CHARACTERIZATION OF A THERMOSTABLE CELLULASES FROM THERMOFLAVIFILUM AGGREGANS SP1 ISOLATED FROM PORING HOT SPRING SABAH, MALAYSIA

NURSHAFRINA AIDA BINTI YAHYA

BIOTECHNOLOGY RESEARCH INSTITUTE

UNIVERSITI MALAYSIA SABAH

2023

CHARACTERIZATION OF A THERMOSTABLE CELLULASES FROM THERMOFLAVIFILUM AGGREGANS SP1 ISOLATED FROM PORING HOT SPRING SABAH, MALAYSIA

NURSHAFRINA AIDA BINTI YAHYA

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

BIOTECHNOLOGY RESEARCH INSTITUTE

UNIVERSITI MALAYSIA SABAH

2023

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL : **CHARACTERIZATION OF A THERMOSTABLE CELLULASES FROM THERMOFLAVIFILUM AGGREGANS SP1 ISOLATED FROM PORING HOT SPRING SABAH, MALAYSIA**

IJAZAH : **SARJANA SAINS**

BIDANG : **BIOTEKNOLOGI**

Saya **NURSHAFRINA AIDA BINTI YAHYA**, Sesi **2018-2023**, mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis ini adalah hak milik Universiti Malaysia Sabah
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/):

☐	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)
☐	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
☒	TIDAK TERHAD	

Disahkan Oleh,

shafrina

NURSHAFRINA AIDA BINTI YAHYA
MZ1821104T

ANITA BINTI ARSAD
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH
(Tandatangan Pustakawan)

Tarikh : 11 Mei 2023

(Dr. Cahyo Budiman)
Penyelia Utama

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries and references, which have been duly acknowledged.

14 December 2022

Nurshafrina Aida binti Yahya

MZ1821104T

CERTIFICATION

NAME	:	**NURSHAFRINA AIDA BINTI YAHYA**
MATRIC NUM.	:	**MZ1821104T**
TITLE	:	**CHARACTERIZATION OF A THERMOSTABLE CELLULASES FROM THERMOFLAVIFILUM AGGREGANS SP1 ISOLATED FROM PORING HOT SPRING SABAH, MALAYSIA**
DEGREE	:	**MASTER OF SCIENCE**
FIELD	:	**BIOTECHNOLOGY**
VIVA DATE	:	**14 DECEMBER 2022**

CERTIFIED BY;

Signature

1. **MAIN SUPERVISOR**
   Dr. Cahyo Budiman

2. **CO - SUPERVISOR**
   Professor Dr. Clemente Michael Wong Vui Ling

3. **CO - SUPERVISOR**
   Dr. Zarina Amin

ACKNOWLEDGEMENTS

In The Name of Allah, Most Gracious, Most Merciful

Assalamualaikum W.B.T

Alhamdulillah to Allah S.W.T that finally, with His blessings and permission as well as the cooperation from those who are involved in this study, I am able to complete it. I would like to say thank you, especially my supervisor, Dr. Cahyo Budiman and also my co supervisor Prof. Dr. Clemente Michael Wong Vui Ling and Dr. Zarina Amin for all the help, support, comments and guidance in helping me to finish this thesis. Furthermore, I would like to say thank you to all other lecturers that provided me with comments and advices to improve my work.

Also not to be forgotten, I would also like to thank you to my parents for their encouragement and financial assistance which had helped me to go through all the obstacles and challenges throughout this research study.

Last but not least, thank you for all of my friends from our team (Protein group) and others for the continuous help and support in completing my thesis.

