

**FUNCTIONAL METAGENOMICS AND CULTURED-
BASED APPROACHES FOR THE DISCOVERY OF
COLD-ACTIVE PROTEOLYTIC ENZYMES FROM THE
ANTARCTIC REGION**

BY

MUHAMMAD ASYRAF BIN ABD LATIP

A thesis submitted in fulfilment of the requirement for the
degree of Doctor of Philosophy (Engineering)

**Kulliyyah of Engineering
International Islamic University Malaysia**

NOVEMBER 2021

ABSTRACT

The Antarctic region is a new frontier as a natural source for bio-prospecting purposes. Its extreme cold temperature may provide unique microbial enzyme characteristics that have valuable potential for industrial and biotechnological applications. However, the majority of microbial cannot be cultured due to their complex structure of life and unfavourable media composition and growth conditions. In this study, two different approaches were designed to ensure the total discovery of proteases that are active and able to work at relatively low temperatures. The first approach was the culture-based technique. Soil samples from the Antarctic region were screened for protease activity on skim milk agar at 4 °C. Bacteria that showed a clear halo zone around the colonies were isolated and identified through 16S rDNA sequencing. Bacteria that showed rapid halo zone formation within 24 hours were further screened for its growth rate condition using one-factor-at-a-time (OFAT). Then, crude protease of each strain was extracted during the late logarithmic phase for enzymatic assay. Strain with significant enzyme activity was optimized using Response Surface Method (RSM) for maximal growth rate. The second approach was the metagenomics approach. Amplicon metagenomics sequencing was performed to analyse the bacteria community and the discovery of cold-active protease was completed through functional metagenomics. A total of 46 bacteria strains with positive protease activity were isolated and phylogenetic analysis showed that 88% were from *Pseudomonas* sp., 9% from *Arthrobacter* sp. and 3% from *Paenibacillus* sp. OFAT result showed that 10 selected strains displayed the highest growth rate at 20 °C, pH 7 and 4% (w/v) of NaCl. The enzymatic assay showed that crude enzyme extracted from strain SC8 exhibited significantly higher activity than the positive control (protease from bovine pancreas) at -20 °C and 20 °C. Furthermore, RSM suggested that the optimized conditions for the growth of strain SC8 were at 20.5 °C, pH 6.83, and 2.05% (w/v) of NaCl. Bacterial community analysis using amplicon metagenomics sequencing showed that each sample (ROB, ROS, and SC) were dominated by the uncultured bacteria. A clone with positive protease activity was isolated and the fosmid was extracted and sequenced. Contigs of NODE_42 with 893 bp showed significant matched to Peptidase M23 and PG binding 1 protein families. In silico analysis showed that this predicted protease exhibited 27.07% similarity of the template enzyme with PDB ID 3SLU. The growth condition of isolated bacteria was influenced by the site surroundings and the coastal environment. These bacteria were categorized as psychrotolerants and the majority of them can tolerate acidic conditions. Optimum bacterial growth conditions were important to maximize enzyme production and activity. Thus, strain SC8 has the potential for bioprospecting for the development of large-scale production of cold-active protease in the future. Overall, for this project, we successfully discovered our targeted cold-active protease. This preliminary research is vital to unearth all potential protease for bioprospecting purposes. In the future, purification of this enzyme is essential to precisely inspect the enzyme activity.

