Production and characterization of DNA ligases isolated from Kogelberg Biosphere metagenomics library

By

Lindiwe Zuma

(Student no. 214500478)

A dissertation submitted in fulfilment of the academic requirements of Masters of Science in Biochemistry, School of Life Sciences College of Agriculture, Engineering and Science, University of Kwa-Zulu Natal, Westville, South Africa

2021

Preface

The research contained in this thesis was completed by the candidate while based in the Discipline of Biochemistry, School of Life Sciences, of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville Campus, South Africa. The research was financially supported by the National Research Foundation. The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

[redacted] 07/10/2021

Student: Lindiwe Zuma Date

[redacted]

Supervisor: Dr Ofentse J. Pooe Date

DECLARATION: PLAGIARISM

I, **Lindiwe Khumbuzile Zuma**, declare that:

(i) The research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) This dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv)This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) Their words have been re-written but the general information attributed to them has been referenced;

b) Where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) Where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) This dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Lindiwe Khumbuzile Zuma _____[redacted]___

Date of submission/acceptance _____

ABSTRACT

Microbial enzymes have been described as an underutilized source of novel enzymes with potential economic advantages. Recently discovered enzymes such as DNA ligase from metagenomic studies, have been shown to achieve great potential in transforming the reagent market specifically in the African continent. Reagent proteins are frequently utilized in the research field widely and are prone to protein degradation and shelf-life reduction. Hence, this study sought to improve biological activity, shelf life and stability of the two DNA ligases identified from Kogelberg Biosphere metagenomics library. Two recombinant DNA ligases expression studies were done using *E.coli* BL21 and purification studies were done subsequently using affinity chromatography. Both recombinant DNA ligases (Ligsv081 & LigpET30) were successfully expressed and purified as homogenous proteins. In this study two approaches were used to enhance the biological DNA ligases, the first approach used was PEGylation. The purified proteins were conjugated to PEG using site-specific PEGylation and non-specific PEGylation. FTIR and UV-VIS spectroscopy were used to analyze the secondary structure of the PEG conjugated DNA ligases. Thermal stability assays were then employed to assess protein stability in the conjugation with PEG. Site-specific PEGylation enhanced ligase activity and reduced the formation of protein aggregates. The second approach involved DNA ligase co-expression in the presence of PfHsp70 or chimeric transcription factor, respectively. Protein co-expression and co-purification assays were conducted. The co-expression and co-purification assays of both proteins with chimeric transcription factor (cTF) were successful, followed by co-expression and co-purification of LigpET30-PfHsp70. Ligation assays were conducted to assess bioactivity of proteins. All DNA ligase complexes were functional and their melting point was increased. Taken together, site-specific PEGylation and protein co-expression with PfHsp70 potentially extended the shelf-life and stability of the proteins. PEGylation strategies and co-expression strategies can potentially be used to enhance reagents in diagnostic and therapeutic tools in molecular biology field.

Keywords: DNA ligases, PEGylation, thermal stability, co-expression

ACKNOWLEDGEMENTS

Firstly, I would like to Thank God for giving me wisdom, patience and grace.

Secondly, I would like to thank my supervisor, Dr Ofentse Jacob Pooe for his constant support and having faith on me to complete this project with success. It has been an honour in working with him. I would like to extend my gratitude towards the Department of Biochemistry (School of Life Sciences), UKZN, for giving me a chance to study for my MSc. degree and I also would like to thank SiiP-DSi-TIA for funding my research study.

To my laboratory members and co-post graduate students; Xolani Mazibuko, Nontokozo Msomi, Nothando Gasa, Dr. Gideon Ayeni, Senzosenkosi Mkhize, Avanthi Moodle and Nonduduzo Mabaso, I am grateful for your help, kindness and friendliness during the entire period of this study.

Last but not least, I would like to thank my family, my father for his support and love that kept me going forward in this journey.

TABLE OF CONTENTS **Page no.**

APPENDICES

LIST OF TABLES

Tables **Page no.**

LIST OF FIGURES

Figures **Page no.**

LIST OF ABBREVIATIONS

ATP -	adenosine triphosphate
BACs-	bacterial artificial chromosomes
CAGR-	compound annual growth rate
CHO-	Chinese ovary
CSIR-	the Council for Scientific and Industrial Research
cTF-	chimeric transcription factor
CV-	column volumes
dNMPD -	deoxynucleoside monophosphate
dNTP -	deoxynucleoside triphosphate
ECL -	chemiluminescene
EDTA -	ethylenediaminetetraacetic acid
FTIR -	Fourier transform infrared
HRP -	horse-raddish peroxidase
Hsp70 -	70 kilo Dalton heat shock proteins
IMAC -	Immobilized Metal-Affinity Chromatography
IPTG -	Isopropyl B-D-thiogalactoside
IR -	infrared
KBR -	Koglberg Biosphere Reserve
kDa -	kilo Dalton
LB -	Luria Burtani medium
mPEG – NTA -	Mpeg-Nitrilotriacetatic acid
NEB -	New England Biolabs
Ni-NTA -	nickel-nitrilotriacetatic acid

Ntase - nucleotidyltransferase

ORF - open reading frame

PCR - polymerase chain reaction

PEG - poly (ethylene) glycol

PVDF - polyvinylidene difluoride

SDS-PAGE - sodium dodecyl-sulfate polyacrylamide gel electrophoresis

TAE - Tris-acetate-EDTA

UV - ultraviolet light

X-gal - 5-bromo-4-chloro-3-Indolyl-beta-D-Galactopyranoside

YAC - yeast artificial chromosome

PLGA - Poly (lactic-co-glycolic acid)

BSA - Bovine Serum Abumin

SASA - Solvent –accessible Surface Area

NLC - Native Chemical Ligation

EPL - Expressed Protein Ligation

LBD - Ligand Binding Domain

MBP - Maltose –Binding Protein

SUMO - Small Ubiquitin related Modifier

AMP - Antimicrobial Peptide

PTM - Post Translation Modification

GPCRs- G protein-coupled receptors

AOX1- Alcohol oxidase 1

SK- Staphylokinase

HPLC- High-Performance Liquid Chromatography

ESI-MS- Electrospray Ionisation Mass Spectrometry

IXC- Ion exchange chromatography

WW- WWP repeating motif

CHAPTER 1: INTRODUCTION

1.1.Introduction

The Kogelberg biosphere reserve complex is located between latitudes 34º 10’ S and 34º 21' S and longitudes 18º 49' E and 19 º 03' E, one hour away from Cape Town at the southern tip of Africa (Gumbi, 2011; Mucina et al., 2006; Tucker, 2013). The Atlantic Ocean which is 7.5 km off a rugged, is where the Kogelberg Biosphere Reserve originates and a rocky shore with glorious interspersed, golden sandy beaches is its identification mark. Additionally, the Kogelberg biosphere reserve is an area recognized for its floral diversity and is thus known as the world’s largest biodiversity hotspot. It is therefore referred to as the heart of the Cape Floral Kingdom and was the first biosphere reserve in southern Africa to be approved by UNESCO. The complex biodiversity of the Kogelberg Biosphere Reserve is home to more than 1, 880 different plant species that are home-based in the folded mountains and highland valleys. The Biosphere Reserve has more than seventy-seven species that are not found anywhere else on earth (Gumbi, 2011; Tucker, 2013).

It hosts approximately 2, 100 species from 20 families that are distinguished mostly for the diversity of the geophyte flora in the Cape. The factors steering diversity of the remarkable Cape flora are still inadequately understood, but factors such as geographical distribution, climatic factors (rainfall quantity and reliability), plant growth form (storage organ size) and soil type diversity or/ abiotic factors have been suggested as possible reasons (Jooste et al., 2019). Oxalis, which is largely populating the Cape is a genus whose abundance is due to successful evolution resulting from its distinct life history that involves geophytic habit, winter flowering, and variable seed strategies. Lately, researchers suggested that the presence of plant-microbial interactions significantly influences the shape, generation and maintenance of ecosystem diversity within the Cape (Jooste et al., 2019). The climate of this area is usually known to be warm and dry in summer, and cool and wet in winter, often with gale-force north-westerly winds and occasional snowfall on the higher peaks, therefore, its classification as a Mediterranean type of climate (Schultz, 2020).

About 100 years ago, bacteriophages were discovered and several studies conducted demonstrated that bacteriophages can play an enormous role in the development of molecular biology and biotechnology (Golkar et al., 2013.; Haq et al., 2012; Petty et al., 2006; Salmond & Fineran, 2015.; Vandamme & Mortelmans, 2018). The interaction mechanism between bacteriophages and their hosts influenced two breakthrough discoveries in biotechnology

(identification of restriction enzymes and CRISPRs) (Jurczak-kurek et al., 2016). The bacterial growth rate is usually mediated by bacteriophages (ecological significance) in soil, as with viruses in other environments, and impact selectively on the diversity of bacterial soil communities. Surprisingly, individual soil niches are influenced by different bacteriophages playing distinctive roles. Additionally, bacteriophages have been recognized for their association in the mediation of biogeochemical properties of soil by controlling the bacterial abundance and hence nutrient cycling capabilities (Allen et al., 2010). Symbioses relationship between roots and bacteria appear to be influenced by phages located within the roots adjacent to the rhizosphere (Batinovic et al., 2019). There are 10^{31} estimated phage particles on pathogens on earth, which is ten-fold more than the bacterial population, making phages the most abundant biological entities in the biosphere (Batinovic et al., 2019).

Over the past ten years, the majority of organisms in the most environments on earth have been studied using metagenomics, which has proven to be a new field of research explaining non-cultured microbes. Furthermore, the science of metagenomics is essentially used for the discovery of new enzymatic activities (Nazir, 2016). Metagenomics is a term that was first coined by Handelsman et al., in 1998. It can be defined as an area of biotechnological research and compiling techniques that allows the exploration of microbial communities of an ecosystem through culture-independent techniques. Metagenomics provides access and allows the collective study of genomes of environmental microorganisms without culturing them (Handelsmanl et al., 1998; Riesenfeld et al., 2004; Sleator et al., 2008). The need for isolation and cultivation of microorganisms is bypassed in metagenomic approaches which allow the exploration of microorganisms that cannot be cultured or are difficult to culture in the laboratory. Consequentially, the utilization of metagenomic approaches is based on the isolation of nucleic acids directly from environmental samples, thus permitting ecological comparison and exploration of ecologies. Also, in the complex environment of the microbial community, the metabolic profile is determined and even facilitates the identification of novel biomolecules from isolated nucleic acids using constructed libraries (Katz et al., 2016).

Direct environmental DNA cloning was firstly proposed by Stahl *et al*., (1985) by cloning of picoplankton DNA in the phage vector to analyse sequences of subsequent 16S rRNA genes (Stahl et al., 1985). The main source for metagenomic library constructions and other metagenomic DNA is the utilization DNA which is highly suitable for cloning and microbial

diversity present in the original sample. Technically, high extreme environments still present complications during isolation of DNA, this is due to the presence of several microorganisms to be lysed. Numerous approaches have been developed for DNA extraction from mesophilic samples which resulted in acquiring or obtaining highly stable nucleases upon cell lysis (Stahl et al., 1985).

Nonetheless, DNA with high quality reserved from a variety of environments such as soil, marine picoplankton, contaminated subsurface sediments, groundwater, hot springs, etc., can be isolated through the utilization of various methods that have been significantly developed. Recovery of novel biomolecules from environmental microbial assemblages was the first main purpose of functionalizing metagenomics. The two techniques utilized for metagenomic libraries screening to develop novel biomolecules recovery from environmental samples are function-based and sequence-based approaches. Functional metagenomics experimental approaches involve acquiring new cultivation methods like meta-transcriptomics, meta-proteomics, meta-metabolomics, and enzyme screening (Ferrer et al., 2007; Rondon et al., 2000; Simon & Daniel, 2011). Metagenomic gene libraries and gene expression are used to directly modify and hydrolyze a specific chemical substrate in the enzyme screening approach (Handelsman, 2004). Isolates of DNA from microbial communities are usually analysed in the functional metagenomics to determine the functionality of the encoded protein. Moreover, functional metagenomics involves DNA fragments cloning, gene expression into a host and enzymatic activity screening.

A DNA sequence is not the only approach that can be used to discover the function of novel enzymes discovered using a function-based approach (Lam et al., 2015). Annotating genomes and metagenomes derived only from a sequence-based approach can be done by using information extracted from function-based analyses. Therefore, functional and sequence-based metagenomics are complements. In the beginning, functional metagenomics involves the construction of a metagenomics library. This step involves the extraction of nucleic acids from the environment which is very challenging due to the high plausibility of introduction of bias in microbial metagenomics. The characteristics such as larger consistent insert size and high cloning efficiency are always preferred when considering vectors such as cosmid and fosmid-based libraries (Satyanarayana et al., 2019). DNA is first extracted from the environmental sample of interest, then size-selected, end-repaired, and ligated to a cos-based vector, allowing packaging by lambda phage for subsequent transduction of *Escherichia coli* (Lam et al., 2015).

Early, functional screening was used in metagenomics studies as a starting point for the identification of the gene of interest for sequencing. The functional approach depends highly on specialized vectors such as bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs) since this approach specifically starts with the cloning of large DNA fragments (Singh et al., 2009). Thus, DNA encoding multiple genes is cloned with a host followed by functional activity screening. Furthermore, host heterologous expression is highly required when conducting a functional approach. The required host is expected be able to conduct an expression process for the cloned genes and some of which might need co-expression of other genes, and the host codon and inserted DNA must also be compatible (Mocali & Benedetti, 2010). The functional approach consists of the major advantage; after a successful heterologous gene expression the gene of interest does not require the recognition from sequence analysis, therefore, permits potentially novel genetic elements to be identified. BACs and YACs potentially serve as the PCR-based amplification target due to the known sequence site and further ease the DNA cloning process by providing target genes that can be simply be sequenced (Mocali & Benedetti, 2010). The sequencing stage is mostly essential stage with which several metagenomic studies recently begin at. The sequencing stage excludes generation of fragments and cloning stages, after fractionating community DNA into smaller pieces then direct sequencing via high-throughput sequencing; this approach is known as shotgun sequencing. The functional metagenomic approach aims in gene identification which codes for the function of interest, therefore, results in the generation of expression libraries containing more than thousands metagenomic clones followed by activity-based screenings.

The traditional metagenomic sequencing method known as Sanger sequencing or chain termination has a relatively lower error rate and yields longer reads that are more than 700 base pairs (bp) (Thomas et al., 2012). Sequence-based screening is significantly required to provide the gene fragment present in the vector and can be used to evaluate the microbial diversity in an ecosystem. The developments made in next-generation sequencing (NGS) influenced the new sequence-based screening protocol. Recently, metagenomes of different microbial communities are assessed by using NGS which involves direct DNA sequencing. Consequently, various tools of bioinformatics are used to assemble and annotate a vast amount of sequencing data generated by NGS of metagenomic DNA (Teeling & Glo, 2012). To increase the frequency of detection of novel open reading frames (ORFs), the combination

of the two metagenomic approaches (function-based and sequence-based) is essentially required.

Both screening techniques involve the cloning of environmental DNA and the construction of small-insert or large-insert libraries. Mostly, a host (*E. coli*) is subsequently employed and transformed with the resulting metagenomic libraries. Several vectors such as plasmids, fosmids, cosmids, and bacterial artificial chromosomes have been used to construct metagenomic libraries. Required characteristics of a selected vector depend on the quality of DNA, target genes, and screening strategy. Small-insert is usually employed in novel biocatalysts identification, which is encoded by a particular gene. However, large-insert libraries are essential for recovery of larger gene clusters coding complex pathways (Peng et al., 2014).

Many novel biocatalysts have been identified through the construction and screening of both metagenomic libraries. Additionally, most frequently recovered biocatalysts until today from metagenomes are lipases and esterases. Peng *et al*., (2014) isolated a novel lipase that is alkaline stable from the metagenomic library, in addition to more than hundreds of sequences encoding lipases that were previously identified. The isolated lipase was obtained from a metagenomic library which was constructed from sediments of marine. This novel lipase was found to be essential in milk fat flavoured products, it can significantly impact a unique and desirable flavour and odour into these products (Peng et al., 2014).

Previously, Segobola *et al*., (2018) conducted a study that involved the exploration of known and unknown bacteria from the Kogelberg Biosphere Reserve Fynbos soil microbial communities. They found the abundant accumulation of Actinobacteria, Proteobacteria, Acidobacteria, Planctomycetes and Bacteroidetes which had also been found by other researchers. Additionally, in the Kogelberg Biosphere reserve soil metaviromes were discovered to be highly populated with bacteriophages that are familiar for infecting bacterial host cells and hijacking the replication process of the host cell (Segobola et al., 2018). Thus, the Kogelberg Biosphere reserve soil metaviromes is significantly populated by bacteriophages with novel genes coding for nucleic acid manipulating enzymes which were modified from this specific biosphere. Moreover, their findings have contributed to the knowledge about the presence of bacteriophages and the recently solved puzzle achieved through metagenomic studies (functional and sequence approaches) on the diversity of nucleic acid manipulating enzyme in the metaviromes of soils from the Kogelberg Biosphere

Reserve. The current value of the global molecular biology market for nucleic acid manipulating enzyme provides incentives to identify novel genetic, functional and structural nucleic acid manipulating enzymes (Segobola et al., 2018).

The increasing demand for microbial enzyme utilization in industries is associated with factors such as the need for cost reduction, natural resources depletion and environmental safety (Choi et al., 2015). Globally, the industrial enzymes market in 2017 was estimated at $7,082 million and this number is expected to develop at a Compound Annual Growth Rate (CAGR) of about 5.7% over the period from 2018 to 2024 to reach almost $10,519 million (Singh et al., 2016). Enzymes are essential in nucleic acid manipulation for research and development in the field of genetic engineering, for instance, restriction endonucleases are used for site-specific cleavage of DNA for molecular cloning and DNA polymerases for DNA amplification by polymerase chain reaction (PCR) (Singh et al., 2016). The application of enzymes is collectively diverse in industries and scientific research, varying from degrading different natural substances in starch processing, detergent and textile industries and the manipulation of nucleic acids in biotechnology research (Li et al., 2012).

Revolutionary recombinant technology and genetic engineering have been rapidly expanding due to the ability of enzymes to manipulate nucleic acids *in vivo* and *in vitro*. These nucleic acid manipulating enzymes can be classified into broad categories such as nucleases, polymerases, topoisomerases and DNA ligases. There is an increasing demand for products able to manipulate and facilitate the analysis of nucleic acids due to the growing interest in recombinant technology, genomic and gene-expression analysis and gene editing. Thus, it is important to find alternative ways to ease over exploitation of enzymes in industries. These enzymes face hurdles such as aggregation which sometimes accelerates protein's shorter shelf-life, affects immune response in humans, toxicity, immunogenicity, reduction of stability (Matthews, 2013), solubility and bioactivity (Hamberg, 2007; Price, 1983; Price et al., 1995; Wilson, 2003). The formation of aggregates (proteins move from monomer to oligomer and then to aggregates) is stimulated by interactions between proteins.

Generally, aggregates develop if the newly synthesized proteins misfold, chaperone molecules fail to refold or initiate degradation of a damaged protein, exposure to denaturing conditions, or less optimal storage conditions (Tripathi, 2016; Tripathi & Shrivastava, 2019). Also, during the production and storage of protein, the aggregate can be formed, thus, it is crucial to ensure the safety and stability of protein formulations (Khodabandehloo & Chen, 2017). Therefore, proteins must be stored well to retain their original structural integrity and

activity for a certain period. However, it should be noted that apart from the storage conditions, the nature of the protein also affects shelf-life.

In cases where the aggregation is irreversible, chemical and physical modifications are employed to render such drawbacks (Zapadka et al., 2017). Therefore, researchers developed protein modification that does not alter the protein's biological activity and secondary structures but only assists protein to recover and maintain solubility, stability, native state, and increase production yield (Sørensen & Mortensen, 2005). Such modifications can be performed in two ways; without engineering target protein, such as testing various temperatures, strain, media, or include molecular chaperones (Sørensen & Mortensen, 2005). Such modification can be performed in expression systems (prokaryotic, eukaryotic, mammalian, or in vitro); (Damaj et al., 2020) and therefore, it is essential to choose the most flexible system to conduct such assessments. Another modification method could be protein PEGylation, engineering fusion technology, which uses alternative technologies and expression fragments. Such modifications have been done to other proteins currently on the market and have been approved by the FDA.

1.2. PROBLEM STATEMENT

Proteins are the most important reagents in research and academic fields, essentially involved in accelerating chemical reactions in laboratory experiments (Hamberg, 2007). These proteins face hurdles such as aggregation which sometimes accelerates protein's shorter shelf-life, affects immune response in humans, toxicity, immunogenicity, reduction of stability (Matthews, 2013), solubility and bioactivity (Hamberg, 2007; Price, 1983; Price et al., 1995; Wilson, 2003). The formation of aggregates is stimulated by interactions between proteins. Generally, aggregates develop if the newly synthesized proteins misfold, chaperone molecules fail to refold or initiate degradation of a damaged protein, exposure to denaturing conditions, or less optimal storage conditions (Tripathi, 2016; Tripathi & Shrivastava, 2019). Furthermore, reagent proteins are less evaluated when it comes to enhancing protein qualities. Therefore, we chose to assess DNA ligase proteins obtained as a gift extracted from Koglberg Biosphere Reserve (KBR) Fynbos soil microbial communities, by Segobola *et al*., (2018). Studies have explored strategies that can assist proteins to overcome such drawbacks through PEGylation and protein co-expression or co-purification methods.

1.3. OBJECTIVE

To recombinantly express, purify and characterize DNA ligases isolated from the Kogelberg Biosphere metagenome library.

1.4. SPECIFIC OBJECTIVES

- To produce enhanced DNA ligases
- To Express and purify DNA ligases and ligase cofactor
- To conduct bioactivity assays in the presence or absence of ligase cofactor
- To evaluate the role of PEGylation on protein activity and thermo stability

1.5.STRUCTURE OF THE DISSERTATION

This dissertation is comprised of six chapters which were organized in a submitted form of articles and manuscripts in preparation as shown below.

1. Chapter 2 (article 1): An overview of protein enhancement systems. This manuscript is in preparation to be submitted.

2. Chapter 3 (article 2): Soluble expression and purification of recombinant DNA Ligase protein in *E.coli* system.

3. Chapter 4 (article 3): Evaluating the Functional Stability of Site-specific Protein PEGylation on Recombinant DNA Ligase. This manuscript has been submitted to Protein and Peptide Letters (BMS-PPL-2021-138).

4. Chapter 5 (article 4): Use of Hsp70 and Chimeric Transcription Factor to Enhance DNA Ligase Stability in *Escherichaia coli.* This manuscript is in preparation.

5. Chapter 6: General Discussion, conclusions and future recommendations

REFERENCES

Allen, B., Willner, D., & Oechel, W. C. (2010). Top-down control of microbial activity and biomass in an Arctic soil ecosystem. *Environmental Microbiology*, *12*, 642–648. https://doi.org/10.1111/j.1462-2920.2009.02104.x.

Batinovic, S., Wassef, F., Knowler, S. A., Rice, D. T. F., Stanton, C. R., Rose, J., Tucci, J., Nittami, T., Vinh, A., Drummond, G. R., Sobey, C. G., Chan, H. T., Seviour, R. J., Petrovski, S., & Franks, A. E. (2019). Bacteriophages in Natural and Artificial Environments. *Pathogens*, 8(100), 1–19. doi:10.3390/pathogens8030100.

Choi, J., Han, S., & Kim, H. (2015). Industrial applications of enzyme biocatalysis : Current status and future aspects. *Biotechnology Advances*, *33*(7), 1443–1454. https://doi.org/10.1016/j.biotechadv.2015.02.014

Damaj, M. B., Jifon, J. L., Woodard, S. L., Vargas-Bautista, C., Barros, G. O. F., Molina, J., White, S. G., Damaj, B. B., Nikolov, Z. L., & Mandadi, K. K. (2020). Unprecedented enhancement of recombinant protein production in sugarcane culms using a combinatorial promoter stacking system. *Scientific Reports*, *10*(1), 1–16. https://doi.org/10.1038/s41598-020-70530-z

Ferrer, M., Golyshina, O., Beloqui, A., & Golyshin, P. N. (2007). Mining enzymes from extreme environments. *Ecology and industrial microbiology*, https://doi.org/10.1016/j.mib.2007.05.004

Golkar, Z., Bagasra, O., & Pace, D. G. (2013). Bacteriophage therapy : a potential solution for the antibiotic resistance crisis. *J Infect Dev Ctries*. 8(2):129-136. https://doi.org/10.3855/jidc.3573

Gumbi, D. P. (2011). The impact of change in climate, human demography, and other social factors on the fire regime of the Kogelberg Nature Reserve. South Africa, Doctoral thesis. University of KwaZulu Natal. 1–114.

Hamberg, A. (2007). Enzyme selectivity as a tool in analytical chemistry. Licenciate thesis. Royal Institute of Technology *(KTH).*

Handelsman, J. (2004). Metagenomics : Application of Genomics to Uncultured Microorganisms. *Microbiology and molecular biology reviews*, *68*(4), 669–685. https://doi.org/10.1128/MBR.68.4.669.

Handelsmanl, J., Rondon, M. R., Goodman, R. M., Brady, S. F., & Clardy, J. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology*, 5(10), 245-249. http://biomednet.com/elecref/10745521005R0245.

Haq, I. U., Chaudhry, W. N., Akhtar, M. N., Andleeb, S., & Qadri, I. (2012). Bacteriophages and their implications on future biotechnology : a review. *Virology Journal*. 9:9 http://www.virologyj.com/content/9/1/9.

Jooste, M., Roets, F., Midgley, G. F., Oberlander, K. C., & Dreyer, L. L. (2019). Nitrogen-fixing bacteria and Oxalis – evidence for a vertically inherited bacterial symbiosis. *BMC Plant Biology*,1–10. https://doi.org/10.1186/s12870-019-2049-7.

Jurczak-kurek, A., Tomasz, G., Nejman-faleń, B., Bloch, S., Dydecka, A., Topka, G., Necel, A., & Jakubowska-deredas, M. (2016). Biodiversity of bacteriophages : morphological and biological properties of a large group of phages isolated from urban sewage. *Scientific reports*, 1–17. https://doi.org/10.1038/srep34338.

Katz, M., Hover, B. M., & Brady, S. F. (2016). Culture - independent discovery of natural products from soil metagenomes. *Journal of Industrial Microbiology & Biotechnology*, *43*(2), 129–141. https://doi.org/10.1007/s10295-015-1706-6.

Khodabandehloo, A., & Chen, D. D. Y. (2017). Particle sizing methods for the detection of protein aggregates in biopharmaceuticals. *Bioanalysis*, 9(3), 313–326. https://doi.org/10.4155/bio-2016-0269.

Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D., & Charles, T. C. (2015). Current and future resources for functional metagenomics. *Frontiers in Microbiology*, *6*, 1–8. https://doi.org/10.3389/fmicb.2015.01196.

Li, S., Yang, X., Yang, S., Zhu, M., & Wang, X. (2012). Technology Prospecting on Enzymes : Application , Marketing and Engineering. *Computational and Structural Biotechnology*, *2*(3), 1–11. https://doi.org/10.5936/csbj.201209017.

Matthews, S. S. (2013). *I*nvestigation into the Effects of PEGylation on the Thermodynamic Stability of the WW Domain. Masters Dissertations. 4280. Brigham Young University. https://scholarsarchive.byu.edu/etd/4280.

Mocali, S., & Benedetti, A. (2010). Exploring research frontiers in microbiology : the

challenge of metagenomics in soil microbiology. *Research in Microbiologoy*. 161 (497e505). https://doi.org/10.1016/j.resmic.2010.04.010.

Mucina, L., Rutherford, M. C., Powrie, L. W., & Ellis, F. (2006). The Vegetation of South Africa, Lesotho, and Swaziland. South African National Biodiversity Institute, Pretoria *Fynbos Biome*. 748-790.

Nazir, A. (2016). Review on Metagenomics and its Applications. *Imperial Journal of Interdisciplinary Research* 2, 277–286.

Peng, Q., Wang, X., Shang, M., Huang, J., Guan, G., Li, Y., & Shi, B. (2014). Isolation of a novel alkaline-stable lipase from a metagenomic library and its specific application for milkfat flavor production. *Microbial Cell Factories*, 13(1), 1–9. https://doi.org/10.1186/1475-2859-13-1.

Petty, N. K., Evans, T. J., Fineran, P. C., & Salmond, G. P. C. (2006). Biotechnological exploitation of bacteriophage research. *TRENDS in Biotechnology*. *25*(1). https://doi.org/10.1016/j.tibtech.2006.11.003

Price, C. P. (1983). Enzymes as Reagents in Clinical Chemistry. *biological science*, 411–422. https://doi.org/10.1098/rstb.1983.0014.

Price, C. P., Campbell, R. S., & Hammond, P. M. (1995). Novel enzymes as reagents. *Clinical Chimica Acta*, *237(1-2)*, 3–16. https://doi.org/10.1016/0009-8981(95)06059-M.

Riesenfeld, C. S., Goodman, R. M., & Handelsman, J. (2004). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology*, *6*, 981–989. https://doi.org/10.1111/j.1462-2920.2004.00664.x.

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., Neil, I. A. N. A. M. A. C., Minor, C., Tiong, C. L. A. I., Gilman, M., Osburne, M. S., Clardy, J. O. N., Handelsman, J. O., & Goodman, R. M. (2000). Cloning the Soil Metagenome : a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. *Applied and environmental microbiology*, *66*(6), 2541–2547. DOI**:** 10.1128/AEM.66.6.2541-2547.2000.

Salmond, G. P. C., & Fineran, P. C. (2015). A century of the phage: past, present and future. *Nature Publishing Group*. 13. 777-786. https://doi.org/10.1038/nrmicro3564

Satyanarayana, T., Kumar, S., Bhavdish, D., & Johri, N. (2019). Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications. *microbial diversity in normal and extreme envirinments*, (2). https://doi.org/10.1007/978-981-13-8487-5_14.

Schultz, B. (2020). A spotted landscape: Threats to leopard, Panthera pardus pardus, & their prey within the Boland Mountain Complex, Western Cape. 1–169. Master of Science. South Africa. Stellenbosch University.

Segobola, J., Adriaenssens, E., Tsekoa, T., Rashamuse, K., & Cowan, D. (2018). Exploring Viral Diversity in a Unique South African Soil Habitat. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-017-18461-0.

Simon, C., & Daniel, R. (2011). Mini-review Metagenomic Analyses : Past and Future Trends. *77*(4), 1153–1161. https://doi.org/10.1128/AEM.02345-10.

Singh, J., Behal, A., Singla, N., Joshi, A., Birbian, N., Singh, S., Bali, V., & Batra, N. (2009). Metagenomics : Concept , methodology , ecological inference and recent advances. *Biotechnol. J.*, 480–494. https://doi.org/10.1002/biot.200800201.

Singh, R., Mehta, M. K., Mittal, A., & Kumar, P. (2016). Microbial enzymes : industrial progress in 21st century. *3 Biotech*, *6*(2), 1–15. https://doi.org/10.1007/s13205-016-0485-8.

