ELUCIDATION OF STRUCTURE AND FUNCTION OF CONSERVED HYPOTHETICAL PROTEINS RELATED TO THERMAL STRESS RESPONSE IN *Pedobacter cryoconitis* BG5 AND *Glaciozyma antarctica* PI12

MAKDI MASNODDIN

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITI MALAYSIA SABAH
2022

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN TESIS

JUDUL : ELUCIDATION OF STRUCTURE AND FUNCTION OF CONSERVED HYPOTHETICAL PROTEINS RELATED TO THERMAL STRESS RESPONSE IN Pedobacter cryoconitis BG5 AND Glaciozyma antarctica PI12

IJAZAH : DOKTOR FALSAFAH

SAYA : MAKDI MASNODDIN **SESI PENGAJIAN :** 2017 - 2022

(HURUF BESAR)

Mengaku membenarkan tesis *(LPSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis 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 AKTA RAHSIA RASMI 1972)
[ ]	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
[/]	TIDAK TERHAD	

(TANDATANGAN PENULIS)

Alamat Tetap: NO.6, LORONG 2 KINGFISHER SULAMAN, 88450 KOTA KINABALU, SABAH.

TARIKH: 15/9/2022

Disahkan oleh:

JACKLYNE JOHANIS @
NORAZLYNNE MOHD JOHAN
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH

(TANDATANGAN PUSTAKAWAN)

Dr. Nur Athirah Yusof

(NAMA PENYELIA)

TARIKH: _____

Catatan:
*Potong yang tidak berkenaan.
*Jika tesis ini SULIT dan TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT dan TERHAD.
*Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana Secara Penyelidikan atau disertai bagi pengajian secara kerja kursus dan Laporan Projek Sarjana Muda (LPSM).

ELUCIDATION OF STRUCTURE AND FUNCTION OF CONSERVED HYPOTHETICAL PROTEINS RELATED TO THERMAL STRESS RESPONSE IN *Pedobacter cryoconitis* BG5 AND *Glaciozyma antarctica* PI12

MAKDI MASNODDIN

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITI MALAYSIA SABAH
2022

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL : **ELUCIDATION OF STRUCTURE AND FUNCTION OF CONSERVED HYPOTHETICAL PROTEINS RELATED TO THERMAL STRESS RESPONSE IN *Pedobacter cryoconitis* BG5 AND *Glaciozyma antarctica* PI12**

IJAZAH : **DOKTOR FALSAFAH SAINS GUNAAN**

BIDANG : **BIOTEKNOLOGI**

Saya **MAKDI MASNODDIN**, Sesi **2017-2022**, mengaku membenarkan tesis Doktoral 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	

MAKDI MASNODDIN
DZ1721007T

Tarikh : 5 Julai 2022

Disahkan Oleh,

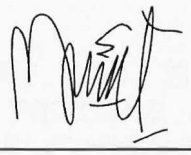
JACKLYNE JOHANIS @
NORAZLYNNE MOHD JOHAN
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH

(Tandatangan Pustakawan)

(Dr. Nur Athirah Yusof)
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.



27 April 2022

Makdi Masnoddin
DZ1721007T

CERTIFICATION

NAME : **MAKDI MASNODDIN**

MATRIC NO. : **DZ1721007T**

TITTLE : **ELUCIDATION OF STRUCTURE AND FUNCTION OF CONSERVED HYPOTHETICAL PROTEINS RELATED TO THERMAL STRESS RESPONSE IN *Pedobacter cryoconitis* BG5 AND *Glaciozyma antarctica* PI12**

DEGREE : **DOCTOR OF PHILOSOPHY IN APPLIED SCIENCE**

FIELD : **BIOTECHNOLOGY**

VIVA DATE : **27 APRIL 2022**

CERTIFIED BY;

Signature

1. **MAIN SUPERVISOR**
   Dr. Nur Athirah Yusof

2. **CO - SUPERVISOR**
   Prof. Clemente Michael Wong Vui Ling

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Dr. Nur Athirah Yusof, for encouragement, guidance, critics and friendship. I am also very thankful to my co-supervisor Professor Dr Clemente Michael Wong Vui Ling for guidance, advice, and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Universiti Malaysia Sabah (UMS) and Ministry of Higher Education (MOHE) for funding my Ph.D study. Malaysian Genome Institute (MGI) and Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM) for providing us with the facilities and equipment for crystal screening work.