خلاصة البحث

تعد منطقة القطب الجنوبي جبهة جديدة كمصدر طبيعي لأغراض التنقيب البيولوجي. قد توفر درجة حرارة البرودة الشديدة خصائص إنزيم جرثومي فريدة لها إمكانات قيمة للتطبيقات الصناعية والتكنولوجيا الحيوية. ومع ذلك ، لا يمكن استزراع غالبية الميكروبات بسبب هيكلها المعقد للحياة وتكوين الوسائط غير المواتية وظروف النمو. في هذه الدراسة ، تم تصميم طريقتين مختلفتين لضمان الاكتشاف الكلي للبروتياز النشطة والقادرة على العمل في درجات حرارة منخفضة نسبيًا. كان النهج الأول هو الأسلوب القائم على الاستنبات البكتيري. تم فحص عينات التربة من منطقة القطب الجنوبي بحثًا عن نشاط الأنزيم البروتيني على أجار الحليب الخالي من الدسم عند درجة حرارة 4 °C. تم عزل البكتيريا التي أظهرت منطقة هالة واضحة حول المستعمرات وتم تحديدها من خلال تسلسل 16S rDNA. تم فحص البكتيريا التي أظهرت تكوين منطقة هالة سريع خلال 24 ساعة بشكل إضافي لمعرفة حالة معدل نموها باستخدام تقنية عامل واحد في كل مرة (OFAT). بعد ذلك ، تم استخراج الأنزيم البروتيني الخام لكل سلالة خلال المرحلة اللوغاريتمية المتأخرة للمقايضة الأنزيمية. تم تحسين السلالة مع نشاط إنزيم كبير باستخدام طريقة الاستجابة السطحية (RSM) لمعدل النمو الأقصى. كان النهج الثاني هو نهج الميتاجينوميات. تم إجراء امبليكون ميتاجينوميات لتحليل مجتمع البكتيريا واكتمل اكتشاف البروتياز النشط البارد من خلال علم الميتاجينوميات الوظيفية. تم عزل 46 سلالة بكتيرية ذات نشاط بروتياز إيجابي وأظهر تحليل النشوء والتطور أن 88٪ كانت من *Pseudomonas sp.* و 9٪ من *Arthrobacter sp.* و 3٪ من *Paenibacillus sp.* أظهرت نتيجة OFAT أن 10 سلالات مختارة أظهرت أعلى معدل نمو عند 20 °C ، ودرجة الحموضة 7 و 4٪ (w/v) من كلوريد الصوديوم. أظهر الفحص الأنزيمي أن الإنزيم الخام المستخرج من سلالة SC8 أظهر نشاطاً أعلى بكثير من التحكم الإيجابي (الأنزيم البروتيني من البنكرياس البقري) عند 20-20 °C و 20 °C. علاوة على ذلك ، اقترح RSM أن الظروف المثلى لنمو سلالة SC8 كانت عند 20.5 °C ، ودرجة الحموضة 6.83 ، و 2.05٪ (w/v) من كلوريد الصوديوم. أظهر تحليل المجتمع البكتيري باستخدام امبليكون ميتاجينوميات أن كل عينة (ROB و ROS و SC) كانت تهيمن عليها البكتيريا غير المستزرعة. تم عزل استنساخ له نشاط بروتياز إيجابي واستخلاص الفوسميد وتسلسله. أظهرت Contigs من NODE_42 مع 893 نقطة أساس معنوية مطابقة لعائلات بروتين Peptidase M23 و PG ملزمة 1 عائلة بروتين. في تحليل السيليكو أظهر أن هذا البروتياز المتوقع أظهر تشابهاً بنسبة 27.07٪ من إنزيم القالب مع PDB ID 3SLU. تأثرت حالة نمو البكتيريا المعزولة بمحيط الموقع والبيئة الساحلية. تم تصنيف هذه البكتيريا على أنها متحملة نفسية ويمكن لمعظمها تحمل الظروف الحمضية. كانت ظروف النمو البكتيري المثلى مهمة لزيادة إنتاج الإنزيم ونشاطه. وبالتالي ، فإن سلالة SC8 لديها القدرة على التنقيب البيولوجي لتطوير إنتاج واسع النطاق للبروتياز النشط البارد في المستقبل. بشكل عام ، بالنسبة لهذا المشروع ، اكتشفنا بنجاح البروتياز النشط البارد. يعد هذا البحث التمهيدي أمرًا حيويًا للكشف عن جميع أنواع البروتياز المحتملة لأغراض التنقيب البيولوجي. في المستقبل ، يعد تنقية هذا الإنزيم ضروريًا لفحص نشاط الإنزيم بدقة.

APPROVAL PAGE

The thesis of Muhammad Asyraf bin Abd Latip has been approved by the following:

Noor Faizul Hadry bin Nordin
Supervisor

Mohd Azrul Naim bin Mohamad
Co-Supervisor

Faridah binti Yusof
Co-Supervisor

Azura binti Amid
Internal Examiner

Fahrul Zaman bin Huyop
External Examiner

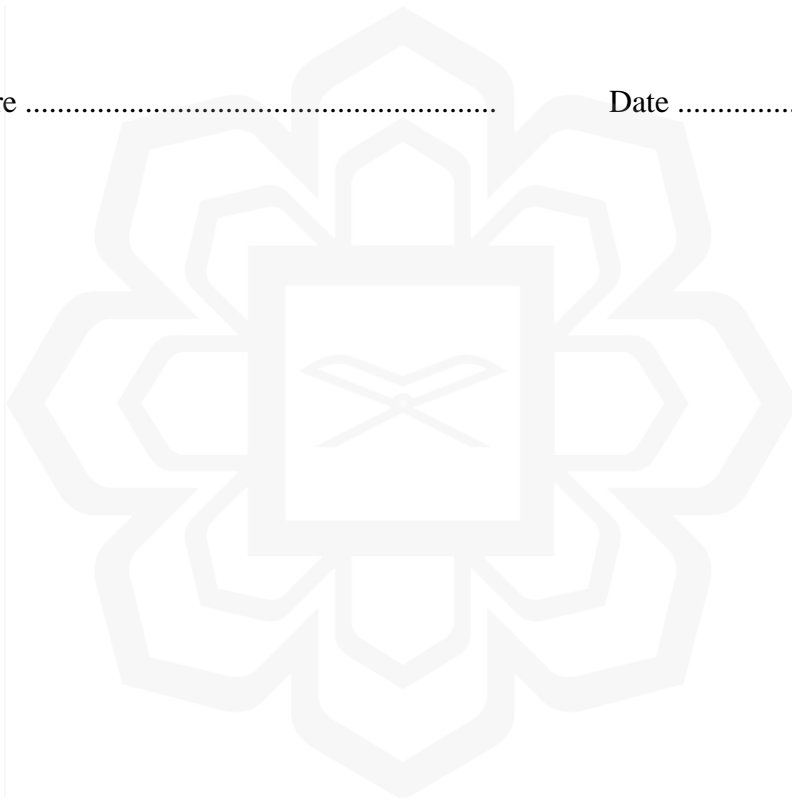
Hassanuddeen bin Abd Aziz
Chairman

DECLARATION

I hereby declare that this thesis is the result of my 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.

