University of Glasgow | School of Life Sciences

=

"Biogenesis of Silk Fibroin *(Bombyx mori)* Production in Bacteria and Baculovirus-Infected Insect Cell System"

Chirapha Butiman

Submitted in fulfilment of the requirements
for the Degree of Doctor of Philosophy

Molecular, Cell and Systems Biology

School of Molecular Biosciences

College of Medical, Veterinary and Life Science

University of Glasgow

2023

ABSTRACT

This research project has established three main hypotheses: 1) bacterial expression systems be used to produce three of the core silk proteins of *Bombyx mori* (P25 glycoprotein, FIBL, and H-chain); 2) baculovirus-insect cell expression systems be used to produce three of the core silk proteins (P25 glycoprotein, FIBL, and H-chain); and 3) H-chain interacts independently from the molecules of silk proteins and other proteins because of its superior properties, such as high self-assembly and adhesive formation. The protein production in the bacterial system of the two lower molecular fibroin proteins (P25 and FIBL, MWs about 30 kDa) was first carried out by labelling with His6x-tagged at the C-terminus, cloned into pET28a, and expressed in two groups of bacterial strains: Group 1 (un-enhanced disulfide bond formation), including three bacterial cells; BL21 (DE3), C41 (DE3), Rosetta™ (DE3)pLysS (Novagen); and Group 2 (enhanced disulfide bond formation), containing one bacteria strain, SHuffle® T7.

The findings revealed that FIBL protein was expressed as inclusion bodies (IBs), which needed to be solubilised using 7M urea to recover the active protein. FIBL protein production was performed using a larger volume of (250mL) LB media with the optimal growth condition of 0.5mM IPTG and induced overnight at 25ºC. The FIBL obtained the secondary structure using circular dichroism. However, the P25 protein expression had the most potential growth in SHuffle® T7 of 0.5mM IPTG and was induced overnight at 25ºC, but its construct was not stable after maintenance at -20ºC for about three months. In the second attempt, the P25 glycoprotein was labelled on both sides with His6x-tagged + 3C protease (N-terminus), whereas TEV protease + His6x-tagged was added (C-terminus) to the pFBDM plasmid and expressed in Sf21 cells (*Spodoptera frugiperda*). The findings showed the system can promote the complex N-linked P25 glycoprotein. The P25 protein expression showed two forms of glycosylation in the Sf21 cells: nonsecretion (semi-glycosylated) and secretion (medium fractions/glycosylated). The semi-glycosylated P25 contained β-pleated sheet (33.4%), un-ordered (31.5%), turns (20.8%), and helix (14.3%), respectively. In contrast, P25 (secreted) consisted of a β-pleated sheet (47.70%), un-ordered (35.05%), turns (17.05%), and helix (0.20%), respectively. This is also the first finding about the P25 glycoprotein secondary structure.

The H-chain is a large subunit (about 350 kDa) with a high crystallinity structure that cannot be easily expressed using the classical cloning technique. Therefore, the better choice is to isolate the H-chain from the native solid posterior silk gland (PSG), *B. mori*. The native H-chain fibroin was satisfactorily isolated from the solid fibroin prepared from the cocoon shells by the first purification with gel filtration chromatography and the second purification using ion exchange chromatography. Three fibroin (H-chain, FIBL, and P25) element interactions and formations among fibroin molecules revealed the assembled relationship as a stoichiometry complex (H(6):FIBL(6):P25(1)), and other complexes were investigated by a pull-down assay. Additionally, the H-chain was employed to form multimers and interactions that have potential with selected human proteins, CLIC1 and CLIC4, in heterodimeric form, a promising finding for novel medical applications. It was interesting that the native H-chain alone always forms homodimers around 720 kDa.

Key words: fibroin, fibroin heavy chain (H-chain), fibroin light chain (FIBL), fibrohexamerin (P25)

ACKNOWLEDGMENTS

Throughout the PhD course of studies in Molecular, Cell and Systems Biology at the University of Glasgow, UK, I have had many great chances of working with the following mentioned scientists and have benefited from their insight. I am grateful to my supervisor, Prof. Cheryl Woolhead and Dr. Sharon Kelly, for providing insightful guidance and knowledge throughout the project and for giving her time to help correct the contents of my thesis. I would like to acknowledge my committee members Prof. Joseph Gray and Dr.Catherine Berry and a special expert in silk fibroin Prof. Motoyoki Sumida. I am thankful for the valuable discussions which were greatly appreciated when fulfilling my research. I sincerely thank all members of many labs, including, “Woolhead Lab’’, “Bulleid Lab’’, and “Walden Lab’’ who kindly gave useful advice and support with regards to instruction on laboratory techniques and answering questions, including **the late Prof. Neil Bullied (various special guidelines from him to overcome the research difficulties)**, Dr. Ojore Oka, Dr.Philip Robinson, Dr. Marcel van Lith, Dr. Martin Rennie, Donald Campbell, Dr. David Steward and Dr. Hazel Bracken. Marie Anne- Pringle, and June Southall.

I must thank my friendly lab mates, Monalissa Halablab, Shaghayegh Sadeghinia, and Junjie Zhu for occasional support in the lab and for the many nice times we had together.

Additionally, thanks goes to my sponsor, Thai Royal Scholarship, Ministry of Science and Technology, Thailand, and also all the staff who provided the supporting documents during my studies both in London and Bangkok - especially Somchai Injorhor.

I want to give tremendous thanks to my workplace, Excellence Centre of Silk Innovation, Mahasarakham University, Thailand, for allowing me to study at University of Glasgow, UK. I am thankful to all the staff of the office of Education Affairs, London, who have been helping and supporting me whilst living in the UK. Importantly, I give thanks to my beloved family, my parents and my Thai friends for their love and support of me being abroad. I hope I have made them proud.

Special thanks to all the people who helped me to improve my scientific writing for this thesis and also made me laugh; Lynley Capon, Saemus Grant, Adityo Paul and Louie Charles Atkins-Turkish.

Thanks to all the friends I have met whilst studying in University of Glasgow; a special mention to Dr. Chris Bouchard, Dr. Angie Sin, Dr.Pornphan Pichai, Lin Bin (Cookie), Jacek Kurek, Dr. Chaitong Churuangsuk, Dr. Zoa Zhigo (Steve), Ryo Iwase, Dr. Pracha Yambanyang, Dr. Patchani Srikhamsuk and Pichaporn Dokmai.

Last, but not least, I would like to thank you for the great Thai food support from another mother (as Thai culture respect) I have met in Arbroath, UK who always cared about my study life in the UK environment.

Finally, thank you to all cells and silkworms who sacrificed their bodies, enabling me to complete this project.

March, 2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

AUTHOR's DECLARATION

"I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work has not been submitted for any other degree at the University of Glasgow or any other institution. "

Printed Name: CHIRAPHA BUTIMAN

Signature: CHIRAPHA BUTIMAN

LIST OF ABBREVIATIONS

~	– About/Approximately
aa	– amino acid
*Ac*NPV	– *Autographa californica* nucleopolyhedrovirus
ASG	– Anterior silk gland
B. mori	– *Bombyx mori*
bp	– Base pair (unit)
BEVS	–Baculovirus expression vector system
BL21	– A commercial competent cell, BL21 (DE3) – Agilent 200131, USA
BME	– Beta-mercaptoethanol/ 2-Mercaptoethanol
Bis-Tris	– 2,2-Bis(hydroxymethyl)-2,2',2"-nitilotriethanol
BSA	– Bovine serum albumin
3C Protease	– HRV-3C Protease Cleavage Enzyme Protein
C-terminus	– Carboxyl-terminus
˚C	– Degree Celsius
cm	– Centimetre
CD	– Circular Dichroism
CLIC1	– Chloride intracellular channel protein1
CLIC4	– Chloride intracellular channel protein4
Con (A)	– Concanavalin A
Cys	– Cysteine
Da	– Dalton
ddH_2O	– Distilled water
DNA	– Deoxyribonucleic acid
DsbC	– Disulfide bond C
DLS	– Dynamic light scattering
DTT	– Dithiothreitol

EDTA – Ethylenediaminetetraacetic acid
E. coli – *Escherichia coli*
Endo-H – Endoglycosidase H
ER – Endoplasmic reticulum
ESI – Electrospray ionisation
FFF – Field flow fractionation
FIBL/L-chain – Fibroin light chain
GFP – Green fluorescent protein
H-chain – Heavy chain
His_6x – Hexahistidine
hr – Hour
HEPES – 4(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
IBs – Inclusion bodies
IEM –2-isocyanatoethyl methacrylate
IP – Immunoprecipitation
IPTG – Isopropyl β- d-1-thiogalactopyranoside
K – Kilo or (1,000)
kDa – Kilodalton
LB – Lysogeny broth
M – Molar
mM – Millimolar
m – metre
mg – Milligram
mL – Millimetre
MW – Molecular weight
MS – Mass spectrometry
MSG – Middle silk gland
min – Minute
N-terminus – Amino terminus
Ni-NTA column – Nickel (Ni^{2+})-Nitriloacetic acid column
nm – Nanometre

O. D – Optical density

P25/*Fhx* – (P25 glycoprotein)/Fibrohexamerin

% – Per cent sign

PCR – Polymerase chain reaction

PNGase F – Peptide-N-Glycosidase F

PSG – Posterior silk gland

psi – Pound per square inch

RNA – Ribonucleic acid

rpm – Revolutions per minute

RI – Refractive index

RT. – Room temperature

SEC – Size Exclusion Chromatography (Gel filtration machine)

SEC/MALS – Size Exclusion Chromatography/multi-angle light scattering

SS bond – Disulfide bond

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sf21 – A cell line developed from ovaries of fall armyworm (*Spodoptera frugiperda*)

TBS – Tris-buffered saline

T7 DNA polymerase – An enzyme used during replication of the T7 bacteriophage

T7 RNA polymerase – RNA from the T7 bacteriophage

TCEP – Tris (2-chloroethyl) phosphate

TEV – Tobacco etch virus

µg – Microgram

µL – Microlitre

µm – Micrometre

CHAPTER I

INTRODUCTION

……

Natural Silks

Silk is a well-known natural fibre having excellent mechanical properties such as superior; strength, transparency and anhydrosity (Tsukawaki et al., 2016) with a production history of over 4,000 years (Hyde 1984: Nilebäck et al., 2017). Silks are part of insect and arachnid nest construction and are commonly defined as biopolymers. Silk proteins are synthesised in the glands of over 100 species of phylum Arthropoda mostly from insects (class insecta) including silkworms and bees and non-insects (class arachnida), spiders, scorpions and mites (Feughelman, 2002). The mulberry silkworm (*Bombyx mori*) is the only species which can be rearing for mass production contrasting to the spider silks, even though mulberry silk fibre has lower strength and less extensibility than the spider silks (Romer & Scheibel, 2008: Rajkhowa et al., 2015). Over 99% of all silk products traded world-wide come from the mulberry silkworm (Peigler, 1993). Silk fibres are produced during metamorphosis for different purposes. The silkworm produces it as part of the larval cocoon spun to protect the pupa inside, whereas spider silks are used in hunting prey and for reproductive purposes (Wingkler & Kaplan, 2001; Sutherland et al., 2010). Also, the composition, structure, and properties of the silk from each of these sources are varied (Qi et al., 2017). For example, spiders can produce seven various kinds of silk with different properties, ranging from strong dragline silk to elastic capture spiral silk (Hayashi et al., 2004). Silk from spiders and other insects are glycine rich, having unique characteristics in strength and extensibility. As a consequence, silks from spiders differ from the mulberry silk (Hakimi et al., 2007). In nature, a fully mature silkworm will spin a cocoon comprising a single long continuous fibre between 600-1,500m in length depending on in the native (polyvoltine) or bivoltine races (Gopinathan, 1992). Spider silks are produced from a range of different glands, ducts, and spigots (Scheibel, 2004) so cannot be easily harvested from the net of orb weaving with diameters ranging from a few centimetres to several meters (Eisoldt et al., 2011). Spiders are not capable of providing long fibres like the mulberry silkworm. It is possible to make up ~ 137 metres from the silk glands and ~12 metres from their webs (Kundu et al., 2013) but the result of spiders yield depends on the species (Bourzac, 2015). There is evidence that spider silk containing a broad range of fibre types has evolved for several ecological factors (Hsia, et al., 2011). As shown in **Figure 1.**, a silkworm *B. mori*, is a complete metamorphosis insect which must go through four stages, egg-larva-pupa-adult in each generation. It will take 55-60 days for a whole life cycle (Xu, et al.,2021; Damrongsakkul, 2020), depending on its strains (polyvoltine or bivoltine).

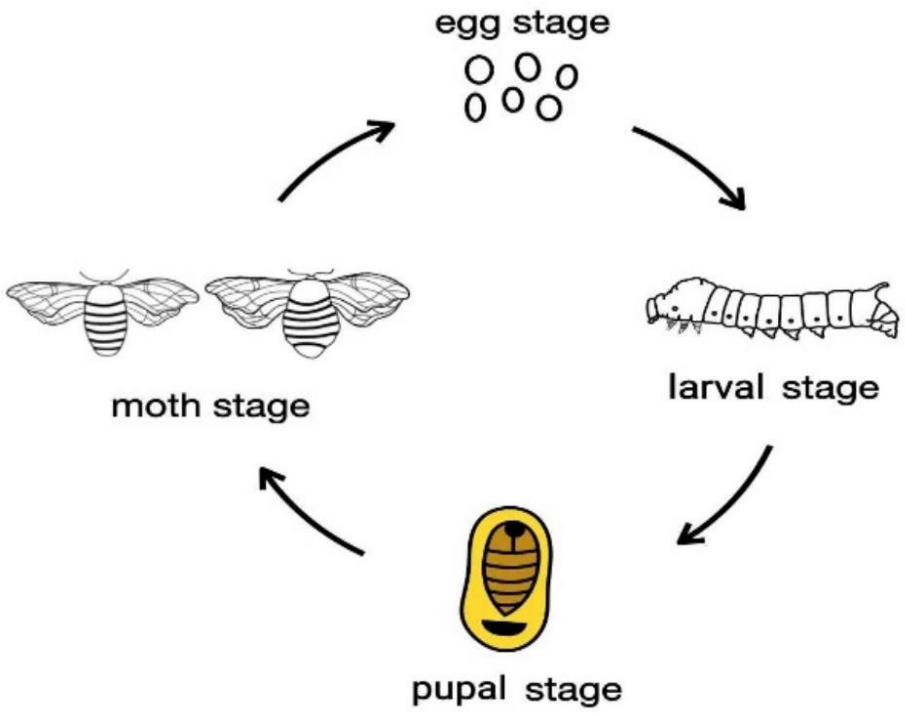


Figure 1. The schematic diagram shows the life cycle of a polyvoltine strain of silkworm (*Bombyx mori*) and its four stages of development over a period of about 45-55 days including, egg, larva, pupa, and adult in one generation (adapted from Damrongsakkul, 2020).

Many previous works have reported on a large amount of silk synthesised in a pair of silk glands of the 5th instar larvae **(Figure 2. A)**. This is the last stage of the silkworm before metamorphosis, during which the cocoon of silk is needed to protect the pupa (Inoue et al., 2000). Silk is primarily composed of two types of proteins, fibroin and sericin. Fibroin forms the main thread core that accounts for approximately 75% of the total silk protein. Sericin (glycoprotein or amorphous protein) provides a glue-like coating that holds two proteins together to form the fibre structure around the pupa called the 'cocoon'. Sericin accounts for about 20-30% of the silk and is primarily synthesised in the middle silk gland **(Figure 2. B)** (Zafar et al., 2015; Wang & Zhang, 2013).

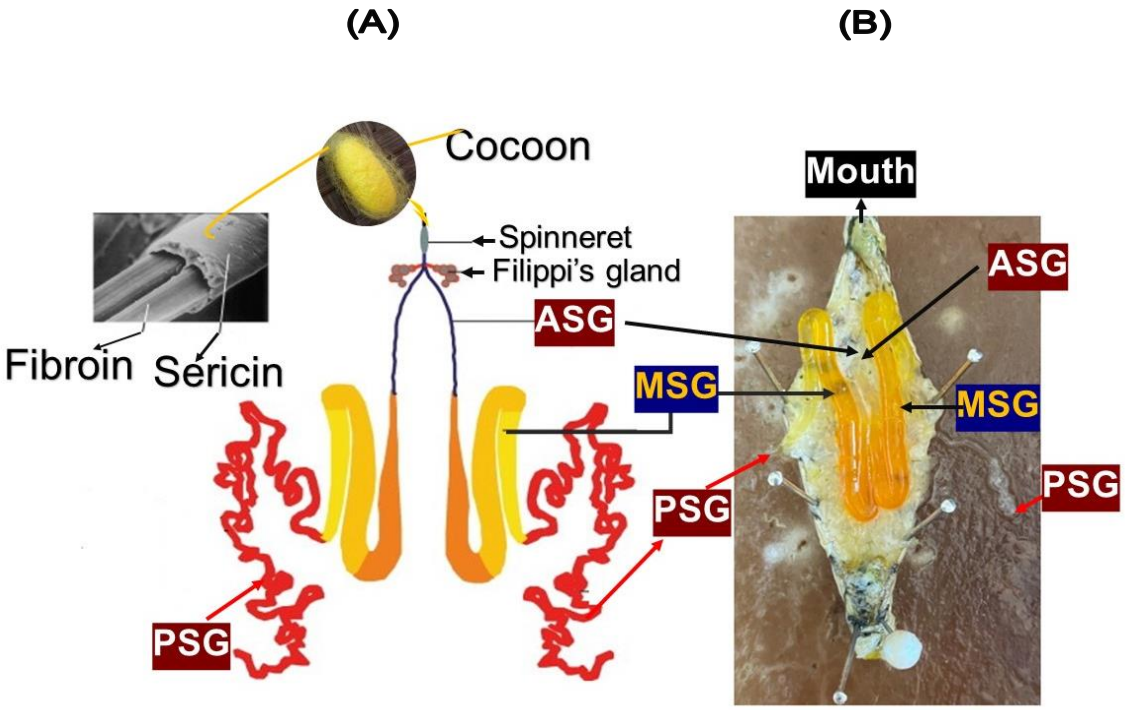


Figure 2. Structural representation of the silk gland in the last instar of a silkworm (*B. mori*) **(A)** modified from Julien et al., 2005**,** where a single silk fibre will be spun in a series through a spinneret at the mouth, whereas an aqueous fibroin complex containing sericin will co-respond to the three main parts from three compartments, anterior silk gland (ASG), middle silk gland (MSG) and the posterior silk gland (PSG). The structure of the silk gland in a dissected silkworm forms a W-shape which has two exocrine silk glands that are aligned on both sides of the silkworm's body **(B**).

Silk Fibroin

Prior studies revealed that fibroin proteins exist in different chromosomes of fibroin genes H-chain= 25th and L-chain (FIBL) =14th, whereas P25 is encoded by the fibrohexamerin gene. Although fibroin genes and the P25 gene are not related in their coding regions, these genes' regulatory systems linked together in PSG cells during the growth stage of the 5th larval instar with very specific and

efficiently coordinated functions in silkworms (Couble et al., 1985). Other than that, the coordinated synthesis of H-chain and FIBL at the 5th instar are related approximately one to one (Takei et al., 1987).

Silk fibroin is an generally insoluble protein in various solvents including water. Silk fibroin is a major protein that has the formation of a macromolecule constructed of about 5,000 amino acids, which is composed of three subunits (Tanaka et al., 1999). In the PSG, silk fibroin is produced and secreted as three major molecule polypeptides; a fibroin H-chain of about 350-390 kDa (Vollkov & Cavaco-Paulo., 2016) and a FIBL of about 30 kDa form a heterodimer, and six dimers are connected with a glycoprotein, P25 to form a large 2.3 MDa elementary structure of silk fibroin as an aqueous solution in the endoplasmic reticulum (ER) (Asakura et al., 2015; Gupta et al., 2015). **As shown in Figure 3,** the formation of fibroin protein mass ratio of H-chain, FIBL and P25 is 6:6:1 (Inoue et al., 1999). The fibroin is accumulated temporarily outside the cells in the glandular lumen of PSG with a very concentrated viscous formation (Feughelman, 2002). Subsequently, this complex will be transported along the MSG where sericin is coated before being spun out through the ASG of the fully grown 5th instar larva which is converted to the single fibre (Asakura et al., 2015).

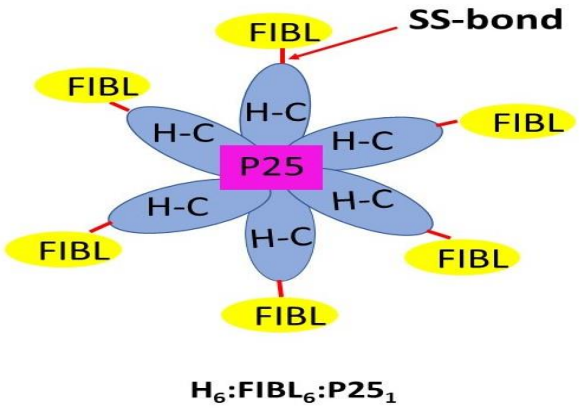


Figure 3. The silk fibroin stoichiometry complex of about 2.3 MDa, designates the elementary unit of fibroin assembled containing covalent bonds with six sets of H-chain- FIBL dimer, and a noncovalent bond of a single P25 glycoprotein.

In **Figure 4A**, the H-chain and L-chain/ FIBL are linked by a single bridge between Cys-172 of the FIBL and Cys-c20 (twentieth residue from the C-terminus) of the H-chain. The H-FIBL linkage complex and P25 are important for the function of fibroin secretion within intracellular and luminar transport in the silk gland (Inoue et al., 2000). The FIBL molecule has a more differentiated amino acid composition, with a non-repetitive amino acid sequence, resulting in naturally occurring hydrophilic and elastic properties (Zafar et al., 2015). **Figure 4B**, the silk fibroin has the mature core fibre of between 10-25 µm (nano fibril) in diameter. Each individual fibre contains the twisted bunches of fibrils ranging from 10-20 nm in diameter (Greving et al., 2012) or 30 nm (Xu et al., 2014). The formation of silk fibroin complex including three subunits in the single fibre in a cocoon.

The secondary structures of silks can be classified into three types including Silk I, Silk II, and silk III. The main crystal structures of fibroin are Silk I (liquid), Silk II (solid) and Silk III (solid) a small,

unstable part. Silk I is a metastable, soluble in water and non-crystalline (random coil and α helical structure) before being spun from the gland in a conformational transition process of Silk I to Silk II (β-sheets structures). Silk III has a trifold helical chain conformation found at the air-water interface. **Figure 5** shows that the structure of silk fibroin, including Silk II formed after being spun, has a regular array of anti-parallel β-sheets (Asakura & Suzuki, 2014; Qi et al., 2017, Nguyen et al., 2019).

The primary structure of fibroin contains seventeen amino acids; 43.75% Glycine, 29.05% Alanine, 11.17% Serine, 1.85% Valine, 0.42% Leucine, 0.53% Isoleucine, 0.90% Threonine, 1.51% Aspartic Acid, 1.07% Glutamic acid, 0.50% Phenylalanine, 5.42% Tyrosine, 0.36% Histidine, 1.90% Arginine, 0.50% Proline, Tryptophan 0.39%, 0.60% Lysine and 0.08% Cysteine (Heslot, 1998; Babu et al., 2013; Zafar et al., 2015). These three residues will form the highly repetitive GAGAGAS motifs which self-assemble into an anti-parallel β-sheet structure, within the crystalline region of the silk fibroin polymer (Marsh et al., 1955). Silk fibroin has three crystal structures: Silk I, Silk II and Silk III. In 1955, Marsh and his colleagues first reported on the Silk II crystal structure as an anti-parallel, bonded β-sheet. This structure was later revealed to have a back-to-back-β-sheet which interestingly consists of adjacent alanyl and glycyl components. This polypeptide arrangement explains why fibroin is insoluble in different solvents.

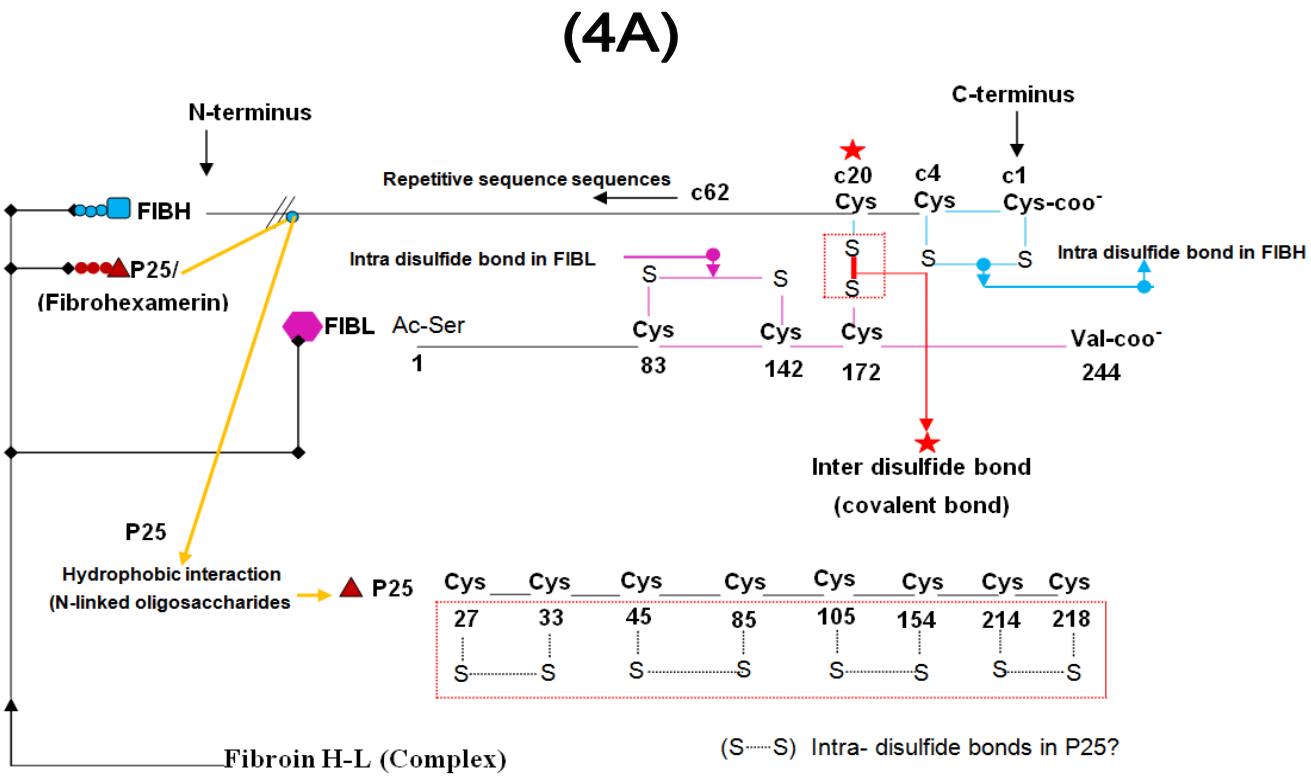


Figure 4A. The inter & intra molecular disulfide bonds of the FIBH & FIBL in fibroin H–L complex with the prediction of intra-disulfide bonds joined together in this complex as the ratio 6:6:1 in the ER in the PSG cells.

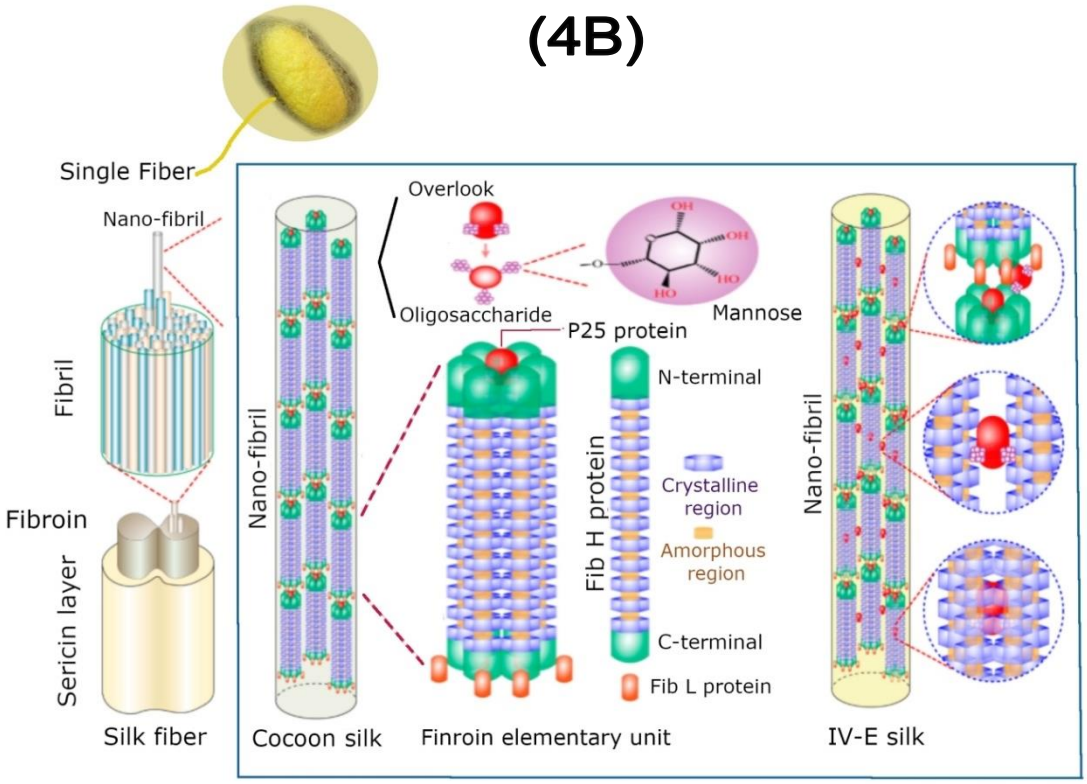


Figure 4B. The fibroin formation by three molecules in the single fibre of a cocoon (adapted from Tanaka et al., 1999; Peng et al., 2019).

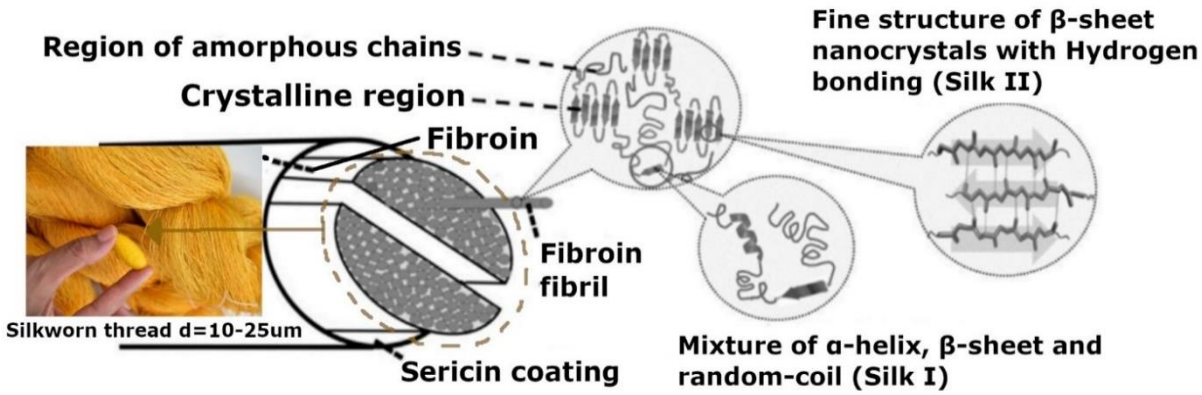


Figure 5. Schematic presentation of silk fibroin structure from a single fibre of a single cocoon. The structures show typical formations of Silk fibroin; Silk I, Silk II and the amorphous and crystalline regions (Adapted from Qi et al., 2017).

P25, an Important Glycoprotein for Silk Fibroin Complex

P25 is a special glycoprotein of fibroin subunit containing 220 amino acids encoded by a gene of five exons which is a different gene from the other two fibroin subunits. As shown in two **Figures 6A-6B**, although P25 is encoded from another gene, there are large blocks of sequence homology part in the 5' flanking regions that enables coordinate transcription with the other fibroin genes. The DNA upstream of P25 gene contains a sequence very similar of a region of fibroin gene 5' flanking DNA which is related to the transcription signal. A large portion of the sequence is dyad symmetric which can potentially form a stem and loop structure (Couble et al., 1985; Chevillard et al., 1986). The P25 protein has four intra-disulfide bonds containing Asn-linked oligosaccharide chains. Recent studies have reported that P25 contains two groups of isoforms, and each group is shown to have two slightly different sizes, 25kDa and 30kDa. According to P25 mass, it occurs in different degrees of glycosylation or glycosylated isoforms. P25 glycoprotein is necessary in maintaining the fibroin complex and it has been thought to have an important chaperon-like role by preventing H-chain degradation (Tanaka et al., 1999; Inoue et al., 2000; Inoue et al.,2004).

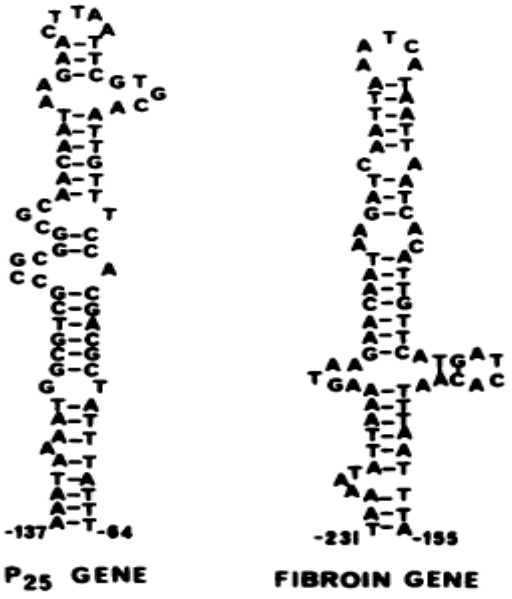


Figure 6A. A comparison of the putative DNA folding of the 5' flanking region between the P25 gene and fibroin gene. Notably, the stem and loop have a homologous sequence as in **Figure 6B**. with 58 matched nucleotides from the upstream region of these two gene. The structures are predicted by the computer program (Couble et al., 1985).

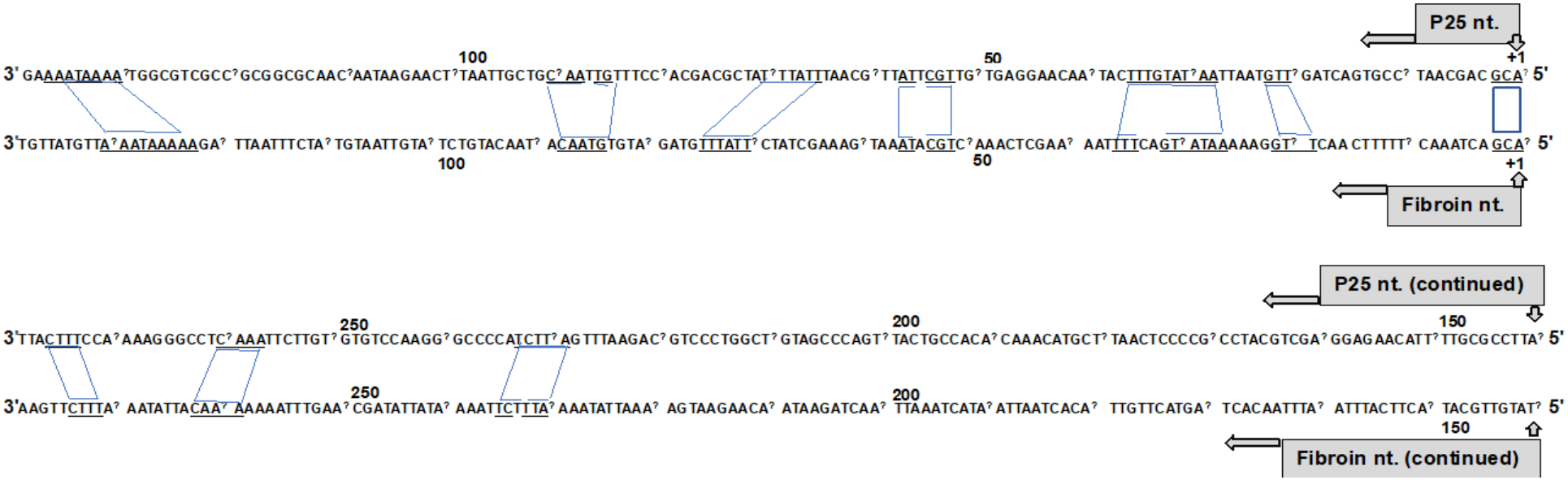


Figure 6B. A comparison of the 5' flanking sequences between the P_{25} and the fibroin genes among 280 nucleotides. The homologies are detected in similar position relative to the cap sites of the genes, which are shown by the blue highlighted lines, mismatches or deletion/insertion are indicated by gaps in the lines (modified from Couble et al., 1985).

Silk Fibroin Applications

Silk, known as the queen of textiles, has been used for royal clothing in China for over 4,000 years. To date, fibroin remains an important material in the high-quality textile industry in many countries because of its special properties, strength, shine, low conductivity, and it is processable from fibres to the solution. Fibroin has become a widely used and also well-studied material. Current applications include bullet-proof vests for soldiers, bowstrings, rag papers, fishing cord, string for making instruments, house decoration, parachute cloth, bicycle tires, bed filling and blankets. More importantly, fibroin has found significant use as a biomaterial in numerous biomedical applications. Fibroin is now extensively used in drug delivery, cell culture, tissue engineering and bone grafting. In fact, the use of fibroin in the medical field is not new, as it was used for surgical purposes in ancient times because of its excellent ability to treat wounds (Moy et al., 1991). However, any application of silk depends on its impressive properties, in particular its biocompatibility with other materials (Vepari & Kaplan, 2007).

Silk Fibroin via Genetic Engineering

Genetic engineering based on various expression systems has been extensively utilised and moved toward improved recombinant fibrous proteins including silk protein (Fahnestock & Irwin, 1997). Much published research has reported that recombinant silk proteins have great potential in various applications as they can be modified into various forms, hydrogels, films, beads and other natural fibres. The same applies to 'normal' silk. A challenge in molecular biotechnology is the discovery of different special silk production techniques through cloning and expressing the repetitive domains.

The molecular approaches will allow us to obtain larger polypeptide structures. Recombinant silks are unlike the traditional *B. mori*. Mostly, recombinant silks are derived from bacterial sources. The most important for recombinant silk protein are those with biocompatible and biodegradable properties, with possible clinical application (Werkmeister & Ramshaw, 2012). Regarding the previous published work, silk protein has mostly been investigated in the *E. coli* as the common expression system of choice. The involved recombinant silk proteins or silk fibroin are shown in **Table 1**.

Table 1. Recombinant chimeric silk and fibrous proteins (Adapted from Werkmeister & Ramshaw, 2012)

Protein combination	Expression System	Authors
Spider silk and elastin	Plant (Tobacco & Potato)	Scheller et al., 2004
Silkworm silk and elastin	*E. coli*	Yao & Asakura, 2003
		Haider et al., 2008
		Teng et al.,2009
		Qiu et al., 2010

Silkworm silk and calcium binding sequence	*E. coli*	Yang et al., 2008a
Silkworm silk, elastin and fibronectin	*E. coli*	Yang et al., 2008b

Over several decades, there have been reports about the three subunit genes of *B. mori* fibroin, H-chain, FIBL and P25, being cloned at the same time and their expression into *E. coli*. Unfortunately, the gene is not consistent due to the highly repetitive gene structure or other possible factors between gene and host (Wingkler & Kaplan, 2001). However, most previous studies are interested in only individual genes and their expression into the bacteria, in particular Fibroin, H-chain. Therefore, this project has the big challenge to investigate all three genes expressed into the selected bacteria and to see how their proteins form into the structure and how they function. This may reveal the mechanism of the silk fibroin complex or biosynthetic pathway and the potential to develop novel biomaterials for other uses.

Prokaryotic Expression Systems

The bacterial expression system - *Escherichia coli* - has been well established as the most commonly used system to express several recombinant proteins. The bacterial expression system can be used for the small motif, especially a protein with an inherent ability to re-fold (Werkmeister & Ramshaw, 2012). *E. coli* is well-known and is an established bio-factory for gene expression. Furthermore *E. coli* is also a micro-organism which offers numerous advantages for use: (i) it has unparalleled fast growth kinetics, (ii) the cells are easily grown, (iii) media sources can be prepared with reasonable cost and (iv) the target genes are easy to transform to the host cells (Rosano & Cecaelli, 2014). *E.coli* is one of the systems successfully used to produce many fibrous proteins. The best known of which is spider silk which has been produced in *E.coli* using genetic engineering. However, this host system still evidences low solubility of the spidroin (Heim et al., 2009).

There are many reasons why *E. coli* is a good choice for the recombinant protein production. Throughout the project, molecular tools and protocols are specifically developed for use in each recombinant protein such as screening of expression plasmids, determining potential vectors or host strains and specific cultivation strategies (Rosano & Ceccarelli, 2014).

In recent years, several basic *E. coli* strains have been used as hosts to overcome the limitations of recombinant protein expression and production. Basically, an *E. coli* expression vector contains the same features as found in any general host, namely, a selection marker (e.g. antibiotic resistance), origin of replication, transcriptional promoter, 5’ untranslated region (5’UTR), and translation imitation site (Jea & Jeon, 2016). Different *E. coli* strains will give different yields and have different advantages and disadvantages depending on the specific biosynthesis situation. Additionally, the most widely applied expression promoter system in a plasmid and other vectors is the phage T7 RNA polymerase which can be recognized only in the promoter on T7 DNA polymerase, but not in promoters on the *E. coli* chromosome (Studier, 1991). For this reason, the T7 expression host, BL21

and its derivatives are the most widely used for heterologous protein production (Rosano & Ceccarelli, 2014).

However, obtaining high yields for recombinant proteins in *E. coli* is challenging and unpredictable, especially when the protein of interest requires post-translational modifications for its bioprocessing, including: disulfide bond forming, phosphorylation, hydroxylation, glycosylation and proteolysis (Schein, 1989; Lobstein et al., 2012). Therefore, a suitable selection of host strains is necessary for a high soluble yield. Mainly, the recombinant proteins are expressed in the *E. coli* cytoplasm, but disulfide bond formation occurs in periplasm space. To deal with that issue, a novel strain called 'SHuffle' has been engineered on *trx*B *gor* (cytoplasmic pathway), and in combination with an engineered signal sequenceless bond isomerase, DsbC. in its chromosome to allow the formation of correctly folded multi-disulfide bonds (Lobstein et al., 2012).

Production, Solubilisation and Purification of Recombinant Proteins in Bacteria

Molecular tools used for the expression of recombinant proteins include bacteria vectors and hosts. It is necessary to have appropriate cDNA and PCR reactions that make the proper plasmid reaction. After ligation of the sequence of interest into the plasmid, this successful ligated plasmid was transformed into the target host on an agar plate. Thus, the selection of the healthiest single colony for further protein production is necessary. Previous work has reported that fermentation of 2 litres of *E. coli* in the complex media will result in about 50-80 g of the wet weight of the cell. The expressed recombinant protein will be approximately 2-5% of the cell. Therefore, there will be about 100-300 mg of the expressed protein in the cells. This can produce soluble proteins with a good yield (>50%) of the total cellular protein whereas insoluble proteins must go straight to the denaturation and refolding steps. The insoluble protein known as aggregation proteins or inactive protein mainly occurs from protein misfolding. This might be the result of many factors such as post-translational modification or crowded and limited environment outside-inside bacterial cells. In addition, it might be caused by the disulfide bonds' formation in the recombinant proteins (Bednarska et al., 2013). However, the way to produce soluble protein from insoluble protein has many challenging strategies. It can also occasionally produce insoluble proteins which require more processing to improve the yield to (5% to 20% of the total cellular protein). Additionally, the small-scale culture can make sufficient protein 10-100 mg. In fact, the recombinant protein forming insoluble or inclusion bodies will require unfolding and refolding back into the native structure.

There are many combination strategies in dealing with the cells to improve the protein purity, such as: solubilisation of inclusion bodies using chaotropic agents (guanidine-HCl or urea), affinity tag and fusion proteins, solubility and location of the protein, determining the isoelectric point, breaking cells, clarifying cell extraction by centrifugation or selective preparation, preparation for the repeat ion-exchange step, performing gel filtration and other methods (Wingfiled, 2016).

Protein N-glycosylation

Protein glycosylation is involved in many biological processes such as function, stability, receptor interaction, immune response, protein secretion and transport which are thought to be shared in all domains of life. Approximately, 70% of eukaryotic proteome is considered to be glycosylated. Each protein can have many sites of glycosylation (Niam et al., 1999; Dell et al., 2010, Cleck et al., 2016). Protein glycosylation is one of the most essential and important post-translational modifications which occurrs in the lumen of ER. There are two types of glycosylation: *N*- and *O*- glycosylation which contain different links of oligosaccharide to the amino acid side chain of the protein (Schoberer et al., 2018). The biosynthetic process in N-glycosylation is more complex than O-glycosylation (Cole & Smith, 1990). The three glycans attached to asparagine (N-linked) and the four most common core structures bound to serine/threoinin (O-linked), are shown in **Figure 7**. However, N-linked oligosaccharides have been characterised as being the majority of the glycoproteins (Hayness, 1998).

N-glycosylation has a complex biosynthetic process that plays a role in the maturation of the protein through the secretory pathway (Galban et al., 2010). The most common type of N-glycosylation is N-acetylglucosamine likened to asparagine (GlcNAcB1-Asn) (Laeza-Reyes et al., 2021). According to the N-glycosylation reported in P25 (a fibroin subunit) it contains three high mannose types of glycan chains associated to the H-chain via noncovalent bond interaction (Inoue et al., 2000).

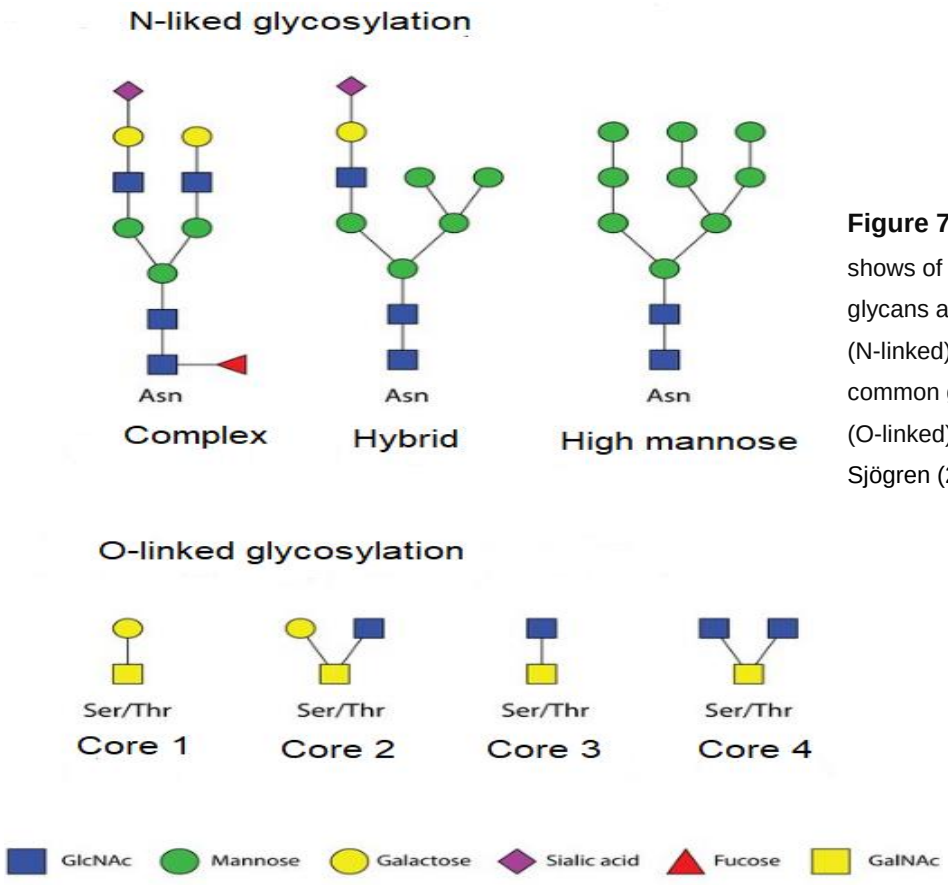


Figure 7. The schematic diagram shows of the most three common glycans attached to asparagine (N-linked) and the most four common glycan to serine/threoinin (O-linked), adapted from Sjögren (2015).

In addition, The N-linked glycans are attached almost exclusively to asparagine residues in a defined N-X-T/S sequon (N:asparagine, X is any amino acid except proline), T/S is threonine or serine. In some uncommon cases, N-glycosylation was also found on non-canonical sequons containing N-X-C (C is glycine, V is valine, G is glycine). The N-Glycosylation process can occur in the folding dynamic with good interaction in the ER. Hence, the N-glycans also tend to depend on the enzymes involved in each step. For Instance, it has been reported that alpha1,3-fucosyltransferase A (FucTA) or the gene related to sialylation and its expression is restricted to specific tissue and development stages such as in silkworms *(B. mori).* The general phenomenon of synthesis of glycosylation is similar among eukaryotic organisms. However, the glycosylation in insects is different from other vertebrates. The example types of N-glycans are found in insects as shown in **Figure 8.** (Walski et al., 2017).

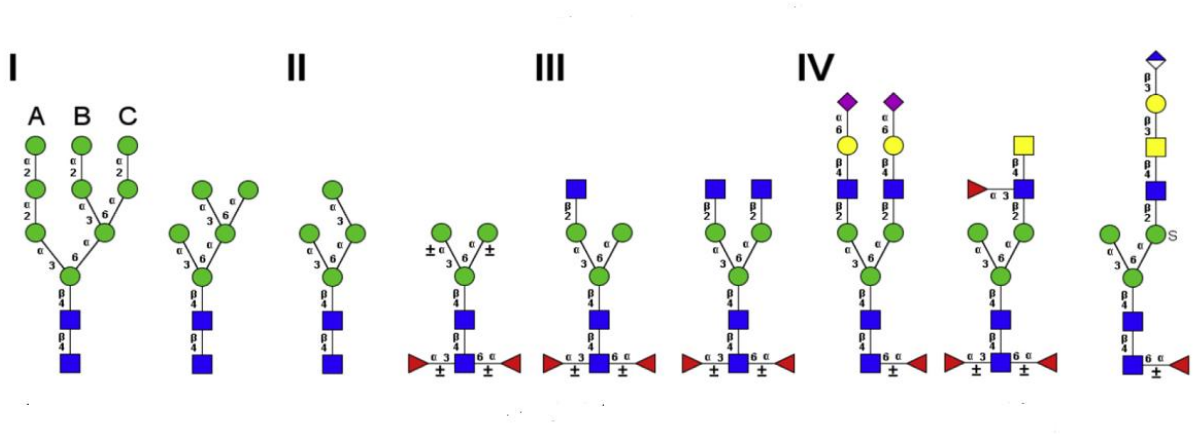


Figure 8. The four different types of N-glycan found in insects; (I) High mannose glycans with nine to five mannose residues, A-C indicate the three antennas of a high mannose N-glycan, (II) Paucimannose glycans containing less than five mannoses and possible core fucoses but no terminal GlcNAc, (III) Common insect complex glycans with terminal GlcNAc residues, (IV) Rare insect complex glycans with various terminal modifications (Walski et al., 2017).

Protein N-glycosylation in the Baculovirus-insect Cell expression System to Recombinant Glycoprotein: An Alternative Way for Recombinant N-linked P25 Glycoprotein

Insect cells are a potential host for baculovirus vector infections that are great for performing the transfecting of oligosaccharide side chains (glycans) to the same site in recombinant proteins as those produced in native protein N-glycosylation. N-glycans are produced in mammalian cells which are more complicated than insect cells (Harrison & Jarvis, 2006). However, the processing of N-linked oligosaccharides on the recombinant products in insect cells are not completely the same as the production in its original expression (Marz, et al., 1995).

The baculovirus expression vector system (BEVS) has the ability to produce the products of interest that require the post-translational modification during the biosynthesis process. BEVS provide many advantages such as high-level expression and a safe system of manufacture of the final products. Furthermore, the system can handle the production of more than one gene (Sokolenko et al., 2012). Additionally, the most common host cells rely on four lepidopteran insect cell lines such as 1) *Spodoptera frugiperda* (Fall armyworm): Sf9 and Sf21, 2) *Trichoplusia ni* (Cabbage looper): BTI-TN-$_5$B$_{1-4}$, marketed as High Five^{TM}, 3) *Drosophila melanogaster* (Fruit fly): S2 and 4) *Bombyx mori* (Japanese silkworm): BM-N and Bm17. Basically, *S. frugiperda* and *T. ni* cell lines are used with the (BEVS) (Zitzmann et al., 2017; Altman et al., 2020).

Protein N-glycosylation Production in the Mammalian Cell Expression System to Recombinant Glycoprotein

Another platform for N-linked glycoprotein production, mammalian cells initially demonstrated promising outcomes with their potential of metabolic machinery to produce and secrete recombinant proteins. The resulting proteins resembled and are more likely to be compatible with humans. Mammalian cells are a desirable choice for a complex type of N-linked glycoprotein (Goh & Ng, 2018). Correct glycan structures are necessary for effective control of pharmaceutical properties in its function of biologics and therapeutic carbohydrates (Tejwani et al., 2018). Different cell types possibly affect the glycosylation products of recombinant proteins. Recently, there was a report of the SARS-CoV2 spike glycoprotein produced in several human cell lines (HEK293, HEK2) and baculovirus-insect system cells (BIC1, BIC2) that had significantly different N-glycosylation sites and type of N-glycans (Wang et al., 2021). Some types of human cell lines are used for the production of approved bio-therapeutics along with Chinese hamster ovary (CHO) cells, NS0 myeloma and Sp2/0 hybridoma mouse cell lines, baby hamster kidney (BHK), human embryonic kidney cells 293 (HEK293) and HT-1080 human cells (Goh & Ng, 2018).

Selective Tools for Characterisations of the Proteins

1) Circular dichroism (CD) spectroscopy

CD is an advanced technique to rapidly characterise the secondary structure, determine conformation and protein folding in proteins. CD spectra for wavelengths less than 240 nm are sensitive to the secondary structure of the polypeptide backbone (France et al., 1992). Data of the Far UV backbone (meaning the amide transition) will be collected from ~190-250 nm (Rodger, 2013). Additionally, the best predictions of secondary structure are obtained when CD spectra are detected to at least 178 nm (Johnson, 1990). CD can measure one or many samples which require 20µg or less in the physiological buffers (without any salts) and the measurement will take 2-3 hours (Greenfield, 2006). Five types of secondary structure can be estimated using CD spectra; α-helix, anti-parallel β-sheet, β-sheet parallel, β-turns (all types) and aperiodic structures (including random coils) (France et al., 1992).

2) Nano-scale liquid chromatographic tandem Mass Spectrometry (nLC-ESI MSMS)

nLC-ESI MSMS is one of the common techniques for identifying peptides from complex mixtures, such as protein digests. This tool is often the best choice for the determination of a very limited number of samples, because it has powerful analysis (Larsen et al., 2021). When the ionisation is connected by electrospray ionisation (ESI) for MS it can improve the sensitivity to the protein coverage. This technique integrates two methods together, mainly the liquid chromatograph to the mass spectrometer (Yang et al., 2007). In the nLC-ESI MSMS the chromatographic separation is combined directly to the MS analysis. Peptides are detected, selected, and fragmented in real time. The column is eluted by spraying into the machine. MS and MSMS scan times are pre-set at the beginning of the run as a compromise between signal intensity and the requirement to analyse as many peptides as possible in a chromatographic window (Shirran & Botting, 2010).

The output of the data file from the instrument will be compared against the NCBI database, using the Mascot search algorithm following the guide from the BSRC mass spectrometry facility, University of St. Andrew, UK.

3) Size Exclusion Chromatography (SEC)/ multi-angle light scattering (MALS)

SEC/MALS is a standard and common tool for characterization of peptides MW, shape, aggregation, oligomerization, interaction, and purity. SEC and field flow fractionation (FFF) is the protein separation tool combined with the MALS for macromolecule characterisation which separates the sample by its size. Light scattering (LS) is commonly used in characterisation of different molecules and particles. When multi-angle light scattering (MALS) is applied for molar mass investigation of macromolecules, including proteins related to Rayleigh theory. The tool has dynamic light scattering (DLS) that can be applied for measuring the hydrodynamic radius of the protein. MALS can analyse the mixed samples of several types of macromolecules containing polymers. In addition, MALS can analyse pure samples due to macro-filtration before loading into the column (Amartely et al., 2018).

Research Aims

1. To investigate the FIBL protein expression, optimisation, purification, and production in bacterial systems
2. To investigate protein expression P25 glycoprotein production, optimisation, purification in bacterial systems and baculovirus-infected insect cell systems
3. To determine the stability, behavior and identify of two lower molecular weight recombinant fibroin subunits, FIBL protein and P25 glycoprotein
4. To determine the recombinant protein secondary structure of FIBL and P25 using Circular-Dichroism technique
5. To investigate the N-glycans attachment in P25 glycoprotein
6. To investigate the complex formation/protein interaction of three molecules fibroin (H-chain from native silk), two recombinant light fibroins (FIBL & P25), and to study the formation of H-chain with human proteins for testing functionality and biomedical engineering applications.

Research Hypotheses

1**)** I hypothesise that bacterial expression systems be used to produce three of the core silk proteins (P25, FIBL and H-chain).

2. I hypothesise that baculovirus-insect cell expression systems be used to produce three of the core silk proteins (P25, FIBL and H-chain).

3. The H-chain interacts independently from the molecules of silk proteins and other protein as its superior properties such as high self-assemble and adhesive formation etc.

BIBLIOGRAPHY,

Altman, R., Chiou, H., Esposito, D. 2020. 'Assembling an effective toolbox of expression systems to support your drug discovery efforts'. *In* Part of the 19th Annual PEPTALK the Protein Science Week, January 21, 2020, San Diego, CA, USA.

Amartely, H., Avraham, O., Friedler, A., Livnah, O. & Lebendiker, M. 2018. 'Coupling multi angle light scattering to Ion Exchange Chromatography (IEX-MALS) for protein characterisation'. *Scientific Reports*. 6907:1-9.

Asakura,T. & Suzuki, T. 2014. 'Silk Fibroin'. *Encyclopedia of Polymeric Nanomaterials*. Springer-Verlag Berlin Heidelberg.

Asakura, T., Okushita, K. & Williamson, M.P. 2015. 'Analysis of the structure of *Bombyx mori* silk fibroin by NMR'. *Macromolecules*. 48:2345-2357.

Babu, K.M. 2013. '3-Structural aspects of silk. In Silk processing and applications'. Woolhead Publishing Series in Textiles. pp56-83.

Bednarska, N.G., Schymkowitz, J., Rousseau, F. & Van Eldere, J. 2013. 'Protein aggregation in bacteria: the thin boundary between functionality and toxicity'. *Microbiology*. 159:1795-1806.

Bourzac, K. 2015. 'Spiders: web of intrigue'. *Nature*. 519:S4-S6.

Chevillard, M., Coubl, P. & Prudhomme, J.C. 1986. 'Complete nucleotide sequence of the gene encoding the *Bombyx mori* silk protein P25 and predicted amino acid sequence of the protein'. *Nucleic acids Research*. 14(15): 6341-6342.

Couble, P., Chevillard, M., Moine, A., Ravel-Chapuis, P. & Prudhomme, J.C. 1985. 'Structural organization of the P25 gene of *Bombyx mori* and comparative analysis of its flanking DNA with that of fibroin gene'. *Nucleic Acids Research*. 13: 1801-1814.

Clerk, F., Reiding, K.R., Jansen, B.C., Kammeijer, G.S.M., Bondt, A. & Wuhrer, M. 2015. 'Human plasma protein N-glycosylation'. *Glycoconjugate Journal*. 33:309-343.

Damrongsakkul, S. 2020. 'Silk Fibroin: Biomaterial for biomedical engineering'. Chulalongkorn University Publisher. Bangkok. 256 Pages. (In Thai).

Dell, A., Galadari, A., Sastre, F. & Hitchen, P. 2010. 'Similarities and different in the glycosylation mechanisms in prokaryotes and eukaryotes'. *International Journal of Microbiology*. Article ID148178. 14 pages.

Fahnestock, S, R., Irwin, S.L., 1997. 'Synthetic spider silk protein and their production in *Escherichai coli*'. *Applied Microbiology Biotechnology*. 47:23:32.

Eisoldt, L., Smith, A. & Scheibel, T. 2011. 'Decoding the secrets of the spider silk'. Materialstoday. 14(3):80-86.

Feughelman, M. 2002. 'Natural protein fibers'. *Journal of Applied Polymer Science*, 83: 489-507.

France, L.L., Kieleczawa, J., Dunn, J.J., Hind, G. & Sutherland, J.C. 1992. 'Structural analysis of an outer surface protein from the Lyme disease spirochete, borrelia burgdorferi, using dichroism and fluorescence spectroscopy'. *Biochimica et Biophysica Acta*. 1120(1):59-68.

Gaban, C., Galban, S., Van Dort, M., Luker, G.D., Bhojani, M.S., Rehemhualla, A. & Ross, B.D. 2010. 'Application of Molecular image'. *Progress in Molecular Biology and Translational Science*. 95:237-298.

Gage, L.P. & Manning, R.F. 1980. 'International structure of the silk fibroin gene of *Bombyx mori*.; I. The fibroin gene consists of a homogeneous alternating array of repetitious crystalline and amorphous coding sequence'. *The Journal of Biological Chemistry*. 255 (19): 9444-9450.

Goh, J.B. & Ng, S.K. 2018. 'Impact of host cell line choice on glycan profile'. *Critical Revies in Biotechnology*. 38(6):851-867.

Greenfield, N.J. 2006. 'Using circular dichroism spectra to estimate protein secondary'. *Nature Protocols*. 1(6): 2876-2890.

Gupta, A.K., Mita K., Arunkumar, K.P. & Nagaraju, J. 'Molecular architecture of silk fibroin of Indian golden silkmoth, *Antheraea assama'*. *Scientific Report* (Online). 5:12706.

Greving, I., Cai, M., Vollrath, F. & Schniepp, H.C. 2012. 'Silk proteins into fibrils studied by atomic force microscopy'. *Biomacromolecules*. 13: 676-682.

Gopnathan, K.P. 1992. 'Biotechnology in sericulture'. *Current Science*. 62(3):283-287.

Haider, M., Cappello, J., Ghandehari, H. & Leong, K. W. 2008.' *In vitro* chondrogenesis of mesenchymal stem cells in recombinant silk-elastinlike hydrogels'. *Pharmaceutical Research*. 25:692–9.

Hakimi, O., Knight, D.P., Vollrath, F. & Vadgama, P. 2007. 'Spider and mulberry silkworm as compatible materials'. *I Composites: Part B: Engineering*. 38: 324-337.

Hayness, P. 1998. 'Phosphoglycosylation: a new structural class of glycosylation'. *Glycobiology.* 8(1):1-5.

Hayashi, C.N., Blackledge, T.A. Lewis, R.V. 2004. 'Molecular and mechanical characterization of Aciniform silk: Uniformity of iterated sequence modules in a novel membrane of the spider silk fibroin gene family'. *Molecular Biology and Evolution*. 21(10):1950-1959.

Harrison, R.L & Jarvis, D.L. 2006. 'Protein N-glycosylation in the baculovirus insect cell expression system and engineering of insect cells to produce 'Mamalianized' recombinant glycoprotein'. *Advances in Virus Research*. 68:159-191.

Heslot, H. 1998. 'Artificial fibrous protein'. *Biochimie*. 80:19-31.

Hsia, Y., Gnesa, E., Jeffery, F., Tang, S. & Vierra, C. 2011.'Spider silk composites and applications'. *In* Metal, ceramic, and polymeric composites for various uses. *IntechOpen (Ebook*). 700pp.

Hyde, N. 1984. The queen of textiles. *National Geographic*. 165:3-49.

Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K. & Mizuno, S. 2000. 'Silk Fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio'. *The Journal of Biological Chemistry*. 275(51): 40517-40528.

Jia, B. & Jeon, C.O. 2016. 'High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspective'. *Open Biology.* (6): 1-17.

Julien, E., Coulon-Bablex, M., Garel, A., Royel, C., Chavanya, G., Prudhomme, J.C. & Couble, P. 2005. '2.11-Silk gland development and regulation of silk protein genes'. *In* Comprehensive Molecular Insect Science. 2: 369-384.

Kundu, B., Rajkhowa, R., Kundu, S.C. & Wang, X. 2013. 'Silk fibroin biomaterials for tissue regenerations'. *Advanced Drug Delivery Reviews*. 65(4):457-470.

Larsen, H.R., Wilson, S.R.& Lundanes, E. 2021. 'Recent advances in on-line upfront devices for sensitive bioanalytical nanoLC methods'. *TrAC Trends in Analytical Chemistry*. 136: (1-10).

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P. & Berkmen, M. 2016. 'SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded protein in its cytoplasm.' *Microbial Cell Factories*. 11(56):1-16.

Nguyen, T.P., Nguyen, Q.V., Nguyen, V.H., Le, T.H., Huynh, V.Q.N.,Vo, D.V., Trinh, Q.T. & Kim, S.Y. 2019. 'Silk fibroin-base biomaterials for biomedical applications: A review'. *Polymers*. 11:1933 (25pp).

Niam, H., Joberty, G., Alfalah, M. & Jacob, R. 1999. 'Temporal association of the N- and O- linked glycosylation event and their implication in the polarized sorting of the intestinal brush border Sucrase-Isomaltase, Aminopeptidase N, and Dipeptidyl Peptidase IV'. *The Journal of Biological Chemistry*. 27(25):17961-17967.

Marz, L., Altmann, F., Staudacher, E. & Kubellka, V. 1999.'Protein glycosylation in insect. Glycoproteins'. *In* Glycoproteins. Elsevier, Amsterdam, 543-563 pp.

Marsh, R.E., Corey,R.B.& Pauling, L. 1955. 'An investigation of the structure of silk fibroin'. *Biochimica et Biophysica Acta*. 16(1): 1-34.

Moy, R.L., Lee, A. & Zalka, A. 1991. 'Commonly used suture materials in skin surgery'. *American Family Physician*. 44(6):2123-2128.

Murphy, A.R.& Kaplan, D.L. 2009. 'Biomedical applications of chemically-modified silk fibroin'. *Journal of Materials Chemistry*. 19(36):6443-6450.

Nilebäck., L., Chouhan, D., Jansson, R., Widhe, M., Mandal, B.B. & Hedhammar, M. 2017. 'Silk-Silk Interactions between silkworm fibroin and recombinant spider silk fusion protein enable the construction of bioactive materials'. *ACS Applied Materials & Interface*. (9): 31634-31644.

Peigler, R.S. 1993. 'Wild silks of the world'. *American Entomologist*. 39(3): 151-162.

Peng, Z., Yang, X., Liu, C., Dong, Z., Wang, F., Wang,X., Hu, W., Zhang, X., Zhao, P & Xia, Q. 2019. 'Structural and mechanical properties of silk from different instars of *Bombyx mori*'. *Biomacromolecules*. 20:1203-1216.

Qi, Y., Wang, H., Wei, K., Yang, Y., Zheng, R.Y., Kim, I.S. & Zheng, K.Q. 2017. 'A review of structure construction of silk fibroin biomaterials from single structures to multi-level structure'. *International Journal of Molecular Science*. 18(3):237.

Qiu, W., Huang, Y., Teng, W., Cohn, C. M., Cappello, J.& Wu, X. 2010. 'Complete recombinant silk-elastinlike protein-based tissue scaffold'. *Biomacromolecules*. 11: 3219–27.

Rajkhowa, R., Kaur, J., Wang, X. & Batchelor, W. 2015. 'Intrinsic tensile properties of cocoon silk fiber can be estimated by removing flaws through reported ensile tests'. *Journal of The Royal Society Interface*. 12:20150177, (1-10).

Rodger, A. 2013. 'Far UV protein Circular Dichroism'. *In* 'Encyclopaedia of Biophysics. Springer, Berlin, Heidelberg.

Romer, L. & Scheibel, T. 2008. 'The elaborate of spider silk'. *Prion*. 2(4):154-161.

Rosano, G.L. & Cecarelli, E.A. 2014. 'Recombinant protein expression in *Escherichia coli* advances and challenges'. *Frontiers in Microbiology*. 5(172):1-17. (Online Journal).

Sari, D., Gupta, K., Balaji, D., Raj, D.B.T.G., Aubert, A., Drncova, P., Garzoni, F., Fitzgerald, D. & Berger, I. 2016. 'The MultiBac baculovirus/ Insect cell expression vector system for producing complex protein Biologics'. *In* Technologies for protein complex production and characterization'. *Spring International Publishing*, Switzerland. pp 199-125.

Scheibel, T. 2004. 'Spider silks: recombinant synthesis, assembly, spinning and engineering of synthetic proteins'. *Microbial Cell Factories*. 3(14):1-10 (10pp).

Scheller, J., Henggeler, D., Viviani, A. & Conrad, U. 2004. 'Purification of spider silk-elastin from transgenic plants and application for human chondrocyte proliferation'. *Transgenic Research*. 13:51–7.

Schein, C.H. 1989. 'Production f soluble recombinant proteins in bacteria'. *Nature*. (7):1141-1149.

Schober, J., Shin, Y.J., Javra, U., Veit, C. & Straser, R. 2018.'Protein glycosylation in the ER'. *Method in Molecular Biology*. 1691:205-222.

Sjögren, J. 2015. ‘Bacterial modification of host glycosylation-in injection biotechnology and therapy’. Doctoral Dissertation, Faculty of Medicine, Lund University, USA. (Unpublished).

Shirran, S.L. & Botting. C.H. 2010. ‘A comparison of the accuracy of iTRAQ quantification by nLC-ESI MSMS and nCL-MALDI MSMS method’. *Journal of Proteomics*. 73:1391-1403.

Studier, F.W. 1991. ‘Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system’. *Journal of Molecular Biology*. 219(5):37-44.

Sokolenko, S., George, S., Wang, A., Tuladhar, A., Andrich, J.M.S. & Aucoin, M.G. 2012. ‘Co-expression vs. co-infection using baculovirus expression vector in insect cell culture; Benefits and drawbacks’. *Biotechnology advance*. 30:766-781.

Sutherland, T.D., Yang, J.H., Weisman, S., Hayashi, C.Y.& Merritt, D.J. 2010. ‘Insect silk: One name, many materials’. *Annual Review of Entomology*. 55: 171-188.

Takei, F., Kikuchi. Y., Kikuchi, A., Mizuno, S., Shimura, K. 1987. ‘Further evidence for importance of the subunit combination of silk fibroin in its efficient secretion from the posterior silk gland cells’. *Journal of Cell Biology*. 105:175–180.

Tanaka, K., Kajiyama, N., Ishikura, K., Waga ,S., Kikuchi, A., Ohomo, K., Takagi, T. & Mizuno, S. 1999. ‘Determination of the site of disulfide linkage between heavy chain and light chain of the silk produced by *Bombyx mori’. Biochemica et Biophysica Acta*. 1432: 92-103.

Teng, W, Cappello, J & Wu, X. 2009. ‘ Recombinant silk-elastinlike protein polymer displays elasticity comparable to elastin’. *Biomacromolecules*. 10: 3028–36.

Tejwan, V., Andersen, M.R., Nam, J.H. & Sharfstein, S.T. 2018.’Glycoengineering in CHO cells: Advance in Systems Biology’. *Biotechnology Journal*. 13(1700234):1-16.

Vepari,C. & Kaplan, D.L. 2007. ‘Silk as a biomaterial’. *Progress in Polymer Science*. 32:991-1007.

Volkov,V. & Cavaco-Paulo, A. 2016. '*In vitro* phosphorylation as tool for modification of silk and keratin Fibrous materials.' *Applied microbiology and Biotechnology*. 100:4337-4345.

Walski, T., De Schutter, K., Van Damme, E.J.M. & Smagghe, G. 2017. 'Diversity and function of protein glycosylation in insect'. *Insect Biochemistry and Molecular Biology*. 83:21-34.

Wang, Y., Wu, Z.,Hu,W.H., Hao, P.L. & Yang, S. 2021. 'Impact of expression cells glycosylation and glycan of the SARS-CoV-2 spike glycoprotein'. *ACS Omega*.6:15988-15999.

Werkmeister, J. & Ramshaw, J.A.M. 2012. 'Recombinant protein scaffolds for tissue engineering'. *Biomedical Biomaterials*. Online journal.7:012002 (29pp).

Wingfield, P.T. 2016. 'Overview of the purification of recombinant protein'. HHS Public Access. *Current Protocols in Protein Science*. Online Journals. 1-50.

Wingkler, S., Kaplan, D.L. 2001. 'Silk produced by engineered bacteria'. *In* Encyclopedia of Materials: Science and Technology. (Second Edition). Pergamon: United Kingdom.

Walski, T., Schutter, K.D., Damme, E.J.M. & Smagghe, G. 2017. 'Diversity and functions of protein glycosylation in insect'. *Insect Biochemistry and Molecular Biology*. 83:21-34.

Xu, G., Gong, L., Yang, Z. & Liu, X.Y. 2014. 'What makes spider silk fibers so strong from molecular –crystallite to hierarchical network structures'. *Soft Matter.* 10: 2116-2123.

Xu, P.Z., Zhang, M.R., Wang, X.Y. & Wu, Y.C., 2021.' Precocious metamorphosis of silkworm larvae infected by *Bm*NPV in the latter half of the fifth instar'. *Frontiers in Physiology*. 12(659972):1-11.

Yang, M., Muto, T., Knight, D., Collins, A. M. & Asakura, T. 2008a. 'Synthesis and characterization of silklike materials containing the calcium-binding sequence from calbindin D9k or the shell nacreous matrix protein MSI60'. *Biomacromolecules*. 9: 416–20.

Yang, M., Tanaka, C., Yamauchi, K., Ohgo, K., Kurokawa, M. & Asakura, T. 2008b. 'Silklike materials constructed from sequences of *Bombyx mori* silk fibroin, fibronectin, and elastin'. *Journal of Biomedical Materials Research Part A*. 84: 353–63.

Yang, Y., Zhang,S., Howe, K., Wilson, D.B., Morser, F., Irwin, D. & Thannhauser, T.W. 2007. 'A comparison of Nlc-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC based protein identification and iTRAQ-based shotgun qualitative proteomics'. *Journal of Biomolecular Techniques*. 18(4):226-237.

Yao, J. & Asakura, T. 2003. 'Synthesis and structural characterization of silk-like materials incorporated with an elastic motif '. *The Journal of Biochemistry*. 133: 147–54.

Zafar, M.S., Belton, D.J., Handby, B., Kaplan, D.L. & Perry, C.C. 2015. 'Functional material features of *Bombyx mori* silk light vs. heavy chain proteins'. *Biomacromolecules*. 16(2): 606-614.

Zitzmann, J., Sprick, G., Weidner, T., Schreiber, C. & Czermak, P. 2017. 'Process optimization for Access): recombinant protein expression in insect cells'. In New insights into cell culture technology. (Open Access): Intechopen/books/5359.

CHAPTER II

MATERIALS AND METHODS

……

Materials and methods are described in this chapter for all three experimental chapters, III, IV, V, including *E. coli* strains information. Thereafter, each experimental chapter is described in detail as follows:

Chapter III: "Observation on the Potential of Recombinant Fibroin Light Chain (*B. mori*) Production in Bacteria System". This chapter contains five main experiments.

1) Vector construction of FIBL gene in pET28a
2) Screening expression and optimisation of FIBL protein (small-scale production)
3) FIBL protein solubilisation, its localisation, and purification
4) Scaled-up expression, solubilisation, and purification for FIBL protein production in bacteria system
5) Characterisation of FIBL protein expression in bacteria

Chapter IV: "Comparative study of recombinant P25 glycoprotein (*B. mori*) in bacteria (Part I) & baculovirus-infected insect cell systems (Part II)". There are two main sections for P25 production. The experiments for P25 production in bacteria system (Part I) and & baculovirus-infected insect cell systems (Part II)" as follows:

Part I:
1) Vector construction for P25 gene (Molecular Engineering)
2) Screening expression and optimisation of P25 protein (small-scale production)
3) P25 Protein solubilisation, purification and its localization in pET28a
4) Scaled-up expression, solubilisation, and purification of P25 protein
5) Characterisations of P25 protein expressed in bacteria

Part II:
1) Vector construction of P25 gene for MultiBac system in pFBDM
2) Bacmid preparation (Molecular Engineering)
3) Generation of baculovirus stocks and P25 protein localisation
4) Scaled-up expression, solubilisation, and purification of P25 protein
5) Two forms of glycosylated P25 protein purifications
6) Characterisation of P25 proteins in baculovirus-infected insect cell systems

Chapter V: "*In vitro* silk fibroin complex, its indigenous interactions and preliminary interactions with human proteins". This chapter covers the experiments using the H-chain, FIBL, and P25 to investigate their interactions. The experiments were shown as follows:

1) Insect (silkworm) preparation, fresh posterior silk glands preparation for solid fibroin and H-chain isolation and purification

2) Investigation of *in vitro* silk fibroin assemble complexes, including four experiments below;

2.1) Observation of the initiation of silk fibroin complexes formation or multimerisation behaviour along its three indigenous proteins (H-chain, FIBL&P25) and interaction with selected human proteins (CLIC1 & CLIC4).

