STRUCTURAL CHARACTERIZATION OF LEGIONELLOSIS DRUG TARGET CANDIDATE ENZYME PHOSPHOMANNOMUTASE FROM LEGIONELLA PNEUMOPHILA STRAIN PARIS: AN IN SILICO APPROACH

Md. Arif Khan[1] and Mohammad Neaz Morshed[2]

[1, 2] Department of Chemistry, Military Institute of Science and Technology (MIST),
Mirpur Cantonment, Dhaka 1216, Bangladesh.
[1]arifkhanbge35@gmail.com, [2]mmneaz@hotmail.com

ABSTRACT

The harshness of legionellosis differs from mild Pontiac fever to potentially fatal Legionnaire's disease. The increasingdevelopment of drug resistance against legionellosis has led to explore new novel drug targets. It has been found thatphosphoglucosamine mutase, phosphomannomutase, and phosphoglyceromutase enzymes can be used as the mostprobable therapeutic drug targets through extensive data mining. Phosphomannomutase isconcerned in a process called glycosylation.The purpose of this study was to predict the potential target of that specific drug. For this,the 3D structure of Phosphomannomutase of Legionella pneumophila (strain Paris) was determined by means ofhomology modeling through Phyre2 and refined by ModRefiner. The designed model was evaluated with a structurevalidation program, for instance, PROCHECK, Verify3D, and QMEAN, for further structural analysis. Secondary structural features were determined through self-optimized prediction method with alignment (SOPMA) and interacting networks by STRING. The analytical result of PROCHECK showedthat 91.0% of the residues are in the most favored region, 8.50% are in the additional allowed region of the Ramachandran plot. Verify3D graph value indicates a score of 0.77 and 0.91 for QMEAN respectively. The findings of this current study along with further extensive investigation may assist drug design against Legionellosis.

Key Words: Legionella pneumophila, Legionellosis, homology modeling.

1.0 INTRODUCTION

A general cause of hospital and community acquired pneumonia is Legionella pneumophila. About 90% cases of legionellosis for this species and the prime serogroup (sg) 1 of L. pneumophila counts for 84% of cases [1]. Legionella isgram negative, aerobic, which is a thin, pleomorphic, flagellated, non-spore forming bacteria present in naturally occurring and artificial water systems and is transmitted to humans by aerosol gulp of air [2].Patients with Legionellosis disease typically have pneumonia, fever, chills, and a cough, which may be dry or may produce sputum. Contagion sources can be identified by comparing environmental and clinical L. pneumophila isolates with a variety of typing methods. Among them, Pulsed-field gel electrophoresis (PFGE) is generally considered to be highly prejudiced [3,4-6,7]. Legionellosis was caused by a single L. pneumophila sg 1 strain in the area of Paris [8].

Phosphomannomutasecatalyzes the chemical reaction alpha-D-mannose 1-phosphate to D-mannose 6-phosphate. This enzyme works on a substrate, alpha-D-mannose 1-phosphate and converts it to D-mannose 6-phosphate. The name of the gene responsible for this enzyme is "Phosphomannomutase 2 (PMM2)". Phosphomannomutase isconcerned in a process

called glycosylation; add the groups of sugar molecules (oligosaccharides) to proteins [10]. Early steps of glycosylation, the PMM enzyme converts a molecule called mannose-6-phosphate to mannose-1-phosphate. Consequently mannose-1-phosphate is converted into GDP-mannose, which can transfer its small sugar molecule called mannose to the growing oligosaccharide chain. Congenital disorder of glycosylation type Ia caused by mutations in the PMM2 gene.Mutations change the structure of the PMM enzyme;however all of the mutations appear to result in reduced enzyme activity. This leads to a shortage of GDP-mannose within cells. For this,there is not enough activated mannose. Incorrect oligosaccharides are produced. The signs and symptoms in CDG-Ia are likely due to the construction of abnormally glycosylated proteins in many organs and tissues [11]. This enzyme participates in fructose and mannose metabolism. The appearance of drug resistance of L. *pneumophila* has led to the search for novel drug targets.