Sleator, R. D., Shortall, C., & Hill, C. (2008). Metagenomics. *Letters in Applied Microbiology*, *47*, 361–366. https://doi.org/10.1111/j.1472-765X.2008.02444.x.

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*(1), 1–8. https://doi.org/10.1186/1475-2859-4-1.

Stahl, D. A., Lane, D. J., Olsen, G. J., & Pace, N. R. (1985). Characterization of a Yellowstone Hot Spring Microbial Community by 5S rRNA Sequences. *49*(6), 1379–1384.

Teeling, H., & Glo, F. O. (2012). Current opportunities and challenges in microbial metagenome analysisça bioinformatic perspective. *Briefings in bioinformatics*, *13*(6), 728–742. https://doi.org/10.1093/bib/bbs039.

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(3), 1–12. doi:10.1186/2042-

5783-2-3.

Tripathi, N. K. (2016). Production and Purification of Recombinant Proteins from *Escherichia coli*. C*hemBioEng Rev*, *3*, 116–133. https://doi.org/10.1002/cben.201600002.

Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins : Expression Hosts and Process Development. *Front.Bioeng. Biotechnol.*, *7(420)*. https://doi.org/10.3389/fbioe.2019.00420.

Tucker, C. M. (2013). Developing Sustainability Indicators for the Kogelberg and Cape West Coast Biosphere Reserves. masters thesis. South Africa. Stellenbosch University. http://scholar.sun.ac.za.

Vandamme, E. J., & Mortelmans, K. (2018). A century of bacteriophage research and applications : impacts on biotechnology , health , ecology and the economy. *J Chem Technol Biotechnol.* 94: 323–342. https://doi.org/10.1002/jctb.5810

Wilson, L. (2003). Enzyme Reactor Design Under Thermal Inactivation. *Critical Reviews in Biotechnology*, *23*(1), 61–93. https://doi.org/10.1080/713609298.

Zapadka, K. L., Becher, F. J., Santos, A. L. G., Jackson, S. E., & Jackson, S. E. (2017). Factors affecting the physical stability (aggregation) of peptide therapeutics. I*nterface Focus* ,7(20170030). 20170030. http://dx.doi.org/10.1098/rsfs.2017.0030.

CHAPTER 2: AN OVERVIEW OF PROTEIN ENHANCEMENT APPROACHES

2.1. Abstract

Enzymes play a powerful role as catalysts with high specificity and activity under mild environmental conditions. Recently, they further emerged as the most important reagents in research and academic fields, essentially involved in the completion of several reactions conducted in laboratory experiments. Significant hurdles, such as degradation, reduced solubility, reduced shelf-life, in-stability, aggregates formation and toxicity are still present. PEGylation and co-expression methods have been used to alleviate these problems. Over the past 40 years, PEGylation techniques have significantly overcome low solubility, increased protein stability, shelf-life, bioactivity and prevented protein aggregates formation. Also, co-expression protein with molecular chaperones has played similar role. This review seeks to provide an updated view on recently discovered impacts of PEG towards protein stability. This followed by an overview of other methods used to enhance proteins.

Keywords: protein stability, PEGylation, co-expression, chaperones, cTF

2.2. Introduction

The physical or chemical interaction between polyethylene glycol (PEG) and bioactive molecules such as peptides, proteins, small molecules, nucleic acids, and nanoparticles is defined as PEGylation (Turecek et al., 2016b). PEGlyation has been associated with therapeutic level benefits such as increasing protein solubility, thermal and chemical stability, reduces toxicity, increase proteolysis resistance, increase hydrodynamic volume and reduce protein aggregation (Pasut & Veronese, 2012). Furthermore, various studies previously revealed that PEG conjugation into proteins has never changed protein structure. Thus, this highlights or poses an unquestionable fact that PEGylation is a hot topic and vital for application in several fields of study (Santos et al., 2018). They are several studies illustrating the effective application of PEGylation such as its application extending half-life of bovine serum albumin (BSA) from 13.6 minutes to 4.5 hours comparatively to non-PEGylated nanoparticles resulting from PEGylation of poly(lactic-co-glycolic acid) (PLGA) nanoparticles that encapsulated bovine serum albumin (BSA) (size of 200 nm and efficiency entrapment of 48.6 %) (Turecek et al., 2016a).

PEG has a non-charged, flexible and hydrophilic backbone with only terminal sites which are accessible for interactions and functionalization. PEGylation processes are classified as non-covalent and covalent PEGylation process depending on the reaction used. Non-covalent PEGylation (recent scarcely employed) exploits hydrophobic and ionic interactions to form complexes between the polymer or protein. The non-specific PEGylation method previously used by David and Abuchowsky in the late 1970s, demonstrated that PEG conjugation on proteins can subsequently reduce protein aggregation, proteolysis and extend protein shelf-life (Lawrence & Price, 2016). Non-covalent PEGylation is referred to as the first generation PEGylation which typically utilizes amine conjugation. Furthermore, the main objective of the first generation PEGylation involves irreversible conjugation (Nucci et al., 1991). There are limitations of widespread use of this method due to the removal of PEG coating.

The first generation PEGylation has evolved into a second generation PEGylation which involves site-specific PEGylation method, thus increases PEGylation specificity between PEG molecules conjugation with particular moieties in the protein (Nucci et al., 1991). In the covalent PEGylation method, fixed chemical bonds are formed which can be site-specific (Dozier & Distefano, 2015). The covalent technique is mostly desireable and can be cost-effective. PEGylation through thiol, N-terminal, enzymatic and histidine tags are some of the methods used to conduct site-specific PEGylation (David & Morbidelli, 2014). The main pathway of site-specific PEGylation is through reversible conjugation which doesn't inhibit conjugate activity. Therefore, the cleavable linkages are utilised to allow temporal attachment of PEG molecules and the conjugates can be released accordingly at a specific time scheduled (Mishra et al., 2016; Nucci et al., 1991).

However, covalent technique has some limitations; not always feasible and may require higher development time. Furthermore, if the molecule presents several of these target-specific sites, the reaction tends to produce PEGylated species that vary in modification degree (number of attached chains) and positional isomerism as it occurs with random PEGylation. The third generation PEGylation is sought to be developed to achieve higher potency and circulation half-life without compromising fast-acting, site specificity and lower dosages (Swierczewska et al., 2015). Overall, this review illustrates the effects of PEGylation on protein stability, summarizes target amino acids for site-specific PEGylation, analytical methods used to characterize PEGylated proteins and also illustrates other methods used to enhance proteins.

2.3. Effects of PEGylation on Protein Stability

Several reports have demonstrated the effectiveness of PEG towards protein's conformational stability. Recently, studies and computational simulations showed how and in what structural context PEGylation influences PEG to increase protein conformational stability substantially (Lawrence & Price, 2016). Protein stability broadly refers to stability against proteases, thermal stability, thermodynamic stability, circulation time in a living system, chemical stability, and structural stability (Liang et al., 2019). Therefore, protein modifications such as protein PEGylation intending to enhance stability are preferably selected because they operate without disrupting protein secondary structures, irrespective of the PEG chain adopted conformation. However, PEGylation strategies have been shown to positively, negatively, or ineffectively affect protein stability.

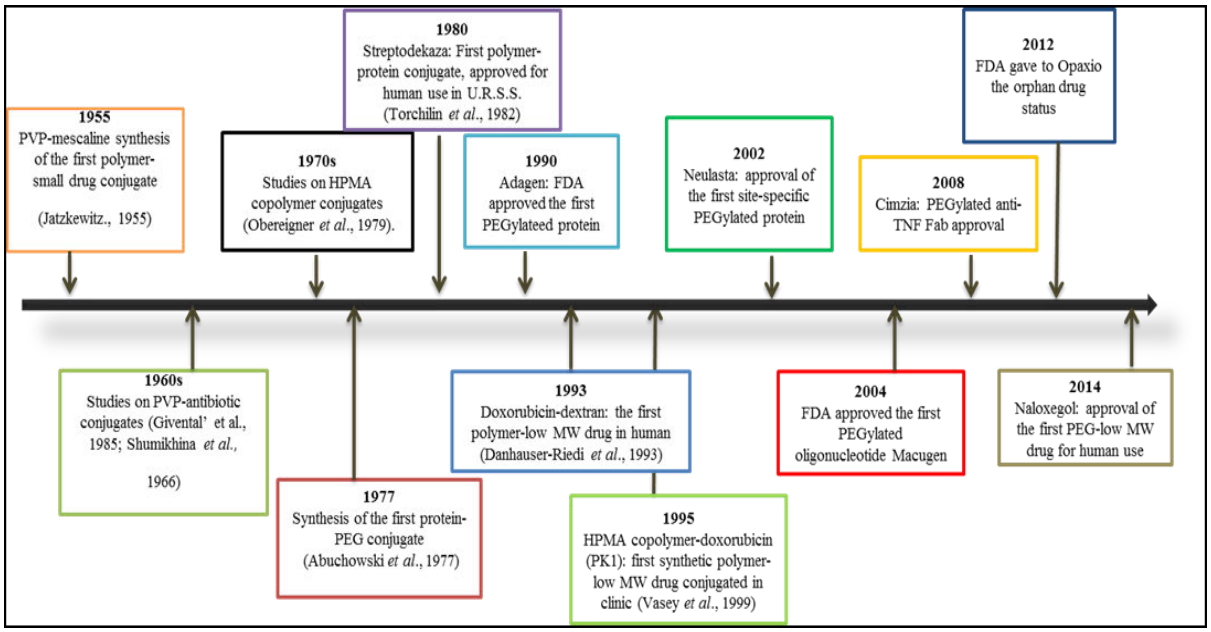

Figure 2.1: Timeline of the most important polymer-drug conjugate developments and biotech-drugs (Grigoletto et al., 2020).

Previously, Lawrence et al., (2014) discovered that the secondary structure motif is not specifically responsible for PEG to induce protein stability. Instead, it was PEG's orientation that was the most influential factor inducing stability (Munasinghe et al., 2021). Furthermore, the study conducted by Abuchowski et al., (1977) showed that amino acids side chains are available for conjugation (Abuchowski et al., 1977). Also, PEG conjugation of BSA showed increase of proteolytic stability, thermal stability, and pH stability was observed. Therefore, their study influenced further understanding of the effects of different PEG (synthetic linear and branched PEG and non-PEG) polymers on the stability and conformation of various

proteins. Similarly, the results obtained showed an increase in half-life and stability (Wright et al., 2019).

Various PEG sizes were used to determine the impact of the PEGylation method on insulin's conformational stability by employing molecular dynamic simulations. The conjugation of PEGs (10, 50, 100, and 200 ethylene oxide units) to insulin was conducted via an amide bond with the e-amino group of LysB29. The solvent-accessible surface area (SASA) was substantially reduced after conjugation of PEG with insulin, and the secondary structure of PEG-insulin conjugate did not change. Furthermore, PEG-protein interactions, including hydrogen bonds and hydrophobic interaction, beneficially excluded water from the surface of insulin. The PEG length caused all these effects, but the molecular weight of PEG beyond 4000 Da did not add any changes (Lawrence & Price, 2016).

Another study previously conducted involved different PEG sizes and demonstrated that the four-unit PEG and PEG with longer chain accelerated folding and reduced unfolding by -0.70+/- 0.04 $kcalmol^{-1}$. The PEG with shorter chains imparted less stability towards WWP domain (WW). The WW domain of the human protein Pin 1 referred to as WW protein, is mainly used as a model to conduct studies. It was preferably selected due to its two folding energetic states, which have been significantly characterized and allow amino acid substitutions at many locations. The WW protein has 34 residues that assist their preparation via solid-phase synthesis of peptides, thus simplifying the linkage of shorter PEG oligomer at a single location (Munasinghe et al., 2019, 2021).

The PEG comprising four ethylene oxide units was attached at position 19 of a single Asn side chain of the WW domain of the human protein. PEG-based stabilization depends highly on the orientation of the side chain at position 19: D-Asn is well tolerated in place of L-Asn at this position, but PEGylation of the D residue does not affect WW conformational stability. Such a result (orientation-dependent) may indicate the fundamental PEG-protein interactions mostly accessible to PEGylated L-Asn but not to PEGylated D-Asn (Chao et al., 2017; Lawrence & Price, 2016).

Several methods have been used to substantially increase conformational stability by PEGylating an Asn residue of WW protein within the reverse turn to accelerate folding and slow unfolding. However, PEGylation protects proteins from proteolysis even though the PEG oligomer is relatively short. Other studies show that direct PEG-OH interactions cannot explain the increased thermodynamic stability; it is likely that nearby OH groups may instead

exert a more indirect influence, involving the network of hydrogen-bound solvent molecules surrounding the protein. Also, the disorder of water molecules around nearby residues was increased by PEG. Further, it stabilized the entropic in origin, with advantageous increases in entropy compensating for unfavorable increases in enthalpy. Lawrence et al., (2014) further reported that WW conformational and proteolytic stability is influenced by both 45- and 4-unit PEG, similarly. Most importantly, the structure-based method can correctly predict a location within the Src SH3 domain present on a beta-sheet protein PEGylation and enhance conformational stability (Lawrence et al., 2014).

The effect of PEG molecular weight was also proven to be capable of decreasing autolysis and increasing the overall stability of chymotrypsin (Wright et al., 2019). Chymotrypsin was conjugated with a different molecular weight of PEG-poly (sulfobetaine methylacrylamide)-block-poly (N-isopropylacrylamide) (pSBAm-block-pNIPAM) (232, 354, and 553 kDa), attempting to enhance pH and thermal stability of chymotrypsin dramatically. Conjugates and native chymotrypsin were incubated at 37^{o}C for eight hours; no significant change was observed for the conjugates while native chymotrypsin lost its 50% initial activity. Similarly, in the incubation of 167mM HCl for three hours, the native enzyme lost 50% of its activity within 30 minutes and all activity after two hours, while conjugates showed residual activity of 60% (Wright et al., 2019).

Additionally, the positional conjugation of WW at 23 with an azido-functionalized four-unit PEG to a propargyloxyphenylalanine residue normally occupied by Tyr, conformational stability increase of PEGylated protein was seen. Also, it was reported that PEG could subsequently increase the strength of the nearby salt bridge, but this effect is not universal. Its precise structural prerequisites are not a simple function of secondary structural context, orientation and distance between PEGylation site and salt bridge, or salt-bridge residue identity (Lawrence & Price, 2016; Xiao et al., 2019).

When the stopped-flow and equilibrium denaturation experiments were used to study the stability of two PEGylated SH3 variants and non-PEGylated counterpart, the non-selective reductive alkylation conditions with PEG-aldehyde were used to prepare tri-PEGylated SH3 variant, these conditions modify the N-terminal a-amine along with two Lys ε-amines (Meng et al., 2012). Also, the mono-PEGylated SH3 variant was produced through alkylation of the N-terminal a-amine with the same PEG-aldehyde. The results showed that tri-PEGylated SH3 was 0.93 kcal/mol stable more than the non-PEGylated counterpart. Also, the PEGylation at

the N-terminal did not substantially change the stability of SH3. The unfolding rate and the SASA of the SH3 folded conformation are directly proportional to an increase in stability (Meng et al., 2012). Furthermore, the residue PEGylation of WW (PrF) strengthened the Glu12-Arg14 salt-bridge by shielding it from the interference of water molecules. The PEGylation site strongly determined protein degradation by a cocktail of proteases. Therefore, the best site characterized by PEG increased protein conformational stability (Draper et al., 2020).

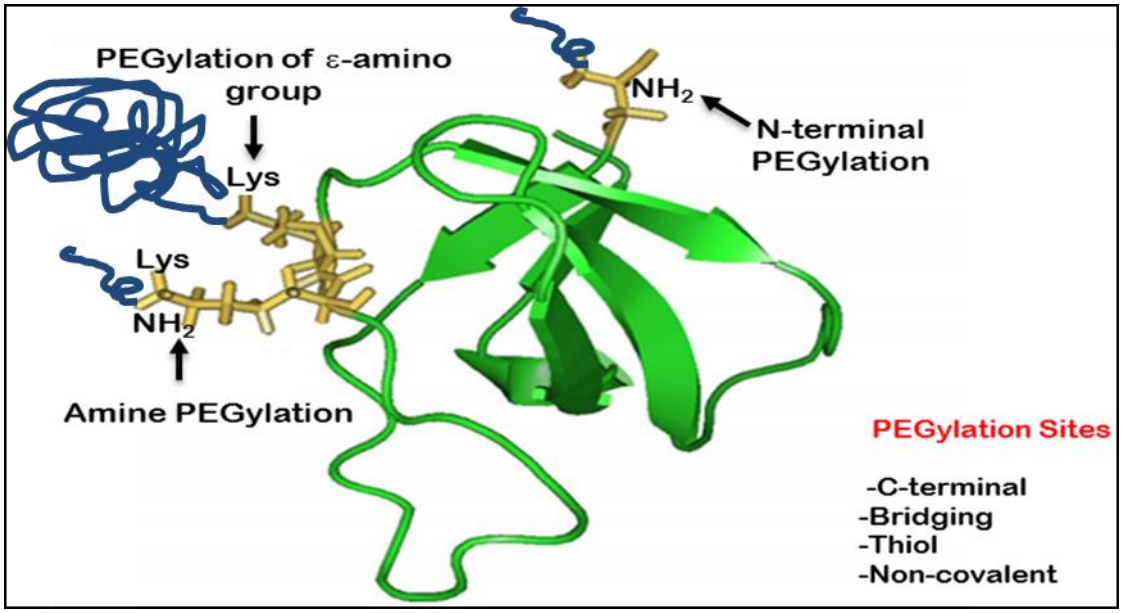

Figure 2.2: Demonstrating the native structure of srcSH3 and three PEGylation sites highlighted with blue (Meng et al., 2012).

Also, Steven R. E. Draper et al. (2020) recently reported that specific modification of Asn residue on the side chain amide nitrogen within the WW domain with a 190 Da monomethoxyPEG (four ethylene oxide units) substantially increased WW conformational and proteolytic stability. In this case, the optimal increase in proteolytic stability was associated with a high increase in protein conformational stability. Furthermore, they found that (depending on the identity of the linker between PEG and protein) alternative PEGylation strategies can effectively alter the impact of PEG on the conformational stability of WW protein. For example, modifying a propargyloxyphenylalanine residue (PrF) at position 19 with a PEG azide via the coppercatalyzed azide-alkyne cycloaddition slightly stabilizes WW relative to the non-PEGylated variant (Draper et al., 2020).

The location, length of PEG and chemistry used to connect PEG with protein influences conformation stability as reported by researcher. Also, different approaches used play a vital role when enhancing their stability.

2.4. Strategies for the Modification of Specific Amino Acids

2.4.1. Targeting Cysteine

Cysteine residues are mostly covered within the protein structure with low apparition frequency. Thus, makes them the most interesting targets for residue-specific modification and reduces cysteine from being accessible to chemical reagents. Furthermore, since it is regarded as rare in nature, therefore they are regularly introduced through genetic engineering (Gunnoo & Madder, 2016; Turecek et al., 2016b). Native chemical ligation (NCL) has been exploited to modify proteins with N-terminal cysteine, this process firstly and reversibly forms a thioester intermediate, followed by a spontaneous shift of S-to-N acyl and end with a production of amide bond (Rosen & Francis, 2017). This method was used in PEGylation of HSA molecule-free Cy34 with PEG-maleimide for the protein sulfhydryl (-SH) groups (highly specific). Furthermore, the method used generated well-defined products, homogeneous products and prevented dimerization. Another technique with an improvement for NCL, which plays a major role in chemical biology and is mostly exploited in several applications, is known as expressed protein ligation (EPL) (Belén et al., 2019; Young et al., 2017).

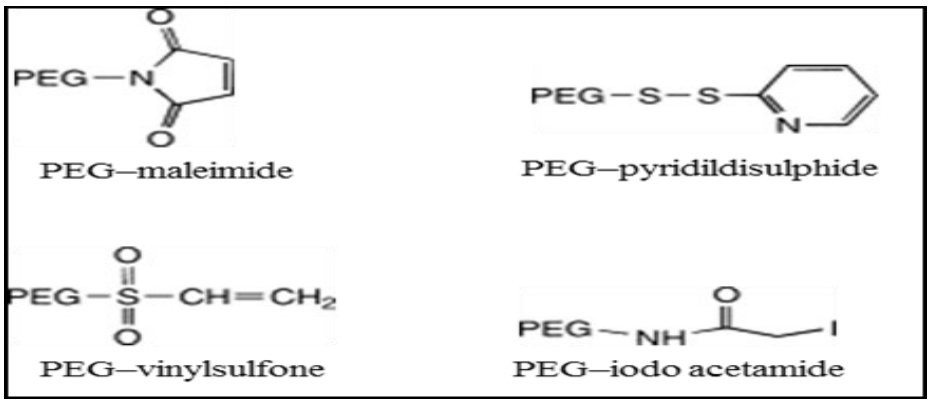

Figure 2.3: PEGs reactive towards a thiol group (Veronese & Pasut, 2005).

Nidetzky and co-workers used this method by PEGylating L-lacate oxidase which retained activity after PEGylation. They mutated serine residue since it was believed to be more susceptible to maleimide PEG, and the results showed a 30% reduction of the activity.

Meanwhile, the PEGylated and unmodified mutant showed a decrease of about 2.5-fold resistance to enzyme activation as compared to the wild types, which demonstrates the difficulty that even small changes in protein structure can cause (Dozier & Distefano, 2015).

2.4.2. Targeting Serine and Threonine

Targeting the N-terminal position of serines and threonine can generate a glyoxylyl group by utilizing a periodate oxidation reaction which can be used in several linkage formations. This reaction is influenced by the susceptibility of 1,2-amino alcohols to periodate oxidation. Previously, Gaertner *et al*., (1996) employed site-specific PEGylation method on N-terminal residue of serine, sodium perioxidate was used for oxidation and conducted oxime ligation with aminooxy and hydrazide PEG derivative (Gaertner et al., 1996). After PEGylation, the modified proteins (interleukin (IL)-8, granulocyte colony-stimulating factor (G-CSF) and IL-1rα) retained their activity (Hoyt et al., 2019). The direct polymer conjugation of tyrosine residues' PEGylation was firstly described by Jones *et al*., (2012). The three components of tyrosine residue modification such as Mannich-type reaction, coupling with diazonium reagents and alkylation at the residue were reported to be the most efficient strategies for tyrosine targeting (Boutureira & Bernardes, 2015). Recently, Mannich-type reaction modification and reactive coloration in fibrous proteins were done, thus confirming their future applications for the reactive process of silk.

2.4.3. Targeting Tryptophan

The Pictet-Spengler reaction with an aldehyde in glacial acetic acid may be used to modify peptides with N-terminal tryptophan residues. This reaction involves N-terminal amino group oxidation to imine, and the cyclic condensation occurs on an aldehyde with the α-amine and the indole side chain of a tryptophan residue, forming a new stable C–C bond (Francis et al., 1996). This method beneficially produces a stable C-C bond in a single step. Sasaki and co-workers employed the Pictet Spengler reaction to label the N-terminal of horse heart myoglobin with an N-terminal glycine by using tryptophan methylester and tryptamine as the coupling partners (Belén et al., 2019; Turecek et al., 2016a).

2.5. Analytical methods for characterization of PEGylated proteins

The number of reports supporting the successful use of PEGylation techniques is increasing. To achieve high-quality product, accurate methods to analyze and provide the most-high level of characterization of molecules should be taken into account. Most of these techniques work better with the combination of the other techniques when characterizing PEGylated proteins, thus getting mostly accurate results. Therefore, this section provides an overview of analytical methods mostly used to characterize PEGlylated proteins and peptides.

2.5.1. High-Performance Liquid Chromatography (HPLC)

HPLC method has been used to quantify and separate both free PEG and PEGylated protein form or PEG conjugates. HPLC further measures molecular weight, mass distribution of polymer and the degree of PEG conjugates (Belén et al., 2019). The utilization of SEC and RP-HPLC previously assessed N-terminal PEGylated EGF, showed the formation of PEGylated macromolecules and N-terminal PEGylation, respectively. Additionally, Brand *et al*., (2015) used HPLC based method to separated N-terminal PEGylated retargeted tissue factor Ttf-NGR, which was then characterized on SDS-PAGE and Western blotting showing one clear band of pure elution fractions of mono-PEGylated protein. The solution of coupling liquid chromatography to mass spectrometry was adopted to improve HPLC performance in PEGylated protein characterization (Belén et al., 2019).

2.5.2. Mass Spectrophotometer (MS)

MS is one of the universal and precise analytic tools employed in protein and PEGlated protein characterization, protein-protein interaction and post-translation modification of proteins. MALDI-TOF MS has been used for decades to characterize mostly the PEGylated proteins to accurately measure molecular weight and PEGylation degree. MALDI further demonstrated and provided excellent information of molecular weight ID regardless of the type and size of PEG used to modify protein. Furthermore, MALDI gives information about heterogeneity, precisely on the distribution of PEG upon protein and the total amount or the information on PEGylation coupling site (Hutanu, 2014).

Previously, MS conjugation with chromatographic separation provided meaningful information concerning heterogeneous protein characterization. But recently, MS has been

conjugated with reversed-phase HPLC to simultaneously allow mass measurements and analyse larger polypeptide chains. Furthermore, several reports discussing the conjugation of MS and MS/MS with non-denaturing separation have been done, mostly conducted in the characterization of complex biopharmaceutical products. Functionally, the native electrospray ionisation mass spectrometry (ESI-MS) tool for online detection for ion exchange chromatography (IXC) usually offers analysis directly without prior collection of fraction and manipulation of sample. The MS detection allows chromatographic separation of species identification based in their masses and gas phase fragmentation patterns, and further permit their conformational integrity monitory in the solution. In another study, Collins and co-workers conducted PEGylation to stabilize formulations of N-terminal amine oxytocin to extend storage or prolonged storage. MS confirmed the conjugation; a clear shift of molecular weight was noticed in the NHS ester polymer to the polymer peptide conjugate. Qin et al., (2012) did an experiment following MALDI-TOF MS utilized comparative analysis to determine modification sites of PEG by mapping between rhGH (recombinant human growth hormone) and PEG-rhGH. Due to challenges (heterogeneity, stability, conformational changes and increased charging) caused by PEGylation in regular use of MS analysis of a protein, different orthogonal MS approaches are required to fully characterize PEG conjugates (Belén et al., 2019; Collins et al., 2016; Meng et al., 2012; Muneeruddin et al., 2017; Yang et al., 2020).

The polydispersity inherent to synthetic polymers is one of the main challenging tasks in the characterization of PEGylated proteins using the MS method by consistently causing higher resolution mass measurements impossible for the protein-polymer conjugation in the size range relevant for biopharmaceutical products, thus reducing the fundamental advantage of MS (Yang et al., 2020).

2.5.3. Dynamic Light Scattering (DLS)

DLS technique is utilized to measure molecular radii of samples and used to evaluate the molecular weight of PEGylated proteins. Recently, the DLS method was used to assess the first large-scale study, which involved the site-specific mono-PEGylation of 15 distinct proteins and characterization of 61 entities in total (Vernet et al., 2016). The decrease in target binding affinities was noted, meanwhile, the hydrodynamic radius was predicted accurately but other sets of proteins their properties were less modelled (Ramberg et al.,

2019). Furthermore, DLS was used to evaluate the size, zeta potentials and physicochemical properties of native and PEGylated Hgh (its site-specific PEGylation was prepared by microbial transglutaminase), and the results indicated that PEGylation can respectively increase or decrease the size and zeta potentials of proteins, enhancing protein stability (Khameneh et al., 2016). The percentage increase of PEG from 0 %, 9 % and 15 % in PLGA NPs was seen increasing particle size from 740 nm to 1170 nm to 1995 nm, respectively. Furthermore, chitosan NPs PEGylation showed similar results where a larger particle size of 290 nm was observed and the non-PEGylated ones had a 200 nm particle size (Sebak, 2018).

Meneguetti *et a*l., (2019) characterized a novel N-terminal PEGylated asparaginase using DLS to show an increase in the hydrodynamic diameter of a PEGylated ASNase which was related to the amount of PEG attached to the protein (Meneguetti et al., 2019). DLS is used diversely among macromolecules, not on PEGylated proteins only, and Ding *et al*., (2005) used DLS to reveal that PEGylation of organic nanotubes dramatically improves the dispersibility of the nanotubes in saline buffer (Ding et al., 2005). They are certain parameters to consider when using DLS such as the presence of larger particles they might be detected during the analysis, close particle in size or a highly polydispersed sample can result in lower resolution, also the viscosity and the density of the particles they cause light absorption by the dispersant interfere with detection (Belén et al., 2019).

2.5.4. Nuclear Magnetic Resonance (NMR)

NMR spectroscopy is useful in characterizing structural conjugates and quantifing PEGylated species in biological complex fluids, also determining the degree of PEGylation, assessing PEGylated higher-order structure of proteins, and the behaviour of free PEG in samples (Ellis et al., 2018; Khandelwal et al., 2019). Recently, Cerofolini *et al*., (2019) used NMR for the structural characterization of larger PEGylated proteins including asparaginase. To increase specificity, NMR was in combination with LC-MS/MS technique and enabled its quantification accuracy of isobaric glycan structure, and this approach can be useful for characterizing the high complexity of PEGylated molecules (Cerofolini et al., 2019).

2.5.5. Immunoenzymatic Assays

ELISA can measure the concentration of PEGylated proteins and study the effectiveness of PEGylation techniques towards protein immunogenicity and the anti-PEG immune response (Tongdee et al., 2020). The pharmacokinetics of Pegasys and PEG-intron were analyzed on sandwich ELISA, by using two mouse monoclonal antihuman IFN-α antibodies that recognize different epitopes of IFN-α. Also the sandwich ELISA was used in the measurement of Neulasta and Mircera (Belén et al., 2019). Su *et al*., (2010) utilized PEG in the production of monoclonal antibodies second-generation (AGP4/ IgM and 3.3/IgG) which is capable of binding subunits of PEG backbone repeatedly containing great affinity compared to those of first-generation AGP3 and E11 (Su et al., 2010). More ranges of specific anti-PEG IgG and IgM monoclonal antibodies for ELISA, FACs, IHC and flow cytometry which is recorded under anti-PEG in the Institute of Biomedical Sciences at Academia Sinica, Taiwan.

2.5.6. Bioinformatics Methods

Computational methods are essentially exploited to engineer, easily design and characterization proteins and further support experimental methodologies, thus save materials and time. Furthermore, it can effectively assist with the proper positioning specific sites for PEGylation (Rouhani et al., 2018). Previously, a bioinformatics study was conducted in which site-specific conjugation of PEG to N- and C-termini of staphylokinase (SK) was conducted, respectively. Four forms of PEGylated staphylokinase obtained were further evaluated to give an insight into the interaction between the PEGylated protein and its receptor. The length of PEG and site of PEGylation of the protein influenced PEG polymer to provide a steric shield towards the protein by wrapping around it (Rouhani et al., 2018). Xu *et al*., (2018) conducted a study to characterize the changes in the molecular level of IFN (representative model system) introduced by several degrees of PEGylation using molecular dynamics simulations (Xu et al., 2018). Simulations generated molecular evidence directly linked to protein and effectively improving stability, retention time, bioavailability and showed bioactivity decrease. Furthermore, provided a vital computational approach improve of PEGylated protein-drug conjugates and their clinical performance (Xu et al., 2018). However, some drawbacks must be resolved in this field such as computational costs in terms

of infrastructure and many times, it could be hard to explain the biological or clinical meaning of features identified using bioinformatics analysis.

