COPYRIGHT© INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

# EXPRESSION, PURIFICATION AND *IN SILICO* CHARACTERIZATION OF BPSL2774, A HYPOTHETICAL PROTEIN FROM *Burkholderia pseudomallei* K96243

BY

SITI MARHAMAH BINTI DRAHAMAN

A thesis submitted in fulfillment of the requirement for the degree of Master of Science (Biotechnology)

Kulliyyah of Science
International Islamic University Malaysia

DECEMBER 2018

# ABSTRACT

Melioidosis is a disease that infects human and animals and can be detrimental in humans. Mortality rate from melioidosis septic shock due to infection from gram-negative *Burkholderia pseudomallei* in endemic regions of Malaysia and Thailand remain high despite available antimicrobial therapy. Different strategies are being utilized to identify essential genes and drug targets in this bacterium for improvement in current antimicrobial therapies. This particular concern is due to the resistance of *B. pseudomallei* to many available and commercial antibiotics as well as the lack exposure about the pathogenicity of this bacterium. In this study, five target genes predicted to be essential for *B. pseudomallei* by transposon-directed insertion site sequencing (TraDIS) technique were selected and amplified using nested polymerase chain reaction (PCR) for subsequent Gateway™ cloning protocols. Currently, positive clones have been verified for one target gene, *BPSL2774* using colony PCR, *Bsr*G1 restriction digest and deoxyribonucleic acid (DNA) sequencing. The essential gene *BPSL2774*, obtained from *B. pseudomallei* K96243 were expressed in *Escherichia coli* BL21(DE3) for protein production. Large scale protein preparations in high density cultures were made according to the auto-induction method to express the soluble target protein. The target protein was successfully separated and expressed in soluble form by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Mass spectrometry analysis shown the soluble BPSL2774 protein was successfully expressed with the correct mass of 35102 kDa. After confirmation of the purified protein identity, *in silico* structural and functional prediction on BPSL2774 protein was performed. Secondary and tertiary structure of BPSL2774 protein was predicted. BLASTp to protein databank (PDB) database showed that BPSL2774 protein have conserved domains of glycosyltransferase GT-B type superfamily. This correlates with the secondary and tertiary structure model of BPSL2774 that displayed two β/α/β Rossmann fold domains with six parallel beta strands found in each domain, indicative of the same fold. By using consensus approach (COACH) meta-server, the top prediction for ligand binding was α-D-glucose, Uridine diphosphate (UDP) and N-Acetylglucosamine (NAG). The refined structural model of BPSL2774 protein validated by Ramachandran plot was used for docking simulations with UDP, GDP and NAG. The docking results from both AutoDock4.2 and AutoDock Vina did not show significant binding affinity of the three tested ligands to *BPSL2774* protein, with binding affinity values ranging from -4.4 kcal/mol to -6.9 kcal/mol. This was consistent with the challenges of characterizing glycosyltransferases due to the various sugar donor and acceptor specificity. 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 were needed to support this. In the future, BPSL2774 protein can be further purified through additional purification steps following initial GST-tagged Affinity Chromatography for subsequent functional assay and biophysical experiments.

