonia* GLK protein and 2QM1 protein of *E. faecalis* amino acid residues.

Structural Modeling of Sequence

The protein sequence (320 residues long) for GLK (E.C.2.7.2.1, 320aa) from *Streptococcus pneumonaie* was retrieved from GenBank FASTA format. The basic local alignment search tool (BLAST) a search at PDB for *Streptococcus pneumonaie* GLK identified a template structure (PDB code: 2QM1 with 2.02 Å resolution) E. faecalis with 65.615% identity (Fig. 7). Sequence alignment revealed that the active site amino acid residue Asp114 in the template was conserved in *Streptococcus pneumonaie* Glk (Asp115). The ligand binding residues in the template Val76, Asp77, Glu102, Pro108, H155 and Gly159 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.

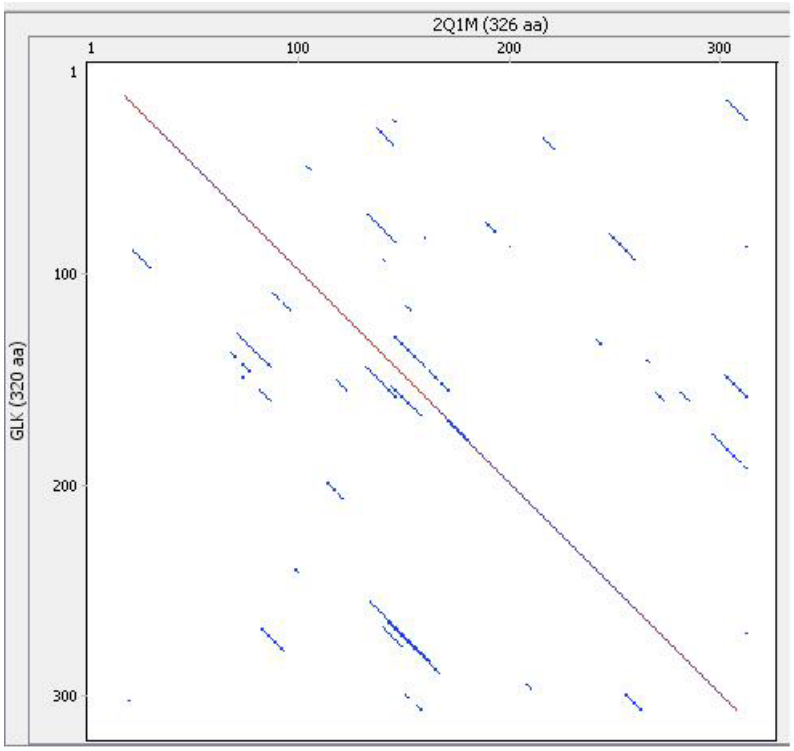


Figure 2 | The sequence similarity GLK with the 2QM1 Protein sequences residues in the Dot plots. Diagonal lines is illustrated 65.615% identities of the template and target sequences

Structural modeled refinement

The homology model of *Streptococcus pneumonaie* GLK was built using Molsoft ICM v3.5software[26]. We then incorporated the Mg^{2+} ion and Zn^{2+} ion from the template structure into the modeled structure (Fig.4). 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 Swissprot Protein database viewer.

Validation

The constructed model of Glucose kinese from *Streptococcus 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 (Fig. 4). The Cα RMSD and the backbone RMSD deviations for the model and the template crystal structure are 0.36 A° and 0.42 A° respectively. The stereo chemical quality of the predicted model was evaluated using

PROCHECK [23] in ADIT. The Ramachandran plot of phi / psi distribution in the model is developed using PROCHECK [23] for checking non-GLY residues at the disallowed regions. Standard bond lengths and bond angles of the model were determined using WHAT IF [24]. 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.