2.6. Limitations of PEGylation

PEGylation has been on the market for over 30 years and is the most broadly used post modification technology with structural drawbacks. Therefore, it has been changing from first-generation to second-generation, which is currently used, and there are new attempts of employing third-generation aiming to increase efficacy. PEG polymers' size and position towards conjugates can effectively affect properties. Other drawbacks of PEG include dispersity index, site-specificity of PEG, and PEGylation's degree (Gupta et al., 2019).

Moreover, the polydispersity of PEG (ranges from 1.01 for low molecular weights: 3-5 kDa up to 1.2 for high molecular weight: 20 kDa) may cause challenges, similarly to dispersity towards PEG conjugates. Additionally, the process mainly preventing proteolytic enzymes from advancing towards PEGylated protein can also refuse a substrate from the protein's active site. Therefore, to prevent such complications and reduce other problems, the active-site protecting agents are used, but PEGylation around the protected site can still occur. Another method was developed requiring proper pH and ionic fortitude; this process involves using an inhibitor linked with an insoluble resin (agarose). Therefore, this will effectively protect the active site as well as the surrounding of it. After removing the inhibitor, the enzyme continued to reflect biological activity towards substrates such as albumin and blood clots (with urokinase). Also, PEGylation therapies have caused side effects on patients by entering vasculature, caused hands and foot syndrome (HFS), mucositis, and rash (Gupta et al., 2019). Other drawbacks have been observed in biotechnology and nanomedicine applications where the receptor binding is decreased due to steric hindrance imposed by the PEG chain's disorder (Ramberg et al., 2019).

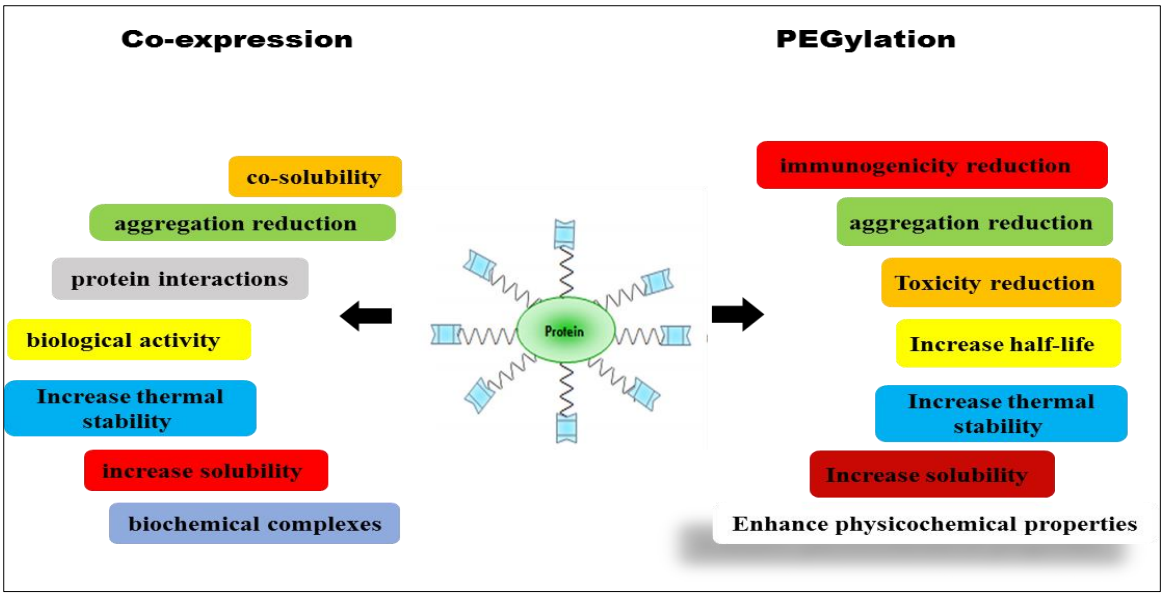

Figure 2.4: potential benefits of protein PEGylation and co-expression.

2.7. Long-standing approaches for protein enhancement

2.7.1. Co-expression studies

Co-expression has been used for the past decade to produce protein complexes with active solubility and stability. With the knowledge of the appropriate components in a complex, potential co-expression partners can be produced and tested. One of the ligand-activated transcription factors is Nuclear receptors which generally regulate development, metabolism and essentially involved in drug discovery. Their ligand-binding domain (LBD) requires the introduction of specific mutation to produce soluble proteins, because when over-expressed in the *E.coli* usually leads to inclusion bodies. Therefore, interacting partners were used to overcome such difficulties through co-expression. The retinoid-X-receptor (RXR) and or steroid receptor co-activator 1 (SRC-1) were previously employed to increase solubility and promoted dimerization of RAR LBD, when expressed without partner protein produced insoluble protein. Furthermore, the co-expression of RAR LBD with RXR resulted in a 1:1 ratio of RAR/RXR with more than 95 % of the expressed proteins in tight heterodimers and the ligand-binding capacity of RAR/RXR heterodimers increased by 5-fold to 10-fold. Co-expression of nuclear receptors with RXR had similar observations, including LXRa/RXRa and PPARc/RZRa (Kerrigan et al., 2011). Another study conducted by Wilson and co-workers, used eukaryotic transcription factors (NF-kB p50 and chimeric transcription factor (cTF) and fused them with T4 DNA ligases. These factors were observed increasing T4 DNA ligase activity by 7-folds (Wilson et al., 2013).

The most utilized co-expression systems are chaperone molecules. They have been employed to enhance protein's stability, solubility, folding, etc. (Sørensen & Mortensen, 2005). Previously, trigger factors were involved in co-overexpression and inhibited aggregate formation in mouse endostatin, human oxygen-regulated protein ORP150, human lysozyme and guinea pig liver transglutaminase. Furthermore, co-overexpression of the GroEL-GroES and DnaK-DnaJGrpE chaperone systems and trigger factors further stimulated soluble expression (Ma et al., 2020). Interaction partner proteins are also categorized and observed as essential partners; hydrophobic residues on protein's surface partially influence insolubility of protein (in the cytoplasm of *E.coli*). Several proteins such as hetero-multimeric suffer from inclusion bodies when expressed without an appropriate binding partner. Previously, it was reported that fusion of heterodimeric complex with thioredoxin influenced subunits of heterodimeric to be soluble (only in the presence of the other) in the solution after proteolytic removal of thioredoxin. Therefore, the interaction partners specifically exist to favour the solubility of target proteins, also the co-expression systems of multiple proteins influence such favours towards target proteins. Soluble expression (in *E. coli)* of the bacteriophage T4 gene 23 product (major capsid protein) required the co-expression of gene product 31 (phage co-chaperonin gp31) (Sørensen & Mortensen, 2005).

Alternatively, protein production yield, solubility (Sørensen & Mortensen, 2005) and folding can be improved through fusion partners or tags. Protein fusion has been approved by United States Food and Drug Administration (FDA) as a promising agent for over 30 years now (Suzuki et al., 2018). Affinity tags are the long-standing tradition for recombinant protein purification, and they have been used to improve protein yield, prevent proteolysis and increase solubility. Furthermore, fusion partners can translocate passenger protein into different locations with less number of proteases within the cell thus, protects a protein from degradation. Such fusion partners are maltose-binding protein (MBP) and small ubiquitin-related modifier (SUMO) can move target protein in the cytosol of *E.coli* to membrane and nucleus, respectively. The MBP and N-utilizing substance A (NusA) are also among the potent solubility enhancing proteins (Costa et al., 2014). Many proteins produce insoluble inclusion bodies during bacterial expression and an only limited amount (25 %) of soluble protein is produced, therefore, fusion tags are introduced into the recombinant construct when *E.coli* is used thus enhance protein solubility (Kosobokova et al., 2016). Alternatively, some tags can be used in the production of toxic proteins; the cellulose-binding modules can be

used as a fusion partner in the production of antimicrobial peptides (AMPs) (Costa et al., 2014).

Table 2.1. Aspects to consider when choosing fusion partners

Aspect	Description
Purpose of fusion	Fusion tags serve different purposes hence it is important to know their target. Hexahistine (His6)-MBP tags are involved in one step where they usually assist with the rapid production of protein/s. the MBP or glutathione-S-transferase function as affinity and increases protein solubility (Costa et al., 2014).
Size and composition of amino acids	This should be considered also when choosing a fusion partner, some proteins require large or small tags depending on the application or purpose (Costa et al., 2014).
Level of production	In structural studies; larger fusion tags are considered for high levels of protein production, whilst in physiological studies; where interactions demand lower production thus require smaller fusion tags (Costa et al., 2014).
Location of tag/s	Tags located at the N-terminal provide reliable context for effective initiation of the translation process, in which fusion proteins take advantage of efficient translation initiation sites on the tag. Furthermore, these tags can be easily removed without leaving any additional residues on the sequence of the target protein (Costa et al., 2014).

2.7.2. Protein modification with Fc-based proteins

Generally, Fc-based proteins are made up of the immunoglobin Fc domain and are usually linked to another peptide. Furthermore, the fused partner can be any other proteinaceous molecule of interest, such as a ligand that activates upon interaction with a cell-surface receptor a peptidic antigen (Ag) against a challenging pathogen protein to identify binding partners assembled in a protein microarray. All fused partners have specific potency, their

interaction with Fc-domain is mainly based on increasing or/ and additional benefits towards their biological activities and pharmacological properties. Furthermore, Fc domains can enhance the solubility and stability of target protein in both *in vitro* and *in vivo*. Fc region allows cost-effective purification with protein-G/A affinity chromatography during manufacture according to technology point of view.

Fc-fusion has been explored for the longest time in pharmacological studies thus researchers decided to extend their application into other fields and systems. Therefore, the Fc domain has been utilized to enhance the expression of the protein in mammalian cells which are not easy to produce. Coupling molecules or proteins (that are hard to produce) to Fc-domains has been reported to increase expression and production of the secreted product. Furthermore, the available Fc-fusion proteins their expression was conducted on Chinese hamster ovary (CHO) cell lines, to ensure correct conformation and post-translational modifications (PTMs) (Sørensen & Mortensen, 2005).

2.8. Conclusion and future recommendations

Protein PEGylation has been growing since its first documentation in 1970s using non-specific PEGylation, till today. As time pass, PEGylation generation also advanced. Currently, researchers are on second-generation and conduct specific synthesis through amino acids present in protein structure. Moreover, researchers are hoping to advance into the third-generation of PEGylation technique in the next few years. This review highlighted the effects of PEGylation toward protein stability which is one of the most important factors for proteins to preserve their biological activity, meanwhile, other factors such as solubility, protein folding, activity, etc., were also enhanced along. PEG enhanced stability through site-specificity or location, changing PEG length and chemistry. Thus, this field requires more development and understanding of what is causing conjugation.

Meanwhile, the co-expression techniques and Fc-fusions are still used to assist proteins or enhance proteins' characteristics. They are several proteins and chaperone systems that have been employed to produce protein complexes. Mostly, chaperones are used to assist protein folding, solubility, activity and stability. Several studies have successfully used protein chaperones for such purposes and still do. Furthermore, other protein partners are used successfully and assist proteins. The co-expression is the most interesting study and should be extended and should be considered to enhance protein solubility, stability and activity.

REFERENCES

Abuchowski, A., McCoy, J. R., Palczuk, N. C., van Es, T., & Davis, F. F. (1977). Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. *Journal of Biological Chemistry*, *252*(11), 3582–3586. https://doi.org/10.1016/s0021-9258(17)40292-4

Belén, L. H., De Oliveira Rangel-Yagui, C., Beltrán Lissabet, J. F., Effer, B., Lee-Estevez, M., Pessoa, A., Castillo, R. L., & Farías, J. G. (2019). From synthesis to characterization of site-selective pegylated proteins. *Frontiers in Pharmacology*, *10*, 1–16. https://doi.org/10.3389/fphar.2019.01450

Boutureira, O., & Bernardes, G. J. L. (2015). Advances in chemical protein modification. In *Chemical Reviews*, 115(5), 2174–2195). American Chemical Society. https://doi.org/10.1021/cr500399p

Brand, C., Fro, M., Ring, J., Schliemann, C., Kessler, T., Mantke, V., Ko, S., Lu, M., Mesters, R. M., Berdel, W. E., & Schwo, C. (2015). Tumor Growth Inhibition via Occlusion of Tumor Vasculature Induced by N - Terminally PEGylated Retargeted Tissue Factor tTF- NGR. *Molecular pharmaceutics.* 321–340. Downloaded https://doi.org/10.1021/acs.molpharmaceut.5b00508

Cerofolini, L., Giuntini, S., Carlon, A., Ravera, E., Calderone, V., Fragai, M., Parigi, G., & Luchinat, C. (2019). Characterization of PEGylated Asparaginase: New Opportunities from NMR Analysis of Large PEGylated Therapeutics. *Chemistry - A European Journal*, *25*(8), 1984–1991. https://doi.org/10.1002/chem.201804488

Chao, S. H., Schäfer, J., & Gruebele, M. (2017). The Surface of Protein λ6-85 Can Act as a Template for Recurring Poly(ethylene glycol) Structure. *Biochemistry*, *56*(42), 5671–5678. https://doi.org/10.1021/acs.biochem.7b00215

Collins, J., Kempe, K., Wilson, P., Blindauer, C. A., Mcintosh, M. P., Davis, T. P., Whittaker, M. R., & Haddleton, D. M. (2016). Stability Enhancing N - Terminal PEGylation of Oxytocin Exploiting Different Polymer Architectures and Conjugation Approaches. *Biomicromolecules*, 17, 2755−2766. https://doi.org/10.1021/acs.biomac.6b00919

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein

solubility , purification , and immunogenicity in Escherichia coli : the novel Fh8 system. *5*, 1–20. https://doi.org/10.3389/fmicb.2014.00063

David, P., & Morbidelli, M. (2014). Process for protein PEGylation. *Journal of Controlled Release*, 180, 134–149. https://doi.org/10.1016/j.jconrel.2014.02.002

Ding, H., Sagar, V., & Agudelo, M. (2015). Enhanced blood – brain barrier transmigration using a novel transferrin embedded fluorescent magneto-liposome nanoformulation. *The journal of biological chemistry*, 280(48). 40235–40240 055101. https://doi.org/10.1088/0957-4484/25/5/055101

Dozier, J. K., & Distefano, M. D. (2015). Site-specific pegylation of therapeutic proteins. In *International Journal of Molecular Sciences*, 16(10). 25831–25864. https://doi.org/10.3390/ijms161025831

Draper, S. R. E., Ashton, D. S., Conover, B. M., Carter, A. J., Stern, K. L., Xiao, Q., & Price, J. L. (2020). PEGylation near a Patch of Nonpolar Surface Residues Increases the Conformational Stability of the WW Domain. *Journal of Organic Chemistry*, *85*(3), 1725–1730. https://doi.org/10.1021/acs.joc.9b02615

Ellis, N., Hattori, C., Cheema, J., Donarski, J., Charlton, A., Dickinson, M., Venditti, G., Kaló, P., Szabó, Z., Kiss, G. B., & Paul, M. (2018). NMR Metabolomics Defining Genetic Variation in Pea Seed Metabolites, 9(1022). https://doi.org/10.3389/fpls.2018.01022.

Francis, G. E., Delgado, C., Fisher, D., Malik, F., & Agrawal, A. K. (1996). Polyethylene glycol modification: Relevance of improved methodology to tumour targeting. *Journal of Drug Targeting*, *3*(5), 321–340. https://doi.org/10.3109/10611869608996824.

Gaertner, H. F., Offord, R. E., Sciences, G., & Avenue, G. (1996). Site-Specific Attachment of Functionalized Poly (ethylene glycol) to the Amino Terminus of Proteins. 7(1), 38–44.

Grigoletto, A., Canato, E., & Pasut, G. (2020). The evolution of polymer conjugation and drug targeting for the delivery of proteins and bioactive molecules. *WIREs Nanomed Nano biotechnol*, 1–33. https://doi.org/10.1002/wnan.1689

Gunnoo, S. B., & Madder, A. (2016). Chemical Protein Modification through Cysteine. *ChemBioChem*, 17, 529–553. https://doi.org/10.1002/cbic.201500667

Gupta, V., Bhavanasi, S., Quadir, M., Singh, K., Ghosh, G., Vasamreddy, K., Ghosh, A., Siahaan, T. J., Banerjee, S., & Banerjee, S. K. (2019). Protein PEGylation for cancer therapy: bench to bedside. *Journal of Cell Communication and Signaling*, *13*(3), 319–330. https://doi.org/10.1007/s12079-018-0492-0

Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L., & Bernardes, G. J. L. (2019). Contemporary approaches to site-selective protein modification. *Nature Reviews Chemistry*, *3*(3), 147–171. https://doi.org/10.1038/s41570-019-0079-1

Hutanu, D. (2014). Trends in Characterization of PEGylated Proteins by Mass Spectrometry. *Modern Chemistry & Applications*, *02*(02), 5–8. https://doi.org/10.4172/2329-6798.1000128

Jones, M. W., Mantovani, G., Blindauer, C. A., Ryan, S. M., Wang, X., Brayden, D. J., & Haddleton, D. M. (2012). Direct Peptide Bioconjugation/PEGylation at Tyrosine with Linear and Branched Polymeric Diazonium Salts. *Journal of the American chemical* society. 134, 7406−7413. dx.doi.org/10.1021/ja211855q

Kerrigan, J. J., Xie, Q., Ames, R. S., & Lu, Q. (2011). Production of protein complexes via co-expression. *Protein Expression and Purification*, *75*(1), 1–14. https://doi.org/10.1016/j.pep.2010.07.015

Khameneh, B., Reza, M., Hassanzadeh-khayyat, M., Mohammadpanah, H., Ghandadi, M., Iranshahi, M., Baratian, A., & Reza, M. (2016). Evaluation of physicochemical and stability properties of human growth hormone upon enzymatic PEGylation. *Journal of Economics, Finance and Administrative Science*, 2–9. https://doi.org/10.1016/j.jab.2016.06.002

Khandelwal, P., Zhang, L., Chimalakonda, A., Caceres-cortes, J., Huang, C., Marathe, P., & Reily, M. D. (2019). Journal of Pharmaceutical and Biomedical Analysis Pharmacokinetics of 40 kDa PEG in rodents using high-field NMR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, *171*, 30–34. https://doi.org/10.1016/j.jpba.2019.03.066

Kosobokova, E. N., Skrypnik, K. A., & Kosorukov, V. S. (2016). Overview of Fusion Tags for Recombinant Proteins. 81(3)

Lawrence, P. B., Gavrilov, Y., Matthews, S. S., Langlois, M. I., Shental-Bechor, D.,

Greenblatt, H. M., Pandey, B. K., Smith, M. S., Paxman, R., Torgerson, C. D., Merrell, J. P., Ritz, C. C., Prigozhin, M. B., Levy, Y., & Price, J. L. (2014). Criteria for selecting PEGylation sites on proteins for higher thermodynamic and proteolytic stability. *Journal of the American Chemical Society*, *136*(50), 17547–17560. https://doi.org/10.1021/ja5095183

Lawrence, P. B., & Price, J. L. (2016). How PEGylation influences protein conformational stability. *Current Opinion in Chemical Biology*, *34*, 88–94. https://doi.org/10.1016/j.cbpa.2016.08.006

Liang, LY, Z., T, K., Bekker, & YX., T. (2019). 乳鼠心肌提取 HHS Public Access. *Physiology & Behavior*, *176*(3), 139–148. https://doi.org/10.1021/acs.bioconjchem.7b00281.PEG-based

Ma, Y., Lee, C., & Park, J. (2020). Strategies for Optimizing the Production of Proteins and Peptides with Multiple Disulfide Bonds. Antibiotics, 9, 541. doi:10.3390/antibiotics9090541

Meneguetti, G. P., Madalena, P., Mariana, K., Obreque, T., Marcello, C., Barbosa, V., Monteiro, G., Helena, S., Farsky, P., Oliveira, A. M. De, Angeli, C. B., Ventura, M., Pessoa-Junior, A., Id, C. D. O. R., Santos, J. H. P. M., Obreque, K. M. T., Barbosa, C. M. V., Monteiro, G., Farsky, S. H. P., … Id, C. D. O. R. (2019). Novel site-specific PEGylated L-asparaginase. *PLoS ONE*, *14*(2), 1–19. https://doi.org/10.1371/journal.pone.0211951

Meng, W., Guo, X., Qin, M., Pan, H., Cao, Y., & Wang, W. (2012). Mechanistic insights into the stabilization of srcSH3 by PEGylation. *Langmuir*, *28*(46), 16133–16140. https://doi.org/10.1021/la303466w

Mishra, P., Nayak, B., & Dey, R. K. (2016). PEGylation in anti-cancer therapy : An overview. *Asian Journal of Pharmaceutical Sciences*, *11*(3), 337–348. https://doi.org/10.1016/j.ajps.2015.08.011

Munasinghe, A., Mathavan, A., Mathavan, A., Lin, P., & Colina, C. M. (2019). Molecular Insight into the ProteinPolymer Interactions in N-Terminal PEGylated Bovine Serum Albumin. *Journal of Physical Chemistry B*, *123*(25), 5196–5205. https://doi.org/10.1021/acs.jpcb.8b12268

Munasinghe, A., Mathavan, A., Mathavan, A., Lin, P., & Colina, C. M. (2021). Atomistic insight towards the impact of polymer architecture and grafting density on structure-dynamics of PEGylated bovine serum albumin and their applications. *Journal of Chemical Physics*, *154*(7). https://doi.org/10.1063/5.0038306

Muneeruddin, K., Bobst, C. E., Frenkel, R., Houde, D., Turyan, I., Sosic, Z., & Kaltashov, I. A. (2017). Characterization of a PEGylated protein therapeutic by ion exchange chromatography with on-line detection by native ESI MS and MS/MS. *Analyst*, *142*(2), 336–344. https://doi.org/10.1039/c6an02041k

Nucci, M. L., Shorr, R., & Abuchowski, A. (1991). The therapeutic value of poly (ethylene glycol) - modified proteins. *Advanced Drug Delivery Reviews*, 6, 133–151

Pasut, G., & Veronese, F. M. (2012). State of the art in PEGylation : The great versatility achieved after forty years of research. *Journal of Controlled Release*, *161*(2), 461–472. https://doi.org/10.1016/j.jconrel.2011.10.037

Ramberg, K. O., Antonik, P. M., Cheung, D. L., & Crowley, P. B. (2019). Measuring the Impact of PEGylation on a Protein-Polysaccharide Interaction. *Bioconjugate Chemistry*, *30*(4), 1162–1168. https://doi.org/10.1021/acs.bioconjchem.9b00099

Rosen, C. B., & Francis, M. B. (2017). Targeting the N terminus for site-selective protein modification. *Nature Chemical Biology*, *13*(7), 697–705. https://doi.org/10.1038/nchembio.2416

Rouhani, M., Khodabakhsh, F., & Norouzian, D. (2018). Journal of Molecular Graphics and Modelling Molecular dynamics simulation for rational protein engineering : Present and future prospectus. *Journal of Molecular Graphics and Modelling*, *84*, 43–53. https://doi.org/10.1016/j.jmgm.2018.06.009

Santos, J. H. P. M., Torres-Obreque, K. M., Meneguetti, G. P., Amaro, B. P., & Rangel-Yagui, C. O. (2018). Protein PEGylation for the design of biobetters: From reaction to purification processes. In *Brazilian Journal of Pharmaceutical Sciences*, 54 (Special):e01009. https://doi.org/10.1590/s2175-97902018000001009

Sebak, A. A. (2018). Limitations of pegylated nanocarriers: Unfavourable physicochemical properties, biodistribution patterns and cellular and subcellular fates. *International Journal of Applied Pharmaceutics*, *10*(5), 6–12.

https://doi.org/10.22159/ijap.2018v10i5.27568

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*, 1–8. https://doi.org/10.1186/1475-2859-4-1

Su, Y., Chen, B., Chuang, K., Cheng, T., & Roffler, S. R. (2010). Sensitive Quantification of PEGylated Compounds by Second-Generation Anti-Poly (ethylene glycol) Monoclonal Antibodies. 525, 1264–1270.

Suzuki, S., Annaka, H., Konno, S., Kumagai, I., & Asano, R. (2018). Engineering the hinge region of human IgG1 Fc-fused bispecific antibodies to improve fragmentation resistance, 1–10. https://doi.org/10.1038/s41598-018-35489-y

Swierczewska, M., Lee, K. C., Lee, S., Swierczewska, M., Lee, K. C., & Lee, S. (2015). What is the future of PEGylated therapies ? What is the future of PEGylated therapies ? *Expert Opinion on Emerging Drugs*, *20*(4), 531–536. https://doi.org/10.1517/14728214.2015.1113254

Tongdee, M., Yamanishi, C., Maeda, M., Kojima, T., Dishinger, J., Chantiwas, R., & Takayama, S. (2020). One-incubation one-hour multiplex ELISA enabled by aqueous two-phase systems. *Analyst*, *145*(10), 3517–3527. https://doi.org/10.1039/d0an00383b

Turecek, P. L., Bossard, M. J., Schoetens, F., & Ivens, I. A. (2016a). PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. *Journal of Pharmaceutical Sciences*, *105*(2), 460–475. https://doi.org/10.1016/j.xphs.2015.11.015

Turecek, P. L., Bossard, M. J., Schoetens, F., & Ivens, I. A. (2016b). PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. In *Journal of Pharmaceutical Sciences*, 105(2), 460–475). Elsevier B.V. https://doi.org/10.1016/j.xphs.2015.11.015

Vernet, E., Popa, G., Pozdnyakova, I., Rasmussen, J. E., Grohganz, H., Giehm, L., Jensen, M. H., Wang, H., Plesner, B., Nielsen, H. M., Jensen, K. J., Berthelsen, J., & Weert, M. Van De. (2016). Large-Scale Biophysical Evaluation of Protein PEGylation E ff ects: In Vitro Properties of 61 Protein Entities. https://doi.org/10.1021/acs.molpharmaceut.6b00049

Veronese, F. M., & Pasut, G. (2005). PEGylation , successful approach to drug delivery REVIEWS. *10*(21), 1451–1458.

Wilson, R. H., Morton, S. K., Deiderick, H., Gerth, M. L., Paul, H. A., Gerber, I., Patel, A., & Ellington, A. D. (2013). Engineered DNA ligases with improved activities in vitro. *Protein Engineering, Design & Selection*, 26(7), 471–478. https://doi.org/10.1093/protein/gzt024

Wright, T. A., Page, R. C., & Konkolewicz, D. (2019). Polymer conjugation of proteins as a synthetic post-translational modification to impact their stability and activity. In *Polymer Chemistry*, 10(4), 434–454). Royal Society of Chemistry. https://doi.org/10.1039/c8py01399c

Xiao, Q., Draper, S. R. E., Smith, M. S., Brown, N., Pugmire, N. A. B., Ashton, D. S., Carter, A. J., Lawrence, E. E. K., & Price, J. L. (2019). Influence of PEGylation on the Strength of Protein Surface Salt Bridges. *ACS Chemical Biology*, *14*(7), 1652–1659. https://doi.org/10.1021/acschembio.9b00432

Xu, D., Smolin, N., Shaw, R. K., Battey, S. R., Tao, A., Huang, Y., Rahman, S. E., & Caylor, M. L. (2018). RSC Advances Molecular insights into the improved clinical performance of PEGylated interferon therapeutics : a molecular dynamics perspective. *RSC Adv*, 8, 2315–2322. https://doi.org/10.1039/c7ra12480e

Yang, S. H., Chen, B., Wang, J., & Zhang, K. (2020). Characterization of High Molecular Weight Multi-Arm Functionalized PEG-Maleimide for Protein Conjugation by Charge-Reduction Mass Spectrometry Coupled to Two-Dimensional Liquid Chromatography. *Analytical Chemistry*, *92*(12), 8584–8590. https://doi.org/10.1021/acs.analchem.0c01567

Young, L. M., Ashcroft, A. E., & Radford, S. E. (2017). Small molecule probes of protein aggregation. *Current Opinion in Chemical Biology*, *39*, 90–99. https://doi.org/10.1016/j.cbpa.2017.06.008

CHAPTER 3: SOLUBLE EXPRESSION AND PURIFICATION OF RECOMBINANT DNA LIGASE PROTEIN IN *E.COLI*

3.1. ABSTRACT

Proteins are amongst the few most fundamental organic macromolecules and can function as enzymes particularly DNA ligases. DNA ligases are involved in the survival of several organisms and are important as an enzyme in research studies. It is important to produce functional effective, well folded and soluble proteins thus, they can fulfil their activities. This study aimed to optimize the production and analyse the functionality of DNA ligase proteins. SDS-PAGE was used to analyse and characterize protein expression and purification. Ligation assays and thermal stability were used to analyse biological activities and thermal stability of proteins. Also, proteins were stored overnight at room temperature to assess stability. Obtained results showed effective bioactivity of DNA ligases during ligation assay. Proteins that were stored overnight were unable to ligate, but after the DNA ligases' amount involved in the ligation assay was increased and results showed successful ligation. Furthermore, thermally denatured proteins were unable to ligate since their native state was compromised. A stable protein preserve its native structure and bioactivity thus can function properly.

Keywords: protein expression, thermal stability, purification, biological activities

3.2. INTRODICTION

A step toward furthering studies concerning the structure and function of heterologous proteins is termed or known as protein expression, which could produce significant amounts of higher demand compounds (enzymes, hormones, agricultural products, vaccines, drugs, etc.) in biotechnological processes (Karbalaei et al., 2020). In the past few years, many diverse proteins were inaccessible due to specific difficulties (relatively expensive) therefore, recombinant DNA technology enabled scientists to produce them (Fakruddin et al., 2013). Proteins are vital in assay reagents and are essential antigens (Konczal & Gray, 2017). In biochemistry, protein expression and production are the most fundamental aspect. Generally, protein expression is conducted through several inexpensive systems such as prokaryotic, eukaryotic, yeast, insect cells, mammalian cells, or *in vitro* systems (Jia & Jeon, 2016).

The universal prokaryotic expression system (*E.coli*) was first used for human insulin production in 1982 (Burnett & Burnett, 2020) and has been utilized for decades for several motives due to their distinct characteristics. This system helps cells grow spontaneously and allows scale alteration, thus allowing the production of both larger and smaller quantities of proteins (Wiseman et al., 2020). Furthermore, *E.coli* has produced several recombinant protein drugs currently on the market, and more than a hundred are still being developed or produced (Burnett & Burnett, 2020). *E.coli* is also used because of its convenience for most laboratories, affordability, diverse vectors, and strains that can maximize protein expression (Konczal & Gray, 2017). Recombinant protein productions are generated in unlimited supplies when *E. coli* is utilized (Fakruddin et al., 2013).

