Pertanika J. Trop. Agric. Sc. 42 (2): 609 - 625 (2019)

PERTANIKA JOURNALS

TROPICAL AGRICULTURAL SCIENCE

Journal homepage: http://www.pertanika.upm.edu.my/

# Partial Purification and Model Structure of BPSL2774, a Hypothetical Protein from *Burkholderia pseudomallei* Predicted to be a Glycosyltransferase

**Siti Marhamah Drahaman[1], Hanisah Ujang[1], Nor Azurah Mat Akhir[2], Noraslinda Muhamad Bunnori[1,3] and Aisyah Mohamed Rehan[1,3]***

*[1]Department of Biotechnology, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, 25200 Kuantan, Pahang, Malaysia*
*[2]Malaysia Genome Institute, Jalan Bangi, 43000 Kajang, Selangor, Malaysia*
*[3]Research Unit Bioinformatic and Computational Biology (RUBIC), Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, 25200 Kuantan, Pahang, Malaysia*

## ABSTRACT

Melioidosis is a disease that infects humans and animals, and can be detrimental in humans. Mortality rate from melioidosis septic shock due to infection from Gram negative *Burkholderia pseudomallei* (*B. pseudomallei*) in endemic regions of Malaysia and Thailand remains high despite available antimicrobial therapy. Multiple strategies are employed to identify essential genes and drug targets in this bacterium to improve current antimicrobial therapies. This is important as *B. pseudomallei* is intrinsically resistant to many commonly used antibiotics. In this study, hypothetical genes predicted to be essential for *B. pseudomallei* by transposon-directed insertion site sequencing (TraDIS) technique were selected. One target gene, *BPSL2774,* has been successfully amplified and cloned from genomic DNA of *B. pseudomallei* strain K96243. Glutathione S-transferase (GST) affinity tag chromatography was performed for partial protein purification. The target protein was successfully expressed in soluble form with satisfactory yield output. Mass spectrometry analysis of 60 kDa Coomassie-stained gel band confirmed the presence of the soluble expressed tagged-target protein, co-purified with *Escherichia coli* chaperonin proteins, possibly due to their interaction with the target protein. BPSL2774 protein have considerable homology to glycosyltransferase GTB type superfamily and RfaB superfamily. On the basis of this similarity, the three-

ARTICLE INFO

*Article history:*
Received: 04 January 2019
Accepted: 11 March 2019
Published: 30 May 2019

*E-mail addresses:*
marhamahdr@gmail.com (Siti Marhamah Drahaman)
hanisah.ujang@gmail.com (Hanisah Ujang)
azurah@mgi-nibm.my (Nor Azurah Mat Akhir)
noraslinda@iium.edu.my (Noraslinda Muhamad Bunnori)
mraisyah@iium.edu.my (Aisyah Mohamed Rehan)
* Corresponding author

ISSN: 1511-3701
e-ISSN: 2231-8542

dimensional structure of BPSL2774 has been modelled and assessed by protein model quality servers. Taking all the results into account, the functional annotation of BPSL2774 protein as a glycosyltransferase is recommended, though future validation from biochemical experiments will be needed to support this.

*Keywords*: *Burkholderia pseudomallei*, BPSL2774 hypothetical protein, glycosyltransferase

## INTRODUCTION

Melioidosis is an infectious disease spread by *Burkholderia pseudomallei* (*B. pseudomallei*), a Gram negative bacterium which resides in contaminated water and soil. Direct contact with the contaminated source either through exposed skin abrasions, inhalation, or ingestion can spread the disease to human and animals. Soil, stagnant water and rice fields are the natural habitat for this bacterium and can be found in endemic regions including Southeast Asia and northern Australia (Chewapreecha et al., 2017; Limmathurotsakul et al., 2016).

Acute cases of melioidosis typically present within 1-21 days after infection, while chronic cases can persist for months (Wiersinga et al., 2006). Death can occur within the first 48 hours due to septic shock, even with optimal antimicrobial chemotherapy given (Holden et al., 2004). Examples of common clinical manifestations include localized abscess formation, metastatic pneumonia, hepatic and splenic abscesses, displaying evidence of bacterial dissemination to distant sites. Melioidosis often affects individuals with underlying medical conditions associated with an altered immune response. The major underlying risk factor for melioidosis are diabetes mellitus, followed by chronic renal disease (Nathan et al., 2018; Wiersinga et al., 2006).

In the past two decades, melioidosis was categorized as an important human infection in Malaysia, Singapore and across the north of Australia (Nathan et al., 2018; Schweizer, 2012; Sim et al., 2018). There are increasing awareness of this disease from other melioidosis-endemic tropical countries i.e. Indonesia (Tauran et al., 2018), Brunei (Pande et al., 2018), Sri Lanka (Corea et al., 2018), Laos (Dance et al., 2018) and Cambodia (Turner et al., 2016). In Malaysia, incidence of melioidosis varies between state, with recorded cases of melioidosis in Kedah, Kelantan, Pahang, Johor, Sabah and Sarawak (reviewed by Nathan et al., 2018). Case fatality varied between 33-54% in four Malaysian case series that included all cases irrespective of bacteraemic status. Cases of bacteraemic melioidosis have higher mortality rates, with up to 63% mortality recorded in Kelantan (Deris et al., 2010).

*Burkholderia pseudomallei* is naturally resistant to many commonly used antibiotics (Holden et al., 2004; Wiersinga et al., 2006). The intrinsic antibiotic resistance is due to the bacterium's physicochemical properties that exclude entry of drug molecules using its lipopolysaccharide component of the cell membrane, enzymatic inactivation, target mutation or efflux from the cell (Rhodes & Schweizer, 2016; Schweizer,

2012). Putative resistance mechanisms for this bacterium that have been reported include the action of seven Ambler class A, B and D β-lactamases, ten multidrug efflux systems and a putative aminoglycoside acetyl transferase (Holden et al., 2004). It secretes lecithinase, lipase, hemolysin and siderophore for its survival and maintenance (Stevens et al., 2002).

The genome of *B. pseudomallei* (strain K96243, a clinical isolate from Thailand was first to be fully sequenced) is known to be one of the largest and most complex genome (Holden et al., 2004). It comprised two chromosomes of 4.07 and 3.17 megabase pairs, respectively. The large chromosome is important for metabolism and growth, whereas the small chromosome encodes accessory functions associated with adaptation and survival (Holden et al., 2004).

