RECOMBINANT COLLAGEN-LIKE PROTEIN BATCH PRODUCTION IN STIRRED TANK BIOREACTORS

BY

ABEIR HUSSEIN MOHAMED GAMEIL

A thesis submitted in fulfillment of the requirement for the degree of Doctor of Philosophy in Engineering.

Kulliyyah of Engineering
International Islamic University Malaysia

SEPTEMBER 2025

ABSTRACT

The collagen family of proteins represents an abundant biopolymer that serves indispensable functions in maintaining the human body. Concerns over the halal/kosher status, sustainability, quality, and immunogenicity of animal-derived collagens have driven efforts to explore alternatives. Among these, recombinant bacterial collagen-like proteins emerged as prospective sustainable sources for various industries. This study investigates the scalability of a recombinant *Rhodopseudomonas palustris* collagen-like protein (RPCLP) production in *E. coli*. Before scale-up, the fermentation medium and process must be optimized to boost the growth and maximize RPCLP concentration. Hence, RPCLP was expressed in *E. coli* BL21(DE3) using the pColdII vector, and induction was accomplished using IPTG and temperature reduction. The optimal M9-casamino acid medium components were determined using a one-factor-at-a-time (OFAT) optimization. The screening of factors (using Taguchi OA) and, subsequently, optimization of selected factor levels (using FCCCD) to maximize RPCLP concentration were conducted in a 2 L bioreactor. Kinetic modeling was investigated by using the Monod and logistic growth models, and Luedeking-Piret models for the glucose consumption and RPCLP production. Nonlinear regression was used to determine kinetic parameters, and ANOVA analyses were conducted to evaluate the models. The fermentation was then upscaled into a 7.5 L stirred tank reactor (STR) via two scale-up criteria, namely constant oxygen mass transfer coefficient (constant k_La) and constant impeller tip speed. The soluble protein, obtained via ultrasonication, was characterized through trypsin digestion and SDS-PAGE, and purified using metal affinity chromatography. MALDI-TOF was also conducted to determine its molecular weight. The designs of the screening and optimization experiments were made by using Design-Expert software. The factors that significantly govern the growth of recombinant *E. coli* expressing *R. palustris* collagen-like proteins in the 2 L bioreactor were identified. The optimum conditions for the growth of the recombinant host cells and the concentration of recombinant collagen-like protein in the 2 L bioreactor were established. Optimum conditions yielded 2 g/L RPCLP, which is moderate compared to typical yields for recombinant collagen-like proteins. In the kinetics study, the Monod and logistic growth models fitted well, with significant R^2 values (0.96 and 0.99, respectively). In contrast, the Luedeking-Piret models generated were not in accord with experimental data for glucose consumption and RPCLP production (R^2 values ≤ 0.7). The kinetic parameters were determined, and the findings suggest that the production is partially growth-associated. The constant impeller tip speed criterion was effective as the maximum tip speed of 0.73 m/s prevents shear damage of the cells and the protein product. Contrarily, the constant k_La criterion was inadequate for scale-up, possibly due to heterogeneous environments caused by improper mixing and mass transfer during k_La measurement, leading to inaccurate estimates. Concentrations of approximately 2.1 to 2.5 g/L RPCLP were obtained in both scales using the constant impeller tip criterion. These concentrations were comparable to a previous study that obtained approximately 1 g/L recombinant Scl2 protein from a 2 L batch culture. A protein band that corresponds to the estimated size of the recombinant RPCLP from the literature was detected via SDS-PAGE. This study is the first to demonstrate successful medium optimization, kinetic modeling, and scale-up of recombinant RPCLP production to a 7.5 L bioreactor. The findings establish a protocol for an enhanced bioprocess to cater to the expanding market demands for sustainable, halal, or vegan collagens.

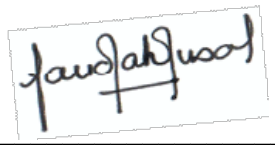
مستخلص البحث

تمثل عائلة البروتينات الكولاجينية بوليمرًا حيويًا وفيرًا يؤدي وظائف لا غنى عنها في الحفاظ على جسم الإنسان. أدت المخاوف بشأن حالة الحلال/الكوشير، والاستدامة، والجودة، والمناعة المحتملة للكولاجين المشتق من الحيوانات إلى دفع الجهود لاستكشاف البدائل، والتي ظهرت منها البروتينات البكتيرية الشبيهة بالكولاجين كمصدر مستدام محتمل لمختلف الصناعات. تبحث هذه الدراسة في قابلية التوسع في إنتاج البروتين الشبيه بالكولاجين (RPCLP) المؤتلف *Rhodopseudomonas palustris* في *E. coli*. قبل التوسع، يجب تحسين الوسط التخميري والعملية لزيادة النمو وتعظيم تركيز بروتين RPCLP. لذلك، تم التعبير عن البروتين الشبيه بالكولاجين من *R. palustris* في الإشريكية القولونية المؤتلفة (BL21) عبر طريقة التعبير البارد باستخدام ناقل pColdII، وتم استخدام IPTG كمحفز مع خفض درجة الحرارة. تم تحسين المكونات المتوسطة لحمض M9-casamino باستخدام نهج العامل الواحد في كل مرة (OFAT). بعد ذلك، تم إجراء فحص متغيرات العملية (باستخدام Taguchi OA) وتبعا تحسين مستويات العوامل المحددة لتركيز البروتين الأمثل (باستخدام FCCCD) في مفاعل حيوي سعة 2 لتر.وأعقب ذلك دراسة حركية لنمذجة العملية في نموذج مونود والنموذج اللوجستي ، بالإضافة إلى نماذج Luedeking-Piret لاستهلاك الجلوكوز وإنتاج .RPCLP. تم تحديد المعلمات الحركية عن طريق الانحدار غير الخطي وأجريت تحليلات ANOVA لتحديد أهمية كل مناسبة. تمت بعد ذلك ترقية عملية التخمير إلى مفاعل خزان مقلب سعة 7.5 لتر (STR) بواسطة معامل نقل الأكسجين الثابت (k_La ثابت) ومعايير زيادة سرعة طرف المكره الثابتة. تم تمييز البروتين القابل للذوبان، الذي تم الحصول عليه عن طريق الموجات فوق الصوتية، من خلال هضم التربسين و SDS-PAGE ، وتنقيته باستخدام كروماتوغرافيا تقارب المعادن. تم إجراء MALDI-TOF أيضًا لتحديد وزنه الجزيئي. تم تنفيذ جميع التصميمات التجريبية بواسطة برنامج Design-Expert. تم تحديد العوامل التي تحكم بشكل كبير نمو الإشريكية القولونية المؤتلفة التي تعبر عن البروتينات الشبيهة بالكولاجين .*R. palustris* في المفاعل الحيوي سعة 2 لتر. تم إنشاء الظروف المثالية لنمو الإشريكية القولونية المؤتلفة وتركيز البروتين الشبيه بالكولاجين المؤتلف في المفاعل الحيوي سعة 2 لتر. أسفرت الظروف المثلى عن تركيز RPCLP متوسط قدره 2 جم/لتر. في دراسة الحركية، كانت بيانات النمو ملائمة لكل من نموذج Monod والنموذج اللوجستي بشكل جيد، مع معاملات تحديد كبيرة (قيم R^2 هي 0.96 و0.99 على التوالي)، ولكن نماذج Luedeking-Piret لم تمثل استخدام الركيزة وتكوين المنتج بشكل جيد.

(قيم $R^2 \leq 0.7$). تم تحديد المعلمات الحركية، وتشير النتائج إلى أن الإنتاج يرتبط جزئيا بالنمو. كان معيار سرعة طرف المكره الثابت فعالاً حيث أن أقصى سرعة للطرف البالغة 0.73 متر/ثانية تمنع تلف القص للخلايا وناتج البروتين. وعلى النقيض من ذلك، كان معيار $k_L a$ الثابت غير كافٍ للتوسع. وربما يرجع ذلك إلى البيئات غير المتجانسة الناتجة عن الخلط غير السليم ونقل الكتلة أثناء قياس $k_L a$ ، مما أدى إلى تقديرات غير دقيقة. تم الحصول على تركيزات تتراوح بين 2.1 و 2.5 جم/لتر تقريبًا من RPCLP في كلا المقياسين باستخدام معيار طرف المكره الثابت. كانت هذه التركيزات مماثلة لدراسة سابقة حصلت على حوالي 1 جم/لتر من بروتين Scl2 المؤتلف من مزرعة دفعة سعة 2 لتر. تم الكشف عن شريط بروتين يتوافق مع الحجم المقدر لـ RPCLP المؤتلف من الأدبيات بواسطة SDS-PAGE. هذه الدراسة هي الأولى التي توضح بنجاح تحسين الوسط والنمذجة الحركية وتوسيع نطاق إنتاج RPCLP المؤتلف إلى مفاعل حيوي سعة 7.5 لتر. تؤسس النتائج لبروتوكول لعملية حيوية معززة لتلبية متطلبات السوق المتزايدة للكولاجين المستدام أو الحلال أو النباتي.

APPROVAL PAGE

The thesis of Abeir Hussein Mohamed Gameil has been approved by the following:



Faridah Yusof
Supervisor

Azlin Suhaida Azmi
Co-Supervisor

Noor Illi Mohamad Puad
Co-Supervisor

Internal Examiner

External Examiner

Chairman

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.

Abeir Hussein Mohamed Gameil

Signature .... Date 26/8/24

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

RECOMBINANT COLLAGEN-LIKE PROTEIN BATCH PRODUCTION IN STIRRED TANK BIOREACTORS

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

By signing this form, I acknowledge that I have read and understood the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Abeir Hussein Mohamed Gameil

…… 26/8/24

Signature Date

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Gracious, Most Merciful.

Alhamdulillah. All gratitude is to Allah for granting me this privilege to embark on this unforgettable journey experience as a researcher and as a person.

I would like to extend my deepest thanks to my main supervisor, Prof. Dr. Faridah Yusof, and co-supervisors Dr. Azlin Suhaida Azmi and Dr. Noor Illi Mohamad Puad, for their unwavering support and invaluable insight, and for providing me with the freedom to chart my way through this experience. I have been truly fortunate to be under the guidance of three wonderful supervisors who provided guidance and financial support as I went through countless hurdles on a personal level as much as on the research level. A special thanks to Dr. Azlin for spending hours to troubleshoot with me in the laboratory.

My sincere gratitude goes to the Infors team at Jenawi Saintifik (Ms. Ashikin, Mr. Sasi, and Mr. Fel) for their timely assistance whenever issues arose with the bioreactor. I am also indebted to Dr. Muhamad Shirwan Abdullah Sani and Dr. Noor Faizul Hadry Nordin for their help and suggestions. I'd also like to thank Dr. Mohd Azmir Arifin (Universiti Malaysia Pahang) for helping me acquire a crucial book, and Dr. Amal Elgharbawy (INHART) for lending me her glucometer.

I'd like to acknowledge the technical assistance provided by Sr. Alifah and Br. Nizam of INHART laboratory, as well as Br. Anuar Arifin, Br. Iznan, Br. Aslan, Br. Nasaruddin, Sr. Suharti, and Br. Hafizul, of the Kulliyyah of Engineering. Your support in the various laboratories has been instrumental in the completion of this research.

Special thanks to Dr. Roslinda Sajari from the Malaysian Rubber Board for her invaluable assistance with the FPLC equipment and for taking the time to help familiarize us with their facility.

I owe everything to the unwavering support, encouragement, prayers, and patience of my family, without whom this achievement wouldn't be possible. I will always be grateful. I'd also like to extend my heartfelt thanks to my colleagues from the CHES department and the INHART laboratory for their assistance and support, especially my comrades in this project – Dr. Faqihah Salleh, Sr. Fatemeh Soroodi, and Sr. Nursyahidatul Azwa Awang. I gratefully acknowledge Sr. Fatemeh Soroodi whose screening study and preliminary shake flask experiments using recombinant *E. coli* BL21(DE3) to express RPCLP provided the foundation for the medium component OFAT optimization.

To all my friends who've been there for me throughout this journey, I thank you for your encouragement and prayers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
ATPS	Aqueous Two-Phase Systems
CCD	Central composite design
CD	Circular Dichroism
CLP	Collagen-like protein
CFD	Computational fluid dynamics
DCW	Dry cell weight
DOE	Design of experiment
DSC	Differential Scanning Calorimetry
ELISA	Enzyme-Linked Immunosorbent Assay
FCCCD	Face-Centered Central Composite Design
FPLC	Fast Protein Liquid Chromatography
FTIR	Fourier Transform Infra-red (spectroscopy)
GCMS	Gas Chromatography Mass Spectrometry
GMO	Genetically Modified Organisms
LB	Lysogeny broth
LCMS	Liquid Chromatography Mass Spectrometry
LS	Large scale
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization Time-of-Flight (mass spectrometry)
NMR	Nuclear Magnetic Resonance
MRI	Magnetic Resonance Imaging
OFAT	One Factor at a Time
ORF	Open Reading Frame
PBD	Plackett Burman Design
PBS	Phosphate Buffered Saline
RPCLP	*Rhodopseudomonas palustris* collagen-like protein
RSM	Response Surface Methodology
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SS	Small scale
STR	Stirred Tank Reactor

LIST OF SYMBOLS

C^*	Liquid *DO* concentration at saturation
C_L	Measured liquid *DO* concentration
v	Impeller tip speed, m/s
D_i	Impeller diameter, m
D_T	Vessel diameter, m
DO	Dissolved oxygen
Fr	Froude number
H_L	Height of liquid, m
H_T	Height of vessel, m
k_La	Volumetric oxygen transfer coefficient, mmol/L.h
N	Impeller speed, rpm
N_A	Flow aeration number
N_{cd}	minimum impeller speed for all medium to be in contact with sparged gas
N_i	Number of impellers
N_P	Power number
OTR	Oxygen transfer rate, mmol/L.h
OUR	Oxygen uptake rate, mmol/L.h
P_g	Gassed power input
Po	Ungassed power input
P_g/V	Gassed power input per unit volume
P_O/V	Ungassed power input per unit volume
Q	Volumetric air flow rate, L/min
Q/V_L	Volumetric air flow rate per unit vessel liquid volume, L/min (vvm)
Re	Impeller-based Reynolds number
T_{mix}	Mixing time, s
v_s	Superficial velocity, cm/s
V_L	Working volume of the vessel, L or m^3
V_T	Total volume of the vessel, L or m^3
X	Cell density, g/L
Y_{xO2}	Cell yield on oxygen, g/mmol/L
ρ	Density, g/cm^3
υ	Liquid effective viscosity, g/cm.s
μ	Growth rate, /h

CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND

Recently, collagen has become a household word that is synonymous with anti-aging, healing, and beauty as it is increasingly found in virtually all nutraceutical and cosmetic products. This is no surprise, as collagen is a wonder product for skin health. Collagen encompasses a large family of ubiquitous proteins constituting various tissues and organs that serve multiple functions in animal systems, providing tensile strength and structure to bones and fascia, tissue repair, and mediating cellular signaling pathways (Srivastava, 2022). These collagen proteins share a common factor, i.e., a unique triple-helical domain made up of tripeptide sequence repeats of glycine (Gly), followed by two amino acid residues. Proline (Pro) and hydroxyproline (Hyp) residues are often at these positions (Xu & Kirchner, 2021). In addition to these unique triple-helices, collagens sport variable domains that may flank either side of the triple-helix (Qiu et al., 2021). Together, the various collagen types support the animal in countless ways, and deficiencies in certain collagen types, whether due to aging or genetic conditions, would manifest as severe issues for the body (Srivastava, 2022).

Collagen boasts a wide variety of natural properties that make it useful for various applications. For instance, its crosslinking and gelling abilities are highly valued (Avila Rodriguez et al., 2018). The gelatin that is obtained by partial hydrolysis of collagen has vast applications in confectionery, health drinks, and other processed products (Tang et al., 2022). Gelatin is also used in drug delivery (Abdullah et al., 2018; Fertala, 2020). However, collagen is not only indispensable to the pharmaceutical, nutraceutical, and cosmetics industries (Sionkowska et al., 2020). The medical field has long used collagen as a biomaterial due to its hemostatic properties, biodegradability, and cell attachment abilities in surgeries and biomedical aids, from catgut sutures to

bandages, prosthetics, scaffolds, and corneal shields (Shenoy et al., 2022). Furthermore, studying collagen in tissue gives scientists a deeper insight into its role in diseases and as a biomarker for disease surveillance, such as in liver fibrosis and cancer (Meurer et al., 2020; Necula et al., 2022).

The collagen market is diverse and expanding and was projected to be worth USD 6.63 billion in 2025 (Avila Rodriguez et al., 2018), but according to a more recent market report by Grand View Research (2023), the global collagen market was estimated at USD 9.115 billion. This growth is driven by numerous factors such as increased consumer awareness and health consciousness (especially in Europe and North America), as well as ease of availability and cost-effectiveness of bovine, vegan, and halal collagen products (for West Asia and Africa). According to Bhadra et al. (2021), the Asia-Pacific region is leading the global market for marine and vegan collagen products. In 2023, the Asia Pacific region dominated with the highest revenue, and this was ascribed to rapidly developing key economies (Grand View Research, 2023). Of the different types, collagen Type I is prevalent due to its anti-aging effect and usage in cosmetics.

The global collagen market was also affected by the advent of the COVID-19 pandemic in 2020. According to Grand View Research (2023), the COVID-19 pandemic has triggered consumer awareness and increased the demand for functional supplements to enhance the immune system and overall well-being. This upward trend in demand is expected to continue. Concurrently, the outbreak of zoonosis related to products of animal origin has shifted attention to alternative sources of collagen. Various market research reports estimated the global collagen market size in 2020 to have been worth around USD 4 to USD 5 billion. Compared to its value in 2016, which was worth USD 3.71 billion, the data indicates a strong growth trajectory that is driven by the global trends towards sustainable production and innovative, environmentally friendly products that encompass multiple industries under its umbrella. These collagen markets can be categorized and studied according to collagen source, industry or applications, product type, and region, as shown in Figure 1.1. Figure 1.2 shows the

percentage of the global collagen market volume by sector of industry, as of 2022, reported by Grand View Research (2023).

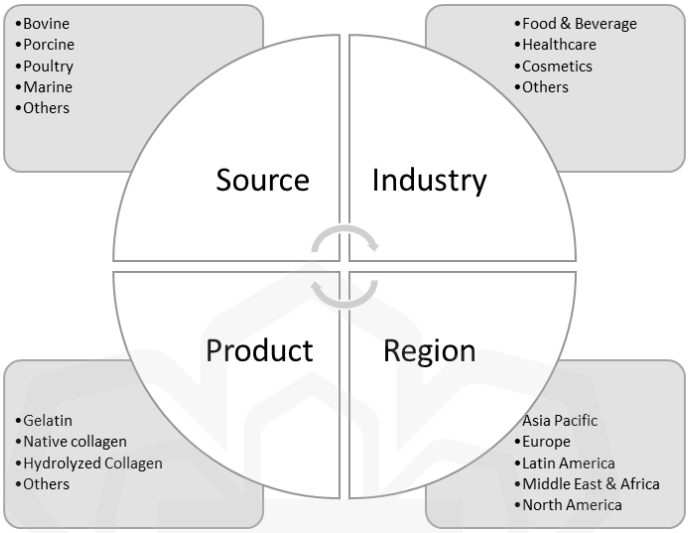


Figure 1.1 Classifications of the collagen market (www.grandviewresearch.com, accessed February 2024)

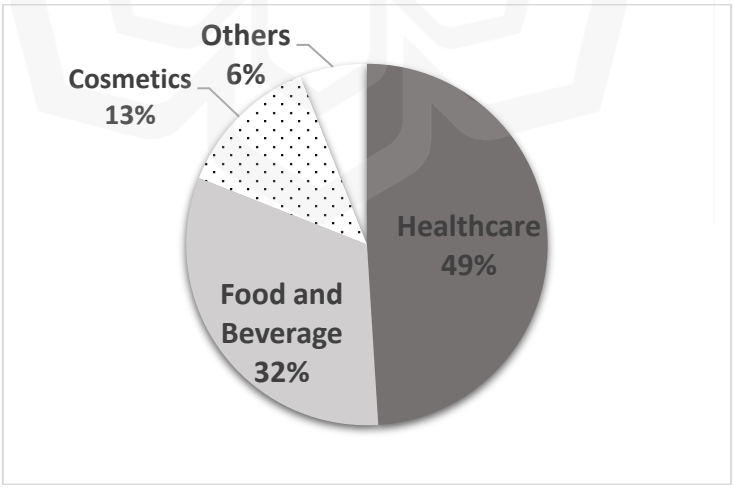


Figure 1.2 Percentage global collagen market share by sector of industry in 2022 (www.grandviewresearch.com, accessed February 2024)

1.2 PROBLEM STATEMENT

Collagen is typically obtained by extraction from the skin, viscera, and bones of various animals, of which bovine, porcine, poultry, and marine collagens are the primary sources. While these animal-sourced collagens dominate the market, concerns about quality, safety, religious, and lifestyle restrictions on consumption have been raised. Growing interests with diverse purposes that aim to tap into different markets have also pushed for alternatives such as synthetic collagen peptides and recombinant human collagens in plants, insects, prokaryotes, and mammalian systems, with varying success due to costly setup, low yields, as well as difficulty of purification and scalability. There is a dire need for a sustainable source of collagen to supply the pharmaceutical, biomedical, food, and cosmetic markets. In addition, characteristics like non-immunogenicity, halal, and kosher compliance are another set of desirable features in collagen. Islamic and Jewish dietary laws strictly forbid the consumption of certain animals, such as pork and unlawfully slaughtered cattle, for instance. Similarly, Hindu dietary laws prohibit the consumption of beef. These dietary laws limit the consumers of collagen products to a few options, emphasizing the need for alternatives (Rakhmanova et al., 2018).

Studies have explored the production of recombinant human collagen due to advantages such as high similarity to native human collagen, batch-to-batch homogeneity, and reduced risks of zoonosis and immunogenicity. A variety of host systems, such as transgenic plants, mammals, insects, insect cell lines, and mammalian cell lines (including human cell lines), have been explored. However, the yields were relatively low compared to conventional animal collagen extraction methods. Recombinant human collagen production was also attempted in bacterial (e.g., *E. coli*) and yeast systems (e.g., *Pichia pastoris*). However, these hosts yielded fragmented and unstable fibers as they lack the ability for post-translational modification of the collagen, unlike mammalian cells (Fertala, 2020; Wang et al., 2017).

Recombinant collagen-like proteins from bacteria such as *Streptococcus pyogenes* have been produced in the laboratory with decent yields. Despite this, *S. pyogenes* was the only species that has been extensively studied out of eight species that have been experimentally investigated and proven to produce collagen-like proteins (CLPs), namely *E. coli, S. pyogenes* (Scl/A and Scl2), *Bacillus anthracis* (BclA), *Legionella pneumophila*, *Clostridium perfringens, Solibacter usitatus, Rhodopseudomonas palustris,* and *Methylobacterium sp 4-46* (Ghosh et al., 2012; Yu et al., 2014). Some bacterial collagen-like sequences are yet to be studied or are only partially characterized in terms of their triple-helical structure. Although numerous species were reported to have collagen-like sequences in their genomes (Rasmussen et al., 2003), the natural expression of these sequences has been demonstrated in only a few cases. This highlights a significant gap between predicted and experimentally verified bacterial CLPs, underlining the need for further studies for potential industrial production and application.

Taken together, these points underscore the need to gain more knowledge and a better understanding of bacterial CLPs. Furthermore, it remains to be seen whether *Streptococcus pyogenes* collagen-like protein 2 (scl2) proteins are the best option for industrial production, in terms of peptide length, stability, and other characteristics. Furthermore, non-pathogenic bacterial strains such as *R. palustris* may have different motifs to explore due to the differences in the function of the CLP in soil bacteria. Thus, it is worth exploring their potential. Together, these challenges and considerations in the production of bacterial collagens have compelled the investigation of *R. palustris* collagen-like protein (RPCLP) production.

The bench-scale production of recombinant *R. palustris* collagen-like protein (RPCLP) using *E. coli* expression host can offer novel insight and protocols that could be applied to pilot and full industrial-scale production of collagen-like proteins. However, several steps must be taken. As each fermentation process involves unique genetic and environmental factors, the factors that affect the growth of the host cells and the production of the desired protein must first be identified by screening and

selection of the most significant factors. Optimizing fermentation conditions at a small scale is crucial to preserve resources and reduce costs. Key factors include expression strain, fermentation medium, and operating conditions. Identifying critical parameters for specific bioreactor fermentation cases is vital. Also, understanding the metabolic responses of the cells is essential, and kinetic studies can provide some basic models of the fermentation before scaling up. Nevertheless, scale-up is complex and cannot be linearly extrapolated from small-scale data. Therefore, several potentially applicable scale-up criteria must be considered and conducted, and may require several attempts before success. This is essential to develop a sustainable and efficient production for recombinant collagen-like proteins, starting from bench-scale optimization to potential industrial-scale application. Successful laboratory-scale production can then be scaled up to industrial levels, ensuring feasible manufacturing.

1.3 HYPOTHESES

It is hypothesized that the production of recombinant *R. palustris* collagen-like protein (RPCLP) in *E. coli* BL21(DE3) can be effectively optimized and scaled up through strategic medium formulation, process parameter optimization, and the application of suitable scale-up criteria, for reasons that are elaborated in Chapter Two. The following specific hypotheses are proposed in line with the research questions and objectives:

1. Optimization of selected components in the M9-casamino acid medium will significantly enhance the yield of recombinant RPCLP by promoting favorable conditions for the growth and protein expression of *E. coli* BL21(DE3).

2. Several physicochemical factors—including inducer concentration, temperature, and pH—are expected to significantly influence the growth and expression efficiency of recombinant *E. coli* in the 2 L bioreactor. Optimization of these factors will lead to improved protein yield and process performance.

3. The kinetics of substrate utilization, cell growth, and RPCLP production in the 2 L bioreactor can be described by a suitable kinetic model (e.g., Monod or logistic growth models), which will allow for better process understanding, control, and prediction of performance.

4. The scale-up of the optimized fermentation process from a 2 L to a 7.5 L bioreactor, when based on maintaining a constant oxygen transfer coefficient (k_La), will better preserve process performance and RPCLP yield than the strategy based on constant impeller tip speed.

5. The recombinant RPCLP produced under optimized and scaled-up conditions will possess favorable biochemical and structural characteristics, rendering it suitable for potential applications in the biomedical, food, and pharmaceutical sectors. These proteins are expected to be halal-compliant and non-immunogenic.

1.4 RESEARCH QUESTIONS

The study addresses the following research questions:

1. What are the optimum M9-casamino acid medium components to maximize recombinant protein yield?
2. What are the factors that significantly govern the growth of recombinant *E. coli* expressing RPCLP in the lab scale?
3. What are the optimum conditions for growth and RPCLP yield from recombinant *E. coli* expressing in the lab scale?
4. What model best fits the fermentation kinetics of substrate, cell growth, and product of the recombinant *E. coli* expressing RPCLP?

5. What scale-up criteria should be maintained to allow the optimized fermentation in the lab scale to be transferred into a medium/ bench-scale that maintains its productivity?

1.5 RESEARCH OBJECTIVES

The overall goal of this research is to upscale the process of recombinant *R. palustris* collagen-like protein production to a 7.5 L bioreactor. This requires optimizing the process parameters in a small-scale (2 L) *E. coli* fermentation to enhance the concentration of recombinant collagen-like protein (RPCLP). The specific objectives of this research are as follows:

1. To optimize selected factors in the medium for maximum recombinant RPCLP yield,
2. To screen for significant variables and optimize the production of collagen-like proteins in a 2 L bioreactor for maximum recombinant RPCLP yield,
3. To investigate the kinetics of production of recombinant RPCLP from *E. coli* BL21(DE3),
4. To up-scale the 2 L bioreactor process into a 7.5 L bioreactor by using constant impeller tip speed and using constant oxygen transfer coefficient (k_La), and
5. To characterize the recombinant RPCLP protein.

1.6 RESEARCH PHILOSOPHY

This research is based on quantitative approaches and is focused on experimental and applied research methodologies to establish a scalable process of RPCLP production via recombinant *E. coli* fermentation. The experiments were designed by reviewing and

incorporating the existing literature on bacterial CLPs and recombinant *E. coli* fermentation, kinetics, and scale-up. To evaluate the findings and make conclusions based on the scientific method, data were gathered from experiments and compared against the current body of literature. The work of Peng et al. (2012) on the scale-up of a recombinant CLP from *S. pyogenes*, using *E. coli*, was regarded as a benchmark for the fermentation and protein concentration. To aid in the design of screening and optimization experiments, a statistical approach was utilized using Design-Expert software version 13. Together, these experiments focus on contributing new knowledge to the practical application of the production of recombinant bacterial collagens.

1.7 RESEARCH METHODOLOGY

In this study, One-Factor-at-a-Time (OFAT) optimization was carried out to optimize the M9 medium components. Then, the screening of process and induction parameters, designed by Design-Expert software version 13, was carried out to determine the significant factors for optimization of RPCLP concentration. The fermentation conditions and their levels were selected from relevant publications on recombinant protein expression, which are detailed in Chapter Two. The study then explores the optimization techniques that would contribute to a higher yield of RPCLP, which is quantified via a glycine-based assay, namely medium component optimization and screening, and optimization of fermentation process parameters, as well as induction parameters. A study of the fermentation kinetics is an essential strategy for understanding this fermentation process and is therefore carried out to model the data, determine kinetic parameters, and elucidate the relationships between biomass growth, glucose utilization, and protein production. After enhancing the yield at the 2 L bench scale bioreactor, it is necessary to consider scale-up parameters such as k_La and impeller tip speed. The choice of these two scale-up criteria stems from findings of the review of literature wherein oxygen transfer coefficient and shear stress significantly influenced the success of the scale-up which is measured by the protein quantitation. The scale-up integrates theories about oxygen transfer and relevant rules of thumb, with

experimentation to validate our strategy. However, it involved and required a lot of iterative refinement to troubleshoot problems and to contribute new insight to the field of bioprocessing and protein production, specifically within the area of recombinant bacterial CLPs.

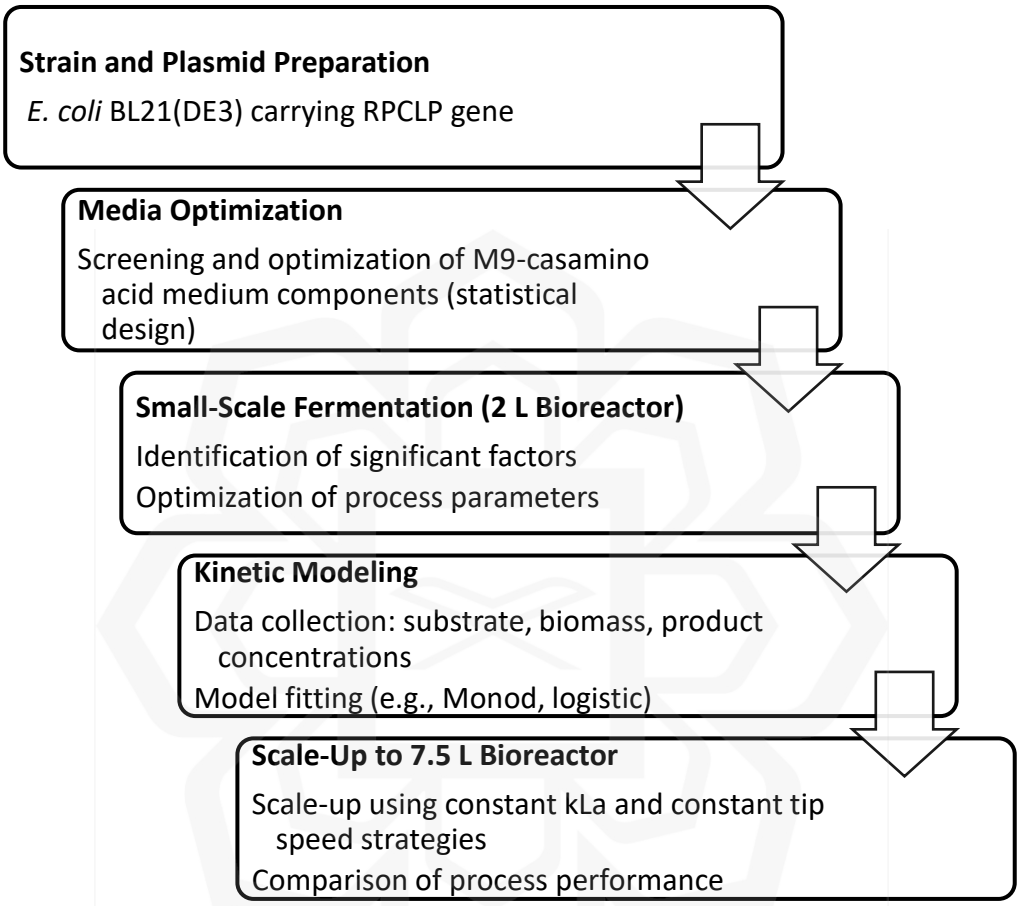


Figure 1.3 Methodology flowchart

1.8 SCOPE AND LIMITATIONS OF THE STUDY

This work focuses on the development, optimization, and scale-up of a bioprocess for the production of recombinant collagen-like protein (RPCLP). Its scope encompasses the following:

1. Medium Optimization: Identification of significant medium components from the literature and their optimization, using OFAT experiments in the shake flask scale, to enhance biomass and protein concentration.
2. Statistical design and execution of experiments for screening and optimization of the fermentation process in a 2 L bioreactor.
3. Fermentation Kinetics: Characterization of growth and production kinetics in shake flasks and bioreactor systems. Modeling of the recombinant *E. coli* fermentation kinetics was limited to the Monod and Logistic growth models, whereas the Luedeking-Piret model was implemented for the kinetics of glucose utilization and RPCLP formation.
4. Bioreactor Scale-Up: Scale-up from 2 L to 7.5 L stirred-tank bioreactors, comparing scale-up strategies based on constant k_La and constant impeller tip speed.
5. Yield Benchmarking: Comparison of RPCLP concentrations with reported bacterial CLPs to establish process competitiveness.
6. Characterization of the recombinant protein is limited to trypsin digestion and SDS-PAGE.

The limitations of the study are as follows:

1. This work is limited to laboratory and bench-scale fermentations; pilot or industrial-scale validation is beyond the current scope.
2. Post-translational modifications (e.g., proline and lysine hydroxylation via enzyme coexpression) are not addressed.
3. Optimization studies of protein separation and purification and functional assays of RPCLP are not included but are highly recommended for future studies.

4. The structural and mechanical properties of RPCLP for specific uses (e.g., scaffolding, hydrogels for biomaterial application) are not explored in this work.
5. Immunogenicity and biocompatibility studies were outside the scope of this investigation.

1.9 SIGNIFICANCE AND NOVELTY OF THE STUDY

There is immense appeal in using recombinant collagen as an alternative to conventional collagen sources for various applications. Enhanced, upscaled bioreactor production of recombinant CLPs from *E. coli* fermentation can demonstrate the potential of this technology and provide for this demand in the market. Recombinant collagen from soil bacteria, such as *R. palustris,* will function as a "blank slate" that can be designed to include motifs and can also be modified chemically to enhance its characteristics for the specific target application whether it is needed as a powder, hydrogel, as a scaffold, as a cosmeceutical or nutraceutical, or for wound dressing. It will also have the advantage of being non-immunogenic, kosher, and halal, thus catering to a larger group of consumers while reducing dependence on animal sources of collagen which are complex, costly, not sustainable, often not halal, and in some cases worrisome because of zoonosis.

Thus far, there is no published analysis of the factors of recombinant CLP production from *R. palustris* in *E. coli* whether screening, optimization, or scale-up in the 2 L and 7.5 L scales, even though the expression of the soluble, intracellular protein has been demonstrated in shake flask culture studies (Peng et al., 2014). Furthermore, no comprehensive work has been dedicated to the kinetics of this particular process.

1.10 THESIS ORGANIZATION

This thesis consists of five chapters. Chapter One sets the background of the study and defines the problem being addressed. It states the main goals and highlights the importance of the study. Chapter Two is dedicated to the review of current, relevant literature pertaining to this research. It includes sections of a systematic review on the process scale-up criteria that have been used in the bioreactor-scale fermentations using *E. coli* to obtain recombinant proteins (Gameil et al., 2021a). Chapter Three details the materials and the techniques applied in the study to achieve its objectives. Chapter Four details the results of this study, describes them, and provides a discussion of all findings. Chapter Five highlights the key findings, and delivers the conclusion of the study as well as relevant recommendations for further investigation.

CHAPTER TWO
LITERATURE REVIEW

2.1 INTRODUCTION

This chapter begins with an introduction to the various collagens, their structures and classification, properties, and sources. After establishing the importance of collagens and the need for novel, sustainable collagen sources a review of previous studies on the production of recombinant collagen is presented. Then, the chapter briefly sheds light on the relevant techniques available for bacterial culture, the design of the experiments for screening of factors and optimization of the expression, the detection, separation, and purification of the CLPs, as well as the scale-up of the process and characterization of the product.

2.2 COLLAGEN

Collagens comprise a vast and indispensable family of proteins that constitute the chief components of the extracellular matrix of metazoa and are among the planet's most ubiquitous biopolymers. The word collagen is coined from two Greek words, wherein "kólla" refers to gum and "gen" means producing, referring to the abundance of collagens in various tissues, holding them together. Collagen is a fibrous, insoluble protein that, in the form of an extracellular matrix, forms the structure of the skin and connective tissues (Silvipriya et al., 2015).

2.2.1 Properties of Collagen, Uses, and Importance

Collagen has a significant importance in the body due to its many properties and uses. Firstly, it demonstrates excellent tensile strength as exhibited by its presence in tendons, bones, cartilage, and fascia among others (Sandhu et al., 2012). It confers elasticity and strength to the skin and contributes to the entrapment, local storage, and delivery of growth factors and cytokines, which helps in wound healing, tissue repair, and organ development (Gelse et al., 2003). Collagen also hinders the absorption of disease-causing microbes and/or their toxins through the skin (Silvipriya et al., 2015). It has a role in the biological functions of the cell (survival, proliferation, and differentiation), in the bones or blood vessels regeneration, and preserves structural integrity (Silvipriya et al., 2015). Due to its many uses, the abundance of various types of collagens in the body is unsurprising. On the other hand, the absence of collagens, as caused by defective collagen genes, causes diseases such as osteogenesis imperfecta, Ehler-Danlos syndrome, epidermolysis bullosa, and Alport syndrome (Gelse et al., 2003). Collagen is also associated with scurvy, SLE (systemic lupus erythematosus), systemic sclerosis, Stickler syndrome, oral submucous fibrosis, and Marfan syndrome (Garrett & Grisham, 2010).

The collagen obtained from various animal sources, mentioned in section 2.2.4, is used in numerous ways. Collagen may be hydrolyzed into gelatin, which is an indispensable polymer in the food and beverage industry, where it is used as a gelling, thickening, water-binding, emulsifying, foaming, and film-forming agent (Bulleid et al., 2000; Hashim et al., 2015). The global collagen market size was worth USD 3.71 billion in 2016 and is expected to be worth USD 6.63 billion by the year 2025 (Avila Rodriguez et al., 2018). Figures 2.1 and 2.2 show the percentage of global collagen market shares for the year 2022 by product type and sources, respectively, as reported by Grand View Research, USA (2023).

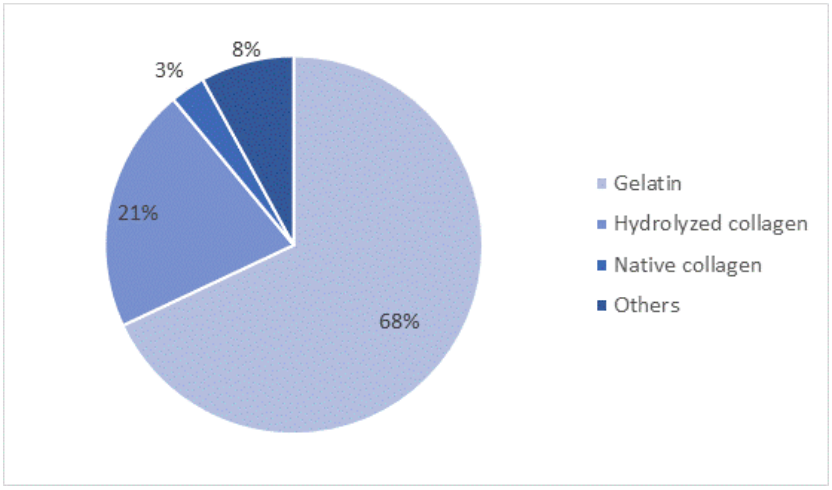


Figure 2.1 Percentage global collagen market share by product type in 2023 (www.grandviewresearch.com, accessed February 2024)

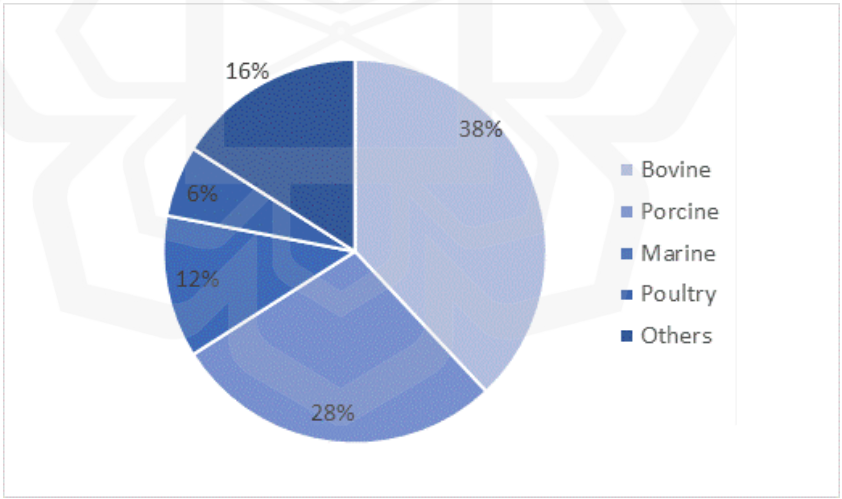


Figure 2.2 Percentage global collagen market share by source in 2023 (www.grandviewresearch.com, accessed February 2024)

In the pharmaceutical industry, collagen finds its use in the form of microparticles and injectable dispersions, and oral supplements for the treatment of photoaging, osteoporosis, and osteoarthritis (Porfírio & Fanaro, 2016; Song et al., 2017). It is also ingested as an anti-aging supplement (Shenoy et al., 2022). Furthermore, it is used in corneal shields, cosmetic dermal fillers, hydrogels, and drug delivery systems due to attributes such as its biodegradability, biocompatibility, low immunogenicity, and cell attachment ability (Wang et al., 2017). Hydrolyzed collagen (gelatin) is also used in the pharmaceutical industry for drug delivery. Bovine gelatin is highly utilized in hard and soft capsules, tablets, syrups, emulsions, and suppositories (Abdullah et al., 2018).

Collagen-based biomaterials are extensively used in tissue engineering, such as in injectable matrices, scaffolds intended for bone regeneration, membranes, and hydrogels which are mainly from fibril-forming collagens. Collagen scaffolds execute numerous roles such as in visualizing cells in models of the nervous system, in the study of cancer cells, and in *ex vivo* organ culture as three-dimensional models for various diseases. In the medical field, collagen scaffolds assist in bone and cartilage regeneration, vascular reconstruction, neuronal regeneration, wound dressings, and surgical suture (Grover et al., 2012; Silvipriya et al., 2015; Wang et al., 2017). Cell-free corneal implants employing recombinant human collagen or collagen-silk (*Bombyx mori*) hybrids can be used to regenerate tissue in patients with corneal blindness (Ahearne et al., 2020). Similarly, collagen-like proteins (CLPs) and recombinant collagens have received considerable attention in the same areas of applications - as hydrogels, scaffolds, stents and grafts, and cosmetic formulations. Several reviews on collagen as a biomaterial and in biomedical applications, in cosmetics and leather production are available (Avila Rodriguez et al., 2018; Chattopadhyay & Raines, 2014; Ferreira et al., 2012; Golser & Scheibel, 2018; Shekhter et al., 2019; Sionkowska et al., 2017). Table 2.1 highlights some of these uses.

Table 2.1 Some applications of recombinant collagens and CLPs in the literature

Collagen type/ form	**Collagen Source**	**Applications**	**Reference**
Recombinant human gelatin	Human collagen	Drug delivery	Cardoso et al. (2017); Fertala (2020)
Recombinant human-like collagen	Human-like collagen	Scaffold for vascular tissue engineering	Zhu et al. (2009)
Recombinant human-like collagen hydrogel	Recombinant human type I and type III collagens	Corneal substitutes	Liu et al. (2008)
Recombinant human-like collagen	Human-like collagen	Hemostatic sponge	Jiang et al. (2017)
Recombinant human collagen	Human collagen, type III	Hemostatic application	Yang et al. (2004)
Collagen-mimetic proteins	Scl2 from *Streptococcus pyogenes* and CLPs from *M. sp 4-46* and *S. usitatus*	Cosmetic formulations	Peng et al. (2015)
Collagen-mimetic proteins	Scl2	Multilayer vascular grafts	Browning et al. (2012)
Variants of scl2 protein containing the sequence GFPGER (Scl2-2) or GFPGEN (Scl2-3)	Scl2 from *Streptococcus pyogenes*	Bioactive hydrogels based on 'designer collagens'	Cosgriff-Hernandez et al. (2010)

2.2.2 Structure and Types of Collagens

Collagens have at least one domain with a three-dimensional structure i.e., a collagen tripeptide triple helix. The conformation of the right-handed collagen triple helix is made up of "three left-handed" alpha chains (Garrett & Grisham, 2010). Only Glycine is found in every third position as there no other amino acid could fit in its place, and

this imposes a repetitive amino acid sequence pattern - (Gly-X-Y) n - and this is the distinctive signature of collagen. Often, X represents proline and Y represents hydroxyproline, and this is a characteristic of animal collagens (Silvipriya et al., 2015). Collagen has three modified amino acids, namely 4-hydroxyproline, 3-hydroxyproline, and 5-hydroxylysine. Hydroxyproline residues confer significant stability to the triple helix through stereoelectronic effects within the helix and/or hydration effects (due to the water molecules binding to the hydroxyproline side chains) and are also deemed essential for collagen self-association and some receptor interactions (Yu et al., 2014). Due to the high content of glycine, proline, and hydroxyproline, collagens do not form α-helices and β-sheets, unlike other proteins. Instead, they form a right-handed triple helix where each strand is arranged in a left-handed helical fashion with 3.3 residues per turn, and a rise of 2.9 Å per residue- a staggered structure (Garrett & Grisham, 2010).

There are at least 28 different known collagens in humans, of at least 46 different polypeptide chains in vertebrates. These may be classified into seven types of collagens, namely fibril-forming collagens, fibril-associated collagens with interrupted triple helices (FACIT), network-forming collagens, anchoring fibrils, transmembrane collagens, basement membrane collagens, and other collagens with unique functions. Fibril-forming collagens of around 1000 amino acids or 300 nm are the most profuse type of collagen and are represented by collagen types I, II, III, V, and XI. Fibril-associated collagens include collagen types IX, XII, XIV, XIX, XX, and XXI. Basement membrane collagens represent collagen type IV. Transmembrane collagens are comprised of collagen type XIII and XVII (Gelse et al., 2003; Ricard-Blum, 2011). Table 2.2 shows some of the common collagen types, their structures, and the tissues they are found in.

Table 2.2 Some common collagen types, their structure, and the tissues they constitute

Type	Molecular species	Tissues
I	[α1(I)]2, α 2(I) [α1(I)]3	Dermis, bone, tendons, ligaments, and the cornea.
II	[α 1(II)]3	Cartilage, vitreous body, and intervertebral discs.
III	[α1(III)]3	Skin, vessel walls, reticular fibers
IV	[α1(IV)]2 α2(IV); α1-α6	Basement membranes (basal lamina) including blood vessels
V	[α1(V)]2, α2(V) [α1(V)]3 [α1(V)]2α4(V) α1(XI)α1(V)α3(XI)	Lungs, cornea, bones, fetal membranes.

Source: Gelse et al., 2003.

2.2.3 Collagen Biosynthesis and Post-Translational Modification

The biosynthesis of collagen involves eight specific enzyme activities including the enzyme Prolyl-4-hydroxylase (P4H) which is critical for triple helix stability. Biosynthesis of the most common abundant (Type I) collagens begins inside the nucleus, where gene transcription, alternative splicing, capping, and polyadenylation, take place, to produce an mRNA. Cell type, growth factors, and cytokines regulate the transcriptional processes. Then, in the rough endoplasmic reticulum or Golgi apparatus, the ribosome-bound mRNA is translated into preprocollagen. A signal peptide helps this molecule to protrude into the lumen of the rough endoplasmic reticulum. This signal peptide is then removed by signal peptidase to give procollagen (Sionkowska et al., 2020).

The biosynthesis of collagen involves many complex post-translational modifications and some of these are exclusive to collagens. Three enzymes perform proline and lysine residue hydroxylation, and these require ascorbate, ferrous ions, molecular oxygen, and 2-oxoglutarate as cofactors. Other enzymes such as protein

disulfide isomerase, and peptidyl-prolyl-cis-trans-isomerase, further process this molecule, followed by procollagen N-proteinase and procollagen C-proteinase. Then, procollagen molecules are packed within the Golgi apparatus into secretion vacuoles, which are then released. In the extracellular space, the molecules undergo processing that depends on the collagen type (Bulleid et al., 2000; Gelse et al., 2003; Ricard-Blum, 2011). Figure 2.3 summarizes the process of collagen biosynthesis from DNA transcription and translation, up to the formation of the collagen fibers.

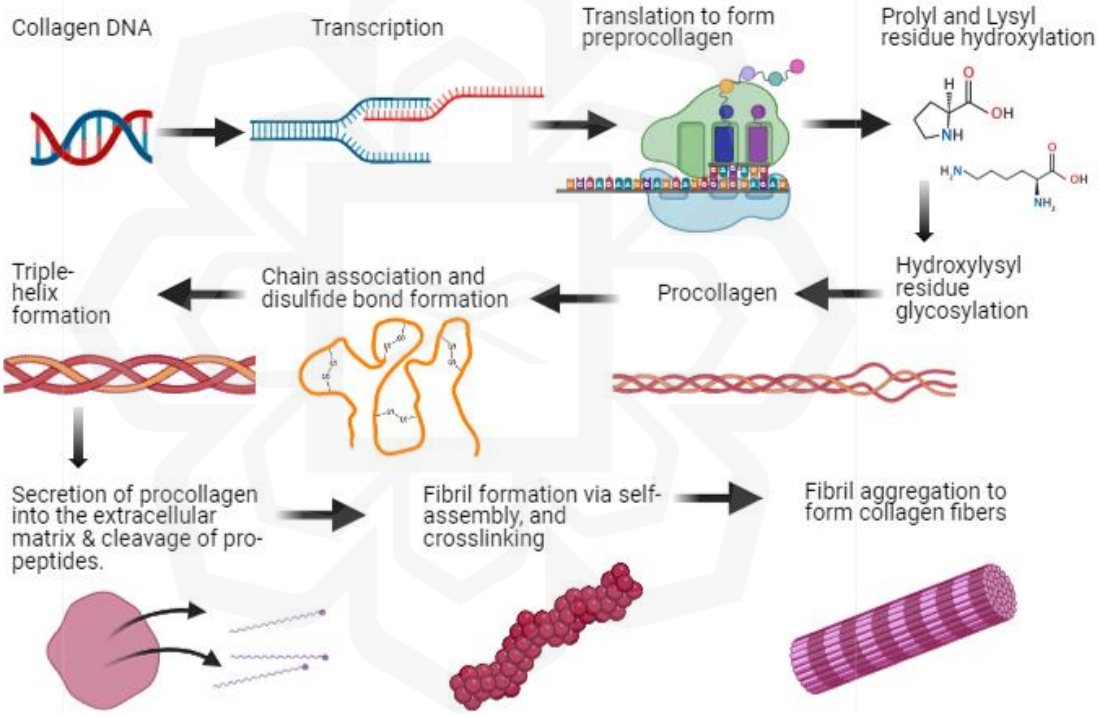


Figure 2.3 Biosynthesis of collagen (Sharma et al., 2008)

As previously mentioned, the biosynthesis of collagen involves unique post-translational hydroxylation, which is done by an iron-containing enzyme called prolyl 4-hydroxylase. The proline residue, α-ketoglutarate, and oxygen interact with prolyl 4-hydroxylase, thereby hydroxylating the proline to hydroxyproline, and the α-

ketoglutarate undergoes oxidative decarboxylation, producing succinate and carbon dioxide. In another reaction that the prolyl 4-hydroxylase enzyme also catalyzes, vitamin C (ascorbic acid) is involved, wherein the enzyme's iron becomes oxidized, thereby inactivating it. Vitamin C (ascorbate) then reduces the enzyme's iron, allowing it to return to its active state. As a deficiency in active prolyl 4-hydroxylase prevents post-translational modification of collagen (i.e., hydroxylation of the proline residues on procollagen), a vitamin C deficiency can thus create defective collagen (Gorres & Raines, 2010).

2.2.4 Collagen Sources

Collagen can be obtained from the bones, cartilage, and skin of various vertebrates as well as the exoskeletons and other structures in invertebrates. Thus, collagen sources include bovine, porcine, avian, amphibian (e.g., giant salamander) (Chen et al., 2021), and marine sources, including crustaceans, echinoderms, fish, jellyfish, mollusks, sea anemone, and sponges (Lukomski et al., 2018). Amputated earthworms also produce a CLP, col4a1, that has wound-healing properties (Du et al., 2021). Other sources include human cadavers (Browne et al., 2013) and human placenta (Karami et al., 2019). Animal collagen has been the main source of collagen, but concern over zoonosis, along with the desire for greater yield and cheaper production, has led to marine and recombinant sources being considered as alternatives to other animal collagens. Even the halal status of collagen and gelatin is an important issue today, as consumer awareness has risen on this matter (Rakhmanova et al., 2018).

Collagen extraction can either be a chemical or an enzymatic process. The collagen extraction process differs for salt-soluble collagen (extraction by saline solution), acid-soluble collagen (acidic extraction), and pepsin-soluble collagen (enzymatic extraction). These methods could positively and negatively affect the qualities of the extracted collagen, including but not limited to its thermal stability,

molar mass, water-retaining, and gel-forming abilities. Recently, ultrasound-assisted extraction (UAE) has also been utilized in the extraction of collagen, and it has led to higher yields without damage to collagen and with a reduced extraction time. The merits and drawbacks of collagen extraction techniques are highlighted in Table 2.3. Disadvantages of extracted animal collagens, such as variability across species and intraspecies, risk of zoonosis, and concerns related to immunogenicity, have prompted the investigation into alternative sources of collagen.

Table 2.3 Advantages and disadvantages of various methods of obtaining collagen

Method	**Advantages**	**Drawbacks**
Animal collagen extraction (acidic, or enzymatic hydrolysis)	Utilizes waste (byproducts of marine and animal slaughter processing) Yield can be high	Complex. May require a long pre-treatment step e.g., defatting, or homogenization. Batch-to-batch variability. Low purity. Costly downstream processing. Zoonosis risk. Immunogenicity risk. Non-compliance with religious restrictions. May affect the collagen's water retention and gel-forming capacity, among others. Collagen denaturation risk depends on the process type.
Ultrasound-assisted animal collagen extraction	Reduced extraction time Increased yield compared to conventional extraction. Retains structural and thermal stability of collagen	Studies on ultrasound-assisted extraction and its effects on the enzymatic digestion of collagen are inadequate. Prolonged ultrasonic application may increase temperatures and shear strength and may eventually cause denaturation of the protein.

De novo collagen synthesis (collagen mimics or mimetic peptides) e.g., solid-phase peptide synthesis	Simple Customizable Resembles and functions like animal collagens e.g. hemostatic hydrogels that are degradable by collagenase (Kumar et al., 2014). Non-inflammatory Consistent and reproducible	Costly Lacking self-assembling capabilities
Bacterial CLPs	Simple Can include novel motifs and domains. Non-immunogenic halal, kosher, suitable for all groups. Sustainable source. Consistency	Lacks post-translational modifications.
Recombinant human collagen expression in yeast, *E. coli*, and plants (maize, tobacco)	No batch-to-batch variability Sustainable supply of collagen	Costly setup and maintenance. Downstream Processing. Complex due to post-translational modification requirements for bacterial culture. Mammalian cell culture can be challenging to scale up.

Source: Schmidt, 2016.

Chemical methods of *de novo* synthesis of short collagen-derived peptides for research and tissue repair applications are another option to produce collagen fragments. However these fragments tend to lack distinguishing collagen characteristics such the collagen triple-helix, hydroxyproline and hydroxylysine residues, and resistance to proteolytic degradation. In addition to that, they synthesized peptides may require additional sequences to thermally stabilize them (Fertala, 2020; Wang et al., 2017).

Besides animal sources, full-length, human collagens have been expressed in yeasts and several transgenic plant, mammalian, and insect hosts. Moreover, several collagen fragments were successfully produced using *E. coli* (Davison-Kotler et al., 2019). One such example is human-like collagen (HLC), which is water-soluble (Fan et al., 2005; Luo & Mu, 2014; Luo et al., 2008) and offers many advantages over conventional animal collagens. For instance, human-like collagen has a low risk of immunogenicity and zoonosis, and it does not have batch-to-batch variability. Furthermore, its composition can be modified to produce novel human-like collagens that have many biomedical uses (Zhu et al., 2009).

2.2.5 Collagen-Like Proteins in Mammals, Bacteria, and Viruses

In addition to the aforementioned sources of collagen, an astonishing number of collagen-like protein (CLP) sequences were discovered in mammalian, bacterial, and viral genomes. Mammalian proteins that have collagen-like domains include membrane proteins, and secreted proteins, such as collectins and ficolins (Ricard-Blum, 2011). While "prokaryotic collagens" differ in residue content and distribution, they exhibit the basic molecular properties that are found in true collagens. In a Uniprot database search, a study identified approximately 18,874 CLPs in bacteria, 695 in viruses, and 157 in archaea (Lukomski et al., 2018). Some examples of these are presented in Table 2.4. Pathogenic *E. coli* produces native CLPs (*Ecl*) consisting of Gly-X-Y repeats and a C-terminal trimerization domain. Pathogenic streptococcal species produce native, homotrimeric Scl proteins on the cell surface, that were found to have fibronectin-binding domains and integrin-binding motifs that enhance their ability to adhere to and attack their hosts and evade the immune system. Also, several of these proteins help with biofilm formation (Ghosh et al., 2012; Lukomski et al., 2017). Some of these bacterial CLPs were produced in *E. coli* with high protein yields and they readily allowed the inclusion of desirable biologically active sequences from mammalian collagens (Brodsky & Ramshaw, 2017; Peng et al., 2012).

Table 2.4 Some prokaryotic collagens in the literature

Source Species/Strain	Protein Name	Reference
Enterohemorrhagic *E. coli* strains.	EPclA, EPclB, EPclC, and EPclD	Ghosh et al. (2012)
Shigella dysenteriae, Shigella boydii BS512.	EPclD	Ghosh et al. (2012)
Clostridium difficile	BclA1	Pizarro-Guajardo et al. (2014)
Clostridium perfringens	-	Xu et al. (2010)
Methylobacterium sp. 4-40	-	
Rhodopseudomonas palustris	-	
Solibacter usitatus	-	
Bacillus anthracis	BclA	Sylvestre et al. (2002)
Streptococcus pyogenes (GAS)	Scl1 and Scl2, or SclA and SclB	Lukomski et al. (2001); Xu et al. (2010); Xu et al. (2002); Rasmussen et al. (2003)
S. equi	Scl C, D, E, F, G, H, I, SclZ.1-5, 7, 12, SclZ.6, 9, 10, FneC, E, F	Lukomski et al. (2017); Karlstrom (2005)
S. zooepidemicus (GCS)		
S. pneumoniae, S. agalactiae (GBS)	Pcla, Phage minor structural protein	Lukomski et al. (2018)
S. pyogenes (GAS), S. equi, and *S. dysgalactiae (GCS).*	Phage-associated hyaluronidas	
S. pneumoniae	PclA	Paterson et al. (2008)
Legionella pneumophila	Lcl	Abdel-Nour et al. (2014)
Burkholderia pseudomallei and *B. mallei*	Bucl	Bachert et al. (2015); Grund et al. (2020)
B. amyloliquefaciens strain FZB42	ClpA, ClpB, ClpC, and ClpD	Zhao et al. (2015)

2.2.6 Recombinant Collagen Production

Successful recombinant collagen production has been realized in many ways. In the past two decades, tremendous progress has been made in terms of collagen production. The collagen can be human or mammalian (natural triple-helical collagens) and is expressed using various expression systems or hosts. Alternatively, collagen can be a CLP or a single-chain collagen-mimicking peptide polymer that is expressed in its original host

or a different expression system. Furthermore, the protein itself could be modified to incorporate more human-like functions (Rutschmann et al., 2014; Wang et al., 2017; Yu et al., 2014).

2.2.7 Recombinant Human Collagens

Recombinant human collagen is another potentially viable collagen production method, with advantages such as high similarity to native human collagen, batch-to-batch homogeneity, and reduced risks of zoonosis and immunogenicity. In addition, this would facilitate the production of less abundant collagen types that would otherwise be too cumbersome to extract from animal tissue. Several studies have been conducted using various host systems from bacteria, yeast, transgenic plants, mammals, insects, insect cell lines, and mammalian cell lines (including human cell lines). Initially, recombinant human collagen production was attempted in bacterial (e.g., *E. coli*) and yeast systems (e.g., *Pichia pastoris*). However, these hosts yielded fragmented and unstable fibers as they lack the necessary ability for post-translational modification of the collagen. The co-expression of enzyme α and β subunits of prolyl-4-hydroxylase (P4H) along with the collagen-coding sequence is possible in yeast and as a result, recombinant human collagen was expressed in yeasts. However, it is expressed only as a homotrimer, and often fragmented (Davison-Kotler et al., 2019; Graumann & Premstaller, 2006). Furthermore, the yields are quite low (Peng et al., 2018). Human prolyl hydroxylase had also been added into *E. coli* fermentation for hydroxylation of proline (Pinkas et al., 2011). While the enzyme is absent in most bacteria, *B. anthracis* has this enzyme, and this could be studied as a potential approach toward hydroxyprolination of recombinant collagen (Yu et al., 2014)

Transgenic plant systems such as the maize (*Zea mays*) and tobacco plants (*Nicotiana tabacum*) have also been explored as expression hosts for recombinant homotrimeric human collagen type I with fairly good results. Collagen I α and β strands

were successfully expressed in transgenic tobacco plants coexpressing propyl-4-hydroxylase (P4H) and lysyl hydroxylase 3 (LH3) enzymes, yielding heterotrimeric collagen that was post-translationally modified and showed identical amino acid sequence, and similar thermostability and protease resistance to native human collagen (Chattopadhyay & Raines, 2014; Davison-Kotler et al., 2019; Shoseyov et al., 2014).

Both whole insect organisms (e.g., transgenic silkworms) and cultured cells (e.g., Clonal High Five insect cells) were transfected with COL1A1 and COL3A1 genes with varying degrees of success. Recombinant human collagen type III production by injecting the cDNA-containing vector into silkworm eggs has also been attempted. However, the protein had inadequate proline hydroxylation due to low P4H enzyme activity in silk glands, and consequently it demonstrated low resistance to protease digestion and a decreased thermostability (Davison-Kotler et al., 2019). As for insect cell lines as hosts, recombinant human collagen type III which was produced within the Clonal High Five (HF) insect cells resulted in a significant intracellular 4-hydroxyproline production. Nonetheless, it was underhydroxylated and had a low thermal stability. On the other hand, including recombinant P4H and ascorbate in the culture medium increased collagen hydroxylation and exhibited enhanced thermostability (Davison-Kotler et al., 2019; Wang et al., 2017).

Transgenic mammals can be used to produce recombinant human collagen type I, such as in mouse embryos. The process resulted in substantial amounts of recombinant human collagen in mouse (*Mus musculus*) mammary glands. The resulting proteins were heterotrimeric, correctly folded, and secreted in a soluble form. Similarly, recombinant collagen type VII has been successfully produced in Chinese hamster ovaries. Human cell lines, including fibrosarcoma cells (HT1080) and embryonic kidney cells (293-EBNA), were also investigated and successfully used for producing recombinant human collagen types I, V, and VII. Although these collagens were identical to their native counterparts, the recombinant collagen yields were often lower compared to those of the other expression systems (Davison-Kotler et al., 2019; Wang et al., 2017), as shown in Table 2.5. Reviews covering various recombinant collagen

expression systems, including hydroxylation of proline and/or lysine, have been conducted over by Fertala (2020), Wang et al. (2017), Gomes and Salgueiro (2022), Shoseyov et al. (2014), and Rutschmann et al. (2014), among others.