Another part of expression systems is the eukaryotic cells which include mammalian and yeast cells. The most common mammalian cell lines are Chinese hamster ovary (CHO) cells. Currently, CHO cells are used to produce biopharmaceutical compounds, monoclonal antibodies, and Fc-fusion proteins. Also, baby hamster kidney, human embryonic kidney 293 and, NS0, SP2/0 (mouse-derived myeloma) cell lines have recently been legalized (Karbalaei, 2020). The advantages of this system are guaranteed protein stability, posttranslational modifications, and proper protein folding (Bulletin & Khan, 2013). However, the limitations of mammalian expression system include growth, which takes longer, the requirement of essential highly-priced nutrients and large-scale production as well the high probability of media contamination with some viruses (Karbalaei et al., 2020).

On the other hand, several yeast strains, including *S. cerevisiae* and *S. pombe,* have been used to investigate the stability and signaling of G protein-coupled receptors (GPCRs). However, for GPCRs overexpression other yeast strains were used such as the methylotrophic yeast and *Pichia pastoris* for structural studies. (Wiseman et al., 2020). The utilization of *P. pastoris* has been feasible because it is easy to manipulate and integrate an expression vector onto it (Ma et al., 2020). Additionally, it allows a high protein expression level to be induced under the tightly controlled alcohol oxidase 1 (AOX1) promoter. It can grow into high cell densities in glycerol and utilize methanol as a carbon source. Every protein system has drawbacks; similarly, *P. pastoris* can cause misfolding of protein through hypermannosylation and can potentially glycosylate protein residues. Furthermore, engineering without these sites (like GPCRs expression) in *P. pastoris* can facilitate crystallization. The yeast system has various advantages: post-translation modifications, rapid high growth of cell densities on a large scale in relatively cheap media.

The recovery of protein from cells becomes the main challenge for all expression systems including *E.coli* where secreted proteins are not present. Therefore, this requires recovery of protein from the cell which is achieved through cell lysis that involves the application of chemical, biochemical, and mechanical methods such as sonication or cell disruption (Labrou, 2014). The purification tasks of biotechnology proteins are very difficult because of the complex mixtures present in products. Therefore, diverse purification methods are utilized in the research laboratory to achieve a high level of purification at the bench scale, and all these methods can be scaled up to the pilot and production levels.

The utilization of affinity tags with recombinant DNA techniques ensures protein modification. Efficient identification, production and, isolation of the protein of interest from the host system are a result of the use of the affinity tags (Young et al., 2012). However, challenges such as protein insolubility, conformation, stability, and structural flexibility, as well as low purification yields, and host cell toxicity are meant to be resolved when microbial hosts are used for expressing recombinant proteins. Fusion tags are incorporated to increase expression yields and influence solubility and native folding which leads to the resolving of challenges of production efficiency and purification of protein; novel tags in combination with affinity techniques increase purification yields and proteases result in tag removal (Young et al., 2012).

This study will only utilize and focus on polyhistidine (polyHis) tags. The recombinant protein purification usually accomplished through the utilization of the polyHis-tags which are the most used affinity tags has led to biophysical and structural studies. Advantages of polyHis-tag include low immunogenicity and it also has a small size of approximately 0.84 kDa -with composition ranging from 3 to 10 His tags in series. Furthermore, polyHis-tags positioned at both or either N- or C-terminus influence protein functionality and, purification methods can be carried out under both native and denaturing conditions. Immobilized Metal-Affinity Chromatography (IMAC) is used in the purification process of recombinant polyHis-tagged proteins. Numerous structural studies have been achieved by the polyHis-tags IMAC purification process. More than 60 % of all protein structures that exist include a polyHis-tag (Derewenda, 2004).

Furthermore, the principle of IMAC allows separation of transition metal ions such as Ni^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+} which are immobilized on a matrix and the metal-binding peptides based on the interaction between the His. The developed of Ni (II)-nitrilotriacetic acid (Ni^{2+}-NTA)

displayed a high affinity for histidine residue which is adjacent to it (Knecht et al., 2009). Thus, it withstands multiple regeneration cycles by providing an inexpensive matrix under stringent conditions. Bound recombinant proteins dissociate through the application of an imidazole gradient which usually ranges from 20 mM to 250 mM for polyHis-tagged fusion proteins with changes in pH and metal chelation (Hefti et al., 2001). Non-specific binding proteins of the host can occur during the purification process which can be resolved by conducting shorter incubation steps and or by utilizing imidazole low concentrations in the column washing step. The presence of imidazole can cause aggregation of protein and affects activity; therefore, it is essential to utilize dialysis to remove residual imidazole following purification. IMAC can be used to purify polyHis-tagged proteins under denaturing conditions and then refolded. This study aimed to express and produce functionally active DNA ligase proteins. DNA ligase enzymes are responsible for repairing and restoring the DNA backbone (Hjerde et al., 2020). DNA ligases play an indispensable role in maintaining genomic integrity and ensuring the completion of the DNA helical structure during DNA replication, recombination, and repair. DNA ligase functional activities include the joining of Okazaki fragments at the replication fork on the lagging strand during DNA replication and catalysing the final steps in base excision repair and non-homologous end-joining repair pathways (Hjerde et al., 2020; Samai & Shuman, 2011; Wilkinson et al., 2001).

3.3. MATERIALS AND METHOD

All laboratory experiments presented in this dissertation were conducted at the University of Kwa-Zulu Natal (Westville), Biochemistry department.

3.3.1. Protein expression

Selection plates were used to plate expression plasmids which were then incubated at 37 °C overnight. Afterwards, a starter culture was produced by selecting a single colony which was re-suspended in liquid culture with a specific antibiotic. A dilution factor of 1:100 was used to inoculate a starter culture into expression media with specific antibiotic which was incubated with shaking at 37 °C while waiting for OD_{600} to reach 0.4 to 0.6, then protein expression was induced by 100 µM IPTG for 6 hours at 37 °C followed by recovery through centrifugation at 13000 g for 4 minutes using a JA14 rotor (Beckman, Optima L-100XP). Then, lysis buffer was used to re-suspend the pellets thus release intracellular proteins. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to

analyse the soluble fractions of protein and visualization of proteins was done with Coomassie stain and confirmed by western blot, subsequently. Polyvinylidene difluoride (PVDF) membrane (0.2 µl pore) was used to transfer proteins and the membrane was incubated into 5 % skim milk in Tris-buffered saline with tween 20 (TBST) for 1 hour to block non-specific binding proteins. Samples were reacted with anti-HIS-tag IgG conjugated with horse-radish peroxidase (HRP from Sigma Germany; diluted 1:1000 in TBST) for one hour in the ice, prior washing for three times. Chemiluminescence (ECL) kit method (Thermo Fisher scientific, SA) was employed to as enhance for protein bands to be detected and developed.

3.3.2. Protein purification

A polyhistidine tag (His-tag) was used in expression of recombinant ligases and transcription factor to facilitate affinity chromatography purification in this study. Furthermore, the HisPur^{TM} Ni-NTA Resin column (ThermoFisher Scientific, SA) was used in the purification of the His-tag proteins following the manufacturers' protocol. Briefly, LEW buffer was used to equilibrate the column followed by loading of 15 ml soluble crude fraction made up of a targeted His-tag proteins. Thereafter, three column volumes (CV) of LEW buffer were used to wash off unbound proteins. Targeted protein was then eluted with elution buffer which was then confirmed with 15 % SDS-PAGE and western blot was used to verify purified protein as state above.

3.3.3. Activity assay of the recombinant fusion protein

Ligation reactions described by Sambrook and Russell (2001) were essentially used to test the activity of purified recombinant DNA ligases and in comparison with commercial T4 DNA ligase from Thermo Fisher Scientific. To test both DNA ligase isolates activity, the mixture or reaction mixture consisted of the following: 1 µl of DNA ligase, digested lambda DNA with restriction enzymes (*HindIII*), commercial cutsmart buffer, nuclease-free water in a 20 µl ligation reaction (ThermoFisher Scientific, SA) at 37 °C (pre-setting) for one hour. Ligation time used was 30 minutes.

3.3.4. Thermal stability

The thermal stability was assessed through pre-incubation of DNA ligase enzymes at 40 °C, 65 °C and 70 °C for 30 minutes and also by overnight storage at 25 °C. The residual activity was evaluated by re-ligating Lambda DNA primarily digested with *HindIII* enzyme. The

reaction mixture for each DNA ligase had the following: 1 µl of DNA ligase enzyme, 2 µl of digested lambda DNA, 2 µl cutsmart buffer, nuclease-free water in a 20 µl ligation reaction. Then the ligation time of 40 minutes at the temperature of 37 °C was done. The ligation product was visualized with 1 % agarose gel electrophoresis.

3.4. RESULTS AND DISUSSION

3.4.1. SDS-PAGE & Western blot analysis

Generally, optimization of protein expression is initiated by assessing the media composition. This process helps in achieving high production yield and cost-efficient recombinant protein. In this study, we decided to use the following medium: Luria broth (LB), 2YT and YT. There are other media that can be used including defined and semi-defined media for protein expression in *E.coli*. The concentration of some nutrients when in excess might inhibit cell growth. For example; the presence of glucose in high concentration may cause an increase in acetate which usually suppress production concentration. Meanwhile, the accumulation of divalent cation supplements increases cell growth and enhances protein production (Tripathi, 2016).

Generally, *E.coli* is a host used to conduct protein expression due to its capacity to grow fast in high cell concentrations, cost-efficient and relatively simple to culture. Producing recombinant proteins in *Escherichia coli* remains the preferred method to this day. One of the tools that helped to produce proteins in *E. coli* is the use of fusion proteins. They increase solubility to avoid the formation of inclusion bodies and facilitate protein purification employing affinity chromatography, which is very practical and high possibility to obtain high purities directly from the cell lysate with just one purification step. In order to take advantage of the flexibility of inducible promoter systems, the optimal point in *E. coli* growth for inducing the expression of the recombinant protein and the optimal inducer concentration should be determined. Thus, in this study IPTG concentrations impact on protein production was assessed(Gomes et al., 2020).

The stained gel pictures are presented (Figure **3.1** to Figure **3.6**) below.

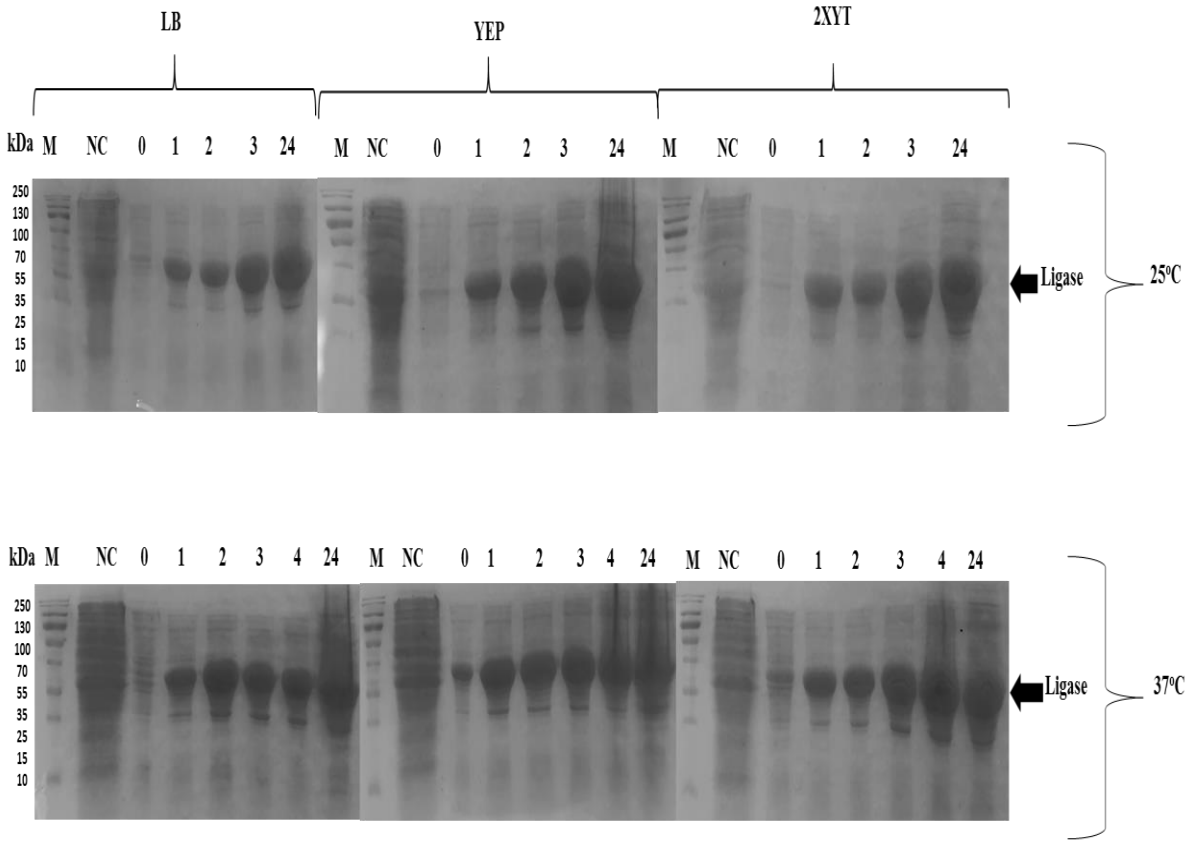

Figure 3. 1: SDS-PAGE showing protein expression in different media at 25 °C and 37 °C. Lane M; marker, Lane NC- *E. coli* BL21 control cells, Lane 0HR and 24HR represent the time-profile sampling at 1 hour and 24 hours post IPTG -induction. LB- Luria broth, YEP-Yeast Extract Peptone and 2xYT-Yeast Extract Tryptone.

A batch cultivation system was used to produce recombinant protein. It should be noted that to produce soluble recombinant proteins, media composition and optimization are critical factors. Therefore, three different medium (LB, YEP & 2xYT) with distinct pH were utilized to produce soluble proteins and conducted at different temperatures(25 °C & 37 °C) to assess which temperature suitable for high yield, reduced aggregates and properly folds produced protein in this study. Higher temperatures promote aggregation reactions determined by the strong temperature dependence of hydrophobic interactions. Therefore, temperature reduction partially eliminates heat shock proteases. Additionally, correctly folded and stable protein is partially influenced by such factors (low temperature). The results show that DNA ligase protein expression is favored by all media used and temperatures, all-time profiles show exponentially increased production of DNA ligase protein with mostly high yield after 24 hours of induction. But the most suitable media showing better yield is YEP at 25 °C. YEP nutrients seem to favor DNA ligase protein expression most by supplying it with required

nutrients at 25 °C without depletion or cell contamination or death for 24 hours. Previously, researchers reported that yeast extract increases protein yield when supplemented in the cultivation media (Tripathi, 2016).

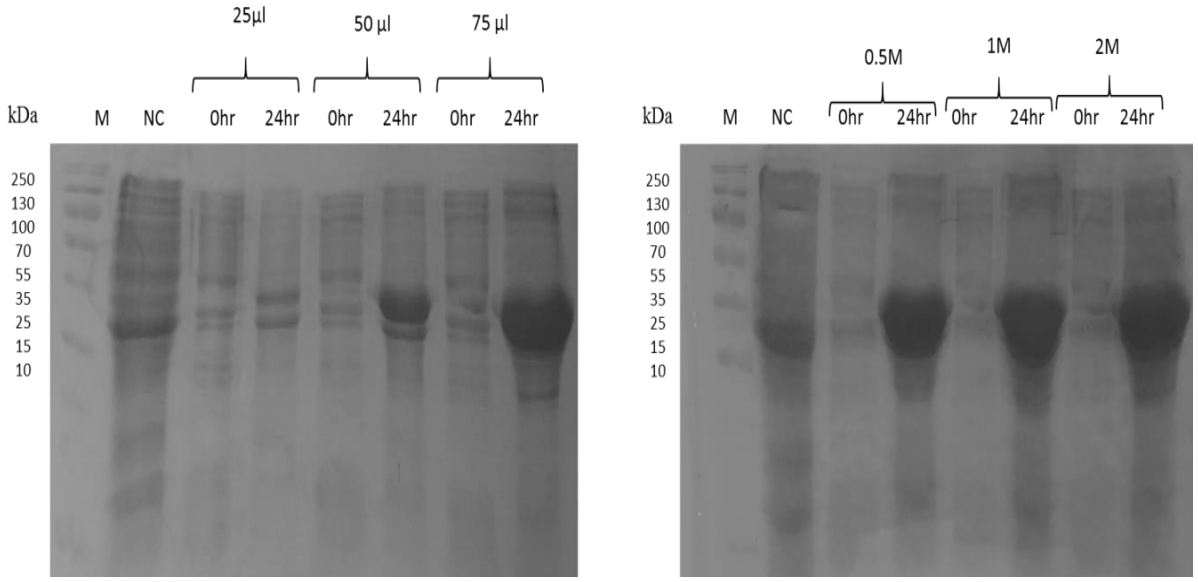

Figure 3. 2: SDS-PAGE (15%) showing protein expression in different amounts of inoculum and increasing concentrations of IPTG. Lane M; marker, Lane NC- *E. coli* BL21 control cells, Lane 0HR and 24HR represent the time-profile sampling at 1 hour and 24 hours post IPTG -induction.

IPTG is usually utilized to induce protein expression when the gene is controlled by lac operon. Further, it is the most widely used in *E. coli* mainly due to its very high expression levels as the target protein can represent up to 50 % of the total cell protein. Level of expression of T7 RNA polymerase from the lacUV5 promoter in *E. coli* strain (λDE3) lysogens can happen in the absence of IPTG leading to leaky expression of heterologous genes placed under the T7 promoter (Gomes et al., 2020). Generally, IPTG is introduced when cells have reached 0.4- 0.6 OD thus to inducing protein expression. The reduction of the growth rate after induction can be caused by the toxic effect of IPTG and or the metabolic burden imposed on the cells due to heterologous gene expression or protein toxicity (Larentis et al., 2014). Therefore, to assess the suitable concentration of IPTG to be used in this study, different concentrations were utilized to assess protein production by collecting samples at zero and after 24 hours.

The results show that overnight cultivation of protein with all concentrations can increase protein production compare to zero hours. And also the inoculum size (75 µl) overnight show a high yield of protein expression. Several factors some assessed above namely; expression strain, IPTG concentration and temperature of induction, affect protein yield and solubility.

Therefore, it is essential to systematically vary these factors to optimize the results. And also, for toxic proteins, expressing at a higher cell density for a shorter period should be considered since may help to increase production yield. It should be noted that one protein (LigpET30) was used for optimization studies and Ligsv081 was used as a control.

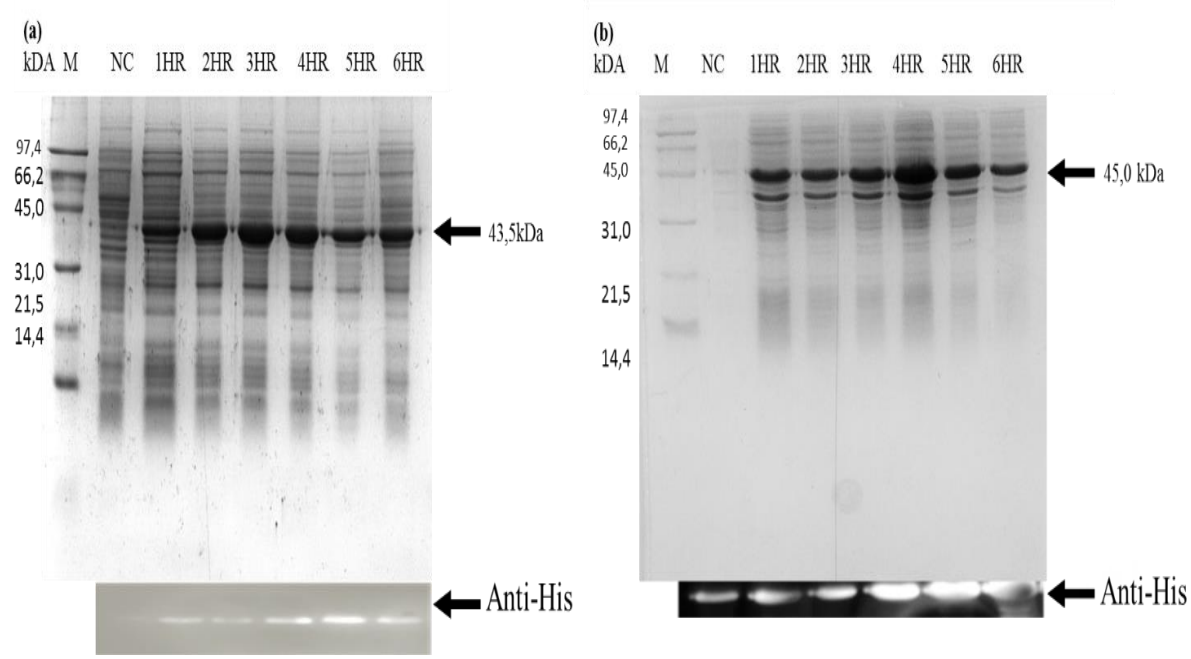

Figure 3. 3: SDS-PAGE (15%) showing IPTG induced expression of (a) Ligsv081 and (b) LigpET30 protein. Lane M; marker, Lane NC- *E. coli* BL21 control cells, Lane 1HR to 6HR represent the time-profile sampling from 1 hour to 6 hours post IPTG -induction.

Expression of DNA ligases, yields soluble proteins with specific molecular weights, these variants were expressed and purified according to methods and materials stated. The large-scale expression of recombinant proteins was done after the optimal expression conditions were investigated in a small scale. Afterward, protein purification was conducted based on the time profile obtained.

It should be noted that Ligsv081 protein was used as a control therefore, we decided not to optimize it.

The two DNA ligase constructs which were obtained as a gift were then designated as Ligsv081 and LigpET30. As demonstrated (Figure **3.3)** all DNA ligases, were efficiently expressed in transformed *E.coli*. The expression vectors pET28a and pET30 were used and accumulated proteins with a molecular weight of 43.5 kDa (Ligsv081) and molecular weight of 45.0 kDa (LigpET30). Further, proteins shown in the above figures were accumulated in a

soluble fraction. Soluble fractions of expressed proteins are vital for the consequent procedure of proteins extraction and purification. Therefore, SDS-PAGE was used to demonstrate proteins in the soluble fractions. The high yield of these protein expressions appeared to be increased when expressed at 37 °C after induction with IPTG, this could be caused by major characteristics of the *E.coli* strain used in this study which is essential for protein expression. Also, both protein accumulation increased at 4HR, followed by the decrease of protein expression due to accumulation in dead cells which caused protein contamination thus protein production decreased.

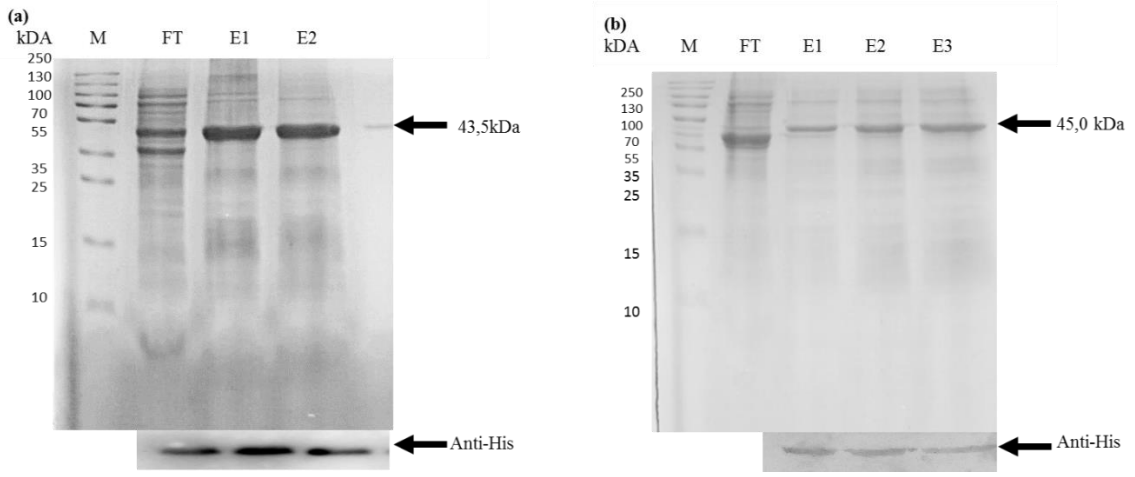

Figure 3. 4: SDS-PAGE (15%) showing purification for DNA Ligsv081 and LigpET30 protein with a buffer containing 250mM imidazole. Lane M; marker (Thermo Science PageRuler plus Pre-stained protein Ladder), Lane FT: Flow-through; Lane E1 & E3 depicts the collected three protein elution samples.

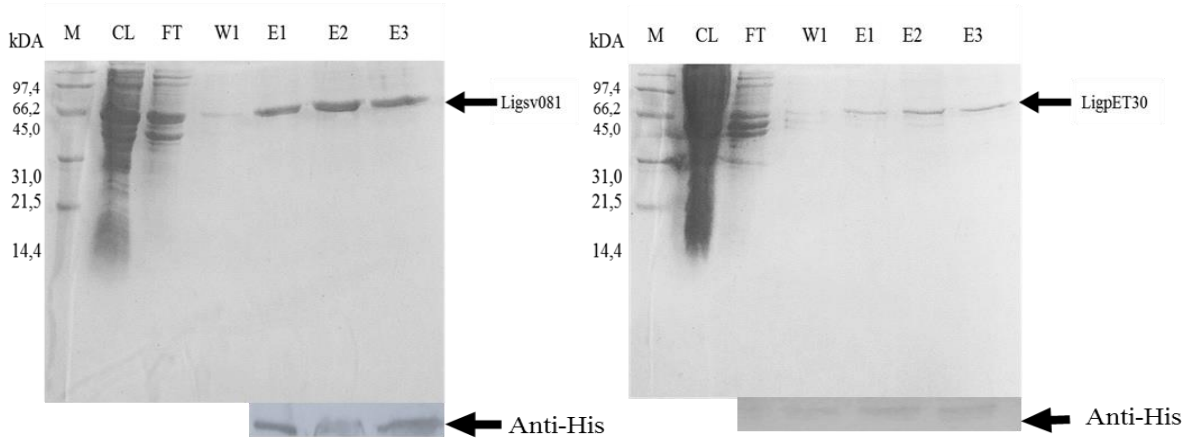

Figure 3. 5: SDS-PAGE (15%) purification for DNA Ligsv081 and LigpET30 protein with buffer containing 200Mm imidazole. Lane M: standard marker, Lane CL: crude lysate, Lane FT: Flow-through; Lane E1- E3 depicts the collected two protein elution samples.

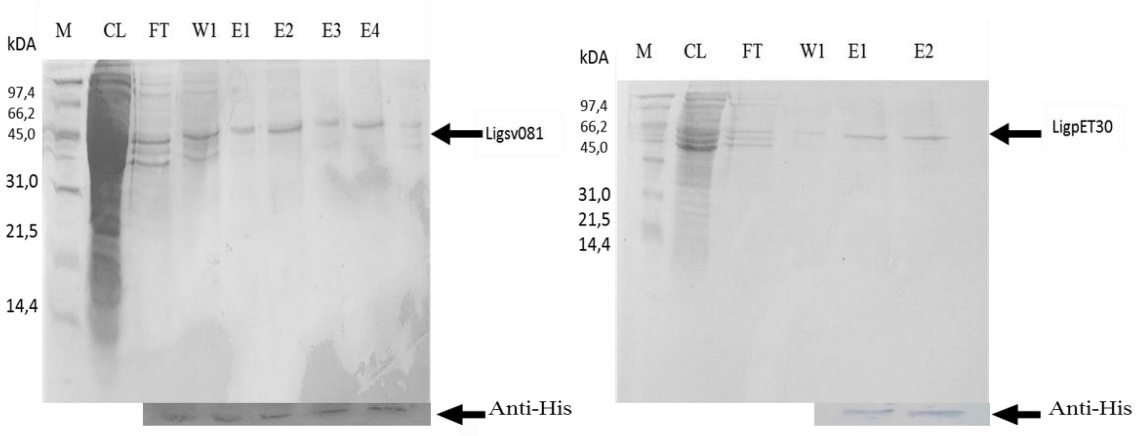

Figure 3. 6: SDS-PAGE (15%) purification for DNA Ligsv081 and LigpET30 protein with a buffer containing 100mM imidazole. Lane M: standard marker, Lane CL: crude lysate, Lane FT: Flow through; Lane E1- E4 depicts the collected two protein elution samples.

These protein constructs had affinity tags (polyHis-tags) added onto their C-terminus so that they could have enhanced protein solubility, stability and expression level, also lower temperatures were considered to maintain their stability. Proteins with low-temperature adaption pose a challenge to work within *in vitro* studies such as maintaining their stability and functional state. Thus, the protein purification process was conducted at 4 °C which ensured a lower proteolysis rate and promoted structural integrity of proteins. Further, obtained proteins were all preserved at a lower temperature (4 °C) for further analysis.

Protein purification essentially utilizes imidazole in wash and elution buffers. Briefly, imidazole has a similar histidine amino acid structure and can compete against histidine binding to Ni-NTA resins. Therefore, higher concentrations of imidazole in the elution buffer play an important role by dissociating proteins with histidine and lower their competitive side to bind to the resin binding sites. Also, imidazole shift binding position of histidine proteins from Ni-NTA resin then target proteins can be removed.

Imidazole concentrations (100 mM to 250 mM) and low pH in the elution buffer easily elute bound proteins after the tags bind immobilized metal through the histidine imidazole ring (Nurjayadi et al., 2019). Figure **3.4** shows protein purification of both proteins. These proteins were eluted with a buffer containing 250 mM imidazole and results of protein purified with lower imidazole concentrations (100 mM to 200 mM) are included (Figure **3.5** & **3.6**) showing few concentrations of purified proteins. The wash buffer had a low concentration of imidazole (25 mM) to remove impurities and reduce the non-specific binding of resin.

Targeted proteins were able to bind on Ni-NTA resin and were successfully purified and verified by the thickness of the protein band in the SDS-PAGE and on Western blot.

3.4.2. Ligation assay

Ligation assay was used to evaluate the functionality of the produced DNA ligase enzymes. Generally, DNA ligase enzymes are essentially involved in the replication and repair of DNA. Therefore, lambda DNA which was primarily digested with *HindIII* restriction enzyme was re-ligated using Ligsv081 and LigpET30 ligation enzymes produced in this study. The ligation procedure was performed in ice-cold (4 °C) followed by the ligation activity conducted at 37 °C for 30 minutes. Figure **3.7** demonstrates successful re-ligation of lambda DNA using produced DNA ligases, therefore, the produced proteins are stable, soluble and biologically active.