In the present, computational analysis of metabolic pathways of the bacteria and host was performed to identify novel drug targets that are non-homologous to Homo sapiens. All enzymes involved in the metabolic pathways of L. pneumophila strain Paris could be searched against the proteome of Homo sapiens, So by these way Phosphoglucosaminemutase, Phosphomannomutase enzymes might be the drug target but this two are vital for common presence of the bacterial metabolic pathways and can be used as potential therapeutic drug target.So, homology modeling will predict the desired function and possible diseases treatment if needed because it has a significant importance on cell metabolism systems.

The present study isaimed to predict the three-dimensional (3D) structure of Phosphomannomutaseby means of homology modeling.

2.0 MATERIALS AND METHODS

2.1 Sequence Retrieval

The amino acid sequences of the enzyme Phosphomannomutase in Legionella pneumophilia (strain Paris) were retrieved from the UniProt Knowledge Base (UniProtKB) database [12]. The accession ID of Phosphomannomutase is Q5X243 and it contains 462 amino acids.

2.2 Secondary Structure Prediction

Secondary structure was predicted by using the self-optimized prediction method with alignment (SOPMA) [13]. Protein’s secondary structural properties are including α helix, 3_{10} helix, Pi helix, Beta bridge, Extended strand, Beta turns, Bend region, Random coil, Ambiguous states and Other states.

2.3 Model Building

Phyre2(Protein Homology/Analogy Recognition Engine) was used to predict the 3D homology model of Phosphoglucosamine mutaseand Phosphomannomutase [29]. The input data of this enzyme was in FASTA format. After model building, it is necessary to furtherrefinement.

2.4 Model Refinement

ModRefiner(http://zhanglab.ccmb.med.umich.edu/ModRefiner/); an algorithm for atomic-level, high-resolution protein structure refinement was used to refine the predicted protein model. Protein sequences were given in the FASTA format. Refinement was done for several times to get the most accurate structure.

2.5 Evaluation And Validation Of The Model

The accuracy and stereo chemical quality of the predicted model was evaluated with PROCHECK by Ramachandran Plot analysis which was done through “Protein structure and model assessment tools” of Swiss-model workspace. The best model was selected on the basis of overall G-factor, number of residues in core, allowed, generously allowed and disallowed regions. ERRAT, Verify3D and QMEAN were used for further analysis of the selected model.

2.6 Protein–protein interaction networking

Protein cooperates with other protein to perform accurate function. STRING was used to identify protein-protein interaction. STRING(http://string-db.org/) is a biological database, which was used to construct PPI network for different known and predicted interactions. At present, the database covers up to 5,214,234 proteins from 1133 organisms [14].

3.0 RESULTS

3.1 Secondary Structure Analysis

Secondary structure analysis is increasing day by day to predict protein function and structure. The secondary structure of Phosphomannomutase was predicted by SOPMA with standard parameters. Secondary structure parameters of Phosphomannomutase are presented in a tabulated form in Table 1 respectively. The graphical secondary structure of Phosphomannomutase is shown in Fig. 1.

3.2 Protein-Protein Interaction (ppi) Network Generation

The interacting partners of Phosphoglucosamine mutase and Phosphomannomutase were determined by STRING (Fig. 2).

3.3 Model Building

Phyre2 was used to build the three-dimensional model of the perspective proteins. Three-dimensional (3D) protein structures provide valuable insights into the molecular basis of protein function.

3.4 Refinement of the Predicted Model

Phyre2 generated model was considered for further refinement through Modrefiner. Refined model is depicted in Fig. 3.

3.5 Model Validation

For validation of the predicted structure Ramachandran plot analysis was done by PROCHECK server. The results of this analysis are depicted in Fig.3 and Table 2. The selected model was verified by using ERRAT, Verify 3D, and 3D match from Softberry. Ramachandran plot was done by PROCHECK to measure the accuracy of protein models. ERRAT, Verify 3D and QMEAN server were used for the verification of protein models and results were tabulated in Table3.