Muhammad Asyraf bin Abd Latip

Signature Date



INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

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

**FUNCTIONAL METAGENOMICS AND CULTURED-BASED
APPROACHES FOR DISCOVERY COLD ACTIVE
PROTEOLYTIC ENZYMES FROM ANTARCTIC REGION**

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

Copyright © 2021 Muhammad Asyraf bin Abd Latip 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 understood the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Muhammad Asyraf bin Abd Latip

.....
Signature

.....
Date

ACKNOWLEDGEMENTS

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

All praise to Allah S.W.T. for giving me the health, strength and guidance to complete this thesis.

First of all, I would like to acknowledge the Yayasan Penyelidikan Antartika Sultan Mizan (YPASM) for financially supporting this exciting project. I would like to thank Kulliyyah of Engineering, International Islamic University Malaysia (IIUM) for facilitating this project. In addition, I would like to thank everyone in the Department of Biotechnology Engineering, who have guided and helped me in completing this project.

I would like to extend my gratitude to Dr Noor Faizul Hadry Nordin, who has not only been a great advisor and mentor but a great friend as well. He taught me more than just how to become a scientist, but also life values from which I have greatly benefited. He always gave me freedom and support to do just about anything I wanted to, inside and outside of the laboratory. Not only he is a great supervisor, but a great person that I have had the pleasure to work with.

I would like to thank Prof. Faridah Yusof and Dr Mohd Azrul Naim Mohamed for their support and advice. They both have been very helpful and accessible whenever I needed them the most. I truly believed that I had the best dissertation committee. I also would like to thank Dr Siti Aisah Alias from National Antarctic Research Centre (NARC) for sharing the Antarctic samples that made this project a possible venture.

Special thanks go to Deni, Oualid and Farah for their assistance, guidance and unconditional moral support. They are my great friends, lab colleagues and PhD mates.

The most important person throughout my PhD journey, my beloved wife who is always by my side in every ups and downs situation. Finally, my deepest gratitude to my dearest father, mother and siblings for their love, support and prayers throughout this chapter of my life. May Allah S.W.T. reward all of them with His mercy and compassion, here and hereafter.

Thank you.

الحمد لله

TABLE OF CONTENTS

Abstract	i
Abstract in Arabic	ii
Approval Page.....	ii
Declaration.....	iv
Copyright Page.....	v
Acknowledgements.....	vi
Table of Contents	vii
List of Tables	x
List of Figures	xii
List of Symbols	xvi
List of Abbreviations	xvii
CHAPTER ONE: INTRODUCTION.....	1
1.1 Background of the Study	1
1.2 Problem Statement.....	3
1.3 Research Questions.....	4
1.4 Research Objectives.....	5
1.5 Significance of the Study.....	5
1.6 Research Hypothesis.....	6
1.7 Research Scope.....	7
1.8 Limitations of the Study	8
CHAPTER TWO: LITERATURE REVIEW.....	9
2.1 Introduction.....	9
2.2 Phylogenetic Study of Bacteria	11
2.3 Advantages of Using Cold-active Enzymes	14
2.4 Biotechnology Applications of Cold-active Enzymes.....	17
2.5 Production of Enzyme	19
2.6 Proteolytic Enzyme.....	20
2.7 Screening of Cold-active Enzymes.....	23
2.8 Culturing Psychrophiles and Psychrotolerants	25
2.9 Metagenomics	26
2.9.1 Amplicon Metagenomics Sequencing.....	28
2.9.2 Functional Metagenomics	31
CHAPTER THREE: RESEARCH METHODOLOGY	37
3.1 Introduction.....	37
3.2 General Flow Chart.....	38
3.3 Chemicals	39
3.4 Materials	39
3.4.1 Preparation of Luria-Bertani (LB) Broth.....	39
3.4.2 Preparation of Luria-Bertani (LB) Agar.....	40
3.4.3 Preparation of LB Broth for Optimization	40
3.4.4 Preparation of Skim Milk (5%) Agar	41