Table 2.5 Recombinant human collagen expression systems

Expression System Type and Host		**Expressed Human Collagen**	**Reported Yield/ Concentration**	**Reference**
Prokaryote	*Escherichia coli*	Type I	1.88 g/L	Xie et al. (2023)
		Type II	0.27 g/L	Luo and Mu (2014)
		Type III	90 mg/L	Rutschmann et al. (2014)
Yeast	*Pichia pastoris*	Type I	1080 g/L	Li et al. (2015)
		Type III	0.7 mg/mL, 0.5 g/L, 2.33 g/L, 4.05 g/l, 1.05 g/L	Fang et al. (2022); Ma et al. (2022); Ma et al. (2014), Xiang et al. (2022), Wang et al. (2014)
	S. cerevisiae	Type III	0.6 g/L	Wang et al. (2017)
Plants (leaves and seeds)	*Nicotiana tabacum*	Type I	200 mg/kg fresh leaves	Rutschmann et al. (2014)
	Zea mays	Type I	NA	Setina et al. (2016);Xu et al. (2011)
Human cell lines	HEK 293 kidney epithelial cells	Type VII	2-5 mg/L	Hou et al. (2015)
	HT1080 fibrosarcoma cells	Type I and Type III	~1 mg/L	Wieczorek et al. (2015)
Other mammals	*Mus musculus* (mammary gland)	Type I	50–200 μg/mL	Davison-Kotler et al. (2019)

	Chinese Hamster Ovaries (CHO)	Type VII	< 0.5 mg/L	Hou et al. (2015); Ramshaw et al. (2014)
Insect	*Bombyx mori* (cells/larvae)	Type II	1 mg/larva	Qi et al. (2016)
		Type III	5 kg per 600 kg cocoon material	Tomita (2011)

2.2.8 Bacterial Collagen-Like Proteins

Over the past three decades, numerous CLPs have been discovered from bacterial genomic databases in microorganisms such as *Klebsiella pneumoniae*, *Bacillus cereus, Bacillus anthracis, Bacillus thuringiensis,* and *Pasteuria ramosa*. These bacterial triple-helices exhibit the repeating sequence of Gly-X-Y that is a characteristic of animal collagens. Some bacterial collagen-like sequences have yet to be studied or are only partially characterized in terms of their triple-helical structure. For instance, there are collagen-like sequences in two *Clostridium taeniosporum* spore appendage proteins, namely protein GP85, which contains 53 Gly-X-Y repeats, and protein CL2, which has 43 such repeats (Walker et al., 2007). Another example is complex triple-helical structures that were found in *Pasteuria ramosa* (McElroy et al., 2011; Mouton et al., 2009). Though numerous species were reported to have collagen-like sequences in their genomes (Rasmussen et al., 2003), the natural expression of these sequences has been demonstrated in only a few cases. This highlights a significant gap between predicted and experimentally verified bacterial CLPs, stressing the need for further structural characterization.

Besides the above-mentioned pathogenic organisms, collagen-like structures were also found in non-pathogenic organisms. The work of Xu and colleagues (2010) reports that an algorithm (Persikov et al., 2005) identified collagen-like structures from three soil bacteria that are not pathogenic, namely *Solibacter usitatus,*

Rhodopseudomonas palustris, and *Methylobacterium sp 4–46*, based on predictions on collagen stability. Out of the hundreds of potential bacteria studied (including some *Clostridium, Shigella, Lactobacillus, Streptococcus, Pseudomonas*, and *Salmonella* species), only eight of those obtained by the algorithm have been further studied and confirmed to have triple-helical collagen-like structures. The proteins obtained all consisted of a single collagen triple-helical domain that is uninterrupted, and they have non-collagenous domains. However, the finding did not reveal any characteristics that differentiated the CLP structures based on pathogenicity.

The bacterial CLPs are often involved with the outer membrane of the microbes, and some show cell adhesive properties. For example, *S. pyogenes* CLPs have a triple helix that extends beyond the cell wall. In this manner, bacterial CLPs resemble certain mammalian transmembrane collagens such as types XIII, XVII, XXIII, and XXV (Franzke et al., 2005; Ricard-Blum, 2011). The ectodomains of these transmembrane collagens, as well as those of several bacterial collagens, exhibit cell adhesion properties. The transmembrane collagens consist of a short N-terminal within the cytsol and a longer "C-terminal ectodomain" that includes multiple triple-helical domains(Yu et al., 2014). Bacterial collagens, in contrast, have a more variable orientation. For example, the *B. anthracis* CLPs have "a C-terminal triple-helix domain and globular domain extending out as hair-like filaments" (Yu et al., 2014). Other similarities between mammalian and bacterial CLPs are in the melting temperatures, resistance to trypsin digestion, as well as having a high calorimetric enthalpy (Yu et al., 2014). Bacterial CLPs do not contain the post-translational hydroxylation of proline and lysine residues to hydroxyproline (Hyp) and hydroxylysine (Hyl). Studies have demonstrated that these hydroxylated amino acids help to stabilize the triple-helical structure of mammalian collagen (Shoulders & Raines, 2009). Hydroxyproline contributes to the stability of mammalian collagen via interchain hydrogen bonding. Despite the absence of these amino acids in bacterial CLPs, the triple helices examined had demonstrated high thermal stability, with melting temperature (T_m) values ranging from 35–39 °C, which are similar to those of mammalian collagens. This high thermal stability is speculated to be on account of various interactions, including electrostatic ones or a

high content of a specific polar amino acid, and thus depends on the amino acid makeup of the specific bacterial collagen (Mohs et al., 2007; Qiu et al., 2021; Yu et al., 2014).

Bacterial collagen-like domain heterogeneity may be related to the diverse environments that the microorganisms had to adapt to. A bioinformatics study showed that the length and amino acid composition of the collagenous domains of prokaryotes are much more heterogeneous and diverse compared to those of animals (Bella, 2014). Another difference is that there is a lack of aggregation in bacterial collagens, and this is due to the absence of hydroxyproline (Peng et al., 2018). Repeating amino acid sequence patterns, of various repeat frequencies, such as GKDGKDGQNGKDGLP sequences in Scl2 from *S. pyogenes*, GPKGEP sequences in CLPs from *M. sp4–46*, and (GPT)5GDTGTT sequences in BclA from *B. anthracis* feature frequently in bacterial collagens, unlike mammalian collagens (Yu et al., 2014). Natural, higher-order structures of bacterial collagens have not been reported to date, though some of them have been observed to aggregate *in vitro*. This general lack of aggregation is another difference between bacterial collagens and mammalian collagen and is believed to be linked to their lack of hydroxyproline (Peng et al., 2018). The genetic distribution, molecular diversity, structures, and functions of bacterial CLPs have been reviewed in a recent journal article (Qiu et al., 2021).

Collagen-like triple-helical domains from the Scl2 gene of *Streptococcus pyogenes* represent perhaps one of the most investigated bacterial CLPs (Cosgriff-Hernandez et al., 2010; Mohs et al., 2007; Peng et al., 2012; Xu et al., 2002; Yoshizumi et al., 2009). Scientists have successfully produced recombinant *S. pyogenes* CLP, named Scl2, in *E. coli* (strain BL21) in shake flasks (Mohs et al., 2007) and then upscaled this to a stirred tank bioreactor, utilizing a high-cell density fed-batch system (Peng et al., 2012). Protein concentrations of approximately 1 g/L were obtained in the flask culture and stirred tank bioreactors. Working in high cell density fed-batch mode and using fully defined media increased concentrations to around 10 g/L. Extending the production at high cell density to 24 hours further increased the concentration to around 19 g/L (Peng et al., 2012), and the resultant concentrations were higher than the average

concentrations (14 g/ L) reported in other studies employing bacterial expression (Yu et al., 2014). The resulting protein was shown to be neither cytotoxic nor immunogenic (Stoichevska et al., 2017). The process itself had many advantages, with high tunability, scalability, low complexity, and low cost, especially when compared to recombinant human collagen production (An et al., 2014).

In addition, the Scl2 protein does not show much interaction with mammalian cells and, therefore, functional domains can be introduced to make the Scl2 protein mimic certain human collagens (Cosgriff-Hernandez et al., 2010; Peng et al., 2018; Peng et al., 2014c; Stoichevska et al., 2017). A study reported on the successful modification of the Scl2 protein by introducing two new binding motifs (i.e., heparin and integrin-binding sequences) from animal collagen-like sequences. It was found that stability decreased slightly with these modifications (Stoichevska et al., 2017). Furthermore, the preparation of constructs of one to four copies of Scl2 collagen domains has been successful in overcoming the drawback of the size of collagen, as Scl2 is only about one-fourth the length of collagen type I, II, or III. However, it was discovered that as the size grew, the collagen yield decreased. In fact, the yield was halved for each construct added (Peng et al., 2014a). Customizable recombinant bacterial collagens will be tremendously useful in the biomedical and pharmaceutical industries if they can be successfully produced in copious quantities. To enable their commercialization, it is important to demonstrate the *E. coli* fermentation process's scalability, its purification at the pilot scale, and its economic viability (Ramshaw et al., 2014).

2.2.8.1 The Role of Bacterial Collagen-Like Proteins

A few studies have attempted to unravel the role of native bacterial CLPs in their respective bacteria, as shown in Table 2.6. In one, the CLPs of *B. amyloliquefaciens*

FZB42 were studied, and it was concluded that they affect the bacteria's biofilm formation and may form an envelope that supports the cell. It was observed that the extracellular matrix was separated slightly from the cells when clpB and clpC were inactivated, suggesting that these CLPs could be needed to maintain the structure of the bacterial extracellular matrix. Furthermore, the shape of the bacteria was also influenced as the wild-type cells of *B. amyloliquefaciens* FZB42 showed a three-dimensional shape with a "jelly-like" coating over the extracellular matrix, whereas the clpB mutants were "flattened" and had a "long rod-like" shape (Zhao et al., 2015).

Similarly, several studies show that two *S. pyogenes* bacterial CLPs have sequences that suggest that they are attached to the cell surface and enable the bacteria to adhere to a number of host proteins (Yu et al., 2014). For instance, "the non-collagenous V-domain of Scl1 may bind to high-density lipoprotein" (HDL, low-density lipoprotein (LDL), factor H (Caswell et al., 2008), complement factor H-related protein 1 (CFHR1), and fibronectin and laminin. The Scl1 protein was found to have a fundamental role in biofilm development by targeting a cellular fibronectin type. Often, this binding can help the bacteria evade the host's immune response, such as phagocytosis or proteolysis. It also enhances adherence to the host cells, such as macrophages, and the extracellular matrix. Also, *S. pyogenes* bacterial CLPs "bind to thrombin-activatable fibrinolysis inhibitors (TAFI, procarboxypeptidase)", allowing them to evade proteolysis. The collagen-like domain of Scl1 "mimics mammalian collagens in interacting with collagen receptor integrins α2β1 and α11β1 via a GLPGER binding site". In doing so, it allows the pathogen's access to host cells and helps in activating intracellular signaling and finally assisting in its exit from the host cell (Yu et al., 2014).

Table 2.6 Localization and function of the bacterial CLPs

CLP	Source	Location	Special Motifs	Function	Reference
BclA and BclB	*B. anthracis*	Exosporium surface proteins thin hair-like surface filaments	The collagen-like region of BclA consists of seventy (70) triplet repeats (GXX) including fifty-four (54) GPT triplets. It has a 21 amino acid motif: (GPT)5GDTGTT that is repeated six times.	BclA is expressed only in the sporulating cells, and it is an important structural component of the filaments that cover the outer layer of the exosporium.	Sylvestre et al. (2002)
BCLPA, BCLPB, BCLPC, BCLPD	*Bacillus amyloliquefaciens* FZB42	outer membrane protein or flagellar proteins Extracellular matrix	-	Give cells a three-dimensional shape with a “jelly-like” coating over the extracellular matrix. CLPs contribute to biofilm formation and are hypothesized to support bacteria by forming a “cross-linked” envelope. Play a significant role in cell auto-aggregation.	Zhao et al. (2015)
Scl1 and Scl2	*Streptococcus pyogenes*	cell surface	SclA and SclB contain the cell wall anchoring motif LAPTGE.	Mediate GAS-cell surface hydrophobicity, which contributes to biofilm formation.	Yu et al. (2014); Yu et al. (2011)

	Pasteuria ramosa			
Epcl	*E. coli*	hypothetical tail fiber proteins	May have a role as a “trimeric phage side-tail protein” that aids in the attachment of phage particles to *E. coli* cells.	Ghosh et al. (2012)
Bucl8	*Burkholderia pseudomallei, Burkholderia mallei*	-	Contains two predicted tandem outer membrane efflux pump domains, an extended extracellular region containing a collagen-like (CL) domain, and a non-collagenous C-terminus (Ct).	(Grund et al., 2020)

Like their mammalian counterparts, bacterial collagens have variable domains in addition to the triple helix. Aside from signaling and evading the immune system, the non-collagenous domains were reported as vital for the collagen-like domain to efficiently fold into the distinctive triple helix. In most mammalian collagens, these trimerization domains are situated on the C-terminal of the triple-helix domain (Khoshnoodi et al., 2006). In *S. pyogenes* Scl1 and Scl2 proteins, the N-terminal globular domains (variable domains) differ in length and amino acid composition across strains, but all have a high α-helix content (Yu et al., 2011). The non-collagenous domains of *S. pyogenes* are needed for the refolding of the triple-helical domain *in vitro* (Yu et al., 2014). The variable domains were also found to bind to the proteins of the extracellular matrix and several plasma components. Similarly, in *B. anthracis*, the C-terminal globular domain composed of β-sheets secondary structure is speculated to be vital for the folding and stability of the BclA triple-helix, whereas its N-terminal non-collagenous domain is essential for attachment to the extracellular matrix (Boydston et al., 2005; Yu et al., 2014). Another noteworthy point about the variable domains is their ability to initiate the proper folding of bacterial collagen-like domains from other species, and this suggests that the variable domains have a versatile nature (Lukomski et al., 2017; Yu et al., 2010). In addition to trimerization ability, the non-collagenous domains of most collagen types are also associated with numerous functions, including angiogenesis, cell proliferation, and apoptosis, such as the NC1 domain of human collagen type IV (Yu et al., 2014).

2.2.8.2 Rhodopseudomonas palustris Collagen-Like Protein

Rhodopseudomonas palustris is a species of purple, non-sulfur bacteria commonly found in soil and water environments. *R. palustris* (strain TIE-1) belongs to the class alpha-proteobacteria (order Rhizobiales or Hyphomicrobiales) and the family Bradyrhizobiaceae. It is a Gram-negative, non-pathogenic, and non-sporulating bacteria. It is a very metabolically versatile microorganism that is capable of being

photoautotrophic (photosynthetic), photoheterotrophic, chemoheterotrophic, as well as chemoautotrophic. It can also absorb atmospheric carbon dioxide and convert it into biomass, fix nitrogen gas into ammonia, and can produce hydrogen gas. It can degrade carbon-containing compounds such as lignin and methanol as well. It grows both in the absence and presence of oxygen (Larimer et al., 2004). *R. palustris* was found to have a gene that encodes for a CLP in a study (Xu et al., 2010). Some of the characteristics of the protein are highlighted in Table 2.7, and its amino acid composition and distribution are shown in Figures 2.4-2.5.

Table 2.7 Characteristics of *R. palustris* CLP

Characteristic	**Value**
Protein Accession Number (UniProtKB)	Q6N3V0
Molecular Weight (YP_0019930)	22.071 kDa
Theoretical molecular weight of the construct	23.5 kDa
Theoretical triple-helical domain length	11.3 kDa
Length	212 amino acids (639 bp)
Isoelectric Point (pH at which solubility is minimal)	9.32
Melting temperature	37 ℃

Sources: Xu et al., 2010; Yu et al., 2014

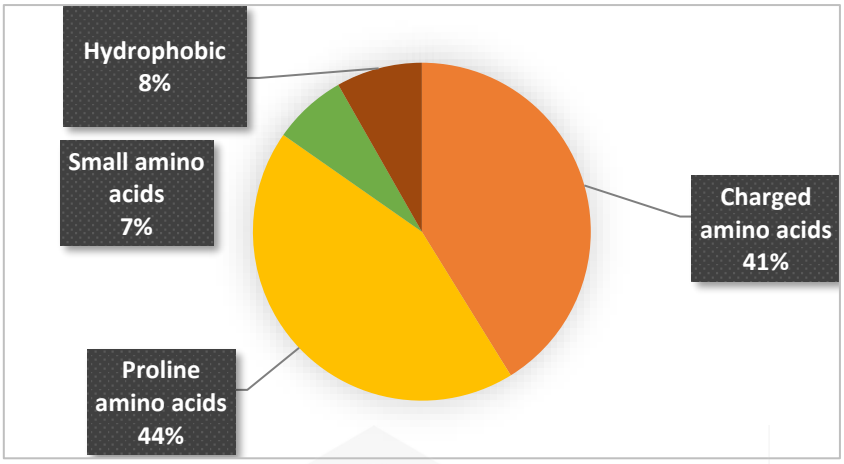
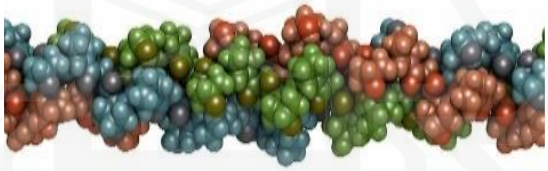


Figure 2.4 Amino acid composition of *R. palustris* CLP (Xu et al., 2010)



WP_012497020.1 ,$(Gly\text{-}Xaa\text{-}Yaa)_n$
Full length: 212 aa
1-9 aa: N-terminal
10-126aa: collagenous domain
127-212aa: C-terminal

Figure 2.5 *Rhodopseudomonas palustris* collagen-like protein length and domains (https://www.ncbi.nlm.nih.gov/nuccore/192288433/?from=4409667&to=4410305)

>NC_011004.1:4409667-4410305 *Rhodopseudomonas palustris* TIE-1, complete genome 639 bp

1 gtggctgaag ctgaatccac ccgccgcgga cgtccggggc cgcagggccc gcgcggacgt

61 cccggagaac cgggccgtcc gggaccgcaa ggacatccgg gccgtcccgg ccccgagggc

121 ccccgaggca agcaaggtcc ggtcggcaag cccggcccgc agggcaaggc cggtccgcaa

181 ggcaaacccg gcatcgccgg caagccgggc cccgatggca agcccggtcc gatcggcccg

241 caaggcaagg ccggcccgca aggtccccgc ggtgaacaag gcctgcgcgg cgagcaaggt

301 cctcgcggtg agcaaggccc gcaggggccg cgtggcgaac agggcccccg gggcgagccc

361 ggccccgccg gggcgctccc ctcgatcgag caggtgatgc cctggctgca cctgatcttc

421 gacgcctacg aagattacaa agcgcagcgc gcccgcgaag cccgcgagct cgaagagcgt

481 ctcgccgccg aagctctcga acaggcggcg cgcgaagccg ccgagcgcga agtcgccgcc

541 gctatcgaag ccgcaaatgc cgaggccgag atcatgctcg acgatgagac gcatgccgag

601 ggcggcaaga aaaagaagaa gcgcaagcac aaggactga

Protein (Calculated: 22.07653 kDa):

MAEAESTRRGRPGPQGPRGRPGEPGRPGPQGHPGRPGPEGPRGKQGPVGKPGPQGKAGPQGKP GIAGKPGPDGKPGPIGPQGKAGPQGPRGEQGLRGEQGPRGEQGPQGPRGEQGPRGEPGPAGAL PSIEQVMPWLHLIFDAYEDYKAQRAREARELEERLAAEALEQAAREAAEREVAAAIEAANAEAEIMLD DETHAEGGKKKKKRKHKD

(https://www.ncbi.nlm.nih.gov/nuccore/192288433/?from=4409667&to=4410305)

The polyhistidine (His-tag), typically composed of six (6) consecutive histidine residues fused to the protein of interest at either the N-terminal or the C-terminal, is widely used in protein detection and purification via immobilized metal affinity chromatography, based on the interaction between the amino acid residues and divalent metal ions on the resin. Imidazole is used to elute the His-tagged protein. Metal affinity chromatography is further discussed under the protein purification and concentration subsection.

2.2.9 Challenges in Recombinant Collagen Production

The greater challenges for a viable large-scale production of recombinant collagens lie in the constraints of yield maximization and post-translational modifications such as hydroxylation of proline into hydroxyproline and lysine into hydroxylysine. Human collagen is 42-54% hydroxylated (Wang et al., 2017). On one hand, there are mammalian, plant, and insect expression systems. While these systems express the recombinant human collagens with post-translational modification to various degrees, they provide relatively poor yields of the recombinant protein and require high maintenance. On the other hand, there are microbial expression systems that offer rapid growth, high yields, easy manipulation, and efficient expression of the target proteins. *E. coli* strain BL21, for example, is an efficient expression host (Bahreini et al., 2014). However, they do not possess the machinery for post-translational modification. Co-translational incorporation of hydroxyproline during fermentation may be a solution to this issue. This has been described by Peng and colleagues (Peng et al., 2018), and they have discovered that the addition of sodium chloride (NaCl) augmented the level of hydroxyproline incorporation via hyperosmotic shock. The optimal concentration of sodium chloride was found to be 200 mM. Brodsky and Kaplan (2013) summarize the challenges of recombinant type I collagen into seven points; namely, post-translational modifications of proline and lysine, improving yield, bioactivity, triple-helix stability, preventing degradation, and the formation of heterotrimers as well as periodic fibrils that mimic those formed in Type I collagen.

2.2.10 Collagen-Like Protein Detection and Quantification

Unlike mammalian collagens, bacterial CLPs, such as RPCLP, often lack post-translational modifications like hydroxylation of proline, limiting the applicability of hydroxyproline-dependent assays. Therefore, CLP detection and quantitation methods

must be selected and/or developed based on other criteria rather than traditional animal collagen biomarkers.

Before the expression of a recombinant protein and its purification are conducted, a protocol to detect and quantify the aforementioned protein must be prepared. This is necessary for the optimization of fermentation and expression conditions, as well as during purification. There are numerous techniques at hand for protein detection, whether qualitative or quantitative. Generally, these may involve chromatography, immunoassays that either rely on a tag's properties (e.g., hexahistidine tags) or are compared to a purified standard, SDS-PAGE, specific functional assays (e.g., enzyme activity), standard colorimetric absorbance methods such as the Bradford assay, and Lowry assay, or fluorescence assays (Yusof, 2015). Animal collagen is detected and quantified through assays that rely on the hydroxyproline content of collagens, such as the Sirius red collagen assay and protease digestion e.g., Proteinase K digestion. In detecting native collagens, their distinctive abundance of hydroxyproline and proline residues is often exploited. Table 2.8 summarizes these methods, as reviewed by Gameil et al. (2021b). From the summarized techniques, only a limited number (including those relying on glycine N-termini or triple-helix specific probes) show promise for non-hydroxylated CLPs.

Emerging technologies such as LC-MS/MS-based label-free quantification have the potential for precise, high-throughput CLP detection. Currently, their application to bacterial CLPs remains underexplored even though they may offer improved specificity over dye-based or colorimetric assays. However, these require significant capital investment to set up.

Table 2.8 Recent collagen detection methods, their principles, advantages and disadvantages, and collagen form.

Method	Collagen form	Principle	Pros	Cons	Reference
Bacterial collagenase digestion + Fluorometric	Recombinant, Animal	3, 4 DHPAA forms a fluorescent complex with N-terminal glycine peptides	Rapid, Selective Sensitive, Not limited to mammalian collagen	Requires a fluorometer with the correct filters	Yasmin et al. (2014)
Bacterial collagenase digestion + Colorimetric	Recombinant, Animal	o-phthalaldehyde	Not limited to mammalian collagen,	Not as selective and sensitive as 3,4 DHPAA	Yasmin et al. (2012) Go et al. (2008)
Collagenase digestion and MALDI-TOF-MS	Denatured collagen	Amino acid sequence	Sensitive	Time-consuming	Nimptsch et al. (2011)
GO-based ssCOLplatform	Collagen peptides	Single-stranded collagen (ssCOL) that has been labelled with a dye can be used as a peptide probe that targets a complementary single-stranded collagen sequence. Graphene oxide can then bind with the probe and alter the fluorescence of the dye. In contrast, the hybridization of the probe with its target collagen peptide GPO retains the fluorescence of the probe.	Sensitive, Selective	Specific probes must be designed for specific peptides	Sun et al. (2016)
Modified SirCol collagen assay (Trypsin digestion)	Mammalian collagen/ hydroxyproline-rich collagen	Sirius red dye stain binds collagen and the intensity of color can be measured by a colorimetric assay.	Simple, Rapid, Relatively cheap.	Only works on collagens containing hydroxyproline residues	Lareu et al. (2010)

Source: Gameil et al., 2021b.

2.2.10.1 Glycine-based Collagen Assays

Previous studies (Boydston et al., 2005; Vandersmissen et al., 2010; Yu et al., 2014) have shown that the CLPs from *Legionella pneumophila*, *Bacillus anthracis* (BclA), and *Streptococcus pyogenes* (Scl1 and Scl2) are susceptible to *Clostridium histolyticum* collagenase digestion, generating short N-terminal glycine peptides. Hence, assays have been developed to quantify CLPs by using this principle (Yasmin et al., 2014; Yin et al., 2002).

2.2.10.1.1 Ninhydrin-based Collagenase Assay

An assay was developed by Yin et al. (2002) as a simple and rapid method to quantitate collagen-like polymer in crude lysates from *E. coli*. Briefly, after harvest, sonication was utilized to lyse the cells. The cell debris was clarified by centrifugation, and removal of excess host proteins was accomplished via a high-temperature (80-100 ºC) precipitation step. Following the precipitation, the supernatant sample was digested by using a *Clostridium histolyticum* collagenase Type VII. This bacterial collagenase cleaves the sequence Pro-X-Gly-Pro between neutral amino acids (X) and glycine, generating many polypeptide fragments from the CLP polymer. These were then incubated with ninhydrin (triketohydrindenhydrate) dissolved in 50% methylcellosolve (ethylene glycol monomethyl ether), which reacts with the free amino groups to form a purple complex. The absorbance of the purple complex at 570 nm wavelength can then be measured and compared against standard curves using bovine serum albumin (BSA) and gelatin (Whittemore, 2005).

2.2.10.1.2 Fluorometric Methods

More recent publications on the detection of native and recombinant collagens have attempted to make use of another unique characteristic of collagen, namely its characteristic triple helix. The collagen triple helix is resistant to protease digestion by pepsin and trypsin but can be fragmented into tripeptides by bacterial collagenase. After that, amine-labeling derivatization can be carried out using various fluorophores to quantitate the peptides containing a free N-terminus or lysine (Yasmin et al., 2014; Yasmin et al., 2018; Yasmin et al., 2015; Yasmin et al., 2012). The reaction with the amines creates a notable increase in fluorescence that is linear in response to increasing protein concentration. Three dyes are commonly used for fluorescence, namely; o-phthalaldehyde (OPA), Fluorescamine, and 3-(4-carboxy benzoyl) quinoline-2-carboxaldehyde (CBQCA). Fluorescamine reacts directly with the amine functional group, whereas OPA and CBQCA dyes require the addition of 2-mercaptoethanol and cyanide, respectively (Noble & Bailey, 2009).

2.2.10.2 Other Methods

2.2.10.2.1 Sirius Red Method

The distinctive abundance of hydroxyproline is extensively used in the detection of collagen via the picrosirius red dye. This is possible because very few other mammalian proteins, such as elastin and argonaute-2, possess similar hydroxyproline-containing collagen-like domains. Sirius red is an anionic dye with sulfonated acid side-chain groups that react with the side-chain groups of basic amino acids, and it can be used to selectively stain collagen tissue (Lareu et al., 2010).

2.2.10.2.2 Bradford Assay

There are many methods to quantify total proteins after cell lysis, such as measuring absorbance at 280 nm, the Biuret method or Lowry estimation, the BCA estimation method, and the Bradford assay. Each of these methods has its own merits, and hence, a suitable method must be selected depending on the protein sample. For instance, measuring the absorbance at 280 nm is a very quick method, but higher concentrations of sodium chloride can interfere with the assay. The main purpose of the Bradford assay is to quantify the total protein concentration in a certain sample via a quick and simple procedure. Similarly, phenolic compounds interfere with the Biuret and Lowry estimation methods. The BCA is another fast method of protein estimation, but lipids, tyrosine, phenolics, and reducing sugars interfere with it. Other methods of quantification, such as infrared absorption, NMR, and thermal decomposition, exist, but they may require specialized equipment or may denature the protein sample (Yusof, 2015).

In this research, the Bradford assay was selected for quantifying the total proteins after cell lysis. Zhang et al. (2009) employed the Bradford total protein assay to quantify recombinant collagen I α1, using bovine serum albumin as the standard. It works on the premise that proteins bind specifically with Coomassie blue G-250 dye, under acidic conditions, producing a color change from red to blue, with a peak absorbance at 595 nm. It is a very rapid, accurate, and sensitive method. Another advantage is that the process is not affected by potassium and magnesium ions, EDTA, Tris, thiols, and carbohydrates. However, some phospholipids or detergents such as sodium dodecyl sulfate (SDS) and Triton-X 100, in high concentrations, may interfere with the assay (Yusof, 2015).

2.2.10.2.3 Coomassie Blue Dye (R250) Metachromasia

Some papers describe this method as a color shift, from blue to pink, observed at maxima 520 to 535 nm, after 24 to 72 hours of the addition of a destaining solution of acetic acid (pH 2-4). The principle for this method is thought to be the closely spaced proline and hydroxyproline residues (Farajollahi et al., 2019).

In summary, despite growing interest in CLPs, standardized detection protocols specifically tailored to them remain lacking. While ninhydrin-based and fluorometric assays are specific due to collagenase digestion, their reliance on heat treatment may compromise the structural integrity of some CLPs. On the other hand, methods such as the Bradford and Coomassie assays lack specificity and are less suitable for selective monitoring during fermentation scale-up. Therefore, the development of structure-specific, non-tag-based quantification methods that do not rely on hydroxyproline content is imperative. There is a paucity of comparative studies evaluating the sensitivity, specificity, and scalability of the different detection methods for recombinant CLPs, and this warrants future investigation.

2.2.11 Separation Techniques (Cell Harvest, Lysis, and Clarification)

As the desired recombinant RPCLP product is soluble (Xu et al., 2010) and is located intracellularly, there is a need to break the cells. Key research on the production of recombinant CLP using *E. coli* shows that the researchers preferred ultrasonication as a cell lysis method, as shown in Table 2.9. Many methods to lyse the cells exist, and they can be categorized into mechanical, chemical, and enzymatic methods.

Mechanical methods include liquid homogenization, such as the use of a French Press, and high-pressure homogenization, whereby cell suspensions are forced through a narrow space, thereby shearing them, or in the latter case by applying a large pressure difference over a small volume. In another approach, repetitive freezing and thawing cycles that are achievable using a freezer, dry ice, or ethanol, disrupt cells through the formation of ice crystals. Sonication is another commonly used technique of mechanical cell disruption wherein pulsed, high-frequency sound waves rupture cells by the formation and implosion of microscopic bubbles that create shock waves, i.e., cavitation. These waves are emitted by a vibrating immersion-type probe. The ultrasonic treatment is typically applied to the heat-sensitive sample while it is immersed in an ice bath. Furthermore, it is applied in multiple, short bursts, with rests in between bursts. While sonication has been preferred in the lab-scale, larger scale studies will require continuous sonification and studies have reported a significant loss in protein yield due to heat generation at prolonged exposure. A recent study has shown that the temperature and mixing can be controlled by optimization of tip immersion depth and analysis by using finite element heat transfer modeling (Ferdous et al., 2021). High-pressure homogenization, on the other hand, offers better scalability with lower shear-induced denaturation rates, though it can initially be capital-intensive (Shehadul Islam, 2017).

Enzymatic digestion refers to the use of enzymes such as lysozyme, lystostaphin, and proteases, to degrade cell walls and membranes, and represents an important technique for various applications such as protein and DNA extraction. The method is lauded for its versatility, high specificity, and gentle. On the downside, enzymatic lysis introduces which can complicate downstream processing. Also, to enhance effectiveness, they are often used in combination with DNases and detergents, and these require optimization. Scientists have also explored combining enzymatic digestion with mild sonication to reduce heat generation for thermolabile recombinant proteins (Shehadul Islam, 2017).

Alternatively, cells may be disrupted by treatment with chemicals that disrupt the integrity of the semi-permeable membrane and cell wall, leading to the release of cell contents. The choice of detergent for cell lysis is determined by the cellular location of the protein product of interest. Generally, non-ionic and zwitterionic detergents such as CHAPS and Triton-X are milder than ionic detergents. They are used to gently solubilize membrane proteins while retaining protein function and structure for enzymatic assays. Another factor that affects the choice of detergent for cell lysis is sample type and cell type. Bacterial cells, specifically *E. coli*, have a cell wall that makes lysis difficult, and hence the disruption needs to be vigorous (Whittemore, 2005).

In addition to the location of protein and choice of detergent, it is imperative to consider the lysis in terms of buffer, pH, salt concentration, and temperature conditions. Increasing the sodium chloride concentration may prevent the copurification of host cell native proteins; however, it will incur costs and additional steps, such as desalting, in the purification phase. Consideration should also be given to the compatibility of the chosen detergent with downstream applications, as well. For instance, additives such as EDTA are incompatible with the purification of His-tagged recombinant protein by a nickel column affinity chromatography and should not be added to the lysis buffer if this purification method is intended. If the detergent needs to be removed, a dialyzable detergent should be chosen for the buffer. Finally, incomplete lysis or excessive subcellular component release can significantly increase the lysate viscosity and hamper the subsequent centrifugation and clarification steps. Hence, inclusion of DNases should be considered (Whittemore, 2005).

Table 2.9 details the separation steps and buffers that have been employed to separate CLPs. Further studies are needed to optimize the CLP separation process. Given the structure of RPCLP, it is plausible that shear stress from aggressive methods like the use of a French press and high temperature generation during lysis may impair its integrity, warranting a combination of gentle lysis strategies where possible. Most studies reported the use of sonication and/or enzymatic lysis (Peng et al., 2014b; Peng

et al., 2012; Xu et al., 2010), and there is limited information on cell lysis processes in relation to RPCLP.

After cell lysis, the cell debris must be separated from the soluble product, and this is easily achieved by centrifugation. After the centrifugation, the cell lysate can be collected from the insoluble fraction (cell pellet). Clarified supernatant can then undergo dialysis, buffer exchange, desalting, or any subsequent steps toward downstream processing. After adjusting to pH 8.0, a second digest with trypsin added to 0.01 mg/mL may be performed for 16 hours at 4 °C to eliminate a major portion of the host proteins (Peng et al., 2014b).

Table 2.9 Lysis methods, parameter settings, and buffers used to obtain the intracellular CLP

Lysis Method	CLP	Buffer composition/Additive (if any)	Lysis Buffer amount	Parameter Settings/ Method	Reference
Mechanical (Sonication)	*Streptococcus pyogenes, Solibacter usitatus, and Methylobacterium sp 4-46*	50 mM Acetate Buffer (pH 2.2)	20 mL / g wet cell paste	5 minutes on ice	Peng et al. (2015)
	Various	50 mM Acetate Buffer (pH 2-7)	2/ g wet cell paste	using Misonix S4000 with Enhance Booster #1 Probe, at 30 A for 5 minutes	Peng et al. (2014b)
	Various	20 mM sodium phosphate buffer, pH 7.4, 500 mM sodium chloride, 10 mM β-mercaptoethanol	NA	4-5 x 1 min Bursts using Ultrasonic processor XL sonicator (Misonix).	Xu et al. (2010)
	Various	20 mM sodium phosphate, pH 8.0	20 mL/g	using Misonix S4000 with Enhance Booster #1 Probe, at 30 A for 6 x 5 minutes on ice.	Peng et al. (2014b), Peng et al. (2014a), Peng et al. (2014c)
	Streptococcus pyogenes	40 mM sodium phosphate, pH 8.0, 1 mM phenylmethylsulfonyl fluoride (PMSF)	20mL / g wet cell paste		Peng et al. (2012)
	Streptococcus pyogenes	40 mM sodium phosphate, pH 8.0	20 mL/ g wet cell paste		Peng et al. (2014c)
Chemical lysis	Streptococcal collagen-like protein (scl)	Bugbuster (Merck)	40 µL Typically, 5 mL/ g wet cell paste, at room temperature	Incubated for an hour	Peng et al. (2014a)

	Burkholderia collagen-like protein 8 (Bucl8)	Tris Buffer, 50 mM NaCl, 2 mM $MgCl_2$, 2% Triton X-100, 10 mM β-Mercaptoethanol, 0.2 mg/ml lysozyme, 1 mL Protease inhibitor (Pierce), 1 mM, 10 µg /mL PMSF.	10 mL/g cell pellet		(Grund et al., 2020)
	Streptococcal collagen-like protein (scl)	Cell Lytic B (Sigma) with 1 mM PMSF and 50 U/mL DNase I	200 µL Typically, 10-20 mL/ g wet cell paste, at room temperature		Lukomski and McNitt (2020)
Mechanical (French Press)	*Streptococcus pyogenes*	NA	NA	NA	Yoshizumi et al. (2009)

Protein solubility is influenced by ionic strength. At lower ion concentrations (typically under 0.5 M), solubility tends to rise as the ionic strength increases. This is because the ions in the solution help shield protein molecules from one another's charges. However, at higher ionic strengths, protein solubility decreases due to the "salting-out" effect. This phenomenon occurs because the salt is excluded from the layer of water near the protein surface, which plays a crucial role in maintaining the solubility and correct folding of the protein. As salt concentration increases, water becomes less available to solubilize proteins, enhancing hydrophobic interactions among protein molecules. Consequently, the protein decreases its surface area by folding to minimize contact with the solvent, and then self-association leads to its precipitation. As folding and precipitation free up bound water molecules, the salting-out process is thermodynamically favorable. As a result, this technique can be used to fractionate proteins according to their solubility in the presence of high salt concentrations (Wang et al., 2007).

The salts that raise the surface tension of water the highest will lead to salting-out, approximately following the Hofmeister series. As ammonium sulfate has a much higher solubility, it makes a suitable reagent for salting out and is thus the most commonly used salt to separate complex proteins via stepwise precipitation. It is typically included in protein purification protocols as a crude step. Ammonium sulfate precipitation can be used in the purification, concentration, as well as folding, and stabilization of protein structures.

Ammonium sulfate may be added incrementally to *E. coli* cell lysates to isolate a recombinant protein of 20 kDa containing no cysteine residues or tags (Arshad et al., 2014). As the precipitation is a result of the reduced solubility and the protein is not denatured, after centrifugation it can be promptly resolubilized using standard buffers such as sodium phosphate buffer. After concentration, the protein can undergo gel filtration whereby the buffer can be exchanged, and the ammonium sulfate is removed. Following initial concentration by ammonium sulfate precipitation, more refined purification was achieved using affinity chromatography, as discussed next.

2.2.12 Protein Purification and Concentration

Many studies were carried out on the purification of collagen using different methods. A systematic literature review reported on 12 studies, of which five studies were conducted using recombinant collagen, four studies using mammalian collagens, and three studies on marine collagens. Out of the five studies, two reported on the purification of bacterial CLPs, while the other studies reported purifying recombinant collagens from other sources. Table 2.10 shows some examples of the methods used to purify collagen irrespective of their sources.

Table 2.10 Summary of selected separation and purification methods of collagens

Separation Strategy	Form of collagen	Advantages	Drawbacks	References
Acid precipitation & proteolysis (pepsin digestion)	Recombinant non-animal collagen	- High selectivity - High purity - Simple - scalable	- Expensive - Low productivity	Peng et al. (2014b)
Nickel Affinity chromatography + Crossflow filtration	Recombinant human collagen. Recombinant bacterial CLP (*Streptococcus pyogenes*).	- Highly selective - Highly pure	- Costly - Low productivity - Difficult scalability - High cost	Rutschmann et al., (2014), Peng et al., (2014a), Xu et al. (2010), Peng et al. (2015), Peng et al. (2012)
Proteolysis (pepsin digestion)	Marine collagen - Sea cucumber (*Bohadschia bivitatta*)	- High yield compared to acid precipitation	- Porcine pepsin was used - Costly	Siddiqui et al. (2013)
Membrane Ultrafiltration & Chromatography	Recombinant collagen	- High yield and purity	- Membrane fouling - Complex process (in four steps)	Zhang et al. (2009)
Pepsin treatment + Crossflow filtration + Sephacryl S200	Recombinant bacterial CLP (*Streptococcus pyogenes*)	- Selective - High purity	- Fouling of membranes	Peng et al. (2018)
Acid precipitation + Pepsin treatment + Crossflow filtration	Recombinant bacterial CLP (*Streptococcus pyogenes*)	- Simple - Cost-effective. - Scalable	- Fouling of membranes	Peng et al. (2014b)
Pepsin extraction + Ultrafiltration (Charged hydrophilic PAN blend membrane)	Marine collagen - Jellyfish	- Eco-friendly - Energy saving - Low maintenance - Reduced fouling of membranes	- Low yield/recovery	Shen et al. (2009)

ATPS (Dextran and PEG system)	Rat tail collagen: Type I	- Simple and benign - Cost-effective - Rapid separation	-not easy to scale up as PEG may corrode equipment	Singh and Tavana (2018)
ATPS (dipotassium hydrogen phosphate and PEG 2000 system)	Recombinant RPCLP	Simple Rapid Cost-effective Higher protein recovery		Awang et al. (2023)

Adapted with modifications from Awang et al. (2020).

2.2.12.1 Chromatography

The principle in chromatography is a separation method that involves two phases, namely the mobile phase and the stationary phase, wherein the mixture is subjected to flow by the mobile phase through the stable surface of the stationary phase (Bryntesson, 2011). Numerous types of chromatography employ this principle, including fast protein liquid chromatography (FPLC), high-pressure liquid chromatography (HPLC), ion-exchange chromatography (IEXC), and affinity chromatography. Affinity chromatography and aqueous two-phase separation are briefly discussed in the subsequent sections as they are relevant to collagen purification and/or the research methodology employed in the next chapter.

Affinity chromatography is a separation technique based on specific interactions between a ligand and the target protein, wherein the target protein binds to the ligand and unbound proteins in the mixture are washed out of the column. Then, the elution process is initiated by modifying the buffer conditions, thereby allowing the target protein to exit the column and to be collected. Affinity chromatography offers advantages such as high resolution and high selectivity. However, the complexity of the

process, high costs, and maintenance are viewed as significant disadvantages, especially for large-scale purifications.

The ligand used for affinity chromatography may be metal, antibodies, enzymes, or other proteins, bound to a resin (Awang et al., 2020). Research conducted by Warner et al. (2007), used nickel columns for affinity chromatography to purify recombinant collagen as a high yield was obtainable compared to other types of metal ions, such as zinc, cobalt, and copper. Consequently, Peng and colleagues (2014) used metal affinity chromatography in the purification of his-tagged recombinant Scl2 collagen. Peng et al. (2012) utilized 20 mM sodium phosphate, 300 mM sodium chloride, and 30 mM imidazole (500 mM in the case of Elution Buffer), pH 8.5, as a binding buffer and a nickel-charged HyperCel-Sepharose metal ion affinity resin.

Although histidine is relatively rare, approximately occurring as 2% of all protein residues, certain native host proteins contain sequences of two or more histidine residues, enabling them to bind to the metal affinity matrix and co-elute with His-tagged recombinant proteins. This nonspecific binding of untagged proteins results in contamination of the final product and this issue tends to be more pronounced in host systems other than *E. coli* which naturally have notably more proteins that have consecutive histidine sequences. *E. coli* tends to produce fewer contaminating His-rich host proteins, making it a suitable host in this context.

Another source of contamination is the formation of disulfide bonds between the protein of interest and other proteins. To address this, using 10 mM 2-mercaptoethanol in the wash and elution buffers typically resolves the problem. Additionally, nonspecific hydrophobic interactions may also cause coelution of unwanted proteins. However, these interactions can be minimized via the addition of low levels of nonionic detergents (e.g., 1% Triton X-100 or Tween 20) in the buffers. Incorporating salt (up to 500 mM NaCl), glycerol (up to 20%), or ethanol (up to 20%) can also help reduce these hydrophobic interactions. The optimum levels of these buffer

components should be determined experimentally based on the specific protein of interest (Rodriguez et al., 2020).

While the use of His_6 tag and affinity chromatography is efficient for laboratory purifications, it is unsuitable in the large scale. Hence, alternative techniques such as aqueous two-phase partitioning and magnetic beads may be sought (Mahmoodi et al., 2019).

2.2.12.2 Aqueous Two-Phase Partitioning

Aqueous two-phase partitioning is a technique used to separate, and to a certain extent, purify whole cells or cell debris from soluble protein. Aqueous two-phase systems (ATPS) are formed when two immiscible polymers, or a combination of a polymer and a salt, or alternatively two salts, are mixed at specific concentrations or a particular temperature (Grilo et al., 2014). Upon the addition of soluble materials, such as a protein, this system is distributed into two phases. Polymer and salt systems are often used in protein purification as they provide a larger contrast in density, lower viscosity, and can offer greater selectivity. Polyethylene glycol (PEG) of various molecular weights is extensively used as a polymer in ATPS, as it is inexpensive. Phosphates and sulfates are commonly used salts in ATPS (Raja et al., 2012). Several factors can affect the partitioning of biomolecules in ATPS, including the molecular weight of the polymer, the concentration of the polymer(s) and/or salt(s), pH value, and temperature (Mahmoodi et al., 2019).

In the purification of collagen, a polymer-polymer aqueous two-phase partitioning method was used to recover a high yield of rat tail collagen (Singh & Tavana, 2018). ATPS is a useful method to separate cells, organelles, proteins, and other biomolecules using mild conditions. Furthermore, it is a simple, cost-effective method compared to other purification methods (Asenjo & Andrews, 2012). The merits of using ATPS include simplicity, efficiency, continuous operation, ease of scale-up,

environmental friendliness, limited denaturation, and low cost (Iqbal et al., 2016). However, to carry out ATPS for a certain protein, using a particular system, the phase diagram must be developed. The ATPS phase diagram, which consists of a binodal curve, provides the concentrations of the components that is required to form a two-phase system that is in equilibrium. The phase diagram for an ATPS for the simple purification of recombinant RPCLP was developed by Awang et al. (2023).

2.3 THE FERMENTATION PROCESS, KINETICS, AND SCALE-UP

2.3.1 Fermentation Media

It is a common practice to see complex media such as lysogeny broth (LB), Tryptone Yeast (TY), and ZYB in cultures on a small scale. Some researchers may opt for semi-defined media instead, such as SOB (Super Optimal Broth), SOC (Super Optimal broth with Catabolite repression), Terrific broth (TB), Super broth (SB), and autoinduction medium, to enhance the protein expression (refer to Table 2.11). In semi-defined media, glucose or glycerol are often used as carbon sources, and complex nitrogen sources such as yeast extract or tryptone are utilized. Along with these, divalent cation salts and phosphate salts are used to enhance cell growth and achieve the desired protein yield (Tripathi, 2016). Recombinant *E. coli* has been cultured in a variety of media such as LB, TB, SB, and M9 minimal medium, and high cell densities were achieved (Tripathi et al., 2009).

Table 2.11 Medium selection for recombinant protein production in *E. coli*

Recombinant protein	Host strain	Medium	Carbon Source	Nitrogen Source	Reference
Recombinant CLP	BL21	2xYT medium and Defined medium	Glucose	-	Peng et al. (2012)
Human-like collagen II	BL21	M9	Glucose	Yeast extract	Luo and Mu (2014)
Reteplase	BL21(DE3)	LB	Catabolizable amino acids	Amino acids	Zare et al. (2019)

Nitrilase	BL21(DE3)	Super Optimal Broth	Glycerol	Tryptone, yeast extract,	Sohoni et al. (2015)
Bromelain	BL21-AI	Autoinduction medium	Glucose, glycerol, lactose	Tryptone, yeast extract, ammonium chloride	Jamaluddin et al. (2014)
Xylanase	DH5α	Defined Medium	Glucose	Ammonium sulfate	Basar et al. (2010)
BoNT/A-Hc botulinum neurotoxin	BL21(DE3)	M9	Glucose	Ammonium sulfate	Yari et al. (2010)
Perdeuterated protein	BL21(DE3)	Modified M9	Glucose	Yeast extract Ammonium chloride	Cai et al. (2016)
Membrane protein	BL21(DE3)	M9-cas	Glycerol	Casamino acids and ammonium chloride	Deacon et al. (2008)

M9 medium, a minimal growth medium, is often used in bacterial cultures and protein expression. Made up of a base formulation that contains salts (M9 salts) and can be supplemented with various amino acids, vitamins, and carbon sources, it is considered inexpensive and has a very low autofluorescence (when excited at 488nm) and very low absorbance. M9-cas medium, a variant of M9 medium containing casamino acids, is also used for the cultivation of *E.coli* (Elbing & Brent, 2019; Pinkas et al., 2011; Xu et al., 2010). The addition of casamino acids is thought to reduce proteolysis (Gellermann et al., 2019) and enhance plasmid yield as their nitrogen is utilized for the synthesis of purine and pyrimidine bases (Dorward et al., 2019). Xu et al. (2010) used M9-cas medium to cultivate the *E. coli* Bl21 strain to produce CLPs from various species of bacteria, including *S. pyogenes*, *R. palustris,* and *M. sp. 4-46* (Mohs et al., 2007; Yoshizumi et al., 2009; Yu et al., 2014). Furthermore, Soroodi (unpublished) has reported on the screening of M9-cas Minimal Medium components to improve the RPCLP expression by *E. coli* BL21(DE3)/pColdII system, induced with IPTG alongside cold-shock induction. Even though cold-shock activation of the cspA gene in the pColdII vector suffices to induce the translation of the gene, it is recommended by the manufacturer (Takara Bio) that IPTG (0.1 up to 1.0 mM) is added to enhance the expression.

According to Soroodi, the concentrations of three of the medium components, namely D-glucose, casamino acids, and disodium hydrogen phosphate, were found to be significant factors. Casamino acids are hydrolysates or products of acid hydrolysis of casein. It has all the essential amino acids except tryptophan. Unlike tryptone, which has some oligopeptides due to incomplete hydrolysis, casamino acids are chiefly free amino acids. It is typically used in microbial growth media as it provides a completely hydrolyzed protein nitrogen source. The relationship of D-glucose and casamino acids concentrations to biomass was directly proportional, whereas the disodium hydrogen phosphate concentration was inversely proportional to cell growth. The lowest concentration of disodium hydrogen phosphate studied was 0.48 M. Thus, these factors must be optimized for optimal cell growth and protein expression (Soroodi, Unpublished). Tables 2.12-2.13 show some of the components and their concentrations as reported in the literature.

Table 2.12 M9 Minimal medium composition in the literature

No	Species/ Strain	Rec. Protein	Medium components	Author
1	*Escherichia coli*	-	M9 medium, 5×: 30 g Na_2HPO_4 15 g KH_2PO_4 5 g NH_4Cl 2.5 g NaCl 15 mg $CaCl_2$ (optional) 2.5 mg $FeSO_4 \cdot 7H_2O$ per liter (optional) Concentrated media were then diluted to 1× with sterile water and supplemented with the following sterile solutions, per liter. 1 mL 1 M $MgSO_4 \cdot 7H_2O$ 10 mL 20% carbon source (sugar or glycerol) Optional: 0.1 mL 0.5% vitamin B1 (thiamine) 5 mL 20% Casamino Acids *or* L amino acids to 40 µg/mL *or* DL amino acids to 80 µg/mL	Elbing and Brent (2019)

2	*Bacillus subtilis strain* SP6	uricase	KH_2PO_4 NH_4NO_3 $MgSO_4.7H_2O$ NaCl $FeSO_4.7H_2O$	Pustake et al. (2019)
3	*E. coli BL21*(DE3)	Glycosyltransferase	M9 salt $Na_2HPO_4.2H_2O$ KH_2PO_4 NaCl NH_4Cl Biotin Thiamine Glucose $MgSO_4$ $CaCl_2$ Trace elements: EDTA $FeCl_3.6H_2O$ $ZnCl_2$ $CuCl_2.2H_2O$ $CoCl_2.6H_2O$ H_3BO_3 $MnCl_2.4H_2O$	(Rudiger & Schwab, 2019)
4	*Escherichia coli*	Human-like collagen II	Minimal Medium: $Na_2HPO_4.12H_2O$ $(NH_4)_2SO_4$ KH_2PO_4 NaCl $MgSO_4·7H_2O$ $CaCl_2$ Thiamine Glucose Trace elements	Luo and Mu (2014)

5	perdeuterated proteins	KH_2PO_4 K_2HPO_4 Na_2HPO_4 K_2SO_4 NH_4Cl Trace elements: $FeSO_4.7H_2O$ $CaCl_2.2H_2O$ $MnCl_2.4H_2O$ $CoCl_2.6H_2O$ $ZnSO_4.7H_2O$ $CuCl_2.2H_2O$ H_3BO_3 $(NH_4)6Mo_7O_{24}$.$4H_2O$ EDTA $MgCl_2$ Thiamine Glucose Yeast extract	Cai et al. (2016)
6	-	$Na_2HPO_4.7H_2O$ KH_2PO_4 NaCl $MgSO_4$ $CaCl_2$ Metal 0.25x Vitamins 0.25x NH_4Cl 0.1% Glucose 1%	Sivashanmugam et al. (2009)

Table 2.13 Glucose, casamino acid, and disodium hydrogen phosphate concentrations used in media

Glucose concentration	Casamino acids (Casein hydrolysate)	Disodium hydrogen phosphate	Reference
1.1 g/L	1.0 g/L	10.1 g/L	Wong et al. (2003)
10 g/L	10 g/L	-	Dorward et al. (2019)
10.0 g/L	10.0 g/L	-	Pérez et al. (2018)
10 g/L	-	-	Valiollah Babaeipour et al. (2017)
6 g/L	1 g/L	37 mM	Trausch and Batey (2015)
15.3 g/L	-	4.25 g/L	Luo and Mu (2014)

2.3.2 Fermentation Process

2.3.2.1 Shake Flask Culture

Shake flask culture represents a crucial small-scale step in the process toward large-scale fermentations. Tasks such as media screening, medium optimization, the screening for a suitable strain and its development, understanding the metabolic pathways, and investigations of the basic growth parameters and growth kinetics are often initiated in shake flasks (50 mL up to 1 L batch cultures) to conserve resources and costs, and also due to their simplicity of handling. Elbing and Brent (2019) recommend the use of an Erlenmeyer flask whose volume is at least five times the volume of the culture and to grow the culture at 37 °C with vigorous agitation (at least 250 rpm) for proper aeration. Monitoring cell numbers is typically accomplished by OD_{600} measurement. Elbing and Brent (2019) stated that a good rule of thumb is that each 0.1 OD unit is approximately 1 x 10^8 cells/mL. Aside from the aforementioned

advantages, shake flasks are useful for the ability to conduct multiple simultaneous tests, such as screening substrates or microorganisms. However, their simplicity of handling also means the inability to continuously measure and control certain fermentation parameters, such as forced aeration and pH control.

2.3.2.2 Stirred Tank Reactors (STR)

Fermentation takes place inside a bioreactor, and there are a variety of options to attain this purpose optimally. Bioreactors provide an environment of efficient means of mixing, mass, and heat transfer between the different phases. Stirred tank and bubble column bioreactors are extensively used in a variety of bioprocesses such as aerobic fermentation. In bubble columns and airlifts, the low shear stress creates suitable conditions for the growth of shear-sensitive cells (Garcia-Ochoa & Gomez, 2009). Stirred tank reactors (STR), on the other hand, are frequently used for the culture of diverse types of bacteria. As the name suggests, the stirred tank reactors are in their simplest form cylindrical vessels with mechanical stirrers that are also known as impellers. The STR provides efficient heat and mass transfer rates and allows for excellent mixing for homogeneity of properties throughout the medium. Homogeneity and bubble dispersion are achieved by powered mechanical agitation. Stirred tank reactors have garnered significant attention in terms of scale-up due to their relatively simple design, unlike other non-mechanically aerated vessels such as bubble columns, packed and fluidized beds, and airlift reactors (Doran, 2012). There are a variety of impellers for different bioprocessing needs. Mixing and transfer are mainly governed by factors such as agitation speed, the number of impellers, impeller type, and the gas flow rate (Garcia-Ochoa & Gomez, 2009). Due to advantages such as its versatility, reliability, and low operation costs, the STR is the leading bioreactor for industrial use (Spier et al., 2012).

2.3.2.3 Bioreactor Modes of Operation

There are three main operational modes used for submerged and liquid fermentation, namely batch or discontinuous, fed-batch or semi-continuous, and continuous. Briefly, batch mode is a partially closed-system operation when no feed and no removal of any liquid is conducted (except for gas exchange, anti-foaming agents, and pH control agents) once the process is set up, until the end of the process (Spier et al., 2012). Figure 2.6 is a schematic of the relationship between cell concentration, the limiting substrate, and dissolved oxygen in a batch process. Fed-batch is a mode where there is continued nutrient feed input into the system, but no removal of medium containing the product(s). Finally, the continuous mode is where both input and removal occur. Table 2.14 summarizes the pros and cons of each mode of operation. All modes of operation have been well-established and well-studied in *E. coli* cultivation, and data from the literature can, therefore, be used as benchmarks for this study.

In this study, production was studied in batch mode, as it is critical to study and establish the early process of fermentation prior to the fed-batch or continuous stage. After gaining insight into the fermentation kinetics, fed-batch may then be used to optimize production and avoid the utilization of substrates that inhibit biomass growth at high concentrations and other issues (Spier et al., 2012).

Table 2.14 Advantages and disadvantages of the three bioreactor modes

Mode	Advantages	Disadvantages
Batch	Simple and easy to operate. Good control of contamination Good biomass accumulation and high biomass yield on glucose Good for strain and culture condition optimization	Closed system, therefore, nutrients are depleted, and byproduct builds up. Long downtime between runs
Fed-batch	Simple and easy to operate and control. High yield of biomass and protein Limited byproduct accumulation Cost-effective	Sterilization is personnel intensive. It is not easy to maintain growth conditions. Variability among batches. Inherent nonlinear dynamic nature
Continuous	Optimized process conditions Higher volumetric productivity Reduced labor costs and reduced vessel costs after reaching a steady state. Reduced vessel downtime for cleaning, filling, sanitization Easy operation and process control at a steady state	Susceptible to long-term bacteriological issues, which require the addition of costly antibiotics, which may result in mutations. More complex downstream processing Difficult to reach a steady state with certain strains

Source: Asenjo & Merchuk, 1994.

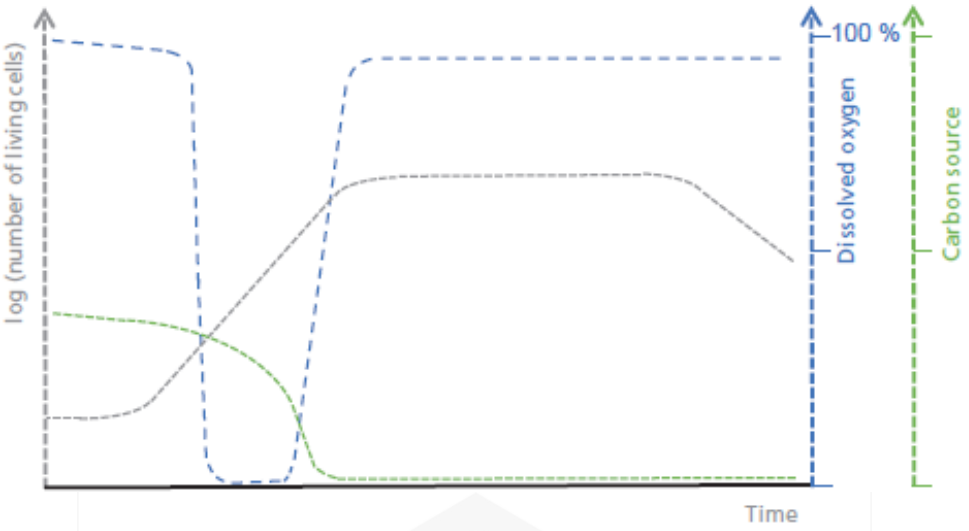


Figure 2.6 Schematic diagram of the relationship between cell concentration, the limiting substrate, and dissolved oxygen in a batch process (INFORS-HT)

2.3.3 *E. coli* as a Recombinant Host

Numerous factors affect the production of fusion proteins in recombinant bacteria. Among these factors, the gene sequence chosen, vector design and selection, choice of host and strain, and fermentation medium logically come to mind. For various reasons elucidated in this section, *E. coli* comes up as a lucrative host.

Escherichia coli is a bacterium in the family Enterobacteriaceae, and this bacterium represents about 1% of the total bacteria in an adult's intestine. It is a rod-shaped, gram-negative bacterium that is a facultative anaerobe and requires oxygen to grow. *E. coli* is a model organism that is well-characterized in biotechnology and has been used as a host for recombinant expression of many products, such as asparaginase, human insulin, human growth hormone, human parathyroid hormone, tissue plasminogen activators, interferons, interleukins, and tumor necrosis factors (Fidanova

et al., 2012; Graumann & Premstaller, 2006; Langlais & Korn, 2006). Simple, well-studied bacteria, like *E. coli,* are genetically engineered to produce these proteins that can be easily harvested in substantial quantities for medical use. For example, an *E. coli* BL21 had been developed for expressing human-like collagen (Fan et al., 2005; Luo et al., 2008).

E. coli grows easily and rapidly in both minimal and rich media, and on both solid and liquid media, with glucose or glycerol as carbon sources, making it an ideal host. Also, it can consume either organic nitrogen (N), from sources such as peptones, or inorganic (N)-salts. In the literature, *E. coli* is frequently grown in lysogeny broth (LB) (Huang et al., 2007; Kitko et al., 2010). Other media used with *E. coli* include M9, Terrific Broth (TB), Super Broth (SB), and NZ amine-Yeast broth (ZYB).

Some of the advantages of using *E. coli* are that it has fast growth (doubling time ~ 20 minutes under optimal conditions), high cell densities are easily attained, and its media are inexpensive (Rosano & Ceccarelli, 2014). Furthermore, there is a wide choice of cloning vectors, and it is easy to control gene expression. *E. coli* has a circular genome of about 4500 kb in length. *E. coli* gives good protein secretion, and larger yields are possible. *E. coli* is usually cultivated in suspension in the stirred tank rate (STR), bubble columns, and airlift reactors. There are also numerous publications citing the successful use of recombinant *E. coli* hosts harboring vectors, such as pCold II, which rely on a thermal shift to a lower temperature (to 15 °C or 25 °C, for example). The pCold series expression vectors have a pUC118 backbone (a pUC18 derivative) with the cspA promoter (Rosano & Ceccarelli, 2014). This is very useful in inducing the production of temperature-sensitive recombinant protein products such as enzymes. The expression of proteins in *E. coli*, when carried out at low temperatures, can enhance their solubility and stability by supporting proper folding (Bartolo-Aguilar et al., 2022).

Although *E. coli* fermentation has many advantages for recombinant protein production, it has its disadvantages as a host expression system. This is mainly the issue

of its inability to perform post-translational modifications such as glycosylation, myristylation, phosphorylation, signal sequence cleavage, and extensive disulfide bond formation, as a prokaryote. Also, as a bacterium, it may have a high endotoxin content, and protein aggregation (misfolded proteins formed in inclusion bodies) may result instead of secreting soluble extracellular proteins (Rosano & Ceccarelli, 2014). In spite of these challenges, *E. coli* remains a versatile host for the production of countless recombinant proteins.

Common *E. coli* strains that are used in the laboratory belong to the K strain and the B strain. BL21(DE3) and its derivatives are extensively utilized for protein expression. The *E. coli* BL21 strain is deficient in the Lon protease and outer membrane protease (OmpT) genes, and this minimizes the target protein degradation (Rosano & Ceccarelli, 2014). This strain also has the hsdSB mutation, which protects from plasmid loss (Rosano & Ceccarelli, 2014). Some of the advantages of using this strain are that it grows faster, the protease deficiency prolongs the protein lifetime, secretes more protein when compared to other laboratory strains, and produces less acetate when glucose (threshold value for *Escherichia coli* at 30 mg/L) is used as the carbon source (Rosano et al., 2019).

E. coli BL21(DE3) is a genetically engineered strain of *E. coli* that is widely used in molecular biology and biotechnology for the production of recombinant proteins. BL21 is known for its compatibility with the T7 promoter system, which allows for controlled and high-level expression of recombinant proteins. It is often used for protein production in molecular biology and biotechnology. DE3 refers to the λ(DE3) prophage that has been integrated into the BL21 genome. The λ(DE3) prophage contains the gene for the T7 RNA polymerase, under the control of the lacUV5 promoter, which is essential for the T7 promoter-driven expression system. This genetic modification is crucial for recombinant protein expression in *E. coli* BL21(DE3). When the lacUV5 promoter is induced by the presence of isopropyl β-D-1-thiogalactopyranoside (IPTG), it leads to the expression of the T7 RNA polymerase, which, in turn, drives the expression of the target protein cloned under the T7 promoter.

The DE3 lysogen allows researchers to tightly regulate the expression of their target genes by inducing expression with an inducer, such as IPTG, and, along with fast growth and high expression levels, makes BL21(DE3) a popular choice for producing recombinant proteins. The T7 promoter and T7 RNA polymerase result in high levels of protein expression.

Table 2.15 presents typical variables in a recombinant CLP production and compares protein concentrations. It shows how the protein concentration improved with the use of a stirred tank bioreactor (Peng et al., 2012). However, it's essential to optimize growth conditions, induction parameters, and other factors to achieve the best results for a particular protein of interest.

Table 2.15 Shake flask and bioreactor culture conditions and protein concentrations

Variable	**Shake flask culture (2 L)**	**Bioreactor culture (2 L)**
Mode	▪ Batch	▪ Batch
Biomass concentration	▪ Initial OD_{600} = 0.3 ▪ Final OD_{600} = 5.6-6.4 ▪ Wet cell paste (CWW) = 8.3-9.7	▪ Initial OD_{600} = 0.25 ▪ Final OD_{600} = 4.3-5.5 ▪ Wet cell paste (CWW) = 5.3-5.8 g/L
Culture volume	▪ 0.5 L	▪ 1.6 L working volume
Scl2 protein concentrations	▪ 0.2-0.3 g/L	▪ 0.7-1.0 g/L
Culture conditions	▪ Inoculum size: 0.5 mL / 500mL ▪ Overnight: 16 hours at 200 rpm, 37 °C ▪ Agitation: 200 rpm ▪ 24 h at 37 °C ▪ Induction with IPTG (0.5 mM) (OD_{600}= 6) ▪ 10 h at 25 °C ▪ 14 h at 15 °C	▪ Inoculum size: NA (see initial OD_{600}). ▪ Foam control: 10% (v/v) polypropylene glycol as antifoam (automated). ▪ pH 7.0 ▪ 2xYT medium, glucose (10 g/L), ampicillin (200 µg /mL). ▪ 20% DO saturation setpoint, supplemented with 5% pure oxygen, 2-step cascade for oxygen control ▪ Agitation speed: 500 to 1200 rpm, Airflow rate: 0.3 L/min to 1.5 L/min, 5 hours at 37 °C. ▪ Induction: IPTG (1 mM) (OD_{600} = 2-3), 7 hours at 25 °C.