Nurshafrina Aida binti Yahya

14 December 2022

ABSTRACT

Thermophilic microorganisms and their enzymes have been utilized in various industrial applications. Cellulases are essential enzymes in various industries and have great significance in present-day biotechnology. Currently, thermostable cellulases have been steadily increasing in demand due to their versatile applications under harsh conditions. Nevertheless, the availability of thermostable cellulases remains limited. Sabah houses hot springs and mud volcanoes that are promising as the sources of indigenous thermophilic bacteria producing thermostable cellulases. This study is aimed to isolate and characterize thermophilic cellulose-degrading bacteria from Sabah hot springs and mud volcanoes. To address, the samples were collected from Poring hot spring and Tawau mud volacanoes and spread onto the carboxymethylcellulose (CMC) agar medium for the screening of cellulose-degrading bacteria. Based on the biocehmical test, SP1 was chosen and undergone whole genome sequencing, assembly and annotation. The gene was then cloned into pET-28a (+) and transformed into *Escherichia coli* BL21(DE3). The expression of this protein was successfully performed by induction of 0.2 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) at 25 °C for overnight. The expressed protein was then successfully purified using Ni-NTA affinity chromatography and size-exclusion chromatography. As results, out of 6 isolated colonies exhibiting cellulose degradation activities, colony SP1 from Poring hot spring displayed the highest cellulolytic activity on crude enzyme at 60 °C. Further, 16S rRNA sequence analysis of SP1 showed closest similarity to *Thermoflavifilum aggregans* (accession no: AM749771), a thermostable bacterium isolated from New Zealand hot spring, with 99.74 % homology. Accordingly, SP1 is designated as *T. aggregans* SP1 strain. To note, the strain SP1 is the first strain from Poring hot spring known to exhibit thermostable cellulolytic activity. Whole genome sequence (WGS) of SP1 strain was then decoded using the Pacific Biosciences Single Molecule, Real-Time sequencing platform which revealed the genome size is 2,874,051 bp with the presence of three (3) genes encoding cellulose-degrading enzymes which might responsible for its celluloytic activity. From the three genes, one gene is from glycosyl hydrolase (GH)-5 and another two genes from GH-9 family members. Due

to the unique properties of GH-5, one gene of GH-5 (designated as CePH4) was selected for further characterization. CePH4 is 1053 bp in size, which encodes a 347-residue polypeptide with theoretical molecular mass of 38 kDa. The sequence analysis of CePH4 indicated that this protein is organized into a catalytic domain of GH5 and an aryl-phospho-beta-D-glucosidase (BglC), with canonical active sites of Glu151 and Glu271. CePH4 was successfully expressed and purified with the yield of 24.22 mg per 1000 mL culture, with apparent size of 38 kDa on SDS. Size exclusion chromatography showed that CePH4 is a monomeric protein. Circular dichroism spectroscopy revealed that CePH4 has melting temperature (Tm) of 80 °C, indicating that this protein is a thermostable enzyme. Further, the specific activity of purified CePH4 against CMC substrate was 7.46 U/mg with the optimum activity at 70 °C and pH 7. Interestingly, up to 80°C, the activity of CePH4 decreased by up to 40 % only. Our findings suggest promising applications of these thermoaerobic bacteria and their potent enzymes for industrial purposes.

Keywords: Thermostable, cellulases, hot springs, whole genome sequence

ABSTRAK

CIRI-CIRI SEL TERMOSTABIL DARIPADA THERMOFLAVIFILUM AGGREGANS SP1 DIASINGKAN DARIPADA PORING HOT SPRING SABAH, MALAYSIA

Mikroorganisma termofilik dan enzimnya telah digunakan dalam pelbagai aplikasi perindustrian. Selulase adalah enzim penting dalam pelbagai industri dan mempunyai kepentingan yang besar dalam bioteknologi masa kini. Pada masa ini, selulase termostabil telah semakin meningkat dalam permintaan kerana aplikasi serba bolehnya di dalam keadaan yang melampau. Namun begitu, ketersediaan selulase termostabil kekal terhad. Sabah menempatkan mata air panas dan gunung berapi lumpur yang menjanjikan sebagai sumber bakteria termofilik asli yang menghasilkan selulase termostabil. Kajian ini bertujuan untuk mengasingkan dan mencirikan bakteria pengurai selulosa termofilik daripada mata air panas Sabah dan gunung berapi lumpur. Untuk menjawabnya, sampel telah dikumpulkan dari mata air panas Poring dan gunung berapi lumpur Tawau dan disebarkan ke medium agar carboxymethylcellulose (CMC) untuk saringan bakteria yang menguraikan selulosa. Berdasarkan ujian biokimia, SP1 telah dipilih dan menjalani penjujukan, pemasangan dan anotasi genom keseluruhan. Gen tersebut kemudiannya diklonkan kepada pET-28a (+) dan dimasukkan ke dalam Escherichia coli BL21(DE3). Ekspresi protein ini berjaya dilakukan dengan induksi 0.2 mM isopropil ß-D-1-thiogalactopyranoside (IPTG) pada suhu 25 °C untuk semalaman. Protein kemudiannya berjaya ditulenkan menggunakan kromatografi afiniti Ni-NTA dan kromatografi pengecualian saiz. Hasilnya, daripada 6 koloni terpencil yang mempamerkan aktiviti degradasi selulosa, koloni SP1 dari mata air panas Poring menunjukkan aktiviti selulolitik tertinggi untuk enzim mentah pada 60 °C. Selanjutnya, analisis jujukan 16S rRNA SP1 menunjukkan persamaan yang paling hampir dengan Thermoflavifilum aggregans (nombor kesertaan: AM749771), bakteria termostabil yang diasingkan daripada mata air panas New Zealand, dengan 99.74 % homologi. Sehubungan itu, SP1 ditetapkan sebagai T. aggregans SP1 strain. Untuk makluman, terikan SP1 ialah terikan pertama dari mata air panas Poring yang diketahui mempamerkan aktiviti selulolitik termostabil. Keseluruhan penjujukan genom strain SP1 kemudiannya dinyahkodkan menggunakan Pacific