3.4.5 Preparation of Chloramphenicol Stock (100 mg/mL)	41
3.4.6 Preparation of Protease Solution (0.2 units/mL)	41
3.5 Samples Collection and Processing	41
3.6 Screening of Proteolytic Enzymes	43
3.7 Phylogenetic Analysis of Culturable Bacteria	44
3.7.1 Bacterial DNA Extraction	45
3.7.2 Polymerase Chain Reaction (PCR)	45
3.7.3 Gel Electrophoresis	46
3.7.4 PCR Purification	47
3.7.5 DNA Quantification	47
3.7.6 Construction of Phylogenetics Tree	49
3.8 Characterization of Bacteria Growth Rate	50
3.8.1 Effect of Temperature on Bacteria Growth Rate	50
3.8.2 Effect of pH on Bacteria Growth Rate	50
3.8.3 Effect of NaCl Concentration on Bacteria Growth Rate	51
3.9 Protease Assay of Crude Enzymes	51
3.9.1 Statistical Analysis	52
3.10 Optimization of Bacterial Growth Rate	53
3.11 DNA Extraction for Metagenomics	54
3.11.1 Preparation of Extraction Buffer	54
3.11.2 Extraction of Environmental DNA (eDNA)	55
3.11.3 Fluorometric DNA Quantification	56
3.12 Amplicon Metagenomics Sequencing Data Analysis	57
3.13 Functional Metagenomics	62
3.13.1 Size Selection of the End-Repaired DNA	63
3.13.2 Recovery of the Size-Fractioned DNA	64
3.13.3 Ligation Reaction	65
3.13.4 Preparation of Phage Dilution Buffer	66
3.13.5 Packaging the CopyControl Fosmid Clones	66
3.13.6 Preparation of EPI300-T1 ^R Plating Strain	67
3.14 Screening of Functional Expression Library	67
3.14.1 Preparation of Skim Milk Agar with Chloramphenicol and Arabinose	68
3.14.2 Purification of Fosmid DNA	68
3.15 Functional Metagenomics Data Analysis	70
3.15.1 Molecular Analysis	73

CHAPTER FOUR: RESULTS AND DISCUSSION 74

4.1 Introduction	74
4.2 Identification of Bacteria with Cold-active Protease	74
4.2.1 Screening of Bacteria with Proteolytic Activity	75
4.2.2 Bacterial Colony Forming Unit (CFU)	76
4.2.3 Polymerase Chain Reaction (PCR) Amplification	79
4.2.4 Bacterial Identification and Phylogenetic Analysis	81
4.2.4.1 <i>Pseudomonas</i> Species	85
4.2.4.2 <i>Arthrobacteria</i> Species	85
4.2.4.3 <i>Paenibacillus</i> Species	85
4.2.5 Enzymatic Assay	86

4.3 Characterizations and Optimization Growth Rate of Bacteria	
Producing Cold-active Protease	88
4.3.1 Effect of Temperature on Growth Rate of Bacteria	89
4.3.2 Effect of pH on Growth Rate of Bacteria	90
4.3.3 Effect of Sodium Chloride (NaCl) Concentration on Growth	
Rate of Bacteria.....	91
4.3.4 Enzymatic Assay of Crude Enzymes	92
4.3.5 Optimization by Response Surface Methodology (RSM).....	94
4.3.6 OFAT vs RSM.....	99
4.4 Metagenomics Studies	99
Part One: Amplicon Metagenomics Sequencing.....	99
4.4.1 High-Molecular-Weight DNA Extraction.....	100
4.4.2 Comparison of Bacterial Diversity and Community	
Composition Between Soils	103
4.4.3 Metagenomics Analysis of 16S rRNA Sequences	107
Part Two: Functional Metagenomics.....	111
4.4.4 Screening of Clones with Proteolytic Activity	111
4.4.5 Functional Metagenomics Data Analysis.....	114
4.4.6 <i>In Silico</i> Analysis and Enzymatic Assay of Predicted Protease	116
CHAPTER FIVE: CONCLUSION AND FUTURE WORKS.....	124
5.1 Conclusion	124
5.2 Future Works	126
REFERENCES.....	127
APPENDIX A: 16S rRNA AMPLICON TAXONOMY ANALYSIS	156
APPENDIX B: BLASTN METAGENOME DATA	243
APPENDIX C: ANNOTATED RESULTS OF NODE.....	249
APPENDIX D: PUBLICATIONS AND CONFERENCES	251

LIST OF TABLES

Table 2.1	Phylogenetic studies of culturable bacteria in the Antarctic region.	12
Table 2.2	Cold active protease have been extracted from various psychrophilic species and their optimal temperature activity.	22
Table 2.3	Comparison of screening and selection technologies.	24
Table 2.4	Screening methods for the enzymes.	24
Table 2.5	Major phyla obtained from 16S rRNA metagenomic studies that were conducted on the Antarctic region.	30
Table 2.6	The important features of QIIME2 in processing 16S rRNA amplicon metagenomics sequences.	31
Table 2.7	Functional metagenomics for novel enzyme studies from the Antarctic region.	36
Table 2.8	Protease activity screening through functional metagenomics approach.	36
Table 3.1	Chemicals and their brands used in this experiment.	39
Table 3.2	Nine different formulas of LB broth were used for optimization analysis.	40
Table 3.3	Coordinate of sample collected along the Southern Victoria Land (Courtesy of J. Symkla).	42
Table 3.4	PCR reaction conditions for a 50 μ L reaction.	45
Table 3.5	Independent variables and their corresponding levels for FCCCD in a response to the bacteria growth rate.	53
Table 3.6	Central Composite Design (CCD) for the chosen parameter.	54
Table 3.7	Standard and sample assay tubes with a total volume of 200 μ L.	56
Table 3.8	The number of reads for each sample before and after preprocessing.	60
Table 3.9	End-Repair reaction conditions of the insert DNA.	63
Table 3.10	Ligation reaction conditions of the insert DNA with pCC1FOS vector.	66