2.2) Determination *In vitro* silk fibroin complexes formation or protein-protein interactions using (Size Exclusion Chromatography and Laser Light Scattering Service, SEC-MALS)

2.3) Determination *in vitro* silk fibroin complexes formation or protein-protein interactions using (Size Exclusion Chromatography, SEC)

2.4) Determination *in vitro* silk fibroin complexes formation or protein-protein interactions using pull-down assay

E. coli strains information

The following genotype of *E. coli* strains were used at various points throughout this project.

DH5α competent cells: *F - Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(rk - , mk +) phoA supE44thi-1 gyrA96 relA1*

BL21(DE3) competent cells: *F– ompT hsdS$_B$* (r_B–, m_B–) *gal dcm* (DE3), Novagen SDS

C41 competent cells: *F – ompT hsdSB (rB- mB-) gal dcm* (DE3)

Rosetta^{TM} (DE3)pLysS competent cells; *F– ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE (Cam^R), Novagen

SHuffle® T7 competent cells; *F´ lac, pro, lacI^q / Δ(ara-leu)7697 araD139 fhuA2 lacZ:T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt::pNEB3-r1-cDsbC* ($Spec^R$, *lacI^q*) *ΔtrxB rpsL150*(Str^R) *Δgor Δ(malF)3* , New England Biolabs® Inc.

Materials and method for chapter III

"Observation on the potential of recombinant Fibroin Light Chain (*B. mori*) production in bacteria system"

The materials and methods involved mainly with FIBL for chapter III were carried out as follows;

1) Vector construction of FIBL gene (Molecular Engineering)

1.1. FIBL Gblocks with TEV+His6-tagged at C-terminus were designed **(Figure 9)** and artificially synthesised by Integrated DNA Technologies, Belgium.

1.2. FIBL Gblocks were amplified to add two restriction zymes using PCR reactions (Primers; the forward primer (with EcoRI added) is 'ATACGCCTA G AATTCATGAAGC' and the revers primer (with HindIII added) is 'ATCGTACGTCGA ATC GG AAG' to increase the concentration of its DNA (transcripts per volume).

1.3. An amplified FIBL gene was cloned (Restriction-Ligation technique) into the carrier vector (pET28a) and a host cell (DH5α™) transformed for making DNA and the gene was sent for verification of its sequence (DNA sequencing and service, University of Dundee, UK).

1.4. Verification sequence of the FIBL construction, the plasmid DNAs containing the insertion (FIBL) and pET28a were transformed into four different competent cells (*E. coli*), two groups of bacterial strains, Group 1 (un-enhanced disulfide bond formation), including three bacterial cells; BL21(DE3), C41 (DE3), Rosetta™ (DE3)pLysS (Novagen) and Group 2 (enhanced disulfide bond formation) containing one bacteria strain, SHuffle® T7. These cells were preserved at -80°C until use for inoculation and further studies for the investigation of protein expression.

2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL)

2.1) A single colony of four transformed FIBL competent cells with two groups of bacterial strains, BL21 (DE3), C41 (DE3), Rosetta™ (DE3)pLysS (Novagen and SHuffle® T7) were picked and tripled in size in a 50mL Falcon tube overnight using 10 mL LB media containing kanamycin.

2.2) The following day, 1mL of the overnight cells in each strain was put into four fresh 50mL Falcon tubes (prepared for the next four IPTG concentrations) and an extra 9 mL of fresh LB-broth was added into each tube with 9µL kanamycin 100µg/mL and grown in the shaker at 30ºC until O.D.600=0.6 for approximately 2-3 hrs (a sample of this culture was retained for analysis as Pre-induction FIBL).

2.3) For optimisation of the two main factors (IPTG & temperature), when the O.D reached about 0.6, each IPTG concentration was then added to each strain in four tubes (0.1, 0.5, and 1.0mM) to induce protein expression. Next, these were grown overnight at different temperatures (15, 20, 25 & 28ºC). The following day, each induction culture was harvested, and the media suspension was put into an Eppendorf tube and spun down at 7K for 5 mins.

2.4) The pellets in each condition were re-suspended in 200µl of 1xTBS then 100µL of this was put into a fresh Eppendorf tube and 100µl of 2x protein sample buffer was added.

2.5) All sample mixtures (pre-post induction) were boiled on the heat block at 85ºC for 15 mins, the DNA sheared using a small syringe, and the proteins expression determined using a bis-tris 4-12% gel (NuPAGE™, ThermoFisher, Scientific, USA) and stained with Coomassie blue. Meanwhile, the rest of the boiled samples were run on a bis-tris gel and the FIBL band detected using western-blotting by adding anti-His6-tagged (Anti-6X His tag® antibody, ABCAM, UK) and visualised using the Licor scanner (LI-COR, ODYSSEY CLx, Biotechnology-UK Ltd). The western blotting protocol was used in various experiments as shown in following 2.6 **(Figure 10).**

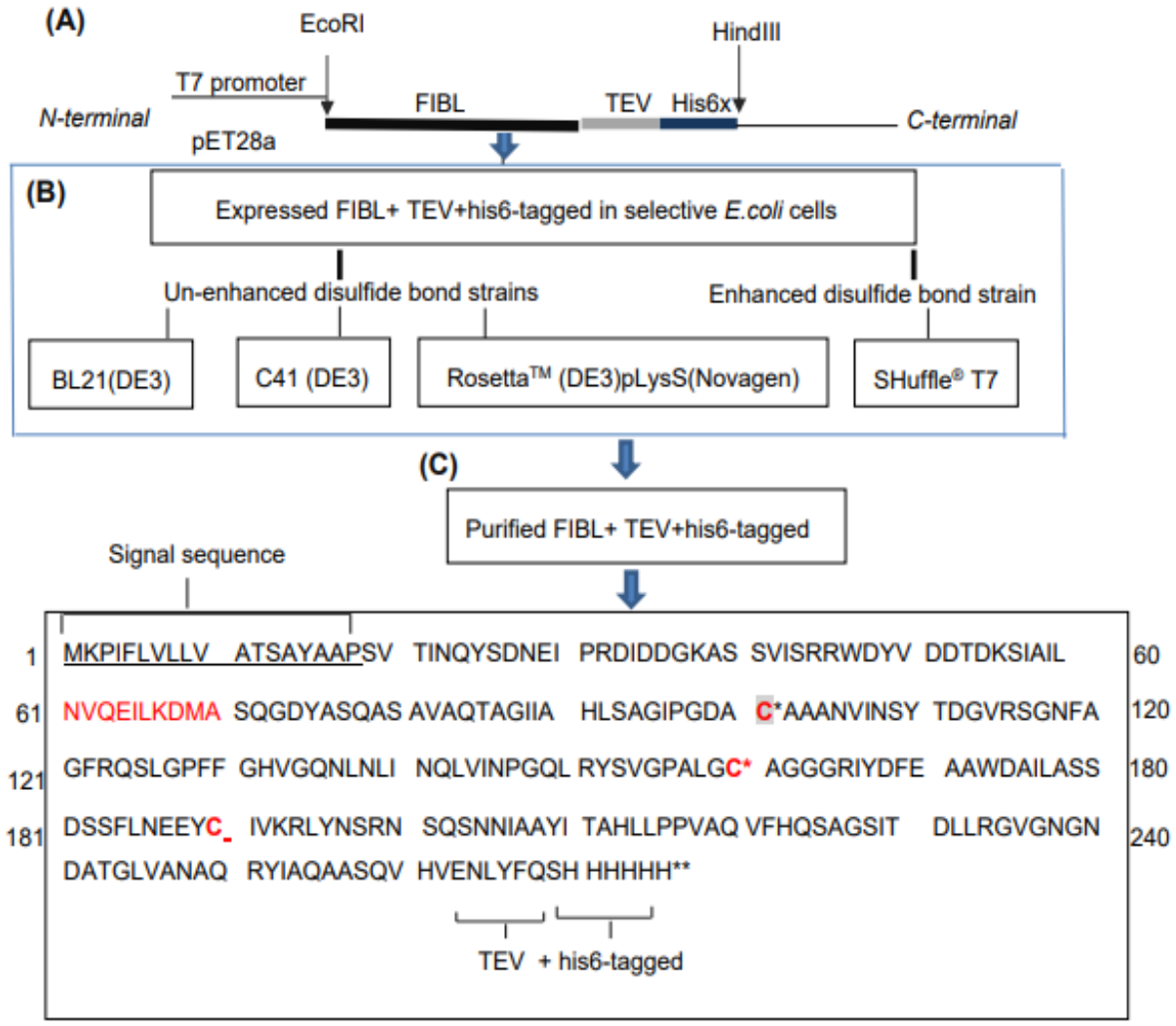


Figure 9. The FIBL gene was constructed in the pET28a with the T7 promoter, and the TEV protease was added at the C-terminus and his6x (his6-tagged) (**A)** and expression in four selective bacterial strains (**B).** The purified recombinant FIBL protein contains 275 aa with the signal sequence from 1-16aa (UniProt, Website 2022). Two different positions of cysteine residues at positions 101 and 160 are linked together as inter-disulfide bond whereas the last Cys residue at position 190 binds to the H-chain when it assembles at the silk fibroin elementary unit (C, red alphabets) in **Figure (C).**

2.6 SDS-PAGE & western blotting protocol

Transfer buffer: NUPAGE TRANSFER BUFFER 20X – Invitrogen (Fisher Scientific, UK)

PBS buffer (10x stock); the following was dissolved in 800mL distilled H_2O, 80g (NaCl), 2.0g (KCl),14.4g (Na_2HPO_4) and KH_2PO4

PBS-T1 solution; 500mL (PBS buffer 1x) and 500µL Tween 20

PBS-T3 solution; 500mL (PBS buffer 1x) and 1.5mL Tween 20

Blocking solution; 1g Marvel milk into a Falcon tube and added 20mL PBS-T1

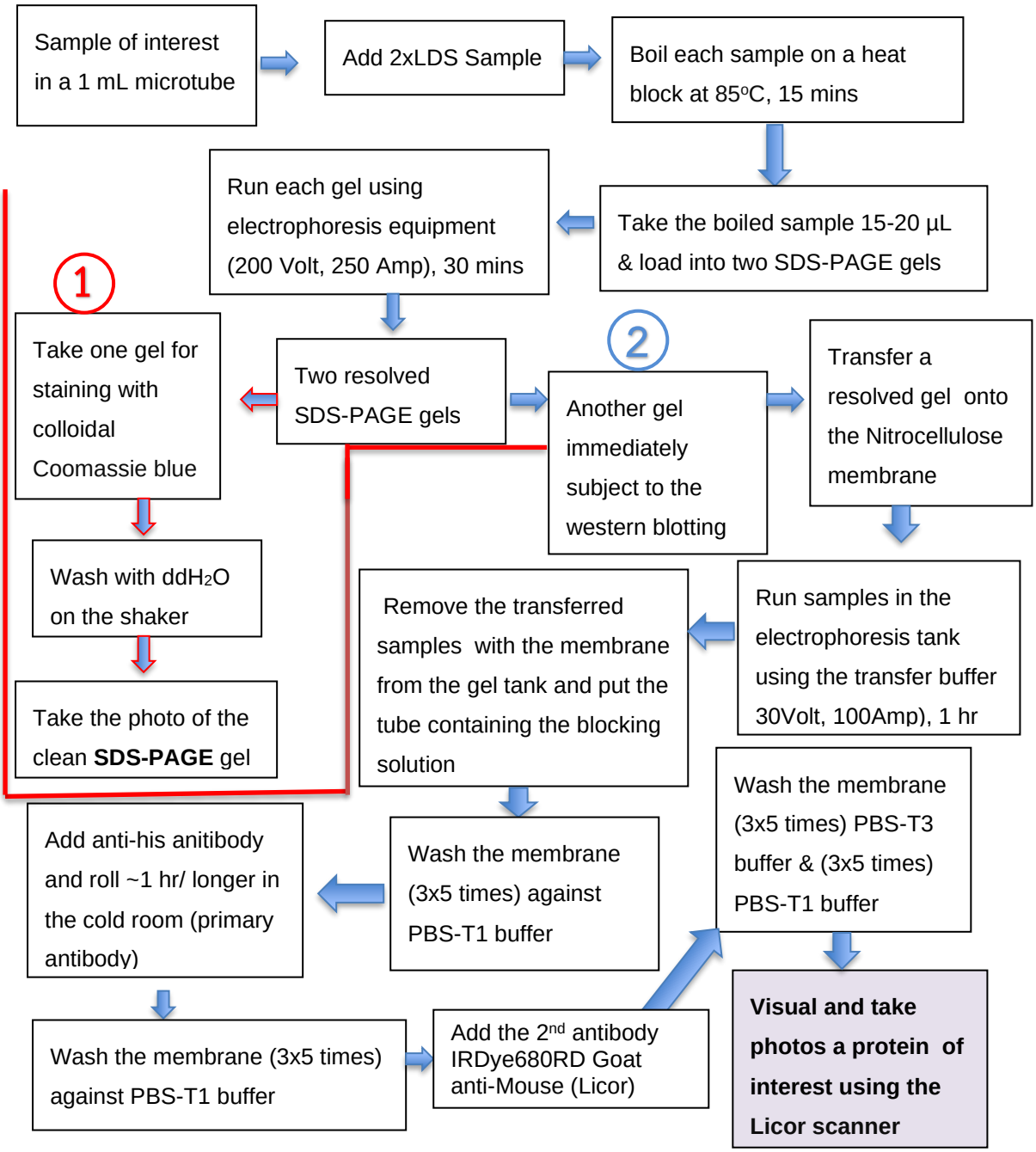


Figure 10. The summary protocol of SDS-PAGE **(1)** & western blotting **(2)** used in this project.

3) FIBL Protein solubilisation, its localisation, and purification (250mL cell culture)

For screening of expression in bacterial strains, two good expression strains of Group 1, BL21 and Rosetta™ (DE3)pLysS, underwent a solubilisation test whereas for Group 2 SHuffle® T7 was used.

3.1 FIBL localisation and solubility test in two bacterial strains from Group 1 & Group 2

A simple solubility test was carried out by collecting the total cell culture 1mL LB and it was spun down at 5,000 rpm. The pellet cells were retained and 1xTBS buffer added, then disrupted six-seven times with the 1mL syringe. The obtained solution was spun down and both fractions were collected for solubility analysis by adding protein buffer and running it on SDS-PAGE and western-blotting. However, for the larger scale localization test, the procedure below was followed:

3.1.1) The harvested cells (FIBL) were collected, and the remaining LB was discarded, then the cells were added and re-suspended in10mL 1xTBS buffer with (1/4 of a tablet, Roche complete, Mini EDTA-Free Protease Inhibitor Cocktail, Germany).

3.1.2) The bacterial cells were disrupted by 2-3 passes through a FRENCH Press (Thermo) at 8,000 psi and subsequent debris (Inclusion bodies, IBs) was pelleted out by centrifugation applying 17,000 rpm at 4ºC for 30 mins. These IBs pellets underwent solubilisation with 7 M urea containing 100mM HEPES, pH 8.0 for analysis by MW on SDS-PAGE and western-blotting for FIBL protein localisation confirmation.

3.1.3) The obtained supernatant fractions from the FRENCH Press were placed into fresh tubes and consequently underwent ultra-centrifugation at 42,000 rpm at 4ºC for 50 mins. The supernatant was then collected in a Falcon tube for Ni-NTA purification (His GraviTrap columns). The Ni-NTA column was washed with wash buffer (40 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 8.0), binding buffer (40mM Tris, 300mM NaCl, pH 8.0), then the resultant supernatant was washed with wash buffer and the FIBL protein eluted against elution buffer (40mM Tris, 300 mM NaCl, 400mM imidazole, pH 8.0).

3.1.4) The samples of purified FIBL protein in the pellets and supernatant were purified using His GraviTrap columns and their quantity (NanoDrop, ND-1000 Spectrophotometer) and quality were determined using SDS-PAGE and western-blotting techniques.

3.2 Optional comparison of non-dialysis & dialysis purification of soluble FIBL expressed in Group 1 (BL21) using Ni-NTA column

In relation to the previous solubilisation experiment with 7M urea at room temperature, the following day, the soluble FIBL, in total volume approximately 5mL expressed in BL21, was obtained by 15,000 rpm centrifugation at 4ºC. Then, the obtained soluble FIBL (5mL) was divided into two portions of 2 & 3mL and immediately forwarded for specific manual purification using two different methods of the Ni-NTA column, with or without dialysis, before being passed through the purification column without dialysis. 2mL soluble FIBL was used for non-dialysis purification, whereas 3mL FIBL was used for a dialysis process, firstly with a cassette (Thermo Scientific, Slide-A-Lyzer R

10K Dialysis Cassettes) and secondly with a magnetic stirrer containing 3L of HEPES buffer, pH 7.6 overnight in a cold room. The following day, both soluble FIBL were purified through the Ni-NTA column. Finally, the eluted purified FIBL protein concentration was measured (0.20mg/mL) and its quality was determined by SDS-PAGE and western-blotting.

4) Scaled-up expression, solubilisation, and purification (500 mL/Flask) Protocol for FIBL protein production in bacteria system

Earlier experiments in FIBL protein screening expression in four different strains revealed that both groups of bacterial strains have the potential for producing an active yield of FIBL. However, at this point, the SHuffle® T7 cells had stronger FIBL expression and production in bacteria due to the aid of the disulfide bond formation process. The SHuffle® T7 cells showed that they can provide the most promising aid to production of FIBL. Therefore, for larger scale protein production and purified protein characterisation, SHuffle® T7 cells were selected for further investigation. FIBL protein was located in the corrected folding of the inclusion bodies. In order to obtain the FIBL protein, experiments were performed rather differently from the small-scale experiments as follows.

4.1) A single colony of FIBL in SHuffle® T7 cells was picked and grown in 10mL LB media containing kanamycin in a 50mL Falcon tube overnight.

4.2) The next day, 10mL of the overnight cells' culture was added into fresh 500mL LB and 500µL kanamycin 100µg/mL was added, then continually grown in the shaker at 30ºC until $O.D._{600}$=0.6 approximately 2-3hrs (a sample of this culture was retained for analysis as Pre-induction FIBL).

4.3) 0.5mM of IPTG was added into the culture *for inducing protein expression when* the O.D reached about 0.6 and it was additionally grown overnight, the temperature maintained at 25ºC. Next day, the induced 500mL culture was split equally into two Nalgene plastic bottles for the cell pellets to be harvested by centrifugation (7K for 5 mins) and the supernatant was discarded**.**

4.4) Two tubes with the results from the previous step (250mL/500 cell culture) and 20mL of 1xTBS were put into a 50mL Falcon tube and 1mg/mL lysozyme was added (to break the bacterial cell walls) and incubated on ice for 20 mins.

4.5) Each cell's mixture was broken with 2-3 passes through a FRENCH Press (Thermo) at 8,000 psi and subsequent debris was pelleted out by centrifugation (17,000 rpm at 4ºC for 30 mins) and the supernatant discarded.

4.6) Two cell pellets obtained as inclusion bodies (IBs) were washed twice against 20mL of 1xTBS by centrifugation (17,000 rpm at 4ºC for 30 mins) and the IBs were preserved at -20ºC until use.

4.7) To optimise solubilisation of the FIBL protein, the obtained IBs (250 mL cell culture) were treated. To obtain the purified FIBL protein, 40mL of 7 M urea and 20mM BME were added to the FIBL IBs, suspended by pipetting up-down, and mixed with the vortex for three hours and subsequently the supernatants were collected, and the pellets discarded.

4.8) The supernatant containing FIBL protein was centrifuged once again (17,000 rpm at 4ºC for 30 mins) for the removal of unwanted contaminants and consequently 50mL of 20mM Tris-HCl,

pH 8.0 containing 0.5mM TCEP for diluting the supernatant's viscosity was added. The solubilised supernatant was then poured through a Ni-NTA purification (His GraviTrap) column. The Ni-NTA column was washed with wash buffer (40mM Tris, 300mM NaCl, 20 mM imidazole, pH 8.0, 0.5mM-TCEP), Binding buffer (40mM Tris, 300mM NaCl, pH 8.0, 0.5mM TCEP), then the solubilised FIBL supernatant was passed, washed with wash buffer and the FIBL protein eluted against elution buffer (40mM Tris, 300 mM NaCl, 400mM imidazole, pH 8.0, 0.5 mM TCEP). Eventually, the concentration and quality of the purified FIBL+TEV+His6-tagged was determined using SDS-PAGE and western blotting techniques.

5) Characterisation of FIBL protein expression in bacteria

The purified FIBL+TEV+His6-tagged was obtained from solubilised IBs pellets and refolded in the Ni-NTA column and subsequent stability, mass spectrometry and Circular Dichroism respectively were determined.

5.1 FIBL protein storage stability

The purified FIBL+TEV+His6-tagged was eluted in Elution buffer (40mM Tris, 300mM NaCl, 400mM imidazole, pH 8.0, 0.5mM TCEP) and initially preserved at -20°C in a freezer until further use. To determine its stability, 600µL of purified FIBL + TEV+His6-tagged was split from the product stored at -20°C and preserved in the refrigerator at 4°C. A sample of 80µL purified FIBL+ TEV+His6-tagged was taken for analysis every single day from the first day until the tenth day. Each day a sample of 2x protein sample buffer was added to FIBL+ TEV+His6-tagged and boiled at 85°C for 15 mins and kept it in the refrigerator until all ten days of samples were ready to be determined with SDS-PAGE and western blotting.

5.2 FIBL protein identification using mass-spectrometry analysis

The concentration (NanoDrop) and quality of the purified FIBL+TEV+His6-tagged in elution buffer (40mM Tris, 300mM NaCl, 400 mM imidazole, pH 8.0, 0.5 mM TCEP) was checked using SDS-PAGE bis-tris 4-12% gel, stained with Coomassie blue, and washed with water. The following day, a photo of the FIBL protein band was taken (used to recheck the whole band of FIBL and sent together with a sample analysis) single specific band of related of FIBL MW (~29.45 kDa) before being cut from a gel and placed into a sterilized tube containing 20µL ddH_2O. This sample was delivered to the BSRC mass spectrometry facility, University of St. Andrews, UK within 1-2 days. As soon as the sample arrived, the gel band was digested with trypsin and the resultant peptides were analysed by nLC-ESI MSMS. The output stored in the data file was compared against the NCBI database, using the Mascot search algorithm (BSRC Mass Spectrometry Facility, http://mass-spec.wp.st-andrews.ac.uk/).

5.3 FIBL protein secondary structure determination using CD technique.

CD spectra measurements were determined by a J-810 Spectropolarimeter (Jasco Inc, Japan) which was utilised to examine the secondary structure of 1mL of purified FIBL+TEV+His6-tagged in 20 mM Tris-HCl, pH 8.0. In looking at its structure, Spectra were recorded in the far-UV region (260 nm - 190 nm). Ideally, the spectra were represented for an average of three individual scans and were corrected for absorbance caused by the buffer. A quartz cuvette with a 0.1cm path length was used for

the measurement. The secondary structure data of the FIBL protein was obtained by means of the structure analysing software 'DicroWeb' (Sreerama, N. & Woody, R.W. 2000) that has been built into the JASCO J-810 spectropolarimeter. The framework for chapter III is shown in **Figure 11.**

Conceptual Framework

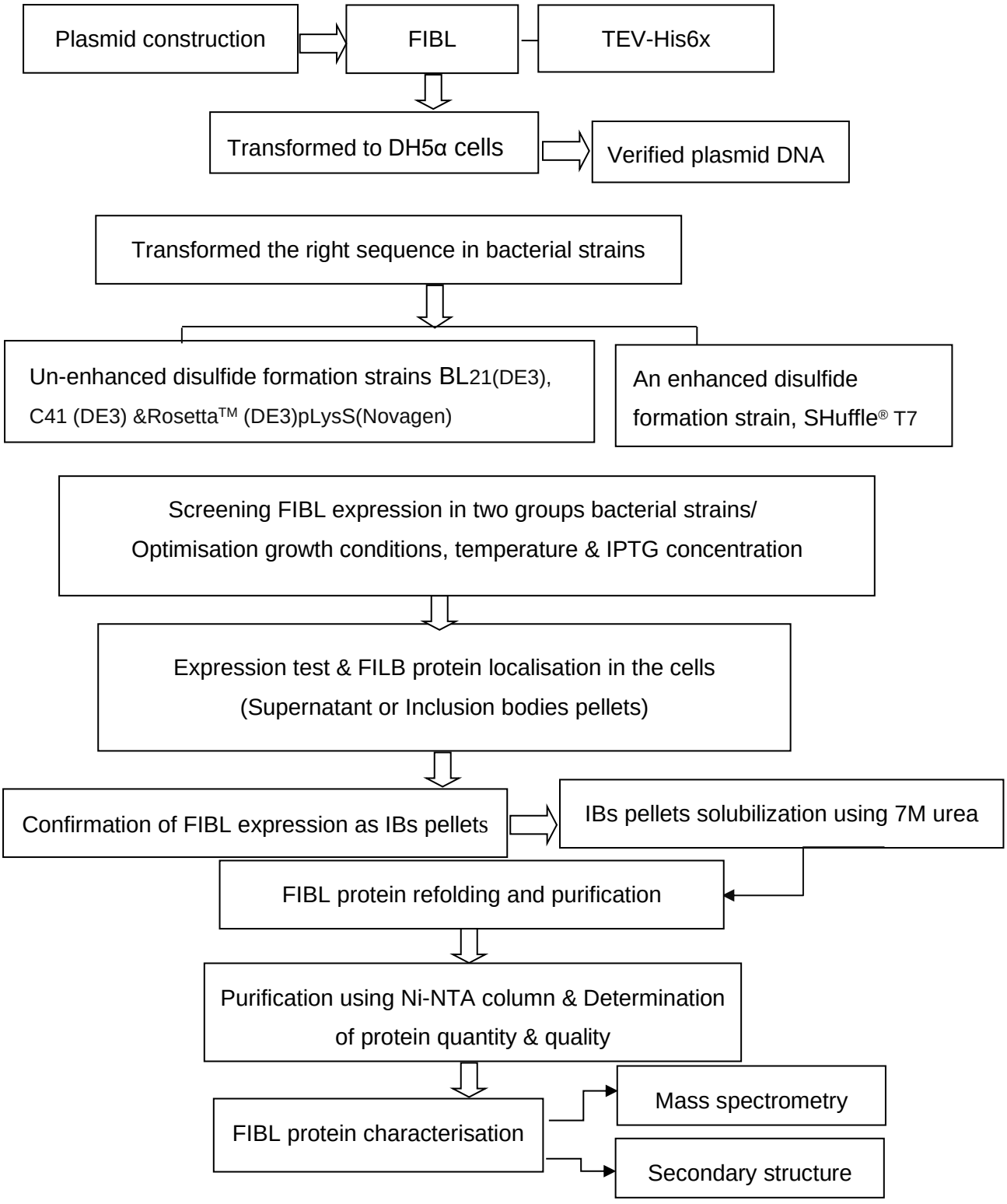


Figure 11. The schematic diagram of FIBL protein expression and production achieved in bacteria system**.**

Materials and method for chapter IV

"Comparative study of recombinant P25 glycoprotein (*B. mori*) in bacteria (Part I) & baculovirus-infected insect cell systems (Part II)"

The materials and methods employed for the P25 protein production were, 1) in bacteria system and 2) in baculovirus-infected insect cell systems as explained in the following section.

Part I: P25 expression in bacterial system

1) Vector construction (Molecular Engineering)

1.1. P25 Gblocks with TEV+His6-tagged at C-terminus were designed **(Figure 12)** and artificially synthesised by Integrated DNA Technologies, Belgium.

1.2. P25 Gblocks were amplified by adding two restriction zymes using PCR reactions (Primers, the forward primer (with EcoRI added) is 'ACG CCT AGG AAT TCA TGC TG' and the reverse primer (with XhoI added) is 'AAA AAA CTC GAG TTA TTA GTG ATG GTG ATG GTG' to increase the concentration of its DNA.

1.3. An amplified P25 gene was cloned (using Restriction-Ligation cloning technique) to the carrier vector (pET28a) and transformed into a host cell (DH5α™). The plasmid DNA rechecked by the ligation test with two restriction enzymes (EcoRI & XhoI). Upon the test, the plasmid DNAs containing the vector and insertion were sent for verification of its sequence (DNA sequencing and service, University of Dundee, UK).

1.4. Verification sequence of the construct, the plasmid DNA containing P25 and vector (pET28a) were transformed into four different competent cells (*E. coli*) organised in two groups of bacterial strains, Group 1 (un-enhanced disulfide bond formation), including three bacterial cells; BL21(DE3), C41 (DE3), Rosetta™ (DE3)pLysS(Novagen) and Group 2 (enhanced disulfide bond formation) containing one bacteria strain, SHuffle® T7. These cells were preserved at -80°C until use for inoculation and further studies for the investigation of protein expression **(Figure 12).**

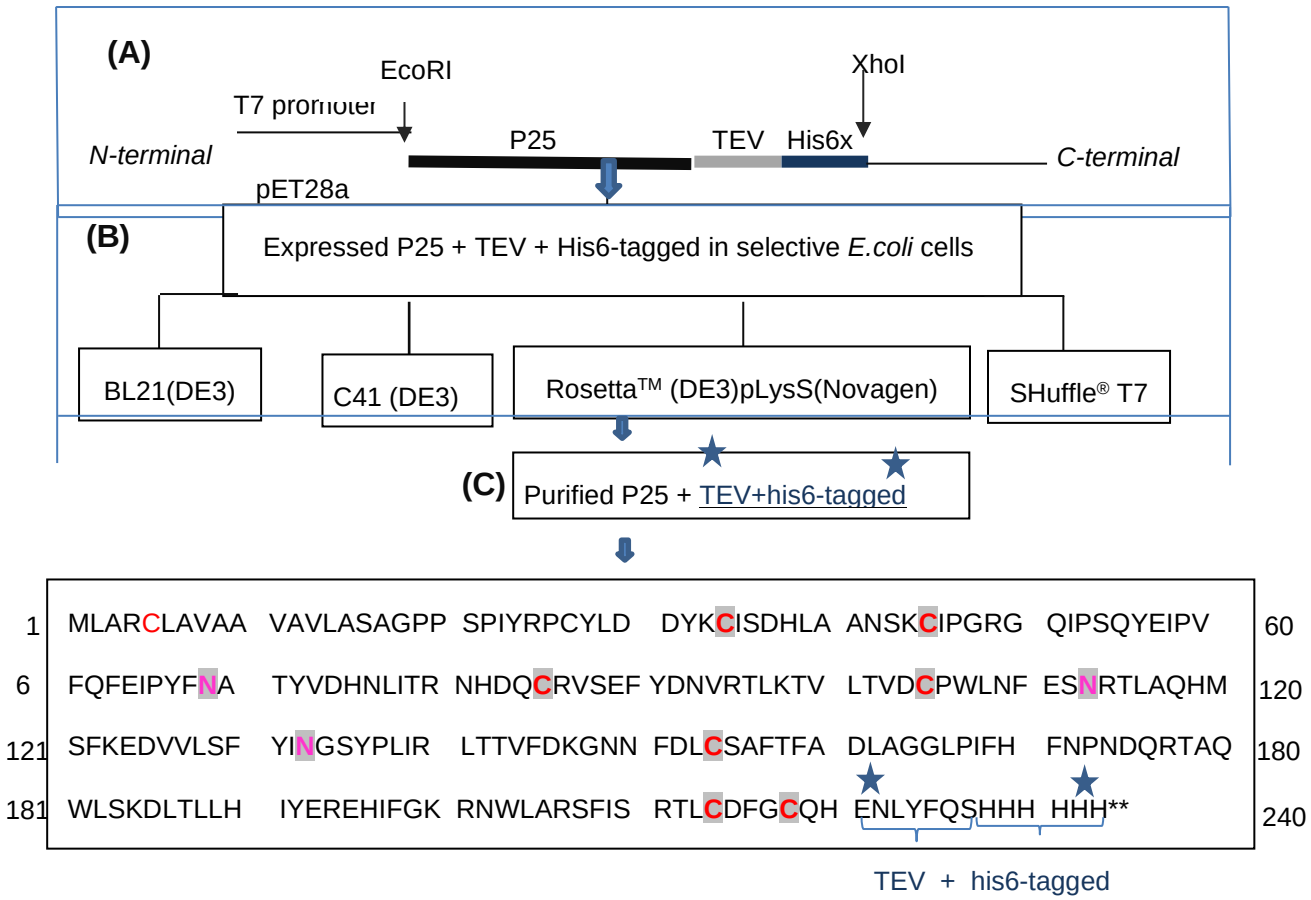


Figure 12. The verified DNA plasmid from to the DH5α cells was maintained as stock (in the refrigerator) **(A)** and was subsequently used for P25 protein expression in the bacterial strains **(B).** The recombinant purified P25 shows amino acids of three N-linked oligosaccharide sites (N, pink highlighted), and eight cysteine amino residues (C, red highlighted) in its full amino acid sequence containing extra amino acids of TEV protease and His6x (His6-tagged) represented as blue stars **(C).**

2) Screening expression and optimisation of P25 protein (small-scale production 10mL)

2.1) A single colony of four transformed P25 competent cells with two groups of bacterial strains, BL21 (DE3), C41 (DE3), Rosetta™ (DE3)pLysS (Novagen and SHuffle® T7) were picked and tripled in size in a 50mL Falcon tube overnight using 10mL LB media containing kanamycin.

2.2) The following day, 1mL of the overnight cells was put into a fresh 50mL Falcon tube and an extra 9mL of fresh LB-broth was added with 9µL kanamycin 100µg/mL and grown in the shaker at 30ºC until $O.D._{600}$=0.6 for approximately 2-3hrs (a sample of this culture was retained for analysis as Pre-induction P25).

2.3) For optimisation of two main factors (IPTG & temperature), IPTG was added to each culture (0.1, 0.5, and 1.0mM) to induce protein expression when the O.D reached about 6.0 and additionally grown overnight at different temperatures (15, 20, 25, & 28ºC). The next day, each induction culture was harvested 1mL, and the media suspension was transferred to an Eppendorf tube and spun down at 7,000 rpm for 5 mins.

2.4) The pellets in each condition were re-suspended in 200ul of 1xTBS then 100µl of this was put into a fresh Eppendorf tube and 100µl of 2x protein sample buffer was added.

2.5) All sample mixtures (pre-post induction) were boiled on the heat block at 85ºC for 15 mins, the DNA sheared using a small syringe, and the proteins expression determined using a bis-tris 4-12% gel (NuPAGE™, ThermoFisher, Scientific, USA) and stained with Coomassie blue. Meanwhile, the rest of the boiled samples were run on a bis-tris gel and the P25 band detected using western-blotting by adding anti-His6-tagged (Anti-6X His tag® antibody, ABCAM, UK) and visualised using the Licor scanner (LI-COR, ODYSSEY CLx, Biotechnology-UK Ltd).

3) P25 Protein solubilisation, purification and its localisation (250mL cell culture)

1) The harvested cells (P25) were collected, and the remaining LB was discarded, then the cells were added and re-suspended in10mL 1xTBS buffer with (1/4 of a tablet, Roche complete, Mini EDTA-Free Protease Inhibitor Cocktail, Germany).

2) The bacterial cells were disrupted using FRENCH Press (Thermo) for 2-3 passes at 8,000 psi and the subsequent debris was pelleted out by centrifugation applying 17,000 rpm at 4ºC for 30 mins. The pellets were retained for solubilisation with 7M urea (P25 protein localisation analysis).

3) The obtained supernatant was placed into fresh tubes and consequently underwent ultra-centrifugation at 42,000 rpm at 4ºC for 50 mins. The supernatant was then collected in a Falcon tube for Ni-NTA purification (His GraviTrap columns). The Ni-NTA column was washed with wash buffer (40 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 8.0), binding buffer (40 mM Tris, 300 mM NaCl, pH 8.0), then the resultant supernatant was washed with wash buffer and the P25 protein eluted against elution buffer (40mM Tris, 300mM NaCl, 400mM imidazole, pH 8.0).

4) The samples of purified P25 protein in the pellets and supernatant were purified using His GraviTrap columns and their quantity (NanoDrop, ND-1000 Spectrophotometer) and quality were determined using SDS-PAGE and western-blotting techniques.

4) Scaled-up expression, solubilisation, and purification (500mL/Flask) Protocol for P25 protein production in bacteria system

Earlier experiments in P25 protein screening expression in four different strains revealed that the SHuffle® T7 cells were the most suitable for P25 expression and production in bacteria due to the P25 glycosylation process. The SHuffle® T7 cells showed that they can provide the most promising aid to production of P25. Therefore, for larger scale protein production and purified protein characterisation, SHuffle® T7 cells were selected for further investigation. P25 protein was located in the corrected folding of the inclusion bodies. In order to obtain the P25 protein from the large-scale culture, experiments were performed rather differently from the small-scale experiments as follows.

4.1) A single colony of P25 in SHuffle® T7 cells was picked for inoculation and grown in 10mL LB media supplemented with kanamycin in a 50mL Falcon tube overnight.

4.2) The next day, 10mL of the overnight cells' culture was added into fresh 500mL LB and 500µL kanamycin 100µg/mL was added, then continually grown in the shaker at 30ºC until $O.D._{600}$=0.6 approximately 2-3hrs (a sample of this culture was retained for analysis as Pre induction P25.)

4.3) 0.5mM of IPTG was added into the culture *to induce protein expression when* the O.D reached about 6.0 and it was additionally grown overnight, the temperature maintained at 25ºC. Next day, the induced 500mL culture was split equally into two Nalgene plastic bottles for the cell pellets to be harvested by centrifugation (7,000 rpm for 5 mins) and the supernatant was discarded**.**

4.4) Two tubes with the results from the previous step (250 mL/500 cell culture) and 20mL of 1xTBS were put into a 50mL Falcon tube and 1 mg/mL lysozyme was added (to break the bacterial cell walls) and incubated on ice for 20 mins.

4.5) Each cell's mixture was broken using FRENCH Press (Thermo) with 2-3 passes at 8,000 psi and subsequently the cell debris was pelleted out by centrifugation (17,000 rpm at 4ºC for 30 mins) and the supernatant discarded.

4.6) Two cell pellets obtained as inclusion bodies (IBs) were washed twice against 20mL of 1xTBS by centrifugation (17,000 rpm at 4ºC for 30 mins) and the IBs were preserved at -20ºC until use.

4.7) To optimise solubilisation of the P25 protein, the obtained IBs (250mL cell culture) were transferred from the freezer -20ºC. To obtain the purified P25 protein, 40mL of 7 M urea and 20mM BME were added into the P25 IBs, suspended by pipetting up-down, and mixed with the vortex for three hours and subsequently the supernatants were collected, and the pellets discarded.

4.8) The supernatant containing P25 protein was centrifuged once again (17,000 rpm at 4ºC for 30 mins) for the removal of unwanted contaminants and consequently 50mL of 20mM Tris-HCl, pH8.0 containing 0.5mM TCEP was added to reduce the supernatant's viscosity. The solubilised supernatant was then poured through an Ni-NTA purification column (His GraviTrap columns). The Ni-NTA column was washed with wash buffer (40mM Tris, 300mM NaCl, 20mM imidazole, pH 8.0, 0.5mM-TCEP), Binding buffer (40mM Tris, 300mM NaCl, pH 8.0, 0.5mM TCEP), then the solubilised P25 supernatant was passed, washed with wash buffer and the P25 protein eluted against elution buffer (40mM Tris, 300mM NaCl, 400mM imidazole, pH 8.0, 0.5mM TCEP). Eventually, the concentration and quality of the purified P25+TEV+His6-tagged was determined using SDS-PAGE and western blotting techniques.

5) Characterisations of P25 protein expressed in bacteria

The purified P25+TEV+His6-tagged was obtained from solubilised IBs pellets and refolded in the Ni-NTA column and subsequent stability, mass spectrometry and Circular Dichroism respectively were determined.

5.1 P25 Protein storage stability

The purified P25+TEV+His6-tagged was eluted in elution buffer (40mM Tris, 300mM NaCl, 400mM imidazole, pH 8.0, 0.5mM TCEP) and initially preserved at -20°C in a freezer until further use. To determine its stability, 600µL of purified P25 + TEV+His6-tagged was split from the product stored at -20°C and preserved in the refrigerator at 4ºC. A sample of 80µL purified P25+ TEV+His6-tagged

was taken for analysis every single day from the first day until the tenth day. Each day a sample of 2x protein sample buffer was added to P25 + TEV+His6-tagged and boiled at 85°C for 15 mins and kept in the refrigerator until all ten days of samples were ready to be determined with SDS-PAGE and western blotting.

5.2 P25 protein identification using mass-spectrometry analysis

The concentration (NanoDrop) and quality of the purified P25+TEV+His6-tagged in elution buffer (40mM Tris, 300mM NaCl, 400mM imidazole, pH 8.0, 0.5mM TCEP) was checked using SDS-PAGE bis-tris 4-12% gel, stained with Coomassie blue, and washed with water. The following day, a photo of the P25 protein band was taken (used to recheck the whole band of FIBL and sent together with a sample analysis) of a single specific band of related P25 MW (~28 kDa) before being cut from a gel and placed into a sterilised tube containing 20µL ddH_2O. This sample was delivered to the BSRC mass spectrometry facility, University of St. Andrews, UK within 1-2 days. As soon as the sample arrived, a gel band was digested with trypsin and the resultant peptides were analysed by nLC-ESI MSMS. The output stored in the data file was compared against the NCBI database, using the Mascot search algorithm (BSRC Mass Spectrometry Facility, http://mass-spec.wp.st-andrews.ac.uk/).

.

5.3 P25 protein secondary structure determination using CD technique.

CD measurements were determined by a J-810 Spectropolarimeter (Jasco Inc, Japan) which was utilised to examine the secondary structure of 1mL of purified P25+TEV+His6-tagged in 20mM Tris-HCl, pH 8.0. In looking at its structure, Spectra were recorded in the far-UV region (260nm - 190nm) and the near-UV region (320nm - 250 nm). Ideally, the spectra represented for an average of three individual scans and were corrected for absorbance caused by the buffer. A quartz cuvette with a 0.1cm path length was used for the measurement. The secondary structure data of the P25 protein was obtained by means of the structure fitting software 'DicroWeb' (Sreerama, N. & Woody, R.W. 2000) that has been built into the JASCO J-810 spectropolarimeter.

Part (II) P25 protein expression in baculovirus-infected insect cell system

1) Vector construction (Molecular Engineering)

1.1. P25 Gblocks with TEV+His6-tagged at the C-terminal used the same stock as described previously in vector construction of P25 in bacterial system.

1.2 The forward primer (with BamHI added) 'AAA AAA ACG CCT AG**G** GAT **CC**A TGC TGG CGC GGT GTC' and (with Not I added) 'AAA AAG CGG CCG CTT ATT AGT GAT GGT GAT GGT G' reverse primers were designed specifically for the insect cell expression vector $^{\text{his6-tagged+3C protease}}$pFBDM (the N-terminal had already been fused) which was obtained from Helen Walden's Lab, University of Glasgow, UK

1.3 The P25 Gblocks were amplified using PCR reactions for substitution in two restriction

enzyme sites, (BamHI) and (Not I), with the original P25 Gblocks which are suitable for substitution with the same restriction enzymes available in the pFBDM vector plasmid.

1.4 An Amplified P25 gene was cloned to the carrier vector (pFBDM) and transformed to a host cell (DH5α ™) for making the plasmid DNA before its sequence was sent for verification (DNA sequencing and service, University of Dundee, UK). The construct for P25 for baculovirus-infected insect cell system as shown in **Figure 13.**

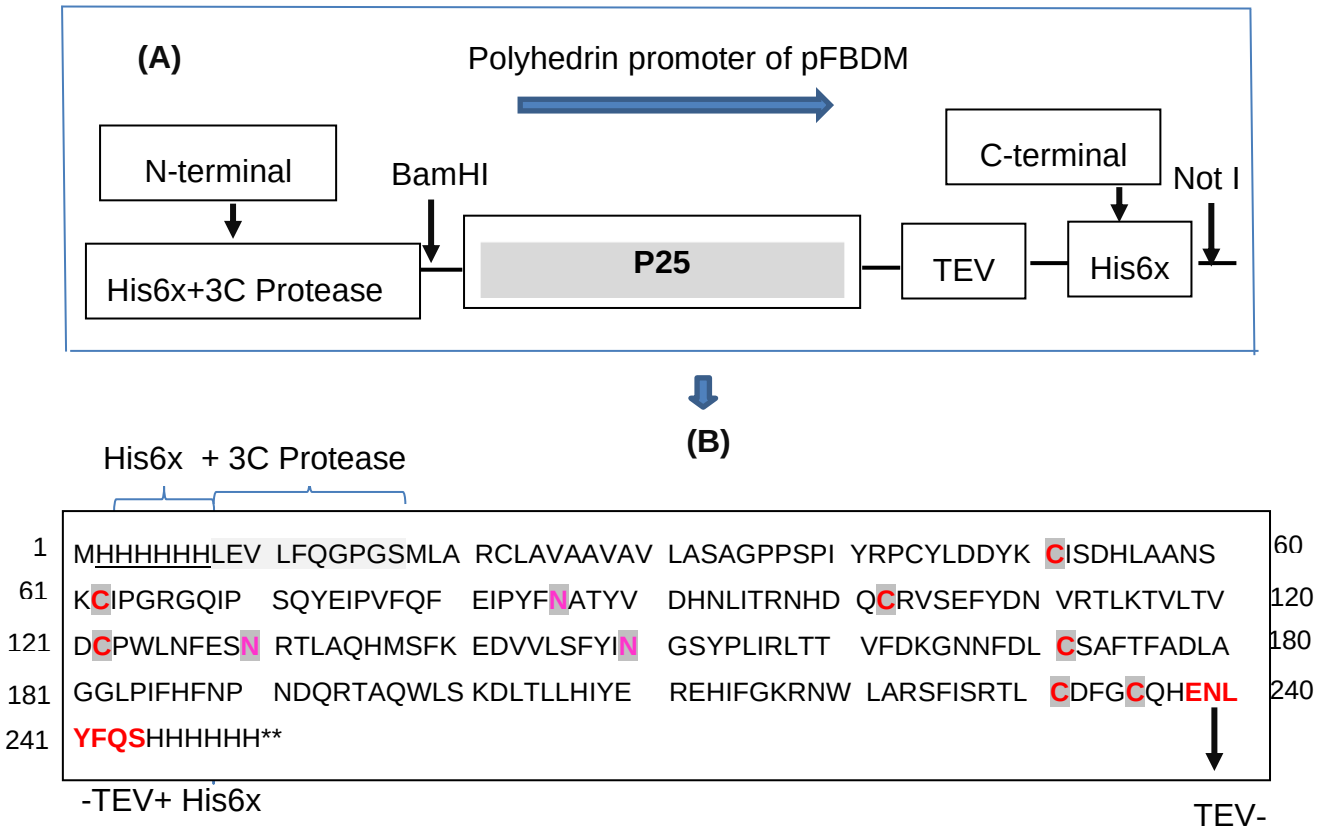


Figure 13. The P25 gene was constructed in the pFBDM with the polyhedrin promoter which was labelled as His6x (His6-tagged) and 3C protease at the N-terminal site and the TEV protease and his6x was added at the C-terminus (**A**). Three P25 *N*-linked oligosaccharide sites (N, pink alphabets), and eight cysteine residues (C, red alphabets) in its full amino acids sequence contain TEV protease and His6x (His6-tagged) **(B).**

2) Bacmid preparation (Molecular Engineering)

Standard transformation briefly explained is where 100µL of EMBacY competent cells are transformed with at least 100ng of P25/pFBDM plasmid and grown with 900µL of LB media incubated/recovered at 37ºC for 4hrs. The incubated cells were plated onto selection LB agar plates (Blue/white Screening technique) and grown at 37ºC until the colonies were noticeable ~24hrs. The right insertion of the P25pFBDM/ EMBacY was visible in white colonies of which two-four colonies were picked and re-streaked onto fresh selection plates and grown at 37ºC until the colonies were visible ~24hrs (to confirm the white colonies). Two confirmed white colonies were grown individually in the

2x2mL LB cultures (50µg/mL Kanamycin and 7µg/mL Gentamycin) at 37ºC for ~24hrs. The following day, *E. coli* cells were pelleted, extracted, and the bacmid DNA purified using Qiagen Kit.

3) Generation of baculovirus stocks and P25 protein localisation

The obtained purified bacmid P25pFBDM/EMBacY was then transfected into Sf21 insect cells /Sf21 cells (*Spodoptera frugiperda*, fall armyworm) for production of a starter virus (P0). This was subsequently used for the next generation of production optimisation and stronger P1 virus for long-term storage in the refrigerator. The next generation, a second baculovirus (P2), was induced from P1 for large-scale expression of recombinant P25 protein. Typically, all three stocks of baculovirus (P0, P1, and P2) and recombinant P25 protein took three days for expression in Sf21 cells. All P25 expressions, Sf21 cells were grown in insect media, Insect-XPRESS™ Protein-free Insect Cell Medium (Lonza).

4) Large-scale expression and P25 protein production

For expression and purification of P25 protein, the Sf21 cells (Wide type, WT) were firstly provided, and its viability determined and activated in the shaker at 110rpm at 27ºC before being transfected with the P2 virus. Generally, large-scale expression of P25 protein (maximum volume/flask was 600mL/medium culture) took 72hrs. Production protocol described in brief is when Sf21 cells at $1x10^6$ cells/mL were firstly added into the plastic flask (2,000 mL), followed by the P2 virus (6mL), 600mL of insect media added and supplemented with 1xAntibiotic-Antimycotin (Anti-Anti, Gibco), respectively. All these steps were conducted in a sterile laminar flow hood. Post-transfection (24-72hrs) of cells viability was measured by Trypan blue staining to monitor infection of P2 in Sf21 cells. This culture was grown in the incubator (110 rpm at 27ºC) and both the cell pellets harvested (semi-glycosylated P25 protein) and the medium suspended (glycosylated P25 protein) ~72hrs by centrifugation 17,000 rpm at 4°C for 15 mins. The cell pellets were re-suspended in 30mL lysis buffer (300mM NaCl, 20mM Tris pH 8.0, 5% glycerol, and 5mM BME containing one tablet of EDTA-free protease inhibitor cocktail) was preserved in the freezer (-20ºC) until later use/ no longer than 2 months. A part of the medium suspension (supernatant) was harvested for further experiments related to this project. The harvested medium (600mL) was re-suspended in a tablet of EDTA-Free Protease Inhibitor Cocktail and preserved in the freezer (~a month) until use for purification.

5) Two forms of glycosylated P25 protein purifications

5.1 Purification of semi-glycosylated P25 protein from the cell pellets using Ni-NTA column

30mL of lysed cell pellets in lysis buffer (300mM NaCl, 20mM Tris pH 8.0, 5% glycerol, and 5mM BME containing one tablet of EDTA-free protease inhibitor cocktail) were gently disrupted manually by thawing (room temperature) and freezing (-20ºC) for three cycles. Only the supernatant of the mixture of lysate cells was then collected by centrifugation twice at 17,000 rpm at 4ºC for 30 mins and the pellets discarded. This subsequent clarified supernatant was purified using Ni-NTA glass column (Bio-Rad). The NTA-resin beads were washed with Wash buffer (40mM Tris, 300mM NaCl, 20mM imidazole, pH 8.0), binding buffer (40mM Tris, 300mM NaCl, pH 8.0), then passed with the supernatant, washed with wash buffer and the P25 protein eluted against elution buffer (40mM Tris, 300mM NaCl, 400mM imidazole, pH 8.0). This resulting purified semi-glycosylated P25 protein was

changed into 20mM Tris-HCl, pH 8.0 (native form) for some further experiments.

5.2 Purification of glycosylated P25 protein from the medium fraction (supernatant) using Ni-NTA column

After the glycosylated P25 protein was secreted out from the transfected Sf21 cells, it was then mixed with the medium culture. 600mL of the harvested medium fraction supplemented in a tablet of EDTA-Free Protease Inhibitor Cocktail was gradually passed through the Ni-NTA glass column or, if it was preserved in the freezer at -20ºC, thawed into the solution before purification using the Ni-NTA column. The purification steps and buffers were used in the same way as described previously for the semi-glycosylated P25 protein process.

5.3 Additional purification of semi-glycosylated and glycosylated P25 proteins using size exclusion chromatography (SEC)

With the two forms of P25 proteins having been purified through Ni-NTA purification column, a size exclusion chromatography machine (AKTA pure systems) was used to further purify these concentrated semi-glycosylated and glycosylated forms of P25 proteins (ranging from 0.7-10mg/mL), by combination of Superdex 75/200 GL column. The purified P25 proteins were eluted in elution buffer (20mM Tris-HCl, pH 8.0, and 2mM TCEP). The eluted proteins corresponding with the peaks were collected and the protein quality was determined using SDS-PAGE bis-tris 4-12% gel stained with the Coomassie blue and western blotting.

5.4 Two forms of semi-glycosylated/ glycosylated P25 protein primary determination

Basically, throughout all experiments to obtain purified P25 proteins, the concentration was measured using NanoDrop and protein quality using SDS-PAGE bis-tris 4-12% gel, stained with the Coomassie blue and western blotting.

6. Characterisation of P25 proteins

6.1 Secondary structure of P25 proteins using CD technique.

Once again, CD measurements were utilised to examine two forms of the secondary structures of 1mL of purified His6-tagged+3C protease+P25+TEV+His6-tagged (in 20mM Tris-HCl, pH 8.0) which were isolated from two parts of culture, pellets/ infected Sf21 cells (semi-glycosylated) and the medium fraction/supernatant (glycosylated).

6.2 Glycan analysis of P25 proteins

6.2.1) Characterisation of glycans between semi-glycosylated and glycosylated P25 proteins using endoglycosidase treatment.

Two forms of purified P25 (semi-glycosylated and glycosylated P25 proteins) after Ni-NTA purification were concentrated and changed into a solution (50mM Tris-HCl, pH 8.0 and 100mM NaCl) using a concentrator and subsequently used for this experiment as folded pattern of P25 (native) whereas an unfolded pattern (denatured) of two forms of purified P25 proteins (20µL) were treated by adding 220µL 8M urea and incubated for 1 hr at room temperature. Both purified P25 proteins with two

patterns, folded and unfolded were further treated with two different enzymes, Endo-H and PNGase- F (New England Biolabs, USA). Treatment with these two enzymes entailed eight reactions in each Eppendorf tube, 0.70mg/mL of semi-glycosylated P25 (native, denatured) and 0.70mg/mL of glycosylated P25 (native, denatured) 2µL of denaturing glycoprotein which was then added and boiled on the heat block at 105ºC for 8 mins. The four boiled reactions of semi-glycosylated P25 (native, denatured) and 0.70mg/mL of glycosylated P25 (native, denatured) were treated with Endo-H by adding 4µL Glycoprotein Buffer3 (10x), 4µL ddH_2O and 2µL Enzyme (Endo-H), respectively. Four other boiled reactions were incubated with PNGase F treatment by adding 4µLGlycoprotein Buffer2 (10x), 4µL of 10% NP-40%, 4µL ddH_2O and 2µL enzyme (PNGase F). All of the reactions were incubated at 37ºC for 1hr and were subsequently spilt into two portions for analysing on a bis-tris 4-10% gel and western blotting. The mobility bands on the gels indicated the release of glycans from the treatments.

6.2.2) Characterisation of glycans between semi-glycosylated and glycosylated P25 protein using lectin binding/ Immunoprecipitation (IP) techniques

Two forms of purified P25 proteins, including semi-glycosylated and glycosylated after Ni-NTA purification, were consequently concentrated and changed into a buffer solution (50mM Tris-HCl, pH 8.0 and 100mM NaCl) which was designated as native state. However, two forms of purified semi-glycosylated and glycosylated P25 proteins were converted to unfolded form (denatured state) by treatment in 8M urea as described previously. To investigate the glycan from semi-glycosylated and glycosylated purified P25 proteins, the lectin (Concanavalin A, Con A) binding technique or immunoprecipitation was used. Briefly, each semi-glycosylated and glycosylated (0.70mg/mL) was provided in four reaction tubes, folded state (native), unfolded state (denatured), IP treatment (repetition 1/ R1), and IP treatment (repetition 2/ R2). 75µL Con A lysis buffer was added to four IP reaction tubes (R1 and R2), followed by 75µL of 20% Sepharose CLB beads and tumbled on the roller for 30 mins at 4ºC. The supernatant from the four reactions was collected by centrifugation at 6,000 rpm at 4ºC, for 1 min before the lectin (Con A, Sepharose beads, Sigma) was added, mixed well and tumbled on the roller for 30 mins at 4ºC. Next, all four reactions were washed (4 times x 5 mins) 6,000 rpm, with 750µL Con A lysis buffer and eventually pellets remained in the tubes. The mobility of these four IP treatments was analysed using SDS-PAGE 4-12% gel, stained with the Coomassie blue and western blotting.

In this chapter, we have shown that P25 protein production in Sf21 cells has many long processes of construction, expression optimization, purification and characterization. The workflow is briefly summarized in **Figure 14.**

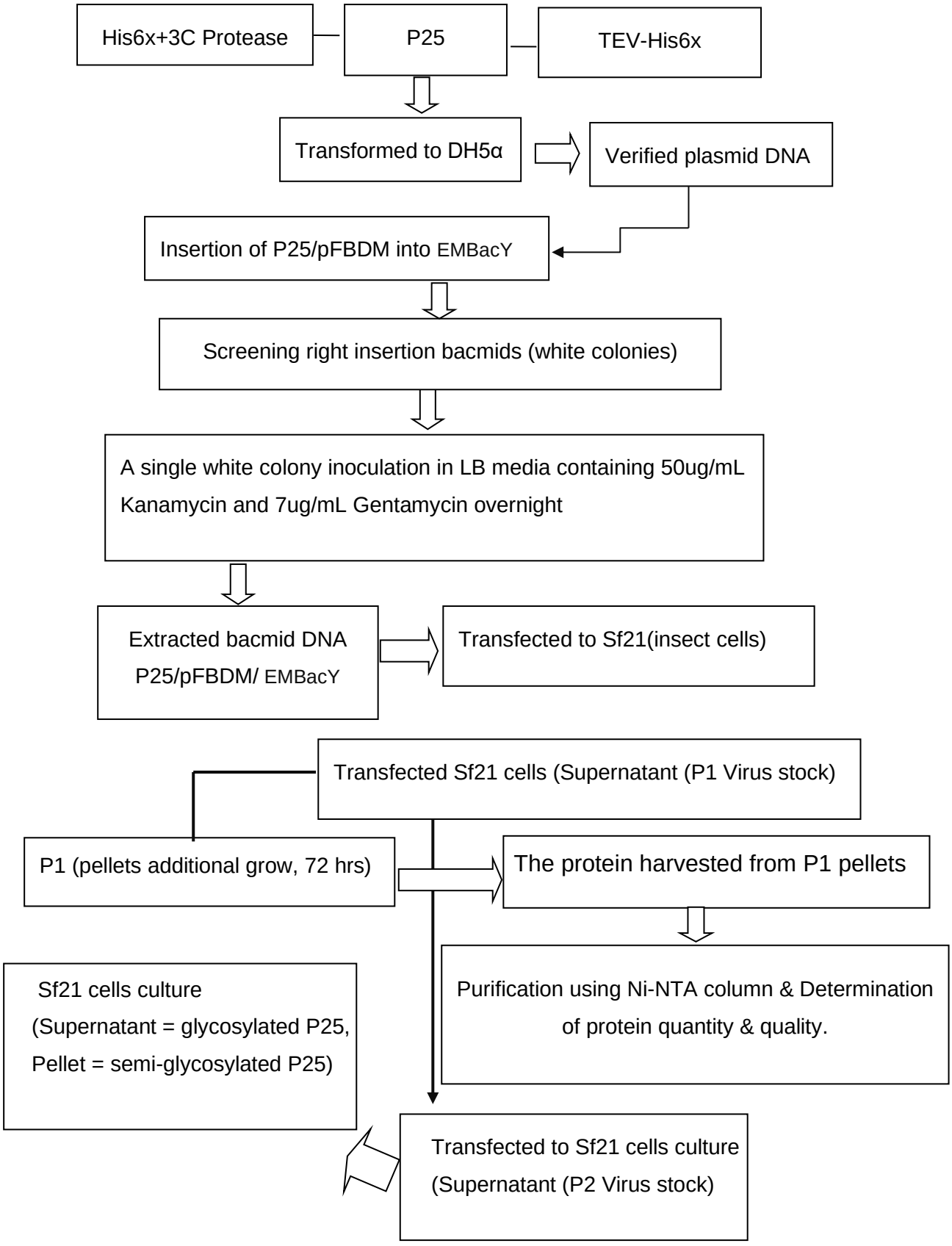


Figure 14. Schematic to explain of the P25 protein production processes starting from the plasmid construction, bacmid insertion P25/pFBDM/EMBacY, transfection of bacmid DNA to Sf21 cells for P1 virus stock, production of P2 virus, and large-scale P25 protein production from virus P2.

Materials and method for chapter V

"*In vitro* silk fibroin complex, its indigenous interactions and preliminary interactions with human proteins"

The materials and methods for chapter V experimentally involved with three fibroin molecules, H-chain, FIBL and P25 are shown as follows:

1) Insect preparation

The *B.mori* silkworm larvae (Thai polyvoltine variety; Nang-Kau) were reared with mulberry leaves at the Center of Excellence for Silk Innovation, Mahasarakham University, Thailand. The fully grown larvae were randomly collected from fifty mixed-sex larvae, on day 5 in the 5th instar, one day before the silk spinning to form a cocoon.

1.1)Fresh posterior silk glands preparation for solid fibroin

Figure 15, fifty silkworm larvae were dissected using a small pair of scissors, the posterior silk gland (PSG) was collected, washed briefly in 1.5% KCl, transferred immediately into ice cold 30% ethanol and incubated for 18hrs at 4°C. The following day, the unwanted outer silk gland cell layer of PSG was squeezed off with fine forceps to harvest only solid fibroin, which was washed several times in distilled water, dehydrated in 50% ethanol, 99.5% ethanol and diethyl ether, respectively (Katagata, et al.,1984). The solid fibroin that was obtained was sent to the author by international parcel post service. The solid fibroin was kept in the desiccator until later use.

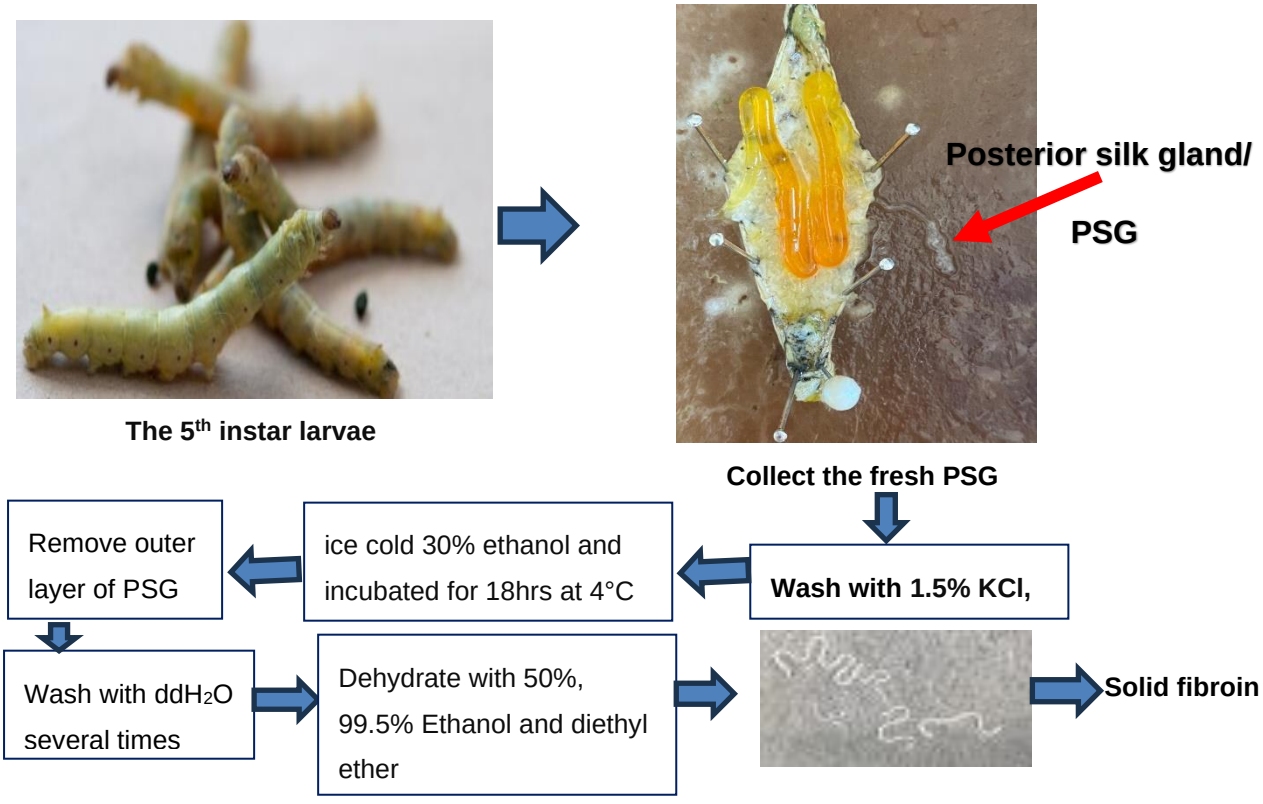


Figure 15. Schematic to describe the solid fibroin preparation from the fresh PSG of the 5th instar larvae.

1.2) H-chain isolation and purification

The solid fibroin was subjected to extraction and isolation using methods of Shimura et al. (1976) & Katagata et al., (1984). Briefly, 10mg of the solid fibroin was dissolved in 1mL of 60%LiSCN. The fibroin solution obtained was then carboxymethylated in 10ml of 1M Tris-HCl, pH 8.0, 20mg of EDTA, and 0.7mL of BME. After carboxymethylation, the fibroin solution was subjected to precipitation by 19% & 30% saturation of $(NH_4)_2\ SO_4$, the pellets dissolved with 60% lithium thiocyanate (LiSCN), dialysed against buffer A (10 kDa MWCO in 5M urea, 20mM Tris-HCl, pH 8.0). Next, the 1st purification of the resulting fibroin solution was achieved using Superdex 75 10/300 GL (GE Healthcare AKTA Pure 25 L FPLC System, USA) and eluted in Buffer A (5M urea in 20mM Tris-HCl, pH 8.0) . Eventually the H-chain molecule of the fibroin solution was obtained by isolating it using anion ion-exchange chromatography with a 1mL HiTrap HP Q column and combined with salt gradients technique (GE Healthcare AKTA Pure 25L FPLC System, USA). The processes to extract H-chain of fibroin from the solid fibroin and its purification are shown in **Figure 16.**

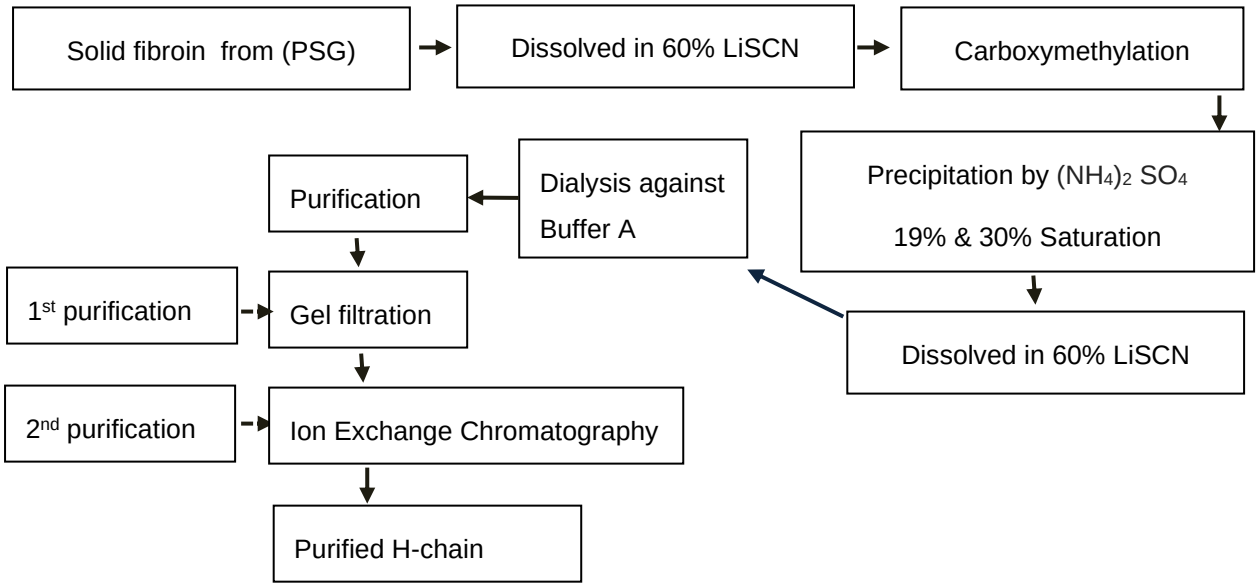


Figure 16. The isolation and purification from the solid fibroin, the H-chain was firstly solubilised in 60% LiSCN, followed by carboxymethylation, precipitation with 19%&30% saturation of $(NH_4)_2SO_4$, the precipitated pellets collected and dissolved in 60% LiSCN once again. The fibroin solutions that were obtained were dialysed in buffer A overnight. The following day, the 1st purification of the fibroin solution was carried out using the gel filtration and the 2nd purification by ion exchange chromatography.

2. Investigation of in vitro silk fibroin assembled complexes

2.1. Observation of the initiation of silk fibroin complexes formation or multimerisation behaviour along its three indigenous proteins (H-chain, FIBL &P25) and interaction with

selected human proteins (CLIC1 & CLIC4).

2.1.1) The main fibroin molecule, H-chain was isolated and purified from the solid fibroin in its partially folded form in Buffer A (5M urea, 20mM Tris-HCl, pH 8.0) or provided in a native form (100mM NaCl, 20mM Tris-HCl, pH 8.0).

2.1.2) Two other fibroin molecules, FIBL and P25 were expressed separately in bacteria and insect cell systems (Materials and Methods Section). FIBL was obtained in SHuffle® T7 cells whereas P25 was obtained in Sf21 cells production. Both purified proteins were immediately preserved in native form (150mM NaCl, 50mM Tris-HCl, pH 8.0) in the refrigerator until use for forming the complex.

2.1.3) The purified native H-chain was used to determine how complexes form with its elements in each incubation condition through five treatments, at room temperature and on ice including, (H-chain alone, H-chain+P25, H-chain+FIBL, FIBL+P25 and 3 subunits) and maintained for 30 minutes in both treatments.

2.1.4) Incubated fibroin complex samples were determined by Native PAGE (without reducing agent) and SDS-PAGE (with two reducing agents, BME or DTT). Both Native PAGE and SDS-PAGE were run on 4-12% Tris-Glycine gels (Invitrogen™ Novex, USA).

2.1.5) In addition, the native H-chain complexes underwent experimentation with two selected human proteins known as intracellular chloride channel proteins (CLIC1 & CLIC4) expressed in bacteria.

2.1.6) Incubated fibroin-CLIC1&4 complexes samples were determined by Native PAGE (without reducing agent) and SDS-PAGE (without reducing agent). Both Native PAGE and SDS-PAGE were run on 4-12% Tris-Glycine gels (Invitrogen™ Novex, USA).

2.2. Determination In vitro silk fibroin complexes formation or protein-protein interactions using (Size Exclusion Chromatography and Laser Light Scattering Service, SEC-MALS)

2.2.1) 200µL of each native fibroin protein complexes, H-chain alone, H-chain+P25, H-chain+FIBL, FIBL+P25, and 3 subunits were placed in 100mM NaCl, 20mM Tris-HCl, pH 8.0 (with the concentration ≥200µg/mL).

2.2.2) To start determination, the protein oligomerization column was flushed overnight in solution with SEC-MALS, 500mLof buffer (100mM NaCl, 20mM Tris-HCl, pH 8.0)

2.2.3) The following day, the 200µL BSA (1-2mg/mL in the SEC buffer) was loaded and each complex was loaded individually each time.

2.2.4). Each loaded protein complex was detected using refractive index (RI) with 15mL of retention volume which takes approximately 30 mins per sample and eventually raw data was obtained and

subjected further for making the chromatogram regions and was analysed/calculated using MALS software OMniSEC software (www.malvern.com/viscotek).

2.3. Determination in vitro silk fibroin complex formation or protein-protein interactions using (Size Exclusion Chromatography, SEC)

Comparison of two forms of protein interaction, (A) native-native fibroin complexes formation vs.(B) partial native-partial native fibroin complexes

(A) native-native fibroin complex formation (fully folded form)

A-1) 1mL of each native fibroin proteins complex, H-chain alone, H-chain+P25, H-chain+FIBL, FIBL+P25, and 3 subunits were placed in 100mM NaCl, 20mM Tris-HCl, pH 8.0 (with the concentration ≥400µg/mL

A-2) To start the manual load, ÄKTA pure25 (GE Healthcare, Little Chalfont, Buckinghamshire, UK) was prepared. The column (Superdex 75 increase 10/300 GL) was cleaned using pump wash with 30mL filtered ddH_2O for about 1hr (flow rate 0.5mL/min), followed by equilibrating the 'column' with 60mL of filtered elution buffer (100mM NaCl, 20mM Tris-HCl, pH 8.0). Then, 1mL of individual native fibroin complex was manually loaded into the column.

A-3) The eluted fractions of each protein were collected corresponding to the time points of the chromatogram. The collected fractions of each protein were run on the SDS-PAGE and stained with Coomassie blue.

(B) partial native-partial native fibroin complex (partially folded form)

B-1) 1mL of each partial native fibroin proteins complex, H-chain alone, H-chain+P25, H-chain+FIBL, FIBL+P25, and 3 subunits were placed in 5M urea, 20mM Tris-HCl, pH 8.0 (with the concentration ≥400µg/mL). The six samples were individually loaded into ÄKTA pure25 and experimented on as described previously in the pattern (A-1-A-3, native-native fibroin complexes formation).

2.4. Determination in vitro silk fibroin complexes formation or protein-protein interactions using pull-down assay

(A) native-native fibroin complex formation (fully folded form)

2.4.1) 50µL of Ni-NTA agarose beads were added into seven separate Eppendorf tubes to immobilise his tagged-recombinant fibroin proteins.

2.4.2) Seven native fibroin complexes were placed on ice, in the Eppendorf tubes (200µL:200 µL V: V), H-chain alone, P25 alone, FIBL alone, H-chain+P25, H-chain+FIBL, P25+FIBL, and 3 subunits, whereas each subunit had a concentration of about 0.300mg/mL. Each tube was incubated on ice for 30 mins.

2.4.3) Incubated fibroin complexes were loaded onto separate Ni-NTA agarose beads and spun down at 400 rpm, 15 mins, at 4°C.

2.4.4) The flow through each complex was collected, all fibroin complexes were then washed

with the 200µL wash buffer (40mM Tris-HCl, pH 8.0, 300Mm NaCl and 40mM Imidazole), and finally each complex was eluted with 50µL elution buffer.

2.4.5) All eluted fibroin complexes were detected on 4-20% reducing SDS-PAGE gels and another gel was subjected to western blotting.