Ramachandran Plot

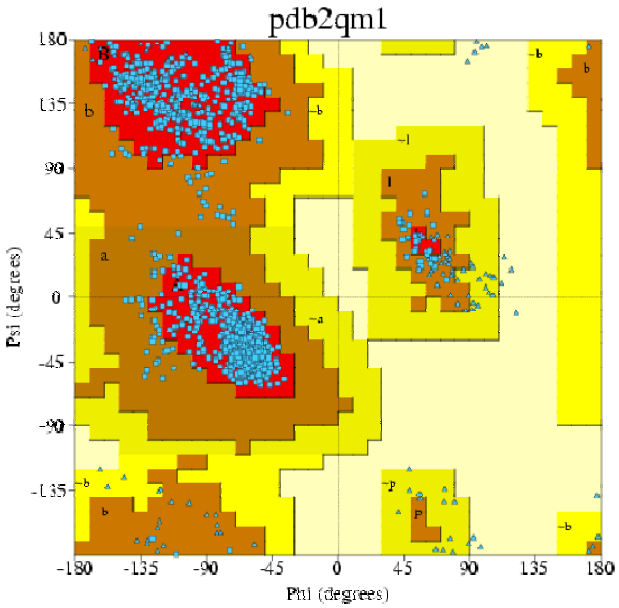
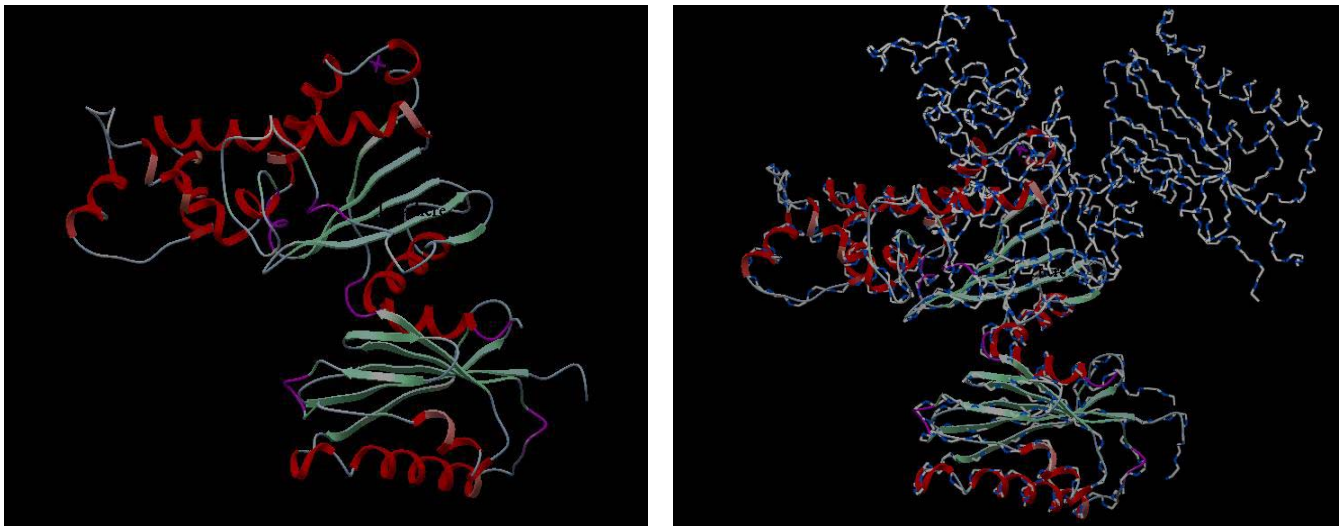


Figure 3 | Ramachandran Plot amino acids clustering determinations of the target

Ramachandran Plot Statistics Details:

		No. of residues	%-tage
Most favoured regions	[A, B, L]	1001	91.9%
Additional allowed regions	[a,b,l,p]	88	8.1%
Generously allowed regions	[a,b,l,p]	0	0.0%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		1089	100.0%
End-residues (excl. Gly and Pro)		8	
Glycine residues		168	
Proline residues		32	
Total number of residues		1297	

Ramachandran plot predicted 2QM1 Template protein model was number of residues in the most of favored regions A, B, L and the percentage of tag is 91.9%, Additional allowed regions is 88 residues and 8.1 % of the tag. GLY “168” residues, PRO “32” residues. The total number of was in 1297 residues presented in the plot. Non - GLY and PRO was 1089 residues. Ramachandran plot value had shown (Fig. 3).