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all of my family members, and I would like to dedicate this thesis to my beloved mother, Jamariah (1953–2010), and sister, Musmah (1978–2021).

Makdi Masnoddin
27 April 2022

ABSTRACT

The genome data of two native Antarctic microbes, *Pedobacter cryoconitis* (bacterium) and *Glaciozyma antarctica* (yeast), highlighted the presence of specific proteins with unique adaptive features. However, most of these proteins were designated as conserved hypothetical proteins (HPs), complicating efforts to understand their cellular functions. Consequently, we aim to identify the structural features of the conserved HPs that were ideal for their function in response to temperature stress. We posit that proteins that respond to thermal stress may have unique structural flexibility and stability that allows them to function under thermal stress, protecting host organisms against cold denaturation and heat aggregation. To address, an *in silico* analysis of HPs was conducted, followed by an *in vitro* approach whereby all selected HPs were cloned, expressed in *Escherichia coli*, purified, and subjected to crystal screening experiment. Purified recombinant proteins were assessed via colorimetric ATPase and citrate synthase aggregation assays and quantitative PCR (qPCR) gene expression analysis. Finally, the three-dimensional (3D) structures of the HPs were constructed and further refined for comparative structure analysis and function relationship clarification. Twelve *P. cryoconitis* and four *G. antarctica* conserved HPs with significant thermal stress response functions that met crystallisation criteria were identified through *in silico* analysis. All target genes were successfully amplified and cloned *in vitro*. Three *P. cryoconitis* (Pcbg5HP1, Pcbg5HP2, and Pcbg5HP12) recombinant proteins were overexpressed in the soluble forms at 16°C and subsequently purified using a two-step purification process. Three recombinant proteins from *G. antarctica* (GaHP2, GaHP3, and GaHP4) were overexpressed in soluble forms at 20°C, but only GaHP2 and GaHP3 were successfully purified. The ATPase assay showed protein activity at 4°C and 25°C for Pcbg5HP1, Pcbg5HP2, Pcbg5HP12, GaHP2, and GaHP3, which thus clarified that protein activity is maintained at low and moderate temperatures. Meanwhile, lower citrate synthase aggregation at 43°C in the presence of either Pcbg5HP1 or GaHP2 suggested the characteristics of chaperone-like activity. The qPCR analysis revealed that these genes were expressed constitutively when cells were exposed to temperatures below or above their optimal growth temperature, indicating their involvement in cellular processes associated with thermal stress. Initial crystal formation was observed when purified Pcbg5HP2 proteins were incubated at 4°C, and reagent optimization revealed the formation of a plate-shaped crystal in reagent 0.2 M potassium sodium tartrate tetrahydrate, 30% PEG/Ion, pH 7.4. This clarified the crystallisation potential of the HPs as predicted by bioinformatics analysis. Furthermore, comparative structural analysis demonstrated that the HPs exhibited cold-adapted traits, most notably increased flexibility in their 3D structures compared to their mesophilic or thermophilic counterparts. Concurrently, the presence of a disulphide bridge and aromatic clusters was attributed to Pcbg5HP1 and GaHP2's unusual protein stability and chaperone activity. Thus, this demonstrated that the HPs examined in this study adopted strategies to maintain a balance between molecular stability and structural flexibility, which contributed to their flexibility and ability to retain protein activities in an extreme environment. Conclusively, this study has established the structure-function relationships of the HPs produced by *P. cryoconitis* and *G. antarctica* and provided crucial experimental evidence indicating their importance in thermal stress response.