CHAPTER III

Observation on the Potential of Recombinant Fibroin Light Chain (*B. mori*) Production in Bacteria System

ABSTRACT

This chapter focuses on the attempt to produce fibroin light chain (L-chain) /FIBL (*Bombyx mori*) in *Escherichia coli***.** The FIBL was constructed with TEV+His6-tagged at the C-terminus in pET28a and transformed into DH5α^{TM}. Verification of the sequence of the expression construct was confirmed by DNA sequencing and expressed in two different groups of *E.coli* strains that supported un-enhanced and enhanced disulfide bond formation. Small-scale expression was optimised using different growth conditions (IPTG concentration, induction temperatures). The findings revealed that the enhanced disulfide bond formation strain, SHuffle® T7 was the most promising strain to produce FIBL $_{TEV+His6\text{-}tagged}$ when induced with 0.5 mM IPTG at 25ºC overnight. However, the expression product accumulated in the inclusion bodies so required solubilisation using 7M urea and refolding in Ni-NTA column to achieve the active state. The protocol used in small-scale production was scalable to produce larger volumes with the identical purity of FIBL $_{TEV+His6\text{-}tagged}$ analysed by SDS-PAGE and western-blotting (~29.45-32 kDa). The stability of FIBL $_{TEV+His6\text{-}tagged}$ showed good quality after seven days when maintained at 4ºC.

The identity of FIBL $_{TEV+His6\text{-}tagged}$ appeared to closely resemble the FIBL produced in a native organism (*B. mori*) when analysed by nLC-ESI MSMS. The most remarkable result was the revelation of the secondary structure by CD analysis in FIBL $_{TEV+His6\text{-}tagged}$ which contains 25.9% Helix1 (α-helix), 14.9% Helix2 (irregular α-helix**),** 8.3% Strand 1(β-sheet), 6.1% Strand2 (anti-parallel), 18.1% Turns (β-turns), and 2% Unordered (others).

Keywords: L-chain/ FIBL, bacteria system, solubilisation, refolding, disulfide bond

INTRODUCTION

Silk fibroin (*B. mori*) is a large, secreted, complex, fibrous protein comprising three different components (H-chain, L-chain (FIBL), and P25). These three fibroin proteins are encoded in different locations i.e. Two fibroin gens; H-chain and FIBL are located in the same fibroin gene but on a different site of chromosomes, H-chain at 25 and FIBL at 14 (Hyodo et al.,1984; Couble et al., 1985) while the P25 protein is derived from the fibrohexamerin gene (*Fhx*) on chromosome 2 (Lu et al., 2020)**.** However, these fibroin genes are expressed co-ordinately (Hyodo et al., 1984; Kimura et al., 1984) and each protein is synthesised individually in PSG (Hyodo, et al., 1984). The FIBL gene was cloned by the Kikuchi et al. (research group) and the sequence was characterised, revealing seven exons and six large introns (Kikuchi et al., 1992). However, the first introns occupies about 60% of the whole gene whereas the other introns together occupy approximately 31%, thus showing that 91% of the gene contains non-coding DNA (Barbosa, 2009). In addition, the multiple sequence alignments of original FIBL (*B. mori*) from two Chinese and Japanese hybrid strains revealed that 51.6% of its introns are conserved regions (Barbosa et al., 2008). The FIBL gene possesses 14,663 nucleotides with an open reading frame of 786 nucleotides encoding a protein with lower molecular mass of ~25-26 kDa from 262 amino acid residue (Yamaguchi et al., 1989, Choi et al., 2002). It is important to note that previous studies revealed that 5' flanking regions of FIBL and H-chain genes have common sequence elements indicating a mechanism for coordinated expression of these two genes (Kikuchi et al., 1992). FIBL protein has a specific development in strict tissue in PSG (as an eukaryotic system) of silkworm larvae during the growth stage, whereas molting stages are inhibited. The FIBL protein is highly expressed especially at the later stage of the 5th instar larvae (Huang, et al. 2021). FIBL protein is the smaller subunit ≈15% if compared to H-chain ≈85% from the total fibroin protein (Liu et al.,2017), even though they have the same molecular ratio, 6:6 (Inoue et al.,2000). FIBL protein contains a more undifferentiated amino acid composition with non-repetitive domains, which allows it to have more hydrophilic properties and be relatively elastic with less or without crystallinity (Zafar et al., 2015). Previous studies demonstrated that the lower molecular weight fibroin molecule 'FIBL' can be secreted into the silk gland lumen in the monomeric form, but not the huge molecule of H-chain which is needed to form the complex before secretion and transportation (Wang et al., 2015).

The FIBL protein is a small subunit crucial for the formation of disulfide linkage (intermolecular disulfide bonds) to the large blocky 'H-chain' heterodimeric molecule (FIBL-H-chain) in the silk fibroin elementary unit (Yamaguchi et al., 1998; Ye et al., 2022). The linkage of a single disulfide bond occurs between a Cys residue at the position 172 in the FIBL protein (18 residues of its signal peptide) and Cys residue at the 20th position from the C-terminus domain of H-chain (Cys-c20) (Tanaka et al., 1999). The role of this relationship between H-chain-FIBL of Lepidoptera is believed to have been maintained throughout 290 million years (Stewart et al., 2022). The six sets of covalent interactions between H-chain protein and FIBL protein (between two cysteines linkage) and one non-covalent attachment with P25 as massive fibroin elementary units are thought to be a biological mechanism for the effective intracellular transport and spinning out for the silk fibre (Kikuchi et al., 1992). That means FIBL and P25 play an important role in accommodating a huge complex of H-chain secretion and transport through the cells, compartments, and spinning process (Takei et al., 1987). Furthermore, the interaction of two

fibroin molecules, H-chain protein and FIBL protein are believed to connect the two main fibroin structures together, including the crystalline (H-chain) and non-crystalline (FIBL) segments (Rahaiee et al., 2019).

The regulation of the FIBL gene expression has been relatively less studied and is poorly characterised (Takei et al., 1985) due to these three silk fibroin genes. It is known that exogenous protein fused with FIBL and GFP or H-chain GFP can be produced in the PSG and eventually incorporated into cocoons (Saotome et al., 2015). There are some previous studies of recombinant proteins, in which the FIBL gene (*B. mori*) was cloned and the GFP gene inserted into exon 7. The chimeric FIBL chain-GFP gene was transferred into the polyhedrin region of *Autographa californica* nucleopolyhedrovirus (*Ac*NPV), and the recombinant virus was used as the targeting vector (the FIBL GFP gene to the FIBL region of the silkworm genome). Female silk moths were used for transfection with the recombinant virus and mated with the normal male silk moths. Genomic DNA from the progenies was screened for the desired targeting process. The findings revealed that the chimeric gene had fused into the FIBL gene and had stable transmission through generations. The chimeric gene was expressed in PFG, and the gene product was spun together in the part of silk cocoons (Yamao et al., 1999). In addition to that, a transgenic silkworm carrying a recombinant FIBL (*B. mori*) fused to GFP (68.7kDa) was successfully hybridized with antimicrobial peptide cecropin B (CEC B) by use of a gene targeting technique (i.e., homologous recombination). This approach indicated that the resulting, cecropin-decorated silk cocoons have anti-bacterial properties (Liu et al., 2015).

More recently, there has been only one report that shows the specific application of original FIBL extraction and isolation from *B. mori* cocoons and consequent conjugation to 2-isocyanatoethyl methacrylate (IEM) associated with (UV-LC precursor) as a resist material which can be used for a potential cellular substrate of the foetal neural stem cell culture (Li et al., 2017).

Nowadays, there have been no reports from research groups on the FIBL protein structure. However, based on FIBL primary structure and its amino acid residues, FIBL protein structure can be predicted using AlphaFold Monomer V2.0 **(Figure 17A)**. In agreement with this, Steward et. al. (2022) reported the full sequence alignment of FIBL and highlighted the number of helical structures and the 3D structure shown in **Figure 17B**. This project of carrying out recombinant FIBL production with a simple host, using suitable engineered modification and effective processing methods, resulted in favourable properties in structure and function, thus providing further aspects for researchers to look into.

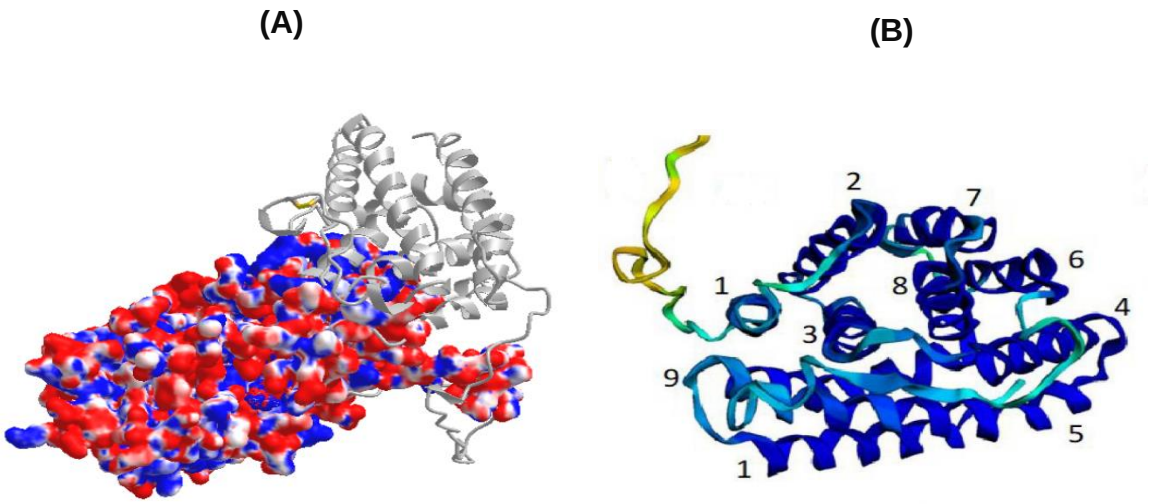


Figure 17. (A), the prediction of the 3D structure of the fibroin light chain (FIBL) protein (*B. mori*) when made by AlphaFold Monomer V2.0 producing a per-residue confide score from P21828.FIBL_BOMMO 3D structure database (UniProt, Website 2022). This AlphaFold database was analysed and modified into the 3D structure using iCn3D: web-based 3D structure viewer (NCBI, website 2022). **(B)**, the nine α-helices of FIBL are numbered and 3D structure of the FIBL was predicted using ColabFold by pLDDT (Steward et al., 2022).

MATERIALS AND METHODS

The main experiments of FIBL production in bacteria were described in chapter II.

RESULTS

1) Vector construction of FIBL gene (Molecular Engineering)

The FIBL Gblocks were successfully amplified by adding EcoRI, HindIII, and TEV+His6-tagged using PCR and used for a classical cloning technique with pET28a.The insertion (FIBL) and the vector (pET28) were individually cut using two restriction enzymes digestion, firstly with EcoRI and secondly with HindIII following the general protocol. The DNA concentration and the DNA quality was subsequently determined by NanoDrop on 1% agarose gels as shown in **Figure 18 (A-B).** Nucleotide sequences of pET28a are 5,369bp whereas the FIBL gene including the TEV+His6-tagged are 752bp. Both the digested FIBL and pET28a DNAs were used for the ligation reaction and transformed into the DH5α competent cells and plated on agar plates containing kanamycin and incubated in the 37^{o}C incubator overnight. The following day, the resultant single colonies were randomly picked and grown on the 10 mL LB supplemented with kanamycin overnight. The next day, the plasmid DNA of each cell culture was extracted, and its primary ligation was rechecked by double digestion enzymes, EcoRI and HindIII. If the agarose gel contained both DNA it meant the cloning was completed and these plasmid DNAs were used for the validation of DNA sequencing **Figure 19.**

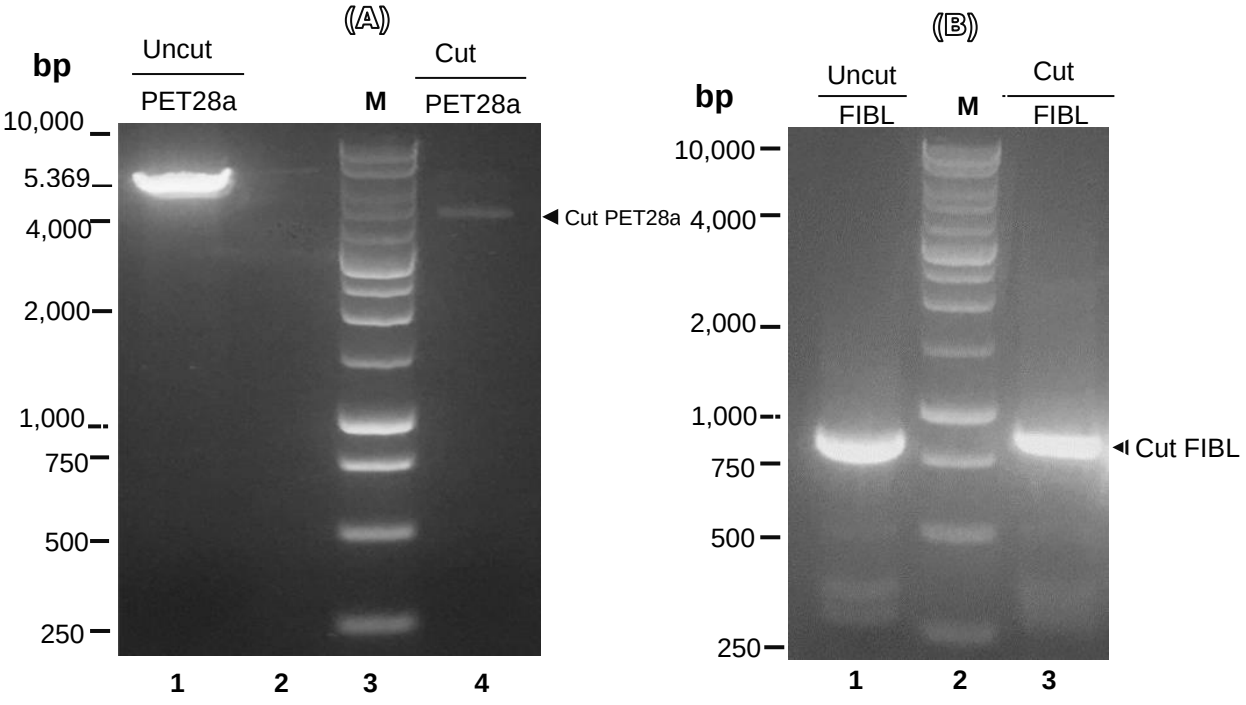


Figure 18. The DNA profile of vector construction between FIBL and pET28a was achieved as a bacteria system. Firstly, amplification of the FIBL Gblocks containing TEV+His6-tagged at the C-terminal (787 nucleotides) used PCR. The purified FIBL DNA and pET28a DNA were consequently used for digestion with EcoRI and HindIII and both reactions of the base pairs of their cut DNAs were measured to assess the DNA concentration and their quality was checked using 1 % agarose gels **(A-B**) and then both digested DNAs were used for Further ligation reaction**.** **Figure 18-A** shows DNA bands of vector, uncut pET28a (Lane 1, control), and cut pET28a, Lane 4 (digested with EcoRI and HindIII). **Figure 18-B** shows DNA bands of gene target, uncut FIBL (Lane 1, control), and Lane3 cut FIBL (digested with EcoRI and HindIII).

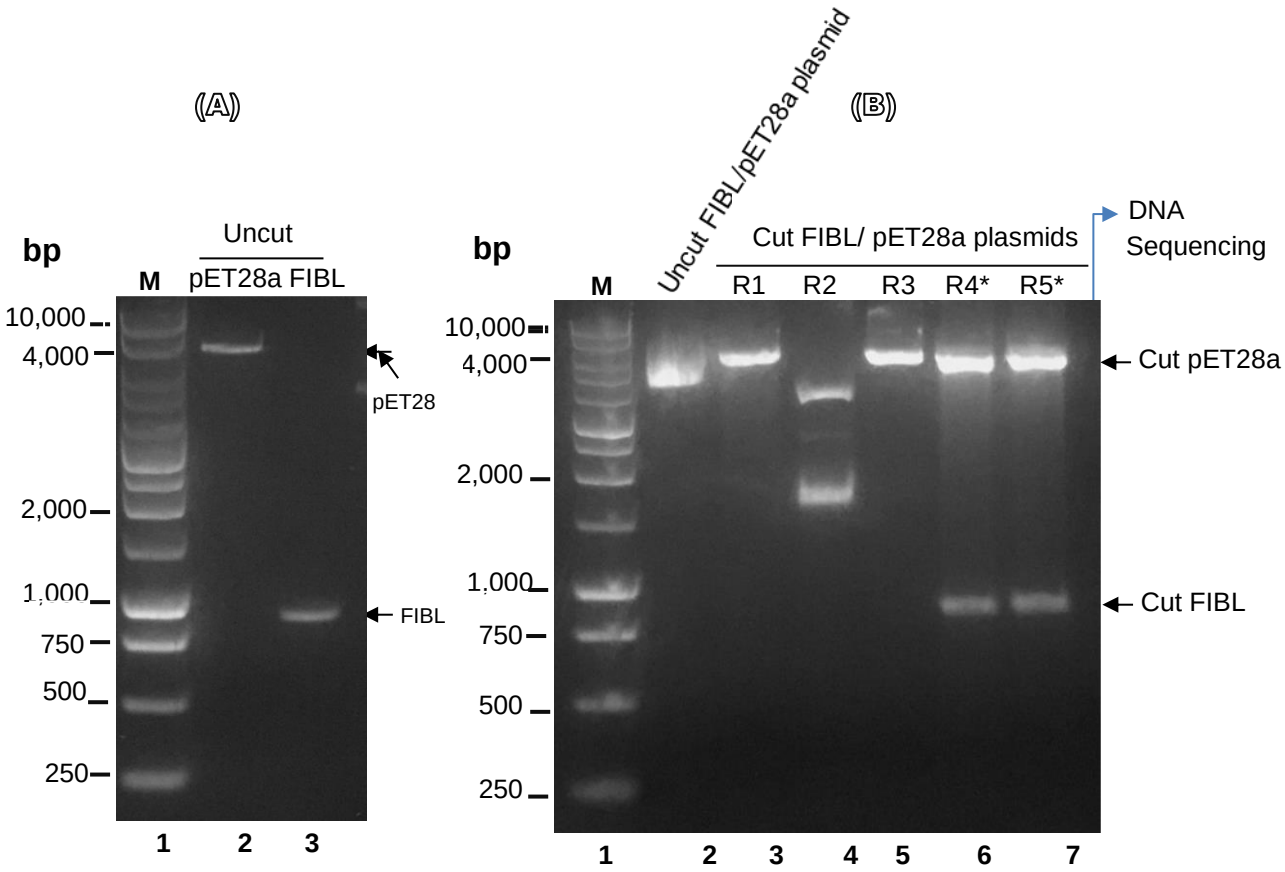


Figure 19. The ligation of the FIBL and pET28a DNAs were transformed into DH5 cells and plated on agar plates containing kanamycin overnight. The next day, in order to verify plasmid DNA after ligation**,** five single colonies were picked and grown overnight in LB at 37ºC with the shaker and subsequently used for DNA extraction using miniPrep. The ligation of these five plasmid DNA was rechecked with digestion enzymes, EcoRI and HindIII, and determined on the 1% agarose gel. An agarose gel **(A)** shows the un-cut DNAs of the vector pET28a (Lane 2) and the FIBL (Lane 3). Among five digested plasmid DNA from the gel, R1, R2, R3, R4, and R5, only two colonies (R4 & R5) containing two bands and right base pairs of both FIBL and pET28a were sent for DNA sequencing **(B).**

RESULTS 2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL) in two groups bacterial strain, cultured at 15, 20, 25, and 28ºC.

Two groups of selected bacterial strains were used in this Chapter; Group 1 (un-enhanced disulfide bond formation), including three bacterial cells BL21 (DE3), C41 (DE3), Rosetta™(DE3) (Novagen). FIBL protein from Group 2 containing one strain, SHuffle® T7 (enhanced disulfide bond formation) was used as a comparison.

(Figure 20-23), expression screening of FIBL was cultured on a small volume (10 mL) of LB media. Group 1 bacterial strains containing FIBL were grown with variations in two culture conditions: 1) temperatures, 15, 20, 25, and 28ºC and 2) IPTG concentrations, 0.1, 0.5 and 1.0 mM. **(Figure 20)** FIBL protein production in three strains BL21, C41 (DE3) and Rosetta™ (DE3)pLysS in all concentrations produced poor expressions at 15ºC. However, **(Figure 20**) the FIBL protein expression improved at 20ºC and there was greater improvement at 25ºC for the two strains, BL21 and Rosetta™ (DE3)pLysS in all three IPTG concentrations. **(Figure 23)** It is interesting that the expression for these two strains in all IPTG inducer concentrations was poor at 28ºC. On the other hand, **(Figure21 – Figure 22)** C41 produced poor expressions at 20-25ºC, but in all IPTG concentrations were surprisingly good at 28ºC **(Figure 23)**. In addition, both strains BL21 and Rosetta™ (DE3)pLysS in all three IPTG inducer concentrations showed good expression at 37ºC, except for C41 (Figure not shown). **(Figure 24 & 25)**. In the screening experiment with a Group 2 bacteria strain, SHuffle® T7 was used to express FIBL protein with the conditions varying a little from Group 1, the two main factors being 1) IPTG concentrations, 0.1, 0.5, 1.0, 2.0 mM (the additional 2.0 mM IPTG making one higher IPTG concentration than Group 1), and 2) temperatures, 15, 20, 25, and 28ºC, respectively. The findings showed that the FIBL protein had good expression in all inducer conditions, temperatures at 15, 20, 25, and 28ºC combined in all IPTG concentrations (0.1, 0.5, 1.0 and 2.0 mM). The good expression could be seen as the intensity bands of FIBL protein at around 29.45 kDa using (SDS-PAGE & western blotting). Most FIBL/ SHuffle® T7 cells (as controls without IPTG induction) expressed at different induction temperatures (15, 20, 25ºC) were obviously presented in the FIBL protein bands, except at 28ºC. It appears that this plasmid can leak and the target bands shown resemble IPTG induction. There was not much difference in the expression of FIBL protein in all four inducer concentrations of IPTG, 0.1, 0.5, 1.0, 2.0 mM, as well as three inducer temperatures (20, 25 & 28ºC) showed a slightly greater expression than those at 15ºC). In additional screening experiments, FIBL was shown to be expressed well in SHuffle® T7 cells for all three IPTG concentrations at 37ºC in all IPTG inducer conditions, but not in C41 (Figure not shown). The summary of screening expression of FIBL in two groups of bacterial strains with different temperature inductions are shown as **Table 2**. This table shows more clearly that FIBL $_{TEV+His6x}$ protein could be expressed well in SHuffle® T7 cells and cultured in LB media in all conditions compared to other selected strains.

RESULTS 2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL) in Group-1 bacterial strain, cultured at 15, 20, 25, and 28 ºC

15 ºC

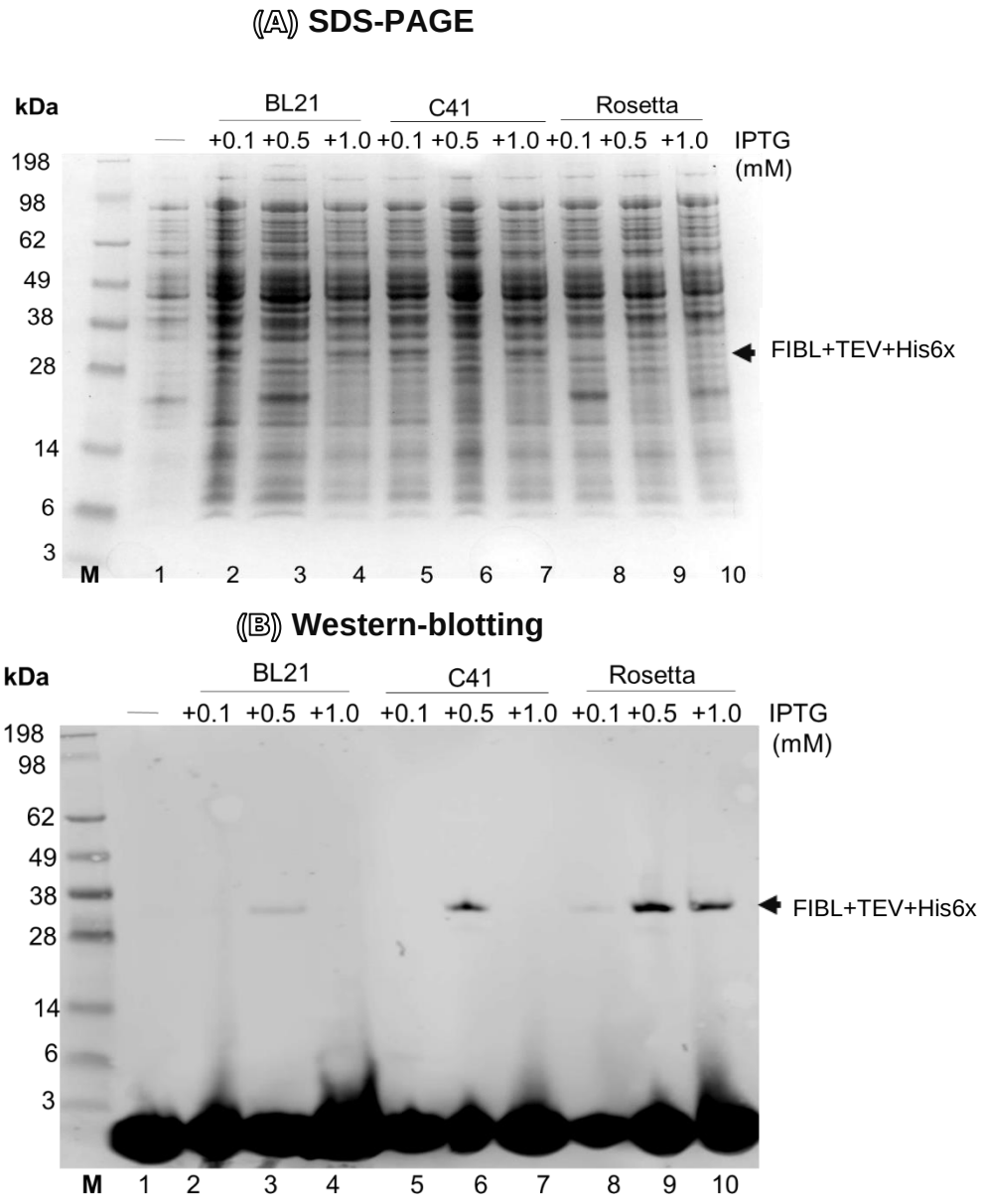


Figure 20. The small-scale induction of FIBL protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ (DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) at 15ºC was performed in 10 mL LB supplemented with kanamycin and the expression maintained overnight on the shaker. To do the expression, firstly a single colony of each strain was picked for inoculation overnight at 37ºC. Next morning, the sample of BL21 inoculation was collected before IPTG, un-induced (-) was added as a control. Then 0.1, 0.5, and 1.0 mM IPTG was added to three tubes of the cell cultures to induce FIBL expression and maintained overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12 %) gels and western-blotting. The FIBL $_{TEV+His6x}$ protein bands were ~29.45-32 kDa.

RESULTS 2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL) in Group-1 bacterial strain, cultured at 15, 20, 25, and 28 ºC (Continued)

20ºC

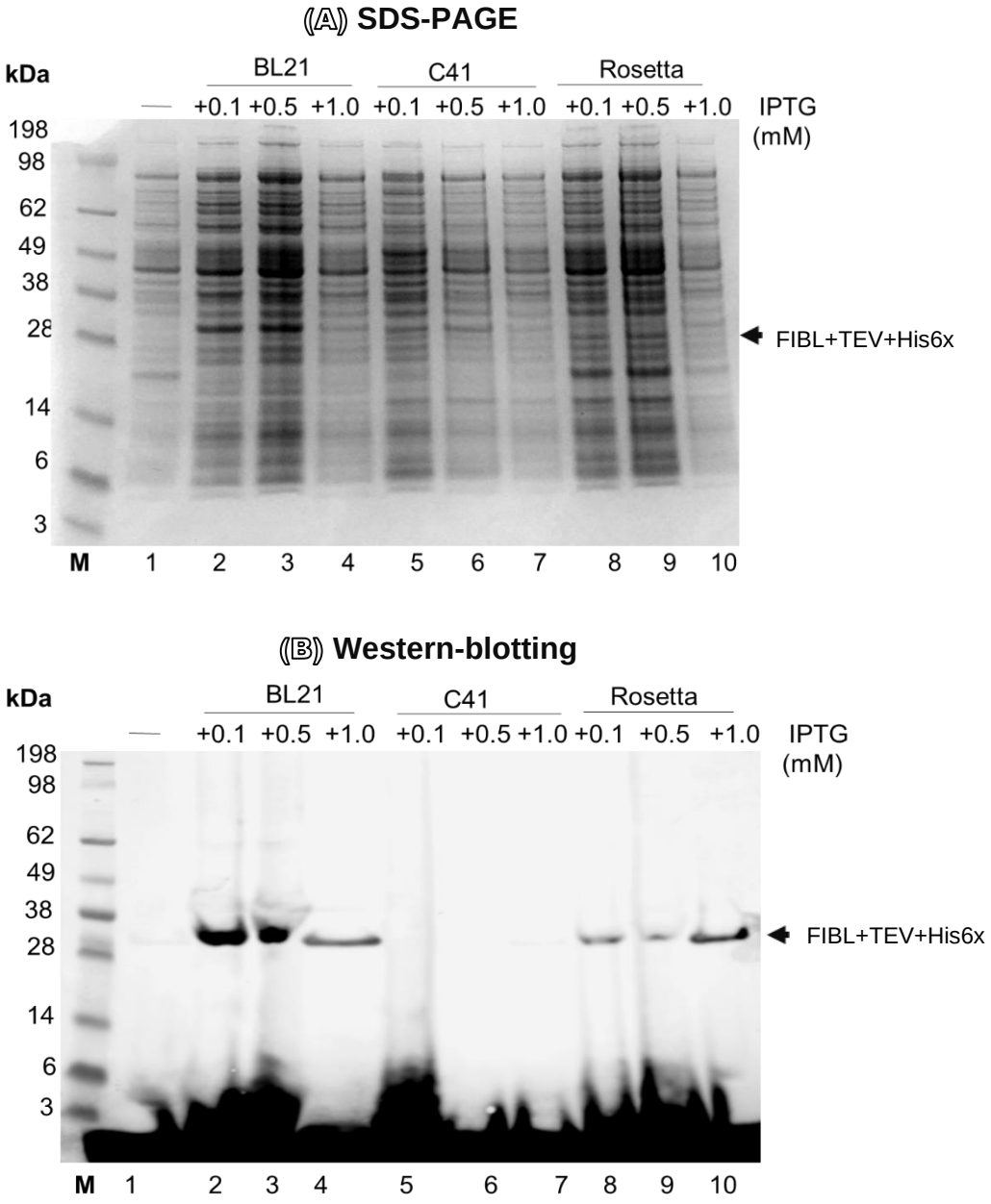


Figure 21. The small-scale induction of FIBL protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ **(**DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) at 20ºC was performed in 10 mL LB supplemented with kanamycin and the expression maintained overnight on the shaker. To do the expression, firstly a single colony of each strain was picked for inoculation overnight at 37ºC. Next morning, the sample of BL21 inoculation was collected before IPTG, un-induced (-) was added as a control. Then 0.1, 0.5, and 1.0 mM IPTG was added to three tubes of the cell cultures to induce FIBL expression and maintained overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12 %) gels and western-blotting. The FIBL $_{TEV+His6x}$ protein bands were ~29.45-32 kDa.

RESULTS 2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL) in Group 1 bacterial strain, cultured at 15, 20, 25, and 28 ºC (Continued)

25ºC

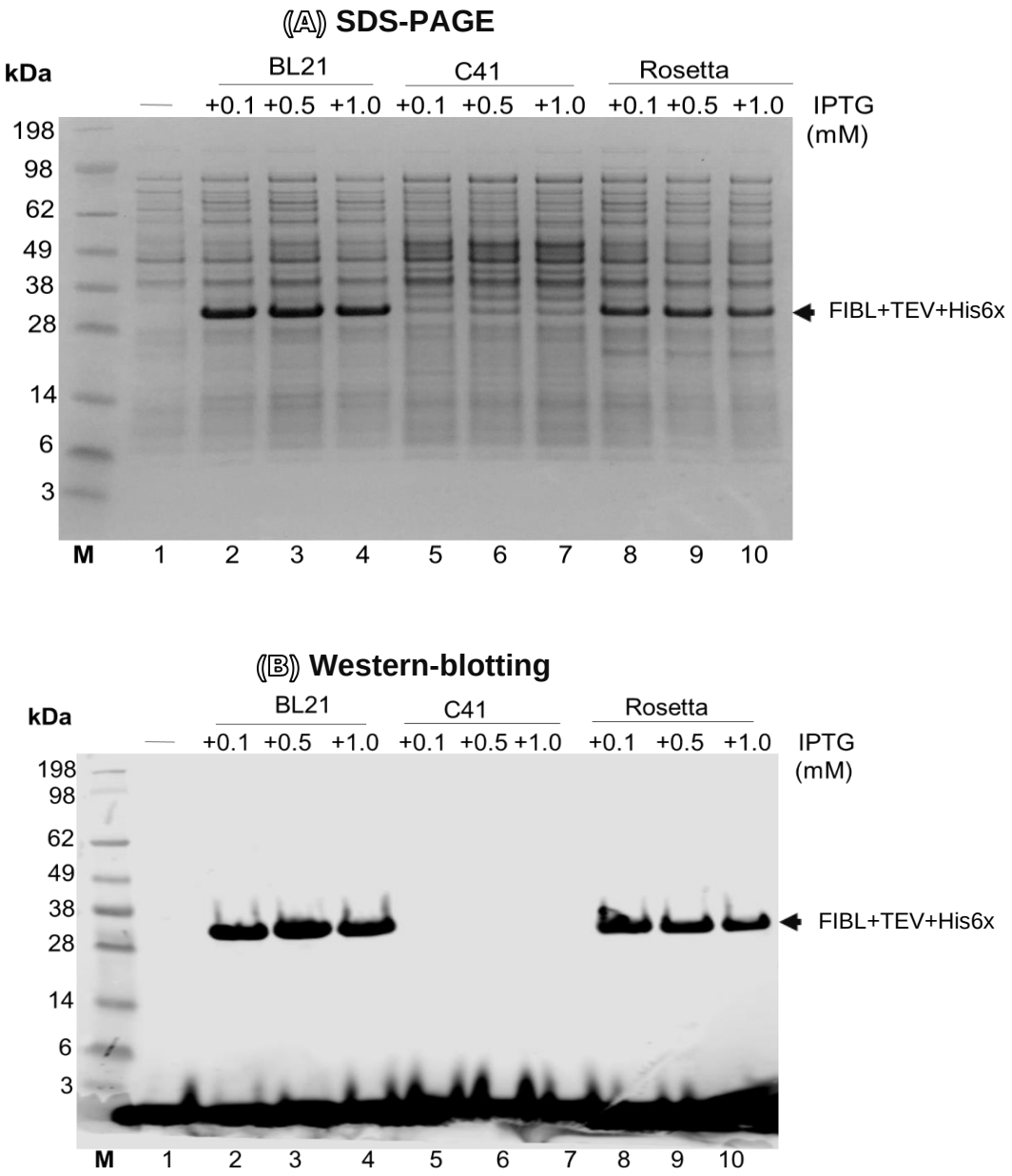


Figure 22. The small-scale induction of FIBL protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ (DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) at 25ºC was performed in 10 mL LB supplemented with kanamycin and the expression maintained overnight on the shaker. To do the expression, firstly a single colony of each strain was picked for inoculation overnight at 37ºC. Next morning, the sample of BL21 inoculation was collected before IPTG, un-induced (-) was added as a control. Then 0.1, 0.5, and 1.0 mM IPTG was added to three tubes of the cell cultures to induce FIBL expression and maintained overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12%) gels and western-blotting. The FIBL $_{TEV+His6x}$ protein bands were ~29.45-32 kDa.

RESULTS 2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL) in Group 1 bacterial strain, cultured at 15, 20, 25, and 28ºC (Continued)

28ºC

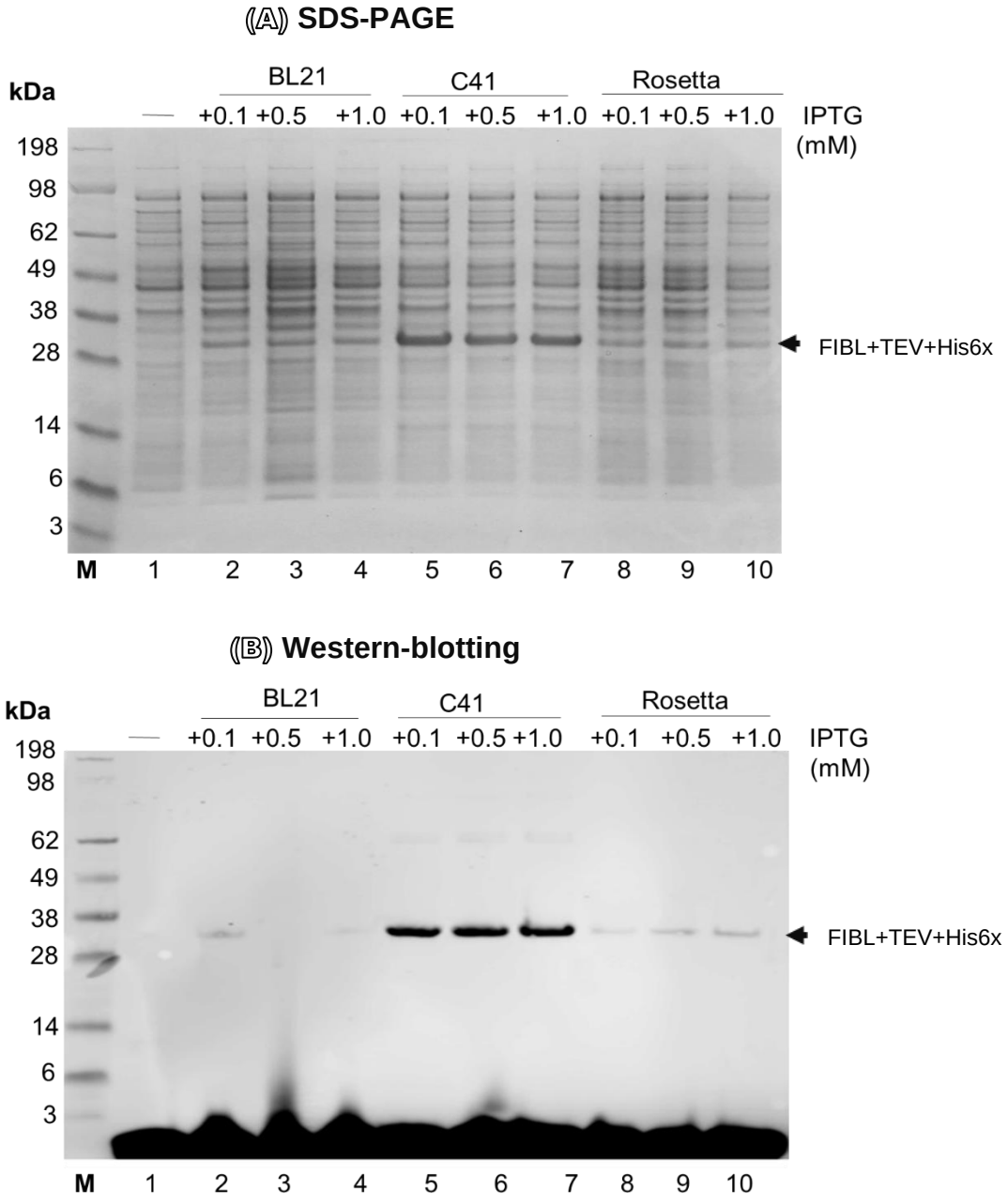


Figure 23. The small-scale induction of FIBL protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ (DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) at 28ºC was performed in 10 mL LB supplemented with kanamycin and the expression maintained overnight on the shaker. To do the expression, firstly a single colony of each strain was picked for inoculation overnight at 37ºC. Next morning, the sample of BL21 inoculation was collected before IPTG, un-induced (-) was added as a control. Then 0.1, 0.5, and 1.0 mM IPTG was added to three tubes of the cell cultures to induce FIBL expression and maintained overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12 %) gels and western-blotting. The FIBL protein bands were ~29.45-32 kDa.

RESULTS 2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL) in Group 2 a bacteria strain, cultured at 15, 20, 25, and 28ºC

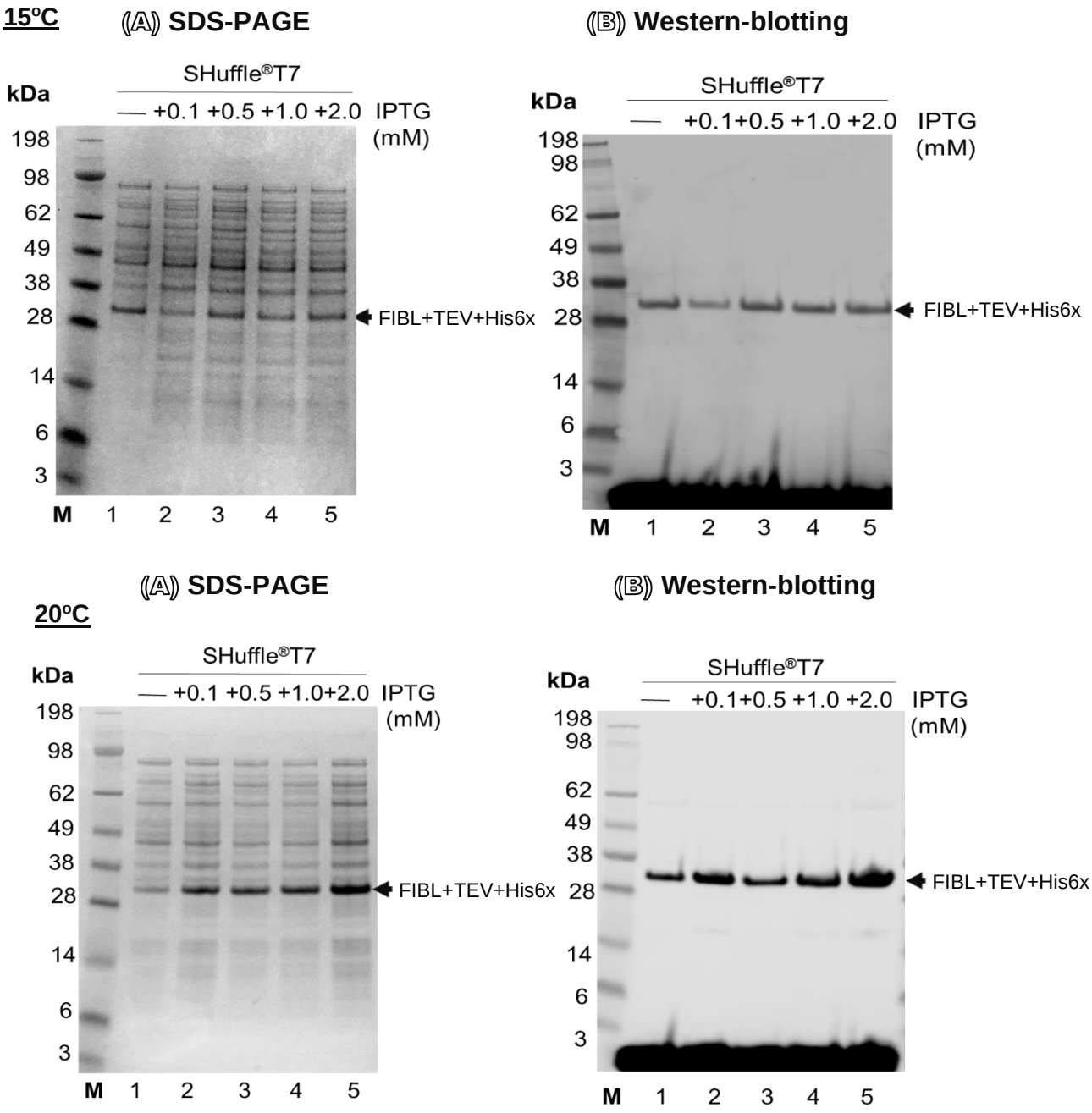
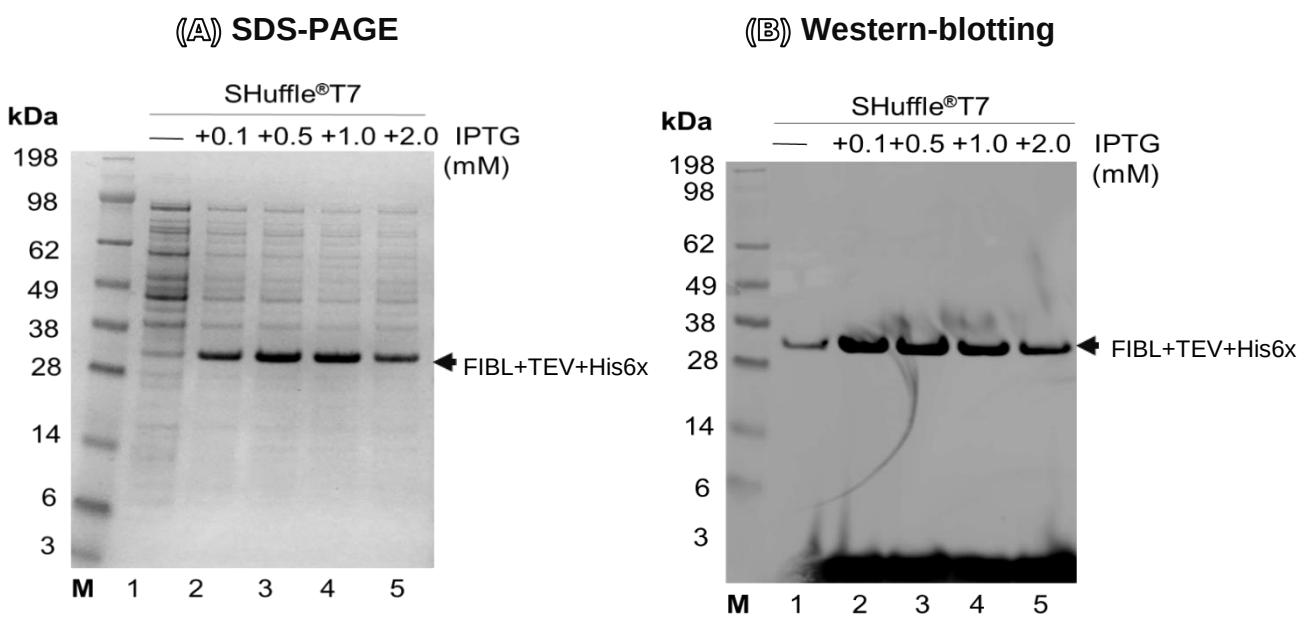


Figure 24. In the comparative small-scale induction of FIBL protein in Group 2, a bacteria strain (SHuffle® T7), with four IPTG concentrations (0.1, 0.5,1.0, and 2.0 mM) at two temperatures 15 or 20ºC, a single colony was inoculated in 10mL LB supplemented with kanamycin and grown overnight on the shaker at 37ºC. Next morning, a tube of the sample SHuffle® T7/25 inoculation was collected for analysis before IPTG, un-induced (-) was added as a control. Consequently, 0.1,0.5, 1.0 & 2.0 mM IPTG was separately added to four other tubes of the cell cultures to induce FIBL expression and maintained at 15 or 20ºC overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12 %) gels and western-blotting. The FIBL protein bands were ~29.45-32 kDa.

RESULTS 2) Screening expression and optimisation of FIBL protein (small-scale production 10 mL) in Group 2 a bacteria strain, cultured at 15, 20, 25, and 28ºC

25ºC



28ºC

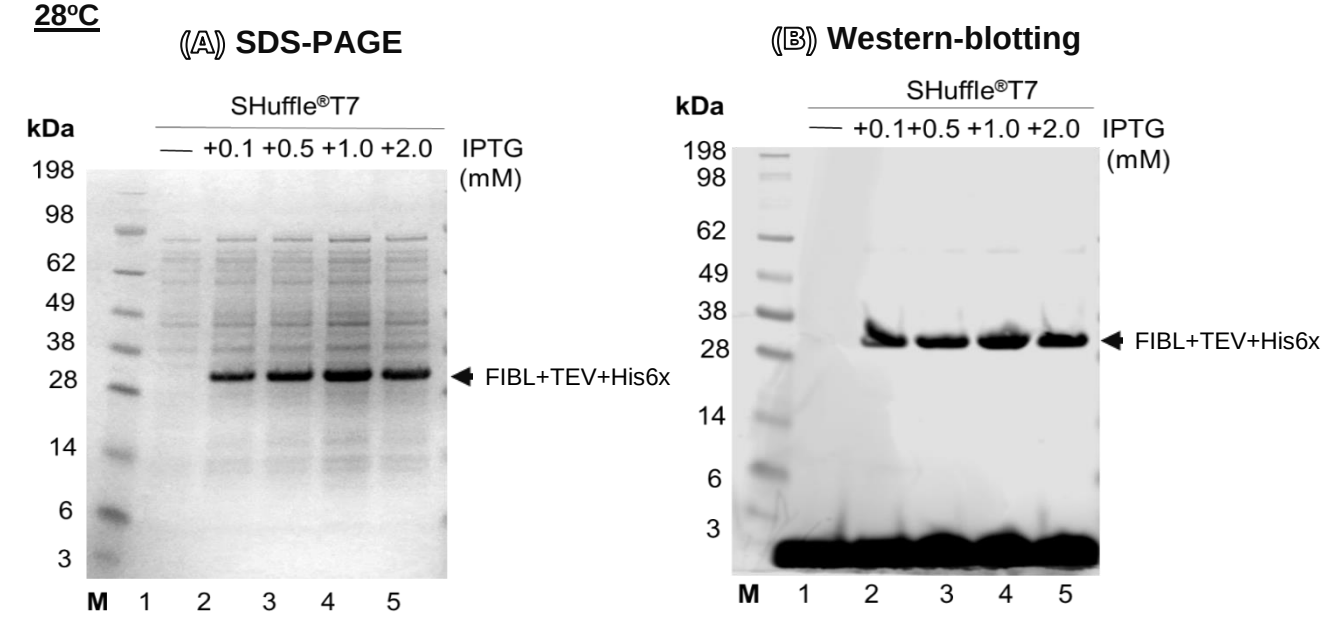


Figure 25. The comparative small-scale induction of FIBL protein in Group 2, a bacteria strain (SHuffle® T7), with four IPTG concentrations (0.1, 0.5,1.0, and 2.0 mM) at two temperatures 25 or 28ºC, a single colony was inoculated in 10mL LB supplemented with kanamycin and grown overnight on the shaker at 37ºC. Next morning, a tube of the sample SHuffle® T7/25 inoculation was collected for analysis before IPTG, un-induced (-) was added as a control. Consequently, 0.1,0.5, 1.0 & 2.0 mM IPTG was separately added to four other tubes of the cell cultures to induce FIBL expression and maintained at 25 or 28ºC overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12 %) gels and western-blotting. The FIBL $_{TEV+His6x}$ protein bands were ~29.45-32 kDa.

RESULTS 2) Summary on screening expression and optimisation of FIBL protein (small-scale production 10 mL) in Group 1 & Group 2 bacterial strains, cultured at 15, 20, 25, and 28 ºC and an additional temperature, 37ºC

Table 2. The summary of FIBL expression among two groups of different bacterial strains

	Temp/IPTG (mM)				Temp/IPTG(mM)				Temp/IPTG(mM)				Temp/IPTG(mM)				Temp/IPTG(mM)			
	15 ºC				20 ºC				25 ºC				28 ºC				37 ºC			
Strain	0.1	0.5	1.0	2.0	0.1	0.5	1.0	2.0	0.1	0.5	1.0	2.0	0.1	0.5	1.0	2.0	0.1	0.5	1.0	2.0
Group1 /non-enhanced capacity of disulfide bonds folding																				
BL21	0	0	/	-	//	//	/	-	///	///	///	-	0	0	0	-	///	///	///	-
C41	0	/	0	-	0	0	0	-	0	0	0	-	///	///	///	-	0	0	0	-
Rosetta	0	/	/	-	/	/	/	-	///	///	///	-	0	0	0	-	/	//	//	-
Group2 -enhanced capacity of disulfide bonds folding																				
SHuffle® T7	//	//	//	//	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///	///

Note of expression scale:

- =not determined
0 =Not expressed/ little expressed not evident
/ =Poor
// =Good
/// = Remarkable good expression (Strong)

RESULTS 3) FIBLprotein localisation, solubilsation, and its purification (250mL cell culture)

Briefly, the results of the previous screening expression of FIBL protein and production in comparative studies of two groups of bacterial strains Group 1) non-enhanced capacity of disulfide bonds folding i.e., BL21, C41 (DE3), Rosetta™ (DE3)pLysS and Group 2) enhanced capacity of disulfide bonds folded in cytoplasm, i.e., SHuffle® T7) were presented on SDS-PAGE & western blotting from the total cells disruption. The same growth conditions of 25ºC and IPTG induction concentration (0.5 mM) were selected for both groups of bacterial strains and maintained overnight in order to investigate the specific FIBL localisation. It was necessary to determine the FIBL protein expression in the host in order to monitor where the active protein existed at specific time points after induction.

3.1 FIBL localisation and solubility test in two bacterial strains from Group 1 & Group 2

(Figure 26) Regarding the screening expression of FIBL protein in three bacterial strains in Group 1, the strong intensity of the target band of FIBL (~29.45-32 kDa) was shown among variant induction growth of both bacterial strains, BL21 and Rosetta™ (DE3)pLysS. Therefore, the localisation and solubility of these two strains were subsequently used for growth determination on a larger scale (250 mL). The experimental findings showed that the induction cells of both strains had a good expression level. The harvested cells were disrupted using the FRENCH Press and centrifuged (15,000rpm). The FIBL protein of both strains (~29.45-32 kDa) were found to be located in the 15KP fraction (15,000 rpm pellet, IBs), but not in the 15KS (15,000 rpm supernatant fraction). However, this obtained 15KS supernatant consequently underwent further determination at the higher centrifugation of 42,000 rpm. The presence of FIBL protein bands were not found in either resulting fractions of bacterial strains, supernatant (42KS) and pellets (42KP).

(Figure 27) The localisation and solubility tests of FIBL protein expressed in Group 2 strain (SHuffle® T7) were carried out using the same process as used for the bacterial strains in Group 1. The findings clearly showed the FIBL protein was located in the 15KP fraction (15,000 rpm pellet, IBs) just as it was in the Group1 bacterial strains (BL21 and Rosetta™ (DE3)pLysS).

3.2 Comparison of non-dialysis & dialysis purification of solubilised FIBL in 7M urea, production in BL21 using Ni-NTA column (Group 1)

Figure 28 (A1-A2) The FIBL protein expressed in the best selected strain from the promising expression in Group 1, BL21. The obtained IBs pellets solubilised in 7M urea were appropriate for providing soluble FIBL protein. This resulting solubilsed FIBL protein was split into two portions with and without dialysis treatment before the purification using Ni-NTA column. The first portion of solubilised FIBL protein without the dialysis showed a single band of purified FIBL protein at ~29.45 kDa (SDS-PAGE) whereas the purified FIBL protein determined by western-blots showed monomer, dimer and trimer bands (~29.45, ~60, &~ 80 kDa). **Figure 28 (B1-B2)** The solubilised FIBL protein (portion two) was dialysed against HEPES buffer, pH 7.6 and purified. It was found that the purified FIBL protein obtained from dialysis showed only expected bands (~29.45-32 kDa) during the purification processes when compared to non-dialysis FIBL purification using both SDS-PAGE and western-blotting. However, the purification of FIBL protein in the 2nd portion (dialysis purification)

presented evidence of only monomeric form (~29.45 kDa) after determination by SDS-PAGE and western-blotting.

3.3 Solubilised FIBL in 7M urea and consecutive purification of FIBL protein using Ni-NTA column, production in SHuffle® T7 (Group 2)

Figure 29 (A) & (B) showed the consecutive solubilisation of fresh IBs, after undergoing production processes with FRENCH Press, and purification of FIBL protein in SHuffle® T7. The experimental findings once again confirmed that the FIBL protein located in the IBs pellets were the same as in previous experiments. These fresh IBs pellets were solubilised and were consequently refolded on the Ni-NTA column. The eluted purified FIBL protein was clearly shown in three forms (~29.45, ~62, &~ 80 kDa).

RESULTS 4) Scaled-up expression, solubilisation, and purification (500 mL/Flask) for FIBL protein production in the most promising bacteria strain, SHuffle® T7 (Group 2)

Figure 30 (A) & (B) At this point in our FIBL protein production efforts, only one strain of bacteria SHuffle® T7 was selected for use in further experiments. It was found that the FIBL protein production showed stable expression, localisation and production. The FIBL protein can be produced with standard IBs solubilisation, and purification with a single band (SDS-PAGE) and its three forms (western-blotting).

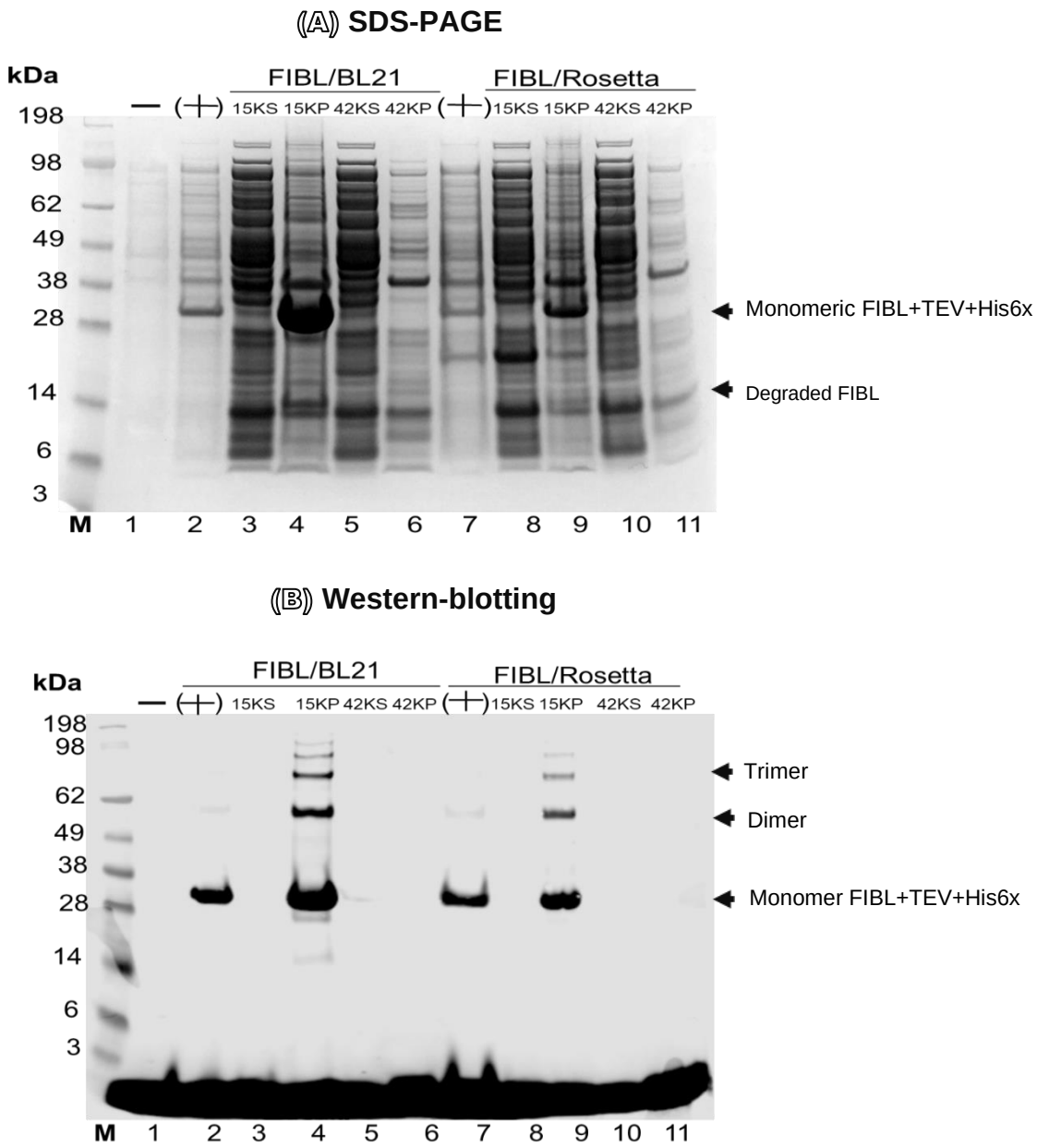


Figure 26. The comparative fractions of FIBL expression in two bacterial strains from Group 1, BL21 and Rosetta, cultured in LB from pre-induction of IPTG in cell (-), Un), post-induction ((+), In) with 0.5 mM IPTG containing Kanamycin at 25ºC were treated overnight on the shaker. The cells were harvested, 1XTBS and a tablet of protease inhibitor were added, and the cells were disrupted with the FRENCH Press. The disrupted cell lysate was then harvested and the 15,000 rpm (15K) supernatant and 15,000 rpm (15K) pellet (Sup 15 K, Pell 15K,) centrifuged at 15,000 rpm for 30 mins at 4ºC. The 15,000 rpm(K) supernatant was consequently centrifuged at 42,000 rpm (42K) for 50 mins at 4ºC and (Sup 142 K, Pell 142K) harvested to determine the FIBL protein localisation **Figure (A-B)**. These different obtained fractions show FIBL located in the 15KP (the pellets from 15,000 rpm as IBs). The expected band of monomeric FIBL $_{TEV+His6x}$ was 29.45~32 kDa.

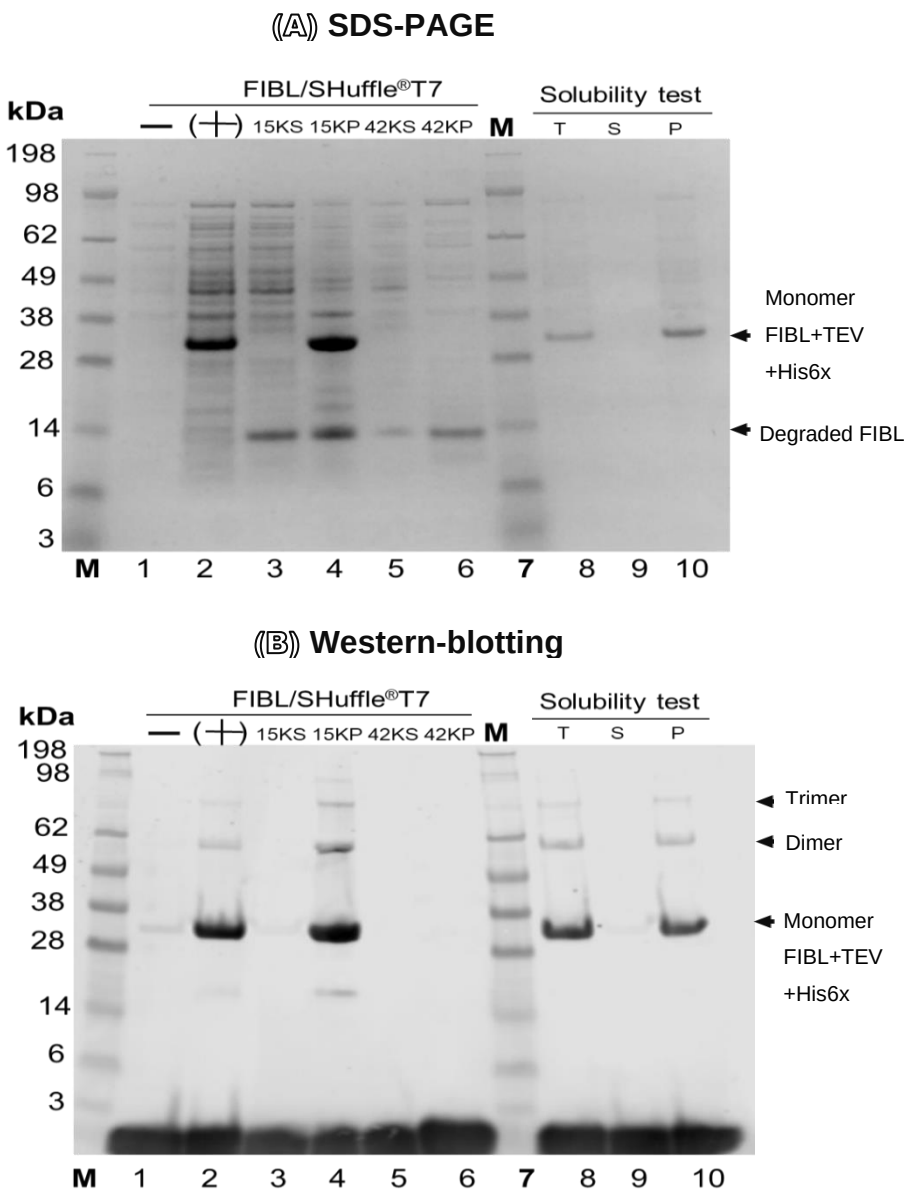


Figure 27. Fractions of FIBL expression in SHuffle® T7, cultured in LB from pre-induction of IPTG in cell (-), Un), post-induction ((+), In) with 0.5 mM IPTG containing Kanamycin at 25ºC were treated overnight on the shaker. The cells were harvested, 1XTBS and a tablet of protease inhibitor were added, and the cells were disrupted with the FRENCH Press. The disrupted cell lysate was then harvested and the 15,000 rpm (15K) supernatant and 15,000 rpm (15K) pellet (Sup 15 K, Pell 15K,) centrifuged at 15,000 rpm for 30 mins at 4 ºC. The 15K supernatant was consequently centrifuged at 42,000 rpm (42K) for 50 mins at 4ºC and (Sup 142 K, Pell 142K) harvested to determine the FIBL protein localization. Figure **(A-B)**. In another test, the induced cells ((+), In) were collected (the cells spun down at 7,000 rpm (K) for 5 mins), 1XTBS added, then the cells re-suspended, disrupted by syringe and eventually the total cell lysate (T) obtained. This total cell lysate was subsequently centrifuged again, and the supernatant (S) and pellets (P) collected. The solubility test showed FIBL in SHuffle® T7 located in the cell pellet determined by SDS-PAGE & western blotting. The expected band of FIBL $_{TEV+His6x}$ was 29.45~32 kDa.

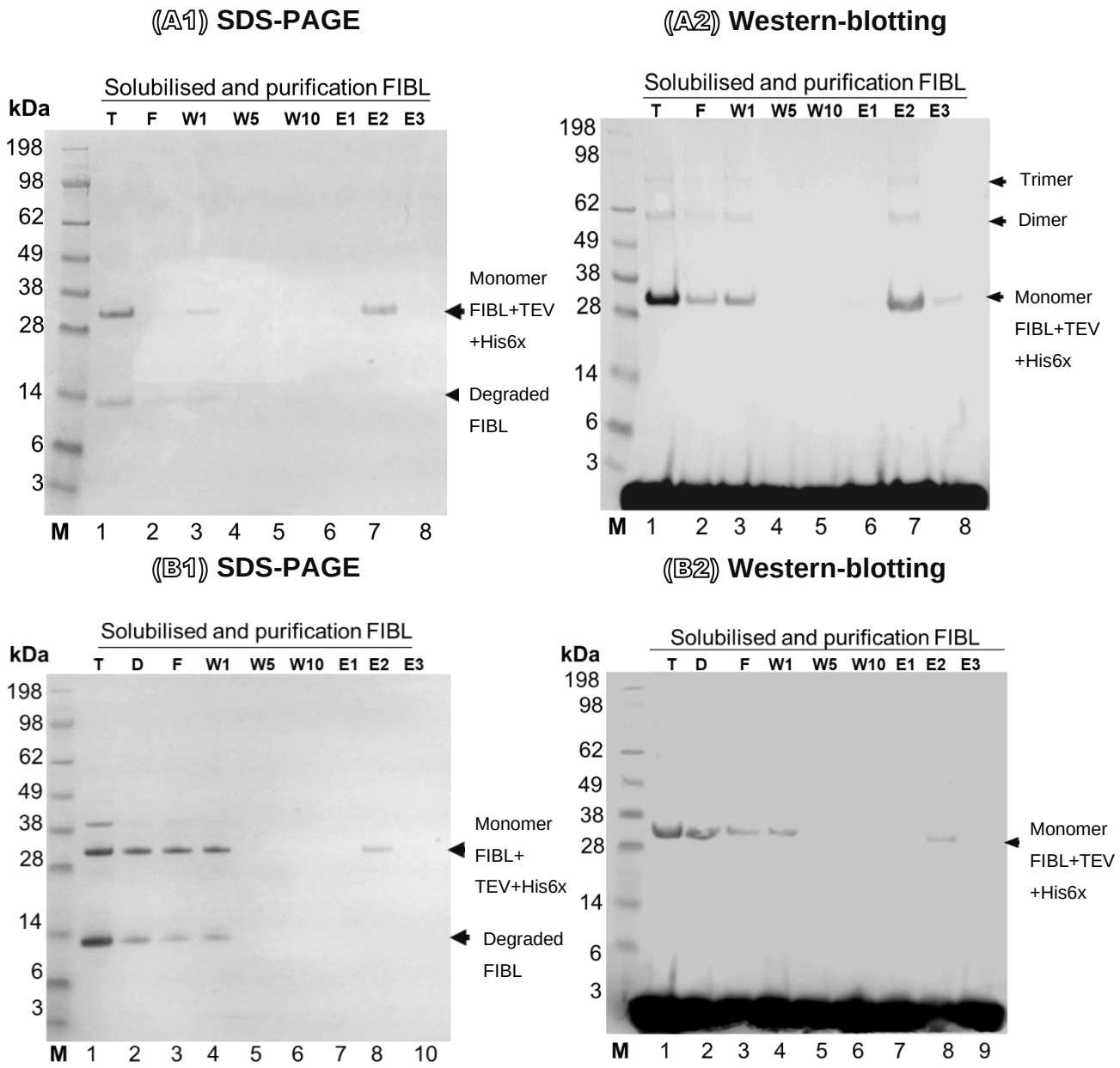


Figure 28. FIBL $_{TEV+His6\text{-}tagged}$ protein expressed in BL21 cells showed the profile of the localisation in the IBs pellet fractions from a small scale 10 mL culture. The supernatant of soluble FIBL (Total soluble (T)) after solubilising IBs pellets overnight in 7M urea at room temperature was collected by centrifugation 15,000 rpm at 4ºC and the remaining pellets discarded. The obtained 5mL of total soluble FIBL protein (T) was split into two portions for comparative purifications with dialysis and non-dialysis. **(A1-A2)** The first soluble 2mL portion of FIBL was immediately forwarded for passing through the Ni-NTA column (F), followed by having the wash buffer flow over the column 10 times (10mL). The three selected washes (W1, W5, and W10) were collected and the final purified FIBL was eluted by three mL elution buffer (E1, E2 and E3). **(B1-B2)** 3mL of the 2nd portion was dialysed with the 3L of HEPES buffer, pH 7.6 overnight in the cold room. The following day, the dialysed soluble FIBL **(D)** was purified using Ni-NTA column in the same way as the first portion. The purified FIBL $_{TEV+His6x}$ bands were determined to be ~ 29.45-32 kDa (Lane 8, SDS-PAGE & western blotting).

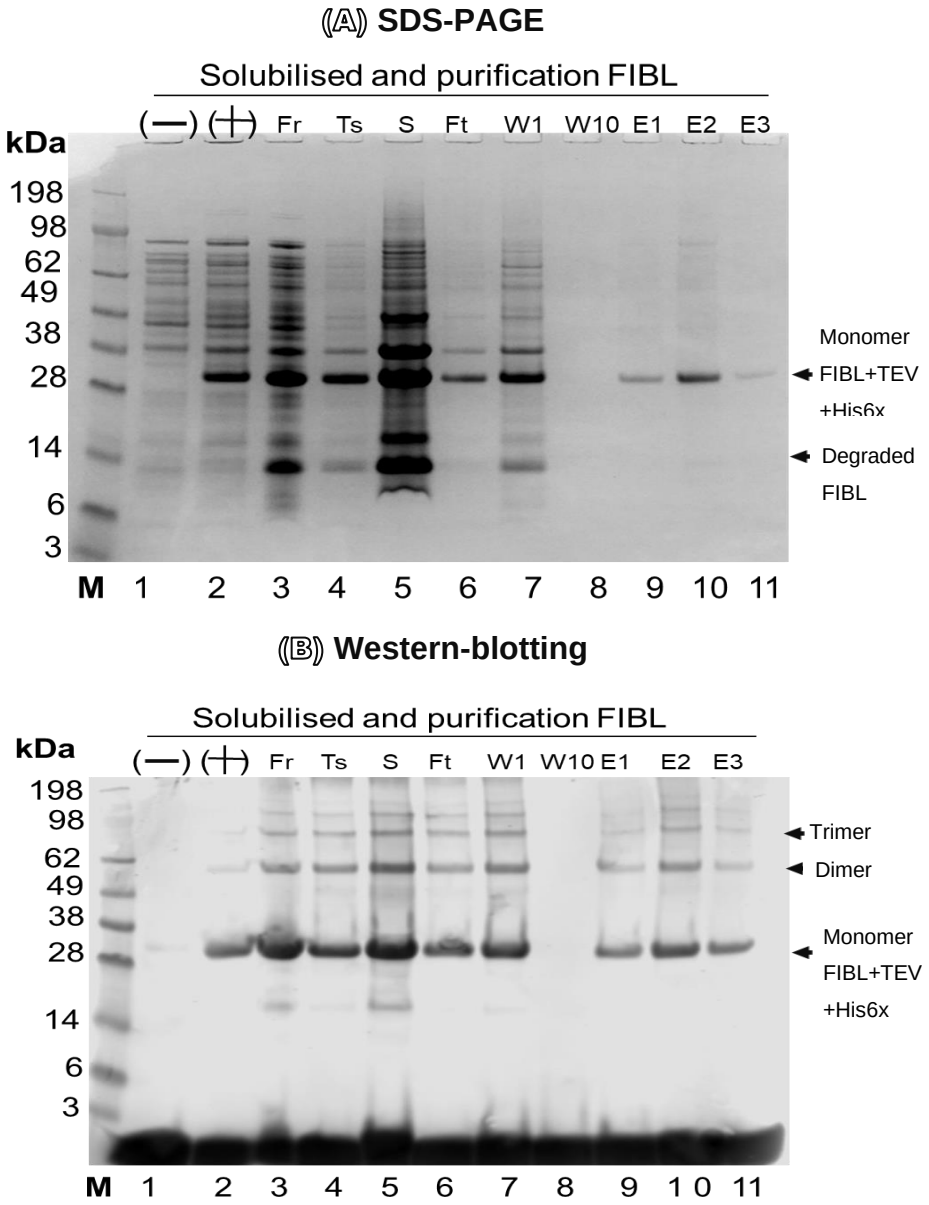


Figure 29. SDS-PAGE **(A)** & western-blotting **(B)** showed the profile of FIBL $_{TEV+His6\text{-}tagged}$ protein expression in SHuffle®T7, solubilisaton and purification. Firstly, a single colony of FIBL $_{TEV+His6\text{-}tagged}$ in SHuffle® T7 was cultured in 10 mL LB media containing kanamycin and the cells grown at 37ºC with a shaker overnight. The following day, the culture was added into 250 mL of fresh LB media and additionally grown until O.D 600 reached up to 0.6 (Un-induced (-) and then induced (+) expression with 0.5 mM IPTG at 25ºC overnight. Then the cells were pelleted out and re-suspended with 1xTBS containing a half tablet of protease inhibitor (EDTA-free). The mixture was disrupted with FRENCH Press (Fr). The resulting IBs pellets were washed with ddH_2O and solubilised with 7M urea containing 100mM HEPES and dissolved overnight. The following day, the total soluble (Ts) was centrifuged to obtain the soluble FIBL $_{TEV+His6\text{-}tagged}$ (S). Followed by His-tag purification, the soluble protein was passed through the Ni-NTA column (Flow through, Ft) and washed with 10 mL of wash buffer and then these washed solutions were collected; (W1&W10), and finally eluted with 3 mL elution buffer (E1, E2 &E3). The purified FIBL $_{TEV+His6x}$ protein of interest is ~29.45-32 kDa.

