Research Article

Molecular Modelling and Structural Analysis of the *Helicobacter pylori* Pseudaminic Acid Biosynthesis UDP-Sugar Hydrolase PseG

Syeda Umme Habiba Wahid*
International Centre for Diarrheal Disease Research, Bangladesh, Dhaka, Bangladesh.
***Corresponding author's E-mail:** suhwahid@gmail.com

Accepted on: 24-09-2016; **Finalized on:** 30-11-2016.

ABSTRACT

Helicobacter pylori is the main etiological agent for*a broad range* of human diseases including gastritis, peptic ulcer and gastric carcinoma. Flagella-driven motility is the prerequisite for *H. pylori* to initiate colonization and development of infection. Glycosylation of flagella with a novel pseudaminic acid (Pse) is crucial for the synthesis of functional flagella. PseG is an inverting UDP-sugar hydrolase and belongs to the glycosyl transferase B super family. It catalyzes the fourth step in the biosynthesis of Pse by hydrolyzing the UDP-2, 4-diacetamido-2, 4, 6-trideoxy-β-L-altropyranose to yield the 6-deoxy-*N*-acetylhexosamine. No crystal-structural information on PseG is available. The 3D structure of PseG was modeledby using Int FOLD, I-TASSER, SWISS-MODEL, Phyre2 and Raptor server. The quality of the predicted models was analyzed based on QMEAN score. The selected model was refined by employing 3Drefine server. The final model was validated withthe Verify3D and ProQ program. PseG is an acidic α, β protein and harbors two domains. The 3D PseG model was used for comparison between known protein structures in the PDB database using the PDBe fold serverto elucidate overall structural fold and identify crucial residues for the catalytic mechanism. The deduced structural information would be useful *in vivo* biochemical studies to design inhibitors to combat multidrug-resistant bacterial infection.

Keywords: *Helicobacter pylori*, glycosylation, modeling, Int FOLD.

INTRODUCTION

Helicobacter pylori is a spiral-shaped, Gram-negative, microaerophilic bacterium that colonizes the stomach mucosa of more than half of the world's population.[1] *H. pylori* is the principal causative agent for a broad range of pathologies including gastric inflammation, peptic ulcers, mucosa-associated lymphoid tissue lymphoma, and gastric cancer.[2, 3]It is the first bacterium classified as a group I (definite carcinogen) for human gastric cancer by IARC (International Agency for Research on Cancer, 1994).[4]At present, there is neither a vaccine against *H. pylori* nor a single medication that can effectively cure the infection. Furthermore, the emergence of multiple-antibiotic resistance in *H. pylori* has complicated the situation.

Flagella-driven motility is the prerequisite for *H. pylori* to initiate colonization and to achieve the highest level of infection.[5]*H. pylori* flagellins are glycosylated with a novel nine-carbon sugar, pseudaminic acid (Pse). Flagellins glycosylation is essential for their assembly and bacterial motility.[6] In *H pylori*, Pse is synthesized by six enzymatic steps.[7] Therefore, the Pse biosynthesis pathway could be exploited to design novel therapeutics to combat *H. pylori* infection. The fourth enzymatic step in the Pse pathway is carried out by an inverting UDP-sugar hydrolase (PseG, EC# 3.6.1.57), which produces the 6-deoxy-*N*-acetylhexosamine by removing the nucleotide moiety from the UDP-2,4-diacetamido-2,4,6-trideoxy-β-L-altropyranose.[7] PseG belongs to the glycosyltransferase B (GT-B) superfamily.[8] Insertional mutagenesis of the gene encoding PesG in *H. pylori* resulted in aflagellated non-motile phenotype and suggested that PseG plays an essential role in the synthesis of functional flagella and bacterial pathogenesis.[6]

