PURIFICATION OF RECOMBINANT COLLAGEN-LIKE PROTEIN FROM *Rhodopseudomonas palustris* EXPRESSED IN *Escherichia coli* USING AQUEOUS TWO-PHASE SYSTEM

BY

NURSYAHIDATUL AZWA BINTI AWANG

A thesis submitted in fulfillment of the requirement for the Master of Science (Halal Industry Science)

International Institute for Halal Research and Training
International Islamic University Malaysia

AUGUST 2022

ABSTRACT

Recombinant collagen-like protein (recCLP) is a collagen-like molecule extracted from microorganisms and expressed in *Escherichia coli* host. The surge in demand for high-quality collagen is due to the multitude of applications in the end-user industries such as pharmaceuticals, food, nutraceuticals and cosmetics. The awareness of the halal status of collagen among Muslim consumers encouraged the discovery of protein from non-mammalian sources, especially from microorganisms. The collagens from microorganisms have great industrial potential because they are free from any zoonotic diseases, contamination and side effect issues. Moreover, the difficulties of the present downstream processing, especially at the purification step, encourage the application of an aqueous two-phase system (ATPS). This study aimed to identify the optimum ATPS conditions for purification of recombinant collagen-like protein expressed in *E. coli* that initially extracted from *Rhodopseudomonas palustris*. Recombinant collagen-like protein from *R. palustris* was purified using the aqueous two-phase system consisting of polyethylene glycol (PEG)/ potassium phosphate. First, the five binodal curves representing five different molecular weights of PEG (1500, 2000, 4000, 6000 and 8000 g/mol) were constructed using the node determination method. Binodal curve that divides the region of two aqueous phases from one phase is important so that a systematic choice of system can be used for portioning experiments. Then, several factors involved in the partitioning behaviour of recCLP such as volume ratio, system pH, the concentration of polymer and salt were studied. The selected ATPS conditions (PEG and salt concentration) were optimised using the response surface methodology (RSM) method. Purification by affinity chromatography was carried out and further compared with ATPS in terms of efficiency and economic aspects to evaluate its potential application as a purification method for recCLP. The binodal curves obtained proved that a high molecular weight of PEG required a low concentration of potassium phosphate to form a two-phase system. As PEG molecular weight increased, the curved was distorted towards the origin. The highest partition coefficient (KE) was found in the system with 26 % (w/w) PEG 2000 and 26 % (w/w) potassium phosphate, making it the best ATPS combination for the OFAT analysis and optimisation process. The range of volume ratio, pH and concentration of PEG and potassium phosphate on the partitioning of recombinant collagen-like protein by ATPS were successfully obtained from OFAT method prior to the optimisation study. Optimisation of ATPS conditions using face-centered central composite design (FCCCD) in Response Surface Methodology (RSM) with 11 runs showed the optimum conditions of ATPS with 24.80 % (w/w) PEG 2000 and 29.20 % (w/w) potassium phosphate with recCLP concentration of 3.23 ± 0.12 mg/mL. Analysis of variance showed the coefficient of determination (R2) were 0.8823, 0.8823, and 0.8193 for fluorescence intensity, the concentration of collagen-like protein and purification factor, respectively. Lastly, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) confirmed the molecular weight of the recCLP, which is 36 kDa. In addition, results showed that ATPS is a low cost, time-saving with a high recovery method that may raise the consideration for substitution of chromatography method. In conclusion, the purification method through ATPS to purify recCLP has high potential, cost-effective, can replace the tedious and expensive downstream processing. Furthermore, this study can serve as a reference and deliver new information for future work.