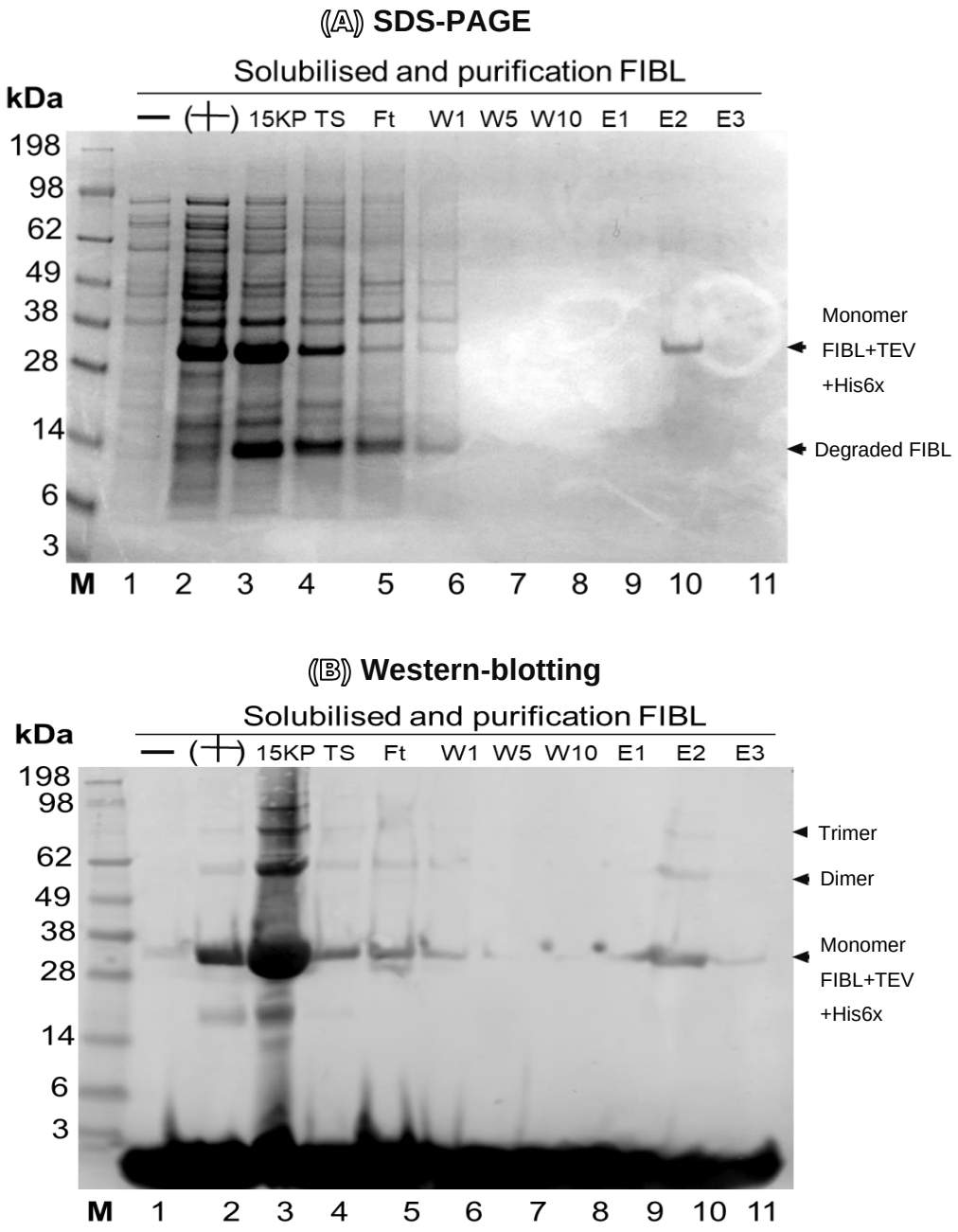


Figure 30. The FIBL $_{TEV+His6\text{-}tagged}$ protein expressed in SHuffle® T7 was shown the production process in LB media through the pre-induction (-), and the post-induction cells with 0.5 mM IPTG(+) at 25ºC overnight. The induced cells were pelleted out, 1XTBS buffer, protease inhibitor and lysozyme added, and the cells disrupted using FRENCH Press. The supernatant fractions were discarded and the retained IBs pelleted by centrifugation (15,000 rpm/15KP). The IBs pellets (15KP) were solubilised with 7M urea into the mix of soluble FIBL protein and the remaining pellets of the total soluble FIBL (TS) were subsequently collected by centrifugation 15,000 rpm (15K). The soluble FIBL was flowed into the Ni-NTA column, the flow through was collected (Ft), washed ten times with wash buffer (Washed 1, W1, washed 5 (W5), washed 10 (W10), and eluted three times (1mL each) by elution buffer, E1, E2, and E3. The purified FIBL $_{TEV+His6x}$ concentration was 0.2 mg/mL which is the expected target bands of FIBL $_{TEV+His6x}$ at ~29.45-32 kDa.

5) Characterisations of FIBL protein expressed in bacteria

5.1 FIBL protein storage stability

(Figure 31) The eluted purified $FIBL_{TEV+His6x}$ (expressed SHuffle® T7) in Elution buffer was preserved at 4ºC to determine its quality ranging from Day1-Day10. The quality changes of $FIBL_{TEV+His6x}$ found that the apparent specific band determined by both SDS-PAGE and western blotting gradually decreased. The $FIBL_{TEV+His6x}$ protein quality can be clearly seen after seven days on the resolution band of monomer, dimer, and trimer with a MW of about 32, 60, and 80 kDa.

5.2 FIBL protein identification using mass-spectrometry analysis

(Appendix, Figure 33 & 35). Briefly, the quality of the eluted purified $FIBL_{TEV+His6x}$ (expressed SHuffle® T7) was visualised and checked on the bis-tris 4-12% gel prior to being determined using the mass-spectrometry technique. The findings showed that the $FIBL_{TEV+His6x}$ protein score resembled the fibroin light chain precursor (*B. mori*) when analysed throughout the process of nLC-ESI MSMS

5.3 FIBL protein secondary structure determination using CD technique.

(Figure 32A & 32B) In order to characterise the secondary structure of $FIBL_{TEV+His6x}$, firstly, the quality and MW of a small sample was checked on SDS-PAGE and western-blotting, the images of which revealed at least three forms of the protein including, 29.45-~32 kDa (monomer), 60.0 kDa (dimer), and 80.0 kDa (trimer), respectively. Then, the remaining secondary structure contents of the verified soluble $FIBL_{TEV+His6x}$ protein were analysed and 25.9% Helix1 (α-helix), 14.9% Helix2 (irregular α-helix**),** 8.3% Strand1(β-sheet), 6.1% Strand2 (anti-parallel), 18.1% Turns (β-turns), and 2% Unordered (others) were found **(Figure 32A & 32B)**.

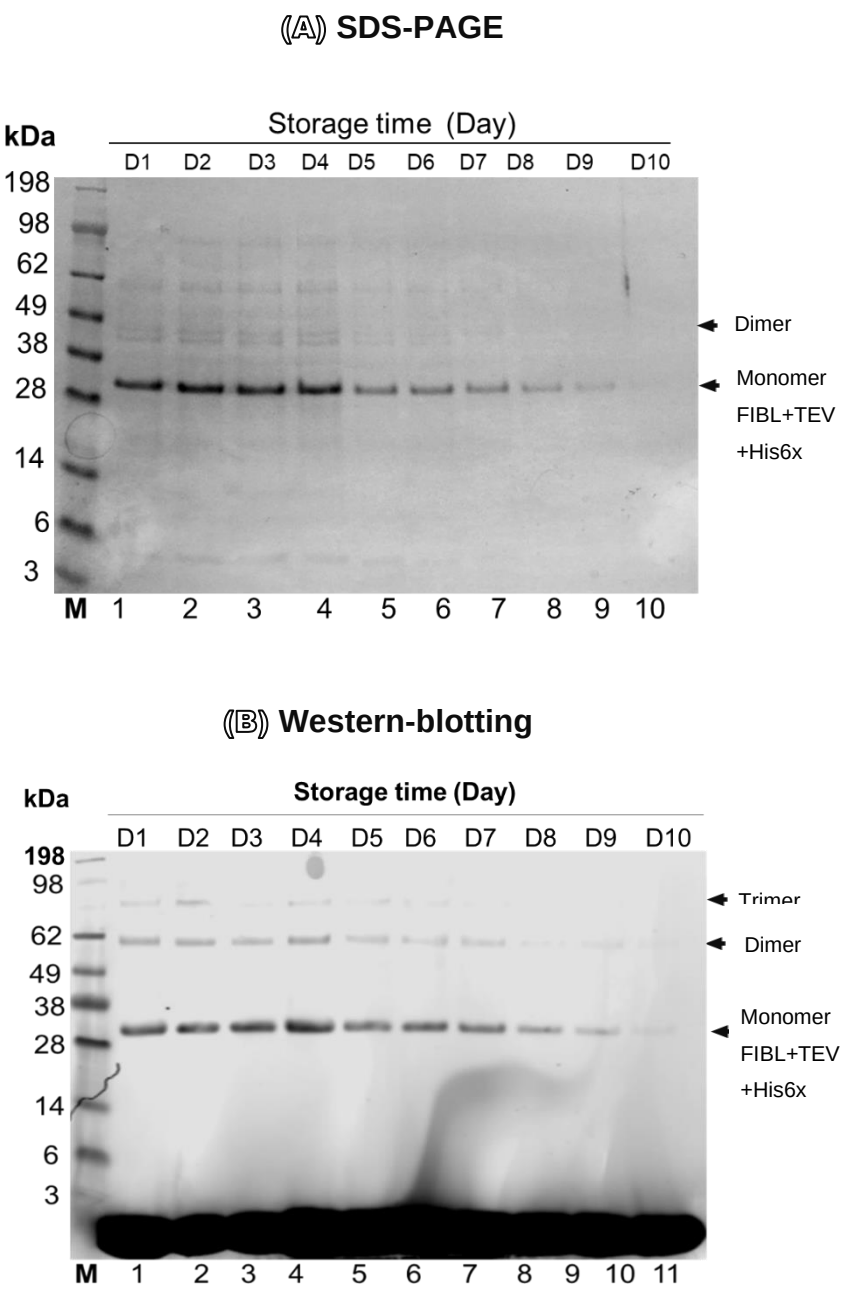
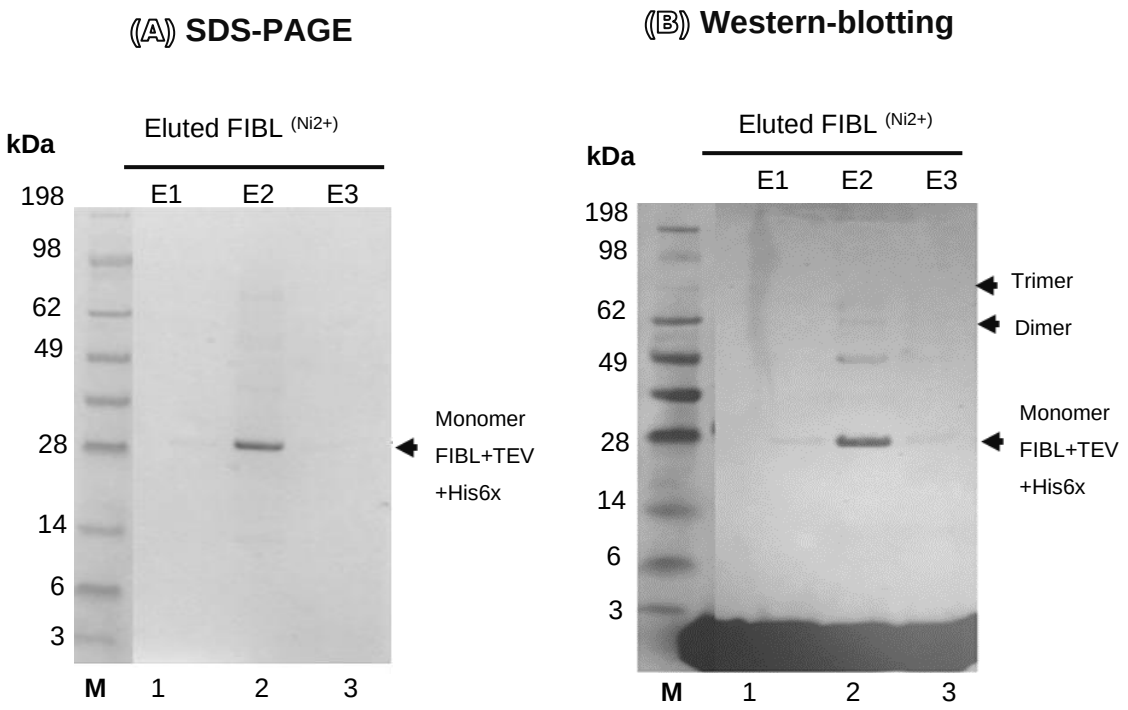
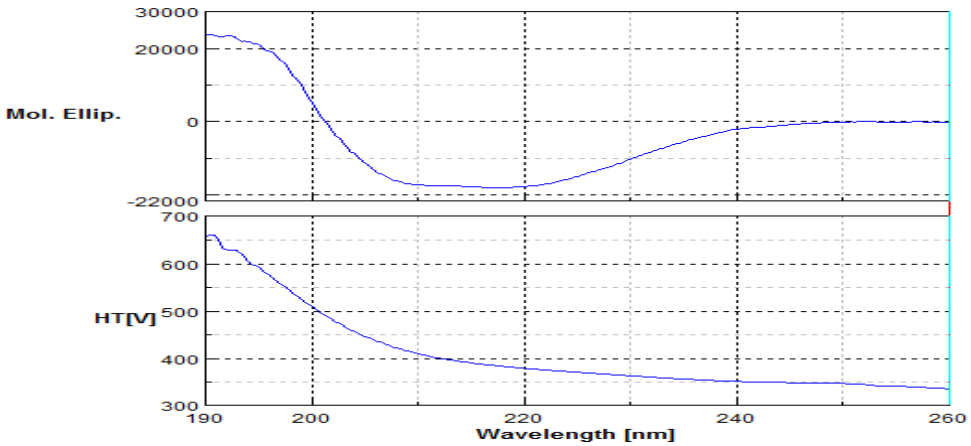


Figure 31. The eluted purified FIBL$_{TEV+His6x}$ protein expressed in SHuffle® T7 was firstly preserved at -20ºC and its stability was consequently tested at 4ºC from the first to the tenth day. Each day, 2x protein buffer was added to a single sample, boiled at 85ºC for 15 mins before being loaded onto the two bis-tris 4-12 % gels (SDS-PAGE). Having finished running the two gels, one SDS-PAGE was stained with Coomassie blue, and another gel was subjected to western blotting. The sample analyses of FIBL revealed that FIBL $_{TEV+His6x}$ would have great quality for about a week when preserved at 4ºC. The FIBL $_{TEV+His6x}$ protein was shown at ~29.45-32 kDa.



(C) Spectra of secondary structure of FIBL protein



(D) Secondary structure contents of FIBL protein

Result	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
1	0.263	0.146	0.088	0.061	0.169	0.274	1.001
2	0.254	0.151	0.078	0.061	0.192	0.261	1
Average	**0.259**	**0.149**	**0.083**	**0.061**	**0.181**	**0.268**	**1.001**

Figure 32. The purified $FIBL_{TEV+His6x}$ was purified using Ni-NTA column and was subsequently concentrated, imidazole was eliminated, and then changed into buffer, 20 mM Tris-HCl pH 8.0, the quality of protein checked using SDS-PAGE **(Figure 32A)** and western-blotting **(Figure 32B)** and consequently the secondary structure was determined using a CD machine. Far-UV spectra of $FIBL_{TEV+His6x}$ has a of concentration 0.6mg/mL measured in 0.01 cm cell path length **(Figure 3C)** the content of these secondary structures, Helix1, Helix2, Strand1, Strand2, Turns, and Unordered was estimated as shown in **Figure (32D).**

CONCLUSION AND DISCUSSION

The fibroin from a native organism (*B. mori*) is a secretory protein (Ohmachi et al., 1982) both in cocoons and in silk fibroins formed and assembled from three subunits, the H-chain, FIBL, and P25 into a molecule of ratio 6:6:1 and secreted into the lumen of the PSG (Inoue et al., 2000). These three subunits are synthesised from different genes as mentioned earlier in the introduction. A large fibroin molecule is synthesised by a large mRNA whereas the other two lower molecules of fibroin are synthesised individually by the translation of two independent mRNAs (Ohmachi et al, 1982). The prior publication revealed that two lower fibroin molecules, both FIBL and P25 glycoprotein, but not the large molecule of H-chain, can be secreted into the lumen gland in a monomeric form, (Tanaka et al., 1999; Long et al., 2015). However, synthesis of the fibroin in the PSG, H-chain and L-chain are commonly associated with a single disulfide bond in a dimeric form (Inoue et al., 2000). Furthermore, using biotechnology to separately produce only the secretory FIBL protein (required as a eukaryotic secretory pathway) without the two H-chain and P25 glycoprotein molecules would be a big challenge according to a few relevant studies (Takei et al., 1984). Previous studies have reported that the genetical modification of L-chain (FIBL) can be more easily cloned and modified than the large molecule fibroin as (H-chain) (Kambe et al., 2007), while other publications reported that there is no crystallite structure in FIBL (Tanaka et., 1999; Yamaguchi et al., 1989). However, genomic sequencing efforts of silk L-fibroins like *B. mori* and its 3D prediction structure, for which scarce experimental information exists, is key to informing our understanding of fibroin evolution (Steward et al., 2022). In this present study, the recombinant FIBL protein/ FIBL protein (in short) was experimented on only in the *E. coli* system. The process of FIBL protein production was conducted throughout with quality checks on the expression construct, validation expression constructs by sequencing confirmation, and identity at the DNA level. In addition, the protein expression was tested with small scale expression and scale up expression in different bacterial strains, temperatures of induction, and concentrations of IPTG. Finally, the secondary structure of FIBL protein was examined using CD analysis. FIBL protein production throughout the whole process was discussed in each next topic.

1) Vector construction of FIBL gene

The successful construction of the FIBL gene containing the TEV+His6x-tagged at the C-terminus in pET28a showed the good expression of DNA level or satisfactory DNA concentration in DH5α™ (the data not shown). The verified sequence of plasmid DNA of FIBL can be compatible in hosts of both groups of bacterial strains, Group1 (un-enhanced disulfide bond formation): BL21(DE3), C41 (DE3), Rosetta™ (DE3)pLysS(Novagen) and Group 2 (enhanced disulfide bond formation): SHuffle® T7.

2) Screening expression and optimisation of FIBL protein

Once plasmid FIBL was transformed into two groups of selected bacterial strains, all strains showed good growth colonies. However, the small scale of FIBL protein expression in different temperatures and IPTG concentrations induction revealed the most promising production in SHuffle® T7, even its induced expression at a low temperature (15ºC) or high temperature (37ºC) (data not shown). Both groups of bacterial strains formed the insoluble FIBL protein as inclusion bodies in the

pellet fractions which required an extra process for solubilisation in order to obtain the active FIBL. The inclusion bodies are aggregated proteins commonly found in recombinant bacteria when hosts are forced to produce heterologous protein species (Garcia-Fruitos et al., 2011). In addition, high level expression of FIBL protein in bacterial cells can lead to disulfide bond formation or it might result in mispairing of cysteines (Ban et al., 2020). However, even though the expression of FIBL in SHuffle® T7 can enhance the disulfide bond formation in the cytoplasm, the FIBL protein still formed the aggregation in the cytoplasm as well as the expression in Group 1 bacterial strains. According to Mergulhao et al. (2005) the explanation is that protein secretion in *E.coli* system is a complex process that attempts to secrete the recombinant protein during the expression. It could have many related obstacles such as incomplete translocation through the inner membrane, insufficient capacity of the export machinery, and proteolytic degradation.

4) Scale-up expression, solubilisation and purification (500mL/Flask) Protocol for FIBL protein production in bacteria system

The experimental findings of the recombinant FIBL expression in two groups of selective *E.coli* showed strains commonly accumulated in inclusion bodies, the cells of which needed to be disrupted and recovered into corrected protein folding. The bacterial cells were lysed by the FRENCH Press and finally FIBL inclusion bodies were solubilised using the selective denaturant 7M urea**.** Singh et al (2015) demonstrated that the high concentration of a chaotropic agent like urea can cause the complete denaturation of the secondary structures which might facilitate the aggregation of FIBL molecules during the refolding processing in Ni-NTA column. The appropriate culture in LB for FIBL protein production would not be so high in volume, about 250-500mL with induction of a lower concentration of 0.5mM IPTG and maintained at 25ºC overnight. The reason for using a lower inducing agent, IPTG and optimum temperature is to reduce the inclusion bodies in the process of expression and the low temperature to reduce the cell growth rate. These two factors can increase the stability and improve protein folding (Rizkia et al., 2014: de Marco et al., 2019). This research project modified the protocol for FIBL protein production from previous studies in relation to solubilisation and refolding. The IBs could be solubilised well in 7M urea without any reducing agents in about 30 mins – 1 hr before being refolded through the Ni-NTA column. The solubilisation of FIBL IBs did not require reducing agents like DTT and BME. These two agents can break a single intra-disulfide bond in the FIBL protein sequence (data not shown). Additionally, if those strong reducing agents are added in the solubilisation process, it reduces FIBL protein so that a single band cannot be seen when visualised on SDS-PAGE. FIBL protein showed a contrast with its solubilisation from other recombinant proteins that require reducing agents to maintain the cysteine residues (Ban et al., 2020).

5) FIBL protein expression and characterisation in bacteria

5.1 FIBL protein storage stability

Determination of the FIBL protein denaturation after the refolding and purification from Ni-NTA column is very important. Upon the test, FIBL protein showed stability in its quality for about a week in the refrigerator. However, after that the resolution of the band on SDS-PAGE gradually decreased. This might commonly involve the proteolytic process as, according to Maurizi (1992),

storage of other soluble proteins (native and recombinant) tends to be susceptible to degradation before or after extraction from the cells. Therefore, after undergoing chromatography, the purified protein of interest should be diluted three – five times and protease inhibitor should be added before storage. It has been suggested that the addition of protease inhibitors is useful for preventing the proteolytic process in recombinant protein (Zhang et al., 2007: Wingfield, 2015).

5.2 FIBL protein identification using mass-spectrometry analysis

Purity of FIBL protein was checked using SDS-PAGE and western-blotting throughout the many experiments discussed in this chapter. Further assurance on checking the identity and purity of FIBL protein was through expression in a host (SHuffle® T7) and then the end of the purification process (Ni-NTA column) was carried out using the nLC-ESI MSMS. Identification of the obtained FIBL protein confirmed that the recombinant technique produced FIBL protein with similarities to the precursor of FIBL in the native organism (*B. mori*). Furthermore, this finding indicates that (SHuffle® T7) could be a promising host to scale up the FIBL protein production in the same way as exploiting the native organism.

5.3 FIBL protein secondary structure determination using CD structure

Light chain fibroin/ FIBL (*B. mori*) contains standard amino acid composition and is characterised as a nonrepetitive sequence (Zhou et al., 2001). FIBL is produced 1:1 to the H-chain whereas the fibroins have a stoichiometry molecule ratio relationship to three subunits, H-chain (6)): FIBL (6): P25(1): (Julien et al., 2005). However, the attention to the FIBL structure is still limited so little is known about the secondary structure of FIBL protein (*B.mori*). A prediction of the secondary structure of FIBL was provided about three decades ago. The prediction used cDNA-derived amino acid sequence and corroborated the analysis using a computer program which included DNASIS. The prediction showed the regions of β-sheet, α-helix and β-turn and others (Yamaguchi et al., 1989). This was the first study to reveal the secondary structure in the purified $FIBL_{TEV+His6x}$ by CD analysis. The findings showed that the average conformation of $FIBL_{TEV+His6x}$ had higher elements of helices (helix & irregular α-helix**)**) about 40.8% higher than those predicted, including strands (β-sheet & anti-parallel) (14.4%), β-turn (18.1%), and others (26.8%). However, the findings from the CD peaks in our study gave evidence of more detail in its elements than in the recent prediction, for example, the prediction of 3D structure of fibroin light chain (*B. mori*) when made by AlphaFold Monomer V2.0 and another prediction of the helices of a 3D FIBL structure using ColabFold by pLDDT (Steward et al., 2022).

BIBLIOGRAPHY,

Barbosa, J.F., Bravo, J.F., Zannatta, D.B., Silva, J.L.C. & Fernandez, M.A. 2009. 'Allelic variability in the third intron of the fibroin light chain gene in *Bombyx mori* (Lepidoptera: Bombycidae)'. *Genetics and Molecular Research*. 8(1):197-206.

Barbosa J.F, Bravo J.P, Takeda K.I, Zanatta D.B, et al. 2008. 'Intrinsic bent DNA colocalizes with the sequence involved in the *Nd*-s^{D} mutation in Fibroin light chain gene of the *Bombyx mori'*. *BMB Reports*. 41: 394-399.

Ban, B., Sharma, M. & Shetty, J. 2020. 'Optimization of methods for the production and refolding of biologically active disulfide bond-rich antibody fragment in microbial hosts'. *Antibodies*. 9:39 (1-18 pages).

Couble, P., Chevillard, M., Moine,A., Ravel-Chapuis, P. & Prudhomme, J.C. 1985. 'Structural organization of the P25 gene of *Bombyx mori* and comparative analysis of its flanking DNA with that of fibroin gene'. *Nucleic Acids Research*. 13: 1801-1814.

Choi, K., H., Goo, T.W., Yun, E.Y., Hwang, J.S., Kang, S.W., Lee, S.M., Sohn, H.D. & Jin, B.R. 2002. 'Molecular cloning and characterization of fibroin light chain from the silkworm Beakok-Jam (*Bombyx mor*i)'. *International Journal of Industrial Entomology*. 5(1): 93-102.

De Marco, A., Ferrer-Miralles, N., Garcia-Fruitos, E., Mitraki, A., Peternel, S., Rinas, U., Trujillo-Roldan, M.A., Valdez-Cruz, N.A., Vazquez, E. & Villaverde, A. 2019. 'Bacterial inclusion bodies are industrially exploitable amyloids'. *FEMS Microbiology Reviews*. 43(1): 53-72.

Garcia-Fruitos, E., Sabate, R., de Groot, N.S., Villaverde, A. & Ventura, S. 2011. 'Biological role of bacterial inclusion bodies: a model for amyloid aggregation'. *FEBS Journal*. 278(14):2419-2427.

Hyodo, A., Yamamoto, T., Ueda, H., Takei, F., Kimura, K. & Shimura, K. 1984. 'Linkage analysis of the fibroin light chain gene in the silkworm, *Bombyx mori*'. *The Japanese Journal of Genetics*. 59: 285-296.

Inoue, S., Tanaka, K.,Arisaka, F., Kimura, S., Ohtomo, K. & Mizono, S. 2000. 'Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio'. *The Journal of Biological Chemistry*. 275 (51):40517-40528.

Julien, E.,Coulon-Bublex, M., Garel, A. Royer, C., Chavancy, G., Prudhomme, J.C. & Couble, P. 2005. 'Silk gland development and regulation of silk protein gene'. *Comprehensive Molecular Insect Science*. 2:369-384.

Kambe, Y., Yamamoto, K., Kojima, K., Tamada, Y. & Tomita, N. 2010. 'Effect of RGDS sequence genetically interfused in the silk fibroin light chain protein on chondrocyte adhesion and cartilage synthesis'. *Biomaterials*. 31(29): 7503-7511.

Kikuchi, Y., Mori, K., Susuki, S., Yamaguchi, K. & Mizuno, S. 1992. 'Structure of *Bombyx mori* fibroin light-chain encoding gene: upstream sequence elements common to the light can heavy chain'. *Gene*. 10:151-158.

Li, Z., Jiang, Y., Cao, G., Li, J.Z., Xue, R. & Gong, C. 2015. 'Construction of transgenic silkworm spinning antibacterial silk with fluorescence'. *Molecular Biology Reports*. 42(1): 19-25.

Liu, W.P., Zhou, Z.T., Zhang, S., Shi, Z.F., Tabarini, J., Lee, W.S., Zhang, Y.H., Corder, S.N.G., Li, X.X., Dong, F., Cheng, L., Liu, M., Kaplan, D.L., Omenetto, F.G., Zhang, G.Z., Mao, Y. & Tao, T.H. 2017. 'Precise protein photolithography (P^3): high performance biopatterning using silk fibroin light chain as the resist'. *Advance Science*. 1700191:1-9.

Lu, F, Wei, Z., Luo, Y., Guo, H., Zhang, G., Xia, Q. & Wang, Y. 2019. 'SilkDB 3.0: visualizing and exploring multiple levels of data for silkworm'. *Nucleic Acids Research*. 48(1):D749-D755.

Long, D., Lu, W., Zhang, Y., Guo, Q., Xiang, Z. & Shao, A. 2014. 'New insight into the mechanism underlying fibroin secretion in silkworm, *Bombyx mor*i'. *The FEBS Journal*. 28(1): 89-101.

Maurizi, M.R. 1992. 'Proteases and protein degradation in *Escherichia coli*.'. *Experimentia*. 48:178-201.

Mergulhao, F.J.M., Summers, D.K. & Monteiro, G.A. 2005. 'Recombinant protein secretion in *Escherichia coli* '. *Biotechnology Advance*. 23:177-202.

NCBI, 2022. 'iCn3D: web-based 3D structure viewer'. National Library of Medicine, National Center for Biotechnology Information. [Online access; 21/10/2022], https://www.ncbi.nlm.nih.gov/ Structure/icn3d/docs/icn3d_about.html.

Ohmachi, T., Nagayama, H. & Shimura, K. 1982. 'The isolation of a messenger RNA coding for the small subunit of fibroin from the posterior silk gland of the silkworm, *Bombyx mori*.' *FEBS LETTERS*. 146(2):385-388.

Rahaieem, S., Zare, M. & Jafari, S.M. 2019. 'Nanostructure of silk fibroin for encapsulation of food ingredients'. *In Chapter 12*, Biopolymer nanostructures for encapsulation purposes. Academic Press. Elsevier Inc. UK. 642pp.

Ritkia, P.R., Silaban, S., Hasan, K., Kamara, D.S., Subroto, T., Soemitro, S., Maksum, I.P. 2014. 'Effect of *Isipropyl-β-D-thiogalactopyranoside* concentration on pretombin-2-recombinan gene expression in *Escherichia coli* '.ER2566, 2015. *Procedia Chemistry*. 17:118-124.

Saotome, T., Hayashi, H., Tanaka, R., Kinugasa, A., Uesugi, S., Tatematsu, K, I., Sezutsu, H., Kuwabara, N. & Asakura, T. 2015. 'Introduction of VEGF or RGD sequence improves revascularization properties of *Bombyx mori* silk fibroin produced by transgenic silkworm'. *Journal of Materials Chemistry B*. 3: 7109-7116.

Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M. & Panda, A.K. 2015. 'Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process'. *Microbial Cell Factories*. 14:41 (1-10 pages).

Sreerama, N. & Woody, R.W. 2000. 'Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set.' *Analytical - Biochemistry.* 287(2): 252-260.

Stewart, R.J., Frandsen, P.B., Pauls, S.U. & Heckenhauer, J. 2022. 'Conservation of three-dimensional structure of Lepidoptera and Trichoptera L-fibroins for 290 years'. *Molecules*. 27(8):5945, 1-18.

Tanaka, K., Kajiyama, N., Ishikura, K., Waga,S., Kikuchi, A., Ohtomo, K., Takagi, T. & Mizuno, S.1999. 'Determination of the site of disulfide linkage between heavy and light chains

of silk fibroin produced by *Bombyx morii* '. *Biochimica et Biophysica Acta (BBA)*. 1432 (1):92-103.

Takei, F., Oyama, F., Kimura, K., Hyyodo, A., Mizuno, S. & Shimura, K. 1985. 'Reduced level of secretion and absence of subunit combination for the fibroin synthesized by a mutant silkworm', *Nd*(2).*The Journal of Cell Biology*. 99:2005-2010.

Takei, F., Kikuchi, Y., Kikuchi, A., Mizuno, S. & Shimura, K. 1987. 'Further evidence for importance of the subunit combination of silk fibroin in its efficient secretion from the posterior silk gland cells' . *Journal of Cell Biology*. 105: 175–180.

UniProt. P21828.FIBL_BOMMO, Fibroin Light Chain. *Bombyx mori* (Silk moth). Gene. Available from https://www.uniprot.org/uniprotkb/P04148/entry, [Access; 22/10/ 2022].

Yamaguchi, K., Kikuchi, Y., Tkagi, T., Kikuchi, A., Oyama, F., Shimura, K. & Mizuno, S. 1989. 'Primary structure of the silk fibroin light chain determined by cDNA sequencing peptide analysis'. *Journal of Molecular Biology*. 210(1): 127-139.

Ye, X., Dai, X., Wang, X., Yu, S., Wu, M., Zhao, S., Ruan, J. & Zhong, B. 2022. Mechanism of silk secretion revealed by protomic analysis of silkworm cocoon with fibroin light chain mutations'. *Journal of Protomics*. 265(104649):1-8.

Yamao M, Katayama N, Nakazawa H, Yamakawa M, Hayashi Y, Hara S, Kamei K & Mori, H. 1999. 'Gene targeting in the silkworm by use of a baculovirus'. *Genes & Development*. 13(5): 511-516.

Wang, H., Wang, L., Wang, Y.L., Tao, H., Yin, W., Sima, Y., Wang, Y.J & Xu, S. 2015. 'High yield exogenous protein HPL production in the *Bombyx mori* silk gland provides novel insight into recombinant expression systems'. *Scientific Reports*. 5(13839):1-1.

Wingfield, P.T. 2016. 'Overview of the purification of recombinant protein'. HHS Public Access. *Current Protocols in Protein Science*. Online Journals. 1-50.

Zafar, M.S., Belton, D.J., Hanby, B., Kaplan, D.L. & Perry, C.C. 2015. 'Functional material features of *Bombyx mori* silk light vs. heavy chain proteins'. *Biomacromolecules*. 16(2): 606-614.

Zhang, Y., Liu, R. & Wu, X. 2007. 'The proteolytic systems and heterologous protein degradation in the methylotrophic yeast *Pichia pastoris*.' *Annals of Microbiology*. 57(4):553-560.

Zhou, C.Z.,Confalonieri, F.,Jacumet, M., Perasso, R., Li, Z.G., Janin, J. 2001. 'Silk fibroin: structural implications of a remarkable amino acid sequence'. *Proteins, Structure, Function, and Genetics*. 44(2):119-122.

APPENDIX CHAPTER III

Identification of FIBL using mass spectrometry

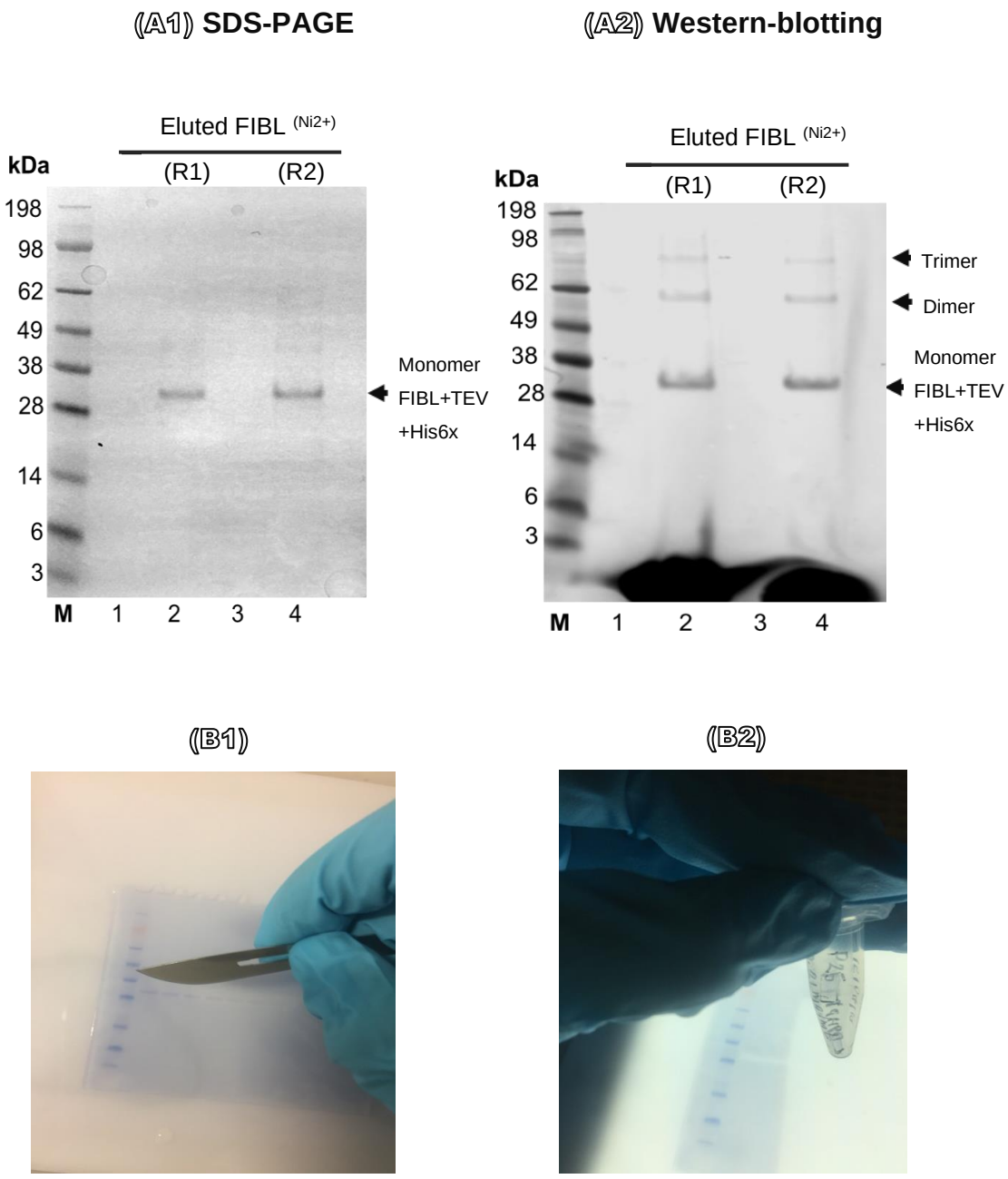


Figure 33. shows the eluted purified soluble FIBL $_{TEV+His6x}$ protein after purification using Ni-NTA column. (**A1**-SDS-PAGE) and (**A2**-western blotting) the quality of MW in a single band of purified FIBL $_{TEV+His6x}$ was checked by two gels of bis-tris 4-12%, one gel was stained with Coomassie blue r-250 for 30 mins and washed with destaining solution overnight and another gel was determined using western-blotting **(A2**). Consequently, the identify of purified FIBL $_{TEV+His6x}$ protein was tested by cutting the cleaned single band of purified FIBL $_{TEV+His6x}$ into the tube and adding 20µL ddH_2O before being sent for mass spectrometry analysis (**B1-B2**).

Identification of FIBL using mass spectrometry (continued)

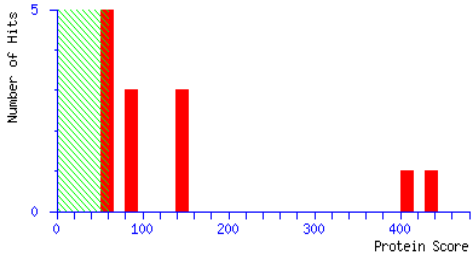
MATRIX SCIENCE Mascot Search Results

```
User           : Chirapha_Butiman
Email          :
Search title   : Chirapha FIBL ESI NCBI all sp, Chirapha FIBL 5uL.mgf
MS data file   : \\cfs.st-andrews.ac.uk\Shared\Chem_PIs\BSRC_Mass_Spec\A MGF REPOSITORY\Data from 5600+\190409\Chirapha FIBL 5uL.mgf
Database       : NCBIprot 20190208 (187857634 sequences; 68510950169 residues)
Timestamp      : 9 Apr 2019 at 17:12:33 GMT
Protein hits   : NP_001037488.1 fibroin light chain precursor [Bombyx mori]
                 P06654.1       RecName: Full=Immunoglobulin G-binding protein G; Short=IgG-binding protein G; Flags: Precursor
                 NP_000412.3    keratin, type I cytoskeletal 10 [Homo sapiens]
                 1AKS_A         Chain A, Crystal Structure Of The First Active Autolysate Form Of The Porcine Alpha Trypsin
                 XP_017519449.1 PREDICTED: trypsin [Manis javanica]
                 AWP19426.1     putative myosin heavy chain fast skeletal muscle-like [Scophthalmus maximus]
                 KFP01285.1     Keratin, type II cytoskeletal 73 [Calypte anna]
                 1OMI_A         Chain A, Crystal Structure Of Prfa,The Transcriptional Regulator In Listeria Monocytogenes
                 XP_027128986.1 LOW QUALITY PROTEIN: myosin heavy chain, fast skeletal muscle [Larimichthys crocea]
                 PWA28320.1     hypothetical protein CCH79_00019005 [Gambusia affinis]
                 XP_006863488.1 PREDICTED: myosin-8 [Chrysochloris asiatica]
                 1CQI_A         Chain A, Crystal Structure Of The Complex Of Adp And Mg2+ With Dephosphorylated E. Coli Succinyl-coa Synthetase
```

	NCBIprot	Decoy	False discovery rate
Peptide matches above identity threshold	20	4	20.00 %
Peptide matches above homology or identity threshold	25	6	24.00 %

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 60 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Peptide Summary Report

1. NP_001037488.1 Mass: 27908 Score: 436 Matches: 5(5) Sequences: 5(5) emPAI: 1.37
fibroin light chain precursor [Bombyx mori]
Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
18	428.2078	854.4010	854.4035	-2.94	0	60	0.048	1	U	R.SGNFAGFR.Q
111	656.7939	1311.5732	1311.5732	0.03	1	72	0.0019	1	U	R.RWDYVDDTDK.S
114	661.3189	1320.6232	1320.6245	-0.95	0	84	0.00029	1	U	R.YSVGPALGCAGGGR.I
119	454.8989	1361.6748	1361.6787	-2.83	1	65	0.032	1	U	R.DIDDGKASSVISR.R
135	807.4035	1612.7925	1612.7917	0.45	0	156	2.1e-11	1	U	R.GVGNGNDATGLVANAQR.Y

Figure 34. a single band of purified FIBL $_{\text{TEV+His6-tagged}}$ was tested and the peptide sequences identified, showing it had the same identity as fibroin light chain precrsor (*B. mori*) by nLc-ESI MSMS analysis.

CHAPTER IV

Comparative Study of Recombinant P25 Glycoprotein (*B. mori*) in Bacteria (Part I) & Baculovirus-Infected Insect Cell Systems (Part II)

……

ABSTRACT (Part I): Expression P25 in Bacteria System

……

The aim of this project was to target the production of recombinant secreted *N*-linked P25 glycoprotein, one of a minor molecule of fibroin subunits (*Bombyx. mori*) in *Escherichia coli* bacteria system. The P25 gene was labelled with TEV protease + His6x-tagged at C-terminus of pET28a and transformed into DH5α cells. Confirmation of DNA plasmid was transformed into two groups of bacterial strains, Group 1 (un-enhanced disulfide bond formation), including three bacterial cells; BL21(DE3), C41(DE3), Rosetta™ (DE3)pLysS(Novagen) and Group 2 (enhanced disulfide bond formation) a bacteria strain, SHuffle® T7. The small-scale production was attempted with a combined variation in two factors of culture conditions; 1) varied temperatures and 2) varied IPTG concentrations. The small-scale process showed that a strain of P25 glycoprotein can be produced only in Group 2, SHuffle® T7 (0.5mM IPTG at 25ºC overnight), which could promote the post-translation modification of all culture conditions. SHuffle® T7 cells were therefore more likely to be useful for producing the active protein (~26.91 kDa/SDS-PAGE ~28kDa). However, in all the conditions, the P25 glycoprotein were expressed in aggregate form as inclusion bodies (IB pellets) which needed to be recovered using 7M urea and the soluble active yield refolded by the Ni-NTA column. This recombinant P25 glycoprotein which was identified by mass spectrometry was found to be similar to that of the original organism (*B. mori*). The quality of purified P25 glycoprotein can be kept at 4ºC for about one week. On the other hand, the construct of P25/pET28a/SHuffle® T7 cells showed that the expression was unstable after the construct maintenance at -20ºC for about three months.

Keywords; P25 glycoprotein, recombinant, Fibroin (*B. mori*), bacteria system, *E.coli*

ABSTRACT (Part II): Expression P25 in Baculovirus-Infected Insect Cell

The P25 gene (*B. mori*) was tagged on both sides, His6x-tagged + 3C protease (N-terminus) whereas TEV protease + His6x-tagged was added (C-terminus) in the pFBDM plasmid. The verified DNA sequence was transformed into MultiBac baculovirus variant cells (EMBacY). The Blue/white screening technique was used to select the right insertion and grown on the LB media. The assured plasmid DNA was transfected into the Sf21 cells (*Spodoptera frugiperda*) throughout the baculovirus-infected insect cell production process.

The findings showed the *N*-linked P25 glycoprotein (P25 glycoprotein) can be expressed well in SF21 cells after 72 hrs. The P25 glycoprotein can be found in both the pellets and medium fractions; semi-glycosylated in the pellet and fully glycosylated in the medium. The pellet cells (non-secreted P25 glycoprotein) were obtained with two combined intact MWs (calculation ~28.89 kDa, SDS-PAGE/ western blotting ~30 & ~32 kDa) which were named as semi-glycosylated P25 protein. However, a medium fraction (secreted P25 glycoprotein) showed a high yield of soluble full glycosylated ~32 kDa (SDS-PAGE and western blotting).

The secondary structures of two forms (semi-glycosylated/glycosylated P25 protein) were determined by circular dichroism (CD). The secondary structure elements of the semi-glycosylated P25 indicated less strength than the glycosylated P25 protein (containing a higher number of strands).

When both glycans of semi-glycosylated and glycosylated P25 glycoprotein were examined using endoglycosidase enzymes treatment (Endo-H) and PNGase F, each enzyme showed both glycoforms had similar mobility to the glycosylate protein. However, the clearest released glycans indicated the complete glycosylation from the PNGase F reactions only in the glycosylated P25 glycoprotein, not in the semi-glycosylated form.

In addition, glycan analysis using immunoprecipitation (IP) combined with Concanavalin A was carried out on both semi-glycosylated and glycosylated P25 glycoproteins. It was found that the glycosylated P25 glycoprotein showed three intact bands (SDS-PAGE) of glycans with different molecular weights (MWs) of attachment (~12, 14, and 26 kDa) whereas these glycans in the glycosylated P25 glycoprotein showed higher MW at around ~30-32 kDa. However, the attachments of these glycans could not be seen in the semi-glycosylated P25 glycoprotein.

Keywords; baculovirus-infected insect cell, glycan, endoglycosidase, secondary structure, immunoprecipitation (IP)

INTRODUCTION

Basically, there are various groups of arthropods that weave different fibrous silk-based protein produced from several types of silk glands or spinning compartments. Many arthropod groups besides silkworms and spiders have independently evolved the ability to produce silk. Arthropod silks are structurally different at the molecular level, containing crystallites of 1) beta-sheets with the protein backbone parallel to the fiber axis, 2) beta-sheet with the backbone perpendicular to the fiber axis 3) alpha-helices as coiled coils 4) collagen triple helices or 5) polyglycine II-like structure. The ordered protein structures in silk can contribute extensively to silk function (Walker et al., 2013). Recently, genetic manipulations have been used to create and design synthetic repetitive silk-like genes that produce the protein sequence which can target specific properties for specific applications (Teule et al., 2009). P25 glycoprotein as a chaperonin-like protein (Inoue et al., 2000), is a minor component among three intact fibroin subunit proteins (H-chain, L-chain or FIBL, and P25 glycoprotein) containing 220 amino acid long peptide encoded by the fibrohexamerin gene (*fhx*). It is composed of a 3.5 Kb single-copy gene with five exons (Couble et al., 1985; Chevillard et al., 1986). The mRNA of P25 codes for 25 kDa from 663 linear bp (NCBI, 2022). The P25 gene is expressed in the posterior silk gland (PSG) and its transcription is repressed at larval moulting stages (Durand et al., 1992). The P25 gene is located on chromosome 2 (Lu et al., 2020) whereas H-chain and FIBL genes are located on chromosomes, 25 and 14 (Couble et al., 1985). The P25 gene is thought to play a key role in fine-tuning the molecular rate of synthesis of fibroin complexes as it regulates the transcription of fibroin genes via SGF-2. The P25 gene promoter is assumed to contain two sets of roles, one as a eukaryotic promoter and secondly as a silk gland specific promoter (Liu et al., 2007: Love et al., 2011: Wu et al., 2021). Among these three fibroin genes, the H-chain gene has been investigated more than the others for the regulation of the expression (Shimizu et al., 2007).

P25 glycoprotein is synthesised and secreted in the PSG cells of silkworms with strict territorial and developmental specificities (Guo et al., 2005) in the same time frame as the H-chain and FIBL, but by different genes. To form the fibroin complex, H-chain-FIBL heterodimers are formed via a single disulfide bond. Six H-chain-FIBL heterodimers and one P25 glycoprotein are associated via hydrophobic interaction (non-covalently linked), to maintain the integrity of the huge fibroin complex with FIBL:H-chain:P25 ratio of 6:6:1 and weight 2.3 MDa (Tanaka et al., 2000).Fibroin (*B. mori*) comprises about 70-75% of the main silk fiber containing three different subunit components, H-chain (large molecule, ~390 kDa), FIBL/L-chain (small molecule ~26kDa) (Rockwood et al., 2011) and P25 glycoprotein/Fibrohexamerin (*Fhx*) (small molecule ~30kDa, Tanaka et al., 1999). This massive secretory elementary unit can be expressed in a stoichiometric ratio (6:6:1) (Inoue, et al., 2000). Full length amino acids of P25 glycoprotein are composed of eight cysteine (Cys) residues which indicates that P25 compacts via its intra-disulfide bonds. (Nony et al.,1995: Tanaka et al., 1999; Lamboni et al., 2018). Information from these Cys residues of P25 glycoprotein indicates that this protein requires post-translational modification for the correct folding, and it might co-interact with *N*-glycosylation for biological functions. After protein synthesis in PSG, P25 glycoprotein, containing *N*-linked oligosaccharide chains, interacts with the H-chain-FIBL complex to form a higher complex of specific conformation during the intracellular transportation and secretion (Tanaka et al., 1999, Ohno et al.,

2013). P25 glycoprotein contains hydrophobic regions which interact with the Gly-Ala heterodimeric linkage (inter-disulfide bond) of the H-chain-FIBL (Inoue et al., 2000; Lamboni et al., 2018). In addition to that, P25 glycoprotein accommodates both the effective secretion of the massive fibroin complex as micelle from PSG into the lumen by reverse pinocytosis and maintains fibroin solubility during luminal transport (Inoue et al., 2000; Kludkiewicz et al., 2009, Zabelina et al., 2021). P25 glycoprotein is not essential for silk formation, but it is thought to affect the morphology of fibroin secretory globules in the lumen of PSG (Zabelina et al., 2021). Glycosylation is another crucial post-translational modification of P25 glycoprotein. In its original organism (*B. mori*), it consists of three *N*-linked oligosaccharide chains with Asn residues at sites, 69, 113, and 133, respectively (Tanaka et al., 1999; Inoue et al., 2000). These act as sites for three *N*-linked glycosylations and are crucial to maintain the huge elementary unit of fibroin complex (2.3 MDa) assembled in the PSG before it is transported to MSG (Inoue et al., 2000, Ohno et al., 2013). P25 glycoprotein can be found in two forms, either a 30 kDa or 27 kDa which are thought to exist in different compositions of oligosaccharide attachments or degrees of glycosylation (Tanaka et al., 1999; Inoue et al., 2000; Inoue et al., 2004). The publication of previous studies of recombinant P25 glycoprotein is still limited. However, there were some interesting studies in which the P25 gene was cloned into pFasBac (donor plasmid) and expressed in a bac-to-bac baculovirus expression system (Sf9 cells). However, this method of obtaining P25 glycoprotein was shown on the SDS-PAGE to carry an incomplete form of glycosylation (Ohno et al., 2013). Fibroin studies show it has well established characterisation in its sequence and secondary structure (Sehanal et al., 2004; Bini et al., 2004). However, biochemical knowledge of the P25 glycoprotein is still unclear and reports of the recombinant P25 glycoprotein in different hosts is very limited including the P25 protein structure. However, the prediction of P25 protein3D structure as shown in **Figure 35**.

Therefore, the recent study can reveal the importance of using biotechnology to produce P25 glycoprotein in two different hosts, bacteria (prokaryotic system) and baculovirus-infected insect cell (eukaryotic system) and give further insights into this fibroin protein and its extensive applications.

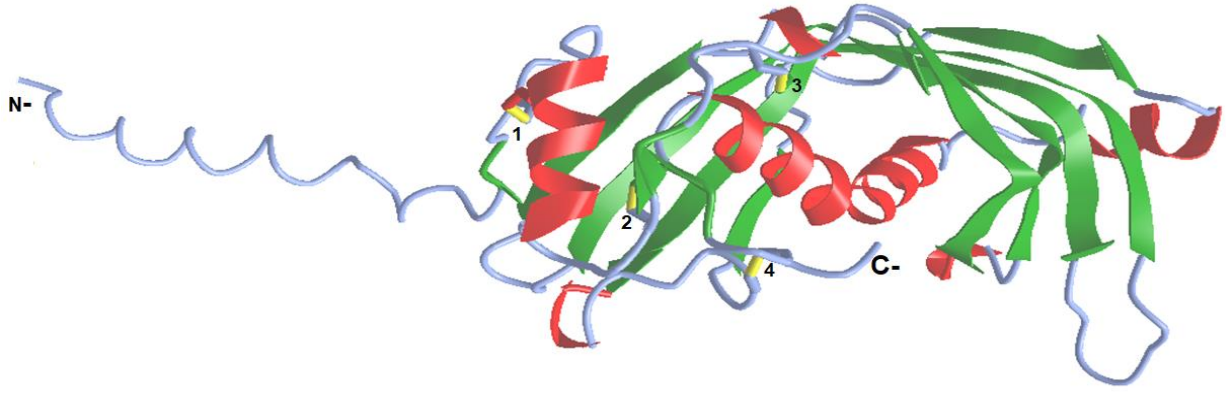


Figure 35. The prediction of 3D structure of P25 glycoprotein (*B. mori*) when made by AlphaFold, produces a per-residue confide score from P04148.SI25_BOMMO 3D structure database (UniProt, Website 2022). This AlphaFold database was analysed and modified into the 3D structure using iCn3D:web-based 3D structure viewer (NCBI, website 2022). The tertiary structure of P25 glycoprotein would contain β-sheet (green), helix (red), and others, random coils, and loop. The four disulfide bounds in its structure are highlighted in yellow.

MATERIALS AND METHODS

The main experiments of FIBL production in bacteria were described in chapter II.

RESULTS: Part (I) (P25 protein expression in bacteria system)

RESULTS 1) Vector construction (Molecular Engineering)

The P25 Gblocks were successfully amplified using a PCR by adding EcoRI and XhoI and classical cloning technique was employed with pET28a.The insertion and vector were double digested with EcoRI and XhoI enzymes following the general protocol and the DNA concentration was subsequently determined by NanoDrop and the DNA quality on 1% agarose gels as shown in **Figure 36 (A-B).** Nucleotide sequences of pET28a are 5,369bp whereas P25 gene including the TEV+His6-tagged are 752bp. Both the digested P25 and pET28a DNAs were employed for the ligation reaction and transformed into the DH5α competent cells and plated on agar plates containing kanamycin and incubated in the 37ºC incubator overnight. The following day, the resultant single colonies were randomly picked and grown on the 10 mL LB supplemented with kanamycin overnight. The next day, the plasmid DNA of each cell culture was extracted and its primary ligation was rechecked by double digestion enzymes, EcoRI and XhoI. If the agarose gel contained both DNA that meant the cloning was completed and these plasmid DNAs were used for the validation of DNA sequencing **Figure 37.**

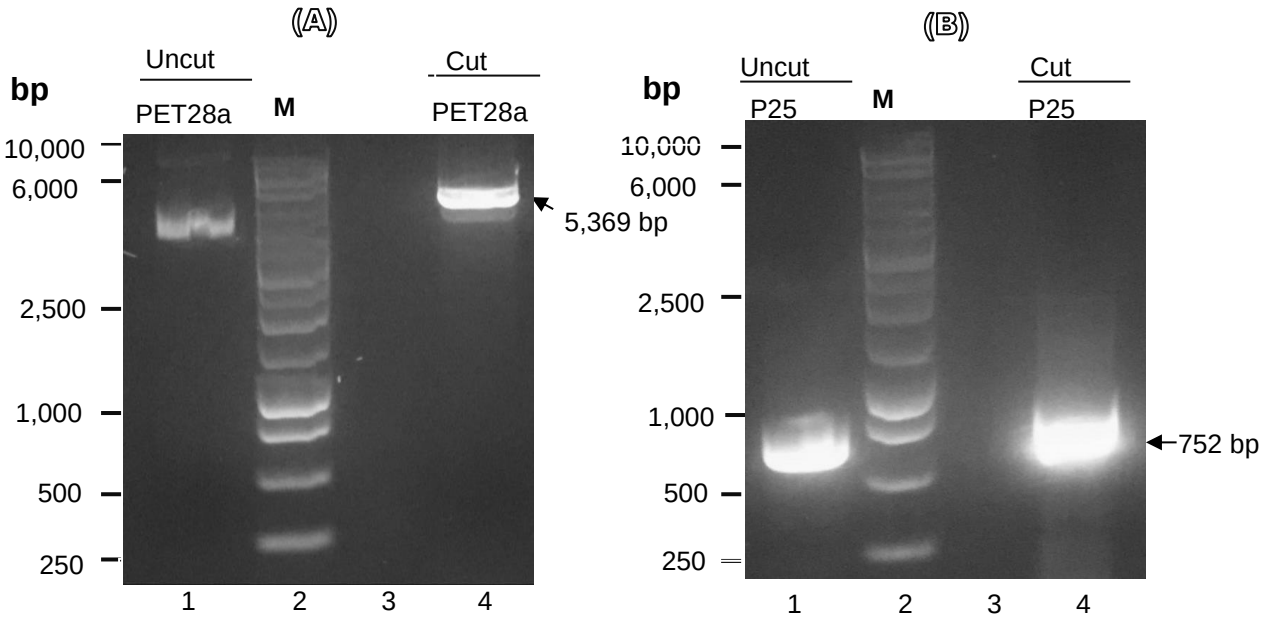


Figure 36. The DNA profile of vector construction between P25 and pET28a was achieved as a bacteria system. Firstly, amplification of the P25 Gblocks containing TEV+His6-tagged at the C-terminal used PCR. The purified P25 DNA and pET28a DNA were consequently used for digestion with EcoRI and XhoI and both reactions of the base pairs of their cut DNAs were measured to assess the DNA concentration and their quality was checked using 1 % agarose gels **(A-B**) and then both digested DNAs were used for ligation**. Figure 36-A** shows DNA bands of vector, uncut pET28a (Lane 1, control), and cut pET28a, Lane 4 (digested with EcoRI and XhoI). **Figure 36-B** shows DNA bands of gene target, uncut P25 (Lane 1, control), and cut P25 (digested with EcoRI and XhoI)

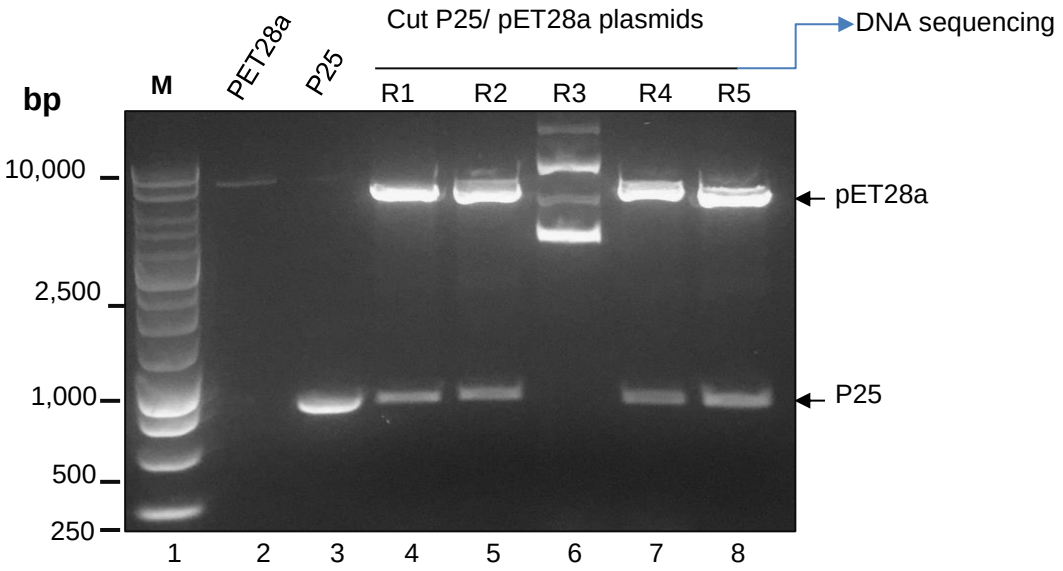


Figure 37. The ligation of the P25 and pET28a DNAs were transformed into DH5 and plated on agar plates containing kanamycin overnight. The next day, in order to verify plasmid DNA after ligation**,** five single colonies were picked and grown overnight in LB at 37ºC with the shaker and subsequently used for DNA extraction using miniPrep. The ligation of these five plasmid DNA was rechecked with digestion enzymes, EcoRI and XhoI, and determined on the gel. Four digested plasmid DNA from the gel, R1, R2, R4, and R5 containing two bands of both P25 and pET28a were sent for DNA sequencing **(C).**

RESULTS 2) Screening expression and optimisation of P25 protein (small-scale production 10 mL) in two bacterial strain groups, cultured at 15, 20, 25, and 28ºC.

There is no previous publication on protocols related to recombinant P25 protein *(B. mori)* production in bacteria. Therefore, the experimental screening expression test and growth optimisation of P25 protein in a selective group, designated as Group 1 (un-enhanced disulfide bond formation), including three bacterial cells BL21 (DE3), C41 (DE3), Rosetta™ (DE3)pLysS(Novagen) was first carried out. However, a comparative study in Group 2, SHuffle® T7 (enhanced disulfide bond formation) was undertaken.

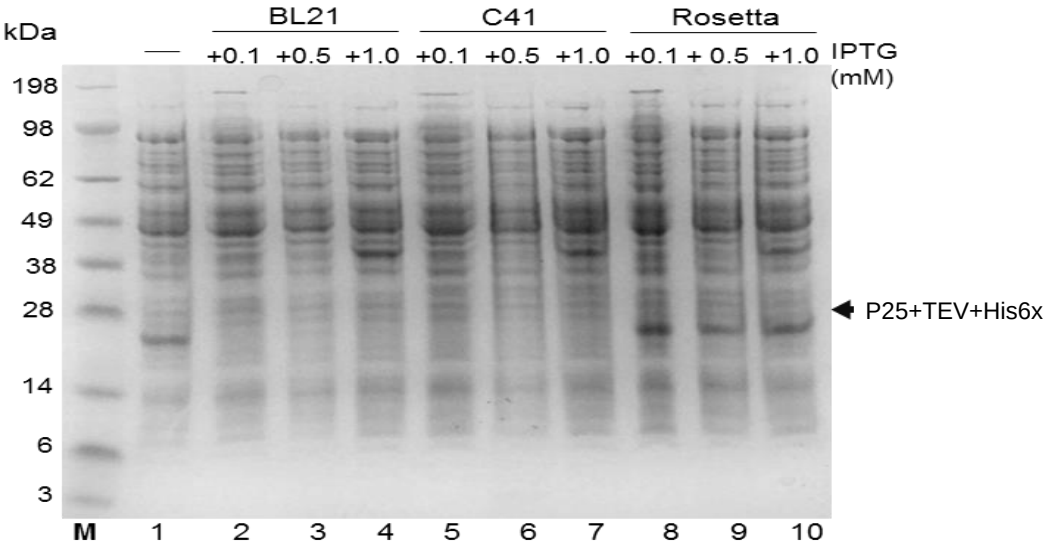
As shown in **Figure 38-41,** briefly, experimental screening was performed on a small scale (10 mL). Group 1 bacterial strains containing P25 were grown under two specific different culture conditions; 1) temperatures, 15, 20, 25, and 28ºC and 2) IPTG concentration, 0.1, 0.5 and 1.0 mM, respectively. All experimental cultures were grown overnight. Two factors of all conditions roughly revealed that P25 protein production in all three strains BL21, C41 (DE3) and Rosetta™ (DE3)pLysS was not so evident

in strong expression, but this expression looked promising when each of the total cells' culture of three strains were disrupted by syringe and the P25 protein band determined using SDS-PAGE and western-blotting. Essentially, P25/C41 western blot images indicated that the target bands of P25 protein had less intensity than the two other strains. There was no best inducer concentration of both factors of IPTG concentrations (between 0.1-1.0 mM) and temperatures (15-25ºC), but P25 protein showed less expression among all three strains when the temperature was increased to 28ºC. However, physically the two strains BL21 and Rosetta™ (DE3)-pLysS cells were more resilient for all conditions than in C41. **(Figure 42 & 43)**, in the experiment with a Group 2 bacteria strain, SHuffle® T7 as a small scale with the conditions varying a little from Group 1, the two main factors being 1) IPTG concentrations, 0.1, 0.5, 1.0, 2.0 mM (added 2.0 mM IPTG only one higher IPTG concentration than Group 1), and 2) temperatures, 15, 20, 25, and 28ºC, respectively. All these cultures were induced overnight before the cells were harvested the following day. Then determination of the P25 protein cell pellet expression was performed both the supernatant and cellular debris were collected for analysis using SDS-PAGE and western blotting. Experimental findings show that all culture conditions of P25/SHuffle® T7 cells were expressed more strongly than in Group 1 bacterial cells when the P25 protein bands were determined using SDS-PAGE and western-blotting. Essentially, the P25 protein had good expression when induced three times at 15, 25, and 28ºC in all IPTG concentrations (0.1, 0.5, 1.0 and 2.0mM). Remarkable intensity bands of P25 protein at around 28 kDa can be seen (SDS-PAGE & western blotting). Some P25/ SHuffle® T7 cells (as controls without IPTG induction) expressed at different induction times (15, 20, 25ºC) were presented in the P25 protein bands. It was indicated that this plasmid can be leaked and the target bands shown resemble IPTG induction There was not much difference in the expression of P25 protein in all four inducer concentrations of IPTG, 0.1, 0.5, 1.0, 2.0 mM, however three inducer temperatures (15, 25 & 28ºC) showed a slightly greater expression than those at 20ºC.

RESULTS 2) Screening expression and optimisation of P25 protein (small-scale production 10 mL) in Group-1 bacterial strain, cultured at 15, 20, 25, and 28ºC

15ºC

(A) SDS-PAGE



(B) Western-blotting

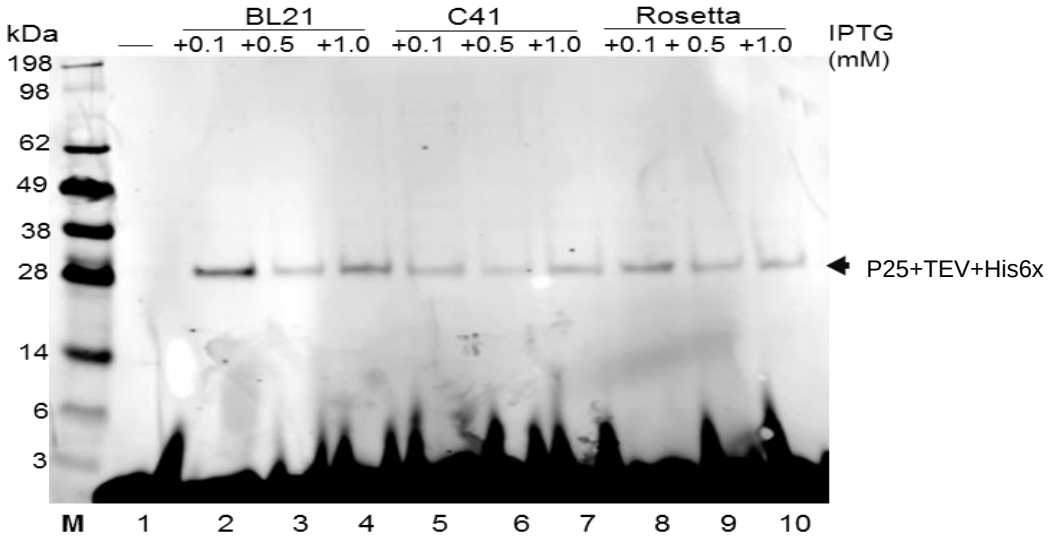


Figure 38. The small-scale induction of P25 protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ (DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) at 15ºC was performed in 10 mL LB supplemented with kanamycin and the expression maintained overnight on the shaker. To do the expression, firstly a single colony of each strain was picked for inoculation overnight at 37ºC. Next morning, the sample of BL21 inoculation was collected before IPTG, un-induced (-) was added as a control. Then 0.1, 0.5, and 1.0 mM IPTG was added to three tubes of the cell cultures to induce P25 expression and maintained overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12%) gels and western-blotting. The P25 protein bands were~28 kDa.

20ºC

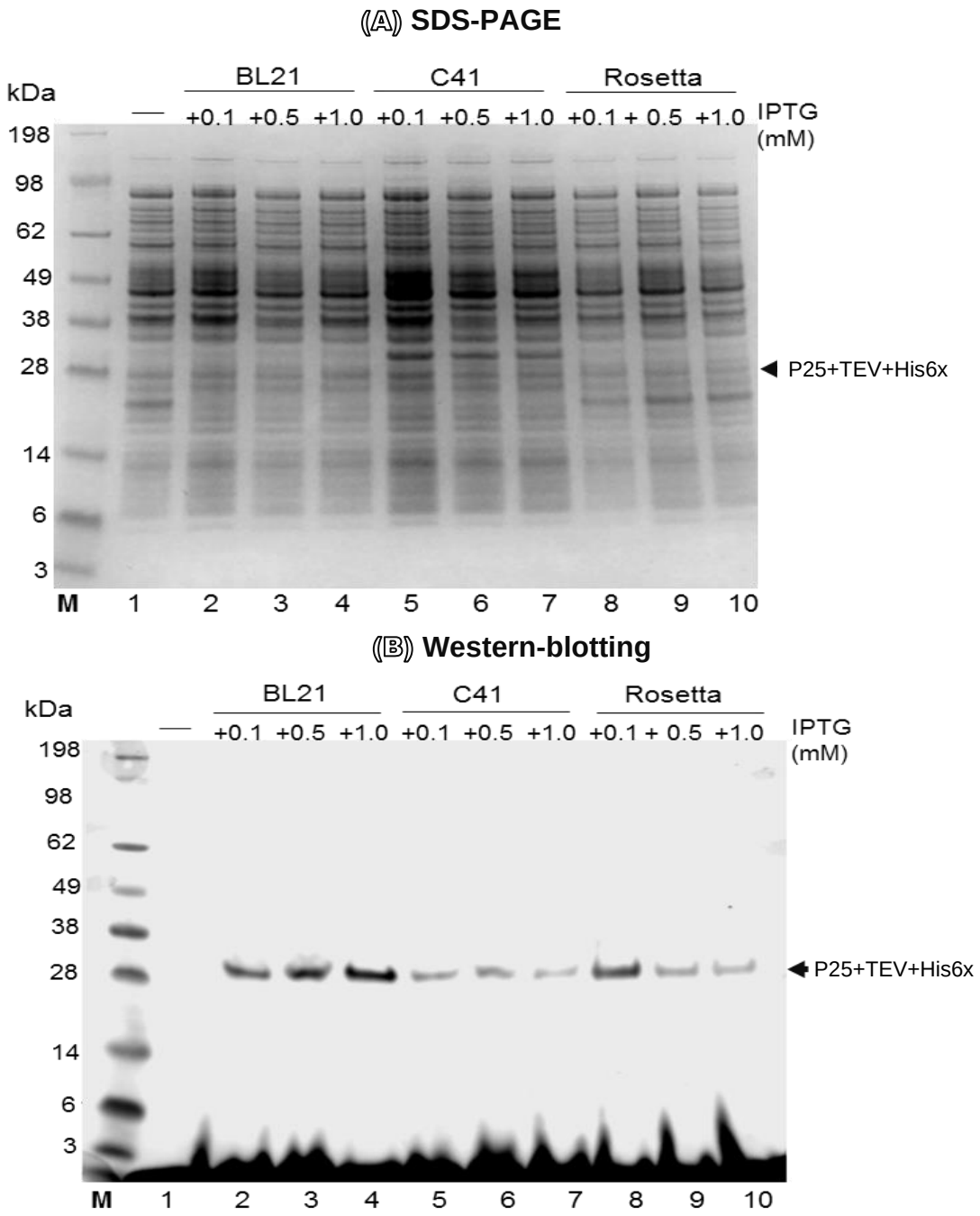


Figure 39. The small-scale ale induction of P25 protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ (DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) at 20ºC was performed in 10mL LB supplemented with kanamycin and the expression maintained overnight on the shaker. To do the expression, firstly a single colony of each strain was picked for inoculation overnight at 37ºC. Next morning, the BL21 inoculation sample was collected before IPTG, un-induced (-) was added as a control. Consequently, 0.1,0.5, and 1.0 mM IPTG was added to three tubes of the cell cultures to induce P25 expression and maintained overnight. Subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12%) gels **(A)** and western-blotting **(B)**. The P25 protein bands were ~28 kDa.

25ºC

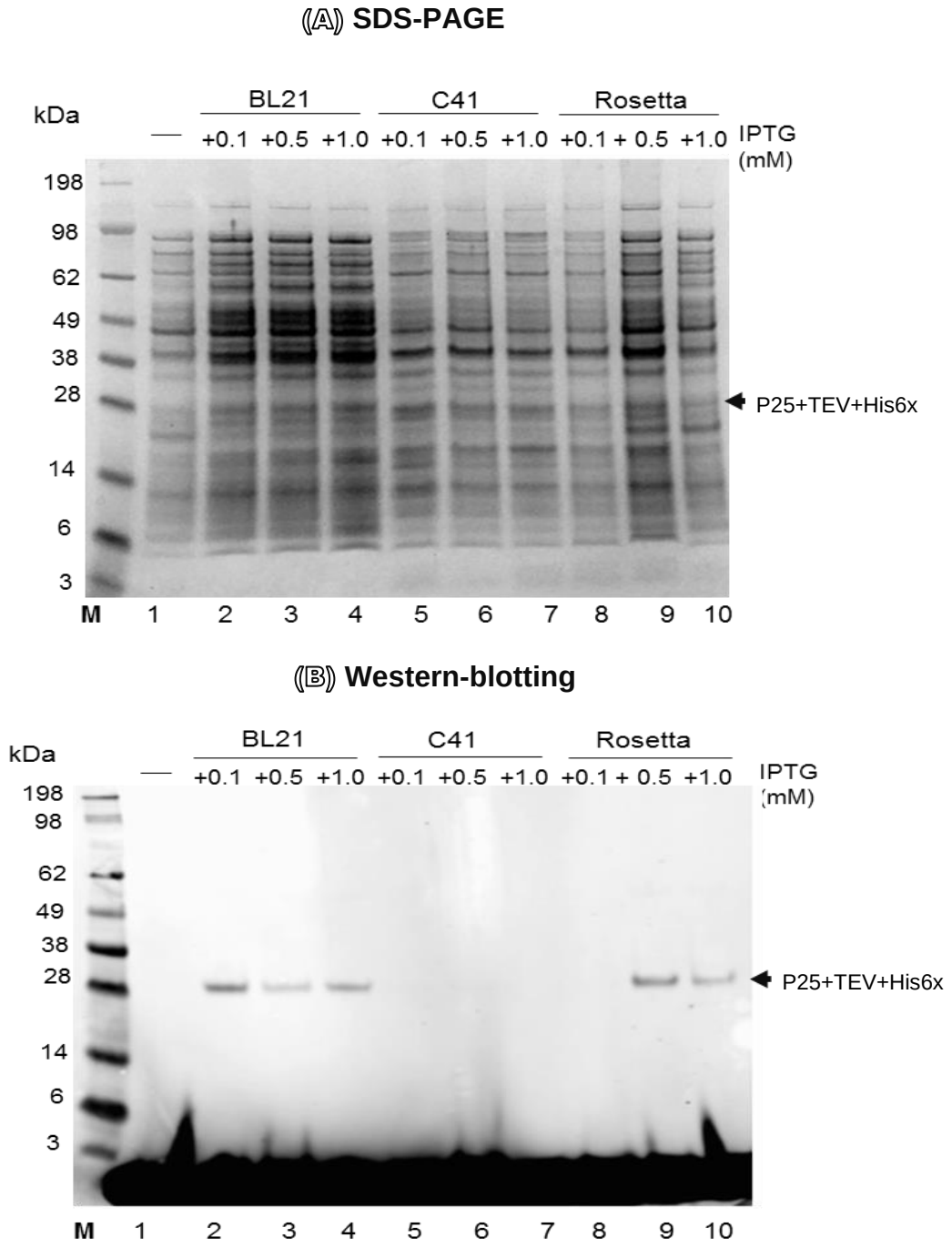


Figure 40. The small-scale induction of P25 protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ (DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) at 25ºC was performed in 10mL LB supplemented with kanamycin and the expression maintained overnight on the shaker. To do the expression, firstly a single colony of each strain was picked for inoculation overnight at 37ºC. Next morning, the sample of BL21 inoculation was collected before IPTG, un-induced (-) was added as a control. Consequently, 0.1, 0.5, and 1.0 mM IPTG was added to three tubes of the cell cultures to induce P25 expression and maintained overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12%) (**A)** gels and western-blotting **(B)**. The P25 protein bands were ~ 28 kDa.