Current research efforts include a focus on prevention of disease and finding ways to reduce mortality and the rate of relapse. A potential vaccination strategy has also been considered using the closely related avirulent *Burkholderia thailandensis* and other attenuated strains. However, this approach was not pursued due to the extensive exposure of both *B. thailandensis* and *B. pseudomallei* to the patients (Cheng & Currie, 2005). A current review on potential melioidosis vaccine candidates indicated that the vaccination strategy required more extensive development and evaluation to protect against multiple routes of disease acquisition, as well to consider risk factors for infection e.g. diabetes (Peacock et al., 2012).

The development of new antimicrobial therapies is emerging, with researchers utilizing different tools to identify essential genes and drug targets to combat melioidosis due to its persistence. Hence, identification of *B. pseudomallei* essential genes and its products may represent excellent targets for development of novel antimicrobial drugs. As an example, one study by Moule et al. (2016) on a transcription accessory protein in *B. pseudomallei*; Tex, had shown that deletion of the particular protein produced a highly attenuated *B. pseudomallei tex* mutant phenotypes. This indicates that identifying essential genes and their subsequent characterization can provide fundamental information on the bacterium survival strategy or pathogenicity.

Transposon library sequencing techniques known as transposon-directed insertion site sequencing (TraDIS) and transposon sequencing (Tn-seq) have been recently used to screen *B. pseudomallei* K96243 bacterial libraries and identify essential genes within the genome (Moule et al., 2014). In this study, hypothetical genes predicted to be essential for *B. pseudomallei* from the TraDIS technique were selected for protein expression and purification. Previously, one target gene; *BPSL2774*, had been successfully amplified and cloned from genomic DNA of *B. pseudomallei* strain K96243 (Drahaman et al., 2016). BPSL2774 protein was expressed in bacterial cells and the soluble phase was utilized for protein purification. Affinity binding tests were performed to confirm expression and solubility. The three-dimensional structure

of BPSL2774 protein was modelled and its active sites predicted to aid in future functional experiments to validate its function as a glycosyltransferase.

## MATERIALS AND METHODS

### Bacterial Cell Culture and Lysis

The cells were cultured using the auto-induction method described by Studier in his 2005 report (Studier, 2005). The gene of interest coding for the target protein had been obtained from genomic DNA of *B. pseudomallei* K96243, cloned into Gateway™ pDEST15 (GST-tagged *BPSL2774*) and transformed into *Escherichia coli* DH5α maintenance strain, which had been kept in glycerol stock at -80°C. Prior to protein expression and purification, the cloned gene was transformed into expression strain, *E. coli* BL21(DE3) competent cell (Life Technologies). The colony was inoculated in a minimal non-inducing medium MDG (25 mM $Na_2HPO_4$, 25 mM $KH_2PO_4$, 50 mM $NH_4Cl$, 5 mM$Na_2SO_4$, 2 mM $MgSO_4$ , 0.1 × trace metals, 0.5% glucose, 0.25% aspartate) added with 100 µg/mL ampicillin before being incubated in 37°C at 200 rpm (approximately 24 *g*), overnight. MDG medium was further inoculated in complex auto-induction medium ZYM-5052 (1% N-Z-amine AS, 0.5% yeast extract, 25 mM $Na_2HPO_4$, 25 mM $KH_2PO_4$, 50 mM $NH_4Cl$, 5 mM $Na_2SO_4$, 2 mM $MgSO_4$, 1 × trace metals, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose) added with 100 µg/mL ampicillin and incubated in 37°C at 200 rpm (approximately 24 × *g*) for 4 hours, followed by at 18°C at 200 rpm, overnight. Then, the cell was harvested by centrifugation at 900 × *g* for 30 minutes at 4°C. The pellet was re-suspended in cold lysis buffer (50 mM Tris-HCl pH8, 100 mM NaCl, 1 mM EDTA). The suspension was sonicated (Omni Sonic Ruptor 400) with sonication condition of 6 × 30 Seconds / 60 Seconds off, Amplitude 40%. After that, the lysed cell was centrifuged at 4500 × *g* for 60 minutes at 4°C.

### Protein Purification

The soluble supernatant fraction was purified using GST Fusion Protein Purification Kit (Genscript). The supernatant was filtered through a 0.25 µm filter fitted to a syringe. 500 µL of settled glutathione resin (1 mL 50% resin slurry) was equilibrated four times with Buffer A (50 mM Tris-HCl pH8, 100 mM NaCl, 1 mM EDTA). The resin was then transferred to a new tube and mixed with the filtered supernatant. The supernatant-resin mixture was incubated for 30 minutes on ice. After 30 minutes, the mixture was transferred into a mini spin column 500 µL at a time and centrifuged at 500 × *g* for 5 seconds. All the flow-through was collected in a sterile tube. This step was repeated until all the supernatant-resin mixture was centrifuged, leaving behind the resin suspension on the spin column. The column containing the resin was then washed 20 times with 400 µL of Buffer A at a time by centrifugation at 500 × *g* for 5 seconds, with each elution fractionated at 800 µL in volume. Finally, the column was eluted 20 times with 400 µL of Buffer B

(50 mM Tris-HCL pH 8.08, 100 mM NaCl, 1 mM EDTA, 2 mM βME, 10 mM reduced glutathione) by centrifugation at 500 × *g* for 5 seconds (each elution fractionated at 800 µL in volume). All flow-through fractions were collected and stored in -20°C for SDS-PAGE analysis. Protein concentrations in each elution fractions were determined using Nanodrop 2000c (Thermo Fisher Scientific).

### Bioinformatics – Homology Identification and Domain Analysis

BPSL2774 protein sequence was retrieved from UniProtKB server (UniProtKB ID Q63R99) (Apweiler et al., 2004). The searches for related protein sequences were conducted using the National Center for Biotechnology (NCBI) with Basic Local Alignment Search Tool (BLAST) against Protein Data Bank (PDB) to find regions of sequence similarity for the functional and evolutionary descriptions (Altschul et al., 1990). Subcellular localisation of BPSL2774 was determined using WoLF PSORT (Horton et al., 2007), Yloc (Briesemeister et al., 2010) and TargetP (Emanuelsson et al., 2000) prediction tools. A trans-membrane helixes Hidden Markov Model (HMM)-based prediction tool, TMHMM (Krogh et al., 2001), was used to predict the presence or absence of trans-membrane domains in BPSL2774 protein. VICMpred tool (Saha & Raghava, 2006) and MP3 tool (Gupta et al., 2014) were used to predict the virulence and pathogenicity of BPSL2774 protein.