# خلاصة البحث

مرض الراعوم أو الميليويدوسيس هو مرض يصيب الإنسان والحيوان وبإمكانه أن يكون ضارًا بالبشر. لا يزال معدل الوفيات بسبب الصدمة الإنتانية للراعوم بسبب عدوى بكتيريا بيركولديريا سودومالي سالبة الجرام في المناطق المستوطنة في ماليزيا وتايلاند مرتفعا على الرغم من توفر العلاج المضاد للميكروبات، ولتحسين العلاجات المضادة للميكروبات الحالية يتم استخدام استراتيجيات مختلفة لتحديد الجينات والأدوية المستهدِفة في هذه البكتيريا. يرجع هذا القلق بشكل خاص إلى مقاومة بكتيريا بيركولديريا سودومالي للعديد من المضادات الحيوية المتاحة والتجارية، بالإضافة إلى قلة المعلومات المتعلقة بالقدرة الإمراضية لهذه البكتيريا. تم في هذه الدراسة اختيار خمسة جينات مستهدفة أساسية لبكتيريا بيركولديريا سودومالي بواسطة طريقة تسلسل موقع الإدخال الموجَّه بالجينات القافزة (TraDIS) وتضخيمها باستخدام تفاعل البلمرة المتسلسل المتداخل (PCR) لبروتوكولات استنساخ Gateway ™ التالية. تم حاليا التحقق من وجود مستنسخات إيجابية لجين مستهدف واحد وهو BPSL2774 باستخدام PCR المستعمرات، وهضم القيد لـ *BisrG1*، وتسلسل الحمض النووي الريبي (DNA). تم التعبير عن الجين الأساسي BPSL2774 الذي تم الحصول عليه من بيركولديريا سودومالي K96243في الإشريكية القولونية BL21 (DE3) لإنتاج البروتين. تم تحضير مستحضرات البروتين على نطاق مضخم في مستنبتات عالية الكثافة وفقًا لطريقة الحث التلقائي للتعبير عن البروتين المستهدف القابل للذوبان. تم فصل البروتين المستهدف بنجاح وتم التعبير عنه بشكل قابل للذوبان عن طريق الفصل الكهربائي لهلام كبريتات دوديكل الصوديوم متعددالأكريلامايد (SDS-PAGE). أظهر تحليل الطيف الكتلي أن البروتين BPSL2774 القابل للذوبان تم التعبير عنه بنجاح بكتلة صحيحة بلغت 35102 كيلو دالتون. بعد التأكد من هوية البروتين المنقى، تم إجراء التنبؤ البنيوي والوظيفي على البروتين BPSL2774. تم أيضا التنبؤ بالبنية الثانوية والثالثية لبروتين BPSL2774. أظهر البحث في قاعدة بيانات البروتينات (PDB) تحت BLASTp أن لدى البروتين BPSL2774 مجالات محمية للفصيلة العليا من نوع جليكوترانسفيريز GT-B، والذي يرتبط بنموذج البنية الثانوية والثالثية لـ BPSL2774 التي أظهرت اثنين من مجالات طية Rossmann β/α/β مع ستة فروع بيتا متوازية في كل مجال، مما يدل على نفس الطية. من خلال استخدام طريقة التوافق (COACH) ميتا-سيرفر، كان التنبؤ الأعلى لاتصال الربيطة هو جلوكوز-α-D، واليوريدين ثنائي الفوسفات (UDP)، و N-أسيتيلجلوكوزامين (NAG). تم استخدام النموذج الهيكلي المكرر لبروتين BPSL2774 الذي تم التحقق منه من خلال مخطط Ramachandran في محاكات الاتصال البروتينية مع UDP و GDP و NAG. لم تظهر نتائج الاتصال في كل من برامج AutoDock4.2 و AutoDock Vina أي تقارب اتصال ملحوظة للربيطات الثلاثة المختبرة لبروتين BPSL2774، وذلك بقيم تقارب اتصال تراوحت من -4.4 كيلو كالوري/مول إلى -6.9 كيلو كالوري/مول. كان هذا في اتساق مع تحديات توصيف جلوكوسيلترانسفيريز بسبب خصوصية التقبل والإعطاء المتنوعة للسكر. مع أخذ جميع النتائج في الحسبان، فإنه يوصى بوضع شرح وظيفي لبروتين BPSL2774 كجلوكوسيلترانسفيراز، على الرغم من أن التحقق من التجارب البيوكيميائية أو تجارب المحاكاة الأكثر عمقا ضرورية لدعم ذلك مستقبلا. بالإمكان لاحقا تنقية بروتين BPSL2774 بشكل أكثر باتباع خطوات التنقية الإضافية بعد التحليل الكروماتوغرافي التماثلي المعلَّم بالـ GST للاختبارات الوظيفية اللاحقة والتجارب البيوفيزيائية.

# APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology)

……..
'Aisyah binti Mohamed Rehan
Supervisor

……..
Noraslinda binti Mohamad Bunnori
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology)

……..
Noor Hasniza binti Md Zin
Internal Examiner

……..
Mohd Nazalan bin Mohamad Najimudin
External Examiner

This thesis was submitted to the Department of Biotechnology and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology)

……..
Mardiana binti Mohd Ashaari
Head, Department of Biotechnology

This thesis was submitted to the Kulliyyah of Science and is accepted as a fulfillment of the requirement for the degree of Master of Science (Biotechnology)

……..
Shafida binti Abdul Hamid
Dean, Kulliyyah of Science

# DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Siti Marhamah binti Drahaman

Signature ..... Date .....

**INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA**

**DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH**

**EXPRESSION, PURIFICATION AND *IN SILICO* CHARACTERIZATION OF BPSL2774, A HYPOTHETICAL PROTEIN FROM *Burkholderia pseudomallei* K96243**

I declare that the copyright holders of this thesis are jointly owned by the student and IIUM.

Copyright © 2017 Siti Marhamah binti Drahaman and International Islamic University Malaysia. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may be used by others in their writing with due acknowledgement.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.

3. The IIUM library will have the right to make, store in a retrieved system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Siti Marhamah binti Drahaman

……..…….. ……..
Signature Date

*My humble effort I dedicate to my loving and supportive parents & parents in law*

*Abah & Ma*
*Abah & Umi*

*Along with all understanding and endurance husband & son*

*Feruza Ashraff & Fathurrahman*

*May this bring good in any way*

# ACKNOWLEDGEMENTS

All praises to Allah, The Almighty, that with His blessings and wills, I am able to complete this project and thesis. I have engaged with many people throughout the process.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Aisyah' binti Mohamed Rehan and my co-supervisor, Dr. Noraslinda binti Mohamad Bunnori for their never-ending support and guidance throughout the project and thesis writing process despite their packed schedule. Also, I would like to express my appreciation to the examiners, for their evaluation and comments on the thesis.

I would like to thank to the collaborator of this project, Dr. Mohd Firdaus Bin Raih and his collagues from Universiti Kebangsaan Malaysia, UKM for their consultation and not to forget to Ministry of Higher Education, MOHE for the funds.

My special thanks go to my colleagues for all the helps, support and motivation in various aspects, either directly or indirectly. Without them, this project would not complete.

Last but not least, I would like to extend my appreciation and gratitude to my family, especially both my parents and my husband, for their faith in me, their prayers, endurance and support which have led to the completion of the project and thesis.

# TABLE OF CONTENTS


Abstract ..... ii
Abstract in Arabic ..... iii
Approval Page ..... iv
Declaration Page ..... v
Copyright Page ..... vi
Dedications ..... vii
Acknowledgements ..... viii
Table of Contents ..... ix
List of Tables ..... xii
List of Figures ..... xiii
List of Symbols ..... xv
List of Abbreviations ..... xvi

**CHAPTER ONE: INTRODUCTION ..... 1**
1.1 Background of the Study ..... 1
1.2 Problem Statement ..... 3
1.3 Resesarch Objectives ..... 4
1.4 Research Hypothesis ..... 5

**CHAPTER TWO: LITERATURE REVIEW ..... 6**
2.1 *Burkholderia pseudomallei* as a Target Bacterium ..... 6
2.1.1 General Characteristics ..... 6
2.1.2 Pathogenicity and Virulence Factor ..... 7
2.1.3 Genomic Study ..... 8
2.2 Melioidosis as a disease of concern ..... 10
2.2.1 Nature of Melioidosis ..... 10
2.2.2 Exposure to Melioidosis ..... 10
2.2.3 Treatment of Melioidosis ..... 12
2.2.4 The Epidemiology ..... 13
2.3 Hypothetical protein in *B. pseudomallei* ..... 14
2.3.1 Proteomic Study on the Hypothetical Protein ..... 16
2.3.2 Target Protein, BPSL2774 ..... 17
2.4 Prediction of essential gene ..... 18
2.5 Gateway cloning method ..... 21
2.6 Protein Expression ..... 24
2.6.1 Protein Production by Auto-induction Method ..... 25
2.6.2 Protein Expression System ..... 25
2.7 Structure Prediction of Protein ..... 27

**CHAPTER THREE: RESEARCH METHODOLOGY ..... 30**
3.1 Research Design ..... 30
3.2 Sample Preparation ..... 31
3.3 Bioinformatics Analysis ..... 31
3.4 Polymerase Chain Reaction (PCR) of DNA Sequences ..... 32