Biosciences Single Molecule, platform penjujukan Masa Nyata yang mempunyai saiz genom 2,874,051 bp dengan kehadiran tiga (3) gen yang mengekod enzim pengurai selulosa yang mungkin bertanggungjawab untuk aktiviti selulolitiknya. Daripada tiga gen tersebut, satu gen adalah daripada glikosil hidrolase (GH)-5 dan dua lagi gen daripada ahli keluarga GH-9. Disebabkan sifat unik GH-5, satu gen GH-5 (ditetapkan sebagai CePH4) telah dipilih untuk pencirian selanjutnya. CePH4 bersaiz 1053 bp, yang mengodkan polipeptida 347-residu dengan jisim molekul teori 38 kDa. Analisis jujukan CePH4 menunjukkan bahawa protein ini disusun ke dalam domain pemangkin GH5 dan aril-phospho-beta-D-glucosidase (BglC), dengan tapak aktif kanonik Glu151 dan Glu271. CePH4 berjaya diekspresikan dan ditulenkan dengan hasil 24.22 mg setiap 1000 mL kultur, dengan saiz ketara 38 kDa pada SDS. Kromatografi pengecualian saiz menunjukkan bahawa CePH4 ialah protein monomerik. Spektroskopi dichroism bulat mendedahkan bahawa CePH4 mempunyai suhu lebur (Tm) 80 °C, menunjukkan bahawa protein ini adalah enzim termostabil. Selanjutnya, aktiviti spesifik CePH4 yang telah ditulenkan terhadap substrat CMC ialah 7.46 U/mg dengan aktiviti optimum pada 70 °C dan pH 7. Menariknya, sehingga 80°C, aktiviti CePH4 berkurangan sehingga 40 % sahaja. Penemuan kami mencadangkan aplikasi menjanjikan bakteria termoaerobik ini dan enzim kuatnya untuk tujuan perindustrian.

Kata kunci: Thermostabil, selulase, air panas, keseluruhan jujukan genom

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

°C	-	Degree celcius
%	-	Percentage
µL	-	Microliter
µmol	-	Micromol
w/v	-	Weight over volume
mM	-	Millimolar
U	-	Unit
ml	-	Mililiter
µg	-	Microgram
kb	-	Kilobase pair
Å	-	Angstrom
kcal/mol	-	kilocalorie per mole
pM	-	picomole
K	-	Kelvin
mg/ml	-	Miligram per milliliter
ml/min	-	Milliliter per minute
kDa	-	Kilodaltons
nm	-	Nanometer
mm	-	Millimeter
bp	-	Base pair
β	-	Beta
α	-	Alpha

LIST OF ABBREVIATIONS

AI	-	Aliphatic index
BglC	-	Aryl-phospho-beta-D-glucosidase
$CaCl_2$	-	Calcium chloride
CBB	-	Coomassie brilliant blue
CBM	-	Carbohydrate-binding module
CDS	-	Protein-coding sequences
CMC	-	Carboxymethyl cellulose
CI	-	Cellulolytic index
$CuSO_4$	-	Copper(II) sulfate
COG	-	Clusters of orthologous groups
GH	-	Glycoside Hydrolase
gDNA	-	Genomic DNA
GO	-	Gene ontology
GRAVY	-	Grand Average Hydropathy
GMQE	-	Global Model Quality Estimation
HGAP4	-	Hierarchical genome-assembly process 4
HGT	-	Horizontal gene transfer
HMF	-	5-hydroxymethylfurfural
HSP	-	Heat shock protein
IPTG	-	Isopropyl -D-1-thiogalactopyranoside
KCl	-	Potassium chloride
KEGG	-	Kyoto Encyclopedia of Genes and Genomes
MD	-	Molecular docking
NaCl	-	Sodium chloride
NADH	-	Nicotinamide adenine dinucleotide
NCBI	-	National Center for Biotechnology Information
OD600	-	Optical density at 600 nm
PDB	-	Protein information bank
PEG	-	Protein encoding gene
pI	-	Isoelectric point
QMEAN	-	Qualitative Model Energy Analysis
Rpm	-	Revolutions per minute
RAST	-	Rapid Annotations using subsystems Technology
RMSF	-	Root means square fluctuation
Sdn Bhd	-	Sendirian Berhad