The catalytic mechanism of a functional homolog of PseG from *Campylobacter jejuni*has been reported by Liu and Tanner.[9]*C. jejuni* PseG (CjPseG) removesthe nucleotide *via* a metal-independent C–O bond cleavage mechanism. Another closely related GT-B super family enzyme from *Escherechia coli* is MurG (EcMurG). This is also an inverting metal- independent UDP-sugar glycosyl transferase and involved in the biosynthesis of peptidoglycan.[8] Crystal structure of *C. jejuni* PseG was solved by X-ray diffraction and consists of two domains.[10] The UDP molecule is accommodated at the long cleft between two domains. The CjPseG and EcMurG showed no detectable sequence similarity (<15%) with HpPseG. To the best of knowledge, no biophysical and structural studies on *H. pylori* PseG have been reported. Elucidation of the 3D structural information of *H. pylori* PseG would be useful to identify novel therapeutics targeting motility to combat multidrug resistance bacterial infection. In the present study, physiochemical parameters, structural information, ligand interactions mechanism and important residues for the catalytic mechanism of *H. pylori* PseG have been analyzed using the modeled 3D structure and also compared with other known structural homologs to elucidate functional information.

MATERIALS AND METHODS

Sequence Retrieval, PhysicochemicalProperties, and Secondary Structure Analysis

The coding sequence for *H. pylori* PseG **(**UniProt ID B6JKQ3; 190 amino-acid residues) was retrieved from NCBI's Protein database. The physiochemical and functional properties of the protein were determined by the Expasy Prot Param Web server (http://www.expasy.org/tools/protparam.html). This program performs theoretical measurement of various physicochemical parameters including molecular mass, iso electric point, extinction coefficient, in stability index, aliphatic index and grand average of hydropathicity (GRAVY). The predicted properties of *H. pylori* PseG are listed in Table 1. The presence of disulfide bond was predicted by employing the tool CYS_REC (http://linux1.softberry.com). The secondary structural components were predicted using the tool SOPMA (Self-Optimized Prediction Method with Alignment).[11]

TM Helix Prediction and Sub cellular Localization

The precise determination of sub cellular localization of a protein is crucial in predicting its cellular function. The subcellular localization of the protein was predicted by using the following servers: Prot Comp B (http://www.softberry.com), PSLpred[12]and PSORTb[13]. The membrane topology, trans membrane helices was predicted using the by TOPCONS server.[14]

Model building, Refinement, and Validation

The PseG 3D model was predicted with the following servers: IntFOLD[15], I-TASSER[16], SWIS-MODEL[17], RaptorX[18] and Phyre2[19]. The quality of all the predicted models was evaluated by the Protein Structure & Model Assessment Tools in the SWISS-MODEL workspace.[17, 20]The best model was selected based on the QMEAN score, which is a linear combination of major geometrical descriptors of protein. The best model was selected for further refined using the 3Drefine server.[21] The finalized model was validated by employing the ProQ (http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi) and Verify 3D program.[22] The PseG model was deposited in the PMDB Protein Model Data Base (PMDB id: PM0080607).

Ligand Docking in PseG

The final 3D PseG model was used for structural analysis and prediction of the ligand binding site using the 3DLigandSite[23]and I-TASSER web server[16].The UPD molecule was docked in the ligand binding pocket. The program Ligplot[24] was employed to represent ligand binding residues of HpPseG.

RESULT AND DISCUSSION

Physicochemicaland Functional Characterization

The physicochemical characteristics analysis of *H. pylori* PseG revealed that the molecular weight is 33.1 kDa and the protein is slightly acidic (computed pI=6.8) (Table 1). The instability index of the protein is greater than 40 which suggest that the protein is less stable in solution. Vetrivel and colleagues have reported that most cyto plasmic proteins serve as a drug target while protein embedded in the membrane proteins are most likely vaccine targets.[25]The protein is likely to be cytoplasmic and harbors no TM helix. The calculated half-life of PseG is 30 h in mammalian reticulocytes *in vitro*, >20 h in yeast *in vivo* and >10 h in *E. coli in vivo*. Computation of the grand average of hydropathicity resulted in a negative index which suggests that PseGis most likely to be insoluble in water. Therefore, expression of the protein for crystallographic and biochemical studies would be difficult.