28ºC

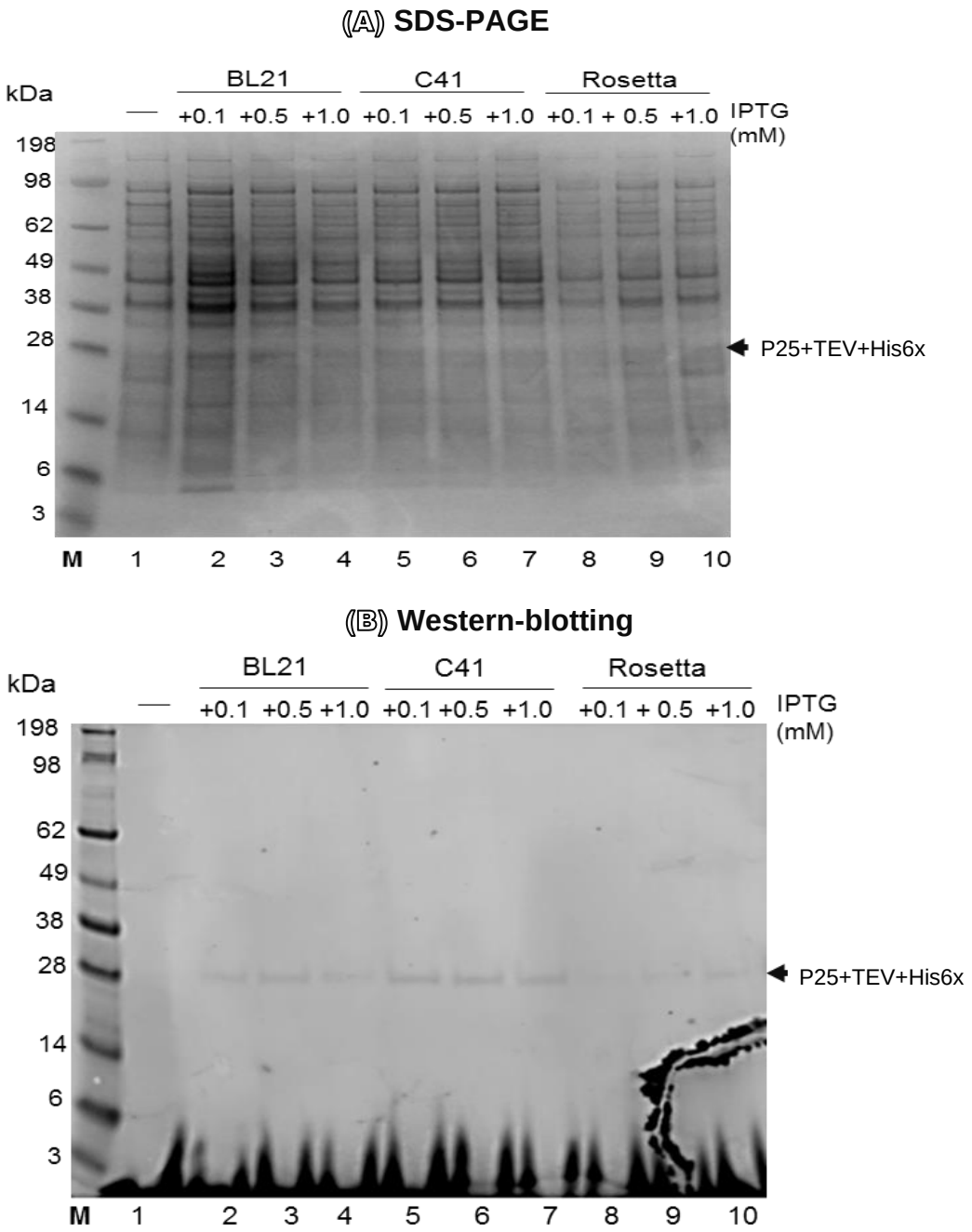
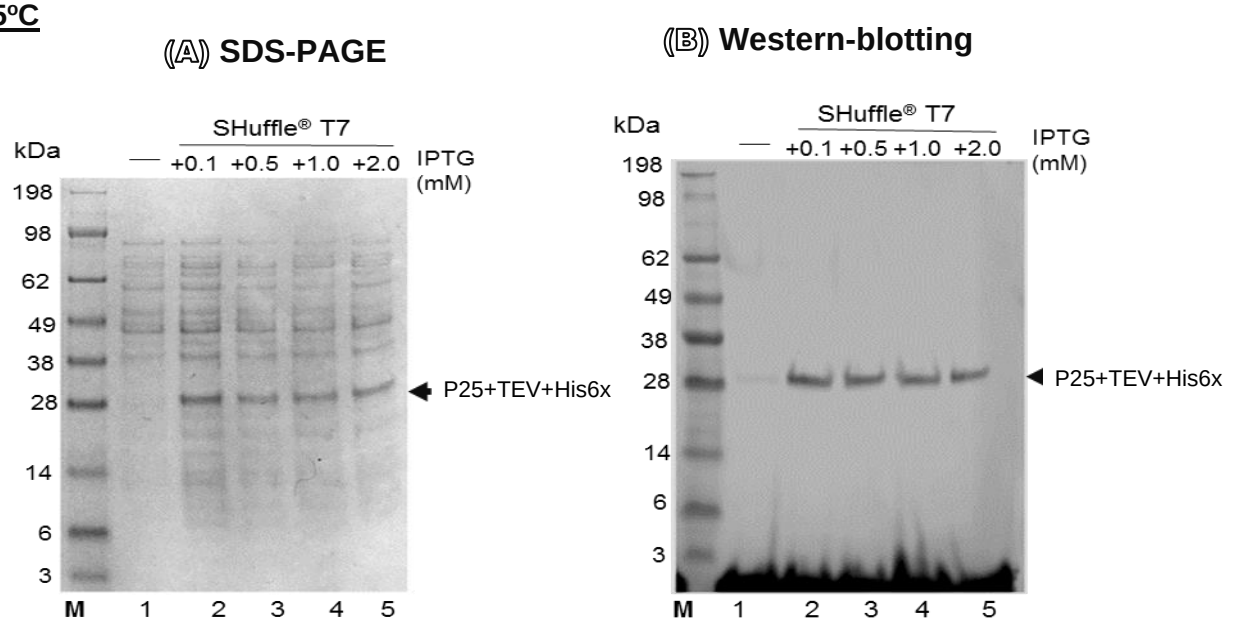


Figure 41. The small-scale induction of P25 protein in Group 1, three bacterial strains, BL21, C41 (DE3) and Rosetta™ (DE3)pLysS with three IPTG concentrations (0.1, 0.5 and 1.0 mM) was performed. To do the expression, firstly a single colony of each strain was picked and inoculated in 10mL LB supplemented with kanamycin and grown at 37ºC overnight on the shaker. Next morning, the sample inoculations of all three strains were collected before IPTG, un-induced (-) was added as a control. Consequently, three IPTG concentrations of 0.1, 0.5, and 1.0 mM IPTG were individually added to the tubes of the cell cultures to induce P25 expression and maintained overnight at 28ºC and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12%) **(A)** gels and western-blotting **(B)**. The P25 protein bands were ~ 28 kDa.

RESULTS 2) Screening expression and optimisation of P25 protein (small-scale production 10 mL) in Group-2 a bacteria strain, cultured at 15, 20, 25, and 28ºC

15ºC



20ºC

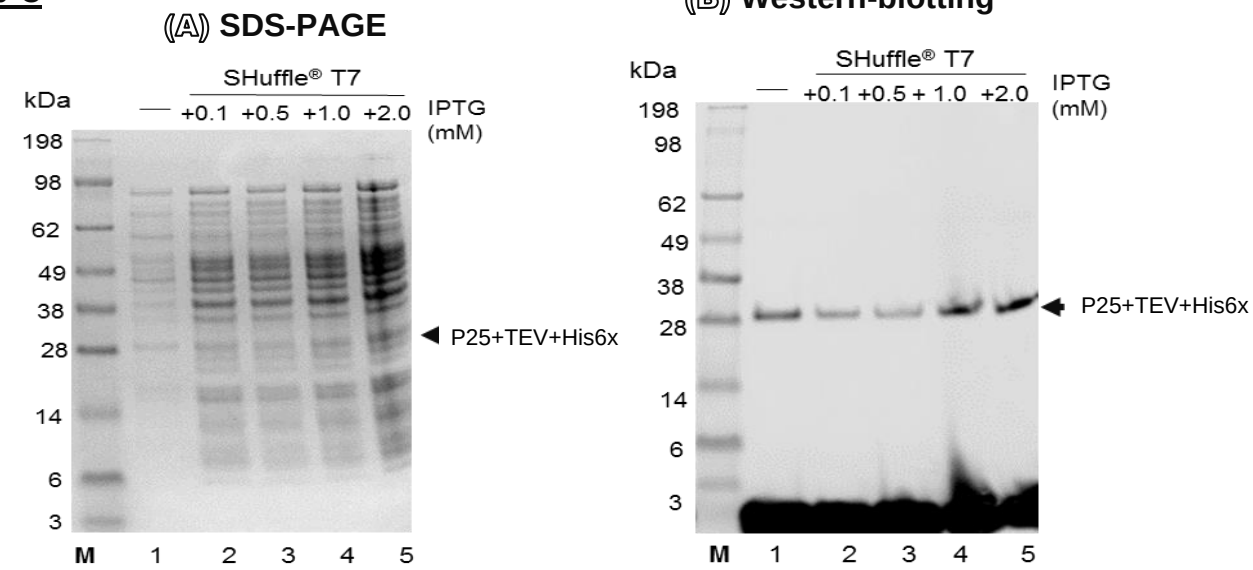


Figure 42. In the comparative small-scale induction of P25 protein in Group 2, a bacteria strain (SHuffle® T7), with four IPTG concentrations (0.1, 0.5,1.0, and 2.0 mM) at two temperatures 15 or 20ºC, a single colony was inoculated in 10mL LB supplemented with kanamycin and grown overnight on the shaker at 37ºC. Next morning, a tube of the sample SHuffle® T7/25 inoculation was collected for analysis before IPTG, un-induced (-) was added as a control. Consequently, 0.1,0.5, 1.0 & 2.0 mM IPTG was separately added to four other tubes of the cell cultures to induce P25 expression and maintained at 15 or 20ºC overnight and subsequently the cell pellets were collected, the total cells disrupted with 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12%) gels **(A)** and western-blotting **(B)**. The P25 protein bands were about 28 kDa.

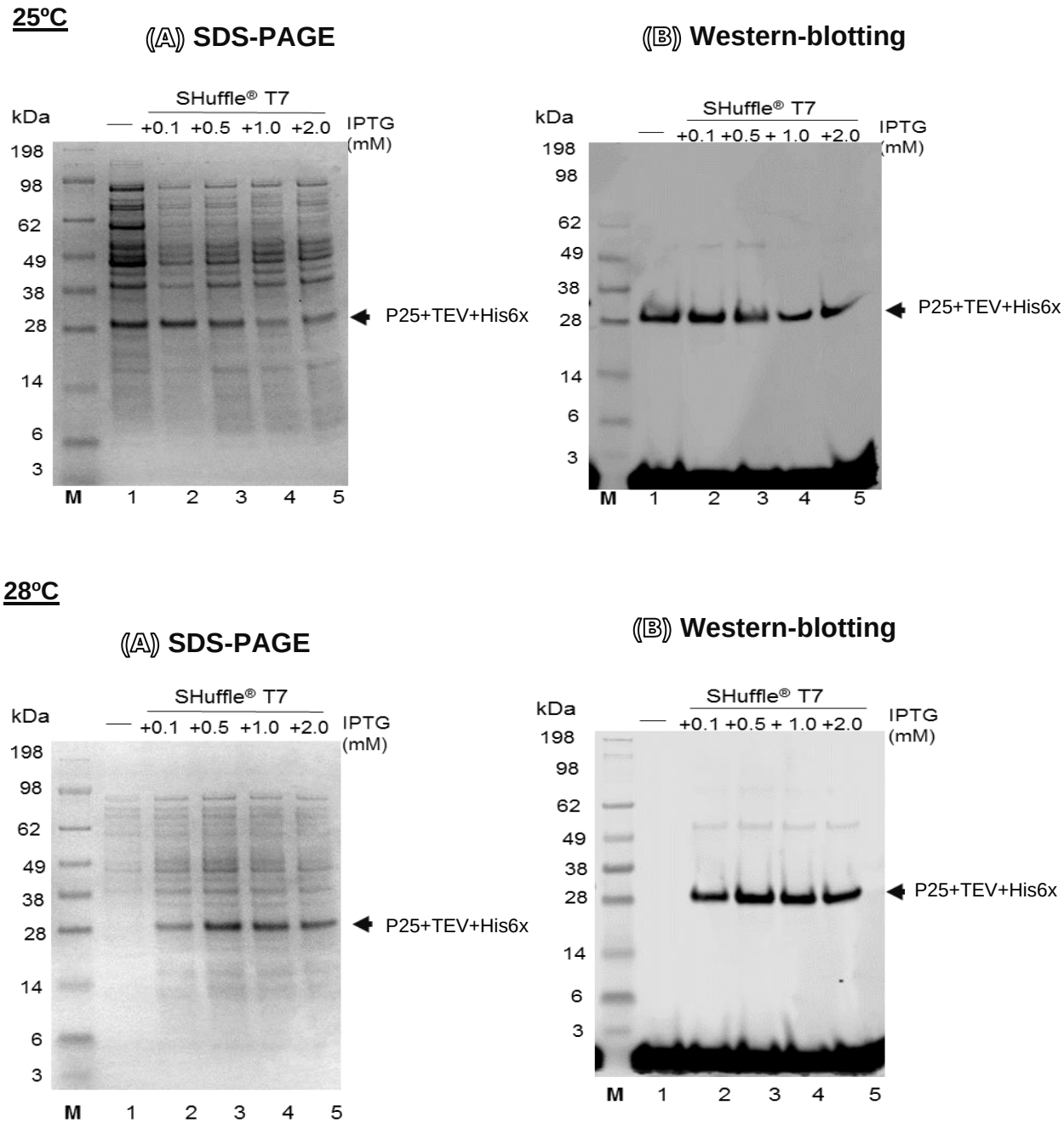


Figure 43. The comparative small-scale induction of P25 protein in Group 2, a bacteria strain (SHuffle® T7), a single colony with four IPTG concentrations (0.1, 0.5,1.0, and 2.0 mM) at two temperatures 25 or 28ºC was inoculated in 10mL LB supplemented with kanamycin and grown overnight on the shaker at 37ºC. Next morning, a tube of the sample SHuffle® T7/ 25 inoculation was collected for analysis before IPTG, un-induced (-) was added as a control. Consequently, 0.1,0.5, 1.0 and 2.0 mM IPTG were separately added to four other tubes of the cell cultures to induce P25 expression and maintained at 25 or 28ºC overnight and subsequently the cell pellets were collected, the total cells disrupted in 1XTBS buffer and analysed with SDS-PAGE (bis-tris 4-12%) gels **(A**) and western-blotting **(B)**. The P25 protein bands were ~ 28 kDa.

RESULTS 3) The method of P5 Protein solubilsation, purification and its localisation (250mL cell culture)

Figure 44, previous experiments with a small scale of the P25 protein production revealed the level of expression in two main groups of *E. coli* strains. Therefore, P25 protein localisation protocol was set up by collecting cell pellets, protease inhibitor added, the cell pellets disrupted and both the supernatant and the inclusion bodies pellets/ IBs were retained. The findings of two groups of bacteria expressions are separately shown below in individual topics 3.1 and 3.2, respectively.

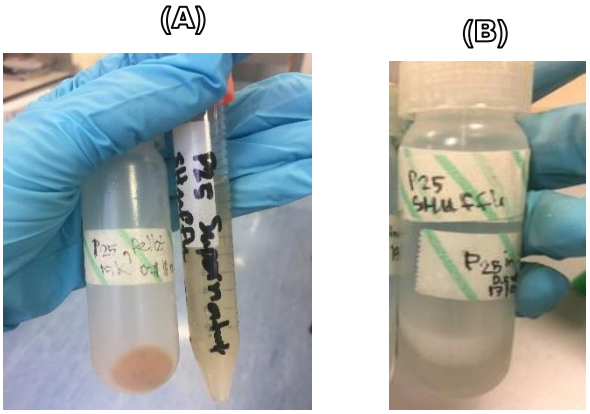


Figure 44. The sample cells of the lysate mixture disrupted by FRENCH Press were obtained by centrifugation (15,000 rpm at 4ºC for 30 mins) in order to collect both the supernatant and inclusion bodies pellets. These two fraction parts were forwarded for P25 protein localisation tests in both groups **(A**). The IBs was being solubilising in 7M urea **(B**).

3.1 Determination of P25 protein localisation in bacteria cells Group 1 non-enhanced capacity of disulfide bonds folding.

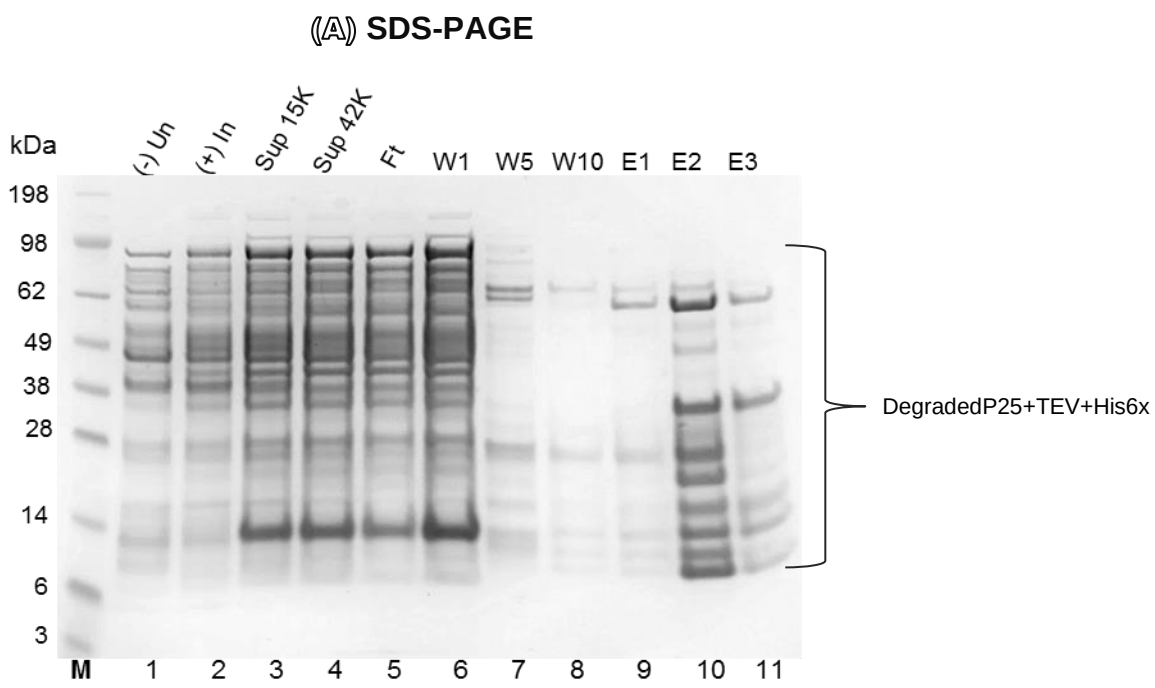
(**Figure 45-46**), from the screening test in Group 1 bacteria cells BL21, C41 (DE3), Rosetta™ (DE3)pLysS in different concentrations of IPTG and temperatures, it was found that two strains BL21, Rosetta™ (DE3)pLysS seemed more promising for expressing P25 protein. Therefore, these two strains were used for further experiments. The P25 protein localisation was firstly monitored in the supernatant (cytoplasm expression) after the cells were disrupted using the FRENCH Press, as well as the pellets (IBs pellet) later being solubilised and purified using Ni-NTA. However, after purification, P25 protein was not present in the supernatant of both strains but multiple bands of degraded purified P25 were obtained after elution with elution buffer. This indicates that P25 was not present in the soluble form of the supernatant from disrupting the cells, but it is found to be in the IBs pellet in insoluble form which is shown in **(Figure 47-48). Figure 4**7, the P25 protein clearly appeared in the IB pellet fractions of both strains with multiple bands of intensity on SDS-PAGE, but more specific bands on the western-blotting. **Figure 48**, the IBs pellet parts from two bacterial hosts, BL21and Rosetta™ (DE3)pLysS were used to solubilise by re-suspending in 7M urea. The findings of each solubilisation from two strains were subsequently refolded and purified through Ni-NTA column. The purified P25 bands on the SDS-PAGE and western-blotting of both strains in BL21and Rosetta™

(DE3)pLysS were very faint, but the purified P25 proteins seemed to be at the right molecular weight of ~28 kDa.

3.2 Determination of P25 protein localisation in a selective bacteria strain Group 2 enhanced the capacity of disulfide bonds folding.

Prior experimental efforts of expressing P25 protein in Group 1 bacterial strains showed little promise. Therefore, an experimental pilot scale in Group 2, SHuffle® T7 was carried out in the same way as the previous strains (Group 1).

(**Figure 49A1-A2** & **49B1-B2**), the experimental findings clearly showed the P25 protein was present in the pellet part (Pell 15K) after disruption of the cells. This presence of P25 was determined by SDS-PAGE and western blotting ~28 kDa. A trace of P25 was also found when the pellet part was determined with the solubility assay. In contrast, obtaining the supernatant of Sup 15K after centrifugation was done by disruption using the FRENCH Press. If this supernatant was employed for purification through the Ni-NTA column and eluted with elution buffer, the eluted P25 protein band (called inactive form) was not homogeneous. These nonspecific bands would be the degraded P25 protein. The P25 band in SDS-PAGE was shown to have lower and higher multiple bands whereas the western blotting band of P25 had lower MW than the actual band of P25 ~28 kDa (**Figure 50**).



(Continued this result and description (western blotting image) in Figure 45B. Next page)

(B) Western-blotting

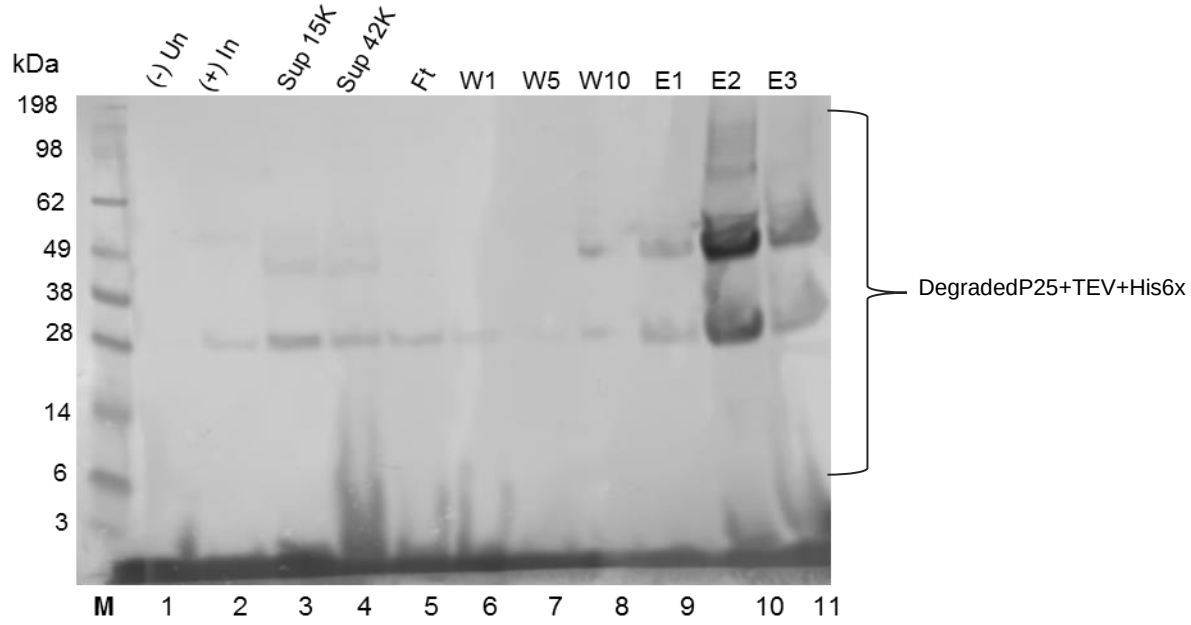


Figure 45. SDS-PAGE **(A)** and Western-blotting **(B)** show the purification profile of P25 expressed in BL21 from the supernatant fraction. The bacterial cells were harvested after the O.D culture reached 6.0 as un-induced IPTG ((-) Un) and after being induced with 0.5 mM IPTG overnight ((+) In). This was followed by cell disruption using the FRENCH Press and the resulting lysated cells were centrifuged for supernatant (15,000 rpm for 30 mins (Sup 15K), then ultracentrifuged for supernatant (42,000 rpm (Sup 42K) for 1hr). With the His GravityTrap purification the first flow through affinity column (Ft), was followed by the 1st wash by Wash buffer (W1), 5th wash (W5), and 10th wash (W10), till the eluted P25 proteins (E1-E3) were evident. The bands were expected to be ~28 kDa.

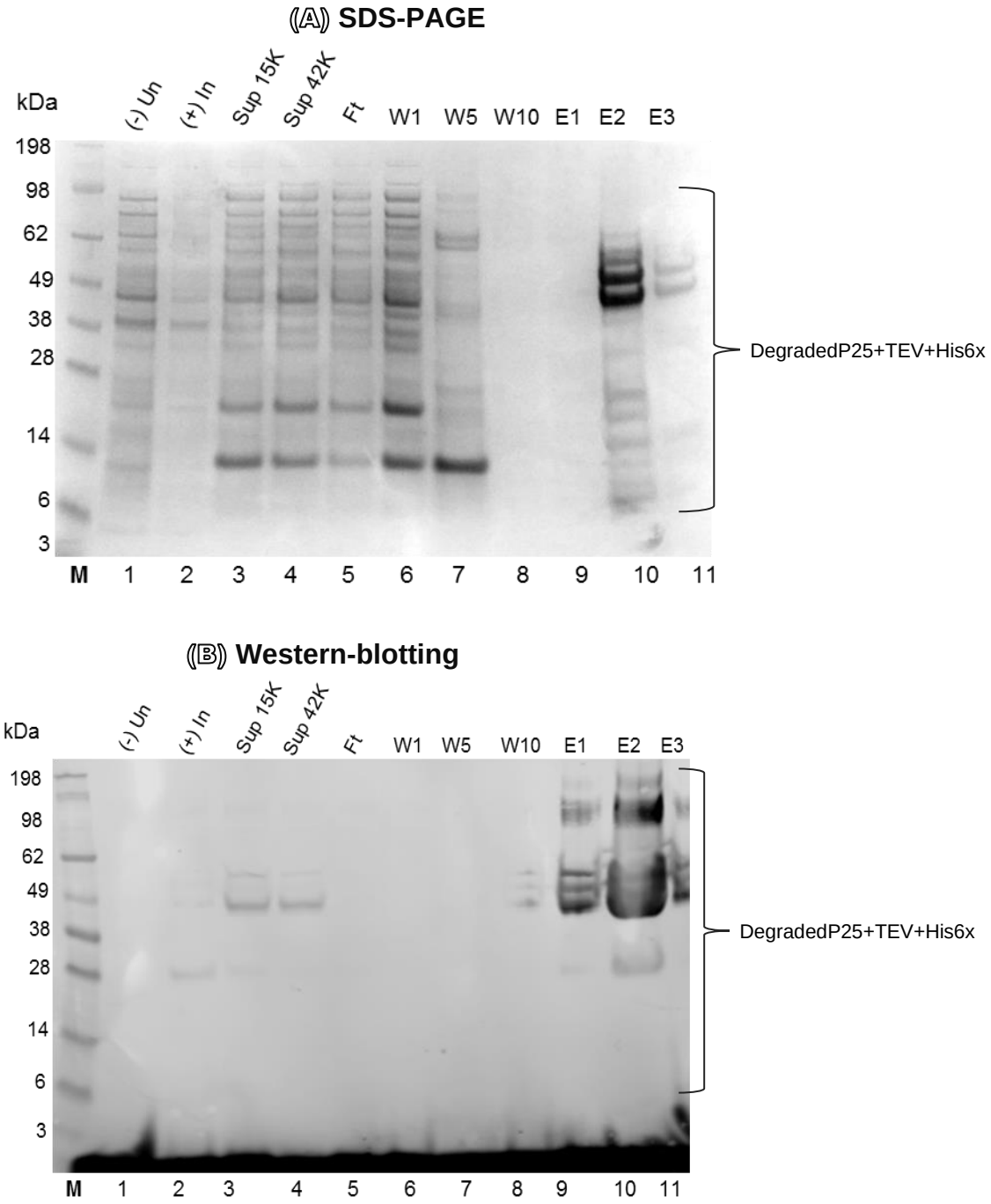


Figure 46. SDS-PAGE **(A)** and western-blotting **(B)** show the purification profile of P25 expressed in Rosetta™ (DE3)pLysS from the supernatant fraction. The bacterial cells were harvested after the O.D culture reached 6.0 as un-induced IPTG ((-) Un) and after being induced with 0.5 mM IPTG overnight ((+) In). This was followed by cell disruption using the FRENCH Press and the resulting lysated cells were centrifuged for supernatant (15,000 rpm for 30 mins (Sup 15K), then ultracentrifuged for supernatant (42,000 rpm (Sup 42K) for 1hr). With the His GravityTrap purification the first flow through affinity column (Ft), was followed by the 1st wash by wash buffer (W1), 5th wash (W5), and 10th wash (W10), till the eluted P25 proteins (E1-E3) were evident. The bands were expected to be ~ 28 kDa.

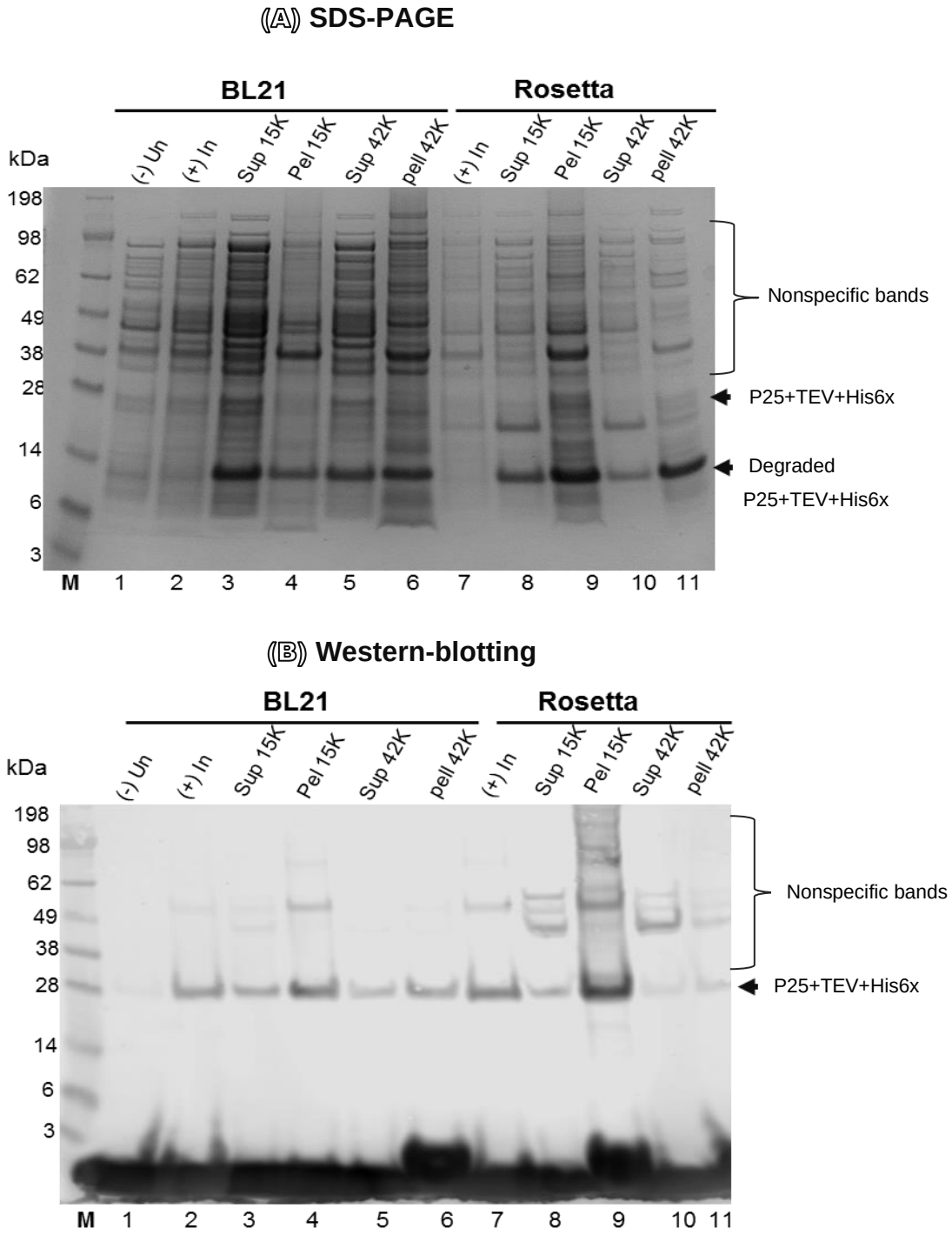


Figure 47. SDS-PAGE **(A)** & western blotting **(B)** of the 250 mL scale culture of P25 expressed in BL21 & in Rosetta™ (DE3)pLysS. The bacterial cells were harvested after the O.D culture reached 6.0 as un-induced IPTG ((-) Un) and after being induced with 0.5 mM IPTG overnight ((+) In). Next, the cell was disrupted by FRENCH Press and the resulting lysated cells were then spun down (15,000 rpm (15K) for 30 mins). Both supernatants of 15 K (Sup 15K) and its pellets (Pell 15K) were then collected. Consequently, the high-speed centrifugation, known as ultracentrifugation at 42,000 rpm (42K) for 1hr was performed and both the resultant supernatant and pellets (Sup 42K & Pell 42K) were collected. The P25 protein solution that was obtained was purified using the His GravityTrap purification process through affinity column. The bands were expected to be ~28 kDa.

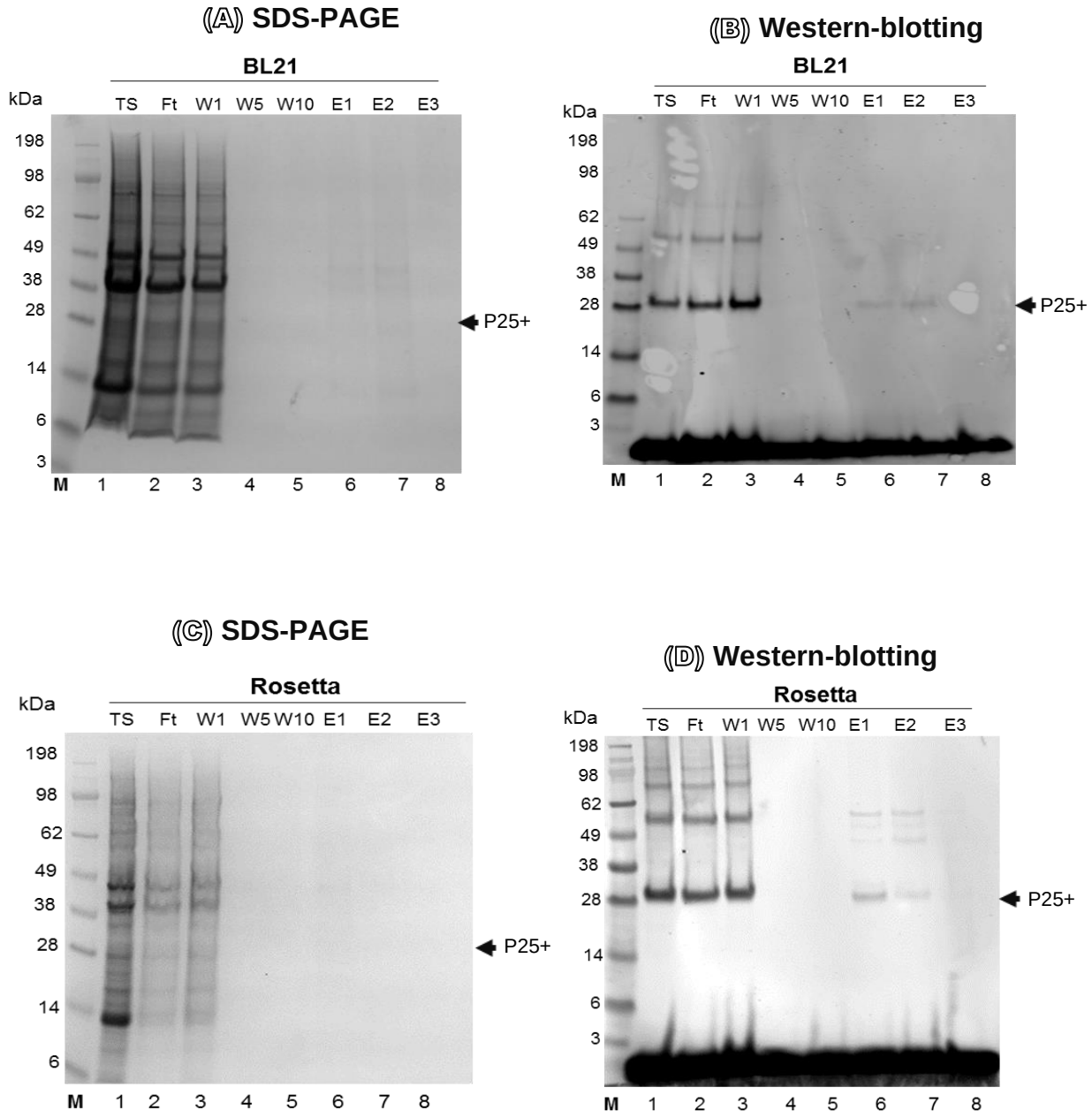


Figure 48. Solubilisation of P25 protein IB pellets (pellets 15 K after disruption by the FRENCH Press) was carried out on both BL21 and Rosetta™ (DE3)pLysS strains using 7M urea. The process used the IB pellets from the cell culture 250mL induced with 0.5 mM IPTG at 25ºC overnight. The IB pellets of both strains that were obtained were solubilised by re-suspension with 7M urea and incubated overnight. The following day, each solubilised P25 protein mixture of the two bacterial strains was pelleted out and the total soluble (TS, Lane 1) was maintained, and subsequently passed through the Ni-NTA column. Each flow through of both strains was collected (Ft) for analysis after the columns had been washed 10 times (1 mL/each) with Wash buffer, W1 (1st washed), W5 (5th washed), W10 (10th washed) respectively. Lastly, each purified P25 protein was eluted with Elution buffer for three mL, E1, E2, and E3 (1mL/each). The selective processes were determined using SDS-PAGE stained with Coomassie blue and western-blotting using anti-His6-tagged antibody **(Figures A-B**, BL21), **C-D** (Rosetta™ (DE3)pLysS). The P25 protein bands were expected to be ~28 kDa.

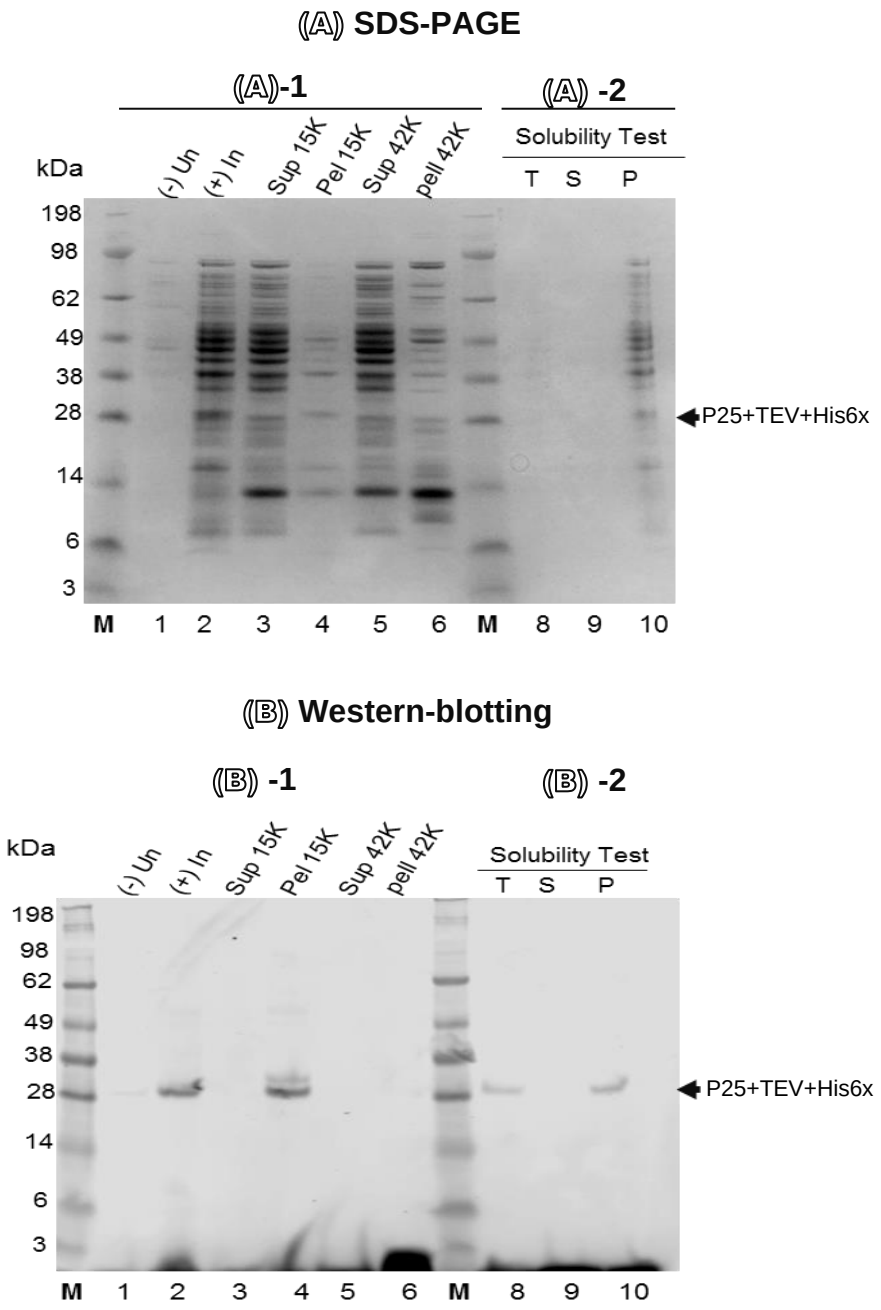


Figure 49. Fractions of P25 expression in SHuffle® T7, cultured in LB from pre-induction of IPTG in cell ((-), Un), post-induction ((+), In) with 0.5 mM IPTG containing Kanamycin at 25ºC were treated overnight on the shaker. The cells were harvested, 1XTBS and a tablet of protease inhibitor were added, and the cells were disrupted with the FRENCH Press. The disrupted cell lysate was then harvested and the 15K supernatant and 15K pellet (Sup 15 K, Pell 15K,) centrifuged at 15,000 rpm for 30 mins at 4ºC. The 15K supernatant was consequently centrifuged at 42,000 rpm (42K) for 50 mins at 4ºC and (Sup 142 K, Pell 142K) harvested to determine the P25 protein localisation **Figure (A1-B1).** In another test, the induced cells ((+), In) were collected (the cells spun down at 7K rpm for 5 mins), 1XTBS added, then the cells re-suspended, disrupted by syringe and eventually the total cell lysate (T) obtained. This total cell lysate was subsequently centrifuged again and the supernatant (S) and pellets (P) collected. The solubility test showed P25 in SHuffle® T7 located in the cell pellet determined by SDS-PAGE & western blotting. The expected band of P25 was ~28 kDa. **Figure (A2-B2)**.

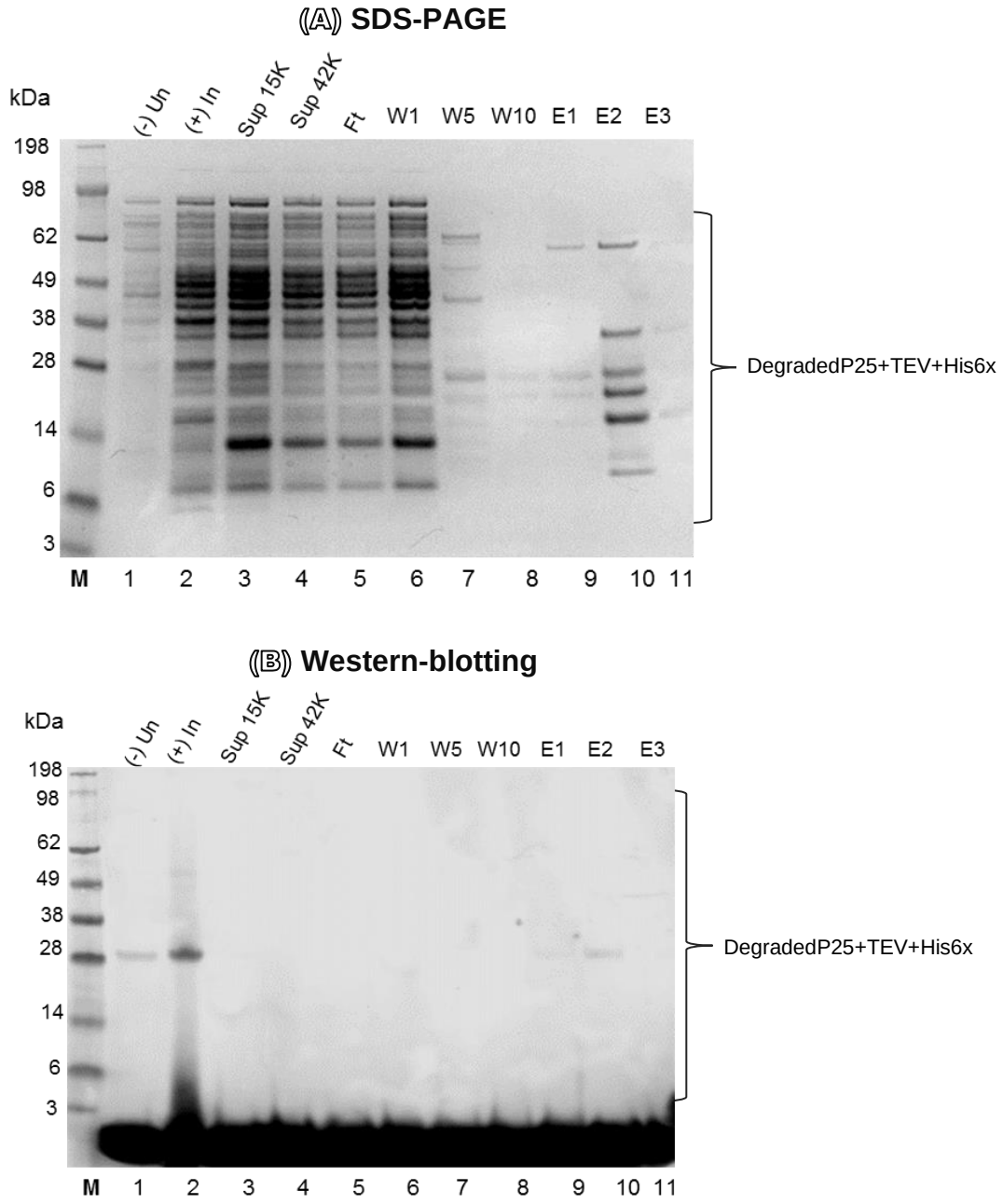


Figure 50. The purification of P25 expressed in SHuffle® T7 from the soluble lysate cell. Initially from the cell disruption (supernatant), cells were collected from the pre-induction cells (un-induced ((-) Un), induced ((+) In with 0.5mM IPTG and maintained at 25ºC overnight on the shaker. The following day, cells were harvested by centrifugation 7K (7,000 rpm) for 5 mins at 4ºC, 1XTBS and a tablet of protease inhibitor added, and the cells disrupted with the FRENCH Press. The supernatant of the disrupted cell lysate was then harvested (15,000 rpm, (Sup 15K) at 30 mins at 4ºC) and was immediately purified using Ni-NTA column protocol, the P25 mixture passed through the column, the flow through (Ft) collected, followed by the column being washed ten times (1mL/each) with wash buffer, W1 (1st washed), W5 (5th washed), W10 (10th washed) and eventually the purified P25 protein was eluted with three mL elution buffer (E1-E3). SDS-PAGE **(A)** and western-blotting **(B)** showed the expected band of P25 ~28 kDa.

4) Scale-up expression, solubilisation, and purification (500 mL/Flask) Protocol for P25 protein production in bacteria system

The P25/SHuffle® T7 showed the greater expression among the bacterial strains in the insoluble form of protein. Therefore, the P25 protein production was cultured in a 500 mL/flask with LB media. The protocol is generated specifically for refolding P25 protein from SHuffle® T7 cells. The process essentially involves collecting only the 15K pellet (IB pellets) for solubilisation with 7M urea containing 100 mM HEPES, pH 8.0 overnight. The supernatant of soluble P25 protein (7M urea containing 100 mM HEPES, pH 8.0) was collected twice by centrifugation at 15,000 rpm. The 50 mM Tris-HCl, pH 8.0 was added to the resulting soluble mixture of P25 protein with 7M urea to reduce its viscosity and incubated at room temp for 30 mins before being passed through the Ni-NTA column using His-tagged purification as described previously. The eluted purified P25 protein showed up as a single band at about 28 kDa **(Figure 51 A-B).**

5) Characterisations of P25 protein expressed in bacterial system

5.1 P25 Protein storage stability

An eluted purified P25 protein in elution buffer (containing a single band) was initially preserved at -20ºC until further use. To test the protein stability, the eluted purified P25 protein was split into 600µL from the -20ºC stock and subsequently maintained at 4ºC. A sample of 80 µL was taken from the stock from the first to the tenth day. 2x of protein sample buffer was added to each sample and boiled at 85ºC for 15 mins and stored in the refrigerator until all ten days of samples were ready to be analysed with SDS-PAGE and western blotting. The images of SDS-PAGE and western blotting showed that P25 protein can be maintained at good quality for a week at 4 ºC. After that, the P25 protein gradually degraded day by day, indicated by the band of P25 protein gradually decreasing after seven days of storage at 4ºC **(Figure 52 A-B)**.

5.2 P25 protein identification using mass-spectrometry analysis

A single band of eluted purified P25 protein was run on the SDS-PAGE gel and cut for analysis via a mass-spectrometry machine. The findings of the mass-spectrometry indicated that the recombinant P25 protein expressed in SHuffle® T7 is similar to the precursor of native fibroin (*B. mori*) if compared to NCBI database **(Figure 53A-C)**.

5.3 P25 protein secondary structure determination using CD technique.

It was found that P25/SHuffle® T7 protein expression in the cells and solubilised from the IB pellets was not stable even when undergoing the same procedures. This phenomenon was recognised after three months of transformation of P25 DNA to SHuffle® T7 cells. In addition, the P25 protein showed high aggregation behaviour when it was changed into different buffers. Because of this inconsistency in expression of P25 SHuffle® T7**,** further analysis for the CD of the secondary structure could not be carried out. However, it was necessary to discover the most suitable buffer for P25 protein. Therefore, four selected buffers, including Acetate buffer pH 5.0, Citrate buffer pH 6.0, HEPES buffer pH 7.0 and Tris-HCl buffer pH 8.0, were used to test the P25 protein reaction using a dynamic light

scattering machine. The findings indicated that P25 protein would have less aggregation in Tris-HCl pH 8.0 **(Figure 54).**

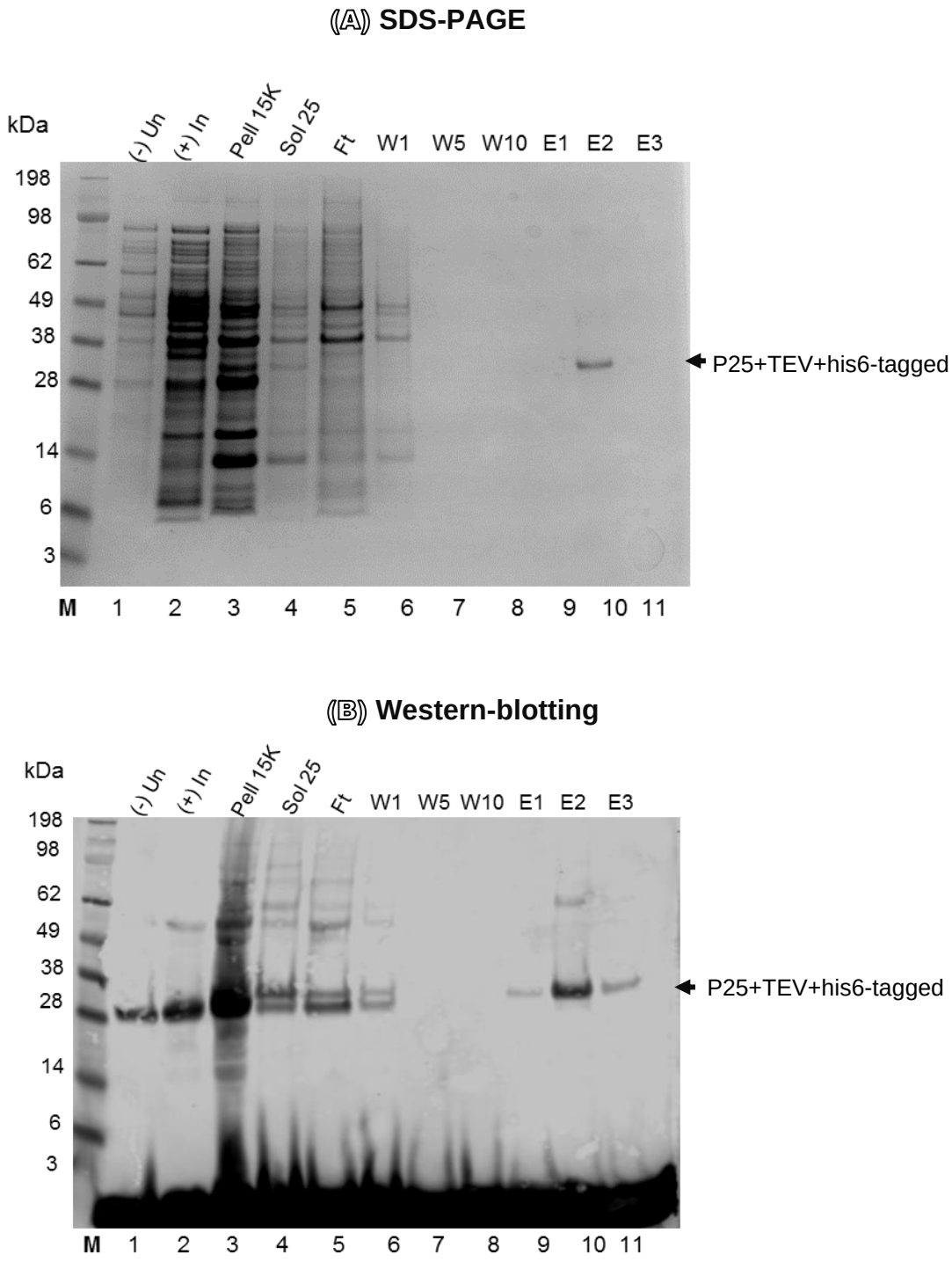


Figure 51. Shows the protein fractions from each process through un-induced cells (Un), induced cells with 0.5 mM IPTG(In), followed by the 15K (15,000 rpm) pellet of P25 in SHuffle® T7 and collected for solubilisation (15,000 rpm,15K). The resulting pellet was solubilised with 7M urea for 3 series and combined as a P25 (So). At this step, the soluble P25 was purified with His-tagged gravity of Ni-NTA column with a final concentration 0.23 mg/mL. SDS-PAGE **(A)** and western-blotting **(B)** revealed the expected target bands of P25 at ~28 kDa

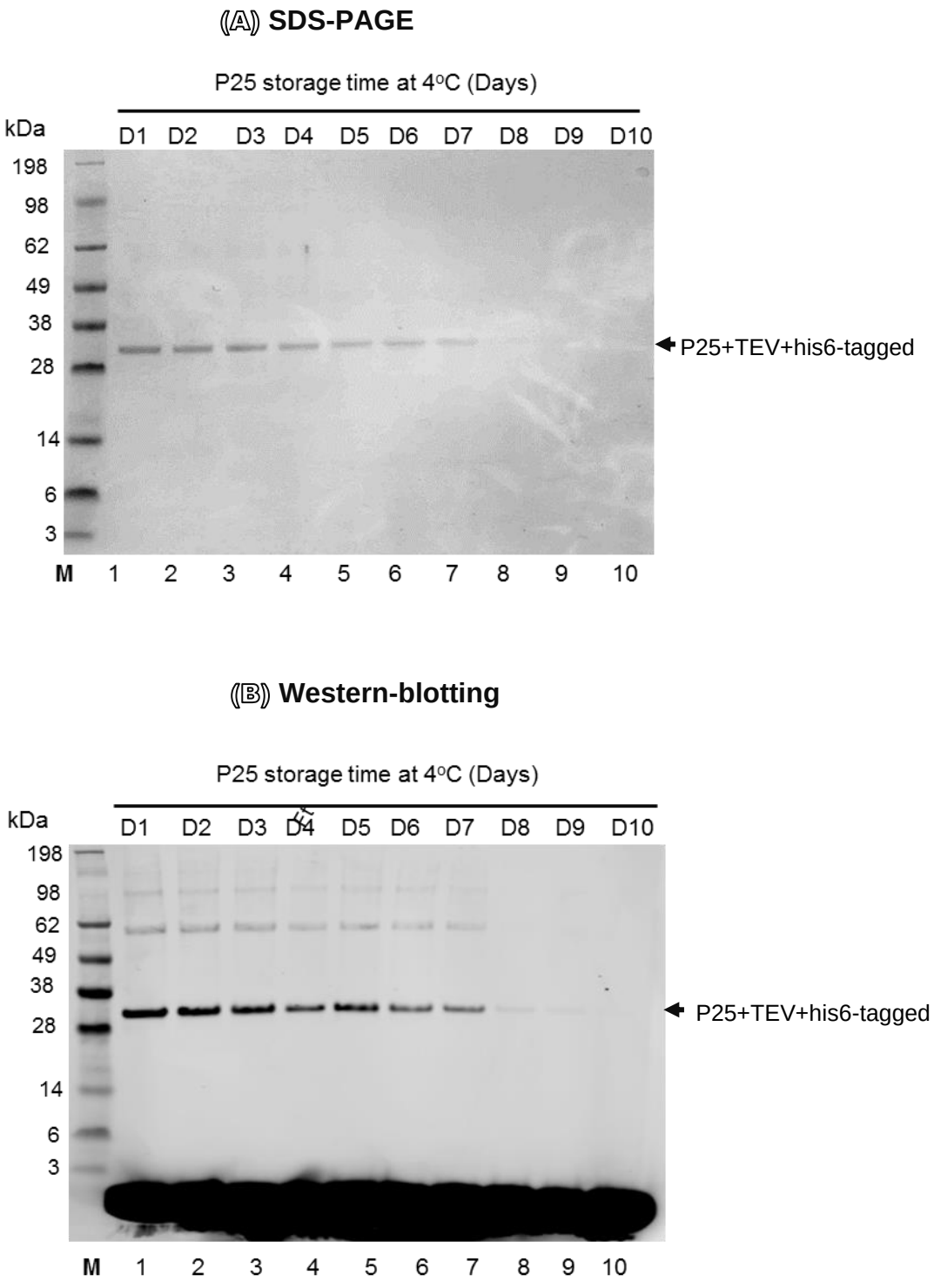
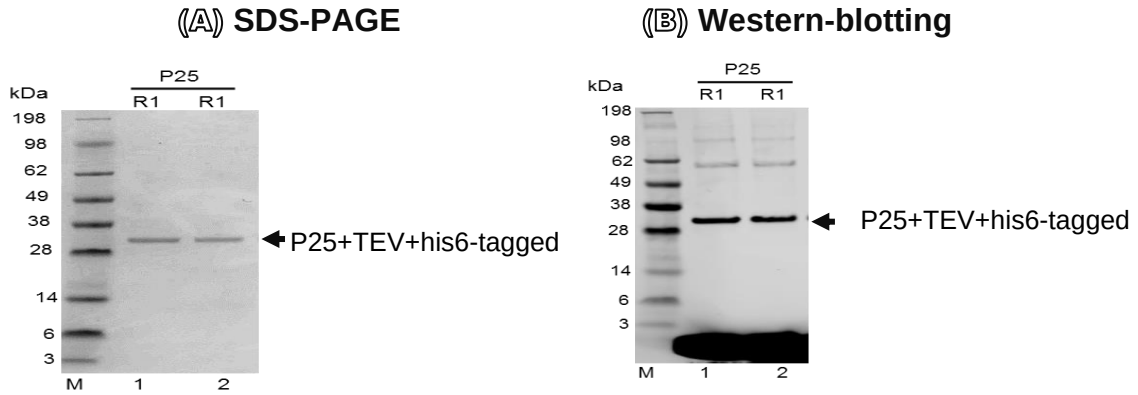


Figure 52. The eluted purified P25 protein expressed in SHuffle® T7 was firstly preserved at -20ºC and its stability was consequently tested at 4ºC from the first to the tenth day. Each day, 2x protein buffer was added to a single sample, boiled at 85ºC for 15 mins before being loaded onto the two bis-tris 4-12 % gels (SDS-PAGE). Having finished running the two gels, one SDS-PAGE **(A)** was stained with Coomassie blue and another gel was subjected to western blotting **(B)**. The P25 protein was present at ~ 28 kDa.



(C) Mas-spec of P25 protein

e-mail :
Search title : Chirapha P25 NCBI ESI, ChiraphaP25_R.mgf
MS data file : \\cfs.st-andrews.ac.uk\Shared\Chem_PIs\BSRC_Mass_Spec\A MGF REPOSITORY\Data from 5600+\190419\ChiraphaP25_R.mgf
Database : NCBIprot 20190208 (187,857,634 sequences; 68,510,950,169 residues)
Timestamp : 22 Apr 2019 at 09:09:54 GMT

Not what you expected? Try the peptide summary.

▸ Search parameters
▸ Score distribution
▸ Modification statistics
▸ Legend

Protein Family Summary

Significance threshold p< 0.05 Max. number of families AUTO
Display non-sig. matches Dendrograms cut at 0
Show Percolator scores
Preferred taxonomy All entries

▸ Sensitivity and FDR (reversed protein sequences)

Protein families 1–10 (out of 15)

10 per page 1 2 Next

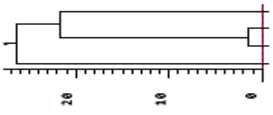
▾1 NP_001037488.1 245 fibroin light chain precursor [Bombyx mori]

		Score	Mass	Matches	Sequences	emPAI
1.1	NP_001037488.1 fibroin light chain precursor [Bombyx mori]	245	27908	6 (6)	6 (6)	1.81

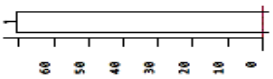
▾6 peptide matches (6 non-duplicate, 0 duplicate)

Query	Dupes	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
182		578.7405	1155.4665	1155.4720	-4.78	0	46	0.0052	▸1	U	R.WDYVDDTDK.S
241		656.7880	1311.5615	1311.5732	-8.89	1	69	0.0025	▸1	U	R.RWDYVDDTDK.S
249		661.3151	1320.6157	1320.6245	-6.64	0	75	1.2e-05	▸1	U	R.YSVGPALGCAGGGR.I
264		670.9073	1339.8000	1339.8075	-5.54	0	78	0.00034	▸1	U	K.SIAILNVQEILK.D
278		454.8993	1361.6760	1361.6787	-1.97	1	58	0.0004	▸1	U	R.DIDDGKASSVISR.R
362		807.3962	1612.7779	1612.7917	-8.59	0	141	1.4e-11	▸1	U	R.GVGNGNDATGLVANAQR.Y

▸ 5 subsets and intersections (12 subset proteins in total)

▸2 

1 NP_006112.3 156 keratin, type II cytoskeletal 1 [Homo sapiens]
2 KFO21255.1 121 Keratin, type II cytoskeletal 1 [Fukomys damare...
4 AAC83410.1 40 epidermal cytokeratin 2 [Homo sapiens]
3 XP_020821861.1 96 LOW QUALITY PROTEIN: keratin, type II cytosk...

▸3 

1 CAA82315.1 149 cytokeratin 9 [Homo sapiens]
2 AAA60544.1 130 keratin 10 [Homo sapiens]

▸4 1AO6_A 145 Chain A, Crystal Structure Of Human Serum Alb...
▸5 1CQI_A 125 Chain A, Crystal Structure Of The Complex Of A...
▸6 1AKS_A 109 Chain A, Crystal Structure Of The First Active Au...

Figure 53. The eluted purified P25 protein expressed in SHuffle® T7 was run on a bis-tris 4-12% gel (SDS-PAGE), **Figure 53: A-B.** The single band of P25 protein was cut and added into the Eppendorf tube containing 20µL of ddH_2O. The P25 protein in the gel was proceeded by trypsin digestion and determined using mass-spectrometry. The output of this recombinant P25 protein after identification was similar to the precursor of silk fibroin light chain (*B. mori*) **Figure 53: C.**

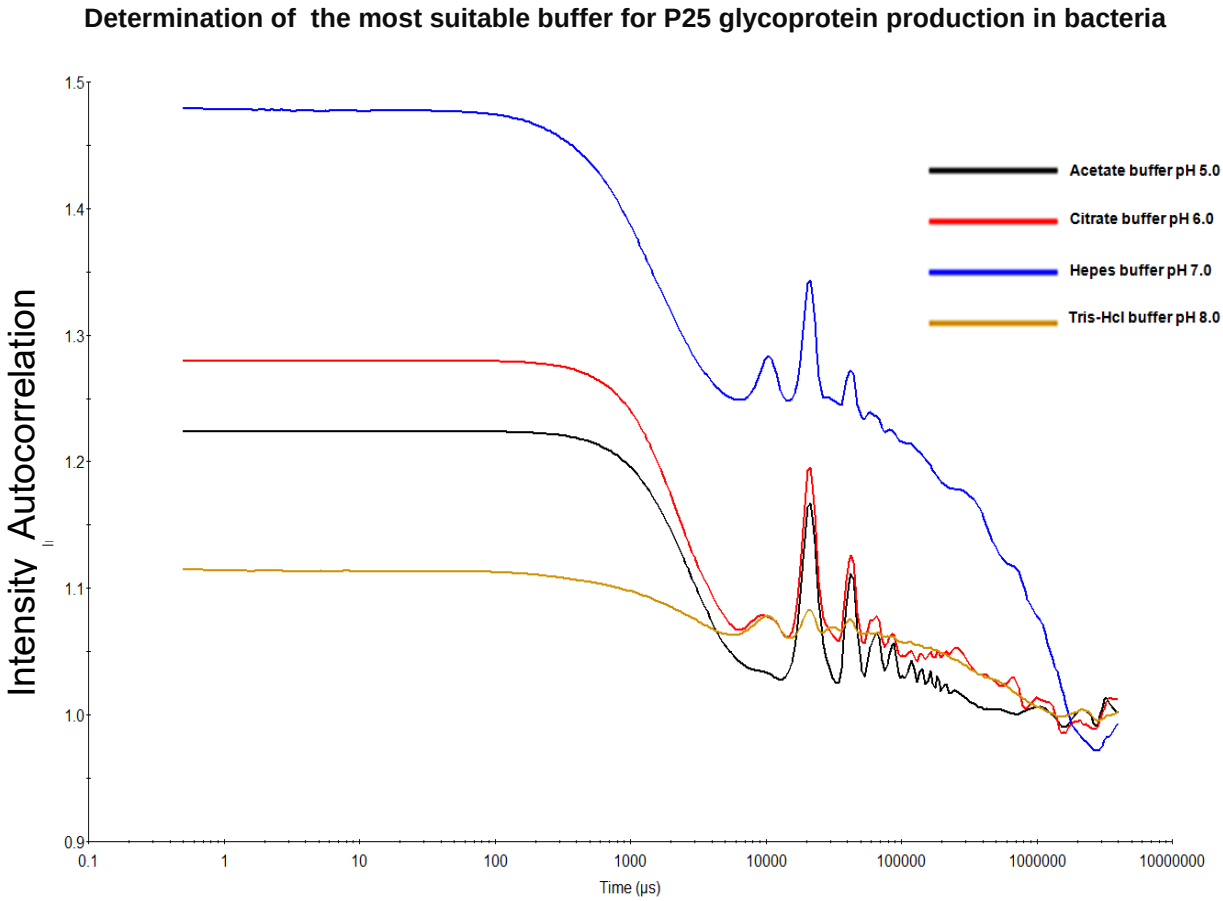


Figure 54. The elution buffer was removed from the eluted P25 protein and replaced with desired buffers. 200µL of Purified 25 protein's behaviour was tested in four different buffers, Acetate buffer pH 5.0, Citrate buffer pH 6.0, HEPES buffer pH 7.0, and Tris-HCl buffer pH 8.0. Each was individually loaded into the dynamic light scattering detection device (Wyatt Technology) which showed the intensity autocorrelation.

RESULTS (PART II)

Part (II) P25 protein expression in baculovirus-infected insect cell system

1) Vector construction and 2) bacmid preparation (Molecular Engineering)

Figure 55, regarding the previous expression of P25 in bacteria with different transfer vectors, the findings showed that the final product of P25 protein containing a His6-tagged after purification appears in many forms (multiple bands) when analysed by SDS-PAGE and only a single band in western-blotting. The unstable forms of P25 indicates that P25 requires the post-translational modification especially the intra-disulfide bonds from eight Cys residues and three sites glycosylation (Tanaka et al., 1999). Therefore, we need to move the expression of P25 in bacteria to the appropriate cell line system (insect cell system) in the selective transfer vector, pFBDM. The P25 cloning into pFBDM was initially achieved by the classic insertion-ligation technique at the polyhedrin promoter. The plasmid DNA was verified in the sequence with the labelling of His6-tagged and 3C cleavage site at the N-terminus of P25. The corresponding P25 construction was shown in Chapter II (materials and methods.

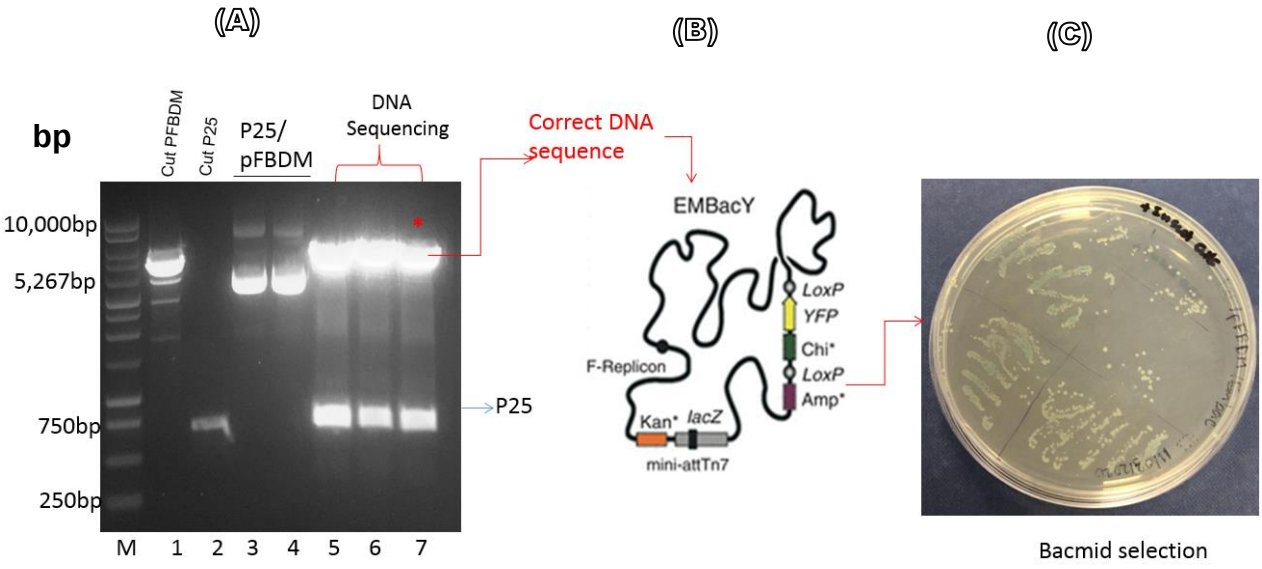


Figure 55. The initial process chain for P25 recombinant protein production using insect cell expression system or BEVS. P25 gene cloning was achieved into the transfer vector pFBDM. Lane 1: Cut pFBDM, Lane 2: Cut P25, Lane 3-4: plasmid DNA (P25/pFBDM), and Lane 5-7: the plasmid DNA was rechecked by double digestion with BamHI and NotI. The last three lanes show the sequence construct **(A)**, followed by the correct sequence plasmid DNA which was transformed into the bacmid cells; EMBacY on an agar plate containing kanamycin, gentamycin, IPTG, and X-gal **(B**) which is a blue-white screen technique **(Figure B** was adapted from Sari et al., 2016). The following day, the bacmid colonies were grown with mostly blue and semi-blue populations **(C).**

RESULTS (PART II) -Continued

3) Generation of baculovirus stocks and P25 protein localisation

3.1) Small scall harvested P25 protein production from the cell pellet P1 & P2 virus (Semi-glycosylated P25 protein)

Figure 56, once the P25 bacmid DNA was purified using QIAprep miniprep used for transfection with the WT Sf21, it was forwarded for virus production in the insect cell system. The successful transfection stock was used to provide a long-term storage stock called 'P1 as described previously (Materials and Method sections in Chapter II). Briefly for P1 production stock (50mL), the Sf21 cells (~0.6x10^6 cells) were put into a disposable flask (250mL), followed by 49mL of Insect cell media, 25µL (1x anti-anti) and finally 0.5 mL of P25 bacmid DNA was added. At this point, after the transfection of bacmid DNA (P25/pFBDM/EMBacY) to Sf21 cells for about 66 hrs, it was determined using an automated cell counter and harvested when the cells infection had about (30 %) viability. Only the supernatant fraction was collected into the 50 mL tube for long term storage (P1stock), preserved in the refrigerator, whereas P1 pellets were re-suspended with 25 mL fresh media supplemented with 25µL of 1x Ant-anti and grown for a further 72 hrs in the shaker to produce P25 protein. At the 72-hours point, the suspension cell culture from P1 cell pellets was harvested and the P25 protein looked at. The samples were then collected and determined using SDS-PAGE and western blotting. The profile of some selected samples through the P25 protein production were additionally grown from the P1 pellets containing two MWs as shown in **Figure 56.**

Figure 57, the purified P25 protein was present with two isoforms of glycoprotein (MW ~30 and 32 kDa).For the production of virus P2 stock (50 mL), the virus stock of P1 was used to transfect the WT Sf21 in order to check the efficacy of P25 glycoprotein expression from virus P1. The Sf21 cells (~0.6x10^6 cells) were put into a disposable flask (250 mL) and additional grown 72hrs to produce P25 protein. At that point, the cell pellets were harvested and the insect cells lysis buffer added and then the cells disrupted using triple thaw cycles and purified through the Ni-NTA column. The samples were collected and determined using SDS-PAGE and western blotting. The results show that the purified P25 protein was present with two isoforms of glycoprotein (MW ~30 and 32 kDa)

RESULTS (PART II) - Continued

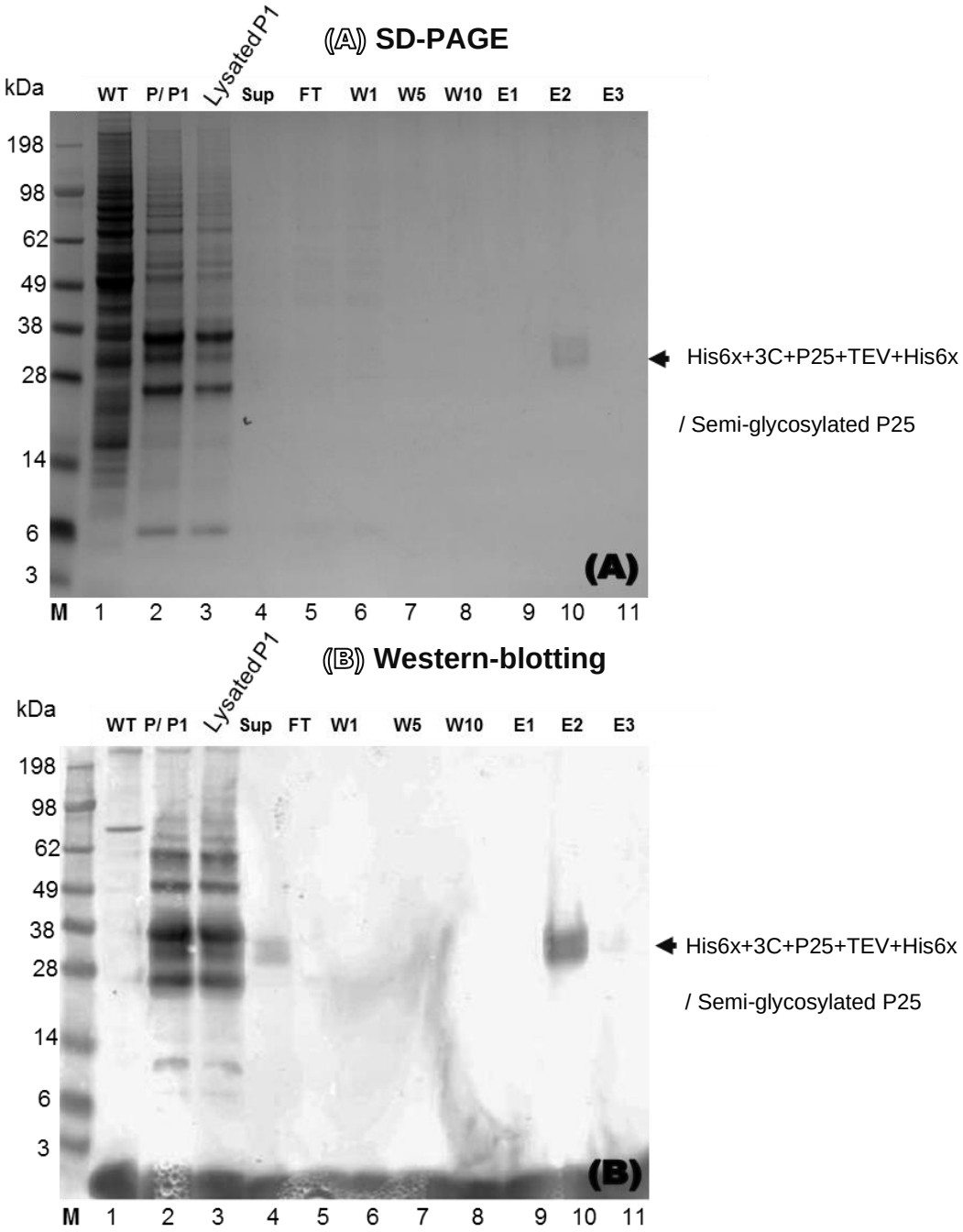


Figure 56. The small-scale protein production of P25/pFBDM/EMBacY after successful transfection and production of virus P1 stock. The transfected P25/pFBDM/EMBacY in Sf21 (insect cells) was spun down at 1,200 rpm for 5 mins, at 4ºC in order to harvest the virus stock (P1) and the pellets were retained for 72 hrs for additional growth for protein production. SDS-PAGE & Western blotting **(A&B)** shows the protein production and purification of P1 pellets using Ni-NTA column. The protein fractions (Lane1: Wild Type-non-transfection:(WT), (2): pellets from P1 (P/P1), (3): P1 pellets lysated by 3 cycles of thawing and freezing, (4): supernatant of P25 after centrifugation at 17K rpm, 30 mins, 4ºC (Sup), (5): flow through from IMAC (FT), (6-8) washed solution (W1, W5 and W10), and eventually the eluted protein (E1, E2, E3) P25 containing His6x+3C at N-terminus and TEV & His6x at C-terminus presents two forms at about 30 and 32 kDa.