3.4.1 Primer Design ..... 32
3.4.2 Gradient PCR ..... 35
3.4.3 Two-steps PCR ..... 35
3.4.4 Purification of PCR Product ..... 37
3.5 Recombinant Cloning using Gateway Cloning System ..... 37
3.5.1 BP Recombination ..... 39
3.5.2 Transformation ..... 39
3.5.3 Plasmid Miniprep Purification ..... 40
3.5.4 LR Recombination ..... 41
3.6 Screening for Positive Clones ..... 41
3.6.1 Colony PCR for Positive Clones ..... 42
3.6.2 BsrG1 Restriction Digest for Positive Clones ..... 42
3.7 Protein Purification ..... 42
3.7.1 Cell Culture ..... 43
3.7.2 Cell Lysis ..... 43
3.7.3 Protein Purification ..... 44
3.8 Gel Electrophoresis ..... 45
3.8.1 DNA Gel Electrophoresis ..... 45
3.8.2 SDS-PAGE Electrophoresis ..... 45
3.9 *In silico* Prediction of BPSL2774 Protein Function ..... 47
3.9.1 Bioinformatics Analysis on the Structural and Functional ..... 47
3.9.2 *In silico* BPSL2774 ..... 47
3.9.3 Docking Experiment ..... 48

**CHAPTER FOUR: RESULT AND DISCUSSION ..... 50**
4.1 Introduction ..... 50
4.2 Selection of Hypothetical Protein Targets ..... 50
4.3 Gene Specific Primer Design ..... 52
4.4 Polymerase Chain Reaction (PCR) Amplification ..... 53
4.5 Directional Cloning using Gateway Vector ..... 58
4.5.1 *BPSL0089* Target Gene Result ..... 58
4.5.2 *BPSL2774* Target Gene Result ..... 61
4.6 Protein Purification and Expression ..... 68
4.6.1 Protein Separation on SDS-PAGE ..... 68
4.6.2 Protein Identification by Mass Spectrometry ..... 71
4.7 *In silico* BPSL2774 Protein Structutre Prediction ..... 74
4.7.1 Physicochemical Properties ..... 74
4.7.2 Prediction of Secondary Structure of BPSL2774 ..... 75
4.7.3 Predicted Structural Model of BPSL2774 ..... 79
4.7.4 Protein Structural and Functional Analysis ..... 86
4.7.4.1 Analysis using BLASTp to PDB Approach ..... 86
4.7.4.2 Analysis using COACH Meta-server Prediction ..... 88
4.7.4.3 Protein-protein Interaction Analysis by STRING ..... 90
4.7.5 Active Site Prediction ..... 92
4.7.6 BPSL2774 Docking Analysis ..... 96

**CHAPTER FIVE: CONCLUSION, LIMITATION AND FUTURE WORK .101**
5.1 Conclusion .....101
5,2 Limitation.....102
5.3 Future Work.....102

**REFERENCES.....104**

**APPENDIX A: PREDICTED ESSENTIAL GENES OF *B. pseudomallei* .....115**
**APPENDIX B: VECTOR MAP AND THE FEATURE .....126**
**APPENDIX C: 50 X TAE BUFFER PREPARATION .....129**
**APPENDIX D: PCR CLEAN-UP KIT (AXYGEN) .....130**
**APPENDIX E: ELECTROCOMPETENT CELL PREPARATION .....131**
**APPENDIX F: AUTO-INDUCING MEDIUM (ZYM) RECIPE.....133**
**APPENDIX G: RESOLVING AND STACKING GEL COMPOSITION.....136**

# LIST OF TABLES


| Table No. | | Page No. |
|---|---|---|
| 3.1 | Gene-specific primers used for the amplification of selected hypothetical protein gene constructs | 34 |
| 3.2 | Generic primers used for the amplification of Gateway recombination sites | 34 |
| 3.3 | PCR Components using Primestar® HS DNA Polymerase (Takara) | 35 |
| 3.4 | First round PCR reaction protocols | 36 |
| 3.5 | Second round PCR reaction protocols | 36 |
| 4.1 | Five selected target genes from TraDIS library | 51 |
| 4.2 | Optimal annealing temperature and PCR product size | 54 |
| 4.3 | DNA purity and concentration of PCR product after purification by AxyPrep PCR Clean-up Kit (Axygen) | 58 |
| 4.4 | Physicochemical properties of BPSL2774 | 75 |
| 4.5 | Generated BPSL2774 structure models from I-TASSER | 81 |
| 4.6 | AutoDock4.2 docking simulation results for BPSL2774 protein | 98 |
| 4.7 | Docking energy and RMSD from AutoDock Vina | 100 |