Analysis of the amino acid composition of PseG revealed that Leucine is the most predominant amino acid (13.4 %) while Tryptophan is the least common (0.3%) (Table 2). The proportion of positively charged residues (Arg + Lys) in PseG is 11.7% (n=34) while the figure of negatively charged residues (Asp + Glu) was 11.0% (n=32). The analysis of disulfide bonding pattern (SS-bond) of Cysteine (Cys) residues using the Cys_Rec tool revealed that there are five Cys residues (C7, C21, C64, C204, and C206) in the protein, and the C64 and C206 could form an SS-bond. Analysis of the secondary structure element revealed that PseG is an αβ protein and consists of mainly α-helices (45.7%), followed by random coil (24.1%), extended β-strands (20%), and β-turn (10%).

Table 1: Physicochemical properties of the *H. pylori* PseG.

Properties	Value
Sequence length	290
Molecular weight (kDa)	33.1
Theoretical pI	6.76
Extinction coefficients ($M^{-1}cm^{-1}$)	26610
Instability index	49.64
Aliphatic index	96.55
Grand average of Hydropathicity (GRAVY)	-0.211

Model Building and Refinement

The 3D models of PseG were predicted using different web servers (Table 3) and the best model was selected based on the QMEAN score and QMEAN Z-score. The QMEAN score is a parameter between 0-1. Structures with QMEAN score equivalent to 1 are similar to X-ray diffracted crystal structure, and the QMEAN Z-scorerepresents that actual quality of the predicted model. Table 3 shows the detailed analysis of the predicted models. The IntFOLD generated a model with aQMEAN and QMEAN Z-scoresof 0.63 and -1.21,respectively (Table 1).[15] The IntFOLD model 1 was further refined and optimized by the 3Drefine program.[21] The calculated QMEAN and QMEAN Z-scores of the optimized model were0.658 and -1.22, respectively (Table 4 and Fig. 1).

Table 2: Amino acid composition of the*H. Pylori* PseG.

Amino Acids	No of residues	Percentage
Ala	16	5.5
Arg	5	1.7
Asn	16	5.5
Asp	9	3.1
Cys	5	1.7
Gln	11	3.8
Glu	24	8.3
Gly	13	4.5
His	9	3.1
Ile	22	7.6
Leu	39	13.4
Lys	27	9.3
Met	5	1.7
Phe	15	5.2
Pro	12	4.1
Ser	24	8.3
Thr	14	4.8
Trp	1	0.3
Tyr	14	4.8
Val	9	3.1

Table 3: Comparison of the predicted models of the *H. pylori* PseG.

No	Program	Model	Total Q-mean score	Q-mean Z score
1	Int FOLD	1	0.631	-1.51
2		2	0.604	-1.80
3		3	0.603	-1.81
4		4	0.566	-2.21
5		5	0.548	-2.41
6	I-TASSER	1	0.536	-2.53
7		2	0.476	-3.18
8		3	0.439	-3.58
9		4	0.440	-3.57
10		5	0.591	-1.95
11	SWIS MODEL	1	0.567	-2.15
12		2	0.440	-3.59
13		3	0.534	-2.29
14	RaptorX	1	0.600	-1.85
15	Phyre2	1	0.480	-3.15

Model Validation

The webserver ProQ and Verify3D server were used to analyze the stereochemical properties of the PseG model. The ProQ server revealed that PseG model has an LGscore of 3.2 which suggests that the quality of the final model is very good. Analysis of the model with Verify3D program resulted that more than 86% amino acids of the model have scored>= 0.2 in the 3D/1D profile. This finding suggests that residues were in the favorable region (Fig. 2).

Table 4: Assessment of the final model using the SWIS-MODEL structure assessment tool.

Scoring function term	Raw score	Z-score
C_beta interaction energy	-88.73	-1.02
All-atom pairwise energy	-6040.55	-1.40
Solvation energy	-9.41	-2.05
Torsion angle energy	-34.58	-2.46
Secondary structure agreement	81.4%	-0.13
Solvent accessibility agreement	77.6%	-0.53
QMEAN6 score	0.658	-1.22