### Structure Prediction, Model Quality Assessment and Active Site Determination

Secondary structure and the three-dimensional structure of the protein were predicted by using I-TASSER (Roy et al., 2010; Yang et al., 2015) and SWISS-MODEL Workspace softwares (Waterhouse et al., 2018). The quality of the predicted structure was determined by the C-score calculation in I-TASSER server and QMEAN Z-score calculation in the QMEAN server (Benkert et al., 2010) available at SWISSMODEL Workspace, as well as verify3D of the UCLA-DOE Structure Evaluation server (Eisenberg et al., 1997). Model structure refinement was performed using ModRefiner algorithm tool (Xu & Zhang, 2011) from I-TASSER database. Ramachandran Plot assessment of the protein 3D model was performed using RAMPAGE server (Lovell et al., 2003). Active site and ligand binding site of BPSL2774 protein was predicted by using COFACTOR and COACH server based on the I-TASSER structure prediction from the I-TASSER website. Besides, metaPocket 2.0 server (Zhang et al., 2011) was used to determine the active site of BPSL2774 protein.

## RESULTS AND DISCUSSION

### Protein Purification and Expression

The transformed GST-tagged BPSL2774 construct was expressed from *E. coli* BL21(DE3) expression strain using auto-induction method for protein purification and expression screening. After lysis of

the sedimented cell pellet, the supernatant containing GST-tagged BPSL2774 protein was purified using GST fusion protein purification kit (Genscript) to isolate the GST-tagged protein. In this study, two cultures of 500 mL and 1 L respectively, of *E. coli* BL21(DE3) expression strain were induced in the auto-induction phase for expression of the target GST-tagged BPSL2774 protein and purified using GST-tagged affinity chromatography. All flow-through fractions, pellet, and resin from the purification step were retained for analysis with SDS-PAGE (Figures 1 and 2).

The combined molecular weight of the target protein (35.1 kDa) and the GST tag (25.5 kDa) is 60.6 kDa. Both SDS-PAGE gels showed the presence of intense band between 60 kDa and 80 kDa marker for Elution 1 and Elution 2 from the 500 mL culture preparation and for Elution 1 and Elution 5 from the 1 L culture preparation.

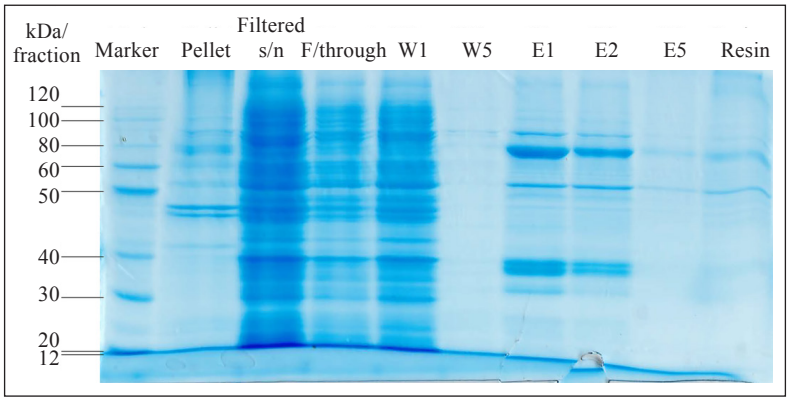


*Figure 1.* SDS-PAGE for fractions from 500 mL culture of *E. coli* BL21(DE3) expression strain. Marker: ProteinRuler® II (12-120 kDa) (Transgen Biotech Co., Ltd.), Pellet: sedimented lysed cells, Filtered s/n: filtered supernatant, F/through: flow-through, W1: first wash fraction, W5: final wash fraction, E1: first elution fraction, E2: elution fraction 2, E5: final elution fraction, Resin: resin sample

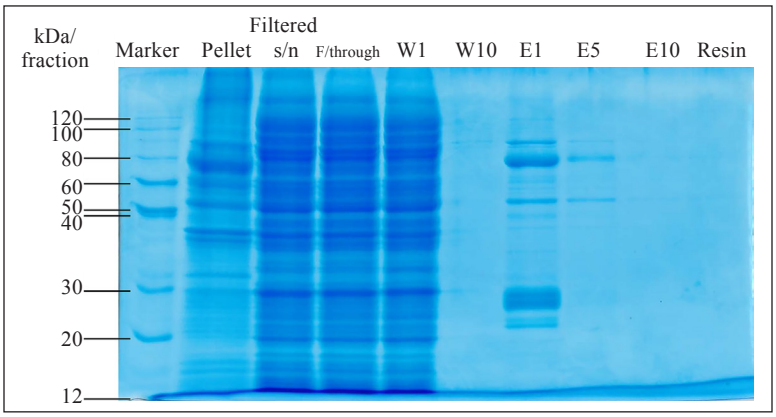


*Figure 2.* SDS-PAGE for fractions from 1 L culture of *E. coli* BL21(DE3) expression strain. Marker: ProteinRuler® II (12-120 kDa) (Transgen Biotech Co., Ltd.), Pellet: sedimented lysed cells, Filtered s/n: filtered supernatant, F/through: flow-through, W1: first wash fraction, W10: final wash fraction, E1: first elution fraction, E5: elution fraction 5, E10: final elution fraction, Resin: resin sample

This indicated a promising possibility that there was a relatively significant soluble overexpression of GST-tagged BPSL2774 protein. This band was then cut from the SDS-PAGE gel and sent for mass spectrometry analysis for confirmation. This band was also observed in the insoluble pellet fraction, markedly for the 1 L culture preparation (Figure 2). This may indicate that at a larger scale protein preparation, there is a tendency for the recombinant GST-tagged BPSL2774 to become insoluble or highly aggregated (otherwise known as inclusion bodies) in *E. coli* host. Future attempts to purify this insoluble fraction of the expressed protein can be made in the denatured form using detergent and refolding method (Yang et al., 2011).

Other protein bands were also observed in the SDS-PAGE gels. Two neighbouring bands at 50 kDa and 80 kDa, as well as two other distinct bands near the 30 kDa mark were observed. These contaminating bands can be deduced to have been eluted together with the target protein as they were not observed in the wash fractions wells. We suspect that these may be contaminating proteins from the host cells, *E. coli* BL21(DE3) that was co-purified with the target protein due to their interaction with BPSL2774. In the future, purification process needs to be followed with several more purification steps e.g. ion exchange chromatography followed by gel filtration to aid in the removal of these contaminants.

For both cultures, high concentrations of the target protein were obtained. The highest protein concentration was for the first 800 µL elution fraction (E1) at 1382 µg/mL from the 500 mL culture, and from the 1 L culture the highest concentration was obtained for the first 800 µL elution fraction (E1) at 910 µg/mL.