Table 4.1	Colony-forming unit (CFU/mL) for each sample on both Skim Milk and Luria-Bertani agar.	77
Table 4.2	Isolation of diverse bacteria from the Antarctic region that can produce cold-active proteolytic enzymes.	82
Table 4.3	Experimental design and results of face-centered central composite design (FCCCD).	96
Table 4.4	Analysis of variance (ANOVA) for response surface quadratic model.	96
Table 4.5	All samples were quantitatively analyzed using a photometer and fluorometer.	101
Table 4.6	The required conditions of samples for 16S rRNA amplicon sequencing.	102
Table 4.7	Fluorometric quantification of amplified products for each sample.	103
Table 4.8	OTU numbers and alpha diversity indices for sample ROB, ROS and SC.	104
Table 4.9	Fifteen OTU with highest total frequency and considered as dominant bacteria strains. The majority highest frequency of OTU found were uncultured bacterium.	109
Table 4.10	Amino acids composition between predictive and template model.	118

LIST OF FIGURES

Figure 2.1	Antarctic location at the South Pole of the Earth and recognized as one of the Earth's continental.	9
Figure 2.2	Blood Falls outflow that is originated below Taylor Glacier at McMurdo Dry Valleys, Antarctic (Mikucki et al., 2009).	10
Figure 2.3	The optimum cold-active enzyme activity (upper panel; blue) and enzyme unfolded (lower panel) shifted to lower temperature compared to mesophilic (red) or thermophilic enzymes. Figure adapted from Feller and Gerday (2003).	16
Figure 2.4	Compare to the lock and key model, in the induced fit model, the active site of the enzymes change conformation to perfectly bind with the substrates (Kokkonen et al., 2019).	17
Figure 2.5	The hierarchy of major taxonomic ranks. The higher level consisted of domain, kingdom, phylum, class, order and family level and the lower level consisted of genus and species level.	28
Figure 2.6	Nine hypervariable regions of 16S rRNA region. Specific primers and barcodes were attached to the V3 and V4 regions to amplify these regions (Dias et al., 2020).	29
Figure 2.7	Illustration of function metagenomics approach (Mirete et al, 2016).	33
Figure 2.8	The overview protocol for CopyControl™ Fosmid Library Production Kit with pCC1FOS™ Vector.	34
Figure 2.9	The vector map for pCC1FOS.	35
Figure 3.1	Map showed the location of the collected sample near the coastal area.	43
Figure 3.2	Flow diagram for screening proteolytic enzyme of Antarctica soil samples.	44
Figure 3.3	Thermo Scientific µDrop™ Plate was used to analyze the purity of the nucleic acid after PCR purification.	48
Figure 3.4	Azocasein consists of coupled diazotized sulfanilamide with the phenolic moieties of casein such a tyrosine and histidine residues (Dong, 2013).	52

Figure 3.5	The graph showed the quality score for forward reads (upper panel) and reverse reads (lower panel). The quality reads below 20 (red line) were discarded for further analysis.	59
Figure 3.6	The gel was cut according to the red line and the end-repair DNA as the insert DNA was avoided from being stained and exposed to UV light.	64
Figure 3.7	FASTQ format created by the Illumina sequencer contained four lines that began with @A00917.	70
Figure 4.1	A) Halo zone formation around the colonies indicated proteolytic activity after 3 months incubation period at 4°C. B) After re-screened to identify colonies with clear halo zone formation after 24 hours at 4°C, 8 strains were selected – BB5, DI22, ROB2, ROS7, ROS8, SC8, SC9 and SC11.	76
Figure 4.2	A) The CFU was counted on Luria-Bertani agar. (B) The CFU with halo zone was counter on skim milk agar. The samples in the graphs were referred to as BB –Botany Bay, CR – Cape Roys, DI – Dunlop Island, MP – Marble Point, ROS – Cape Roberts, ROS – Cape Ross and SC – Spike Cape.	77
Figure 4.3	Gel electrophoresis of bacterial DNA was extracted using the boiling method. All extracted DNA of 46 cold-active proteolytic bacteria showed a similar smearing DNA band.	80
Figure 4.4	PCR amplification gel electrophoresis of 16S rRNA gene fractions. All PCR products of 46 cold-active proteolytic bacteria showed a clear single DNA band at a 1.5 kb DNA ladder.	81
Figure 4.5	Neighbour-joining phylogenetic tree of isolated bacteria producing protease was constructed using MEGAX. <i>Saccharolobus shibatae</i> B12 was used as a control as an out of bacteria group. The optimal tree with the sum of branch length = 1.57172218 is shown. Bar, 0.1 substitutions per nucleotide position. The cluster subtree consisted of 19 OTUs (SC13, SC14, SC11, SC10, DI25, SC5, SC9, DI23, SC2, SC8, DI22, MP12, SC7, BB5, MP7, SC6, SC3, MP2, BB16).	84
Figure 4.6	Protease activity of eight selected bacteria strains after 24 hours incubation period at 20 °C and 200 rpm. <i>P. aeruginosa</i> (PA) was used as a positive control. Results were expressed as mean ± SEM. * p < 0.05	87
Figure 4.7	Effect of different temperature levels on the bacteria growth rate. The pH of the media and NaCl concentration were maintained at 7 and 1%, respectively.	90