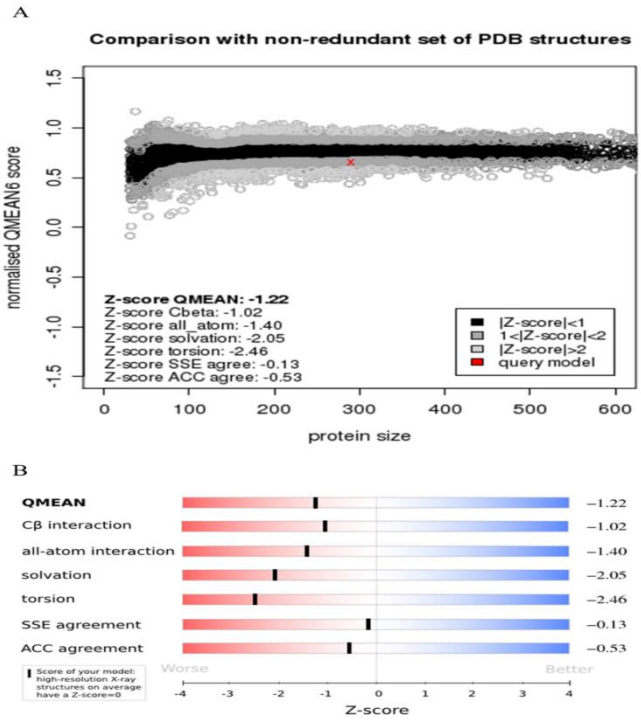

Figure 1: Quality analysis of the final refined *H. pylori* PseG 3D model. (A) The QMEAN quality analysis for the predicted *H. Pylori* PseG model. The predicted model (Generated from IntFOLD program)is presented by a red star. (B) The QMEAN z score of the optimized final model.

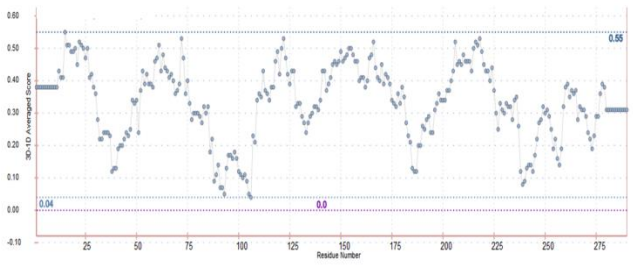

Figure2: The 3D scores for the modeled*H. Pylori* PseG structure.

The Overall Structure of *H. pylori* PseG and Comparison to Other Homologs

The PseG is an αβ protein and consistsof two domains: N-terminal domain (residues 1-129), and C-terminal domain (residues 137-290) connected by a short helix (residues130-136.) The N-terminal domain has a parallel six beta sheet (β2-β1-β3-β4-β5-β6) flanked by an α helices (α1) on one side and two α-helices (α2 and α3) on the other face (Fig. 3). The C-terminal domain harbors a parallel five beta sheet (β8-β7-β9-β10-β11) flanked by four α helices (α7-α8-α9-α11) on one side and three α helices(α5-α6-α10) on the other face. The two domains are linked by a short helix (α4) (Fig. 3).

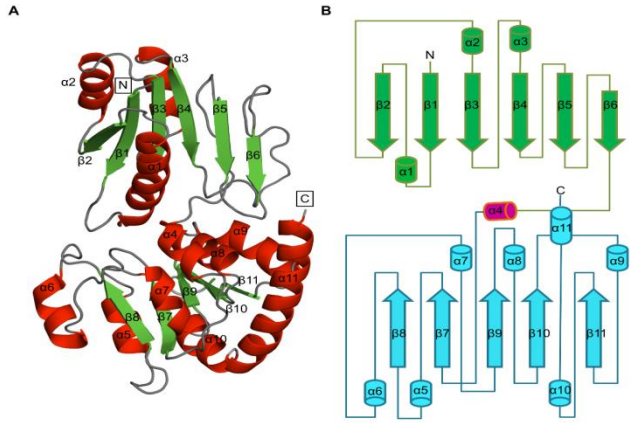

Figure3: The overall fold of *H. pylori* PseG. (A) Stereo diagram of the structure of the PseG molecule. β-strands and α-helices are represented as arrows and coils, and each element of the secondary structure is labeled as in text. The figure was prepared using PyMOL.[26](B) The topology of secondary structure elements of PseG. The N-terminal domain, domain-linking α-helix and C-terminal domain are colored as green, pink and cyan, respectively.