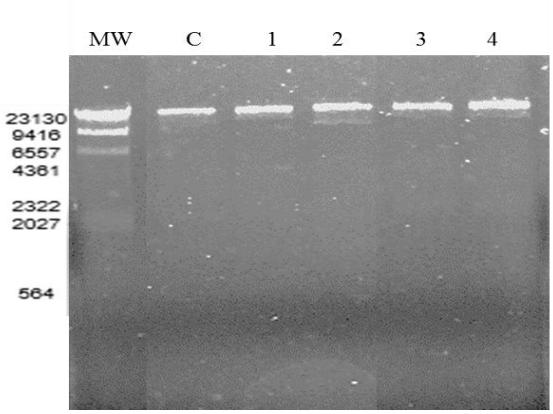

Figure 3. 7: Agarose gel ligation assay (2%). Lane MW-molecular weight- *HindIII* lambda DNA marker; lane C- T4 DNA ligase (control); lane1 & lane2- Ligsvo81 and lane3 & lane 4- LigpET30.

3.4.3. Thermal denaturation

Thermal denaturation is a major hurdle in storage of proteins also during reaction processes. Furthermore, it is essential conducted to evaluate the stability and to know the melting temperature of the proteins, by altering their environmental temperature. Some proteins can maintain their native structures and function properly at increased temperatures. However, most proteins fail to function. Therefore, this section was conducted to assess the stability and shelf life of the two DNA ligases.

The two proteins were stored overnight at room temperature to evaluate effectiveness, stability and shelf life. The ligation assay was then conducted and agarose gel results are shown below in Figure **3.8** and Figure **3.9**.

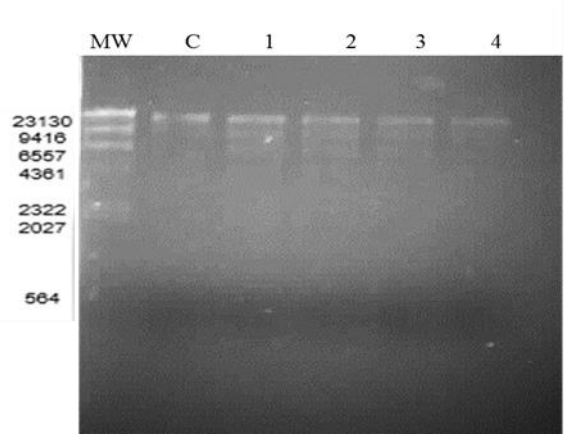

Figure 3. 8: Ligation activity of DNA ligases stored overnight at room temperature. Lane MW (molecular weight)- *HindIII* lambda DNA marker; lane C- T4 DNA ligase (control); lane 1 & lane 2- Ligsvo81 and lane 3 & lane 4- LigpET30.

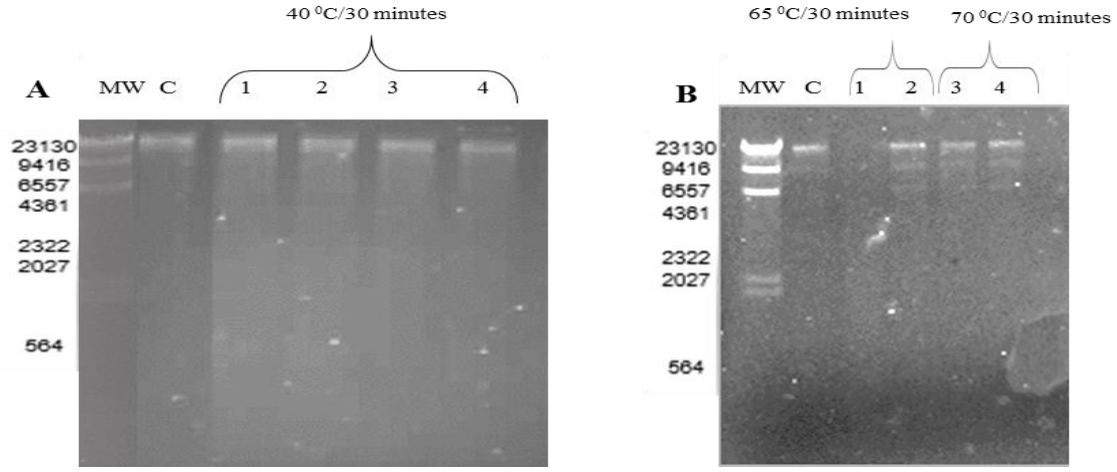

Figure 3. 9: Agarose gel (2%) showing thermal stability of DNA ligases. **A:** Lane MW (molecular weight)- *HindIII* lambda DNA marker; Lane C- T4 DNA ligase (control); Lane 1 & Lane 2- Ligsvo81; Lane 3 & Lane 4- LigpET30. **B**: Lane MW- molecular weight- *HindIII* lambda DNA marker; Lane C- T4 DNA ligase (control); Lane 1 & Lane 2- Ligsvo81; Lane 3 & Lane 4- LigpET30.

The results obtained above show successful and partial ligation of lambda DNA. The storage of DNA ligases at room temperature did not entirely dis-stabilize DNA ligase proteins since they were able to ligate after room storage overnight. Heat denaturing reduced protein functionality and also denatured it and resulted in the formation of smears on other lanes, however, some lanes show successful ligation. It is important to keep protein in their native environment conditions to maintain three-dimensional structures. Also, changing their microenvironment conditions should be avoided since might generate disruption in the protein interactions, causing denaturation of protein leading to protein unfolding and inactivation. Various stress conditions such as thermal denaturation of protein interfere with the folded conformation of the protein and hydrophobic residues become exposed to water, resulting in a thermodynamically unfavorable state. Therefore, the proper temperature to preserve these proteins that should be considered is 4 °C.

3.5. CONCLUSION

Most recombinant proteins in the market were produced in *E.coli* which is still the preferred system to produce protein on a smaller or larger scale. Also, *E.coli* showed the ability to assist protein optimization in this study. The protein optimization can be extended by introducing other factors such as amino acids during protein expression. In this study, we were able to produce functional proteins and assess their biological activities. Proteins are widely utilized most in industries and research fields as catalysts or enzymes therefore, it is important to produce soluble proteins with preserved biological activities and stability. Furthermore, the storage buffer and temperature determine the formation of aggregates and the ability of proteins to perform successfully, therefore it is important to store proteins in an environment that favors their quality, characteristics and native structure.

REFERENCES

Bulletin, A. P., & Khan, K. H. (2013). Gene Expression in Mammalian Cells and its Applications. *Advanced Pharmaceutical Bulletin*, *3*(2), 257–263. doi: 10.5681/apb.2013.042.

Burnett, M. J. B., & Burnett, A. C. (2020). Therapeutic recombinant protein production in plants: Challenges and opportunities. *Plants, People, Planet*, *2*(2), 121–132. https://doi.org/10.1002/ppp3.10073

Derewenda, Z. S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. *Methods*, *34*, 354–363. https://doi.org/10.1016/j.ymeth.2004.03.024

Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A., & Hossain, M. N. (2013). Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant *E. coli*. *ISRN Biotechnology*, *2013*(3), 1–7. https://doi.org/10.5402/2013/590587

Gomes, L., Monteiro, G., & Mergulh, F. (2020). The Impact of IPTG Induction on Plasmid Stability and Heterologous Protein Expression by Escherichia coli Biofilms. *Int. J. Mol. Sci*, 21, 576; doi:10.3390/ijms21020576.

Hefti, M. H., Toorn, C. J. G. V. V. Der, Dixon, R., & Vervoort, J. (2001). A Novel Purification Method for Histidine-Tagged Proteins Containing a Thrombin Cleavage Site. *Analytical Biochemistry*, 185, 180–185. https://doi.org/10.1006/abio.2001.5214

Hjerde, E., Maguren, A., Rzoska-Smith, E., Kirby, B., & Williamson, A. (2020). DNA ligases of Prochlorococcus marinus: An evolutionary exception to the rules of replication. *BioRxiv*, 1–7. https://doi.org/10.1101/2020.05.11.089284

Jia, B., & Jeon, C. O. (2016). High-throughput recombinant protein expression in *Escherichia coli*: Current status and future perspectives. *Open Biology*, *6*(8). https://doi.org/10.1098/rsob.160196

Karbalaei, M. (2020). Pichia pastoris : A highly successful expression system for optimal synthesis of heterologous proteins. *J Cell Physiol*,1–15. https://doi.org/10.1002/jcp.29583

Karbalaei, M., Rezaee, S. A., & Farsiani, H. (2020). Pichia pastoris: A highly successful

expression system for optimal synthesis of heterologous proteins. In *Journal of Cellular Physiology*, 235(9), 5867–5881. Wiley-Liss Inc. https://doi.org/10.1002/jcp.29583

Knecht, S., Ricklin, D., Eberle, A. N., & Ernst, B. (2009). Oligohis-tags : mechanisms of binding to Ni 2 R -NTA surfaces. *J. Mol. Recognit*, 22: 270–279. https://doi.org/10.1002/jmr.941

Konczal, J., & Gray, C. H. (2017). Streamlining workflow and automation to accelerate laboratory scale protein production. *Protein Expression and Purification*, *133*, 160–169. https://doi.org/10.1016/j.pep.2017.03.016

Labrou, N. E. (2014). Protein Purification : An Overview Chapter 1 Protein Purification : An Overview. *Methods in Molecular Biology*, 1129, https://doi.org/10.1007/978-1-62703-977-2.

Larentis, A. L., Fabiana, J., Quintal, M., Esteves, S., Vareschini, D. T., Vicente, F., Almeida, R. De, Galvão, M., Galler, R., & Medeiros, M. A. (2014). Evaluation of pre-induction temperature , cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E . coli using shaking flasks and microbioreactor. *BMC Research Notes.* 1–13.. doi:10.1186/1756-0500-7-671

Ma, Y., Lee, C., & Park, J. (2020). Strategies for Optimizing the Production of Proteins and Peptides with Multiple Disulfide Bonds. Antibiotics, 9, 541, doi:10.3390/antibiotics9090541.

Nurjayadi, M., Afrizal, R., Hardianto, D., & Agustini, K. (2019). Variations of binding, washing, and concentration of imidazole on purification of recombinant Fim-C Protein Salmonella typhi with Ni-NTA Resin. *Journal of Physics: Conference Series*, *1402*(5). https://doi.org/10.1088/1742-6596/1402/5/055055

Samai, P., & Shuman, S. (2011). Structure-function analysis of the OB and latch domains of Chlorella virus DNA ligase. *Journal of Biological Chemistry*, *286*(25), 22642–22652. https://doi.org/10.1074/jbc.M111.245399

Tripathi, N. K. (2016). *Production and Purification of Recombinant Proteins from Escherichia coli*. *3*, 116–133. https://doi.org/10.1002/cben.201600002

Wilkinson, A., Day, J., & Bowater, R. (2001). Bacterial DNA ligases. *Molecular Microbiology*, *40*(6), 1241–1248. https://doi.org/10.1046/j.1365-2958.2001.02479.x

Wiseman, D. N., Otchere, A., Patel, J. H., Uddin, R., Pollock, N. L., Routledge, S. J., Rothnie, A. J., Slack, C., Poyner, D. R., Bill, R. M., & Goddard, A. D. (2020). Expression and purification of recombinant G protein-coupled receptors: A review. *Biomolecules*, 10, 1262; doi:10.3390/biom10091262.

Young, C. L., Britton, Z. T., & Robinson, A. S. (2012). Recombinant protein expression and purification : A comprehensive review of affinity tags and microbial applications. *Biotechnol. J,* 7, 620–634. https://doi.org/10.1002/biot.201100155.

CHAPTER 4: EVALUATING THE FUNCTIONAL STABILITY OF SITE-SPECIFIC PROTEIN PEGYLATION ON RECOMBINANT DNA LIGASE

4.1. ABSTRACT

Reagent proteins such as DNA ligases play a central role in the global reagents market. DNA ligases are routinely used and are vital in academic and science research environments. Their major functions include sealing nicks by linking the 5'-phosphorylated end to a 3'-hydroxyl end on the phosphodiester backbone of DNA, utilizing ATP or NADP molecules as an energy source. The current study sought to investigate the role of protein PEGylation on purified recombinant DNA ligases. We expressed recombinant DNA ligases using *E. coli* and subsequently purified using affinity chromatography. The produced proteins were conjugated to site specific PEGylation or nonspecific PEGylation. FTIR and UV-VIS spectroscopy were used to analyze secondary structures of the PEG conjugated DNA ligases. Thermal stability assays were employed to assess protein stability in the presence of PEG. In this study, both recombinant DNA ligases were successfully expressed and purified as homogenous proteins. Site-specific PEGylation promoted secondary structure stability, enhanced ligase activity, and reduced the formation of protein aggregates. Therefore, site-specific PEGylation can potentially be explored as a means to enhance stability of recombinant proteins over time during storage.

Keywords: DNA ligases; site-specific PEGylation; recombinant proteins; thermal degradation; protein PEGylation; PEG conjugation

4.2. INTRODUCTION

DNA ligases are regarded as the workhorse of molecular biology, playing critical roles in DNA cloning experiments. DNA ligases can be classified as ATP-dependent or NAD-dependent (Hjerde et al., 2020b). DNA ligases consist of two domains, namely, the nucleotidyltransferase domain and an OB domain which are common to all DNA ligases and are involved in nucleic acid binding (Samai & Shuman, 2011). These domains are mainly made up of amino acids which are essential within the ligation pathway, for example, such as lysine which is involved in the first step of the ligation pathway, where AMP covalently links to lysine (Wilkinson et al., 2001). Furthermore, amino acids such as Phe-75, Met-83, Thr-84, Lys-173, and Arg-176 form NTase domain side chains associated with the DNA-binding

surface and aid in forming the ligase clamp around the nicked DNA (Samai & Shuman, 2011). The first step of the ligation pathway is where a conserved lysine residue of ligase enzyme experiences covalent adenylation by the nucleotide donor, which releases pyrophosphate or nicotinamide mononucleotide, and ligase-(lysyl-N)-AMP linkage is formed. The second step is where nicked DNA is activated for nucleophilic attack by the AMP moiety transfer. The third step involves new phosphodiester bond formation on the backbone of DNA where AMP is released. Conserved nucleotidyl transferase motifs have been observed to play an essential role in these chemical steps (Hjerde et al., 2020a). However, frequently used reagent proteins, such as DNA ligases, are prone to protein degradation and reduced shelf life (Pothukuchy et al., 2018).

PEGylation is the most widely used and functional post-production modification technique, defined as the covalent conjugation of polyethylene glycol (PEG) to protein (Ramirez-Paz et al., 2018). The covalent attachment of PEG onto proteins has been reported to enhance physicochemical properties and produce PEG conjugated proteins without compromising protein secondary structures (González-Valdez et al., 2012). Furthermore, PEGylation strategies have positive attributes such as reducing protein aggregation because of steric repulsion present on the surface of PEGylated products, increasing thermal and long-term stability (Kusterle et al., 2008; Ramirez-Paz et al., 2018). It has also been known to alter *in vitro* activities, toxicity, and immunogenicity reduction (Zhao et al., 2015).

Proteins have attachment sites such as -NH_2, -COOH, -SH, and -SS- available for conjugation with the functional group on a PEG (Kusterle et al., 2008). Site-specific PEGylation conjugation is the most desirable form of PEGylation, as PEG will be confined to one specific site leading to the overall retainment of protein structure and bioactivity conservation (Ramirez-Paz et al., 2018). Recently, it was reported that PEGylation also increases the half-life of proteins by masking the protein surface to prevent antibodies, proteases, and antigen processing in cells (Kumari et al., 2020). This improves the solubility and biological distribution of proteins (Veronese & Pasut, 2005). Furthermore, site-specific PEGylation can be attached to one protein molecule, and there is no active site shielding of the protein in more than one chain of PEG, which causes protein deactivation (Ramirez-Paz et al., 2018). In addition, protein modification through site-specific PEGylation produces homogenous and well-defined conjugates without compromising the overall biological activity of the conjugate (Sadiki et al., 2020). Hence, this study aimed to produce biologically active DNA ligases and extend their shelf life. DNA ligases were used in this study as a

model because of their biological importance for all organisms and their central role in the global reagent market.

4.3. MATERIALS AND METHOD

4.3.1. Protein Expression and Purification

Segobola and colleagues (2018) donated *E. coli* BL21 cells clones expressing two DNA ligases designated as Ligsv081 and LigpET30 (Segobola et al., 2018). Selection plates were individually streaked with *E. coli* BL21 cells harbouring the expression plasmids which were then incubated at 37 °C on YT media (50 µg/ml kanamycin) overnight. Afterwards, a starter culture was produced by selecting a single colony which was re-suspended in liquid YT media culture with 50 µg/ml kanamycin. A dilution factor of 1:100 was used to inoculate a starter culture into expression media with 50 µg/ml kanamycin which was incubated with shaking (200 rpm) at 37 °C while waiting for OD600 to reach 0.4 to 0.6. After this, protein expression was induced by 100 µM IPTG for 6 hours at 37 °C followed by recovery through centrifugation at 13000 *g* for 10 minutes using a JA14 rotor (Beckman, Optima L-100XP). Then, a lysis buffer was used to re-suspend the pellets thus release intracellular proteins. All proteins investigated in this study were expressed with a N-terminal polyhistidine tag (His-tag) to facilitate affinity chromatography purification, similar to previous studies (Opoku et al., 2019; Pooe et al., 2017; Zininga et al., 2015, 2017). The HisPur™ Ni-NTA Resin column was used in the purification of the His-tagged proteins following the manufacturer's protocol. Briefly, LEW buffer was used to equilibrate the column followed by loading of 15 ml soluble clarified lysate made up of targeted His-tag proteins. Thereafter, three column volumes (CV) of LEW buffer were used to wash off unbound proteins. The targeted protein was eluted with elution buffer and confirmed with 15% SDS-PAGE. Western blot was used to verify the purified protein (Opoku et al., 2019; Pooe et al., 2017).

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse the soluble fractions of protein and visualization of proteins was done using the coomassie stain and confirmed by western blot analysis. Polyvinylidene difluoride (PVDF) membrane was used to transfer proteins and the membrane was incubated in 5% skim milk in Tris-buffered saline with tween 20 (TBST) for one hour to block non-specific binding proteins. Samples were incubated with anti-HIS-tag IgG conjugated with horse-radish peroxidase (HRP from Sigma Germany; diluted 1:1000 in TBST) for one hour at 4 °C, then

washed for three times. The chemiluminescence (ECL) kit method (Thermo Fisher scientific, SA) was employed to detect recombinant DNA ligases as per the manufacturer's guidelines.

4.3.2. Preparation of His-Tag-Specific PEGylated DNA Ligases

Already prepared mPEG NTA and poly(ethylene glycol), Mv- 8, 000 was used. An amount of 20 mg/ml was dissolved into a phosphate buffer (5 mM sodium phosphate, 15 mM sodium chloride and 1 mM EDTA, pH 8). Purified DNA ligase protein (500 µg/ml) was added to the buffer (50 mM sodium phosphate, 150 mM sodium chloride and 10 mM EDTA, with controlled pH 6.7). Then 60 µl of PEG solution was mixed with 400 µl of buffer containing the protein solution at room temperature for 25 minutes. The two ligases were independently left to conjugate to either mPEG-Nitrilotriacetic acid or polyethyleneglycol 8 000 at 25 °C for three hours followed by buffer exchange through dialysis tubing. Protein post-PEGylation purification was performed using a HisPurTM Ni-NTA Resin column (ThermoFisher) following the manufacturer's protocol. Briefly, LEW buffer (20 mM NaH_2PO_4 and 0.4 mM NaCl) was used to equilibrate the column followed by loading of a PEG and protein solution of the targeted PEGylated proteins. Thereafter, three column volumes (CV) of LEW buffer were used to wash off unbound proteins. The targeted protein was eluted with elution buffer (20mM NaH_2PO_4, 0.5 NaCl and 300 mM imidazole). Obtained conjugates of proteins were then stored at 4 °C before analysis.

4.3.3. Measurement of Particle Size

All samples of DNA ligase proteins were diluted in a ratio of 1:1000 with ddH_2O followed by measurement of particle size and zeta potential in a Nanosight NS-500 (Malvern Instruments, Worcestershire, UK) at 25 °C. Particle sizes were all calculated by NTA software v3.0 using the Stokes-Einstein equation and the Smoluchowski equation was used to estimate the zeta potential (D. Singh & Singh, 2021).

4.3.4. Fourier Transform Infrared (FTIR) Spectroscopy Protein Analysis

For FTIR measurements, 2 µl of Ligsv081, Ligsv081-non-specific-PEG, Ligsv081 site-specific-PEG, LigpET30, LigpET30 non-specific-PEG and LigpET30 site-specific-PEG each were diluted at 2.7 mg/ml in 10 mM sodium acetate buffer pH 7.5, then deposited on the

diamond plate of the single reflection of a universal ATR Sampling Accessory. The samples were cooled to room temperature before collecting the spectra in the region of 400 cm^{-1} to 4000 cm^{-1} (Ndlela et al., 2020).

4.3.5. Evaluating Protein Thermal Stability

An external water bath (connected to the UV-VIS) was used to change the temperature of the cell holder in contact with the specific cuvette used to read the proteins. The water bath temperature range was 25 °C to 60 °C. The temperature in which the protein unfolding usually takes place is in the range 45 °C to 60 °C, therefore in this range temperature increase was controlled with 1 °C increments.

4.3.6. Cloning assay

Blue-white screening method was used to assess DNA ligase activities in a vector with insert during ligation assay. After ligation, the transformed cells were recovered in LB medium for one hour at 37 ^{o}C, followed by spread on LB-agar plate containing kanamycin, IPTG and X-gal. White colonies showing successful ligation were counted on plates after incubation at 37 ^{o}C for 16 hours.

4.4. RESULTS AND DISCUSSION

4.4.1. DNA ligase purification

The first conjugation of protein and a polymer involved PEG and bovine serum albumin (BSA), which caused a positive explosion of protein-PEG conjugation populations and opened a new area of interest in protein post-modification strategies (Zhao et al., 2015). Our work extends this field. Our work aimed to extend the shelf life of the produced recombinant DNA ligase and evaluated the biophysical characterisation of PEG conjugated and unmodified DNA ligases. Both N-terminal ligases were produced with a polyHis-tag positioned at either N-terminus, thus impacting the protein purification processes which are usually conducted under native and denaturing conditions, simultaneously. The purified proteins were later conjugated to PEG and stored at 4 °C for further analysis (Figure **4.1**).

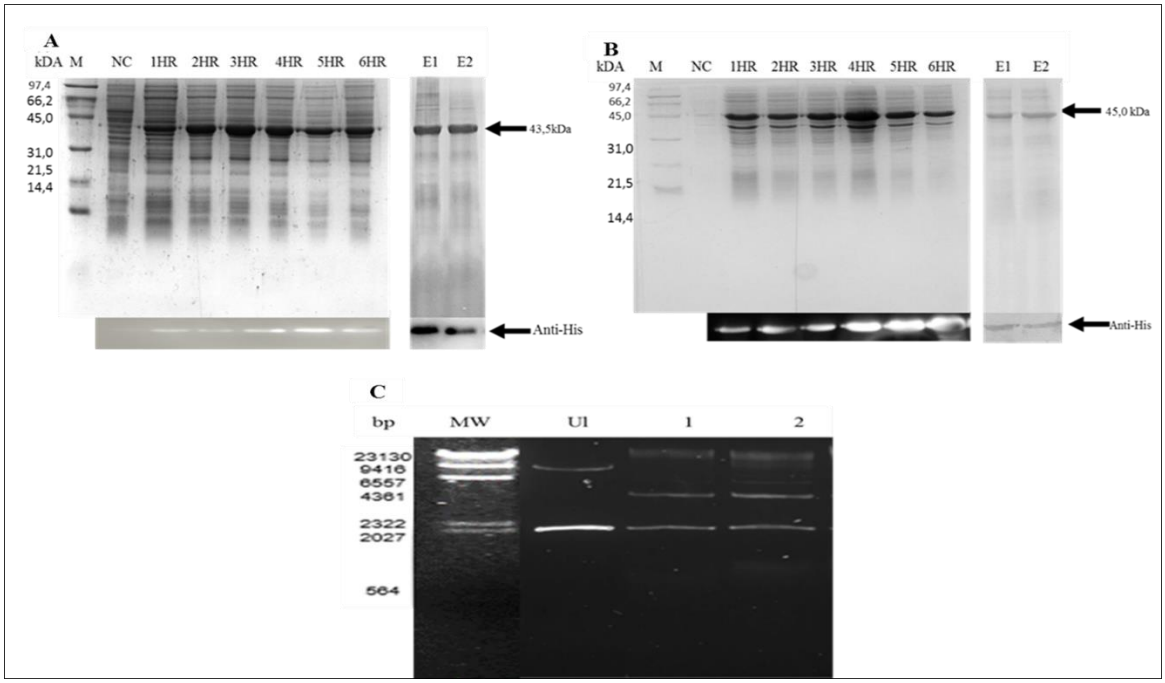

Figure 4. 1: A and **B**: SDS-PAGE showing IPTG induced expression and purification for DNA Ligsv081 and LigpET30 proteins. Lane M- marker (Thermo Science PageRuler plus Pre-stained protein Ladder), Lane NC- *E. coli* BL21 control cells, Lanes 1HR to 6HR represent the time profile sampling from 1 hour to 6 hours post-IPTG induction. Lanes E1 and E2 depict the collected two protein elution samples. **C:** Ligation of blunt ended vector and insert using Ligsv081 and LigpET30 DNA ligases. Lane MW- lambda DNA *HindIII* digested, Lane Ul- un-ligated vector to insert and Lanes 1 and 2- ligated products.

4.4.2. Fourier Transform Infrared (FTIR) Spectroscopic Studies

FTIR spectra were used to evaluate the role of PEG in relation to the DNA ligases' protein secondary structures which are known to determine biological activities and stability of every protein. FTIR utilizes two known absorption bands (amide I and amide II) to determine the biophysical characteristics of a protein's secondary structure and they are located at the wavelengths of 1700 cm^{-1} to 1600 cm^{-1} and 1600 cm^{-1} to 1500 cm^{-1}, respectively. The protein secondary structure is mainly observed by the presence of C=O stretching vibrations of the peptide bonds, which are modulated by the secondary structure (α-helix, β-sheet, β-turn and β- antiparallel) attached, with the in-phase bending of the N-H bond and stretching of the C-N bond. Amide II which is more complex and derived from in-plane N-H bending and the C-N stretching vibration. Amide I is also sensitive but less utilised for protein structure.

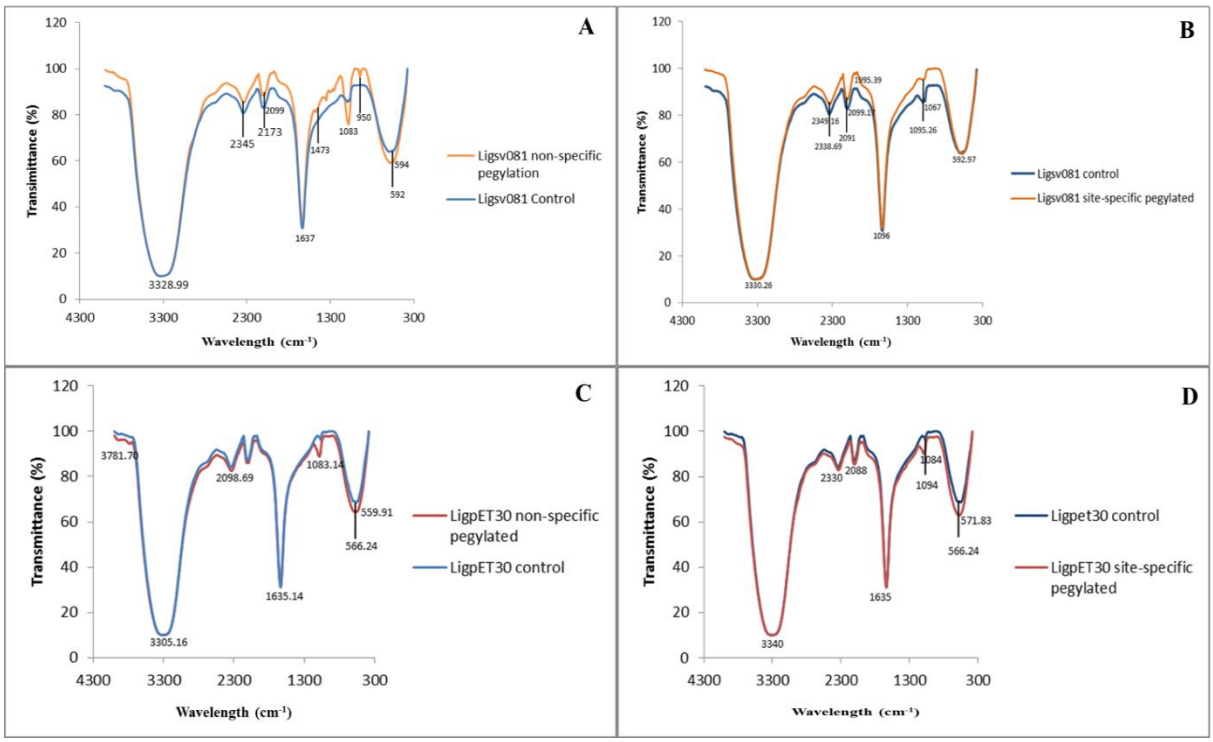

Figure 4. 2: FTIR spectra showing PEGylated and non- PEGylated proteins. FTIR analysis results of DNA ligases at 400 cm^{-1} to 4000 cm^{-1}. **(A)** Ligsv081 control & Ligsv081 non-specific PEGylated. (**B**) Ligsv081 control & Ligsv081 site-specific PEGylated. (**C)** LigpET30 control & LigpET30 non-specific PEGylated. LigpET30 control & LigpET30 site-specific PEGylated (**D**).

FTIR spectrum of DNA ligase Ligsv081 and LigpET30 unmodified and DNA ligase Ligsv081 and LigpET30 derivatives (Figure **4.2**) were obtained before and after mono-PEGylation and non-specific PEGylation. The absorption peak at 3570 cm^{-1} to 3200 cm^{-1} assigned to the hydroxyl group (OH) of the amino acid present in these proteins decreased after PEG conjugations. Band absorption at 2 330.02 cm^{-1}, 2 329.79 cm^{-1} and 2 330.14 cm^{-1} belongs to NH of amino acids which is also present in these proteins. Furthermore, peaks at 2200 cm^{-1} to 2000 cm^{-1} showed that there is likely C-N- stretching in protein or DNA ligases. Band absorption at 1090 cm^{-1} to 1020 cm^{-1} is assigned to primary amine of DNA ligases. Band absorption at 1680 cm^{-1} to 1620 cm^{-1} is assigned to the aromatic structure of amino acid/s. Peak absorption at 1140 cm^{-1} to 1070 cm^{-1} is assigned to the stretching vibrations of C-O-C- found in PEG. Also, the hydroxyl group of PEG has an absorption peak at 3781.70 cm^{-1} (Salomane et al., 2021). The peaks shown in the spectra validated protein stability post PEGylation.

The FTIR spectra shows distinct peak absorption of methylene –C-H- bending at 1351.22 cm^{-1} which belongs to PEG. The absorption band at 1301.62 cm^{-1} is assigned to primary or

secondary OH- in plane bend (Figure **4.2**). The peak absorption at 950.71 cm^{-1} shows -CH_2- out of plane bend. The presence of PEG increased the transmittance percentage due to weak bond formation due to the PEGylation process.

