SPECIFICITY OF THE SH3 DOMAIN IN STAPHYLOCOCCAL ENDOLYSIN ENDO88 THROUGH *IN VITRO* AND *IN SILICO* MUTAGENESIS

By

THAM HONG YUN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

January 2025

FBSB 2023 11

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

SPECIFICITY OF THE SH3 DOMAIN IN STAPHYLOCOCCAL ENDOLYSIN ENDO88 THROUGH *IN VITRO* AND *IN SILICO* MUTAGENESIS

By

THAM HONG YUN

January 2025

Chairman : Adelene Song Ai Lian, PhD

Faculty : Biotechnology and Biomolecular Sciences

Bacteriophages produce endolysins at the end of their lytic cycle to degrade bacteria host cell walls and release progeny virions. Endolysins are being extensively studied as antimicrobial alternative due to rising antimicrobial resistance and as biosensors for their high specificity. Endolysins typically consist of enzymatically active domains (EADs) at the N-terminus, which degrade the bacterial peptidoglycan layer, and a C-terminal cell wall binding domain (CBD), which binds to specific peptidoglycan components and influences endolysin lytic host range. This study investigates the role of the CBD in lytic host range by analyzing the impact of amino acid modifications introduced through site-directed mutagenesis in the CBD of Endo88, an endolysin derived from *Staphylococcus aureus* phage 88. The mutations were designed based on two strategies; (i) comparison of primary structure of the SH3 domains with different host range, and (ii) comparison of the tertiary structure of predicted binding site of endolysins with different host range. Five mutant endolysins were generated, cloned,

and expressed in *Escherichia coli*, and their lytic and binding activities against an array of bacteria were compared to those of the wild-type Endo88. Lytic activity was evaluated using turbidity reduction assay which was optimized by Biophysical characterization of Endo88 including different buffers, temperature and pH. It was ascertain that Endo88 had optimum lytic activity in Tris buffer without NaCl, was active at pH of 6.0-11.0 and possessed lytic activity up to 37°C but quickly loses activity at 45°C. The binding activity of endolysins to various bacteria was assessed via a GFP-bound CBD flow cytometry fluorescence assay. The results revealed that while the mutations affected lytic and binding activity significantly, it did not alter the host range of Endo88. This suggests that CBD modifications may not necessarily affect host specificity but do influence binding activity and even lytic efficiency, despite the CBD having no lytic activity and the Endo88 being a modular protein. Further, flow cytometry analysis indicated no direct correlation between binding and lytic activity. This aligns with findings from previous studies, which have shown that some endolysins can retain lytic activity even in the absence of a CBD. Although the CBD is generally considered essential for maximal lysis, the data from this study suggest that binding may not always be critical for determining the host range. Nevertheless, the reduction in lytic activity caused by CBD mutations points to the CBD's auxiliary role in enhancing overall lysis in endolysin. Molecular docking of two mutants and Endo88 provided insights into binding site interactions with various peptidoglycan structures, which could aid in predicting host range alterations through endolysin engineering. This study highlights the intricate relationship between the CBD and EAD in modulating endolysins efficiency. Understanding this interplay could be contributory in developing more effective endolysins to combat antimicrobial resistance, especially in developing targeted phage endolysin therapies and diagnostics via biosensor development. Finally, this study contributes to the limited knowledge on

endolysin host range since most studies focusses on the lytic efficiency rather than the host-range.

Keywords: Binding specificity, Cell-wall binding domain (CBD), Endolysin, Lytic host range, Molecular docking.

SDG: GOAL 3: Good Health and Well-Being.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KEKHUSUSAN DOMAIN SH3 DALAM ENDOLISIN STAFILOKOKAL ENDO88 MELALUI MUTAGENESIS *IN VITRO* DAN *IN SILICO*

Oleh

THAM HONG YUN

Januari 2025

Pengerusi : Adelene Song Ai Lian, PhD

Fakulti : Bioteknologi dan Sains Biomolekul

Bakteriofaj menghasilkan endolisin pada penghujung kitaran lisis mereka untuk menguraikan dinding sel bakteria hos dan melepaskan virion progeni. Endolisin sedang dikaji secara meluas sebagai alternatif antimikrobial disebabkan oleh peningkatan rintangan antimikrobial dan sebagai biosensor kerana kespesifikan tinggi mereka. Endolisin biasanya terdiri daripada domain aktif enzimatik (EAD) pada terminus-N, yang menguraikan lapisan peptidoglikan bakteria, dan domain pengikat dinding sel terminal C (CBD), yang mengikat komponen peptidoglikan tertentu dan mempengaruhi julat hos lisis endolisin. Kajian ini mengkaji peranan CBD dalam julat hos lisis dengan menganalisis kesan pengubahsuaian asid amino yang diperkenalkan melalui mutagenesis terarah tapak dalam CBD Endo88, satu endolisin yang diperoleh dari faj *Staphylococcus aureus* 88. Mutasi telah direka berdasarkan dua strategi; (i) perbandingan struktur utama domain SH3 dengan julat hos yang berbeza, dan (ii) perbandingan struktur tertiari tapak ikatan yang diramalkan bagi endolisin dengan

julat hos yang berbeza. Lima endolisin mutan telah dihasilkan, diklon dan diekspresikan dalam *Escherichia coli*, dan aktiviti lisis serta aktiviti pengikatan mereka terhadap pelbagai bakteria dibandingkan dengan Endo88 jenis liar. Aktiviti lisis dinilai menggunakan ujian pengurangan kekeruhan yang dioptimumkan melalui pencirian biofizikal Endo88 termasuk penimbal, suhu, dan pH yang berbeza. Didapati bahawa Endo88 mempunyai aktiviti lisis optimum dalam penimbal Tris tanpa NaCl, aktif pada pH 6.0-11.0, dan mempunyai aktiviti lisis sehingga 37°C tetapi dengan cepat kehilangan aktiviti pada 45°C. Aktivitit pengikatan endolisin kepada pelbagai bakteria dinilai melalui ujian kekalimantangan aliran sitometri CBD yang terikat GFP. Keputusan menunjukkan bahawa walaupun mutasi memberi kesan signifikan kepada aktiviti lisis dan pengikatan, ia tidak mengubah julat hos Endo88. Ini mencadangkan bahawa pengubahsuaian CBD mungkin tidak semestinya mempengaruhi kespesifikan hos tetapi mempengaruhi aktiviti pengikatan dan bahkan kecekapan lisis, walaupun CBD tidak mempunyai aktiviti lisis dan Endo88 adalah protein modular. Selain itu, analisis sitometri aliran menunjukkan tiada korelasi langsung antara aktiviti pengikatan dan lisis. Ini selaras dengan penemuan kajian sebelum ini, yang menunjukkan bahawa beberapa endolisin boleh mengekalkan aktiviti lisis walaupun tanpa CBD. Walaupun CBD secara amnya dianggap penting untuk lisis maksimum, data dari kajian ini mencadangkan bahawa pengikatan mungkin tidak selalu kritikal untuk menentukan julat hos. Namun begitu, pengurangan dalam aktiviti lisis yang disebabkan oleh mutasi CBD menunjukkan peranan auksiliari CBD dalam meningkatkan keseluruhan lisis dalam endolisin. Dok molekul bagi dua mutan dan Endo88 memberikan pandangan tentang interaksi tapak pengikatan dengan pelbagai struktur peptidoglikan, yang boleh membantu dalam meramalkan perubahan julat hos melalui kejuruteraan endolisin. Kajian ini menyoroti hubungan rumit antara CBD dan EAD dalam memodulasi kecekapan endolisin.

Pemahaman hubungan ini boleh menyumbang kepada pembangunan endolisin yang lebih berkesan untuk melawan rintangan antimikrobial, terutamanya dalam membangunkan terapi endolisin faj yang disasarkan dan diagnostik melalui pembangunan biosensor. Akhir sekali, kajian ini menyumbang kepada pengetahuan terhad mengenai julat hos endolisin memandangkan kebanyakan kajian menumpukan pada kecekapan lisis berbanding julat hos.

Kata Kunci: Kespesifikasi pengikatan, Domain pengikat dinding sel (CBD), Endolisin, Pemodelan Homologi, Julat hos lisis.

SDG: MATLAMAT 3: Kesihatan Baik dan Kesejahteraan

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Adelene Ai-Lian Song. I am very lucky to have her as my supervisor; she has always been helpful and a true motivator whenever I felt lost in the studies. I would also like to thanks my co-supervisors, Prof. Datin Paduka Dr. Khatijah, Assoc. Prof. Dr Tan Geok Hun, and Prof. Dr. Asif Mohammad Khan, who guided me throughout the journey. Without their support, I would not have made it through.

I would also like to thank those who lent me a hand whenever I encountered bottlenecks in my research, especially Ms Chong from Perdana University, Assoc. Prof. Dr. Sy Bing Choi, and Dr Frederick, Jojo, Assoc. Prof. Dr. Eddie, Dr Mizi. Also, thanks to my family, who have provided unwavering support throughout this journey.

Of course, also my fellow lab mates,

The knowledge we share, and the potluck we prepared,

Party in the lab, and the reagents we shared,

Experiments run late, Fatimah will be there.

From pipettes to tubes, every step paired,

With data and deadlines, we never run away,

LB broth to be autoclave, do you have extra space.

A thousand words of appreciation, one heartfelt thanks say it all – Thank you!

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Adelene Song Ai Lian, PhD
Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Chairman)

Datin Paduka Khatijah binti Yusoff, PhD
Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Member)

Mohammad Asif Khan, PhD
Professor
University of Doha for Science & Technology, Qatar
(Member)

Tan Geok Hun, PhD
Associate Professor
Faculty of Agriculture
Universiti Putra Malaysia
(Member)

_____
ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 8 May 2025

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AH	Ambiguous host range
AIR	Ambiguous restraint
AMD	Amidase domain
ATCC	American Type Culture Collection
AVANA	Antigenic Variability Analyser
BCA	Bicinchoninic acid assay
BHI	Brain Heart Infusion
BSA	Bovine serum albumin
CBD	Cell wall binding domain
CFU	colony forming unit
CHAP	Cysteine, Histidine-dependent amidohydrolases/peptidase
Cryo-EM	EM-Cryo-electron microscopy
DNA	deoxyribonucleic acid
EAD	Enzymatically active domain
EDTA	Ethylenediaminetetraacetic acid
Fwd	Forward
GFP	Green-fluorescent protein
GlcNAc	N-acetylglucosamine
HADDOCK	High ambiguity driven protein-protein DOCking
HCSS	Highly conserved serotype-specific sequence
HCL	Hydrochloric acid
IMPs	Immunomagnetic particles
IPTG	β-D-1-thiogalactopyranoside

LB	Luria Bertani
MAFFT	Multiple Alignment using Fast Fourier Transform
MD	Molecular dynamic
MH	Medium host range
MI	Mutual Information
MRS	De Man, Rogosa and Sharpe broth
MRSA	Methicilin-resistant Staphylococcus aureus
MSA	Multiple sequence alignment
MurNAc	N-acetyl muramic acid
NCBI	National Center for Biotechnology Information
NH	Narrow host range
NMR	Nuclear magnetic resonance
NaCl	Sodium chloride
OD	Optical density
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PG	Peptidoglycan
PLA	Plate Lysis assay
PLIP	Protein-Ligand interaction profiler
PVDF	Polyvinylidene difluoride membrane
RE	Restriction enzyme
Rev	Reverse
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SH3	Src Homology 3
SOE-PCR	Splicing by overlap extension PCR

TAE	Tris base, acetic acid and EDTA
TBST	Tris-buffered Saline with Tween-20
TRA	Turbidity reduction assay
U	unit
UIR	Un-ambiguous restraint
WH	Wide host range
g	g force
Ml	milliliters
mM	millimolar
ng	nanogram
µL	microliter
µg	microgram
µM	micromolar

CHAPTER 1

INTRODUCTION

1.1 Background

The emergence of antimicrobial resistant bacteria (superbug) such as Methicilin-resistant *Staphylococcus aureus* (MRSA) poses a significant challenge in healthcare settings and the community (N. A. Turner et al., 2019). The spread of superbug is mainly attributed to the misuse of antibiotics (Harkins et al., 2017), and the overreliance to antibiotic, which exerts selective pressure favouring the development of resistant strains. The emergence of the SARS-CoV-2 pandemic has heightened the demand for antibiotics in managing secondary bacterial infections in critically ill Covid-19 patients, thus further exacerbating the antimicrobial resistance crisis (Arshad et al., n.d.; Garg, 2021; Pandak et al., 2023; Shah et al., 2023). The surge in demand is likely contributed by nosocomial infections among these patients. Therefore, the need for alternative antimicrobial compounds to antibiotics is crucial in combating antibiotic resistance. The prospective antimicrobial activity of endolysins has been studied intensively over recent years, and the result has been promising (Chandran et al., 2022; Gondil et al., 2020; Rahman et al., 2021). There are currently two endolysin-based commercialized products which are Staphefekt™ [developed by Micreos (https://www.staphefekt.com/en/)] and Artilysin [developed by Lysando (https://www.lysando.com/overview.html)]. Staphefekt mainly targets MRSA (Totté et al., 2017), while Artilysin mainly targets multidrug resistance Gram-negative bacteria (Briers et al., 2014).

Endolysins are bacteriophage-encoded peptidoglycan hydrolases which hydrolyze the bacterial peptidoglycan layer. These enzymes are produced within the host cell at the end of the phage lytic cycle. Their primary function is to lyse the host cell from within, leading to bacterial lysis and the subsequent release of virion progeny (Loessner, 2005; Schmelcher, Donovan, et al., 2012). This process is facilitated by another phage protein, holin, which forms a pore in the inner bacterial membrane and allows passage of endolysin to the peptidoglycan layer (Lin et al., 2017; Loessner et al., 1999; I. N. Wang et al., 2000). When applied externally, recombinant endolysins exhibit the same effect, without the need of holin (Haddad Kashani et al., 2017; Melo et al., 2018), thus holding promise for therapeutic applications as an alternative to antibiotics (Gondil et al., 2020; Heselpoth et al., 2017). It is noteworthy that phages itself typically display a relatively narrow lytic host range (Rahman et al., 2021), targeting specific strains within species. In contrast, recombinant endolysin often demonstrate a broader host range, enabling them to lyse bacteria across species or genera (Abdelkader et al., 2019; Maciejewska et al., 2018; Murray et al., 2021; Yoong et al., 2004).

Gram-positive phage endolysin mostly exists as modular proteins consisting of several functional domain, including the N-terminal enzymatically active domain (EAD) and the cell wall binding domain (CBD) (Haddad Kashani et al., 2017; Maervoet & Briers, 2016). The EAD is responsible for the degradation of the peptidoglycan (PG), while the CBD recognizes and facilitates endolysin binding to bacterial peptidoglycan (PG), leading to host lysis (Bateman & Rawlings, 2003; Gutiérrez et al., 2018). The CBD is believed to play a pivotal role in determining the lytic host range and affinity of the endolysin (Y. Chang & Ryu, 2017; Loessner, 2005; Son et al., 2018). Staphylococcal phage endolysin typically consist of two EADs - the Cysteine, Histidine-dependent

amidohyrolases/peptidases (CHAP) domain and the Amidase domain (AMD), along with a CBD – the Src Homology 3 (SH3) domain (Benešík et al., 2018; Gutiérrez et al., 2018; Jarábková et al., 2015). Notably, some staphylococcal phage endolysin lacks the AMD in their structure (Y. Chang & Ryu, 2017).

The SH3 domain is ubiquitous across various organisms and was first found for its involvement in protein-protein signaling within eukaryotic cells (Mayer & Eck, 1995). This domain is also commonly present in staphylococcal phage endolysins, and is proposed to confer host range specificity to the endolysin (S. C. Becker, Foster-Frey, et al., 2009; Loessner, 2005). Specifically, the SH3 domain binds to the interpeptide bridge within bacterial peptidoglycan which varies among bacterial species/genera (Gründling & Schneewind, 2006; Mitkowski et al., 2019; Schleifer & Kandler, 1972). In the case of *Staphylococcus aureus*, studies suggested that the SH3 domain recognizes the pentaglycine interpeptide bridge (Y. Chang & Ryu, 2017; Gonzalez-Delgado et al., 2020; Gu et al., 2014; Jarábková et al., 2015).

Despite high similarities in sequences and structure, many endolysins including SH3-domain containing endolysins exhibit varying degree of lytic host range. Some have a narrow host range, with observed capability of lysing within species, while other have a wider host range, capable of lysing across genera (Huang et al., 2015; Melo et al., 2018; Yokoi et al., 2005; Yoong et al., 2004). The mechanism by which the CBD influences endolysin lytic host range remains elusive. Current research predominantly involves CBD domain swapping, truncation or addition, but no specific functional or structural motif that determine the host range have been identified within the CBD.