ABSTRAK

PENCIRIAN STRUKTUR DAN FUNGSI PROTEIN HIPOTETIK TERPULIHARA YANG BERKAITAN DENGAN TINDAK BALAS TERHADAP HABA DALAM Pedobacter cryoconitis BG5 DAN Glaciozyma antarctica PI12

Data genom dua mikroorganisma Antartika, Pedobacter cryoconitis *(bakteria) dan* Glaciozyma antarctica *(yis), mengandungi protein dengan ciri penyesuaian unik. Namun, sebahagian besar protein ini dikategorikan sebagai protein hipotetik (HP) terpulihara, yang merumitkan pemahaman tentang fungsinya. Oleh itu, kajian ini bertujuan untuk menentukan ciri struktur HP terpulihara yang berkaitan dengan fungsinya. Hipotesis kajian adalah protein ini mungkin mempunyai fleksibiliti dan kestabilan struktur tersendiri yang membolehkan ia terus berfungsi dan melindungi organisma perumah daripada suhu melampau. Analisis* in silico *telah dijalankan, kemudian analisis in vitro di mana semua HP yang dipilih telah diklon, diekspres dalam* Escherichia coli*, ditulenkan, dan digunakan untuk eksperimen penghabluran. Protein rekombinan yang telah ditulenkan, diuji melalui ujian kolorimetrik ATPase dan pengagregatan sitrat sintase, serta analisis kuantitatif ekspresi gen PCR (qPCR). Pada akhirnya, struktur tiga dimensi (3D) HP dibina dan dimurnikan untuk perbandingan struktur dan penentuan fungsi. Dua belas HP daripada* P. cryoconitis *dan empat HP daripada* G. antarctica *dengan fungsi berkaitan perubahan haba dan sesuai untuk penghabluran telah ditentukan melalui analisis* in silico*. Semua gen sasaran berjaya diamplifikasi dan diklon secara* in vitro*. Tiga protein rekombinan* P. cryoconitis *(Pcbg5HP1, Pcbg5HP2, dan Pcbg5HP12) telah diekspreskan secara larut pada 16°C dan ditulenkan melalui purifikasi dua peringkat. Tiga protein rekombinan* G. antarctica *(GaHP2, GaHP3, dan GaHP4) berjaya diekspreskan dalam bentuk larut pada* 20°C*, tetapi hanya GaHP2 dan GaHP3 berjaya ditulenkan. Ujian ATPase menunjukkan aktiviti protein Pcbg5HP1, Pcbg5HP2, Pcbg5HP12, GaHP2, dan GaHP3 pada 4°C dan 25°C. Ini menunjukkan protein tersebut kekal aktif pada suhu rendah dan sederhana. Agregasi sitrat sintase yang lebih rendah pada 43°C dengan kehadiran Pcbg5HP1 atau GaHP2 pula menunjukkan sifat seperti pengantar. Analisis qPCR menunjukkan gen-gen tersebut diekspres secara berterusan apabila sel terdedah kepada suhu di bawah atau di atas suhu pertumbuhan optimum, justeru mengesahkan ia terlibat dalam proses tindak balas terhadap perubahan haba. Pembentukan hablur dikesan pada protein tulen Pcbg5HP2 pada suhu inkubasi 4°C. Seterusnya pembentukan hablur berbentuk plat dihasilkan dengan optimasi bahan uji 0.2 M kalium natrium tartrat tetrahidrat, 30% PEG/Ion, pH 7.4. Ini mengesahkan potensi penghabluran HP yang diramalkan oleh analisis bioinformatik. Analisis struktur HP menunjukkan ciri-ciri penyesuaian pada suhu sejuk, terutamanya fleksibiliti yang tinggi pada struktur 3D berbanding protein mesofilik atau termofilik. Kewujudan ikatan disulfida dan kelompok aromatik pula boleh dikaitkan dengan kestabilan dan aktiviti chaperon bagi Pcbg5HP1 dan GaHP2. Ini menunjukkan bahawa HP-HP dalam kajian ini menggunakan strategi kestabilan molekul dan fleksibiliti struktur, yang menyumbang kepada fleksibiliti dan keupayaan mereka untuk mengekalkan aktiviti dalam persekitaran ekstrim. Kajian ini telah menyelesaikan hubungan antara struktur dan fungsi HP-HP yang dihasilkan oleh* P. cryoconitis *dan* G. antarctica *dan menghasilkan data eksperimen yang menunjukkan kepentingannya dalam tindak balas terhadap perubahan haba.*

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

°C	-	Degree Celsius
kDa	-	Kilo Dalton
mL	-	Mili liter
µL	-	Micro liter
µM	-	Micro gram
mM	-	Mili molar
mg	-	Mili gram
nm	-	Nano meter
mA	-	Mili ampere
pH	-	Potential of hydrogen
bp	-	Base pairs
kb	-	Kilo base pairs