A B

Figure 4 | A - GLK Modeled Protein . B-Superpose of the predicted model of GLK from *S. pneumoniae* onto to the template PDB ID: 2QM1.

Analysis of the structure to buried inside the catalytic cleft

Accessible surface area (ASA) analysis of the predicted GLK model showed the active site amino acid Asp “115” with zero ASA value (Fig. 5) and is.

Table 1 | Accessible surface area of Protein amino acids Distance (A^o) between the residues

S.No	Amino acid	Distance (A^o)
1	Pro	109
2	His	156
3	Gly	161
4	Cys	179
5	Cys	181
6	Cys	186
7	Val	77
8	Asp	78
9	Glu	103
10	Asn	114
11	Ala	116
12	Asn	117
13	Gly	146
14	His	168

Some of the ligand binding residues and metal ion binding residues were found to have high ASA values was Pro109, His156, Gly161, Cys179, Cys181, Cys186 and some others were found to have low ASA values was Val77, Asp78, Glu103, Asn114, Ala116, Asn117, Gly146 and His168. Residues with high ASA values are on the surface of the cleft and those with low values are buried inside (Table. 1).

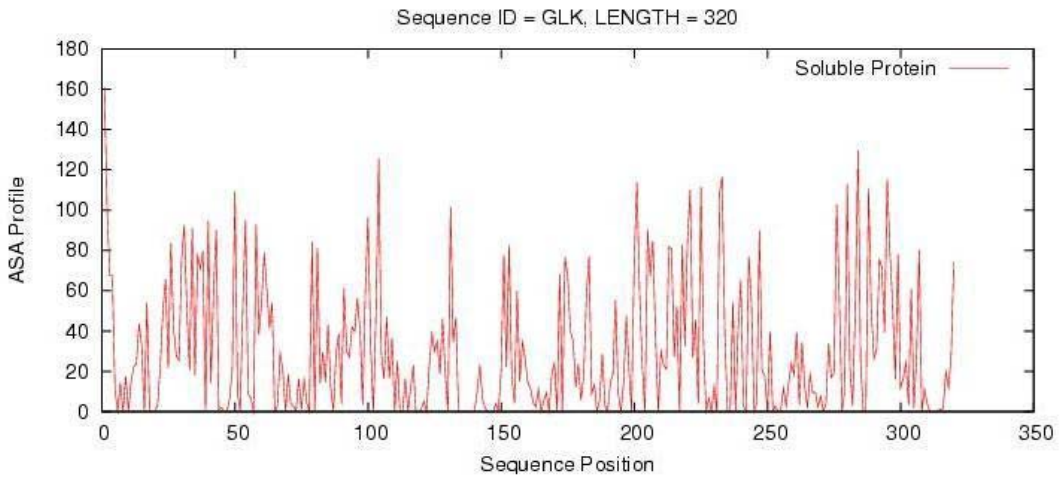


Figure 5 | ASA analysis of the developed model of GLK from *S. pneumoniae*.