While endolysins are mostly studied for its antimicrobial properties, it is also worth mentioning that the CBD of endolysins is also used to develop biosensors for detection of bacteria, especially in diagnostics against pathogens (Al-Hindi et al., 2022; Cho et al., 2021). As such, the host-range of endolysins is important depending on its application. In therapy, endolysins should be specific enough that it does not harm host beneficial microbes yet not too specific that it is only effective against certain strains or serovar. In diagnostics, customizable endolysins which can detect either the species or genus level depending on the needs, is desirable.

Thus, the current study aimed to elucidate the relationship between endolysin-peptidoglycan binding and target host range by determining if there are specific factors, structure or motifs within the CBD of an endolysin which affects host-range. A staphylococcal endolysin, Endo88 was chosen as the model protein, since staphylococcal endolysins are among the most studied and established endolysin. Endo88 was subjected to a series of mutations with the aim to alter the host range and to elucidate the factors which affects host-range.

1.2 Problem statement

The structure and function of CBD is similar amongst phage's endolysin, yet different endolysins possess different degrees of lytic host range. The underlying mechanism by which the CBD influences the host range specificity of the endolysin remains unclear. Current research predominantly involves CBD domain swapping, truncation or addition, but no specific functional or structural motifs that determine the host range have been identified within the CBD. Therefore, this study hypothesized that mutation introduced into the SH3 domain of the *Staphylococcus aureus* phage 88 endolysin will

alter its binding specificity and lytic host range, resulting in measurable differences in lytic activity and bacterial binding affinity as compared to the wild-type endolysin.

1.3 Objectives

The general objective of this study is to elucidate the relationship between endolysin-peptidoglycan binding and target host range by altering the lytic host range through a mutagenesis approach. Specific objectives include:

I. To design and validate targeted mutations of the wild-type *Staphylococcus aureus* phage 88 endolysin CBD through bioinformatics approaches for potential alteration of endolysin target host-range.

II. To optimize the cloning and expression conditions for the wild-type Staphylococcal phage 88's endolysin and mutants endolysin in *E. coli*.

III. To evaluate the lytic activity and host-range of wild-type and mutant endolysins through quantitative assays.

IV. To assess the correlation between lytic activity and binding affinity of mutant endolysins to various bacterial strains using CBD-GFP fusion protein.

CHAPTER 2

LITERATURE REVIEW

2.1 Bacteriophage endolysin

Endolysin is a lytic enzyme within the peptidoglycan hydrolases (PGHs) enzyme family (Jarábková et al., 2015). It is produced at the end of a bacteriophage/ phage lytic cycle. Prior to lysis, a protein, holin, creates pores in the inner bacterial membrane, allowing endolysins access to the peptidoglycan layer (Loessner et al., 1999; I. N. Wang et al., 2000; Young, 1992). The peptidoglycan layer will then be digested by endolysins which results in cell lysis, subsequently releasing newly formed virion progeny (Gutiérrez et al., 2018; Loessner et al., 1998). This process, known as "Lysis-from-within", involves endolysins lysing the cell internally which is a natural phenomenon (Young, 1992). Conversely, "Lysis-from-without" is a human-mediated phenomenon which occurs when externally applied recombinant endolysins result in bacterial cell death without the presence of holin since the endolysins is directly in contact with the cell wall when applied exogenously (Abedon, 2011). In view of the ability to act externally, the endolysin are considered as potential alternatives to antibiotics (Haddad Kashani et al., 2017).

Endolysins can be either globular or modular protein, with modular endolysins being more common in Gram-positive phages (Haddad Kashani et al., 2017; Maervoet & Briers, 2016). Modular endolysin consist of two main domains with distinct function, including the Enzymatically Active Domain (EAD) and the Cell wall Binding Domain (CBD). The EAD is responsible for digesting the peptidoglycan layer, whereas CBD

facilitates the binding of endolysin to bacterial peptidoglycan, and is believed to confer lytic specificity to the endolysin (Loessner, 2005; Son et al., 2018).

Unlike Gram-negative bacteria, Gram-positive bacterial PG display more diverse PG compositions and arrangements, particularly within the peptide stem region (Bouhss et al., 2002; Schleifer & Kandler, 1972; Vollmer et al., 2008). This diversity in peptidoglycan components may create selective pressure, resulting in the varieties in EAD and CBD observed in Gram-positive phage endolysin (Figure 2.1) (Oechslin et al., 2022; Schmelcher et al., 2015).

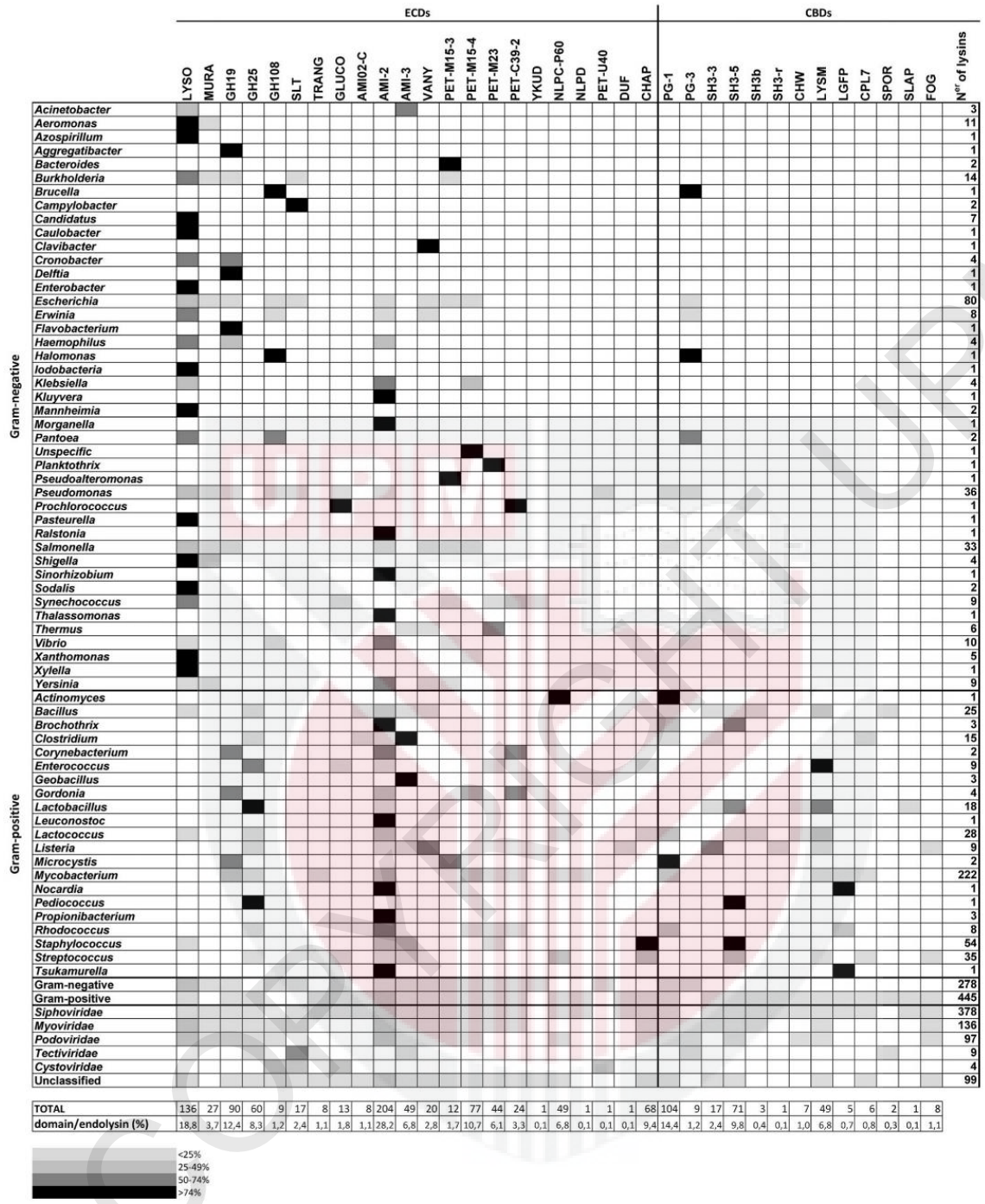


Figure 2.1: Diversity of EAD and CBD in phage endolysins. Total of 727 putative endolysins were analysed and annotated (Adapted from Oliveira et al., 2013).

Despite conservation in biological function, the EAD and CBD are diverse amongst bacteria in terms of target site and structure. This reflects the specificity and versatility of endolysins in lysing their host. This modular architecture of endolysin is believed to have evolved through homologous recombination between phages infecting the same host (Oechslin et al., 2022), or heterologous recombination between phages and their bacteria host (García et al., 1988). The diversity in modular endolysin could be attributed to evolutionary pressure on phages to adapt to diverse and ever-changing bacterial cell wall structures (Labrie et al., 2010). Moreover, the presence of various EAD and CBD domains and their diverse combinations in endolysins provide possibilities for host-range expansion, including cross-genus lysis (Liu et al., 2023; Vázquez et al., 2021). For example, although CHAP and SH3 domains are predominant in staphylococcal phage endolysins, a small number of other bacteria genus have been found to acquire these domains (Vázquez et al., 2021). The biological diversity of these combination remains unknown. One of the speculations are that the exchange of domains occurred at some point during the past evolution, followed by evolutionary selective pressure that favored certain combinations of lytic systems. These favored combinations persisted in phages due to their fitness advantage (Oechslin et al., 2022; Oliveira et al., 2013). However, logically, in evolution, the fitness cost will always be prioritized, and domains that are favored by the selective pressure (components on peptidoglycan) will be inherited (Oechslin et al., 2022). This explains why some domains are prevalent in an endolysin targeting a specific genus, while others are unique to certain endolysins. For instance, while CHAP and SH3 domain are commonly found in staphylococcal phage's endolysin (S. C. Becker, Foster-Frey, et al., 2009; Vázquez et al., 2021), the CW_binding_1 CBD in

streptococcal phage endolysins are rarely found in endolysins from other genus of phages (Bustamante et al., 2017; Vázquez et al., 2021).

2.2 Modular Mechanism of Gram-Positive Phage Endolysins in Peptidoglycan Lysis

Bacteriophage have evolved specialised mechanisms to lyse their bacterial hosts and release virion progeny (Young, 1992). In Gram-positive phages, this process involves the coordinated action of holins and endolysins (Jacob et al., 1957; Leprince et al., 2022; Loessner et al., 1999; I. N. Wang et al., 2000). Holins are small transmembrane proteins that perforate the bacterial inner membrane at the end of the phage lytic cycle, allowing endolysins to acess the peptidoglycan (C. Y. Chang et al., 1995; Gründling et al., 2000). Holins are able to regulate the lysis timing which is crucial, as premature lysis resulted in the realease of immature virion progeny, whereas delayed lysis may slow down infection cycles to a new host and reduce phage propagaiton (Abeysekera et al., 2022; Leprince et al., 2022). Once access is granted, the EAD catalyzes the cleavage of peptidoglycan bonds, while the CBD facilitates recognition and attachment to specific cell wall structures. This modular architecture enhances substrate specificity and lytic efficiency. Upon peptidoglycan degradation, the loss of structure intergrity results in increased osmotic pressure, ultimately causing bacterial lysis and release of virion progeny (Jarábková et al., 2015; Liu et al., 2023; Schmelcher, Donovan, et al., 2012).

2.2.1 Synergistic interactions between EADs and CBDs

Although the function of EADs in bacterial lysis is well characterized, the role of CBDs remains controversial (Phothichaisri et al., 2022). Some studies suggest that CBD removal does not impair, and may even enhance lytic activity, indicating that CBD binding is not always necessary for cell wall degradation (Cheng & Fischetti, 2007; Donovan, Dong, et al., 2006). Conversely, other studies demonstrated reduction in lysis efficiency upon CBD removal, supporting its role in enhancing endolysin lytic activity (S. C. Becker, Dong, et al., 2009; Gu et al., 2014; Huang et al., 2015). While domain truncation studies provide insight into the role of individual domain, growing evidence suggests that EAD and CBD exhibit structural interdependence (Oechslin et al., 2022). The modular nature of endolysins implies that interactions between the EAD and CBD may modulate enzymatic activity synergistically.

For instance, PlySK1249 a endolysin with Amidase doamin and CHAP, also contains a LysM as CBD. While direct lytic was not observed on CHAP domain alone, it exihibit a dechaining effect on *Streptococcus dysgalactiae*. (Oechslin et al., 2013). A similar effect was observed in CHAP domain in Cse autolysin from *Streptococcus thermophiles* (Layec et al., 2009).

Additionally, truncation studies of staphylococcal endolysins revealed that the amidase domain alone was inactive (Benešík et al., 2018; Horgan et al., 2009). However, its consistent presence in many staphylococcal phage endolysins suggests and evolutionarily conserved function (S. C. Becker, Foster-Frey, et al., 2009; Y. Chang & Ryu, 2017; Jarábková et al., 2015). In some case, amidase domain in

LysSA12 and LysSA97 was enzymatically inactive, but its presence enhanced the binding of CBD (Son et al., 2018). Similar observations were reported in phi11 (Donovan, Lardeo, et al., 2006) and E-LM12 (Costa et al., 2020).

2.2.2 Types of Enzymatically Active Domains (EAD) and Cell wall Binding Domain (CBD)

Due to the high variability in PG composition, different endolysins have evolved different EAD and CBD to target different bond within the PG (Broendum et al., 2018; Oechslin et al., 2022). These enzymes are classified based on the type of bond they cleave within the peptidoglycan layer:

(I) *N*-acetylglucosaminidase target the reducing end of the GlcNAc within the glycan chain. Example endolysin: streptococcal C1 phage endolysin, PlyC (McGowan et al., 2012).

(II) *N*-acetylmuramidase target the reducing side of the MurNAc. Example endolysin: pneumococcal Cp-1 phage endolysin, Cpl-1 (Hermoso et al., 2003); *Listeria* phage P40 endolysin, Ply40 (Romero et al., 2018).

(III) Transglycosylases, disrupt the PG recycling by forming non-reducing *N*- acetyl 1,6-anhydromuramic acid and *N*-acetyleglucosamine (Suvorov et al., 2008). Example endolysin: Phage phiKZ endolysin, Gp144 (Fokine et al., 2008; Paradis-Bleau et al., 2007).

(IV) Metal-dependent PG amidases and **endopeptidase** cleave the amide bond between the MurNAc sugar backbone and the peptide crosslink or cleave the peptide bonds between the stem peptide and the peptide crosslink (endopeptidase). Example endolysin: LambdaBa02 *Bacillus* phage endolysin, PlyL (Low et al., 2005); *Bacillus*

subtilis prophage endolysin, XlyA (Low et al., 2011); LysGH15 (Gu et al., 2014); LysK (S. C. Becker, Dong, et al., 2009; Sanz-Gaitero & van Raaij, 2021), Ply500 (Korndörfer et al., 2008).

(V) Cysteine, histidine-dependent Aminohydrolase/peptidase (CHAP) **domains** are not named based on the specific bond they cleave in the peptidoglycan, but on their catalytic mechanism dependent on the cysteine and histidine residue in their active site (Bateman & Rawlings, 2003). Two commonly reported target site for CHAP are amidase (*e.g. L*-muromoyl-L-alanine amidase) (Bateman & Rawlings, 2003; Nelson et al., 2006) and endopeptidase (*e.g. D*-alanyl-glycyl endopeptidase) (S. C. Becker, Dong, et al., 2009; J. Z. Lu et al., 2006), some even comprising of both activities (Linden et al., 2015).

The CBD is ubiquitous and diverse among phage endolysins, especially in Gram positive phage endolysins. There are 6 common groups of CBD found in phage endolysins (Jarábková et al., 2015):

(I) Lysin Motif (LysM), which was observed to target the *N*-acetylglucosamine (GlcNAc) residues in the glycan strand of the peptidoglycan (Garvey et al., 1986; Ohnuma et al., 2008);

(II) **Peptidoglycan Binding Domain (PG binding)**. Protein family commonly consists of three-helical bundle (Dideberg et al., 1982; Krogh et al., 1998). It was proposed to only target D-Asn in PG crosslink, which are found in prophage's endolysins from *Lactobacillus casei* BL23 (Lc-Lys and Lc-Lys-2 (Regulski et al., 2013).