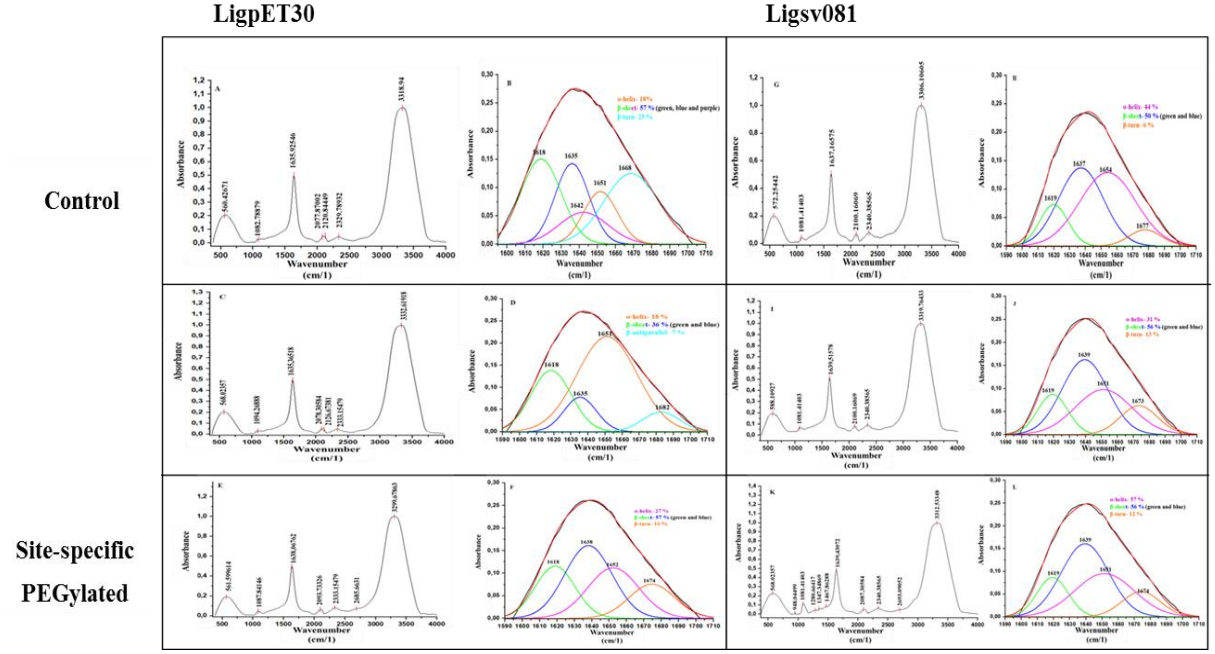

Figure 4.3. FTIR spectra showing PEGylated and non- PEGylated proteins. FTIR analysis results of DNA ligases at 400 cm^{-1} to 4000 cm^{-1}. **(A)** LigPET30 control, site-specific PEGylated & Ligsv081 non-specific PEGylated. (**A-F**). Ligsv081 control, site-specific PEGylated & LigpET30 non-specific PEGylated (**G-L**). Amide I at 1600 cm^{-1} to 1700 cm^{-1} (**B**, **D**, **F**, **H**, **J** and **L**) and amide band peaks at 400 cm^{-1} to 4000 cm^{-1} representing protein secondary structure (**A**, **C**, **E**, **G**, **I** and **K**).

The FTIR readings of the secondary structure proteins (Figure **4.3**) contributing to the amide I band are as follow: α-helix (1660 cm^{-1} to 1650 cm^{-1}), β-sheet (1637 cm^{-1} to 1614 cm^{-1}), β-turn (1678 cm^{-1} to 1670 cm^{-1}) and β- antiparallel (1691 cm^{-1} to 1680 cm^{-1}). Unmodified DNA ligase (LigpET30) had a spectral readings at 1619 cm^{-1} and 1637 cm^{-1} for β-sheet, 1654 cm^{-1} for α-helix, and 1677 cm^{-1} for β-turn. Due to PEG conjugation with LigpET30 the spectra absorption shifted to 1639 cm^{-1} (LigpET30 non-specific PEGylated and site-specific PEGylated) for β-sheet. Furthermore, the spectra reading for α-helix and β-turn decreased to 1651 cm^{-1} and 1674 cm^{-1}, respectively.

The amide region of unmodified Ligsv081 had readings at 1650 cm^{-1} for α-helix and 1618 cm^{-1} and 1635 cm^{-1} for β-sheet. The PEG conjugation shifted band readings of Ligsv081 to 1651 cm^{-1} and 1652 cm^{-1} for α-helix. Furthermore, for β-sheet the band reading increased to 1638 cm^{-1} for site-specific PEGylated Ligsv081. However, for non-specific PEGylated Ligsv081 there was no change in band reading. The β-turn for Ligsv081 site-specific

PEGylated was observed at 1674 cm^{-1} and β-antiparallel reading at 1682 cm^{-1} was observed after non-specific PEGylation.

The FTIR spectra were able to identify absorption bands of chemical bonds and functional groups of amino acids present in these proteins and also showed interaction or bonding of DNA ligases with PEG (weak bonding). FTIR also demonstrated a stable physical state or secondary structure of proteins which is important in maintaining functional optimal folded form and good quality of proteins. Moreover, this is essential to store proteins so they can retain their original structural integrity and activity for longer periods.

4.4.3. Assessing Functionality of PEG-conjugated DNA ligases

Ligation efficacy studies were conducted to investigate the functional status of the newly PEGylated proteins. Figure **4.4** represents the ligation products using the purified DNA ligases. The number of white colonies obtained after ligation of the vector to insert the DNA ligases and PEG conjugated DNA ligases produced in this study. The PEG conjugated proteins showed an increased number of colonies, compared to un-conjugated DNA ligases. Furthermore, site-specific PEGylation produced more colonies than non-specific PEGylation (Figure **4.4**) which possibly translates to increased numbers of viable cells and functional activity of DNA ligase proteins.

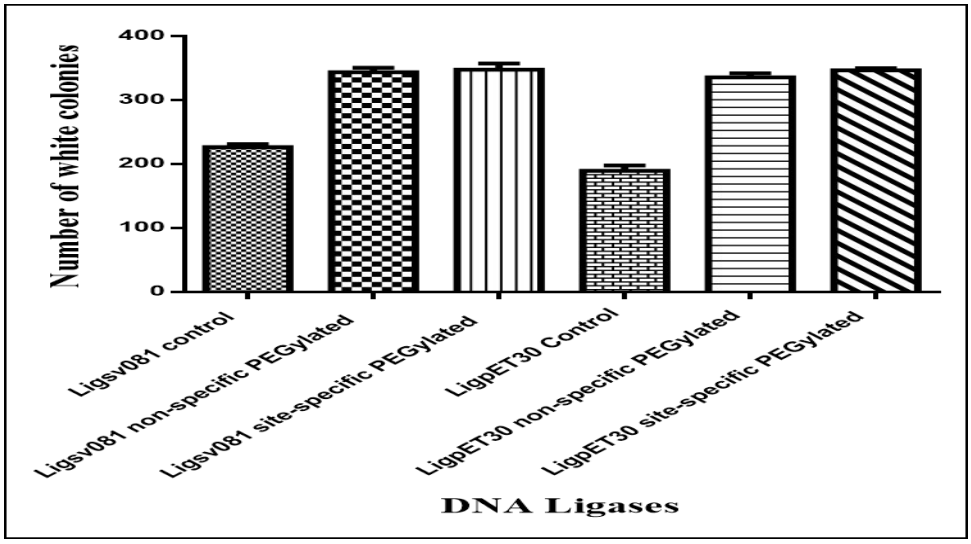

Figure 4.4: Cloning assay. Blunt-ended insert and standard vector were ligated together and transformed into *E. coli* Bl21. Colony counts are means of independent triplicates.

The number of white colonies recorded, (Figure **4.4**) were observed and counted on plates after ligation of the insert to the standard vector. Transformed cells were recovered on LB medium (at 37 °C for one hour) followed by plating on LB-agar plates. All DNA ligases and their derivatives were able to ligate a vector to insert and the produced variable colony count obtained had the statistical significance of $p < 0.05$.

4.4.4. Thermal denaturation studies

The results shown were performed on a UV-VIS Cary spectrophotometer (Figure **4.5**). DNA ligases and PEG conjugated DNA ligases were heated from 25 °C to 60 °C. The temperature in which the protein unfolding usually takes place is in the range 45 °C to 60 °C, therefore in this range temperature increase was controlled with 1 °C increments. As the temperature increased, the OD also increased. Temperatures higher than the melting temperature of unfolded proteins generally showed loss of ordered native structure, where unfolding and breakage of covalent bonds present on proteins occurs until proteins become unstable and completely denatured (Fatima & Khan, 2007). For all the tested samples, instability and degradation increased between temperatures of 50 °C and 60 °C. Thus, 50 °C represents PEG modified proteins' melting point. UV-VIS spectrometry was conducted to understand conformational changes of the proteins' secondary structure in the presence of PEG. Mostly, proteins in solution absorb ultraviolet light at 280 nm which is called absorbance maxima (Grigoryan et al., 2020). Additionally, absorbance maxima is influenced by aromatic amino acids, namely, tyrosine, tryptophan and phenylalanine that are mainly present in the secondary structures of proteins. The PEG conjugates in this study demonstrated potential to improve and increase of the shelf-life. Figure **4.5c** (**A**) and **4.5** (**B**) represent the observed UV-visible spectra at 60 °C for PEGylated and non-PEGylated DNA ligases at secondary structure level site-specific PEGylated proteins, showing that PEGylation marginally reduced the protein degradation rate and conferred thermal-degradation protection (Figure **4.5**).

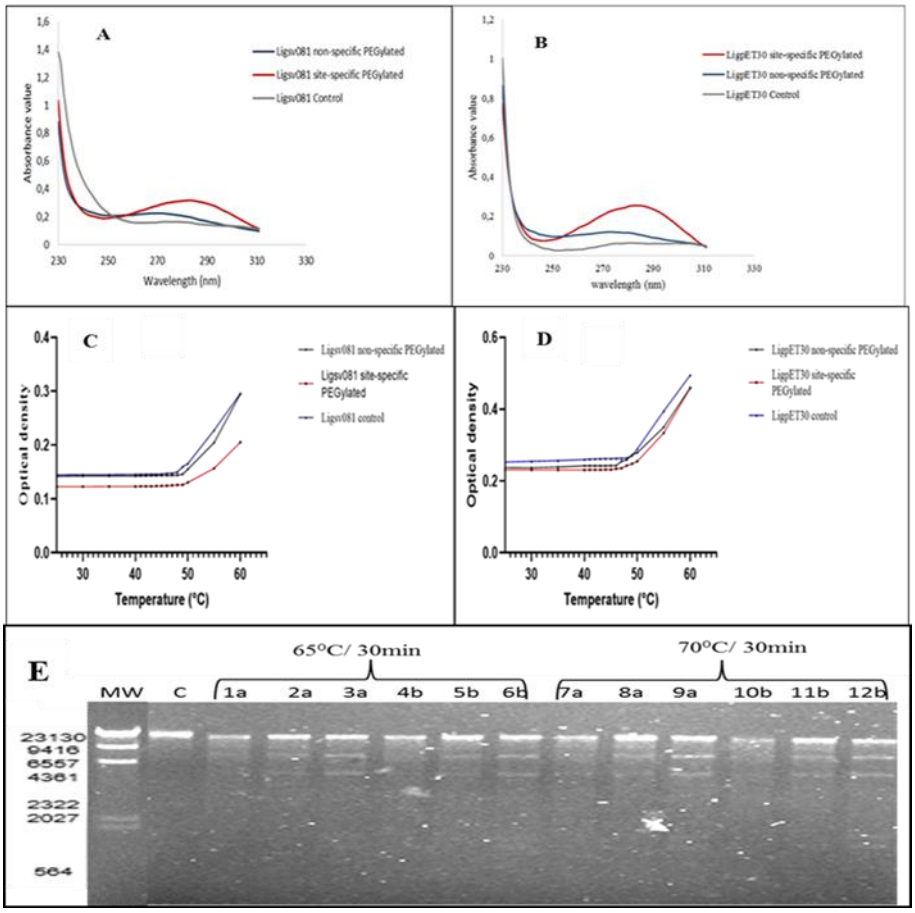

Figure 4.5: Thermal denaturation assays of PEG conjugation and unconjugated studies. Optical density assay for Ligsv081 DNA ligase protein & Ligsv081 PEG conjugate DNA ligase proteins (**A**). LigpET30 DNA ligase protein and LigpET30 PEG conjugate DNA ligase proteins (**B**). UV-visible spectra analysis (**C** and **D**). Ligation activity of thermal denatured DNA ligases (**E**). Lane MW- molecular weight; C- control (T4 DNA ligase); Lane 1a- Ligsv081 control, Lane 2a- Ligsv081 non-specific PEGylated, Lane 3a Ligsv081 site-specific PEGylated, Lane 4b- LigpET30 control, Lane 5b- LigpET30 non-specific PEGylated, Lane 6b- LigpET30 site-specific PEGylated, Lane 7a- Ligsv081 control, Lane 8a- Ligsv081 non-specific PEGylated, Lane 9a- Ligsv081 site-specific PEGylated, Lane 10b- LigpET30 control, Lane 11b- LigpET30 non-specific PEGylated, Lane 12b- LigpET30 site-specific PEGylated.

Figure **4.5** (**E**) demonstrates successful ligation activities of unmodified DNA ligases and PEG conjugated DNA ligase derivatives after they were all subjected to high temperatures (65 °C and 70 °C for 30 minutes) to investigate their thermal denaturation. Several factors including temperature or heat affect the stability and folding or unfolding processes of proteins (Grigoryan et al., 2020). Unfolded and unstable proteins have a reduced biological activity in Figure **4.5** (**E**), with partial ligation of lambda DNA caused by denaturation of DNA ligases before they were used to perform ligation reaction. Additionally, reduced functionality occurred due to the ability of temperature to unfold and disrupt the covalent bonding present in these proteins, thus altering their functionality. In this study, shelf life of all proteins was assessed through storage or incubation at 4 °C for seven days and fourteen

days. After seven days of incubation, proteins were still stable and aggregation was prevented or reduced by PEG. However, fourteen days of incubation at 4 °C led to an increase in particle size. This could be due to the buffer used to store the site-specific PEGylated proteins. Nevertheless, site-specific PEGylation conferred protection to the purified proteins during heating.

Table 4.2: Particle size of un-modified and PEG conjugated DNA ligases

	Particle size (nm)	
DNA ligases	**7-days incubation**	**14-days incubation**
Ligsv081 control	143.3 +/- 45.9	229.9 +/- 23.1
Ligsv081 non-specific PEGylated	112.6 +/- 7.6	721.1 +/- 24.0
Ligsv081 site-specific PEGylated	112.5 +/- 3.1	465.9 +/- 19.5
LigpET30 control	161.2 +/- 4.1	247.9 +/- 17.0
LigpET30 non-specific PEGylated	89.4 +/- 6.0	595.3 +/- 18.3
LigpET30 site-specific PEGylated	96.9 +/- 1.3	210.7 +/- 10.8

Table **4.2** shows data regarding the proteins (unmodified and PEG conjugates) stored in a refrigerator at 4 °C for approximately seven days and fourteen days before samples were analysed using the Nano-Sight NTA technique. The protein storage was conducted to assess the effectiveness of PEG towards the shelf life of protein. The particle size increase over the 14 days incubation had no statistical significance ($p = 0.52$). The particle size increased due to the formation of aggregates related to surface charge and hydrophobicity change that developed after long storage. Furthermore, the formation of aggregates might have caused by the pH conditions of the solution. pH is commonly known to affect density, surface charge distribution, the degree of collapse of protein structure and protein molecule interaction in the solution. Another possible reason could be that other amino acids in the produced proteins are highly soluble than others, thus different concentrations are required for the development of aggregates.

Generally, the increase of PEGylation degree increases protein shell life by increasing its hydrodynamic radius. Also, it is essential to know the degree of PEGylation contributing

towards hydrodynamic volume increase of a protein. Therefore, in this study, it is possible that the PEG degree decreased with storage. Thus, PEG molecules were inhibited and did not interact with hydrophobic patches of protein, resulting in the formation of aggregates. The modified and unmodified proteins produced were functionally stable and produced a higher number of viable colony counts. Furthermore, PEGylation appeared to assisted proteins to retain their native structure.

CONCLUSION

Taken together in our hands these results demonstrate that site-specific PEGylation had minimal activity to extend the shelf-life. Furthermore, site-specific and non-specific PEGylation enhanced functional activity. Thus, site-specific PEGylation can potentially be explore as a means to enhance the biological activity of various reagent proteins used in diagnostic. The modified and unmodified proteins produced were functionally stable and produced a higher number of viable colony counts. Furthermore, PEGylation appeared to assisted proteins to retain their native structure. Reduced functionality occurred due to the ability of temperature to unfold and disrupt the covalent bonding present in these proteins, thus altering their functionality.

REFERENCES

Chiti, F., & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, *75*, 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901

Bhadra S, Pothukuchy A, Shroff R, Cole AW, Byrom M, Ellefson JW, et al. (2018) Cellular reagents for diagnostics and synthetic biology. PLoS ONE 13(8): e0201681. https://doi.org/ 10.1371/journal.pone.0201681

Fatima, S., & Khan, R. H. (2007). Effect of polyethylene glycols on the function and structure of thiol proteases. *Journal of Biochemistry*, *142*(1), 65–72. https://doi.org/10.1093/jb/mvm108

González-Valdez, J., Rito-Palomares, M., & Benavides, J. (2012). Advances and trends in the design, analysis, and characterization of polymer-protein conjugates for "pEGylaided" bioprocesses. In *Analytical and Bioanalytical Chemistry* (Vol. 403, Issue 8, pp. 2225–2235). https://doi.org/10.1007/s00216-012-5845-6

Grigoryan, K., Zatikyan, A., & Shilajyan, H. (2020). Effect of monovalent ions on the thermal stability of bovine serum albumin in dimethylsulfoxide aqueous solutions. Spectroscopic approach. In *Journal of Biomolecular Structure and Dynamics*. Taylor and Francis Ltd. https://doi.org/10.1080/07391102.2020.1743759

Hjerde, E., Maguren, A., Rzoska-Smith, E., Kirby, B., & Williamson, A. (2020a). DNA ligases of Prochlorococcus marinus: An evolutionary exception to the rules of replication. *BioRxiv*, 1–7. https://doi.org/10.1101/2020.05.11.089284

Hjerde, E., Maguren, A., Rzoska-Smith, E., Kirby, B., & Williamson, A. (2020b). DNA ligases of Prochlorococcus marinus: An evolutionary exception to the rules of replication. In *bioRxiv*. bioRxiv. https://doi.org/10.1101/2020.05.11.089284

Kumari, M., Sahni, G., & Datta, S. (2020). Development of Site-Specific PEGylated Granulocyte Colony Stimulating Factor With Prolonged Biological Activity. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.572077

Kusterle, M., Jevševar, S., & Porekar, V. G. (2008). Size of pegylated protein conjugates studied by various methods. *Acta Chimica Slovenica*, *55*(3), 594–601.

Ndlela, S. S., Friedrich, H. B., & Cele, M. N. (2020). Effects of modifying acidity and

reducibility on the activity of NaY zeolite in the oxidative dehydrogenation of n-octane. *Catalysts*, *10*(4). https://doi.org/10.3390/catal10040363

Opoku, F., Govender, P. P., Pooe, O. J., & Simelane, M. B. C. (2019). Evaluating Iso-Mukaadial Acetate and Ursolic Acid Acetate as Plasmodium falciparum Hypoxanthine-Guanine-Xanthine Phosphoribosyltransferase Inhibitors. *Biomolecules*, *9*(12). https://doi.org/10.3390/biom9120861

Pooe, O. J., Köllisch, G., Heine, H., & Shonhai, A. (2017). Plasmodium falciparum Heat Shock Protein 70 Lacks Immune Modulatory Activity. *Protein & Peptide Letters*, *24*(6), 503–510. https://doi.org/10.2174/0929866524666170214141909

Ramirez-Paz, J., Saxena, M., Delinois, L. J., Joaquín-Ovalle, F. M., Lin, S., Chen, Z., Rojas-Nieves, V. A., & Griebenow, K. (2018). Site-specific PEGylation crosslinking of L-asparaginase subunits to improve its therapeutic efficiency. *BioRxiv*. https://doi.org/10.1101/317040

Sadiki, A., Vaidya, S. R., Abdollahi, M., Bhardwaj, G., Dolan, M. E., Turna, H., Arora, V., Sanjeev, A., Robinson, T. D., Koid, A., Amin, A., & Zhou, Z. S. (2020). Site-specific conjugation of native antibody. *Antibody Therapeutics*, *3*(4), 271–284. https://doi.org/10.1093/abt/tbaa027

Salomane, N., Pooe, O. J., & Simelane, M. (2021). Iso-mukaadial acetate and ursolic acid acetate inhibit the chaperone activity of Plasmodium falciparum heat shock protein 70-1 Iso-mukaadial acetate and ursolic acid acetate inhibit the chaperone activity of Plasmodium falciparum heat shock protein 70-1. *Cell Stress and Chaperones.* https://doi.org/10.1007/s12192-021-01212-6

Samai, P., & Shuman, S. (2011). Structure-function analysis of the OB and latch domains of Chlorella virus DNA ligase. *Journal of Biological Chemistry*, *286*(25), 22642–22652. https://doi.org/10.1074/jbc.M111.245399

Segobola, J., Adriaenssens, E., Tsekoa, T., Rashamuse, K., & Cowan, D. (2018). Exploring Viral Diversity in a Unique South African Soil Habitat. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-017-18461-0

Singh, D., & Singh, M. (2021). *Hepatocellular-Targeted mRNA Delivery Using*

Functionalized Selenium Hepatocellular-Targeted mRNA Delivery Using Functionalized Selenium Nanoparticles In Vitro. *February*. https://doi.org/10.3390/pharmaceutics13030298

Veronese, F. M., & Pasut, G. (2005). PEGylation, successful approach to drug delivery. *Drug Discovery Today*, *10*(21), 1451–1458. https://doi.org/10.1016/S1359-6446(05)03575-0

Wilkinson, A., Day, J., & Bowater, R. (2001). Bacterial DNA ligases. *Molecular Microbiology*, *40*(6), 1241–1248. https://doi.org/10.1046/j.1365-2958.2001.02479.x

Zhao, W., Liu, F., Chen, Y., Bai, J., & Gao, W. (2015). Synthesis of well-defined protein-polymer conjugates for biomedicine. *Polymer*, *66*, A1–A10. https://doi.org/10.1016/j.polymer.2015.03.054

Zininga, T., Makumire, S., Gitau, G. W., Njunge, J. M., Pooe, O. J., Klimek, H., Scheurr, R., Raifer, H., Prinsloo, E., Przyborski, J. M., Hoppe, H., & Shonhai, A. (2015). Plasmodium falciparum hop (PfHop) interacts with the Hsp70 chaperone in a nucleotide-dependent fashion and exhibits ligand selectivity. *PLoS ONE*, *10*(8), e0135326. https://doi.org/10.1371/journal.pone.0135326

Zininga, T., Pooe, O. J., Makhado, P. B., Ramatsui, L., Prinsloo, E., Achilonu, I., Dirr, H., & Shonhai, A. (2017). Polymyxin B inhibits the chaperone activity of Plasmodium falciparum Hsp70. *Cell Stress and Chaperones*, *April*, 1–9. https://doi.org/10.1007/s12192-017-0797-6

CHAPTER 5: USE OF HSP70 AND CHIMERIC TRANSCRIPTION FACTOR TO ENHANCE DNA LIGASE STABILITY IN *ESCHERICHIA COLI*

5.1. ABSTRACT

The use of reagent proteins in research fields has increased and most chemical reactions in laboratory experiments require their utilization. Due to hindrances such as low shelf-life, increase of aggregates over time of storage, this study intended to investigate the role of Hsp70 and cTF stabilizing factor towards DNA ligase proteins. DNA ligase functionally seal nicks by linking 5'-phosphorylated end to a 3'-hydroxyl end on the phosphodiester backbone of DNA, utilising ATP or NADP molecules as an energy source. We co-expressed DNA ligases with Hsp70 and cTF in *E.coli*, respectively. Co-purification was conducted using affinity chromatography. The bioactivity of produced protein complexes was evaluated by conducting ligation studies. Thermal stability assays were utilized to evaluate the shelf-life and stability of the complex proteins. In this study, the co-expression of DNA ligases with Hsp70 and cTF was successfully conducted. Hsp70 enhanced folding, stability and shelf-life of DNA ligase protein. However, cTF couldn't enhance stability or shelf-life of target proteins, thus further studies can be conducted. Therefore, co-expression with Hsp70 can potentially increase stability and shelf-life of proteins for the longest time.

Keywords*: DNA ligase, co-expression, co-purification, Hsp70, cTF*

5.2. INTRODUCTION

The widespread application of reagent proteins keeps spreading especially in research fields, influenced by their specificity and effectiveness. The main natural development of enzymes (reagent proteins) is to use them to accelerate chemical reactions required for the survival of organisms and to complete several assays in the field. Enzymatic methods face several drawbacks including small shelf-life in solution, less possibility to use them in the presence of proteinases, others became less stable or unstable after a single use (Wong et al., 2020). Another principal drawback of proteins is the deactivation of their bioactivity even at relatively low temperatures (Gibbs et al., 2005). However, such shortcomings need to be overcome, and there are several approaches such as protein co-expression that can be used in such events.

The co-expression method is widely employed to reconstruct multi-protein complexes and to functionally determine co-solubility, co-purification, protein interactions, biochemical complexes, increase solubility and biological activity of protein partners. Several systems have been recognized efficiently conducting co-expression however, *E.coli* is still the most outstanding, eases to manipulate and rapidly grow (Ahmad et al., 2018; Vincentelli et al., 2011). Moreover, *E.coli* can allow co-expression using a single vector or several vectors that are co-transformed and capable of harbouring one or several promoters controlling one or more genes in a single-gene expression cassette (Chesnokov et al., 1999; Diebold et al., 2011). The component proteins constitutively co-exist as a complex due to instability that might sometimes arise when existing alone (Yamashita et al., 2017).

An identical replication origin and antibiotic resistance marker prevent stable proliferation propagation of plasmid pairs within a similar cell, which requires the removal of one or both open reading frames (ORFs) from a new plasmid to achieve co-expression (Scheich et al., 2007). Several methods including the use of plasmids containing different selectable markers with compatible origins of replication, one plasmid containing two ORFs under the control of separate promoters, or a single plasmid containing ORFs arranged in a polycistronic message have been employed, but all these methods have drawbacks. Therefore, accordingly, a facile and general method was developed which utilizes *E.coli* for protein co-expression. This method requires the sets of ORFs in identical expression plasmids, and these plasmids are required to contain a 61-nucleotide sequence which is called a link (Alexandrov et al., 2004). This permits the two identical plasmids with different ORFs to be linked together through the head to tail in a single tandem and propagated in *E.coli*. The presence of the link, in turn, further increases the joining of the two plasmids and generalizes the method for any pair of ORFs (Alexandrov et al., 2004). Recently, co-expression of distinct target proteins with chaperone systems such as Hsp70, GroEL-GroES and DnaK-DnaJGrpE have been performed targeting inclusion bodies, aggregation, preventing premature folding of the newly synthesized protein and stimulate protein solubility (Ma et al., 2020).

The chaperone systems have been mostly exploited for several years till today, they are known to be cooperative and the most favorable strategies to consider for co-expression. The co-overexpression of molecular chaperones strategy reduces the formation of inclusion bodies and recently has been used to improve the solubility of proteins. Encoded chaperones within *E.coli* determine protein folding, however, other chaperones prevent aggregation of proteins. The trigger factor chaperones and proteins interact immediately after the protein

leaves *E.coli* ribosomes (Sørensen & Mortensen, 2005). Moreover, trigger factors associate with newly synthesised protein to prevent premature folding by protecting exposed hydrophobic patches. Thereafter, the protein folds properly instantly after it got released from trigger factors. Hsp70 chaperone family-DnaK utilizes non-native proteins and aggregation-prone conformation to inhibit inclusion body formation by reducing aggregates and promote proteolysis of misfolded proteins. They are several essential heat shock proteins such as Hsp100 chaperone family which facilitates protein solubility and disaggregate proteins, and also GroEL (Hsp60 chaperone family) has been seen positively impacting disaggregation and inclusion body formation (Ma et al., 2020).

Previously, trigger factors were involved in co-overexpression and inhibited aggregate formation in mouse endostatin, human oxygen-regulated protein ORP150, human lysozyme and guinea pig liver transglutaminase. Furthermore, co-overexpression of the GroEL-GroES and DnaK-DnaJGrpE chaperone systems and trigger factors further stimulated soluble expression (Ma et al., 2020). The co-expression of Hsp70 with a plasmodial protein (PfGCHI) enhanced solubility, quality and functionality of the PfGCHI (Makhoba et al., 2016), and also in this study, we intend to obtain similar results and enhance the stability and shelf-life of DNA ligase proteins.

Thus far, chaperone molecules continue to work best as expected; GroEL/ES, Dnak/J-GrpE and trigger factor were also significantly observed enhancing solubility of co-type NHase and they further reduced industrial production costs by assisting expression systems to produce activated and soluble products (Makhoba et al., 2015). Furthermore, the nitrilase (from Penicilium marneffei) solubility and activity were increased after co-expressed with GroEL-GroES. Similar results (including functional expression) were also obtained on human interferon-gamma when co-expressed with GroEL-GroES significantly (Xu et al., 2020). Additionally, the fusion expression of another chaperone from R.ruber (RrGroEL) with NHase improved the activity and stability of the enzyme, while MtGroEL1 and MtGroEL2 showed refolding activity (independent from GroES). Therefore, chaperones such as RrGroEL and RrGroES, their characteristics and functions should be examined and be considered for utilization with enzymes such as nitrilase or enzymes with promising future in the industrial application (Xu et al., 2020).

A rare protein fusion or protein co-expression study was done by Wilson *et al* (2013) where seven DNA binding proteins including eukaryotic transcription factors; NF-kB p50 and

chimeric transcription factor (cTF) were fused with T4 DNA ligases. The study reported that all these factors and DNA binding proteins increased the T4 DNA ligase activity by seven-folds. Comparatively, the most active variants viewed were p50- fused to T4 DNA ligase and T4 DNA ligase fused to cTF. With approximately 160 % factor, the ligase fused to cTF out-performed T4 DNA ligase in the cloning assay of the blunt end (vector-on-insert), and the construction of a library for Illumina sequencing (Wilson et al., 2013). Therefore, this study intended to investigate the role of Hsp70 and cTF stabilizing factor towards DNA ligase proteins. DNA ligases play an indispensable role in maintaining genomic integrity and ensuring the completion of the DNA helical structure during DNA replication, recombination, and repair. Their functional activities include the joining of Okazaki fragments at the replication fork on the lagging strand during DNA replication and catalysing the final steps in base excision repair or non-homologous end-joining repair pathways (Hjerde et al., 2020; Samai & Shuman, 2011; Wilkinson et al., 2001).