# LIST OF FIGURES


| Figure No. | | Page No. |
|---|---|---|
| 2.1 | Schematic circular diagrams of the large (left) and small (right) chromosomes of *B. pseudomallei* genome | 9 |
| 2.2 | The range of clinical manifestations after infection with *B. pseudomallei* | 11 |
| 2.3 | Conserved domain of BPSL 2774 in blastp to PDB | 17 |
| 2.4 | Simplified TraDIS method | 20 |
| 2.5 | The Gateway Cloning | 22 |
| 2.6 | Nested PCR and recombinant cloning using the Gateway™ Technology | 23 |
| 3.1 | Summary of methodology workflow | 30 |
| 3.2 | Overview of the Gateway systems | 38 |
| 4.1 | First round PCR products (Standard PCR) for target gene constructs | 55 |
| 4.2 | Standard second round PCR products for target gene constructs | 56 |
| 4.3 | Screening for positive clones in BP recombination for BPSL0089 gene | 60 |
| 4.4 | Screening for positive clones in BP recombination for BPSL2774 gene | 62 |
| 4.5 | BLASTN alignments between Query and Subject K96243 | 64 |
| 4.6 | Multiple sequence alignment using ClustalWe | 65 |
| 4.7 | Screening for positive clones in LR recombination for BPSL2774 | 67 |
| 4.8 | SDS-PAGE gel of 50 mL culture (A) and 500 mL culture (B) | 70 |
| 4.9 | Mascot score histogram of mass spectrometry analysis | 71 |

| | | |
|---|---|---|
| 4.10 | Protein summary report from mass spectrometry analysis for the third hit (labelled Q63R99) showing the presence of target protein BPSL2774 | 72 |
| 4.11 | Secondary structure prediction of BPSL2774 generated using PDBsum server | 77 |
| 4.12 | Topology from secondary structure of BPSL2774 generated from PBDsum server | 78 |
| 4.13 | Top 10 threading templates used by I-TASSER for modeling BPSL2774 protein | 80 |
| 4.14 | Predicted three-dimensional structure model of BPSL2774 | 84 |
| 4.15 | Model validation studies of BPSL2774 by Ramachandran's plot using Rampage server | 85 |
| 4.16 | Alignments of BPSL2774 and three glycosyltransferases from *B. anthracis* | 87 |
| 4.17 | Predicted ligand binding sites and possible binding ligands for raw unrefined BPSL2774 Model1.pdb structure using COACH meta-server | 89 |
| 4.18 | Predicted ligand binding sites and possible binding ligands for refined BPSL2774 structure (refined.pdb) using COACH meta-server | 90 |
| 4.19 | Network analysis results for BPSL2774 protein | 91 |
| 4.20 | Network analysis results for BPSL2774 protein using STRING tool, with narrower required interaction score set | 92 |
| 4.21 | Active sites of the predicted 3D structure of the target protein as determined by metaPocket server | 94 |
| 4.22 | Predicted ligands that bind to the predicted active site by I-TASSER, and superimposed on active sites prediction spheres from metaPocket server | 95 |

# LIST OF SYMBOLS

| | | |
|---|---|---|
| % | - | Percent |
| °C | - | Degree Celsius |
| ×g | - | Times gravity |
| bp | - | Base pair |
| fmol | - | Femto mole |
| g | - | Gram |
| kb | - | Kilo base pair |
| kDa | - | Kilo Dalton |
| L | - | Liter |
| Mb | - | Mega base pair |
| µL | - | Microliter |
| µm | - | Micrometer |
| µM | - | Micro molar |
| mA | - | Milli Ampere |
| mg | - | Milligram |
| mL | - | Milliliter |
| ng | - | Nano gram |
| nm | - | Nano meter |
| RPM | - | Revolutions per minute |
| U | - | Unit |
| V | - | Volt |