(III) Choline-binding Modules. It recognizes choline rich regions on the cell wall of *Streptococcus pneumonia* (Buey et al., 2007; Hermoso et al., 2003, 2007; Maestro & Sanz, 2016)

(IV) Clostridial Hydrophobic with Conserved Tryptophan (CHW). The domain family was first observed targeting the cell wall of *Clostridium acetobutylicum* (Sullivan et al., 2007). It was also found on *Lactobacillus* phage A2 (Conserved domain: smart00728) and *Lactococcus* phage 949 (Conserved domain: PF07538/IPR006637) (Oliveira et al., 2013).

(V) SRC Homology 3 (SH3). Mostly found in *Staphylococcus* phage endolysin (Jarábková et al., 2015), including Endo88 (Chandran et al., 2022) used in this study and described below. It can also be found in *Streptococcus* spp. (Huang et al., 2015; Ji et al., 2015; Tang et al., 2013), *Enterococcus* (Yoong et al., 2004) and *Bacillus* spp. (Etobayeva et al., 2018; Son et al., 2012).

2.3 Domain swapping/shuffling in endolysins

Due to the vast diversity of EAD and CBD available in phage endolysins, researchers have attempted to construct optimised synthetic endolysins through domain swapping/shuffling (Table 2.1). This strategy has also been widely used to investigate endolysin host range by exchanging domains between different endolysins (Table 2.1) (Dong et al., 2015; Jarábková et al., 2015; Son et al., 2021; H. Yang et al., 2014).

While domain swapping/shuffling is often associated with engineered constructs to enhance specificity or activity, there is no direct evidence that this process occurs naturally. However, under *in vitro* conditions, studies have shown that phages co-infecting the same bacterial host can exchange genetic materials, including endolysin genes, through homologous recombination between virulent phage and prophage

(Oechslin et al., 2022). Although these recombination events may not necessarily lead to direct domain swapping, they suggest a potential mechanism by which phages could acquire and rearrange endolysin domains over evolutionary timescales. Given the vast diversity of endolysin domains discovered, the acquisition of specific domain combinations may be influenced by the host peptidoglycan composition (Oliveira et al., 2013). This raises the possibility that natural selection, rather than frequent recombination, plays a greater role in shaping the modular architecture of endolysins in response to variations in bacterial cell wall structure.

Table 2.1: Chimeric endolysin with enhanced/extended lytic host range.

Endolysin	**Lytic host range of the parental endolysin (genus)**	**References**
Parental N-terminal Ply187_CHAP	*Staphylococcus aureus, S. albus, S.epidermidis,*	(Bateman & Rawlings, 2003; Dong et al., 2015; Loessner et al., 1999)
Parental C-terminal LysK_SH3	*Staphylococcus* spp.	(O'Flaherty et al., 2005)
Mutant Ply187CHAP – LyskSH3	*Staphylococcus* spp. (Ten-fold increase)	(Mao et al., 2013; P. K. Singh et al., 2014)
Parental N-terminal Ply187_CHAP	*Staphylococcus aureus, S. albus, S.epidermidis,*	(Bateman & Rawlings, 2003; Dong et al., 2015; Loessner et al., 1999)
Parental C-terminal PlyV12_SH3	*Staphylococcus* spp., *Streptococcus* spp., *Enterococcus* spp.	(Yoong et al., 2004)
Mutant Ply187CHAP - PlyV12SH3	*Staphylococcus* spp., *Streptococcus* spp., *Enterococcus* spp.	(Dong et al., 2015)

Table 2.1: Continued

Parental N-terminal Lys168_CHAP	*Enterococcus faecalis, Enterococcus faecium*	(Fernandes et al., 2012; Proença et al., 2012)
Parental C-terminal Lys87_CBD	*Staphylococcus, Micrococcus, Bacillus, Enterococcus*.	(Fernandes et al., 2012)
Mutant Lys168CHAP - Lys87CBD	*Staphylococcus* spp., *Streptococcus pyogenes, Enterococcus* spp.	(Fernandes et al., 2012)
Parental N-terminal Lys170_AMD	*Enterococcus faecalis, Enterococcus faecium*	(Fernandes et al., 2012; Proença et al., 2012)
Parental C-terminal Lys87_CBD	*Staphylococcus, Micrococcus, Bacillus, Enterococcus*.	(Fernandes et al., 2012)
Mutants Lys170AMD – LysCBD	*Staphylococcus* spp., *Streptococcus pyogenes, Entrococcus* spp.	(Fernandes et al., 2012)
Parental N-terminal LambdaSa2_endopeptidase	*Streptococcus* sp. *Staphylococcus* sp. (weak activity)	(Donovan & Foster-Frey, 2008; Pritchard et al., 2007)
Parental C-terminal Lysostaphin, SH3b like domain	*Staphylococcus* spp.	(Bastos et al., 2010; S. C. Becker et al., 2008; DeHart et al., 1995; Gonzalez-Delgado et al., 2020; Recsei et al., 1987)
Mutant LambdaSa2_E Lyso-SH3b	*Streptococcus* sp. *Staphylococcus* sp. (increased in lytic activity)	(S. C. Becker, Foster-Frey, et al., 2009; Schmelcher, Korobova, et al., 2012)

Table 2.1: Continued

Parental N-terminal LambdaSa2_endopeptidase	*Streptococcus* sp. *Staphylococcus* sp. (weak activity)	(Donovan & Foster-Frey, 2008; Pritchard et al., 2007)
Parental C-terminal LysK_SH3	*Staphylococcus* spp.	(S. C. Becker, Dong, et al., 2009; Horgan et al., 2009; Mao et al., 2013; O'Flaherty et al., 2005)
Mutant LambdaSa2_endopeptidase - LysKSH3	*Streptococcus* sp. *Staphylococcus* sp. (increased in lytic activity)	(S. C. Becker, Foster-Frey, et al., 2009; Schmelcher, Korobova, et al., 2012)
Parental N-terminal Ply187_CHAP	*Staphylococcus aureus, S. albus, S.epidermidis,*	(Bateman & Rawlings, 2003; Dong et al., 2015; Loessner et al., 1999)
Parental C-terminal LysK_SH3	*Staphylococcus* sp.	(S. C. Becker, Dong, et al., 2009; Horgan et al., 2009; Mao et al., 2013; O'Flaherty et al., 2005)
Mutant Ply187CHAP LysKSH3	*Staphylococcus* sp., *Streptococcus dysgalactiae* (weak activity)	Mao et al., 2013; O'Flaherty et al., 2005; Schmelcher et al., 2015)
Parental N-terminal B30 lysin_endopeptidase	*Streptococcus agalactiae, Streptococcus dysgalactiae, Streptococcus uberis*	(Donovan, Dong, et al., 2006; Pritchard et al., 2004)
Parental C-terminal Lysostaphin, SH3b like domain	*Staphylococcus* spp.	(Bastos et al., 2010; S. C. Becker et al., 2008; DeHart et al., 1995; Gonzalez-Delgado et al., 2020; Recsei et al., 1987)
Mutant B30 lysin_endopeptidase – Lysp-SH3b	*S. aureus, Streptococcus agalactiae, Streptococcus dysgalactiae, Streptococcus uberis*	(Donovan, Dong, et al., 2006)

For instance, the domain swapping between LysK and Ply187. Ply187 was first studied and found in *S. aureus* bacteriophage 187 (Loessner et al., 1999). It consists of a N-terminal CHAP domain and a C-terminal glucosaminidase domain with an unknown CBD (Bateman & Rawlings, 2003; Loessner et al., 1999). LysK from staphylococcal phage K features two EAD (CHAP & AMD) and one CBD (SH3), displaying high lytic activity against *Staphylococcus* .spp (MRSA, *S. epidermidis, S. saprophyticus, S. chromogenes, S. capitis, S. hominis, S. haemolyticus, S. caprae, S. hyicus*) (O'Flaherty et al., 2005), whereas the C-terminal truncated Ply187 exhibits enhanced activity against *S. aureus* compared to the full-length Ply187 (Loessner et al., 1999). The fusion of the CHAP from Ply187 and the CBD from LysK resulted in ten-fold increase in lytic activity against *S. aureus* (Mao et al., 2013), and was also shown to work effectively *in vivo* (P. K. Singh et al., 2014). Beside enhanced staphylolytic activity, chimeric Ply187 also exhibited weak lytic activity against *Streptococcus dysgalactiae* (Mao et al., 2013). It was uncertain that where the ability to lyse *S. dysgalactiae* originated from, as both Ply187_CHAP (Dong et al., 2015) and LysK (Mao et al., 2013) were unable to lyse it.

Another fusion example involved the CBD of *Enterococucs* phage endolysin PlyV12 (Yoong et al., 2004), which was fused to the Ply187_CHAP. This fusion resulted in increased lytic activity against *S. aureus* and extended the host range to *Streptococcus dysgalactiae, Streptococcus agalactiae,* and *Streptococcus pyogenes*, as showed by plate lysis assay. Despite no lysis zone was observed for *Enterococcus faecium* and *Enterococcus faecalis* in the plate lysis assay, lytic activity was observed through turbidity reduction, albeit reduced activity as compared to PlyV12 (Dong et al., 2015). Additionally, Lys168 and Lys170 are also derive from *Enterococcus* phage,

with lytic host range specifically target against *Enterococcus* spp, predominantly *E. feacalis* and *E. faecium* (Proença et al., 2012). In the studies by Fernandes et al, the EAD of these endolysin was fused with the CBD of Lys87, a staphylococcal phage endolysin. These chimeras demonstrated a wider lytic host range against *Enterococcus* spp, *S. pyogenes*, and *Staphylococcus* spp (Fernandes et al., 2012).

2.4 CBD as potential candidate in biosensor development

Antibodies are commonly used for bacterial detection, but their production remains laborious and time-consuming. Although phage-based biosensors offer high specificity and stability, bacterial resistance poses challenges for their future development as the it might affect the binding of phage (Al-Hindi et al., 2022). In contrast, CBDs have gained attention as promising alternatives in biosensors development due to their strong binding capabilities (Cho et al., 2021). Genetic engineering enables the attachment of different probes (E.g. GFP protein) on CBD easily, enhancing their versatility as candidate in biosensor development. One of the example was the CBD from Ply500, an endolysin from *Listera monocytogenes* (Gaeng et al., 2000; Loessner et al., 2002). The CBD was fused with GFP, and was successfully used to detect *L. monocytogenes* ScottA cells from a mixture solution of different bacteria, including *Staphylococci*, *Enterococci*, *Bacilli*, and *Lactococci* (Figure 2.2) (Loessner et al., 2002). Furthermore, this CBD was utilised in the development of a CBD-based electrochemical impedance biosensor. In the studies by Tolba et al., 2012, the CBD of Ply500 was immobilised on a gold-coated sensor chip, allowing rapid detection of *Listeria* cells in both buffer and artificially contaminated milk. This method demonstrated a detection limits of 10^4 CFU/mL in buffer and 10^5 CFU/mL in milk using the electrochemical impedance spectroscopy.

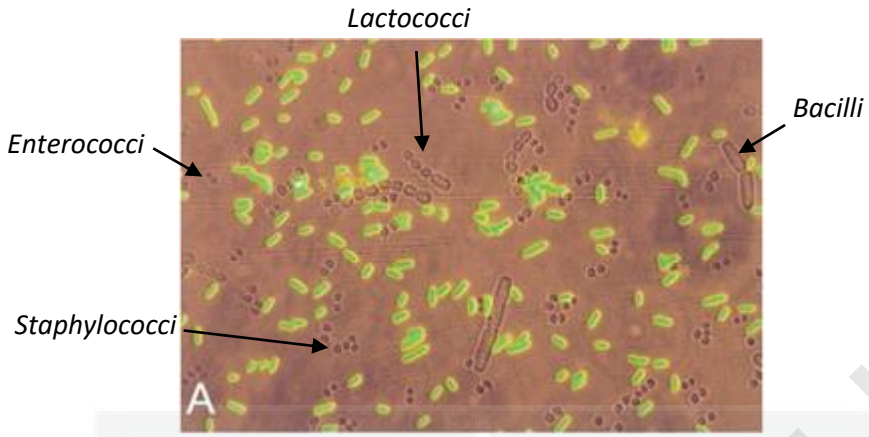


Figure 2.2 Superimposed phase-contrast microscopy and fluorescent microscopy image. *L. monocytogens* ScottA cells labelled with GFP-CBD in a mixture of different bacteria (*Staphylococci*, *Enterococci*, *Bacilli*, and *Lactococci*. (Photo reproduced from Loessner et al., 2002).

S. aureus is a notorious pathogen associated with both nosocomial infection and foodborne diseases (K. Becker et al., 2014; Chaibenjawong & Foster, 2011; Hennekinne et al., 2012). Conventional methods detecting this pathogen include biochemical coagulase test, Methichillin-resistance *S.aureus* (MRSA)-selective chromogenic agar (Wassenberg et al., 2010; Xu et al., 2016), antibody agglutination tests (French, 2009; Jafri et al., 2000), and PCR-based detection of the *mec*A gene (Francois et al., 2003; Grisold et al., 2002; Jonas et al., 2002; Reischl et al., 2000; Ryffel et al., 1992). However, these approaches are laborious and time-consuming, PCR can be difficult to assess in some developing countries (Ahmed et al., 2014; Hallak et al., 2016; Wassenberg et al., 2010). Therefore, there is a pressing need for a rapid and cost-effective diagnostic alternatives.

In recent years, several studies have focused on the development of CBD-based biosensors for pathogen detection (Rahman et al., 2021). For instance, a CBD-based

magnetic enrichment immuneassay was developed to detect *S. aureus* in contaminated milk (Yu et al., 2016). In this assay, immunomagnetic particles (IMPs) coated with antibodies targeting *S. aureus* surface protein A were used to capture and concentrate cells from the milk samples. Following this, CBD fused with a probe (Red fluorescence) was used to detect the cell. This technique achieved a detection limit of 4×10^3 CFU/mL within two hours (Yu et al., 2016). Moreover, flow cytometry assay could be one of the options for rapid and sensitive detection. In the studies of Costa et al., 2020, *S. aureus* was successfully detected in horse blood sample and also buffer by use of flow cytometry assay. However, these technique often require costly equipment and advanced knowledge which can be a disadvantage in developing countries. In light of this, a cost-effective detection assay for *B. cereus* could be adapted for *S. aureus* detection. A nitrocellulose-based lateral flow assay was developed for *Bacillus cereus* using the CBD of endolysins LysB4. This method involves dipping the nitrocellulose membrane into a *B. cereus*-containing solution, allowing bacterial cells to adhere to the immobilized CBD on the membrane (Kong et al., 2017), Adapting this approach for *S. aureus* could help in the development of a rapid and affordable detection kit.

The principle of these techniques ultimately relies on the binding ability and host range of the CBD.However, the exact mechanism of CBD binding remains unclear. Gaining a deeper understanding of the CBD could provide an alternative way to modify the host range of endolysin, which also would be crucial for biosensor development.

2.5 Staphylococcal phage endolysins

Staphylococcal phage endolysins typically consists of the Cysteine, Histidine-dependent amidohydrolase/peptidase (CHAP) domain, the Amidase domain (AMD) and the SRC Homology 3 (SH3) domain (Benešík et al., 2018; Gutiérrez et al., 2018; Jarábková et al., 2015). However, not all staphylococcal phage endolysins possess two EADs (Chang & Ryu, 2017). Also, not all staphylococcal endolysin have SH3 domain as the CBD. (Table 2.2).

Table 2.2: Grouping of Staphylococcal endolysin based on the domain composition.