4.0 DISCUSSION

The UniProt Knowledge Base (UniProtKB) delivers an authoritative resource for protein sequences and functional information. Sequences of Phosphomannomutase of L pneumophilia (strain Paris) was obtained from UniProtKB [16]. Secondary structure of modeled Phosphoglucosamine mutase and Phosphomannomutase were analyzed by SOPMA. It reveals that there have no significant differences between these two protein. Phosphoglucosamine mutase contains 42.64% alpha helix,18.80% extended strand, 9.89% Beta turn and 29.69% random coil [Table 1]. Phosphomannomutase contains 41.56% alpha helix, 18.40% extended strand, 8.23% Beta turn and 31.82% random coil [Table 1].

Protein-protein interaction of Phosphomannomutase, was generated through STRING is presented in Fig. 2. STRING forecasts a confidence score, 3D structures of protein and Protein domains. STRING utilizes references from UniProt and predicts functions of different interacting protein. PPI network demonstrates that Phosphomannomutase interacts with ten proteins among which dfp, Ipp2431, Ipp2946, Ipp0373 and Ipp0821 are hypothetical protein; dut stands on deoxyuridine 5'-triphosphate nucleotidohydrolase which is involved in nucleotide metabolism. It produces dUMP, the immediate precursor of thymidine nucleotides and it decreases the intracellular concentration of dUTP so that uracil cannot be incorporated into DNA; rmlAstands on glucose-1-phosphate thymidylyltransferase which catalyzes the formation of dTDP-glucose, dTTP and glucose-1-phosphate, as well as its pyrophosphorolysis. The next one is glmS, stands on glucosamine-fructose-6-phosphate aminotransferase which catalyzes the first step in hexosamine metabolism, converting fructose-6P into glucosamine-6P using glutamine as a nitrogen source.

Homology modelling of this unique and essential metabolic protein was done by using Phyre2 in order to obtain three-dimensional structures of those. Three-dimensional protein structure gives important insights about the molecular basis of protein function and thereby allows an effective design of experiments [17]. That is why, in the understanding and manipulation of biochemical and cellular functions of proteins, the high resolution 3D structure of a protein is the main key [18].

ModRefiner derived refined model of Phosphomannomutasewere analyzed. In initial

model, the percent of residues in favored region were 84.0%whereas 91.7% in the final model.

After that, PROCHECK was used to measure the accuracy of protein models. Parameter comparisons of these proteins were made with well-refined structures that have similar resolution. Ramachandran plot and Ramachandran plot statistics is shown in Fig.3 and Table 2. Ramachandran Plot Statistics of Phosphomannomutase revealed that most of the amino acid residues (above 90 % of amino acid residues) were present in most favored regions. Thus, the protein model was very good as all of the residues were within the limits of Ramachandran plot.

Verification is also done by Errat, Verify 3D and Qmean server. Errat uses a quadratic error function to characterize and differentiate between correctly and incorrectly a determined region of protein structures based on characteristic atomic interaction [19]. The overall quality of model of Phosphomannomutase is 71.492. Verify 3D graph value of Phosphomannomutaseis 0.77indicates that environmental profile ofPhosphomannomutase is quite good [20-23]. On the basis of a linear combination of six structural descriptions the QMEAN (Qualitative Model Energy Analysis) scoring function estimates the global quality of the models. The local geometry model analysis is done by a torsion angel potential over three consecutive amino acids and the quality of the model can be compared to reference structure of high resolution obtained from X-ray crystallography analysis through Z score. QMEAN Z-score provides an estimation of the "degree of nativeness" of the structural features observed in a model and indicates that the model is of comparable quality to experimental structures [24].

The assessing of long range interactions are carried through secondary structure specific distance dependent pair wise residue level potential. A solvation potential describes the burial status of the residues. SSE and ACC agreement ensures the quality assessment between the predicted and calculated secondary structure and solvent accessibility [9].

The respective values of Z-scores of C_β interaction energy, solvation energy, torsion angle energy, secondary structure, and solvent accessibility of Phosphomannomutaseare -0.39, 0.68, -1.23, 1.91 and 0.98 shown in Table 3. The overall QMEAN score for Phosphomannomutaseis 0.91. QMEAN generated results conferPhosphomannomutase as a qualified model for further drug target scopes [25].