Source: Peng et al., 2012.

2.3.4 Fermentation Parameters Affecting Biomass and Product

Many physical and chemical factors influence the growth of *E. coli* and the expression of a certain recombinant protein. For example, a very low oxygen uptake rate can lead to dormancy of *E. coli* cells. Conversely, when it is too high, acetate, which inhibits biomass, is produced as a byproduct (De Mey et al., 2007). Other examples include temperature, pH, agitation, and incubation time. Manipulating growth parameters, such as pH, and media contents can affect the solubility of recombinant proteins.

2.3.4.1 Inoculum Age and Size

Inoculum development in terms of its age and size are important variables for recombinant protein production from *E. coli* as inoculum age and size can affect the length of the lag phase. If the inoculum size is small, the production cycle is extended. On the other hand, if the inoculum size is too large, microorganisms would enter the exponential phase quickly, and the production phase proceeds simultaneously. Excessive inoculation is undesirable and typically, an inoculum size of 1-5 % is used for bacteria. However, the cell density as well as the inoculum age must also be considered. Inoculum age is the age (in hours) of the cells in the seed culture at inoculation. A young inoculum age will extend the fermentation period due to slow growth. Conversely, an aged inoculum will affect the recombinant strain stability and lead to a decline in the production of the protein. Inoculating at a late logarithmic growth phase is generally favorable. However, many studies reported on inoculation at the mid-log phase with great success. The incubation time is another parameter that when optimized can improve biomass yield and decrease costs. Different researchers have obtained maximum yield production in different periods depending upon the microorganism and culture conditions. A long incubation time, for instance, causes the nutrients in the medium to decrease thereby reducing growth (Yanti, 2018).

2.3.4.2 Temperature

Temperature significantly influences fermentation kinetics, chiefly through its effect on reaction rates as described by the Arrhenius equation. Enzymes within the cells require a certain temperature range to function, beyond which the proteins are denatured and below which the proteins are inactive. Also, temperature affects the structure of cell constituents and membrane components. Generally, *E. coli* grows in a temperature range between 23 °C to 40 °C. The optimal temperature for cell growth is 37 °C. However, this may not be the most suitable temperature for an optimal rate of protein

production. Hence, the post-induction temperature must also be considered, while taking into account the method of induction.

Hajihassan et al. (2018) found that in addition to dissolved oxygen (DO) concentration and post-induction time, post-induction temperature also significantly affected the production of recombinant human-nerve growth factor (NGF). For the specific case of recombinant bacterial collagen-like proteins, the post-induction temperatures chosen by Peng and colleagues (2012, 2014b) were a dual strategy of 25 for 10 hours flask culture for ten hours followed by 15 °C for 14 to 15 hours for flask culture. For the STR, seven hours at 25 °C were followed by 10 hours at 15 °C. The reason for these low induction temperatures was the pCold choice of vector which activated protein expression at low temperatures.

2.3.4.3 Inducer Concentration

Another factor that needs to be optimized is the inducer concentration. In this work, IPTG serves as the inducer, and therefore similar fermentations were studied to determine a range for the IPTG screening levels and ultimately optimize the IPTG concentration to maximize the desired protein yield. The IPTG concentration range typically used in the induction ranges from 0.05 mM up to 2.0 mM beyond which IPTG is considered toxic to cells (Bahreini et al., 2014). IPTG can significantly inhibit growth in some cases (Restaino et al., 2018). As such, IPTG concentration is also frequently optimized for each specific fermentation. Peng et al. (2012) studied Scl collagen-like protein expression by using 0.5 mM IPTG but later studies by the research team have used 1 mM IPTG for the induction (Peng et al., 2014b).

2.3.4.4 Induction Point

Aside from the inducer concentration, the induction point is critical to maximize protein yield. This is because recombinant protein production is a metabolic burden on the recombinant bacteria and can result in a reduced growth rate and cell density. If induction occurs too early in the fermentation, before the exponential growth phase, low cell density and low productivity ensue. Jamaluddin (2015) reported that recombinant bromelain productivity in the bioreactor was sensitive to inducer addition and the recombinant bromelain activity was optimal when the induction was conducted during the mid-log phase. Peng et al. (2014b) successfully induced Scl collagen-like protein expression in *E. coli* BL21(DE3) harboring pET vector by induction at the mid-log phase (OD_{600} = 0.6). Xu et al. (2010) successfully expressed several CLPs including RPCLP in *E. coli* BL21(DE3) by induction at the mid-log phase (OD_{600} = 0.9-1.0), which is ideal since the bacteria are rapidly growing at this phase.

2.3.4.5 Agitation Speed

Agitation or mixing is a physical operation that is done to reduce non-uniformities or non-homogeneous regions in fermentation by eliminating gradients of nutrient concentration, oxygen concentration, temperature, and other properties. Low agitation may cause inadequate mixing of nutrients and lead to low cell productivity. On the other hand, high aeration rates may inhibit cell and protein product yields due to high shear stress. Thus, agitation affects the actual pH, temperature, dissolved oxygen (DO) content, and substrate concentration of a fermentation process (Azmi et al., 2015).

2.3.4.6 Dissolved Oxygen

Dissolved oxygen (DO) content is an essential control parameter that may significantly impact the aerobic growth of *E. coli*, as well as the formation of a recombinant product. Limitations in dissolved oxygen concentration have been reported to both increase and decrease the expression of various recombinant proteins (Zare et al., 2019), and as such the monitoring and control of dissolved oxygen are essential during the fermentation. However, the dissolved oxygen level is difficult to maintain due to the low solubility of oxygen in the fermentation media. Furthermore, the actual level of dissolved oxygen depends on factors such as temperature, pressure, and salinity.

The concentration of the dissolved oxygen in the medium depends on the rate of oxygen transfer across the gas-liquid interface, and this is a function of hydrodynamic conditions in the bioreactor. The concentration also depends on the rate at which the oxygen is transferred into the cells, and then its consumption rate for the purposes of growth, maintenance, and production (Garcia-Ochoa & Gomez, 2009). Recombinant *E. coli* requires a higher aeration capacity, as noted in a study on the effects of dissolved oxygen and its mass transfer on the overexpression of a gene (Bhattacharya & Dubey, 1997). Despite this, there were cases where the recombinant product of interest favored low oxygen conditions, possibly due to a lower cell density and lower acetate accumulation (Zare et al., 2019).

The DO level may be enhanced by improving the rate of oxygen transfer through the value of the oxygen mass transfer coefficient (k_La), by means such as increasing sparge rates, agitation speed, or vessel pressure, or by lowering the temperature to reduce oxygen uptake by the cells. The oxygen mass transfer coefficient is further discussed in section 2.7. For *E. coli*, dissolved oxygen is typically maintained above a lower limit of 10% to 30% saturation oxygen concentration value, and this can be achieved by increasing airflow rate and agitation speed, and oxygen enrichment (Valiollah Babaeipour et al., 2017; Junker, 2004). Previous studies have reported on the

significance of oxygen level, aeration, and agitation rate on the cell growth rate in liquid media in bioreactors (Wang et al., 2016; Wang et al., 2016; Zhong, 2010), as well as on the *E. coli* strain and induction conditions (Hajihassan et al., 2018).

2.3.4.7 pH

The pH of a culture affects the expression levels of certain genes, oxygenation and uptake rates, cell growth and structure, and the activities of its various enzymes. Therefore, cellular growth and the synthesis of a certain product such as membrane and periplasmic proteins and metabolic enzymes, as well as the secretion of proteins are all influenced by pH levels (Zare et al., 2019). According to Doran (2012), the maximum growth rate is usually maintained over 1-2 pH units but it often plunges with larger changes. The pH also affects the product formation in anaerobic cultures and can change the maintenance energy needs (Zare et al., 2019). *E. coli* is a neutrophile, favoring pH values of 7.4- 7.8 within its cytoplasm, and a 5-9 environmental pH range (Kitko et al., 2010). The importance of pH, in addition to selection of time of harvest, was highlighted in a study where control of pH and aeration conditions in 1 L fermenter cultures using autoinduction obtained higher yields compared to those with uncontrolled pH and aeration conditions (Deacon et al., 2008).

2.3.4.8 Other Considerations

Finally, it is important to consider and study interactions between several factors in attaining a high protein yield. As the fermentation and production of recombinant proteins involves a complex biological system, some of these factors combine to create

synergistic interactions whilst others form antagonistic effects. The use of statistical tools such as response surface methodology can help identify these interactions which might be overlooked in OFAT studies. Past studies have proven useful in guiding the screening and optimization of fermentation parameters for certain strains, some of these studies have been selected and are presented in Tables 2.16-2.17.

Table 2.16 Parameters for bioreactor fermentation of recombinant *E. coli* for protein production

Product	***E. coli* strain**	**Scale**	**Parameters (* - means optimized at a shake flask scale)**	**Observations**	**Reference**
Lipase	DH5α	5 L STR 3.5 L working volume	Agitation (200-350 rpm) Aeration Inoculum size	Maximum lipase activity at: speed: 300 rpm aeration:1 L/min inoculum size: 1.5 % 1.7 X biomass enhancement 5 X lipase activity enhancement	Kumar et al. (2017)
Xylanase	DH5α	2 L STR 1 L working volume 10% (v/v) inoculum	Dissolved oxygen tension (DOT) 0-40%	The optimal glucose and ammonium sulfate levels were 10 g/L1 and 2 g/L, respectively. The growth of *E. coli* DH5α and the enzyme production was inhibited by limited oxygen, i.e. DOT level was controlled at 0%. Xylanase production was enhanced at 20% DOT level, though growth was not significantly improved. High xylanase production (1784.57 U/mL) was obtained.	Basar et al. (2010)
	BL21 (DE3)	Shake flask	IPTG (0.001-1 mM) Temperature (16-37 ℃) Arabinose supplementation	0.002 mM 20 ℃	Chang et al. (2017)
Nitrilase	JM109	7 L STR 5 L working volume 10% (v/v) inoculum	pH (6-8) Agitation (200-500 rpm) (Impeller tip speed: 0.8-1.99 m/s) Aeration (0.2-0.66 vvm)	Initial pH 7 (without control) 400 rpm 0.4 vvm	Jain et al. (2012)
Lysostaphin	TOP10F8	5 L STR 3 L working volume 10% (v/v)	Agitation (400-800 rpm) pH (5-7) Temperature (25-37 ℃)	400 rpm pH 6 37℃ 2X enzyme production efficiency was attained	Szweda et al. (2014),Luo and Mu (2014)
Reteplase	BL21(DE3)	2 L STR	Temperature (23-37 ℃) Agitation (200-400 rpm) pH (6.5-8.5)	32 ℃ 210 rpm pH 8.4	Zare et al. (2019)
Human lysozyme	SHuffle T7 Express	3 L	Induction temperature (16, 22, 30, 37 ℃)	pH 6.5 Temperature 37 ℃ 0.5 mM IPTG	Lamppa et al. (2013)

				20-hour expression	
Human LOX-1	BL21(DE3) Origami Rosetta-gami 2 Rosetta-gami 2 pLysS	5 L	Temperature (18, 25, 30, 37 ℃) Inducer Concentration (0.25, 0.5, 1, 2mM)		Hu et al. (2017)
rPDT	Rosetta-gami 2 LB	5 L 3 L pH 7.2 37℃	IPTG (25, 12.5, 200 µM) Temperature (20, 30, 40 ℃) DO (10, 30, and 50 %)	IPTG concentration : 12.5 µM Temperature: 20℃ DO: 30%	Koopaei et al. (2018)
Aspartase	BL21	1 L working 10 % (v/v)	Agitation (130-270 rpm) Aeration (03-1.7 vvm) Induction time (3-7 h) 37 ℃ pH 7	250 rpm 1.25 vvm 6 h 7 X increase	Singh and Yadav (2013)
Japanese encephalitis virus protein	BL21(DE3)	100 L 85 L	1 mM IPTG 0.25-0.6 vvm 100-250 rpm pH 7 37 ℃ DO 30 %	Yield is ~124mg of highly purified and biologically active EDIII protein per 100 g of biomass.	Tripathi et al. (2012)
Human Growth Hormone	BL21	5 L & 30 L	Oxygen supply modes: Air vs. high-purity oxygen	Productivity using high-purity oxygen: 2 X in 5 L 4 X in 30 L	Shang et al. (2009)
1,3-β-glucanase	BL21	3 L 37℃	Agitation (30, 60, 90, 120, 180, 240 rpm) Aeration (0.18, 0.36, 0.71, 1.43, 2.14 vvm)	135 rpm 0.8 vvm	Zaslona et al. (2015)

2.4 DESIGN OF THE EXPERIMENTS

The design of experiments is an essential component of this study, and hence considerable effort was invesed in selecting suitable designs for each stage of the research methodology. In this research, the optimization of the process involved one-factor-at-a-time (OFAT) optimization, statistical multivariate optimization using Design of Experiment software, as well as knowledge-based selections from the literature. OFAT was used to optimize M9-cas medium components in the shake flask culture, which have been selected from screening experiments in the literature (Soroodi, unpublished). Statistical multivariate optimization, namely Taguchi and Face-centered central composite designs, was used to design screening and optimization experiments for bioreactor fermentation.

2.4.1 One Factor at a Time (OFAT) Optimization

OFAT refers to optimization that is conducted by varying one factor while keeping other factors fixed. Its advantages rely on the simplicity of the application as well as the ability to observe and analyze the effects of the modified variable directly, without the need for complex analysis. However, this means that the interactions between each factor and the effect of this interaction are not clearly observed or taken into consideration in the optimization. OFAT is less favored in studies of complex systems due to its inability to detect interactions. However, OFAT is often used to obtain preliminary or estimated values for the optimization to obtain a suitable range of values for a more complex and thorough optimization.

2.4.2 Response Surface Methodology (RSM)

Design of Experiment (DOE) is an important aspect of process development where the aim is to select points wherein the response is evaluated. In any process, several independent input variables have the potential to affect the process's performance or yield, measured using an output variable or termed the response. Experimental runs are thus designed to study and estimate the relationship of the response as a function of these various input variables. Often, the relationship may be explained by a linear or quadratic regression model equation, though it will not fit exactly. Then, the lack of fit or the inadequacy of the model must be measured to evaluate the model's effectiveness.

One of the tools under the umbrella of DOE is Response surface methodology (RSM) which refers to a collection of statistical tools and approaches that are used to develop, enhance, and optimize industrial processes. RSM provides a visual presentation of the response surface either as a 3D surface plot or several contour plots where the response is plotted against only one variable, keeping all other variables fixed. Each contour would correspond to a specific height of the response surface (Myers et al., 2018).

Commonly used response surface designs of experiments include factorial design, fractional factorial design, Plackett-Burman Design (PBD), Box-Behnken Design and Central Composite Design (CCD). One of the advantages of using RSM is the reduction in the number of experimental runs that must be conducted to obtain sufficient data. Various design tools are currently available for the construction of response surface experiments for both screening and optimization purposes. DOE software, such as Design-Expert and Minitab, also facilitate detailed statistical data such as analysis of variance (ANOVA), model equations, and plots showing the interactions between factors.

Table 2.17 shows some examples of screening and optimization studies for recombinant products using *E. coli* fermentation. In some cases, response surface methodology (RSM) was employed in the optimization of the fermentation parameters at the laboratory scale. As shown in Table 2.17, RSM can be implemented to develop a statistical model for the individual and interactional effects of fermentation parameters including medium composition, pH, cultivation time, or temperature. Based on a statistical model, the optimization of a few parameters at once using RSM rather than "one at a time" provides a practical approach to the fermentation process that is composed of many variables that may interact with each other and these interactions need to be determined for an accurate model.

Table 2.17 Design of screening and optimization studies for recombinant products using *E. coli*

***E. coli* strain**	**Medium**	**Parameters Studied**	**Design**	**Runs**	**Observations**	**Reference**
BL21(DE3)	Modified M9	Time after induction (h) OD_{600} of induction $(NH_4)_2SO_4$ (g/L) Temperature (℃) Glucose (g/L) $MgSO_4.7H_2O$ (g/L) K_2HPO_4/KH_2PO_4 ratio Trace-elements solution (mL) IPTG concentration (mM)	Taguchi Screening	16 runs	Significant factors: Time after induction (h) $MgSO_4.7H_2O$ (g/l) K_2HPO4/KH_2PO_4 ratio IPTG concentration (mM)	Yari et al. (2010)
		Time after induction [11-15 h] $MgSO_4.7H_2O$ [2-4 g/L] K_2HPO_4/KH_2PO_4 ratio [1-2] IPTG concentration [0.3-0.7 mM)	Taguchi Optimization (Qualitek-4 software)	9 runs	Optimal levels: Time after induction = 15 g/L $MgSO_4.7H_2O$ = 4 g/L a K_2HPO_4/KH_2PO_4 ratio of 2 0.7 mM IPTG Concentration: 52.1 mg/L BoNT/A-Hc	
BL21(DE3)	LB	Incubation temperature Peptone and yeast extract concentration, Cell density (OD_{600}) before induction Inducer (IPTG) concentration	Fractional Factorial Design Screening	8 runs	Significant factors: IPTG concentration Ca^{2+} ion concentration Induction time	Beigi et al. (2012)

		Induction time Ca^{2+} ion concentrations IPTG concentration Ca^{2+} ion concentration Induction time	Optimization using full factorial CCD	19 runs	Optimal levels: 0.0089 mM IPTG 1.3 mM Ca^{2+} ion concentration OD_{600} = 2 Maximum SVP activity (522 U/mL) was achieved in 0.5% peptone and yeast extract for 24 h at 30℃.	
BL21(DE3)	NA	Medium (LB, LBM, TB, TBlac, TBmod, TBaim, ZYB, ZYBbuff, SOC, SB, SBmod, SBaim, SBenrich, NBS, NBSmod, M9, Ries) Medium composition Initial medium pH (5.5-9.5, at 0.5 increments) Incubation temperature (25, 30, 37 or 42 °C) Culture volume to flask volume ratio (3:1, 5:1, 10:1, and 20:1) Agitation speed (150, 200, 250 rpm) IPTG concentration (0.1, 0.5, 1 and 3 mM) Elapsed fermentation time at induction (2, 4, 6, 9, and 16h) Induction period	OFAT	NA	medium (LB, LBM, TB, TBlac, TBmod, TBaim, ZYB, ZYBbuff, SOC, SB, SBmod, SBaim, SBenrich, NBS, NBSmod, M9, Ries) medium composition initial medium pH (5.5-9.5, at 0.5 increments) incubation temperature (25, 30, 37 or 42 °C) culture volume to flask volume ratio (3:1, 5:1, 10:1, and 20:1) agitation speed (150, 200, 250 rpm) IPTG concentration (0.1, 0.5, 1 and 3 mM) Elapsed fermentation time at induction (2, 4, 6, 9, and 16h) Induction period	Collins et al. (2013)

BL21(DE3)	Basal medium	OD_{600} (0.3-0.9) IPTG (0.1-1.0 mM) Temperature (18- 36 ℃) Incubation time (4-20 h) Medium composition: Yeast Extract (5-10 g/L) Tryptone (10-20 g/L) Fructose (1-5 g/L)	Fractional Factorial, Plackett-Burman, and Definitive Screening Design,	16, 12, and 22 respectively	Significant factors: Temperature incubation time tryptone OD_{600}	(Uhoraningoga et al., 2019)
		Temperature Incubation time Tryptone OD_{600}	CCD	28	Optimal levels: OD_{600} = 0.55 Temperature= 26 ℃ incubation time=12 h tryptone, 15 g/L. A 2.62X increase in recombinant β-glucosidase production was attained.	
TB1	Modified M9	Medium components Inoculum age Inoculum size Volume of the medium Induction time Inducer concentration, Initial pH of the medium Induction temperature	OFAT	-	Medium: M9 with 3% yeast extract 19 h 6% 50 mL 21 h 0.7 mM IPTG pH 6.6 37 ℃	Lu et al. (2015)

					The concentration of rMBP-NAP increased from 59 mg/L to 592 mg/L after optimization.	
NA	Modified M9 with yeast extract	Inoculum volume (400-600 mL) Induction time (2-4 h) Inoculum age (10-14 h) Aeration rate (10-20 L/min) Fermentation volume (5-7 L) Fermentation pressure (0.03-0.06 MPa) pH (6.2-6.8) DO concentration % (20-40)	Plackett-Burman Screening	12	Significant factors: Induction time Inoculum age pH	Zhang et al. (2010)
		induction time (2-4 h) inoculum age (10-14 h) pH (6.2-6.8)	Box-Behnken Optimization	15	induction time 3.2h inoculum age 12.6 h pH 6.7 An average of 9.68 g/L HLC III - which was 80% higher than the concentration of 5.36 g/L before optimization.	
JM 109	LB	Bioreactor fermentation conditions: Agitation speed (200- 500 rpm) pH (6.0 - 8.0) aeration (0.2 - 0.6 vvm)	OFAT	NA	400 rpm agitation pH 7.0 uncontrolled 0.4 vvm aeration	Jain et al. (2012)

2.4.3 Taguchi Orthogonal Array Design

The Taguchi Orthogonal Array design is a simple approach involving a system of arrays that allow for the maximum number of main effects to be estimated using a minimal number of experiments, in an unbiased manner. Fontani (2003) recombinant *Helicobacter pylori* neutrophil-activating protein production applied the Taguchi Robust Design to optimize medium components with a small number of experiments (four runs). Mel et al. (2006) used the Taguchi OA design to study lactic acid fermentation using *Lactobacillus rhamnosus* in a 2 L stirred tank reactor batch process. The three factors studied were pH, dissolved oxygen, and agitation speed. The design was selected for its reproducibility and the effects of the factors were studied via four experimental runs. From the main effects and correlations, the findings were later used to determine optimal fermentation conditions. Similarly, Hajinia (2012) used the Taguchi L4 Orthogonal Array to design the optimization of recombinant hGM-CSF fermentation factors, namely the IPTG concentration, carbon source, and post-induction time. A study by Rao et al. (2008) reviewed the use of the Taguchi methodology as a statistical tool in the field of biotechnology and stated that it allows for rapid and accurate gathering of data for low-cost design and production. The review also cited studies comparing the method to response surface methodology (RSM), concluding that despite the RSM having certain advantages, the results were comparable, and the Taguchi method was much faster. In this study, the Taguchi method was chosen for its efficiency in identifying significant variables with minimal experimental runs, especially at early stages of process screening.

2.5 RECOMBINANT PROTEIN PRODUCTION CHALLENGES

In a bioreactor, various reactions are taking place, and several factors are at play affecting the fermentation process. For an aerobic process, the fermentation medium

type, its composition, its viscosity, oxygen content, pH, agitation speed and shear, and induction temperature, are some of the factors that affect biomass production, product yield, and activity of the recombinant protein (Eiteman & Altman, 2006; Graumann & Premstaller, 2006; Zhou et al., 2018). Changing the culture conditions is a convenient way to alter the growth of *E. coli* and directly affects the yield of the recombinant protein. Therefore, the kinetics of recombinant *E. coli* fermentation must be carefully studied in order to optimize the concentration and activity of the recombinant CLP products. The kinetics of cell growth, substrate utilization, and product formation are further elaborated in section 2.6.

For an enhanced production that meets market demands, the microbial fermentation process must be transferred from laboratory-scale bioreactors into pilot-scale and industrial-scale bioreactors (Azmi et al., 2015). However, this transfer is not straightforward. Key challenges are summed up in the excerpt:

"A common problem with large-scale protein production is the metabolic overload and stress since plasmid maintenance and heterologous protein expression burdens upon the host cell. Especially in high cell density fermentation, mixing problems, and resulting nutrient, pO_2, and pCO_2 gradients may even enhance cellular stress and favor culture heterogeneity. Glucose feed and dissolved oxygen concentrations govern the formation of acetate which may be toxic for the cells. Apart from host/vector-specific factors, plasmid retention may be dependent on fermentation parameters and nutritional status" (Graumann & Premstaller, 2006).

In the scale-up of *E. coli* fermentation processes, it is necessary to consider conditions that would provide adequate oxygen and nutrient mass transfer to the cells, reduce the shear stress experienced by the cell, and allow sufficient removal of products from cells. Furthermore, the stability of the product must be considered, as correct protein refolding at a large scale remains a major bottleneck. Another factor that must be considered is the desired form of the protein, for example in terms of solubility. Fermentation and induction parameters can affect the solubility of the protein. A review reported that bacterial cultures at 37 °C resulted in the formation of inclusion bodies,

while incubation at a lower temperature (30 °C) led to the production of active and soluble proteins (Rosano & Ceccarelli, 2014). Furthermore, longer induction durations at low temperatures (15 - 25 °C) rather than 37 °C were found to be conducive to the formation of correctly folded, stable, soluble protein and thus elevated target protein yields (Huang et al., 2021). Large-scale purification has been another challenge as noted in the case of Scl2 production. Fortunately, a method has been developed and optimized that involves acid precipitation and proteolysis of residual host proteins (Peng et al., 2014b).

2.6 KINETIC STUDIES AND MODELING

For industrial biotechnology, accurate mathematical models aid in the rational design of microbes and the production processes in which they are utilized. Mathematically modeling the kinetics of fermentation represents an important tool in describing complex microorganism behavior and understanding the biochemistry of metabolic regulation, and in predicting and controlling problem fermentations, and it has technical, economic, and physiological consequences in the analysis of experimental data, reactor design, and optimization of operating conditions to produce a target bioproduct (Almquist et al., 2014).

The growth of a microbe is a unique process that is measured by the increase in the constituents of the cells. As microbial growth is typically rapid, it is realized that cell growth and cell division are inseparable for microbes and thus it is more practical to analyze population dynamics. A thorough understanding of the growth kinetics of the host cells allows for the fine-tuning of the fermentation process until the optimal concentration of the target product is attained (Gameil et al., 2023).

The growth rate is directly proportional to the concentration of the cell, and this cell concentration can be analyzed via several methods which may be classified into direct and indirect methods. Direct methods involve quantifying the cell concentration by mass or number density via the dry weight, wet cell paste, turbidometry (optical density at 600 nm), or colony counts. In the dry cell weight (DCW) method, a similar approach is used but the sample is centrifuged at 9000 rpm for 10 minutes and washed three times with distilled water, and placed in an oven to dry at 105 ℃ till a constant mass is achieved. The disadvantages of this method are that it is time-consuming and not very sensitive for the case of bacteria. This method is not suitable for bacterial growth and is often used for fungi. Alternatively, turbidometric methods such as OD_{600} are rapid, easy, and sensitive. On the other hand, indirect methods measure the concentration of total proteins, metabolic activity products, or DNA content. Microbial growth kinetics then elucidate the relationship between the specific growth rate and the substrate concentration. Thus, microbial growth kinetics are affected by the specific culture conditions. Furthermore, from the relationship between product synthesis and cell growth, a fermentation process may then be described as either growth-associated, non-growth-associated, or mixed-growth associated.

Mathematical modeling of bioprocesses began with Monod's model of microbial growth, proposed in 1950. It is centered around the Monod equation which has the same form as the Michaelis–Menten equation, with both equations assuming that a single rate-limiting enzyme-catalyzed step controls the growth rate. Unlike the Michaelis-Menten equation which is based on theoretical considerations, the Monod equation was derived empirically. Despite its simplicity, the equation works well for a large number of steady-state and dynamic situations. Over time, many modifications of Monod's model such as Moser, Contois, and Teissier, have emerged to include consideration for maintenance requirements, inhibition, and other special features of fermentation processes. For most engineering applications, the predicted values from the models tie in satisfactorily with those of experimental data.

With progress, kinetic models have been classified into structured and unstructured models and segregated and unsegregated models. Compared to unstructured kinetic models, structured models are more idealistic and complex. They can explain complex microbial systems at the molecular level. However, the simpler unstructured kinetic models, such as the Monod model and Luedeking–Piret (LP) model, have been used more frequently for practical reasons (Ali et al., 2017). The growth, substrate consumption, and product formation rates can be expressed either in terms of volumetric (per unit volume of the bioreactor) or specific rates. The volumetric rates are important properties in the design and operation of industrial-scale fermentation processes whereas specific rates are relevant in the evaluation of the cell metabolism (Noorbatcha, 2011).

2.6.1 Growth Kinetics

A typical microbial growth curve is composed of four distinctive segments, namely the lag phase, log or exponential growth phase, stationary phase, and death phase. Sometimes two additional phases are considered, namely the acceleration and deceleration phases, flanking the log phase. In order to acclimatize itself to the new environment, the microorganism first undergoes a lag phase under which it modifies its metabolic pathways and biosynthesizes new cell components and/or products to survive (Rolfe et al., 2012). It is a stage of non-productive growth or non-replication. Often, the log phase is considered to be the governing phase of the economics of a bioprocess. However, the lag phase can play a very important role too, especially in fed-batch mode. This is because the length of the lag phase should be minimized for cost-effective fermentation. For economics, both the times of active production of the desired product and the nonproductive periods associated with the lag phase, inoculation, cleaning, sterilization, filling, and emptying of the bioreactor as well as heating and cooling should be considered. The lag phase length is influenced by the growth medium characteristics, the ratio of growth medium to inoculum volume, as well as the microorganism type and inoculum age (Hill & Thatcher, 2014). The log phase, on the

other hand, is a period of fast exponential growth and replication that occurs at a constant rate. As substrate are depleted and conditions become unfavorable for growth, bacterial growth ceases but the bacteria are metabolically active and compete for nutrients. Due to the rate of growth and the rate of bacterial death being balanced, the viable cell number plateaus. If the bacterial death rate is greater, however, the growth curve enters the death phase and the cell number declines.

Cells that contain plasmids grow slower than non-plasmid-containing cells. Over time, in long fermentation, more cells lose their plasmids and this plasmid instability can be due to several factors such as the plasmid type, the medium composition (culture environment), and the metabolic state of the host cells and their growth rate (Doran, 2012; Rosano & Ceccarelli, 2014). Hence, this is another reason for understanding the kinetics of the fermentation processes, particularly those involving plasmid-carrying recombinant bacteria.

2.6.1.1 The Monod Model

The Monod model is a valuable, frequently used, mathematical tool in microbial kinetics and plays a central role in understanding and optimizing microbial processes in biotechnology. It describes how the specific growth rate of a microorganism is influenced by the concentration of the limiting substrate available in the growth medium. The Monod model assumes that the growth of the microorganism is primarily limited by the availability of a single, specific substrate. This substrate is usually a carbon source or another nutrient necessary for growth.

The Monod model for growth is expressed in Equation 2.1 as follows.

$$\mu = \mu_{max} \frac{S}{Ks+S} \qquad \text{Equation 2.1}$$

Where μ represents the specific growth rate of the microorganism (/h), μ_{max} is the maximum specific growth rate that the microorganism can achieve when the substrate concentration is not limiting its growth (/h), S is the concentration of the limiting substrate (usually the carbon source) in the medium (e.g., glucose) (g/L or mM), and K_s is the half-saturation constant (or Monod constant) which represents the substrate concentration at which the specific growth rate is half of μ_{max}.

As the substrate concentration (S) increases and approaches K_s, the specific growth rate (μ) also increases rapidly but eventually levels off as it approaches the maximum specific growth rate (μ_{max}). When the substrate concentration is much lower than the half-saturation constant, K_s, growth is limited by substrate availability, and the specific growth rate is roughly proportional to the substrate concentration. In contrast, when S is much higher than K_s, growth is not limited by substrate concentration. At very low substrate concentrations ($S << K_s$), the growth rate is negligible. A smaller K_s indicates a higher affinity for the substrate. The Monod constant (K_s) is thus an important parameter that characterizes the affinity of the microorganism for the substrate. It quantifies how sensitive the microorganism is to changes in the substrate concentration. A lower K_s indicates that the microorganism can grow at lower substrate concentrations, while a higher K_s means that higher substrate concentrations are required for growth.

Equation 2.2 for the modeling of growth rate in batch fermentation is based on the Monod model:

$$\frac{dX}{dt} = \mu . X = \mu_{max} \frac{S}{Ks+S} . X \qquad \text{Equation 2.2}$$

where X is cell concentration and μ is the specific growth rate defined in Equation 2.1.

Despite its popularity, the Monod equation can be unsatisfactory for modelling growth in certain cases, such as describing growth on multiple substrates, and in describing the lag and acceleration phases of batch cultures (Noorbatcha, 2011). In the event where the limiting substrate is not easily identified, substrate-independent kinetics

of growth could be obtained by implementing a suitable modified Monod model equation or by using the Logistic growth model (Ali et al., 2017). The logistic growth model is elaborated in the next subsection.

2.6.1.2 The Logistic Growth Model

The logistic growth model is a well-known model for bacterial growth, including recombinant *E. coli*. It is based on the idea that natural bacterial populations initially grow exponentially but eventually reach a carrying capacity where the growth levels off due to limited resources or inhibitory factors due to cell concentration (Azmi et al., 2018). It is frequently used to model substrate-independent growth kinetics for processes where information on the limiting carbon substrate is unavailable, e.g., in complex media, or when the growth does not depend mainly on a limiting substrate.

The logistic model equation (also known as the Pearl-Velhurst equation), is expressed as:

$$\mu = k\,(1 - (\frac{X}{Xmax})) \qquad \text{Equation 2.3}$$

and the growth rate is :

$$\frac{dX}{dt} = \mu.X = k\,(1 - (\frac{X}{Xmax}))X \qquad \text{Equation 2.4}$$

where X represents biomass concentration (g/L), k represents the initial specific growth rate (h^{-1}), and X_{max} represents the maximum cell concentration (g/L).

2.6.2 Substrate Utilization and Product Formation Kinetics

Simple, unsegregated and unstructured models that have been derived from the Monod equation are frequently utilized in preliminarily defining the cell growth and fermentation kinetics patterns and parameters in all operation modes of fermentation processes. The Monod model equation for growth expressed in Equations 2.1-2.2 can be input into the rate of change of substrate concentration with respect to time as follows:

$\mu X = (\mu_{max}.S.X)/(K_s + S)$ Equation 2.5

$r_s = -ds/dt = \mu X/Yx/s + m_s X$ Equation 2.6

wherein, K_s is the Monod cell growth saturation coefficient (g/L), μ_{max} represents the maximum specific growth rate (/h), and m_s represents the substrate utilization rate constant related to cell maintenance.

In the case of fermentation processes with product formation, the rate of formation is described by the mass balance:

$dP / dt = q_p X – (F_o / V) - K_{pd} P$ Equation 2.7

wherein P is the product concentration, q_p is the rate of product formation (synthesis), F_o / V represents the removal of the product, and K_{pd} is the rate of product decay or denaturation (1/h).

Assuming that the product is stable and not removed from the bioreactor, this equation then becomes:

$dP / dt = q_p X$ Equation 2.8

and, the specific rate of product formation, q_p can be expressed as :

$q_p = 1 / X (dP/dt)$ Equation 2.9

The product formation itself can be characterized as one of three categories, namely, growth-associated, partially growth-associated, or non-growth associated product formation. Growth-associated product formation is directly related to cell growth and is described as:

$$dP / dt = \mu X / Y_{p/x} \quad \text{Equation 2.10}$$

wherein $Y_{p/x}$ is the product yield coefficient (g product per g DCW).

Some processes involve partially-growth associated product kinetics, as in the case of lactic acid fermentation, wherein the product is formed during a part of the growth phase only. This rate of product formation is described by adding a non-growth associated term as follows:

$$dP / dt = \alpha.dX / dt + \beta.X \quad \text{Equation 2.11}$$

whereby α and β are coefficients representing the growth and non-growth associated product formation terms, respectively. Generally, the equation is useful in mixed-growth fermentation systems, and it becomes Equation 2.12 when the product formation is associated with the utilization of the carbon source:

$$dS / dt = \mu .(X/Y_{x/s})+m.X \quad \text{Equation 2.12}$$

which can be rewritten as:

$$1 / X (dS/dt) = \mu/ Y_{x/s} + m \quad \text{Equation 2.13}$$

As product formation can be expressed as:

$$q_p = Y_{p/s} . q_s \quad \text{Equation 2.14}$$

This equation becomes:

$$q_p = (Y_{p/s} / Y_{x/s}) \mu + mY_{p/s} \quad \text{Equation 2.15}$$

Thus, it is observed that α and β in Equation 2.12 became the terms $Y_{p/s} / Y_{x/s}$ and $mY_{p/s,}$ respectively. These parameters help us in understanding growth efficiency and product formation efficiency.

Finally, non-growth-related products are products that are not essential for the growth of the organism, such as certain microbial toxins and agents of symbiosis and are referred to as secondary metabolites. In such cases, product formation is not related to biomass growth and therefore Equation 2.7 must be used (Noorbatcha, 2011).

The Luedeking-Piret model expands on the Monod model by considering both growth-associated and non-growth-associated product formation. This model is useful for predicting product yields during fermentation and in describing the relationship between biomass growth, product formation, and byproduct formation in a microbial fermentation process. However, in some cases, this basic model may not fully capture the complexities of the system. To address this, several modifications and extensions have been proposed over the years to make it more versatile and applicable to a wider range of scenarios. For instance, the classical Monod equation links cell growth and substrate utilization and linearly relates the yield coefficient (Y_{XS}) to the specific growth rate (μ) and the rate of substrate consumption but does not consider non-growth-related substrate utilization i.e., cell maintenance. Hence, later equations emerged to include maintenance, for example, by including the maintenance rate (m):

$$\frac{dS}{dt} = -\frac{1}{Y_{XS}}\frac{dX}{dt} - m\,X = -\frac{\mu}{Y_{XS}}X - m\,X \qquad \text{Equation 2.16}$$

The Luedeking-Piret model-based equations for the rate of substrate consumption ($\frac{dS}{dt}$) and the rate of product formation ($\frac{dP}{dt}$) are given by:

$$\frac{dS}{dt} = -m\frac{dX}{dt} - n.X \qquad \text{Equation 2.17}$$

$$\frac{dP}{dt} = \alpha\frac{dX}{dt} + \beta X \qquad \text{Equation 2.18}$$

where X is the biomass concentration (g/L), S is the substrate concentration (g/L), P is the product concentration (g/L), α represents the growth associated product formation constant, β represents the non-growth associated product formation constant, m represents growth associated substrate utilization constant (g/g), n represents the non-growth associated substrate utilization constant (g/g.h), k_e is the maintenance coefficient for cells (g/g.h), Y_{XS} is the yield of cell mass from substrate (g/g), and Y_{PS} is the yield coefficient of product from substrate (g/g).

Determining substrate utilization constants *m* and *n* allows understanding of the substrate utilization in terms of active biomass growth and cellular maintenance and metabolism. Understanding growth and non-growth related substrate utilization then allows for the prediction of substrate requirements. Determining product formation constants α and β helps in quantifying the relationship between biomass growth and product formation for a specific bioprocess. The β value may be zero, negative, or positive, as the Luedeking-Piret model is related to growth-associated or non-growth-associated product formation. If the product formation rate depends only on biomass growth rate and not on biomass as such, β becomes 0. A negative β value suggests that the product being formed is also being consumed. If β is positive, it indicates non-growth-associated product formation continues even during stationary phase. If the magnitude of α is bigger than β, the production of *P* is more strongly linked to cell growth. Once the model parameters have been estimated, the Luedeking-Piret equation can be used to optimize process conditions, such as substrate feeding rates, to maximize product yield or productivity while minimizing substrate consumption. Thus, from the values of of α and β, a fermentation process may be described as growth-associated ($\alpha > 0, \beta \approx 0$) , non-growth-associated ($\alpha \approx 0, \beta > 0$), or mixed-growth associated ($\alpha > 0, \beta > 0$) (Carranza-Saavedra et al., 2021).

2.6.3 Tools for Kinetics Parameter Estimation

While the field of kinetic modeling for *E. coli* batch fermentation has witnessed significant advancements over the past decade, the Monod, logistic, and Luedeking-Piret models remain indispensable tools towards understanding growth kinetics, product formation, and substrate utilization. With the integration of computational tools like MATLAB and techniques such as ODE45, Levenberg-Marquardt algorithm, and nonlinear regression, researchers have the means to simulate and optimize fermentation

processes with greater precision and efficiency. These are discussed further in Chapter Three.

2.7 SCALE-UP

Bioprocesses are usually developed in three phases or scales, from the bench scale, where basic screening procedures are conducted, to the pilot scale, where the optimal operation conditions are established; and finally, the plant scale, where an economically viable production is set up. Once a recombinant product of interest has been successfully produced at the bench scale or laboratory scale, with sufficient productivity, the engineering process is taken to the next step, which is scale-up. Scale-up is an important aspect of any bioprocess. It is the process of successfully reproducing the operation of a small viable bioprocess at a larger scale production in larger quantities to meet the demands of the industry (Spier et al., 2012). While it may seem simple for chemical processes, the scale-up of bioprocesses is often not a straightforward process. This is because they involve live cells, which are very complex and are not limited to the production of our desired product alone. Furthermore, with an increase in the volume of fermentation, heterogeneities become more apparent due to deficient mixing. These gradients affect microorganisms as they experience different zones of concentration. Consequently, these become more apparent as significant differences in productivity, quality, and/ or yield of our desired products (Xia et al., 2016). There are, therefore, many considerations to make to successfully upscale a bioprocess.

Traditionally, the methods of scale-up rely on the theory of similarity. It is not possible to apply geometric, dynamic, thermal, mass, and biochemical similarity altogether as a whole to the scale-up of bioreactors. However, similarity theory has allowed the development of scale-up methods. Scale-up methods can be classified into fundamental models, dynamic analysis, rules of thumb, scale-down approach/regime analysis, and trial and error.

Fermentations, in general, can be assessed by using the following interrelated physical characteristics and transport phenomena: mixing time, shear, pH, heat transfer, and mass transfer. As pH and heat transfer do not constrain scale-up, they are often considered separately. Of the remaining characteristics, one or some of the following parameters are often kept constant for scale-up:

1. The geometrical and configurational similarity of the bioreactor,
2. Volumetric oxygen transfer coefficient, k_La,
3. Maximum shear,
4. Power input per unit volume of liquid,
5. Volumetric gas flow rate per unit volume of liquid,
6. Superficial gas velocity,
7. Mixing time,
8. Impeller Reynolds number, Re_i, and
9. Momentum factor (Asenjo & Merchuk, 1994).

There are four commonly used scale-up strategies for geometrically similar bioreactors based on identifying a criterion, namely – a constant volumetric oxygen mass transfer coefficient (k_La), constant volumetric power input (ratio of agitation power input, P, to volume of liquid, V), constant maximum shear stress (given by the tip speed of the impeller), and constant mixing time. Usually, a combination of two to three strategies is applied for a particular scale-up, and often k_La is one of these because of its importance in aerobic fermentations. For instance, Pérez et al. (2018) attempted upscaling the production of P46k protein in recombinant *E. coli* K12 GCC366 cells from the 2.5 L scale to the 70 L scale, by implementing four (4) scale-up strategies. The dry cell weight (DCW) and protein production were highest for constant k_La (refer to Table 2.18). In contrast, cell and protein concentrations showed a pronounced decrease when the constant Reynolds number (*Re*) criterion was used, and this was due to the resultant change in impeller speed. With the exception of the constant *Re* criterion, the scale-up criteria gave comparable impeller speeds. Nonetheless, the constant k_La

criterion achieved scale-up without increasing the shear stress on the cells, and thus it was favored.

In another publication, Deniz et al. (2015) studied the scale-up of bioethanol fermentation using *E. coli* KO11 by testing three scale-up criteria, namely constant mixing time, constant impeller tip speed, and constant oxygen mass transfer coefficient. The maximum bioethanol concentration attained using constant mixing time, 23.42 g/L, 18.13 g/L using constant tip speed, and the bioethanol concentration was lower than both these values using the constant oxygen mass transfer coefficient scale-up criterion of 0.0012 s^{-1}. As the fermentation was viscous, mixing time proved to be the best criterion in this case, as homogeneity was important. Garcia-Ochoa and Gomez (2009); Garcia-Ochoa et al. (2020); Garcia-Ochoa et al. (2010), and Marques et al. (2010) provide important reviews of the oxygen transfer and various scale-up criteria, and how they impact the scale-up. To this point, no scale-up study that is based on a scale-up criterion rather than trial and error has been published for bacterial CLP production. In fact, the only publication that involved a scale-up study was by Peng et al. (2012), but it does not provide details on scale-up strategies and parameters.

Table 2.18 Scale-up criteria and their applications

Parameter	Scale-Up Criteria	Application
Aeration number (N_A)	$(N_A)_1 = (N_A)_2$ $N_A = Q/(nD_i^{\,3}) = Q/[(nD_i)(D_i^{\,2})]$	This criterion is usually applied for cases where expense on stirring is key.
Impeller Reynolds number	$(N_i.Re)_1 = (N_i.Re)_2$ $Re = \rho N D_i^2 / \upsilon$	In cases where the rate of heat removal from the fermentation broth to the bioreactor vessel's cooling jacket becomes critical, for example, when working with thermophiles. This is because the impeller *Re* affects the heat transfer coefficient.
Impeller tip speed	$(N_i.D_i)_1 = (N_i.D_i)_2$ N_2 agitation speed in scale-up N_1 agitation speed in scale-down $Di,_1$ impeller diameter of scale-down $Di,_2$ impeller diameter of scale-up $v_{tip} = \pi N D_i$	This is used when handling shear-sensitive cells whereby a maximum shear rate is allowed to prevent possible irreversible shear damage to host cells inside the bioreactor. When scale-up based on constant tip speed is attempted, *P/V* and mixing time will decrease. It is also utilized for cases where the cells form dense masses and impede the culture, and therefore the minimum shear rate needed to keep the cells separate has to be calculated.
Mixing Time	$N_2 = N_1 (D_1 / D_2)^{1/4}$ N_2 is agitation speed in scale-up, N_1 is agitation speed in scale-down, D_1 is impeller diameter of scale-down, and D_2 is impeller diameter of scale-up	This is often useful for viscous fermentation broths wherein ensuring that the materials are well-mixed promptly is a top priority.
Power Per Unit Liquid Volume (P/V_L)	$\left(\frac{P}{V.L}\right)_1 = \left(\frac{P}{V.L}\right)_2$ $P/V \approx N^3/D_i^2$ or $P/V \approx N^3/D_i^2$	Mammalian cells cannot handle a lot of power introduced into the culture media as it can cause small eddies that will shear the fragile cell membranes. The majority of aerobic fermenter systems have been scaled up on the k_La basis and a few on the *P/VL* basis.

	P is the power supplied, V is the volume of the bioreactor, N is the agitation speed, and D_i is the impeller diameter.	
Superficial Gas Velocity (V_s)	$V_s = Q_{gas} / A_v$ V_s - superficial gas velocity Q_{gas} is the gas volumetric flow rate, and Av is the inside cross-sectional area of the vessel.	Increasing the superficial gas velocity causes an increase in foam generation
Volumetric oxygen transfer coefficient	$(K_L a)_1 = (K_L a)_2$	This criterion is usually applied to aerobic systems where oxygen concentration is most important and affects the metabolism of the microbial cell. The bulk of aerobic fermenters were upscaled on this basis.

Source: Gameil et al., 2021a.

2.7.1 Constant Oxygen Mass Transfer Coefficient

Oxygen is a key substrate for aerobic processes. However, it has low solubility in fermentation broths. Its solubility in distilled water is only about 7 mg L^{-1} at 30 °C (Vendruscolo et al., 2012), whereas its solubility in M9 minimal medium (with 5 g/L glucose) = 0.264 mM at 37 °C) and a continuous supply is needed (Patnaik et al., 1992). Often, oxygen becomes the limiting substrate in the aerobic reaction. It is important that the oxygen concentration remains above the critical oxygen concentration value (C_{crit}) so that the rate of oxygen uptake by the cells is independent of the oxygen level. As such, the oxygen transfer rate of a bioprocess must be estimated at critical points in the fermentation to optimize the operation and scale-up of a bioprocess. For this reason, comparable oxygen transfer is often used as a basis for the scale-up of a bioprocess.

According to the two-film theory of oxygen transfer from the gas phase to the liquid medium, the rate of oxygen transfer depends on the k_La value and the driving force for mass transfer (Shuler & Kargi, 2001). The mass transfer coefficient (k_La) indicates the efficiency of oxygen supply to the microbes and is an important scale-up factor. During cell culture, oxygen is transferred from a gas phase (a bubble or headspace) to the liquid phase from where it can then be transported and utilized by the microbial cells. This process is a flux across a barrier which can thus be expressed as the product of the driving force (difference in concentrations) and a transfer coefficient (k_La). The oxygen flux is given by Equation 2.19.

$N_{O2} = k_L A(C^* - C_L)$ Equation 2.19

Where k_L is the liquid mass transfer coefficient in *m/s*, *A* is the interfacial area for mass transfer in m^2, and *c** is the equilibrium oxygen concentration in mol/m^3.

Then, the oxygen transfer rate per unit of reactor volume, Q_{O2}, is given by Equation 2.20.

$$Q_{O2} = \frac{N_{O2}}{V} k_L a(C^* - C_L)$$ Equation 2.20

$$\frac{dC_L}{dt} = k_L a(C^* - C_L) - q_{O2}$$ Equation 2.21

Where $k_L a$ is the oxygen mass transfer coefficient c^* is the equilibrium oxygen concentration in mol/m^3, c_L is the equilibrium oxygen concentration in mol/m^3, and q_{O2} is the specific oxygen uptake rate.

It is difficult to determine the value of the specific surface area for small bubbles in a bioreactor. Hence, for practical purposes, the oxygen mass transfer coefficient, k_L, and the specific exchange surface area, a, are not considered separately, but as a product, i.e., $k_L a$, as in Equation 2.21.

The $k_L a$ value is a useful parameter that can help in describing and understanding oxygen transfer in a bioreactor that is dependent on several physical and chemical factors. The $k_L a$ value is dependent on the reactor design (size), material properties, and operation factors, such as impeller design, temperature, aeration, and agitation rates. The impeller type, its speed, and location all influence the impeller tip speed and there is a positive correlation between $k_L a$ and impeller tip speed. Contrary to expectations, however, increasing the number of impellers may not enhance $k_L a$ values as gas passing through the upper impellers is smaller than the lower impellers, so the additional gas dispersion is insignificant at the expense of higher power consumption (Mel et al., 2010). The sparger characteristics, namely pore size, pore number, and surface area, all affect $k_L a$ due to their effect on bubble size, gas velocity, and flow rates. Increasing temperature inversely affects oxygen solubility in the culture medium and causes C^* values to drop so that the driving force for mass transfer (C^*- C_L) is reduced. This in turn causes $k_L a$ to drop (Mel et al., 2010). Airflow, on the other hand, does not generally exert much effect on $k_L a$ except at very low sparging rates. Similarly, increasing the number of impellers may not affect $k_L a$ favorably since power consumption increases but overall gas dispersion does not change much (Mel et al., 2010).

Systems in which a microorganism is induced to produce a certain product, such as the production of recombinant collagens, may require higher aeration capacities than

their counterparts. This was shown in a study by Bhattacharya and Dubey (1997) wherein it was observed that the measured k_La values were higher for induced fermentations compared to uninduced fermentations under similar conditions of aeration (2.5 vvm) at various agitation speeds (250, 500, 650, 750, 1000, 1200 rpm). However, this is not always the case. For instance, studies by Rao (2009) and Kapat et al. (2001) showed that glucose oxidase production was not dependent on k_La, but the agitation speed and aeration were still important factors in the process. Figure 2.7 is a schematic that shows the main factors that influence k_La values in a bioreactor process

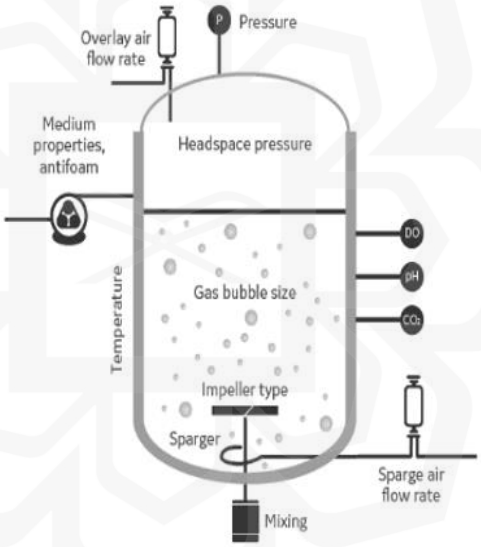


Figure 2.7 Key factors that influence k_La values in a bioreactor process (Cytiva Life Sciences)

Determination of k_La values is very challenging because of the complicated nature of gas-liquid hydrodynamics (Spier et al., 2012). However, a correlation for k_La values has been long established, relating it to power density or gassed power consumption per unit volume of broth (P_g/V_L), and superficial gas velocity, v_s:

$k_La \propto (P_g/V_L)^{a1}.(v_s)^{b1}$ Equation 2.22

The constants a_1 and b_1 depend on the geometry of the vessel and stirrers, medium composition (e.g., liquid viscosity), properties of the microorganism, and experimental methodology. This correlation between k_La and P_g/V_L and its variants has been widely used in the past decades. While some empirically derived expressions are available for predicting k_La values in non-Newtonian fluids, it is difficult to account for all of the variables that can affect the k_La. As such, k_La studies need to be performed to specifically investigate the unique bioreactor conditions and its value determined experimentally by any of several chemical, enzymatic, and physical methods, for example by a reducing agent such as sodium sulfite, and adsorption of carbon dioxide. Usually, physical determination methods i.e., static, and dynamic gassing out are used in scale-up. The underlying principle of these methods is that in integrating**:**

$$\frac{dC_{AL}}{dt} = k_L a(C^*_{AL} - C_{AL}) \qquad \text{Equation 2.23}$$

to obtain:

$$\int_{C_{AL_1}}^{C_{AL2}} \frac{1}{(C^*_{AL} - C_{AL})} dC_{AL} = k_L a \int_0^t dt \qquad \text{Equation 2.24}$$

$$ln\left(\frac{C^*_{AL} - C_{AL1}}{C^*_{AL} - C_{AL2}}\right) = k_L a.t \qquad \text{Equation 2.25}$$

a plot of $ln\left(\frac{C^*_{AL} - C_{AL1}}{C^*_{AL} - C_{AL2}}\right)$vs t is generated, and it will result in a straight line with k_La as its gradient and the intercepting at the origin.

The static gassing out method, often used with nonaerobic microorganisms, is carried out by sparging the vessel with nitrogen gas, thus displacing the oxygen. Variations in dissolved oxygen concentration (DO) are then monitored using a (polarographic) DO probe. The DO is allowed to fall to 0% saturation, and then nitrogen gas flow is stopped. Air is then sparged into the vessel at a known flow rate. Variations in dissolved oxygen concentration (DO) are then monitored with respect to time. Finally, a plot of $ln\left(\frac{C^*_{AL} - C_{AL1}}{C^*_{AL} - C_{AL2}}\right)$vs t gives a line with a slope k_La and an intercept at the origin (Azmi et al., 2015).

In the case of aerobic organism(s), k_La may be determined via the dynamic gassing out method whereby initially, the air supply to the vessel is suspended temporarily. The decrease in dissolved oxygen concentration (DO) concentration is monitored for a certain length of time via a (polarographic) DO probe before the air supply is restored. The subsequent rise in DO concentration with time after the resumption of aeration is monitored and noted (Azmi et al., 2015; Jain et al., 2012). The method involves measuring the DO of the fermentation broth during a short interval where aeration is interrupted to determine k_La. Experiments on prolonged oxygen starvation suggested that the OTR will vary if prolonged (2–10 minutes) oxygen starvation occurs, below the critical biological oxygen concentration (Rao, 2009). The dynamic gassing out method is the most prevalent method employed for k_La and requires the determination of oxygen transfer rate (*OTR*) and oxygen uptake rate (*OUR*), respectively (Zhou et al., 2018). It has the advantages of being rapid and internally consistent. Unlike the static gassing out method, dynamic gassing out uses cells in the medium and it is much more precise compared to other methods of estimating k_La. The drawbacks of this classical method are mainly the accounting for surface aeration at high speeds and the slow release of bubbles in highly viscous fermentation broths (Rao, 2009). An improved method, which takes the mass transfer between the headspace and the broth into account and involves bubbling nitrogen gas through the broth when air is shut off was developed to allow a more accurate measurement of k_La for high cell density cultures using low or medium agitation (Damiani et al., 2014).

In the sodium sulfite method, the bioreactor is filled with 1 M sodium sulfite solution. When the aeration is turned on, the sodium sulfite reacts quickly with the dissolved oxygen in the presence of a 1 mM catalyst (often copper (II) or Cobalt (II) ions) to produce sulfate. After ten (10) minutes of aeration, the airflow is stopped, and samples (2 mL) are taken every 30 minutes. The reproducibility of the results is affected by batch-to-batch variability in sodium sulfite purity. The sulfite depletion rate is determined from concentrations at various time points, and k_La may be determined from Equation 2.26-2.27, whereas C* can be obtained by using Equation 2.28.

$-dC\mathrm{Na_2SO_3}/dt = 2k_La C^*$ Equation 2.26

$k_La = -\mathrm{dC_{Na2SO3}}/\mathrm{dt}/2C^*$ Equation 2.27

$C^* = y\mathrm{O2}P/H/\Sigma C\mathrm{i}$ Equation 2.28

Recent studies have questioned the widespread applicability of constant k_La to the scale-up of fermentations and criteria such as the oxygen uptake rate (OUR) or dissolved oxygen (DO) levels are being utilized instead, especially for complex broths and shear-sensitive fermentations (Gameil et al., 2021a). Some of these cases are shown in Table 2.19. However, using constant DO as a scale-up criterion is not recommended as it does not reflect the actual oxygen transfer and uptake rates. It is not reliable and does not reflect the complexity of the hydrodynamics as the fermentation scale changes. Furthermore, there is a lack of standardized protocols for k_La measurement in emerging bioreactors such as single-use bioreactors and micro-bioreactors. Therefore, studies have emphasized the importance of combining multiple criteria, for example using constant k_La and maximum shear rather than depending on a single parameter (Espinosa Perez et al., 2018).

2.7.2 Constant Impeller Tip Speed

A constant impeller tip speed is another commonly used criterion in scale-up. This is because it implies a comparable homogeneity for the up-scaled bioprocess. Numerous factors such as agitation speed, the number of impellers, impeller type, and the gas flow rate into the reactor influence mixing and transfer. However, while an adequate oxygen transfer rate is desirable, compromise in terms of shear stress to cells must be controlled. This criterion is also particularly useful for shear-sensitive fermentations such as those

of branched yeast and filamentous bacteria. The maximum shear rate is indirectly controlled via the maximum impeller tip speed, which is a function of agitation speed and impeller diameter to prevent shear damage that would in turn negatively-affect protein production (Gameil et al., 2021a). Hence this scale-up criterion was employed in this study to ensure that the shear stress is not compromised by the scale-up. Equation 2.29 relates the impeller tip speed, v, to the agitation speed, and impeller diameter. Equation 2.30 relates the agitation speeds (N_1 and N_2) at both scales to the respective impeller diameters and vessel volumes when keeping the impeller tip speeds constant during scale-up.

$$v = \pi(\frac{N}{60})\, D_i \qquad \text{Equation 2.29}$$

$$\frac{N2}{N1} = \frac{D_{i,1}}{D_{i,2}} = \sqrt[3]{\frac{V_1}{V_2}} \qquad \text{Equation 2.30}$$

Where v is the impeller tip speed in m/s, N is the agitation speed in rpm, D_i is the impeller diameter in m, and V is the volume in m^3.

The effect of impeller tip speed on nitrilase production in recombinant *E. coli* JM109 was examined in a previous study, as shown in Table 2.16 (Jain et al., 2012). In the study, the optimal cell concentration and nitrilase activity were obtained at 400 rpm agitation speed, corresponding to 1.59 m/s impeller tip speed, whereas lower agitation speeds negatively affected the final biomass concentration and product activity as the oxygen transfer rate decreased. Furthermore, it was also observed that an agitation speed greater than 400 rpm likewise decreased the cell concentration as well as the nitrilase activity, establishing that the 400 rpm speed was optimal. Nuge et al. (2023) used a constant impeller tip speed scale-up criterion, and it performed better compared to the constant P/V criterion in the 1:10 scale-up of recombinant phytase in *E. coli* DH5α. However, the details of the scale-up parameters were not specified. As a general guideline, it is assumed that cell damage may occur above an impeller tip speed value of 3.2 m/s, though the exact value depends on broth rheology, among other factors (Azmi et al., 2015; Cascaval et al., 2011; Lee, 2009; Marques et al., 2010). However,

for production-level fermenters, it was reported that calculated tip speeds are usually greater than 3 m/s (Azmi et al., 2015; Marques et al., 2010). Finally, Imamoglu and Sukan (2013) recommended this criterion for the scale-up of recombinant organism fermentations, especially when the scale-up factor is small. Often, the P_g/V_L value is lowered, adversely affecting the aeration efficiency and so, this criterion must be coupled with a constant k_La criterion (Gameil et al., 2021a).

Some of the findings using various scale-up strategies are presented in Table 2.19.

Table 2.19 Scale-up for recombinant protein production in *E. coli*

Protein	**Scale-up strategies**	**Scales compared ***		**Target variables**	**Mode**	**Observation(s)**	**Reference**
		SS	**LS**				
P46k protein	constant D_i constant k_La constant P/V constant Re	2.5 (1.5)	70 (50)	Cell dry weight (g/L) P64k Protein Production (mg/L)	Batch	Cell dry weight and protein production improved with all scale-up criteria, when compared to the small scale, except in the case of constant Re where it declined. The production was highest for constant k_La.	Pérez et al. (2018)
Phosphotriesterase-like lactonase	constant DO	2.5	150	Enzyme production (mg/L) Final biomass (gCWW/L)	Fed-batch	Increased concentrations of the enzyme (37.6 to 39.7) and biomass (108 to 130) were obtained. Undesirable formation of acetate was observed. Product inhibition occurred and thus IPTG was substituted with galactose.	Restaino et al. (2017)
GCSF	geometric analogy constant DO	2 (1)	13(8)	Final cell density (DCW/l)	Batch	Cell density and recombinant protein concentration declined significantly due to acetate accumulation and improper mixing and distribution. Plasmid stability	Babaeipour et al. (2017)

				GCSF concentration (g/L)		Oxygen concentration	
Human-like collagen	constant DO	12.8 (6)	30 (15.4)	growth (DCW g L/L) and human-like collagen (g/L)	Fed-Batch	Acetate production	Ma et al. (2014)
CMCase	NA (trial and error)	7 (5)	100 (70)	Production of cells (g/L) CMCase (U/mL)	Batch	Oxygen transfer. Increasing inner pressure provided a higher oxygen transfer rate and this enhanced production to 5.43 g/l from 3.83 g/l and to 880.2 U/ml from 683.8 U/mL.	Lee et al. (2013)

*The values in the brackets indicate working volumes whereas those outside the brackets refer to the vessel volumes. SS refers to small scale and LS refers to Large Scale. Source: Gameil et al., 2021a.

The transport characteristics of bioreactors include heat, hydrodynamic, and mass transfer properties, but the prevalent scale-up strategies are still rules of thumb, as well as trial and error in practice as the effects of hydrodynamics on the growth and metabolism of a specifc microorganism are often unpredictable (Schmidt, 2005). Constant k_La has been successfully utilized in the scale-up of some aerobic cultures and it can be applied to oxygen-limited cultures. As can be seen from Table 2.19, the control and optimization of the oxygen transfer is frequently necessary to achieve high product yields. An extremely low supply of oxygen often leads to production of toxic by-products whereas an oxygen oversupply can inhibit formation of the desired product. For instance, oxygen toxicity was associated with lower enzyme activity (Jain et al., 2012). While kinetic models and scale-up approaches can be useful guides for selecting a scale-up strategy, they cannot reflect the complexity of the conditions that govern the fermentation scale-up and its success and may need further optimizations on each scale.

2.7.3 Scale-Up Strategies: Recent Advances and Challenges

While maintaining constant volumetric oxygen transfer coefficient (k_La) has traditionally been the primary criterion for aerobic fermentation scale-up, recent research suggests that this approach alone may be insufficient for optimizing productivity in complex bioprocesses. Studies using advanced computational fluid dynamics (CFD) and scale-down bioreactors have demonstrated that factors such as heterogeneous oxygen distribution, shear stress gradients, and mixing inefficiencies significantly influence cell physiology and product formation. These factors are influenced by local flow variables and recent studies have highlighted the lack of and need for investigations into flow variables (Delvigne & Noorman, 2017; Li et al. 2018; Nadal-Rey et al., 2022; Nadal-Rey et al., 2021; Makwana, 2023). Machine learning algorithms and artificial intelligence tools have also been employed to predict optimal scale-up parameters by analyzing large datasets from multi-scale bioreactor experiments, offering a promising tool for accelerating process development (Duong-Trung et al., 2022).

In addition to the application of CFD and machine learning algorithms to scale-up, emerging real-time monitoring technologies have permitted the dynamic adjustment of operating conditions based on oxygen uptake rates and other metabolic indicators rather than relying on static k_La values. Despite advances, however, standardized methods for measuring k_La in non-conventional bioreactors remain deficient, and this makes cross-study comparisons and industrial applications more challenging (Xia et al., 2016).