Group	Domain composition	Phage containing the same group endolysin
Type I	EAD (CHAP) and CBD (no homology 1)	42E, 187, SAP-26, SAP090B, phi13, phiBU01, phiPV83, phiNM3, phiN315, tp310–3, P954, and JS01
Type II	EAD (CHAP) and CBD (SH3 domain)	44HJD, 66, P68, SAP-2, S13, S24-1, GRCS, vB_SauM_Remus, vB_SauM_ Romulus, SA5, and ISP
Type III	EAD (CHAP, amidase-2 (N-acetylmuramoylL-alanine amidase)) and CBD (SH3 domain)	11, 29, 37, 52A, 55, 69, 80, 85, 88, 92, 676Z, A3R, A5W, DW2, JD007, Fi200W, MSA6, ×2, EW, P4W, GH15, K, MCE-2014, Sb-1, SP5, SP6, SA12, SA13, Staph1N, S25-3, S25-4, phiMR11, phiMR25, phiNM2, phiSauS-IPLA88, phiSA012, phiIPLA-C1C, phiIPLA-RODI, and 812
Type IV	EAD (CHAP, amidase-3 (N-acetylmuramoylL-alanine amidase)) and CBD (SH3 domain)	3 A, 53, 96, 77, 80alpha, ROSA, TEM123, TEM126, tp310–1, tp310–2, SMSAP5, StauST398–4, LH1, PVL, phiNM, phi12, phiSauS-IPLA35, phiPVL108, phi2958PVL, phiPVL-CN125, phiNM1, phiNM4, phiSa119, phiSLT, phi5967PVL, phi7247PVL, phi7401PVL, and YMC/09/04/R1988
Type V	EAD (CHAP, amidase-3 (N-acetylmuramoylL-alanine	phiETA, phiETA2, phiETA3, 71, StauST398–1, StauST398–5, and SA97
Type VI	EAD (peptidase_M23, amidase-2 (N-acetylmuramoylL-alanine amidase)) and CBD (SH3 domain)	2638A

Type I: Endolysins containing CHAP as EAD and a putative CBD with no homology to any known CBD; Type II: Endolysin with only one EAD – CHAP domain and a SH3 domain; Type III: Two EAD – CHAP & amidase-2 (*N*-acetylmuramoyl-L-alanine amidase) and a SH3 domain; Type IV: Two EAD – CHAP & amidase-3 (*N*-acetylmuramoyl-L-alanine amidase) and SH3 domain; Type V: Two EAD – CHAP & amidase-3 and a putative CBD with no homologues; Type VI: Two EAD – peptidase_M23 & amidase-2, and a SH3 domain. Yellow highlight**:** Endo88 –Model endolysin used in this study.

Reproduced from (Y. Chang & Ryu, 2017)

2.5.1 Role of SH3 in endolysins host-range

While the SH3 domain does not exhibit lytic activity, they are suggested to recognise and bind to the interpeptide bridge within the bacterial PG (Gründling & Schneewind, 2006; Gu, Lu, et al., 2011; Mitkowski et al., 2019). Interestingly, the SH3 domain's binding target was first elucidated in a bacterial lysin, lysostaphin, rather than in a staphylococcal phage endolysin. Lysostaphin is a peptidoglycan hydrolase found in *Staphylococcus simulans*. It is involved in cell division and shares SH3 domain homologues with *Staphylococcus* phage endolysin (Broendum et al., 2018; H. Kim & Seo, 2023; Mayer, 2001). The lysostaphin SH3 domain recognise the pentaglycine interpeptide bridge as the recognition site (Gonzalez-Delgado et al., 2020; Hirakawa et al., 2009; J. Z. Lu et al., 2006; Mitkowski et al., 2019; Tossavainen et al., 2018), which differs between bacterial species/genus (Schleifer & Kandler, 1972; Vollmer et al., 2008). This observation aligns with findings from another homologue of Lysostaphin – ALE-1, originally found in *Staphylococcus capitis* ALE-1 (Hirakawa et al., 2009; J. Z. Lu et al., 2006). Notably, the suggested binding target site for the SH3 domain in phage endolysins is also the interpeptide bridge (Gu et al., 2014; Gu, Lu, et al., 2011).

The CBD was suggested to play a crucial role in mediating the host range and affinity of the endolysin (Liu et al., 2023; Loessner et al., 2002). The changes in lytic host range was observed through experiments where replacing the SH3 domain with those from other phage endolysins led to altered lytic activity and host range (S. C. Becker, Foster-Frey, et al., 2009). This indicated that the presence of the SH3 domain specifically affects the lytic activity and host range of the endolysin. A study by Oliveira et al., 2019 reported that approximately 96 % (analysis from 205

Staphylococcus phage) of staphylococcal phage endolysins contain only one binding domain from the SH3 family, indicating a conserved recognition strategy among the endolysins for the staphylococcal peptidoglycan layer. Intriguingly, the SH3-containing endolysin from staphylococcal phage exhibit varying lytic host range. Some endolysins are limited to lysing within a single species, while some others have broader host ranges, capable of lysing bacteria across different genera (Huang et al., 2015; Melo et al., 2018; Yokoi et al., 2005; Yoong et al., 2004). Despite having similar SH3 functions and sequences among different SH3-containing endolysins, studies suggest it exhibited differences in lytic host-range, as shown in Table 2.3.

Table 2.3: Lytic host-range of different SH3-containing endolysins

Origin	Endolysin	Domain architecture	Bacteria tested	Lytic host range	Binding assay of truncated SH3 domain	Ref.
Staphylococcus aureus bacteriophage ϕMR11	MV-L	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp., *Enterococcus* spp., *Streptococcus* spp., *Bacillus* spp.	*Staphylococcus* spp.	null	(Rashel et al., 2007)
Staphylococcus warneri M prophage (ϕWMY)	LysWMY	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp., *Enterococcus hirae*. *Bacillus* sp., *Lactobacillus* spp., *Lactococcus diacetylactis*, *Pediococcus pentosaceus*, *Micrococcus luteus*	*Staphylococcus* spp., *Lactococcus diacetylactis*, *Pediococcus pentosaceus*, *Micrococcus luteus*	null	(Schmelcher et al., 2015; Yokoi et al., 2005)
Staphylococcus bacteriophage: vB_SauM-LM12 (LM12)	E-LM12	[CHAP] – [PGRPs] – [SH3]	*Staphylococcus* spp., *Enterococcus* spp., *Streptococcus pneumonia*, *Bacillus* spp. *Listeria monocytogenes*,	*Staphylococcus aureus*	*Staphyloc-occus* spp.	(Melo et al., 2018)
Staphylococcus phage P68	P68	[CHAP] – [SH3_5]	*Staphylococcus* spp.	*Staphylococcus* spp.	*Staphyloc-occus aureus* *did not tested on other species	(Schmelcher et al., 2015, p. 44; Vybiral et al., 2003)
Staphylococcus aureus (NCTC 8325) phi11 prophage	Phi11	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp., *Streptococcus agalactiae*	*Staphylococcus* spp.	null	(Donovan, Lardeo, et al., 2006; X. Wang et al., 1991)

Table 2.3: Continued

Origin	Endolysin	Domain architecture	Bacteria tested	Lytic host range	Binding assay of truncated SH3 domain	Ref.
Staphylococcus phage G15	LysGH15	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp., *Streptococcus pneumonia, Bacillus subtilis*	*Staphylococcus aureus*. *did not tested on other species	*Staphyloc-occus* spp.,	(Gu et al., 2014; Gu, Xu, et al., 2011)
Staphylococcus phage SA12	Lys-phiSA012	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp., *Streptococcus agalactiae*	*Staphylococcus* spp.	null	(Fujiki et al., 2018; Nakamura et al., 2020)
Staphylococcus phage GRCS	LysGRCS	[CHAP] – [SH3_5]	*Staphylococcus* spp., *Enterococcus* spp., *Streptococcus* spp., *Bacillus pumilus*	*Staphylococcus* spp.	null	(Linden et al., 2015)
Staphylococcus phage 812	LysF1	[CHAP] – [SH3_5]	*Staphylococcus* spp., *Streptococcus uberis, Micrococcus luteus*	*Staphylococcus* spp.	*Staphyloc-occus* spp., *Streptoco-ccus uberis,*	(Benešík et al., 2018)
Staphylococcus phage phiH5	LysH5	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp., *Streptococcus pneumonia, Enterococcus* sp., *Bacillus cereus, Lactobacillus* spp., *Lactococcus* spp., *Leuconostoc* spp., *Clostridium* spp., *Listeria* spp.	*Staphylococcus* spp.	null	(Obeso et al., 2008)

Table 2.3: Continued

Origin	Endolysin	Domain architecture	Bacteria tested	Lytic host range	Binding assay of truncated SH3 domain	Ref.
Staphylococcus haemolyticus prophage ΦSH2	PhiSH2	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp.	*Staphylococcus* spp. (Not lysing *S. epidermidis*)	null	(Schmelcher et al., 2015; Schmelcher, Korobova, et al., 2012; Takeuchi et al., 2005)
Staphylococcus phage K	LysK	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp., *Streptococcus* spp., *Lactobacillus* spp., *Listeria monocytogenes*, *Rhodococcus equi*	*Staphylococcus* spp.	null	(S. C. Becker, Dong, et al., 2009; S. C. Becker et al., 2008; Gill, 2014; O'Flaherty et al., 2005; Schmelcher et al., 2015)
Staphylococcus phage 80 alpha	phi80 α	[Ami_3] – [SH3_5]	*Staphylococcus* spp.	*Staphylococcus* spp.	null	(Christie et al., 2010; Schmelcher et al., 2015)
Staphylococcus phage 2638A	2638A	[NlpD] – [Ami] – [SH3]	*Staphylococcus* spp.	*Staphylococcus* spp.	null	(Abaev et al., 2013; Kwan et al., 2005; Schmelcher et al., 2015)
Staphylococcus phage Twort	Twort	[CHAP] – [Ami_2] – [SH3_5]	*Staphylococcus* spp.	*Staphylococcus* spp.	null	(S. C. Becker et al., 2015; Loessner et al., 1998; Schmelcher et al., 2015)

Table 2.3: Continued

Origin	Endolysin	Domain architecture	Bacteria tested	Lytic host range	Binding assay of truncated SH3 domain	Ref.
Staphylococcus phage SA4	LysSA4	[CHAP] – [SH3_5]	*Staphylococcus* spp.	*Staphylococcus* spp.	null	(Mishra et al., 2013)
Streptococcus suis prophage	PlySs	[CHAP] – [SH3b]	*Staphylococcus* spp., *Streptococcus* spp.,	*Staphylococcus* spp., *Streptococcus* spp., *Enterococcus* spp.	*Streptococcus suis*. *did not tested on other genus	(Gilmer et al., 2017; Huang et al., 2015)
Enterococcus bacteriophage Φ1	PlyV12	[N1PC_P60] – [SH3]	*Staphylococcus* spp., *Streptococcus* spp., *Enterococcus* spp.	*Staphylococcus* spp., *Streptococcus* spp., *Enterococcus* spp.	null	(Yoong et al., 2004)
Bacillus phage TsarBomba	PlyTB40	[Ami_3] – [SH3_5]	*Bacillus* spp., *Lysinb. Sphaericus, Paenib. Polymyxa, Streptococcus pyogenes* (binding assay only)	*Bacillus* spp. *Lysinb. Sphaericus* (weak activity), *Paenib. Polymyxa* (weak activity)	*Bacillus* spp.	(Etobayeva et al., 2018)
Bacillus phage Phrodo	PlyP56	[L-alanyl-D-glutamate endopeptidase] – [SH3b]	*Bacillus* spp., *Lysinb. Sphaericus, Paenib. Polymyxa, Streptococcus pyogenes* (binding assay only)	*Bacillus* spp.	*Bacillus* spp.	(Etobayeva et al., 2018)

Table 2.3: Continued

Origin	Endolysin	Domain architecture	Bacteria tested	Lytic host range	Binding assay of truncated SH3 domain	Ref.
Bacillus phage Nigalana	PlyN74	[Ami_3] – [SH3b]	*Bacillus* spp., *Lysinb. Sphaericus, Paenib. Polymyxa, Streptococcus pyogenes* (binding assay only)	*Bacillus* spp.	*Bacillus* spp.	(Etobayeva et al., 2018)
Bacillus phage BPS13	LysBPS13	[Ami_3] – [SH3]	*Staphylococcus* spp., *Streptococcus* sp., *Enterococcus* spp., *Listeria* sp.	*Bacillus* spp.	null	(J. Park et al., 2012)

PGRPs: Peptidoglycan recognition proteins. N1pc_P60: Protein superfamily consists of protein function similar to N1pC – an *E. coli* membrane-associated lipoprotein and P60 – murein hydrolases of *Listeria monocytogenes* (Anantharaman & Aravind, 2003). Null: not tested

From the table 2.3 it can be observed that the host range varies among SH3-containing endolysins, with some even exhibiting an extended host range. For example, LysWMY (Schmelcher et al., 2015; Yokoi et al., 2005) displayed a wider host range by lysing *Staphylococcus* spp, *Lactobacillus diacetylactis, Pediococcus pentosaceus* and *Micrococcus luteus*. In contrast, other endolysins have a much narrower host range, often restricted to species level. For instance, MV-L demonstrated staphylococcal-specific activity when tested against different genus of bacteria (Rashel et al., 2007). Similarly, LysF1 (Benešík et al., 2018), a natural mutant derived from LysK (S. C. Becker, Dong, et al., 2009), possessd only staphylolytic activity. However, domain truncation studies involves GFP-fused SH3 domain revealed that the SH3 domain from LysF1 was able to bind to some *Staphylococcus* spp and *Streptococcus uberis* (Benešík et al., 2018). Aside from staphylococcal phage endolysin, SH3 domains can also be

found in other genera. PlyV12 from *Enterococcus* phage Φ1 exhibited cross-genus lytic activity, including *Staphylococcus* spp., *Streptococcus* spp., and *Enterococcus* spp. (Yoong et al., 2004). Conversely, SH3-containing endolysins from *Bacillus* species often exhibit a rather narrower host range restricted to *Bacillus* spp. For example, LysBPS13 was shown to specifically target *Bacillus* spp. (J. Park et al., 2012). Additionally, while the binding activity of PlyTB40, PlyP56 and PlyN74 were restrict to *Bacillus* spp, a weak lytic acitivity on *Lysinb. Sphaericus* and *Paenib. Polymyxa* could be observed (Etobayeva et al., 2018).

2.5.2 Structural study of the SH3 domain

The SH3 domain was initially defined in human proteins and is associated with protein-protein signaling (Jarábková et al., 2015; Mayer et al., 1988). This domain is particulary known for its ability to recognize proline-rich sequences, specifically the PXXP as the core motif (Proline-x-x-Proline) (Asbach et al., 2012; Kaneko et al., 2008),or even lipid molecules (Sipeki et al., 2021). Beside eukaryotic proteins, SH3 domains are ubiquitous across various organisms, including bacterial lysostaphin (Gonzalez-Delgado et al., 2020; Mitkowski et al., 2019) and phage endolysin (S. C. Becker, Foster-Frey, et al., 2009; Gu, Lu, et al., 2011).

Despite their presence in diverse organisms, the SH3 domains maintain a conserved structure which includes the characteristic β-barrel fold consists of five or six β-strands arranged into two anti-parallel β-sheets, along with two key structure: arginine-threonine loop (RT) loop and Src loop (Figure 2.3A). While the RT and Src loop have been extensively studied in eukaryotic SH3 domains, their roles in phage endolysins remain insufficiently explored. The RT loop is a flexible region that forms part of the

binding site (Figure 2.3B), and its flexibility has been suggested to influence binding specificity (Arold et al., 1998; Longshore-Neate et al., 2024). The Src loop, on the other hand, commonly forms a hydrophobic pocket within the binding site, which are critical to ligand recognition (D'Aquino & Ringe, 2003; Longshore-Neate et al., 2024). Interestingly, in bacterial lysostaphin, the binding site is also located near the RT and Src loops (Figure 2.3B), suggesting potential structural and functional similarities.

While bacterial SH3 domain are known to bind the peptides within the PG layer (J. Z. Lu et al., 2006; Mitkowski et al., 2019; Tossavainen et al., 2018), SH3 domain derived from phage endolysins possess a similar mechanism (Benešík et al., 2018; Gu et al., 2014; Tamai et al., 2014). To the best of our knowledge, most studies investigating SH3 domain interactions with the PG layer were primarily reliant on wet lab experiments (e.g. Nuclear magnetic resonance, NMR). However, some computational approaches, including molecular docking have also been employed, yielding comparable results (Hirakawa et al., 2009; S. Wang et al., 2025). Given their efficiency and cost-effectiveness, *in silico* analyses as recommended as complementary tools for studying SH3-PG interaction, particularly with the advancement of AI-based tools (e.g. AlphaFold), which further enhance the accuracy and reliability of structural predictions and interactions.