5.0 CONCLUSION

The 3D structure of Phosphomannomutase of Legionella pneumophila (strain Paris) through in silico approach has been identified. The results might assist in developing drugs againststrain Paris.This in silico approach can be further utilized in drug design to identify putative drugtargets for other clinically significant pathogens.

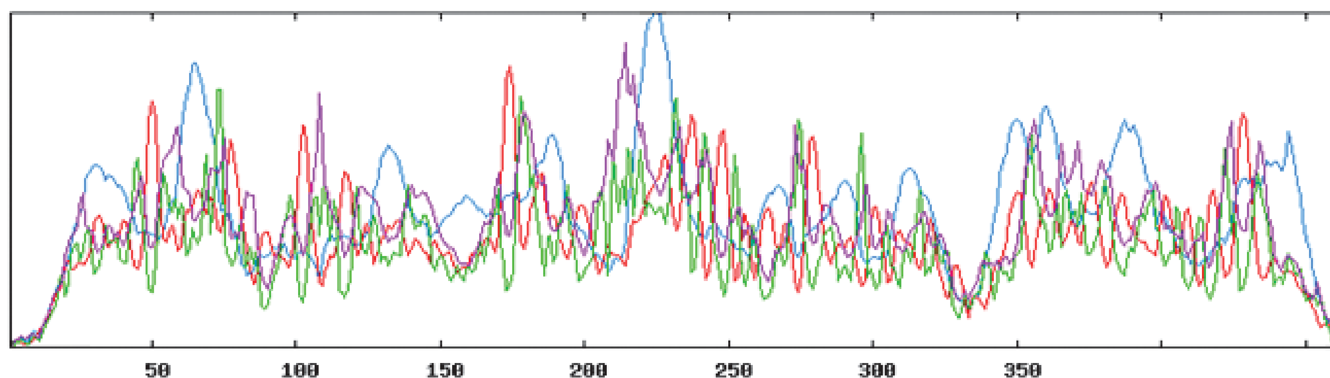

Fig 1. Predicted secondary structure of Phosphomannomutase. Here, helix is indicated by blue, while extended strands and beta turns are indicated by red and green, respectively.