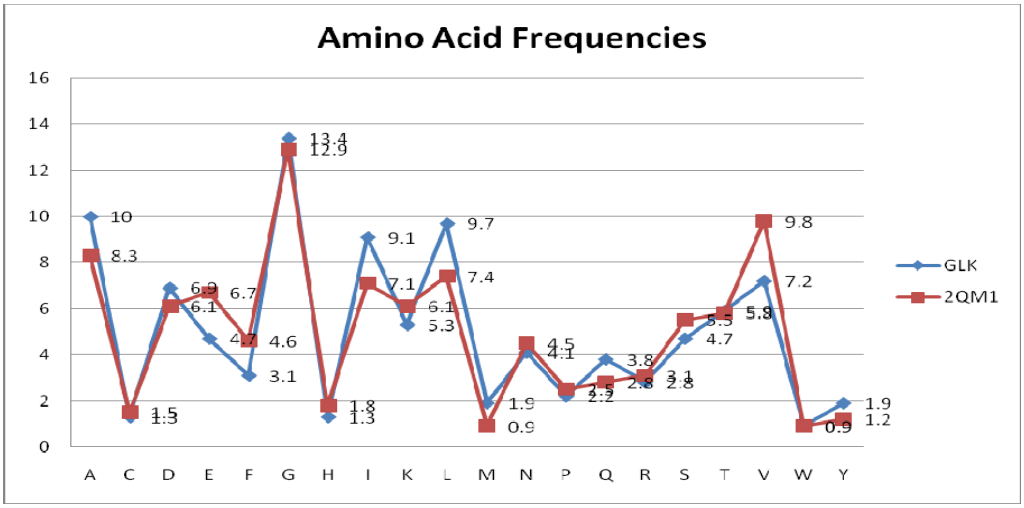


Figure 6 | Frequencies of the GLK and 2QM1 Protein sequences , percentages proteins GLK-2QM1 residues A (10.0 - 8.3%), C (1.3 - 1.5%), D (6.9 - 6.1%), E(4.7 - 6.7%), F (3.1 - 4.6%), G (13.4 - 12.9%), H (1.3-1.8), I (9.1 - 7.1%), K (5.3-6.1%), L(9.7-7.4), M (1.9-0.9%), N(4.1-4.5%), P(2.2-2.5%), Q(3.8-2.8%), R(2.8- 3.1), S(4.7-5.5%), T(5.9-5.8%), V(7.2-9.8%), W(0.9-0.9%) and Y(1.9-1.2%).

Table 2| Frequencies of the GLK and 2QM1 Protein sequences

FREQUENCIES OF AMINO ACID RESIDUES				
	GLK		2QM1	
Single letter	A. No	Percentage (%)	A. No	Percentage (%)
A	32	10.0	27	8.3
C	4	1.3	5	1.5
D	22	6.9	20	6.1
E	15	4.7	22	6.7
F	10	3.1	15	4.6
G	43	13.4	42	12.9
H	4	1.3	6	1.8
I	29	9.1	23	7.1
K	17	5.3	20	6.1
L	31	9.7	24	7.4
M	6	1.9	3	0.9
N	13	4.1	15	4.5
P	7	2.2	8	2.5
Q	12	3.8	9	2.8
R	9	2.8	10	3.1
S	15	4.7	18	5.5
T	19	5.9	19	5.8
V	23	7.2	32	9.8
W	3	0.9	3	0.9
Y	6	1.9	4	1.2

METHODS

The study was performed on an Intel ® Core ™ 2 Duo CPU E4600 @ 2.40GHz + 2.40GHz with Windows vista operating system. Sequence alignment was performed with PIR -Protein Information Resource [25] (Fig. 7) software and homology modeling was carried out using Molsoft -ICM v3.5 [26] comparative modeling software. Energy minimization of the developed model was performed using Swiss PDB Viewer[12]**.** Protein structure checks were conducted using the - Adit validation server[13]. Protein structural images were developed using Rasmol[16] **.**The color illustrates is conserved residues which are not involved in active site, ligand binding and metal binding are shaded in tan color. Conserved residues which are involved in active site are shaded in blue. Conserved residues involved in ligand binding are shaded in red and conserved residues involved in metal binding are shaded in orange.

Glucose Kinese (GLK) *Streptococcus pnemoniae* template **Sequence from the Swissprot**

>NP_345173/glk / *S. pnemoniae*

MMSQKIIGID LGGTSIKFAI LTTAGEIQGK WSIKTNILDE GSHIVDDMIE SIQHRLDLLG
LAAADFQGIGMGSPGVVDRDKGTVIGAYNLNWKTLQPIKQKIEKALGIPFFIDNDANVA
ALGERWMGAGDNQPDVVFMTLGTGVGGGIVAEGKLLHGVAGAAGELGHITVDFDQPI
SCTCGKKGCLETVASATGIVNLTRRYADEYEGDAALKRLIDNGEEVTAKTVFDLAKEG
DDLALIVYRNFSRYLGIACANIGSILNPSTIVIGGGVSAAGEFLLQGVQKVYDENSFPQVR
TSTKLALATLGNDAG VIGAASLVLQ