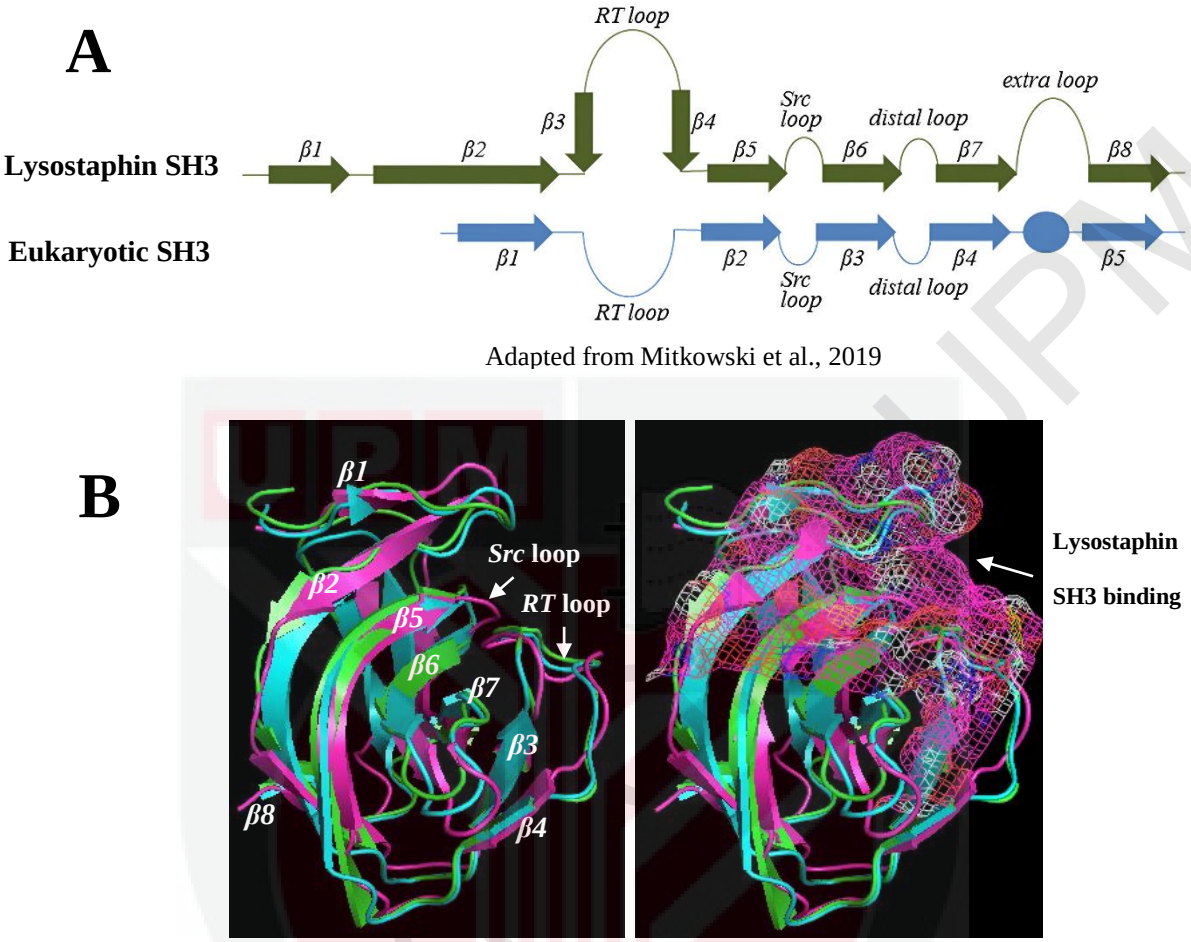


Figure 2.3: Conserved structural motifs of SH3 domain across different organisms. The SH3 domain exhibits a conserved β-sheeted barrel structure, connected by the Src loop and RT loop. **(A):** Cartoon representation of the lysostaphin SH3b (Green) and eukaryotic SH3 domain (Blue). **(B):** Superimpose 3D structure of the lysostaphin SH3 domain (PDB: 6RK4), LysGH15 phage endolysin SH3 domain (PDB: 2MK5) and Endo88 SH3 domain (Chandran et al., 2022), demonstrating the conserved architecture of SH3 domains aross different species.

2.6 *Stapylococcus* Phage 88 and its endolysin (Endo88)

To investigate the specificity of the cell wall binding domain of stapylococcus endolysins, the staphylococcal phage 88 endolysin, Endo88 was selected as a candidate for mutagenesis studies. Endo88 was isolated from *Staphylococcus* Phage 88 (ATCC 33742-B1), which targets *S. aureus* subsp. *aureus* PS88 (ATCC 33742). According to American Type Culture Collection (ATCC), this strain is a clinical isolate from New York.

Bioinformatic characterization has previously been performed for Endo88 (Chandran et al., 2022). In brief, the endolysin encoded by the Staphylococcal Phage 88 (YP_240699), denoted as Endo88, is encoded in ORF006 and comprises of 491 amino acids. According to NCBI Conserved Domain Database (CDD), Endo88 belongs to the Peptidoglycan recognition proteins (PGRPs). It consists of a Cysteine, Histidine-dependent Amidohydrolase/Peptidase (CHAP) domain at residues 31 – 119 (Accession number: pfam05257; E-value: 2.96e-08), an Amidase (Ami_2) domain at residues 198 – 322 (Accession number: smart00644; E-value: 1.05e−24), and a SH3_5 domain at residues 395–460 (Accession number: pfam08460; E-value: 2.03e−21) (Figure 2.4).

Figure 2.4: Domain architecture of Endo88. Endo88 consists of two EAD – CHAP & AMD, and one CBD – SH3.

Amino acid sequence of Endo88 has high similarity with MV-L (BAF33253.1, similarity: 99.8%; Identity: 99.6%) (Rashel et al., 2007, 2008) and LysH5 (YP_002332536.1, similarity: 96.5%; Identity: 94.7%) (Rodríguez-Rubio et al., 2012), but not with LysGH15 (ADG26756.1, similarity: 50.6%; Identity: 38.2%) (Gu et al., 2014; Gu, Lu, et al., 2011), LysK (YP_009041293, similarity: 50.8%; Identity: 38.2%) (S. C. Becker et al., 2008) and Twort (AAX92311, similarity: 55.3%; Identity: 41.9%) (S. C. Becker et al., 2015) (Figure 2.5). The residues forming the catalytic site in CHAP (C32, H95, E111 & N113) and the SH3 binding site (N390, Y392, G393 & T394) are conserved among the endolysin (Gu et al., 2014).

Figure 2.5: Multiple sequence alignment of Endo88 with other SH3-containing staphylococcal phage endolysin. The alignment consists of Endo88 (YP_240699), LysGH15 (ADG26756.1), MVL (BAF33253.1), LysK (YP_009041293), Twort (AAX92311.1) and LysH5 (ACJ64589.1). Blue, orange and green bar represent the CHAP, Amidase, and SH3 domains, respectively. The conserved region are denoted as "*", semi-conserved region are denoted as "+" and scores were given for other column (higher number mean more conserve). The residues forming the catalytic site in CHAP (C32, H95, E111 & N113) and the SH3 binding site (N390, Y392, G393 & T394) are indicated by the red and black box, respectively. (Adapted from Chandran et al., 2022).

2.6.1 Target sites of Staphylococcal endolysin 88 domains on bacterial peptidoglycan

The precise site of action for Endo88's CHAP domain is unknown, but it likely possesses *D*-alanyl-glycyl endopeptidase activity (Figure 2.6), as it shares >99% identity with the CHAP domain of the phi_P24556 and MV-L endolysins, which site of action have been experimentally confirmed (Navarre et al., 1999; Rashel et al., 2007). On the other hand, the AMD is presumed to exhibit *N*-acetylmuramoyl-L-alanine amidase activity. However, several studies suggested that the AMD may not significantly contribute to the lytic activity of the endolysin (S. C. Becker, Dong, et al., 2009; Costa et al., 2020; Fujiki et al., 2018; Navarre et al., 1999; Son et al., 2018). As indicated by its name, the CHAP domain relies on cysteine and histidine residues, along with a third residue, to form a catalytic triad in its active site (Gu et al., 2014; Sanz-Gaitero et al., 2014). These catalytic residues are highly conserved among staphylococcal phage endolysins (Chandran et al., 2022).

Staphylococcus aureus cell wall consists of a highly cross-linked peptidoglycan (PG) layer, which provides structural integrity and protection (Vollmer, 2008; Vollmer & Seligman, 2010). As shown in Figure 2.6 the main glycan back bone consists of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) linked by β-1,4 linkage, with peptide stems crosslinking these backbones to form a rigid structure (Garde et al., 2021; S. J. Kim et al., 2015). Notably, in Gram positive bacteria, the peptide crosslink varies across species/genera (Garde et al., 2021; S. J. Kim et al., 2015; Schleifer & Kandler, 1972).

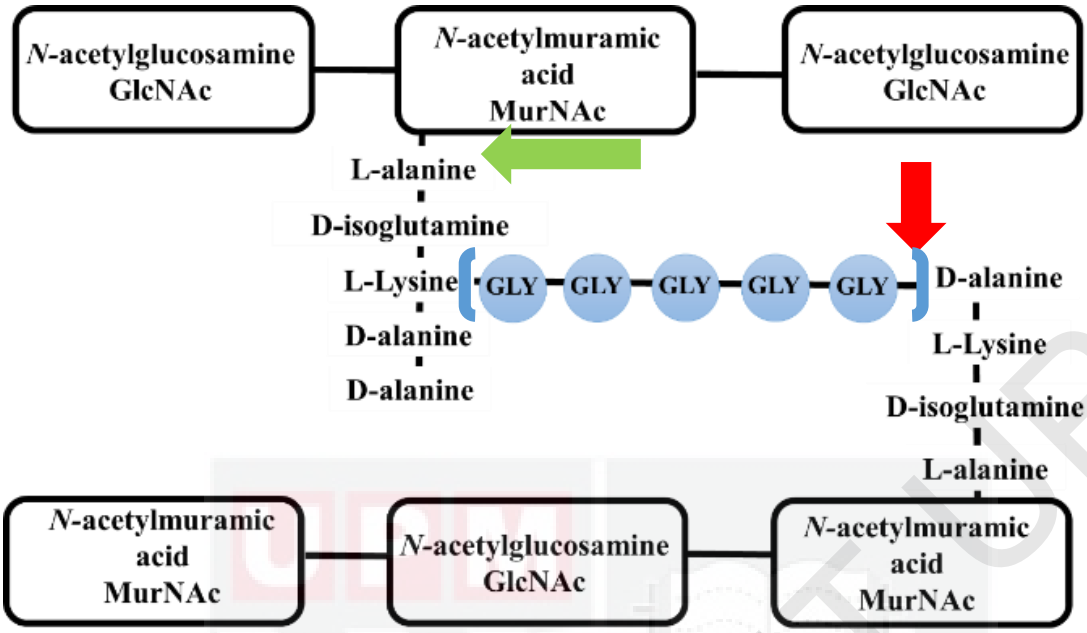


Figure 2.6: Illustration of peptidoglycan in *S. aureus*. The repeating unit of *N*-acetylglucosamine (GlcNAc) and *N*-acetyl muramic acid (MurNAc) polymerize through beta-1, 4-glycosidic bonds to form a glycan strand. Between the glycan strands are peptide stems, which consist of short amino acid chains attached to the MurNAc residues. These peptide stems are crosslinked through pentaglycine (or an interpeptide bridge in general) that link to the adjacent peptide stem. Red arrow: Putative CHAP target site - *D*-alanyl-glycyl endopeptidase; Green arrow: *N*-acetylmuramoyl-L-alanine amidase; Blue bracket: Pentaglycine chain.

CHAPTER 3

DEVELOPMENT OF MUTATIONS BASED ON THE PRIMARY SEQUENCE

3.1 Introduction

Despite the similar functions and sequence in SH3 domains among various SH3-containing endolysins, studies suggest they exhibit differences in lytic host range. This observation implies that while there is a conserved mechanism for the SH3 domains to recognize staphylococcal PG, variation within the amino acid sequence of SH3 domains may contribute to the differences in host range. Such variations are typically caused by mutations, which play a pivotal role in evolution, particularly in relation to the structure and functionality of proteins. Studies have demonstrated that proteins can acquire new functions through mutations in the amino acid sequence without compromising the original function, thereby driving evolution (Jayaraman et al., 2022; Sotomayor-Vivas et al., 2022). Mutation can occur in both conserved region and non-conserved region. While conserved positions are undoubtedly critical for the protein function, mutations in non-conserved positions can also affect protein functionality (Gloor et al., 2005).

Mutations are thought to occur along with phage evolution (Oechslin et al., 2022). It is noteworthy that during phage evolution, mutations have been detected in virion proteins, particularly those related to host attachment (Ceyssens et al., 2006; Paterson et al., 2010). Intriguingly, beside host attachment proteins, several point mutations in endolysin have also been identified (Oechslin et al., 2022). For instance, studies have

indicated that phages with a broader host range may require a longer adaptation period with their host to achieve long-term co-evolution rather than spontaneous adaptation. This extended adaptation period allows multiple mutations to occur gradually during the co-evolution process (Chin et al., 2022; Gibson et al., 2020; Hall et al., 2011). In contrast, if phages are grown with non-evolving hosts, the selective pressure is low, resulting in fewer mutations in the amino acid sequence (Chin et al., 2022; Gibson et al., 2020; Hall et al., 2011).

In the work of Oechslin et al. 2022, phages were genetically modified to replace their original endolysin gene with endolysins targeting different bacteria. This modification was intended to observe adaptive mutations in the foreign endolysin as it interacts with the same host. During the experimental evolution assay, several point mutations were detected in both conserved and non-conserved regions of the endolysin's CBD as the phages were adapting to their host. This observation leads to the speculation that staphylococcal phages may evolve their SH3 domains to alter their host-range, thereby increasing their chances of survival. Therefore, in this chapter, the goal was to identify amino acids that could alter the host range of the Endo88 endolysin using a bioinformatics approach targeting differences in the primary structure of Endo88 in comparison with other SH3-containing endolysins.

This chapter describes the first objective of the study which was to design and validate targeted mutations of the wild-type *Staphylococcus aureus* phage 88 endolysin CBD through bioinformatics approaches for potential alteration of endolysin target host-range. In this chapter, primary sequences of endolysin SH3 domains were compiled and analysed for differences in sequences with possible correlation to its host-range.

Shannon's entropy and mutual information (MI) analysis were used to identify differences in amino acids in the SH3 domain, comparing Endo88 with endolysins of different lytic host range. This was performed by means of a bioinformatics tool - Antigenic Variability Analyser (AVANA). These analyses resulted in two regions with high MI value and low entropy value which were chosen for mutation (Mutants A1 and A2). Additionally, the multiple sequence alignment of staphylococcal phage endolysin's SH3 domain was also analysed manually to identify distinct differences in the sequences of the SH3 domain between endolysins with different lytic host range (Mutant M).

3.2 Methodology

3.2.1 Bioinformatic characterization of Endo88 (YP_240699).

The Sequence of Endo88 was retrieved from GenBank database (accession numbers: YP_240699). The functional conserved domain was predicted using NCBI Conserved Domain Database (NCBI CDD) and Pfam database. Multiple sequence alignment was performed on Endo88 with other SH3-containing staphylococcal phage endolysins with Multiple Alignment using Fast Fourier Transform (MAFFT) alignment tool (https://tcoffee.crg.eu/) to study the conserved amino acid.

3.2.2 Identification of highly conserved serotype-specific sequence (HCSS) with information theory using Antigenic Variability Analyser (AVANA).

The overview for the workflow to develop the mutants using AVANA is summarized in Figure 3.1. AVANA is a bioinformatics software which calculates the variability in a set of protein sequences alignment and identifies mutation in different sequence sets. Although AVANA is typically used to determine the antigenicity of vaccine candidates, this bioinformatics tool was employed here because it utilizes an algorithm capable of identifying gene sequences that are unique within certain defined groups. In this case, the algorithm was used to identify sequences unique within defined endolysin host range groups.

In brief, two local databases denoted as NCBI Local Database and Literature Local Database were generated to compile a comprehensive list of the SH3 domain of SH3-containing phage endolysin amino acid sequences. These two databases were then joined together and analysed using AVANA (Miotto et al., 2008).