5.3. MATERIALS AND METHOD

5.3.1. Preparation of Chemical competent cells encoding DNA ligase

The LB plate was used to select a single colony encoding DNA ligase to prepare a starter culture of cells. The LB media (10 ml) was inoculated with a single colony and grown overnight at 37 °C (160 rpm). The overnight starter culture was inoculated into 100 ml LB media and grown by shaking at 37 °C until OD_{600} reached 0.4 to 0.5 Abs. The cells were then chilled immediately on ice for 20 to 30 minutes with swirling, occasionally. Then, the culture was plitted into 50 ml centrifuge tubes pre-chilled into ice followed by cells harvest by centrifugation at 1000 g (4000 rpm in the Beckman JA-10 rotor) for 7 minutes (4 °C). The supernatant was discarded and the pellet was re-suspended with ice cold 0.1 $MgCl_2$ of the original volume and stored in ice for 15 minutes. Thereafter, the cells were centrifuged at 4000 rpm for 7 minutes (4 °C), afterwards the supernatant was discarded. The ice cold 0.1 M of $CaCl_2$ was used to re-suspend the cells with half volume of the original and incubated in ice for 15 minutes. Thereafter, the pellet was re-suspended with 1 ml of ice cold 0.1 M $CaCl_2$ and 20 % of glycerol. Thereafter, sterilized and ice pre-chilled microfuge tubes (1.5 ml) were used to aliquote cells which were then stored into the -80 °C freezer.

5.3.2. Transformation of electro-competent cell encoding DNA ligase

Approximately, 10 ng of purified DNA of transcription factor cTF and PfHsp70 were added to 50 µl of ice chilled competent cells of DNA ligase (Ligsv081) and DNA ligase (LigpET30), respectively. Then they were stored in ice for ten minutes and the water bath was turned on until 42 °C temperature was reached. The transformation was conducted by placing the competent cells with DNAs into water bath for 30 seconds. Thereafter, the cells were immediately chilled into ice for one minute and one-hour incubation in 1 ml of LB media with shaking at 37 °C (160 rpm). Afterward, the cells were plated onto LB agar plates with two respective antibiotics (Ampicillin and Kanamycin) and incubated at 37 °C overnight. It should be noted that PfHsp70 was co-expressed with LigpET30 only, and Ligsv081 was used as a control.

5.3.3. Protein co-expression

A single colony of complex DNA ligase was inoculated into 10 ml of LB broth supplemented with 50 µg/ml kanamycin and 100 µg/ml ampicillin followed by incubation at 37 °C overnight in a shaker at 160 rpm. The overnight inoculum was then sub-cultured into 150 ml LB broth with 50 µg/ml of kanamycin and 100 µg/ml ampicillin which was then subjected to a temperature of 37 °C with shaking until the culture reached optical density between 0.4 to 0.6 Abs (600 nm). The induction of protein expression was conducted by adding 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Post-induced samples were recovery through centrifugation at 13000 g for 3 minutes on a JA14 rotor (Beckman, Optima L-100XP) and intracellular proteins were released through re-suspending with lysis buffer (10 mM Tris-HCl, pH 7.4, 300 mM NaCl, 10 mM imidazole). A sodium dodecyl sulfatepolyacrylamide gel electrophoresis (SDS-PAGE) (15%) was used to analysed protein expression profile of 1 hour to 24 hours run at 100 V, 25 mA for 1 hour. The Coomassie brilliant blue was used to stain the gel SDS-PAGE gel. The western blot was performed to confirm the expression of protein; Polyvinylidene difluoride (PVDF) membrane (0.2 µl pore) was used to transfer proteins. Furthermore, to block non-specific binding, the membrane was incubated into 5 % skimmed milk for 1 hour. Anti-HIS-tag IgG conjugated with horse-radish peroxidase (HRP from Sigma Germany; diluted 1:1000 in TBS) was used to react with samples for 1 hour in ice, followed by three times wash with TBST. Chemiluminescence (ECL) kit method (Thermo Fisher Scientific, SA) was employed to as enhance protein bands to be detected and developed (Zininga et al., 2015).

5.3.4. Protein co-purification

The affinity chromatography purification was facilitated using polyhistidine tag (His-tag) in the co-expression of recombinant ligases and transcription factor of cTF, with minor modifications (Pooe et al., 2017; Salomane et al., 2021). The His-tagged proteins were purified using the HisPur^{TM} Ni-NTA Resin column (Thermo Fisher scientific) following the manufacture's protocol. Briefly, the column was equilibrated with LEW (7.8 g NaH_2PO_4 and 17.5 g $NaCl_2$). Adjust pH using in NaOH buffer followed by 15 ml loading of soluble crude fraction made up of His-tagged proteins. Afterward, the LEW buffer (three column volumes-CV) was used to wash off unbound proteins. The elution buffer (7.8 g NaH_2PO_4, 17.5 g NaCl, 17.0 g imidazole with adjusted pH using in NaOH) was used to elute target protein, which was then verified with SDS-PAGE (15 %) and western blot was done as stated above in **5.3.4.**

5.3.5. Activity assay of the co-expressed protein

Ligation reactions described by (Sambrook et al., 2009) were essentially used to test the activity of purified recombinant DNA ligases and co-expressed DNA ligases with cTF and PfHsp70 in comparison with commercial T4 DNA ligase from Thermo Fisher Scientific. To test DNA ligase isolates activity, the mixture or reaction mixture consisted of the following: 1 µl of DNA ligase recombinant, digested lambda DNA with restriction enzyme (*HindIII*), commercial cutsmart buffer, and nuclease-free water in a 20 µl ligation reaction (Thermo Fisher Scientific, SA) at 37 °C (pre-setting) for one hour. Varied ligation times were used from 30 minutes and 1-hour temperature (25 °C). Ligations were carried out by performing ligation (vector to an insert) of pET28 with an insert digested with *XhoI* and *NdeI* to demonstrate test co-expressed or co-produced DNA ligases.

5.3.6. Investigating the role of Hsp70 and cFT using agar plates thermal stability assay

Co-expressed variants of Ligsv081-cTF, LigpET30-cTF and LigpET30-PfHsp70 were grown overnight at 37 °C in LB media containing 25 mM of kanamycin and 50 mM of ampicillin. The inoculum grown overnight was transferred into fresh media of LB and shake under the same growth conditions and at the mid-log growth phase, the cells were induced with 1 mM IPTG. This followed by Shaking until the cells reached OD_{600} = 2.0. Afterward, the cells

were standardized to a density of 0.2 OD_{600} before spotted on LB agar plates containing necessary antibiotics and 20 mM IPTG and incubated at 37 °C and 43.5 °C overnight. The co-purified proteins and DNA ligases which were used as negative control were incubated at 43.5 °C for 30 minutes followed by centrifugation to obtain supernatant and soluble products which were then characterized on SDS-PAGE and Western blot.

5.4. RESULTS AND DISCUSSION

5.4.1. Co-expression and co-purification of protein complexes

Most high-yield protein produced by *E. coli* accumulate as insoluble aggregates of inclusion bodies therefore, it is essential to solubilize and refold their structure thus, restore bioactivity (Tripathi, 2016). Co-expression is a process that is mainly employed for such activities toward target protein. Over the years co-expression with chaperone systems mostly grants prestigious functionality and improved proteins. Currently, most proteins face several hurdles such as reduced shelf-life, stability, aggregation, etc. especially reagent proteins. Therefore, it is important to consider or explore more the co-expression techniques so they can be utilized as an alternative in proteins and protein therapy.

Co-expression and co-purification of DNA ligases with PfHsp70 and cTF were successfully performed as shown below in Figure **5.1a** and Figure **5.1b**. The presence of both recombinant proteins meant to be co-expressed and co-purified simultaneously are characterized on the SDS-PAGE and confirmed on Western blot. The co-expression in this study generated DNA ligase derivatives which were designated as Ligsv081-cTF, LigpET30-PfHsp70 and LigpET30-cTF. The results show protein complexes of DNA ligase with PfHsp70 and the band of LigpET30 shown on SDS-PAGE is well folded.

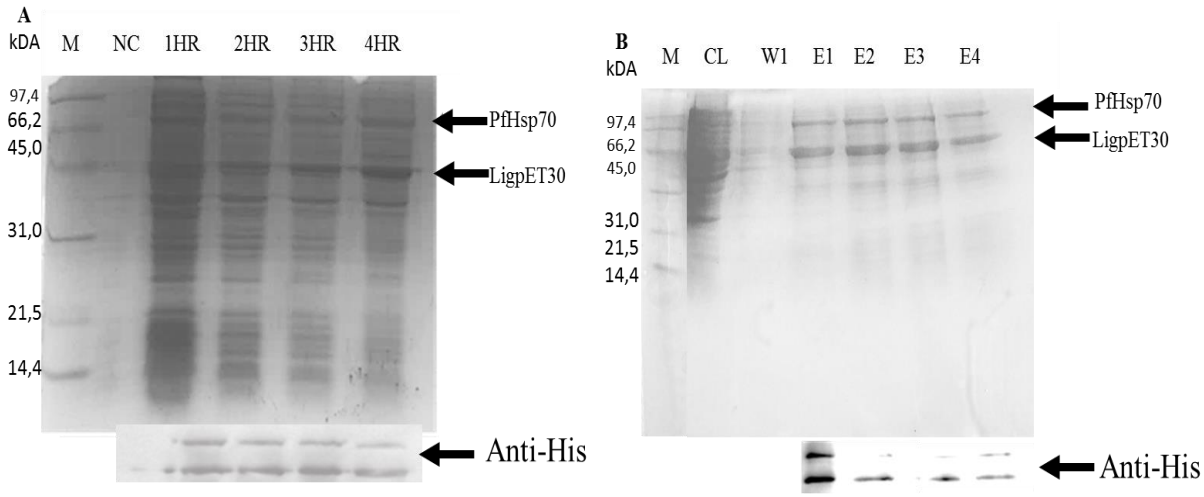

Figure 5.1a: SDS-PAGE showing IPTG induced co-expression and purification of DNA ligase (LigpET30) with PfHsp70. **A**: Lane M; marker, Lane NC; *E. coli* BL21 control cells, Lane 1HR to 4HR represent the time-profile sampling from 1 hour to 4 hours post IPTG -induction. **B**: Cl; crude lysate, W1; wash one and E1 to E4; eluted protein.

Heat shock proteins (PfHsp70) are essentially involved in the protein folding or assembling of newly synthesized proteins. Moreover, PfHsp70 assists misfolded and aggregated proteins to refold, and control the activity of regulatory proteins. PfHsp70s regulate properly as central component to catalyse folding processes, therefore they specifically acquire housekeeping functions that control proofreading of protein structures and repair detected misfold (Mayer & Bukau, 2005). It is essential for PfHsp70 to interact with hydrophobic peptide segments (found in proteins controlled by ATP) properly so that all these activities could be completely achieved. Furthermore, PfHsp70 has three related activities in the folding of non-native proteins such as stimulating folding or aggregated protein to the native state, solubilization and refold properly all aggregated proteins (Mayer & Bukau, 2005). according to the results in Figure **5.1a**, PfHsp70 essentially folded properly LigpET30 and increased its protein production yield. The concentration of 250 mM imidazole in elution buffer was used to elute target proteins.

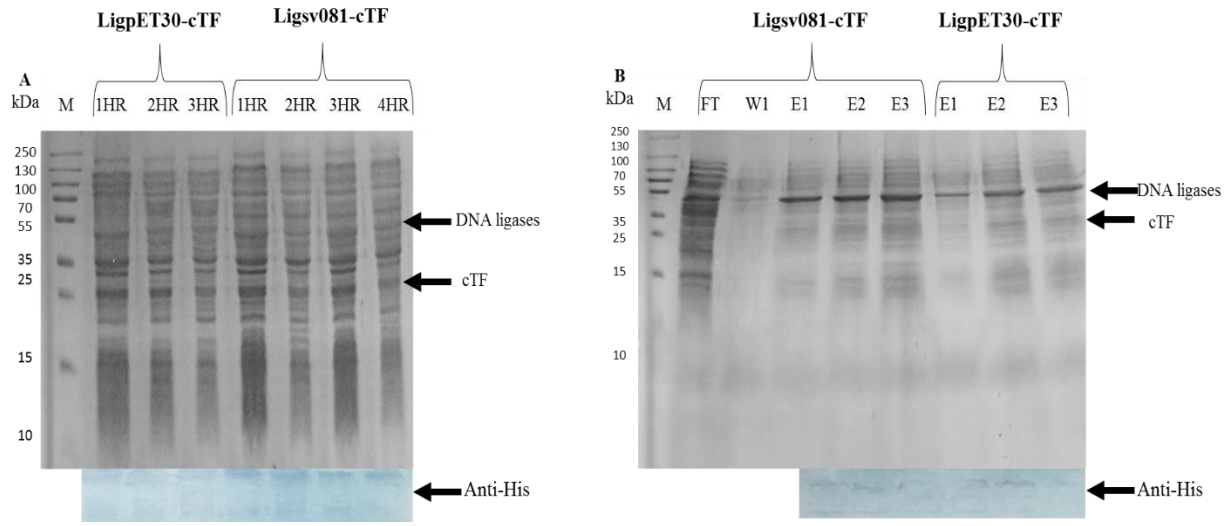

Figure 5.1b: SDS-PAGE showing IPTG induced co-expression and purification of DNA ligases (Ligsv081 & LigpET30) with cTF. **A**: Lane M; marker, Lane 1HR to 3HR & 1HR to 4HR represent the time-profile sampling from 1 hour to three hours & 1 hour to 4 hours post IPTG-induction for LigpET30-cTF and Ligsv081-cTF. **B**: Ft; flow-through, W1; wash one and E1 to E3; eluted protein.

According to Figure **5.1b**, the production of co-expressed proteins with cTF was successful even though the concentration of cTF produced after purification was smaller. The concentration of 250 mM imidazole was used to elute purified target proteins. After the proteins were obtained through purification were stored at a lower temperature for further analysis.

5.4.2. Ligation activity of co-expressed proteins

The co-expressed protein complexes were purified and ligation activity was conducted followed by results analysis on agarose gel as shown below in Figure **5.2**. The results demonstrate successful ligation of lambda DNA which was firstly restricted with *HindIII* then re-ligated with LigpET30-PfHsp70 **(A)** and Ligsv081-cTF and LigpET30-cTF as shown (**B**). Afterward, the standard vector and insert were ligated and plated on agar plates, the number of obtained white colonies are displayed in the graph (Figure **5.2**). All DNA ligase complexes show higher number of colony counts compare to ligase controls. The yielded colony counts results are statistically significant with $p = 0.12$.

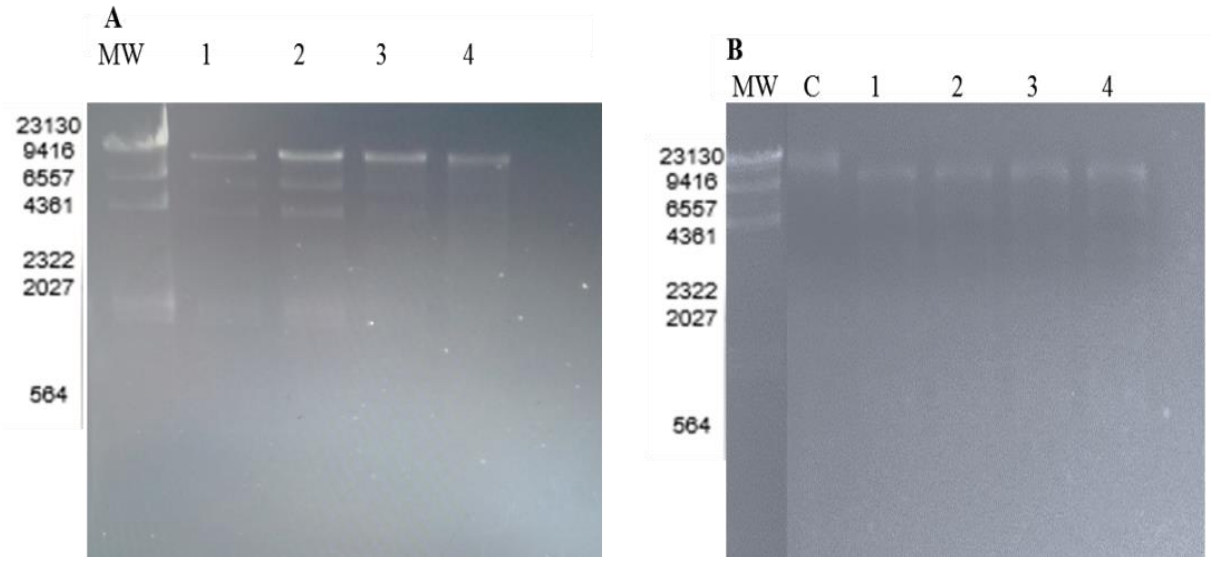

Figure 5.2: Agarose gel showing ligation activity of co-expressed proteins. **A**: Lane 1- Lambda DNA digested with *HindIII* as a marker; Lane 1 to 4 LigpET30-Hsp70. **B**: C- T4 DNA ligase (control), Lane 1 & 2- Ligsv081-cTF and Lane 3 & 4- LigpET30-cTF.

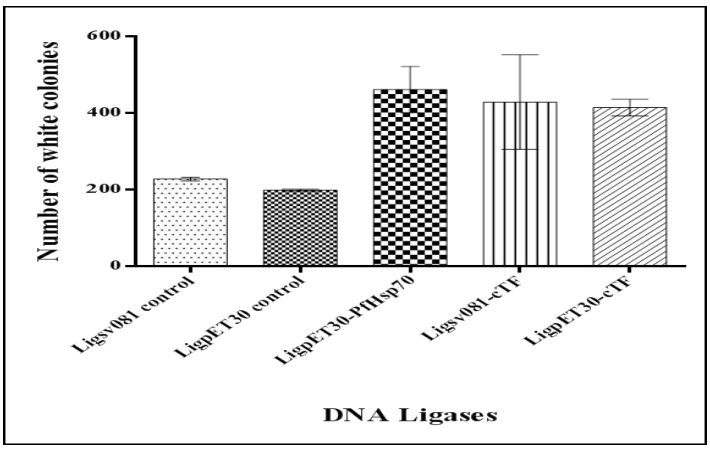

Figure 5.3: Cloning assay. Blunt-ended insert and standard vector were ligated together and transformed into *E. coli* Bl21. Colony counts are means of independent triplicates.

5.4.3. Thermal denaturation of co-expressed proteins

An external water bath (connected to the UV-VIS) was used to change the temperature of the cell holder in contact with a specific cuvette used to read the samples or proteins. The water bath of temperatures between 25 °C and 45 °C was increased with 5 °C and 50 °C to 60 °C. While, the temperature in which the protein unfolding usually takes place in range from 40

°C to 50 °C; the temperature increase was controlled with smaller 1 °C steps. Therefore the calibration curve of the temperature present in the cuvette vs OD is shown in Figure **5.4**.

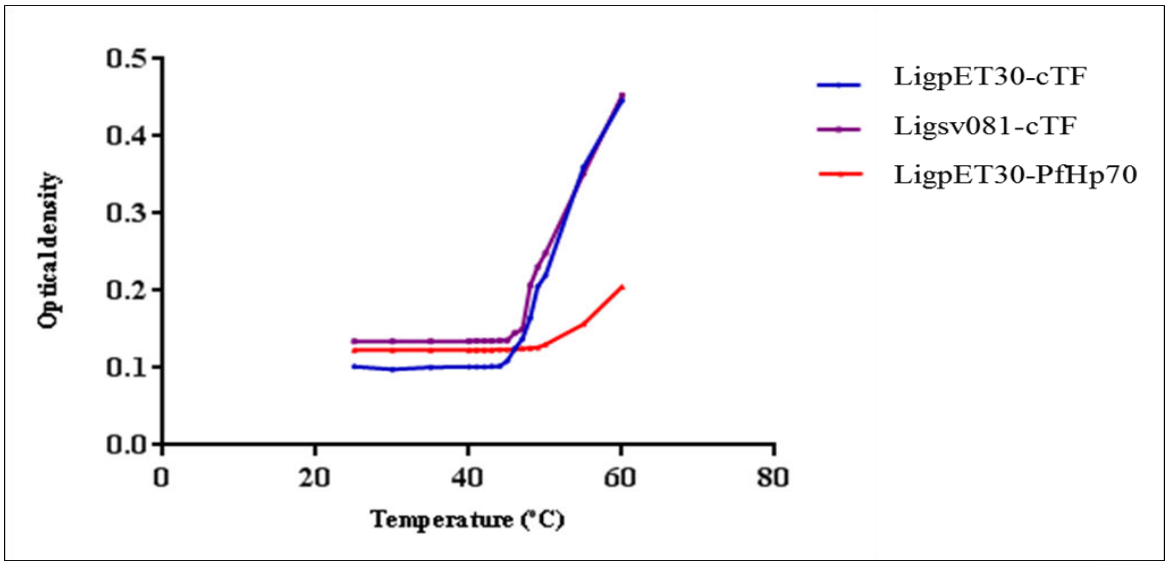

Figure 5.4: Thermal denaturation of LigpET30-cTF DNA, Ligsv081-cTF DNA ligase protein and LigpET30-PfHsp70.

Co-expressed protein (LigpET30-PfHsp70, LigpET30-cTF and Lgsv081-cTF) results in Figure **5.3** above, showing that the melting point of proteins has slightly increased in all proteins. LigpET30-cTF and Ligsv081-cTF their melting point shifted to 47 °C, and for LigpET30-PfHsp70 which we decided to store for three months in the refrigerator (4 °C) thus to assess its shelf life and stability, the melting point obtained was 51 °C. Thermal denaturation gives an insight of protein unfolding events and temperature range in which protein can adapt, especially temperatures above optimal temperature for protein activity (Modarres et al., 2015). Co-expression studies were also done to enhance sites of protein structure that are highly sensitive to temperature. After passing the melting point, proteins start to unfold and denature as temperature increases. The thermal stability assessment of DNA ligases which were not co-expressed, the values were not consistent therefore it was decided not to record them in the graph. Thermal stability and bioactivity of proteins can be deactivated even at relatively low temperatures (Bommarius & Broering, 2009; Gibbs et al., 2005).

5.4.4. Evaluation of protein complexes' shelf-life and stability

Co-expressed protein complexes and DNA ligase proteins were stored for four months at 4 °C to assess the role of PfHsp70 and cTF towards protein's shelf-life and stability. The results on Figure **5.4.4 (A)** show that co-expressed and DNA ligase cells were able to grow at 43.5 °C overnight, mostly at their high concentration. The cells that were lagated with complex DNA ligase co-expressed with PfHsp70 were able to survive even at their low concentration due to the capability of PfHsp70 to elevate protection effectively in various physiological and environmental insults towards proteins. Co-purified and purified proteins were also heated at 43.5 °C for 30 minutes, followed by characterization on the SDS-PAGE and western blot (**B**). The co-expressed complex of LigpET30-PfHsp70 clear bands was visible on Western blot but other proteins were not visible on both SDS-PAGE and Western blot, meaning long storage reduced their shelf-life and melted or denatured during heating, and cTF was not able to stabilize these proteins. Alternatively, PfHsp70 played a vital role by protecting its complex partner protein from denaturation and instability, and further extended shelf-life.

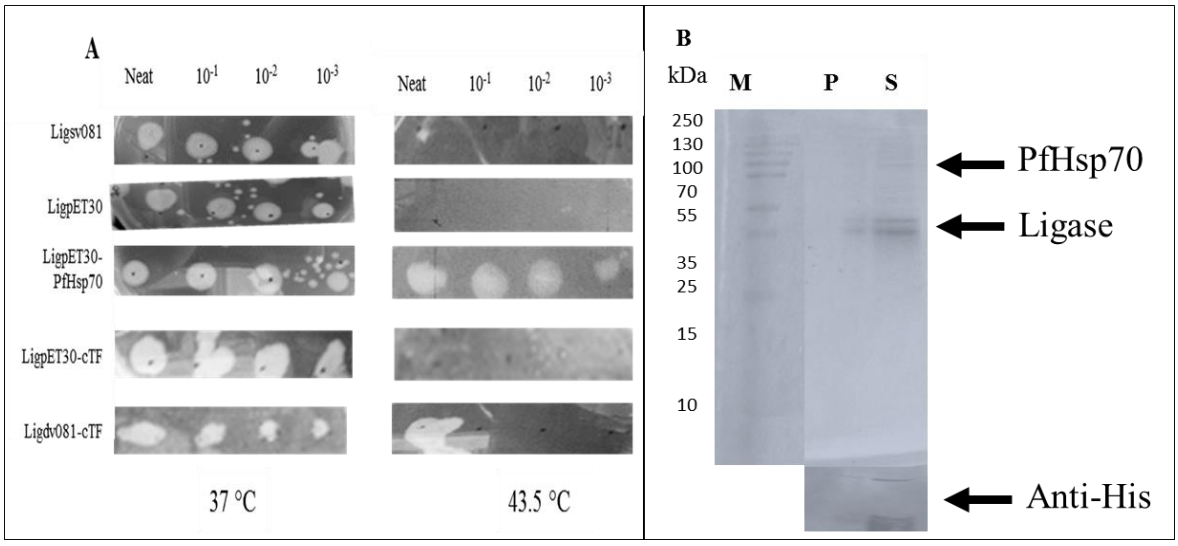

Figure 5.5: Thermal denaturation of the co-expressed complex of DNA ligases on agar plates and SDS-PAGE.

Figure **5.5** (**A**) is showing results of standard vector and insert ligated by complex DNA ligases, afterwards the ligated samples were spotted on agar plates followed by incubation at 43.5 °C overnight. The bacterial cells *E. coli* (DE3) were used to transform ligated product with the known optimum temperature of 37 °C. Increasing growth temperature can cause

stress to bacteria, thus more responsibility for DNA ligase to conduct more DNA repairing events and also more replication forks can be active. Amino acids present in the hydrophobic residue also play an important role by assisting DNA ligase to retain its activity at non-permissive temperatures. The non-permissive temperature may result into the disruption of biochemical and biophysical structure of protein thus, reduced the protein activity (Modarres et al., 2015). The presence of PfHsp70 may have caused the growth of viable cells after temperature increase on plates.

5.5. CONCLUSION

In conclusion, enhancing protein thermal stability is important considering protein functionality is not compromised. In this study the co-expression of DNA ligase withPfHsp70 had a positive impact on the growth of viable cells and stabilised protein activity by preventing disruption of biochemical and biophysical structure of protein. Furthermore, PfHsp70 reduced protein aggregation after introduced into temperatures above protein melting point. Protein co-expression can be used essentially to increase protein stability without altering the functionality of the proteins.

REFERENCES

Ahmad, I., Nawaz, N., Darwesh, N. M., ur Rahman, S., Mustafa, M. Z., Khan, S. B., & Patching, S. G. (2018). Overcoming challenges for amplified expression of recombinant proteins using Escherichia coli. *Protein Expression and Purification*, 144, 12–18. https://doi.org/10.1016/j.pep.2017.11.005.

Alexandrov, A., Vignali, M., Lacount, D. J., Quartley, E., Vries, C. De, Rosa, D. De, Babulski, J., Mitchell, S. F., Schoenfeld, L. W., Fields, S., Hol, W. G., Dumont, M. E., Phizicky, E. M., & Grayhack, E. J. (2004). A Facile Method for High-throughput Co-expression of Protein Pairs. *Molecular and Cellular Proteomics*, *3*(9), 934–938. https://doi.org/10.1074/mcp.T400008-MCP200.

Bommarius, A. S., & Broering, J. M. (2009). Established and novel tools to investigate biocatalyst stability Established and novel tools to investigate biocatalyst stability. *Biocatalysis and Biotransformation*, 23(3/4): 125-139. https://doi.org/10.1080/10242420500218877.

Chesnokov, I., Gossen, M., Remus, D., & Botchan, M. (1999). Assembly of functionally active Drosophila origin recognition complex from recombinant proteins. *Genes and development*, 1289–1296.

Diebold, M., Fribourg, S., Koch, M., Metzger, T., & Romier, C. (2011). Deciphering correct strategies for multiprotein complex assembly by co-expression : Application to complexes as large as the histone octamer. *Journal of Structural Biology*, *175*(2), 178–188. https://doi.org/10.1016/j.jsb.2011.02.001.

Gibbs, P. R., Uehara, C. S., Neunert, U., & Bommarius, A. S. (2005). Accelerated Biocatalyst Stability Testing for Process Optimization. *Biotechnol*, 21, 762−774.

Haffke, M., Marek, M., Pelosse, M., Diebold, M., Schlattner, U., Berger, I., & Romier, C. (2015). Chapter 4 Characterization and Production of Protein Complexes by Co-expression in *Escherichia coli*. *Methods in Molecular Biology*, *1261*, 63–89. https://doi.org/10.1007/978-1-4939-2230-7.

Hjerde, E., Maguren, A., Rzoska-Smith, E., Kirby, B., & Williamson, A. (2020). DNA ligases of Prochlorococcus marinus: An evolutionary exception to the rules of replication. *BioRxiv*, 1–7. https://doi.org/10.1101/2020.05.11.089284.

Johnston, K., & Marmorstein, R. (n.d.). Co-Expression of Proteins in E. coli Using Dual Expression Vectors. 205(18), 205–213.

Ma, Y., Lee, C., & Park, J. (2020). Strategies for Optimizing the Production of Proteins and Peptides with Multiple Disulfide Bonds. *Antibiotics*, 9, 541. doi:10.3390/antibiotics9090541.

Makhoba, X. H., A, B., Coertzen, D., Zininga, T., L-M, B., & A, S. (2016). Use of a Chimeric Hsp70 to Enhance the Quality of Recombinant Plasmodium falciparum S - Adenosylmethionine Decarboxylase Protein Produced in *Escherichia coli*. *PLoS ONE* ,1–21. https://doi.org/10.1371/journal.pone.0152626.

Makhoba, X. H., Pooe, O. J., & Mthembu, M. S. (2015). Molecular Chaperone Assisted Expression Systems: Obtaining Pure Soluble and Active Recombinant Proteins for Structural and Therapeutic Purposes. *Journal of Proteomics & Bioinformatics*, *08*(09), 212–216. https://doi.org/10.4172/jpb.1000371.

Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, *62*(6), 670–684. https://doi.org/10.1007/s00018-004-4464-6.

Pooe, O. J., Köllisch, G., Heine, H., & Shonhai, A. (2017). Plasmodium falciparum Heat Shock Protein 70 Lacks Immune Modulatory Activity. *Protein & Peptide Letters*, *24*(6), 503–510. https://doi.org/10.2174/0929866524666170214141909.

Salomane, N., Pooe, O. J., & Simelane, M. (2021). Iso-mukaadial acetate and ursolic acid acetate inhibit the chaperone activity of Plasmodium falciparum heat shock protein 70-1 Iso-mukaadial acetate and ursolic acid acetate inhibit the chaperone activity of Plasmodium falciparum heat shock protein. *Cell Stress and Chaperones, 70-1*. https://doi.org/10.1007/s12192-021-01212-6.