The search of PDB database using the PDBe Fold server revealed that CjPseG (Z-score =15.4, PDB code. 3hbm,[10])showed the highest similarity followed by *Pseudomonas aeruginosa* MurG (Z-score= 8.7, PDB code. 3s2u,[27]) and *E. coli* MurG (Z-score= 7.1, PDB code.1nlm,[8]). HpPseG, CjPseG, PaMurG and EcMurG harbor very similar fold [root-mean-square deviations of 1.36, 2.6 and 2.7 Å for the pairwise superimposition of 233, 238 and 232 C_α atoms from CjPseG, PaMurG and EcMurG, respectively] (Fig .4).

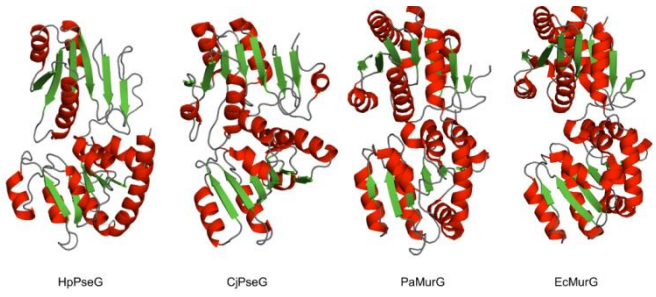

Figure 4: Comparison of *H. pylori* PseG to other bacterial glycosyltransferase B family proteins. α-helices, loops and β-strands are colored red, gray and green, respectively

However, HpPseG showed very limited sequence homology (<18% identity) with other GT-B super family proteins.

Prediction of Ligand Binding Pocket and Ligand Docking

The ligand binding site in HpPseG structurewas predictedusing the 3DLigandSite[23] and I-TASSER servers[16].The predicted ligand binding site is located between the two domains. The UDP-moiety of its cognate ligand (UDP-6-deoxy-AltdiNAc)was docked at the putative ligand binding site. Detail analysis of the modeled complex revealed that residues from the helix α1, α6, α7, strand β7, β7- α5 loop, and β8-α6 loop interact with the UDP moiety of the ligand.

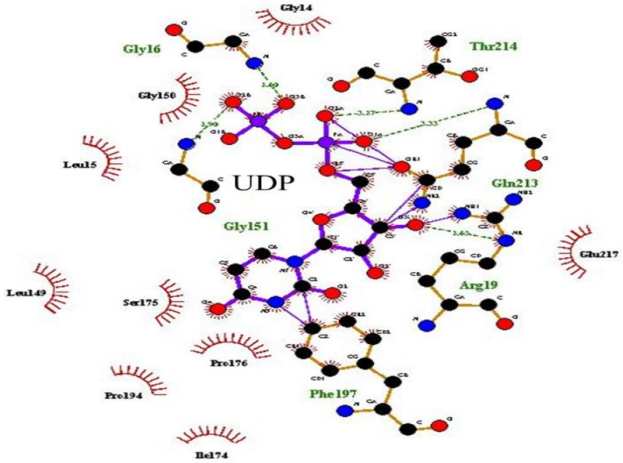

Figure 5: Ligplot results showing amino acids forming the ligand binding pocket of HpPseG. Interacting residues of PseG with docked Uridine di phosphate molecule are presented. Residues making vander Waals contacts with PseG are indicated by arcs with spokes radiating toward the ligand moieties they contact.

Figure 5 shows residues that form interactions with the UDP molecule. The UDP forms hydrogen bonds and van der Waal's interactions with six and nine residues, respectively. Mutagenesis study in closely related functional homolog CjPseG revealed that Histidine 17 residues is crucial for substrate recognition and serve as an active site basein for CjPseG.[10]Super imposition of HpPseG model on CjPseG structure showed that HpPseG has a His17 on the equivalent position. This finding suggests that HpPseG would likely to follow the similar catalytic mechanism of CjPseG.