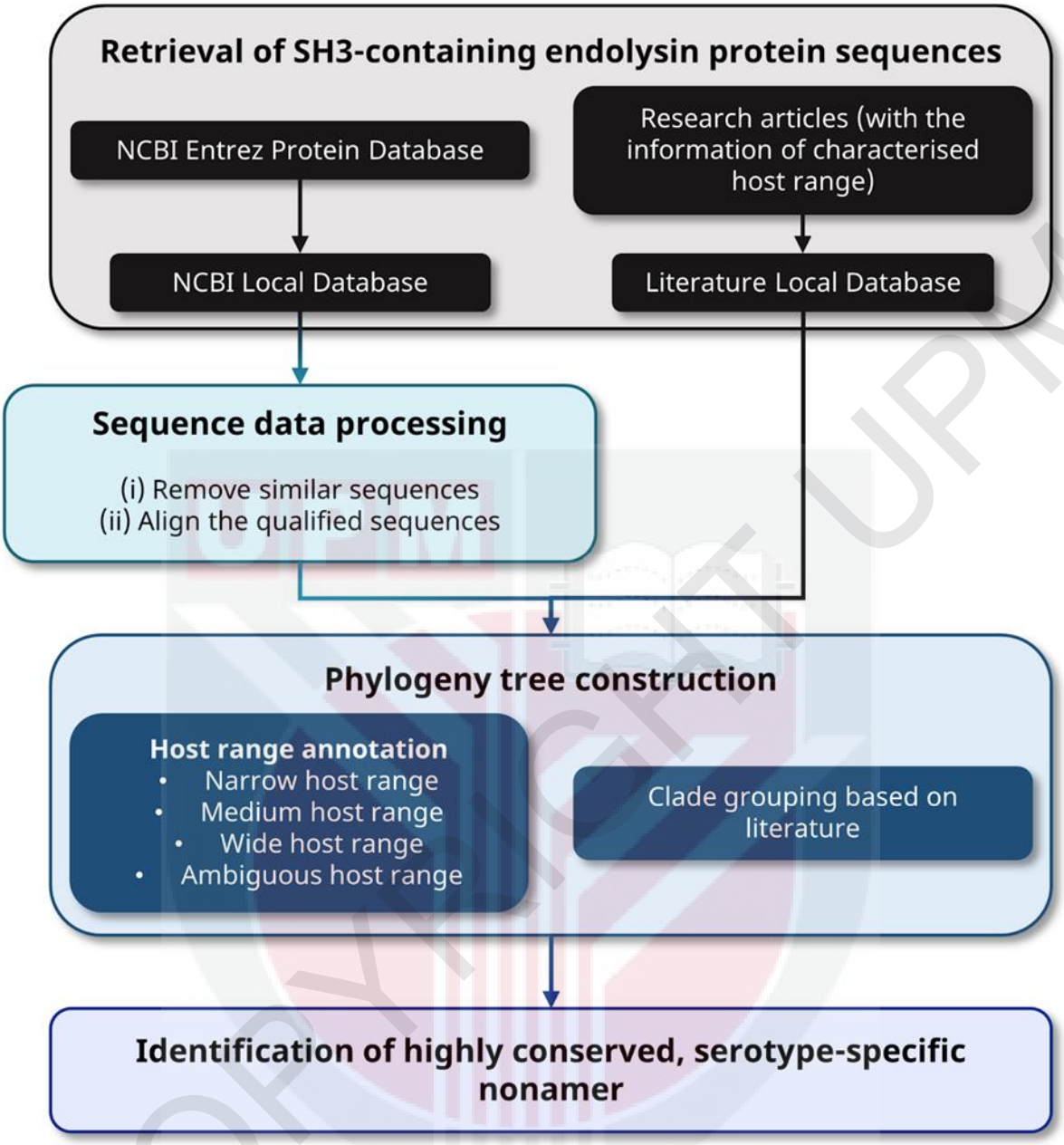


Figure 3.1: Overview of the workflow used to design mutations using AVANA based on the primary structure of endolysins.

3.2.2.1 Construction of NCBI local database

The NCBI Local database comprised of sequences retrieved from the non-redundant (nr) database, primarily derived from genome sequencing data, including those with or without experimental characterisation. Construction of this database involved submitting amino acid sequences of Endo88 (YP_240699) to NCBI BLASTp which was filtered based on several parameters, including Non-redundant protein sequence (nr) database, Organism: Viruses (Taxid: 10239), E- value (<0.05), Identity (>20%), Similarity (>50%), Query coverage (>50%), and Alignment score (>50%). To reduce sampling bias, CD-Hit tool (W. Li & Godzik, 2006) was used to eliminate redundant sequences (sequences that have 100% identity to Endo88), and manual screening was performed to remove hypothetical or partial proteins. A total of 100 sequences were initially retrieved, with 92 remaining after screening.

3.2.2.2 Construction of Literature local database.

The Literature Local Database was compiled using phage endolysin sequences reported in published research articles which had their lytic host-range experimentally characterized. These endolysin sequences were compiled, annotated and categorized into four groups: (I) Narrow host-range (NH), indicating endolysins that specifically lyse within species (e.g. lyse only *S. aureus*); (II) Medium host-range (MH): Endolysin which lyse within the same genus but across species (e.g. capable of lysing both *S. aureus* and *S. epidermidis*); and (III) Wide host-range (WH), encompassing endolysins with ability to lyse across genus (e.g. targeting both *S. aureus* and *Streptococcus pyogenes*). However, many of the reported literature focused primarily on the lytic activity instead of the lytic host-range of the endolysin and only tested lysis in the primary host. These

endolysins were grouped as (IV) Ambiguous host-range (AH). In total, 24 sequences with annotated host-range were compiled to construct the Literature Database.

3.2.2.3 Multiple sequence alignment (MSA) with the database.

The amino acid sequence of Endo88 (YP_240699) was aligned and compared to other homologous endolysin. Two multiple sequence alignments were performed. The first alignment combined the sequences from the BLASTp search and the sequences retrieved from the literature. Subsequently, the alignment was subjected to phylogenic tree and AVANA analysis. The second alignments included only the sequences retrieved from literature. This alignment was to observe any distinct differences in the amino acid sequence, amongst the SH3 domains of staphylococcal phage endolysins with annotated host range for manual design of one of the mutants. MAFFT, an alignment tool was used (Katoh et al., 2002) for both alignmenst and was manually inspected to ensure no misalignments were present.

3.2.2.4 Phylogenetic tree analysis.

The aligned sequences from the combined database (NCBI local databse & Literature local database) were retrieved. The sequence corresoponding to the SH3 domain were extracted for further analysis. Notably, a few additional amino acids upstream (`NKYGT`) of the SH3 domain were also included in the extraction, as they are suggested to contribute to ligand binding (Gonzalez-Delgado et al., 2020; Gu et al., 2014; Mitkowski et al., 2019). The combined dataset underwent phylogenetic analysis using the Maximum Likelihood method in MEGAx software (Kumar et al., 2018). Clades

within the resulting phylogenetic tree were annotated and grouped according to their predicted host range (NH/MH/WH/AH) derived from the Literature Database.

3.2.2.5 Designing of Mutant A1 and A2 with AVANA.

By leveraging these two distinct databases, AVANA was employed to analyse the amino acid sequences of endolysins with different lytic host-range. Despite its conventional usage in identifying antigenic regions in vaccine development, AVANA's algorithm is suitable in this context. In the current study, AVANA was used to detect sequences differences between the clade annotated as having a putative Wide host range (WH) and the clade representing a putative Medium host range (MH), where Endo88 was positioned.

AVANA measure the diversity of sequences within and across group/clade by combining Shannon's entropy and mutual information (MI) analysis (Chong & Khan, 2019; Gloor et al., 2005; Paninski, 2003, 2003; Shannon, 1948). Shannon entropy is a measure from information theory, to determine how much variation exists at each position in the sequence. High entropy values mean that a particular amino acid position changes frequently (i.e., it's variable), which could indicate a region is not conserved. Mutual information looks at how changes in one part of the protein sequence might be linked to changes in another part. This helps to identify co-evolving positions—places where mutations in one region depend on or are related to mutations in another, which may suggest their functional and structural importance. By comparing endolysins with different host ranges, nonamers were chosen from region where MI values were high and entropy was low.

Prior to calculation, three gaps were removed in the multiple sequence alignment of combined database to prevent interference (Chong & Khan, 2019; Paninski, 2003). Entropy calculations were performed for overlapping nonamers, using a window size of nine amino acids (e.g. Calculating the sequence diversity from amino acid 1 – 9, followed by 2 – 10, and so on). The Shannon's formula is expressed as :-

$$H(x) = -\sum_{i=1}^{n(x)} pi,x \; Log \log 2 \, (p_{i,x})$$

Where *H(x)* represent the entropy of the nonamer peptide, *pi,x* is the probability of a specific nonamer peptide (*i*), starting at a particular amino acid position (*x*) (Chong & Khan, 2019). A low entropy value indicates that the region of the sequence is highly conserved within the group, while a high entropy value suggests the region is less conserved. The entropy values were then incorporated into the Mutual information (MI)'s formula to analyse the mutual dependency between the MH group and the WH group. The formula is expressd as follows (Gloor et al., 2005; Paninski, 2003):

$$\mathrm{MI}(\mathrm{WH}, \mathrm{MH}) = \mathrm{H}(\mathrm{MH}) + \mathrm{H}(\mathrm{WH}) - \mathrm{H}(\mathrm{WH}, \mathrm{MH})$$

H (WH, MH) represent the joint entropy between the MH and WH group, while H (MH) and H (WH) denote the entropy calculated from the MH and WH groups, respectively. Mutual information (MI) values range from 0 to 1, where higher values indicated significant differences in the nonamers between the two groups. Ultimately, two nonamers with high MI values and low entropy values were selected for mutation. A high MI value and low entropy value indicated that these nonamers are specific and

less diverse within a group. These selected nonamer sequences were then substituted into the wild-type (Endo88) at corresponding amino acid position, resulting in two mutants denoted as “A1” and “A2”.

3.2.3 Manual designing of Mutant M

Apart from AVANA, another approach was conducted based on the multiple sequence alignment of SH3-containing staphylococcal endolysins from the Literature local database. After the phylogeny tree was constructed, staphylococcal endolysin with the annotated host-range were retrieved from the tree and aligned with MAFFT. Subsequently, the aligned sequences were manually inspected to identify distinctive amino acid regions within each host-range group. The region was then replaced into Endo88 at the corresponding amino acid position and the mutant generated was denoted as “M1”.

3.3 Results & Discussion

By comparing endolysins with different host ranges, nonamers were chosen from region where MI values were high and entropy was low. Before AVANA analysis, the host range of all the endolysins sequences (from NCBI database and literature database) were annotated based on phylogenetic tree analysis.

3.3.1 Phylogenetic Tree analysis

The SH3 amino acid sequences from both the NCBI Database (92 sequences remained after filtering) and the Literature Database (24 sequences with annotated host range) were merged, aligned and a phylogeny tree was constructed (Figure 3.2). The clades were annotated based on host range information from the sequences in the Literature Database, which had been experimentally verified and reported in literature. The hypothesis of the experimental design was that SH3 sequences with high similarity will have similar host range. Endo88 was predicted to possess a medium host range (MH) as it was clustered within the MH clade. The clades without annotation either did not belong to the *Staphylococcus* genus but possessed the SH3 as CBD or did not have sequences from the Literature Database, thus the host-range was unknown.

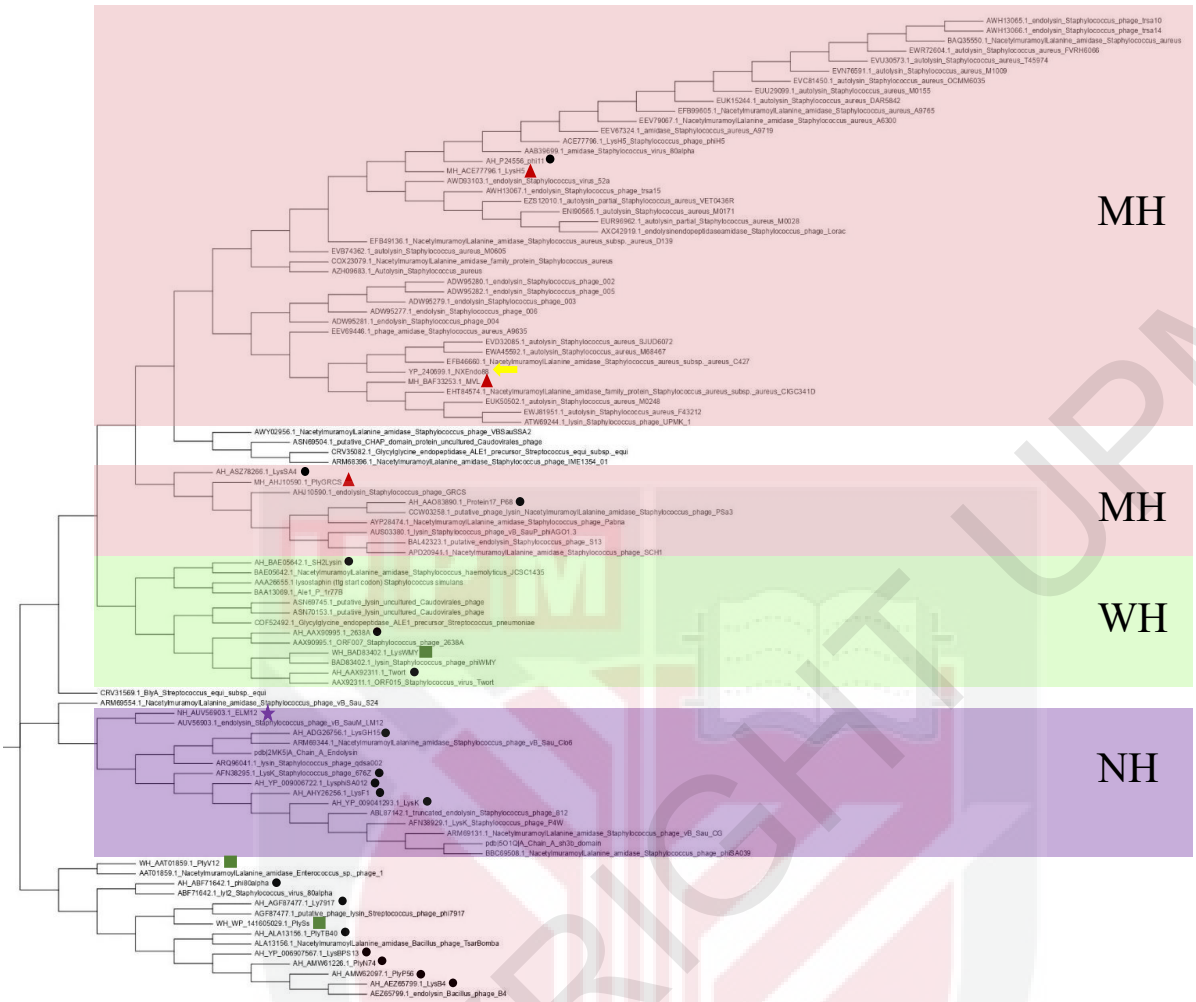


Figure 3.2: An unrooted maximum likelihood phylogenetic tree of SH3 domain sequences from both databases. Dots with different colour denotes sequences with annotated host range from the Literature Database: Black dots: AH endolysins; Dark green: WH endolysins; Dark purple: NH endolysins; Dark red: MH endolysins. Yellow arrow = Wild-type Endo88. Based on these annotated endolysin host-range, the clades were assigned their respective host-range as such: Red box: Medium host range (MH) clade; Green box: Wide host range (WH) clade; Purple box: Narrow host range (NH) clade. Literature Database

3.3.2 AVANA

The HCSS is a nonamer with sequence that is unique within a group, and also less diverse within the group (Chong & Khan, 2019). In the phylogeny tree, the reference endolysin – LysWMY (BAD83402.1) from the previously designed mutation (Mutant M), formed a WH clade with other sequences from the NCBI database. It was expected that the nonamers derived from the WH groups would be unique for the WH group, and might potentially alter the host range of Endo88. By use of the AVANA analysis, HCSS within the WH group of SH3-containing endolysin was identified by comparing with the MH group.

Using AVANA, a total of 58 HCSS nonamers were identified within the WH group, with MI > 0.5, and average entropy of 337 (Highest: 3.75; Lowest: 2.26) (Figure 3.3). Two best potential candidates with the combined highest MI value and lowest entropy value were chosen for mutation (`FTPNTD-IITRTTGP` & `SGVLKAGQTIHYDEVMK`) were chosen for mutation (Table 3.1). These sequences had the lowest Shannon Entropy value which means that this sequence was conserved within the WH group in comparison with the MH group. It also had the highest MI value, which meant that the sequence is more specific within the WH group. These sequences were then replaced into Endo88 at the same amino acid position, resulting in two mutants (A1 and A2) with the expectation of the host-range of Endo88 changing from MH to WH (Figure 3.4).