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| $A_{260}$ | - | Absorbance at the wavelengths of 260 nm |
| $A_{280}$ | - | Absorbance at the wavelengths of 280 nm |
| $A_{260}/A_{280}$ | - | Ratio of absorbance at 260 nm and 280 nm |
| $A_{260}/A_{230}$ | - | Ratio of absorbance at 260 nm and 230 nm |
| AHLs | - | Acyl-homoserine-lactones |
| Ara | - | L-arabinose |
| *B. mallei* | - | *Burkholderia mallei* |
| *B. pseudomallei* | - | *Burkholderia pseudomallei* |
| *B. thailandensis* | - | *Burkholderia thailandensis* |
| BLAST | - | Basic Local Alignment Search Tool |
| BSA | - | *Burkholderia* secretion apparatus |
| CDs | - | Coding sequences |
| DNA | - | Deoxyribonucleic acid |
| *E. coli* | - | *Escherichia coli* |
| EDTA | - | Ethylenediaminetetraacetic acid |
| et al. | - | et alia: and others |
| EtBr | - | Ethidium Bromide |
| GC-content | - | Guanine-Cytosine content |
| QS | - | Quorum sensing |
| IDT | - | Integrated DNA Technologies |
| IIUM | - | International Islamic University Malaysia |
| LB | - | Luria-Bertani medium |
| LPS | - | Lipopolysaccharide |
| NaOH | - | Sodium Hydroxide |
| O-PS | - | O-polysaccharide |
| ORF | - | Open Reading Frame |
| PCR | - | Polymerase Chain Reaction |
| PDB | - | Protein Data Bank |
| rTEV | - | Recombinant TEV |
| RPM | - | Revolutions per minute |
| TAE | - | Tris-Acetate-EDTA |
| TE | - | Tris-EDTA |
| TTSS | - | Type III secretion system |
| Tm | - | Melting temperature |
| Tris-Cl | - | Tris-Chloride |
| UKM | - | Universiti Kebangsaan Malaysia |
| UV | - | Ultraviolet |

# CHAPTER ONE

# INTRODUCTION

## 1.1 BACKGROUND OF THE STUDY

Melioidosis, also called Whitmore's Disease, is an infectious disease spread by *B. pseudomallei*, a gram negative bacterium which resides in contaminated water and soil. Direct contact with the contaminated source can spread the disease to human and animals. Soil, stagnant water and rice fields are the habitat for this bacterium and can be found in endemic regions including Southeast Asia and northern Australia (Limmathurotsakul et al., 2016; Chewapreecha et al., 2017). It secretes lecithinase, lipase, hemolysin and siderophore for its survival and maintenance (Stevens et al., 2002). In the past two decades, melioidosis was categorized as an important human infection in Malaysia, Singapore and across the north of Australia. This is of particular concern for *B. pseudomallei* as it is naturally resistant to many commonly used antibiotics. The intrinsic resistance mainly contributed by the exclusion of the antibiotic from the bacterium by restriction of the cell envelope and lipopolysaccharide (Rhodes & Schweizer, 2016). The mechanism of resistance in the genome of strain K96243 is also aided by the presence of seven Ambler class A, B and D β-lactamases, ten multidrug efflux systems and a putative aminoglycoside acetyl transferase (Schweizer, 2012). The genomic study of *B. pseudomallei* K96243 has shown that it comprised two chromosomes of 4.07 and 3.17 megabase pairs, respectively (Holden et. al, 2004). The large chromosome is important for metabolism and growth, whereas the small chromosome is useful in adaptation and survival (Holden et al., 2004).

A vaccine to protect against this bacterium is yet to be developed and the proteomics involved in the pathogenicity of *B. pseudomallei* is still need to be discovered. 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 such as diabetes (Peacock et al., 2012). A potential vaccination strategy has also been considered using the closely-related avirulent *B. 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 et al., 2004).