## Mass Spectrometry of Purified Samples from SDS-Page Gel

The Coomassie-stained protein gel band near the 60 kDa mark was cut and sent to First BASE Laboratories Sdn. Bhd. for further analysis using mass spectrometry.

Mass spectrometry analysis of the purified samples in SDS-PAGE confirmed the presence of the target protein, BPSL2774 as the third hit with a score of 390 and 27% protein sequence coverage (Figures 3 and 4 respectively). By performing BLASTp search to non-redundant database using the mass spectrometry result, it showed 100% sequence identity to BPSL2774 hypothetical protein from *B. pseudomallei* K96243 (NCBI Reference Sequence: WP 004550046.1) (Figure not shown).

The two upper hits found to be present at 60 kDa were *E. coli* chaperonin GroL protein and *E. coli* chaperonin GroEL protein, with less protein sequence coverage, at 14% and 9% respectively. In protein synthesis, molecular chaperones are commonly present to interact with new proteins as they form their final structure (Rosano & Ceccarelli, 2014). As the target protein was purified from *E. coli* expression system, it could be expected that some protein chaperones from the host were co-purified together with the target protein. In order to obtain the target protein in high purity, further purification steps would be required (e.g. GST-tag removal, ion exchange chromatography followed by gel filtration).

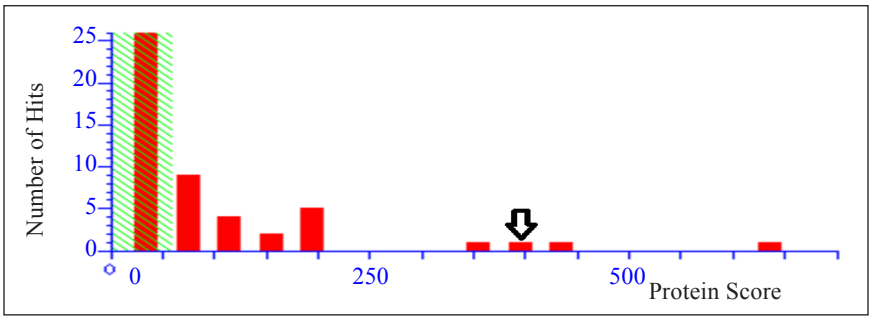


*Figure 3.* Mascot Score histogram of mass spectrometry analysis. Ions score is 10*Log(P), where P is the probability that the observed match is a random event. Individual ions scores > 59 indicate identity or extensive homology ($p<0.05$). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits. The target protein BPSL2774 score at 390 (labeled with an arrow) is the third histogram from the right of graph

3. Q63R99 Mass: 35102 Score: 390 Matches: 7(3) Sequences: 7(3)
tr|Q63R99|BPSL2774 Q63R99_BURPS Uncharacterized protein n=3 Tax_Id=272560 [Burkholderia pseudomallei (strain K96243)]
Check to include this hit in error tolerant search or archive report

| Query | Observed | Mr(expt) | Mr(calc) | Delta | Miss | Score | Expect | Rank | Unique | Peptide |
|---|---|---|---|---|---|---|---|---|---|---|
| 13 | 949.6718 | 948.6645 | 948.4818 | 0.1828 | 0 | 37 | 12 | 1 | U | R.FVIHDYR.S |
| 18 | 1026.7838 | 1025.7765 | 1025.5770 | 0.1995 | 0 | 27 | 96 | 1 | U | R.AWNAKPALR.I |
| 22 | 1112.8352 | 1111.8279 | 1111.6138 | 0.2141 | 0 | 62 | 0.03 | 1 | U | R.HVEGLVFVGR.V |
| 25 | 1140.7867 | 1139.7794 | 1139.5611 | 0.2183 | 0 | 59 | 0.065 | 1 | U | K.ASYPEIAAYR.D |
| 35 | 1480.9202 | 1479.9129 | 1479.6379 | 0.2750 | 0 | 99 | 5.2e-06 | 1 | U | R.GFDVHDGDASAYAR.F |
| 52 | 2037.4723 | 2036.4650 | 2036.1055 | 0.3596 | 0 | 61 | 0.0091 | 1 | U | R.SHVTGAAIFDELLPLAQVR.A |
| 54 | 2062.4114 | 2061.4041 | 2061.0419 | 0.3623 | 0 | 44 | 1 | 1 | U | R.DSLVLVGQPEPEIYEAFR.H |

*Figure 4.* Peptide summary report from mass spectrometry analysis for the third hit (labelled Q63R99) showing the presence of target protein BPSL2774

## Homology Identification and Domain Analysis

Through BLASTP search of the PDB database for BPSL2774, the protein is shown to have conserved domains of glycosyltransferase GTB type superfamily and RfaB superfamily (glycosyltransferase involved in cell wall biosynthesis). Glycosyltransferases catalyse the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds (Breton et al., 2006). It has 11% sequence similarity to two glycosyltransferase enzymes from *Bacillus anthracis* strains (PDB ID: 3MBO A and 2JJM A). Likewise, BPSL2774 has also been included and mapped into the Burkholderia Ortholog Group #BG016035 (downloadable listing of the group members is available at http://www.burkholderia.com/orthologs/list?id=374551) in which glycosyltransferases across *Burkholderia* species or strains are grouped together (Winsor et al., 2008). The target protein is shown to have conserved domains of glycosyltransferase GTB type superfamily (Figure 5).

From WoLF PSORT, Yloc, TargetP and TMHMM prediction tools, BPSL2774 protein was predicted to be located within the cytoplasm with no trans-membrane domain evident. It is also predicted to

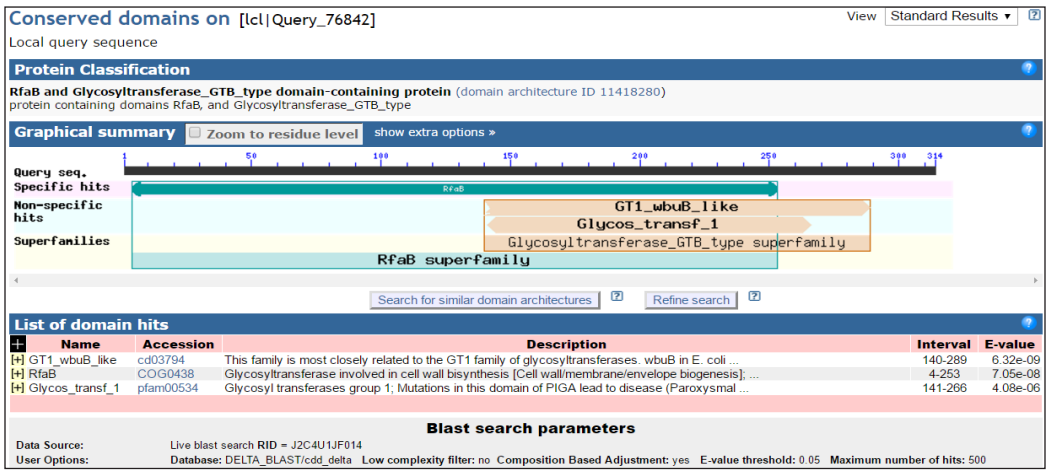


*Figure 5.* Conserved domain found on BPSL2774 through DELTA-BLAST search of the PDB database

be involved in metabolic processes by VICMpred and is not involved in the pathogenicity of *B. pseudomallei* by MP3 software (data not shown).