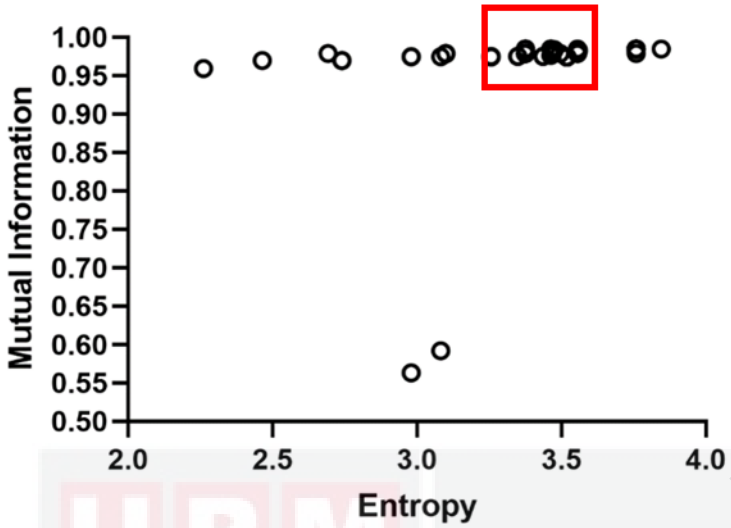


Figure 3.3: Scatter plot of Mutual Information (MI) versus Entropy. Nonamers with highest MI value and low entropy value were chosen and concatenated (Red box).

Table 3.1: Nonamers with highest MI value and lowest entropy value within the 58 nonamers.

Position	MI value	Nonamers	Entropy	Combined Nonamers	Average MI value	Average Entropy
16	0.9848	PNTDIITRT	3.375459	FTPNTD-IITRTTGP	0.9848	3.41273
17	0.9848	NTDIITRTT	3.375459			
18	0.9848	TDIITRTTG	3.375459			
19	0.9848	DIITRTTGP	3.375459			
13	0.9848	FTPNTDIIT	3.462415			
14	0.9848	TPNTDIITR	3.462415			
15	0.9848	PNTD-IITR	3.462415			
34	0.9848	SGVLKAGQT	3.555533	SGVLKAGQTIHYDEVMK	0.9848	3.7451
35	0.9848	GVLKAGQTI	3.757925			
37	0.9848	LKAGQTIHY	3.757925			
38	0.9848	KAGQTIHYD	3.757925			
39	0.9848	AGQTIHYDE	3.757925			
40	0.9848	GQTIHYDEV	3.757925			
36	0.9848	VLKAGQTIH	3.844882			
41	0.9846	QTIHYDEVM	3.757925			
42	0.9846	TIHYDEVMK	3.757925			

The nonamers with high MI value and low Entropy value were chosed and concatenated, yielding two potential candidates for mutation.

Figure 3.4: Nonamers selected as mutational candidate. Two amino acid nonamers "FTPNTD-IITRTTGP" & "SGVLKAGQTHYDEVMK" with high MI value and low entropy values were output by the AVANA analysis (Red box) and chosen for mutation. A high MI value indicates that the region is specific within a group (WH group), while a low entropy value indicates that the region is less diverse in the WH group.

3.3.3 Manually design based on MSA (Mutant M)

The amino acid sequences of the SH3 domain of staphylococcal phage endolysins from the Literature Database were aligned and observed manually. The lytic host-range of the endolysins were annotated based literature reports. As depicted in Figure 3.5, the region (Red box) in LysWMY_BAD83402.1 display distinct differences compared with the MH group, specifically the wild-type Endo88. The regions were then replaced into Endo88 at the same amino acid position. LysWMY was annotated as a endolysin exhibiting wide lytic host range based on the lytic assay (Yokoi et al., 2005), while Endo88 was predicted as medium host range from the phylogeny tree (Figure 3.2).

The hypothesis underlying this substitution is that variations in the SH3 domain influence the lytic host range of endolysins. LysWMY, has been reported to exhibit a wide host range, contains the "SWTTSDG" motif which is not present in any of the MH sequences which instead contains the "GYTWE/QG", this region was replaced in

Endo88 with the intention to investigate if this mutation may similarly expand its lytic host range. The approach is similar with the AVANA mutants, which aims to assess whether specific amino acid differences contribute to host specificity. However, while AVANA facilitate large-scale comparative analysis between WH and MH groups, this approach was identified through manual screening, focus exclusively on the literature database which all analysed suquences had annotated host range.

Figure 3.5: Multiple sequence alignment of SH3 domain of annotated staphylococcal phage endolysin. The distinctive region (Red box) was then substituted into Endo88 at the corresponding amino acid position. "**:**" denotes conserved regions; "**.**" denotes semi-conservative regions; "*****" denotes fully conserved regions. MH preceding the protein accession number denotes Medium Hosta-Range while WH denotes Wide Host-Range.

In summary, two mutations were designed through AVANA analysis. Top two nonamers "FTPNTD-IITRTTGP" and "SGVLKAGQTHYDEVMK" were selected for mutation with highest MI value and lowest entropy value. Another mutation was determined through manual screening in literature database. WH group endolysin containing "SWTTSDG" motif is absent in all MH sequences, which instead contain the "GYTWE/QG" motif. Therefore, "SWTTSDG" motif was selected as the third mutation, denoted as "M".

CHAPTER 4

CLONING AND EXPRESSION OF WILD-TYPE ENDO88 AND MUTANTS DESIGNED BASED ON PRIMARY STRUCTURE

4.1 Introduction

Recombinant phage endolysins have shown to inhibit a variety of pathogens when applied externally to the cell, making them potential candidates for antibiotic alternatives in treating Gram-positive bacterial infections (Fischetti, 2008; Loessner, 2005). Given that promising role, in this chapter, the cloning of Endo88 in *E. coli* is described. Initially, we cloned *Endo88* with a N-terminal his-tag (denoted as *BXEndo88*, with *BamH*I & *Xho*I flanking the gene) which was not expressed in the solubilised fraction. After much effort in optimising the re-folding and solubilisation of the protein from the pellet, the protein was not functional. Thus, we re-cloned the *Endo88* with a C-terminal his-tag (denoted NXEndo88, with *Nco*I & *Xho*I flanking the gene) which was expressed in the soluble fraction. Using the conditions optimised for the NXEndo88 (which was after that just referred to as Endo88), the mutants designed in the previous chapter (A1, A2 and M1) was similarly cloned and expressed. The Endo88 was fully characterised for lytic activity in various biophysical conditions including different buffers, pH and temperatures. Using the optimised conditions, lytic activity and host range of Endo88 and mutants were characterised and compared.

Escherichia coli is commonly utilized for recombinant protein production due to its well-established expression system and various commercially available strains and compatible plasmid vector. The combination of *E. coli* B strain (e.g. BL21(DE3)) with pET plasmid vector is widely adopted. However, overexpression often results in the

formation of inclusion bodies, which are expressed insoluble protein structure (mostly unfunctional) found in the host cytoplasm (Palmer & Wingfield, 2004). Due to mis-folding during overexpression, the functionality of these proteins are often compromised. To restore its biological activity, purification protocols involving solubilization and refolding can be used (S. M. Singh & Panda, 2005; Z. Yang et al., 2011). However, these processes often yield low and are laborious. Hence, the expression of soluble protein is preferable.

This chapter describes the second and third objectives of the study which was to clone and express the endolysin, subsequently analyse the lytic host range. Firstly, the cloning and expression of wild-type Endo88 was optimized prior to cloning of the mutants. Different cloning strategies and buffer adjustments were implemented to ensure the expression of soluble and functional Endo88. After the expression of soluble proteins, the activity and stability of the endolysin under different biochemical conditions such as concentration of Sodium Chloride (NaCl), temperature, and pH were also investigated. The optimized expression conditions were then applied to mutant constructs, and their lytic host range was evaluated and compared to wild-type Endo88 using turbidity reduction assay.

4.2 Methodology

4.2.1 Bacterial strains and growth condition

The plasmid and bacterial strains used in this study are listed in Table 4.1. *Escherichia coli* BL21(DE3) was sub-cultured from plate colonies and grown in Luria-Bertani broth (LB broth). The culture was then incubated at 37 °C, 200 rpm for 16-18 hours. Recombinant *E. coli* Top 10 and BL21(DE3) containing pET28a expression plasmid was grown in LB broth supplemented with 50 µg/mL Kanamycin as the selection marker, and incubated at 37 °C, 200 rpm for 16-18 hours. Pathogens were grown in different condition. *Staphylococcus aureus* PS88, *Staphylococcus epidermidis* (ATCC 12228), *Staphylococcus hominis* and *Enterococcus faecalis* (ATCC 29212) were grown in Brain Heart Infusion broth (BHI broth) and incubated at 37 °C, 200 rpm for 16-18 hours. *Streptococcus pyogenes* (ATCC 19615) and *Streptococcus agalactiae* were grown in BHI broth and incubated statically at 30 °C for 16-18 hours. *Bacillus subtilis* (ATCC 6633), *Bacillus cereus* (ATCC 11778) and *Micrococcus luteus* (ATCC 10240) were grown in TSB broth and incubated at 37 °C, 200 rpm for 16-18 hours. *Lactobacillus plantarum* was grown in De Man, Rogosa and Sharpe broth (MRS broth) and incubated statically at 30 °C for 16-18 hours.

Table 4.1: List of plasmid and bacteria used in this study

Bacteria strain and plasmid used	Features	Sources
E. coli Top 10	Cloning host for plasmid pET28a amplification	Novagen, USA
E. coli BL21(DE3)	Expression host for pET28a. Protease deficient protein expression host containing gene for T7 RNA polymerase. Suitable expression host for pET28a.	Novagen, USA
E. coli Origami B(DE3)	Expression host for pET28a	Novagen, USA
Lactococcus lactis NZ9000	Expression vector for pNZ8048	MoBiTec, Germany
pET28a	Expression plasmid vector carry an N-terminal His-tag/Thrombin cleavage site/ T7.tag configuration plus an optional C-terminal His-tag.	Merck, Bioscience
Staphylococcus aureus PS88	Clinical isolates from New York City	ATCC 33742
Staphylococcus epidermidis	FDA strain use for food testing	ATCC 12228
Staphylococcus hominis	Clinical isolate isolated from Human Breast milk	Gifted by Dr Nurulfiza Mat Isa, UPM (Hassan et al., 2016)
Streptococcus pyogenes	Clinical isolate from the pharynx of a child	ATCC 19615
Streptococcus agalactiae	Isolated from Tilapia fish	(Megat Mazhar Khair et al., 2023)
Enterococcus faecalis	Isolated from urine	ATCC 29212
Bacillus subtilis	Test microorganism for food, media, and pharmerceutical testing	ATCC 6633
Bacillus cereus	Test microorganism for food, media, and pharmerceutical testing	ATCC 11778
Micrococcus luteus	Isolated from air and involves in pharmaceutical research	ATCC 10240
Lactobacillus plantarum	Isolated from pandan	UPMC 267 (Jalilsood et al., 2015)

4.2.2 Plasmid extraction

Plasmid extraction was done using FavorPrep Plasmid DNA Extraction Mini kit (Favorgen, Taiwan). A 3 mL overnight bacterial culture was transferred and pooled into a 1.5 mL microcentrifuge tube by centrifuging at 12 000 *x g* for 3 minutes to obtain the cell pellet and discard the supernatant. Washing step was performed by resuspending the cell pellet with 1 mL Phosphate-buffered saline (PBS) and centrifugation to remove the supernatant. The cell pellets were then resuspended completely with 250 µL FAPD1 Buffer containing 50 µg/mL RNase A. After that, 350 µL of FAPD3 Buffer were added and gently inverted 5 – 10 times to ensure complete neutralization of the mixture. The tube containing white precipitate was centrifuged for 10 minutes at 13 000 *x g*, 4 °C. After centrifugation, the supernatants were then carefully transferred to FAPD column with collection tube and centrifuged for 30 seconds at 11 000 *x g*. Four hundred microliter of W1 Buffer were then added and centrifuged for 30 seconds at 11 000 *x g*. Second washing step was performed by adding 750 µL of Wash Buffer to FAPD column and centrifuging for 1 minute at 11 000 *x g* after which the flow-through solution was discarded, followed by additional centrifugation for 7 minutes at 13 000 *x g* to removed residual liquid in the column. Then, the plasmid DNA was eluted by adding 50 µL of warm water and centrifuging for 5 minutes at 13 000 *x g*. The extracted plasmid DNA was then stored at 4 °C for short-term usage and -20 °C for long-term storage.

4.2.3 Cloning of endolysin (*Endo88*) gene.

The plasmid pET28a (Novagen) served as the expression vector for Endo88 in this study. The nucleotide sequence of the *S. aureus* Phage 88 (ATCC 33742-B1) endolysin (RefSeq ID: YP_240699; ORF006) was obtained from the genomic sequence (GenBank ID: NC_0070631) of Phage 88 (Kwan et al., 2005). The ORF006 was synthetically synthesized (IDT, Singapore) and inserted into the pUCIDT plasmid provided by the company. Then, 250 µL of FAPD2 Buffer were added to the synthesized plasmid and gently invert the tube 5 – 10 times and incubated at room temperate for 5 minutes for storage.

4.2.4 Primers design

Two different sets of primers (Table 4.2) were designed to introduce RE sites flanking the *Endo88* gene using Phusion High-Fidelity DNA polymerase (ThermoFisher, USA). BXEndo88 primer sets contained the *Bam*HI and *Xho*I, while the NXEndo88 primer sets contained the *Nco*I and *Xho*I restriction site. Moreover, the annealing temperature for the primers were predicted using Tm calculator web tool provided by ThermoFishe Scientific (https://www.thermofisher.com/my/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html).

Table 4.2: Primer sets used to clone *NXEndo88* and *BXEndo88*.

NXEndo88	NXEndo88-Fwd	5'- CATG *CCATGG* CA **ATGCAAGCAAAATTAACTAAAAAAG** -3'
	NXEndo88-Rev	5'- CC *CTCGAG* GCC **AGGTGACTTATGGGGAGAAATCAGT** -3'
BXEndo88	BXEndo88-Fwd	5'- CGC *GGATCC* GCG **ATGCAAGCAAAATTAACTAAAAAAG** -3'
	BXEndo88-Rev	5'- CC*CTCGAGG***ACTGATTTCTCCCCATAAGTCACCT**-3'

Italic: Restriction site; Bold: Endolysin sequence

4.2.5 Polymerase Chain Reaction (PCR)

The primers used to amplify the Endo88 gene are listed in Table 4.2. The PCR reaction was conducted according to the manufacturer's instructions for Phusion DNA polymerase kit (ThermoFisher, USA). Prior to amplification, the PCR profile was optimized with gradient PCR (gPCR) to determine the optimal annealing temperature for the primers. Three different annealing temperatures (55 °C, 60 °C, 65 °C) were used to assess the optimum annealing temperature for the primers. Next, conventional PCR reaction was carried out in a total reaction of 20 µL reaction mixture using the Master cycler Gradient (Eppendorf, USA). The reaction mixture included 1X of 5X Phusion HF Buffer which provided in the kit, 0.5 µM of forward and reverse primers, 200 µM of 10 mM dNTP, 0.02 U/µL of Phusion High-Fidelity DNA Polymerase and 1 µg of DNA template. The reaction was performed with settings as follows: (Step 1) pre-denaturation at 98 °C for 3 minutes, 1 cycle, (Step 2) denaturation at 98 °C for 10 seconds, annealing at 60 °C for 30 seconds, extension at 72 °C for 30 seconds, 35 cycles, and (Step 3) final extension at 72 °C for 5 minutes (1 cycle). Subsequently, the amplified products were analysed through agarose gel electrophoresis.

4.2.6 Agarose Gel Electrophoresis

Agarose gel electrophoresis was conducted to separate the size of the DNA bands. A 1% agarose gel was prepared by dissolving 0.25 g agarose powder (Vivantis, Malaysia) in a conical flask with 25 mL of 1x TAE (Tris-acetate, EDTA) buffer. The mixture was microwaved until boiling (approximately 1 minutes), swirled and re-heated until fully dissolved. After a brief cooling under running tap water, 2 µL of nucleic acid stain (ViSafe Red Gel Stain, Vivantis Malaysia) was added and mixed

well. The mixture was then poured into the pre-assembled gel cast with comb and allowed to solidify. The solid gel was immersed into the gel tank filled with 1X TAE running buffer. Prior to loading into the wells, the sample was mixed with 6x loading dye (ThermoScientific, USA). Gel electrophoresis was carried out at 90 V for 40 minutes or until the loading dye reached the bottom of the gel. The bands were then visualized using Benchtop UV Transilluminator (DaiggerScientific, Malaysia) and the images were captured using Amersham Imager 600 (GE Healthcare,USA).

4.2.7 DNA purification

After PCR, two DNA purification processes were employed based on the specific application using FavorPrep GEL/PCR Purification Kit (Favorgen, Taiwan).