The scale-up of recombinant bacterial collagen-like protein (CLP) has remained largely underexplored. Additionally, existing bioreactor studies on CLP production, such as Peng et al. (2012), do not report detailed scale-up strategies or parameter optimization, indicating a significant knowledge gap. Therefore, it is necessary to integrate multi-criteria scale-up approaches with modern monitoring and modeling tools in order to advance recombinant protein bioprocesses involving bacterial CLPs.

2.8 PROTEIN CHARACTERIZATION

A recombinant protein may undergo structural changes that make it different from a native protein, during the course of the fermentation, and these changes must be investigated and controlled. The end use of the recombinant protein also imposes several standards on its formulation and purity. Hence, it is important to characterize the recombinant protein through accurate, reliable, and reproducible methods. Numerous characterization methods have been reported for recombinant collagens and relevant ones have been summarized as in Table 2.20. Together, these techniques can be used to confirm structural integrity, purity, and functionality of recombinant collagen and CLPs and are indispensable for downstream application.

Table 2.20 Characterization methods employed in recombinant collagen studies

Characterization Method	Recombinant Collagen	Observation/ Key Findings	Reference
Circular Dichroism Spectroscopy	Collagen from *R. palustris*	Stability: typical collagen-like features at 4 ℃. Partial degradation of triple-helix is suspected as 220 nm peak was negative and the Rpn (to estimate triple-helix content) value was low. Refolding: Complete refolding *in vitro* was observed	Xu et al. (2010)
	Collagen from *S. pyogenes* (incorporated with Hyp)	Typical collagen features, with a maximum at 220 nm and minimum at 198 nm. The Rpn values of the triple-helical domain were similar to those in mammalian collagen, indicating that complete folding of the triple helix occurred. Thermal stability was slightly higher for hydroxyproline-included CLP by 0.6-1 ℃, as compared to that with no hydroxyproline inclusion. The samples were also able to refold.	Peng et al. (2018); Yoshizumi et al. (2009)
Differential scanning calorimetry	Collagen from *S. pyogenes*	High enthalpy values were obtained though lower than those for animal collagens which could be related to hydrogen bonding.	Yoshizumi et al. (2009);Xu et al. (2002)
Proteolytic Digestion (Trypsin Digestion)	Collagen from *R. palustris*	This soluble protein contains Trypsin resistant triple-helical collagen domains of actual size 13.335.87 Da (11,275.48 Da theoretical size)	Xu et al. (2010)
Proteolytic Digestion (Various proteases)	Various CLPs	The soluble CLPs contain trypsin-resistant triple-helical collagen domains.	Xu et al. (2010)
MALDI-TOF MS	Collagen from *S. pyogenes*(Scl2)	Protein purity was confirmed.	Yoshizumi et al. (2009)
Rotary shadowing and Electron microscopy	Collagens from *S. pyogenes*	Scl1 and Scl2 proteins formed a "lollipop-like structure" composed of a rod-like collagen-like domain and a globular variable domain.	Xu et al. (2002)
Dynamic light scattering	*Collagen from S. pyogenes (Scl2)*	Self-association and the formation of aggregates of samples at a low temperature (4 °C) and neutral pH were observed.	Yoshizumi et al. (2009)
LC-ESI-/MS	Collagens from *S. pyogenes* (Scl1 and Scl2)	Amino-terminal sequencing and mass determination confirmed the identity of the proteins.	Xu et al. (2002)

2.8.1 SDS-PAGE for Preliminary Detection of CLPs

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is one of the most extensively used laboratory methods to separate and visualize proteins. Polyacrylamide gel electrophoresis itself is a technique where macromolecules and proteins are differentiated according to their electrophoresis mobility which is affected by the molecule's length, conformation, and charge. Sodium dodecyl sulfate is used to linearize polypeptides, giving them an even distribution of charge, and conferring a negative charge which allows them to migrate in the gel, according to their length (Kurien & Scofield, 2012). Smaller proteins or monomers migrate further through the gel than the larger dimers and protein aggregates. The proteins in the gel are then stained with a dye such as Coomassie Blue R-250, Amido Black, Fast Green, Congo Red, Direct Red 81 (Goldring, 2018). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used extensively in the literature to confirm the presence of the collagen-like domain and intact triple helix of the recombinant CLPs from various bacterial DNA constructs (Vandersmissen et al., 2010; Xu et al., 2010; Peng et al., 2015). Typically, 4-12% or 12% Bis-Tris gels were used, with MES running buffer, using 180 V constant voltage conditions for an hour. The sample buffer contained 5% (v/v) 2-mercaptoethanol (Peng et al., 2014a; Peng et al., 2014b; Peng et al., 2015; Xu et al., 2010). SDS-PAGE is also a precursor to western blotting, where antibodies can be used to detect and quantify the protein of interest.

2.8.2 Resistance to Enzymatic Proteolysis

As native triple-helical structures were found to resist enzymatic proteolysis by trypsin, chymotrypsin, pepsin, enzymatic proteolysis of samples that is followed by SDS-PAGE can be used as an indicator of intact structure (Yu et al., 2014). Trypsin is a protease that is secreted by the pancreas and hydrolyzes amide bonds at the carboxylic groups of arginine and lysine. Pepsin is a protease found in the stomach of humans and certain

animals that initiates protein digestion at a naturally acidic medium with an optimal pH value of around 2 and is denatured at neutral pH. Papain is a protease secreted by papaya (*Carica papaya)* and it mediates hydrolysis cleaving peptide bonds of basic amino acids, leucine, or glycine, and hydrolyzing esters and amides. It works at a wide pH range of 3 to 9 (Jon, 2012), and can withstand high temperatures, with an optimal 65 ℃ temperature (Singh et al., 2019). Bromelain is a cysteine protease secreted by the fruit and stem of the pineapple *(Ananas comosus)* that exhibits proteolytic activity in a wide pH range, with optimal enzyme activity around pH 7.0, and 50 to 60 ℃, for stem bromelain (Varilla et al., 2021).

In several studies, trypsin digestion was done to ascertain the stability and integrity of the triple helix (Yu et al., 2014). For example, protease digestion of human-like collagen was studied using pepsin, trypsin, and bromelain (Ma et al., 2017). Protease digestion was also studied for CLPs from *Bacillus anthracis* (Boydston et al., 2005), *Legionella pneumophila* (Vandersmissen et al., 2010), and *Solibacter usitatus* (Peng et al., 2014b). In another study, trypsin digestion of RPCLP was conducted for one hour at room temperature, using 1:1000 (protein/enzyme ratio), at pH 8. SDS-PAGE showed a band of trypsin-resistant species at around the 15 kDa MW mark whereas the purified undigested band is around 25 kDa (Xu et al., 2010). Similarly, studies by (Peng et al., 2014b) carried out experiments on trypsin, chymotrypsin, papain, bromelain, and pepsin digestion of several CLP constructs followed by SDS-PAGE to determine the extent of protease digestion if any. In some of these constructs, e.g., *Methylobacterium sp. 4-46* CLP, the location of the protease digestion cleavage sites was also determined. Results show that pepsin and papain were particularly efficient at removing residual proteins from cell lysates. On the other hand, trypsin and bromelain weren't as effective (Peng et al., 2014b).

As the enzymes used for the proteolysis vary, optimal pH values are selected and the methods of the proteolysis also vary –the incubation temperature, the length of the incubation period, the inhibition method, and the inhibitor used to stop the proteolysis e.g., trifluoroacetic acid (TFA), phenylmethylsulfonyl fluoride (PMSF), and

tosyl phenylalanyl chloromethyl ketone (TPCK), as well as other additives that may be added. Table 2.21 shows some of these combinations that have been applied to collagen-like proteins, including enzyme: protein ratio, incubation time, incubation temperature, pH, and inhibition method. Even though studies on bromelain and ficin were conducted, the details were not reported fully (Peng et al., 2014b).

Table 2.21 Methods of protease digestion reported for collagen-like proteins

Enzyme	Protein: Enzyme Ratio	pH	Incubation Temperature	Incubation Time	Inhibition Method	Reference
Trypsin	10 µg/mL	8	4 °C	24 h	NA	Peng et al. (2015)
	1:100	NA	Room temp.	1 hour	PMSF	Xu et al. (2010)
	10 µg/mL	8	15 °C	1 hour	PMSF	Xu et al. (2002)
	15 µg:15 ng	7.4	10-35 °C	2 hours	2X Laemmli Buffer	Rutschmann et al. (2014)
Chymotrypsin	10 µg/mL	8	4 °C	24 h	NA	Peng et al. (2014b)
Pepsin	10 µg/mL	2.2	4 °C	16 h	Adjust pH to 7	Peng et al. (2014a), Peng et al. (2014b), and Peng et al. (2015).
Papain	10 µg/mL	6.5	4 °C	24 h	NA	Peng et al. (2015) and Peng et al. (2014b).
Bromelain	NA	NA	NA	NA	NA	Peng et al. (2014b)
Ficin	10 µg/mL	NA	NA	NA	NA	Peng et al. (2014b)

NA indicates that the information is not available or not reported in the publication(s).

2.8.3 Western Blotting

Proteins that have been resolved on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE) gels can be adsorbed onto microporous membranes by use of an electric current and this is termed western blotting (WB). This powerful approach allowed the detection and characterization of numerous proteins, especially those that are present in small quantities. Various methods have been used for detecting proteins on membranes including the use of Congo red dye along with alkaline phosphatase- and horseradish peroxidase-antibody conjugates (Kurien & Scofield, 2012).

2.9 Research Gap Analysis

While significant progress has been made in the area of recombinant collagen production, several gaps exist that impede further advancement of RPCLP production. A critical evaluation of existing literature reveals sevearl areas where knowledge is incomplete, inconsistent, or lacking altogether. This subsection identifies and discusses these gaps, thereby establishing the foundation for the objectives and highlighting the importance of this study.

Although bacterial collagen-like proteins (CLPs) have emerged as promising alternatives to animal-derived collagens, several gaps remain in the current literature regarding *Rhodopseudomonas palustris* collagen-like protein (RPCLP):

i. While the demand for safe, sustainable, and halal collagen sources is increasing, scalable bioprocesses for producing such CLPs, including RPCLP, remain underdeveloped or unexplored.
ii. There have been no published efforts to optimize culture media specifically for enhancing RPCLP production, despite the known impact of media composition and optimization on recombinant protein yields.
iii. There is limited research and data on the expression, yield, and downstream processing of RPCLP in recombinant systems, with most studies focusing on CLPs from *Streptococcus* or *Clostridium* species.
iv. Kinetic models describing the fermentation and RPCLP production dynamics in *E. coli* are absent, limiting understanding of biomass, substrate, and product relationships during fermentation.
v. No scale-up studies for bacterial collagen-like protein production exist; this is a novel area needing focused research. More specifically, no existing studies report the bioprocess scale-up of RPCLP production in bioreactor systems using *E. coli*, particularly at working volumes such as 2 L or 7.5 L.
vi. Comparative studies evaluating different bioreactor scale-up strategies (e.g., via constant oxygen transfer coefficient or impeller tip speed) for RPCLP production have not been conducted.

2.10 SUMMARY

Research interest in recombinant collagen has grown tremendously over the past three decades, and many key contributions have been made that expand our understanding of recombinant collagen and CLPs. Simultaneously, efforts were made to produce and

purify these proteins and apply them in the development of biomaterials as well as other fields. The principal milestones can be considered to have been:

i. Discovery of collagen-like sequences in bacteria,
ii. Recombinant expression of various constructs of some of the said proteins in E. coli and their characterization (protease digestion, circular dichroism, isoelectric point, pH stability, and solubility),
iii. Identification of challenges in recombinant collagen production that can be addressed by the use of these bacterial collagens and attempts to overcome them (monomers to dimers, hydroxyprolination, hydroxylysination, etc.),
iv. Recombinant protein recovery and purification method improvement,
v. Large-scale production (2 L culture in the bioreactor),
vi. Characterization to understand their mechanism of stability and properties (mostly carried out on Scl1 and Scl2 proteins from S. pyogenes), and
vii. Identification of motifs and application studies (wound healing, cytotoxicity, etc.) and designs of novel scaffolds and biomaterials.

This chapter attempted to provide a comprehensive review of available studies, data, and methodologies in relation to bacterial collagen-like protein production, to explain the selection and rationale of the methodology and techniques that are detailed in Chapter 3. By addressing several critical gaps, the thesis establishes a framework for the sustainable, cost-effective, and scalable production of RPCLPs for future functional applications. These critical gaps are:

i. First demonstration of RPCLP production in *E. coli* at bioreactor scale, screening and optimization of the process to maximize RPCLP yield,
ii. First medium optimization study for RPCLP,
iii. First kinetic modeling and scale-up comparison for this system.

CHAPTER THREE
MATERIALS AND METHODS

3.1 INTRODUCTION

This chapter describes the methodology that was followed to realize the objectives of this research. To achieve the first objective, One-Factor-at-a-Time (OFAT) was carried out to optimize the medium component. Then, the screening of process and induction parameters, designed by Design-Expert software version 13, was carried out to obtain the significant factors for optimization of RPCLP concentration, and further optimization was carried out until it was successful. Cultivation conditions and their ranges were selected based on available relevant literature on recombinant protein expression unless otherwise explained. For the third objective, a kinetic study and modeling were carried out. The scale-up processes represent the fourth objective of this study. Finally, the characterization of the recombinant protein was carried out, as the fifth objective.

3.2 FLOWCHART

The flowchart of the methodology is illustrated in Figure 3.1 and is explained in detail in the subsequent sections of this chapter.

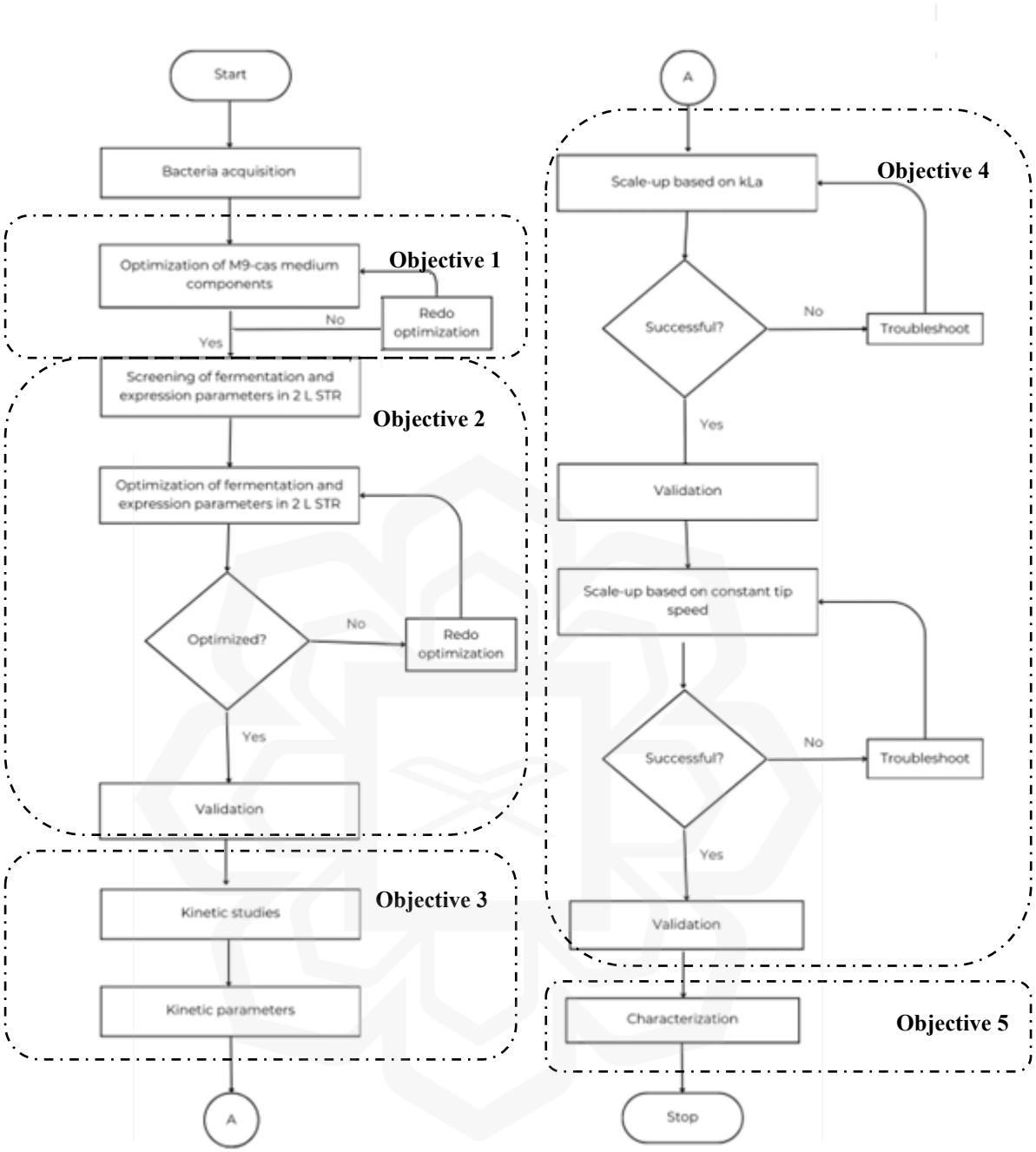


Figure 3.1 Overall flowchart of the methodology

3.3 MATERIALS AND EQUIPMENT

The methodology for this study involved fermentation of recombinant *E. coli* at the lab scale (2 L bioreactor) and then scale-up to a 7.5 L bioreactor (Labfors, Infors-HT Inc., Bottmingen-Basel, Switzerland). Recombinant *E. coli* strain BL21(DE3), with the DNA encoding for the CLP sequence, was grown in M9-casamino acid medium (constituent chemicals listed in Appendix I with manufacturer) with 100 µg/mL ampicillin as an antibiotic to select the plasmid-bearing *E. coli* cells (Chan et al., 2010). M9-casamino acid medium was selected as the medium for the shake flask culture based on its success as a medium for producing various CLPs using the *E. coli* BL21 strain (Xu et al., 2010). The medium components were optimized by OFAT in this study, on the basis of medium screening results obtained by a fellow researcher, Fatemeh Soroodi, from the same laboratory and CLP research project at IIUM. Cold-shock induction using the pColdII vector (Takara Bio) was applied. The expression of the recombinant protein was induced using isopropyl β-D-thiogalactopyranoside (IPTG). The full list of materials and equipment used is provided in Appendices I-II.

3.4 ACQUISITION OF RECOMBINANT *E. coli*

3.4.1 Design of the DNA Construct

The genetic engineering component of this study was carried out by Soroodi (unpublished). Briefly, the sequence for the RPCLP DNA fragment was obtained from the NCBI website (NC_011004.1), and the desired construct was synthesized by Genscript (Hong Kong) in the pUC57 plasmid. The DNA construct is depicted in Figure 3.2.

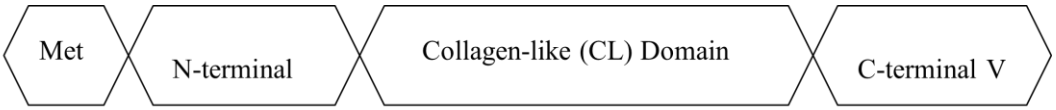


Figure 3.2 The construct made up of a collagen-like (CL) domain and a variable domain (V)

Recombinant DNA sequence:

This DNA sequence codes for a 212-amino acid sequence that corresponds to a molecular weight of 22.08 kDa (Xu et al., 2010). The sequence is as shown.

ATGGCTGAAGCTGAATCCACCCGCCGCGGACGTCCGGGGCCGCAGGGCCCG
CGCGGACGTCCCGGAGAACCGGGCCGTCCGGGACCGCAAGGACATCCGGG
CCGTCCCGGCCCCGAGGGCCCCCGAGGCAAGCAAGGTCCGGTCGGCAAGCC
CGGCCCGCAGGGCAAGGCCGGTCCGCAAGGCAAACCCGGCATCGCCGGCA
AGCCGGGCCCCGATGGCAAGCCCGGTCCGATCGGCCCGCAAGGCAAGGCC
GGCCCGCAAGGTCCCCGCGGTGAACAAGGCCTGCGCGGCGAGCAAGGTCCT
CGCGGTGAGCAAGGCCCGCAGGGGCCGCGTGGCGAACAGGGCCCCCGGGG
CGAGCCCGGCCCCGCCGGGGCGCTCCCCTCGATCGAGCAGGTGATGCCCTG
GCTGCACCTGATCTTCGACGCCTACGAAGATTACAAAGCGCAGCGCGCCCG
CGAAGCCCGCGAGCTCGAAGAGCGTCTCGCCGCCGAAGCTCTCGAACAGGC
GGCGCGCGAAGCCGCCGAGCGCGAAGTCGCCGCCGCTATCGAAGCCGCAA
ATGCCGAGGCCGAGATCATGCTCGACGATGAGACGCATGCCGAGGGCGGC
AAGAAAAAGAAGAAGCGCAAGCACAAGGACTGA

```
        10         20         30         40         50         60
MAEAESTRRG RPGPQGPRGR PGEPGRPGPQ GHPGRPGPEG PRGKQGPVGK PGPQGKAGPQ

        70         80         90        100        110        120
GKPGIAGKPG PDGKPGPIGP QGKAGPQGPR GEQGLRGEQG PRGEQGPQGP RGEQGPRGEP

       130        140        150        160        170        180
GPAGALPSIE QVMPWLHLIF DAYEDYKAQR AREARELEER LAAEALEQAA REAAEREVAA

       190        200        210
AIEAANAEAE IMLDDETHAE GGKKKKKRKH KD
```

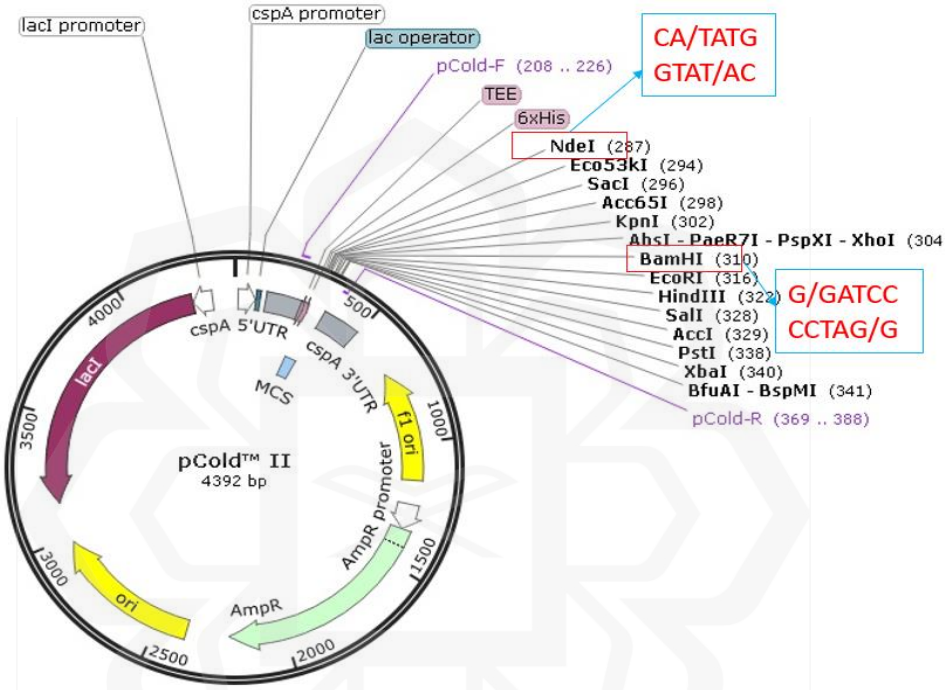


Figure 3.3 Plasmid vector map for pCold II (SnapGene 7.2.1)

The DNA construct in pUC57 was transferred into pColdII for expression in *E. coli* BL21(DE3) strain via several steps. First, the pUC57 was transformed into *E. coli* DH5α. To amplify the quantity of vectors, the clones were grown in a 100 mL culture medium, and the plasmids were extracted and purified using Qiagen kits (Germantown, US). After the plasmids were purified, the pUC57 and empty pColdII vectors underwent

restriction enzyme double digestion, gel extraction, and DNA ligation to sub-clone the construct into the pColdII vector containing an N-terminal His_6 tag (Takara Bio, Shiga, Japan), using 5′ *Nde*I and 3′ *Bam*HI restriction sites. Positive clones were confirmed via polymerase chain reaction (PCR). Next, a selected positive clone was transformed into *E. coli* BL21(DE3) cells (New England Biolabs) and, after verification, these cells were cultured and used to prepare glycerol stocks for storage at -80 °C. A batch of these glycerol stocks was acquired to fulfill the objectives of this research.

DNA sequencing was also carried out to confirm the presence of the correct insert of the RPCLP gene within the recombinant pColdII vector (Soroodi, unpublished). For this, 2 µL of forward and reverse primers for *R. palustris*, as well as the recombinant pColdII, were sent to 1st Base Asia for DNA sequencing. The method utilized was Sanger sequencing. The generated sequence was then compared against the whole *R. palustris* genome available on the NCBI database the identity percentage was obtained. The nucleotide sequence alignment between the complete genome of *R. palustris* (Subject) from NCBI (NC_011004.1) and the generated sequence (Query) is shown in Appendix III. The identity percentage between the Query and Subject value of 100% indicates that the transformation was successful.

In addition to DNA sequencing, the correct reading frame (ORF) of the insert was verified by using the NCBI ORF Finder tool, translation, and checking the amino acid sequence. The full sequence, including vector backbone and fusion tag regions, was submitted. The tool identified six (6) possible ORFs corresponding to the forward and reverse reading frames. The correct ORF was then determined based on the position on the forward strand, the presence of the expected ATG start codon, the predicted amino acid length (212 bp), and the His-6 tag. The translated amino acid sequence of the selected ORF matched the expected sequence via the BLAST. This confirmed that

the insert was in the correct reading frame, without premature stop codons (TAA, TAG, TGA), and properly fused with the His- tag.

3.4.2 Colony PCR and Agarose Gel Electrophoresis

Colony PCR, a quick method to determine the presence or absence of the inserted DNA into plasmids in bacterial colonies, was used to amplify the recombinant DNA-containing *E. coli*. The primers, shown in Table 3.1, were designed manually, and the Multiple Primer Analyzer (www.thermofisher.com) online tool was used to verify the primers. The reaction mixture composition is detailed in Table 3.2. The optimized conditions for the PCR, developed by Soroodi (unpublished), were as in Table 3.3.

Table 3.1 Design of Primers

	Sequence	**T_m °C**	**CG %**	**A**	**T**	**C**	**G**	**Extinction coefficient (1/(mol.cm)**	**Molecular weight (g/mol)**	**nmol**	**µg / OD_{260}**
FW	catatggctgaagctgaatccaccc	72	52	7	5	8	5	236600	7611	4.2	32.2
RE	cctcagtccttgtgcttgcgcttct	74	56	1	10	9	5	207800	7541.9	4.8	36.3

Source: Soroodi (Unpublished).

Table 3.2 PCR reaction mixture components

Component	Volume	Final Concentration
Forward primer	5 μL	1 μM
Reverse primer	5 μL	1 μM
Distilled water	15 μL	-
Master mix	25 μL	
Sample	(1 colony/1 μL)	-
Total	50 μL	

Table 3.3 Optimized PCR conditions

Step	Process	Temperature (°C)	Time (min)
1	Initial denaturation	95	3:00
2	Denaturation	95	0:30
3	Annealing	66	0:30
4	Extension	72	1:00
5	Cycle	Go to step 3, 30X	-
6	Final extension	72	7:00
7	Storage	4	∞

Total PCR time: approximately 1 hour, 20 minutes

Source: Soroodi, unpublished.

Agarose gel electrophoresis was used to detect the presence of the recombinant collagen-like DNA in the cell lysate after transformation and polymerase chain reaction. A 1 % agarose gel containing 1 μL Florosafe dye (1st Base, Axil Scientific, Singapore)

was prepared. Next, a pre-stained DNA ladder (2.5 µL) was loaded into the first well (1st Base, Axil Scientific, Singapore). Then, 7 µL of each sample was loaded into the other 1-mm wells. The agarose gel electrophoresis was then carried out in a Bioer Mini GE-100 system using 0.5x Tris/Borate/EDTA (TBE) buffer (Axil Scientific, Singapore). The agarose gel electrophoresis conditions were set as follows: 220 V, and 25 minutes. *E. coli* cell lysates from cell cultures that do not contain the recombinant collagen-like DNA were used as a control.

3.4.3 Preparation of Glycerol Stocks

To prepare glycerol stocks, 500 µL of the overnight culture was added to 500 µL of 50% glycerol in a 2 mL screw top tube or cryovial and gently mixed. The labeled glycerol stock vials were then immediately frozen at -80 °C.

3.4.4 GMO Ethics Approval and Safety

After examination of relevant documentation and careful consideration by the IIUM Biosafety and Biosecurity Committee (IBBC), this project has been exempted from submission to the Department of Biosafety Malaysia. However, all investigators involved with the project were required to undergo Biorisk Management Training to ensure appropriate measures were taken in the laboratory during handling and disposal of the recombinant bacteria.

3.4.5 Shake Flask Fermentation Techniques

3.4.5.1 Ampicillin-selective Culture Plates

Recombinant *E. coli* strain BL21(DE3) glycerol stocks were used to inoculate culture plates containing LB Vegitone (Sigma Aldrich) agar (35 g/L) with 50 μL/ml ampicillin for positive selection of plasmid-containing cells. The plates were incubated overnight, and the colonies were later used to prepare overnight flask cultures.

3.4.5.2 Overnight Cultures for Inoculation

Lysogeny broth (LB) is a rich medium that is mainly used for bacterial culture. Disposable inoculum loops were used to transfer colonies from the agar plates into the fresh LB Vegitone medium (20 g/L, Sigma Aldrich) containing 50 μL/mL ampicillin. Overnight cultures shake flasks. at 28 ℃ temperature and 160 rpm agitation, for 14-16 hours. The OD_{600} of the culture was measured before inoculation for shake flask experiments. For visibly turbid cultures, the OD_{600} of the sample's 10-fold dilution was measured. Cultures with optical density values between 3 and 5 were used for the experiments (Elbing & Brent, 2019).

3.4.5.3 Sample Analyses

Hourly sampling was conducted for each flask, taking 100 µL of each flask culture into a cuvette containing 900 µL medium. The absorbance at 600 nm wavelength (OD_{600}) was then determined using a Biophotometer, considering the 10 × dilution where necessary, and recorded. The OD_{600} was converted to dry cell weight (g/L) based on a standard curve given in Appendix VII (Hajihassan et al., 2018). The natural logarithm of biomass concentration/OD_{600} was plotted against time (ln X vs time). The specific growth rate (μ) was calculated from the slope of this growth profile. The doubling time (t_d) was obtained by using Equation 3.1.

$$t_d = \frac{\ln 2}{\mu} \qquad \text{Equation 3.1}$$

3.5 RECOMBINANT PROTEIN SEPARATION AND PURIFICATION TECHNIQUES

3.5.1 Cell Harvest and Crude Protein Recovery

The product was obtained by mechanical methods of disruption (sonication) following cell harvest via centrifugation (Ahmad-Raus et al., 2010). The cells were first pelleted at 4696 x g by means of refrigerated centrifugation (Heraeus Multifuge X1R, Thermo Scientific). The supernatant (spent media) was then discarded, and the cell pellets were stored at -20 °C (Jamaluddin, 2015).

Frozen cell pellets stored at -20 °C were gently allowed to thaw to room temperature. These were then lysed using a probe-type sonicator (230 V 50 Hz, Labsonic M, Sartorius). First, lysis buffer (20 mM sodium phosphate, pH 7.4, 0.5 M sodium chloride, chilled) was used to resuspend the cell pellets at a ratio of 3 mL buffer per gram of wet cell weight (Xu et al., 2010). A five-minute burst cycle pulsating every 0.5 seconds at 40% amplitude was applied to the sample which is kept on ice throughout the process. The lysed cells were then clarified via centrifugation to separate cell debris (Heraeus Multifuge X1R, Thermo Scientific, USA). Tables 3.4 and 3.5 summarize the centrifugation settings and the sonication settings, respectively. Finally, the supernatant (cell lysates) was placed into either a sterile centrifuge tube before further analysis and chilled or is stored at -20 °C (Jamaluddin, 2015).

Table 3.4 Centrifugation parameters and set points

Centrifugation Parameter	**Setpoint**
Speed	4696 × g (to harvest cells)
	12, 000 rpm (to clarify samples)
Temperature	4 °C
Time	30 minutes

Table 3.5 Sonication parameters and set points

Sonication Parameter	Setpoint
Amplitude	40% (230 V, 50 Hz)
Mode	Burst cycling (0.5)
Time	5 minutes

3.5.2 Protein Sample Desalting and Buffer Exchange

Sample desalting and/or buffer exchange using a 5-mL HiTrap column FPLC (Cytiva, previously GE Healthcare) and the AKTA PrimePlus Fast Protein Liquid Chromatography system (GE Healthcare)'s HiTrap Desalting template. The buffer preparation and entire HiTrap Desalting protocol was following the AKTA Fast Protein Liquid Chromatography system user manual.

3.5.3 Protein Concentration Technique

Protein concentration was carried out using VivaSpin (Cytiva, previously known as GE Healthcare) centrifugal concentrators. Vivaspin concentrators are disposable ultrafiltration devices that can be used in sample concentration, desalting, and buffer exchange. An upper chamber for the sample is separated from the lower compartment by a semipermeable membrane to size-exclude the target molecule. Centrifugation then

forces the solvent (liquid) to go through the membrane and into the lower chamber. The target molecules are retained and do not pass through the membrane.

To use the Vivaspin concentrator, the most appropriate membrane cut-off size (MWCO) was determined to be 10 kDa as the manufacturer recommends an MWCO at least 50% smaller than the molecular size of the species of interest for maximum recovery. The ultrafiltration spin column concentrator (6 mL, PES, MWCO 10 kDa, Cytiva) was filled with up to maximum volumes (as recommended by the manufacturer). After ensuring that the lid was fully seated, the fully assembled concentrator was inserted into the appropriate centrifuge and the centrifugation speed was set as recommended. The centrifugation time was set, based on typical values for various proteins and concentrators recorded in the user manual. After the samples were concentrated, the sample was recovered from the bottom of the concentrate pocket with a pipette. The filtrate tube was then sealed and stored.

3.5.4 Affinity Chromatography (Small-scale Purification)

The presence of the N-terminal His_6 tag in the construct is useful for the purification of the recombinant CLP via metal affinity chromatography. The recombinant CLP was purified by using the Histidine-tag purification application template of the Akta PrimePlus Fast Protein Liquid Chromatography (FPLC) system, monitored using Prime View software (GE Healthcare Bioscience, Canada). A 1-mL HisTrap HP column prepacked with pre-charged Ni2+ Sepharose High Performance resin (Cytiva) was used. As the protein samples had a 6xHis-tag, the clarified supernatant was taken to the

binding buffer. Imidazole is utilized in the elution buffer as a competitor for the elution of histidine-tagged proteins. Imidazole is often added, in low concentrations, to the sample and binding buffers to reduce the binding of unwanted proteins, and thus increase the purity (Cytiva, previously GE Healthcare). The binding buffer used was 20 mM sodium phosphate, 0.5 M sodium chloride, and 5 mM imidazole, pH 7.4). The elution buffer was made up of 20 mM sodium phosphate, 0.5 M sodium chloride, and 0.5 M imidazole solution, pH 7.4 (Xu et al., 2010). All buffers were filter sterilized and degassed before use.

After washing and priming the system and column according to the AKTA Fast Protein Liquid Chromatography system user manual, the system was set to His Tag Purification HisTrap in the Template mode and then a sample volume of 2 mL was injected at the injection valve and the purification was allowed to run until complete. Prime View software was used to detect the fractions containing the His-tagged protein of interest. The fractions containing the eluted His-tagged recombinant CLP were collected for further analysis.

3.5.5 Total Soluble Protein Assay (Bradford Assay)

Total soluble protein concentration from the crude samples was determined, by use of the Quickstart Bradford Protein Assay (Biorad, USA) which was carried out at room temperature (Bradford, 1976). Clarified cell lysate (5 µL) was carefully added to the protein assay reagent (250 µL), in a 96-well plate, and they were mixed well. A blank for the protein assay was prepared by adding lysis buffer (5 µL) to the reagent dye (250 µL) (Jamaluddin, 2015). The experiment was also repeated using 2 X dilutions of the cell lysate. The protein concentration was estimated at 595 nm optical density using Multiskan™ GO spectrophotometer (Thermo Scientific, USA) after incubation for ten minutes at room temperature. Readings were done in triplicates and average readings were used. Standards of Bovine serum albumin (BSA) (Pierce Thermo Scientific) ranging from 0.25 to 1 mg/mL were used to plot a linear standard curve that was used to estimate the protein concentration. The standard curve obtained is presented in Appendix VII.

3.6 RECOMBINANT PROTEIN DETECTION AND QUANTIFICATION

Denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of the cell lysate was used to detect the presence of the recombinant CLPs after cell harvest. Detection of the product and its quantification were carried out via the glycine-based, fluorometric collagen assay as described by Yasmin and colleagues (Yasmin et al., 2014).

3.6.1 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

The protein-containing samples were diluted twice before loading samples for SDS-PAGE. Mini gels consisting of 12% resolving gel and 4% stacking gel were prepared for the experiments. To prepare the sample, the protein sample or cell lysates, and the 2X Laemmli sample buffer (containing 5% v/v β-mercaptoethanol) (Bio-Rad Laboratories Inc., USA) were mixed at a ratio of 1:1. This mixture was denatured for 5 minutes in a water bath at 95 °C (Xu et al., 2002). A 10 µL sample was loaded into each well of the gel. For the protein ladder, a 2 µL pre-stained protein marker was loaded into a well (Peacock Pre-stained Protein Marker, PM2610, Biotium). The polyacrylamide gel electrophoresis was then carried out using 1X Tris/Glycine/SDS-buffer (Bio-Rad Laboratories Inc., USA) in the Mini-Protean system (Bio-Rad Laboratories Inc., USA). The power settings were as follows: constant voltage, 120 V voltage, 400 mA current, and 60 minutes run duration. Gel staining was done using a Ready Blue Protein gel stain (Sigma-Aldrich) for 30 minutes followed by de-staining with double-distilled water, thrice, for 5 minutes each time, with mild agitation (30 rpm) on a rocker (Stuart Scientific, UK). In some experiments, the gel was first fixed and stained using Ready Blue Protein Gel stain (Sigma Aldrich) or Coomassie Blue stain (R250). Then it was destained using a fixing solution of 10% acetic acid, 50% methanol, and 40% water (Chan). The protein bands were visualized using a Gel Documentation system (Biorad, USA).

ImageJ software (NIH) was used to carry out densitometric analysis on the gel image to quantify the protein content of the band. First, the total density of the molecular weight marker lane was measured using the image analysis software, and then the total density of the other lanes present in the SDS gel was noted. For each lane, the gel background was subtracted. Then, the corrected density of the protein band was divided

by the corrected density for the entire lane and multiplied by 100% (Alonso Villela et al., 2020).

3.6.2 Western Blotting

In a separate experiment, after the SDS-PAGE run, the unstained, electrophoresed gel bands were immediately transferred onto a nitrocellulose membrane using 100 V, 700 mA, run for 60 minutes, at 4 °C, using Biorad equipment (Biorad, USA). Then, the membranes were treated with 50 mL blocking buffer (5% skim milk (analytical grade, Merck)) in Tris Buffered Saline Tween 20 transfer buffer (TBST) for 60 minutes, at room temperature, with constant rocking. The membrane was rinsed in TBST and then equilibrated in the blocking buffer for 15 minutes with constant rocking. The membrane was then incubated in a blocking buffer containing mouse anti-hexahistidine tag primary antibody (GeneTex, 1:5,000) at 4 °C temperature for an hour with constant rocking. After several washes with TBST buffer, it was incubated in a blocking buffer containing the secondary antibody, goat anti-mouse-HRP conjugated antibody (GeneTex, 1:1,000) at 4 °C temperature for an hour with shaking. To visualize blots, a fresh working solution was prepared by mixing equal volumes (900 µL) of 3,3'-diaminobenzidine (DAB) stain stock solution, imidazole buffer solution, and substrate (hydrogen peroxide) in 40 mL deionized water (peroxidase DAB stain kit, Nacalai Tesque, Japan). The blot was incubated with the working solution after washing with TBST, for an hour at ambient temperature or until the image develops. The brown, stained blot was rinsed under running water to wash off excess detection reagent and decrease background.

3.6.3 Fluorescence Detection of Collagens using DHPAA

A rapid method is required to verify the expression of the desired protein and quantify it. As the CLPs do not contain the unique hydroxylysine and hydroxyproline residues that distinguish collagens from other proteins, existing collagen assays such as the Sirius Red or the Fast Green assay cannot be implemented in this study. Hence, an easier way, besides SDS-PAGE, was sought for confirmation. The method of Yasmin et al. (2014), which is a selective assay that uses enzymatic reactions (bacterial collagenase) and uses a novel fluorogenic reagent (3, 4- dihydroxyphenylacetic acid) that reacts with N-terminal Glycine-containing peptides was utilized.

Bacterial collagenase and all other chemicals for the assay were obtained from Merck, except for 3, 4-dihydroxyphenylacetic acid, and borate buffered saline which were obtained from Sigma. Briefly, the dye solution was prepared (0.75 mM 3,4-DHPAA in water) and aliquoted into 1 mL tubes. The sodium periodate solution (1.25 mM sodium periodate in water) was also prepared separately and aliquoted into 1 mL tubes. Bacterial collagenase was dissolved in 50 mM Tris buffer, pH 7.5 with 5 mM calcium chloride, to a concentration of 1.0 mg/mL, and stored at -20 ℃ until use. To dilute the enzyme solution, this buffer was used in a later step, and the enzyme solution was then aliquoted into 1 mL tubes. A 100 µL mixture of 125 mM sodium borate buffer (pH 7.5) and 5 mM calcium chloride, and 30 µL of water was prepared and aliquoted into 50 mL tubes. Sample solutions (100 µL) were prepared using the cell lysates. To plot the standard curve, a stock standard solution of collagen (3 mg/mL) was first prepared. Next, a diluted stock is prepared by mixing 5 µl of the collagen stock solution and 295 µL distilled water to get a diluted stock with a concentration of 50 µg/mL.

Collagen standards of different concentrations (0, 15, 30, 50 µg/mL) were then prepared in PCR tubes and labeled carefully. The digest reaction mixture (labeled as digest mix) was prepared, comprising the buffer, distilled water, and enzyme solution at a ratio of 10 µL borate buffer: 2 µL enzyme: 3 µL water. Figure 3.4 shows the subsequent flow of steps.

The sample solution (10 µL) of standard collagen or cell lysates was mixed with 15 µL of the digest mix. Sample blank wells contained collagen standards (10 µL) and buffer (15 µL) only, therefore they lacked the digest mix. The resultant mixture (25 µL) was incubated at 37 ℃ for 60 minutes. Then, 25 µL of 0.75 mM 3,4-DHPAA in water, 25 µl of 125 mM sodium borate (pH 8.0), and 25 µL of 1.25 mM of sodium periodate in water were added to it, and the resultant mixture (100 µL) was incubated at 37 ℃ for 10 min. Samples were transferred carefully to their respective wells in a black 384-well, flat-bottomed plate (Corning). All wells were designed in triplicates and average readings were used. The fluorescence spectrophotometer was set to the correct plate, mode, and wavelengths (Varioskan LUX Multimode Reader, Thermo Scientific, USA). The fluorescence intensity of each well was measured using a spectrofluorometer at $\lambda_{excitation/emission}$ = 375/465 nm (SkanIt Software version 6.1.1). The diagram in Figure 3.5 summarizes the detection steps.

The results of the standard collagen samples at various concentrations, after subtracting the sample blanks, were used to develop a collagen standard curve (ΔFL against standard concentrations), as shown in Appendix VII. When the sample fluorescence value was higher than the fluorescence reading for the 50 µg/mL (equivalent to 5 mg/dL, or 50 ppm) collagen standard, the sample was diluted in water and the assay was repeated.

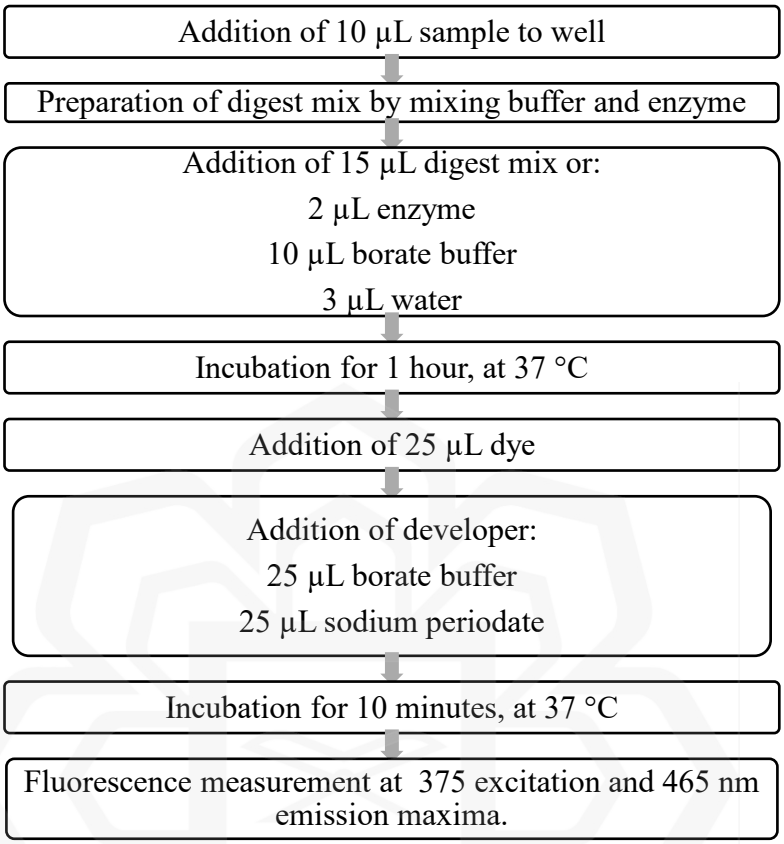


Figure 3.4 Collagen assay method for a 384-well plate assay based on Yasmin et al. (2014)

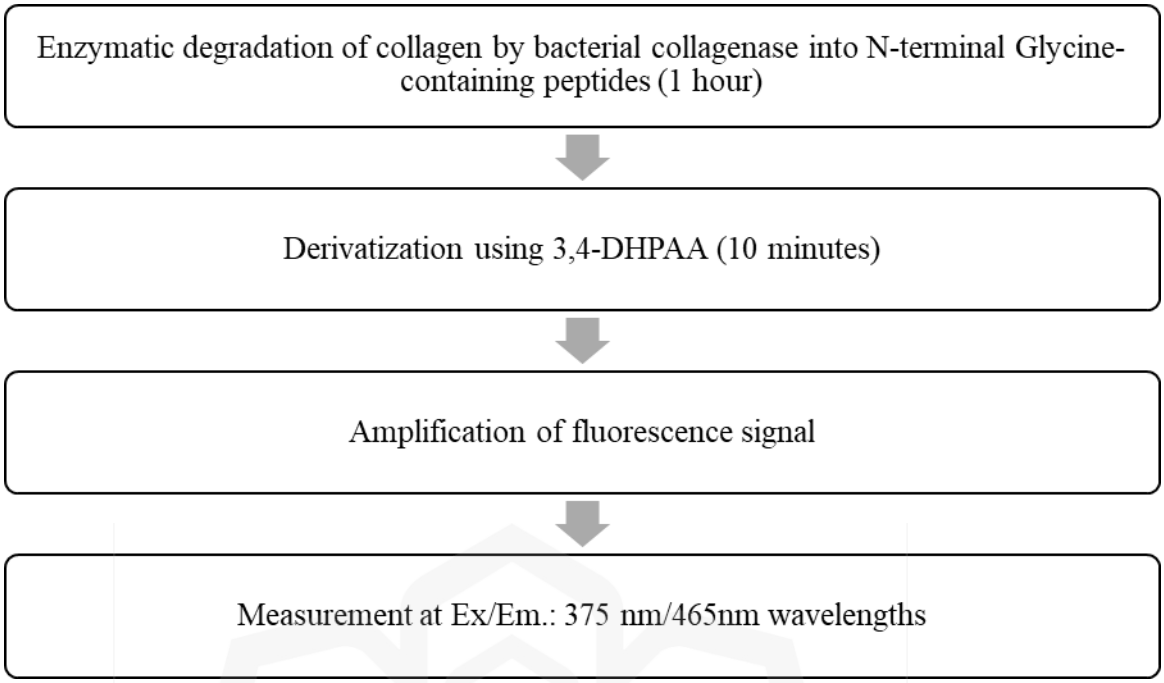


Figure 3.5 Stages of the recombinant collagen detection assay based on Yasmin et al. (2014)

3.7 OBJECTIVE 1: OPTIMIZATION OF MEDIUM COMPONENTS

As mentioned in the previous chapter, there is a need to optimize the mineral as well as organic medium components of M9-casamino acid media to produce the recombinant CLP. This was accomplished by modifying one factor at a time while keeping the others constant. Table 3.6 shows the OFAT design. Fifteen experimental runs, with triplicates for each run, were performed in shake flask culture. A 2.5 mL volume of seed culture with an OD_{600} of 3-5 was used to inoculate 50 mL M9-casamino acid medium that has been supplemented with ampicillin (50 µL/mL) in 250-mL shake flasks (unbaffled) (Gameil et al., 2023). The shake flask cultures (initial OD_{600} ~0.25 to 0.3) were grown at 37 °C temperature and 200 rpm speed in an incubator shaker. Sampling was done

every 20 minutes for the first 2 hours or until induction, and then hourly for 12 hours. The outputs studied were cell OD_{600}, total protein using the Bradford assay, and SDS-PAGE (Gameil et al., 2023). Induction was carried out at OD_{600} value between 0.8 – 1 (around 2 hours after inoculation), by reducing the temperature to 25 ℃, and setting the agitation speed to 220 rpm, before the addition of 1 M IPTG solution (42.5 µL) for a final IPTG concentration of 1 mM (Gameil et al., 2023). The lengths of the pre-induction phase and the induction phase were two (2) hours at 37 ℃ and 24 hours at 25 ℃, respectively (Peng et al., 2012).

Table 3.6 Design for OFAT optimization for M9-casamino acids media

Varied Parameter	**Levels**	**Fixed Parameters**
Disodium hydrogen phosphate concentration final concentration in M9 salt	0.0704 M, 0.14 M, and 0.2113 M.	Glucose concentration: 0.5% Casamino acids concentration: 0.5%
Glucose concentration final concentration in M9 medium	1%, 2.5%, 5%	Disodium hydrogen phosphate concentration: 0.4 M Casamino acids concentration: 0.5%
Casamino acids concentration in M9 medium	0.5%, 0.75%, 1%	Disodium hydrogen phosphate concentration: 0.4 M Glucose concentration: 1%

For the induction of protein expression, the cells were cultured in LB Vegitone Media with ampicillin (50 μg/mL) at 37 °C and 200 rpm speed for 1 to 2.5 hours until the optical density reached a value between 0.8 and 1 (Elbing & Brent, 2019; Xu et al., 2010). The temperature of the culture was then reduced to 20 °C and 1 mM IPTG was added to induce the RPCLP expression, based on the induction methods of Peng et al. (2014b). IPTG serves as an inducer by mimicking lactose which induces the transcription of the lac operon by binding to the lac repressor, allowing T7 RNA polymerase to bind to the lac operon (lacO), thereby inducing transcription of the gene that is under the control of the lac operon. After 22 hours incubation at 20 °C, the cells were harvested by centrifugation. Colony PCR followed by agarose gel electrophoresis was run again to detect the presence of positive clones at the end of the fermentation. BL21(DE3) cells that did not contain the vector were cultured and used as negative controls.

3.8 BIOREACTOR FERMENTATION AND PROTEIN EXPRESSION

For the fermentation of the recombinant *E. coli* and expression of the CLPs in stirred tank bioreactors, a 2 L stirred tank bioreactor (Labfors 4, Infors-HT Inc., Bottmingen, Switzerland) was employed, using a working volume of 1.0 L of M9-cas medium containing 100 μL/mL ampicillin, and a 5% inoculum size. Glucose (10 g/L) was used as the carbon source. Foaming was controlled via the addition of 10% (v/v) SE15 antifoam (Sigma Aldrich). The inoculum was added to the bioreactor to attain an OD_{600} value of 0.3. After inoculation, fermentation was maintained at 35 °C, 1 vvm, and 300 rpm, until the culture almost reaches the mid-log phase. Then, the culture was cooled to 25 °C to activate the cold shock component of the vector and induce protein expression

using cold-shock and 1 mM IPTG (Peng et al., 2012). After seven hours of cultivation at 25 °C, the cells were harvested by centrifugation.

3.8.1 Bioreactor System, Components, and Preparation

The bioreactor system consisted of a double-jacketed borosilicate, cylindrical, round-bottom, glass vessel. The agitation was achieved via a top-driven motor and mechanically sealed agitator shaft where two six-bladed Rushton impellers are mounted. The vessel was equipped with three baffles, an air pipe, a Pt-100 temperature sensor, pH, and polarographic pO_2 probes as well as four peristaltic pumps for the addition of media, an acid (20% phosphoric acid) and a base (10% ammonia solution) (for pH regulation), and an antifoam solution. Online measurement of pH was accomplished by using a sterilizable glass pH electrode (Mettler-Toledo, Switzerland). Online measurement of dissolved oxygen was accomplished by means of a sterilizable membrane-type pO_2 electrode (Mettler Toledo, Switzerland). Air was supplied via a gas mixer connected to a rotameter and air compressor and a cylinder of purified oxygen gas (99.7%).

Figure 3.6 Setup of the 2 L bioreactor, Labfors, INFORS HT, Switzerland

Figure 3.7 Setup of the 7.5 L bioreactor, Labfors, INFORS HT, Switzerland

3.8.1.1 Sterilization of the Vessel, Calibration, and Media Preparation

The bioreactors and accessories used are shown in Figures 3.6 and 3.7, and Appendix VI. The pH electrode was first calibrated at two points (pH 4.0 and 7.0) by using two standard buffers (Fisher Scientific, UK). These electrodes were then placed in their appropriate slots in the vessel besides other bioreactor accessories including a Super Safe sampler (Infors-HT Inc., Bottmingen, Switzerland). Membrane filters (0.2 μM, Pall, USA) were appropriately attached at the gas inlet and exhaust points, and they were replaced periodically to ensure sterile conditions. Similarly, the acid, base, and antifoam reservoir outlet tubing were attached with 0.45 μM membrane filters (Pall, USA). All the tubing was then closed using clips to avoid spillage of the medium as a result of built-up pressure build-up in the autoclave. All exposed openings of the port plate and the membrane filters were loosely plugged with non-absorbent cotton wool and then wrapped with aluminum foil. The exhaust outlet (fitted with a filter) was left unclamped.

After preparing the vessel, the fermentation media was poured into the vessel. For a 1.0 L working volume, the media was prepared as in Appendix V. For the 7.5 L bioreactor, a 5.0 L working volume was used. The 250 mL reservoir bottles for acid and base were filled with 50 mL distilled water. On the other hand, the antifoam bottle was filled with 10% (v/v) SE-15 anti-foaming agent (Sigma-Aldrich, USA). Then, the bioreactor vessel was sterilized in an autoclave at 121°C.

Once the sterilization was completed, the vessel was left to cool, and then the heat-sensitive medium components such as glucose, thiamine, and ampicillin were

aseptically added into the bioreactor. Next, the acid and base bottles were filled with 20% phosphoric acid and 10% ammonia solution. The acid, base, and antifoam pump tubing were then primed.

The dissolved oxygen (pO_2) electrode was then polarized by connecting it to the online controller system of the bioreactor tower for a minimum of two hours before inoculation. The dissolved oxygen probe was calibrated by adjusting the temperature of the bioreactor to 37 °C and using an agitation speed of 300 rpm. The aeration was set to its equivalent airflow rate in terms of liters per minute (L/min), after conversion from volume of air to volume of medium ratio per minute (vvm). The aeration rate was fixed at 2 L/min for all experiments for both 1 L cultures (2 vvm) and 5 L cultures (0.4 vvm), unless otherwise stated. After the DO reading stabilizes (in about 30 minutes), the pO_2 is controlled via a cascade control sequence, as shown in Table 3.7 and the control sequence code given. Other parameters were set as in Table 3.8, according to the respective experimental run. A 1 mL sample of the uninoculated medium was withdrawn and its optical density at 600 nm was measured and noted, for use as a sample blank.

Table 3.7 One-step cascade program to control dissolved oxygen (DO) levels in the bioreactor

Cascade Parameter	**ON/OFF**	**Minimum**	**Maximum**	**Setpoint**
Agitation	ON	100 rpm	680 rpm	300 rpm

The oxygen control cascade sequence code executed in IRIS 6.0 was as follows:

```
#0, pO2 check, 5
//delay until po2 actual value is below required setpoint
IF(pO2.v>pO2.sp) {seq=0}
IF(pO2.v<=pO2.sp) {seq=1}
#1 cascade stirrer, 20
IF(pO2.v<=pO2.sp)(stirrer.sp+=50)
IF(pO2.v>pO2.sp+30)(stirrer.sp-=10)
IF(pO2.v>pO2.sp+10 AND stirrer.sp>680)(stirrer.sp=680)
IF(pO2.v>pO2.sp AND stirrer.so<380)(stirrer.sp=380)
Seq=1
```

The initial temperature of the culture (37 °C) was maintained until it reached the early mid-log phase after which it was lowered for induction, as detailed in Table 3.8. All fermentation parameters were keyed in and monitored online, and the data was stored, via IRIS v5.0 software (Infors-HT, Bottmingen, Switzerland).

Table 3.8 Fermentation control setpoints

Parameter	**Value**	**Reference**
Mode	batch	-
Fermenter	2 L	-
Working Volume	1 L	-
Inoculum size (%, v/v)	5 %	Lu et al. (2015)
Inoculum age (hours)	16 hours	Peng et al. (2012)
Pre-induction Temperature (℃)	37.0	Peng et al. (2012)
pH	7.0	Peng et al. (2012)
DO saturation (%)	≥ 20	Hajihassan et al. (2018)
Aeration setpoint (L/min)	2.0	Zhou et al. (2018)
Impeller speed setpoint (rpm)	300	Deniz et al. (2015)

3.8.1.2 Inoculum Preparation and Inoculation

A 5% (v/v) inoculum size was chosen for all cultivations in the bioreactor (Lu et al., 2015) to reduce the lag phase and initiate a faster cell growth. To prepare the 50 mL inoculum, fresh recombinant *E. coli* colonies on an ampicillin-containing LB Vegitone agar plate were used to inoculate a sterile LB Vegitone broth (Sigma-Aldrich, USA) that had been supplemented with ampicillin (100 μg/mL), in a 250 mL flask. The flask was then incubated overnight at 37 °C temperature and 300 rpm agitation speed (final OD_{600} ~ 3-5) using a rotary shaker (Sartorius). This inoculum (OD_{600} = 3 - 5) was transferred into a sterile inoculum bottle (250 mL) and then added aseptically to the bioreactor. The inoculation time is recorded as the beginning of the fermentation time (t_o). A sample is withdrawn at this point (t_o) to determine the initial cell density in the bioreactor at the beginning of the fermentation. The inoculum OD_{600} at inoculation was typically 0.25- 0.3.

3.8.1.3 E. coli Culture Sampling

During fermentation, a 10 mL sample was withdrawn aseptically through the Super Safe sampler, using a sterile syringe, for OD monitoring and other analyses, on an hourly basis. The sampling was then divided as follows: 2 mL for measurement of biomass concentration, OD_{600}; 2 mL for the recombinant CLP assays; 2 mL for glucose assay, and 2 mL for dry cell weight (DCW) measurement in g/L.

3.8.2 Cell Separation and Protein Recovery

After the fermentation was stopped, the culture was collected from the 2 L and 7.5 L bioreactors into sterile 1 L and 5 L bottles, respectively. *E. coli* cells were harvested using sterile, 500 mL polypropylene bottles (Eppendorf) by centrifugation at 4000 rpm, 4 °C for 30 minutes using a refrigerated centrifuge (Heraeus Multifuge X1R, Thermo Scientific, USA) or by centrifugation of culture media in sterile 1 L bottles at 5000 x g, at 4 °C, for 15 minutes using a centrifuge (Avanti J-26S XPI, Beckman Coulter, USA). The supernatant (waste media) was discarded. The bottles were carefully labeled and weighed before and after discarding the supernatant, and the wet cell weight was noted. The wet cell pellets were stored at -20 °C temperature until further use. For cell lysis, cell pellets were resuspended in a phosphate buffer (20 mM disodium hydrogen phosphate buffer containing 0.5 M sodium chloride, pH 7.4) at a ratio of 20 mL per gram of wet cell pellet, as recommended by Peng et al. (2012). Then, sonication was carried out on ice, at 40% amplitude, 10 seconds ON, 20 seconds off, half bursts for 5 minutes per sample, using an ultrasonic probe (Labsonic, Sartorius).

3.8.3 Large-scale Purification (Aqueous Two-Phase Separation)

On the large scale, the recombinant protein product was purified via aqueous two-phase separation (ATPS) followed by freeze-drying. The aqueous two-phase separation system was prepared by mixing optimized quantities of the polymer (polyethylene glycol, PEG 2000), at 24.8% (w/w), and the salt (dipotassium hydrogen phosphate), at 29.23% (w/w), in the form of potassium phosphate buffer at pH 7.0 and 10% (w/w) of

crude extract (cell lysate) (Awang et al., 2023). Distilled water was added to attain a total weight of 2.0 g. The mixture was mixed well and then centrifuged at 1000 x g speed, at 25 °C, for 10 minutes (Heraeus Multifuge X1R, Thermo Scientific) to allow for phase separation. The top (PEG-rich) and bottom (salt-rich) phases were then collected separately for analysis. The recombinant CLP favors the PEG-rich top phase (Awang et al., 2023).

3.8.4 Freeze-drying (Lyophilization)

Lyophilization of CLP samples recovered from aqueous two-phase separation was accomplished using a 1.5 L freeze dryer (Thermo Fisher Scientific, USA). The samples to be freeze-dried were placed overnight at -80 °C in 50 mL graduated Falcon tubes. The tubes containing the samples were then freeze-dried at -50 °C for 24 hours. The freeze-dried samples obtained were then ground into a fine powder and stored in airtight containers at -20 °C (Mohd Arshad, 2016).

3.8.5 Removal of PEG 2000 from the Aqueous Solution

The removal of PEG 2000 from aqueous solutions containing the protein after ATPS was accomplished via the use of activated carbon to adsorb the polyethylene glycol followed by filtration using a sterile 0.45 µm filter to separate activated carbon from the solution, according to the method described by (Gadjos, 2007).

3.9 OBJECTIVE 2: SCREENING OF PARAMETERS AND DESIGN OF EXPERIMENT

The screening of different physical variables was conducted using the Taguchi Orthogonal Array L9 (3 x 4) design. This design was selected to identify the important fermentation process components with respect to their main effects as there are few interactions between variables. It is useful in screening the effects of process variables on concentration while minimizing the number of experiments, saving time and resources. Four variables of the fermentation process were studied, namely induction temperature, dissolved oxygen (DO) concentration, induction point, and inducer (IPTG) concentration (Table 3.9). All four factors chosen in the present investigation were tested at three levels (Jamaluddin, 2015; Peng et al., 2012), and then statistically analyzed for significance. A total of nine (9) experimental runs were performed, according to the experimental design presented in Table 3.10. Analysis of data was carried out using the Design-Expert 13 and Minitab statistical software version 21.4.2.0 (https://app.minitab.com/).

Table 3.9 Levels of each parameter for the Taguchi OA Screening Design

Factor	**Units**	**Minimum (-1)**	**Mean (0)**	**Maximum (+1)**
Dissolved oxygen (DO)	%	20	30	40
Induction point	h	2	2.5	3
Induction temperature	℃	15	20	25
IPTG (inducer) concentration	mM	1	1.25	1.5

Table 3.10 Runs with actual values for each screening factor (version 21.4.2.0)

Run	IPTG conc. (mM)	Induction Temperature (℃)	Dissolved oxygen (%)	Induction point (h)	Response
1	1.00	15	20	2.0	
2	1.00	20	30	2.5	
3	1.00	25	40	3.0	
4	1.25	15	30	3.0	
5	1.25	20	40	2.0	
6	1.25	25	20	2.5	
7	1.50	15	40	2.5	
8	1.50	20	20	3.0	
9	1.50	25	30	2.0	

The signal-to-noise (S/N) ratio is a measure of robustness, which can be used to identify the factor levels that minimize the effect of noise on the response. The Taguchi design gives the following options for the signal-to-noise ratio calculation for each factor level combination, namely Larger is better, Nominal is best, and Smaller is better. In this study, the Larger is Better methodology was selected as the goal of the experiments was to maximize the concentration, and its formula is given by:

$$S/_N = -10\ \log_{10} \frac{\Sigma(\frac{1}{Y^2})}{n}$$

where Y represents the responses for the given factor level combination and n is the number of responses in the factor level combination.

Although the experiment was initially designed using Design-Expert 13, the design and analysis of data was transferred to the Minitab app version 21.4.2.0 as the Minitab user interface was more flexible and provided more direct statistical analysis

and visualization options. Then, the signal-to-noise (S/N) ratio was used as a measure of the amount of variation present or deviation from the desired value in the response. The main effects of each factor and the ranking of the parameters based on significance were done by setting S/N ratio calculations to “larger-is-better” for high concentrations of RPCLP. Subsequently, the findings were used to optimize the fermentation parameters as detailed in the next section.