Target sequence 2QM1 of the *Enterobacteria faecalis*

>2Q1M/PDB/

SSNAXDKKIIGIDLGGTTIKFAILTTDGVVQQKWSIETNILEDGKHIVPSIIESIRHRIDLYN
MKKEDFVGIGMGTPGSVDIEKGTVVGAYNLNWTTVQPVKEQIESALGIPFALDNDANV
ALGERWKGAGENNPDVIFITLGTGVGGGIVAAGKLLHGVAGCAGEVGHVTVDPNGFDC
TCGKRGCLETVSSATGVVRVARHLSEEFAGDSELKQAIDDGQDVSSKDVFEFAEKGDHF
ALMVVDRVCFYLGLATGNLGNTLNPDSVVIGGGVSAAGEFLRSRVEKYFQEFTFPQVR
NSTKIKLAELGNEAGVIGAASLALQFSKEK

Glucose Kinese (GLK) sequence alignment and model building:

The protein sequence of *Streptococcus pneumoniae glucose* kinese (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 *Enterobacteria faecalis* with a 65.615% sequence identity (Fig. 7) to the query sequence. We then used the structure described in 2QM1 as a template for model building. We used Genious software to generate an alignment of the query and template. The GLK model from *Streptococcus pneumoniae* was superposed on to template 2QM1 and the Cα and back bone RMSD values using Molsoft ICM Pro[26] algorithm and online server SUPERPOSE [19]**.** The model was further checked with Ramachandran plot at PROCHECK[13, 23 and] WHAT IF [24]. Accessible surface area prediction using ASAP was performed [20]**.** The two protein amino acids frequencies of the sequence percentage values shown (Table. 2).

```
s-w opt: 1414  Z-score: 1732.4  bits: 328.8 E(): 1.1e-94 Smith-Watermanscore:
1414;  65.615% identity (65.823% ungapped)in 317 aa overlap (4-320:6-321)

                    10        20        30        40        50
GLK       MMSQKIIGIDLGGTSIKFAILTTAGEIQGKWSIKTNILDEGSHIVDDMIESIQHRLDL
              .:::::::::::.:::::::: : .: ::::.::::..:.::: ..:::::.::::
2Q1M   RYXXDKKIIGIDLGGTTIKFAILTTDGVVQQKWSIETNILEDGKHIVPSIIESIRHRIDL
                  10        20        30        40        50        60

          60        70        80        90       100       110
GLK     LGLAAADFQGIGMGSPGVVLRDKGTVIGAYNLNWKTLQPIKQKIEKALGIPFFIDNDAANV
         ..   :: :::::.:: :: .:::::.:::::::: :.::.:..::.:::::: .::::::
2Q1M    YNMKKEDFVGIGMGTPGSVDIEKGTVVGAYNLNWTTVQPVKEQIESALGIPFALDNDANV
                  70        80        90       100       110       120

        120       130       140       150       160       170
GLK     AALGERWMGAGDNQPDVVFMTLGTGVGGGIVAEGKLLHGVAGAAGELGHITVDFDQPISC
        ::::::: :::.:::::.:.:::::::::::: ::::::::: :::.::::::  . ..:
2Q1M    AALGERWKGAGENNPDVIFITLGTGVGGGIVAAGKLLHGVAGCAGEVGHVTVD-PNGFDC
              130       140       150       160       170

        180       190       200       210       220       230
GLK     TCGKKGCLETVASATGIVNLTRRYADEYEGDAALKRLIDNGEEVTAKTVFDLAKEGDDLA
        ::::.:::::::::.: ..:. ..:. ::. ::. ::.:..:..: ::..:..:: .:
2Q1M    TCGKRGCLETVSSATGVVRVARHLSEEFAGDSELKQAIDDGQDVSSKDVFEFAEKGDHFA
      180       190       200       210       220       230

        240       250       260       270       280       290
GLK     LIVYRNFSRYLGIACANIGSILNPSTIVIGGGVSAAGEFLLQGVQKVYDENSFPQVRTST
        :.:       :::.: .:.:. ::::..:::::::::::::: . :.: ..: .::::::.::
2Q1M    LMVVDRVCFYLGLATGNLGNTLNPDSVVIGGGVSAAGEFLRSRVEKYFQEFTFPQVRNST
      240       250       260       270       280       290

        300       310       320
GLK     KLALATLGNDAGVIGAASLVLQ
        :. :: :::.:::::::::.::
2Q1M    KIKLAELGNEAGVIGAASLALQFSKEK
      300       310       320     326
```

Figure 7 | Pairwise sequence alignment result of GLK from *Streptococcus pneumoniae* and the template PDB ID: 2QM1 from *E. feacalis*. The [:] is described much similarity. The [.] Symbol described somewhat similarity.