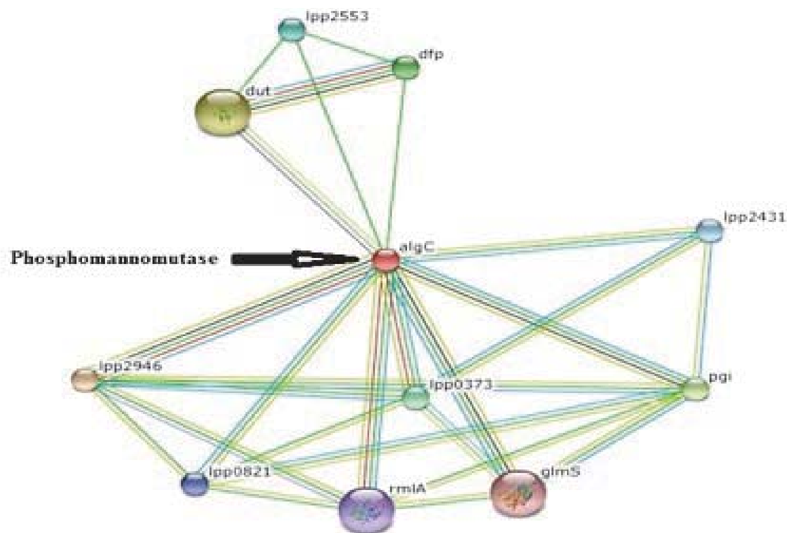

Fig 2. PPI network of Phosphomannomutase

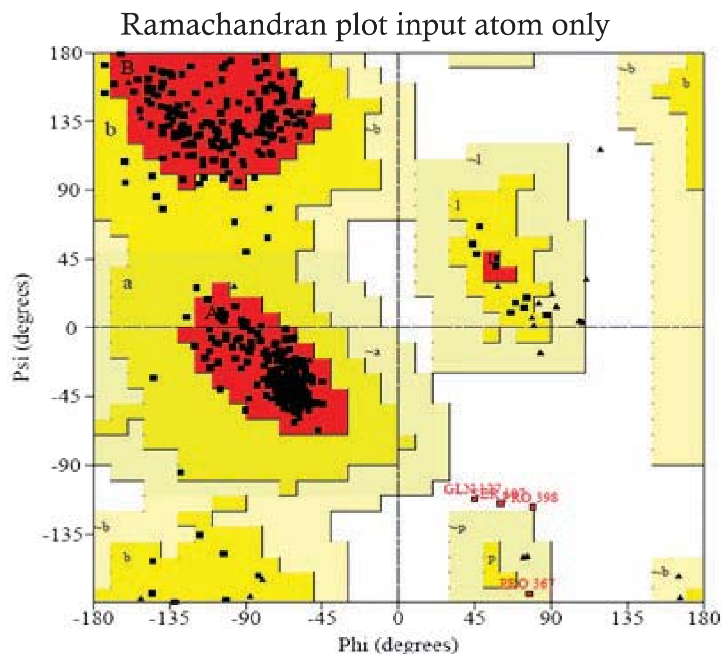

Fig 3. Ramachandran plot analysis of predicted proteins through Swiss model Workshop

Fig 4. Refined model of the proteins Phosphomannomutase

Table 1. Calculated secondary structure elements of Phosphomannomutase of Legionella pneumophilaby SOPMA.

Secondary structure	No of Residue	Percentage (%)
Alpha helix (Hh)	192	41.56%
3_{10} helix (Gg)	0	0.00%
Pi helix (Ii)	0	0.00%
Beta bridge (Bb)	0	0.00%
Extended strand (Ee)	85	18.40%
Beta turn (Tt)	38	8.23%
Bend region (Ss)	0	0.00%
Random coil (Cc)	147	31.82%
Ambigous states (?)	0	0.00%
Other states	0	0.00%

Table 2. Ramachandran plot analysis of Phosphomannomutase

Ramachandran plot statistics	Phosphomannomutase	
	Residue	%
Residues in the most favored regions [A,B,L]	365	91.0%
Residues in the additional allowed regions [a,b,l,p]	34	8.5%
Residues in the generously allowed regions [a,b,l,p]	0	0.0%
Residues in the disallowed regions [xx]	2	0.5%
Number of non-glycine and non-proline residues	401	100%
Number of end residues (excl. Gly and Pro)	2	
Number of glycine residues	35	
Number of proline residues	19	
Total number of residues	457	

Table 3. Z-scores of QMEAN for individual component

Scoring function term	Z-score
	Phosphomannomutase
C_β interaction energy	0.39
All atom pairwise energy	0.94
Solvation energy	0.68
Torsion angle energy	-1.23
Secondary structure agreement	1.91
Solvent accessibility agreement	0.98
QMEANscore	0.91

References

[1] V. L. Yu, J. F. Plouffe, M. C. Pastoris, J. E. Stout, M. Schousboe, A. Widmer, J. Summersgill, T. File, C. M. Heath, D. L. Paterson, and A. Chereshsky, Distribution of Legionella species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey, J InfectDis, 186, (2002), 127-128.

[2] R. M. Atlas, Legionella: from environmental habitats to disease pathology, detection and control, Environ Microbiol, 1, (1999), 283-293.

[3] N. K. Fry, S. Alexiou-Daniel, J. M. Bangsborg, S. Bernander, M. CastellaniPastoris, J. Etienne, B. Forsblom, V. Gaia, J. H. Helbig, D. Lindsay, P. C. Lück, C. Pelaz, S. A. Uldum, and T. G. Harrison, A Multicenter evaluation of genotypic methods for the epidemiologic typing of Legionella pneumophilaserogroup 1: results of a pan-European study, ClinMicrobiol Infect, 5, (1999), 462-477

[4] P. C. Lück, J. H. Helbig, V. Günter, M. Assman, R. Blau, H. Koch, and M. Klepp, Epidemiologic investigation by macrorestriction analysis and by using monoclonal antibodies of nosocomial pneumonia caused by Legionella pneumophilasero group 10, J ClinMicrobiol, 32, (1994), 2692-2697.