LIST OF ABBREVIATIONS

HP	-	Hypothetical proteins
3D	-	Three-dimensional
2D	-	Two-dimensional
1D	-	One-dimensional
FPLC	-	Fast protein liquid chromatography
SEC	-	Size exclusion chromatography
NCBI	-	National Center for Biotechnology Information
BLAST	-	Basic Local Alignment Search Tool
IPTG	-	Isopropyl β-D-1-thiogalactopyranoside
DUFs	-	Domain of unknown function
HMM	-	Hidden Markov Model
pI	-	Isoelectric point
ROS	-	Reactive oxygen species
SDS-PAGE	-	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
TEMED	-	Tetramethylethylenediamine
RAST	-	Rapid Annotation using Subsystem Technology
LIC	-	Ligation Independent Cloning (LIC)
LB	-	Luria-Bertani
UV/VIS	-	UltraViolet-Visible

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Background

Antarctica is a fascinating region with a diverse range of climatic extremes and is generally considered the centre of the Earth's climate and marine circulation systems (Convey & Peck, 2019). Two previously discovered and characterised native Antarctic microbes, *Pedobacter cryoconitis* and *Glaciozyma antarctica*, have generated research interest due to their exceptional capacity to thrive in the extreme Antarctic ecosystem (Margesin et al., 2003; Turchetti et al., 2011). According to past studies, the facultatively psychrophilic *P. cryoconitis* culture can grow at temperatures ranging from 1 to 25 °C (Wong et al., 2011). Meanwhile, the psychrophilic yeast *G. antarctica* has been observed to be temperature tolerant between -12°C and 20°C (Boo et al., 2013; Soon et al., 2018).

As a means of surviving and adapting to the Antarctic climate, these extremophiles produce a wide range of biologically important proteins, particularly those involved in the thermal stress response (Song et al., 2017; Yusof et al., 2016; Wong et al., 2011). Many interesting discoveries about *P. cryoconitis* have been made regarding genes encoding resistance to cold stress and heavy metals, as well as industrially valuable enzymes (Lee et al., 2016). Meanwhile, genomic analysis of *G. antarctica* revealed numerous protein-coding genes associated with cold tolerance, such as antifreeze proteins and fatty acid desaturases (Firdaus-Raih et al., 2018). However, despite the availability of whole genome sequences for both microorganisms, no detailed description of their stress response mechanisms has been documented (Wong et al., 2013, 2019). This problem is complicated by the fact that 35% of the protein-coding genes in *P. cryoconitis* (Lee et al., 2016) and

37% of the protein-coding genes in *G. antarctica* (Firdaus-Raih et al., 2018) were classified as hypothetical proteins (HPs). Furthermore, 82 of the 319 transcripts that are currently unique to *G. antarctica* have been identified as having unknown functions (Bharudin et al., 2018; Wong et al., 2019). These functionally unknown proteins may be involved in significant aspects of this microorganism's biological function. Previous research has demonstrated that a set of proteins with unknown functions is vital in the physiological regulation and cold adaptation of psychrophilic microorganisms (Koh et al., 2017; Teoh et al., 2021). Similarly, recent research suggested that the HPs in *P. cryoconitis* and *G. antarctica* were important in the early stages of cold and freeze stress, though these findings have yet to be validated (Soon et al., 2018; Wong et al., 2019). This presents an opportunity for new discoveries to be made in order to gain a better understanding of their distinctive properties for adaptation mechanisms.

Proteins are versatile macromolecules that are required for the cellular adaptation system to function properly. Although psychrophiles share basic cold-adaptation strategies, different species have been shown to adopt different approaches for tolerating and surviving thermal stressors (Boo et al., 2013; Collins & Margesin, 2019; Firdaus-Raih et al., 2018). As every protein is unique, characterization at the physiological and biochemical level is critical for unravelling their molecular mechanisms and realising their full biotechnological potential. Several discoveries have been made primarily as a result of structural biology research, including drug design, vaccine development, and the discovery of industrially important enzymes (Eisenstein et al., 2000; Maveyraud & Mourey, 2020; Muhammed & Aki-Yalcin, 2019). More specifically, structural and functional analysis of HPs from a variety of pathogenic species has resulted in a greater understanding of disease mechanisms, diagnostics, symptom treatment, and vaccine development (Islam et al., 2015; Sen & Verma, 2020). By elucidating the structure and function of the HPs, researchers will gain knowledge about new protein pathways and cascades, allowing us to better understand the protein mosaic and determine protein-protein interactions.