SEM	-	Scanning electron microscope
SDS-PAGE	-	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
sHSP	-	Small heat shock proteins
SMRTbell	-	Single-molecule, real-time bell
TIM	-	Triose-phosphate isomerase
Tm	-	Transition temperature
UV CD	-	Ultraviolet circular dichroism
YASARA	-	Yet Another Scientific Artificial Reality Application
$ZnSO_4$	-	Zinc sulphate
3D	-	Three-dimensional

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Lignocellulosic biomass is the world's most economical and highly renewable natural resource. The development of renewable energy converted from lignocellulosic biomass as an alternative to fossil fuel is ultimately essential for the survival of humans (Zoghlami & Paes, 2019). Apart from its application as a material for producing renewable energy, there has also been increased interest in developing lignocellulosic biomass-derived platform chemicals for biobased polyurethane applications. Technologies have been developed to produce various platform chemicals, such as sugar alcohols, organic acids, furfural, and 5-hydroxymethylfurfural (HMF), from lignocellulosic biomass via biorefining technologies (Kohli *et al.*, 2019). For these purposes, cellulose degrading enzyme or cellulases (E.C. 3.2.1.4) is required to process lignocellulosic biomass. The enzymes work through hydrolysis of cellulose to its simple sugar for further conversion to bioproducts through fermentation or other chemical processes. In nature, complete hydrolysis of cellulose is done by the three main cellulases types: endoglucanase, exoglucanase, and beta-glucosidase, which converts cellulose into sugar (glucose) (Jayasekara & Ratnayeke, 2019).

Nowadays, cellulases have been gaining broad interest in the industry due to their ability to replace inorganic catalysts for more environmentally friendly industrial processes and minimize carbon emissions. In Malaysia, cellulase enzymes are majorly

utilized in the industry related to edible oil and palm oil products, the oleochemical industry, the detergent industry, and the food and beverages, manufacturing, animal feed, and baking industries (Kiew *et al*., 2012). Notably, Sabah is known to be the biggest producer of palm oil products. In 2021, national production reached 26.48 million tonnes, with more than 6.29 million tonnes being contributed by Sabah (Malaysian Palm Oil Board, 2021). This huge industry leads to a huge number of by-products or cellulosic wastes, which can be converted into value-added products such as bioethanol and sustainable bio-based products (Kunasundari *et al*., 2016; Goh *et al*., 2010). Such processes are usually triggered by cellulose degrading enzymes produced by specific bacteria and microorganisms.

A critical bottleneck to the wide industrial use of hydrolytic enzymes is that they are often required to perform under harsh conditions, including temperatures above 40 °C. As a result, the enzyme stability against elevated temperatures is a crucial prerequisite before the broad industrial use of this type of enzyme is realized (Santos & da Costa, 2002). Accordingly, thermostable enzymes and microorganisms have been topics of much research during the last two decades (Singh *et al*., 2016). In industrial applications, isolated thermostable enzymes are today dominating over thermostable microorganisms. In this respect, the thermostable enzymes undergo some purification process from target microorganisms. An enzyme or protein is classified as thermostable when it exhibits optimum activity between 60 - 80 °C (Vieille & Zeikus, 2001). In some cases, enzymes also exhibited optimum activity at temperatures higher than 80 °C, which is then classified as hyperthermophilic (Viele *et al*., 1996). Active at high temperatures, thermophilic and hyperthermophilic enzymes typically do not function below 40 °C. Meanwhile, as producers of thermostable enzymes, thermostable microorganisms are usually characterized by their optimum growth temperature range of 50 - 80 °C (Vieille & Zeikus, 2001).

In nature, fungi tend to produce more cellulases than bacteria, but cellulases produced by bacteria are better catalysts as they encounter less feedback inhibition. Moreover, bacteria not only grow rapidly compared to fungi, allowing for higher