Sambrook, J., Russell, D. W., Sambrook, J., & David, W. (2009). Protocol Blunt-end Cloning of PCR Products. 2006(2), 2–5. https://doi.org/10.1101/pdb.prot3830.

Samai, P., & Shuman, S. (2011). Structure-function analysis of the OB and latch domains of Chlorella virus DNA ligase. *Journal of Biological Chemistry*, *286*(25), 22642–22652. https://doi.org/10.1074/jbc.M111.245399.

Scheich, C., Ku, D., Soumailakakis, D., & Delbru, M. (2007). Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Research*, *35*(6). https://doi.org/10.1093/nar/gkm067.

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*, 1–8. https://doi.org/10.1186/1475-2859-4-1.

Tripathi, N. K. (2016). Production and Purification of Recombinant Proteins from Escherichia coli. *ChemBioEng Rev*, 3, 116–133. https://doi.org/10.1002/cben.201600002.

Vincentelli, R., Cimino, A., Geerlof, A., Kubo, A., Satou, Y., & Cambillau, C. (2011). High-throughput protein expression screening and purification in Escherichia coli. *Methods*, *55*(1), 65–72. https://doi.org/10.1016/j.ymeth.2011.08.010.

Wilkinson, A., Day, J., & Bowater, R. (2001). Bacterial DNA ligases. *Molecular Microbiology*, *40*(6), 1241–1248. https://doi.org/10.1046/j.1365-2958.2001.02479.x.

Wilson, R. H., Morton, S. K., Deiderick, H., Gerth, M. L., Paul, H. A., Gerber, I., Patel, A., & Ellington, A. D. (2013). Engineered DNA ligases with improved activities *in vitro*. *Protein Engineering, Design & Selection*, 26(7), 471–478. https://doi.org/10.1093/protein/gzt024.

Wong, J. X., Ogura, K., Chen, S., & Rehm, B. H. A. (2020). Bioengineered Polyhydroxyalkanoates as Immobilized Enzyme Scaffolds for Industrial Applications. *Biotechnol*, 8:156. https://doi.org/10.3389/fbioe.2020.00156.

Xu, C., Tang, L., Liang, Y., Jiao, S., Yu, H., & Luo, H. (2020). Novel Chaperones Rr GroEL and Rr GroES for Activity and Stability Enhancement of Nitrilase in Escherichia. *Molecules*, 25, 1002. doi:10.3390/molecules25041002.

Yamashita, A., Nango, E., & Ashikawa, Y. (2017). METHODS AND APPLICATIONS A large-scale expression strategy for multimeric extracellular protein complexes using Drosophila S2 cells and its application to the recombinant expression of heterodimeric ligand- binding domains of taste receptor. Protien science, *26*, 2291–2301. https://doi.org/10.1002/pro.3271.

Zininga, T., Makumire, S., Gitau, G. W., Njunge, J. M., Pooe, O. J., Klimek, H., Scheurr, R.,

Raifer, H., Prinsloo, E., Przyborski, J. M., Hoppe, H., & Shonhai, A. (2015). Plasmodium falciparum hop (PfHop) interacts with the Hsp70 chaperone in a nucleotide-dependent fashion and exhibits ligand selectivity. *PLoS ONE*, *10*(8), e0135326. https://doi.org/10.1371/journal.pone.0135326.

CHAPTER 6

6.1. GENERAL DISCUSSION

In 2015, the global ligase enzyme market was valued at $503 million and it is expected to reach $918 million by 2026 (https://www.360researchreports.com/enquiry/request-sample/13938709). Ligases are generally defined as enzymes that catalyse the joining together of the two DNA molecules coupled in the presence of ATP/NADP molecule functioning as energy source (Williamson, 2019). Thus, DNA ligases play a significant role as DNA manipulating enzymes and the enzyme reagent market in general. The importance of DNA ligases in DNA maintenance and development of cells causes them to be more considerable in the development of antimicrobial agents and cancer therapeutics (Ellenberger & Tomkinson, 2008). Nicks or breaks found in the phosphodiester backbone of DNA are sealed by DNA ligases by linking 5'-phosphorylated end to a 3'-hydroxyl end (Nair et al., 2008).

In the present study, the two DNA ligases evaluated were extracted from Kogelberg Biosphere Reserve (KBR) Fynbos soil microbial communities, by Jane *et al*., (2018). Using the microbial community as a source for enzymes is advantageous as bacteria have a broad range of catabolic activities, provide their enzymes with unusual stability characteristics since they can thrive in extreme conditions (Choi et al., 2015), can achieve a great yield of products and the ease of engineering and expressing modified proteins to improve specificity and stability (Price et al., 1995). Results obtained proved that DNA ligases share functional groups and chemical structures. Furthermore, the protein major factors such as the stability, aggregation and bioactivity of both DNA ligases were increased by PEG and co-expression technique forming complexes with PfHsp70 and cTF. PEGylation and protein co-expression techniques have been used before successfully. Unfortunately, neither of the two techniques could extend the shelf life of DNA ligase enzymes after storage for approximately 14 days incubation. This could have been caused by the pH value of the storage buffer used during shelf life assessment (pH usually determines the charge of the protein), the degree of PEGylation and the concentration of proteins used.

The present study gives an insight on how to increase the two enzymes of DNA ligases and their activity which essentially involves the survival of every organism by specifically participating in the DNA replication, repairing and recombination which is taking place at the cellular level. Furthermore, the concentration increase of proteins with PEGs conjugation;

resulted in the production of new DNA ligase derivatives. These DNA ligase derivatives were produced with an increased melting point, less aggregate formed and were able to ligate vector to insert. All proteins utilized in this study require proper storage which is the lower temperature to prevent sudden damagers or reduction of biological activities of proteins (Gokarn, 2003).

PEG molecules assisted the two DNA ligase enzymes to assume their 3D tertiary structure. Generally, the 3D structure appears when the hydrophilic residues, polypeptides and polar side of the protein are exposed to the polar aqueous environment. Whereas, the non-polar hydrophilic residues are shield from exposure to water. However, thermal denaturation exposed hydrophobic residues and thermodynamic unfavourable conditions occurred. Thermo stability assessed on the UV-VIS spectrum showed the increase of melting temperature (where the protein is half native and half unfolding) of LigpET30-PfHsp70. The increase in the melting temperature leads to an increase in protein stability. Additionally, by changing the pH of the solution and adding additives in this study PEG molecules were added and changed the melting temperature of the proteins, thus, PEG enhanced the stability of the two DNA ligase enzymes. The thermal denaturation of conjugates showed partial ligation of lambda DNA during ligation assay, this occurred since usually PEG molecules aggregate at elevated temperatures depending on the molecular weight (Huberty, 2014).

During the 14 days incubation, particle size increased as proteins aggregated seeking to minimize unfolding due to the exposure of the hydrophobic surface area to water. Usually, PEG molecules are expected to protect proteins from the exposure of their hydrophobic surface area and hinder protein-protein interactions during storage and further present steric hindrance. Resulting in a reduction of protein aggregates. Conjugates aggregated since the PEGylation degree used was lower and did not enhance shelf life. Therefore, the degree of PEGylation is the most important aspect and it is essential to know the contribution of PEGylation towards the increase of the hydrodynamic volume of a protein (Gokarn, 2003; Wang et al., 2010).

It is essential to begin by finding ways that increase concentration and which will not decrease functional activities of proteins such as usage of PEG which is FDA approved. Also, co-expression of targeted proteins with other proteins with necessary characteristics will assist with improving activities of the other protein can be attempted. Protein thermo-stability is the most essential aspect in industries that use proteins mostly, therefore, it is important to identify ways to increase protein thermal stability and shelf-life. In this study, these aspects were inconclusive.

Protein enhancement through PEGylation and co-expression techniques are highly recommended techniques that may lead to optimized proteins which can retain their bioactive characteristics and increases protein's shelf-life over a long storage period on the reagent proteins. Future considerations, supplementing media with metals should be assessed to see how they can assist protein with folding, thermal stability, bioactivity and other characteristics. Furthermore, the evaluation of thermal stability of protein conjugated using kinetic assays should be considered.

REFERENCES

Choi, J., Han, S., & Kim, H. (2015). Industrial applications of enzyme biocatalysis : Current status and future aspects. *Biotechnology Advances*, *33*(7), 1443–1454. https://doi.org/10.1016/j.biotechadv.2015.02.014

Ellenberger, T., & Tomkinson, A. E. (2008). Eukaryotic DNA Ligases : Structural and Functional Insights. https://doi.org/10.1146/annurev.biochem.77.061306.123941

Gokarn, Y. R. (2003). Hydrodynamic behavior and thermal stability of a PEGylated protein : Studies with hen egg lysozyme. Doctoral Dissertations. 192. https://scholars.unh.edu/dissertation/192

Huberty, W. (2014). Characterization of a water soluble , non-ionic , helical poly (α-amino acid). LSU Doctoral Dissertations. 2390. https://digitalcommons.lsu.edu/gradschool_dissertations/2390.

Nair, P. A., Koster, D. A., Shuman, S., & Dekker, N. H. (2008). Dynamics of phosphodiester synthesis by DNA ligase. *PNAS*, 105(19), 1–6. www.pnas.orgdoi10.1073pnas.0800113105.

Price, C. P., Campbell, R. S., & Hammond, P. M. (1995). Novel enzymes as reagents. *Clinica Chimica Acta*, *237*, 3–16.

Wang, W., Nema, S., & Teagarden, D. (2010). Protein aggregation — Pathways and influencing factors. *International Journal of Pharmaceutics*, 390, 89–99. https://doi.org/10.1016/j.ijpharm.2010.02.025

Williamson, A. (2019). NAR Breakthrough Article Structural intermediates of a DNA – ligase complex illuminate the role of the catalytic metal ion and mechanism of phosphodiester bond formation. *Nucleic Acids Research*, *47*(14), 7147–7162. https://doi.org/10.1093/nar/gkz596

APPENDICES

Appendix A: Supplementary methods

A.1: Agarose DNA gel electrophoresis

DNA sample mixtures (final concentration of 1 X buffer) were prepared using 10 X DNA sample buffer followed by loading of mixture into agarose gel wells. TAE buffer (2 M Tris-Acetate, 0.05M EDTA, pH 8.3) 1 X was used to dissolved agarose 1 % (w/v) to prepare agarose gels which were further stained with ethidium bromide (0.05% w/v) to visualized DNA. Thereafter, the gel with loaded DNA was subjected to the electrophoresis (100 V) for 1 hour. The DNA was visualized by placing the gels on a short wave ultra-violet transilluminator.

A.2: Isolation and purification of DNA from agarose gels

The agarose gels were employed to run digested DNA samples and further excided with sterilized scalpel blade from the gel under short wave ultra-violet illumination. The gel was kept at minimal exposure time under UV and transferred into eppendorf tubes. The Zymoclean purification kit was used to purify excised DNA following manufacturer's protocol. DNA eluated with 30 µl dH_2O was further used or stored at -20 °C until further use.

A3: Bacterial strains and growth conditions

Growth media used in this study were all autoclaved (121^0C, at 0.1 MPa pressure) for 20 minutes and cooled to 55 C. The addition of appropriate antibiotics and inducers prior to use was done where necessary. **Table A.1** & **A.2** shows buffer and media preparations.

Table A. 1: Antibiotics and inducers used in this study

Antibiotics and inducers	**Preparation**
100 mg/ml Ampicillin	100 µg of Ampicillin sodium salt (Sigma) dissolved in 1ml of distilled water
50 mg/ml Kanamycin	50 µg of Kanamycin sulfate (Sigma) dissolved in 1ml of distilled water
1 M Isopropyl-β-Dthiogalactopyranoside (IPTG)	2.83 g Isopropyl B-D-thiogalactoside (IPTG) (Invitrogen) dissolved in 10 ml distilled water

Table A. 2: Buffers solutions and media used in this study

Buffer and medium	Composition (1L)	pH
Buffers 10 X Agarose gel running buffer	108 g Tris base; 55 g Boric acid; 7.45 g EDTA	8.3
Tris-HCl buffer	6.1 g Tris base adjust pH using 32 % HCl	7.0
10 X SDS-PAGE running buffer solution	30.0 g of Tris base, 144.0 g of glycine, and 10.0 g of SDS	8.3
Coomassie staining solution	10 % (v/v) acetic acid; 0.003 % (w/v) coomassie brilliant blue G	~1.0
SDS de-staining solution	10 % (v/v) acetic acid and 10 % (v/v) methanol	N/A
LEW buffer	7.8 g NaH_2PO_4 and 17.5 g NaCl. Adjust pH using in NaOH	8.0
1 X elution buffer	7.8 g NaH_2PO_4, 17.5 g NaCl, 480.5 g Urea, 17.0 g imidazole. Adjust pH using in NaOH	8.0
Media Luria Burtani medium (LB)	10 g Tryptone, 5 g Yeast extract and 10 g	7.0
Luria Burtani Agar	10 g Tryptone, 5 g Yeast extract, 10 g NaCl and 15 g agar	7.0

Appendix B: Supplementary Data

Table B. 1: DNA ligases and DNA ligase derivatives showing zeta potential and protein concentration

DNA ligases	Protein concentration (mg/ml)	Zeta potential (mV)
Ligsv081 control	0.670	-36.0 +/- 1.5
Ligsv081 non-specific PEGylated	0.525	-32.6 +/- 0.3
Ligsv081 site-specific PEGylated	1.032	-33.5 +/- 0.6
LigpET30 control	0.297	-32.1 +/- 0.2
LigpET30 non-specific PEGylated	0.265	-36.5 +/- 0.5
LigpET30 site-specific PEGylated	0.719	-29.3 +/- 0.6

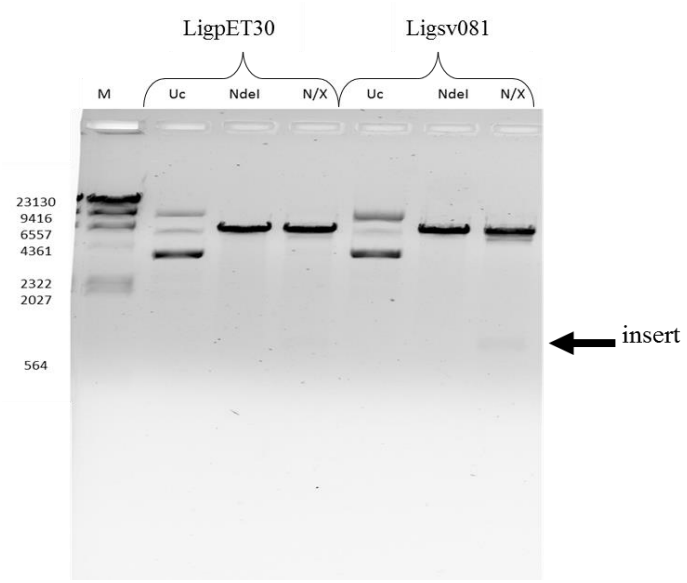

Figure B. 1: Agarose gel (2%) showing DNA products of DNA ligases. Lane M- molecular marker lambda DNA, Uc- uncut DNA products of DNA ligases, lane *NdeI*- single digested DNA with *NdeI* and lame N/X- double digested DNA with *NdeI* and *XhoI*.

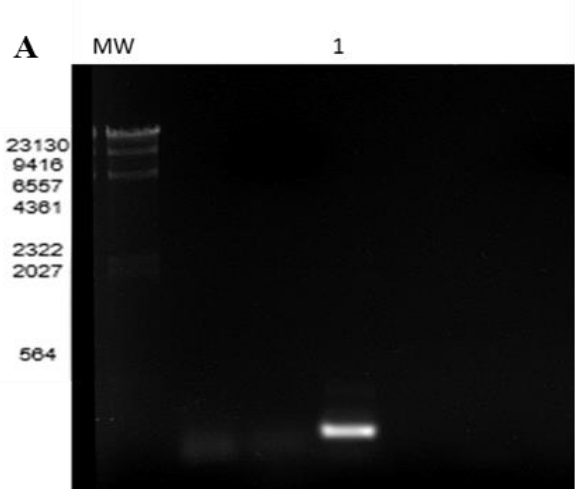

Figure B. 2: Agarose gel (2%) showing PCR product of transcription factor cTF and DNA ligase. **A**: MW; molecular marker lambda DNA and Lane 1; cTF.

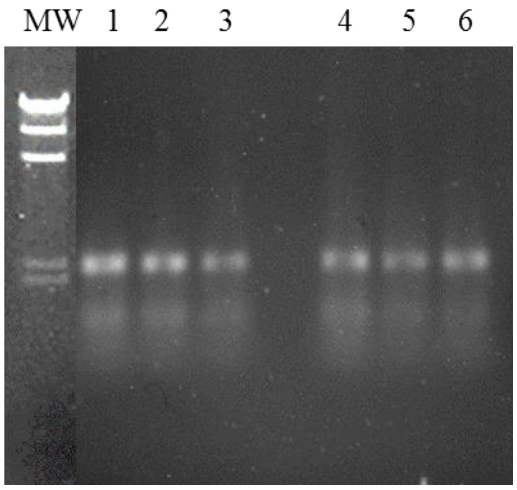

Figure B. 3: PCR amplification of DNA ligases and PEGylated DNA ligases. Lane MW- molecular marker lambda DNA (HindIII); lane 1- LigpET30 control; lane 2- LigpET30 non-specific PEGylated; Lane 3- LigpET30 site-specific PEGylated; Lane 4- Ligsv081 control; Lane 5- Lisv081 non-specific PEGylated and Lane 6- Ligsv081 site-specific PEGylated.

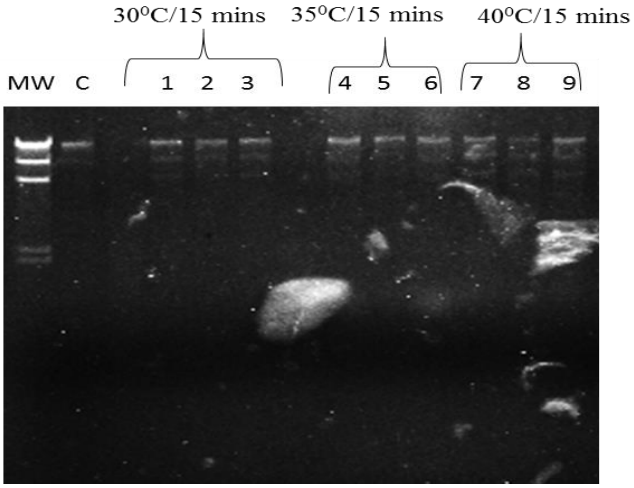

Figure B. 4: Agarose gel ligation assay (2%). Lane MW-molecular weight; lane c- T4 DNA ligase control; lane 1,4,7-Ligsv081; lane 2,5,8- LigpET30 and lane 3,6,9- LigpET30-PfHsp70.

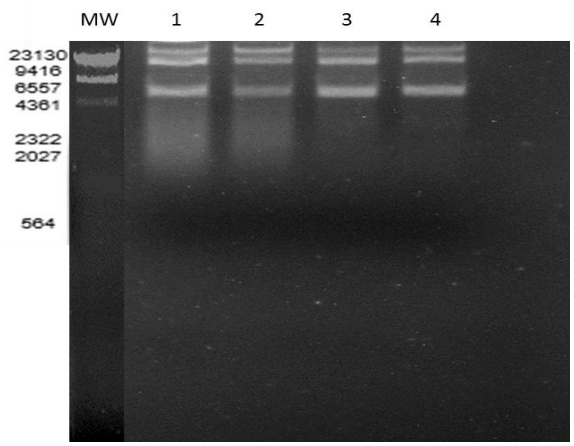

Figure B. 5: ligation activity of PEG conjugated DNA ligase proteins. Lane MV- Molecular maker, lane 1- LigpET30 non-specific PEGylated, Lane 2- LigpET30 site-specific PEGylated, lane 3- Ligsv081 non-specific PEGylated and Ligsv081 site-specific PEGylated.

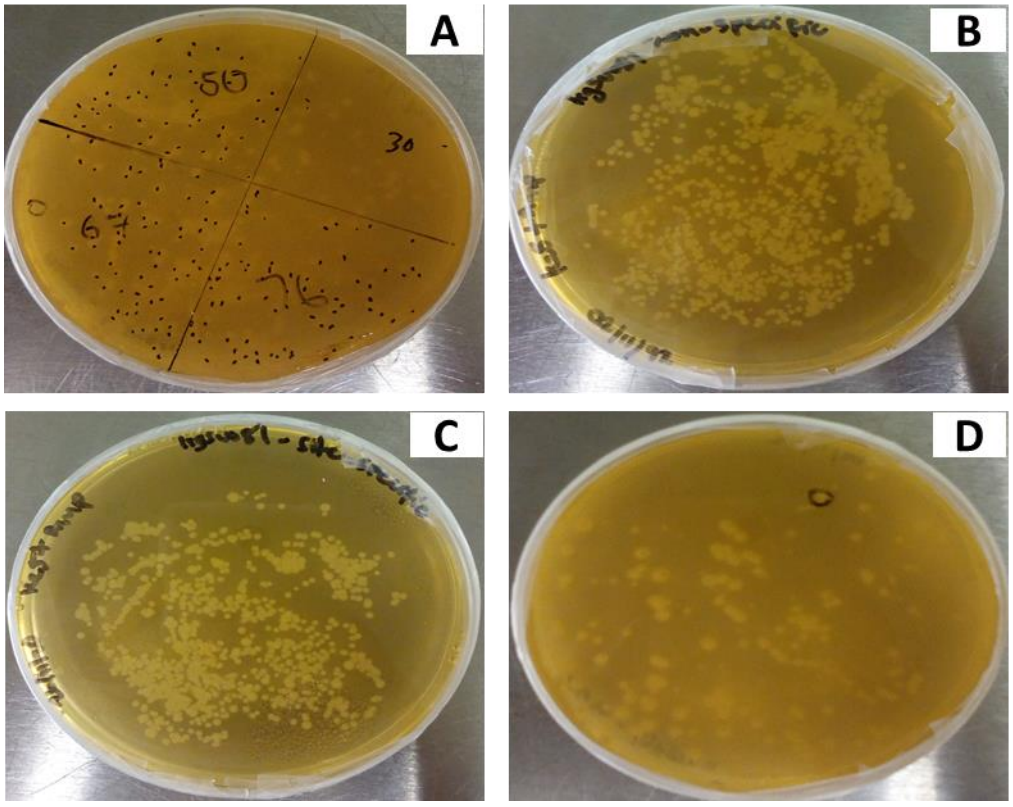

Figure B. 6: transformation efficiency of PEGylated and unmodified DNA ligases. **A**- DNA ligase- Ligsv081 control, **B**- Ligsv081 non-specific PEGylated, **C**- ligsv081 site-specific PEGylated and **D**- DNA ligase- LigpET30 control.

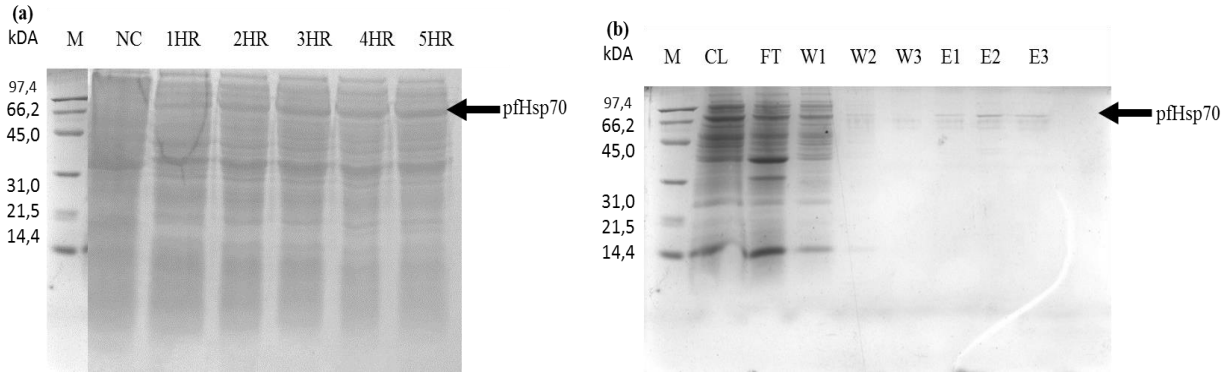

Figure B. 7: SDS-PAGE (15%) showing IPTG induced expression and purification for PfHsp70. **(a) :** Lane M; marker, Lane NC- *E. coli* BL21 control cells, Lane 1HR to 5HR represent the time –profile sampling from 1 hour to 6 hours post IPTG -induction. **(b):** Lane M: standard marker, lane CL: crude lysase; Lane FT: Flow through; lanes W1 to W3: wash one to wash 3; Lane E1 to E3 depicts the collected three protein elution samples.

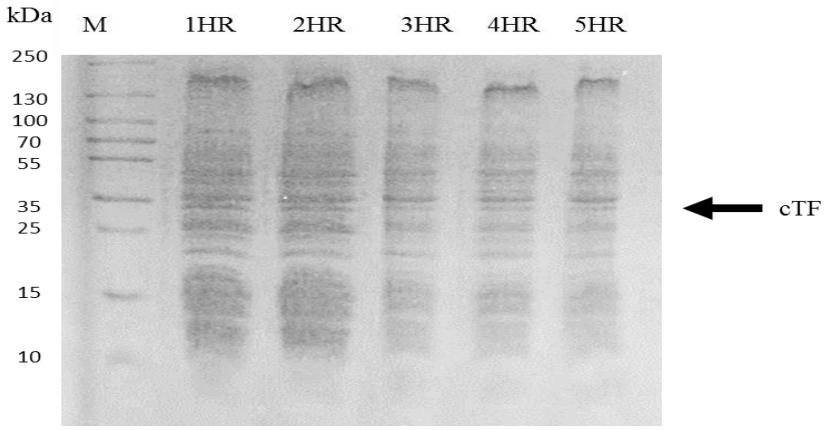

Figure B. 8: SDS-PAGE (15%) showing IPTG induced cTF protein expression. Lane M; marker, Lane 1HR to 5HR represent the time-profile sampling from 1 hour to 5 hours post IPTG -induction.

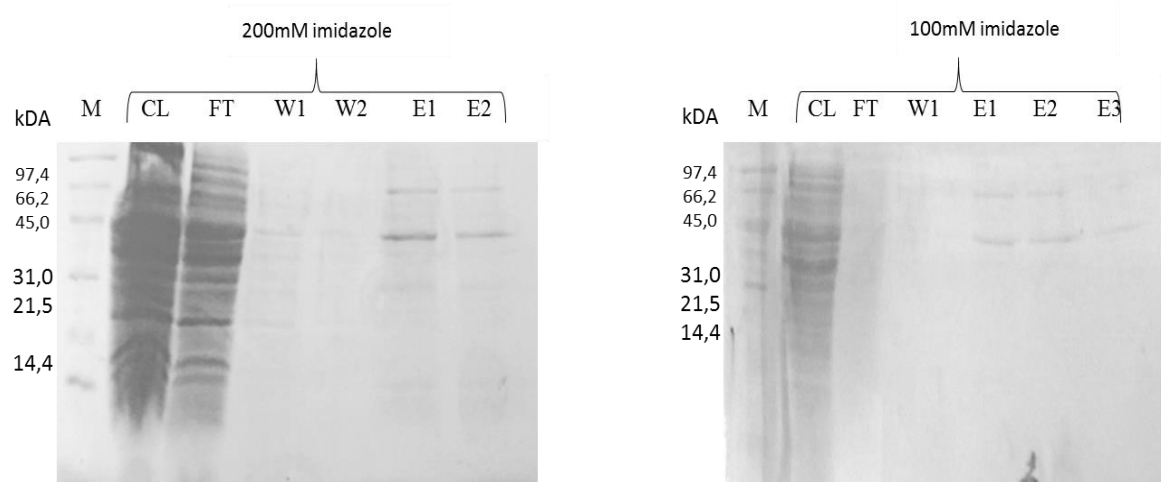

Figure B. 9: SDS-PAGE (15%) purification for LigpET30-PfHsp70 protein. Lane M: standard marker, Lane CL: crude lysate, Lane FT: Flow through; Lane E1- E4 depicts the collected two protein elution samples.

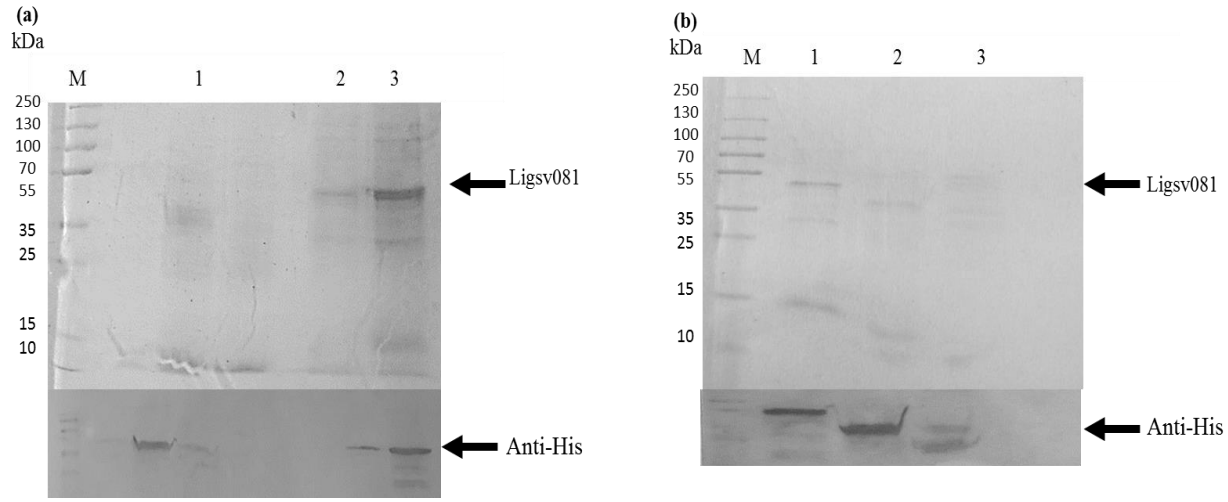

Figure B. 10: SDS-PAGE (15%) showing purified DNA ligase proteins (Ligsv081 & LigpET30) and non- and site-PEGylation. **(a)**: Lane M; marker (Thermo Science PageRuler plus Pre-stained protein Ladder), Lane 1; Ligsv081 control, lane 2; Ligsv081-non-specific PEGylated and lane 3; site-specific PEGylated. (**b)**: Lane M; marker, lane 1; LigpET30 control, Lane 2; ligpET30-non-specific PEGylated and LigpET30-site-specific PEGylated.

Appendix C: Reagents used in this study

All reagents were bought from Inqaba Biotec (South Africa) unless stated otherwise. Bacteriological agar, tryptone, sodium chloride and yeast extract were purchased from Merck (South Africa). The GeneJETM Plasmid Miniprep Kit, DNA size markers, T4 DNA ligase, Protein Page Ruler molecular weight markers and restriction enzymes were obtained from Thermo Scientific (South Africa). Oligonucleotide primers for polymerase chain reaction (PCR) were synthesised by Inqaba Biotec (South Africa). Protino® Ni-TED kit was purchased from Macherey Nagel (Germany). Antibiotics were purchased from Sigma-Aldrich (USA).