## Three-Dimensional Model of BPSL2774

The initial model of BPSL2774 from I-TASSER server has a C-score of 0.13, with a relatively average confidence in the quality of the predicted model (Figure 6). Two ModRefiner refinement runs were performed to refine the initial model. Figure 7 showed the Ramachandran plot of the initial model and the final refined model. The final Ramachandran Plot Statistics from RAMPAGE server showed 92.3% residues in the most favoured regions, with two residues in the disallowed regions (ARG220 and GLY299). An estimated 89.8% of the amino acid had an averaged 3D-1D score of more than 0.2 using Verify3D Plot (figure not shown), indicating that the environment profile of the model was acceptable. The QMEAN (Qualitative Model Energy ANalysis) scoring function provided an estimate of the quality of the model with QMEAN6 score of 0.62 (Figure 8). The QMEAN6 score range from 0 to 1, with one being considered to be a model of good quality (Benkert et al., 2010).

The generated model of the BPSL2774 protein tertiary structure from I-TASSER was found to have two domains which is similar to GT-B type superfamily of glycosyltransferase. It consists of two β/α/β Rossmann fold domains with six parallel beta strands found in each domain and the two domains are less tightly associated (Lairson et al., 2008). Only two structural folds, GT-A and GT-B, have been identified for the nucleotide sugar-dependent enzymes, but other folds are now appearing for the soluble domains of lipid phosphosugar-dependent glycosyltransferases (Lairson et al., 2008). They play essential roles in biosynthesis pathways of oligo- and poly-saccharides, as well as protein glycosylation and formation of valuable natural products (Schmid et al., 2016).

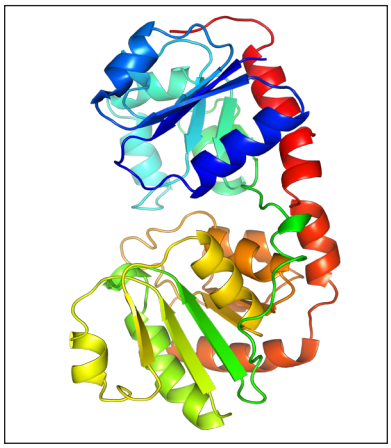

*Figure 6*. Predicted three-dimensional structure of BPSL2774, with N-terminal coloured blue and C-terminal coloured red

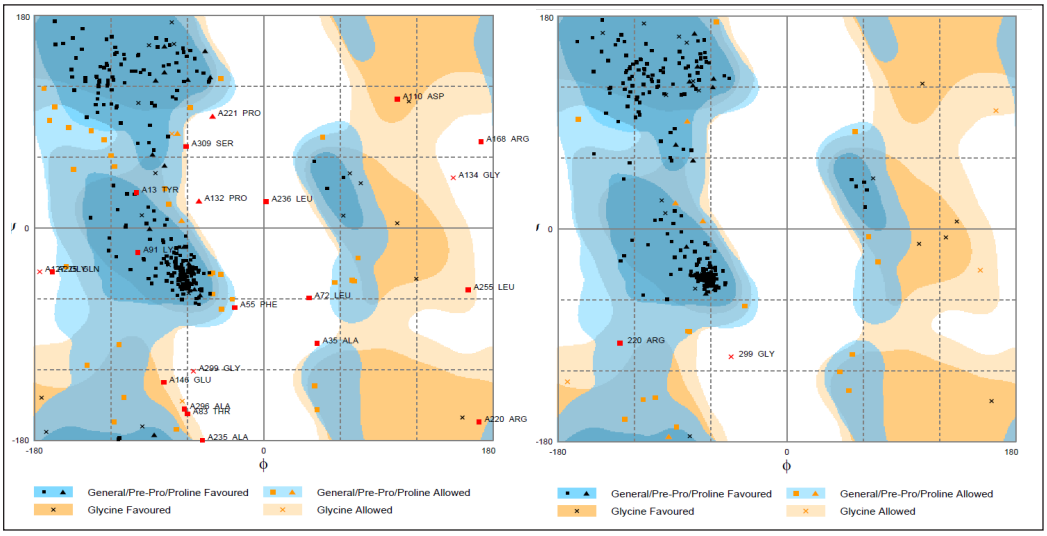


*Figure 7*. Ramachandran plot analysis of modeled structure of the original model (left) and the final refined model (right)

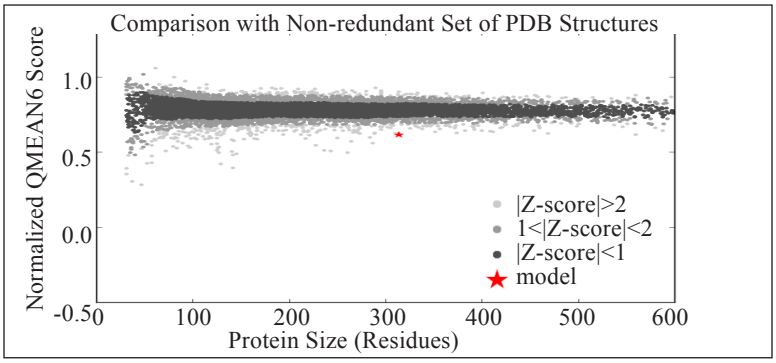


*Figure 8*. Graphical presentation of estimation of absolute quality of model with QMEAN