Figure 4.8	Effect of different levels of pH in a response to the bacteria growth rate. The incubation temperature and NaCl concentration were maintained at 20 °C and 1%, respectively.	91
Figure 4.9	Effect of concentration of sodium chloride on the bacteria growth rate. The incubation temperature and pH of the media were maintained at 20 °C and 7, respectively.	92
Figure 4.10	Shifting of white colour skim milk to clearer media showed that proteolytic activity occurred in the media..	93
Figure 4.11	Results of crude protease assay at three different temperatures. SC8 showed the highest proteolytic activity at the lowest temperature. * $p < 0.05$	93
Figure 4.12	3D surface and contour graph showed the interaction of three independent variables with the response to the bacteria growth rate.	98
Figure 4.13	The extracted eDNA showed smearing bands for all samples ROB, ROS and SC.	100
Figure 4.14	Clear single band for each sample showed that the amplification of the V3-V4 region was successful.	103
Figure 4.15	Rarefaction curve of bacterial 16S rRNA sequences for three different samples.	104
Figure 4.16	Venn diagram of bacterial communities with shared and unique OTU among 3 different samples.	106
Figure 4.17	Stacked bar illustrated the different percentages of bacteria distribution between the samples.	107
Figure 4.18	(A) The vectors were blunt-ended but not 5'-phosphorylated and the DNA insert was blunt-ended and 5'-phosphorylated. Higher molar of vector than DNA insert would create linear dsDNA. (B) The uncut purified fosmid showed a single band, while the cut fosmid showed two separated bands.	112
Figure 4.19	The PCoA-tetranucleotide frequency of the assembled contigs. Contigs (NODE) outside of the eclipse were potential candidates for fosmid insert.	115
Figure 4.20	Alignment of NODE_67 to a fragment of the <i>E. coli</i> reference genome where mobile genetic elements (MGEs) were detected in this region.	116

- Figure 4.21 (A) Pfam result analysis showed that two domains section, Peptidase_M23 (green) and PG_binding_1 (red). (B) Analysis of conserved domain through NCBI showed that the selected contig's protein sequence hit specific domain Peptidase M23 and PG binding 1 superfamilies. 117
- Figure 4.22 A) Full amino acid sequence (297) of predicted protease (Pre_pro). B) Sequence alignment of predicted protease (Pre_pro) and the template enzyme (3SLU). There was not much difference that can be observed in the overall secondary structure of both enzymes (***) is conserved HXH motif). 120
- Figure 4.23 Predicted protease homology structure (blue) and the template enzyme (3SLU) (brown). Large deletion (red) was observed when both enzymes aligned together. 122
- Figure 4.24 The activity of the predicted protease displayed a similar profile to the crude enzyme extracted from SC8. Results were expressed as mean \pm SEM. * $p < 0.05$ 123
- Figure 5.1 Silverskin was a thin, shiny layer composed mainly of protein elastin, which was harder and chewier to eat (Goldwyn & Blonder, 2017). 125