ملخص البحث

البروتين الشبيه بالكولاجين (recCLP) هو جزيء شبيه بالكولاجين يتم استخراجه من الكائنات الدقيقة ويعبّر عنه في مضيف *Escherichia coli*. وتعزى الزيادة في الطلب على الكولاجين عالي الجودة إلى تعدد التطبيقات في صناعات المستخدمين النهائيين مثل المستحضرات الصيدلانية والأغذية والمكملات الغذائية ومستحضرات التجميل. وشجّع الوعي بحالة الحلال للكولاجين لدى المستهلكين المسلمين على اكتشاف البروتين من مصادر غير ثديية ، ولا سيما من الكائنات الدقيقة. وللكولاجين من الكائنات الدقيقة إمكانات صناعية كبيرة لأنها خالية من أي أمراض حيوانية أو تلوث أو آثار جانبية. وبالإضافة إلى ذلك، فإن الصعوبات التي تواجه عملية المعالجة الحالية، وخاصة في خطوة التنقية، تشجّع على تطبيق نظام مائي من مرحلتين (ATPS). وتهدف هذه الدراسة إلى تحديد الظروف المثلى لنظام ATPS لتنقية البروتين الشبيه بالكولاجين المعاد تركيبه المعبّر عنه في *E. coli* والمستخرج بدايةً مِنْ *Rhodopseudomonas palustris*. تم تنقية البروتين الشبيه بالكولاجين المستخرج من *R.palustris* باستخدام النظام المائي ثنائي الطور المكوّن من البولي إيثيلين جليكول (PEG)/فوسفات البوتاسيوم. أولا ، تم بناء المنحنيات الثنائية الخمسة التي تمثل خمسة أوزان جزيئية مختلفة من PEG (1500 ، 2000 ، 4000 ، 6000 ، 8000 جم/مول) باستخدام طريقة تحديد نقطة تقاطع المدارين. المنحنى الثنائي الذي يقسم المنطقة من مرحلتين مائيتين عن مرحلة واحدة هو أمر مهم بحيث يمكن استخدام الاختيار المنهجي للنظام لتقسيم التجارب. وبعد ذلك، تمت دراسة عدة عوامل تؤثر على سلوك تقسيم recCLP مثل نسبة الحجم والأس الهيدروجيني للنظام وتركيز المبلمر والملح. تم تحسين الظروف المختارة لـ ATPS (PEG وتركيز الملح) باستخدام منهجية سطح الاستجابة (RSM). وأُجريت عمليات تنقية باستخدام كروماتوغرافيا التقارب ، كما أُجريت مقارنات أخرى مع ATPS من حيث الكفاءة والجوانب الاقتصادية لتقييم إمكانية تطبيقها كأسلوب لتنقية البروتين الشبيه بالكولاجين (recCLP). وأثبتت المنحنيات الثنائية التي تم الحصول عليها أن الوزن الجزيئي العالي من PEG يتطلب تركيزًا منخفضًا من فوسفات البوتاسيوم لتشكيل نظام ثنائي الطور. أدت زيادة الوزن الجزيئي PEG إلى انحراف المنحنى نحو نقطة الأصل. وقد وُجد أعلى معامل للتقسيم في النظام مع 26%(كتلة/كتلة) من PEG 2000 وَ 26% (كتلة/كتلة) من فوسفات البوتاسيوم، مما يجعله أفضل تركيبة ATPS لكلٍّ من تحليل OFAT والأمثليّة . ودُرس تأثير نسبة الحجم والأس الهيدروجيني وتركيز فوسفات البوتاسيوم على تقسيم البروتين الشبيه بالكولاجين المعاد تركيبه بواسطة الـ ATPS. وقد أظهر الاستخدام الأمثل لظروف الـ ATPS باستخدام التصميم المركزي المركب المركز على الوجه (FCCCD) في منهجية منهجية سطح الاستجابة (RSM) مع 11 دورة أظهرت الظروف المثلى لـ ATPS مع 24.80%(كتلة/كتلة) من PIG 2000 و 29.20% (كتلة/كتلة) من فوسفات البوتاسيوم مع تركيز recCLP بقيمة 3.23±0.12 ملجم/مل. وأظهر تحليل التباين أن قيم معامل التحديد (R^2) كانت 0.8823 و 0.8823 و 0.8193 بالنسبة لكثافة الفلورسنت وتركيز البروتين الشبيه بالكولاجين ومعامل التنقية، على التوالي. وأخيرا ، أكّد الفصل الهلامي الكهربائي (SDS-PAGE) الوزن الجزيئي لـ recCLP ، والتي تبلغ 36 كيلودالتون. وبالإضافة إلى ذلك ، أظهرت النتائج أن الـ ATPS هي طريقة منخفضة التكلفة ومقتصدة للوقت مع استرداد عالٍ قد تثير النظر في الاستعاضة عن طريقة الكروماتوغرافيا. وختامًا، فإن طريقة استخدام طريقة ATPS لتنقية recCLP تنطوي على إمكانات عالية وفعالة من حيث التكلفة، ويمكن أن تحل محل التجهيز الشاق والمكلف للمعالجة النهائية.وعلاوة على ذلك، يمكن لهذه الدراسة أن تكون مرجعًا وأن تقدم معلومات جديدة للعمل في المستقبل.