CONCLUSION

This present study presents physicochemical properties, secondary structure analysis, functional information and crucial catalytic residues of *H. pylori* PseG by using a broadrange of bioinformatics tools. The derived properties and structure information would serve as a fundamental basis to design *in vivo* biochemical assay for better understanding the catalytic mechanism and molecular biology of GT-B super family proteins. This 3D model of *H. pylori* PseG could be useful for future drug discovery studies targeting bacterial motility to combat multidrug-resistantbacterial infection.

REFERENCES

1. Dunn B, Cohen H, Blaser MJ. *Helicobacter pylori*. Clinical Microbiology Reviews, 10, 1997, 720-741.
2. Fox JG, Wang TC. Inflammation, atrophy and gastric cancer. Journal of Clinical Investigation, 117, 2007, 60-69.
3. Suerbaum S, Michetti P. *Helicobacter pylori* infection. The New England Journal of Medicine, 347, 2002, 1175-1186.
4. International Agency for Research on Cancer.IARC Monographs on the evaluation of carcinogenic risks tohuman. 61, 1994, 177-240.
5. Ottemann KM, Lowenthal AC. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. Infection and Immunity, 70, 2002, 1984-1990.
6. Schirm M, Soo E, Aubry A, Austin J, Thibault P, Logan S. Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. Molecular Microbiology, 48, 2003, 1579-1592.
7. Schoenhofen IC, McNally DJ, Brisson JR, Logan SM. Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: Synthesis from UDP-*N*-acetylglucosamine by a single enzymatic reaction. Glycobiology, 16, 2006, 8C-14C.
8. Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S. Crystal structure of the MurG: UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. Proceedings of the National Academy of Sciences U S A, 100, 2003, 845-849.
9. Liu F, Tanner ME. PseG of pseudaminic acid biosynthesis a UDP-sugar hydrolase as a masked glycosyltransferase. Journal of Biological Chemistry, 281, 2006, 20902-20909.
10. Rangarajan ES, Proteau A, Cui Q, Logan SM, Potetinova Z, Whitfield D, Purisima EO, Cygler M, Matte A, Sulea T. Structural and functional analysis of *Campylobacter jejuni* PseG a UDP-sugar hydrolase from the pseudaminic acid biosynthetic pathway. Journal of Biological Chemistry, 284, 2009, 20989-21000.
11. Geourjon C, Deleage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer Applications in the Biosciences, 11, 1995, 681-684.
12. Bhasin M, Garg A, Raghava G. PSLpred: Prediction of subcellular localization of bacterial proteins. Bioinformatics, 21, 2005, 2522-2524.
13. Nancy YY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 26, 2010, 1608-1615.
14. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Research, 43, 2015, W401-407.
15. McGuffin LJ, Atkins JD, Salehe BR, Shuid AN, Roche DB. IntFOLD: An integrated server for modelling protein structures and functions from amino acid sequences. Nucleic Acids Research, 43, 2015, W169-W173.
16. Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nature Protocols, 5, 2010, 725-738.
17. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research, 42, 2014, gku340.
18. Källberg M, Margaryan G, Wang S, Ma J, Xu J. RaptorX server: A resource for template-based protein structure modeling. Methods in Molecular Biology, 1137, 2014, 17-27.
19. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, 10, 2015, 845-858.
20. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22, 2006, 195-201.
21. Bhattacharya D, Cheng J. 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. Proteins, 81, 2013, 119-131.
22. Eisenberg D, Lüthy R, Bowie JU. Verify3D: Assessment of protein models with three-dimensional profiles. Methods in Enzymology, 277, 1997, 396-404.
23. Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Research, 38, 2010, W469-473.
24. Laskowski RA, Swindells MB. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. Journal of chemical information and modeling, 51, 2011, 2778-2786.
25. Vetrivel U, Subramanian G, Dorairaj S. A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. The HUGO Journal 5, 2011, 25.
26. Schrodinger L. The PyMOL molecular graphics system, version 1.3 r1. 2010. There is no corresponding record for this reference, 2015.
27. Brown K, CM Vial S, Dedi N, Westcott J, Scally S, DH Bugg T, A Charlton P, MT Cheetham G. Crystal structure of the *Pseudomonas aeruginosa* MurG: UDP-GlcNAc substrate complex. Protein & Peptide Letters, 20, 2013, 1002-1008.

Source of Support: Nil, **Conflict of Interest:** None.