## Active Site Prediction

The active site of the protein as predicted by metaPocket server is as shown in Figure 9. The top two metaPocket clusters were shown in Figure 9, with the first cluster having a Z-score of 23.82 of six pocket sites and the second cluster with two pocket sites and a Z-score of 1.35. I-TASSER server suite provides biological annotation of the target protein by COFACTOR and COACH programs. The two top predictions for BPSL2774 for ligand binding were uridine diphosphate (UDP) for a glycosyltransferase function (based on its PDB hit, 3mboE, a glycosyltransferase from *B. anthracis*) with the highest C-score of 0.33. This glycosyltransferase structure from *B. anthracis* is involved in bacillithiol (a novel low-molecular-weight thiol) biosynthetic pathway (Parsonage et al., 2010). The second highest C-score of 0.25 was N-acetylglucosamine (NAG) ligand, derived from clusters of PDB hit to 5e9uA, which is also a glycosyltransferase from *Streptococcus gordonii* (Figure 10, Table 1). This glycosyltransferase is involved in O-glycosylation reactions, which has a critical role for biogenesis and modification of adhesins in *streptococci* and *staphylococci* bacteria (Chen et al., 2016). In general,

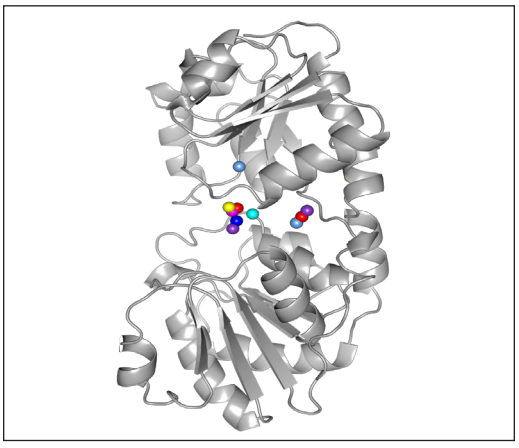

*Figure 9.* Active sites of the predicted 3D structure of the target protein as determined by metaPocket server. The color of the spheres indicated the active site predicted by MetaPocket method (red ball), PASS method (actinium ball), LIGSITE method (magenta ball), FPocket method (potassium ball), SURFNET method (wheat ball), GHECOM method (yellow ball) and ConCavity method (blue ball)

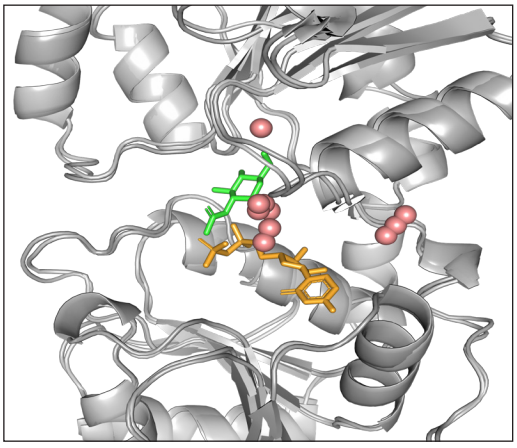

*Figure 10.* Predicted ligands that bind to the predicted active site by I-TASSER, and superimposed on active sites prediction spheres from metaPocket server (spheres colored pink). Model structure is colored gray, UDP ligand is colored orange, and NAG ligand is colored green

Table 1
*Amino acid residues for predicted ligand binding sites from I-TASSER prediction suite*

| Ligand | PDB Hit | Ligand Binding Site Residues |
|---|---|---|
| UDP | 3mboE | 144,145,146,151,197,198,199,202,227,228,229,232 |
| NAG | 5e9uA | 67,68,119,223,224,225,226,227,228 |

models with C-score > −1.5 have a correct fold (Roy et al., 2010). The results from both metaPocket and I-TASSER active sites prediction corresponds well and located in the crevice between the two Rossman fold domains.

## Glycosyltransferase Function in Pathogenic Bacteria

Glycosyltransferases catalyze glycosidic bond formation using sugar donors containing a nucleoside phosphate or a lipid phosphate leaving group (Breton et al., 2006). Most commonly, the donor sugar substrate is activated in the form of nucleoside diphosphate sugars e.g. UDP galactose or GDP mannose. However, other forms such as nucleoside monophosphate sugars, lipid phosphates, and unsubstituted phosphates are also used. Glycosyl transfer frequently occurs to the nucleophilic oxygen of a hydroxyl substituent of the acceptor, but it can also occur to nitrogen nucleophiles, such as in formation of N-linked glycoproteins; sulfur nucleophiles, such as in formation of thioglycosides in plants; and also in carbon nucleophiles, such as in C-glycoside antibiotics (Lairson et al., 2008) The product of this reaction may be a growing oligosaccharide, a lipid, or a protein (Breton et al., 2006).

Some pathogenic *Streptococcus* and *Pasteurella* bacteria have capsules that contain nonimmunogenic hyaluronan, which protects them against a mammalian host's immune system. The mammalian body possesses an abundance of hyaluronan, which means that any significant response against the bacterial hyaluronan capsule could cause widespread autoimmune complications in the host. Hyaluronan is synthesized by the glycosyltransferase hyaluronan synthase (DeAngelis, 1999).

In Gram negative bacteria, glycosyltransferases are valuable in the formation of lipolysaccharides (LPS), i.e. the major cell-surface component protecting the bacterium from extracellular threats (Cote & Taylor, 2017). In Gram negative *Neisseria meningitidis* and *Haemophilus influenzae*, small molecular inhibitors of LgtC glycosyltransferase are being analysed as potential anti-virulence drug candidates (Xu et al., 2018). The *B. pseudomallei* genome carries four large polysaccharide loci, all of which have been demonstrated to play a role in virulence in vivo; these encode the type I O antigen polysaccharides (O-PS) capsule, the type II O-PS LPS, and two additional clusters defined as type III O-PS and type IV O-PS (Moule et al., 2016).

In other pathogenic bacteria such as *Escherichia coli*, *Salmonella enterica* and *Shigella dysenteriae*, glycosyltransferases along with other proteins play significant roles to ensure their survival in mammalian host, namely in the modification processes of bacteria's protein, enterobactin. This protein can successfully compete for iron, Fe (II) binding against the host, which is important to the pathogen's ultimate survival (Fishbach et al., 2006).

Three genes predicted to be glycosyltransferases in *B. pseudomallei; BPSS2167, BPSS2248* and *BPSL1444* have recently been identified as newly discovered

genes involved in in vivo virulence with roles in different stages of *B. pseudomallei* pathogenesis, including extracellular and intracellular survival (Moule et al., 2016). In the same 2016 paper, Moule suggested that "the role of polysaccharides in *B. pseudomallei* infections is even more complex than has been previously described", as these three genes identified in their screen did not belong to any of the four polysaccharide clusters (Moule et al., 2016).