RESULTS (PART II) -Continued

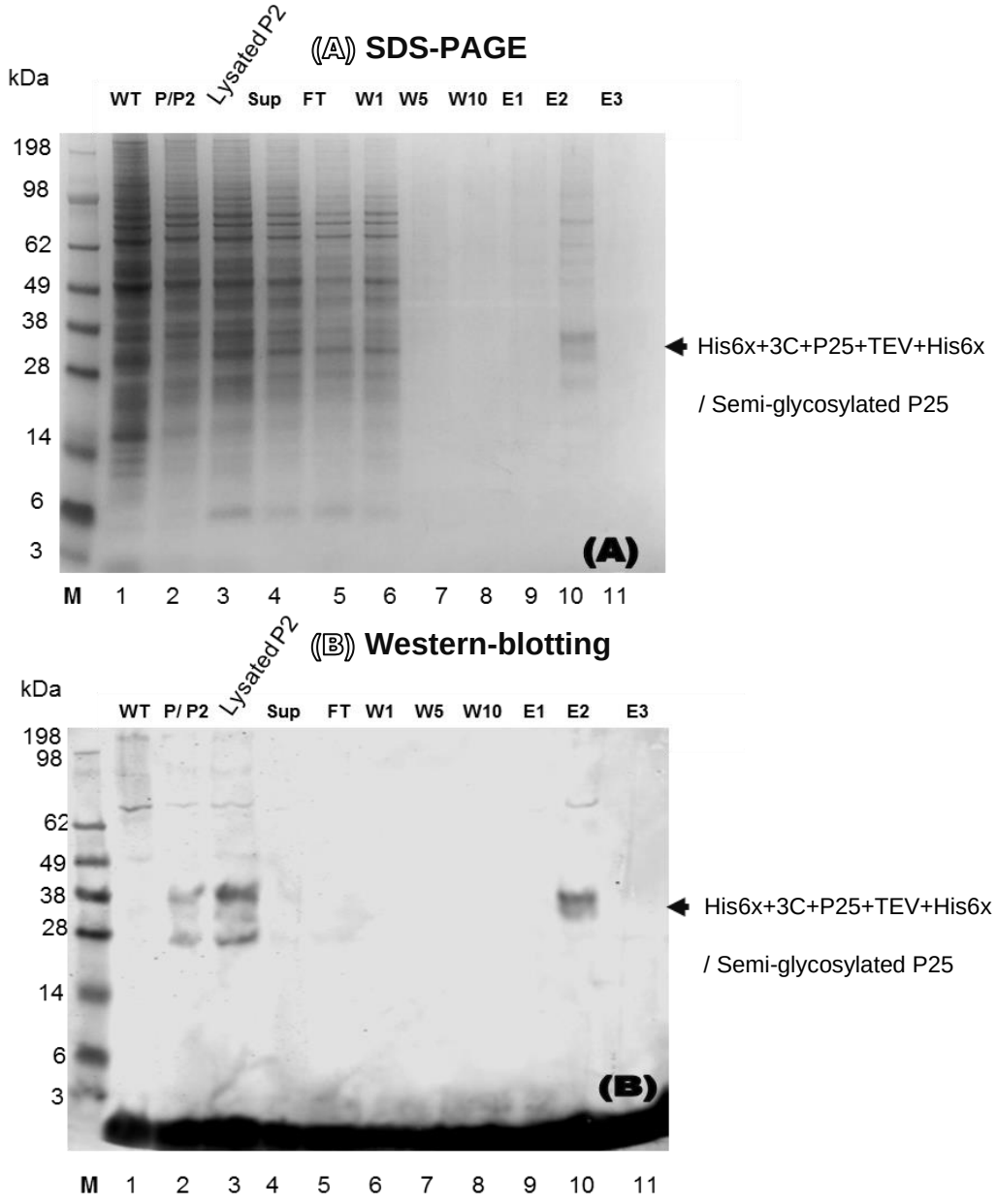


Figure 57. The small-scale protein production from P2 pellets of P25/pFBDM/EMBacY after successful transfection and production as stock of virus P1. The fresh Sf21 cells ($0.6x10^6$ cells) were added into a clean flask with 1x anti-anti and grown for virus production for 60hrs and the supernatant harvested (as P2 virus stock) whereas the pellets were re-suspended with the fresh media containing 1x anti-anti for a further 12 hrs for protein production. Samples of each step of the virus production and protein production were collected the for SDS-PAGE & western blotting **(A&B).** The images show the protein production and purification of P2 pellets using Ni-NTA column. The protein fractions (Lane1: Wild Type-non-transfection:(WT), (2): pellets from P2 (P/P2), (3): Lysated P2 pellets re-suspended in lysis buffer (300 mM Nacl,40Mm Tris, pH 8.0, 10% glycerol 2.5 mM BME and protease inhibitor) and then thawed and frozen for 3 cycles, (4): supernatant of P25 after centrifugation at 17K rpm, 30 mins, 4ºC (Sup), (5): flow through from Ni-NTA column (FT), (6-8) washed solution (W1,W5 and W10), and eventually the eluted protein (E1, E2, E3). P25 containing His6x+3C at N-terminus and TEV & His6x at C-terminus presented two forms at about 30 and 32 kDa.

RESULTS (PART II) - 4) Larg-scale expression and 5) two forms glycosylated P25 purification

The two result sections (4&5) have been connected in order to be better explained. **(Figure 58)**, the optimised P1 and P2 viruses verified P25 protein expression from previous small-scale experiments. In the final stage of production, the protein was obtained from only the purified cell pellets which were selected initially. It was found that the cell pellets from the culture after purification contained a P25 protein having two forms of glycoprotein (MW ~30-32 kDa). On the other hand, the medium fraction needed to be investigated to clarify the P25 protein expression and secretion, so the medium fraction from the production process was experimentally monitored in this section. Firstly, the production of P25 protein started from the Sf21 cells which had been transfected with P2 virus and subsequently were grown in larger containers (600mL x 2 flask) at 27ºC for 72hrs. At each time point, 24, 48, and 72 hrs, the cells were counted to monitor the cells infection and photographed. **(Figure 59),** finally, at the 72 hrs production time point, the two fractions in the culture flask, including cell pellets and culture medium, were collected separately by centrifugation at 10K rpm, 4ºC for 30 mins. The insect cell lysis buffer was added to the obtained cell pellets which were then persevered at -20ºC for consequent gentle disruption (triple thaw cycle) and then purification using Ni-NTA column. Regarding the obtained medium fraction, 1 mL of the sample was collected from the culture flask and the P25 protein content measured by adding 50µL of Bradford reagent. The medium fraction contained about 13 mg/mL. the protein content measured again using Nano Drops. As in the Bradford testing, it was found that the medium fraction after 72 hrs contained the protein because the sample had changed from yellow to blue. The medium fractions could be preserved in the freezer (-20ºC or -80ºC) for about one and half months without any preservative agents or could be immediately purified using Ni-NTA protocol. This indicated that after secretion (glycosylated) P25 protein was more persistent than P25 existing in the cells (semi-glycosylated).

RESULTS (PART II) - Continued

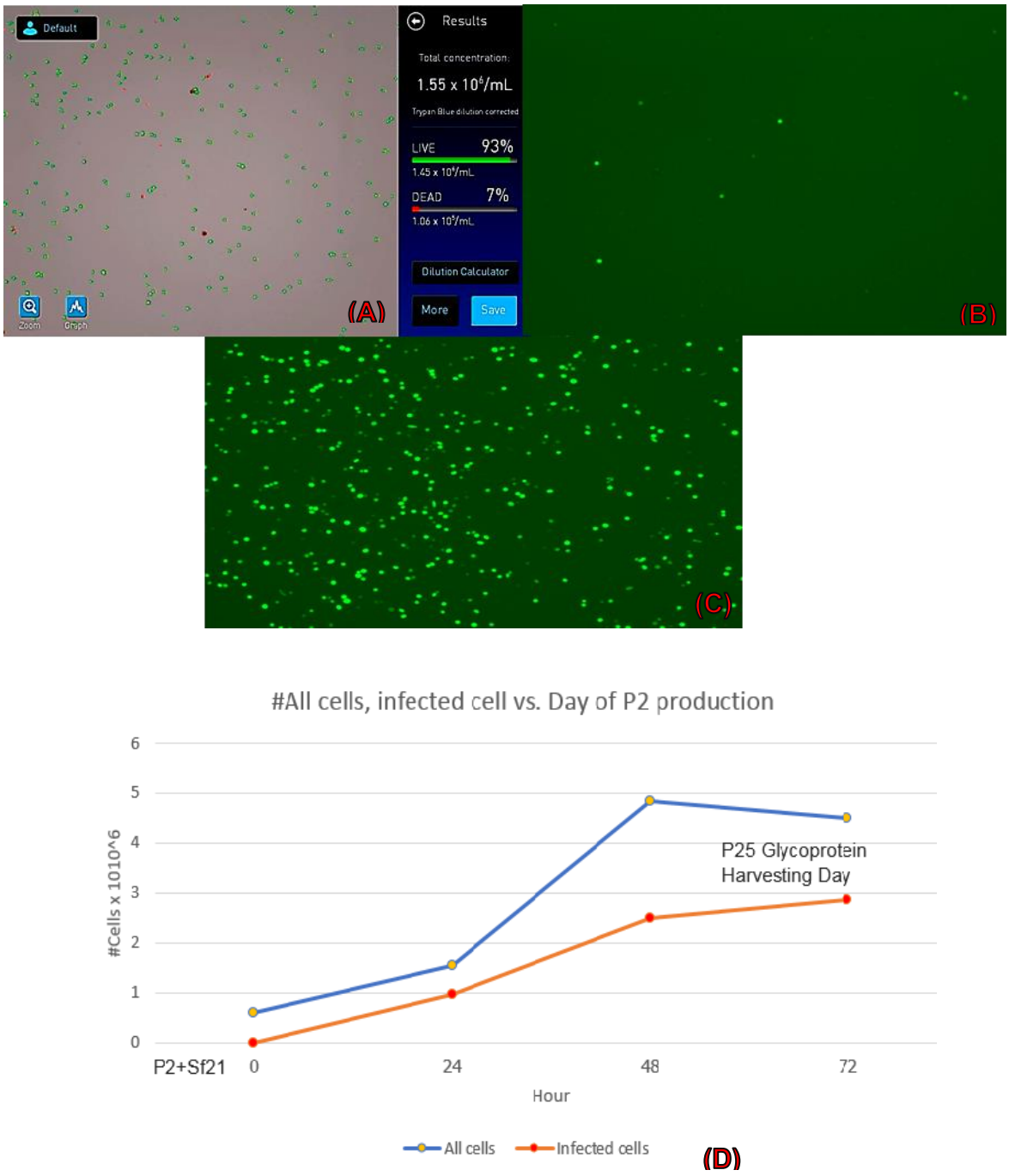


Figure 58. Infection developmental stages of P25/pFBDM/EMBacY (from P2 stock) after infection of Sf21 (WT) (~0.6x10^6 cells), 600mL with insect media in a plastic flask (2,000 mL) containing 1x anti-anti (Gibco Antibiotic-Antimycotic) were monitored. The Sf21 cells after infection were collected at an early stage of infection, 24 hrs**. (A),** middle stage of infection, 48 hrs. **(B)**, and full stage of infection, 72 hrs. **(C)**. At each time point of protein production, ranging from 24, 48, and 72 hrs, the suspension media was collected and the number of whole cells while being infected were measured using the cell counting machine (Countess II FL Automated Cell Counter, ThermoFisher Scientific). Cell numbers of three stages infection (Early, Middle, and Full) were measured and plotted **(D)**.

RESULTS (PART II) - Continued

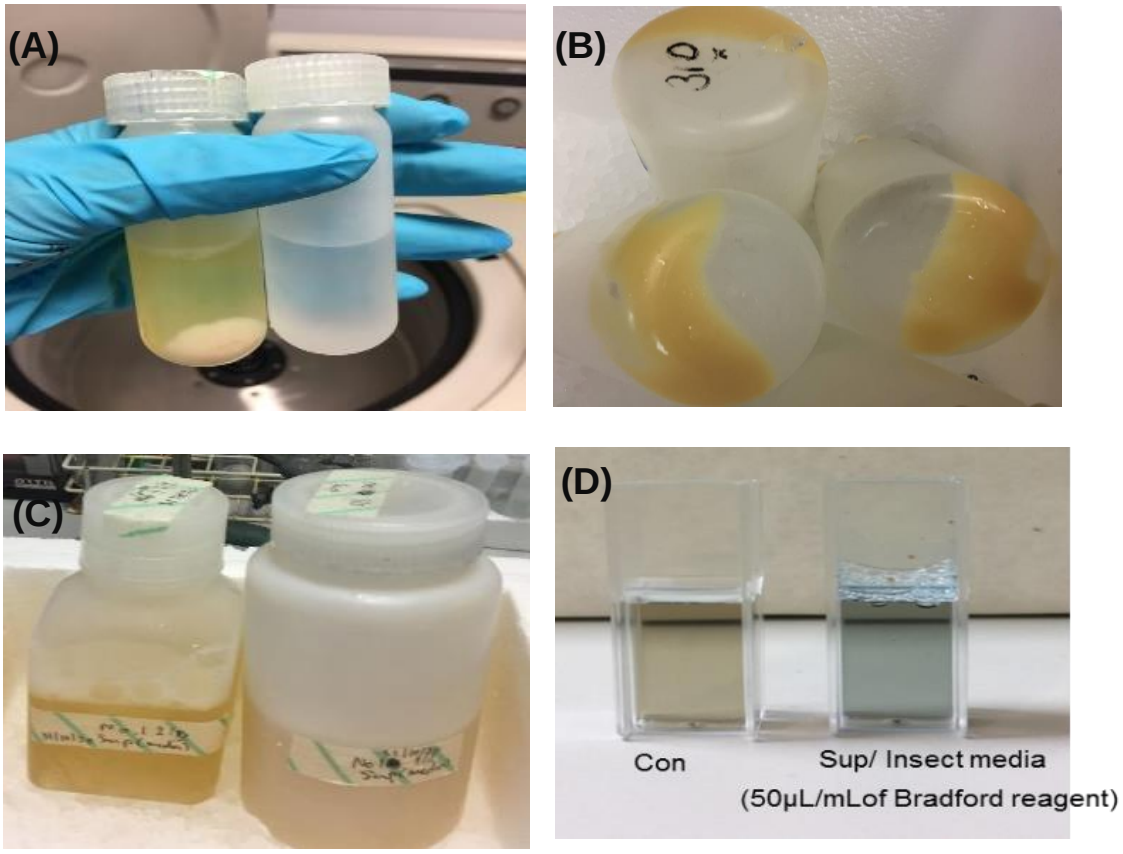


Figure 59. The larger scale culture was carried out in a plastic flask (2,000 mL) followed by adding Sf21 (WT) (~0.6x10^6 cells) 600 mL of insect cell media, 1x anti-anti (Gibco Antibiotic-Antimycotic), and finally 5mL of P2 virus. This cell culture was maintained on a shaker at 27ºC, 110 rpm, 4ºC for 72 hrs. At the end of this time point, the cell pellets and the medium fraction were harvested by centrifugation at 10,000 rpm, 4ºC, for 30 mins. **Figure (A).** Firstly, the insect cells lysis buffer was added to the harvested cell pellets and kept at -20ºC until being used for purification through the Ni-NTA process **Figure (B).** Part of the medium fractions were isolated from the supernatant after centrifugation and maintained in the freezer (-20ºC or -80ºC) until used for further purification using the Ni-NTA process **Figure (C)** and these medium fractions samples were taken from the bottle primarily to measure the P25 protein content using Bradford reagent. The control Bradford reagent (con) remained the same colour (reddish brown) whereas the tube including medium fraction and Bradford reagent turned blue **Figure (D).**

5) Two forms of glycosylated P25 protein purifications

Many previous experimental findings, from both small-scale and large-scale P25 protein production in suspension insect cell media, have revealed that P25 protein expression in Sf21 cells present two forms of glycosylation in two different parts of the culture. The semi-glycosylated P25 protein (~30&32 kDa) as non-secreted protein can be found in the cell pellets whilst glycosylated P25 protein (32 kDa) as secreted protein can be harvested from the medium fractions. The results of these two forms of P25 glycoproteins expressed in large scale were shown as follows.

5.1) Large-scale protein purification from the cell pellets (Semi-glycosylated P25 protein) and 5.3 additional purification of semi-glycosylated P25 protein from the cell pellet using SEC and protein determination

The method to purify semi-glycosylated P25 protein from the cell pellets on a large scale was described previously in this chapter. Attempts at protein purification on a large scale involved two steps; 1) using Ni-NTA column and 2) using the gel filtration machine or size exclusion chromatography. The image of the findings of 1st semi-glycosylated P25 protein purification is shown in **Figure 60 (A-B)** revealing the semi-glycosylated P25 protein in multiple bands, whereas the western-blotting image confirmed the two MW of semi-glycosylated P25 protein was 30&32 kDa. When the 2nd purification was eluted in Buffer A through the gel filtration, SDS-PAGE and western blotting revealed that the multiple bands diminished in the presence of semi-glycosylated P25 protein with the right MW bands (30&32 kDa) **Figure 61 (A-B)**.

5.2) Large-scale protein purification from the medium fractions (Glycosylated P25 protein) and 5.4) glycosylated P25 protein determination

Figure 62- 63, the medium fraction purifications revealed that a high concentration of glycosylated P25 protein was more likely to have high aggregations when glycosylated P25 was concentrated with concentrations over 10 mg/mL. SDS-PAGE & western blotting (**Figure 62, A-B)** images show the lower concentration of medium fractions before purification with few multiple bands over 32 kDa. These multiple bands revealed disulfide bond formations when the glycosylated P25 protein had higher concentrations in the same environment. The 1st purification using Ni-NTA column revealed the multiple bands of purified glycosylated P25 protein as well as the 2nd purification using gel filtration **(Figure 63)**. These multiple bands of glycosylated P25 protein can be slightly reduced by adding a reducing agent (DTT) as shown in **Figure 63 (B-C**).

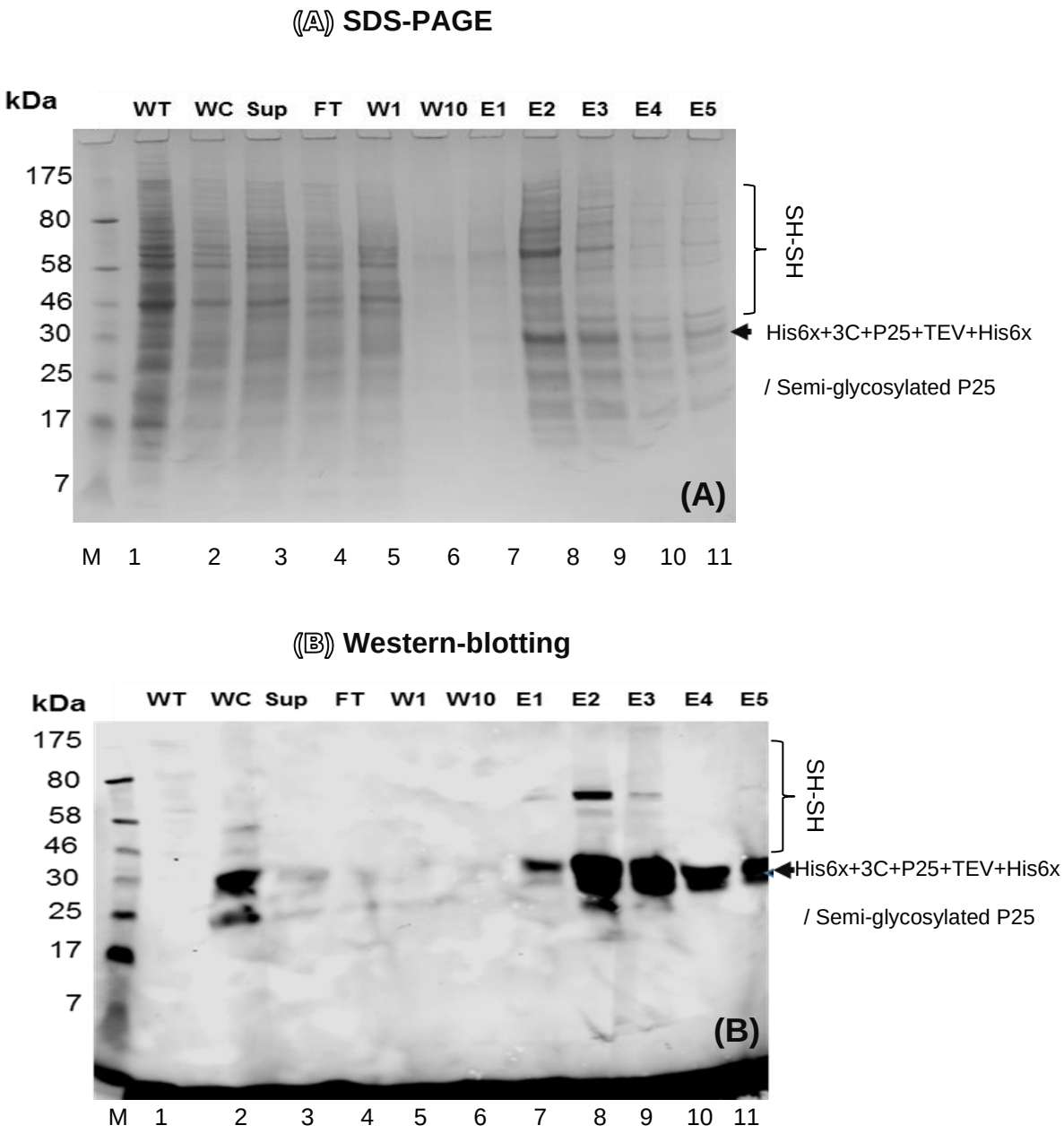


Figure 60. P25/pFBDM/EMBacY (P2 virus stock) was used to infect the larger scale culture. The method was carried out in a plastic flask (2,000 mL) with Sf21 cells (WT) (~0.6x10^6 cells), then 600 mL of insect cell media, 1x Anti-anti (Gibco Antibiotic-Antimycotic), and finally P2 virus from the stock added. The infected cells were grown in the closed system (incubator with shaker at 27ºC, 110 rpm, 72 hrs.). At the end, the fully infected cell pellets (72 hrs) were harvested with the centrifugation (10,000 rpm, 4ºC, 30 mins). The medium fractions were collected for further analysis whereas the resulting pellets were re-suspended with the insect cells lysis buffer, the cells then disrupted using the triple thaw cycle before further centrifugation at 17,000 rpm, 30 mins, 4ºC to collect the obtained supernatant. Finally, this cleaned supernatant was immediately forwarded to purification through the Ni-NTA column and was then eluted with elution buffer. After purification, the samples were collected and determined using two SDS-PAGE gels, one gel stained with the Coomassie blue (**A)** and another gel was subjected to western blotting (**B)**. Lane1-11 were Sf cells wild type (WT), whole cells (WC), Supernatant 17,000 rpm (Sup), flow through (FT), wash 1st and wash 10th (W1 and W10), and eluted P25 protein 1mL each (E1, E2, E3, E4, and E5).

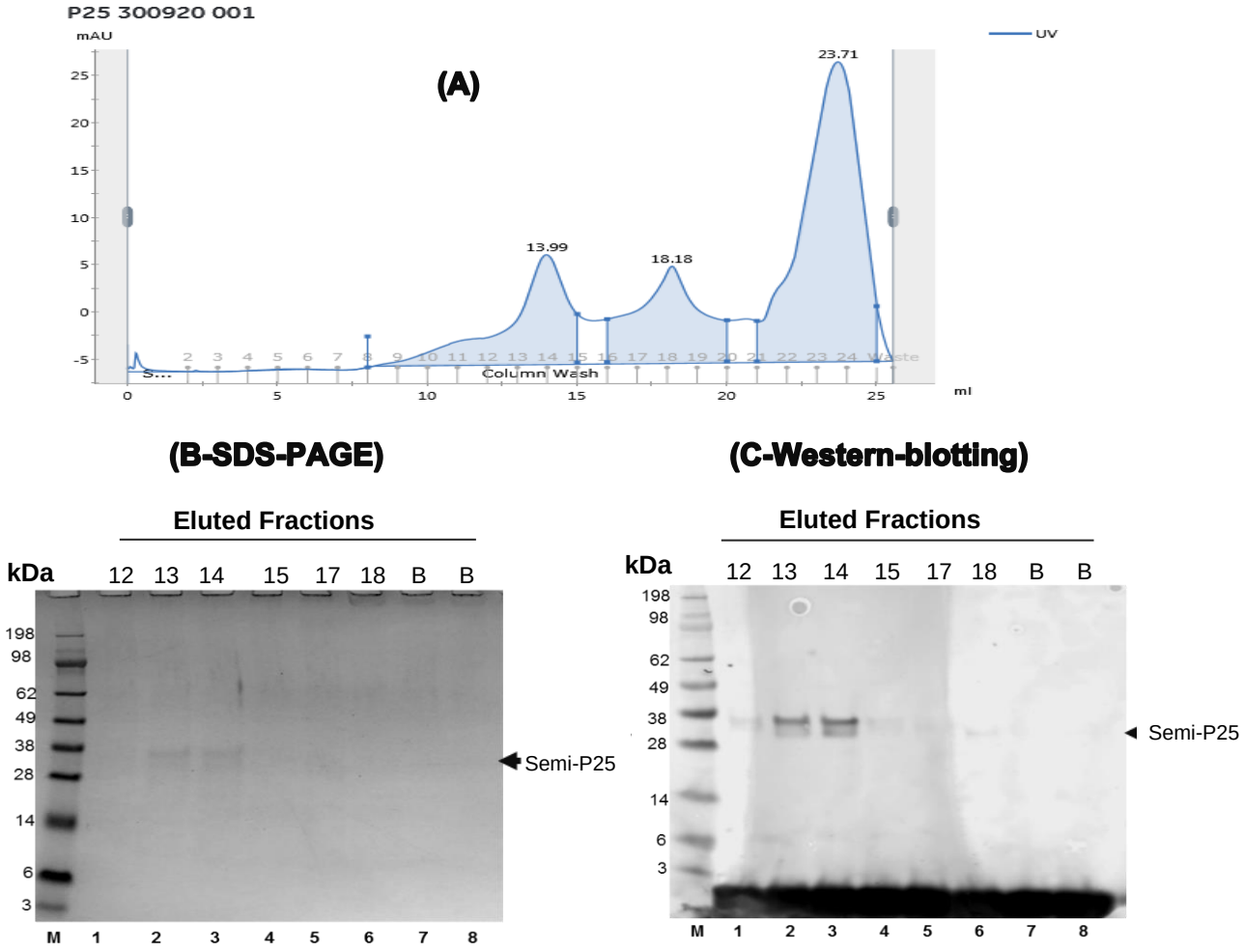


Figure 61. The purified P25 semi-glycosylated protein after purification through the Ni-NTA process was concentrated and changed into Buffer A (20 mM Tris-HCl, pH 8.0, 5M urea). The obtained P25 glycoprotein (0.73 mg/1mL) was further purified by loading 1mL into the size exclusion chromatography machine (AKTA pure systems, using Superdex 75/200GL). Consequently, the eluted fractions of each peak **(A)** were collected and the target (bands P25 protein) from the specific MW were determined. The P25 protein fractions related to the SEC peak were detected with SDS-PAGE **(B)** & western blotting **(C)** and the expected bands of P25 protein with the N-terminus and C-terminus tagged were detected by anti-6xHis-tagged antibody at ~ 30-32 kDa.

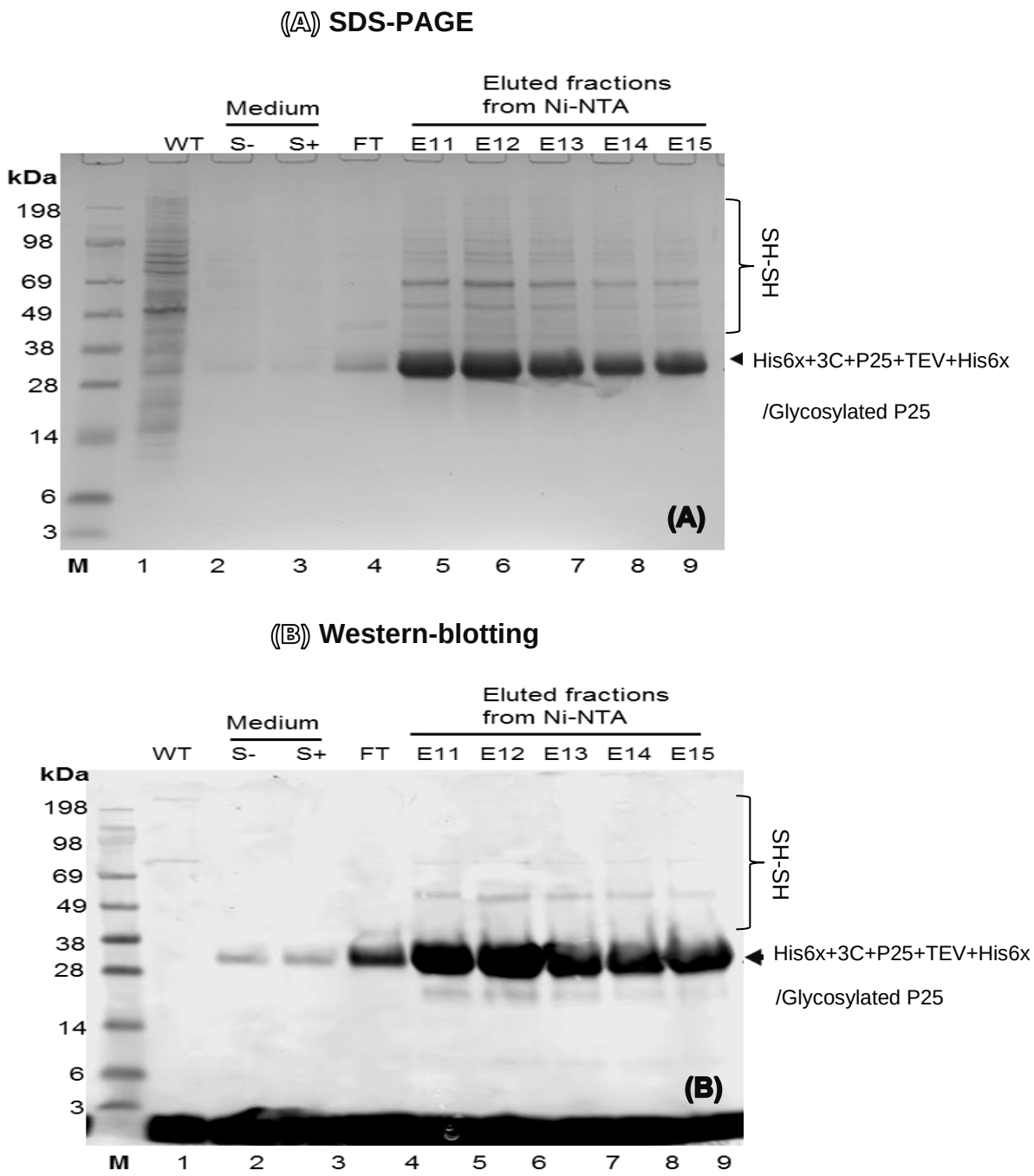


Figure 62. P25 glycosylated protein from the large-scale production was recovered from the medium fraction of baculovirus-infected Sf21 cells culture after being harvested by centrifugation at 10,000 rpm, 4ºC for 30 mins. This medium suspension was again centrifuged at 10,000 rpm, 30 mins, to remove the debris and other contaminants. The cleaned supernatant was collected into a 1L beaker, kept cold on ice, and gradually passed through the glass Ni-NTA column (Bio-Rad). Unbound protein (Flow through, FT) was collected, the column washed in 100 mL (wash buffer), then the bound P25 glycoprotein was eluted 25 times, the selection shown from (E1-E15)/mL/time. The eluted purified fractions were separated on the gels together with wild type whole lysate (WT), medium fractions without the lysis buffer (S-) and with the lysis buffer (S+), Figure (C-D). The secreted P25 glycoprotein with (N-terminus& C-terminus 6x-Histagged was about 30 kDa). **Figure 62-A** results from SDS-PAGE and **Figure 62-B** results after western-blotting with the target bands at ~32 kDa.

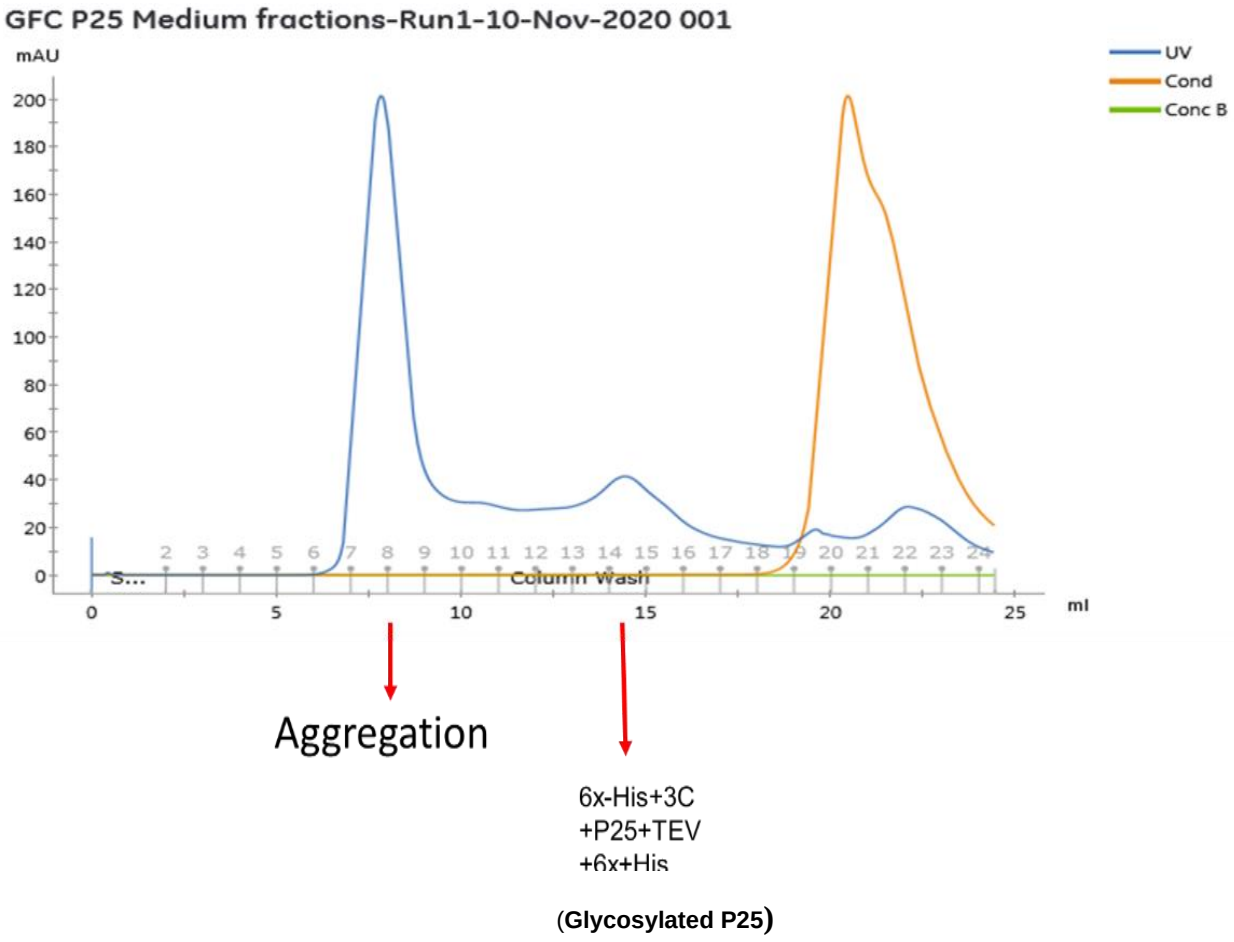


Figure 63. The chromatogram of purified secreted glycosylated P25 protein (medium fraction) after Ni^{2+} purification was concentrated and changed into buffer A (20 mM Tris-HCl, pH 8.0, and 5 M urea). The obtained concentrated P25 glycoprotein (14 mg/mL) was loaded into the Superdex 75 increase 10/300 GL and the eluted fractions which corresponded to the peaks (F7-F10) were collected and the semi-glycosylated P25 purified (F14-F15) for further determination with and without a reducing agent (DTT).

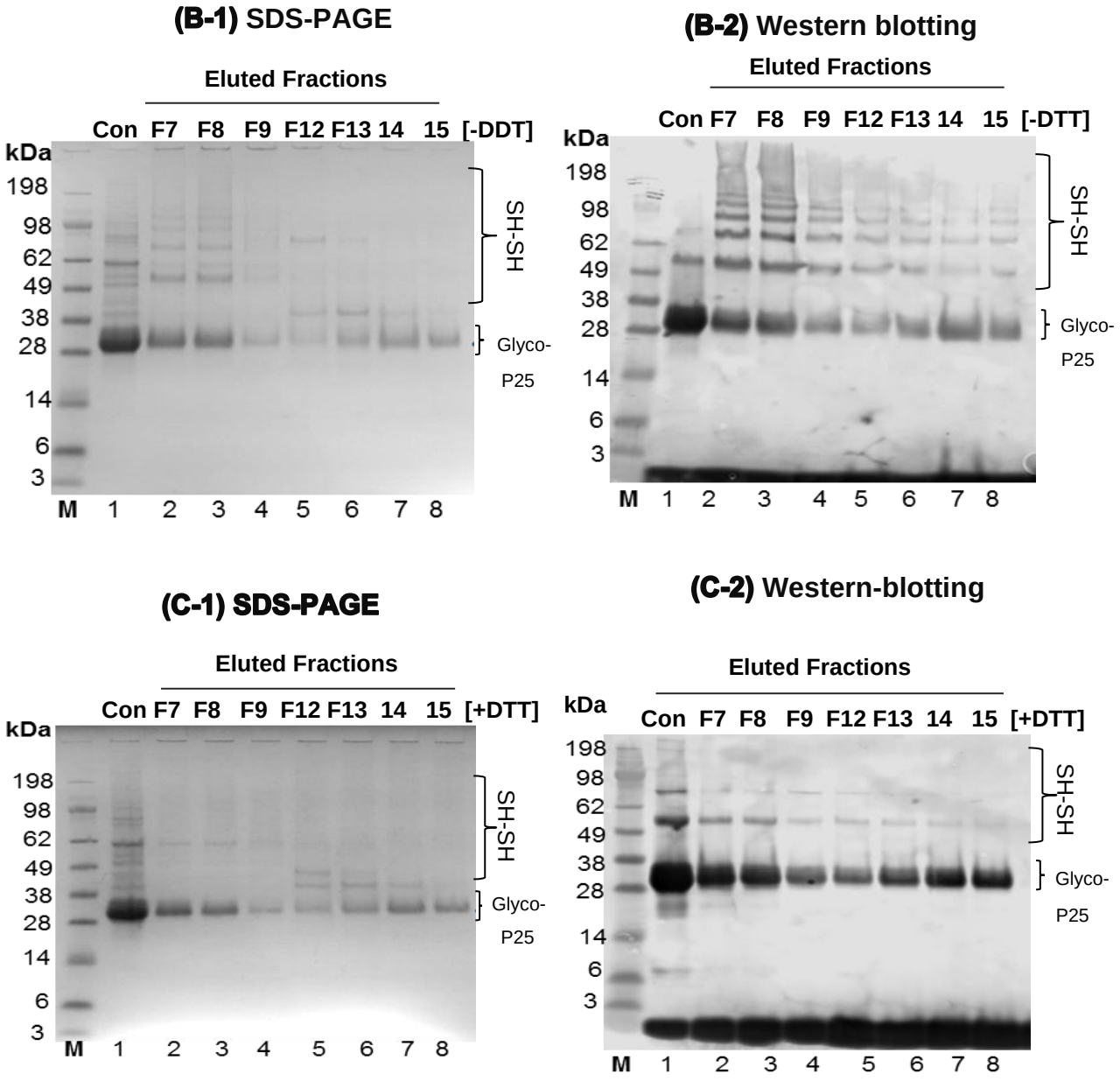


Figure 64 (B-C). The 2nd purified secreted glycosylated P25 protein (medium fraction) after Ni^{2+} purification was concentrated and changed into buffer A (20 mM Tris-HCl, pH 8.0, and 5 M urea). The obtained concentrated P25 glycoprotein (14 mg/mL) was loaded into the Superdex 75 increase 10/300 GL and eluted fractions corresponding to the peaks were collected. To check the obtained purified P25 protein quality, the collected eluted fractions were separated on two bis-tris 4-12% gels, one of which was subjected to Western blotting **Figure (B1-B2, SDS-PAGE & western blotting). T**he samples were mixed only with 2x protein sample buffer (without reducing agents, -DTT) whereas in **Figure (C1-C2,SDS-PAGE & western blotting)** to 2x protein sample buffer containing reducing agents (+DTT was added to the samples. The secreted P25 glycoprotein with (N-terminus& C-terminus 6x-Histagged is ~ 30kDa).

RESULTS (PART II) - Continued

6. Characterisation of P25 protein

Findings from several repeated experiments confirmed that P25 protein expression in Sf21 cells always obtained two forms of glycoproteins in the infection culture after 72hrs, 1) cell fractions (non-secreted glycoprotein/semi-glycosylated, 30-32kDa) and 2) medium fractions (secreted glycoprotein/semi-glycosylated, 32kDa). Each native state of the secondary structure pattern of P25 glycoprotein was subsequently characterised using far-UV. Furthermore, the effect of enzyme treatments on both P25 glycoproteins was studied using two selective endoglycosidase enzymes (Endo-H, and PNGase F), and the number of glycans were determined by using Immunoprecipitation and Con A, respectively.

6.1 Secondary structure of P25 proteins using CD technique.

1) Secondary structures of semi-glycosylated P25 protein (non-secreted)

(Figure 65 A-D), the purified semi-glycosylated P25 protein isolated from the cell pellets underwent two types of purification process; Ni-NTA column and gel filtration machine. The eluted fractions of purified semi-glycosylated P25 protein including two MW (30-32 kDa) were run on SDS-PAGE and western blotting gels to check the protein quality **(Figure 65 A-B)**. The bands of cleanest semi-glycosylated P25 protein (Fraction 13-15) were used to predict the secondary structures. **(Figure 65 C)**. Far-UV spectra indicated that semi-glycosylated P25 contains helixes (Helix1-Helix2, ~14%), strands (Strand1-Strand2, ~33.4%), turns (Turns, ~20.9%), and unordered (Unordered, ~31.5%) **(Figure 65 D)**.

2) Secondary Structures of glycosylated P25 protein (secreted)

(Figure 64 A-D), after imidazole was eliminated from the eluted purified glycosylated P25 protein using the Ni-NTA column in Elution buffer, it was then changed into 20mM Tris-HCl, pH 8.0 buffer (which is more compatible with the CD machine). This glycosylated P25 protein (20 mM Tris-HCl ,pH 8.0) was concentrated with a concentrator (10K kDa, MWCO) and its quality was determined on the SDS-PAGE and western-blotting gels **(Figure 66 A-B)**. A band of concentrated glycosylated P25 protein (Lane 3) was used to predict the secondary structures. **(Figure 66 C**), Far-UV spectra indicated that glycosylated P25 contains helixes (Helix1-Helix2, ~0.00%), strands (Strand1-Strand2, ~47.7%), turns (Turns, ~17.1%), and unordered (Unordered, ~35.1%) **(Figure 66 D).**

(Figure 67, A-B), additionally, the tertiary structure of both purified semi-glycosylated P25 protein and glycosylated P25 protein were individually determined using near-UV (250-320). The spectra from each P25 glycoprotein were not shown in a tertiary formation. However, it seems glycosylated P25 protein had a more organised shape of spectra than that of semi-glycosylated P25 protein.

6.2 Glycan analysis of P25 proteins

6.2.1) Characterisation of glycans between semi-glycosylated and glycosylated P25 proteins using endoglycosidase treatment.

(Figure 68 & Figure 69, A-C), the achievement of P25 protein production in this project established new protocol. Therefore, it is necessary to clarify and characterize the two forms of

P25 glycoproteins, semi-glycosylation and glycosylation, with two selective related endoglycosidase enzymes (Endo-H and PNGase F). The states of both semi-glycosylated and glycosylated P25 proteins (folded/native) and unfolded/denatured) were compared using two enzyme treatments. The findings of the enzyme treatments showed that the mobility of the bands between semi-glycosylated and glycosylated P25 proteins were the same. To extend that, the bands of mobility (SDS-PAGE and western-blotting) of Endo-H treatments both native semi/glycosylated showed little change from the control (nontreated). On the other hand, the bands mobility after PNGase F treatments of both native semi/glycosylated P25 protein were remarkably changed from the control (nontreated). Similarly, both denatured semi/glycosylated show similar mobility bands after the same enzyme treatment, but obviously different mobility when compared with two different enzyme treatments. Notably, both enzyme treatments with denatured two semi/glycosylated did not show multiple bands over the protein target.

6.2.2) Characterisation of glycans between semi-glycosylated and glycosylated P25 protein using lectin binding/ Immunoprecipitation (IP) techniques

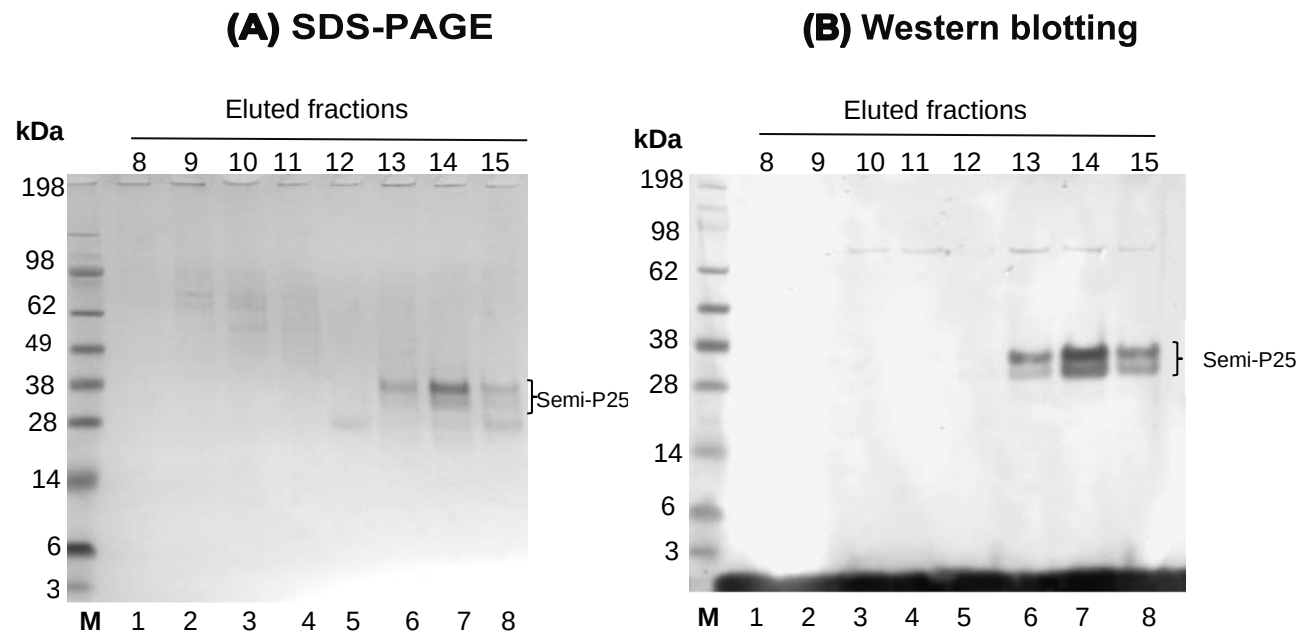
P25 found in the silk fibroin of the native silkworm (*B. mori*) is an *N*-linked glycoprotein, one minor ratio bound to each elementary fibroin unit containing three oligosaccharide chains. In this project, a recombinant P25 glycoprotein was obtained (semi/glycosylated) through production in two different fractions of culture. This method was essential for looking at these two glycoproteins using a lectin binding method known as Con (A). Con (A) is a tetramer lectin that can recognise and bind to a specific class of glycoprotein from the protein samples. Briefly described, both purified semi/glycosylated P25 protein including two states of protein (folded or native)/ unfolded or denatured) were treated with Con (A) following the IP protocol (Material and Methods section).

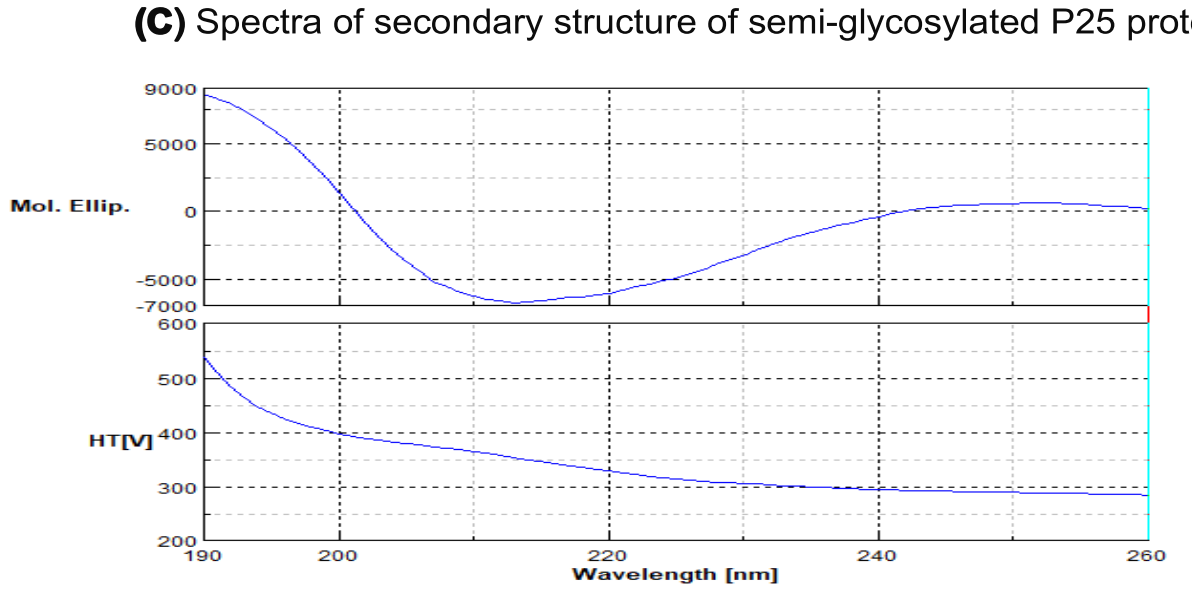
1) IP results of semi-glycosylated P25 protein

Both reactions of semi-glycosylated P25 protein (non-secreted) including native and denatured states were treated with Con (A) and finally determined with SDS-PAGE and wesstern blotting. The IP experiments found that no released glycans were present in semi-glycosylated P25 proteins including native and denatured states when compared to the control. However, slight IP bands of P25 glycosylated protein (~30-32kDa) were present in each IP treatment both native and denatured states. In addition, the IP images (SDS-PAGE & western-blotting) of semi-glycosylated P25 native state showed clearer bands of two MWs (~30-32kDa) than in denatured states **(Figure 70, A-D)**.

2) IP results of glycosylated P25 protein

Both reactions of glycosylated P25 protein (secreted), including native and denatured states, were treated with Con (A) and finally determined with SDS-PAGE and western blotting. The IP experiments found that released glycans were present in glycosylated P25 proteins including native and denatured states but with different MWs (Glycans number 0-3rd) from the glycosylated protein when compared to the control. These IP images show both glycosylated P25 proteins, native and denatured can be bound to the Con (A) and three glycans are present at ~12, 14, and 26 kDa, whereas after IP, glycosylated P25 proteins were ~30-32kDa (one glycosylated P25 protein band) (**Figure 71, A-D).**

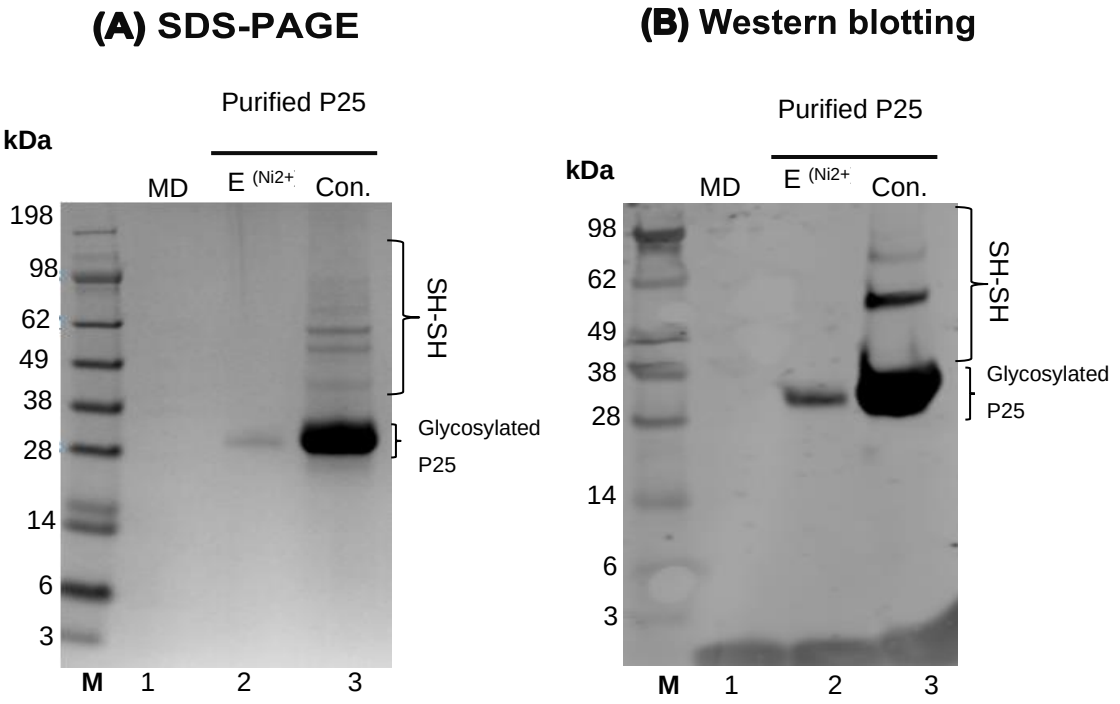


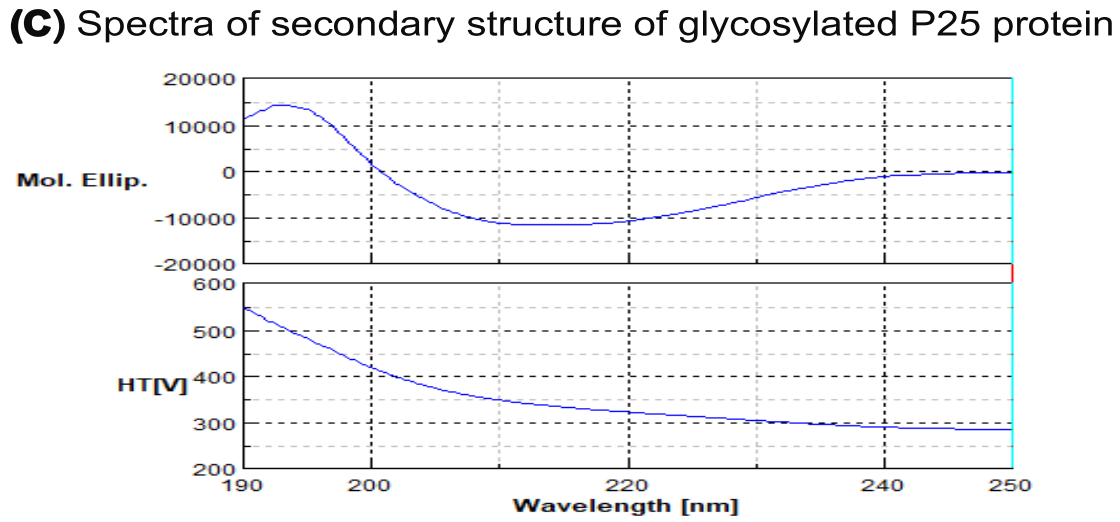


(D) Secondary structure contents of semi-glycosylated P25 protein

Result	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
1	0.082	0.066	0.214	0.112	0.200	0.325	0.999
2	0.064	0.074	0.223	0.118	0.217	0.304	1
Average	0.073	0.070	0.219	0.115	0.209	0.315	0.9995

Figure 65. The purified semi-glycosylated P25 protein isolated from the cell pellets was purified in two steps, firstly through an Ni-NTA column and then by a gel filtration machine and was subsequently changed into a buffer, 20 mM Tris-HCl ,pH 8.0, then the quality of protein checked in SDS-PAGE **(Figure A)** and western-blotting before **(Figure B)** the secondary structure was determined using a CD machine. Far-UV spectra of semi-glycosylated P25 protein has concentration of 0.3 mg /mL measured in a 0.01cm cell path length **(C)** the content of secondary structures of these spectra, Helix1, Helix2, Strand1, Strand2, Turns, and Unordered were estimated as shown in **Figure (D).**

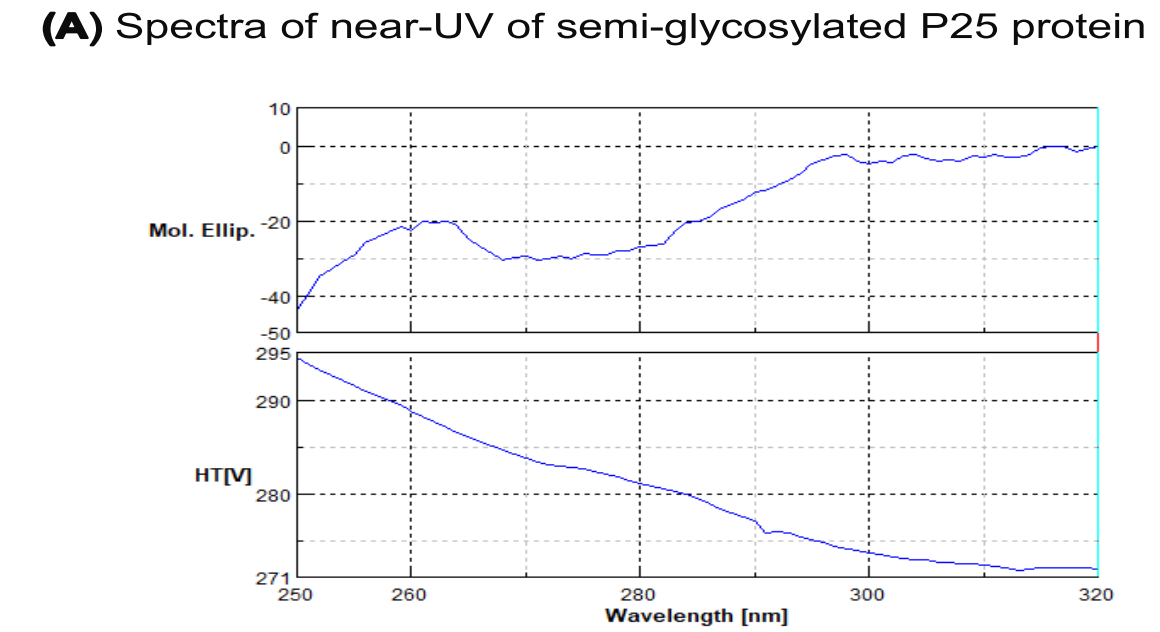




(D) Secondary structure contents of glycosylated P25 protein

Result	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
1	0.000	0.000	0.358	0.145	0.146	0.351	0.999
2	0.000	0.003	0.310	0.141	0.195	0.350	1
Average	**0.000**	**0.001**	**0.334**	**0.143**	**0.1705**	**0.3505**	**0.9995**

Figure 66. The purified glycosylated P25 protein isolated from the medium fraction was purified by Ni-NTA column and was subsequently concentrated, imidazole eliminated, and then changed into buffer, 20 mM Tris-HCl, pH 8.0, the quality of protein checked using SDS-PAGE **(Figure A)** and western-blotting **(Figure B)** and consequently the secondary structure was determined using a CD machine. Far-UV spectra of glycosylated P25 protein has a concentration of 1.0 mg /mL measured in a 0.01 cm cell path length **(Figure C)** the content of these secondary structures, Helix1, Helix2, Strand1, Strand2, Turns, and Unordered was estimated as shown in **Figure (D).**



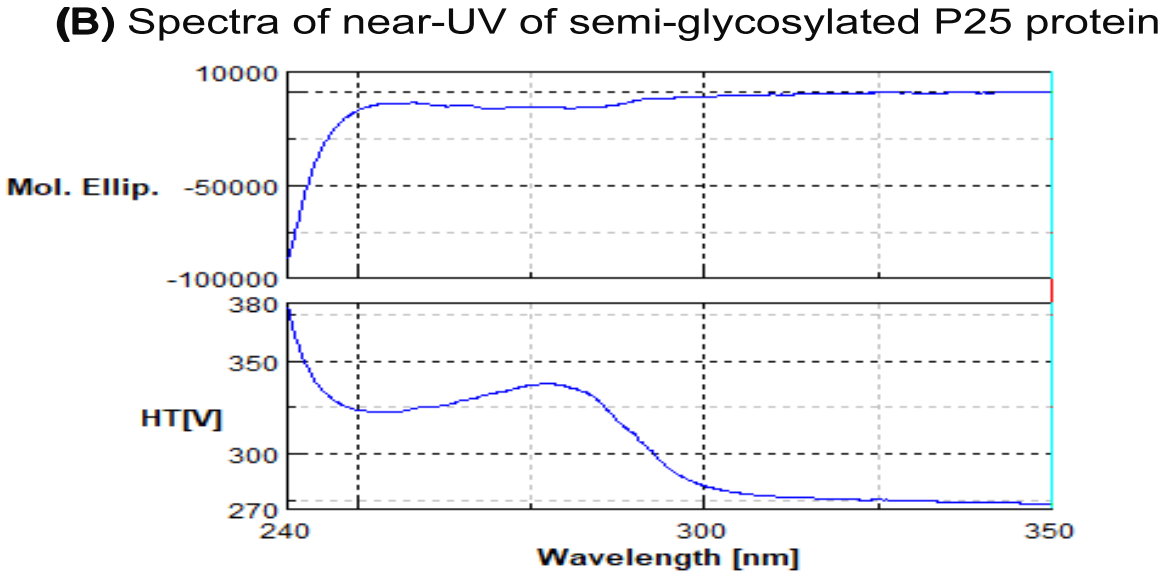


Figure 67. Comparative investigation was undertaken using near-UV (~200-400 nm) of semi-glycosylated P25 protein concentration 0.3 mg/mL in a 0.2 cm cell path length **(Figure A**) and near-UV spectra of glycosylated P25 protein at 2 mg /mL in a 0.5 cm cell path length **(Figure B)** loaded separately into a CD machine.

Incomplete P25 glycoprotein expressed in the insect cells (no glycan attached to P25 yet). P25 protein interacts with Endo-H & PNGase F.

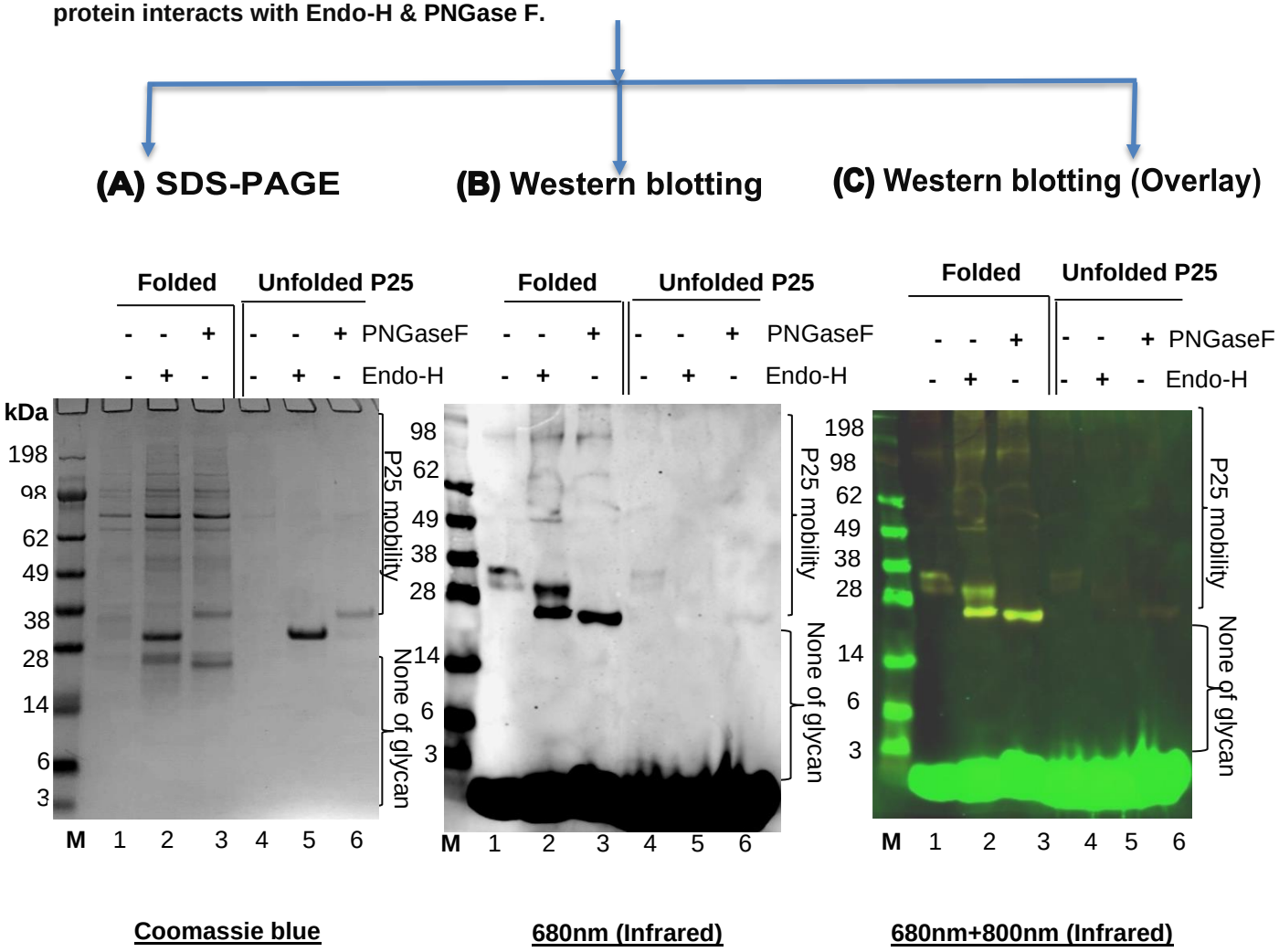


Figure 68. The mobilities of semi-glycosylated P25 protein was compared in two endoglycosidase enzymes treatments, (Endo-H & PNGase F) which were carried out on both folded & unfolded forms. To provide folded P25 semi-glycoprotein, the Ni^{2+}purified of secreted P25 glycoprotein was concentrated (0.70 mg/mL) by a concentrator (10 kDa, MWCO) and changed into Tris-HCl buffer (50 mM Tris-HCl, pH 9.0, and 95 mM NaCl). After that, the concentrated purified 25 glycoprotein was used in its folded form, whereas unfolded P25 semi-glycoprotein was prepared by taking 20µLfolded form into the Eppendorf tube and 220µL 8 M urea was added, and it was then incubated for 1hr at room temperature. Folded and unfolded P25 were treated either in 2 µL Endo-H or 2 µL PNGase F, incubated at 37ºC, 1hr. Eventually, both enzyme reactions were analysed by bis-tris 4-12% gels (SDS-PAGE **(A)** & western blotting (**B-C**)). The western blots used different wavelengths, **(B)** licor channel (680R), and overlay channel 800R (C). No Glycans were seen in either Endo-H and PNGase F treatments of folded or unfolded forms (**Figure A-C).** The panel 680nm **(B**) used one secondary antibody in the western-blot protocol whereas the panel of 680nm+800nm **(C)** used two secondary antibodies added into the transferred membrane. The images of P25 protein in yellow and green in the right lane are degraded proteins revealed using the Odyssey® infrared Imaging system. The western blot **(B&C**) did not have Biotinylated Con A added into the protocol. The glycoprotein cannot be seen clearly when using antibodies in the two infrared channels. The folded forms, but not in the unfolded form of P25 protein mobilties can be seen in the three panels. The IP Con A assay will indicate the glycans of P25 glycoprotein in the next experiment.

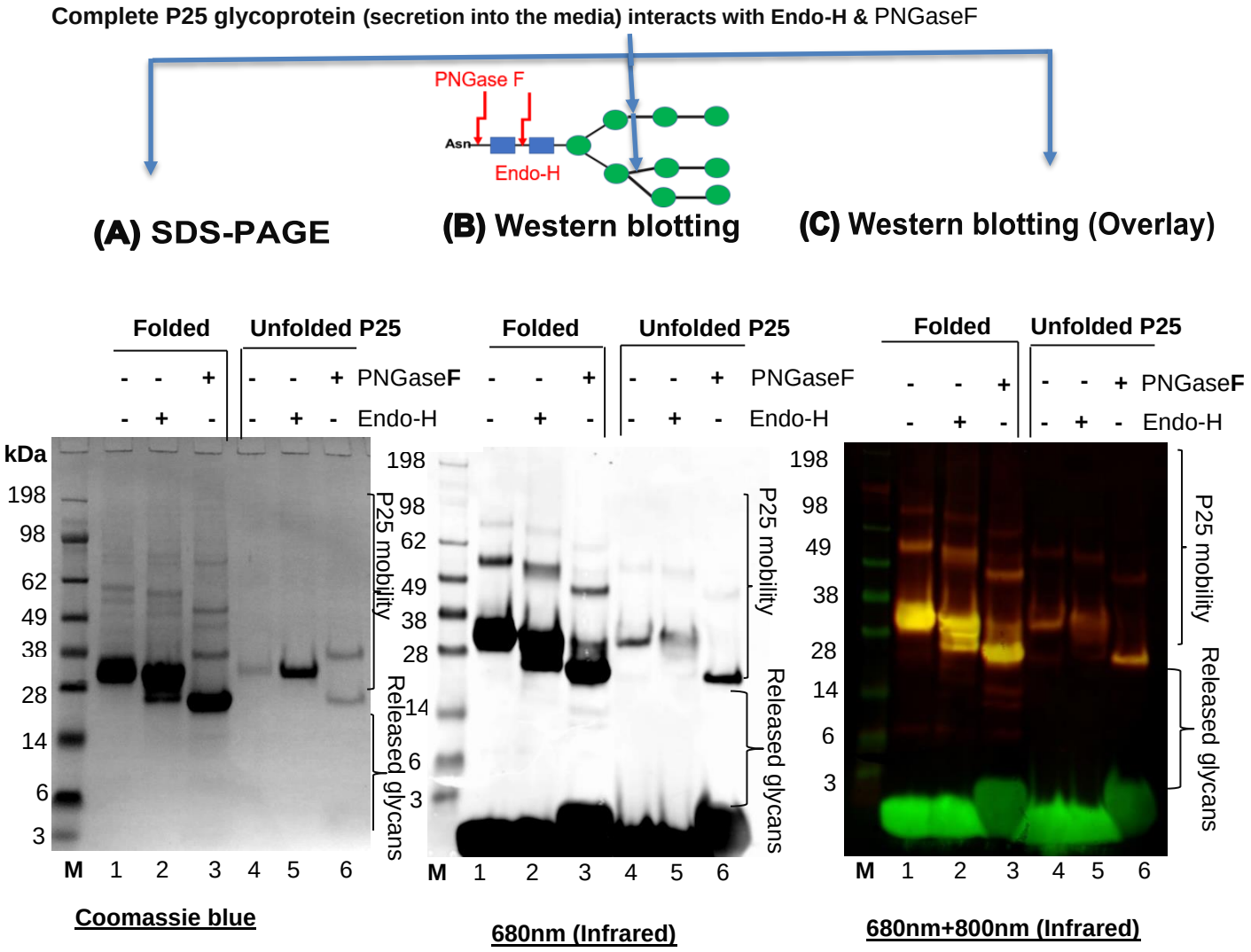


Figure 69. The mobilities of glycosylated P25 protein compared after two endoglycosidase enzyme treatments, (Endo-H & PNGase F) carried out on both folded & unfolded forms. To provide folded P25 glycoprotein, the Ni^{2+}purified of secreted P25 glycoprotein was concentrated (0.70 mg/mL) by a concentrator (10K kDa, MWCO) then changed into Tris-HCl buffer (50 mM Tris-HCl, pH 9.0, and 95 mM NaCl). After that, the concentrated purified 25 glycoprotein was used as folded form whereas unfolded P25 glycoprotein was prepared by taking 20µLfolded form into an Eppendorf tube and adding 220µL 8 M urea, then being incubated for 1hr at room temperature. Folded and unfolded P25 were treated either in 2µL Endo-H or 2 µL PNGase F, incubated at 37ºC,1 hr. Finally, both enzyme reactions were analysed by bis-tris 4-12% gels (SDS-PAGE) **(A) and** western blotting (**B-C**). The western blots used different wavelengths, **(B)** licor channel (680R), and overlay channel 800R **(C).** The most obvious released glycans were in PNGase F of folded forms (**Figure A-C**). The panel 680nm (**B**) used one secondary antibody in the western-blot protocol whereas the panel of 680nm+800nm **(C)** used two secondary antibodies added into the transferred membrane and the images are revealed using the Odyssey® infrared Imaging system. The mobilities of the protein of interest are in yellow, whereas the bottom green lanes are degraded proteins. The western blot **(B&C)** did not have Biotinylated Con A added into the protocol. Therefore, the glycoprotein cannot be seen clearly when using the antibodies in the two infrared channels. In all three panels, the attachment of thicker glycans on the P25 protein (folded form) can be seen. The IP Con A assay will indicate the glycans of P25 glycoprotein next experiment.