.

APPROVAL PAGE

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

…….
Prof. Ts. Dr. Azura Binti Amid
Supervisor

DR. ZATUL IFFAH BINTI MOHD ARSHAD
Pensyarah Kanan
Fakulti Teknologi Kejuruteraan Kimia dan Proses
Universiti Malaysia Pahang

…….
Dr. Zatul Iffah Binti Arshad
Co-Supervisor

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

…….
Prof. Dr. Hamzah Mohd Salleh
Examiner

…….
Prof. Dr. Norasikin Bt Othman

External Examiner

This thesis was submitted to the International Institute for Halal Research and Training (INHART) and is accepted as a fulfilment of the requirement for the degree of Master of Science (Halal Industry Science).

…….
Prof. Dr. Yumi Zuhanis Has-Yun Bt. Hashim, Deputy Dean, International Institute for Halal Research and Training (INHART)

…….
Prof. Dr. Hamzah Mohd Salleh
Dean, International Institute for Halal Research and Training (INHART)

DECLARATION

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

Nursyahidatul Azwa Binti Awang

Signature Date

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

PURIFICATION OF RECOMBINANT COLLAGEN-LIKE PROTEIN FROM *Rhodopseudomonas palustris* EXPRESSED IN *Escherichia coli* USING AQUEOUS TWO-PHASE SYSTEM

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

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

Affirmed by Nursyahidatul Azwa Binti Awang

Signature Date

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Compassionate, the Most Merciful. Alhamdulillah, all praises be to Him for the chance to complete this study.

First of all, I would like to convey my warm-hearted appreciation and gratitude to my supervisor, Prof. Ts. Dr. Azura Amid and co-supervisor Dr.Zatul Iffah Arshad for the invaluable guidance, comments, ideas, patience and endless support in finishing this project.

My acknowledgment also goes to INHART and Biotechnology Engineering Department team for providing the facilities to complete my laboratory works from 2019 till 2021. Grateful acknowledgment extended to Ms. Nazira, Science Officer of Institute of Advance Technology, UPM for permitting us to use Multimode reader. Sincere thanks go to technical staffs and lecturer, Sr. Alifah, Bro. Nizam, Dr. Shirwan and Bro. Anuar, for their help in providing the equipment needed.

Furthermore, heartfelt appreciation to my faithful and hardworking colleagues Faqihah Salleh, Abeir Hussin, Fatemeh Soroodi, Nurjehan Ezzatul and Nurfatin Azma who have supported and helped me a lot whenever I found difficulties in completing this project. Last but not least, I am utterly grateful to my amazing parents and family members for the endless love, prayers, and support throughout my period of study. Enormous help from those who indirectly contributed to this project, your kindness means a lot to me. Thank you very much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

g	gram
g	gravitational force
h	hour
l	litre
kda	kilo Dalton
mg	milligram
mL	millilitre
min	minute
rpm	revolutions per minute
%	percentage
w/w	weight over weight
w/v	weight over volume
v/v	volume over volume
°C	degree celcius
µl	micro litre

LIST OF ABBREVIATIONS

ADH	alcohol dehydrogenase	His	histidine
ANOVA	analysis of variance	IgG	immunoglobulin
ATPS	aqueous two-phase system	NCBI	National Center for Biotechnology Information
BSA	bovine serum albumin	IgG	immunoglobulin
BSE	bovine spongiform encephalopathy	OFAT	one factor at a time
CLP	collagen-like protein	PCR	Polymerase chain reaction
recCLP	Recombinant collagen-like protein	PEG	Polyethylene glycol
DEAD	diethyl aminoethyl cellulose	PF	Purification factor
FCCCD	face-centered central composite design	RSM	response surface methodology
FL	fluorescence intensity	Scl	Streptococcal collagen-like protein
FMD	foot-and-mouth disease	SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel
FPLC	fast protein liquid chromatography	TSE	transmissible spongiform encephalopathies
GRAS	Generally regarded as safe	Vr	volume ratio

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF STUDY

Collagen is the most abundant protein in mammals and the primary structural protein in all animals, comprising about 25-30 % of total animal proteins (Kiew & Don, 2013). It is the main structural material of the extracellular matrix of all connective tissues such as skin, bones, ligaments, tendons, and cartilage as well as interstitial tissues of all parenchymal organs (Felician et al., 2018). It is defined by a unique structure known as triple-helical conformation that consists of three polypeptide chains (α-chains) super-coiled together around a common axis to give a rope-like structure (Figure 1.1(a)) (Gould, 2016; Knupp & Squire, 2003; Yd et al., 2013). Figure 1.1(b) illustrates the repeating structure of amino acid of collagen Gly-Xaa-Yaa with the unusual abundance of three amino acids: glycine (Gly), proline, and hydroxyproline (Knupp & Squire, 2003; Lukomski et al., 2018). Due to its unique structure, collagen plays a big role in fibril formation, mechanical properties and responsible for the interaction with a wide range of molecules (Brodsky & Ramshaw, 1997).