[5] J. M. Pruckler, L. A. Mermel, R. F. Benson, C. Giorgio, P. K. Cassiday, R. F. Breiman, C. G. Whitney, and B. S. Fields, Comparison of Legionella pneumophila isolates by arbitrarily primed PCR and pulsed-field Gel electrophoresis: analysis from seven epidemic investigations, J ClinMicrobiol, 33, (1995), 2872-2875.

[6] D. Schoonmaker, T. Heimberger, and G. Birkhead, Comparison of ribotyping and restriction enzyme analysis using pulsed-field gel electrophoresis for distinguishing Legionella pneumophila isolates obtained during a nosocomial outbreak, J ClinMicrobiol, 30, (1992), 1491-1498.

[7] M. J. Struelens, Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems, ClinMicrobiol Infect, 2, (1996), 2-11.

[8] C. Lawrence, M. Reyrolle, S. Dubrou, F. Forey, B. Decludt, C. Goulvestre, P. Matsiota-Bernard, J. Etienne, and C. Nauciel, 1999. Single clonal origin of a high proportion of Legionella pneumophila serogroup 1 isolates from patients and the environment in the area of Paris, France, over a 10-year period, J ClinMicrobiol, 37, (1999), 2652-2655.

[9] L. A. Kelley, S.Mezulis, C. M. Yates, M. N.Wass and M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nature Protocols, 10, (2015), 845–858.

[10] L. Jolly, F. Pompeo, J. van Heijenoort, F. Fassy, D. Mengin-Lecreulx, Autophosphorylation of phosphoglucosaminemutase from Escherichia coli, J Bacteriol, 182, (2000), 1280-5.

[11] S. Grünewald, The clinical spectrum of phosphomannomutase 2 deficiency (CDG-Ia), BiochimBiophysActa, 1792, (2009), 827-34.

[12] M. Magrane, and UniProt Consortium, UniProt Knowledgebase: a hub of integrated protein data, Database,2011, (2011), bar009.

[13] C. Geourjon, and G. Deléage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, ComputApplBiosci, 11, (1995), 681–4.

[14] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, STRING v9. 1: protein-protein interaction networks, with increased cover-age and integration,Nucleic Acids Res, 41,(2013), D808–15.

[15] O. Trott, and A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry, 31, (2010), 455-461.

[16] P. Palaga, L. Nguyen, U. Leser, J. Hakenberg, High-performance information extraction with Alibaba, EDBT ACM, 360, (2009), 1140-1143.

[17] S. Fidanova, and I. Lirkov, 3D Protein Structure Prediction, JAnaleleUniversitatii de Vest Timisoara Seria Matematica Informatica, 2, (2009), 33-46.

[18] L. A. Kelley, and M. J. Sternberg. Protein structure prediction on the Web: a case study using the Phyre server, NatProtoc, 4,(2009), 363-71.

[19] C. Colovos, and T. O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, ProteinSci, 2, (1993), 1511-9.

[20] D. Premalatha, P. Ravindra, L. V. Rao, Homology modeling of putative thioredoxin from Helicobacerr pylori, Indian J Biotechnol, 6, (2007), 4859.

[21] J. U. Bowie, R. Lüthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science, 253, (1991), 164-70.

[22] P. Benkert, M. Künzli, T. Schwede, QMEAN server for protein model qualityestimation, Nucleic Acids Res, 37, (2009); 10-4.

[23] D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, J MolBiol,292, (1999), 195-202.

[24] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics,27, (2011), 343- 50.

[25] P. Benkert, T. C. Tosatto, D. Schomburg, QMEAN: A comprehensive scoring function for model quality assessment, Proteins, 71, (2008), 261-77.