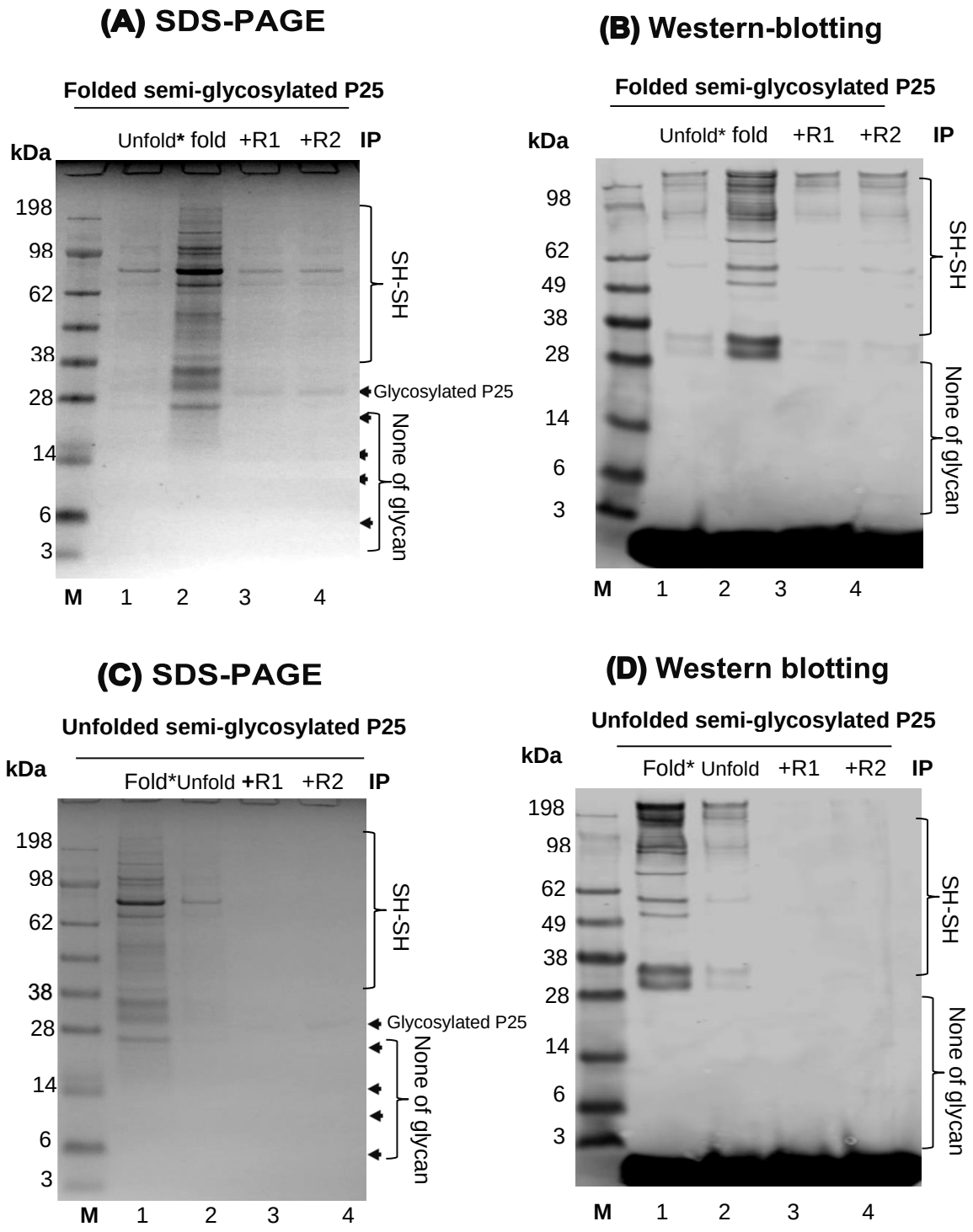


Figure 70. Lectin binding & IP (Con A) was performed on folded and unfolded forms of semi-glycosylated P25 protein. The reactions initially used the Ni^{2+}purified un-secreted P25 glycoprotein concentrated (0.70 mg/mL) by a concentrator (10kDa MW) in Tris-HCl buffer (50 mM Tris-HCl, pH 9.0, and 95 mM NaCl). Folded form control reactions **(A-B)** are Lane 1 (Control), Lane 2 (unfolded form, treated 8M urea) whereas unfolded form control reactions **(C-D)** are Lane 1 (Control), Lane 2 (unfolded form). Part of the IP reactions of folded and unfolded forms of semi-glycosylated P25 protein contained two reactions, each of which were briefly treated with con A lysis buffer, CL2B beads, and Con A Sepharose beads. The Con A recognised the *N*-link oligosaccharides and the semi-glycosylated protein which was not bound and there were no related glycans (Glycan, 0-3rd) in Lanes 3 & 4 of both folded and unfolded forms.

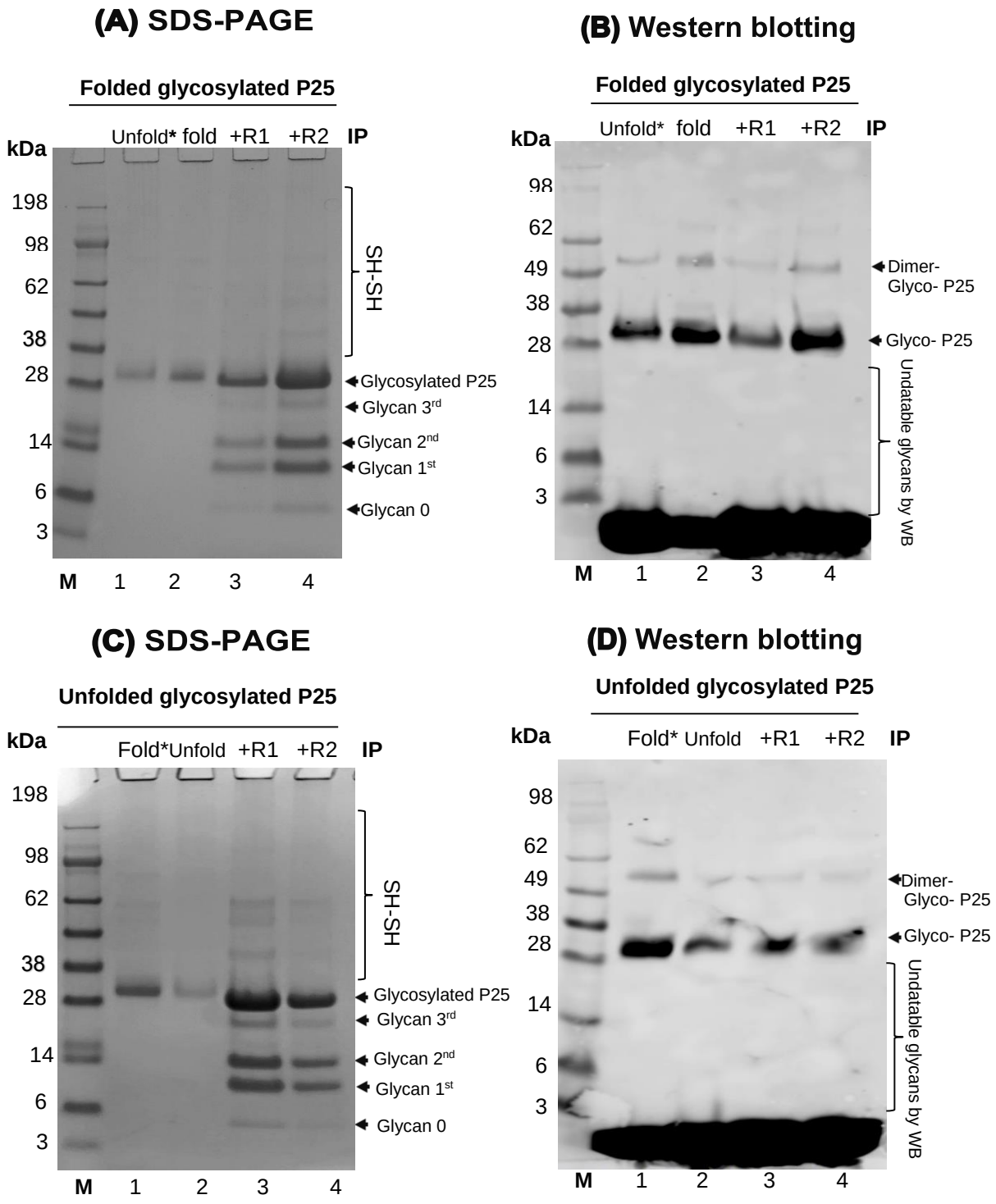


Figure 71. Lectin binding & IP (Con A) was performed on folded and unfolded forms of glycosylated P25 protein. The reactions initially used the Ni^{2+}purified secreted P25 glycoprotein concentrated (0.70 mg/mL) by a concentrator (10kDa MW) in Tris-HCl buffer (50 mM Tris-HCl, pH 9, and 95 mM NaCl). The folded form control reactions **(A-B)** are Lane 1 (Control), Lane 2 (unfolded form, treated with 8M urea) whereas the unfolded form control reactions **(C-D)** are Lane 1 (Control), Lane 2 (unfolded form). Part of IP reactions of folded and unfolded forms of glycosylated P25 protein contained two reactions, each of which were briefly treated with con A lysis buffer, CL2B beads, and Con A Sepharose beads. The Con A recognised the *N*-link oligosaccharides and the bound glycosylated protein and related glycans (Glycan, 0-3rd) in Lanes 3 (R1) & 4 (R2), two replications) including both folded and unfolded forms.

CONCLUSION AND DISCUSSION

1. General discussion of P25 protein

In a native organism (silkworm, *B. mori*), the P25 glycoprotein is a eukaryotic protein which is synthesised and secreted in small amounts because of its biological function within the fibroin complex (only one minor segment in the formation of an elementary fibroin unit). It has been suggested that P25 glycoprotein acts as a chaperone that facilitates the transport and secretion of H-fibroin (Tanaka et al., 1999). P25 glycoprotein is characterised by its distinctive *N*-glycan structure referred to as Asn-linked high-mannose type oligosaccharide chains. (Yamaguchi & Uchida, 1996). The distinctive *N*-glycan structure referred to as insect type is 3 mannoses (Man3) with α1,3- and α1,6-fucose residues (M3FF). (H Kajiura , 2021). However, previous studies reported that insects with *N*-glycan have active α1,3-fucosyltransferase (FUCT) and α1,6-FUCT which are particularly distributed in plants and mammals (Haltiwanger & Lowe, 2004; Kajiura et al., 2021). Additionally, protein glycosylation is a common and important post-translational modification among numerous proteins, involved in many biological processes such as cell adhesion, signal transduction, immune response, and inflammatory reaction (Haltiwanger & Lowe, 2004). *N*-linked glycosylation mostly occurs in eukaryotes and archaea, but it is uncommon in bacteria (Mizukami et al., 2018). It is important to note that individual proteins commonly have a hidden relationship between *N*-linked glycosylation and disulfide bonds which is referred to as a co/posttranslational modifications (PTMs). These two PTMs are mostly conserved and co-exist together in secreted or membrane proteins proving they directly interact with each other (Bakshi et al., 2022).

N-linked glycosylation was believed to be specifically unique to eukaryotes. However, it has now been found that the *N*-linked glycoproteins are involved in all domains of life such as archaea and bacteria. For instance, a research group, 'Fisher and colleagues', have enhanced and expanded the *N*-linked glycoproteins pathway to *E. coli* by adding a glycosylation tag as an extra strategy in the "glycoengineering toolbox" to help complete the *N*-linked glycosylation for the production of recombinant glycoproteins (Fisher et al., 2011). Another idea for the recombinant glycoprotein, the baculovirus-insect cell system, is likely to be more advantageous for producing proteins from the eukaryotic system. This system can serve as post-translational modification of *N*-linked glycosylation. In general, insect protein glycosylation pathways are simpler than the glycosylation of higher eukaryotes (Shi & Jarvis., 2007). Recently, the recombinant P25 glycoprotein (*B. mori)* has been commercially produced from the amino acid sequence 17-220 (without the signal sequence 1-16, Uniprot, Website 2022) in many hosts such as, yeast, *E. coli*, *In vivo* biotinylation in *E. coli*, baculovirus, and mammalian cells (Cusabio, Website 2022). These are available on the online market but are expensive.

Recombinant P25 protein (*B. mori*) is not a simple protein, but it is one of the glycoproteins whose behaviour and related challenging issues needs to be better understood. Not a lot of research has gone into this protein in the past. However, in this project P25 glycoprotein, an interesting subunit, has been revealed to be the one key fibroin proteins which helps to explain the silk fibroin mechanism and how it can be used in many ways, including for medical applications. Production of P25 glycoprotein

in two hosts (bacteria & baculovirus-insect cell) showed different qualities in its final active form as soluble P25 glycoprotein. This is explained in the following conclusion and discussion.

2. P25 protein production in bacteria (Prokaryotic system)

Based on its origin, P25 glycoprotein has eight Cys residues and it could possibly have four pairs of disulfide bonds. This glycoprotein could be named as a “rich disulfide bond” or “multi-disulfide bond” protein. However, there is no previous evidence as to whether these disulfide bonds in P25 glycoprotein are consecutive or non-consecutive types. Furthermore, P25 glycoprotein has been characterised as a glycoprotein containing three *N*-linked (Asn) residues in its full amino sequence. Thus, there is the challenge of finding the right host for suitable post-translational modification of these two types of residues containing (Cys and Asn) in its sequence. Hence, this project started the recombinant P25 glycoprotein production process with the most simple and standard host system for cloning the P25 gene fused with the C-terminus tagged (TEV protease and His6x tagged) and expressed in selective commercial bacterial strains. These selected bacterial strains (Gram-negative bacteria, *E coli*) make up the two groups as mentioned earlier, Group 1 (un-enhanced disulfide bond formation), including three bacterial cells; BL21(DE3), C41 (DE3), Rosetta™ (DE3)pLysS(Novagen) and Group 2 (enhanced disulfide bond formation) containing one bacteria strain, SHuffle® T7.

1) P25 glycoprotein production in Group 1 (un-enhanced disulfide bond formation), included three bacterial cells; BL21(DE3), C41 (DE3), Rosetta™ (DE3)pLysS(Novagen). Surprisingly, the findings showed that all P25 glycoprotein production in Group 1 with four different bacterial strains were located in the debris of the cell pellets (IBs). It could explain why the P25 glycoprotein might be overexpressed in the intracellular production leading to the formation of insoluble aggregates or IBs (Burdette et al., 20128) which might be caused by cellular stress. The resulting IBs are thought to be misfolded and biologically inactive (Sørensen & Mortensen., 2005). These IBs therefore needed to be recovered by solubilisation with the selected denaturant (7M urea containing 100mM HEPES, pH 8.0) in order to be bioactive P25 glycoprotein. Even the final eluted protein yield in this group of P25/pET28a, after being run through the Ni-NTA column and expressed in all four strains of different optimisation, culture conditions, temperature and IPTG concentrations, resulted in a low quality of protein when determined with (multiple bands) of SDS-PAGE and western-blotting. The experiments confirmed that the P25 glycoprotein expressed in these un-enhanced disulfide bonds strains is unable to glycosylate the protein. Furthermore, during the protein folding process in these four bacterial strains, completed disulfide bonds and *N*-linked glycosylation were not able to be formed through the P25 glycoprotein production because P25 glycoprotein has a type of eukaryotic protein containing eight cysteines which form disulfide bonds in the native state. However, these bacterial strains may not support the proper pairing of disulfide bonds (Tyedmers et al., 2010). According to Berkmen (2012), disulfide bonds can serve structures, catalytic and signaling roles. However, this highlights that four bacterial strains Group-1 (un-enhanced disulfide bond formation) does not provide a metabolic system or a suitable environment for disulfide bond formation. Landeta et al. (2018) reported that disulfide bond formation in *E. coli* requires a thioredoxin-family protein such as DsbA whose role is to act as an oxidant to catalyse sulfur-sulfur bond formation between pairs of cysteines in substrate protein.

2) P25 production in Group 2 (enhanced disulfide bond formation) containing a bacteria strain, SHuffle® T7. The strategy of experimental efforts in this project was to find a more suitable host than Group 1 bacterial strains because not much progress was possible using this strain of the soluble bioactive yield of P25 glycoprotein. Thus, SHuffle® T7 was used to improve the P25 glycoprotein bioactive yield. It was found in the small-scale production that P25 glycoprotein was more strongly expressed in almost all culture conditions with variation of different temperatures (15, 20, 25 and 28ºC) and IPTG concentrations. In all these conditions P25 protein was found to be located in IB pellets. However, on the larger scale, P25 glycoprotein was selectively expressed and produced at 25ºC with 0.5 mM IPTG induction. The P25 glycoprotein that was obtained showed a good quality of purified P25 glycoprotein after solubilisation against urea and refolding in Ni-NTA column. A single band of P25 glycoprotein was evident after SDS-PAGE and western blotting. The stability of P25 protein when stored at (~4ºC) was found to be about a week. The single band of P25 glycoprotein was identified by spectrometry technique as being similar to the precursor of fibroin light chain (*B. mori*). Light scattering determination showed that P25 glycoprotein was best maintained in a solution or buffer of about pH 8.0. However, after about 3 months of being maintained in the freezer (~20ºC), the purified protein of P25 construct no longer was in good quality, as the P25 gene showed itself to be unstable. Although the P25 protein was treated using the same protocol as previously, the resultant yield showed multiple bands after refolding by Ni-NTA column and visualised with SDS-PAGE and western blotting.

Hence, there was no observable secondary structure of P25 protein expressed in SHuffle® T7 cells using the CD technique. It would indicate that SHuffle® T7 could enhance the expression and folding of P25 glycoprotein most effectively because SHuffle® T7 cells have a genetically modified cytoplasm engineered from the selected *E. coli* strain that helps to oxidize cysteines within the P25 glycoprotein to form disulfide bonds correctly and expressed in the cytoplasm cells *dsb*C (Anton et al., 2016). It has previously been reported that, in the bacteria system, the *N*-linked glycosylation pathway occurs independently from protein translocation machinery and fully folded protein can be glycosylated (Kowarik et al., 2006). Although at the beginning of P25 protein production it seemed likely to be able to obtain a better-quality glycoprotein in SHuffle® T7 than in any other of the Group 1 strains, glycosylation of P25 glycoprotein in this selective strain, SHuffle® T7, produced an unstable protein expression resulting in poor quality of protein homogeneity. Ghadari et al (2012) reported that this is the most frequent occurrence in modification of recombinant proteins, and it could have a huge effect on the biological activity. Nevertheless, there was a research group (Hug et al., 2011) who investigated production of recombinant glycoprotein *Haemophilus pyloni* Furosyltransferase and successfully produced engineered *E. coli* to support post-translational modifications.

3. P25 glycoprotein production in baculovirus-insect cell system (Eukaryotic system)

The original fibroin proteins, natural untagged (H-chain, FIBL, and P25) which are secreted from the posterior silk gland cells of silkworms (Dong et al., 2013), belong to the order Lepidoptera. However, this recombinant P25 glycoprotein is labelled with a total of twelve His-tags on both sides of the N-terminus and C-terminus. The P25 glycoprotein includes His6x + 3C Protease at (N-terminus) and P25+ TEV+ His6x (C-terminus). The P25 glycoprotein including His6x + 3C Protease+P25+ TEV+ His6x, was successfully expressed and produced in Sf21 cells (Lepidoptera). According to Lemaitre et

al. (2019) this achievement would suggest that the P25 glycoprotein could be suitable for production in the baculovirus-insect cell system as it contains the appropriate post-translational modifications due to correctly folded formation. Location of P25 glycoprotein accumulation was determined from the final cell culture. Two types of P25 glycoprotein can be obtained from the harvested infected cell culture and the medium fraction after 72 hrs. 1) The semi-glycosylated P25 glycoprotein (non-secretion) was obtained from the soluble fraction of cell lysates (disrupted cell pellets) after centrifugation whereas 2) the glycosylated P25 (secretion) was obtained in the medium fractions. In this project, the level of secreted P25 glycoprotein in the medium fraction was greater than 10mg/mL in each batch. Moreover, in agreement with Lemaitre et al. (2019) the infected Sf21 cells were able to be grown in serum-free media which facilitated purification of the secreted P25 glycoprotein from the infected media. Experiments found that a high level of fully glycosylated P25 glycoprotein needed to be secreted from the cells into the medium. This is in line with what Rendić et al., (2008) reported; that glycan chains of insect cells were produced and secreted into the medium fractions, and some glycoproteins can be found in the supernatant of these insect cells.

Regarding P25 glycoprotein purification, it was necessary to add protease inhibitor as needed to the semi-glycosylated P25, but not to the glycosylated P25 glycoprotein during the purification process. The semi-glycosylated P25 seemed to be degraded more easily by the proteolytic process than the glycosylated P25 glycoprotein. This indicates that glycosylation and protein production in P25 are correlated. This is explained further in the secondary structure and glycan analysis discussed below. This project's findings on the secondary structure of semi-glycosylated and glycosylated P25 glycoprotein confirmed that the secondary structure elements of glycosylated P25 had stronger conformation than semi-glycosylated P25, as shown by the increased number of strand1/β-sheet (11.5%). There were significantly fewer helical structures (α-helix) present in glycosylated P25 than in semi-glycosylated P25. However, glycosylated P25 showed a remarkable increase in conformation of strand1 (β-sheet, ~11.5%) compared to the semi-glycosylated form. After full glycosylation and secretion into the medium fraction, secondary structure analysis confirmed the conformation transition of P25 glycoprotein from the cells to the medium fractions. The conformation elements of glycosylated P25 protein in medium fractions clearly demonstrate that glycosylation can occur in Sf21 cells before secretion into the suspension medium. Glycosylated P25 in the medium fraction shows a complex glycoprotein containing glycan attachments. Prior to this work, the secondary structure of P25 glycoprotein had not been reported. This project has made a significant discovery regarding these two types of secondary structures of P25 glycoprotein.

When undertaking the glycan analysis, two endoglycosidase enzymes treatments, (Endo-H & PNGase F) were used to look at the glycosylation processing event. The findings revealed that the Endo-H can affect some parts of the semi-glycosylated P25 glycoprotein both folded and unfolded forms. However, the PNGase F was able to release the glycan only from the glycosylated P25 glycoprotein seen clearly by the overlaying of two wavelengths 680R and l 800R channels combined with the western blotting. On the other hand, in the semi-glycosylated treatment using these two enzymes, the glycans attachment in this glycoprotein seemed not have been completed. Thus, there was no evidence of glycans being released when using the 680R and l 800R wavelength channels.

In a further glycan analysis using lectin binding/Immunoprecipitation (IP) techniques, the findings clearly showed the three attachments of all glycans to the glycosylated P25 glycoprotein as was previously reported by Tanaka et al., (1999) Inoue et al., (2000), but they did not report evidence of the glycans MW mobility on SDS-PAGE as this current ground-breaking work has done.

Based on the above improved results, there is strong evidence that the baculovirus-insect cell system could have the high efficacy for stable recombinant P25 glycoprotein.

FURTHER DIRECTIONS

1) The glycans attachment of P25 glycoprotein need to be identified.
2) The disulfide bonds formation of P25, whether non-seductive or consecutive linkage, should be investigated.

BIBLIOGRAPHY,

Anton, B.P., Fomenkov, A., Raleigh, E.A. & Berkmen, M. 2016. 'Complete genome sequence of engineered *Escherichia coli* SHuffle strains and their wild-type parents'. *Genome Announcements*. 4(2): e00230-16.

Bakshi, T., Pham, D., Kaur, R. & Sun, B. 2022. 'Hidden relationship between N-glycosylation and disulfide bonds in individual proteins'. *International Journal of Molecular Science*. 23(3742):1-25.

Burdette, L.A., Leach, S.A., Wang, H.T. & Tullman-Ercek. 2018. 'Developing gram-negative bacteria for the secretion of heterologous proteins'. *Microbial Cells Factories*. 17(196):1-16.

Berkmen, M.,2012. 'Production of disulfide bonds proteins in *Escherichia coli'. Protein Expression and Purification*. 82:240-251.

Bini, E, Knight D.P & Kaplan, D.L, .2004. 'Mapping domain structures in silks from insects and spiders related to protein assembly'. *Journal of Molecular Biology*. 335(1): 27–40.

Couble, P., Chevillard, M., Moine,A., Ravel-Chapuis, P. & Prudhomme, J.C. 1985. 'Structural organization of the P25 gene of *Bombyx mori* and comparative analysis of its flanking DNA with that of fibroin gene'. *Nucleic Acids Research*. 13: 1801-1814.

CUSABIO. 'Recombinant *Bombyx mori* Fibrohexamerin P25.' Available from https://www.cusabio.com /Recombinant-Protein/Recombinant-*Bombyx-mori*-Fibrohexamerin-P25--308531.html [Access 20th July, 2022].

Chevillard M, Couble P, Prudhomme J C, 1986. 'Complete nucleotide sequence of the gene encoding the *B. mori* silk protein P25 and predicted amino acid sequence of the protein'. *Nucleic Acids Researc*h. 14, 6341–6342.

Dong, Z., Zhao, P., Wang, C., Zhang, Y., Chen, J., Wang, X., Lin, Y., Xia, Q. 2013. 'Comparative proteomics reveal diverse functions and dynamics changes of *Bombyx mori* silk proteins spun from different development stages'. *Journal of Proteome Research*. 12: 5213-5222.

Fisher, A.C., Haitjema, C.H., Guarino, C., Celik, E., Endicott, C.E., Reading, C.A., Merritt, J.H., Ptak, A.C., Zhang, S. & Delisa, M.P. 2011. 'Production of secretory and extracellular N-linked glycoproteins in *Escherichia coli*'. *Applied and Environmental Microbiology*. 77(3): 871-881.

Ghaderi, D., Zhang, M., Hurtado-Ziola, N. & Varki, A. 2012. 'Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation'. *Biotechnology & Genetic Engineering Reviews*. 28:147–175.

Hug, I., Zheng, B., Reiz, B., Whittal, R.M., Fentabil, M.A., Klassen, J.S. & Feldman, M.F. 2011. 'Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein'. *Journal of Biological Chemistry*. 286(43): 37887-37894.

Landeta, C., Boyd, D. & Beckwith, J. 2018. 'Disulfide bond formation in prokaryotes'. *Nature Biology.* 3:270-280.

Love, N.R.; Thuret, R.; Chen, Y.Y.; Ishibashi, S.; Sabherwal, N.; Paredes, R.; Alves-Silva, J.; Dorey, K.; Noble, A.M.; Guille, M.J.; et al. 2011. 'pTransgenesis: A cross-species, modular-transgenesis resource'. *Development*. *138*: 5451–5458.

Liu, W.B., Gong, C.L., Xue, R.Y., Zhou, W.L., Zhu, X.X. & Cao, G.L. 2007. 'Cloning and analysis Fhx/P25 gene promoter of *Bombyx mori* fibroin protein'. *Zoological Research*. 28(1):17-23. (in Chinese).

Khudkiewicz, B., Takasu, Y., Fedic, R., Tamura, T., Sehnal, F. & Zurovec, M. 2009. 'Structure and expression of the silk adhesive protein Ser2 in *Bombyx mori'*. *Insect Biochemistry and Molecular Biology*. 39:938-946.

Kajiura, H., Miyauchi, R., Kakudo, A., Ohashi, Misaki, R. & Fujiyama, K. 2021. ' *Bombyx mori* β1,4-*N*-acetylgactosaminyltransferase possesses relax donor substrate specifically in *N*-glycan synthesis'. *Scientific Reports*. 11(5505):1-15.

Kowarik, M, Numao, S, Feldman, M.F, Schulz, B.L., Callewaert, N., Kiermaier, E., Catrein, I., Aebi, M. 2006. ' *N*-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase'. *Science.* 314 (5802):1148-1150.

Haltiwanger, R. S. & Lowe, J. B. 2004. 'Role of glycosylation in development'. *Annual Review of Biochemistry*. 73: 491–537.

Inoue, S., Tanaka, K.,Arisaka, F., Kimura, S., Ohtomo, K. & Mizono, S. 2000. 'Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio'. *The Journal of Biological chemistry*. 275 (51):40517-40528.

Inoue, S., Tanaka, K., Tanaka, H., Ohtomo, K., Kanda, T., Immuru, M., Quan, G.X., Kojima, K., Yamashita, T., Nakajima,T., Taira, H., Tamura, T. & Mizono, S. 2004. 'Assembly of the silk fibroin elementary unit in endoplasmic reticulum and a role of L-chain for protection of α 1.2-mannose residues in N-linked oligosaccharide chains of fibrohexamerin/P25'. *European Journal of Biochemistry*. 271:356-366.

Lambony, L., Souho, T., Osi, A.& Yang, G. 2018. 'Silk protein: A natural resource for biomaterials'. *In* Bioinspired Materials Science and Engineering. Wiley, USA. 389pp.

Lu, F, Wei, Z., Luo, Y., Guo, H., Zhang, G., Xia, Q. & Wang, Y. 2019. 'SilkDB 3.0: visualizing and exploring multiple levels of data for silkworm'. *Nucleic Acids Research*. 48(1): D749-D755.

Lemaitre, R.P., Bogdanova, A., Borgonavo, B., Woodruff,J.B. & Drechsel, D.N. 2019. 'FlexiBAC: a versatile open-source baculovirus vector system for protein expression, secretion, and proteolytic processing'. *BMC Biotechnology*. 19(20):1-11.

Mizukami, A. & Caron, A.L. 2018. 'Platforms for recombinant therapeutic glycoprotein production'. *In* Chapter 1, Recombinant glycoprotein production, methods, and protocols. Hamana Press, USA. 285pp.

NCBI, 2022. '*Bombyx mori* silk protein P25 (P25), mRNA'. National Library of Medicine, National Center for Biotechnology Information. [Online access; 28/06/2022], mcbi.nih.gov/nucore/nm-001145941.1.

NCBI, 2022. 'iCn3D:web-based 3D structure viewer'. National Library of Medicine, National Center for Biotechnology Information. . [Online access; 21/10/2022], https://www.ncbi.nlm.nih.gov/ Structure/icn3d/docs/icn3d_about.html.

Nony, P., Prudhomme, J.C & Couble, P. 1995. 'Regulation of the P25 gene transcription in the silk gland of *Bombyx mori*'. *Biology Cell.* 84:43–52.

Ohno, K., Sawada, J., Takiya, S., Kimono, M.,Matsumoto,A., Tsubota, T., Uchino, K., Hui, C., Sezutsu, H., Honda, H. & Suzuki, K. 2013. 'Silk gland factor-2, in fibroin gene transcription consists of LIM homeodomain, LIM-interacting and single-stranded DNA-binding proteins'. *The Journal of Biological Chemistry*. 288(44):31581-31591.

Rendić, D, Wilson, I.B.H. & Paschinger, K. 'The glycosylation capacity of insect cells'. *Croatica Chimica Acta*. 81(1):7–21.

Rockwood, D.N., Preda, R.C., Yücel, T., Wang, X., Lovett, M.L.& Kaplan, D.L. 2011. 'Materials fabrication from *Bombyx mori* silk fibroin', *Nature Protocols*. 6(10):1612-1631.

Sari, D., Gupta, K., Balaji, D., Raj, D.B.T.G., Aubert, A., Drncova,P., Garzoni, F., Fitzgerald,D. & Berger, I. 2016. 'The MultiBac baculovirus/ Insect cell expression vector system for producing complex protein Biologics'. *In* Technologies for protein complex production and characterization. Spring International Publishing, Switzerland. pp 199-125.

Sehnal F. & Zurovec M. 2004.' Construction of silk fiber core in lepidoptera'. *Biomacromolecules* 5(3): 666–674.

Shi, X. & Jarvis, D.L. 2007. 'Protein *N*-glycosylation in the baculovirus-insect cell system'. *Current target*. 8(10):116-1125.

Shimizu, K., Ogawa, S., Hino, R., Adachi, T., Tamita, M., Yoshizato, K. 2007. 'Structure and function of 5'-flanking regions of *Bombyx mori* fibroin heavy chain gene: Identification of a novel transcription element with a homeodomain protein-binding motif'. *Insect Biochemistry and Molecular Biology*. 37(7):713-725.

Sørensen, H.P. & Mortensen, K.S. 2005. 'Soluble expression of recombinant protein in the cytoplasm of *Escherichia coli*'. *Microbial Cell Factories*. 4(1):1-8.

Sreerama, N. & Woody, R.W. 2000. 'Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set'. *Analytical Biochemistry.* **287(2)**: 252-260.

Tanaka, K., Inoue, S. & Mizuno, S. 1999. 'Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by *Bombyx mori'*. *Insect Biochemistry and Molecular Biology*. 29:269–276.

Teule. F, Cooper, A.R., Furin, W.A., Bittencourt, D., Rech, E.L., Brooks, A. & Lewis, R.V. 2009. 'A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning'. *Nature Protocols*. 4(3):341-355.

Tyemers, J., Mogk, A. & Bukau, B. 2010. 'Cellular strategies for controlling protein aggregation'. *Nature Review Molecular Cell Biology*. 11:777=788.

UniProt. PO4148.SI25_BOMMO, Fibrohexamerin. *Bombyx mori* (Silk moth). *Gene*. Available from https://www.uniprot.org/uniprotkb/P04148/entry, [Access; 9/10/ 2022].

Walker, A.A., Church, J.S., Woodhead, A.L. & Sutherland, T.D. 2013.'Silverfish silk is formed by entanglement of randomly coiled protein chains'. *Insect Biochemistry and Molecular Biology*. 43(7):572-579.

Wu, M., Ruan, J., Ye, X.G., Zhao, S., Tang, X.L., Wang, X.X., Li, H.P. & Zhong, B.X. 2021. 'P25 gene knockout contributes to human epidermal growth factor production in transgenics silkworms'. *International Journal of Molecular Sciences*. 22(2709): 1-12. https://doi.org/ 10.33901 /ijms22052709.

Yamaguchi, H & Uchida, M, 1996. 'A chaperone-like function of intramolecular high-mannose chains in the oxidative refolding of bovine pancreatic RNase B'. *The Journal of Biochemistry*. 120,474-477.

Zabelina, V., Takasu, Y., Sechadova, H., Yonemura, N., Nakajima, K., Sezutsu, H., Sery, M., Zurovee, M., Sehnal, F. & Tamura, T. 2021. 'Mutation in *Bombyx mori* fibrohexamerin (P25) gene causes reorganization of rough endoplasmic reticulum in posterior silk gland cells and alters morphology of fibroin secretory globule in the silk gland lumen'. *Insect Biochemistry and Molecular Biology*. 135:1-6.

CHAPTER V

In vitro Silk Fibroin Complex, Its Indigenous Interactions and Preliminary Interactions with Human Proteins

ABSTRACT

This chapter focuses on the interactions 1) of three fibroin molecules (*Bombyx mori*) (H-chain, L-chain/FIBL and P25) within their indigenous formation and 2) with two other selected human proteins as known Chloride Intracellular Ion Channel proteins, (CLIC1 and CLIC4). The large molecule, H-chain was obtained from the posterior silk gland (PSG) of a Thai male silkworm on the last day of the 5th instar larvae. Two other lower MW subunits 1) FIBL $_{\text{TEV+his6-tagged}}$ and 2) $_{\text{his6-tagged+3C}}$P25 $_{\text{TEV+his6-tagged}}$ were achieved by recombinant technology and expressed in bacteria and baculovirus-infected insect cell system. The findings showed that purified H-chain isolation and purification protocol was successfully established. The silk fibroin complex within its indigenous molecules (H-chain, FIBL, and P25) could potentially form both homo and heterodimers. This complex specifically contains the H-chain (both native without the urea and partially folded including 5M urea) which can be seen from the Native PAGE (H-chain (~900kDa), FIBL (~32kDa), and P25 (~30kDa). However, the SDS-PAGE showed the H-chain (~350kDa), FIBL (~32kDa), and P25 (~30kDa). Additionally, the H-chain could also be formed with other molecules (CLIC1 and CLIC4).

This molecular mass of native H-chain was first identified in this work, indicating that native chain's natural behaviour as found in the native organism is more stable in its dimer formation. The native state of a fibroin complex containing H-chain, FIBL and P25 (6:6:1) would have about 4.6MDa (dimeric form) and 2.3MDa (monomeric form). SEC-MALS results showed the varied native MWs of five fibroin complex formations 1) H-chain alone (~2.97 x10^6 Da), 2) H-chain+P25 (~7.35 x10^9 Da), 3) H-chain+FIBL, 4) FIBL+P25 (1.35 x10^9 Da) and 5) three mixed subunits (H-chain+FIBL+P25, undetectable MWs formation due to the limitation of the machine not for the over huge molecule).

Three fibroin molecules *in vitro* formation complex were able to be seen using agarose Ni-NTA beads assay. When these three molecules, including H-chain (not tagged), and two lower MWs with their tagged FIBL $_{\text{TEV+his6-tagged}}$ and $_{\text{his6-tagged+3C}}$P25 $_{\text{TEV+his6-tagged , and}}$ were eluted into the solution at the same time, this fibroin presented three intact bands on SDS-PAGE (H-chain (~350kDa), FIBL (~32kDa), and P25 (~30kDa). These three mixed formed intact fibroins had different MWs when compared with the three intact molecules from the native organism appeared on SDS-PAGE, (H-chain (~350kDa), P25 (~30kDa), and FIBL (26kDa).

Keywords; PSG, SEC-MALS, SEC, native H-chain, Native PAGE

INTRODUCTION

It has been reported that three-components, H-chain, FIBL and P25 in the fibroin complex comprise the common factor among silk-producing insects in Lepidoptera (Tanaka et. al., 2001). However, Fibroin *(B. mori*) is composed of two different genes which produce three fibroin proteins; H-chain and FIBL which are synthesised in the same gene, but in a different position on the chromosomes (H-chain=25th and L-chain/FIBL=14th, and P25 which is encoded by the fibrohexamerin gene (Couble et al., 1985). In the recent study, the fibroin genes involved in the mature instar larvae silkworm were investigated (4th and 5th instar) in order to obtain a better understanding of silk protein synthesis such as transcription factors (Y—box and poly A-binding proteins). These various silk formation regulations play important roles in fibroin synthesis, including nucleotide binding, ATP binding and protein folding (Ma et al., 2021). Natural silk fibroin is a fibril-forming protein made up and secreted specifically from the cells of posterior silk gland, PSG into its lumen content (Toshima, 2005)**.** Fibroin is a huge protein complex, about 2,300kDa, containing three subunits, one main subunit H-chain (~350kDa), and two other small molecule subunits, FIBL (~26kDa), and P25 glycoprotein (~30kDa).The elementary complex of fibroin is 6:6:1 molar ratio of H-chain, FIBL and P25 constituting six sets of heterodimer of the H-chain linked to a disulfide bond, FIBL, and one molecule of P25 (Inoue et al., 2000). The two minor proteins, FIBL and P25, are thought to be secreted in the monomeric molecule, but a massive macromolecule of H-chain could be easily transported and secreted by forming a complex (Wang et al., 2015). Fibroin elementary complexes contain heterodimers which are absolutely crucial for highly efficient secretion of the fibroin out of the PSG cells (Takei et al., 1984). Basically, these large complexes do not interact with the membranes, but the interaction might exhibit the physical role of secretion via cell membranes on the PSG. Previously, there were two reports which mentioned that silk fibroin is secreted by exocytosis (Sasaki et al., 1980 & 1981). These fibroin proteins are secreted as an aqueous solution to make a single fiber by shear stress within the narrowing part of the silkworm's mouth using the spinneret and applying tension. Fibroin proteins synthesis is controlled by various factors containing a complex hierarchical structure to eventually form a unity fiber and contributes to other aspects of fibroin formation. Therefore, studying their subunits *in vitro* formation may help us to better understand the specific biological self-assembly mechanism involved. Various biomolecular techniques could reveal the fibroin formation as its native mechanism before being spun out and these could link their functions to future uses.

MATERIALS AND METHODS

The main experiments of FIBL production in bacteria were described in Chapter II.

RESULTS 1. Insect preparation

1.1) Fresh PSG preparation for solid fibroin

Based on the method of Katagata et al. (1984) **(Figure72)**, the solid fibroin was initially prepared for extraction and isolation in its partially folded form of H-chain in 5M urea, 20mM Tris-HCl, pH 8.0. To do that, the solid fibroin was prepared as described previously as a fibroin solution that generally exhibits three intact fibroin proteins (H-chain, FIBL and P25) altogether by experimentation in various processes as shown in **Figure 72.** Then the intact fibroin solution in its partially folded form was firstly purified with the SEC elution fractions corresponding to the right peaks and the present bands were collected and checked on the SDS-PAGE.

Figure 72. Fibroin solution preparation process; silkworm larva (5th instar, day 5th before spinning) was dissected for collection of the posterior silk gland **(A**) and washed with KCl, treatment with different concentrations of ethanol and ether. Then, the solid fibroin was obtained for solubilisation with 60% LiSCN **(B**). Solid fibroin 5mg/ mL was dissolved in 60% LiSCN at the starting point **(C)** and after being solubilised and shaken with vortex for about 1hr, the solid fibroin became a solution which underwent carboxymethylation and further purification using SEC and anion Ion Exchange Chromatography, respectively **(D)**.

1.2) H-chain isolation and purification from solid fibroin

Fibroin solution was precipitated with 19% & 30% saturated in $(NH_4)_2SO_4$), purified by SEC and the corresponding bands were determined by 10% SDS-PAGE gels **(Figure 73A-D**). As shown in **Figure 72**, the H-chain can be isolated by the modified method of Katagata et al. (1984) integrated with the isolation by ion exchange chromatography (IEC). To get the single band of H-chain, the solid fibroin (PSG) should be carboxymethylated before being isolated with IEC **(Figure 74).** If the solid fibroin (PSG) was not carboxymethylated, the eluted fraction of the three subunits of fibroin protein from IEC remained intact **(Figure 75).**

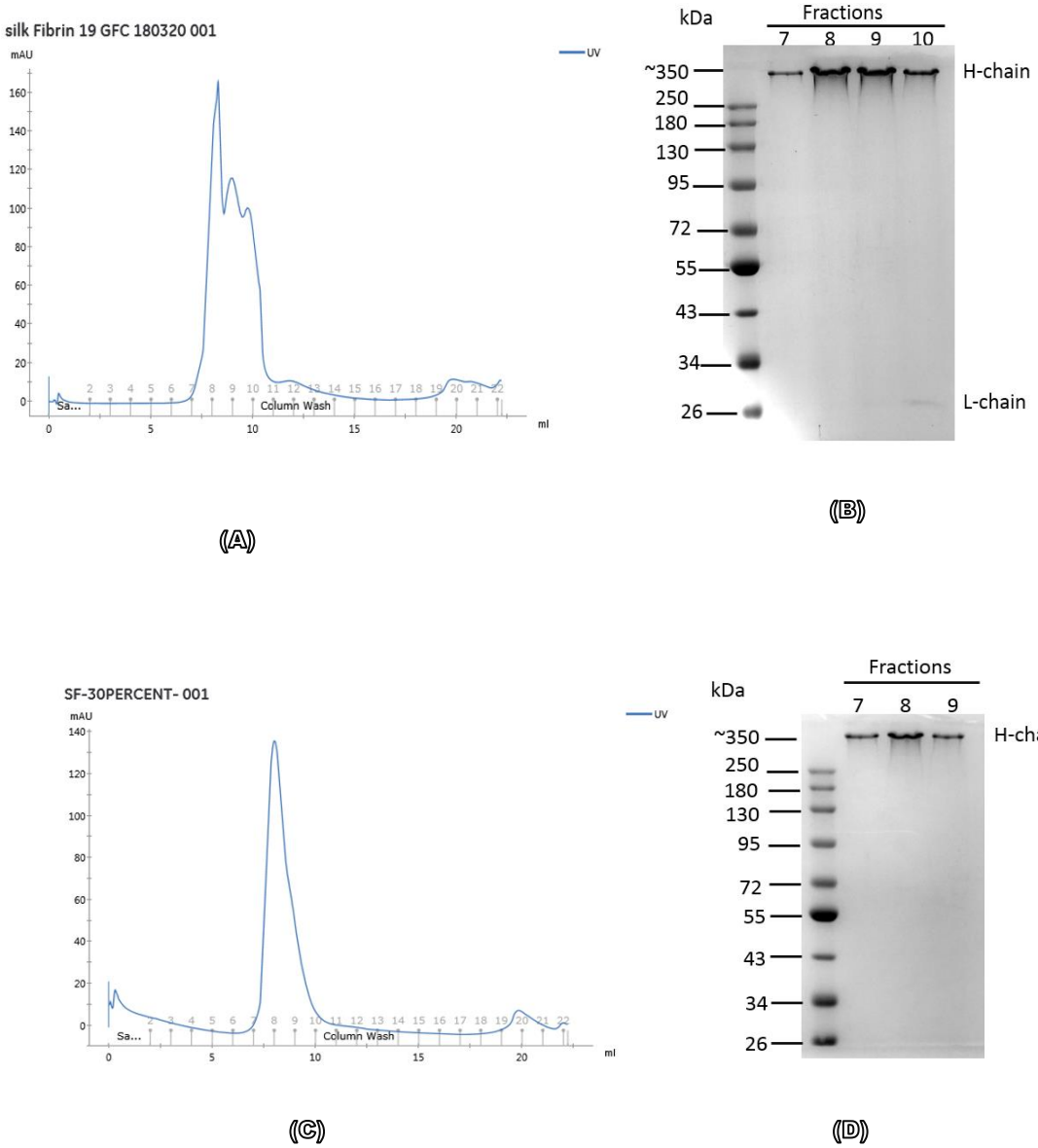


Figure 73. **(A)** Silk fibroin solution in 60% LiSCN was carboxymethylated, precipitated with 19% **(A)** saturated $(NH_4)_2SO_4$) at room temperature, then the pellet was collected for 19% saturation. The supernatant went immediately for higher precipitation with 30% **(C)** saturated $(NH_4)_2SO_4$) at room temp overnight. The following day, the pellet was collected and solubilised with 2mL of 60% LiSCN, and dialysed against buffer A. The resulting fibroin was concentrated into 1 mL and injected into the column (Superdex 200, 10/300 GL). After running for 30 mins, the eluted fractions; 7-9 **(A & C)** were collected and the corresponding peaks were identified by 10% gel SDS-PAGE for 2hrs 300 V, 30 Amp, and stained with Coomassie blue. The SDS-PAGE reveals the H-chain band at about 350 kDa in all fractions **(19% B)& (30% D)**.

1. 2.1) The 2nd H-chain purification from partially folded form (with carboxymethylating)

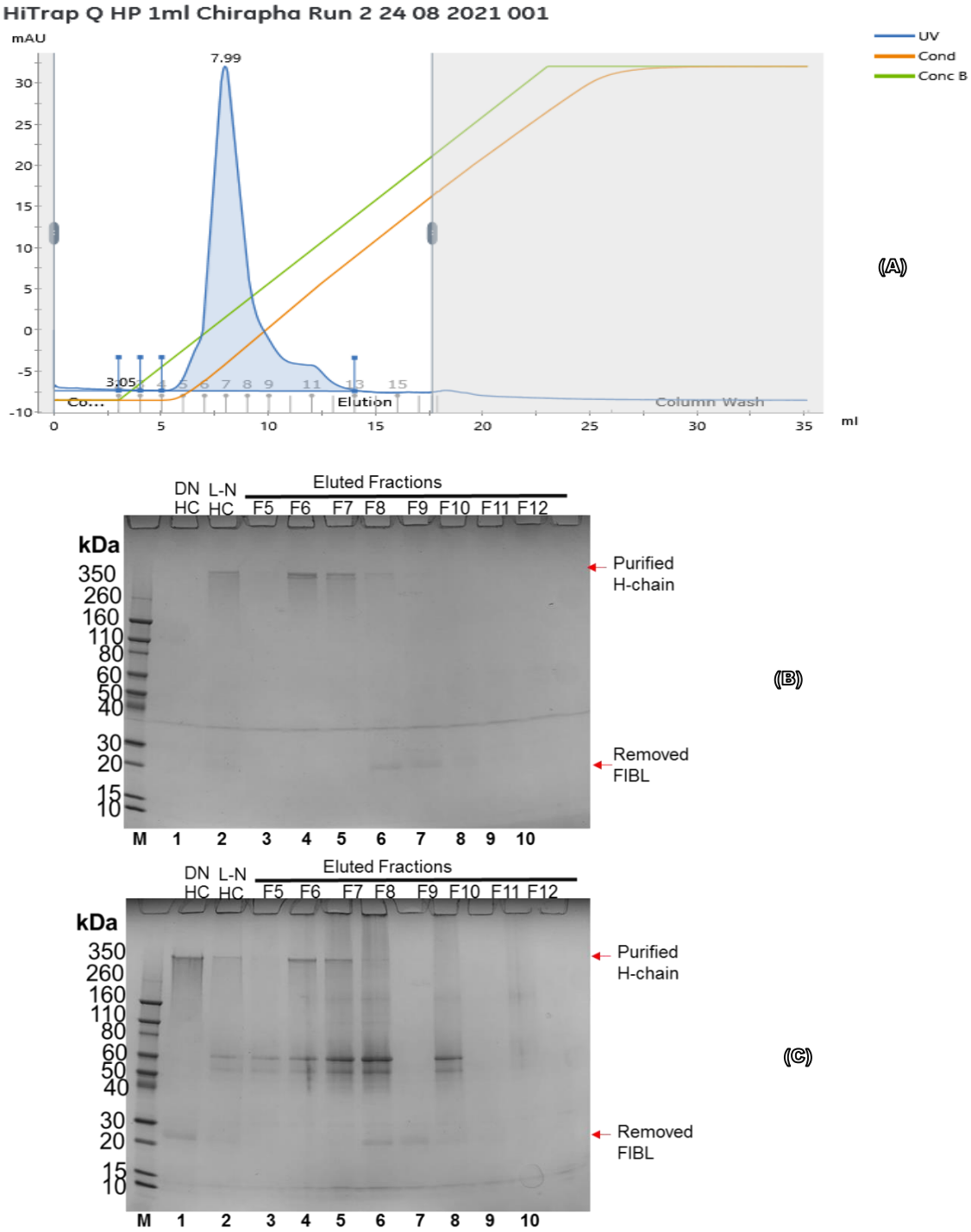


Figure 74. The partially folded purified H-chain, 0.400mg/mL in 5M urea, 20mM Tris-HCl, pH 8.0 after the 1st purification using SEC so that the urea was desalted out and changed into a native buffer (fully folded) in 100mM NaCl, 20mM Tris-HCl, pH 8.0. Consequently, the native H-chain was purified a 2nd time using anion Ion Exchange Chromatography by salt gradients method with a 1mL HiTrap HP Q column (**Figure 74: A**). The H-chain fractions were collected corresponding to the peaks and immediately a sample buffer, 20mM DTT, was added, boiled and run on reducing 4-20% SDS-PAGE gel **(Figure 74: B)**. All samples were kept for 7 days and run on 4-20% SDS-PAGE gel **(Figure 74: C).**

1.2.2) The 2nd H-chain purification from partially folded form (without carboxymethylating)

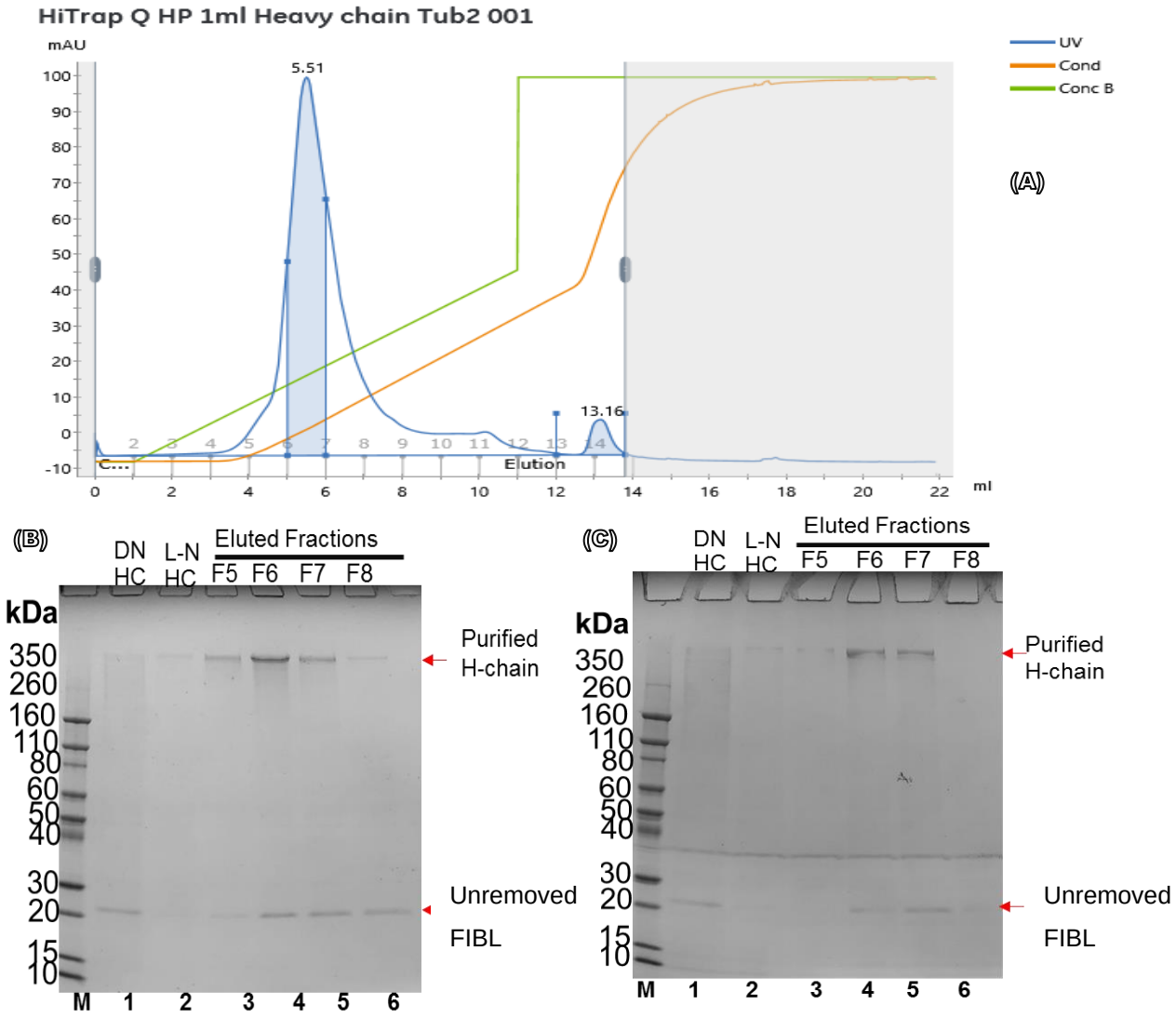


Figure 75. The fibroin solution solubilized from solid fibroin (1.0 mg/mL in 5M urea, 20mM Tris-HCl, pH 8.0) was subjected to the method of Katagata, et al. (1984) but without carboxylmethylation and without purification using SEC. The urea in the fibroin solution was desalted out and changed into a native buffer (fully folded) in 100mM NaCl, 20mM Tris-HCl, pH 8.0. Consequently, the native H-chain was purified a 2nd time using anion Ion Exchange Chromatography by salt gradients method with a 1mL HiTrap HP Q column **(Figure 75: A**). The H-chain fractions were collected corresponding to the peaks and immediately a sample buffer, 20mM DTT, was added, boiled and run on reducing 4-20% SDS-PAGE gel **(Figure 75: B)**. All samples were kept for 7 days and run on 4-20% SDS-PAGE gel **(Figure 75: C).**

RESULTS (Continued)

2. Investigation of in vitro silk fibroin assemble complexes

2.1. Observation of the initiation of silk fibroin complexes formation or multimerisation behaviour along its three indigenous proteins (H-chain, FIBL & P25) and interaction with selected human proteins (CLIC1 & CLIC4).

Figure 76, to start formation of fibroin complexes, the quality of two lower purified recombinant fibroin proteins (FIBL & P25) containing His6-tag were checked on a SDS-PAGE gel and western blotting, whereas the quality of the protein from a cocoon shell of purified H-chain isolated from the PSG was compared and checked only on a SDD-PAGE gel.

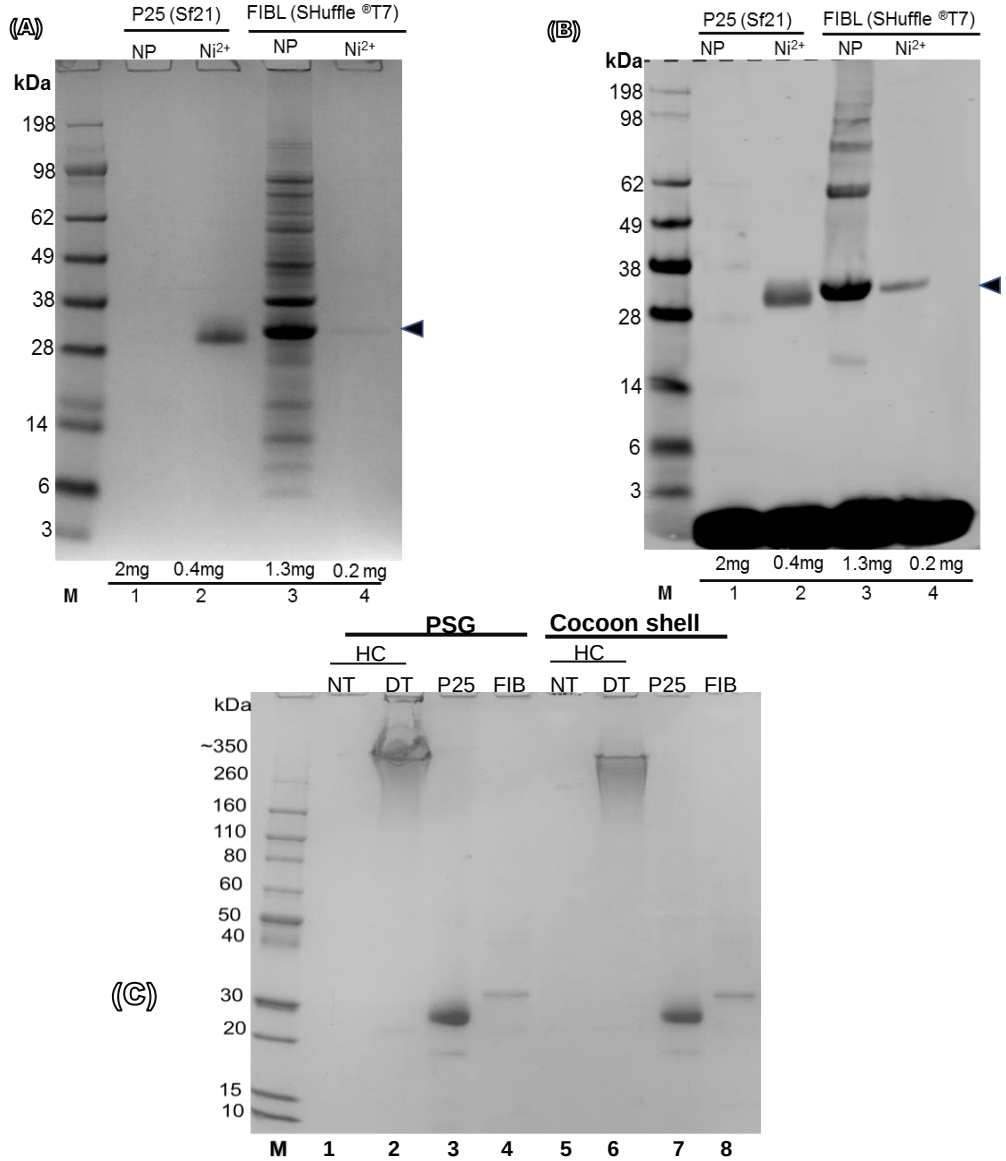


Figure 76. Samples of non-purified (NP) and purified P25 expressed in Sf21 and FIBL expressed in SHuffle® T7 named as (Ni^{2+}) proteins with His6-tagged after purification from Ni-NTA column (**A-B,** Bis-Tris 4-12%) were determined by SDS-PAGE and western blotting with anti-6x Histag antibody, whereas H-chain (HC) was isolated from PSG and a cocoon shell without the tag determined only on a SDS-PAGE (**C,** Novex,4-20% gel).

2.1 Observation of the initiation of the silk fibroin complexes formation (Continued)

Table 3. Measurement of silk fibroin proteins complexes formed overnight (stored in the refrigerator)

Parameters	H-chain from PSGs (Saturation 19%)	
	Protein concentration (mg/mL)	
	On ice (incubation)	RT (incubation)
HC alone/ (H-chain alone)	0.67	0.68
HC+P25	0.80	0.59
HC+FIBL	0.72	0.56
FIBL+P25	0.25	0.25
HC+P25+FIBL	0.32	0.31

When the time (1 hour) for completing incubation of the *in vitro* formation complexes with both treatment on ice and at room temperature (RT) **(Table 3)** had ended, the samples of each condition was investigated using multimers of Native-PAGE **(Figure 77A)**. However, fibroin complexes were also detected using non-reducing SDS-PAGE **(Figure 77B).** Furthermore, very interesting reactions were found in the sample formation complexes, especially when the reductant sample buffers DTT and BME, were added, mixed and run on the gel. Interestingly, it was found that the H-chain alone, HC+P25, HC+FIBL and HC+P25+FIBL formed multimers of the same size over 720 kDa (Native PAGE, **Figure 77 (A),** but when the complexes were separated with SDS-PAGE, the HC alone was not shown on the SDS-PAGE (**Figure 77B**), In addition, it was very interesting that the H-chain was not present when the reducing agents (DTT or βME) were added into the sample buffer before loading onto the gels, SDS-PAGE **(Figure 78. A&B**).

On the native gel **(Figure 77 A**), it seems that whenever there are only native H-chains in the reaction, they will stick to other native H-chains to form a homodimer above 720 kDa. However, when two different subunits, the native FIBLs (L-chains), were added into the reaction with native H-chains, another phenomenon arises - the H-chain will stick to them as a single disulfide bond, to form heterodimers.

In two different subunits, there are H-chains and FIBLs together, thus the H-chains could bind to other H-chains and link to FIBLs in the subunit. In three different subunits, there are H-chains, FIBLs, and P25 together, so maybe the H-chains stick to other H-chains and stick to FIBLs and associate to a P25 in the subunit. On the other hand, no dimer bands appear on the native gel, if the reaction does not include H-chains, but has two lower molecular weight components, FIBL and P25.

(Figure 79), it was found that H-chain can be formed the homo-dimer with two fibroin subunits and two human protein CLIC1 and CLIC4.

2.1. Observation of the initiation of the silk fibroin complexes formation (Continued)

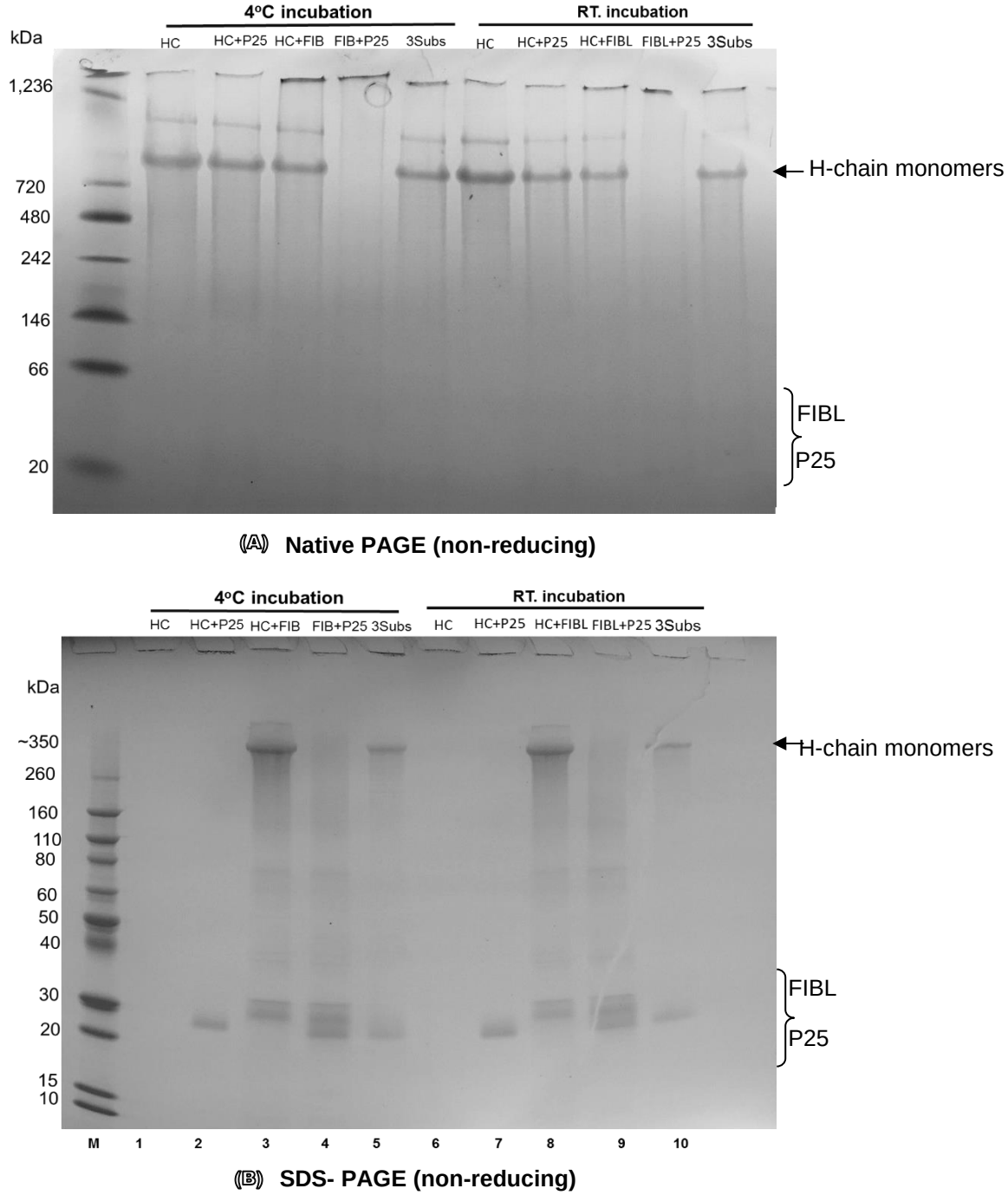
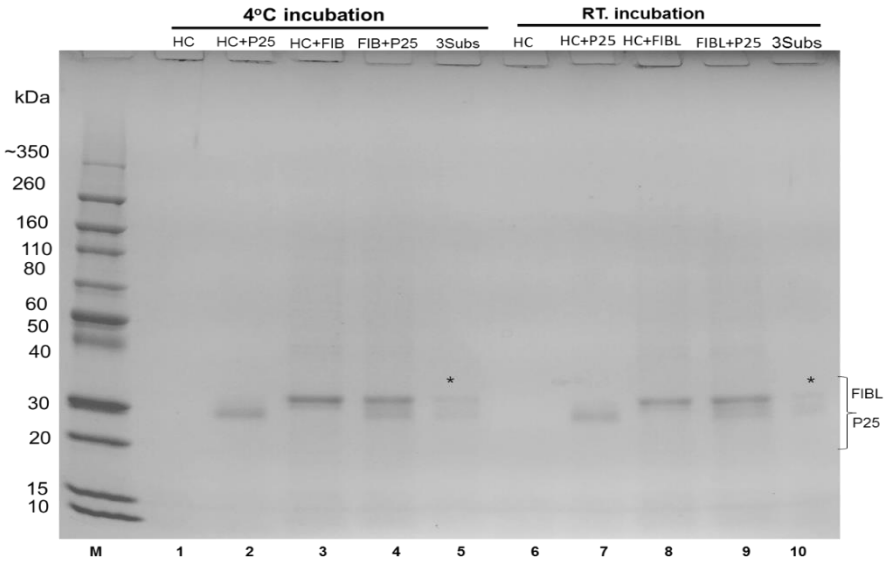


Figure 77. Three purified fibroin subunits were prepared in the same buffer (150mM NaCl, 50mM Tris-HCl, pH 8.0). Five reactions were set in each treatment (incubated at 4ºC or at room temperature), Heavy chain alone (HC), Heavy chain+P25 (HC+P25), Heavy chain +Fibroin light chain (HC+FIBL), Fibroin light chain+P25 (FIBL+P25), and Heavy chain +Fibroin light chain+P25 (3 subunits). The reactions were incubated either on ice (4ºC) or at room temperature (RT) for 1 hour. Consequently, the samples from each reaction were treated with the sample buffer without any reducing agent and determined by Native PAGE **(Figure 77 A**) and all reactions were collected, mixed with the sample buffer, boiled at 85ºC for SDS-PAGE **(Figure 77 B)**, Novex, 4-20%.

RESULTS (Continued)

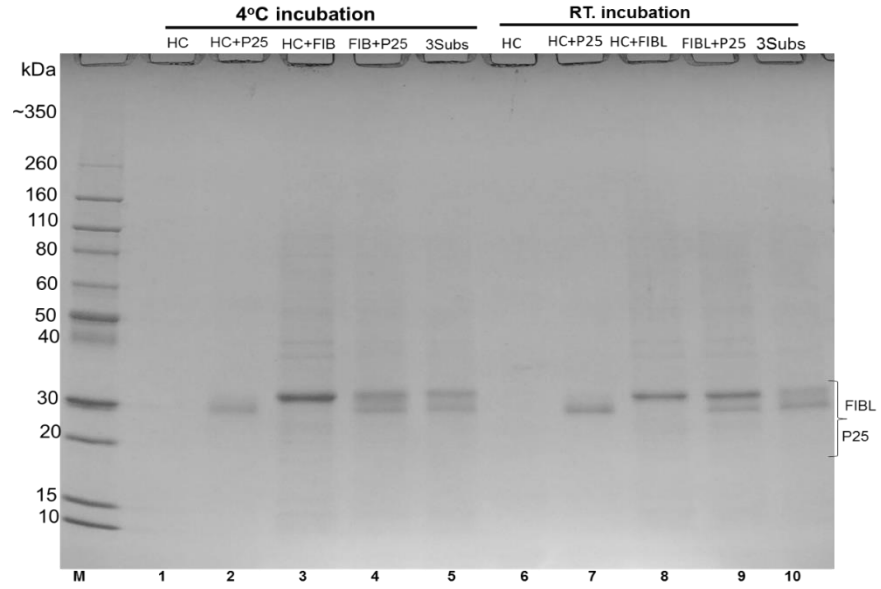
2.1. Observation of the initiation of silk fibroin complexes formation (Continued)

1) Reducing with βME



(A) SDS- PAGE (reducing with BME)

2) Reducing with DTT



(B) SDS- PAGE (reducing DTT)

Figure 78. Three purified fibroin subunits were prepared in the same buffer (150mM NaCl, 50mM Tris-HCl, pH 8.0). Five reactions were set in each treatment (incubated at 4^0C or at room temperature), Heavy chain alone (HC), Heavy chain+P25 (HC+P25), Heavy chain +Fibroin light chain (HC+FIBL), Fibroin light chain+P25 (FIBL+P25), and Heavy chain +Fibroin light chain+P25 (3 subunits). The reactions were incubated either on ice (4^0C) or at room temperature (RT) for 1 hr. After the treatments, all samples were maintained in the refrigerator overnight Following day, the samples from each reaction were treated with the sample buffer containing 10mM (βME) **(Figure 78 A)** and all reactions were collected, mixed with the sample buffer containing 10mM (DTT), boiled at 85^0C, separated by SDS-PAGE, Novex, 4-20% **(Figure 78 B).**

RESULTS (Continued)

2.1. Observation of the initiation of silk fibroin complexes formation between three indigenous proteins (H-chain, FIBL & P25) and selected human proteins (CLIC1 & CLIC4).

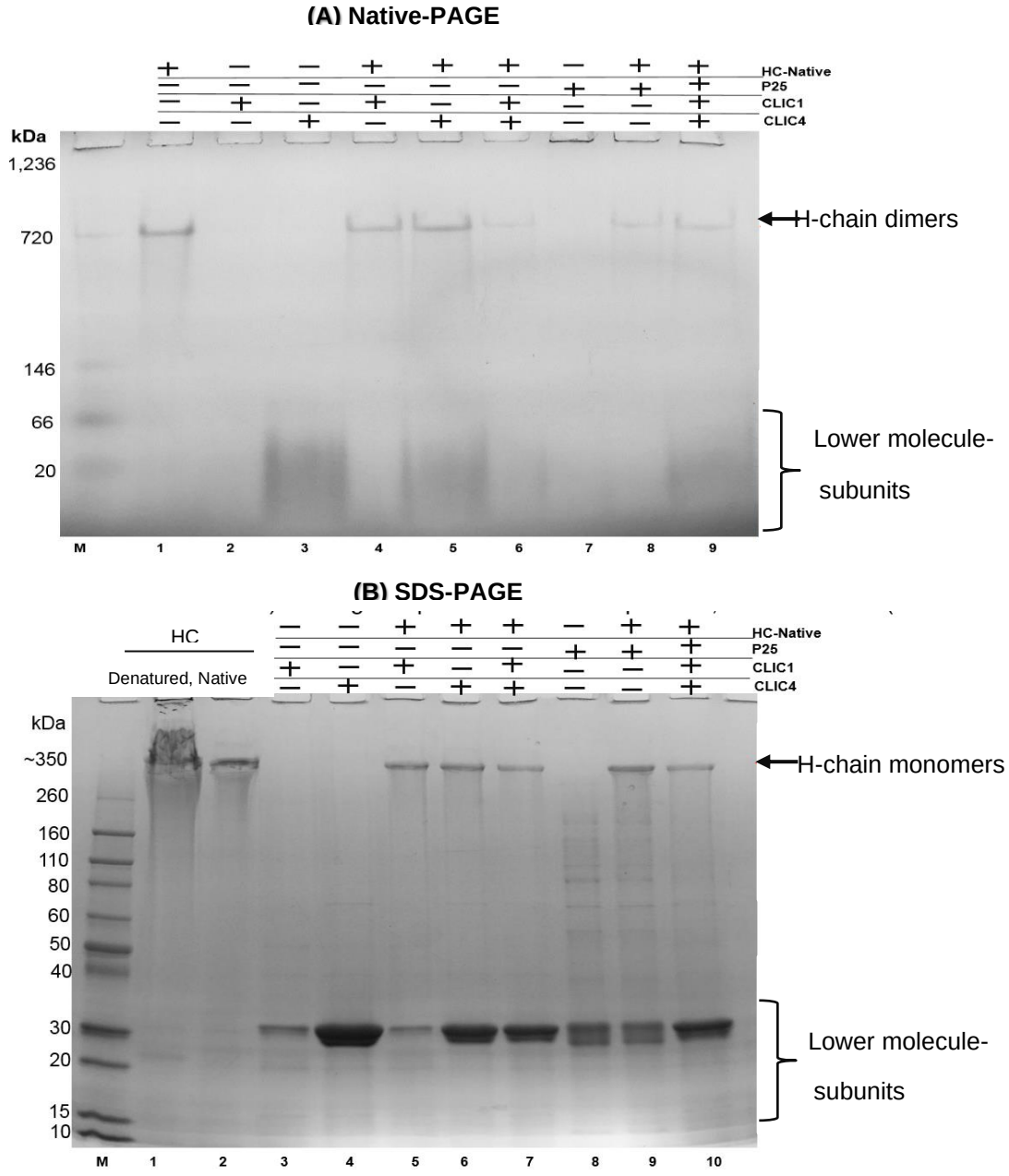


Figure 79. The purified H-chain fibroin in buffer (150mM NaCl, 50mM Tris-HCl, pH 8.0) was isolated from the posterior silk gland. Multimer reactions between H-chain and human proteins, CLIC1 and CLIC4 were set in each treatment (incubated at 4^0C, for 1 hour). **Figure 79: A**, from lane 1-9, Heavy chain alone (HC), CLIC1 alone, CLIC4 alone, H-chain+CLIC1, H-chain+CLIC4, H-chain+CLIC1+CLIC4, P25 (cellular) alone, H-chain+P25 (cellular), H-chain+P25 (cellular)+CLIC1+CLIC4, respectively. Consequently, the samples from each reaction were treated with the sample buffer without any reducing agent and determined by Native PAGE (**Figure 79 A)** and all reactions were collected and mixed with the sample buffer, boiled at 85^0C with SDS-PAGE, Novex, 4-20% (**Figure 79 B).**

RESULTS (Continued)

2.2. Determination in vitro native silk fibroin complexes formation or protein-protein interactions using (Size Exclusion Chromatography and Laser Light Scattering Service, SEC-MALS)

Six solutions of native fibroin complexes (H-chain alone, H-chain+P25, H-chain+FIBL, FIBL&P25, and 3 subunits of fibroins) were incubated on ice until the macromolecular formation systems were measured with SEC-MALs using BSA 200µL (1-2mg/mL in the SEC buffer) as the control. The fibroin protein interactions and their behavior were tested and molecule mass of these complexes was measured by the SEC-MALS system using a light scattering and refractive index (RI) detector. The data and findings are shown in **Table** 4 & **(Figure 80-84).**

The BSA (standard protein) was first loaded and the mass measured **(Figure 80).** The peaks starting from the retention volume of 4-6 mL, show the aggregate of the BSA, the retention of these volumes occurring when the proteins were initially run. The two retention volumes of 7.0 and 7.5mL were trimer and dimer, whereas, the retention volume of 8.19 mL indicated the monomer of BSA at 68,893 Da (68.90 kDa) **(Figure 80).**

Each fibroin complex, after incubation, including H-chain alone, H-chain+P25, H-chain+FIBL, FIBL & P25, and 3 subunits of fibroin was individually loaded into the SEC-MALS system as described earlier. **Figure 79**, the molecular mass of the native H-chain alone was determined first, and it was found that native H-chain behaviour alone formed a massive aggregation of the complex. However, it could also be the massive complex of a macromolecule with a retention volume of 6.14-7ml (2.97 $x10^6$ Da).