Collagen can be derived from various sources such as land animals, marine animals, birds, and microorganisms. Mammalian collagen (mainly bovine and porcine) is the major industrial source of collagen owing to the multitude of applications in industries. Recently, instead of focusing on animal-derived collagen, most studies focused on non-mammalian sources such as bacteria and marine sources. The rise of potential disease transmission and religious issues triggered researchers to come up with an alternative to animal collagen. Therefore, for the past 10 years, more than 100 putative collagen-like proteins (CLPs) have been discovered from various bacterial genomes and eight of them have been recombinantly expressed in *Escherichia coli* (Xu et al., 2010). Collagen-like protein (CLP) in microorganism or bacterial collagen has been identified based on the repeating signature (Gly-Xaa-Yaa)n sequence

characteristic of triple-helix. Furthermore, their uniqueness is that bacterial collagen has a different amino acid structure from animal collagen but is able to function well as animal-derived collagen (Xu et al., 2014). Bacterial collagen has a big potential to be further developed and commercialised due to the high demand for collagen in various industries such as the pharmaceutical, biomedical, food, and cosmetic industry. This study is focusing on collagen from microorganisms as they are purer and disease-free compared to animal collagen.

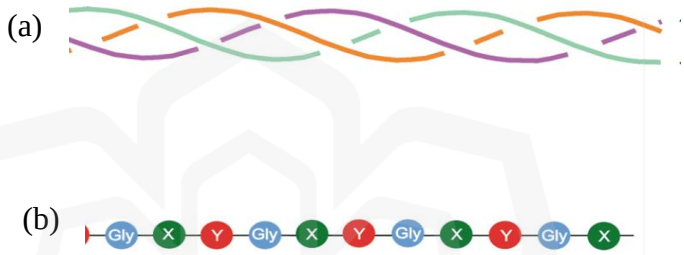

Figure 1.1 (a) The Collagen Triple-Helix Structure (b) The Repeating Signature Triplets Gly-Xaa-Yaa in Collagen Amino Acid Sequences (Knupp & Squire, 2003)

In the recent era, recombinant DNA technology comes under the spotlight of researchers and scientists due to tremendous advancement and various applications in the medical, pharmaceuticals, and agriculture industry (Khan et al., 2016). Recombinant technology for the production of collagen from either plant or bacterial sources is currently under investigation and has been used in some animal studies (Peng et al., 2012; Shilo et al., 2013; Wang et al., 2013). This technology would allow the production of non-animal collagen in an animal-free system with the improvement of final products in terms of purity, yield, and safety (Gould, 2016). To keep up, there is a need to develop an efficient and cost-effective downstream process (Ratanapongleka, 2013; Rosa et al., 2010). Recently, a technique known as aqueous two-phase systems (ATPS) has become the alternative method for the purification of biomolecules which can reduce the number of stages as well as the overall cost (Raja et al., 2011).

The main purpose of this study is to purify the recombinant collagen-like protein from bacteria using an efficient and cost-effective method known as an aqueous-two-phase system (ATPS). Several potential factors can affect the performance of ATPS and they were investigated using the one factor at a time (OFAT) and optimised using Response Surface Methodology (RSM) method.

1.2 PROBLEM STATEMENT

The demand for high-quality collagen is due to the multitude of applications in industries such as pharmaceuticals, food, and cosmetic industry. Collagen is normally extracted from human sources or animal sources mainly from pigs and cows (Vázquez et al., 2016). However, Muslims and Jews are prohibited from consuming collagen from pigs whereas bovine sources are prohibited for Sikhs and Hindus (Eriksson et al., 2013). At present, the awareness of halal authentication among the consumers has been spread to other industries such as pharmaceuticals and cosmetics. The halal status of collagen is depending on the origin of the raw sources and the process conditions. Nonspecific collagen is highly suspected to contaminate with porcine elements and haram for used by the Muslim consumers.