3.10 OBJECTIVE 2: OPTIMIZATION AND EXPERIMENT DESIGN

The significant variables from the screening were selected and optimized by means of response surface methodology (RSM), using face-centered central composite design (FCCCD) of the experiment, to boost the concentration of recombinant protein. Two factors were selected based on the findings of the screening experiments, and three suitable levels for each of them were determined from the screening results as shown in Table 3.11. The parameter of the induction point was fixed at three (3) hours into the fermentation and the induction temperature was fixed at 20 ℃ for all experiments, based on the optimal levels determined by the screening. The statistical software, Design-Expert (Version 13, State-Ease Inc., USA) was used to obtain the face-centered central composite (FCCCD) experimental design matrix shown in Table 3.12. A total of eleven (11) experimental runs were designed, with three (3) center points. All the measurements were done in triplicates, and the average concentrations of protein were presented as the response. Statistical analysis including Analysis of Variance (ANOVA), model construction, and response surface graphs were all performed using the Design-Expert software.

Table 3.11 The independent variables and their corresponding levels for FCCCD

Factor	Name	Units	Minimum (-1)	Mean (0)	Maximum (+1)
A	IPTG concentration	mM	1.25	1.5	1.75
B	Dissolved oxygen	%	10	15	20

Table 3.12 Runs with coded values for each optimization factor (Design-Expert 13)

Run	IPTG Concentration (mM)	Dissolved Oxygen (%)	RPCLP Concentration (g/L)
1	10	1.25	
2	20	1.25	
3	10	1.75	
4	20	1.75	
5	10	1.5	
6	20	1.5	
7	15	1.25	
8	15	1.75	
9	15	1.5	
10	15	1.5	
11	15	1.5	

3.11 OBJECTIVE 3: FERMENTATION KINETICS

For the kinetic study, fermentation process parameters such as pH, induction temperature, aeration, and agitation in relation to high recombinant protein concentration were monitored to study their effects on biomass production. Experiments were carried out at the shake flask culture level, in triplicates, with fermentation details according to the methods described previously. Briefly, shake flasks containing 50 mL optimized M9-casamino acids medium were inoculated with *E. coli* BL21(DE3) seed cultures that had been grown overnight (OD_{600} ~ 3-5). After growth for 1.5-2 hours (OD_{600} ~ 0.8-1) at 37 °C at 200 rpm, the RPCLP expression was induced by reducing the temperature of the culture to 20 °C, and addition of 1 mM IPTG. After increasing the agitation speed to 250 rpm, the culture was maintained at this temperature overnight. As control experiments, flasks that contained recombinant *E. coli* were cultured but not induced.

3.11.1 Sample Analyses

Hourly sampling was done offline using a spectrophotometer, in duplicates, during the fermentation process. A 10 mL sample of culture media was withdrawn using a sterile syringe and the Super Safe sampling valve from the bioreactor through aseptic techniques. The optical density of the sample (OD_{600}) was noted, its glucose concentration measured, and then the sample was stored at -80 ℃ for centrifugation, cell lysis, and recombinant protein quantification using the fluorometric assay described in Section 3.6.3.

3.11.2 Biomass Concentration

Cell density was investigated by turbidometry using a spectrophotometer at 600 nm optical density (OD_{600}). To plot biomass concentration against fermentation time, biomass determination must be carried out using the wet cell paste mass and the dry cell weight. In the wet cell paste method, the culture media was centrifuged according to the conditions mentioned previously for cell harvest. The supernatant (spent media) was discarded, and the cell pellet mass was determined. The cell concentration was by determined by absorbance at OD_{600} (Biophotometer, Eppendorf, Germany), which allowed for rapid monitoring of *E. coli* growth. In addition, dry cell weight (DCW) was estimated in g/L. Samples were transferred into microwaveable pans of known masses and dried for 5 minutes in a microwave to determine the DCW. All samples were diluted with a cell-free medium to ensure that the absorbance reading was below 1. A calibration curve of DCW (in g/L) against OD_{600} was plotted to estimate the DCW. This calibration curve is presented in Appendix VII.

3.11.3 Glucose Concentration Measurement

A glucose assay kit (Cell Biolabs, USA) was employed to estimate the glucose levels in the medium throughout the culture,. In this colorimetric assay, glucose is oxidized by glucose oxidase to produce D-gluconic acid and hydrogen peroxide. The hydrogen peroxide is subsequently detected using a highly specific probe. Horseradish peroxidase catalyzes the reaction between the probe and hydrogen peroxide, which interact in a 1:1 ratio. After incubation for 30-45 minutes, both the samples and glucose standards of

known concentration are measured using a standard 96-well ELISA plate reader (Multiscan Go, Thermo Scientific, USA) at OD = 540 nm. The intensity of the OD is proportional to glucose concentration. A glucose standard curve was prepared according to manufacturer protocol and this curve is presented in Appendix VII.

All reagents were thoroughly mixed before use. First, the stock Glucose Standard solution (400 mM) was diluted with 1X Assay Buffer at a ratio of 1:10, by adding 5 µL of the glucose standard stock to 45 µL of the 1X Assay Buffer, to yield a 40 mM, 50 µL Glucose Solution. Next, the 40 mM Glucose Solution was used to prepare a series of the Glucose standards as in Table 3.13.

Table 3.13 Glucose concentrations for the preparation of the standards for the glucose assay standard curve.

Standard Tubes	**40 mM Glucose Solution (µL)**	**1X Assay Buffer (µL)**	**Glucose (µM)**	**Glucose (mg/dL)**
1	4	1596	100	1.8
2	250 of Tube #1	250	50	0.9
3	250 of Tube #2	250	25	0.45
4	250 of Tube #3	250	12.5	0.225
5	250 of Tube #4	250	6.25	0.113
6	250 of Tube #5	250	3.13	0.056
7	250 of Tube #6	250	1.56	0.028
8	0	250	0	0

To perform the assay, 50 µL of each glucose standard or samples were transferred into the wells of a 96-well microtiter plate. Then, 50 µL of the Reaction Mix was added to each well and the contents of each well were mixed thoroughly. The 96-well plate was then incubated for 30-45 minutes at 37ºC, protected from light. Finally, the plate was transferred to a spectrophotometric microplate reader (Multiscan Go Reader, Thermo Scientific, USA), and the optical density (OD) was measured at 540 nm wavelength. All the assay readings were performed in triplicates. The standard curve was first plotted, and the absorbance values for the samples were then used to determine the glucose concentration from the standard curve and its equation, taking the dilutions made into account.

3.11.4 Kinetic Models

With the assumption that glucose was the only limiting substrate and that the yield coefficient of cells from the substrate, *Y*xs, is constant throughout the process, two unstructured batch kinetic models were chosen. The logistic model of cell growth was used alongside the Monod model, to consider the effect of substrate consumption. These were combined with the Luedeking-Piret model for glucose utilization and RPCLP production (Ali et al., 2017; Ariff et al., 2015; Garnier & Gaillet, 2015). These model combinations are henceforth referred to as the Monod-Luedeking Piret (MLP) and the Logistic-Luedeking Piret (LLP), respectively. The Monod equation was used for the specific growth rate, μ, and is given by Equation 3.2.

$$\mu = \mu_{max} \frac{S}{Ks+S} \qquad \text{Equation 3.2}$$

wherein μ_{max} represents the maximum specific growth rate, S represents substrate (glucose) concentration (g/L), and K_S is the half-saturation constant for glucose.

The logistic model equation used is given by:

$$\mu = k\,(1 - (\frac{X}{Xmax})) \qquad \text{Equation 3.3}$$

$$\frac{dX}{dt} = \mu.X \qquad \text{Equation 3.4}$$

The Luedeking-Piret model equations used for the rates of substrate (glucose) consumption ($\frac{dS}{dt}$) and RPCLP production ($\frac{dP}{dt}$) were:

$$\frac{dS}{dt} = -m\frac{dX}{dt} - n.X \qquad \text{Equation 3.5}$$

$$\frac{dP}{dt} = \alpha\frac{dX}{dt} + \beta X \qquad \text{Equation 3.6}$$

A modified Luedeking-Piret model equation was also used for the rate of glucose utilization ($\frac{dS}{dt}$) that accounts for substrate uptake for growth, maintenance, as well as product formation (Rajendran & Thangavelu, 2008).

$$\frac{dS}{dt} = -\frac{1}{Y_{XS}}\frac{dX}{dt} - \frac{1}{YPS}\frac{dP}{dt} - k_eX \qquad \text{Equation 3.7}$$

where S is substrate concentration (g/L), X is biomass concentration (g/L), P is CLP concentration (g/L), α is the growth-associated CLP production constant, β is the non-growth associated CLP production constant, m is growth-associated glucose consumption constant (g/g), and n is the non-growth associated glucose consumption constant (g/g g), k_e is the maintenance coefficient for cells (g/g.h), Y_{XS} is the yield of cell mass from substrate, Y_{PS} is the yield coefficient of product from substrate, k is initial specific growth rate (h^{-1}), and X_{max} is the maximum cell concentration (g/L).

3.11.5 Initial Value Estimation

To solve the ordinary differential equations representing the models, the estimates of the initial values for each parameter in the model must be determined. As the solution involves iteration, the better the initial estimates are, the faster the convergence to the fitted value will be. On the other hand, convergence to erroneous final values can happen if the initial estimates are poor. The maximum specific growth rates, μ_{max}, were determined after the specific growth rate became constant. The optical density (OD_{600}) values of the cultures were determined every 30 minutes. The value of X_{max} is determined from experimental data. For the determination of initial values for the kinetic parameters and yield constants, experimental data was input into Microsoft Excel software, and the natural logarithms of the OD_{600} data were used to plot a time profile and the slope of the resultant line was taken as μ_{max}. Using Michelis-Menten kinetics; for batch growth culture, the specific growth rate (μ) is equal to the maximum specific growth rate μ_{max} (Doran, 2012). Meanwhile, the saturation constant or half-velocity constant (K_s) was calculated from measured μ_{max} and residual glucose concentrations, when $\mu_{max}/2$ (Doran, 2012). Values for observed yields such as Y_{XS}, and Y_{PS}, were first determined by plotting graphs of experimental data in Microsoft Excel. The value of n can be obtained from the stationary phase of the growth curve at which it is assumed that (dX/dt = 0) and ($X = X$m). Thus, n becomes -dS/dt/Xm. However, the model plots obtained using these values performed poorly in fitting the experimental data. This is supported by cases in the literature where transformations of the nonlinear rate equations to linear forms, such as Lineweaver-Burk linearization, have certain limitations making them undesirable (Chu et al., 2011). Therefore, the initial values were determined by observation (e.g., X_{max}) or calculation (e.g., μ_{max} and yield coefficients) by evaluating the models at the initial values, using Microsoft Office Excel software Office Excel (Microsoft, Redwood, WA). These estimates became the initial

values for solving the ordinary differential equations (ODEs) and fitting the data in Matlab software (MathWorks, Natick, MA).

3.11.6 Modeling and Statistical Analyses

Matlab software (R2022a) was used for plotting data and to analyze fermentation kinetics. Then, the *lsqcurvefit* function together with the *ode45* function of MATLAB was used to simulate and fit the experimental data to various models using the values of kinetic parameters that were estimated from the kinetic analysis of experimental data such as μ_{max}, $Y_{X/S,}$ and $Y_{P/S}$. The *ode45* function solves the differential equations using the Runge-Kutta formula. This was achieved by minimization of the sum of squares or quadratic differences between experimental and model-predicted values, using the nonlinear least-squares Levenberg-Marquardt method. The MATLAB codes are provided in Appendix VIII.

The coefficient of determination (R^2) was determined to evaluate the data fitting to the kinetic models. An R^2 value that is close to 1 indicates that the values obtained for the model are reliable and the data fits well (Ali et al., 2017). Analysis of Variance (ANOVA) was also carried out to evaluate the models. The statistical F-value is a ratio of the relative regression variation/relative residual variation. An F-value that is significantly greater than the critical F value suggests that the regression model is applicable to the data (Ali et al., 2017).

3.11.7 MATLAB Curve Fitting Toolbox

Finally, the Curve Fitting Toolbox of MATLAB (7.10.0.499) (R2010a) was used to simulate and fit various models using the values of kinetic parameters that were gained from the kinetic analysis such as μ_{max}, $Y_{XS,}$ and Y_{PS}. The toolbox has been useful in numerous kinetics studies, model-fitting, and determination of kinetic parameters (Ariff et al., 2015). To use the Curve Fitting App's custom equation feature, the equations cannot be in the differential equation form. Hence, the equations were integrated into the forms given by Equations 3.8- 3.15 (Ariff et al., 2015; Garnier & Gaillet, 2015; Limoes, 2016; Mullai et al., 2013; Rajendran & Thangavelu, 2008; Vinayagam et al., 2015).

For the Monod-Luedeking Piret model, the following equation was used:

$$X = X_0 e^{\mu_{max}\frac{S}{Ks+S}t} \qquad \text{Equation 3.8}$$

For the Logistic-Luedeking Piret model, the following equations were used:

$$X(t) = \frac{X_{max}}{1+(\frac{X_{max}}{Xo}+1)e^{kt}} = \frac{Xoe^{kt}}{1-\frac{Xo}{X_{max}}(1-e^{kt})} \qquad \text{Equation 3.9}$$

$$S_t = S_0 - m\,M'(t) - n\,N'(t) \qquad \text{Equation 3.10}$$

$$P_t = \alpha\,A'(t) + \beta\,B'(t) \qquad \text{Equation 3.11}$$

Wherein,

$$M' = Xo\left\{\frac{e^{kt}}{1-\frac{Xo}{Xmax}(1-e^{kt})} - 1\right\} \qquad \text{Equation 3.12}$$

$$N' = \frac{X_{max}}{k} \ln\left\{1 - \frac{Xo}{X_{max}}(1 - e^{kt})\right\}$$ Equation 3.13

$$A' = \frac{Xoe^{kt}}{1 - \frac{Xo}{k}(1 - e^{kt})}$$ Equation 3.14

$$B' = \frac{X_{max}}{k} \ln\left\{1 - \frac{Xo}{X_{max}}(1 - e^{kt})\right\}$$ Equation 3.15

where X represents biomass concentration (g/L), S is the glucose concentration (g/L), and P is CLP concentration (g/L). α is the growth-associated CLP production constant and β is the non-growth associated CLP production constant. m is the growth-associated glucose consumption constant (g g^{-1}), and n is the non-growth associated glucose consumption constant (g g^{-1} g^{-1}). Finally, k_e is the maintenance coefficient for cells (g/g.h), Y_{XS} is the yield coefficient of cell mass from substrate, and Y_{PS} is the yield coefficient of product from substrate.

3.12 OBJECTIVE 4: SCALE-UP

In aerobic fermentation processes, as in this case, oxygen supply frequently becomes a limiting factor that restricts growth. This is due to the low solubility of oxygen in the fermentation broth as well as its rapid depletion by fast-growing *E. coli* in high cell-density cultures (Gill et al., 2008). This makes the effect of gas-liquid mass transport a key factor in the scale-up of aerobic fermentation (Yasin et al., 2015). Consequently, scale-up in aerobic fermentation is often performed on the basis of maintaining the volumetric mass transfer coefficient (k_La) or the oxygen transfer rate (OTR) constant (Doran, 2012; Schmidt, 2005).

In this study, the scale-up of the 2 L bioreactor into the 7.5 L bioreactor was attempted by using constant impeller tip speed as well as constant volumetric transfer coefficient (k_La), by a trial-and-error approach as detailed by Azmi et al. (2015). The characteristics and specifications of the bioreactors used are detailed in Table 3.14.

Table 3.14 Characteristics of the 2 L and 7.5 L bioreactors

	Small scale	**Large Scale**
Model	Labfors 4, INFORS-HT (NW115)	Labfors 4, INFORS-HT
Vessel volume, V_T (m^3)	0.002	0.0075
Working volume. V_L (m^3)	0.0010	0.005
Vessel height (mm)	205	465
Vessel diameter (mm)	115	150
H_T/D_T ratio	1.8:1	3.1
Impeller diameter, D_i (mm)	46	54
Impeller type	6-blades Rushton turbine	6-blades Rushton turbine
D_i/D_T ratio	0.4	
Number of impellers	2	2
Baffles	Yes (3)	Yes (3)
Drive	Mechanically stirred	Mechanically stirred
Sterilization	Autoclaved	Autoclaved

3.12.1 Constant k_La

The scale-up criterion of keeping k_La constant can be achieved by varying and determining aeration and agitation rates for the 7.5 L scale to keep the k_La value constant. In this study, the k_La values were determined at various agitation rates in both the 2 L scale and the 7.5 L scale, to determine a correlation for the k_La values using agitation rates at both scales. This way, the range of agitation rates for the 7.5 L scale can be determined, keeping the k_La a constant in the scale-up.

Measuring k_La is crucial to monitoring and optimizing aeration. Its value has to be determined empirically as it depends on many factors. The dynamic gassing-out technique was used to determine the oxygen transfer coefficient, k_La, for this study (Jain et al., 2012). The method relies on the assumption that a steady state of dissolved oxygen concentration has been reached in the bioreactor, after inoculation. Initially, the air supply is turned on and inoculation is carried out. Aeration is achieved by using a sparger as well as agitation by the impellers. The aeration rate was fixed at 2 L/min for all experiments at both scales. Simultaneously, dissolved oxygen is being taken up by the cells. It is assumed that the DO level remains constant (C*) as there is a steady state of oxygen transfer and oxygen consumption.

Based on the dynamic mass balance for the batch fermentation, an equation for changes in *DO* can be established:

$$\frac{dCO_2}{dt} = OTR - OUR = K_La * CO_2 - CO_2 - QO_2.X \qquad \text{Equation 3.16}$$

Where:

C_{O2} is the dissolved oxygen concentration of the fermentation broth (g/L), C^*_{O2} is the dissolved oxygen concentration in equilibrium with oxygen partial pressure (g/L), Q_{O2} is the specific oxygen uptake rate per unit mass (g/g·h), X is cell concentration (g/L), and k_La is the volumetric oxygen transfer coefficient (1/h).

The method involved two stages, namely a "gas-out" and a "gas-in" phase, which were used to measure oxygen uptake rate (*OUR*) and the oxygen transfer rate (*OTR*), respectively. In the "gas-out" stage, the air inflow was briefly disconnected, and the dissolved oxygen (*DO*) concentration dropped, as recorded by the *DO* sensor. The oxygen uptake rate (*OUR*) was then determined by the rate of change of the *DO* concentration. Here, Equation 3.16 becomes:

$$\frac{dCO_2}{dt} = -OUR = -QO_2.X \qquad \text{Equation 3.17}$$

The "gas-out" stage was kept brief and dissolved oxygen (*DO*) concentration was kept above the critical oxygen concentration, C_{crit} (assumed to be 20% DOT) to ensure that the oxygen uptake rate (*OUR*) was almost constant, and no cell damage occurs. During the "gas-in" stage, the airflow inlet was resumed, and the DO concentration was observed to increase. Then, the oxygen transfer rate (*OTR*) was determined as follows:

$$OTR = \frac{dCO_2}{dt} + OUR \qquad \text{Equation 3.18}$$

Furthermore, Equation 3.16 can be rearranged into:

$$CO_2 = -1 * K_La * \frac{dCO_2}{dt} + CO_2 + QO_2.X \qquad \text{Equation 3.19}$$

The underlying principle of these methods is that in integrating:

$$\frac{dC_{AL}}{dt} = k_L a(C^*_{AL} - C_{AL}) \qquad \text{Equation 3.20}$$

we get:

$$\int_{C_{AL_1}}^{C_{AL2}} \frac{1}{(C^*_{AL} - C_{AL})} dC_{AL} = k_L a \int_0^t dt \qquad \text{Equation 3.21}$$

$$ln\left(\frac{C^*_{AL} - C_{AL1}}{C^*_{AL} - C_{AL2}}\right) = k_L at \qquad \text{Equation 3.22}$$

In this study, the dynamic gassing out method was employed to determine $k_L a$ at various agitation rates within the range used during cascade control to control the dissolved oxygen in the 2 L bioreactor. As the bioreactor did not reach the 680 rpm initially set as the maximum in Table 3.7, the range studied was between 300 rpm and 540 rpm, as shown in Table 3.15 for the 2 L bioreactor. As for the 7.5 L bioreactor, the range was set to 200 rpm to 400 rpm, based on studies at similar scales employing *E. coli* (refer to Table 2.16).

The measurement was performed immediately after inoculation of the bioreactor, at 37 ºC fermentation using the optimized fermentation conditions. As the experiments were performed early in the fermentation, oxygen enrichment was not used. The slope of DO (dCL/dt) was determined at several points when the air supply was resumed. Nonlinear data bounding the critical sections of the curves used for the calculations were intentionally not used because it is not accounted for by the simplistic models. These models do not take the dynamics of the DO probe into account, which are the cause of the nonlinearities such as the brief lag when the air supply is restored before the DO starts to rise and factors related to the DO probe itself. A plot of $ln\left(\frac{C^*_{AL} - C_{AL1}}{C^*_{AL} - C_{AL2}}\right)$ vs t was generated, resulting in a linear plot that intercepts at the origin, and $k_L a$ *is* represented by its slope.

Table 3.15 Speed settings for each point to determine k_La

Bioreactor	Agitation Speed (rpm)		
2 L	300	420	540
7.5 L	200	300	400

Source: Islam et al. (2008)

3.12.2 Constant Impeller Tip Speed

A constant impeller tip speed is another commonly used criterion in scale-up. This is because it implies a comparable homogeneity for the up-scaled bioprocess. Mixing and transfer are influenced by numerous factors such as agitation speed, the number of impellers, impeller type, and the gas flow rate into the reactor. However, while an adequate oxygen transfer rate is desirable, compromise in terms of shear stress to cells must be controlled, and hence this scale-up criterion was employed in this study. Keeping the impeller tip speed (v) constant at both scales implies that we can use Equation 3.22 to calculate the agitation speed, N_2, in rpm, for the large-scale bioreactor:

$$v = \frac{N2}{N1} = \frac{D_{i,1}}{D_{i,2}} \qquad \text{Equation 3.23}$$

where N_1 is the agitation speed for the small-scale bioreactor, N_2 is the agitation speed for the large-scale bioreactor, $D_{i,1}$ is the impeller diameter for the small-scale bioreactor, and $D_{i,2}$ is the impeller diameter for the large-scale bioreactor.

3.12.3 Validation of Choice of Scale-Up Strategy

The success of the scale-up processes is typically established when experimental results from both the small- and large-scale fermentations show that there is no significant difference between them while maintaining a scale-up criterion constant (Xia et al., 2016; Zhou et al., 2018).

3.13 OBJECTIVE 5: CHARACTERIZATION OF THE RECOMBINANT PROTEIN

Proteolytic digestion was done to characterize the structural integrity of the RPCLP protein, and the methodology used (including the rationale behind it) is briefly described in Section 3.13.1. Protein identification via MALDI-TOF/TOF LCMS was also attempted, as detailed in the subsequent section (3.13.2).

3.13.1 Proteolytic Digestion

The resistance of the collagen triple helix to enzymatic hydrolysis, such as trypsin digestion, was used as an indicator of its structure. Trypsin is a protease that is secreted by the pancreas and hydrolyzes amide bonds at the carboxylic groups of arginine and lysine. As Trypsin is derived from non-halal porcine sources, an animal-free alternative using Bovine trypsin made in maize seed named (TrypZean, Sigma-Aldrich) was employed for comparison of the digestion. The recombinant bromelain was obtained from Prof. Dr. Azura Amid (INHART, IIUM). Recombinant VCL protein standards were extracted from wet cell paste by sonication in 40 mM sodium phosphate buffer pH

8.0, 1 mM phenylmethylsulfonyl fluoride (PMSF) using 20 mL buffer per gram of cell paste (Peng et al., 2012).

3.13.1.1 TrypZean Digestion

Supernatant samples containing the recombinant protein dissolved in 50 mM disodium phosphate binding buffer, pH 8.0, were digested with Trypzean (at an enzyme: protein ratio, 1:20) at 37 °C for four hours. Aliquots were removed after 60 minutes, and the reaction was halted by adding phenylmethylsulfonyl fluoride (1 mM final concentration).

3.13.1.2 Trypsin Digestion

Supernatant samples containing the recombinant protein dissolved in 50 mM disodium phosphate binding buffer, pH 8.0, were digested with bovine pancreatic trypsin (Sigma Aldrich) (at an enzyme: protein ratio, 1:1000) at 4 °C for 24 hours. The reaction was halted by the addition of 1 mM phenylmethylsulfonyl fluoride (1 mM final concentration) (Peng et al., 2015).

3.13.1.3 Pepsin Digestion

Supernatant samples containing the recombinant protein dissolved in binding buffer were adjusted to pH 2.0 with hydrochloric acid, and porcine stomach pepsin (Sigma Aldrich) was added (enzyme: protein ratio, 1:1000) at 4 °C for 16 hours. The reaction was halted by adjusting the pH to 7.5 (Peng et al., 2014a; Peng et al., 2014b; Peng et al., 2015).

3.13.1.4 Bromelain Digestion

Supernatant samples containing the recombinant protein dissolved in binding buffer, pH 8.0, were digested using recombinant bromelain (at an enzyme: protein ratio, 1:1000) at 4 °C for 24 hours. Aliquots were removed after 60 minutes, and the reaction was halted by adding phenylmethylsulfonyl fluoride (1 mM final concentration) (Peng et al., 2014b; Renu, 2019). After the digestion reactions were stopped, each reaction mixture was analyzed using denaturing SDS-PAGE. Then, the samples were analyzed by denaturing SDS-PAGE using an undigested CLP sample as a reference, in addition to the 12-band Peacock Prestained Protein Marker (PM2610, Biotium). The band intensities were measured and compared with an undigested sample and trypsin sample, and this was used as a control (Yu et al., 2012).

3.13.2 Protein Identification by LCMS (MALDI-TOF/TOF)

Protein Identification by LCMS was conducted following SDS-PAGE in-gel digestion and trypsin. For the MALDI-TOF/TOF-MS analysis method, the peptide samples were mixed with an equal amount of matrix solution containing 10 mg/mL R-cyano-4-hydroxycinnamic acid in 0.1% TFA/50% ACN and spotted onto a 384-well stainless steel MALDI target plate (Applied Biosystems, Foster City, CA). The system used for the analysis of spotted samples was the ABI 4800 Proteomics Analyzer MALDI TOF/TOF mass spectrometer (Applied Biosystems). Both MS and MS/MS spectra were recorded in the combination mode and submitted for database searching. For protein ID, combined MS and MS/MS data were submitted via GPS Explorer (version 3.6, Applied Biosystems) to the Mascot server (version 2.0, Matrix Science) for database searching. The databases used were Swiss Prot and Uniprot (organism: *Rhodopseudomonas palustris*, last accessed: 2023.05.05).

3.14 SUMMARY

This chapter describes the materials that were used and the methods that were used to realize the objectives of this research. The flowchart of the methodology is presented, and the methodology for each objective is explained in detail. To achieve the first objective, One-Factor-at-a-Time (OFAT) optimization was carried out to optimize the medium component. Then the screening of process and induction parameters was carried out using a Taguchi L9 OA experiment design to obtain the significant factors for optimization of concentration. The two most significant factors were then optimized for maximum concentration of recombinant CLP, using the Face-centered Central

Composite Design, completing the second objective. For the third objective, a kinetic study and modeling were carried out, with the help of MATLAB software. The scale-up process represents the fourth objective of this study, and it was carried out until it was successful using two scale-up criteria, namely constant impeller tip speed and constant oxygen mass transfer. Finally, the characterization of the recombinant CLP was carried out via the use of SDS-PAGE, trypsin digestion, and MALDI-TOF. In the next section, the findings of the research are presented, described, and analyzed according to the objectives described here. The key findings are highlighted and discussed, leading to the conclusion from the results.

CHAPTER FOUR
RESULTS AND DISCUSSION

4.1 INTRODUCTION

In this section, the findings and their respective interpretations are presented according to the objectives.

4.2 COLONY PCR TO CONFIRM PRESENCE OF RECOMBINANT DNA

As can be seen in the agarose gel of the PCR products (Figure 4.1), the thick bands obtained corresponded to approximately 700 base pairs between the 500 and 750 bp DNA ladder bands. The recombinant DNA has been confirmed to be present in the cell pellets, before fermentation and cell rupture, via PCR and gel electrophoresis. The recombinant DNA has been confirmed to be present in the cell pellets, before lysis, via PCR and gel electrophoresis.

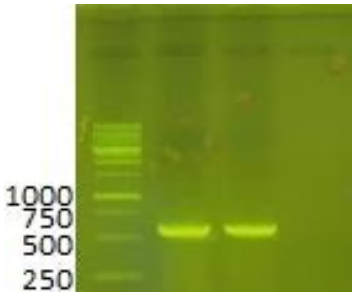


Figure 4.1 Agarose gel (1%) confirms the presence of the recombinant DNA encoding for the CLP from *R. palustris*. Lane 1 is the 10 kbp DNA ladder, Lanes 2 and 3 show the colony PCR product for RPCLP-containing *E. coli* BL21(DE3), giving a clear band at around 700 bp, and Lane 4 is the control PCR product of *E. coli* without the vector.

4.3 GROWTH CURVES

The growth profiles (Figures 4.2-4.3) indicated that the M9-cas medium formulation, after screening of components using the Plackett-Burman design carried out by Soroodi (Unpublished), was a suitable medium for the growth of the recombinant *E. coli*. The culture grown in M9-cas medium showed maximum OD_{600} values between 5.9 and 6.7 at the stationary phase, which were higher than those obtained using LB medium (OD_{600} = 5.3 to 5.4), as can be seen by comparing the growth profiles (Figures 4.2 – 4.3). The lag phases observed were also shorter in the M9-cas medium, as can be seen in Figures 4.2-4.3. Furthermore, the μ_{max} values were estimated for the log phase, and it was observed that they were higher for the culture grown in M9-cas medium (μ_{max} = 0.33 – 0.74 h^{-1}) than for LB (μ_{max} = 0.22 - 0.23 h^{-1}), indicating that the cells were able to grow rapidly in M9-cas medium.

A comparison of protein induction using LB and M9-cas media (Figures 4.2 - 4.3) shows that the IPTG slightly slowed down the growth of the bacteria, as can be

seen from the log phase slopes. This is due to the metabolic burden imposed by the induction as well as the toxicity of IPTG, which would have hampered the growth. In addition, induction was carried out at a shift to a lower temperature (from 37 °C to 20 °C), and this low temperature would slow down the bacterial growth as well. For LB, the final maximum optical density is relatively unaffected by the IPTG, even though the growth was slowed down. As for the M9-cas, the growth was slower than in the uninduced culture. Again, stress due to the induction conditions was expected to slow down the growth in the log phase. However, the final maximum biomass optical density attained was higher for the induced culture. This could be due to the casamino acid components in the M9-cas medium that supported the cell under stress and lower temperatures. Baez (2018) showed that *E. coli* BL21 cells turn off the production and transport of amino acids when they are available in the nutrient medium in the form of casamino acids, and this allows the cells to direct their metabolism to the proteome efficient fermentation pathways, such as growth and providing large amounts of protein for other metabolic activities. The lower temperature itself keeps the cells at optimal conditions for a longer duration, leading to an increased cell density at the end of the fermentation (Cai et al., 2016). This is supported by the length of the log phase in the induced fermentations (Figures 4.2-4.3). Overall, the growth with induction favored the M9-cas medium.

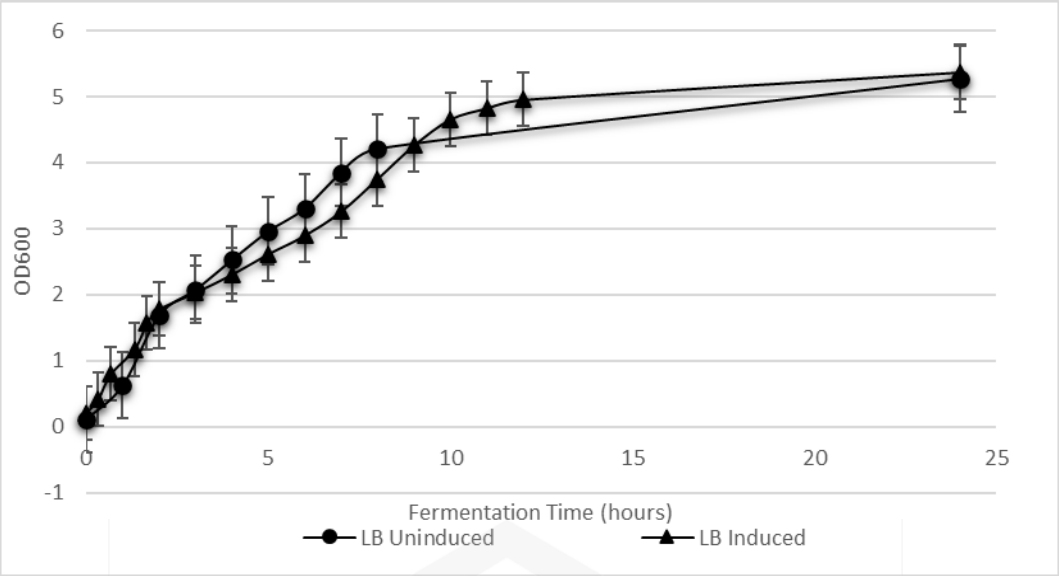


Figure 4.2 Growth curves of the recombinant *E. coli* BL21(DE3) in LB Vegitone medium at 37 °C, 200 rpm. The triangles show the growth of the bacteria induced with 1 mM IPTG, 20 °C, at OD_{600}= 0.8-1 to express RPCLP. The circles show the growth curve without induction. Data represent mean ± SD of three replicates.

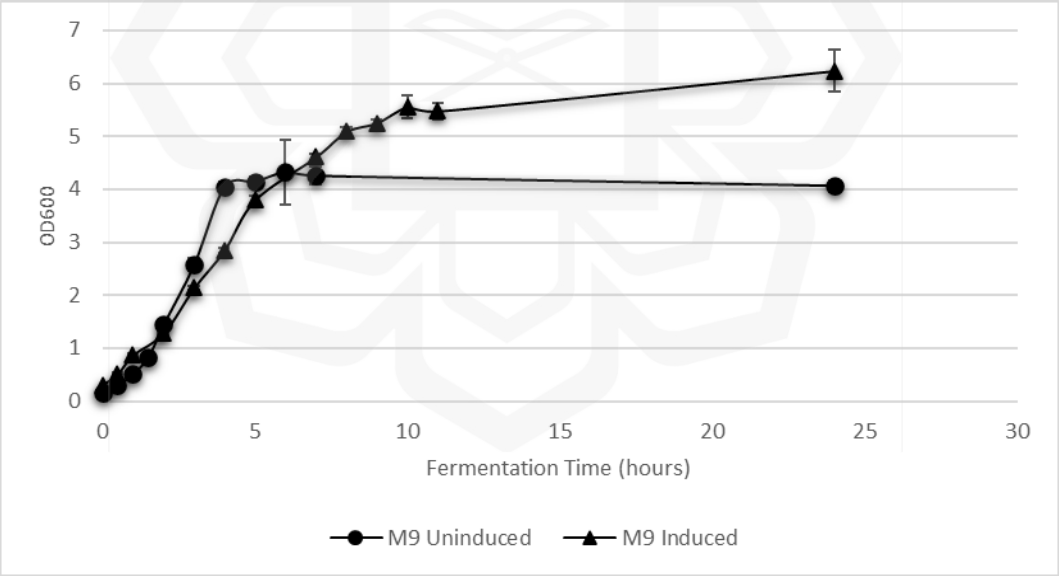


Figure 4.3 Growth curve of the recombinant *E. coli* BL21(DE3) in M9 medium at 37 °C, 200 rpm. The triangles show the growth of the bacteria induced with 1 mM IPTG, 20 °C, at OD_{600}= 0.8-1 to express RPCLP. The circles show the growth curve without induction. Data represent mean ± SD of three replicates.

The Bradford assay was used to quantify the total protein in each run. The results are shown in Figure 4.4. The induced culture in the M9-cas medium generated a higher total protein concentration compared to the uninduced culture in the same medium. This was expected from the induction of the CLP as well as the higher biomass obtained, as can be observed in Figure 4.3. On the other hand, the induced culture in LB had a lower total protein concentration compared to the uninduced culture in the same medium, even though they had attained similar final cell optical densities. This could be due to a leaky basal expression and/or expression occurring due to the lowering of the temperature before the addition of IPTG at the induction point. In the case of M9-cas, glucose is used in the medium, which leads to suppression of the expression in the uninduced fermentation. Overall, the concentration of total protein was higher in M9-cas medium fermentation.

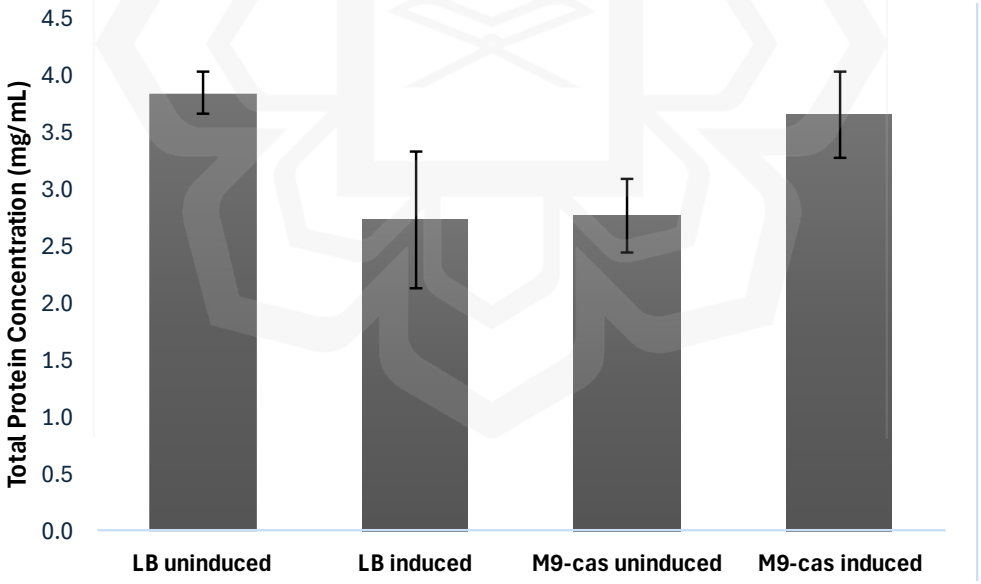


Figure 4.4 Total Protein Content as estimated by the Bradford assay for recombinant *E. coli* grown in LB and M9, when RPCLP production is induced and uninduced.

4.4 OBJECTIVE 1: MEDIUM COMPONENT OFAT

4.4.1 Effect of Disodium Hydrogen Phosphate Concentration on Growth

The growth profiles of *E. coli* BL21(DE3) grown on M9-cas medium of various concentrations of disodium hydrogen phosphate (provided in Table 4.1) are presented in Figure 4.5. From the growth curves (Figure 4.5), it can be observed that the lowest concentration (0.2 M) shows the least growth, as can be seen from the slope of its log phase ($\mu \cong 0.42\ h^{-1}$), although its maximum optical density reached (at the stationary phase) was 6.48. Conversely, the middle and highest concentrations studied showed a less significant difference in their maximum growth rates ($\mu \cong 0.44\ h^{-1}$) and maximum OD_{600} values (6.14 and 5.90, respectively). At a glance, these three values seem close together, and it is difficult to draw a conclusion. However, inspection of the growth profile indicates a distinguishable and undesirable lag phase in the case of the lowest concentration of disodium hydrogen phosphate. Combining these findings with the data from collagen concentration led to 0.4 M being determined as the optimal salt concentration in the M9 minimal salt solution.

Table 4.1 The concentration of disodium hydrogen phosphate salt in M9 salts used in the OFAT optimization

Level of Na_2HPO_4 in the M9 salt	**Low**	**Medium**	**High**
Concentration (in M)	0.2	0.3	0.4

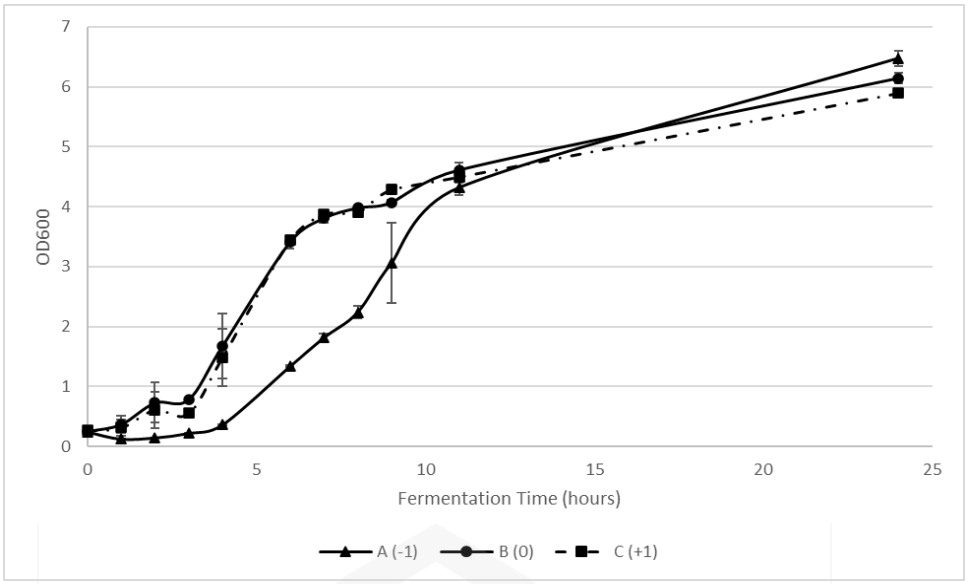


Figure 4.5 Growth curves of the recombinant *E. coli* BL21(DE3) for optimization of disodium hydrogen phosphate concentration (OFAT) in M9-cas media. Data represent mean ± SD of three replicates.

There are several different M9 recipes available in the Cold Spring Harbor Protocols, Helmholtz Zentrum München (https://www.helmholtz-muenchen.de/), and SubtiWiki. In these M9 minimal medium recipes, typical concentrations of the disodium hydrogen phosphate salt range from 0.2 M (in 5X salt solution) (Elbing & Brent, 2019) to 0.423 M (in 10X salt solution). A commercially available SX M9 Minimal Salts solution (Catalog number: A1374401, Gibco) uses 25.6 g sodium phosphate (dibasic) heptahydrate, which is the equivalent of a 0.0477 M 1X concentration. The manufacturer claims that cell growth rates decrease at lower or higher osmolarity and that the M9 medium has a lower osmolarity (~0.25 Osm/L) than *E. coli* needs to optimally grow (~0.3 Osm/L), allowing for tight control over the bacterial growth. Similarly, 5X M9 Minimal Salts (M9956, Sigma Aldrich) use 64 g/L $Na_2HPO_4 \cdot 7H_2O$ which corresponds to a 0.0477 M 1X concentration. Hence, the optimal value obtained in this study conforms with similar research work findings and reiterates the importance

of sodium hydrogen phosphate salt in *E. coli* growth and recombinant protein production media. The uptake of phosphate is important as it supplies an essential nutrient (phosphate) to the bacteria, along with its buffering capacity that ensures a stable pH for the bacteria to thrive.

4.4.2 Effect of Glucose Concentration on Growth

The growth patterns using the various concentrations of glucose (Table 4.2) were very similar in terms of both maximum growth rate (log phase, ($\mu \cong 0.88$, 0.97, and 0.91 h^{-1} in order of increasing glucose concentration) and final optical density (stationary phase, $OD_{600} \cong$ 7-8), as can be seen from Figure 4.6. It was therefore difficult to determine which concentration is the best from the growth profiles alone. The experimental runs were repeated several times. Earlier experimental runs indicated that the lowest concentration (1%) is the optimal concentration among those studied. Later experimental runs, on the other hand, point towards the middle concentration (2.5%) being faster (steeper slope/growth rate in log phase) and overall higher stationary phase optical density. A final set of experiments revealed a similar pattern, and thus it was considered adequate to use the lowest glucose concentration for further experiments, as the cell growth rate attained at this glucose concentration was very high. This would be cost-effective and also prevent issues such as the accumulation of growth inhibitors, such as acetate formation, caused by very high uptake of glucose in the medium, caused by acetyl-CoA being diverted from the TCA cycle towards acetate formation (Luo & Mu, 2014). In this minimal medium, glucose serves to provide the correct carbon precursors and the energy required for the synthesis of biomass. However, the presence of an organic source of nitrogen, i.e., casamino acids, could provide some carbon source

as well, and thus glucose serves as an energy supplier (Luo & Mu, 2014). The same concentration of glucose was used successfully by Peng et al. (2012) in the 2 L bioreactor production of CLP from *S. pyogenes* in *E. coli.*

Table 4.2 Levels of glucose concentration used in the OFAT optimization

Level of Glucose concentration	**Low**	**Medium**	**High**
Concentration of glucose (%)	1	2.5	5
Concentration of glucose (g/L)	10	25	50

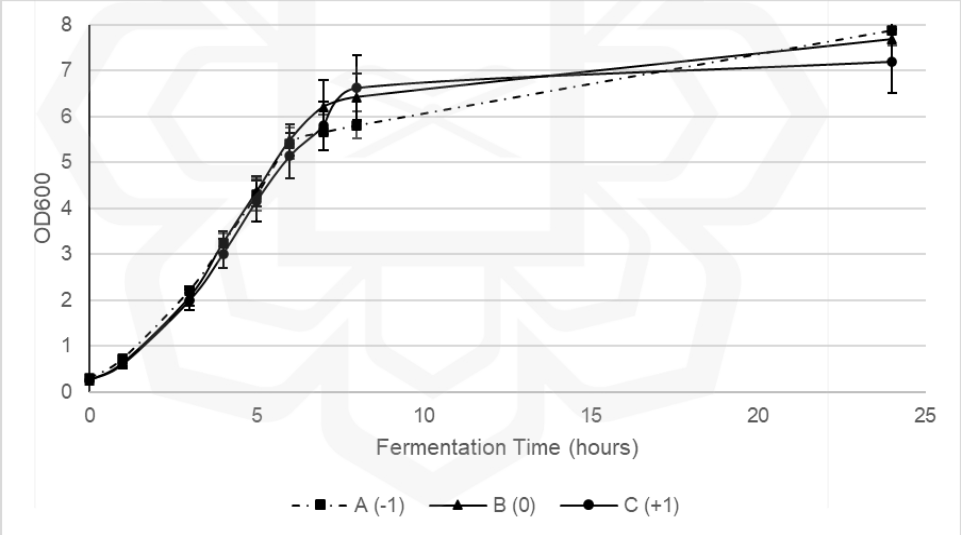


Figure 4.6 Growth curves of the recombinant *E. coli* BL21(DE3) for optimization of glucose concentration (OFAT) in M9-cas media. Data represent mean ± SD of three replicates.

4.4.3 Effect of Casamino Acids Concentration on Growth

The growth profiles of *E. coli* BL21(DE3) grown on an M9-cas medium of various concentrations of casamino acids are shown in Figure 4.7. The concentrations are given in Table 4.3. The growth profile for the lowest concentration (5 %) shows the least growth, as can be seen from the slope of its log phase ($\mu \cong 0.24\ h^{-1}$) and its maximum optical density reached (at the stationary phase, $OD_{600} = 7.6$). On the other hand, the middle and highest concentrations studied showed less significant differences in their maximum growth rates ($\mu \cong 0.24$ and $0.25\ h^{-1}$) and maximum OD_{600} values (8.07 and 8.29).

Table 4.3 Levels of casamino acids concentration used in the OFAT optimization

Levels	**Low**	**Medium**	**High**
Concentration of casamino acids (%)	5	7.5	10
Concentration of casamino acids (g/L)	0.5	0.75	1

The results for OFAT of casamino acids showed that an increase in the concentration of casamino acids significantly enhanced growth patterns, as can be seen in Figure 4.7. This can be attributed to the casamino acids being a rich source of organic nitrogen. Organic nitrogen sources were found to enhance growth and protein production in other studies as well (Uhoraningoga et al., 2019; Zamani et al., 2015; Zhang, 2009). This may be due to other nutrients and growth enhancers present in them that increase the availability of biosynthetic precursors and/or enhance plasmid stability and reduce proteolysis (Gellermann et al., 2019), as well as being a source of carbon for

biomass growth. However, the marginal increase between the growth of flasks B (0) and C (+1) comes at a significant cost, as casamino acids make up a significant proportion of the total cost of the M9 medium components. Hence, the middle level of concentration B (0) corresponding to 7.5 g/L casamino acids was considered adequate and was employed in further experiments. Studies such as Trausch and Batey (2015) and Wong et al. (2003) have successfully applied media containing 10 g/L glucose and 1 g/L casein hydrolysates (casamino acids) for the culture of *E. coli* producing recombinant proteins. Others, such as Espinosa Perez et al. (2018), have utilized concentrations of up to 10 g/l casamino acids.

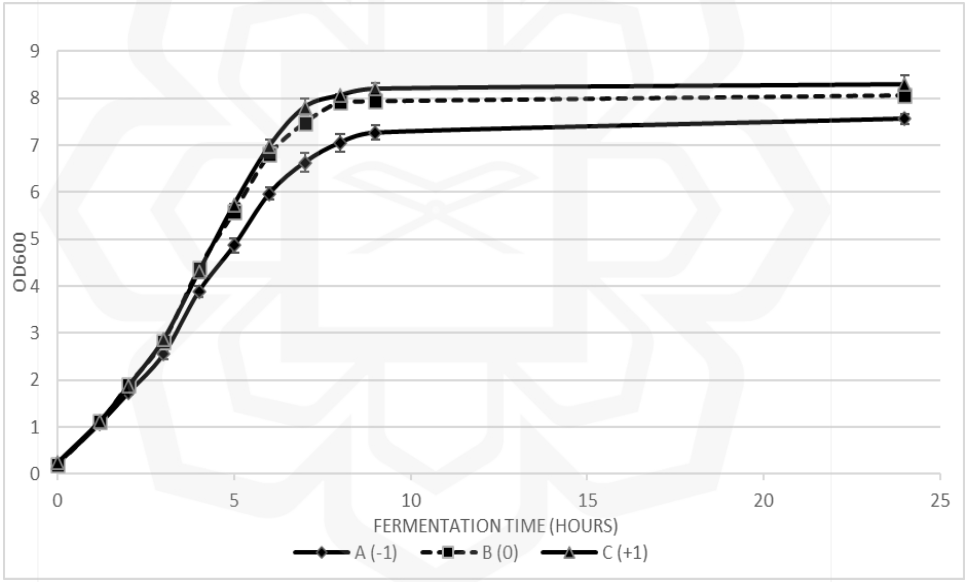


Figure 4.7 Growth curves of the recombinant *E. coli* BL21(DE3) for optimization of casamino acid concentration (OFAT) in M9-cas media. Data represent mean ± SD of three replicates.

4.4.4 Effects of Medium Components on Total Protein

The total protein content has been assayed via the Bradford assay, and the results (Figure 4.8) showed that the highest total protein content was obtained from samples that grew in the highest concentration of salt (0.4 M), and the lowest concentration of glucose (1%) and the middle level of casamino acids (7.5%). Previously, the growth profiles concluded that the optimal M9-cas has 0.4 M for disodium hydrogen phosphate, 1% or 10 g/L glucose, as well as 7.5% casamino acids concentrations. Since these results were similar, they were adopted as the optimal level of each medium component.

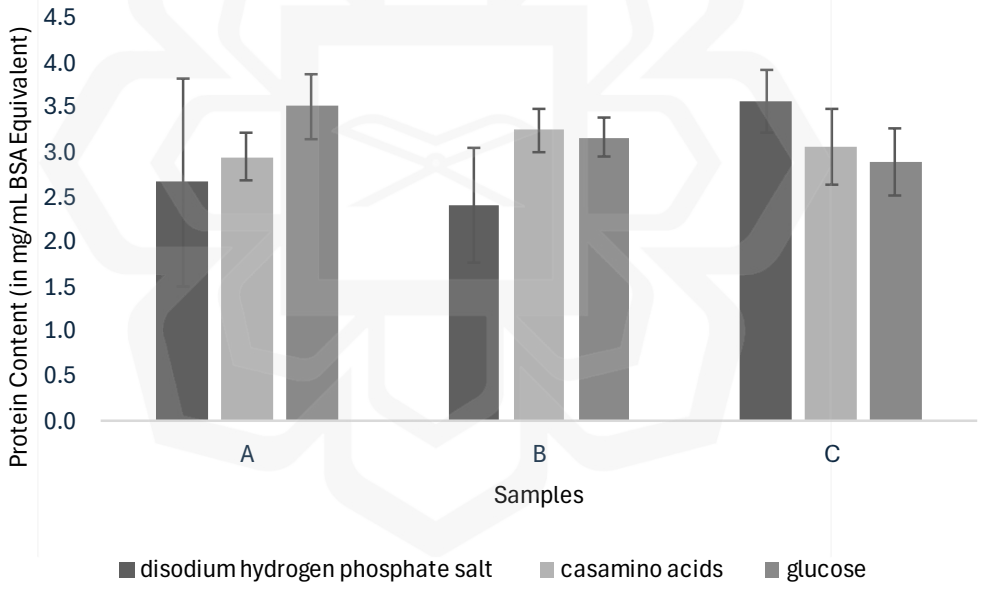


Figure 4.8 Total Protein Content (BSA equivalent) for each flask. The highest total protein content achieved is around 3.5 mg/mL BSA equivalent, achieved using the highest concentration of disodium hydrogen phosphate salt (C) at 0.4 M, the lowest concentration of glucose (A) at 1%, and the middle level (B) of casamino acids (7.5%).

4.4.5 Effects of Medium Components on Collagen-Like Protein

Later on, the CLP content was assayed via a fluorometric glycine-based assay, and from these, the highest concentrations were obtained at media composition: 2.5% glucose, 0.4 M disodium hydrogen phosphate salt concentration, and 10% casamino acids concentration, as can be seen in Figure 4.9. The case for 1% glucose and 7.5% casamino acids concentrations has been discussed in the previous subsections of this chapter, namely 4.4.1 – 4.4.4. These results match the findings from the growth curves and reflect the influence of biomass on CLP concentration, although kinetic studies are necessary to unravel this relationship in more detail, as was done in Section 4.6 of this chapter.

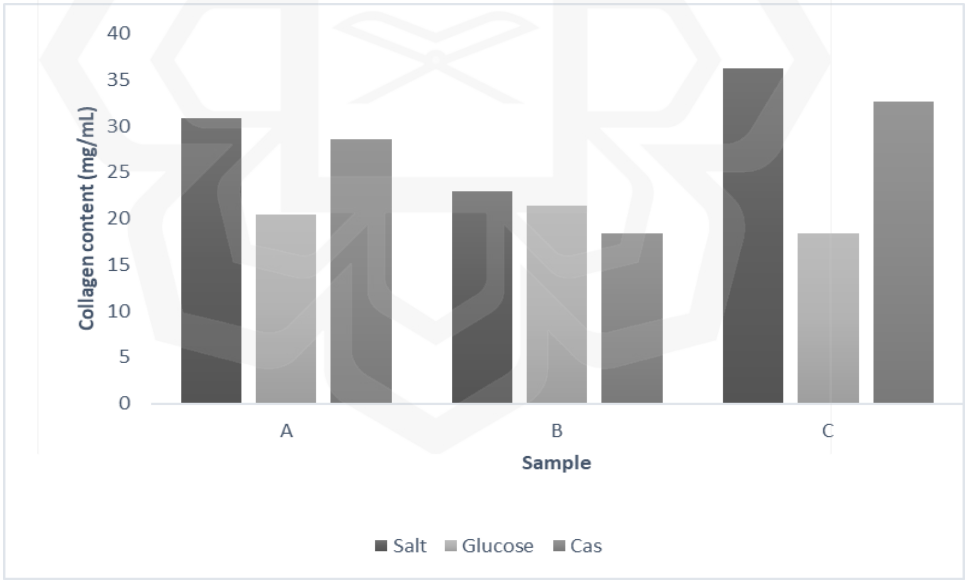


Figure 4.9 Total collagen content for each flask, where A shows the lowest concentration, B shows the middle concentration, and C shows the highest concentration.

4.4.6 Validation and Conclusion

In conclusion, the use of an M9 minimal medium using glucose as the carbon source, supplemented with thiamine and casamino acids has proven to be a good choice of basal medium for the growth of *E. coli* cells harboring the gene for RPCLP. Merging the collagen assay data, Bradford assay data for total proteins, and cell growth profiles, it was concluded that the following solution gives optimum growth and total protein concentration:

- 0.4 M $Na_2HPO_4.2H_2O$ (71.196 g per liter of M9 salt solution),
- 1% glucose (i.e., 10 g/L medium), and
- 7.5% (i.e. 7.5 g/L medium) casamino acids.

The results showed that the optimized medium is a well-balanced nutrient source. After the medium component optimization, there was a measurable increase in maximum biomass OD_{600} from an average of 6.24 (in M9-cas) and 5.37 (in LB) to 8.18 (in optimized M9-cas), total protein concentration, as well as growth rate, compared to the original M9-medium recipe as well as Lysogeny Broth (Vegitone). The use of a defined and optimized medium containing glucose as the carbon source has both reduced the fermentation cost and the medium complexity, simplifying the downstream processing, among other advantages. Table 4.4 summarizes the findings of the validation as compared to the original medium formulation (given in Appendix IV) and LB (Vegitone).

Table 4.4 Validation of medium component optimization

	LB (Vegitone)	**M9-cas**	**Optimized M9-cas**
Biomass OD_{600}	5.37	6.24	8.18
Biomass (DCW) g/L	4.60	5.22	6.61
Maximum specific growth rate	0.22	0.33	1.75
total protein concentration (mg/mL)	2.742	3.669	~ 22

4.5 OBJECTIVE 2: SCREENING AND PROCESS OPTIMIZATION IN 2 L BIOREACTOR

4.5.1 Taguchi L9 Design (Screening)

Data from the screening runs statistically designed using the Taguchi Orthogonal Array L9 (3^4) design of the experiment were used to identify the significant investigated fermentation process parameters. The highest protein concentration was obtained in Run 8, at 5.41 g CLP per L of culture, at Dissolved Oxygen Tension (DOT) ≥ 20%, three (3) hours induction point, using 1.5 mM IPTG and at 20 °C. In contrast, Run 4 gave the lowest CLP concentration (0.78 g/L culture). The results show that an induction temperature of 20 °C was the best among the three points, and an IPTG concentration of 1.5 mM gave the highest protein concentration, though this is inconsistent, as the 1 mM performed better than the 1.25 mM, as seen from Table 4.5. Similarly, high protein concentrations were obtained only when dissolved oxygen was at the extremes. Induction time was best at three (3) hours, which was the highest level studied here. This may be due to a certain population size being reached before induction, which can lead to a higher protein concentration.

The main effects of each factor and the determination of significant factors for the production of RPCLP were determined by setting S/N ratio calculations to "larger-is-better" for high concentration of RPCLP. The calculated S/N ratios and means are given in Table 4.6. The maximum S/N ratio (14.66) was obtained in Run 8 and the lowest obtained (-2.15) was in Run 4. Run 8 had the lowest level of dissolved oxygen tension, and the highest levels for induction point and IPTG concentration. Table 4.7 provides the ranking of the four factors in terms of S/N ratios and means, with 1 being the most significant and 4 being the least significant factor.

Table 4.5 Taguchi L9 OA: Factors, levels, and Larger is Better S/N Ratio (Minitab v. 21.4.2.0)

Run	**Factor 1** **A: IPTG** **mM**	**Factor 2** **B: IND** **deg C**	**Factor 3** **C: DO** **%**	**Factor 4** **D: IND** **hour**	**Response 1** **RPCLP** **g/L**	**S/N RATIO**
1	1.00	15	20	2.0	2.43	7.6997
2	1.00	20	30	2.5	1.12	1.0218
3	1.00	25	40	3.0	2.52	8.0312
4	1.25	15	30	3.0	0.78	-2.1462
5	1.25	20	40	2.0	1.72	4.7106
6	1.25	25	20	2.5	1.23	1.7910
7	1.50	15	40	2.5	1.34	2.5342
8	1.50	20	20	3.0	5.41	14.6608
9	1.50	25	30	2.0	1.76	4.9332

Table 4.6 Response table for S/N ratios and means (Minitab v. 21.4.2.0)

	Signal to Noise (S/N) Ratios					**Means**				
Level	1	2	3	Delta	Rank	1	2	3	Delta	Rank
IPTG	5.584	1.452	7.376	5.924	2	2.024	1.243	2.837	1.594	3
TEMP	2.696	6.798	4.918	4.102	4	1.515	2.751	1.838	1.236	4
DO	8.05	1.27	5.092	6.781	1	3.021	1.224	1.86	1.798	1
IND	5.781	1.782	6.849	5.066	3	1.97	1.231	2.903	1.672	2

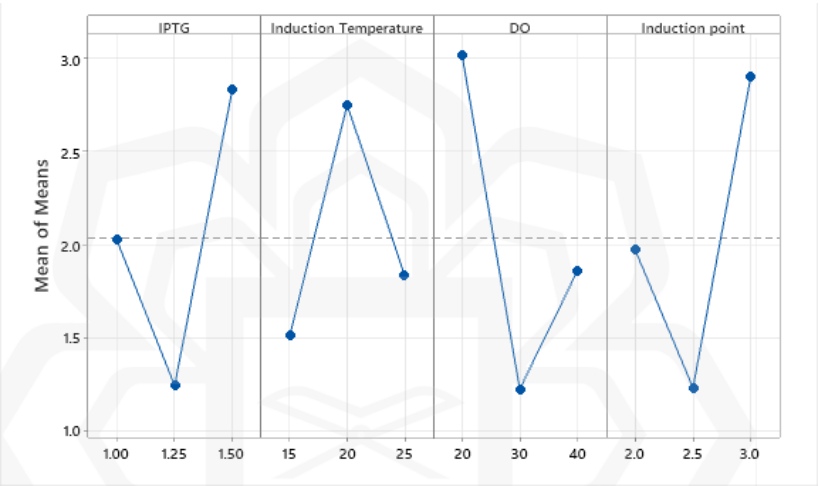


Figure 4.10 Main effects plot for means (Taguchi L9, Minitab v. 21.4.2.0)

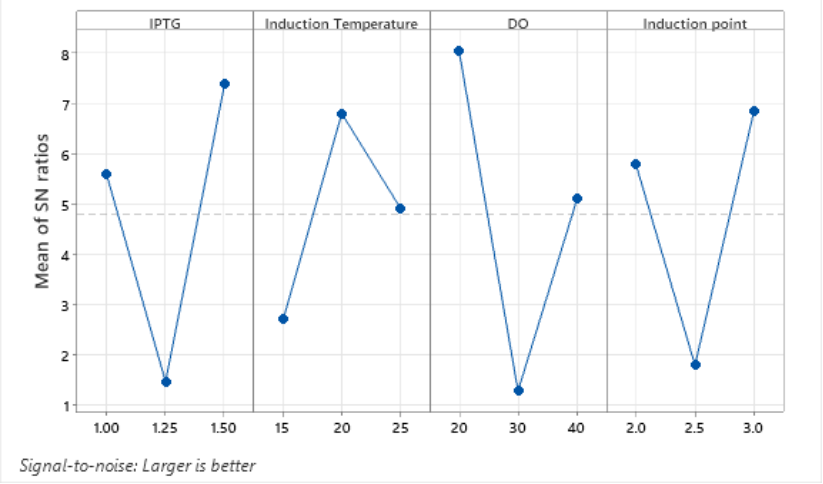


Figure 4.11 Main effects plot for S/N Ratios (Minitab v. 21.4.2.0)

Figures 4.10 and 4.11 illustrate the main effects of each factor in terms of means and S/N ratios, respectively. Ranked by S/N ratios, the most significant factor was dissolved oxygen tension, followed by IPTG concentration, induction point, and lastly induction temperature. Similarly, ranked by the means, the most significant factor was dissolved oxygen tension, followed by induction point, IPTG concentration, and lastly induction temperature. ANOVA could not be applied to the L9 design, unlike the L27 design which gives extra degrees of freedom (dF), as can be seen from Table 4.7. The residuals follow a normal distribution as shown by the normal probability plot (Figure 4.12) which follows a linear trend.

Table 4.7 ANOVA analysis for means

Source	DF	Seq SS	Adj SS	Adj MS	F	P
IPTG	2	3.8109	3.81086	1.90543	*	*
Induction Temperature	2	2.4638	2.46379	1.23189	*	*
DO	2	4.9852	4.98520	2.49260	*	*
Induction point	2	4.2145	4.21450	2.10725	*	*
Residual Error	0	*	*	*		
Total	8	15.4743				

* indicates that it cannot be determined.