Next, the retention volume of a mass of a mixed native solution of HC+P25 was revealed by the SEC-MALS system to have a retention volume of 6.38-7.7mL at about (7.35 $x10^9$Da). A lot of peaks indicated that it was not a good monomer formation (**Figure 82**). Following this, a mixed native molecule of HC+FIBL was found to have a huge complex with a retention volume of 6.28mL at approximately 1.35 $x10^9$Da or 1.35 million kDa **(Figure 83**). Then, a mixed native subunit of FIBL+ P25 showed interesting results. Two molecular weights presented together, the peak (1) with a retention volume of 6.07mL had a big molecule of about 1.00$x10^7$Da (1,000 kDa) whereas at the peak (2) it was 11.08mL, comprising a small molecule of 32kDa. At WtFr (Peak) value, the protein existing in peak 1 was (5%) while in peak 2 it was (95%). This phenomenon might be due to an unfolded 32kDa protein or it could be a long peptide which needs to be studied further **(Figure 84).** However, a mixed native three molecule fibroin had a very high aggregation which could be passed through the filter of the SEC-MALS system but the actual mass of the massive complex of the native form, as found in the posterior silk gland of a native organism, could not be measured.

Table 4. The illustration of *In vitro* complexes as solution among three different fibroin molecules

No. of Complex	Fibroin subunit	Number of Subunits	Concentration (mg/mL)
No.1	Heavy chain alone	1	0.465
No.2	Heavy chain+P25	2	0.200
N0.3	Heavy chain+FIBL	2	0.272
No.4	FIBL+P25	2	0,272
No.5	Heavy chain+FIBL+P25	3	0,212

2.2.1) Standard protein (BSA)

Acq. Date	Aug 26, 2021 - 11:17:44
ID	BSA_1
File name	2021-08-26_11;17;44_BSA_1_01.vdt
dA/dc	0.0000
Calculated	0.6395
Recovery	94.6077
dn/dc	0.1850
Method	BSA 1-0011.vcm

Peak	1
Ret Vol (mL)	8.191
Mw (Da)	68,893
Mw/Mn	1.072
Wt Fr (Peak)	1.0000
Rg(w) (nm)	0.00
RI Area (mVmL)	16.98
UV Area (mVmL)	0.00

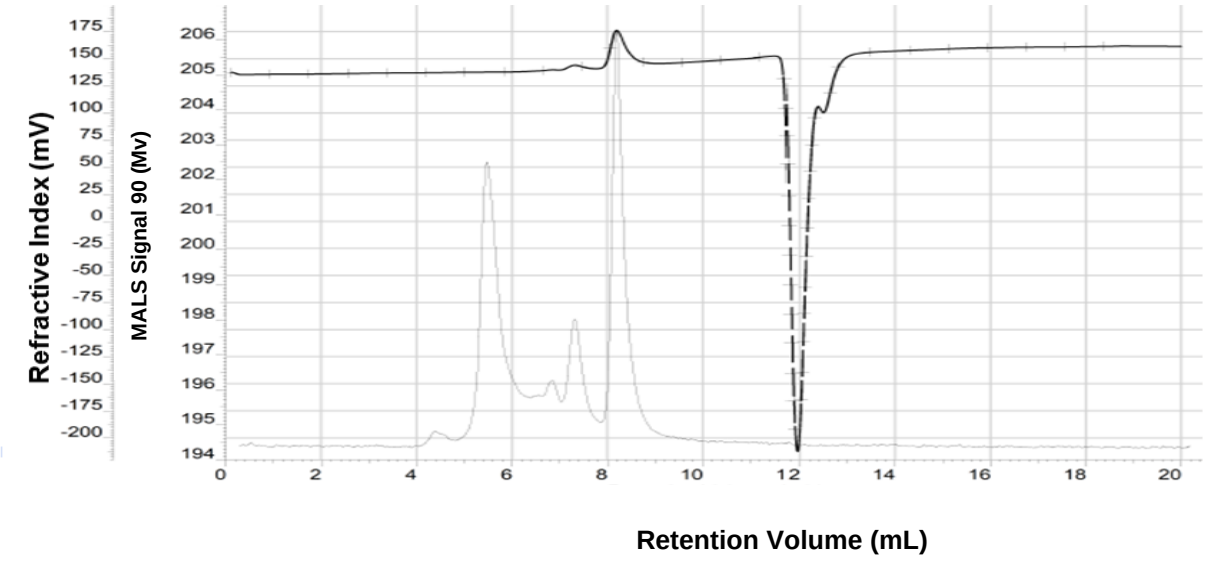


Figure 80. The chromatogram of BSA (Blank) was measured using SEC-MALS. Solid black is the MALS channel and the dotted line is the RI channel.

2.2.1) Determination of native H-chain solution alone using SEC-MALS

Acq. Date	Aug 26, 2021 - 11:44:53
ID	Heavy Chain
File name	2021-08-26_11;44;53_Heavy_Chain_01.vdt
dA/dc	0.0000
Calculated	2.2590
Recovery	485.8004
dn/dc	0.1850
Method	BSA 1-0003.vcm

Peak	1	2
Ret Vol	6.147	12.042
Mw	2.961 e 6	491
Mw/Mn	2.765	-1097.762
Rg(w)	66.36	0.00
Wt Fr (Peak)	0.0470	0.9530
RI Area	3.03	61.52
UV Area	0.00	0.00
MALS Area (90°)	6.91	0.07

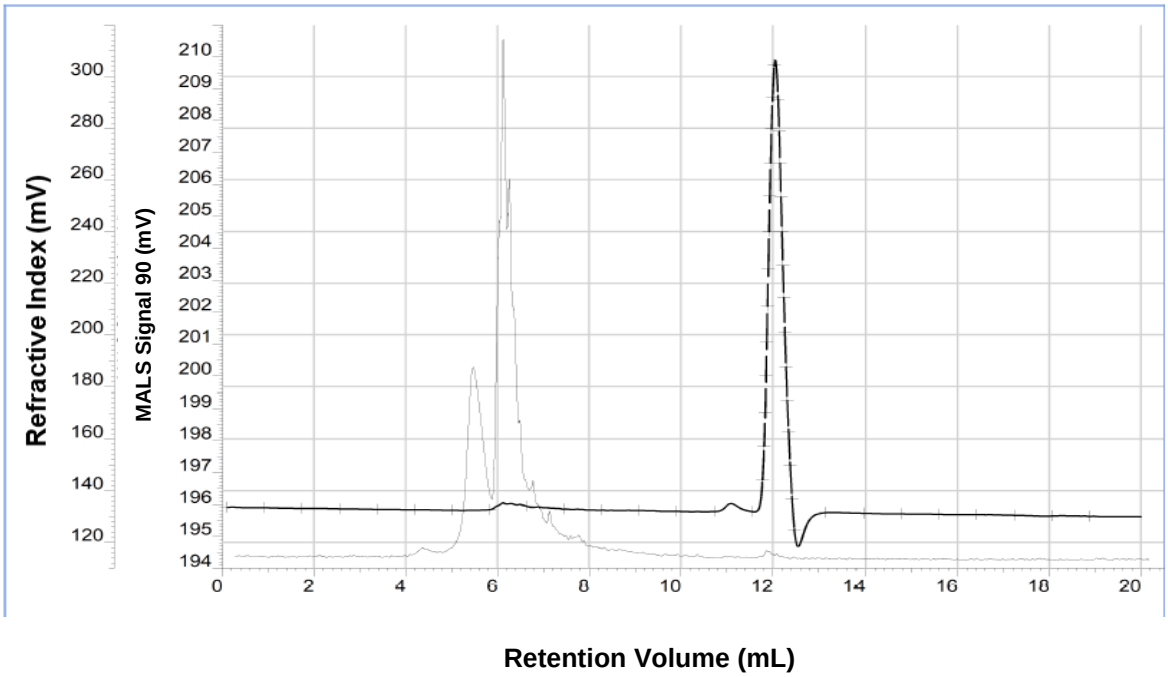


Figure 81. The chromatogram of Native H-chain alone (as control) in which the protein behavior and molecular weight was measured using SEC-MALS. Solid black is the MALS channel and the dotted line is the RI channel.

2.2.2) Determination of native H-chain+native P25 interactions using SEC-MALS

Acq. Date	Aug 26, 2021 - 12:12:11
ID	Heavychain + P25
File name	2021-08-26_12;12;11_Heavychain_+_P2
dA/dc	0.0000
Calculated	0.0972
Recovery	48.1236
dn/dc	0.1850
Method	BSA 1-0003.vcm

Peak	1	2
Ret Vol (mL)	6.383	11.107
Mw (Da)	7.349 e 9	1
Mw/Mn	4054.478	1.000
Wt Fr (Peak)	0.4269	0.5731
Rg(w) (nm)	64.12	1.00
RI Area (mVmL)	1.20	1.61
UV Area (mVmL)	0.00	0.00

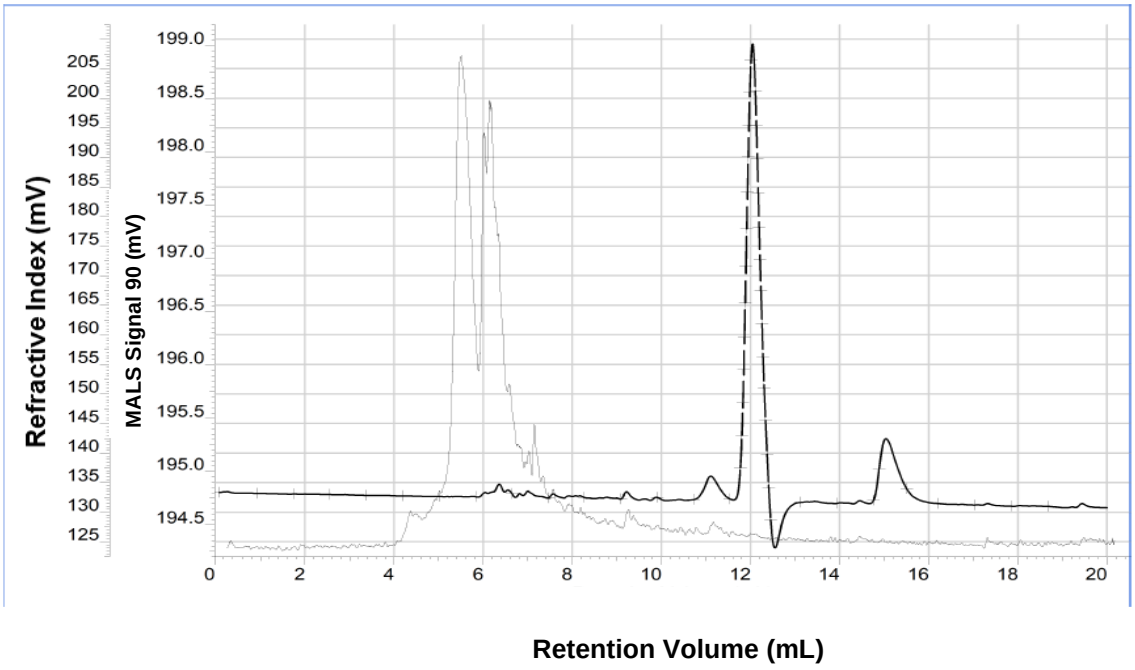


Figure 82. The chromatogram of a mixture of native H-chain+P25 where the interactions and molecular weight were measured using SEC-MALS. Solid black is the MALS channel and the dotted line is the RI channel.

2.2.3) Determination of native H-chain+native FIBL interactions using SEC-MALS

Acq. Date	Aug 26, 2021 - 12:35:20
ID	Heavy chain +FIBL
File name	2021-08-26_12;35;20_Heavy_chain_+FIBL_0
dA/dc	0.0000
Calculated	0.1013
Recovery	37.2332
dn/dc	0.1850
Method	BSA 1-0003.vcm

Peak	1	2
Ret Vol (mL)	6.280	11.077
Mw (Da)	1.352 e 9	1
Mw/Mn	336.031	1.000
Wt Fr (Peak)	0.2623	0.7377
Rg(w) (nm)	62.87	1.00
RI Area (mVmL)	0.77	2.16
UV Area (mVmL)	0.00	0.00

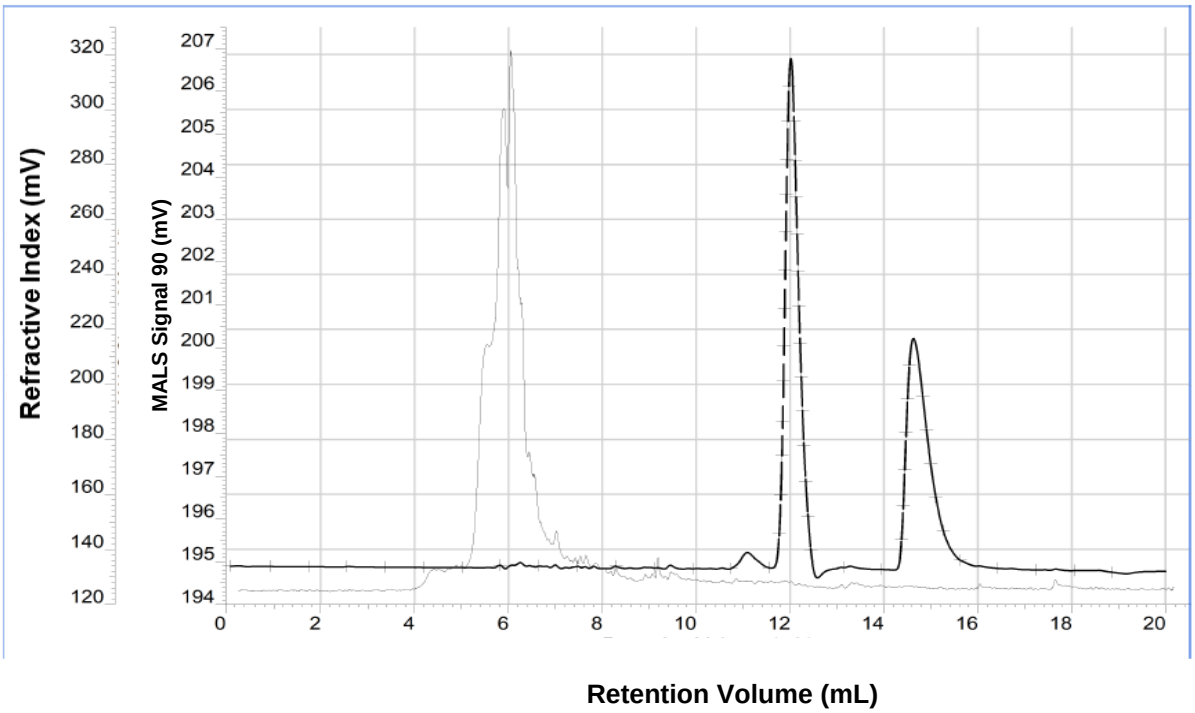


Figure 83. The chromatogram of a mixture of native H-chain+FIBL where the interactions and molecular weight were measured using SEC-MALS. Solid black is the MALS channel and the dotted line is the RI channel.

2.2.4) Determination of native FIBL+ native P25 interactions using SEC-MALS

Acq. Date	Aug 26, 2021 - 13:38:45
ID	FIBL + P25
File name	2021-08-26_13;38;45_FIBL_+_P25_01.vdt
dA/dc	0.0000
Calculated	0.0896
Recovery	32.9306
dn/dc	0.1850
Method	BSA 1-0003.vcm

Peak	1	2
Ret Vol (mL)	6.068	11.080
Mw (Da)	1.003 e 7	32,988
Mw/Mn	1.023	1.354
Wt Fr (Peak)	0.0511	0.9489
Rg(w) (nm)	27.27	142.77
RI Area (mVmL)	0.13	2.46
UV Area (mVmL)	0.00	0.00

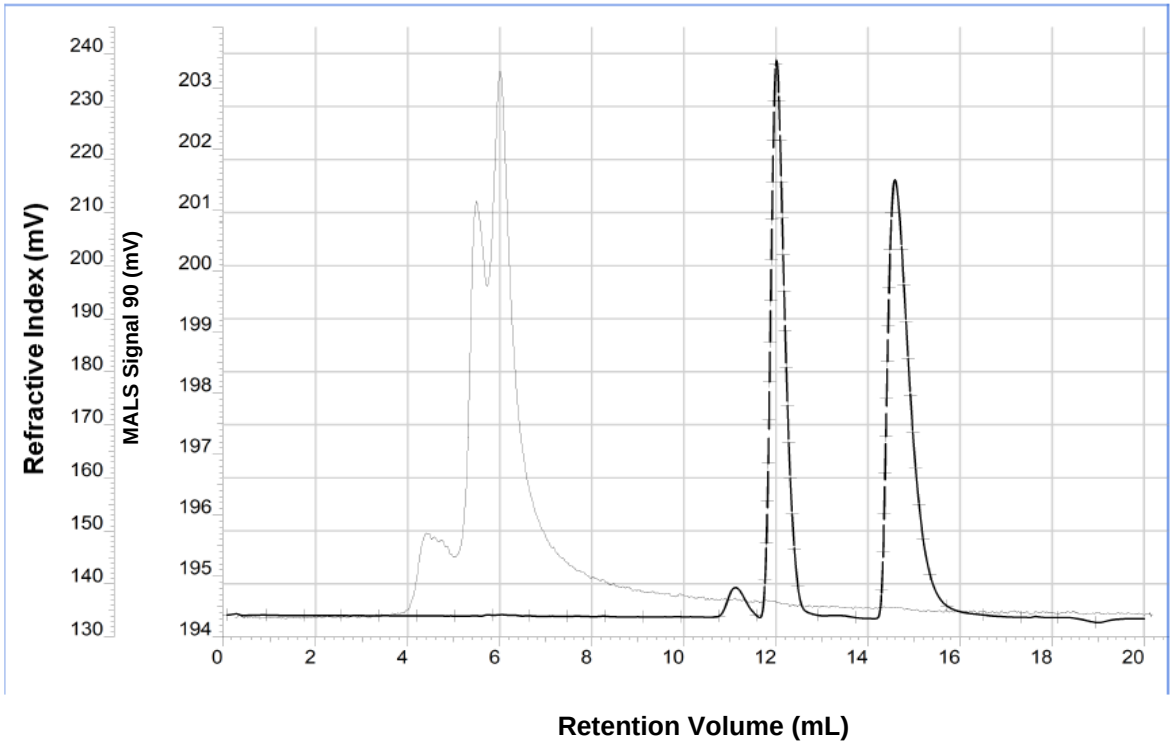


Figure 84. The chromatogram of a mixture of native FIBL+P25 in which the interactions and molecular weight were measured using SEC-MALS. Solid black is the MALS channel and the dotted line is the RI channel.

2.3) Determination in vitro silk fibroin complexes formation or protein-protein interactions using (Size Exclusion Chromatography, SEC)

A selective approach to the study of the interaction of three fibroin molecules was mainly undertaken separately on two protein states: all three molecules as natives and all three molecules as partial folds. These two forms of complexes were analysed using the SEC. The findings firstly are shown from the native state interaction as native-native fibroin complexes formation (fully folded form) in **Figures 85A-85B** and followed by partially folded form **(Figure 86**).

(2.3-A) native-native fibroin complexes formation /(fully folded form)

This experiment investigated the behaviour of the three fibroin molecules, H-chain, FIBL and P25 when these three native molecules were mixed into the same tube, incubated and loaded into the Superdex 75 10/300 GL column. The eluted fractions corresponding to relevant peaks were collected and determined using a reducing SDS-PAGE gel (Novex, 4-20%) as shown in **Figure 85 A&85B**.

(85 A)

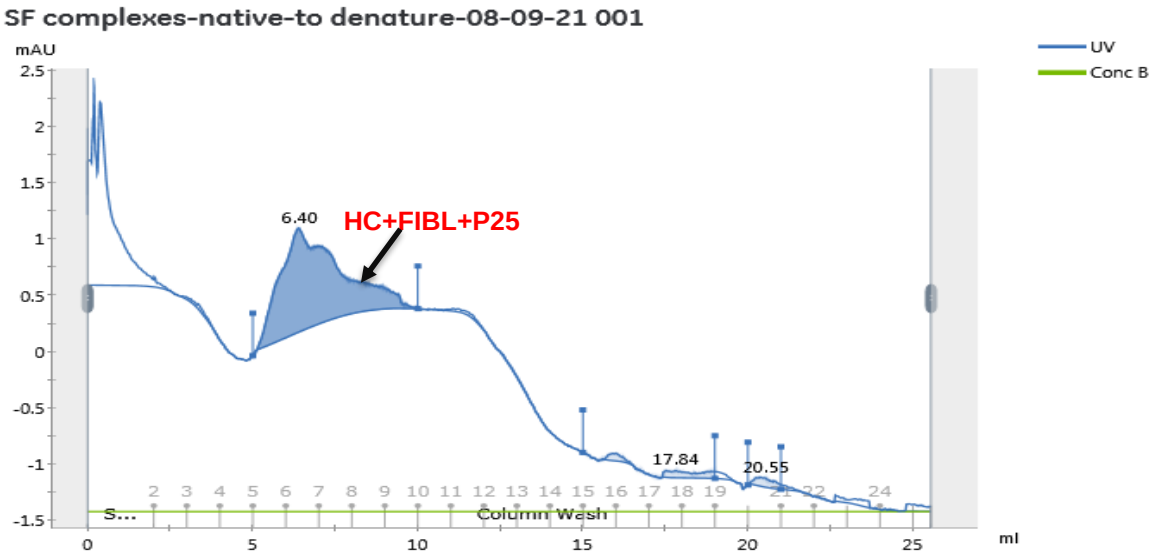


Figure 85: The chromatogram of 3 mixed subunits initially comprising only a H-chain isolated from the PSG using SEC. The fully fold purified H-chain was collected and eventually isolated with anion ion exchange chromatography as a native form. Secondly, the fibroin complex was prepared and mixed with three native molecules, H-chain+ FIBL+P25 and later this complex was changed into the partially folded (5M urea, 20mM Tris-HCl, pH 8.0) and eventually 1mL of the concentrated three subunits complex was loaded into the Superdex 75 10/300 GL column **(Figure 85A).**

(85 B)

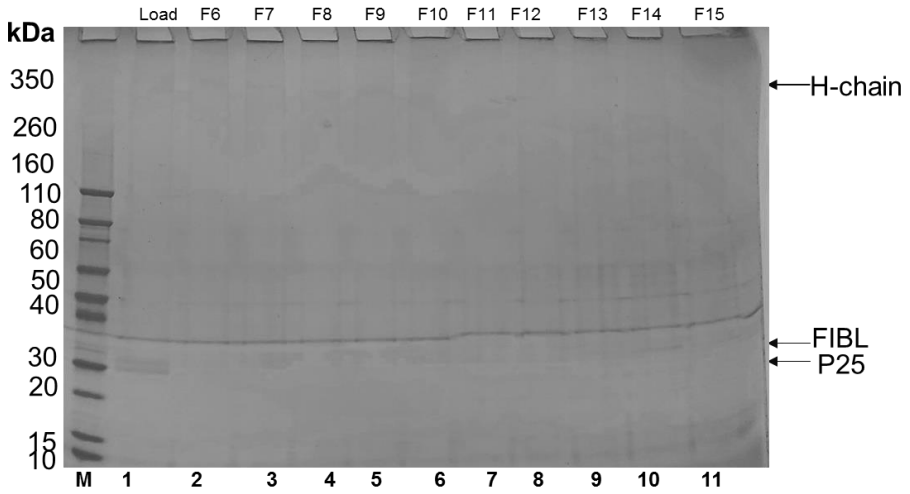


Figure 85 (continued). The elution fractions corresponding to the relevant peaks (ranging F6-F15 from **Figure 85A),** were collected and run onto a reducing SDS-PAGE gel, Novex, 4-20% **(Figure 85B).**

(2.3-B) partial native-partial native fibroin complexes /partially folded form).

The quality of two forms (partially folded & fully folded) of three single molecules of fibroin proteins (H-chain alone, P25 alone, FIBL alone, and mixed three subunits) were firstly determined and bands appeared on the SDS-PAGE gel (Novex 4-20%) before loading only partial native protein complexes, NT into the size exclusion chromatography (Superdex 75 increase 10/300 GL) as shown in **Figure 86.**

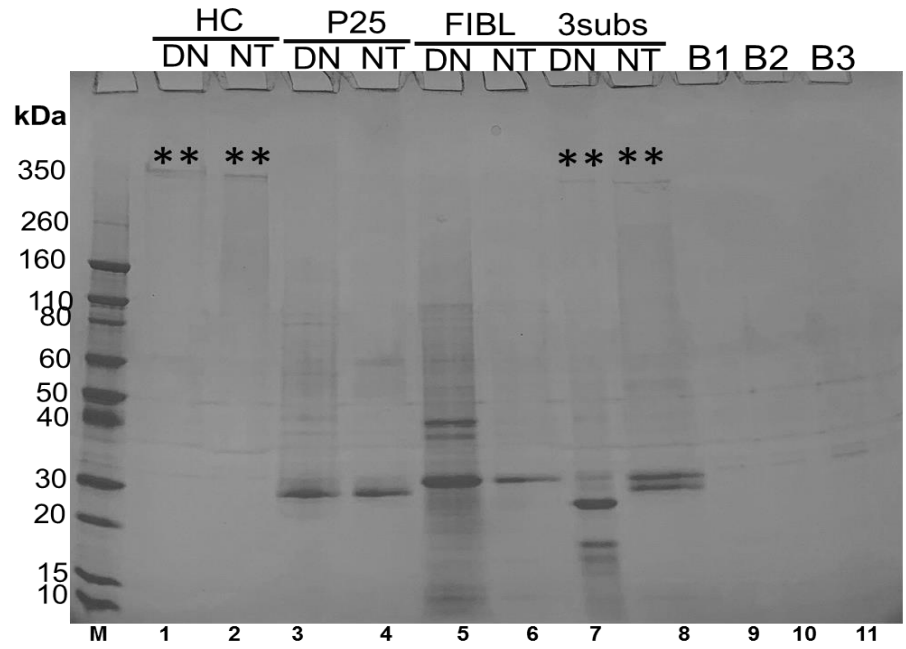


Figure 86. Comparison of two versions of three single molecules of fibroin proteins between denatured form as partial native form and native form (NT) including, H-chain (HC, Lane 1&2), P25 (Lane 3&4), Fibroin light chain (FIBL, Lane 5&6), three mixed molecules of fibroin proteins together (Lane 7&8), and B1-B3 (Control (no sample), blank) were run onto a reducing SDS-PAGE gel (Novex, 4-20%). Two forms of fibroin proteins, DN samples were prepared in 5M urea, 20mM Tris-HCl, (pH 8.0) and NT samples were prepared in 20mM Tris-HCl, (pH 8.0).

Next, six partial native fibroin protein complexes, H-chain alone, H-chain+P25, H-chain+FIBL, FIBL+P25, and three subunits were prepared (1mL each sample) in 5M urea, 20mM Tris-HCl, pH8.0. Individual loads of each single complex were inserted into the column (Superdex 75 increase 10/300 GL). The eluted fractions of each complex corresponding to the peaks were collected so that the protein concentrations could be measured as shown in **Figures 87A-87D**. A partial native H-chain alone was first run and eluted from the points 5-10mL.

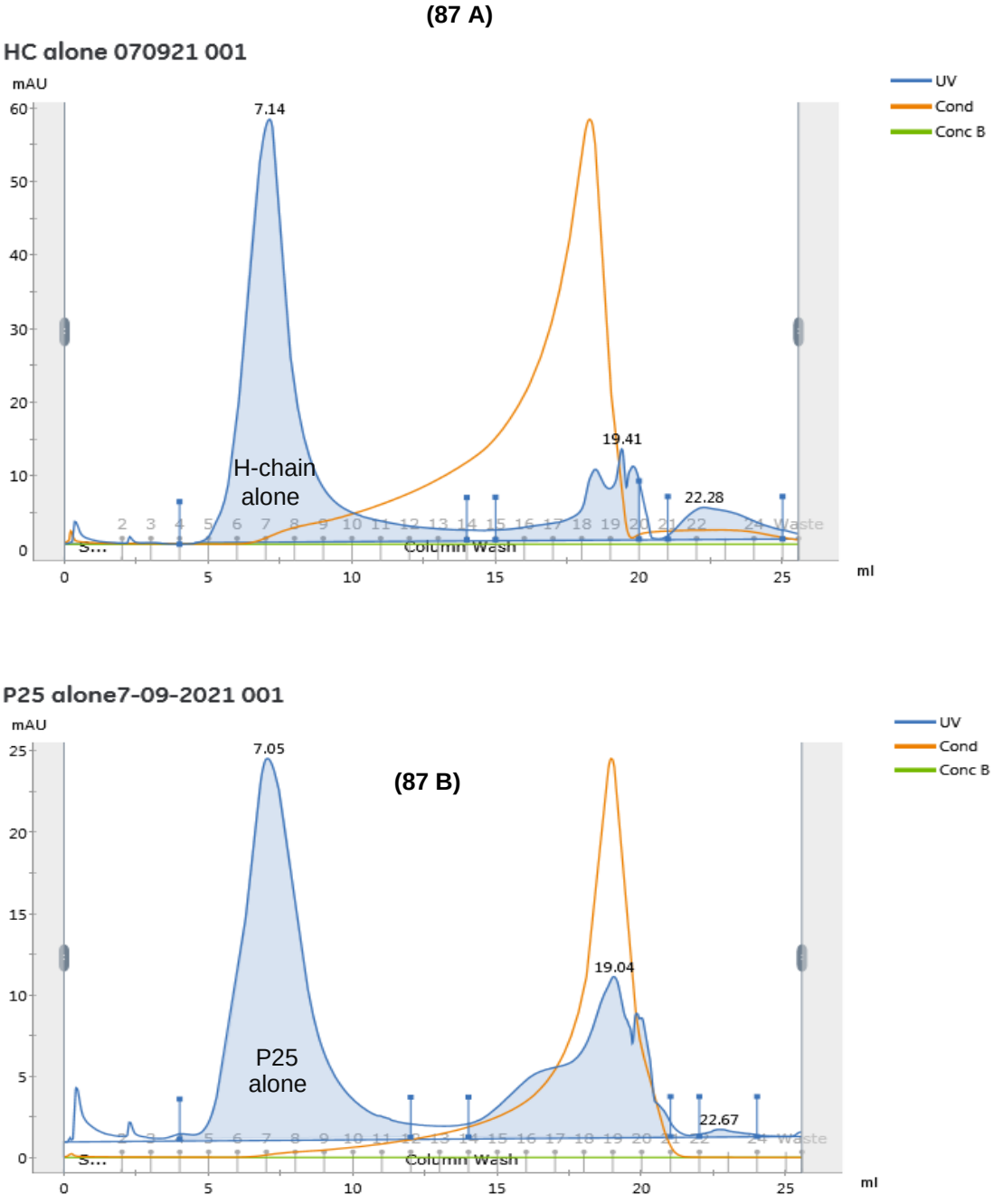


Figure 87. The chromatogram of 1mL of partial native states (5M urea, 20mM Tris-HCl, pH 8.0) of fibroin molecules, H-chain alone and P25 alone were examined using SEC through the column (Superdex 75 increase 10/300 GL). The eluted fractions indicate the partial native H-chain **(Figure 87A)** and partial native P25 alone **(Figure 87 B)** were presented ranging from fractions 5-10mL.

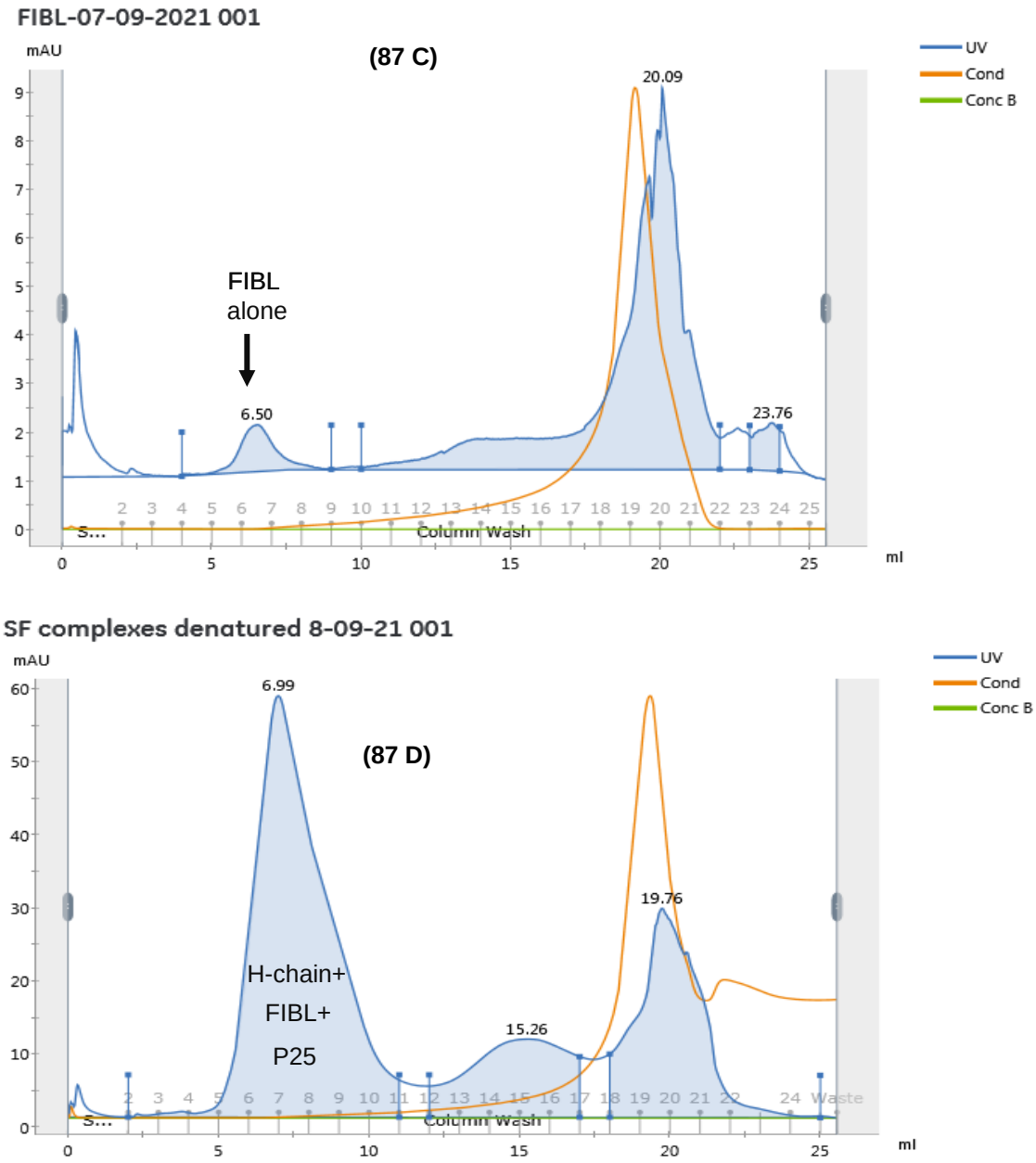


Figure 87. The chromatogram of 1mL of partial native states (5M urea, 20mM Tris-HCl, pH 8.0) of two forms of fibroin complexes, FIBL alone and 3 molecules mixed together (H-chain+FIBL+P25) were examined using SEC through the column (Superdex 75 increase 10/300 GL). The eluted fractions indicating the partial native FIBL alone **(Figure 87C)** and partial native, H-chain+FIBL+P25 (**Figure 87D)** were presented ranging from fractions 5-10mL.

2.4) Determination in vitro silk fibroin complexes formation or protein-protein interactions using pull-down assay

(A) native-native fibroin complexes formation /(fully folded form)

It has been shown from the experimental findings in this chapter, that the formation, presentation and interaction of its three fibroin molecules (H-chain, P25, and FIBL) are different. The native H-chain is the main molecule that can form the complex with other proteins, CLIC1 and CLIC4. Therefore, to better understand fibroin formation and the interactions of these three molecules, another

approach was needed. Hence, the pull-down method was used as the last assay in this chapter to look at the fibroin molecules' interaction.

Ideally, both the lower molecules FIBL and P25 included his6-tagged, but only the H-chain from the native organism was not tagged. So, if the H-chain interacted with both of the lower two molecules, it would be bound to be a complex after passing through the Ni-NTA beads. To prove this, seven native fibroin complexes were provided, 1) H-chain alone, 2) P25 alone, 3) FIBL alone, 4) H-chain+P25, 5) H-chain+FIBL, 6) P25+FIBL and 7) three subunits. The pull-down method was described previously in the materials and methods section. All samples, including four fibroin molecules for non-passing through the Ni-NTA column (denatured H-chain, native H-chain alone, native P25 alone, and native FIBL alone) and seven native eluted complexes (H-chain alone, P25 alone, FIBL alone, H-chain+P25, HC+FIBL, P25+FIBL, and three subunits were shown in **(Figure 88A-88B)** and the bands of fibroin molecules with and without pull-down passing through the Ni-NTA column were compared.

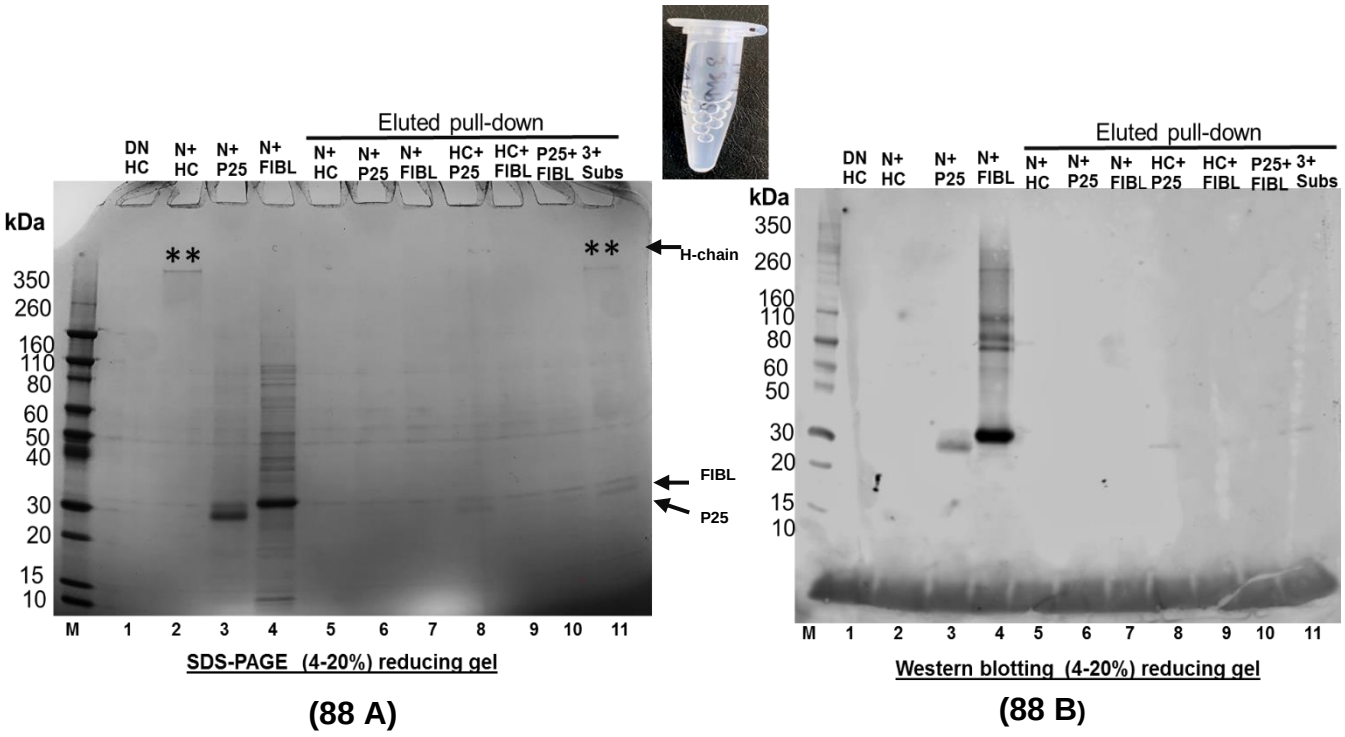


Figure 88. The chromatogram of fibroin molecules which were determined with and without pull-down and passed through the Ni-NTA column. Then, replicate samples were run on two 14-20% SDS-PAGE gels with the sample buffer containing 20mMDTT, one gel stained with the Coomassie blue **(Figure 88A)** and another gel was subjected to western-blotting with the anti-his6-tagged **(Figure 88B**). Two parts of samples were M (marker), un-pull-down Lane1-4 (1) Denatured H-chain:DN-HC, 2) Native H-chain:N+HC, 3) Native P25:N+P25, and 4) Native FIBL (N+FIBL) and eluted pull-down samples were Lane 5-11, 5) Native H-chain:(N+HC), 6) Native P25: (N+P25), 7) Native FIBL: (N+FIBL), 8) H-chain+P25: (HC+P25), 9) H-chain+ FIBL: (HC:FIBL), 10) P25+FIBL:(P25:FIBL) and 11) H-chain+P25+FIBL: (3+Subs), respectively.

CONCLUSION AND DISCUSSION

Prior studies have clarified that SF (*B. mori*) is secreted from the PSG as a huge molecular complex, about 2.3 MDa, composed of six heterodimeric disulphide-linked H-chain and FIBL proteins and a single P25 (Inoue et al., 2004). The protein interactions can be described via a combination of Vander Walls forces, hydrogen bonds, and hydrophobic and steric effects, which are the electrolytes' predominant source of interaction (Walker-Taylor & Jones, 2005). Therefore, the presence of fibroin proteins in native organisms as natural molecules in blends may improve cell adhesion due to the increase of protein binding sites (Silva et al., 2005).

Recent studies have shown that the superior mechanical properties of silk fibroin are obviously caused by the nanostructure of β-sheet crystals which are formed naturally by the self-assembly mechanisms, *in vivo* (Keten et al., 2010: Lu et al.,2012). Experimentally, an H-chain of fibroin reveals very distinctive roles in mechanical properties (Yang et.al, 2020). Naturally in PSG, H-chain and FIBL are covalently attached together by a single disulfide bond from two C-terminus regions of each molecule between (Cys-c20) of H-chain and Cys-172 of FIBL as herterodimeric interaction. Nonetheless, each chain end possesses many Cys residues which may have a chance to connect within the same chain (intra-molecular disulfide bond (s) between the C-terminus and N-terminus regions. Between H-chain there may be the possibility of forming intermolecular disulfide bond (s) (Yamaguchi et al., 1989: Ha et al., 2005).

The H-chain is composed of four different parts including, N-terminus regions, C-teminus region, 11 spacer regions, and 12 large repeat domains. The N and C terminal and 11 spacer regions contribute toward the hydrophilicity pattern of fibroin, whereas the 12 large bulk domains consisting of the repeating hexapeptides GAGAGS and GAGAGY sequences give the protein its hydrophobic nature (Dubey et al., 2015) as shown in **Figure 89.** In addition, Ji et al (2009) reported that the hydrophilic spacer 'GTGSSGFGPYVAN (H)GGYSG YEY' in the H-chain of silk fibroin is likely to be involved in the binding of ferric ions, and His, Asn and Tyr residues are considered to be the potential binding sites. It is commonly stated in many previous publications that H-chain has a size of 350-390 kDa in SDS-PAGE. It is in fact a dimer in its native form which would have a high mass ≈900 kDa. However, the monomeric form could occur if H-chain is subjected to high temperatures or treated with denaturing agents. This has been confirmed in several repetitions of experiments which have been shown in the result sections.

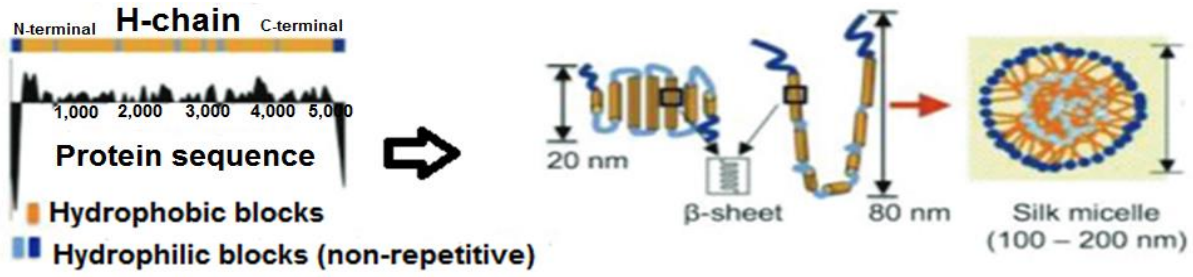


Figure 89. Schematic model of silk fibroin folding (*B.mori*) demonstrates the hydrophobicity pattern of H-chain with its possible formations and micelle assembly of silk fibroin in water (Adapted from Jin & Kaplan, 2003: Holland et al., 2003).

1) H-chain isolation and purification from solid fibroin

The intact fibroin including three molecules was able to be prepared from the solid fibroin modified from Katagata et al., 1984 before isolating only H-chain. Two other lower molecules were removed using two combinations, 1) size exclusion chromatography (eluted purified protein as partially folded (5M urea, 20mM Tris-HCl, pH 8.0) which can be stored long term in the refrigerator at about 4^0C until use, and 2) anion ion exchange chromatography was used for the final purification using salt gradient technique to elute purified native H-chain (100mM NaCl, 20mM Tris-HCl, pH 8.0). It was found that the native H-chain could not be preserved for long inthe refrigerator because its behaviour as a single molecule was not stable. Furthermore, either the H-chain alone could not be detected or no bands appeared on the native gel orSDS-PAGE after being kept overnight in the refrigerator (data not shown). This result would indicate that H-chain always requires to be formed as a dimer form.

2) Observation of the initiation of the silk fibroin complexes formation or multimerisation behaviour along its three indigenous proteins (H-chain, FIBL &P25) and interaction with selected human proteins (CLIC1 & CLIC4).

It is interesting to note from our findings in this project that H-chain alone was identified using Native PAGE (Invitrogen™, Novex™, 4-20%, USA) at about 920 kDa as always forming dimers. This was previously seen by 3-16% Native PAGE, but unidentified in Song et al., 2021. However, the SDS-PAGE images showed monomeric form of H-chain ~350kDa as found in many previous studies. In regard to dimers formation of H-chain, it seems the molecular weight of the Native H-chain is different from that of the denatured form though no other reports have been made about this. In addition, when the fibroin complexes containing native H-chain such as H-chain alone, H-chain+P25, H-chain+FIBL, presented as 3 mixed subunits (H-chain+P25+FIBL) were incubated at RT temperature and on ice and then run on SDS-PAGE (±) reducing agents (DTT, BME), the outcome was that all complexes containing H-chain disappeared when run on both reduced gels, but not in a non-reduced gel.

Nevertheless, the observation of fibroin complexes with other proteins (selective recombinant human proteins, CLIC1& CLIC4), The Native-PAGE and SDS-PAGE revealed that the H-chain could form two different dimers: chiefly the H-chain would be able to form homodimers when only the H-chain existed in the reaction, whereas the heterodimers could be linked when the reaction contained H-chain together with CLIC1/CLIC4, or possibly with CLIC1+ CLIC4. The H-chain dimers were detected at approximately 900 kDa (Native-PAGE) and the lower molecular weight molecules, CLIC1 or CLIC4 at around 30 kDa (Native PAGE & SDS-PAGE). The findings indicate that H-chain has more abundant hydrophobic amino acid residues containing a massive molecular weight (~ 900kDa) and has the high potential of exchanging copolymer block with amorphous parts. This structural arrangement of hydrophobic and copolymer blocks in (H-chain, *B. mori*) provides a unique and superior compatibility for self-assembly with other suitable materials for different aspects of innovative uses (Altman et al., 2003; Rubezic et al., 2020).

Following the experiments of fibroin molecules formation within its components and of H-chain ability to form complexes with two selective human proteins (CLIC1 & CLIC4), each fibroin protein and its complexes in solutions were determined using Native-PAGE and SDS-PAGE to monitor its behaviour. It could be clearly seen that the native H-chain, or the Native H-chain complexes formed with other molecules after 24 hours in the refrigerator, could not be detected. This might be because the native form of H-chain alone or the native fibroin complexes including native H-chain deteriorated easily or were destroyed by a proteolytic process. That means Native H-chain fibroin molecules degraded faster than the other two, FIBL and P25. During observation of the complexes formation, we concluded that anti-protease may need to be added. Basically, a macromolecule of native H-chain requires the dimers to be formed in native organism (FIBL) or also formed *in vitro* system with possible Cys residue in other molecules or formed via interlinkage with its own in order to make it more stable for transporting and secreting. However, the two lower molecules FIBL and P25 can be secreted into the lumen in the monomeric form (Wang et al., 2015).

The behaviour of the fibroin complexes of the H-chain was also observed with a different number of components; H-chain alone, H-chain+P25, H-chain+FIBL or the three molecules mixed. When samples of all complexes containing the H-chain were taken and the sample buffer containing 20mM DTT or BME were added, no H-chain bands were detected on the SDS-PAGE. They were, however, detected in complexes containing FIBL or P25. There are two possible reasons for this: 1) the H-chain may form a large molecule with massive aggregation with reducing agents such as BME or DTT. This aggregate might not enter the gradient gel 4-20% SDS-PAGE and does not appear on the resultant gel and 2) H-chain may undergo aggregation in the presence of reducing agents BME or DTT in the presence of FIBL or P25, or in the presence both of FIBL of P25. It is likely that Native H-chain is more susceptible to reducing agents than the other two molecules, FIBL of P25.

3) Determination in vitro native silk fibroin complexes formation or protein-protein interactions using (Size Exclusion Chromatography and Laser Light Scattering Service, SEC-MALS)

An investigation into the molecular weight of the native fibroin proteins and its indigenous molecules (protein-protein complexes) in solutions was conducted. This was performed using the combination of SEC for separating the protein's size with MALS to determine the number and type of the protein in solution (i.e, monomer, native oligomers or aggregate). This approach relates to the elution volumes of the analysis to estimate the molecular mass. Five native fibroin solutions including 1) H-chain alone, 2) H-chain+P25, 3) H-chain+FIBL, 4) FIBL+P25 and 5) three mixed subunits (H-chain+FIBL+P25) were individually run in the SEC-MALS system. The SEC-MALS results showed that the native heavy chain alone in the solution showed very high aggregation ~2.97 x10^6 Da, whereas two other fibroin complexes showed higher molecular weight than a single native H-chain such as H-chain+P25 (~7.35 x10^9 Da, not such great monomeric forms) and H-chain+FIBL (1.35 x10^9 Da, a huge aggregation peak). In previous studies, H-chain+P25 interacted together within a hydrophobic interaction. Tanaka et al. (1999) suggested that when P25 associates only with H-chain in the absence of H-chain+FIBL linkage, the oligosaccharide content of P25 decreased because there was no formation of the H-chain+FIBL disulfide bond.

H-chain+FIBL would generally link together with a single heterodimer as an organism in order to transport and secrete in the PSG lumen (Tanaka et al., 1999: Wang et al., 2015). When these two H-chain+FIBL molecules were mixed into the solution, massive macromolecules resulted as soon as they were added into the same tube. However, with SEC-MALS the result of two lower molecular weights P25+FIBL showed that both molecules were not likely to engage with each other which correlates to previous publications that found in nature P25 is not linked together with FIBL, but P25 generally interacts noncovalently to H-chain+FIBL (heterodimer) to form the complex (Tanaka et al., 1999). In addition to that, three mixed fibroin molecules (H-chain+P25+FIBL) in the solution could not pass through the SEC filter column (0.200um) twice due to high aggregation. Therefore, three subunits (multimer) of fibroin proteins in the solution were not successfully detected using SEC-MALS.

4) ***Determination in vitro silk fibroin complexes formation or protein-protein interactions using (Size Exclusion Chromatography, SEC)***

The value of native fibroin molecules molecular weights (M_w) was shown from the previous SEC-MALS determination. It was totally different to the value determined with SDS-PAGE method. Therefore, further investigation using only SEC to gain more insight into fibroin molecule interactions and behaviour was undertaken so both natural and partially natural states could be shown.

4-A) native-native fibroin complexes formation /(fully folded form)

When fibroin molecules, H-chain (native organism), P25 (*E. coli*), and FIBL (insect cell, Sf21) were individually prepared in different platforms, basically each molecule behaved differently after purification and protein preservation. It was found that both purified recombinant lower molecular weight

molecules (native), P25 and FIBL are more stable in monomeric form in the refrigerator (5 -7 days) which showed the same formation of both molecules as found in the PSG (Tanaka et al., 1999: Wang et al., 2015), but not in the native H-chain alone which has naturally fastest degradation even in the refrigerator. Similarly, Watanabe et al. (2004) revealed that in the time course of hydrolysis of liquid fibroin by purified fibroinase, H-chain was hydrolysed faster if compared to L-chain determined by SDS-PAGE of 10% gel. Also, many previous publications reported that H-chain was prepared and isolated from the PSG or fresh cocoons as partially folded (containing 5M urea) and determined by SDS-PAGE (35±25kDa) but there were no previous reports of the native molecular weight using Native-PAGE. To obtain SDS-PAGE results, H-chain samples must be boiled or denatured before observation in many fibroin protocols (Yamada et al., 2001: Inoue et al., 2000; Katagata et al., 1984: Shimura et al., 1976).

The chromatogram result of native-native fibroin complex containing three main components in the solution revealed that once all three fibroin molecules were mixed together in the same tube, a native state macromolecule formed as a huge fibroin complex which could not pass through the filter of SEC-MALS system. Therefore, to avoid that problem, in this experiment the native fibroin complex (three molecules) was changed to partially folded in (5M urea, 20mM Tris-HCl, pH 8.0) before being loaded into the SEC system. Both the chromatogram result and SDS-PAGE indicated that when the fibroin complex was treated with the transition from native state (20mM Tris-HCl, pH 8.0) to partial denatured state (5M urea, 20mM Tris-HCl, pH 8.0) it was almost completely degraded.

4-B) Partial native-partial native fibroin complexes /partially folded form).

Conformation transition of fibroin molecules with a wide range of urea concentration, is still limited. However, the pepsin molecules retain the natural conformation between 0-5M urea, partially folded at 6M, and strongly unfolded from 7-10M (Rho et al., 2019). Thus, the definition of partially folded fibroin subunits is designated at a urea concentration of 5M.

In the first attempt to observe the behaviour of three fibroin molecules compared with different mixes of its partially folded subunits (5M urea, 20mM Tris-HCl, pH 8.0), SEC was performed using (Superdex 75 increase 10/300 GL) column. The condition remained the same with the buffer from mixing the complex, incubation in the solution and then subjected to the SEC column. The findings indicated the interaction of the *in vitro* complex formation within three fibroin molecules showed the same peak after being run through the SEC chromatogram. However, it would be clearer if these samples could be run in another column (e.g. Superdex GL10/300) to compare with Superdex 75 increase 10/300 GL). Interestingly, the results confirmed that fibroin molecules could still be active even in partially folded form which is a great potential property for other uses as a bio-polymer.

5)Determination in vitro silk fibroin complexes formation or protein-protein interactions using pull-down assay

Some points related to the findings of fibroin molecules interaction can be explained as a result of all the above experiments. Nevertheless, the findings of pull-down assay provide more evidence to

show how the native fibroin forms after its native molecules were mixed in the solution as the massive complex. In particular, the native fibroin complex containing one, two, or three molecules, H-chain, FIBL, and P25 was experimented on through incubation, washed with the wash buffer, and eventually eluted through the Ni-NTA column (eluted buffer). The native eluted samples of the fibroin molecules (three subunits) from pull-down assay were subject to detection using SDS-PAGE analysis including H-chain~350kDa, FIBL~32kDa, and P25~30kDa), These intact components were in a different order range of molecular mass as *in vivo* of the native organism which was analysed with the same SDS-PAGE method, H-chain (~350kDa), P25 (~30kDa),and FIBL (26kDa) by Wöltje et al. (2021).

This pull-down finding indicated that even native H-chain (without the tag) was isolated from the PSG, and can interact with its indigenous recombinant molecules (with his6-tagged) e.g. two molecule (H-chain+P25) or (H-chain+FIBL), three molecules (H-chain, P25+FIBL) to form the complex during the incubation and the complex was bound to the Ni-NTA and eluted by centrifugation with elution buffer. These eluted fibroin complexes were confirmed by SDS-PAGE, but in western-blotting no band of H-chain was shown due its having no antibody detection. The diagram of pull-down assay of fibroin complexes is shown in **Figure 90.**

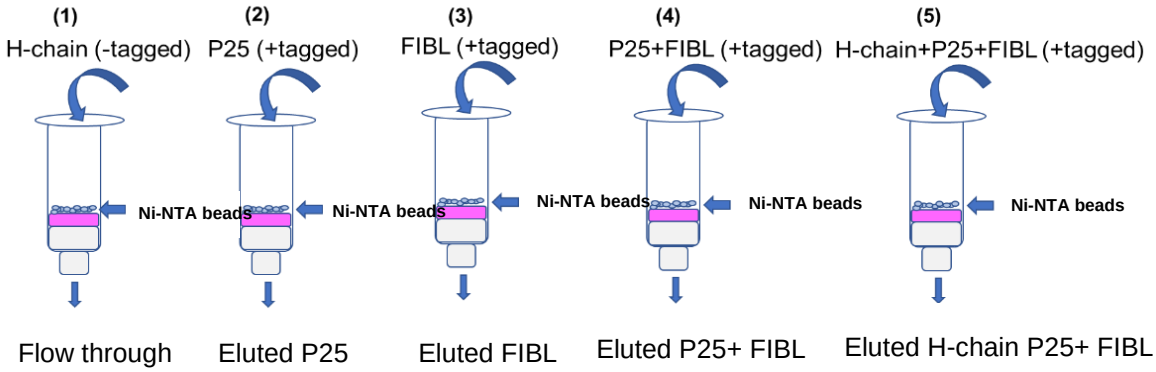


Figure 90. The Schematic diagram of fibroin molecules which formed into different complexes components and were determined using pull-down assay. The eluted solutions were further subjected to SDS-PAGE and western-blotting with the his6-tagged antibody.

FURTHER DIRECTIONS

To fully understanding the transition between the natured and denatured states of H-chain or other lower molecular weight fibroin molecules, it is necessary to look at the kinetics of the transition between any intermediate states. Further study is especially needed to determine the possible role of intermediate states in the denaturation of native H-chain (Neet & Timm, 1994). Native H-chain has an excellent quality for forming both homo and hetero dimers which is very useful for designing bio-polymers in future applications.

BIBLIOGRAPHY,

Altman, G.H., Diaz, F.,Jakuba,C.,Calabro, T., Horan, R.L., Chen,J., Lu, H., Richmond, J. & Kaplan, D.L. 2003. 'Silk-base biomaterials'. *Biomaterials* 24:401-416.

Couble, P., Chevillard, M., Moine, A., Ravel-Chapuis, P. & Prudhomme, J.C. 1985. 'Structural organization of the P25 gene of *Bombyx mori* and comparative analysis of its flanking DNA with that of fibroin gene'. *Nucleic Acids Research*. 13: 1801-1814.

Ha, S.W., Gracz, H.S., Tonelli, A.E., & Hudson, S.M. 2005. 'Structural study of irregular amino acid sequence in the heavy chain of *Bombyx mori* silk fibroin'. *Biomacromolecules* 6:2563-2569.

Holland, C., Numato, K., Rnjak-Kovaccina, J. & Sieb, F.P. 2019. 'The biomedical use of silk: Past, Present, Future'. Advanced Healthcare Materials, *Progress Report* 8: 1-26.

Inoue, S., Tanaka, K.,Arisaka, F., Kimura, S., Ohtomo, K. & Mizono, S. 2000. 'Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio'. *The Journal of Biological chemistry*. 275 (51):40517-40528.

Inoue, S., Tanaka, K., Tanaka, H., Ohtomo, K.,Kanda, T., Imamura, M., Quan, G.X., Tamura, T. & Mizuno, S. 2004. 'Assembling of the silk fibroin elementary unit in endoplasmic reticulum and role of L-chain for protection of alpha1,2-manose residue in N-linked oligosaccharide chains of fibrohexamerin/P25'. *European Journal of Biochemistry.* 271(2):356-366.

Ji, D., Deng, Y.B. & Zhou, P. 2009. 'Folding process of silk fibroin induced by ferric and ferrous ions'. *Journal of Molecules Structure*. 938(1-3):305-310.

Jin, H.J. & Kaplan, D.L. 2003. 'Mechanism of silk processing in insect and spiders'. *Nature* 1057-1061.

Katagata, Y., Kikuchi, A. & Shimura, K. 1984. 'Characterization of the crystalline-region peptides prepared from the posterior silk gland fibroin'. *Journal of Sericultural Science of Japan* 53(2): 165-174.

Keten, S., Xu, Z., Ihle, B., Buehler, MJ. 2010. 'Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk'. *Nature Materials*. 9(4): 359-367.

Lu,Q., Zhu, H.,Zhang, C.,Zhang, F., Zhang,B. & Kaplan, D.L. 2012. 'Silk self-assembly mechanisms and control-from thermodynamics to kinetics'. *Biomacromolecules*. 13(3):826-832.

Ma, Y.M., Luo, Q., Qu, Y., Tang, Y.Y., Zeng, W.H., Wang, H.M., Hu, J. & Xu, H. 2021. ' New insights into the proteins interacting with the promoters of silkworm fibroin genes'. *Scientific reports*, 11:15880. https://doi.org/10.1038/s41598-021-95400-0.

Neet, K.E. & Timm, D.E., 1994. 'Conformational stability of dimeric proteins: Quantitative studies by equilibrium denaturation'. *Protein Science* 3:2167-2174.

Rho, Y., Kim, J.H., Min, B. & Jin, K.S. 2019. 'Chemically denaturation structures of porcine pepsin using small angel x-ray scattering'. *Polymers*. 11(12):1-17.

Rubezic, M.Z., Krsttio,A.B., Stankovic,H.Z., Ljupkovic,R.B.,Randelovic, M.S., & Zarubica, A.R. 2020. 'Different type of biomaterials: structure and application: A short review'. *Advance Technology*. 9(1): 69-79.

Sasaki, S. & Nakagaki, I. 1980. 'Secretory mechanism of fibroin, a silk protein, in the posterior silk gland cells of *Bombyx mori*.' *Membrane Biochemistry* 3:37-47.

Sasaki, S., Nakajima, E., Fujii-Kuriyama, Y. & Tashiro, Y. 1981. 'Intracellular transport and secretion of fibroin in the posterior silk gland of the silkworm *Bombyx mori*'. J*ournal of Cell Science* 50:19-44.

Shimura, K., Kikuchi, A., Ohtomo, K., Katagata, Y.& Hyodo, A. 1976. 'Studies on silk fibroin in *Bombyx mori*, I fraction of fibroin prepared from the posterior silk gland'. *The Journal of Biochemistry* 80:693-702.

Silva, S.S, Santos, M.I., Coutinho, O.P., Mano, J.F. & Reis, R.L. 2005. 'Physical properties and biocompatibility of chitosan/soy blended membranes'. *Journal of Materials Science: Materials in Medicine.* 16: 575–579.

Song, K., Wang, Y., Dong, W., Li, Z.Z., He, H.W., Zhu, P. & Xia, P.Z. 20210. 'Silkworm spinning: the programmed self-assembly from natural silk fibroin to superfibre'. *BioRxiv Reprint*. https://doi.org/10.1101/2021.03.08.434386

Takai, F., Oyama, F., Kimura, K., Hyodo, A., Mizono, S. & Shimura, K. 1984. 'Reduced level of secretion of subunit combination for the fibroin synthesized by mutant silkworm, *Nd* (2)'. *The Journal of Cell Biology*. 99: 2005-2010.

Tanaka, K., Inoue, S. & Mizuno, S. 1999. 'Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by *Bombyx mori*'. *Insect Biochemistry and Molecular Biology*. 29: 269–276.

Tanaka, K & Mizuno, S. 2001. 'Homologues of fibroin L-chain and P25 of *Bombyx mori* are present in *Dendrolimus spectabilis* and *Papilio xuthus* but detectable in *Antheraea yamam*'. *Insect Biochemistry and Molecular Biology*. 31:665-667.

Toshima, Y. 2005. 'Unusual protein behaviour illustrated with silk fibroin'. *Biochimica Biophysica Acta (BBA)-Biomembranes*. 1713(1): 1-4.

Wang, H., Wang, L., Wang, Y.L., Tao, H., Yin, W., Sima, Y.H., Wang, Y.J., Xu, S.Q. 2015. 'High yield exogenous protein HPL production in the *Bombyx mori* silk gland provides novel insight into recombinant expression system'. *Scientific Reports*, 5:13839: 1-11. https://doi.org/ 10.1038/srep13839

Walker-Taylor, A. & Jones, D.T. 2005. 'Computational methods for predicting protein-protein interactions'. *In* Proteomics and Protein-Protein Interactions: Biology, Chemistry, Bioinformatics, and Drug Design. Springer: Boston, MA, USA, 89–114 pp.

Watanabe, M., A. Yura, M. Yamanaka, K. Kamei, S. Hara & Sumida, M. 2004. 'Purification and characterization of fibroinase, a cathepsin L-like cysteine proteinase, from the silk gland in the

fourth instar *Bombyx mori* Larva at the fourth molt period, stage D2'. *Journal of Insect Biotechnology and Sericology* 73: 61-70.

Wöltje, M., Kölbel, A., Aibibu, D. & Cherif, C. 2021.' A fast and reliable process to fabricate regenerated silk fibroin in solution from degummed silk in 4 hours'. *International Journal of Molecular Science*. 22 (19): 1-16.

Yamada, H., Nakao, H., Takasu, Y. & Tsubachi, K. 2001. 'Preparation of undegraded native molecular fibroin solution from silk cocoon'. *Material Science & Engineering: C*. 14(1-2): 41-46.

Yamaguchi, K., Y. Kikuchi, T. Takagi, A. Kikuchi, F. Oyama, K. Shimura & Mizuno, S. 1989. 'Primary structure of the silk fibroin light chain determined by cDNA sequencing and peptide analysis'. *Journal of Molecular Biology.* 210: 127–139.

Yang, K., Guan, J., Shao, Z., Ritchie, R.O. 2020. 'Mechanical properties and toughening mechanism of natural silkworm silks and their composites'. *Journal of the Mechanical Behaviour of Biomedical Materials* .110: 1-6.

CHAPTER VI

GENERAL CONCLUSION AND DISCUSSION

……

Mulberry silk is an insect protein fibre that can be commercially produced on a large scale rearing at a low cost (Mi et al., 2023). The natural native fibroin is composed of three fibroin proteins: H-chain, FIBL, and P25. Most studies have been based on experiments with the intact fibroin (*B. mori*) containing these three molecules to generate useful biomaterials. Only a few publications have reported on the isolation of H-chain and FIBL and the study of the characteristics of the material in its isolation. H-chain and FIBL are non-toxic and provide good adhesion to the cells (Zafar et al., 2015), although it is generally believed that the main properties of these three subunits result from the H-chain (Yonesi et al., 2021). This PhD research project took on the challenge of isolating the H-chain from the solid posterior silk gland (PSG) of the late 5^{th} instar larvae in its native state and forming fibroin complexes with two recombinant lower molecular weights of the fibroin proteins, FIBL and P25. In the native fibroin protein synthesis, the mRNA levels from these genes and the corresponding fibroin protein accumulation in the silk gland involved the silkworm larvae from the 1^{st} to the 5^{th} stages. It also depended on the role of transcription factors in the regulation of the silk genes and the growth of hormones such as Juvenile Hormone (JH), which regulates the silk gland and the development of the silkworm (Zhoa et al., 2015).

In this study, the biogenesis of the two recombinant fibroin proteins, FIBL and P25, was produced in heterologous hosts (FIBL produced in *E. coli* and P25 produced in Sf 21 cells). Although these molecular formations were produced in different systems, they were recognised as forming the fibroin complex discussed in Chapter V. The characterisation of the H-chain (350 kDa) using the amplification of the polymerase chain reaction is not easy. This is because it contains many repeated units with a long Glycine-Alanine-rich sequence. However, the advancement of genetic engineering like Cas9-based would help further studies of the silk core fibroin. The basic fibroin unit (H-chain-FIBL-P25 with a 6:6:1 molar ratio) is the elementary molecular complex for efficient secretion of silk fibroin from the posterior silk gland to other compartments in the silkworm body (Inoue et al., 2004). Previous studies of the mutation of the fibrohexamerin gene (P25) in *B. mori* revealed this can cause rough endoplasmic reticulum in the PSG and affect the morphology of the fibroin secretion in the PSG lumen. However, the cocoons obtained from the P25 mutation showed very little of this effect (Zabelina et al., 2021). With the FIBL mutation (Inoue et al., 2005) reported that FIBL is essential in forming the disulfide linkage between the H-chain, which is crucial for the intracellular transport and secretion of the fibroin complex. However, the disulfide linkage did not occur in the partial deletion of the FIBL gene. The mutant silkworm produced recombinant protein in a molecular ratio equal to the fibroin H-chain, this being about half of the total molecules of the PSG silk proteins, but the study did not show the effect on the quality of the cocoon.

This project did not determine the secretion of the fibroin complex after the formation of its indigenous three molecules. Therefore, the secretion of the fibroin still needs to be elucidated in further studies. During the experimental formation of fibroin proteins, bubbles could be seen occurring while adding each subunit to the microtube tube. This clearly indicated that the fibroin complexes were active, rather than if there were no bubbles occurring when each fibroin molecule was added to the tube. Chloride ion channels: CLIC1 and CLIC4 were the desired proteins for use in this project because they are a human plasma protein and expressed in intracellular organelles involved with the development of cancers (Peretti et al., 2015). The experimental findings in Chapter V found that both CLIC proteins interacted with the partial native H-chain. This could be an important point for further studies. The body of knowledge of individual fibroin proteins arising this project could also be investigated further in order to find ways of forming complexes using various other proteins to form new superior properties in biomaterials such as protein from spider silks or other wild silks. In addition, it would also be interesting to further investigate production of the three fibroin proteins synthesis in the same host or system at the same time more extensively.

In conclusion, this project found that in the bacterial system, the FIBL construct was more stable in pET28a, which was not the case for P25. Baculovirus-insect cell expressions showed potential as a host for producing a large volume of the active glycosylated P25 protein. The glycan numbers attached to the P25 protein were revealed clearly by SDS-PAGE, but these glycan types were not identified in this PhD project due to time constraints. Both forms of P25 protein, semi-glycosylated (non-secretion from the Sf21 cells) and glycosylated (secretion from the Sf21 cells), were active in the formation of the fibroin complexes. This project clearly gave us a greater understanding of the behaviour of fibroin proteins (H-chain, FIBL, and P25) and how to apply them properly for many uses.

BIBLIOGRAPHY,

Inoue,S.,Tanaka K., Tanaka, H., Ohtomo k., Kanda T., Imamura, M., Quan G.X., Kojima,K., Yamashita, T., Nakajima,T., Taira, H., Tamura, T., Mizuno, S. 2004. ‘Assembly of the silk fibroin elementary unit in endoplasmic reticulum and a role of L-chain for protection of alpha1,2-mannose residues in N-linked oligosaccharide chains of fibrohexamerin/P25’. *European Journal of Biochemistry*. 271, 356–366.

Mi, S., Zhou, Y., Ma, S., Zhou, X., Xu, S., Yang, Y., Sun, Y., Xia, Q., Zhu, H., Wang, S., Tain, L. & Meng, Q. 2023. ‘High strength and ultra-tough whole spider silk fibers spun from transgenic silkworms’. *Matter*. 6:3661-3683.

Peretti, M., Agelini, M., Savalli, N., Florio, T., Yuspa, S.H. & Mazzanti, M. 2015. ‘Chloride channels in cancer: focus on chloride intracellular channel 1 and 4 (CLIC1 and CLIC4) proteins in tumor

development and as novel therapeutic'. *Biochimica et Biophysica Acta (BBA)-Biomembranes*. Part B. 1848(10):2523-2531.

Yonesi M., Garcia-Nieto, M., Guinea, G.V., Panetsos, F., Perez-Rigueiro & Gonzalez-Nieto. 2021. 'Silk fibroin: An ancient material for repairing the injured nervous system'. *Phamaceutics*. 13:429.

Zabelina, V., Takasu, Y., Sehadova, H., Yonemura, N., Nakajima, K., Sezutsu, H., Sery, M., Zurovec, M., Sehnal, F., Tamura, T. 2021. 'Mutation in *Bombyx mori* fibrohexamerin (P25) gene cause recognization of rough endoplasmic reticulum in posterior silk gland cells and alters morphology of fibroin secretory globules in the silk gland lumen'. *Insect Biochemistry and Molecular Biology*. 135, 103607:1-10.

Zarfar M., Belton, D.J., Hanby, B., Kaplan, D.L. & Perry, C.C. 2015. 'Functional material features if *Bombyx mori* silk light versus heavy chain protein'. *American Chemical Society*. 16(2):606-614.

Zhao, X.M., Liu, C., Jiang, L.J., Li, Q, Q.Y., Zhou, M.T., Cheng, T.C., Mita, K. & Q.Y. 2015. 'A juvenile hormone transcription factor Bmdimm-Fibroin H chain pathway is involved in the synthesis of silk protein in silkworm, *Bombyx mori*'. *The Journal of Biological Chemistry*. 290(2):972-986.

Inoue,S., Kanda, T., Imamura, M., Quan, G.X., Kojima, K., Tanaka, H., Tomita, M., Hina, R., Yoshizato, K., Mizuno, S & Tamura, T. 2005. ' A fibroin-deficient silkworm mutant, ND-S^{D}, provides an efficient system for producing recombinant proteins'. *Insect Biochemistry and Molecular Biology*. 35(1):51-59.