Discussion

Homology modeling was designed and developed for *Streptococcus pneumoniae* glucose kinese enzyme 3D structural model using Molsoft - ICM v3.5 [26] because three dimensional structures are not available as yet at PDB. The structure of *S. pneumoniae* Glucose kinese enzyme is important for establishing its molecular function. Glucose kinese enzyme 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* Glucose kinese enzyme. The sequence similarity is 65.615% with the template and reliability of the predicted model thus generated using Molsoft ICM v3.5 [26] 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 *Streptococcus pneumoniae* glucose kinese enzyme using predicted model. Ramachandran plot predicted 2QM1 Template protein model was number of residues in the most of favored regions A, B, L and the percentage of tag is 91.9%, Additional allowed regions is 88 residues and 8.1 % of the tag. Gly "168" residues, Pro "32" residues. The total number of was in 1297 residues presented in the plot. Non - Gly and Pro was 1089 residues (Fig. 3). The Dot plot predicted of the GLK sequence and 2QM1 protein sequences was sixty five percentage identities (Fig. 2). Solvent accessible surface area predicted of the glucose kinese enzyme 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. Importance of solvent exposed catalytic residues in molecular function.

References

1. Dagan, R. Vaccine *19 Suppl 1: S9–S16.* PMID 11163457.
2. Johnston , R. B., Jr, *Rev. Infect., Dis.* **13**, S509-S517 (1991)
3. Musher, D. M, *Clin. Inect. Dis.,* **14**, 801-807 (1991)
4. Neralla, S. & Meyer, K.C. *Drugs Aging.,* **21**,851- 864 (2004)
5. Alexander, J. E. (1994) Infect. *Immun.,* **62**,5683-5688
6. Novak, R. *Nature,* 399:590-593 (1999)
7. Dagan, R. *Curr. Opin. Otolaryngol. Head Neck Surg.*, **12**, 488-494 (2004)
8. McCormick, A.W. *Nat. Med.,* **9**, 424-430 (2003)
9. Cohen, J. *Science,* **265**, 1371-1373 (1994)
10. Thompson, J.D. *Nucl. Acids. Res.,* **22**, 4673- 4680 (1994)
11. Guex, N. & Peitsch, M.C. *Electrophoresis,* **18**,2714-2723 (1997)
12. Laskowski, R. A. *J. Appl. Cryst.,* **26**,283-291(1993)
13. Vriend, G. *J. Mol. Graph.,* **8**,52-56 (1990)
14. Sippl, M.J. *Proteins,* **17**,355-362 (1993)
15. www.umass.edu/microbio/rasmol/
16. http:// www.ncbi.nlm.nih.gov/
17. Altshul, S.F. J.Mol.Biol., **215**,403-410 (1990)
18. http://wishart.biology.ualberta.ca/SuperPose/
19. http://ccb.imb.uq.edu.au/ASAP
20. Ryan, K,J., Ray, C,G.*McGraw Hill*. ISBN 0-8385-8529-9. (2004)
21. Pericone., Christopher, D., Overweg., Karin., Hermans., Peter, W. M., Weiser. & Jeffrey, N. *Infect Immun* **68** (7), 3990–3997 (2000)
22. Regev-Yochay, G., Trzcinski, K., Thompson, C.M., Malley, R. & Lipsitch, M. *J Bacteriol* **188** (13), 4996–5001. PMID 16788209. (2006)
23. Laskowski, R.A., Mac Arthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK a program to check the streochemical quality of protein structure *J.Appl.crystallogr*. **26**, 283-291(1993).
24. Vriend, G. WHAT IF: amolecular modeling and drug design program. *J.Mol.Graph*. **8**,526-529 (1990).
25. http://pir.georgetown.edu/cgi-bin/pairwise.pl.
26. http://www.molsoft.com/icm_pro.html.