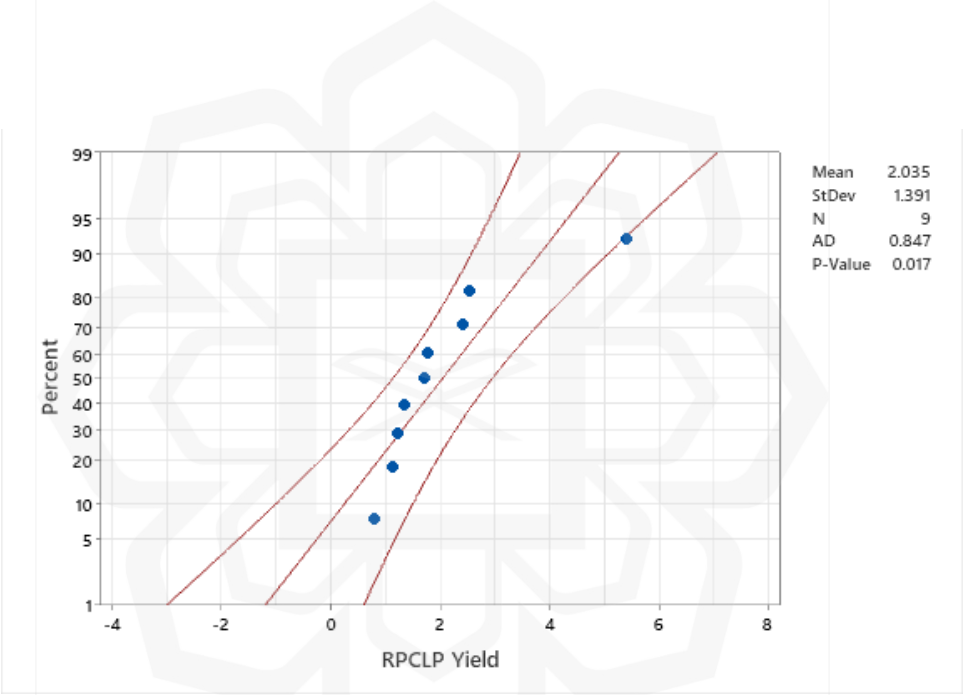


Figure 4.12 Normal probability plot of concentration data (95% confidence interval) (Minitab v. 21.4.2.0)

4.5.1.1 Effect of Dissolved Oxygen Concentration

The effect of the dissolved oxygen concentration on recombinant *E. coli* cells producing recombinant proteins has been the subject of numerous studies. Limitations in dissolved oxygen concentration have been reported to both increase and decrease the expression of various recombinant proteins (Zare et al., 2019), and as such the monitoring and control of dissolved oxygen are essential during the fermentation. Furthermore, the deficiency of dissolved oxygen could lead to acetate formation and/or accumulation, interfering with methionine biosynthesis, and thereby inhibiting recombinant protein expression (Luo & Mu, 2014). In these experiments, the minimum dissolved oxygen setpoint in the bioreactor was varied to determine the optimal value that would enhance recombinant CLP concentration. Recombinant *E. coli* generally requires a higher aeration capacity (Bhattacharya & Dubey, 1997), but as there were cases where the recombinant product of interest favored low oxygen conditions (Zare et al., 2019), it has to be investigated for the specific fermentation.

4.5.1.2 Effect of IPTG Concentration

The reported effects of the IPTG inducer concentration on recombinant protein expression by *E. coli* vary in the literature. The IPTG concentration range typically used in induction ranges from 0.05 mM up to 2.0 mM beyond which IPTG is considered toxic to cells (Bahreini et al., 2014), with 1 mM being frequently used and then reduced concentrations being considered to enhance protein solubility (Choi & Geletu, 2018). IPTG can significantly inhibit growth in some cases (Restaino et al., 2018). In the

present study, it is observed that the cells are not very sensitive to IPTG concentrations, and as such a higher concentration (1.5 mM) was needed to increase the cell concentration.

4.5.1.3 Effect of Induction Point

The induction point was found to be another important factor in recombinant protein production. This is because recombinant protein production is a metabolic burden on the recombinant bacteria and can result in a reduced growth rate and cell density. If it is done hastily, it would result in low cell density and lead to low productivity. Jamaluddin (2015) had previously reported that recombinant bromelain productivity in the bioreactor was sensitive to inducer addition and the recombinant bromelain activity was optimal when the induction was conducted during the mid-log phase. In the present study, it can be observed that the late induction point favored a higher RPCLP concentration. However, this effect was not as pronounced and significant as the effect of dissolved oxygen on the RPCLP concentration.

4.5.1.4 Effect of Induction Temperature

Increasing the cultivation temperature has been shown to favor cell growth whereas during the induction phase decreasing the temperature usually enhances the solubility of the protein expressed and prevents the formation of inclusion bodies (Koopaei et al., 2018; Yari et al., 2010). Hajihassan et al. (2018) found that in addition to DO concentration and post-induction time, induction temperature also significantly affected the production of recombinant human β-nerve growth factor (NGF). In the present study, induction temperature ranked last among all the factors studied. This suggests that while it is important to consider the temperature at which the induction was conducted, the range studied (15 °C to 25 °C) represented a favorable range of conditions for the production of the protein and the effect of temperature was not as significant as the dissolved oxygen concentration and other factors.

Table 4.8 Results of the numerical optimization (Taguchi L9, Design-Expert 13)

Factor	**Optimal Level**
Minimum Dissolved Oxygen concentration setpoint	20%
IPTG concentration.	1.5 mM
Induction Point	3 hours
Induction Temperature	20 °C

In summary, screening experiments designed using the robust Taguchi Orthogonal Array were analyzed via the Minitab app and it was found that the minimum dissolved oxygen, inducer (IPTG) concentration, and induction point were the most statistically significant factors. Data follows normal distribution as all points lie within

the 95% confidence interval and therefore the statistical analyses and models are significant. From the numerical optimization guided by the Design-Expert software version 13.0, for maximizing concentration, the optimal conditions are as in Table 4.8, with the expected concentration being 5.41 g/L. Subsequently, these findings were incorporated into the design of the optimization experiments as detailed in the next section using minimum dissolved oxygen and inducer (IPTG) concentrations. The parameter of the induction point was fixed at three (3) hours into the fermentation, and the induction temperature was fixed at 20 ℃ for all experimental runs based on the optimal levels shown in Table 4.9.

4.5.2 Face-centered Central Composite Design (Optimization)

The statistical software, Design-Expert (Version 13, State-Ease Inc., USA) was used to obtain the face-centered central composite (FCCCD) experimental design matrix shown in Table 4.9. A total of 11 experimental runs were designed, including 3 center points. Statistical analysis including Analysis of Variance (ANOVA), model construction, and response surface graphs were all performed using the Design-Expert software.

Table 4.9 Runs with coded values for each optimization factor (Design-Expert 13)

Run	Dissolved Oxygen (%)	IPTG Concentration (mM)	RPCLP Concentration (g/L culture medium)
1	10	1.25	1.225
2	20	1.25	1.528
3	10	1.75	1.554
4	20	1.75	2.082
5	10	1.5	0.860
6	20	1.5	1.475
7	15	1.25	1.244
8	15	1.75	1.832
9	15	1.5	1.508
10	15	1.5	1.292
11	15	1.5	0.990

The maximum CLP concentration was 2.08 g/L and was obtained in Run 4 and the lowest was 0.86 g/L, obtained in Run 5. Run 4 had the highest level of dissolved oxygen (20 %) and IPTG concentration (1.75 mM). Run 5 had the lowest level of dissolved oxygen (10 %) and IPTG concentration (1.0 mM). Compared to the screening results where the maximum CLP concentration was 5.41 g/L and the lowest was 0.78 g/L, the highest concentrations appear to be lesser than that obtained during the screening. However, the optimization runs were conducted by keeping the fermentation time fixed at seven (7) hours and inducing at three (3) hours into the fermentation, compared to the screening runs where the fermentation was for eight (8) hours. This led to a lower cell density and in turn affected the production of RPCLP protein as kinetic models were established and the relationship between biomass accumulation and RPCLP product formation was determined to be mixed growth associated, as discussed in Section 4.6.

4.5.2.1 Quadratic Model, ANOVA, and Fit Statistics

The quadratic model equation in terms of the coded factors is given in Equation 4.1. The coded equation allows the determination of the relative impact of the factors by comparing the factor coefficients. The equation in terms of actual factors is also given here, where the values must be specified in the original units for each factor, to predict the concentration. A contour plot and the 3D response surface for the model are shown in Figure 4.13.

RPCLP concentration = 1.23 + 0.2409A + 0.2453B + 0.0562*A*B - 0.0141A^2 + 0.3559B^2 Equation 4.1

RPCLP concentration = 12.73282 – (0.002371*DO) – (16.77520*IPTG concentration) + (0.044946*DO*IPTG concentration) – (0.000562*DO^2) + (5.69408*IPTG concentration^2) Equation 4.2

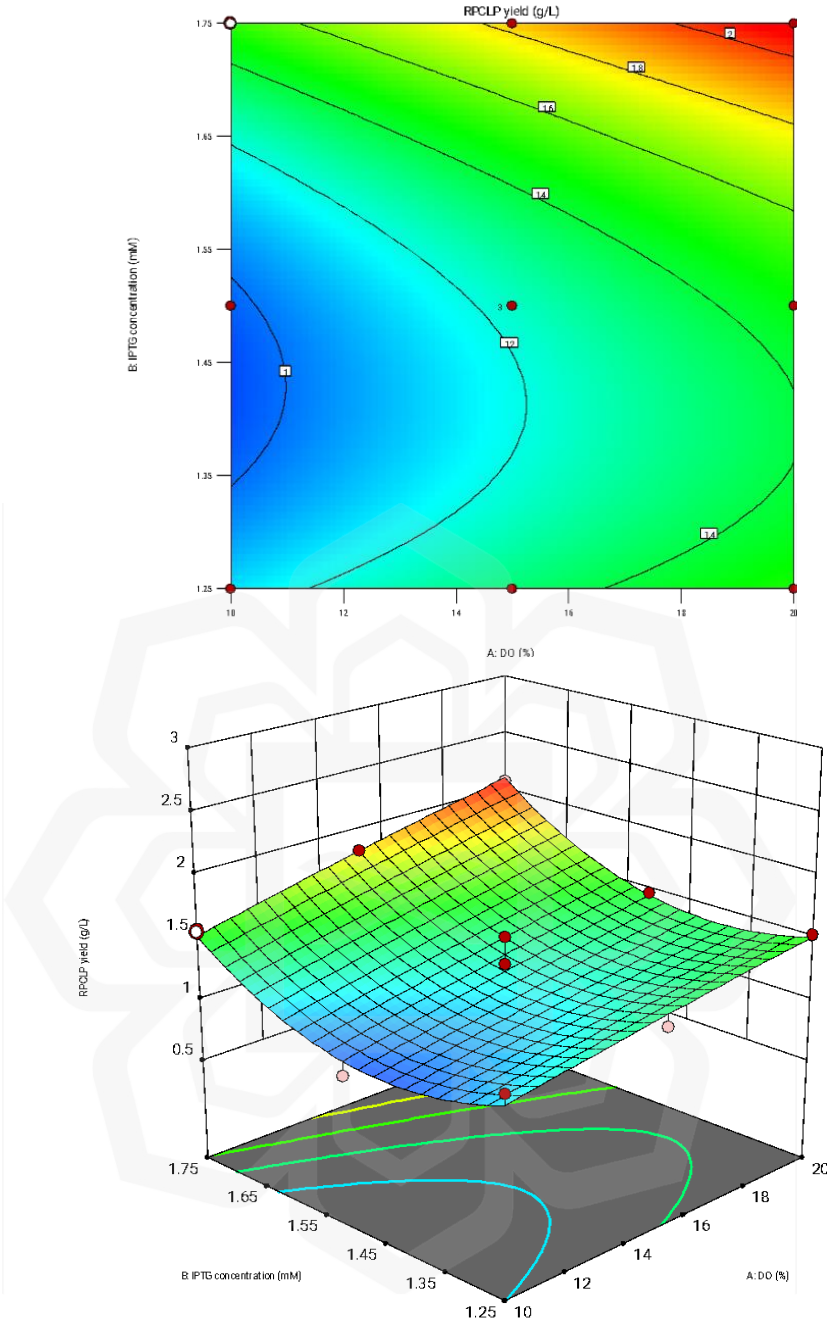


Figure 4.13 Quadratic model visualization (Contour plots and 3D surface)

Table 4.10 Statistical analysis of the quadratic model showing the significant terms A, B, and B^2 with *p-values* of 0.0243, 0.0228, and 0.0281, respectively. The model is significant and the Lack of Fit is not significant (Design-Expert v. 13).

Source of Variation	**Sum of Squares**	**Degree of Freedom**	**Mean Square**	**F-value**	**p-value (Prob >F)**	**Significance**
Model	1.06	5	0.2121	6.19	0.0335	significant
A-DO	0.3483	1	0.3483	10.16	0.0243	significant
B-IPTG concentration	0.3611	1	0.3611	10.54	0.0228	significant
AB	0.0126	1	0.0126	0.3684	0.5704	*not significant*
A^2	0.0005	1	0.0005	0.0146	0.9085	*not significant*
B^2	0.3208	1	0.3208	9.36	0.0281	significant
Residual	0.1713	5	0.0343			
Lack of Fit	0.0357	3	0.0119	0.1755	0.9048	not significant
Pure Error	0.1356	2	0.0678			
Cor Total	1.23	10				

From Table 4.10, the Model *F-value* of 6.19 means that the model is significant. There is a 3.35% probability that an *F-value* this large could be attributed to noise. *P-values* less than 0.050 are obtained when model terms are significant, whereas values greater than 0.100 show that the model terms are not significant. In this case, *A*, *B*, and B^2 are the significant model terms with *p-values* 0.0243, 0.0228, and 0.0281, respectively. There were no statistically significant interactions observed between the variables. The Lack of Fit *F-value* of 0.18 indicates that the lack of fit is not significant. There is a 90.48% probability that this Lack of Fit F-value could be attributed to noise. Unlike OFAT, the response surface methodology enables observation of the effects of interactions between factors. As such, interactions between factors were observed and measured, as can be seen from the model coefficients. According to the model, the

interactions between DO and IPTG concentrations were not significant (*p-values* > 0.05) in this fermentation.

Fit statistics for the model are presented in Table 4.11. The Predicted R^2 of 0.54 and the Adjusted R^2 of 0.72 are in agreement, i.e., the difference is less than 0.2. Adequate Precision represents the signal to noise ratio and here its value was determined to be 8.33. As a ratio greater than 4 is desirable, the value obtained indicates an adequate signal and that the model can be used.

Table 4.11 Fit statistics showing an R^2 value of 0.86 and Adequate Precision of 8.33 (Design-Expert v. 13)

Parameter	**Value**	**Parameter**	**Value**
Std. Dev.	0.1851	R^2	0.8609
Mean	1.42	Adjusted R^2	0.7218
C.V. %	13.06	Predicted R^2	0.5377
		Adequate Precision	8.3326

In this study, the residuals were shown to follow a normal distribution as shown by the normal probability plot (Figure 4.14) which follows a linear trend. The Cook's Distance values were obtained for each experimental run, as shown in Figure 4.15. Runs with a Cook distance between 0.5 and 1, as in Run 7, are considered moderately influential. In contrast, runs with a Cook's distance greater than 1 are considered to be influential as they can potentially skew the analysis results. In this case, all runs had Cook's Distance values less than 1.

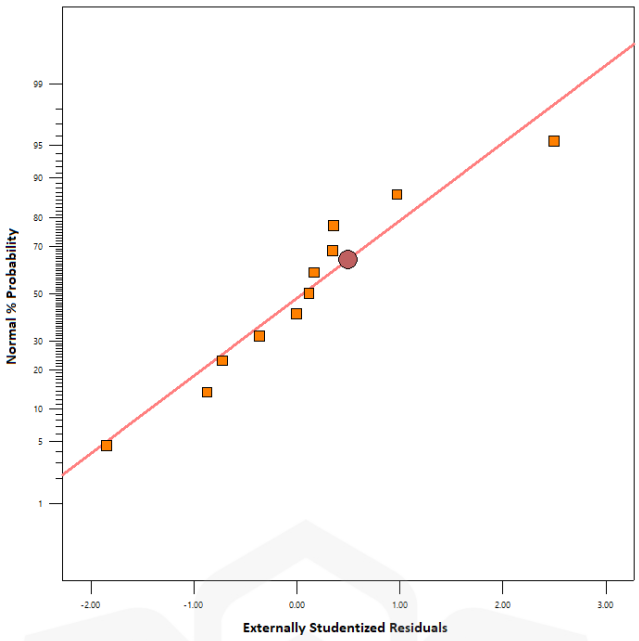


Figure 4.14 Normal probability plot (Design-Expert v. 13)

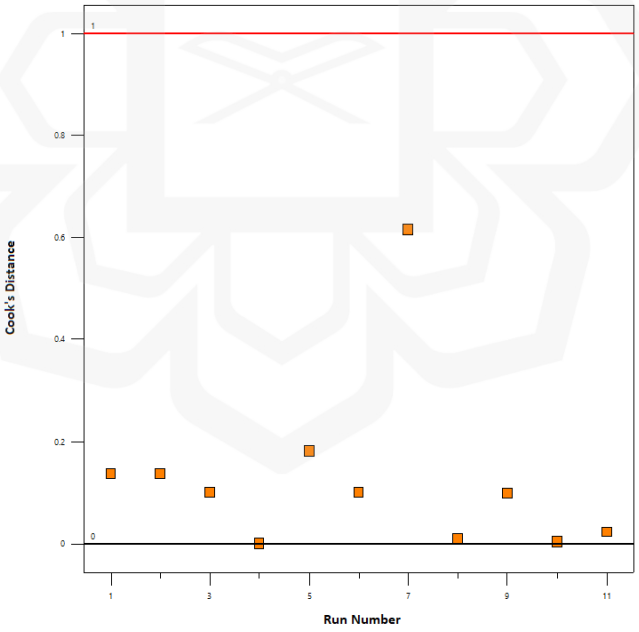


Figure 4.15 Cook's Distance for various runs (Design-Expert v. 13)

4.5.2.2 Validation of Results

Numerical optimization for maximizing RPCLP concentration was performed in the Design-Expert software version 13.0, as shown in Table 4.12, according to the FCCCD model. As all the solutions were very similar to each other, as shown in Table 4.13, and therefore only one solution was selected for the validation experiments. Under these conditions (minimum DO = 20 %, IPTG concentration = 1.75 mM), the optimal RPCLP concentration was predicted to be 2.1 g/L. The optimal conditions were validated experimentally as in Table 4.14.

Table 4.12 Numerical optimization constraints (Design-Expert v. 13)

Parameter	**Unit**	**Lower Limit**	**Upper Limit**	**Goal**	**Weight**
RPCLP concentration	g/L	0.860407	2.08189	Maximize	+++++
Minimum DO	%	10	20	In range	(5)
IPTG concentration	mM	1.25	1.75	In range	

Table 4.13 Solutions 1 to 5 of the Numerical Optimization (Design-Expert 13)

Number	**DO tension** (%)	**IPTG concentration (g/L)**	**Theoretical RPCLP concentration (g/L)**	**Desirability**
1	19.697	1.747	2.086	1.000
2	19.922	1.748	2.104	1.000
3	19.747	1.747	2.088	1.000
4	20.000	1.750	2.115	1.000
5	19.740	1.750	2.100	1.000

Table 4.14 Validation results of the optimization using FCCCD

	DO tension %	IPTG concentration (g/L)	Soluble, recombinant collagen-like protein concentration (g/L)
Experimental value (Solution #4)	20	1.75	2.129
Expected value			2.115

The fermentation growth profile was also analyzed, and it was noted that the cells experienced rapid growth during the first three (3) hours. The absence of a lag phase indicates that the cells were able to adjust quickly to the fermentation medium and optimal process conditions. This rapid growth is explained by the *E. coli* growing at its optimal temperature of 37 °C, controlled optimal pH 7.0, and the availability of a well-defined medium with adequate concentration of carbon and nitrogen sources, as well as supplements and micronutrients. The protein accumulation profile at optimal process conditions is not shown here and is discussed separately under the kinetic modeling in Section 4.6.

4.5.3 Conclusion

The Taguchi L9 Orthogonal Array design was used to screen and rank the most significant factors in the protein production in the 2 L bioreactor. These were, in order of decreasing significance, dissolved oxygen tension, IPTG concentration, induction point, and induction temperature. Response Surface Methodology (face-centered

central composite design) was used to optimize the two most significant factors affecting the CLP concentration and develop a model for their relationship. Statistical analyses show that the model is significant, with minimum dissolved oxygen (*A*) and IPTG concentration (*B*), as well as B^2, being the significant model terms, as indicated by *p-values* less than 0.0500. On the other hand, the Lack of Fit was not significant. There were no statistically significant interactions observed between the factors. All the data were within the normal probability distribution. Validation of results using the optimization solution shows that the model was able to predict the RPCLP concentration well, with a 2.13 g/L RPCLP protein concentration.

4.6 OBJECTIVE 3: KINETIC MODELING

To facilitate the fermentation process optimization for high RPCLP concentration and scale-up, insight into its fermentation kinetics is necessary. These could be further split into kinetics of substrate utilization using glucose as the carbon source (in M9 casamino-acids medium, kinetics of biomass (*E. coli*) growth, and kinetics of RPCLP production. Besides that, fermentation kinetics can be used to obtain useful kinetic parameters such as various types of yield constants. In this way, fermentation kinetics can address how the process yield is improved after process transfer from the shake flask to the benchtop bioreactor scale. For these purposes, two unstructured growth models were proposed- the Monod model and the logistic model to evaluate the experimental data and approximate their respective kinetic parameters.

The growth curve of the recombinant *E. coli* BL21(DE3) exhibits the distinctive bacterial growth characterized by a short lag, a clear log phase that is indicative of rapid growth, and a stationary stage where the specific growth rate stabilizes after eight (8) hours of fermentation (Figure 4.16). Similar growth patterns have been reported for recombinant protein-producing *E. coli* (Ariff et al., 2015). The time profile of substrate concentration (Figure 4.17) shows a marked drop in glucose levels following the inoculation of the medium. The product concentration profile with time establishes the accumulation of RPCLP, but the increase is notably low and incompatible with the expected results (Figure 4.18). Induction of the protein expression with IPTG starts several hours after the inoculation, at the mid-log phase. It was anticipated that product accumulation would increase steadily after the induction. However, the product concentration curve does not show this gradual increase. In general, however, product concentration is proportional to the cell growth during the fermentation period. The highest biomass concentration attained was 8.25 g/L, whereas the maximum RPCLP concentration attained was a modest 2 mg/L. A possible explanation for low concentration values could be due to the glucose and collagen assays themselves or the degradation of samples during the fermentation. Some of the results detailed in this section have been published recently, and the publication is cited in this section.

4.6.1 Kinetic Modeling using MATLAB Functions

From Figure 4.16, it can be observed that the recombinant *E. coli* BL21(DE3) growth is well represented by the Monod-Leudeking Piret (MLP) model. On the other hand, the glucose utilization and RPCLP production models (Figure 4.17- 4.18) studied did not fit the data well (Gameil et al., 2023). Similarly, for the Logistic-Luedeking Piret (LLP)

model (Figure 4.19), it fits the experimental data for growth well but does not fit the glucose utilization and RPCLP production (Figures 4.20 - 4.21).

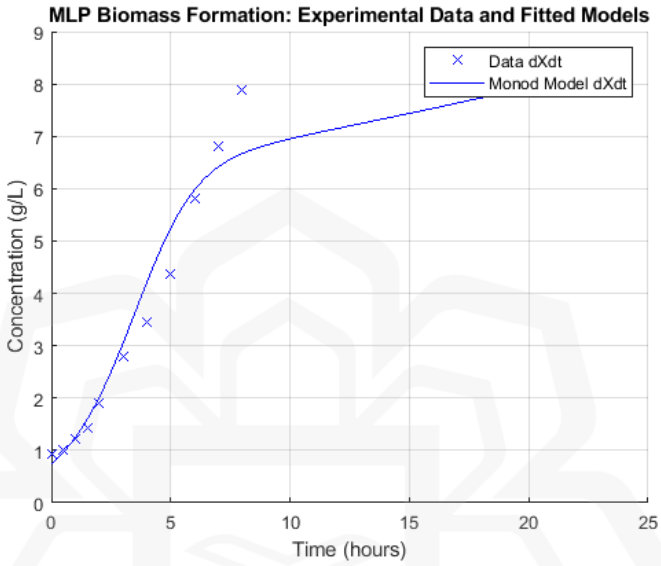


Figure 4.16 Biomass formation data fitting using MLP model (MATLAB)

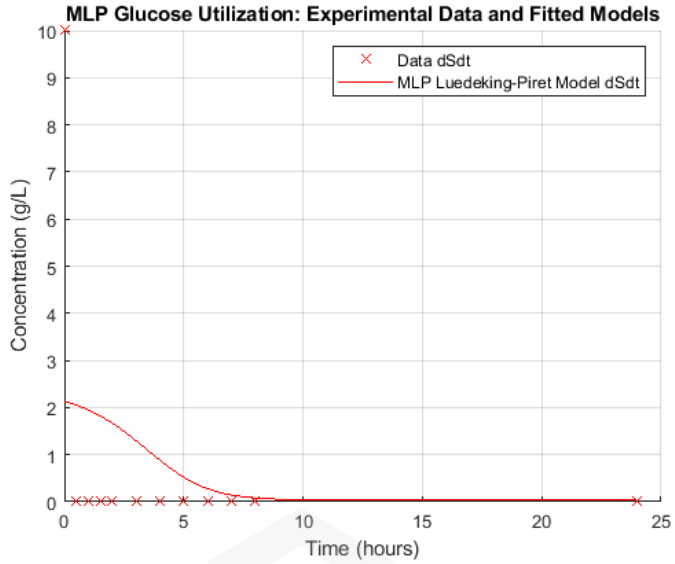


Figure 4.17 Glucose utilization data fitting using MLP model (MATLAB)

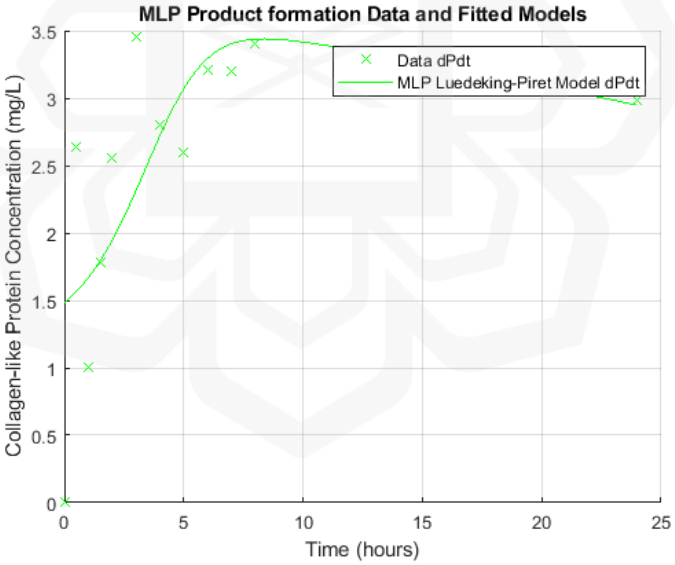


Figure 4.18 Product formation data fitting using MLP model (MATLAB)

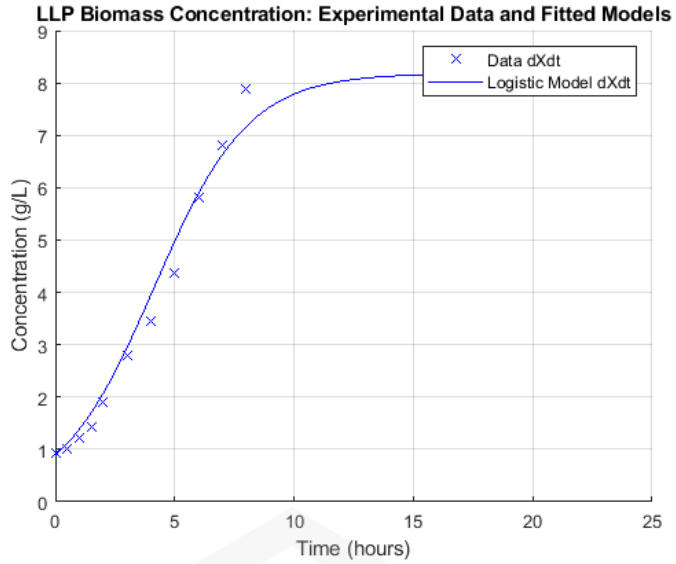


Figure 4.19 Biomass formation data fitting using MLP model (MATLAB)

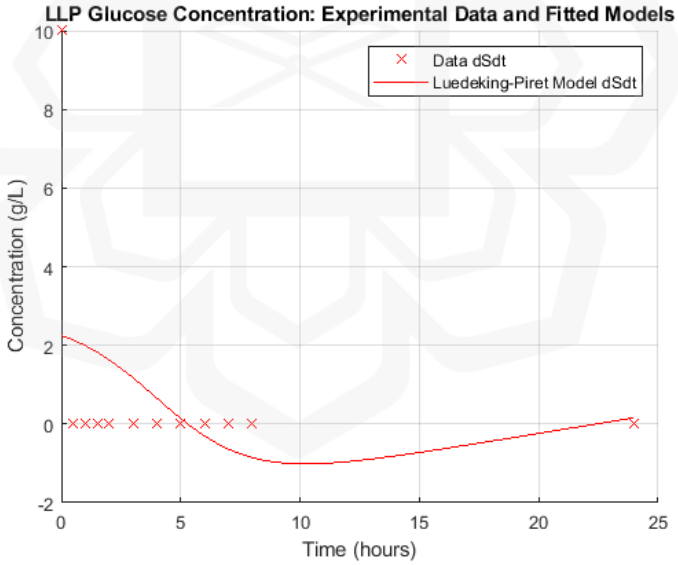


Figure 4.20 Glucose utilization data fitting using LLP model (MATLAB)

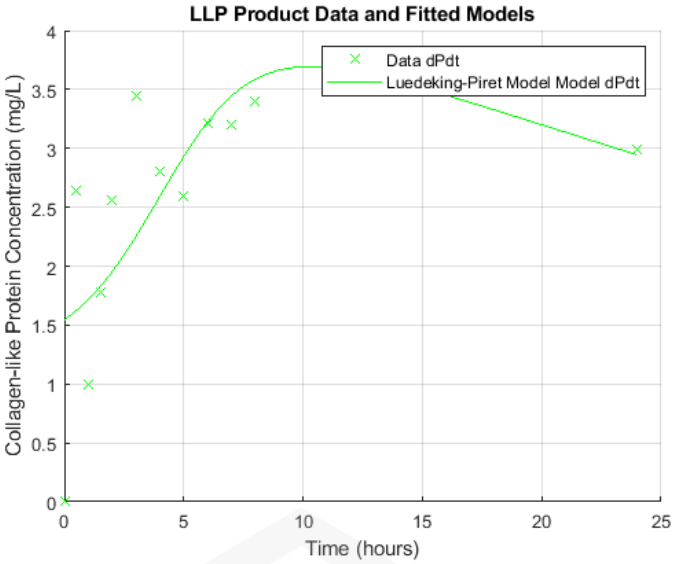


Figure 4.21 Product formation data fitting using LLP model (MATLAB)

4.6.2 Statistical Analysis

The coefficient of determination (R^2) values observed with the experimental and predicted results, in addition to the statistical analyses using ANOVA, are presented in the subsequent figures (Figures 4.23 – 4.26) and tables (Tables 4.15 – 4.20). The data follows a normal probability distribution (Figures 4.23 and 4.26). The highly significant coefficient of determination (R^2) values observed with the experimental and predicted results, as shown by Tables 4.15 and 4.18, in addition to the statistical analyses using ANOVA (Tables 4.16 and 4.19), indicate that the regression models were significant with large F values. The p-values were less than 0.05 for both variables in the Logistic model, indicating both terms were significant. On the other hand, the intercept for the Monod model had a p-value greater than 0.05 and this indicates that that variable was not statistically significant.

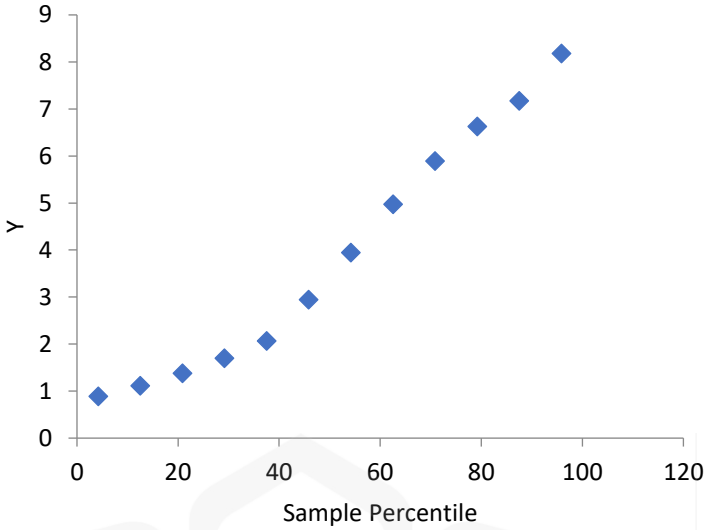


Figure 4.22 Monod normal probability plot

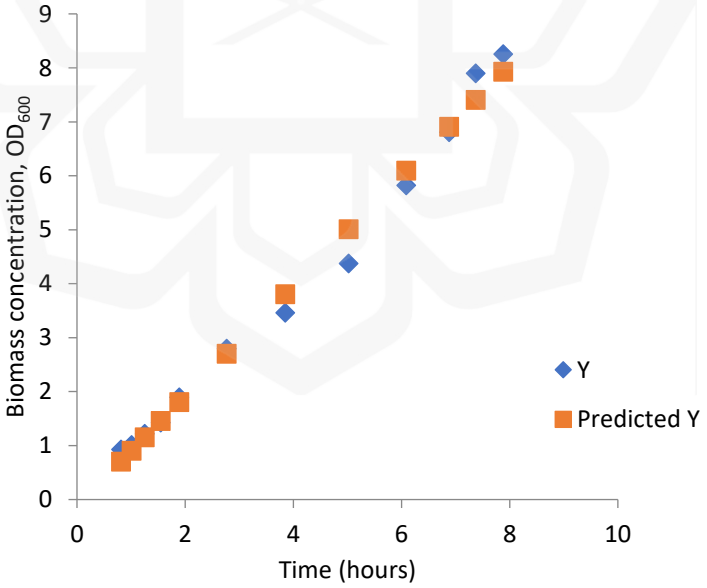


Figure 4.23 Monod model line fit plot

Table 4.15 Regression statistics (Monod model) showing R^2 value of 0.96

Parameter	Value
Multiple R	0.981461
R Square	0.963266
Adjusted R Square	0.959592
Standard Error	0.526225
Observations	12

Table 4.16 ANOVA analysis (Monod model)

	df	SS	MS	F	Significance F
Regression	1	72.61324231	72.61324231	262.2246	1.67194E-08
Residual	10	2.769124125	0.276912412		
Total	11	75.38236643			

Table 4.17 Fit statistics and confidence intervals (Monod model)

	Coefficients	Standard Error	t Stat	p-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.2998	0.2679	1.1189	0.2893	-0.2972	0.8968	-0.2972	0.8968
X Variable	0.9347	0.0577	16.1934	1.67E-08	0.8061	1.0633	0.8061	1.0633

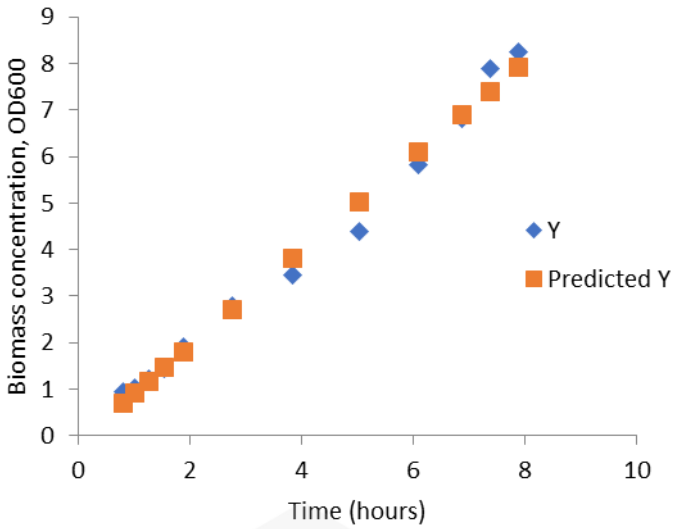


Figure 4.24 Logistic model line fit plot

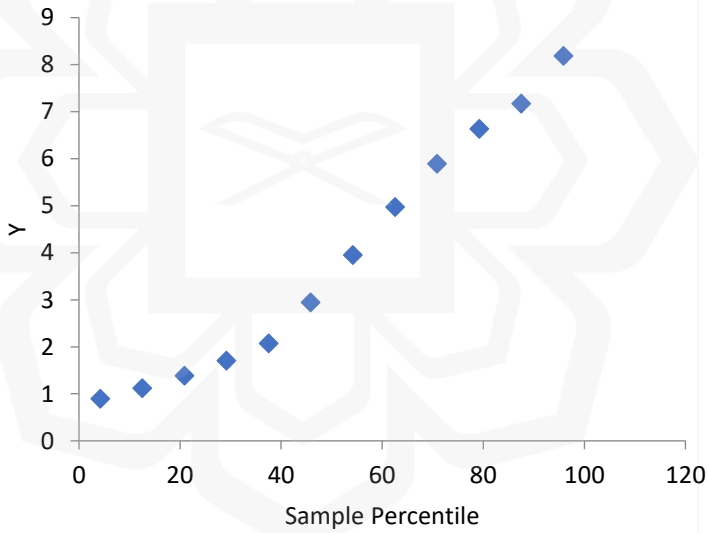


Figure 4.25 Logistic normal probability plot

Table 4.18 Regression statistics (Logistic model) showing a high R^2 value of 0.99

Parameter	Value
Multiple R	0.99384
R Square	0.98771
Adjusted R Square	0.98648
Standard Error	0.30122
Observations	12

Table 4.19 ANOVA analysis (Logistic model)

	df	SS	MS	F	Significance F
Regression	1	72.93385	72.93385	803.81	7E-11
Residual	10	0.907352	0.090735		
Total	11	73.84121			

Table 4.20 Fit Statistics and confidence intervals (Logistic model)

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.3276	0.1534	2.13603	0.0584	-0.014	0.6694	-0.0141	0.6694
X Variable	0.9367	0.0330	28.3515	6.93E-11	0.8631	1.0103	0.8631	1.0103

Both the Monod and the logistic model were suitable for fitting the growth data, with significant coefficients of determination (R^2) values of 0.96 and 0.99, respectively,

as shown in Tables 4.17 and 4.20. However, Luedeking-Piret models were a poor fit for the glucose utilization and RPCLP formation data (correlation coefficient ≤ 0.7) (Statistical analysis results not shown). The results could be attributed to the observed scattering of experimental data for substrate and product concentration. This scattering could be attributed to the selectivity and/or sensitivity of methods in measuring low concentrations of the growth-limiting substrate as well as the assay for product concentration (Gameil et al., 2023).

4.6.3 Kinetic Modeling using the MATLAB Curve Fitter app and Statistical Analysis

In using the MATLAB Curve Fitter app, the logistic growth model is given in Figure 4.26. The logistic model growth kinetic parameters were estimated to be k = 0.53 h^{-1}, X_{max} = 8.54 g/L, and X_0 = 0.72 g/L. As for the goodness of fit, the results of the R-squared value and Adjusted R-squared values for the model were both 0.99 as shown in Table 4.21.

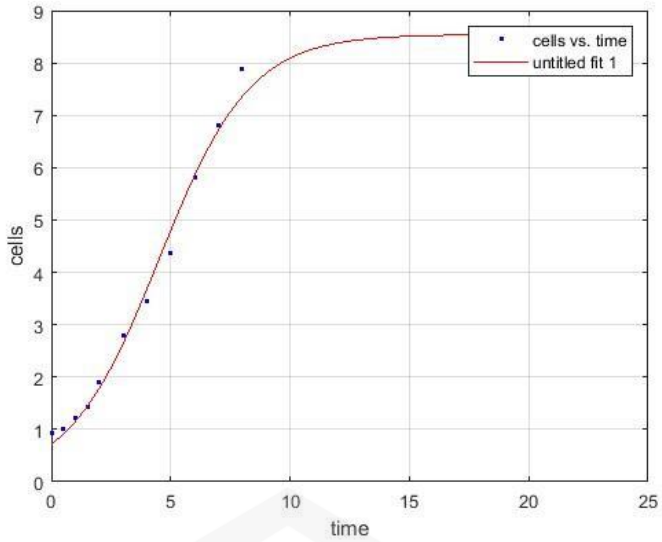


Figure 4.26 Curve Fitting of growth data using the Logistic Model (MATLAB R2022a Curve Fitter)

Table 4.21 The goodness of fit for Logistic Model in MATLAB Curve Fitter

Parameter	**Value**
SSE	0.7017
R^2	0.99156
Adj. R^2	0.9897
RMSE	0.27923

Using the MATLAB Curve Fitter app, the Logistic-Luedeking-Piret product formation model is shown in Figure 4.27. It estimated kinetic parameters to be μ_{max} = 0.70 h^{-1}, X_{max} = 6.69 g/L, and X_0 = 6.57 g/L. Fit statistics are given in Table 4.22. The R-squared value for the model was 0.48, indicating that it was a very poor fit for the data.

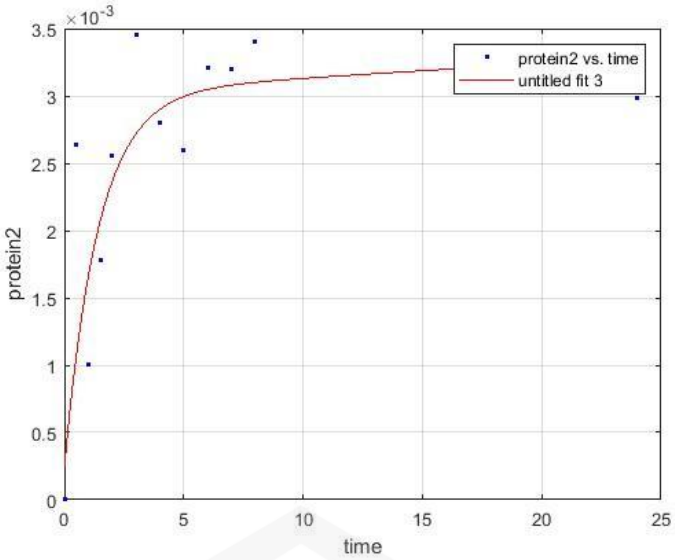


Figure 4.27 Luedeking-Piret product formation model - MATLAB Curve Fitter

Table 4.22 Goodness of Fit (Luedeking-Piret product formation model - MATLAB Curve Fitter)

Parameter	**Value**
SSE	4.008 e-0.6
R^2	0.6682
Adj. R^2	0.4786
RMSE	0.007567

The Luedeking-Piret model for glucose utilization kinetics is shown in Figure 4.28. As for the goodness of fit results, the R-squared and Adjusted R-squared values for the model were 1, as shown in Table 4.23.

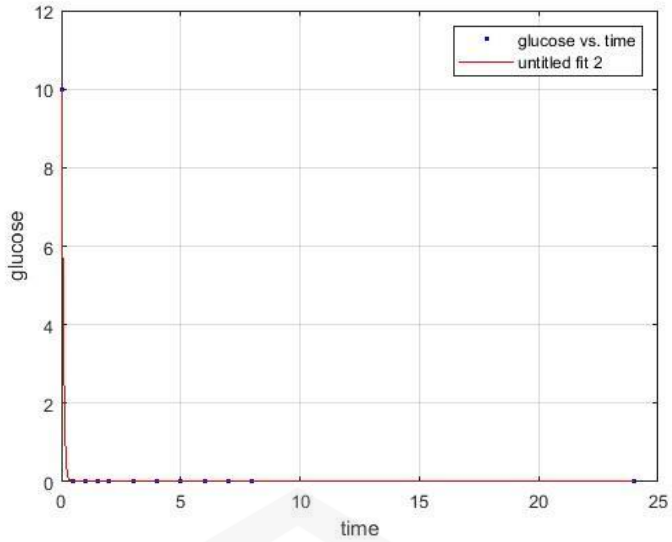


Figure 4.28 Luedeking-Piret glucose consumption model - MATLAB Curve Fitter

Table 4.23 The goodness of fit (Luedeking-Piret model of glucose consumption - MATLAB Curve Fitter)

Parameter	**Value**
SSE	0.0002785
R2	1
Adj. R^2	1
RMSE	0.006813

Fitting using the MATLAB Curve Fitting toolbox achieved correlation coefficient values of 0.7 and 1 for the Luedeking-Piret models product formation and glucose utilization kinetics respectively.

4.6.4 Estimation of Kinetic Parameters

All the nonlinear kinetic models mentioned have been fitted to the data of the relation between bacterial growth rate and concentration of the substrate with glucose as the limiting factor of the culture of *Escherichia coli*. The parameters of these models have been estimated using the Gauss-Newton Method of estimation. Estimated parameters for the models along with their confidence intervals have been presented in Table 4.24 for the MATLAB functions and Table 4.25 for the MATLAB Curve Fitter app.

The Monod kinetic parameter values of μ_{max} = 1.75 and K_s = 4.82 g/L, were in accord with those reported by other studies for *E. coli BPPTCC EGRK2* in glucose substrate, specifically μ_{max} = 1.694 and K_s = 6.629 g/L (Limoes, 2016). A study by Kovarova showed *E. coli* ML30 on 500 mg/L glucose substrate reported kinetic parameters to be μ_{max} = 0.76 h^{-1} dan K_s = 0.0328 mg/L. The logistic model growth kinetic parameters were estimated to be k = 0.507 h^{-1}, and X_{max} = 8.18 g/L. Poccia et al. (2014) also estimated the growth model of *E. coli* JM 109 and BL 21 strains on multi-substrate media, and they obtained μ_{max} = 0.53 h^{-1} for JM109, while for BL 21 the value of μ_{max} was 0.76 h^{-1} which is considerably higher. From these data, it can be seen that μ_{max} and K_s values for this *Escherichia coli* BL21(DE3) on glucose substrate are much higher than other *E. coli* strains in general. This could be due to it being recombinant and having a high substrate consumption. In addition, the high value of μ_{max} may be due to the glucose content (10 g/L) in the fermentation medium as *E. coli* grows faster on glucose media than other carbon and energy sources and can achieve a doubling time (t_d) of 20 minutes at 37 °C. However, the high K_s value also indicates that it requires a high substrate concentration to achieve a high growth rate. One of the observations was that the cell growth was high even though the glucose concentration depleted sharply, according to experimental data (Figures 4.26 and 4.28). This might be attributed to casamino acids found in the M9 medium used which can serve as a nitrogen source as

well as a carbon source. Casamino acids are a rich, complex source of nutrients that promote cell growth (Baez, 2018). Unfortunately, the monitoring of amino acid concentrations throughout the study could not be carried out during the experiment to elucidate the uptake of casamino acids by the cells and its effects on growth and product formation.

Concerning Luedeking-Piret model parameters, under both MLP and LLP models, the growth and non-growth associated product formation coefficients (α and β) are nonzero suggesting that the CLP biosynthesis is partially associated with growth (Wu et al., 2014). A negative β suggests that the product is being consumed or degraded. However, this could be a result of experimental data scattering which could be attributed to the selectivity and/or sensitivity of methods in measuring low concentrations of the growth-limiting substrate as well as the assay for product concentration. The low product concentrations also suggest that the product may be unstable and was degraded (Gameil et al., 2023).

On the other hand, the values of some of the kinetic parameters approximated by the MATLAB Curve Fitter approach were very far from those expected. This could be due to the inability of the MATLAB Curve Fitting App to consider all experimental data and solve all model equations provided simultaneously, and instead, it has solved each independently, resulting in three separate sets of kinetic parameters as shown in Table 4.25. Using the MATLAB Curve Fitter app, the Logistic-Luedeking-Piret substrate utilization kinetic model estimated parameters to be μ_{max} = 19.68 h^{-1}, X_{max} = 8.55 g/L, S_0 = 14.71 g/L and X_0 = 1.94 g/L. The experimental S_0 was 10 g/L, so this value for it was not far off. Similarly, the values of X_0 and X_{max} obtained from the model were similar to those determined from experimental data. However, the value for μ_{max} was very different from the expected range of values from the literature as well as experimental data. The value of m (3.568) was much larger than the value of n (7.096 x 10^{-5}), suggesting that much of the substrate utilization is used for growth (Gameil et al., 2023). Along with the much higher magnitudes of α (0.985) compared to β (1.631x10^{-6}) (refer to Table 4.25), the findings support a growth-associated process for CLP production and glucose consumption which is consistent with those of many other recombinant protein fermentations in existing literature (Hua et al., 2006; Rajendran & Thangavelu, 2008; Vinayagam et al., 2015). As for recombinant human-like collagen production in *E. coli*, using 30 g/L glucose in batch culture in a 6.5 L working volume fermenter reported 0.65 h^{-1} maximum specific growth rate, and yield coefficient of cells from the substrate, Y_{XS} = 0.51 g/g. For the fed-batch process, the yield coefficient of cells from the substrate, Y_{XS} = 0.51, and the α *and* β values were 0.526 and 5.88 x 10^{-3} respectively (Hua et al., 2006).

Table 4.24 Kinetic parameters estimated by the models using MATLAB functions

Kinetic Parameter	**Units**	**Microbial growth model**		**Substrate utilization model**		**Product formation model**	
		Monod	Logistic	MLP	LLP	MLP	LLP
μ_{max}	h^{-1}	1.75	-	1.75	-	1.75	-
K_S	g/L	4.82	-	4.82	-	4.82	-
Y_{XS}	g/g	-	-	2.69	1.796638	2.69	1.796638
Y_{PS}	g/g	-	-	0.02	0.02	0.02	0.02
α	g/g	-	-	-	0.000365	0.0004	0.000365
β	g/g.h	-	-	-	-0.000008	-0.000009	-0.000008
k_e	g/g.h	-			-0.012240	-0.004831	-0.012240
k		-	0.507146	-			
γ		-	0.122203	-		-	0.122203

Table 4.25 Model kinetic parameters as determined for each model separately by MATLAB Curve Fitter and their 95% confidence intervals

Parameter/Model	Logistic	Luedeking-Piret Substrate	Luedeking-Piret Product
μ_{max}	0.5267	19.68	0.7036
	(0.4499, 0.6035)	(-6.722e+04, 6.726e+04)	(-0.4402, 1.847)
X_0	0.7197	1.944	6.574
	(0.5163, 0.923)	(-7.482e+04, 7.482e+04)	(-3.061e+05, 3.062e+05)
X_{max}	8.544	8.552	6.694
	(7.961, 9.127)	(-5.619e+05, 5.619e+05)	(-3.173e+05, 3.174e+05)
S_0	NA	14.71	NA
		(-3.019e+05, 3.019e+05)	
coefficients	NA	m= 3.568	a = 0.985
		(-2.021e+05, 2.021e+05)	(-823.3, 825.3)
	NA	n = 7.096e-05	b = 1.631e-06
		(-4.663, 4.663)	(-0.07733, 0.07734)

4.6.5 Conclusion

The Monod and logistic-based models resulted in reasonable fits for the growth experimental data. The highly significant coefficient of determination values observed with the experimental and predicted results (R^2 values 0.96 and 0.99, respectively), in addition to the statistical analyses using ANOVA, indicate that the regression models

were significant. The growth-associated and non-growth-associated product formation constants, α and β, indicate that the RPCLP production is partially associated with growth. As the Luedeking-Piret model was inadequate to describe the product formation and substrate consumption, improvements to the model are necessary (Gameil et al., 2023). Biomass concentration and product concentrations followed a similar time profile in the 2 L bioreactor. As a whole, the kinetic parameters estimated give useful indicators that can be used to further enhance the fermentation process design, optimization, and scale-up.

4.7 OBJECTIVE 4: SCALE-UP FROM THE 2 L TO 7.5 L BIOREACTORS

To achieve Objective 4 of this study, experiments were run to determine the change in k_La upon scale-up when the agitation speed was varied while keeping other fermentation process parameters constant. The resulting data helped in determining the steps that need to be taken to keep the k_La value constant at the higher scale, thereby ensuring sufficient oxygen mass transfer, and the RPCLP concentration was determined at the higher scale. The aeration rate at the 7.5 L scale was fixed at 2 L/min (0.4 vvm).

4.7.1 Effect of Scale-up Based on Constant Oxygen Mass Transfer Coefficient

The measurement of the oxygen mass transfer coefficient to determine k_La at various agitation rates was carried out through the dynamic method in two stages, namely a "gas-out" and a "gas-in" phase, which were used to measure oxygen uptake rate (*OUR*)

and the oxygen transfer rate (*OTR*), respectively. In the "gas-out" stage, the airflow inlet was interrupted, and the dissolved oxygen (*DO*) concentration dropped due to cellular respiration, as recorded by the *DO* sensor. The "gas-out" stage was kept short and dissolved oxygen (*DO*) concentration was maintained above the critical oxygen concentration, C_{crit} (assumed to be 20% DOT) to ensure that the oxygen uptake rate (*OUR*) was almost constant, and no cell damage occurs. During the "gas-in" stage, the airflow inlet was restarted, and the DO concentration increased. The dissolved oxygen (DO) concentration was stable at 100%. Air supply was stopped at this condition and the decrease in DO was observed, as depicted in Figure 4.29. The DO concentration decreased from 100% to ~30% during this period. Then, the air supply was reconnected, and the DO was recorded. It was observed that the DO increased to 60-85% and remained stable. The results were plotted and shown in Figure 4.30-4.31.

$$ln\left(\frac{C_{AL}^{*}-C_{AL1}}{C_{AL}^{*}-C_{AL2}}\right) = k_L at$$ Equation 4.3

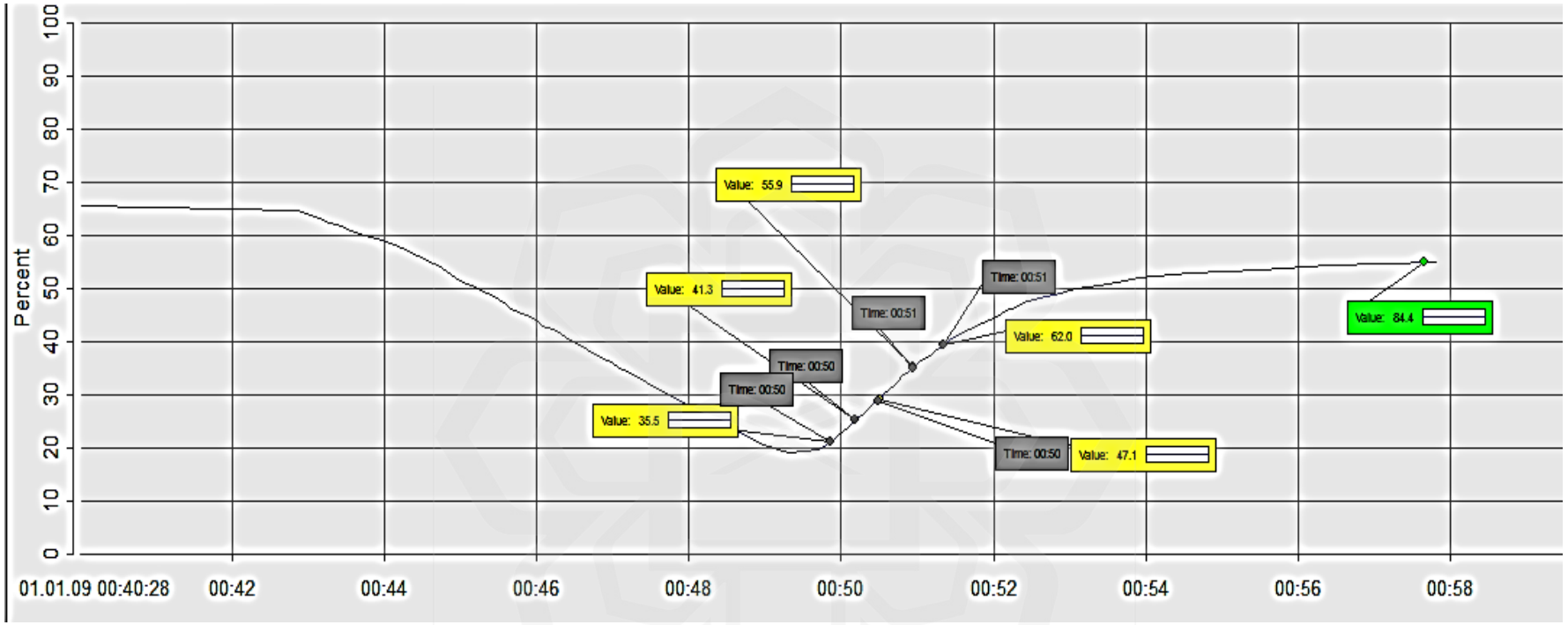


Figure 4.29 DO profile during the dynamic gassing out experiment

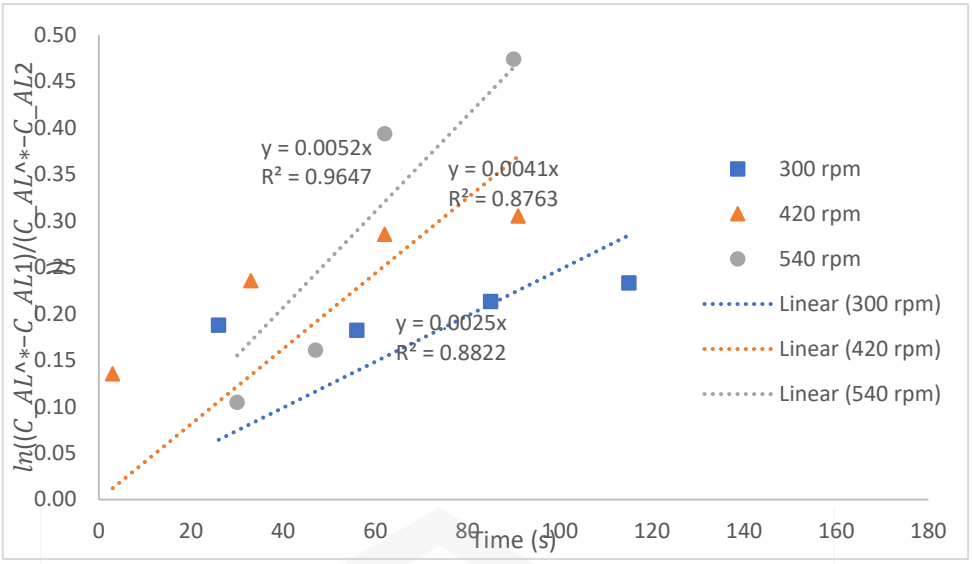


Figure 4.30 Determination of k_La values in the 2 L bioreactor at various agitation speeds (0.0025 s^{-1}at 300 rpm, 0.0041 s^{-1} at 420 rpm and 0.0052 s^{-1} at 540 rpm)

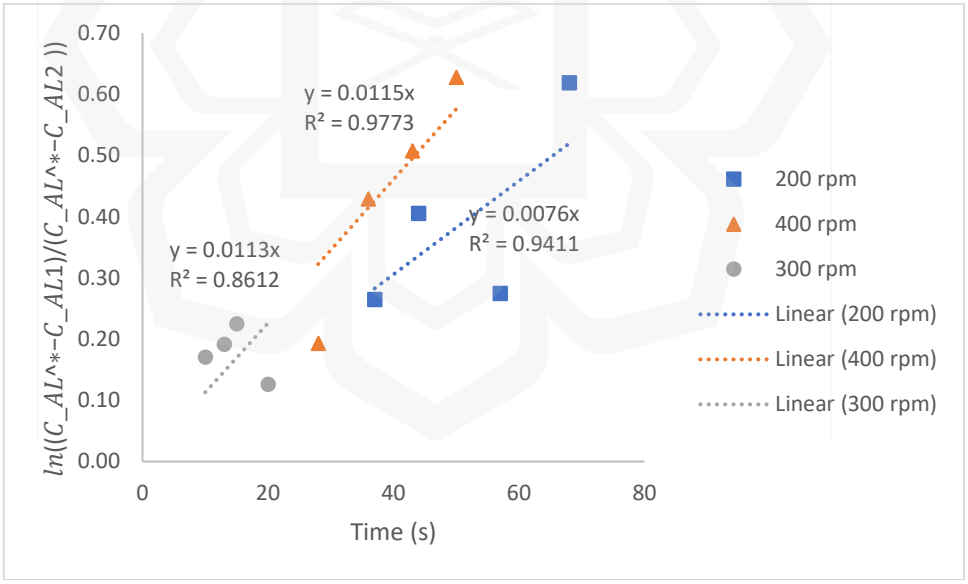


Figure 4.31 Determination of k_La values in the 7.5 L bioreactor at various agitation speeds (0.0076 s^{-1} at 200 rpm, 0.0113 s^{-1} at 300 rpm, and 0. 0115 s^{-1}at 400 rpm)

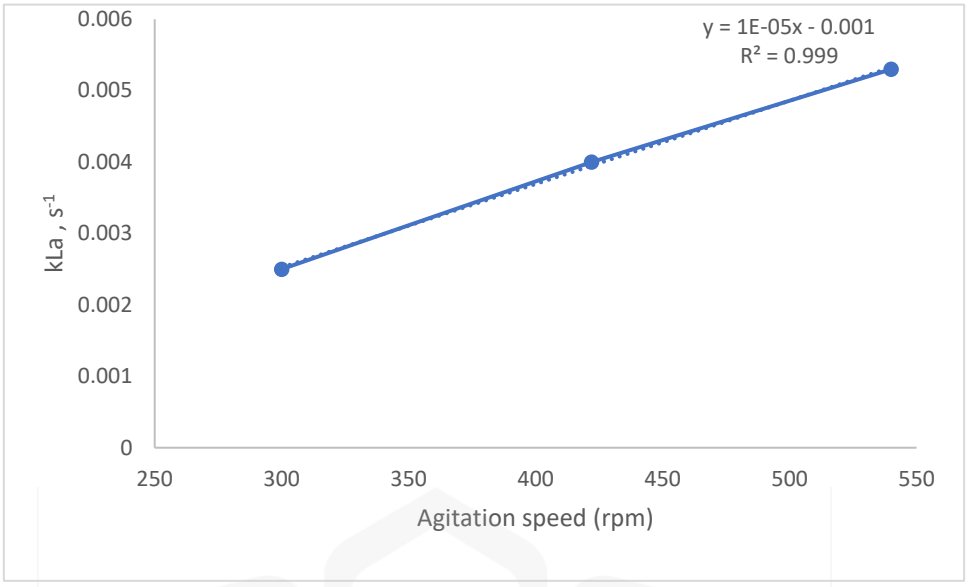


Figure 4.32 Effect of increasing the agitation speed on k_La value (2 L STR)

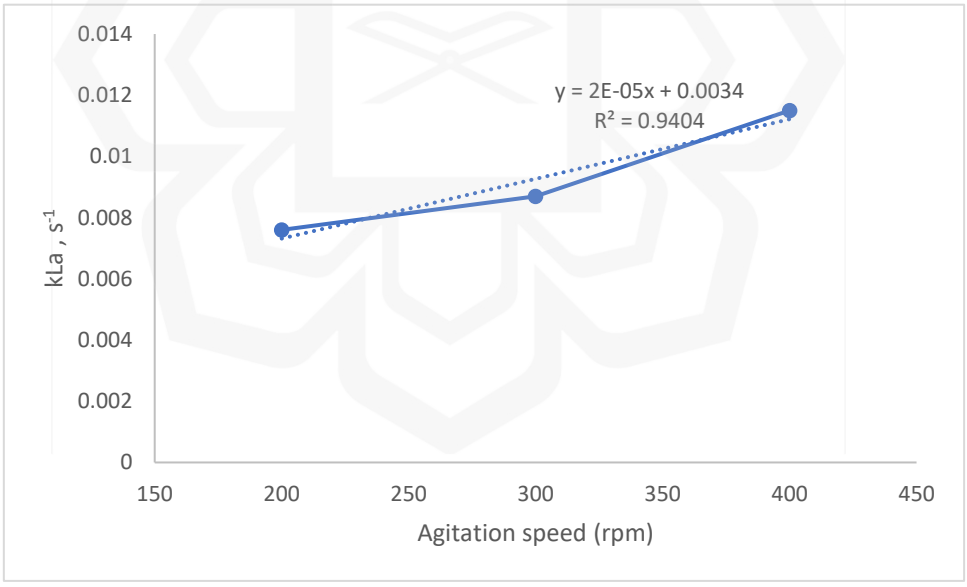


Figure 4.33 Effect of increasing the agitation speed on k_La value (7.5 L STR)

The graphs in Figures 4.32 and 4.33 show the k_La values (slope of each given equation of the form $y=mx$) as a linear function of various agitation speeds in the 2 L and 7.5 L bioreactor. The slopes of these graphs represent the volumetric mass transfer coefficient. It is apparent that the k_La values are affected significantly by the variation of the agitation speed and increasing the agitation speed increased the k_La values proportionally. As the agitation speed was increased from 300 rpm to 540 rpm, in the 2 L bioreactor, the k_La values also increased from 0.0025 to 0.0052 s^{-1}. Using this range and the correlation for the large-scale k_La and agitation values (shown in Figure 4.33), the range of agitation speeds for the higher scale was calculated to be 0 to 90 rpm. An agitation speed of 90 rpm corresponds to 0.2545 m/s impeller tip speed. By contrast, an agitation speed range of 300 rpm to 540 rpm in the 2 L bioreactor corresponds to 0.72 to 1.30 m/s impeller tip speed, using the equation given in Figure 4.32. Thus, using constant k_La as a criterion for scale-up has resulted in a two- to five-fold reduction in impeller tip speed. Considering this as the maximum tip speed means no increase in shear would be observed. The impeller agitation speed values obtained were similar to those reported by (Zaslona et al., 2015) who found that 135 rpm and 0.8 vvm were optimal conditions at the 3 L scale for the production of 1,3- β glucanase using *E. coli* BL21 (refer to Table 2.16).

In this study, the constant k_La criterion failed to provide sufficient oxygen levels to sustain the fermentation beyond the first 30 minutes of the experiment. This could be attributed to erroneous estimates of k_La that significantly underestimated the agitation speed required to sustain the fermentation's dissolved oxygen levels. Perhaps, it would be more useful to consider the change in aeration flow rate for the 7.5 L bioreactor and estimate the k_La value by using empirical correlations. The RPCLP concentration obtained from the fermentation was not determined as the experiment was discontinued after dissolved oxygen levels fell below 5% within the first 30 minutes of the

experiment. Thus, using the k_La scale-up criterion, at the constant value of 0.0052 s^{-1}, the impeller tip speed decreased significantly, and the scale-up was not successful.