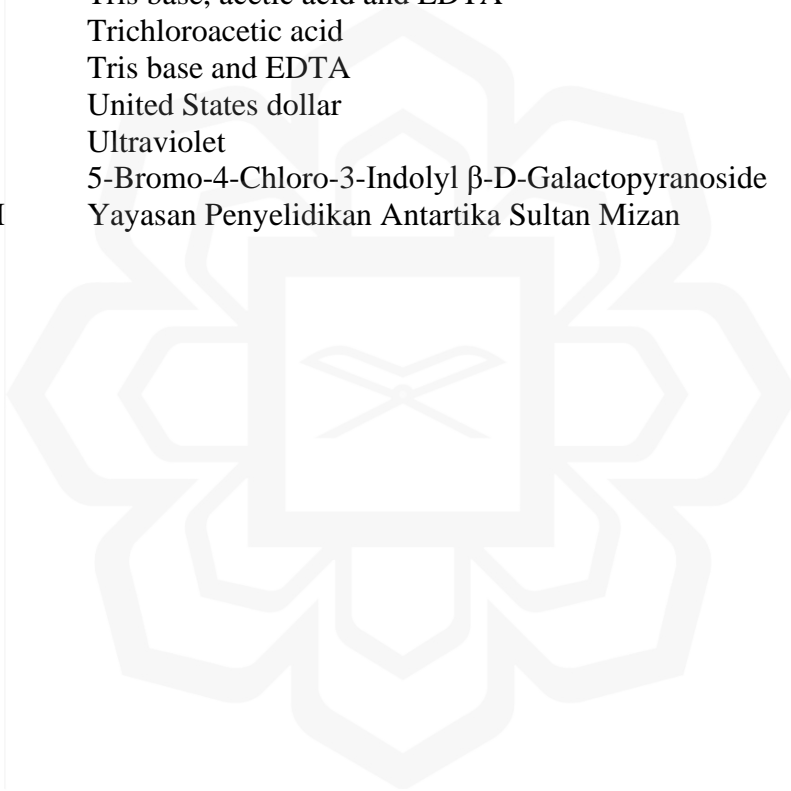
LIST OF SYMBOLS

%	Percentage
±	Plus minus
°C	Degree Celsius
μg	Microgram
μL	Microliter
μM	Micromolar
∞	Infinity
Å	Ångström
bp	Base pair
cm	Centimeter
g	Gram
kb	Kilo base pair
L	Liter
M	Molar mass
mg	Milligram
min	Minute
mL	Milliliter
mm	Millimeter
mol	Mole
nm	Nanometer
pg	Picogram
pmol	Picomole
rpm	Revolutions per minute
s	Second
V	Volt
x g	Relative centrifugal force
α	Alpha
β	Beta

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
ASV	Amplicon sequence variants
ATS	Antarctic Treaty System
ATCC	American Type Culture Collection
BAC	Bacteria alkaline phosphatase
BACs	Bacterial artificial chromosomes
BAS	British Antarctic Survey
BLAST	Basic Local Alignment Search Tool
CFU	Colony-forming unit
CIP	Calf-intestine alkaline phosphatase
CMC	Carboxymethyl cellulose
CTAB	Cetyl trimethylammonium bromide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
dsDNA	Double-stranded DNA
EC	Enzyme Commission
eDNA	Environmental DNA
EDTA	Ethylenediaminetetraacetic acid
FCCCD	Face-centered central composite design
Fe (III)	Ferric ion
Gly	Glycine
GPMAW	Guide to free protein properties calculator
HCl	Hydrochloric acid
KOH	Potassium hydroxide
KOPRI	Korea Polar Research Institute
LB	Luria-Bertani
LMP	Lower melting point
MgCl ₂	Magnesium chloride
MGE	Mobile genetic elements
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NaCl	Sodium chloride
NARC	National Antarctic Research Center
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
NIPJ	National Institute of Polar
OD	Optical density
OFAT	One factor at a time
ORF	Open reading frame
OTU	Operational taxonomic unit
PA	<i>Pseudomonas aeruginosa</i>
PCoA	Principal Coordinates Analysis
PCR	Polymerase chain reaction
PDB	Phage Dilution Buffer
PKC	Palm kernel cake
PVP	Polyvinylpyrrolidone

QIIME2	Quantitative Insights into Microbial Ecology
RAST	Rapid Annotations using Subsystems Technology
RFLP	Restriction fragment length polymorphism
RFU	Raw fluorescence units
RMSD	Root mean square deviation
ROB	Cape Roberts
ROS	Cape Ross
rRNA	Ribosomal ribonucleic acid
rSAP	Shrimp alkaline phosphatase
RSM	Response surface method
SEM	Standard error mean
SO ₄ ²⁻	Sulfate
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
SC	Spike Cape
TAE	Tris base, acetic acid and EDTA
TCA	Trichloroacetic acid
TE	Tris base and EDTA
USD	United States dollar
UV	Ultraviolet
X-gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside
YPASM	Yayasan Penyelidikan Antartika Sultan Mizan



CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

An enzyme is a biocatalyst that increases the chemical reaction rate by lowering the activation energy. The enzyme is important to accelerate the process of reactions. There are large numbers of reactions that need these biocatalysts to operate. The usage of biocatalyst in Malaysia is essential in many bio-industries and experimental processing. Therefore, a large number of enzymes must be imported to meet this growing demand and the cost of these imported enzymes is quite expensive. Annually, this country, Malaysia has spent about USD 3.5 million to import various types of enzymes, mostly from Belgium, Denmark and the Netherlands (Ibrahim, 2008). Although enzymes have several advantages in terms of efficiency, selectivity and environmental friendliness compared to the chemical catalyst, they are costly. Manufacturing or production of enzymes on a large scale to reach the demand for industrial biotechnology is almost non-existent in Malaysia. Many foreign companies like Novozymes and Danisco/Genencor dominate our local biocatalyst market (Biotechcorp, 2009). Realizing the importance and demand of these biocatalysts, it is high time for Malaysia to start the scaling-up process and increase our local enzyme production.

The major challenges to initiating enzyme production for industrial purposes in this country are high capital investment and high production costs at the early stage. However, if this county can manufacture the enzymes locally, these enzymes can be sold at a cheaper price and reduce the dependency on imported enzymes. The sources of the enzyme are from living things like plants, animals and microorganisms. However, due to their rapid growth and higher yield of production, microorganism becomes the

major source of enzyme, especially for commercialization purpose compared to plants and animals. Furthermore, the process of extracting enzymes from microorganisms is easier and less time-consuming. It can be easily controlled chemically and physiologically. If the cost of enzyme production can be reduced, this will benefit in terms economically, as the selling price will be much cheaper. Generally, there are four stages in producing enzyme industrially that include selection of enzyme, formulating the medium, process of production and purification of the enzyme. For the process of production, most industries in advanced countries used submerged and solid-state fermentation (Liu & Kokare, 2017). This process is preferable because of its lower production cost and lesser risk of contamination compared to others. Besides, the production of the enzyme can be increased by optimizing the conditions for microorganism growth or by gene modification.