The identification of essential genes in *B. pseudomallei* is crucial for understanding the cellular mechanism and potential targets selection for the development of new or improved antibiotics (Song et al., 2014). Single gene detection, transposon mutagenesis, genetic footprinting and antisense RNA techniques are used for the prediction and discovery of essential genes (Guo et al., 2015). These techniques may not always be feasible however, as they require large investment of time and resources (Acencio & Lemke, 2009). Recently, next-generation sequencing approach has enabled the screening of mutagenesis so that the transposon insertion sites can be located efficiently. This is useful for analysis of the mutants in the transposon-directed insertion site sequencing (TraDIS) and transposon sequencing (Tn-seq) techniques. These techniques enable the compilation of the essential gene important for antimicrobial development. Many candidate open reading frame (ORF) are being identified with no known function from numerous genomes sequencing projects. Experimental and computational approaches may help to identify and characterize these proteins (Chen et al., 2006). New discoveries are

possible by determining whether the expression of unknown genes during infection is relevant to the process of disease. Moule and his colleagues utilize TraDIS techniques to list and compile the essential genes encoding for hypothetical proteins or conserved hypothetical proteins (Moule et al, 2014). BPSL 2774 is one of the hypothetical proteins in *B. pseudomallei* that are predicted to be essential through TraDIS method (Moule et al, 2014).

In this study, essential genes encoding for hypothetical proteins from *B. pseudomallei* are targeted for cloning, over expression and characterization to improve fundamental knowledge on the bacterium's survival strategy. A screening process was performed by amplifying five selected target genes predicted to be essential by TraDIS method using PCR for recombinational cloning by Gateway™ cloning system. Recombinational cloning allows DNA segments to be transferred into a variety of vector backgrounds for protein expression and functional analysis of these genes (Hartley et al., 2000). One of the commonly used vectors is Gateway™ vectors. It is suitable for the cloning of ORFs for studies in protein interaction mapping, structure determination and protein localization (Walhout, 2000). Then, a small-scale culture and protein expression on the successfully cloned target gene was conducted. The successfully expressed target protein was further analyzed for the predicted function using various *in silico* structural and functional analysis including secondary and tertiary structure prediction, active site prediction as well as docking study with the expected ligand molecules.

## 1.2 PROBLEM STATEMENT

Mortality from melioidosis septic shock, caused by infection from gram-negative *B. pseudomallei* remains high despite appropriate antimicrobial therapy. Thus, the

development of new antimicrobial therapies is ongoing, with researchers utilizing different strategies to identify essential genes and drug targets to combat melioidosis. A study of essential genes of *B. pseudomallei* would be of great value due to its resistance to the commonly used antibiotics and the products might be significant for novel antimicrobial drugs production. Using TraDIS technique, a set of essential genes that are needed for bacteria development has recently been predicted and identified (Moule et al., 2014). This list provides researchers a good starting point to analyze and characterize potential functional genes, against which to ultimately develop potential novel antibiotics. Therefore, biochemical and biophysical characterization can be completed based on the expression screening of the essential genes.

## 1.3 RESEARCH OBJECTIVES

The study aimed to achieve the following objectives:

**General objective:**

To characterize *BPSL 2774* gene of *Burkholderia pseudomallei*.

**Specific objectives:**

i. To amplify five target genes listed as essential genes by transposon-directed insertion site sequencing (TraDIS) technique using PCR.

ii. To clone the successfully amplified target genes from *B. pseudomallei* using Gateway™ cloning system.

iii. To conduct protein purification and expression on the successfully cloned target genes.

iv. To perform *in silico* structural and functional prediction using bioinformatics.

## 1.4 RESEARCH HYPOTHESES

The hypotheses are as follows:

i. Potential target genes from *B. pseudomallei* can be screened and characterized for confirming the prediction from Transposon-Directed Insertion site Sequencing (TraDIS).

ii. *BPSL 2774* gene from *B. pseudomallei* has glycosyltransferase activity.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 *BURKHOLDERIA PSEUDOMALLEI* AS A TARGET BACTERIUM