4.7.2 Effect of Scale-up Based on Constant Tip Speed

The scale-up using the constant tip speed criterion (Espinosa Perez et al., 2018) was performed according to Equation 4.4, where the agitation rate, N_2, in rpm, for the large-scale bioreactor is a function of tip speed and impeller diameter.

$$N_2 = N_1 \frac{D_{i,1}}{D_{i,2}} \qquad \text{Equation 4.4}$$

where N_1 is the agitation speed for the small-scale bioreactor, N_2 is the agitation speed for the large-scale bioreactor, $D_{i,1}$ is the impeller diameter for the small-scale bioreactor, and $D_{i,2}$ is the impeller diameter for the large-scale bioreactor. The aeration rate for the 7.5 L bioreactor was fixed at 2 L/min (0.4 vvm).

From the values of the impeller diameters for both the 2 L and 7.5 L bioreactors (as shown in Table 4.26) and using 300 rpm - 540 rpm as the agitation speed in the 2 L bioreactor, the agitation speed for the large bioreactor was determined to be 255.56 – 460 rpm, corresponding to a constant 0.73 m/s impeller tip speed. As our fermenter has an impeller tip speed that is much lower than the recommended guidelines of around 3 m/s discussed in Chapter 2, it is held that shear effects would not damage our cells and protein product. This agitation speed range corresponds to a k_La range of *0.0025- 0.0052* s^{-1}. While the constant impeller tip speed criterion does possess some advantages in the case of scale-up involving shear-sensitive microorganisms, there were cases where it was not inadequate in scaling up aerobic fermentations (Espinosa Perez et al., 2018).

Table 4.26 Characteristics of Bioreactors at 2 L and 7.5 L scale

	Small scale	**Large Scale**
V_T, Vessel Volume (m^3)	0.002	0.0075
VL, Working volume (m^3)	0.001	0.005
VL / V_T	0.5	0.6667
Vessel Height, H (m)	0.240	0.465
Vessel internal diameter, D (m)	0.115	0.150
H_T/D_T ratio	2.1	3.1
Impeller diameter, D_i (m)	0.046	0.054
Impeller type	6-blade Rushton turbine	6-blade Rushton turbine
D_i/D_T ratio	0.4	0.36
Number of impellers	2	2
Power (W)	150	150

4.7.3 Validation of Choice of Scale-up Strategy

To validate the scale-up strategy, the experimental results of the small- and large-scale bioreactors were examined. The recombinant *E. coli* BL21(DE3) growth and RPCLP expression were similar at both scales, as shown by Table 4.27, hence validating the selected scale-up strategy.

Table 4.27 Validation results of the scaleup, showing the RPCLP concentration in the 2 L and 7.5 L bioreactors, which were approximately 2.1 and 2.5 g/L, respectively.

	Concentration of RPCLP per L cell culture (g/L)
Experimental result (7.5 L)	2.4927
Experimental result (2 L)	2.129

Fermentation was carried out in the 7.5 L bioreactor using the estimated agitation and flow rate values for the constant k_La criterion (n = 300 rpm and Q = 2 l min^{-1} or 0.4 vvm). There was a sharp increase in the optical density and consequently the DCW) from the beginning of the fermentation until the end. From the log phase, the maximum value of the specific growth rate was determined to be μ_{max} = 1.90 h^{-1}. After seven (7) hours, the fermentation was stopped, and the maximum optical density reached was OD_{600} = 10.36. From the standard curve (Appendix VII), this final OD_{600} corresponds to a biomass concentration of approximately 8.18 g/L dry cell weight. This is considerably higher when compared to (Peng et al., 2012) who reported 5.3 to 5.8 g/L wet cell pastes from the batch fermentation of constructs *of Streptococcus pyogenes* collagen-like protein 2 using a defined medium. This cell density can be further enhanced for the pilot scale by developing and establishing a fed-batch mode of operation. Once more, compared to (Peng et al., 2012) who reported 0.7 to 1.0 g/L volumetric productivity for the batch fermentation, the results are higher, even though the total fermentation time was reduced considerably.

4.7.4 Conclusion

Scale-up of the batch recombinant CLP was successfully achieved using the constant impeller tip speed criterion. The scale-up conditions based on constant impeller tip speed were able to sustain the fermentation with a constant DO (20%) throughout the fermentation. This yielded 79.8 g/L total protein concentration. Hence, scale-up based on constant impeller tip speed was successful. A maximum specific growth rate of 1.90 /h was achieved, and a final biomass OD_{600} of 10.36 which corresponds to 8.18 g/L was attained at the end of the fermentation. A maximum RPCLP concentration of 2.5 g/L was obtained from the fermentation. These values were all higher than those obtained in the 2 L bioreactor. The total RPCLP production increased 1.17-fold when data from both scales were analyzed. The concentrations are higher than 0.7 to 1.0 g/L concentrations of recombinant Scl2 protein from the batch 2 L scale fermentation of *E. coli* BL21 as reported by Peng et al. (2012). The biomass yield on initial glucose as substrate (10 g/L) and the product yield on this substrate values were 0.82 g biomass/g substrate and 0.25 g product/g substrate, respectively. Furthermore, the scale-up did not compromise the integrity of the protein, as shown by the SDS-PAGE profile of cell lysates as well as trypsin-digested lysate samples. On the other hand, the scale-up conditions based on constant k_La did not sustain the dissolved oxygen concentration beyond the first 30 minutes (low agitation speed). This was contrary to expectations and hypothesis but can be attributed to the fact that the method relies heavily on the accurate measurements of k_La and is affected by several factors including the DO probe itself.

4.8 OBJECTIVE 5: CHARACTERIZATION OF THE RECOMBINANT COLLAGEN-LIKE PROTEIN

SDS-PAGE was used to confirm the presence and size of the recombinant RPCLP after cell lysis and after affinity chromatography was carried out as mentioned in Section 3.5. The structural stability of the triple helix was examined via protease digestion and SDS-PAGE was used to profile this digestion. MALDI-TOF was also carried out using an excised gel band containing the RPCLP which underwent trypsin digestion. The findings of these methods are discussed in the subsequent sections.

4.8.1 SDS-PAGE Analysis

Figures 4.34-4.36 show the SDS-PAGE analyses of cell lysates and purified protein from affinity chromatography without protease digestion. The first lane shows the molecular weight marker, and the second lane shows the control cell lysates (no induction with IPTG). The subsequent lanes show cell lysates of induction with IPTG. No expression was observed in an uninduced *E.coli* BL21(DE3) host containing the recombinant vector. Similarly, no expression was observed in the induced culture of BL21(DE3) cells that were transformed with an empty pColdII vector. The observed protein band corresponds to the protein size of ~25 kDa and matches the predicted protein size. According to Xu et al. (2010), the calculated/theoretical molecular weight of a CLP made up of *R. palustris* V and CL (collagen) like domains is around 22.1 kDa. However, the actual size of the RPCLP obtained using SDS-PAGE, according to Xu et al. (2010) was 23.5 kDa (around the 25 kDa peptide marker band) for the purified protein and the size of the CL band after trypsin digestion was around 13.335 kDa

(around the 15 kDa peptide marker band), larger than the predicted CL size of 11.275 kDa.

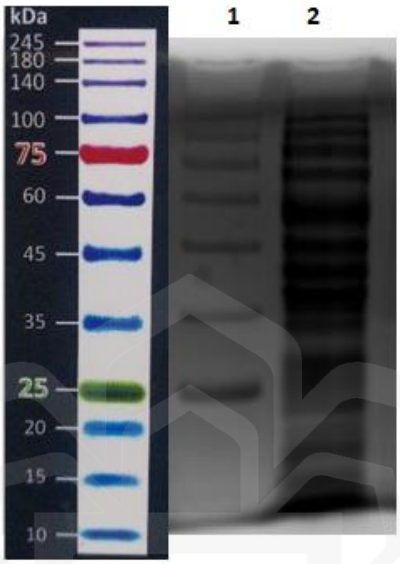


Figure 4.34 SDS-PAGE of cell lysate. Lane 1 – protein marker, Lane 2 RPCLP-containing cell lysate (lysed in sodium phosphate buffer, pH 7.4).

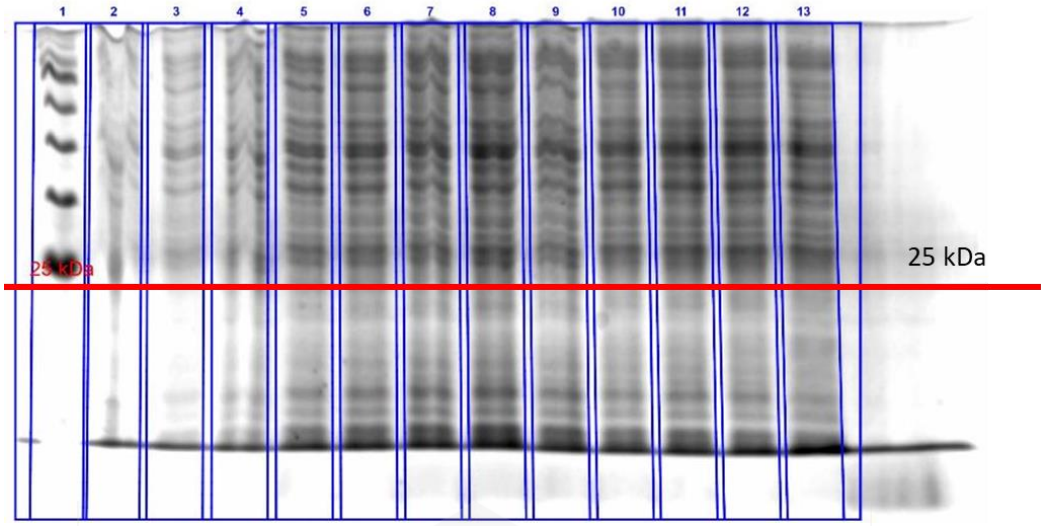


Figure 4.35 SDS-PAGE shows a time profile of the accumulation of the product in a shake-flask culture with induction. Lane 1 – protein marker, Lanes 2 to 13 – cell lysates at 0 h, 0.5 h, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 7 h, 8 h, 9 h, and 24 hours of fermentation.

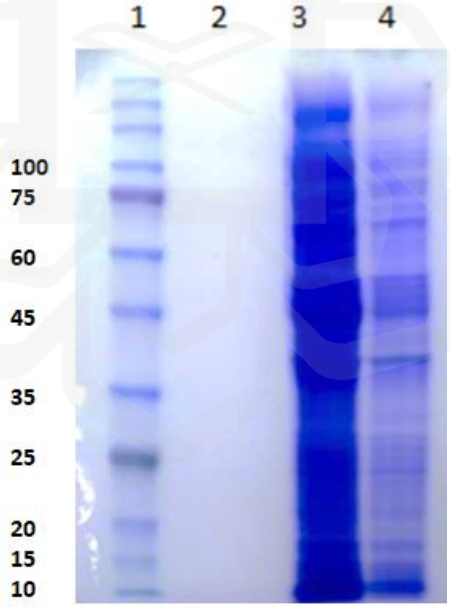


Figure 4.36 SDS-PAGE gel showing overexpression of RPCLP. Lane 1 – protein marker, Lane 2 – unused, Lane 3 – cell lysate, Lane 4 – diluted cell lysate (clearer bands).

4.8.2 Protease Digestion

Figure 4.37 depicts SDS-PAGE gels with trypsin digestion. In this study, bands between 20-25 kDa on all cell lysate lanes were observed. Faint bands at around 20 kDa on all pepsin digested sample lanes as well as 10 kDa band region were observed. For trypsin, on the other hand, no distinguishable bands around the 10 kDa mark were observed, indicating that the CLP was cleaved by trypsin, leaving the triple-helical collagen-like domain intact. However, a protein band around 20 kDa that is present in Lane 4 is absent from Lanes 5 to 7, indicating that it was digested by trypsin. Trypsin digestion had been previously done to ascertain the stability and integrity of the triple helix (Yu et al., 2014). In the literature, trypsin digestion of RPCLP was carried out for one (1) hour at room temperature, using a 1:1000 (protein/enzyme ratio). It was reported that SDS-PAGE showed a band of trypsin-resistant species at around 15 kDa MW mark, whereas the purified undigested band is around 25 kDa (Xu et al., 2010). Therefore, it was expected that similar bands would be seen in the SDS-PAGE gel in our study.

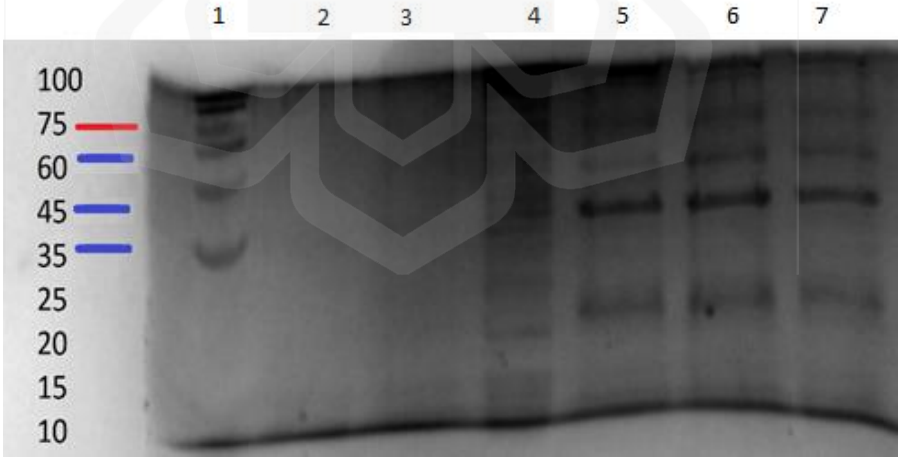


Figure 4.37 SDS-PAGE gel analysis of trypsin digested lysates samples (Lane 1 - protein marker, Lanes 2 & 3 – unused, Lane 4 – undigested lysate sample, Lanes 5 to 7 – Trypsin digested lysates).

Figures 4.38-4.39 depict pepsin digestion using cell lysates prepared using two different lysis buffers, namely sodium phosphate buffer (20 mM disodium hydrogen phosphate buffer containing 0.5 M sodium chloride, pH 7.4) and sodium acetate buffer 50 mM acetate buffer (pH 2.2). As can be seen from the gels, using the acetate buffer removed a significant portion of the host proteins, but it distorted the SDS-PAGE gel. Faint bands are seen between the 20 - 25 kDa marks for all cell lysates that contain RPCLP. The results of the bromelain and TrypZean digestion did not show any differences from the cell lysates and were excluded from this discussion. Cell lysates that had most host proteins removed via sodium acetate buffer underwent isoelectric precipitation (at pH 9.3). A very faint band between 20 and 25 kDa is observed.

Figure 4.38 SDS-PAGE following pepsin digestion is shown in Lanes 5 and 6. Lane 1 shows the protein marker, Lanes 2 (2X dilution), 8 (5X dilution), and 10 (undiluted) show the cell lysate in sodium phosphate buffer. Lanes 3 and 7 show the cell lysate in sodium acetate buffer. Lane 4 is uninduced cell lysate in sodium acetate buffer that was digested with pepsin. Lane 9 is the result of isoelectric precipitation (at pH 9.3).

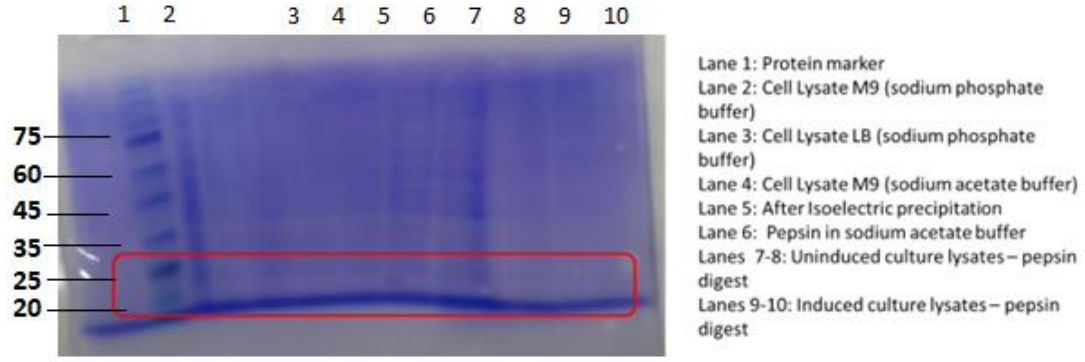


Figure 4.39 SDS-PAGE of pepsin digestion and isoelectric precipitation. Lane 1 is the protein marker. Lanes 2 and 3 show cell lysates in sodium phosphate buffer in M9 and LB, respectively. Lane 4 shows cell lysates in sodium acetate buffer. Lane 5 is the cell lysate after isoelectric precipitation. Lane 6 shows pepsin in sodium acetate buffer as a control. Lanes 7 and 8 show pepsin digests of uninduced cell lysates as a control. Lanes 9 and 10 show pepsin digests of induced cell lysates.

4.8.3 Western Blotting

Western blotting was done to further confirm the identity of the protein by probing with a mouse anti-hexahistidine tag primary antibody and a goat anti-mouse-HRP conjugated secondary antibody. A single prominent band that corresponds to 24-25 kDa molecular weight was expected. Although SDS-PAGE analysis showed a band, no bands were observed when western blotting was carried out. Numerous reasons can explain this, and to test them, several troubleshooting avenues were explored, such as adjusting the antibody-to-antigen ratio, increasing the protein abundance by loading more sample, precipitation, using fresh samples that have not been degraded, and changing incubation time and temperature. It is possible that the antibody may have degraded. Unfortunately, treatments with protease inhibitors, using different buffers and membrane types, and including a positive control were not carried out due to financial and time constraints.

4.8.4 Affinity Chromatography

The N-terminal His_6 tag present in the construct allowed for the purification of the recombinant CLP via metal affinity chromatography using a 1-mL HisTrap HP column prepacked with pre-charged Ni Sepharose High Performance resin (Cytiva) and Akta PrimePlus Fast Protein Liquid Chromatography (FPLC) system. Figures 4.40 and 4.41 show the chromatograms obtained. Initially, the protein of interest (RPCLP) was eluted in Fractions 16 and 17 (Figure 4.40). Later experiments after the maintenance of the equipment eluted the protein in Fractions 11, 12, and 13 (Figure 4.41). However, the latter's peak is notably much lower than the initial experimental runs. After obtaining the recombinant protein via chromatography, the cell lysate and purified samples were resolved on a 4-12% SDS–PAGE gel to analyze the protein purity. The protein of interest appeared as a faint band with a molecular weight between 20 and 25 kDa, confirming the presence of the His-tagged RPCLP.

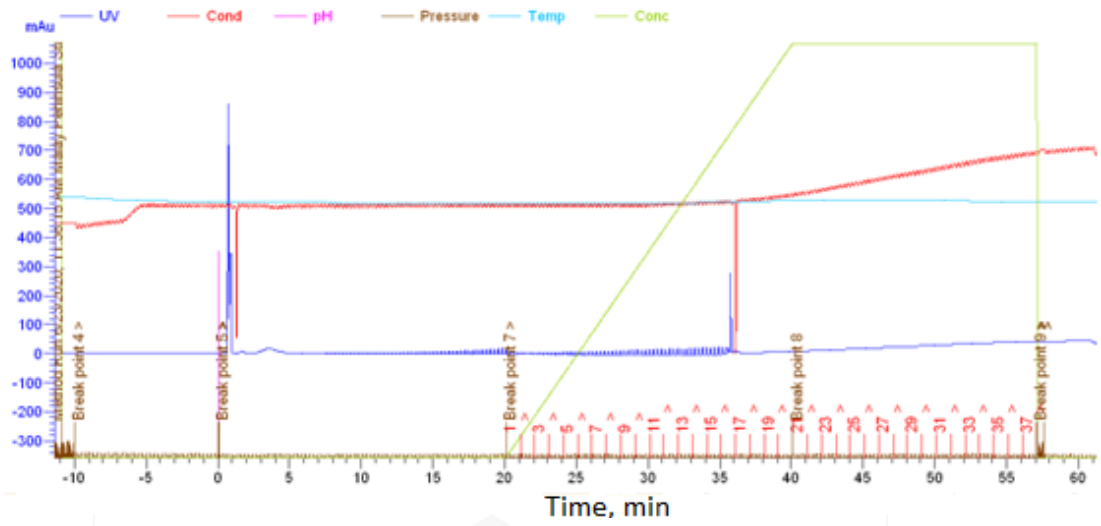


Figure 4.40 Chromatogram obtained from affinity chromatography using a Ni-charged His-Trap column showing purified RPCLP protein peak around fractions 16 and 17.

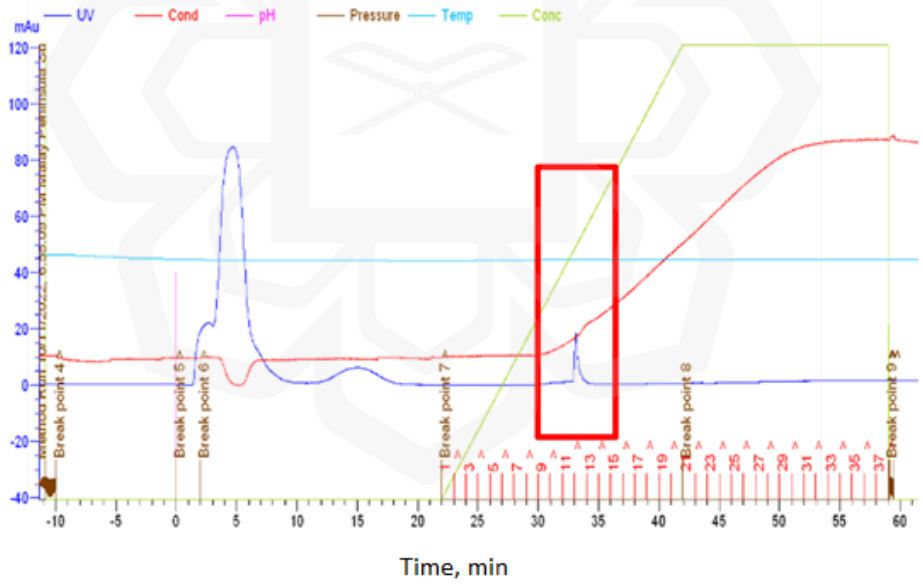


Figure 4.41 Chromatogram obtained from affinity chromatography using Ni-charged His-Trap column and 1 mL/minute flow rate showing purified RPCLP protein peak (A_{280}) at fractions 12 and 13.

4.8.5 MALDI-TOF

Protein identification by Matrix-assisted Laser Desorption/ Ionization-Time of Flight mass spectrometry was conducted following SDS-PAGE in-gel tryptic digestion. The RPCLP theoretical molecular weight was calculated to be around 23.5 kDa, based on the sequence, including an 11.3 kDa triple-helical section that is resistant to trypsin digestion. Table 4.28 shows the results of the MALDI-TOF mass spectrometry.

Table 4.28 MALDI-TOF mass spectrometry results show a 16.7 kDa fragment that corresponds to *R. palustris* DUF1320 domain-containing protein.

Database	**Proteins detected**	**Accession number**	**Molecular Weight**	**Protein Score** **C. I. %**
SwissProt	*E. coli* : Superoxide dismutase	sp|P0AGD3|SODF_ECOLI	21.3096 kDa	96.961
UniProt	*Rhodopseudomonas palustris* : DUF1320 domain-containing protein	tr|A0A323UIU7|A0A323UIU7_RHOPL	16.7027 kDa	0

While MALDI-TOF did not detect RPCLP, it detected a 16.7 kDa fragment that corresponds to *R. palustris* DUF1320 domain-containing protein. It is possible that this is, in fact, a product of trypsin digestion of RPCLP. If so, the band obtained from the SDS-PAGE and subjected to in-gel tryptic digestion and MALDI-TOF corresponding

to approximately 23.5 kDa molecular weight would be our protein of interest (RPCLP). However, NCBI's Blast tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) revealed no significant similarities between the DUF1320 domain-containing protein and the RPCLP protein. There may be restriction sequence sites within the protein that have fragmented the protein's triple helix upon treatment with proteases, preventing its detection via MALDI-TOF. Unfortunately, further experiments to fine-tune the MALDI-TOF could not be carried out due to financial constraints.

4.9 SUMMARY OF FINDINGS

1. The significant components of the M9-casamino acids medium were optimized via OFAT optimization.
2. The factors that significantly govern the growth of recombinant *E. coli* expressing RPCLPs in the 2 L bioreactor were identified.
3. The optimum conditions for the growth of recombinant *E. coli* and the concentration of recombinant CLP in the 2 L bioreactor were discovered. Under the optimized conditions, the model gives a modest RPCLP concentration of around 2 g/L.
4. The fermentation kinetics of substrate, cell growth, and product of the recombinant *E. coli* expressing RPCLP were modeled. While the Monod model fitted the biomass growth data very well, the Logistic-Luedeking-Piret model provided a better fit for substrate utilization and CLP production kinetics.
5. The optimized 2 L-scale fermentation was transferred into a 7.5 L-scale process that maintains productivity.

6. The presence of a protein band that corresponds to the estimated size of the recombinant RPCLP from the literature was verified via SDS-PAGE. However, western blotting did not obtain this band.

Although the concentrations were modest compared to other recombinant bacterial CLPs in the literature, where concentrations reached 1 g/L culture medium, the findings of this research are still useful for narrowing the gaps in our study of soil bacterial CLPs and represent the first study of RPCLP production optimization, kinetics study, as well as cultivation at the bench-scale bioreactor levels of 2 L and 7.5 L. The findings have demonstrated that the process is scalable.

CHAPTER FIVE
CONCLUSION

5.1 CONCLUSION

Obtaining high concentrations of Halal recombinant CLPs at a large scale is an appealing prospect. The objectives of this research sought to develop an optimized novel CLP production in 2 L and 7.5 L bioreactors via two scale-up criteria, namely maintaining the oxygen transfer rate and constant impeller tip speed. The novelty of this research lies in the optimization of the medium components and fermentation process, investigation of the kinetics of recombinant *E. coli* growth and RPCLP formation, as well as the scale-up of the process from 1 L to 5 L working volumes. The outcomes of this project were as follows:

1. The significant components of the M9-casamino acids medium were optimized via a One Factor at a Time optimization. It was concluded that the following solution gives optimum growth and total protein concentration:

    - 0.4 M $Na_2HPO_4.2H_2O$ (71.196 g per liter of M9 salt solution),
    - 1% glucose (i.e., 10 g/L medium), and
    - 7.5% (i.e. 0.075 g/L medium) casamino acids.

1. The factors that significantly govern the growth of recombinant *E. coli* expressing RPCLPs in a 2 L bioreactor were identified via screening experiments using a Taguchi L9 Orthogonal Array design.
2. The optimum conditions for the concentration of recombinant CLP in a 2 L bioreactor were determined via optimization experiments using a Face Centred

Central Composite Design. Under the optimized conditions, the model gives an RPCLP concentration of 2.13 g/L, which is similar to that estimated by the optimization model.

3. The fermentation kinetics of substrate, cell growth, and product of the recombinant *E. coli* expressing RPCLP were modeled. Furthermore, the kinetic parameters were estimated. While the Monod model fitted the biomass growth data very well, the Logistic-Luedeking-Piret model provided a better fit for substrate utilization and the CLP production kinetics. The characteristics of the product formation concerning growth were also elucidated from the estimated constants.
4. The optimized 2 L-scale fermentation was successfully transferred into a 7.5 L-scale process that maintains productivity via a constant impeller tip speed scale-up criterion.
5. The presence of a protein band that corresponds to the estimated size of the recombinant RPCLP from the literature was verified via SDS-PAGE. However, western blotting did not reveal this band.

Although the RPCLP concentrations were modest (2.13- 2.49 g/L), the findings were comparable to other recombinant bacterial CLPs that have been reported, where concentrations reached 1 g/L culture medium, the findings of this research serve to deepen our understanding and narrow the gaps in our study of soil bacterial CLPs, and represent the first study of RPCLP production optimization, describing the kinetics, as well as cultivation at the bench scale bioreactor levels of 2 L and 7.5 L. The outcomes of this study, therefore, contribute to our understanding and the development of novel collagens.

In keeping with the Sustainable Development Goals established by the United Nations, the production of collagen contributes to the advancement of knowledge to

improve health and well-being, primarily in wound healing, anti-aging, and tissue engineering, as in SDG 3. These products can be designed to be eco-friendly. Furthermore, its sustainable production promotes the efficient utilization of resources and waste reduction (SDG 12). Finally, collagen production can create decent employment opportunities in biotechnology and entrepreneurship and will lead to economic growth (SDG 8).

5.2 RECOMMENDATIONS

The following recommendations are advised for future work on the scaled-up bioreactor production of RPCLP in recombinant *E. coli*:

1. The development of empirical correlations for the $k_L a$ determination should be explored and validated with experimental data.
2. Further scaling-up studies need to be conducted by cultivation at higher scale-up factors, such as 10 and 100 times, to enhance the production before commercialization.
3. Evaluation of the yield and economic aspects of fed-batch and continuous processes in the bioreactor should be explored.
4. The use of software that can automatically and directly select a mathematical model that fits the experimental data and model validation may be able to better approximate kinetic parameters.
5. Fine-tuning of the separation methodology needs to be implemented for the scalability of downstream processes where FPLC would be impractical.

ACKNOWLEDGEMENT OF FINANCIAL SUPPORT

This study was financed by the Ministry of Higher Education (MOHE), Malaysia, through the Transdisciplinary Research Grant Scheme (TRGS/1/2018/UIAM/01/1/3). I would like to thank the Ministry of Higher Education (MOHE) for the financial support in carrying out this research. I would also like to thank the Research Management Centre at IIUM for their assistance in facilitating all the grant-related matters.

REFERENCES

Abdel-Nour, M., Duncan, C., Prashar, A., Rao, C., Ginevra, C., Jarraud, S., Low, D. E., Ensminger, A. W., Terebiznik, M. R., & Guyard, C. (2014). The Legionella pneumophila collagen-like protein mediates sedimentation, autoaggregation, and pathogen-phagocyte interactions. *Appl Environ Microbiol*, *80*(4), 1441-1454.

Abdullah, M. S. P., Noordin, M. I., Ismail, S. I. M., Mustapha, N. M., Jasamai, M., Danik, M. F., ... & Shamsuddin, A. F. (2018). Recent advances in the use of animal-sourced gelatine as natural polymers for food, cosmetics and pharmaceutical applications. *Sains Malaysiana*, *47*(2), 323-336.

Ahmad-Raus, R., Mel, M., Mohd-Abdullah, S. N., & Yusoff, K. (2010). Cell rupture of recombinant Escherichia coli using high-pressure homogenizer. . Journal of Applied Sciences, 10(21), 2717-2720.

Ahearne, M., Fernández-Pérez, J., Masterton, S., Madden, P. W., & Bhattacharjee, P. (2020). Designing scaffolds for corneal regeneration. *Advanced Functional Materials*, *30*(44), 1908996.

Ali, M. K., Serge, H., Nawel, O., Radia, C., & Noreddine. (2017). Kinetic models and parameters estimation study of biomass and ethanol production from inulin by Pichia caribbica (KC977491). *African Journal of Biotechnology*, *16*(3), 124-131.

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., & Jirstrand, M. (2014). Kinetic models in industrial biotechnology - Improving cell factory performance. *Metab Eng*, *24*, 38-60.

Alonso Villela, S. M., Kraiem, H., Bouhaouala-Zahar, B., Bideaux, C., Aceves Lara, C. A., & Fillaudeau, L. (2020). A protocol for recombinant protein quantification by densitometry. Microbiologyopen, 9(6), 1175-1182.

An, B., Abbonante, V., Xu, H., Gavriilidou, D., Yoshizumi, A., Bihan, D., Farndale, R. W., Kaplan, D. L., Balduini, A., Leitinger, B., & Brodsky, B. (2016). Recombinant Collagen Engineered to Bind to Discoidin Domain Receptor Functions as a Receptor Inhibitor. *J Biol Chem*, *291*(9), 4343-4355.

An, B., Kaplan, D. L., & Brodsky, B. (2014). Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Front Chem*, *2*, 40.

Ariff, A. B., Nelofer, R., Abdul Rahman, R. N. Z. R., & Basri, M. (2015). Kinetics and modelling of batch fermentation for the production of organic solvent tolerant

and thermostable lipase by recombinant E. coli. Turkish Journal of Biochemistry, 40(4), 298-309.

Arshad, Z. I., Amid, A., Yusof, F., Jaswir, I., Ahmad, K., & Loke, S. P. (2014). Bromelain: an overview of industrial application and purification strategies. *Appl Microbiol Biotechnol*, *98*(17), 7283-7297.

Asenjo, J. A., & Andrews, B. A. (2012). Aqueous two-phase systems for protein separation: phase separation and applications. *J Chromatogr A*, *1238*, 1-10.

Asenjo, J. A., & Merchuk, J. C. (1994). *Bioreactor System Design*. CRC Press.

Avila Rodriguez, M. I., Rodriguez Barroso, L. G., & Sanchez, M. L. (2018). Collagen: A review on its sources and potential cosmetic applications. *J Cosmet Dermatol*, *17*(1), 20-26.

Awang, N. A., Amid, A., & Arshad, Z. I. (2020). Method for purification of collagen: A systematic review. *AsPac J. Mol. Biol. Biotechnol.*, *28*(3), 99-112.

Awang, N. A., Amid, A., & Arshad, Z. I. (2023). Rhodopseudomonas palustris collagen-like recombinant protein purification using an aqueous two-phase system. *IIUM Engineering Journal*, *24*(1).

Azmi, A. S., Che Aziz, N. A., Mohamad Puad, N. I., Halim, A. A., Yusof, F., & Yusup, S. (2018). Chlorella vulgaris logistic growth kinetics model in high concentrations of aqueous ammonia. *IIUM Engineering Journal*, *19*(2), 1-9.

Azmi, A. S., Sulaiman, S., & Amin, N. F. M. A. (2015). Scaling-Up Recombinant Enzyme Fermentation. In A. Amid (Ed.), *Recombinant Enzymes-From Basic Science to Commercialization* (pp. 99–113). Springer.

Babaeipour, V., Mofid, M. R., Khanchezar, S., Faraji, F., & Abolghasemi, S. (2017). Bench-scale Overproduction and Purification of recombinant GCSF in Escherichia coli fed-batch process. *Journal of Applied Pharmaceutical Science*, *7*(8), 149-155.

Bachert, B. A., Choi, S. J., Snyder, A. K., Rio, R. V., Durney, B. C., Holland, L. A., Amemiya, K., Welkos, S. L., Bozue, J. A., Cote, C. K., Berisio, R., & Lukomski, S. (2015). A Unique Set of the Burkholderia Collagen-Like Proteins Provides Insight into Pathogenesis, Genome Evolution and Niche Adaptation, and Infection Detection. *PLoS One*, *10*(9), e0137578.

Baez, A. (2018). Effect of amino acids on transcription and translation of key genes in E. coli K and B grown at a steady state in minimal medium. New Biotechnology.

Bahreini, E., Aghaiypour, K., Abbasalipourkabir, R., Goodarzi, M. T., Saidijam, M., & Safavieh, S. S. (2014). An optimized protocol for overproduction of recombinant protein expression in Escherichia coli. *Prep Biochem Biotechnol*, *44*(5), 510-528.

Bartolo-Aguilar, Y., Chavez-Cabrera, C., Flores-Cotera, L. B., Badillo-Corona, J. A., Oliver-Salvador, C., & Marsch, R. (2022). The potential of cold-shock promoters for the expression of recombinant proteins in microbes and mammalian cells. *J Genet Eng Biotechnol*, *20*(1), 173.

Basar, B., Mohd-Shamzi, M., Rosfarizan, M., Puspaningsih, N. N. T., & Ariff, A. B. (2010). Enhanced production of thermophilic xylanase by recombinant Escherichia coli DH5α through optimization of medium and dissolved oxygen level.

Beigi, L., Karbalaei-Heidari, H. R., & Kharrati-Kopaei, M. (2012). Optimization of an extracellular zinc-metalloprotease (SVP2) expression in Escherichia coli BL21 (DE3) using response surface methodology. *Protein Expr Purif*, *84*(1), 161-166.

Bella, J. (2014). A first census of collagen interruptions: Collagen's own stutters and stammers. *J Struct Biol*, *186*(3), 438-450.

Bhadra, B., Sakpal, A., Patil, S., Patil, A., Date, A., Prasad, V., & Dasgupta, S. (2021). A Guide to Collagen Sources, Applications and Current Advancements. *Systematic Bioscience and Engineering*, *1*(2):67–8.

Bhattacharya, S. K., & Dubey, A. K. (1997). Effects of dissolved oxygen and oxygen mass transfer on overexpression of target gene in recombinant E. coli. *Enzyme and Microbial Technology*, *20*(5), 355-360.

Boydston, J. A., Chen, P., Steichen, C. T., & Turnbough, C. L., Jr. (2005). Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of Bacillus anthracis. *J Bacteriol*, *187*(15), 5310-5317.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72, 248-254.

Brodsky, B., & Kaplan, D. L. (2013). Shining light on collagen: expressing collagen in plants. *Tissue Eng Part A*, *19*(13-14), 1499-1501.

Brodsky, B., & Ramshaw, J. A. (2017). Bioengineered Collagens. *Subcell Biochem*, *82*, 601-629.

Brodsky, B., Thiagarajan, G., Madhan, B., & Kar, K. (2008). Triple-helical peptides: an approach to collagen conformation, stability, and self-association. *Biopolymers*, *89*(5), 345-353.

Browne, S., Zeugolis, D. I., & Pandit, A. (2013). Collagen: Finding a solution for the source. *Tissue Eng Part A*, *19*(13-14), 1491-1494.

Browning, M. B., Dempsey, D., Guiza, V., Becerra, S., Rivera, J., Russell, B., Hook, M., Clubb, F., Miller, M., Fossum, T., Dong, J. F., Bergeron, A. L., Hahn, M., & Cosgriff-Hernandez, E. (2012). Multilayer vascular grafts based on collagen-mimetic proteins. *Acta Biomater*, *8*(3), 1010-1021.

Bryntesson, M., Hall, M., & Lacki, K. (2011). *Chromatography method* (U.S. Patent No. 7,901,581). U.S. Patent and Trademark Office.

Bulleid, N. J., John, D. C., & Kadler, K. E. (2000). Recombinant expression systems for the production of collagen. *Biochem Soc Trans*, *28*(4), 350-353.

Cai, M., Huang, Y., Yang, R., Craigie, R., & Clore, G. M. (2016). A simple and robust protocol for high-yield expression of perdeuterated proteins in Escherichia coli grown in shaker flasks. *J Biomol NMR*, *66*(2), 85-91.

Cardoso, F. M. C., Petrovajova, D., & Hornakova, T. (2017). Viral vaccine stabilizers: status and trends. *Acta Virol*, *61*(3), 231-239.

Carranza-Saavedra, D., Sánchez Henao, C. P., & Zapata Montoya, J. E. (2021). Kinetic analysis and modeling of L-valine production in fermentation batch from E. coli using glucose, lactose and whey as carbon sources. *Biotechnology Reports*, *31*, e00642.

Cascaval, D., Galaction, A.-I., & Camarut, S. (2011). Scale-up of Aerobic Stirred Bioreactors Using the Mixing Time Criteria 1. Simulated Broths. *Chem. Biochem. Eng. Q.* , *25*(1), 43–54

Caswell, C. C., Barczyk, M., Keene, D. R., Lukomska, E., Gullberg, D. E., & Lukomski, S. (2008). Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins alpha2beta1 and alpha11beta1. *J Biol Chem*, *283*(52), 36168-36175.

Chan, S. W. P. (2012). *Fabrication of modular human collagen and collagen variants in yeast*. University of California, Irvine.

Chan, S. W. P., Hung, S.-P., Raman, S. K., Hatfield, G. W., Lathrop, R. H., Da Silva, N. A., & Wang, S.-W. (2010). Recombinant Human Collagen and Biomimetic Variants Using a De Novo Gene Optimized for Modular Assembly. *Biomacromolecules*, 11, 1460–1469.

Chang, S., Guo, Y., Wu, B., & He, B. (2017). Extracellular expression of alkali tolerant xylanase from Bacillus subtilis Lucky9 in E. coli and application for xylooligosaccharides production from agro-industrial waste. *Int J Biol Macromol*, *96*, 249-256.

Chattopadhyay, S., & Raines, R. T. (2014). Review collagen-based biomaterials for wound healing. *Biopolymers*, *101*(8), 821-833.

Chen, X., Jin, W., Chen, D., Dong, M., Xin, X., Li, C., & Xu, Z. (2021). Collagens made from giant salamander (Andrias davidianus) skin and their odorants. *Food Chem*, *361*, 130061.

Chu, K. H., Feng, X., Kim, E. Y., & Hung, Y.-T. (2011). Biosorption Parameter Estimation with Genetic Algorithm. *Water*, 3(1), 177-195.

Collins, T., Azevedo-Silva, J., da Costa, A., Branca, F., Machado, R., & Casal, M. (2013). Batch production of a silk-elastin-like protein in E. coli BL21(DE3): key parameters for optimisation. *Microb Cell Fact*, *12*, 21.

Cosgriff-Hernandez, E., Hahn, M. S., Russell, B., Wilems, T., Munoz-Pinto, D., Browning, M. B., Rivera, J., & Hook, M. (2010). Bioactive hydrogels based on Designer Collagens. *Acta Biomater*, *6*(10), 3969-3977.

Cox, M. M., & Nelson, D. L. (2008). *Principles of biochemistry*. W. H. Freeman and Company.

Cytiva Life Sciences. 7 Factors That Affect Oxygen Transfer to Cells in Bioreactors. 2023. Available from: https://www.cytivalifesciences.com/en/us/solutions/bioprocessing/knowledge-center/7-factors-that-affect-oxygen-transfer-to-cells-in-bioreactors [Retrieved February 2024].

Damiani, A. L., Kim, M. H., & Wang, J. (2014). An improved dynamic method to measure kL a in bioreactors. *Biotechnol Bioeng*, *111*(10), 2120-2125.

Davison-Kotler, E., Marshall, W. S., & García-Gareta, E. (2019). Sources of collagen for biomaterials in skin wound healing. *Bioengineering*, *6*(3), 56.

De Mey, M., De Maeseneire, S., Soetaert, W., & Vandamme, E. (2007). Minimizing acetate formation in E. coli fermentations. *J Ind Microbiol Biotechnol*, *34*(11), 689-700.

Deacon, S. E., Roach, P. C., Postis, V. L., Wright, G. S., Xia, X., Phillips, S. E., Knox, J. P., Henderson, P. J., McPherson, M. J., & Baldwin, S. A. (2008). Reliable scale-up of membrane protein over-expression by bacterial auto-induction: from microwell plates to pilot scale fermentations. *Molecular Membrane Biology*, *25*(8), 588-598.

Delvigne, F., & Noorman, H. (2017). Scale-up/scale-down of microbial bioprocesses: A modern light on an old issue. *Microbial Biotechnology, 10*(4), 685–687.

Deniz, İ., Imamoglu, E., & Vardar Sukan, F. (2015). Evaluation of scale-up parameters of bioethanol production from Escherichia coli KO11. *Turkish Journal of Biochemistry*, *40*(1), 74-80.

Doran, P. M. (2012). *Bioprocess Engineering Principles* (Second ed.). Academic Press.

Dorward, A., O'Kennedy, R. D., Folarin, O., Ward, J. M., & Keshavarz-Moore, E. (2019). The role of amino acids in the amplification and quality of DNA vectors for industrial applications. *Biotechnol Prog*, *35*(6), e2883.

Du, C., Li, Y., Xia, X., Du, E., Lin, Y., Lian, J., Ren, C., Li, S., Wei, W., & Qin, Y. (2021). Identification of a novel collagen-like peptide by high-throughput

screening for effective wound-healing therapy. *International Journal of Biological Macromolecules*, *173*, 541-553.

Du, Y. H., Wang, M. Y., Yang, L. H., Tong, L. L., Guo, D. S., & Ji, X. J. (2022). Optimization and Scale-Up of Fermentation Processes Driven by Models. *Bioengineering (Basel, Switzerland)*, *9*(9), 473.

Duong-Trung, N., Born, S., Kim, J. W., Schermeyer, M.-T., Paulick, K., Borisyak, M., Cruz-Bournazou, M. N., Werner, T., Scholz, R., Schmidt-Thieme, L., Neubauer, P., & Martínez, E. (2022). When bioprocess engineering meets machine learning: A survey from the perspective of automated bioprocess development. *Biochemical Engineering Journal, 190,* 108764.

Eiteman, M. A., & Altman, E. (2006). Overcoming acetate in Escherichia coli recombinant protein fermentations. *Trends Biotechnol*, *24*(11), 530-536.

Elbing, K. L., & Brent, R. (2019). Growth of E. coli in Liquid Medium. *Curr Protoc Mol Biol*, *125*(1), e81.

Fan, D. D., Luo, Y., Mi, Y., Ma, X. X., & Shang, L. (2005). Characteristics of fed-batch cultures of recombinant Escherichia coli containing human-like collagen cDNA at different specific growth rates. *Biotechnol Lett*, *27*(12), 865-870.

Fang, J., Ma, Z., Liu, D., Wang, Z., Cheng, S., Zheng, S., Wu, H., Xia, P., Chen, X., Yang, R., Hao, L., & Zhang, Y. (2022). Co-expression of recombinant human collagen α1(III) chain with viral prolyl 4-hydroxylase in Pichia pastoris GS115. *Protein Expression and Purification*, *201*, 106184.

Farajollahi, S., Dennis, P. B., Crosby, M. G., Slocik, J. M., Pelton, A. T., Hampton, C. M., Drummy, L. F., Yang, S. J., Silberstein, M. N., Gupta, M. K., & Naik, R. R. (2019). Disulfide Crosslinked Hydrogels Made From the Hydra Stinging Cell Protein, Minicollagen-1. *Front Chem*, *7*, 950.

Ferreira, A. M., Gentile, P., Chiono, V., & Ciardelli, G. (2012). Collagen for bone tissue regeneration. *Acta Biomater*, *8*(9), 3191-3200.

Fertala, A. (2020). Three decades of research on recombinant collagens: reinventing the wheel or developing new biomedical products?. *Bioengineering*, *7*(4), 155.

Ferdous, S., Dopp, J. L., & Reuel, N. F. (2021). Optimization of E. coli tip-sonication for high-yield cell-free extract using finite element modeling. *AIChE Journal*, 67(10), e17389.

Fidanova, S., Roeva, O., & Ganzha, M. (2012, September). ACO for parameter settings of E. coli fed-batch cultivation model. In *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 407-414). IEEE.

Fontani, S. (2003). Studies on the maximization of recombinant Helicobacter pylori neutrophilactivating protein production in Escherichia coli: application of Taguchi robust design and response surface methodology for process optimization. *World Journal of Microbiology& Biotechnology*, *19*, 711-717.

Franzke, C. W., Bruckner, P., & Bruckner-Tuderman, L. (2005). Collagenous transmembrane proteins: recent insights into biology and pathology. *J Biol Chem*, *280*(6), 4005-4008.

Gadjos, L. (2007). Elimination of Polyethylene Glycol from Aqueous Solution Using Activated Carbon. *Polish J. of Environ. Stud.*, 16(3), 385-388.

Gameil, A. H. M., Yusof, F., Azmi, A. S., & Puad, N. I. M. (2021a). Process scale-up criteria in production of recombinant proteins in E. coli: A systematic review. *Chemical and Natural Resources Engineering Journal (Biological and Natural Resources Engineering Journal)*, *5*(2), 37-61.

Gameil, A. H. M., Yusof, F., Azmi, A. S., & Mohamad Puad, N. I. (2021b). Progress in the detection and quantification of collagens: a review. In IOP Conf. Ser.: Mater. Sci. Eng., 1192(1), Kuala Lumpur, Malaysia.

Gameil, A. H. M., Yusof, F., Azmi, A. S., & Mohamad Puad, N. I. (2023). Modeling of E. coli growth, glucose consumption, and recombinant collagen-like protein formation kinetics. IIUM Engineering Congress Proceedings, 1 (1), 74–78.

Garcia-Ochoa, F., & Gomez, E. (2009). Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview. *Biotechnol Adv*, *27*(2), 153-176.

Garcia-Ochoa, F., Gomez, E., & Santos, V. E. (2020). Fluid dynamic conditions and oxygen availability effects on microbial cultures in STBR: An overview. *Biochemical Engineering Journal*, *164*, 107803.

Garcia-Ochoa, F., Gomez, E., Santos, V. E., & Merchuk, J. C. (2010). Oxygen uptake rate in microbial processes: An overview. *Biochemical Engineering Journal*, *49*(3), 289-307.

Garnier, A., & Gaillet, B. (2015). Analytical solution of Luedeking-Piret equation for a batch fermentation obeying Monod growth kinetics. *Biotechnol Bioeng*, 112(12), 2468-2474.

Garrett, R. H., & Grisham, C. M. (2010). *Biochemistry* (4th ed. ed.). Brooks/Cole, Cengage Learning.

Gellermann, P., Schneider-Barthold, C., Bolten, S. N., Overfelt, E., Scheper, T., & Pepelanova, I. (2019). Production of a recombinant non-hydroxylated gelatin mimetic in Pichia pastoris for biomedical applications. *Journal of functional biomaterials*, *10*(3), 39.

Gelse, K., Poschl, E., & Aigner, T. (2003). Collagens--structure, function, and biosynthesis. *Adv Drug Deliv Rev*, *55*(12), 1531-1546.

Ghosh, N., McKillop, T. J., Jowitt, T. A., Howard, M., Davies, H., Holmes, D. F., Roberts, I. S., & Bella, J. (2012). Collagen-like proteins in pathogenic E. coli strains. *PLoS One*, *7*(6), e37872.

Gill, N. K., Appleton, M., Baganz, F., & Lye, G. J. (2008). Quantification of power consumption and oxygen transfer characteristics of a stirred miniature bioreactor for predictive fermentation scale-up. *Biotechnol Bioeng*, 100(6), 1144-1155.

Go, K., Horikawa, Y., Garcia, R., & Villarreal, F. J. (2008). Fluorescent method for detection of cleaved collagens using O-phthaldialdehyde (OPA). *J Biochem Biophys Methods*, *70*(6), 878-882.

Goldring, J. P. D. (2018). The Roles of Acetic Acid and Methanol During Fixing and Staining Proteins in an SDS–Polyacrylamide Electrophoresis Gel. In B. T. Kurien & R. H. Scofield (Eds.), *Protein Gel Detection and Imaging: Methods and Protocols* (Vol. 1853).

Golser, A. V., & Scheibel, T. (2018). Routes towards novel collagen-like biomaterials. *Fibers*, *6*(2), 21.

Gomes, V., & Salgueiro, S. P. (2022). From small to large-scale: a review of recombinant spider silk and collagen bioproduction. *Discover Materials*, *2*(1), 3.

Gorres, K. L., & Raines, R. T. (2010). Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol*, *45*(2), 106-124.

Graumann, K., & Premstaller, A. (2006). Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnol J*, *1*(2), 164-186.

Grand View Research, Inc., USA. Collagen Market Estimates and Forecasts to 2030. 2023. Available from: https://www. grandviewresearch.com/industry-analysis/collagen-market [Retrieved February 2024].

Grilo, A. L., Raquel Aires-Barros, M., & Azevedo, A. M. (2014). Partitioning in Aqueous Two-Phase Systems: Fundamentals, Applications and Trends. *Separation & Purification Reviews*, *45*(1), 68-80.

Grover, C. N., Cameron, R. E., & Best, S. M. (2012). Investigating the morphological, mechanical and degradation properties of scaffolds comprising collagen, gelatin and elastin for use in soft tissue engineering. *J Mech Behav Biomed Mater*, *10*, 62-744.

Grund, M. E., Choi, S. J., McNitt, D. H., Barbier, M., Hu, G., LaSala, P. R., Cote, C. K., Berisio, R., & Lukomski, S. (2020). Burkholderia collagen-like protein 8, Bucl8, is a unique outer membrane component of a putative tetrapartite efflux pump in Burkholderia pseudomallei and Burkholderia mallei. *PLoS One*, *15*(11), e0242593.

Hajihassan, Z., Tilko, P. G., & Sadat, S. M. (2018). Improved Production of Recombinant Human beta-NGF in Escherichia coli - a Bioreactor Scale Study. *Pol J Microbiol*, *67*(3), 355-363.

Hajinia, E., Fatemi, S. S. A., Karkhane, A. A., Safekordi, A. A., & Yakhchali, B. (2012). Optimization of secretory expression of recombinant hGM-CSF in high cell density cultivation of recombinant Escherichia coli using Taguchi statistical method. *Iranian Journal of Biotechnology*, *10*(4), 275-280.

Hashim, P., Mohd Ridzwan, M. S., Bakar, J. , & Mat Hashim, D. (2015). Collagen in food and beverage industries. *International Food Research Journal*, *22*(1), 1-8.

Hill, C. G., & Thatcher, W. (2014). *An Introduction to Chemical Engineering Kinetics and Reactor Design* (Second Edition ed.). John Wiley & Sons, Inc.

Hou, Y., Guey, L. T., Wu, T., Gao, R., Cogan, J., Wang, X., Hong, E., Vivian Ning, W., Keene, D., Liu, N., Huang, Y., Kaftan, C., Tangarone, B., Quinones-Garcia, I., Uitto, J., Francone, O. L., Woodley, D. T., & Chen, M. (2015). Intravenously Administered Recombinant Human Type VII Collagen Derived from Chinese Hamster Ovary Cells Reverses the Disease Phenotype in Recessive Dystrophic Epidermolysis Bullosa Mice. *J Invest Dermatol*, *135*(12), 3060-3067.

Hu, W., Xiang, J. Y., Kong, P., Liu, L., Xie, Q., & Xiang, H. (2017). Expression and Characterization of a Single-Chain Variable Fragment against Human LOX-1 in Escherichia coli and Brevibacillus choshinensis. *J Microbiol Biotechnol*, *27*(5), 965-974.

Huang, C. J., Peng, H. L., Patel, A. K., Singhania, R. R., Dong, C. D., & Cheng, C. Y. (2021). Effects of lower temperature on expression and biochemical characteristics of HCV NS3 antigen recombinant protein. *Catalysts*, *11*(11), 1297.

Huang, J., Wong Po Foo, C., & Kaplan, D. L. (2007). Biosynthesis and Applications of Silk-like and Collagen-like Proteins. *Polymer Reviews*, *47*(1), 29-62.

Imamoglu, E., & Sukan, F. V. (2013). Scale-up and kinetic modeling for bioethanol production. *Bioresour Technol*, *144*, 311-320.

Infors-HT, The Cookbook, Available from: http://www.infors-ht.com.cn/uploadfile/2018/0515/INFORS-HT_cookbook_en.pdf [Retrieved February 2024].

Islam, R. S., Tisi, D., Levy, M. S., & Lye, G. J. (2008). Scale-up of Escherichia coli growth and recombinant protein expression conditions from microwell to laboratory and pilot scale based on matched kLa. *Biotechnol Bioeng*, 99(5), 1128-1139.

Iqbal, M., Tao, Y., Xie, S., Zhu, Y., Chen, D., Wang, X., Huang, L., Peng, D., Sattar, A., Shabbir, M. A., Hussain, H. I., Ahmed, S., & Yuan, Z. (2016). Aqueous two-phase system (ATPS): an overview and advances in its applications. *Biol Proced Online*, *18*, 18.

Jain, D., Meena, V. S., Kaushik, S., Kamble, A., Chisti, Y., & Banerjee, U. C. (2012). Production of Nitrilase by a Recombinant Escherichia coli in a Laboratory Scale Bioreactor. *Fermentation Technology*, *1*(1), 1-4.

Jamaluddin, M. J. A. (2015). *Production of intracellular recombinant bromelain by Escherichia coli BL21-AI using shake flask and stirred tank bioreactor.* [Master's thesis, International Islamic University Malaysia].

Jamaluddin, M. J. A., Amid, A., Azmi, A. S., & Othman, M. E. F. (2014). Screening of Important Autoinduction Medium Composition for High Biomass Production of E. coli Expressing Recombinant Bromelain. *Journal of Pure and Applied Microbiology*, *8*, 741-750.

Jiang, X., Wang, Y., Fan, D., Zhu, C., Liu, L., & Duan, Z. (2017). A novel human-like collagen hemostatic sponge with uniform morphology, good biodegradability and biocompatibility. *J Biomater Appl*, *31*(8), 1099-1107.

Jon. (2012). Papain, a Plant Enzyme of Biological Importance: A Review. *American Journal of Biochemistry and Biotechnology*, *8*(2), 99-104.

Junker, B. H. (2004). Scale-up methodologies for Escherichia coli and yeast fermentation processes. *J Biosci Bioeng*, *97*(6), 347-364.

Kapat, A., Jung, J.-K., & Park, Y.-H. (2001). Enhancement of glucose oxidase production in batch cultivation of recombinant Saccharomyces cerevisiae: optimization of oxygen transfer condition. *Journal of Applied Microbiology*, *90*(2), 216–222.

Karami, A., Tebyanian, H., Sayyad Soufdoost, R., Motavallian, E., Barkhordari, A., & Nourani, M. R. (2019). Extraction and Characterization of Collagen with Cost-Effective Method from Human Placenta for Biomedical Applications. *World J Plast Surg*, *8*(3), 352-358.

Karlstrom, A. (2005). *Collagen-like proteins in horse pathogenic Streptococcus equi* (No. 2005: 99). [Dissertation, Swedish University of Agricultural Sciences]. Uppsala.

Khoshnoodi, J., Cartailler, J.-P., Alvares, K., A., V., & Hudson, B. G. (2006). Molecular recognition in the assembly of collagens: Terminal noncollagenous domains are key recognition modules in the formation of triplehelical protomers. *J. Biol. Chem.*, *281*, 38117–38121.

Kitko, R. D., Wilks, J. C., Garduque, G. M., & Slonczewski, J. L. (2010). Osmolytes contribute to pH homeostasis of Escherichia coli. *PLoS One*, *5*(4), e10078.

Koopaei, N. N., Khadiv-Parsi, P., Khoshayand, M. R., Mazlomi, M. A., Kebriaeezadeh, A., Moloudian, H., Solhi, R., & Aminian, M. (2018). Optimization of rPDT fusion protein expression by Escherichia coli in pilot scale fermentation: a statistical experimental design approach. *AMB Express*, *8*(1), 135.

Kumar, R., Banoth, L., Banerjee, U. C., & Kaur, J. (2017). Enantiomeric separation of pharmaceutically important drug intermediates using a Metagenomic lipase and optimization of its large scale production. *Int J Biol Macromol*, *95*, 995-1003.

Kumar, V. A., Taylor, N. L., Jalan, A. A., Hwang, L. K., Wang, B. K., & Hartgerink, J. D. (2014). A nanostructured synthetic collagen mimic for hemostasis. *Biomacromolecules*, *15*(4), 1484-1490.

Kurien, B. T., & Scofield, R. H. (2012). Common artifacts and mistakes made in electrophoresis. *Methods Mol Biol*, *869*, 633-640.

Lamppa, J. W., Tanyos, S. A., & Griswold, K. E. (2013). Engineering Escherichia coli for soluble expression and single step purification of active human lysozyme. *J Biotechnol*, *164*(1), 1-8.

Langlais, C., & Korn, B. (2006). Recombinant Protein Expression in Bacteria. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine* (pp. 1609-1616).

Lareu, R. R., Zeugolis, D. I., Abu-Rub, M., Pandit, A., & Raghunath, M. (2010). Essential modification of the Sircol Collagen Assay for the accurate quantification of collagen content in complex protein solutions. *Acta Biomater*, *6*(8), 3146-3151.

Larimer, F. W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M. L., Pelletier, D. A., Beatty, J. T., Lang, A. S., Tabita, F. R., Gibson, J. L., Hanson, T. E., Bobst, C., Torres, J. L., Peres, C., Harrison, F. H., Gibson, J., & Harwood, C. S. (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nat Biotechnol*, *22*(1), 55-61.

Lee, E. J., Lee, B. H., Kim, B. K., & Lee, J. W. (2013). Enhanced production of carboxymethylcellulase of a marine microorganism, Bacillus subtilis subsp. subtilis A-53 in a pilot-scaled bioreactor by a recombinant Escherichia coli JM109/A-53 from rice bran. *Mol Biol Rep*, *40*(5), 3609-3621.

Lee, T. S. (2009). A methodological approach to scaling up fermentation and primary recovery processes to the manufacturing scale for vaccine production. *Vaccine*, *27*(46), 6439-6443.

Li, L., Fan, D., Ma, X., Deng, J., & He, J. (2015). High-level secretory expression and purification of unhydroxylated human collagen α1(III) chain in Pichia pastoris GS115. *Biotechnology and Applied Biochemistry*, *62*(4), 467-475.

Limoes. (2016). Kinetic study of Escherichia coli BPPTCC-EgRK2 to produce recombinant cellulase for ethanol production from oil palm empty fruit bunch. *IOP Conference Series: Earth and Environmental Science* (Vol. 141, No. 1, p. 012016). IOP Publishing.

Liu, W., Merrett, K., Griffith, M., Fagerholm, P., Dravida, S., Heyne, B., Scaiano, J. C., Watsky, M. A., Shinozaki, N., Lagali, N., Munger, R., & Li, F. (2008). Recombinant human collagen for tissue engineered corneal substitutes. *Biomaterials*, *29*(9), 1147-1158.

Lu, J., Song, Q., Ji, Z., Liu, X., Wang, T., & Kang, Q. (2015). Fermentation optimization of maltose-binding protein fused to neutrophil-activating protein from Escherichia coli TB1. *Electronic Journal of Biotechnology*, *18*(4), 281-285.

Lukomski, S., Bachert, B. A., Squeglia, F., & Berisio, R. (2017). Collagen-like proteins of pathogenic streptococci. *Mol Microbiol*, *103*(6), 919-930.

Lukomski, S., Bachert, B. A., Squeglia, F., & Berisio, R. (2018). Collagen-like proteins of pathogenic streptococci. *Mol Microbiol*, *103*(6), 919-930.

Lukomski, S., & McNitt, D. H. (2020). Expression and Purification of Collagen-Like Proteins of Group A Streptococcus. *Methods Mol Biol*, *2136*, 163-179.

Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, E. A., & Musser, J. M. (2001). Identification and characterization of a second extracellular collagen-like protein made by group A Streptococcus: control of production at the level of translation. *Infect Immun*, *69*(3), 1729-1738.

Luo, Y., & Mu, T. (2014). Preparation of a low-cost minimal medium for engineered Escherichia coli with high yield of human-like collagen II. *Pak J Pharm Sci*, *27*(3 Suppl), 663-669.

Luo, Y. E., Fan, D. D., Shang, L. A., Shi, H. J., Ma, X. X., Mi, Y., & Zhao, G. F. (2008). Analysis of metabolic flux in Escherichia coli expressing human-like collagen in fed-batch culture. *Biotechnol Lett*, *30*(4), 637-643.

Ma, L., Chai, Y., Wu, T., & Wang, M. (2017). Order of stability for proteolysis sites of a bacterial collagen-like protein. *The Journal of Biochemistry*, *162*(3), 227-235.

Ma, L., Liang, X., Yu, S., & Zhou, J. (2022). Expression, characterization, and application potentiality evaluation of recombinant human-like collagen in Pichia pastoris. *Bioresources and Bioprocessing*, *9*(1).

Ma, X. X., Fan, D. D., Zhu, C. H., Shang, Z. F., & Mi, Y. (2014). Optimization of fermentation medium for collagen production of recombinant Pichia pastoris during induction phase. *Journal of Chemical and Pharmaceutical Research*, *6*(7), 1802-1809.

Mahmoodi, S., Pourhassan-Moghaddam, M., Wood, D. W., Majdi, H., & Zarghami, N. (2019). Current affinity approaches for purification of recombinant proteins. *Cogent Biology*, *5*(1), 1665406.

Makwana, Mandar Narayan. (2023). *Using computational fluid dynamics and optical sensor technology to scale cell culture platforms* (KGI Theses and Dissertations no. 28). [Doctoral Dissertation, Keck Graduate Institute of Applied Life Sciences].

Marques, M. P. C., Cabral, J. M. S., & Fernandes, P. (2010). Bioprocess scale-up: quest for the parameters to be used as criterion to move from microreactors to lab-scale. *Journal of Chemical Technology & Biotechnology*, *85*(9), 1184-1198.

McElroy, K., Mouton, L., Du Pasquier, L., Qi, W., & Ebert, D. (2011). Characterisation of a large family of polymorphic collagen-like proteins in the endospore-forming bacterium Pasteuria ramosa. *Res Microbiol*, *162*(7), 701-714.

Mel, M., Karim, M. I. A., Jamal, P., Mohamed Salleh, M. R., & Zakaria, R. A. (2006). The influence of process parameters on lactic acid fermentation in laboratory scale fermenter. *Journal of Applied Sciences*, *6*(10), 2287-2291.

Mel, M., Karim, M. I. A., & Salleh, H. M. (2010). The Evaluation of kLa Values for Recombinant Escherichia coli Fermentation Producing β-Glucuronidase Enzyme. *Journal of Applied Sciences*, *10*(4), 325-330.

Meurer, S. K., Karsdal, M. A., & Weiskirchen, R. (2020). Advances in the clinical use of collagen as biomarker of liver fibrosis. *Expert review of molecular diagnostics*, *20*(9), 947–969.

Milano, F., Masi, A., Madaghiele, M., Sannino, A., Salvatore, L., & Gallo, N. (2023). Current trends in gelatin-based drug delivery systems. *Pharmaceutics*, *15*(5), 1499.Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., & Brodsky, B. (2007). Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J Biol Chem*, *282*(41), 29757-29765.

Mouton, L., Traunecker, E., McElroy, K., Du Pasquier, L., & Ebert, D. (2009). Identification of a polymorphic collagen-like protein in the crustacean bacteria Pasteuria ramosa. *Res Microbiol*, *160*(10), 792-799.