Extremozymes are enzymes or biocatalysts that are extracted from extremophiles microorganisms such as bacteria, archaea, protozoa, algae, fungi, viruses, and helminths. These microorganisms inhabit extreme environments and pose enzymes that can functionally work in their surrounding environment. There are several categories of extremophiles such as thermophiles, psychrophiles, acidophiles, alkaliphiles, halophiles, piezophiles, osmophiles and radiophiles. These extremozymes are more efficient than genetically modified enzymes when applied in harsh industrial conditions because of their natural catalytic reaction in extreme environments. Currently, researchers are actively mining these extremozymes because of their advantages and potential to be applied in biotechnological or industrial applications.

The Antarctic is one of the regions that are rich and diverse with microorganism communities. Scientists have discovered that these microorganisms can produce enzymes that are capable to work efficiently at a lower temperature. Therefore, they can

survive under extremely cold conditions. However, the biodiversity of this microorganism is less explored so far. Identifying and isolating for a specific cold-active enzyme with higher stability is required. Besides, through genetic modification and strain improvement, this cold-active enzyme would play a more important role in various biotechnological industries and applications.

Protease is one of the hydrolase subclass enzymes that breakdowns protein into smaller fragments or amino acids with the addition of water molecules. In future, this valuable potential of cold-active protease would lead the enzyme market greater than thermostable enzymes. It is very useful for harnessing and bioprospecting this enzyme due to its greater potentials and chances for industrial usage in the future.

1.2 PROBLEM STATEMENT

A large number of enzyme strains have been identified via culture-based approaches. However, most of these culturable bacteria were categorized into similar phylogenetic affiliation (Vester et al., 2014). Axiomatically, bacteria are one of living things that also have a very complex life structure. Not all bacteria can be cultured. Therefore, to ensure better coverage of screening cold-active proteolytic enzymes, metagenomics approaches are initiated. Initially, amplicon metagenomics sequencing is implemented to analyze the total microbial diversity of the sample including culturable and uncultured bacteria. While functional metagenomics would allow the total genome to be analyzed in the suitable host as recombinant enzymatic screening for its functionality. These culture-independent approaches can screen these targeted enzymes from other bacteria groups.

Enzymes are well-known biocatalyst that works under a specific range of environmental conditions. The usage of enzymes is increasing especially in industrial

applications due to their environmentally characteristics. As the output demand increase, the harshness of the industrial process also increases to produce more products. Consequently, extremozymes were explored to meet the harshness conditions process where their predecessor was incapable. A cold-active enzyme is also classified as extremozymes. This enzyme enables the system to run at low temperatures. The majority of commercial or industrial enzymes are mesophilic that required high temperatures for efficient catalytic activity. Therefore, the system needs to be heated for the enzyme to function effectively. In the interest of profits, it is considered time and energy-consuming. In addition, a higher temperature is also needed to deactivate this enzyme through the denaturation process. Consequently, introducing higher temperatures in the system will always initiate undesirable chemical reactions and loss of volatile compounds. Therefore, the cold-active enzyme has valuable potential as an alternative to overcome these limitations of these commercial or industrial enzymes.

1.3 RESEARCH QUESTIONS

The study aimed to answer the following questions:

- 1- What are the potential bacteria that can produce cold-active proteases in the Antarctic soil?
- 2- What are the optimal growth conditions of the selected bacteria that can produce cold-active protease?
- 3- Are there any unculturable bacteria and functional cold-active protease that can be identified using the metagenomics approaches?

1.4 RESEARCH OBJECTIVES

The study aimed to achieve the following objectives:

- 1- To isolate and identify bacteria from Antarctic regions that can produce cold-active proteolytic enzymes based on the 16S rRNA gene.
- 2- To optimize the growth conditions of selected strains that can produce the cold-active proteolytic enzyme.
- 3- To analyze uncultured bacteria diversity of the sample through amplicon metagenomics sequencing and identify clones with cold-active protease through functional metagenomics.

1.5 SIGNIFICANCE OF THE STUDY

This study is important as a preliminary study for bioprospecting of cold-active protease from Antarctic regions. The discovery of cold-active protease in this study will include both culturable and non-culturable modes. Hence, the total discovery of cold-active protease could be assessed and analysed from the Antarctic samples.

Initially, this study will isolate and identify the bacteria that can produce cold-active protease enzymes. These bacteria have the potential as a candidate for high-scale production of cold-active protease for specific industrial and biotechnology applications. In addition, the discovery of cold-active protease via metagenomics and functional metagenomics studies not only will provide the taxonomic status of unculturable bacteria but also the potential novel cold-active protease. The terms novel could be clarified in many ways such as enzyme structure, function and working conditions.