### 2.1.1 General Characteristics

The genus *Burkholderia* was first described in 1950 by Walter Burkholder (Burkholder, 1950). It is made up of 43 species of non-sporing, motile, bacillus and gram-negative bacterium. Under gram stain, *B. pseudomallei* displays bipolar staining, giving it a 'safety pin' like appearance.The organism is easily recovered on standard culture medium but may be misidentified as *Burkholderia cepacia, Pseudomonas stutzeri* or other *Pseudomonas* species (Currie, 2010). The organism is oxidase positive due to its usage of glucose in its oxidative pathway. Furthermore, the ability of *Burkholderia* to utilize L-arabinose enables the distinction between *B. thailandensis* ($Ara^+$) and *B. pseudomallei* ($Ara^-$) (Wuthiekanun et al., 1996). *B. pseudomallei* can be cultured aerobically and it shows white colonies on most agar media after 24-48 hours incubation at 37 ºC. Ashdown's medium or modified versions are commonly used. They initially produce smooth colonies, followed by dry or wrinkled colonies after further incubation (Puthucheary, 2009). It is a remarkably diverse species that are found in various localities, ranging from contaminated soils to the respiratory tract of humans (Vellasamy et al., 2012) which include environmental, clinical and agro-biotechnological relevance. This genus was shown to have distinct lineages from the large groups of plant-associated and saprophytic bacterial species and human pathogen, *B. cepacia* complex (Estrada et al., 2013). Most species in this genus are

harmless and non-pathogenic with a few exceptions that are able to cause severe and life threatening infections to humans and animals. One of the pathogenic species is *B. pseudomallei* which are commonly found in soil and water in Southeast Asia, Northern Australia, Central America and South America (Currie et al., 2008). This bacterium is the pathogen of melioidosis, an unusual bacterial epidemic characterized by abscesses in tissues and organs.

### 2.1.2 Pathogenicity and Virulence Factor

Several features of melioidosis suggest that *B. pseudomallei* is a facultative intracellular pathogen. It is inherently resistant to many antibiotics such as penicillin, ampicillin, first-generation and second-generation cephalosporin, gentamicin, tobramycin, streptomycin, and polymyxin and can cause latent infection (Wiersinga et al., 2012). Identified virulence factors for the pathogenesis of this bacterium includes its cell surface polysaccharides and lipopolysaccharide, adhesins for host cell adherence, secretions systems, e.g., Type III secretion system found in Gram-negative pathogens, actin-based intracellular motility and a variety of secreted factors (Stone et al., 2014). Once it has entered the intracellular compartment, *B. pseudomallei* is able to escape from endocytic vacuoles and move within the cytoplasm and enter neighbouring cells by inducing actin rearrangement, leading to the formation of actin tails and membrane protrusions (Shalom et al., 2007). Among the mechanism of antibiotic resistance documented in *B. pseudomallei* is the permeability of the cell envelope, efflux from the cell as well as altered target sites (Schweizer, 2012). Consistent with this, *B. pseudomallei* has been shown to survive and multiply within non-phagocytic cells, macrophages and free-living amoebae (Inglis et al., 2000).

It is also predicted that the survival of *B. pseudomallei* in the immune system are the main components in the pathogenesis of melioidosis. In addition, their survival and persistence in the environment as well as in the host offer a notable example of bacterial adaptation. The quorum sensing system, type III secretion system [TTSS] gene clusters, type VI secretion systems, capsular polysaccharide and surface O-polysaccharides are the possible multiple virulence factor for *B. pseudomallei* (Wiersinga et al., 2006). However, further study on the impact for each virulence factor is still needed (Wiersinga et al., 2012). Recently, several environmental *B. pseudomallei*-like organisms were formally classified as *B. thailandensis* which are not correlated with human disease and are avirulent in the Syrian golden hamster animal model (Coenye & Vandamme, 2003) and useful for examining *B. pseudomallei* in detail due to the genomic similarity for both bacteria (Yu et al., 2006). In addition, *B. thailandensis* can be used as potential vaccine with approximately 50% defence against *B. pseudomallei* in guinea pigs (Iliukhin et al., 2002). However, using living *B. thailandensis* as a vaccine is not preferred as it can cause pneumonia-derived sepsis (Wiersinga et al., 2008).

### 2.1.3 Genomic Data

The *B. pseudomallei* with strain K96243 originated from Thailand and is the first *B. pseudomallei* strain to have its whole genome sequenced. The genome is quite complex, where it is comprised of two chromosomes of 4.07 and 3.17 megabase pairs respectively (Figure 2.1, Holden et al, 2004). The genome consists of a variety of genes that are responsible for the survival in extreme conditions as well as for pathogenicity. A core set of 2590 genes is shared between *B. pseudomallei* and other members of the *Burkholderia* genus (Wiersinga et al., 2012). Within the prototypic *B.*