Mohd Arshad, Z. I. (2016). *Recombinant bromelain purification, powder formulations and its applications.* [Doctoral dissertation, International Islamic University Malaysia].

Mullai, P., Rene, E. R., & Sridevi, K. (2013). Biohydrogen production and kinetic modeling using sediment microorganisms of Pichavaram mangroves, India. *Biomed Res Int*, 2013, 265618.

Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M., & Kowalski, S. M. (2018). Response Surface Methodology: A Retrospective and Literature Survey. *Journal of Quality Technology*, *36*(1), 53-77.

Nadal-Rey, G., McClure, D. D., Kavanagh, J. M., Cassels, B.,Cornelissen, S., Fletcher, D. F., & Gernaey, K. V. (2022). Computational fluid dynamics modelling of hydrodynamics, mixing, and oxygen transfer in industrial bioreactors with Newtonian broths. *Biochemical Engineering Journal*, *177*, 108265.

Nadal-Rey, G., McClure, D. D., Kavanagh, J. M., Cornelissen, S., Fletcher, D. F., & Gernaey, K. V. (2021). Understanding gradients in industrial bioreactors. *Biotechnology Advances*, *46*, 107660.

Necula, L., Matei, L., Dragu, D., Pitica, I., Neagu, A., Bleotu, C., Diaconu, C. C., & Chivu-Economescu, M. (2022). Collagen family as promising biomarkers and therapeutic targets in cancer. *International journal of molecular sciences*, *23*(20), 12415.

Nimptsch, A., Schibur, S., Ihling, C., Sinz, A., Riemer, T., Huster, D., & Schiller, J. (2011). Quantitative analysis of denatured collagen by collagenase digestion and subsequent MALDI-TOF mass spectrometry. *Cell Tissue Res*, *343*(3), 605-617.

Noble, J. E., & Bailey, M. J. A. (2009). Guide to Protein Purification. In *Methods in Enzymology* (2nd Edition ed., Vol. 463).

Noorbatcha, I. A. (2011). Modeling the Growth Kinetics of Escherichia coli Fermentation in Bioreactors. In *Bioprocessing of Recombinant E. coli producing beta-glucuronidase enzyme*. IIUM Press, Kuala Lumpur, Malaysia, pp. 115-123.

Nuge, T., Hayyan, A., M., E. A. A., Mohd Salleh, H., Yap, J. Y., Kamarudin, A. F., Hizaddin, H. F., Hashim, Y. Z. H.-Y., Liu, X., Saleh, J., Daoud, J. I., M., A. A. S., Alhumaydhi, F. A., Zulkifli, M. Y., Mohd Nor, M. R., & Al Abdulmonem, W. (2023). Enhanced large-scale production of recombinant phytase in E. coli DH5a: Medium components optimization and thermodynamic studies. *Journal of Molecular Liquids*, *370*, 120965.

Paterson, G. K., Nieminen, L., Jefferies, J. M., & Mitchell, T. J. (2008). PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol Lett*, *285*(2), 170-176.

Patnaik, R., Roof, W. D., Young, R. F., & Liao, J. C. (1992). Stimulation of glucose catabolism in Escherichia coli by a potential futile cycle. *J Bacteriol*, *174*(23), 7527-7532.

Peng, Y. Y., Howell, L., Stoichevska, V., Werkmeister, J. A., Dumsday, G. J., & Ramshaw, J. A. M. (2012). Towards scalable production of a collagen-like protein from Streptococcus pyogenes for biomedical applications. *Microbial Cell Factories*, *11*(1), 146.

Peng, Y. Y., Nebl, T., Glattauer, V., & Ramshaw, J. A. M. (2018). Incorporation of hydroxyproline in bacterial collagen from Streptococcus pyogenes. *Acta Biomater*, *80*, 169-175.

Peng, Y. Y., Stoichevska, V., Howell, L., Madsen, S., Werkmeister, J. A., Dumsday, G. J., & Ramshaw, J. A. (2014a). Preparation and characterization of monomers to tetramers of a collagen-like domain from Streptococcus pyogenes. *Bioengineered*, *5*(6), 378-385.

Peng, Y. Y., Stoichevska, V., Madsen, S., Howell, L., Dumsday, G. J., Werkmeister, J. A., & Ramshaw, J. A. (2014b). A simple cost-effective methodology for large-scale purification of recombinant non-animal collagens. *Appl Microbiol Biotechnol*, *98*(4), 1807-1815.

Peng, Y. Y., Stoichevska, V., Schacht, K., Werkmeister, J. A., & Ramshaw, J. A. (2014c). Engineering multiple biological functional motifs into a blank collagen-like protein template from Streptococcus pyogenes. *J Biomed Mater Res A*, *102*(7), 2189-2196.

Peng, Y. Y., Stoichevska, V., Vashi, A., Howell, L., Fehr, F., Dumsday, G. J., Werkmeister, J. A., & Ramshaw, J. A. (2015). Non-animal collagens as new options for cosmetic formulation. *Int J Cosmet Sci*, *37*(6), 636-641.

Pérez, R. E., Suárez, J. G., Diaz, E. N., Silva Rodríguez, R., Caballero Menéndez, E., Balaguer, H. D., & Musacchio Lasa, A. (2018). Scaling-up fermentation of Escherichia coli for production of recombinant P64k protein from Neisseria meningitidis. *Electronic Journal of Biotechnology*, *33*, 29-35.

Persikov, A. V., Ramshaw, J. A., & Brodsky, B. (2005). Prediction of collagen stability from amino acid sequence. *J Biol Chem*, *280*(19), 19343-19349.

Persikov, A. V., Xu, Y., & Brodsky, B. (2004). Equilibrium thermal transitions of collagen model peptides. *Protein Sci*, *13*(4), 893-902.

Pinkas, D. M., Ding, S., Raines, R. T., & Barron, A. E. (2011). Tunable, post-translational hydroxylation of collagen Domains in Escherichia coli. *ACS Chem Biol*, *6*(4), 320-324.

Pizarro-Guajardo, M., Olguin-Araneda, V., Barra-Carrasco, J., Brito-Silva, C., Sarker, M. R., & Paredes-Sabja, D. (2014). Characterization of the collagen-like exosporium protein, BclA1, of Clostridium difficile spores. *Anaerobe*, *25*, 18-30.

Porfírio, E., & Fanaro, G. B. (2016). Collagen supplementation as a complementary therapy for the prevention and treatment of osteoporosis and osteoarthritis: a systematic review. *Revista Brasileira de Geriatria e Gerontologia*, *19*(1), 153-164.

Pustake, S. O., Bhagwat, P. K., & Dandge, P. B. (2019). Statistical media optimization for the production of clinical uricase from Bacillus subtilis strain SP6. *Heliyon*, *5*(5), e01756.

Qi, Q., Yao, L., Liang, Z., Yan, D., Li, Z., Huang, Y., & Sun, J. (2016). Production of human type II collagen using an efficient baculovirus-silkworm multigene expression system. *Mol Genet Genomics*, *291*(6), 2189-2198.

Qiu, Y., Zhai, C., Chen, L., Liu, X., & Yeo, J. (2021). Current Insights on the Diverse Structures and Functions in Bacterial Collagen-like Proteins. *ACS Biomater Sci Eng*, *9*(7), 3778-3795.

Raja, S., Murty, V. R., Thivaharan, V., Rajasekar, V., & Ramesh, V. (2012). Aqueous Two Phase Systems for the Recovery of Biomolecules – A Review. *Science and Technology*, *1*(1), 7-16.

Rajendran, A., & Thangavelu, V. (2008). Evaluation of Various Unstructured Kinetic Models for the Production of Protease by Bacillus sphaericus MTTC511. *Engineering in Life Sciences*, *8*(2), 179-185.

Rakhmanova, A., Khan, Z. A., Sharif, R., & Lu, X. (2018). Meeting the Requirements of Halal Gelatin: a mini review. *MOJ Food Process Technol, 6*(6), 477-482.

Ramshaw, J. A., Werkmeister, J. A., & Dumsday, G. J. (2014). Bioengineered collagens: emerging directions for biomedical materials. *Bioengineered*, *5*(4), 227-233.

Rao, G. (2009). Dynamic measurement of the volumetric oxygen transfer coefficient in fermentation systems. *Biotechnol Bioeng*, *104*(5), 841-853.

Rao, R. S., Kumar, C. G., Prakasham, R. S., & Hobbs, P. J. (2008). The Taguchi methodology as a statistical tool for biotechnological applications: a critical appraisal. *Biotechnol J*, *3*(4), 510-523.

Rasmussen, M., Jacobsson, M., & Bjorck, L. (2003). Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J Biol Chem*, *278*(34), 32313-32316.

Renu, I. C., & Thampi, B. H. (2019). Studies on collagen hydrolysis by pineapple (Ananas comosus) stem bromelain. *Int J Pharm Biol Sci*, *9*, 118-26.

Restaino, O. F., Borzacchiello, M. G., Scognamiglio, I., Fedele, L., Alfano, A., Porzio, E., Manco, G., De Rosa, M., & Schiraldi, C. (2018). High yield production and purification of two recombinant thermostable phosphotriesterase-like lactonases from Sulfolobus acidocaldarius and Sulfolobus solfataricus useful as bioremediation tools and bioscavengers. *BMC Biotechnol*, *18*(1), 18.

Restaino, O. F., Borzacchiello, M. G., Scognamiglio, I., Porzio, E., Manco, G., Fedele, L., Donatiello, C., De Rosa, M., & Schiraldi, C. (2017). Boosted large-scale production and purification of a thermostable archaeal phosphotriesterase-like lactonase for organophosphate decontamination. *J Ind Microbiol Biotechnol*, *44*(3), 363-375.

Ricard-Blum, S. (2011). The collagen family. *Cold Spring Harb Perspect Biol*, *3*(1), a004978.

Rodriguez, E. L., Poddar, S., Iftekhar, S., Suh, K., Woolfork, A. G., Ovbude, S., Pekarek, A., Walters, M., Lott, S., & Hage, D. S. (2020). Affinity chromatography: A review of trends and developments over the past 50 years. *J Chromatogr B Analyt Technol Biomed Life Sci*, *1157*, 122332.

Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thompson, A., Cameron, A. D., Alston, M., Stringer, M. F., Betts, R. P., Baranyi, J., Peck, M. W., & Hinton, J. C. (2012). Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol*, *194*(3), 686-701.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol*, *5*, 172.

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Science*, *28*(8), 1412-1422.

Rudiger, J., & Schwab, W. (2019). Improving an Escherichia coli-based biocatalyst for terpenol glycosylation by variation of the expression system. *J Ind Microbiol Biotechnol*, *46*(8), 1129-1138.

Rutschmann, C., Baumann, S., Cabalzar, J., Luther, K. B., & Hennet, T. (2014). Recombinant expression of hydroxylated human collagen in Escherichia coli. *Appl Microbiol Biotechnol*, *98*(10), 4445-4455.

Sandhu, S. V., Gupta, S., Bansal, H., Singla, K., & Yadav, N. S. (2012). Collagen in Health and Disease. *Journal of Orofacial Research*, *2*, 153-159.

Schmidt, F. R. (2005). Optimization and scale up of industrial fermentation processes. *Appl Microbiol Biotechnol*, *68*(4), 425-435.

Schmidt, M. M. (2016). Collagen extraction process. *International Food Research Journal*, *23*(3), 913-922.

Setina, C. M., Haase, J. P., & Glatz, C. E. (2016). Process integration for recovery of recombinant collagen type I alpha1 from corn seed. *Biotechnol Prog*, *32*(1), 98-107.

Shang, L., Tian, P. Y., Kim, N. J., Chang, H. N., & Hahm, M. S. (2009). Effects of Oxygen Supply Modes on the Production of Human Growth Hormone in Different Scale Bioreactors. *Chemical Engineering & Technology*, *32*(4), 600-605.

Sharma, S. R., Poddar, R., Sen, P., & Andrews, J. T. (2008). Effect of vitamin C on collagen biosynthesis and degree of birefringence in polarization sensitive optical coherence tomography (PS-OCT). *African Journal of Biotechnology*, *7*(12), 2049-2054.

Shehadul Islam, M., Aryasomayajula, A., & Selvaganapathy, P. R. (2017). A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines*, *8*(3), 83.

Shekhter, A. B., Fayzullin, A. L., Vukolova, M. N., Rudenko, T. G., Osipycheva, V. D., & Litvitsky, P. F. (2019). Medical Applications of Collagen and Collagen-Based Materials. *Curr Med Chem*, *26*(3), 506-516.

Shen, J.-n., Li, D.-d., Jiang, F.-y., Qiu, J.-h., & Gao, C.-j. (2009). Purification and concentration of collagen by charged ultrafiltration membrane of hydrophilic polyacrylonitrile blend. *Separation and Purification Technology*, *66*(2), 257-262.

Shenoy, M., Abdul, N. S., Qamar, Z., Al Bahri, B. M., Al Ghalayini, K. Z. K., Kakti, A., ... & Alqarni, M. A. (2022). Collagen structure, synthesis, and its applications: A systematic review. *Cureus*, *14*(5).

Shoseyov, O., Posen, Y., & Grynspan, F. (2014). Human collagen produced in plants: more than just another molecule. *Bioengineered*, *5*(1), 49-52.

Shoulders, M. D., & Raines, R. T. (2009). Collagen structure and stability. *Annu Rev Biochem*, *78*, 929-958.

Shuler, M. L., & Kargi, F. (2001). *Bioprocess Engineering: Basic Concepts*. Prentice Hall.

Siddiqui, Y. D., Arief, E.M., Yusoff, A., Hamid, S. S. A, Norani, T. Y., & Abdullah, M. Y. S. (2013). Extraction, purification and physical characterization of collagen from body wall of sea cucumber (Bohadschia bivitatta). *Journal of Environmental Health*, *4*(2), 53-65.

Silvipriya, K., Kumar, K., Bhat, A., Kumar, B., John, A., & Lakshmanan, P. (2015). Collagen: Animal Sources and Biomedical Application. *Journal of Applied Pharmaceutical Science*, 123-127.

Singh, P. K., Shrivastava, N., & Ojha, B. K. (2019). Enzymes in the Meat Industry. In *Enzymes in Food Biotechnology* (pp. 111-128). Academic Press.

Singh, R. S., & Yadav, M. (2013). Enhanced production of recombinant aspartase of Aeromonas media NFB-5 in a stirred tank reactor. *Bioresour Technol*, *145*, 217-223.

Singh, S., & Tavana, H. (2018). Collagen Partition in Polymeric Aqueous Two-Phase Systems for Tissue Engineering. *Front Chem*, *6*, 379.

Sionkowska, A., Adamiak, K., Musiał, K., & Gadomska, M. (2020). Collagen Based Materials in Cosmetic Applications: A Review. *Materials*, *13*(19), 4217.

Sionkowska, A., Skrzyński, S., Śmiechowski, K., & Kołodziejczak, A. (2017). The review of versatile application of collagen. *Polymers for Advanced Technologies*, *28*(1), 4-9.

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., & Li, Q. (2009). Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci*, *18*(5), 936-948.

Sohoni, S. V., Nelapati, D., Sathe, S., Javadekar-Subhedar, V., Gaikaiwari, R. P., & Wangikar, P. P. (2015). Optimization of high cell density fermentation process for recombinant nitrilase production in E. coli. *Bioresour Technol*, *188*, 202-208.

Song, H., Zhang, S., Zhang, L., & Li, B. (2017). Effect of orally administered collagen peptides from bovine bone on skin aging in chronologically aged mice. *Nutrients*, *9*(11).

Spier, M. R., de Souza Vandenberghe, L. P., Pedroni Medeiros, A. B., & Soccol, C. R. (2012). Application of Different Types of Bioreactors in Bioprocesses. In P. G. Antolli & Z. Liu (Eds.), *Bioreactors: Design, Properties and Applications*. Nova Science Publishers.

Srivastava, A. (2022). *Collagen: Collagen is a glue that holds everything together*. Blue Rose Publishers.

Stoichevska, V., An, B., Peng, Y. Y., Yigit, S., Vashi, A. V., Kaplan, D. L., Werkmeister, J. A., Dumsday, G. J., & Ramshaw, J. A. (2016). Formation of multimers of bacterial collagens through introduction of specific sites for oxidative crosslinking. *J Biomed Mater Res A*, *104*(9), 2369-2376.

Stoichevska, V., Peng, Y. Y., Vashi, A. V., Werkmeister, J. A., Dumsday, G. J., & Ramshaw, J. A. (2017). Engineering specific chemical modification sites into a collagen-like protein from Streptococcus pyogenes. *J Biomed Mater Res A*, *105*(3), 806-813.

Sun, X., Fan, J., Ye, W., Zhang, H., Cong, Y., & Xiao, J. (2016). A highly specific graphene platform for sensing collagen triple helix. *J Mater Chem B*, *4*(6), 1064-1069.

Sutherland, T. D., Peng, Y. Y., Trueman, H. E., Weisman, S., Okada, S., Walker, A. A., Sriskantha, A., White, J. F., Huson, M. G., Werkmeister, J. A., Glattauer, V., Stoichevska, V., Mudie, S. T., Haritos, V. S., & Ramshaw, J. A. (2013). A new class of animal collagen masquerading as an insect silk. *Sci Rep*, *3*, 2864.

Sylvestre, P., Couture-Tosi, E., & Mock, M. (2002). A collagen-like surface glycoprotein is a structural component of the Bacillus anthracis exosporium. *Molecular Microbiology 45*(1), 169–178.

Szweda, P., Gorczyca, G., Filipkowski, P., Zalewska, M., & Milewski, S. (2014). Efficient production of Staphylococcus simulans lysostaphin in a benchtop bioreactor by recombinant Escherichia coli. *Prep Biochem Biotechnol*, *44*(4), 370-381.

Tang, C., Zhou, K., Zhu, Y., Zhang, W., Xie, Y., Wang, Z., ... & Xu, B. (2022). Collagen and its derivatives: From structure and properties to their applications in food industry. *Food Hydrocolloids*, *131*, 107748.

Tomita, M. (2011). Transgenic silkworms that weave recombinant proteins into silk cocoons. *Biotechnol Lett*, *33*(4), 645-654.

Trausch, J. J., & Batey, R. T. (2015). Design of modular "plug-and-play" expression platforms derived from natural riboswitches for engineering novel genetically encodable RNA regulatory devices. *Methods Enzymol*, *550*, 41-71.

Tripathi, N. K. (2016). Production and purification of recombinant proteins from Escherichia coli. *ChemBioEng Reviews*, *3*(3), 116-133.

Tripathi, N. K., Sathyaseelan, K., Jana, A. M., & Rao, P. V. L. (2009). High yield production of heterologous proteins with Escherichia coli. *Defence Science Journal*, *59*(2).

Tripathi, N. K., Shrivastava, A., Biswal, K. C., & Rao, P. V. (2012). Development of a pilot-scale production process and characterization of a recombinant Japanese

encephalitis virus envelope domain III protein expressed in Escherichia coli. *Appl Microbiol Biotechnol*, *95*(5), 1179-1189.

Uhoraningoga, A., Kinsella, G. K., Frias, J. M., Henehan, G. T., & Ryan, B. J. (2019). The statistical optimisation of recombinant beta-glucosidase production through a two-stage, multi-model, design of experiments approach. *Bioengineering*, *6*(3), 61.

Vandersmissen, L., De Buck, E., Saels, V., Coil, D. A., & Anne, J. (2010). A Legionella pneumophila collagen-like protein encoded by a gene with a variable number of tandem repeats is involved in the adherence and invasion of host cells. *FEMS Microbiol Lett*, *306*(2), 168-176.

Varilla, C., Marcone, M., Paiva, L., & Baptista, J. (2021). Bromelain, a group of pineapple proteolytic complex enzymes (Ananas comosus) and their possible therapeutic and clinical effects. A summary. *Foods*, *10*(10), 2249.

Vendruscolo, F., Rossi, M. J., Schmidell, W., & Ninow, J. L. (2012). Determination of oxygen solubility in liquid media. *ISRN Chemical Engineering*, *2012*, 1-5.

Vinayagam, R. (2015), Vytla, R. M., & Chandrasekaran, M. (2015). Development of a simple kinetic model and parameter estimation for biomass and nattokinase production by Bacillus subtilis 1A752. *Austin J. Biotechnol. Bioeng*, *2*(1), 1036-1040.

Walker, J. R., Gnanam, A. J., Blinkova, A. L., Hermandson, M. J., Karymov, M. A., Lyubchenko, Y. L., Graves, P. R., Haystead, T. A., & Linse, K. D. (2007). Clostridium taeniosporum spore ribbon-like appendage structure, composition and genes. *Mol. Microbiol.* , *63*, 629–643.

Wang, D., Wang, C., Wu, H., Li, Z., & Ye, Q. (2016). Glutathione production by recombinant Escherichia coli expressing bifunctional glutathione synthetase. *J Ind Microbiol Biotechnol*, *43*(1), 45-53.

Wang, L., Fan, D., He, J., L, Z., & Zhu, C. (2014). A new strategy for secretory expression and mixed fermentation of recombinant human collagen α1 (III) chain in Pichia pastoris. *Biotechnol Bioprocess Eng*, *19*(5), 916-924.

Wang, T., Lew, J., Premkumar, J., Poh, C. L., & Win Naing, M. (2017). Production of recombinant collagen: State of the art and challenges. *Engineering Biology*, *1*(1), 18-23.

Wang, W., Liu, Q. J., & Cui, H. (2007). Rapid desalting and protein recovery with phenol after ammonium sulfate fractionation. *Electrophoresis*, *28*(14), 2358-2360.

Wang, Y., Li, Q., Zheng, P., Guo, Y., Wang, L., Zhang, T., Sun, J., & Ma, Y. (2016). Evolving the L-lysine high-producing strain of Escherichia coli using a newly developed high-throughput screening method. *J Ind Microbiol Biotechnol*, *43*(9), 1227-1235.

Warner, L. R., Blasick, C. M., Brown, R. J., & Oxford, J. T. (2007). Expression, purification, and refolding of recombinant collagen alpha1(XI) amino terminal domain splice variants. *Protein Expr Purif*, *52*(2), 403-409.

Whittemore, B. A. (2005). *Recombinant collagen production optimization in Escherichia coli* (Doctoral dissertation, Massachusetts Institute of Technology).

Wieczorek, A., Rezaei, N., Chan, C. K., Xu, C., Panwar, P., Bromme, D., Merschrod, S. E., & Forde, N. R. (2015). Development and characterization of a eukaryotic expression system for human type II procollagen. *BMC Biotechnol*, *15*, 112.

Wong, I., García, M. A., Rodríguez, I., Ramos, L. B., & Olivera, V. (2003). Fermentation scale up for production of antigen K88 expressed in Escherichia coli. *Process Biochemistry*, *38*(9), 1295-1299.

Wu, X., Jiang, S., Zhang, M., Luo, S., Li, X., Pan, L., Zheng, Z., & Liu, M. (2014). A new method studying the kinetics of l-lactic acid production by pellets Rhizopus oryzae in semi-continuous fermentation. *Annals of Microbiology*, 65(3), 1473-1480.

Xia, J., Wang, G., Lin, J., Wang, Y., Chu, J., Zhuang, Y., & Zhang, S. (2016). Advances and practices of bioprocess scale-up. *Adv Biochem Eng Biotechnol*, 152, 137-151.

Xiang, Z.-X., Gong, J.-S., Shi, J.-H., Liu, C.-F., Li, H., Su, C., Jiang, M., Xu, Z.-H., & Shi, J.-S. (2022). High-efficiency secretory expression and characterization of the recombinant type III human-like collagen in Pichia pastoris. *Bioresources and Bioprocessing*, *9*(1).

Xie, W., Wu, Q., Kuang, Z., Cong, J., Zhang, Q., Huang, Y., ... & Xiang, Q. (2023). Temperature-controlled expression of a recombinant human-like collagen I peptide in Escherichia coli. *Bioengineering*, *10*(8), 926.

Xu, C., Yu, Z., Inouye, M., Brodsky, B., & Mirochnitchenko, O. (2010). Expanding the family of collagen proteins: Recombinant bacterial collagens of varying composition form triple-helices of similar stability. *Biomacromolecules*, *11*, 348–356.

Xu, X., Gan, Q., Clough, R. C., Pappu, K. M., Howard, J. A., Baez, J. A., & Wang, K. (2011). Hydroxylation of recombinant human collagen type I alpha 1 in transgenic maize co-expressed with a recombinant human prolyl 4-hydroxylase. *BMC Biotechnol*, *11*, 69.

Xu, Y., Keene, D. R., Bujnicki, J. M., Hook, M., & Lukomski, S. (2002). Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J Biol Chem*, *277*(30), 27312-27318.

Xu, Y., & Kirchner, M. (2021). Collagen mimetic peptides. *Bioengineering*, *8*(1), 5.

Yang, C., Hillas, P. J., Baez, J. A., Nokelainen, M., Balan, J., Tang, J., Spiro, R., & Polarek, J. W. (2004). The application of recombinant human collagen in tissue engineering. *BioDrugs*, *18*(2), 103-119.

Yanti, N. A., Ahmad, S. W., & Muhiddin, N. H. (2018, December). Evaluation of inoculum size and fermentation period for bacterial cellulose production from sago liquid waste. In *Journal of Physics: Conference Series* (Vol. 1116, No. 5, p. 052076). IOP Publishing.

Yari, K., Fatemi, S. S., & Tavallaei, M. (2010). Optimization of the BoNT/A-Hc expression in recombinant Escherichia coli using the Taguchi statistical method. *Biotechnol Appl Biochem*, *56*(1), 35-42.

Yasin, M., Jeong, Y., Park, S., Jeong, J., Lee, E. Y., Lovitt, R. W., Kim, B. H., Lee, J., & Chang, I. S. (2015). Microbial synthesis gas utilization and ways to resolve kinetic and mass-transfer limitations. *Bioresour Technol*, 177, 361-374.

Yasmin, H., Kabashima, T., Rahman, M. S., Shibata, T., & Kai, M. (2014). Amplified and selective assay of collagens by enzymatic and fluorescent reactions. *Sci Rep*, *4*, 4950.

Yasmin, H., Rahman, M. S., & Shibata, T. (2018). Facile and Selective Determination of Dipeptides Using 3-Methylcatechol as a Novel Fluorogenic Reagent. *International Journal of Peptide Research and Therapeutics*, *25*(2), 583-589.

Yasmin, H., Rahman, M. S., Shibata, T., Kabashima, T., & Kai, M. (2015). Sensitive and selective determination of peptides, PG and PGP, using a novel fluorogenic reagent 4-chlorobenzene-1, 2-diol. *Chemical Papers*, *69*(4), 504-509.

Yasmin, H., Shibata, T., Rahman, M. S., Kabashima, T., & Kai, M. (2012). Selective and sensitive determination of peptides using 3,4-dihydroxyphenylacetic acid as a fluorogenic reagent. *Anal Chim Acta*, *721*, 162-166.

Yin, J., Tomycz, L., Bonner, G., & Wang, D. I. (2002). A simple and rapid assay of collagen-like polymer in crude lysate from Escherichia coli. *J Microbiol Methods*, *49*(3), 321-323.

Yoshizumi, A., Fletcher, J. M., Yu, Z., Persikov, A. V., Bartlett, G. J., Boyle, A. L., Vincent, T. L., Woolfson, D. N., & Brodsky, B. (2011). Designed coiled coils promote folding of a recombinant bacterial collagen. *J Biol Chem*, *286*(20), 17512-17520.

Yoshizumi, A., Yu, Z., Silva, T., Thiagarajan, G., Ramshaw, J. A., Inouye, M., & Brodsky, B. (2009). Self-association of streptococcus pyogenes collagen-like constructs into higher order structures. *Protein Sci*, *18*(6), 1241-1251.

Yu, Z., An, B., Ramshaw, J. A., & Brodsky, B. (2014). Bacterial collagen-like proteins that form triple-helical structures. *J Struct Biol*, *186*(3), 451-461.

Yu, Z., Brodsky, B., & Inouye, M. (2011). Dissecting a bacterial collagen domain from Streptococcus pyogenes: sequence and length-dependent variations in triple helix stability and folding. *J Biol Chem*, *286*(21), 18960-18968.

Yu, Z., Kabashima, T., Tang, C., Shibata, T., Kitazato, K., Kobayashi, N., Lee, M. K., & Kai, M. (2010). Selective and facile assay of human immunodeficiency virus protease activity by a novel fluorogenic reaction. *Anal Biochem*, *397*(2), 197-201.

Yu, Z., Visse, R., Inouye, M., Nagase, H., & Brodsky, B. (2012). Defining requirements for collagenase cleavage in collagen type III using a bacterial collagen system. *J Biol Chem*, *287*(27), 22988-22997.

Yusof, F. (2015). Purification of Recombinant Protein for Industrial Use. In *Recombinant Enzymes - From Basic Science to Commercialization* (pp. 61-80).

Zare, H., Mir Mohammad Sadeghi, H., & Akbari, V. (2019). Optimization of fermentation conditions for reteplase expression by Escherichia coli using response surface methodology. *Avicenna J Med Biotechnol*, *11*(2), 162-168.

Zaslona, H., Trusek-Holownia, A., Radosinski, L., & Hennig, J. (2015). Optimization and kinetic characterization of recombinant 1,3-beta-glucanase production in Escherichia coli K-12 strain BL21/pETSD10 - a bioreactor scale study. *Lett Appl Microbiol*, *61*(1), 36-43.

Zhang, C., Baez, J., Pappu, K. M., & Glatz, C. E. (2009). Purification and characterization of a transgenic corn grain-derived recombinant collagen type I alpha 1. *Biotechnol Prog*, *25*(6), 1660-1668.

Zhang, C., Fan, D., Shang, L. a., Ma, X., Luo, Y. e., Xue, W., & Gao, P. (2010). Optimization of fermentation process for human-like collagen production of recombinant Escherichia coli using response surface methodology. *Chinese Journal of Chemical Engineering*, *18*(1), 137-142.

Zhao, X., Wang, Y., Shang, Q., Li, Y., Hao, H., Zhang, Y., Guo, Z., Yang, G., Xie, Z., & Wang, R. (2015). Collagen-like proteins (ClpA, ClpB, ClpC, and ClpD) are required for biofilm formation and adhesion to plant roots by Bacillus amyloliquefaciens FZB42. *PLoS One*, *10*(2), e0117414.

Zhong, J.-J. (2010). Recent advances in bioreactor engineering. *Korean Journal of Chemical Engineering*, *27*(4), 1035-1041.

Zhou, Y., Han, L. R., He, H. W., Sang, B., Yu, D. L., Feng, J. T., & Zhang, X. (2018). Effects of agitation, aeration and temperature on production of a novel glycoprotein GP-1 by Streptomyces kanasenisi ZX01 and scale-up based on volumetric oxygen transfer coefficient. *Molecules*, *23*(1), 125.

Zhu, C., Fan, D., Duan, Z., Xue, W., Shang, L., Chen, F., & Luo, Y. (2009). Initial investigation of novel human-like collagen/chitosan scaffold for vascular tissue engineering. *J Biomed Mater Res A*, 89(3), 829-840.

APPENDIX I MATERIALS

Process/Step	Chemicals/Reagents	Formula	Manufacturer
Fermentation medium	See M9 formulation in Appendix IV.		
Antifoaming agent	SE 15	-	Sigma Aldrich
Overnight culture medium	Luria Bertani, Vegitone	-	Sigma Aldrich
pH sensor calibration	pH indicator solutions	-	Merck
pH sensor storage solution	Potassium chloride (3M)	KCl	Merck
Acid for pH adjustment	20% phosphoric acid	H_3PO_4	Merck
Base for pH adjustment	25% ammonia solution	NH_4OH	Merck
Lubricant for bioreactor mechanical seal	Glycerin	$C_3H_8O_3$	Hanseler AG, Switzerland
Cell lysis buffer	Disodium hydrogen phosphate	Na_2HPO_4	Merck
	Sodium dihydrogen phosphate	NaH_2PO_4	Merck
	Ethylenediaminetetracetic acid (EDTA)	$C_{10}H_{16}N_2O_8$	Sigma
	L-cysteine		Sigma
ATPS Purification	PEG (Polyethylene Glycol) 1500		Merck
	PEG 2000		Merck
	PEG 4000		Merck
	PEG 6000		Merck
	PEG 8000		Merck
	Dipotassium hydrogen phosphate	K_2HPO_4	Merck
	Potassium dihydrogen phosphate	KH_2PO_4	Merck
Collagen detection Assay	Bacterial (Clostridium hystolyticum) collagenase	-	

	3, 4-dihydroxyphenylacetic acid	$C_8H_8O_4$	
	50 mM Tris Buffer (pH 7.5)		
	Calcium chloride	$CaCl_2$	
	Sodium hydroxide	NaOH	
	Boric acid	H_3BO_3	
	Sodium tetraborate	$Na_2[B_4O_5(OH)_4].nH_2O$	
	Sodium periodate	$NaIO_4$	
Bradford Assay reagent	Quickstart Bradford Protein Assay	-	BioRad, Germany
SDS-PAGE and Western Blotting	30% Acrylamide/Bis	-	(Bio-Rad Laboratories Inc., USA)
	Tetramethylethylenedimine (TEMED)	$C_6H_{16}N_2$	(Bio-Rad Laboratories Inc., USA)
	10% APS	$(NH_4)_2S_2O_8$	(Bio-Rad Laboratories Inc., USA)
	10% SDS	$NaC_{12}H_{25}SO_4$	(Bio-Rad Laboratories Inc., USA)
	Stacking Gel Buffer 0.5 M Tris-HCl, pH 6.8	$NH_2(CH_2OH)_3HCl$	(Bio-Rad Laboratories Inc., USA)
	Resolving Gel Buffer 1.5 M Tris-HCl, pH 8.8	$NH_2(CH_2OH)_3HCl$	(Bio-Rad Laboratories Inc., USA)
	Laemmli Sample buffer	65.8 mM Tris-HCl, Glycerol 26.3% (w/v), SDS 2.1% (w/v), Bromophenol Blue 0.01% (w/v), pH 6.8,	(Bio-Rad Laboratories Inc., USA)
	1x Tris/ Glycine/SDS-Running buffer	25 mM Tris, 192 mM Glycine, SDS 0.1% (w/v), pH 8.3.	(Bio-Rad Laboratories Inc., USA)
	β-mercaptoethanol	C_2H_6OS	BioRad, Germany
	Bromophenol blue	$C_{19}H_{10}Br_4O_5$	BioRad, Germany
	Glycine	$C_2H_5NO_2$	Merck
	Glycerol	C_3H_7OH	Omnipur
	Sodium thiosulfate	$Na_2S_2O_3$	Sigma, USA

	Acetic acid	$C_2H_4O_2$	Sigma, USA
	Methanol	CH_3OH	Sigma, USA
	Coomassie blue dye	$C_{45}H_{44}N_3NaO_7S_2$	Merck
	pre-stained protein marker	-	(PageRuler, Pierce Thermo Scientific, Rockford, IL, USA
	Imperial protein stain reagent	-	(Pierce Thermo Scientific, Rockford, IL, USA)
	Skim Milk	-	
	Anti-6X Histag antibody	-	GeneTex
	ultrapure water	-	-
Trypsin Digestion	Trypzean Enzyme	-	Sigma Aldrich
	Bovine pancreatic trypsin		Sigma Aldrich
Pepsin Digestion	Porcine stomach pepsin	-	Sigma Aldrich
Glucose assay kit	colorimetric glucose assay kit II	-	Cell Biolabs Inc, USA
Agarose Gel Electrophoresis	Agarose		Axil Scientific, Singapore
	TBE Buffer		
	Florosafe DNA stain		
	DM3100 DNA ladder (10 kbp)		
PCR	Master Mix	-	Axil Scientific, Singapore
	Primers	-	

APPENDIX II LIST OF EQUIPMENT

Name	Specifications	Model, Manufacturer
Autoclave		HV-130, Hiclave, Hirayama
Mass Balance		Ohaus, Pioneer, China
Biophotometer		Eppendorf
Bioreactor	2 L	Labfors, Infors HT, Switzerland
Bioreactor	7.5 L	Labfors, Infors HT, Switzerland
Blood Glucose Monitor		OneTouch, Select Simple, LifeScan Europe
Centrifuge		MiniSpin, Eppendorf
Centrifuge		Heraeus Multifuge X1R, Thermo Scientific
Centrifuge		Sigma 1-14, Sartorius
Chiller	4 ℃	Berjaya
Circulator chiller		Alpha RA8, Lauda
		LabTech
Freezer	-20 ℃	Thermo Scientific
Freezer	-80 ℃	Thermo Scientific
Gel electrophoresis set		Bioer Mini GE-100 , Hangzhou Bioer Technology Co. Ltd
High Performance Centrifuge		Avanti J-265 XPI, Beckman Coulter
Hot plate with Magentic stirrer		C-Mag HS7, Ika
		Fisher Scientific
		PC-420D, Corning
Incubator		Ecotron ,Infors HT
Incubator shakers		Infors HT and
		Sartorius
Ice Maker		Hoshizaki

Lab dryer		FDD-720, Protech
Laminar flow cabinet		AHC-4D1, ESCO
Microwave oven	42 L	Panasonic
Nitrogen cylinder	47 L	Fuelink Sdn Bhd.
Oxygen cylinder	47 L, 99.7% Purity	Fuelink Sdn Bhd.
PCR machine		5020 Arktik, Thermo Fisher Scientific
pH/ion meter		SevenCompact S220 Mettler Toledo
SDS PAGE set		Biorad, USA
Spectrofluorometer		Tecan Infinite 200 Pro
Ultrasonic Homogenizer		Labsonic (Sartorius Stedim Biotech)
Microplate reader		Multiscan Go, Thermo Scientific, USA
Vacuum filtration set		Disposable type
Vacuum pump		Sartorius Stedim
Vortex mixer		EVM-6000, Erla
Water bath		Isotemp 220, Fisher Scientific
		SUB Aqua 26 Plus, Grant
Water distiller		Automatic Water Stills (Double distillation stills system), Aquatron A4000D, Stuart Bibby Scientific

APPENDIX III DNA SEQUENCING

```
# Length: 638
# Identity:     638/638 (100.0%)
# Similarity:   638/638 (100.0%)
# Gaps:           0/638 ( 0.0%)
# Score: 3190
#
#
#=======================================

EMBOSS_001         2 TGGCTGAAGCTGAATCCACCCGCCGCGGACGTCCGGGGCCGCAGGGCCCG     51
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         2 TGGCTGAAGCTGAATCCACCCGCCGCGGACGTCCGGGGCCGCAGGGCCCG     51

EMBOSS_001        52 CGCGGACGTCCCGGAGAACCGGGCCGTCCGGGACCGCAAGGACATCCGGG    101
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001        52 CGCGGACGTCCCGGAGAACCGGGCCGTCCGGGACCGCAAGGACATCCGGG    101

EMBOSS_001       102 CCGTCCCGGCCCCGAGGGCCCCCGAGGCAAGCAAGGTCCGGTCGGCAAGC    151
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       102 CCGTCCCGGCCCCGAGGGCCCCCGAGGCAAGCAAGGTCCGGTCGGCAAGC    151

EMBOSS_001       152 CCGGCCCGCAGGGCAAGGCCGGTCCGCAAGGCAAACCCGGCATCGCCGGC    201
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       152 CCGGCCCGCAGGGCAAGGCCGGTCCGCAAGGCAAACCCGGCATCGCCGGC    201

EMBOSS_001       202 AAGCCGGGCCCCGATGGCAAGCCCGGTCCGATCGGCCCGCAAGGCAAGGC    251
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       202 AAGCCGGGCCCCGATGGCAAGCCCGGTCCGATCGGCCCGCAAGGCAAGGC    251

EMBOSS_001       252 CGGCCCGCAAGGTCCCCGCGGTGAACAAGGCCTGCGCGGCGAGCAAGGTC    301
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       252 CGGCCCGCAAGGTCCCCGCGGTGAACAAGGCCTGCGCGGCGAGCAAGGTC    301
```

```
EMBOSS_001      302 CTCGCGGTGAGCAAGGCCCGCAGGGGCCGCGTGGCGAACAGGGCCCCCGG    351
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      302 CTCGCGGTGAGCAAGGCCCGCAGGGGCCGCGTGGCGAACAGGGCCCCCGG    351

EMBOSS_001      352 GGCGAGCCCGGCCCCGCCGGGGCGCTCCCCTCGATCGAGCAGGTGATGCC    401
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      352 GGCGAGCCCGGCCCCGCCGGGGCGCTCCCCTCGATCGAGCAGGTGATGCC    401

EMBOSS_001      402 CTGGCTGCACCTGATCTTCGACGCCTACGAAGATTACAAAGCGCAGCGCG    451
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      402 CTGGCTGCACCTGATCTTCGACGCCTACGAAGATTACAAAGCGCAGCGCG    451

EMBOSS_001      452 CCCGCGAAGCCCGCGAGCTCGAAGAGCGTCTCGCCGCCGAAGCTCTCGAA    501
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      452 CCCGCGAAGCCCGCGAGCTCGAAGAGCGTCTCGCCGCCGAAGCTCTCGAA    501

EMBOSS_001      502 CAGGCGGCGCGCGAAGCCGCCGAGCGCGAAGTCGCCGCCGCTATCGAAGC    551
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      502 CAGGCGGCGCGCGAAGCCGCCGAGCGCGAAGTCGCCGCCGCTATCGAAGC    551

EMBOSS_001      552 CGCAAATGCCGAGGCCGAGATCATGCTCGACGATGAGACGCATGCCGAGG    601
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001      552 CGCAAATGCCGAGGCCGAGATCATGCTCGACGATGAGACGCATGCCGAGG    601

EMBOSS_001      602 GCGGCAAGAAAAAGAAGAAGCGCAAGCACAAGGACTGA    639
                    |||||||||||||||||||||||||||||||||||||||
EMBOSS_001      602 GCGGCAAGAAAAAGAAGAAGCGCAAGCACAAGGACTGA    639
```

APPENDIX IV ORIGINAL RECIPE FOR M9 MEDIUM

Materials (Elbing & Brent, 2002):

- 5X Salt Stock
- Na_2HPO_4
- KH_2PO_4
- NH_4Cl
- NaCl
- $CaCl_2$ (optional)
- Additional Components for 1X Working Solution
- $MgSO_4.7H_2O$
- Carbon source (sugar or glycerol)
- 0.5% vitamin B1 (thiamine)
- 20% Casamino Acids or L amino acids to 40 µg /mL or DL amino acids to 80 µg/mL
- Antibiotic

Procedure (Elbing & Brent, 2002):

5X Salt stock

To a 2-liter flask, add:

- 30 g Na_2HPO_4
- 15 g KH_2PO_4
- 5 g NH_4Cl
- 2.5 g NaCl
- 15 mg $CaCl_2$ (optional)
- 1 liter of high-quality distilled water

Once the ingredients are added, heat while stirring until the components are completely dissolved. Pour the solution into smaller bottles with loosened caps and the autoclave at 15 lb/in^2 for 15 min.

* If you wish to add antibiotics or nutritional supplements, do this only after the autoclave cycle is complete, as the high temperature may destroy these components. Wait until the bottle is less than 50 °C (warm to touch) and then add the components. After the bottles cool to below 40 °C, the caps can be tightened, and the concentrated medium stored indefinitely at room temperature.

1× Working Solution

To make 1× working solution, the 5× media should be diluted to 1× with high quality sterile distilled water.

Add the following sterile solutions

1 mL 1 M $MgSO_4.7H_2O$

10 mL 20% carbon source (sugar or glycerol)

Typically, several additional components are also added to make a complete medium.

vitamin B1(thiamine) 0.5% 0.1 mL

Casamino Acids 5 mL 20% or L amino acids to 40 µg/mL or DL amino acids to 80 µg/mL

Antibiotic

10× M9 salt:

Components	**Molecular Weight**	**Total volume = 250 mL**	**Total Volume= 500 mL**
Na_2HPO_4	141.96	15 g	30 g
KH_2PO_4	136.09	7.5 g	15 g
NaCl	58.44	1.25 g	2.5 g
NH_4Cl	53.49	2.5 g	5 g
dH_2O		250 mL	500 mL

Adjust pH to 7.2 before adding all water

6. If too Acidic : add NaOH
7. If too Basic: Add HCl

#Autoclave

M9 minimal media (100 mL)

Components	**Volume**	**Final Concentration**
10x M9 minimal salts	10 mL	1x
20% glucose	2 mL	0.4%
1 M $MgSO_4$	200 µl	2 mM
Thiamine HCl (1 mg/mL)	100 µl	1 µg
Biotin (1 mg/mL)	100 µl	1 µg
100x Trace Elements	1 mL	1x
1 M $CaCl_2$	10 µl	0.1 mM
Sterile dH_2O	86.6 mL	-
Ampicillin	100 µl	50 µg/mL

APPENDIX V MODIFIED M9-Cas MEDIUM COMPOSITION

Table 5.2 M9 Medium Composition as optimized by Soroodi et al. (Unpublished)

Component	**Chemical Formula**	**Preparation volume**		**Manufacturer**
Distilled water		100ml	1 L	-
Disodium hydrogen phosphate	Na_2HPO_4	-	6	Merck
Potassium dihydrogen phosphate	KH_2PO_4	-	3	Merck
Sodium chloride	NaCl	-	0.5	Merck
Ammonium chloride	NH_4Cl	-	1	Merck
Magnesium sulfate	$MgSO_4$	0.24	-	Merck
Thiamine		0.001	-	Sigma-Aldrich
Calcium chloride	$CaCl_2$	0.00111	-	Merck, Sigma-Aldrich
Casamino acid				Merck
D-glucose	$C_6H_{12}O_6$			Merck
Ampicillin	-			Roche
Trace elements	-			-

Components	Volume	Final Concentration
10× M9 minimal salts	10 mL	1×
20% glucose	2 mL	0.4%
1 M $MgSO_4$	200 µl	2 mM
Thiamine HCl (1 mg/ml)	100 µl	1 µg
Biotin (1 mg/ml)	100 µl	1 µg
100× Trace Elements	1 mL	1×
1 M $CaCl_2$	10 µl	0.1 mM
Sterile dH_2O	86.6 mL	-
Ampicillin	100 µl	50 µg/ml

Components	Molecular Weight	Total volume = 250 mL	Total Volume= 500 mL
Na_2HPO_4	141.96	15 g	30 g
KH_2PO_4	136.09	7.5 g	15 g
NaCl	58.44	1.25 g	2.5 g
NH_4Cl	53.49	2.5 g	5 g
dH_2O		250 mL	500 mL

APPENDIX VI- SETUP OF THE 2 L AND 7.5 L BIOREACTORS

Figure 5.1 Setup of the 2 L bioreactor, Labfors, INFORS HT, Switzerland

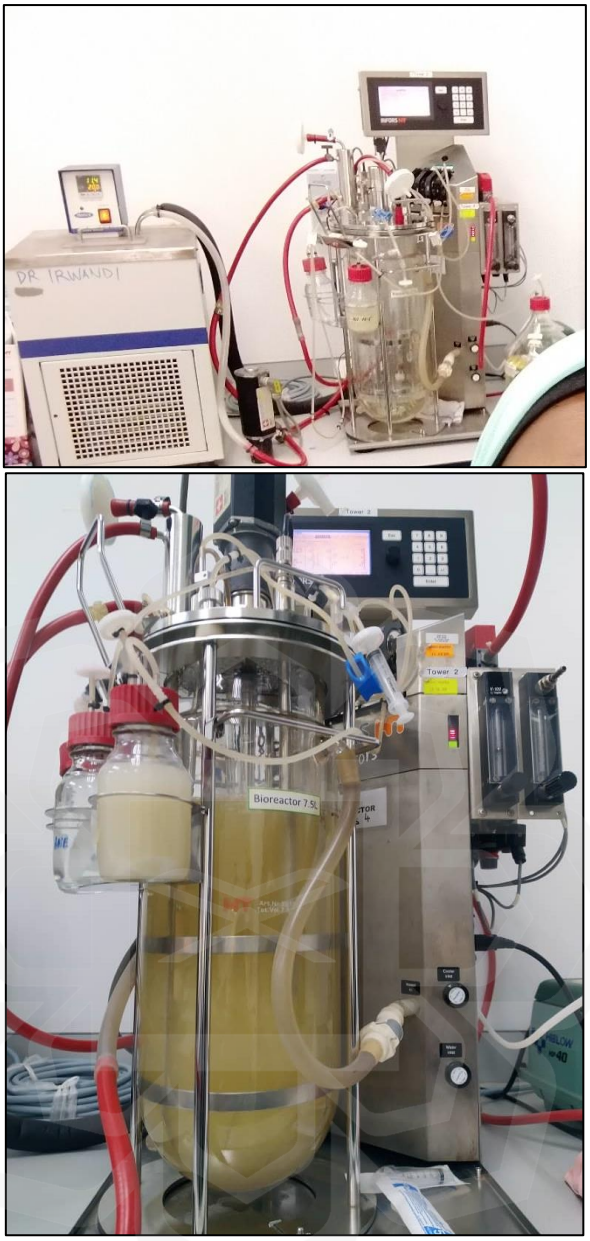


Figure 5.2 Setup of the 7.5 L bioreactor, Labfors, INFORS HT, Switzerland

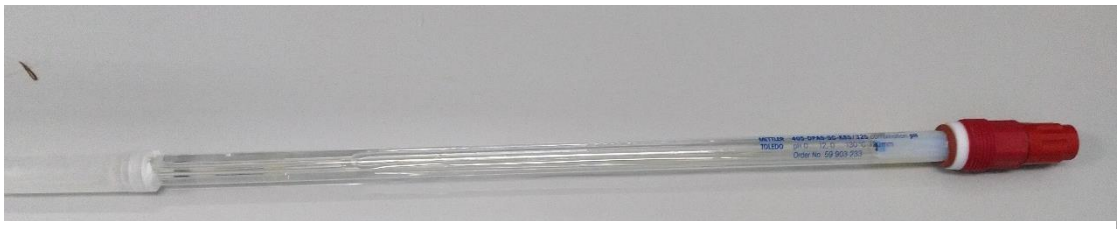


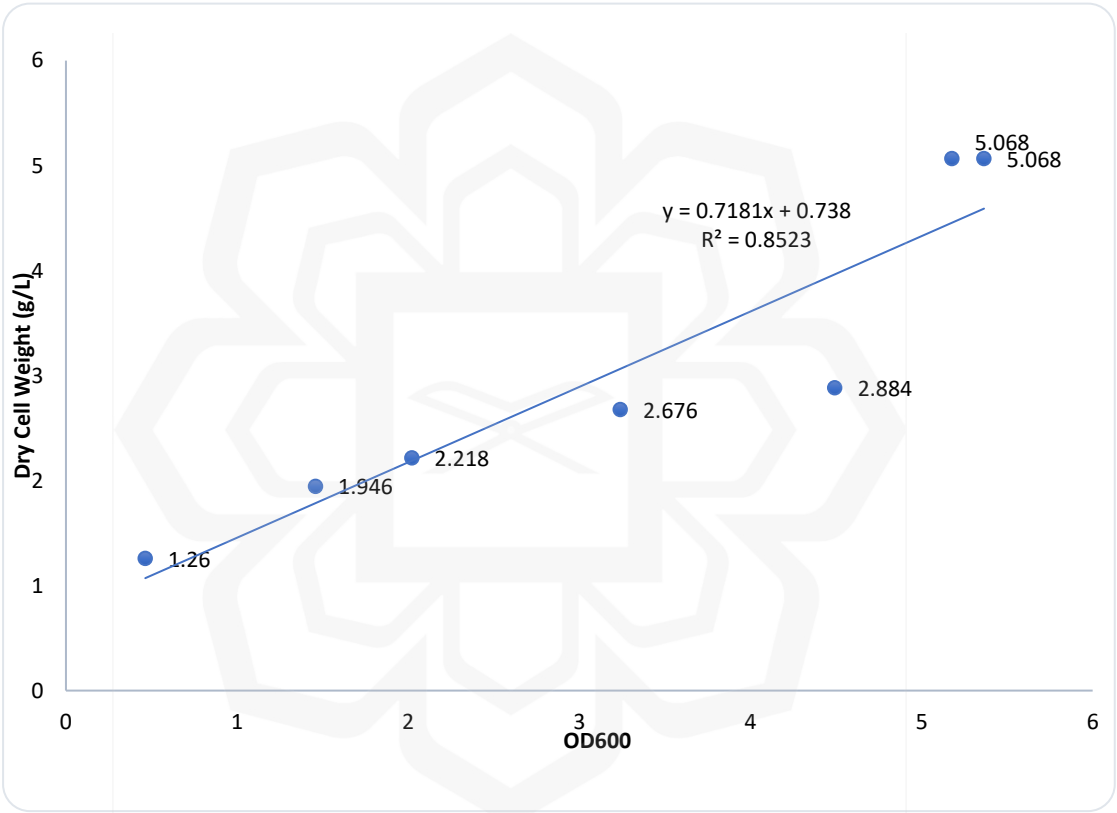
Figure 5.3 pH sensor (L=325mm), Mettler-Toledo

Figure 5.4 pO_2 electrode, Mettler-Toledo

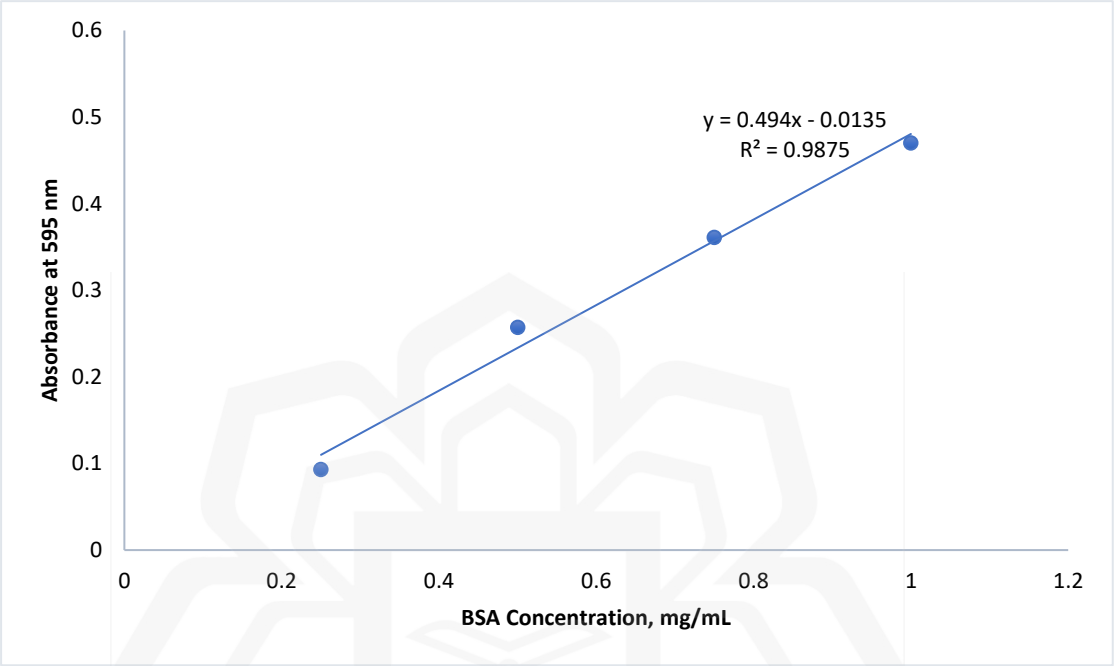
Figure 5.5 Super Safe Sampler for aseptic sampling, INFORS HT, Switzerland

APPENDIX VII STANDARD CURVES

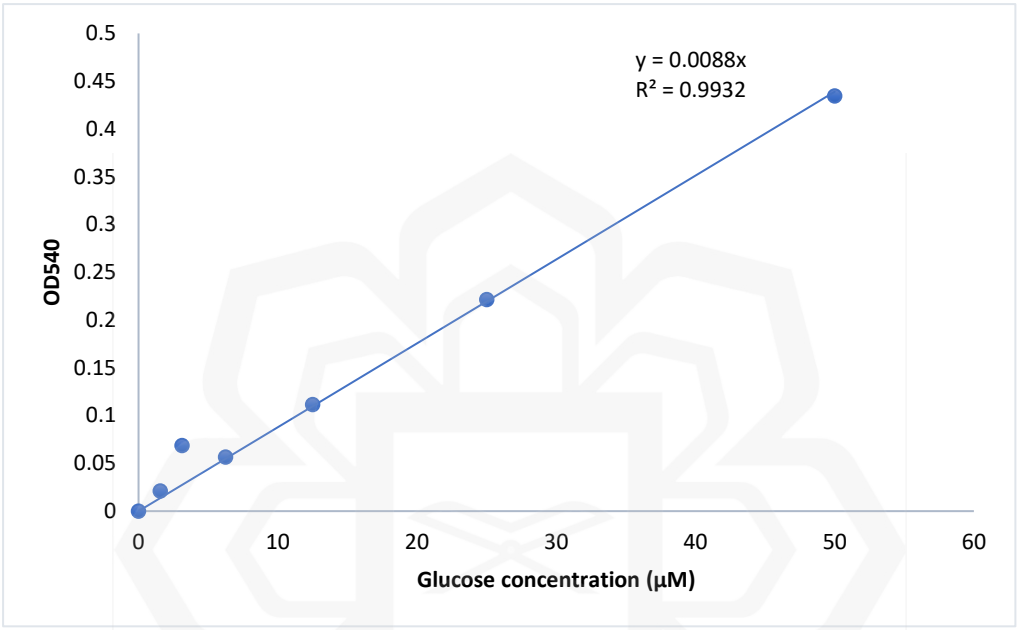
Standard curve of DCW (in g/L) vs Absorbance (OD_{600})



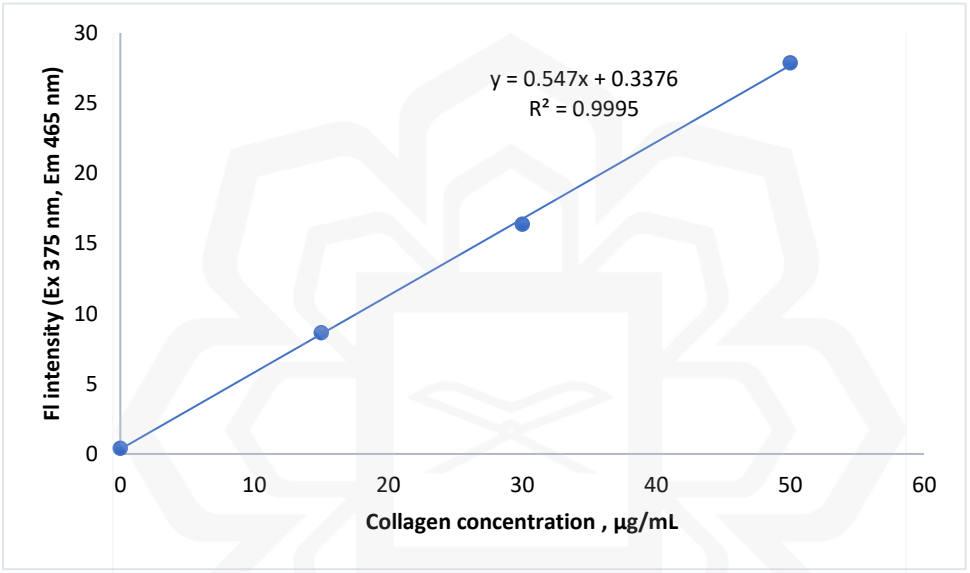
Standard curve of Absorbance (OD_{595}) vs Bovine Serum Albumin (BSA) concentration (µM)



Standard curve of Absorbance (OD_{540}) vs glucose concentration (µM)



Standard curve of fluorescence intensity vs collagen concentration



APPENDIX VIII- MATLAB CODES

General function call

```
time    = [0 0.5 1 1.5 2 3 4 5 6 7 8 24]';
cells   = [0.9283 1.0108 1.2169 1.4294 1.8954 2.7951 3.4599 4.3761 5.8232
6.8051 7.8964 8.2519]';
glucose = [10 0.01988 0.019554 0.0203 0.0204 0.00193 0.0196 0.0197 0.0180
0.0137 0.0126 0.0041]';
protein = [0.0000 1.5671 0.6701 1.0961 1.5221 2.0091 1.6531 1.5411 1.8791
1.8731 1.9841 1.7561]';   % in mg/l
protein2= (0.001)*protein;
data    = [cells glucose protein2];
mmax = 0.63;            % /h  0.6334
yxs  = 1.6;           % g/g  1.598007193
ksx  = 0.4;           % g/L  0.004, 0.01922
yps   = 0.02;         % g/g  0.0198489
kp   = 200;    % g/L
%Pmax = 0.002;   % g/l  0.2009084*0.001
Pmax= max(protein2);
bet   = 1/(max(cells));
%Ypx=yps./yxs;
B0=[mmax yxs ksx yps kp Pmax bet data(1,1) data(1,2) data(1,3)]';
%B0 = rand(10,1)*100;
options = optimoptions('lsqcurvefit','Algorithm','levenberg-marquardt');lb
= []; ub = [];
[B3,resnorm3,residual3,exitflag3,output3,Lmda3,Jmat3] =
lsqcurvefit(@mixedmodel,B0,time,data,lb,ub,options);
t = linspace(min(time),max(time),250);
Mixedfit= mixedmodel(B3,t);
mixedcells=Mixedfit(:,1); mixedsubs=Mixedfit(:,2);
mixedprod=1000*Mixedfit(:,3);
new=mixedmodel(B3,time);
% plotting the data and the models
figure; title('Biomass Data and Fitted Models');hold on; grid;
xlabel('Time (hours)'); ylabel('Biomass Concentration (g/L)');
plot(time,cells,'bx',t,mixedcells,'b-');
figure; title('Substrate Data and Fitted Models');hold on; grid;
legend('Data dXdt','Monod Model dXdt');
xlabel('Time (hours)'); ylabel('Biomass and Glucose Concentration (g/L)');
plot(time,glucose,'rx',t,mixedsubs,'r-');
legend('Data dSdt','Luedeking-Piret Model dSdt');
figure; title('Product Data and Fitted Models');hold on; grid;
ylabel('Collagen-like Protein Concentration (mg/L)');
plot(time,protein,'gx',t,mixedprod, 'g-');
legend('Data dPdt','Logistic Model dPdt');
% statistics/information about fit
%cint3 = nlparci(B3,residual3,Jmat3);
%fprintf('The mixed model  has residual norm %f,\n',resnorm3);
%fprintf('Confidence interval is : %c \n',cint3);
```

```
%fprintf('New parameters: umax = %f,yxs =  %f, ksx = %f, b= %f, yp = %f, kp
= %f, Pmax = %f, qp = %f,\n',B3);
fprintf('\nExperimental data %f \n',new(:,2));
```

Logistic-Luedeking-Piret (LLP) model

```
% logistic_model codes the system of differential equations describing the
Logistic model:
function S = logistic_model(B, t)
x0 = B(8:10);         % initial values of x0, s0, p0
[T,SX] = ode45(@DifEq, t, x0);
      function dS = DifEq(t,x)
          umax=B(1); Yxs=B(2);  Ksx = B(3);
          Yps= B(4); Ksp = B(5); Pmax = B(6); qpmax = B(7);
          %initial guesses
                    X=x(1); Sub=x(2); P=x(3);
          xdot    = zeros(3,1);
          k=1.0;
          xm=8.25;

          xdot(1)=k.*X.*(1-(X./xm));        % biomass formation

          xdot(2)=-(xdot(1)./Yxs)-(xdot(3)./Yps);   % substrate
consumption

          xdot(3)= qpmax.*(Sub./(Ksp+Sub)).*X.*(1-P/Pmax);  % product
formation

          dS = xdot;
      end
S=SX;
end
```

Monod-Luedeking-Piret (MLP) model

```
% mixedmodel codes the system of differential equations describing the
Monod-Luedeking-Piret model:

function S = mixedmodel(B, t)
x0 = B(8:10);         % initial values of x0, s0, p0
[T,SX] = ode45(@DifEq, t, x0);
      function dS = DifEq(t,x)
     umax=B(1); Yxs=B(2);  Ksx = B(3);
          Yps= B(4); Ksp = B(5); Pmax = B(6); bet = B(7);
     X=x(1); Sub=x(2); P=x(3);
         xdot=zeros(3,1);
         u=(umax .* Sub) ./ (Ksx + Sub);
         Ypx=Yps./Yxs;
```

```
        xdot(1)=u.*X;                                % monod dX/dt,

        xdot(2)=-1.*((xdot(1))./Yxs)-((xdot(3))./Yps);   % luedeking dS/dt

        xdot(3)=(Ypx.*(xdot(1))) + (bet.*X);              % luedeking dP/dt

       % xdot(3)= qpmax.*(Sub./(Ksp+Sub)).*X.*(1-(P./Pmax)); % dP/dt
        dS = xdot;
    end
S=SX;
end
```

LIST OF PUBLICATIONS & PRESENTATIONS

1. Gameil, A. H. M., Yusof, F., Azmi, A. S., & Mohamad Puad, N. I. (2023). Modeling of *E. coli* growth, glucose consumption, and recombinant collagen-like protein formation kinetics. IIUM Engineering Congress Proceedings, 1 (1), 74–78.

2. Gameil, A. H. M., Yusof, F., Azmi, A. S., & Mohamad Puad, N. I. (2021). Process scale-up criteria in production of recombinant proteins in *E. coli*: A systematic review. *Biological and Natural Resources Engineering Journal*, *5*(2).

3. Gameil, A. H. M., Yusof, F., Azmi, A. S., & Mohamad Puad, N. I. (2021). Progress in the detection and quantification of collagens: a review. In IOP Conf. Ser.: Mater. Sci. Eng., 1192(1), Kuala Lumpur, Malaysia.

4. KERICE 2020 (Video participation)

5. 3-Minute Thesis competition 2020 (Video participation)

6. 3-Minute Thesis competition 2021 (Video participation)