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 ٤ °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 ٢٤ 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 للمقايسة 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 ٤٦ سلالة بكتيرية ذات نشاط بروتياز إيجابي وأظهر تحليل النشوء والتطور أن ٨٨٪ كانت من *Pseudomonas* sp. و ٩٪ من *Arthrobacter* sp. و ٣٪ من *Paenibacillus* sp. أظهرت نتيجة OFAT أن ١٠ سلالات مختارة أظهرت أعلى معدل نمو عند ٢٠ °C، ودرجة الحموضة ٧ و ٤٪ (w/v) من كلوريد الصوديوم. أظهر الفحص الأنزيمي أن الإنزيم الخام المستخرج من سلالة SC8أظهر نشاطًا أعلى بكثير من التحكم الإيجابي (الأنزيم البروتيني من البنكرياس البقري) عند -٢٠ °C و ٢٠ °C. علاوة على ذلك ، اقترح RSM أن الظروف المثلى لنمو سلالة SC8 كانت عند ٢٠.٥ °C ، ودرجة الحموضة ٦.٨٣ ، و ٢.٠٥ ٪ (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 ٨٩٣ نقطة أساس معنوية مطابقة لعائلات بروتين Peptidase M23 وPG ملزمة ١ عائلة بروتين. في تحليل السيليكو أظهر أن هذا البروتياز المتوقع أظهر تشابهًا بنسبة ٢٧.٠٧ ٪ 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 البروتيز 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.

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

بسم الله الرحمن الرحيم

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 الله

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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	*Pseudomonas aeruginosa*
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-existed 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.