## CONCLUSION

The GST-tagged BPSL2774 target protein was able to be expressed in soluble form from high density cultures and has been partially purified using affinity chromatography. Higher protein purity can be achieved through further purification steps following the initial GST-tagged affinity chromatography. *Escherichia coli* chaperonin proteins from the *E. coli* host system was found to be co-purified along with the target protein. The purified protein however is at acceptable purity and at sufficient concentration for use as samples in functional assays, e.g. a fluorescence-based or bioluminescence-based glycosyltransferase assay in the near future. Due to the challenges to determine both the sugar donor and acceptor for a GT of unknown function, *in silico* approaches were performed to annotate the structure and function of BPSL2774 protein. The quality of the refined model was verified by using Ramachandran plot. Through preliminary docking runs using AutoDock 4.2 (Morris et al., 2009) and AutoDock Vina suite (Trott & Olson, 2010) on predicted ligands UDP and NAG to BPSL2774 structure model, satisfactory docking results were obtained (results not shown). It was worth to note that the bioinformatics structural and functional annotation predictions all pointed towards BPSL2774 functioning as a glycosyltransferase. Taking all the results into account, the functional annotation of BPSL2774 protein as a glycosyltransferase is recommended, though future validation from biochemical experiments or a more exhaustive docking simulation experiments will be needed to support this.

## ACKNOWLEDGEMENT

We would like to thank the IIUM Research Management Centre, all laboratory staff at Kulliyyah of Science, International Islamic University Malaysia and the Malaysian Ministry of Higher Education (MOHE). This research was funded by the Research Acculturation Grant Scheme (RAGS) Phase 1/2014 from the Ministry of Education (MOE) – Ref: RAGS/1/2014/SG05/UIAM//2 (RAGS 14-036-0099) and IIUM RIGS research grant (RIGS16-312-0476).

## REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410. doi: 10.1016/S0 022-2836(05)80360-2

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., ... Martin, M. J. (2004). UniProt: The universal protein knowledgebase. *Nucleic Acids*

*Research*, *32*(suppl_1), D115-D119. doi: 10.1093/nar/gkh131

Benkert, P., Biasini, M., & Schwede, T. (2010). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343-350. doi: 10.1093/bi oinformatics/btq662

Breton, C., Šnajdrová, L., Jeanneau, C., Koča, J., & Imberty, A. (2006). Structures and mechanisms of glycosyltransferases. *Glycobiology*, *16*(2), 29R-37R. doi: 10.1093/glyc ob/cwj016

Briesemeister, S., Rahnenführer, J., & Kohlbacher, O. (2010). YLoc—An interpretable web server for predicting subcellular localization. *Nucleic Acids Research*, *38*(suppl_2), W497-W502. doi: 10.1093/nar/gkq477

Chen, Y., Seepersaud, R., Bensing, B. A., Sullam, P. M., & Rapoport, T. A. (2016). Mechanism of a cytosolic O-glycosyltransferase essential for the synthesis of a bacterial adhesion protein. *Proceedings of the National Academy of Sciences*, *113*(9), E1190-E1199. doi: 10.1073/pnas.1600494113

Cheng, A. C., & Currie, B. J. (2005). Melioidosis: Epidemiology, pathophysiology, and management. *Clinical Microbiology Reviews*, *18*(2), 383-416. doi: 10.1128/CMR.18.2. 383-416.2005

Chewapreecha, C., Holden, M. T., Vehkala, M., Välimäki, N., Yang, Z., Harris, S. R., ... Bizet, C. (2017). Global and regional dissemination and evolution of *Burkholderia pseudomallei*. *Nature Microbiology*, *2*(4), 16263. doi: 10.1038/nmicrobiol.2016.263

Corea, E. M., de Silva, A. D., & Thevanesam, V. (2018). Melioidosis in Sri Lanka. *Tropical Medicine and Infectious Disease*, *3*(1), 22. doi: 10.3390/tropicalmed3010022

Cote, J., & Taylor, E. (2017). The glycosyltransferases of LPS core: A review of four heptosyltransferase enzymes in context. *International Journal of Molecular Sciences*, *18*(11), 2256. doi: 10.3390/ijms18112256

Dance, D. A., Luangraj, M., Rattanavong, S., Sithivong, N., Vongnalaysane, O., Vongsouvath, M., & Newton, P. N. (2018). Melioidosis in the Lao People's Democratic Republic. *Tropical Medicine and Infectious Disease*, *3*(1), 21. doi: 10.3390/tropicalmed3010021

DeAngelis, P. L. (1999). Hyaluronan synthases: Fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cellular and Molecular Life Sciences CMLS*, *56*(7), 670-682. doi: 10.1007/s000180050

Deris, Z. Z., Hasan, H., & Suraiya, M. N. S. (2010). Clinical characteristics and outcomes of bacteraemic melioidosis in a teaching hospital in a northeastern state of Malaysia: A five-year review. *The Journal of Infection in Developing Countries*, *4*(07), 430-435. doi: 10.3855/jidc.491

Drahaman, S. M., Raih, M. F., Bunnori, N. M., & Rehan, A. M. (2016). *Cloning and expression of hypothetical protein targets in* Burkholderia pseudomallei *by transposon-directed insertion site sequencing (TraDIS) technique.* Retrieved December 13, 2018, from http://conference.kuis.edu.my/icpr/img/16.docx

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, *277*, 396-404. doi: 10.1016/S0076-6879(97)77022-8

Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, *300*(4), 1005-1016. doi: 10.1006/jmbi.2000.3903

Fischbach, M. A., Lin, H., Liu, D. R., & Walsh, C. T. (2006). How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nature*

*Chemical Biology*, *2*(3), 132. doi: 10.1038/nchembio771

Gupta, A., Kapil, R., Dhakan, D. B., & Sharma, V. K. (2014). MP3: A software tool for the prediction of pathogenic proteins in genomic and metagenomic data. *PloS One*, *9*(4), e93907. doi: 10.1371/journal.pone.0093907

Holden, M. T., Titball, R. W., Peacock, S. J., Cerdeño-Tárraga, A. M., Atkins, T., Crossman, L. C. & Sebaihia, M. (2004). Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(39), 14240-14245. doi: 10.1073/pnas.0403302101

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, *35*(suppl_2), W585-W587. doi: 10.1093/nar/gkm259

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567-580.

Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyltransferases: Structures, functions, and mechanisms. *Annual Review of Biochemistry*, *77*, 521 – 555. doi: 10.1146/annurev.biochem.76.061005.092322

Limmathurotsakul, D., Golding, N., Dance, D. A., Messina, J. P., Pigott, D. M., Moyes, C. L., ... Hay, S. I. (2016). Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. *Nature Microbiology*, *1*(1), 1-5. doi: 10.1038/nmicrobiol.2015.8

Lovell, S. C., Davis, I. W., Arendall III, W. B., De Bakker, P. I., Word, J. M., Prisant, M. G., ... Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function and Bioinformatics*, *50*(3), 437-450. doi: 10.1002/prot.10286

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785-2791. doi: 10.1002/jcc.21256

Moule, M. G., Hemsley, C. M., Seet, Q., Guerra-Assunção, J. A., Lim, J., Sarkar-Tyson, M., ... Wren, B. W. (2014). Genome-wide saturation mutagenesis of *Burkholderia pseudomallei* K96243 predicts essential genes and novel targets for antimicrobial development. *MBio*, *5*(1), e00926-13. doi: 10.1128/mBio.00926-13

Moule, M. G., Spink, N., Willcocks, S., Lim, J., Guerra-Assunção, J. A., Cia, F., ... Bancroft, G. J. (2016). Characterization of new virulence factors involved in the intracellular growth and survival of *Burkholderia pseudomallei*. *Infection and Immunity*, *84*(3), 701-710. doi: 10.1128/IAI.01102-15

Nathan, S., Chieng, S., Kingsley, P., Mohan, A., Podin, Y., Ooi, M. H., ... How, S. H. (2018). Melioidosis in Malaysia: Incidence, clinical challenges, and advances in understanding pathogenesis. *Tropical Medicine and Infectious Disease*, *3*(1), 25. doi: 10.3390/tropicalmed3010025

Pande, K., Abd Kadir, K., Asli, R., & Chong, V. H. (2018). Melioidosis in Brunei Darussalam. *Tropical Medicine and Infectious Disease*, *3*(1), 20. doi: 10.3390/tropicalmed3010020

Parsonage, D., Newton, G. L., Holder, R. C., Wallace, B. D., Paige, C., Hamilton, C. J., & Claiborne, A. (2010). Characterization of the N-acetyl-α-d-glucosaminyl l-malate synthase and deacetylase functions for bacillithiol biosynthesis in *Bacillus anthracis*. *Biochemistry, 49*(38), 8398-8414. doi: 10.1021/bi100698n

Peacock, S. J., Limmathurotsakul, D., Lubell, Y., Koh, G. C., White, L. J., Day, N. P., & Titball, R. W. (2012). Melioidosis vaccines: A systematic review and appraisal of the potential to exploit biodefense vaccines for public health purposes. *PLoS Neglected Tropical Diseases*, *6*(1), e1488. doi: 10.1371/journal.pntd.0001488

Rhodes, K. A., & Schweizer, H. P. (2016). Antibiotic resistance in *Burkholderia* species. *Drug Resistance Updates*, *28*, 82-90. doi: 10.1016/j.drup.2016.07.003

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*, *5*(172), 1-17. doi: 10.3389/ fmicb.2014.00172

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725. doi: 10.1038/npro t.2010.5

Saha, S., & Raghava, G. P. S. (2006). VICMpred: An SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition. *Genomics, Proteomics and Bioinformatics*, *4*(1), 42-47. doi: 10.1016/S1672-0229(06)60015-6

Schmid, J., Heider, D., Wendel, N. J., Sperl, N., & Sieber, V. (2016). Bacterial glycosyltransferases: Challenges and opportunities of a highly diverse enzyme class toward tailoring natural products. *Frontiers in Microbiology*, *7*, 182. doi: 10.3389/fmicb.2016.00182

Schweizer, H. P. (2012). Mechanisms of antibiotic resistance in *Burkholderia pseudomallei*: Implications for treatment of melioidosis. *Future Microbiology*, *7*(12), 1389-1399. doi: 10.2217/fmb.12.116

Sim, S., Ong, C., Gan, Y., Wang, D., Koh, V., Tan, Y., ... Ye, A. (2018). Melioidosis in Singapore: Clinical, veterinary, and environmental perspectives. *Tropical Medicine and Infectious Disease*, *3*(1), 31. doi: 10.3390/tropicalmed3010031

Stevens, M. P., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., & Galyov, E. E. (2002). An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Molecular Microbiology*, *46*(3), 649-659. doi: 10.1046/j.1365-2958.2002.03190.x

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, *41*(1), 207-234. doi: 10.1016/j.pep.2005.01.016

Tauran, P. M., Wahyunie, S., Saad, F., Dahesihdewi, A., Graciella, M., Muhammad, M., ... Pratiwi, D. I. N. (2018). Emergence of melioidosis in Indonesia and today's challenges. *Tropical Medicine and Infectious Disease*, *3*(1), 32. doi: 10.3390/tropicalmed3010032

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455-461. doi: 10.1002/jcc.2 1334

Turner, P., Kloprogge, S., Miliya, T., Soeng, S., Tan, P., Sar, P., ... Turner, C. (2016). A retrospective analysis of melioidosis in Cambodian children, 2009–2013. *BMC Infectious Diseases*, *16*(1), 688. doi: 10.1186/s12879-016-2034-9

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... Lepore, R. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296-W303. doi: 10.1093/nar/gky427

Wiersinga, W. J., Van der Poll, T., White, N. J., Day, N. P., & Peacock, S. J. (2006). Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. *Nature Reviews*

*Microbiology*, *4*(4), 272. doi: 10.1038/ nrmicro1385

Winsor, G. L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M. D., & Brinkman, F. S. (2008). The Burkholderia Genome Database: Facilitating flexible queries and comparative analyses. *Bioinformatics*, *24*(23), 2803-2804. doi: 10.1093/bioinformatics/ btn524

Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, *101*(10), 2525-2534. doi: 10.1016/j.bpj.2011.10.024

Xu, Y., Cuccui, J., Denman, C., Maharjan, T., Wren, B. W., & Wagner, G. K. (2018). Structure-activity relationships in a new class of non-substrate-like covalent inhibitors of the bacterial glycosyltransferase LgtC. *Bioorganic and Medicinal Chemistry*, *26*(11), 2973-2983. doi: 10.1016/j.bmc.2018.03.006

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7. doi: 10.1038/nmeth.3213

Yang, Z., Zhang, L., Zhang, Y., Zhang, T., Feng, Y., Lu, X., ... Wang, X. (2011). Highly efficient production of soluble proteins from insoluble inclusion bodies by a two-step-denaturing and refolding method. *PloS One*, *6*(7), e22981. doi: 10.1371/journal.pone.002 2981

Zhang, Z., Li, Y., Lin, B., Schroeder, M., & Huang, B. (2011). Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics*, *27*(15), 2083-2088. doi: 10.1093/bioinformatics/btr331