The three-dimensional (3D) protein structures that mediate biochemical interactions must be evaluated to understand biological processes at the system

level (Hauri et al., 2019). For structural determination experiments, a high level of soluble protein amenable to downstream processing is required. Heterologous expression Implementing a suitable expression host, such as *Escherichia coli*, is one of the most effective strategies to optimise the production of high-quality pure proteins (Ahmad et al., 2018). Subsequent protein purification is then accomplished by a fast protein liquid chromatography (FPLC) system through the two-step purification of recombinant proteins via nickel affinity chromatography followed by size exclusion chromatography (SEC) (Kim et al., 2015). Due to the capability of X-ray crystallography to provide extremely detailed structural information, it has become indispensable in protein structure determination (Maveyraud & Mourey, 2020). However, crystallisation is a difficult process with a low success rate, and not all proteins crystallise (Holcomb et al., 2017). In the absence of an experimentally determined structure, comparative or homology modelling can generate a useful 3D model of a protein that is linked to at least one known protein structure. To achieve this ambitious target, expensive and time-consuming structural determination experiments are complemented with theoretical approaches. Bioinformatics aids in the prediction of structures from genomic data and comparative structural modelling. Recent advances in omics technologies have provided an excellent insight into the molecular basis behind cold adaptation processes, enabling the theoretical assessment of several physicochemical parameters to identify the function and composition of previously uncharacterized proteins (Li et al., 2011; Naqvi et al., 2015). Thus, by combining structural knowledge of proteins with functional annotation tools, previously uncharacterized proteins can be elucidated (Jez, 2017).

The present study was designed to evaluate the conserved HPs associated with thermal stress responses in *P. cryoconitis* and *G. antarctica* in order to establish a better understanding of their adaptation mechanisms. While gene sequences provide important information, they are devoid of information on uncharacterized proteins, making it challenging to determine functionally significant sequences. Moreover, there is a limited number of studies addressing the properties of thermal stress proteins produced by psychrophilic microorganisms (Song et al., 2017; Wong et al., 2019). To date, there are no reports that focus on the functional and structural analysis of conserved HPs in *P. cryoconitis* and *G.*

roadblocks in the study of HPs, such as the limitation of publicly available public databases and the costly and labour-intensive experimental procedures for structure determination. The findings of this work will provide a new perspective and scientific approach to understand how proteins behave in terms of flexibility, stability, and dynamic conformations under thermal stress.

1.3 Research Questions

As mentioned, the following research questions arise:

1. Do the hypothetical proteins found in the genomes of *P. cryoconitis* and *G. antarctica* play an important role in thermal adaptation?
2. How are the hypothetical proteins different from other known proteins?
3. Do the hypothetical proteins acquire certain characteristics in the structure that allow them to function at extreme temperatures?

1.4 Research Objectives

The purpose of this research is to elucidate the function and structure of HPs involved in the thermal stress response in *P. cryoconitis* and *G. antarctica*, as well as to understand the relationship between protein molecular architecture and function under cold-adapted conditions. To accomplish these goals, molecular biology techniques will be integrated with omics technology. The scope of the study comprises the *in silico* analysis of existing genomic data and the collection of *in vitro* experimental data for protein characterization. In an attempt to achieve the research target, the main objectives are:

1 To screen for conserved hypothetical proteins related to thermal stress.

2 To conduct *in vitro* analysis of the selected hypothetical proteins and resolve their functions.

3 To determine the relationship between the protein's molecular structure and its function.

1.5 Expected Outcome

The conserved HPs that respond to thermal stress may have specific structural attributes that contribute to their flexibility and stability. This enhanced their capacity to maintain their activities under thermal stress, thus protecting host organisms against cold denaturation and heat aggregation. The results of this research will add to our understanding of the cold-adapted protein's structural and functional properties, thus revealing its real potential for biotechnological applications.