Besides, another challenge in producing collagen from animal sources is the emergence of potential disease transmission, contamination, and side effect issues to the consumers. Production of animal-recombinant collagen is complex and requires additional processes results in the high cost of production. Therefore, recombinant technology has elevated the production of recombinant collagen from bacteria.

In recent years, recovery and purification of recombinant protein are the major challenges due to its complexity and high cost. The purification process is the most crucial step and makes up more than 70% of downstream processing costs (Diamond & Hsu, 1992; Goja et al., 2013). The conventional methods consist of several unit operations that cause the high cost of operation and maintenance (Raja et al., 2011). For

instance, precipitation needs to be combined with another process such as chromatography processes, which are complex, tedious, time-consuming, and often produced low yields (Cao & Xu, 2008). Furthermore, the high unit operation involved caused loss of target molecules that will result in a low yield of the product.

Therefore, in this study, the aqueous two-phase system is proposed as an alternative for the purification method. This method is very simple, mild, and free from protein denaturation due to the high water content and stabilising effect supplied by the polymers (Asenjo & Andrews, 2012). Thus, this method can maintain the native structure of proteins. Polyethylene glycol (PEG) is the commonly used polymer in ATPS as it is available at a low cost and able to form a two-phase system when reacts with other neutral polymers as well as salts (Raja et al., 2011).

1.3 RESEARCH QUESTIONS

The research questions in this study are as following:

i. What is the most suitable molecular weight and concentration of polymer that able to separate CLPs?
ii. What is the most suitable concentration of salt that able to separate CLPs?
iii. What are the optimum conditions for purification of recombinant collagen-like protein using aqueous two-phase system?

1.4 RESEARCH OBJECTIVES

The objectives of this project are as the following:

i. To identify suitable concentration and molecular weight of polymer for purification of recombinant collagen-like protein (CLP) by binodal curve determination.
ii. To identify suitable salt concentration for purification of collagen-like protein (CLP) using binodal curve determination.
iii. To optimize the purification method for collagen-like protein (CLP) by response surface methodology (RSM).

1.5 RESEARCH SCOPE

The present work focused on the application of aqueous two-phase system as the purification method for recombinant collagen-like protein. Firstly, five different binodal curves that represents the five molecular weight of polyethylene glycol (PEG) were developed using node determination method. After that, the most suitable molecular weight of PEG and concentration of PEG and salt required to purify recCLP were identified.

One Factor at A Time (OFAT) technique was adopted in this study as a tool for screening of optimum range of the selected parameters; molecular weight of polyethylene glycol (1500, 2000, 4000, 6000 and 8000), volume ratio (0.33-3.5), concentration of PEG (20- 32) % (w/w), concentration of salt (20-32) % (w/w) and pH (6.0-8.0). The responses involved were fluorescence intensity, concentration of collagen-like protein and purification factor.

Afterwards, the optimisation of ATPS conditions was done using Face Centered Central Composite Design (FCCCD) in Response Surface Methodology (RSM) where

the factors are concentration of PEG and salt. The responses involved were fluorescence intensity, concentration of recombinant collagen-like protein and purification factor. In addition, purification of recCLP using affinity chromatography was conducted and the result obtained were compared with ATPS purification method.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter starts with the general introduction of collagens including its structure, biosynthesis, function, stability, types of collagens that have been discovered followed by their applications in various industries. It also includes the details of the purification method involved in the purification of collagen such as the theory, mechanisms, advantages and disadvantages.

2.2 COLLAGEN

2.2.1 Structure of Collagen

Collagen is the main structural protein that is ubiquitously found in the extracellular matrix of animals, including all vertebrates and invertebrates that comprised of amino acids such as glycine (Gly), proline (Pro), and hydroxyproline (Figure 2.1) (Dutson, 1976; Yamazaki et al., 2010). The most common collagen, type 1, contains amino acid sequence of Gly-Xaa-Yaa triplets that are commonly occupied with proline (Pro) and hydroxyproline (Hyp) in Xaa and Yaa position, respectively (Gelse et al., 2003a; Krishnan & Perumal, 2013; Ricard-blum, 2011). The prevalence of Gly-Pro-Hyp sequence in collagen triplets is about 10.5% (Domene & Wajid, 2016; Gorres & Raines, 2010). These triplets are bound together to form polypeptide chains that twisted together in the form of a triple helix structure (Brodsky & Ramshaw, 1997).