(I) PCR clean-up: This method was employed when only a single bright band was observed on electrophoresis gel with the correct size. First, the PCR product was pooled into a 1.5 mL microcentrifuge tube and 5 volumes of FADF Buffer (100 µL PCR product + 500 µL FADF Buffer) were added. Then, the mixture was transferred into the FADF Column and centrifuged for 30 seconds at 11 000 *x g*, and the flow through was discarded. After that, 750 µL of Wash Buffer was added into the FADF column, centrifuged for 30 seconds at 11 000 *x g* and the flow through was discarded. The column was then dried by centrifuging again for 5 minutes at 17 000 *x g*. Then, PCR product was eluted by adding 40 µL of warm distilled water in the column and centrifuged for 5 minutes at 17 000 *x g*. The extracted PCR product was then store at 4 °C for short-term usage and -20 °C for long-term storage.

(II) Gel excision purification: In cases where the target band was bright and of the correct size multiple bands were presented on the electrophoresis gel even after

optimization, gel excision purification was used to purify the target gene. Additionally, this step was also applied to both restriction enzyme digestion (RE digestion) and Splicing by Overlapping PCR (SOE-PCR) which is elaborated below. First, the target band was excised from the electrophoresis gel and weighed. Then, 500 µL of FADF Buffer was then added and mixed by vortex. The mixture was then incubated at 60 °C and vortexed every 2 – 3 minutes until the gel slice fully dissolved into the solution. After that, the dissolved gel solution was then transferred to FADF column and centrifuged for 30 seconds at 11 000 *x g* after which the steps are as described in the PCR clean-up protocol above. Finally, the concentration and purity of purified products were measured using NanoPhotometer (Implen, Germany).

4.2.8 Restriction Enzyme (RE) digestion

The DoubleDigest Calculator web tool provided by ThermoFisher Scientific (https://www.thermofisher.com/my/en/home/brands/thermo-scientific/molecular-biology/thermo-scientific-restriction-modifying-enzymes/restriction-enzymes-thermo-scientific/double-digest-calculator-thermo-scientific.html) was used to determine the optimal double digestion reaction conditions for the target genes and pET28a. Two different set of primers were designed to introduce restriction (RE) sites flanking the Endo88 gene (Denoted as: *BXEndo88* and *NXEndo88*). The recipe for the reaction is listed below (Table 4.3 and Table 4.4). In brief, total 50 µL of reaction mixture was incubated at 37 °C, overnight and enzyme deactivation was done at 80 °C, 10 minutes. The double digested DNA products were then purified using the gel excision purification method which was mentioned in section 4.2.1.7.

Table 4.3: Double-digestion recipe for BXEndo88

Ingredients	Final amount in 50 µL
*Bam*HI	2 µL(10 U)
*Xho*I	1 µL(5 U)
10x Tango Buffer	10 µL (2x)
DNA	1 µg
dH_2O	Top up to 50 µL

Table 4.4: Double-digestion recipe for NXEndo88

Ingredients	Final amount in 50 µL
*Nco*I	1 µL(5 U)
*Xho*I	1 µL(5 U)
10x Tango Buffer	10 µL (2x)
DNA	1 µg
dH_2O	Top up to 50 µL

4.2.9 Ligation of pET28a with endolysin gene

Ligation of the double digested target gene and pET28a were done using T4 DNA Ligase (Vivantis, Malaysia). A ligation ratio (1:5) was calculated using the ligation calculator web tool provided by Qiagen (link), and the recipe for ligation is listed in Table 4.5. The mixture was then incubated in 4 °C overnight.

Table 4.5: Ligation recipe for double digested gena and pET28a

Ingredients	Final amount in 50 µL
dH_2O	Top up to 20 µL
10x T4 buffer	2 µL (1x)
Vectors	150 ng
Inserts	205 ng
T4 Ligase	0.5 µL (1 U)

4.2.10 Preparation of *E. coli* BL21(DE3) competent cell

Preparation of *E. coli* BL21(DE3) competent cells was done using the $CaCl_2$ treatment method (Chan et al., 2013). Initially, a single bacterial colony was inoculated into 10 mL LB broth and incubated at 37 °C, 200 rpm for 16 – 18 hours. The overnight culture was then sub-cultured again into fresh LB broth and grown until reaching the log-phase (OD_{600}: 0.5 ~ 0.6). After incubation, the cells were harvested by centrifugation at 4000 *x g* for 8 minutes at 4 °C. After the supernatant was discarded, the cell pellets were gently resuspended in 10 mL of ice-cold 0.1 M $CaCl_2$ and incubated on ice for 30 minutes. Subsequently, the culture was centrifuged again at 4000 *x g* for 8 minutes at 4 °C and gently resuspended again with 2 mL storage buffer (0.1 M $CaCl_2$, 15 % glycerol). The competent cells were then aliquoted into a sterile 1.5 mL microcentrifuge tubes, with a total volume of 100 µL each and stored at -80 °C.

4.2.11 Transformation

Transformation of *E. coli* BL21(DE3) competent cells were done using the heat-shock method. Five microliter of the ligation mixture was gently added into competent cells, and the mixture was incubated on ice for 10 minutes. After incubation, the cultures were then heat-shocked at 42 °C for 30-50 seconds and then incubated back on ice for 5 minutes. One milliliter of LB broth was then added to the culture and allowed to grow at 37 °C, 250 rpm for 1 hours. The cultures were then plated on LB agar supplemented with 50 µg/mL Kanamycin and incubated at 37 °C overnight. The putative positive *E.coli* BL21(DE3) transformants were initially verified by their ability to grow on LB agar supplemented with 50 µg/mL Kanamycin. The colonies were then picked and proceeded with colony PCR screening.

4.2.12 Colony PCR screening

Colony PCR screening was used to screen transformants which harbored recombinant pET28a (*Endo88*). The PCR reaction was performed using *Taq* DNA polyemerase (1st Base, Malaysia) and a pair of primers targeting upstream and downstream of the multiple cloning side (pETF2_Fwd: 5'- TCCCGCGAAATTAATACGAC-3'; pETR_Rev: 5'- GCTGAGCAATAACTAGCATA -3'). The reagents are listed below (Table 4.6) and the template preparation was slightly different from conventional PCR. After transformation, the colonies were labelled alphabetically and streaked onto a fresh LB agar supplemented with 50 µg/mL Kanamycin and incubated overnight at 37 °C. Using a sterile pipette tip, the cultures were swabbed and immersed into a PCR tube containing 20 µL sterile dH_2O. The tubes were then boiled for 10 minutes and spinned down with benchtop mini-centrifuge to separate the debris from the supernatant. The 20 µL of the solution was then used as the template for colony PCR.

Table 4.6: Recipe for colony PCR screening

Ingredients	Final amount in 25 µL
dH_2O	Top up to 25 µL
10x *Taq* buffer	1X
dNTP Mix	0.2 mM
Forward primer	0.1 µM
Reverse primer	0.1 µM
Taq DNA polymerase	1.25 U
$MgCl_2$	2.5 mM
Template	0.5 µL

4.2.13 Sanger sequencing of insert

The same pairs of primers in the colony screening PCR were used for DNA sequencing. The purified recombinants plasmid, along with the primers, were sent to 1st Base Company for sequencing services. Subsequently, the sequencing results were aligned with the expected sequence to ensure the absence of mutations. All sequencing results are appended in Appendix A.

4.2.2 Protein expression and purification

BXEndo88 and NXEndo88 were expressed at two different temperature (18 °C and 37 °C) and different concentrations of β-D-1-thiogalactopyranoside (IPTG) to check the solubility of the expressed endolysin. Overnight culture of *E. coli* BL21(DE3) harboring recombinant pET28a harbouring *BXEndo88* or *NXEndo88* were cultured in LB broth supplemented with 50 µg/mL Kanamycin at 37 °C, 200 rpm until reaching mid-log phase (OD_{600}: 0.4 – 0.6). Subsequently, gene expression was induced by adding different concentrations of IPTG (0.05 mM, 0.1 mM, 0.5 mM, 1 mM), and the bacteria were incubated at 200 rpm for 18 hours at 18 °C and 4 hour at 37 °C respectively. After induction, the cells were harvested by centrifugation (4000 *x g*, 4 °C for 10 minutes), washed two times with PBS and resuspended in Tris buffer (50mM Tris-HCl, pH 7.5; 150 mM NaCl) at a ratio of 1:10 (1g of cell pellet to 10 mL of buffer). The cells were ruptured physically using sonication (QSONICA, USA) for 5 minutes (30 seconds sonication separated by 30 seconds rest) in an ethanol ice bath.

4.2.2.1 Purification of BXEndo88 (Inclusion body) under denaturing condition.

The protocol of protein purification under denaturing condition was adapted from other studies (Schmelcher, Korobova, et al., 2012; S. M. Singh & Panda, 2005). After *BXEndo88* was induced at 37 °C, 4 hours, the inclusion bodies obtained from the pellet fraction were treated with 8 M urea and incubation on ice for 30 minutes. Then, the mixture underwent centrifugation (4000 *x g*, 20 minutes, 4 °C) to separate the supernatant from the pellet. The supernatant was then proceeded with purification under denaturing conditions using cOmplete His-tag Purification Resin column gravitational dripping method (Merck), with decreasing concentrations of urea in the washing steps (6, 4 and 2 M) and elute with elution buffer (250 mM imidazole, 50 mM Tris-HCl, 150 mM NaCl). Lastly, the buffer in the eluted sample was exchanged for Tris buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) using Pierce Protein Concentrator with a polyethersulfone membrane and a 5k molecular cutoff (Thermo Fisher Scientific).

4.2.2.2 Purification of NXEndo88 (Soluble fraction) under native condition

NXEndo88 was induced at 18 °C, 200 rpm for overnight and proceeded with sonication. Then, the mixture underwent centrifucation (4000 x *g*, 20 minutes, 4°C to obtain the soluble fraction (supernatant). Subsequently, the soluble fraction underwent filtration with a 0.22 µM syringe filter (Merck). NXEndo88 in the soluble fraction was then purified using cOmplete His-tag resin column (Merck) with gravitational dripping method. Elution was done with elution buffer (250 mM imidazole, 50 mM Tris-HCl, 150 mM NaCl). Lastly, the buffer in eluted sample was exchanged for Tris buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) using Pierce Protein Concentrator with a

polyethersulfone membrane and a 5k molecular cutoff (MWCO) (Thermo Fisher Scientific).

4.2.2.3 Lysis buffer optimisation for protein stability

Different lysis buffers (Table 4.7) were tested during sonication to determine the effect of lysis buffer on the solubility of expressed protein (Anthony C.Grabski, 2009; Hjertén et al., 1988; Kragh-Hansen et al., 1998; Son YJ et al., 2008; Vagenende et al., 2009, 2009; Y. Wang et al., 2009). Two gram of cell pellet (g) to 20 mL lysis buffer (mL) (1:10) and proceed with sonication. After sonication, the cell lysate was incubated on ice for 30 minutes and centrifuged (4000 *x g*, 20 minutes, 4 °C) to separate the soluble fraction (supernatant) and pellets (inclusion bodies). The total protein of the each fractions were then quantified using Bicinchoninic acid assay (BCA) and analysed using SDS-PAGE.

Table 4.7: Optimisation of lysis buffer for sonication to reduce formation of inclusion body.

Buffer A	Buffer B	Buffer C	Buffer D	Buffer E
150 mM NaCl	150 mM NaCl	150 mM NaCl	150 mM NaCl	150 mM NaCl
50 mM Tris-HCl (pH 7.4-7.6)	50 mM Tris-HCl (pH 7.4-7.6)	50 mM Tris-HCl (pH 7.4-7.6)	50 mM Tris-HCl (pH 7.4-7.6)	50 mM Tris-HCl (pH 7.4-7.6)
	0.1% SDS		0.1% SDS	0.1% SDS
		0.05% Beta-mercaptoethanol	0.05% Beta-mercaptoethanol	0.05% Beta-mercaptoethanol
dH_2O	dH_2O	dH_2O	dH_2O	dH_2O

4.2.2.4 Bicinchoninic acid assay (BCA)

Protein concentration was quantified using the microplate procedure in Pierce BCA Protein Assay Kit (ThermoFisher, USA). The diluted albumin (BSA) standards and protein standard curve were prepared according to manufacturer's manual. The working reagent (WR) was prepared in 50:1 ratio (Reagent A: B) and 200 µL of the mixture was then aliquoted into each well in a 96-well plate. After that, 25 µL of samples or BSA were then added and resuspended into the well and incubated for 30 minutes at 37 °C. Lastly, the plate was then measure at 570 nm absorbance. All the samples were measured in triplicate.

4.2.2.5 Sodium dodecyl sulphate polyacrylamide Gel Electrophoresis (SDS-PAGE) analysis.

All proteins were analysed using SDS-PAGE prepared using the Mini-PROTEAN Tetra Handcast system (Bio-rad, USA) (Gallagher & Wiley, 2012). The gel casted contained 15 % resolving gel and 4 % stacking gel and the detailed composition of all buffers and materials used in SDS-PAGE is provided in Appendix B After assembling the casting apparatus, resolving gel solution was loaded into the apparatus until 75 % full. Butanol was pipetted into the glass cassette to ensure the separating gel produced a smooth and straight surface on top of the separating gel. After polymerization, the glass cassette was then washed with distilled water to remove the remaining butanol and dried with blotting paper to remove water. Subsequently, 4 % stacking gel was then gently loaded on top of the resolving gel and 1.0 mm gel-comb (10 wells or 15 wells) was inserted into the fully loaded stacking gel. After polymerization, the comb was removed gently and the glass cassette was then transferred to the Mini-PROTEAN Tetra vertical chamber (Bio-rad, USA) and filled with 1X SDS running buffer. Before

loading into the well, protein samples were pre-treated by mixing with (4x loading buffer, and boiled for 10 minutes. The mixtures were then loaded into the well and ran together with a 5 – 245 kDa protein ladder (1st Base, Malaysia) at 100 V for 90 minutes or until the dye reached the gel end. After electrophoresis, the stacking gel was removed using a gel releaser. The resolving gel was stained with Coomassie blue staining solution by immersing the gel into the staining solution and microwaved until it started bubbling. The gel was then incubated with slight agitation for 2 hours. After staining, the stained gel was gently washed by distilled water and immersed into Destaining solution and incubated overnight or until the bands can be clearly seen.

4.2.2.6 Western blotting analysis

After SDS-PAGE, the gel was proceeded to wet Western blotting analysis to detect the His-tag embedded on the endolysin. The detailed composition of all buffers and materials used in western blot is provided in Appendix B. A transfer sandwich was prepared using Mini Trans-blot Module (Bio-Rad, USA) by stacking two pieces of Whatman 3 MM filter paper which were pre-soaked in ice-cold 1x Western blot transfer buffer. Then, a polyvinylidene difluoride membrane (PVDF) (GE Healthcare Life Science, USA) was activated by soaking into 100 % methanol before stacking on top of the two filter papers, followed by the SDS-PAGE gel and another two piece of filter papers. The staking process was done by soaking the materials in ice-cold transfer buffer and gently pressing to remove bubble trapped between the sandwich. Subsequently, the sandwich was transferred into the blotting tank filled with ice cold 1x Transfer buffer and the protein was blotted at 100 V, 90 minutes in cold room. After blotting was completed, the PVDF membrane was proceeded to blocking step by soaking the membrane with 1 % bovine serum albumin (BSA) and incubated at room

temperature for 1 hours. After blocking, the membrane was then washed three times with Tris-Buffered Saline with 0.1 % Tween-20 (TBST) for 10 minutes. Then, the membrane was then incubated with 1:1000 primary antibody; mouse Anti Hia-tag Monoclonal Antibody (Thermo Scientific, USA) at 4 °C, overnight. Next, the membrane was washed three times, five minutes each and incubated with 1:50 000 secondary antibody; goat anti-mouse IgG peroxidase-conjugated antibody (Thermo Scientific, USA) for 1h at room temperature. Another washing step (three times, 5 minutes each) was performed after the secondary antibody incubation. Finally, the bands were then visualized using SuperSignal West Dura substrate (Thermo Scientific) and viewed under Amersham imager 600 (GE Healthcare Life Science).

4.2.3 Protein functionality assessments

The functionality of the endolysin was assessed by the lytic ability on *Staphylococcus aureus* PS88, through well diffusion plate lysis assay and turbidity reduction assay.

4.2.3.1 Plate lysis well-diffusion assay (PLA)

Method of lytic activity assessment was adapted from Chandran et al. 2022 with slight modifications. *S. aureus* PS88 overnight culture was inoculated into a 10 mL fresh BHI broth and incubated at 37 °C, 200 rpm until mid-log phase (OD_{600}: 0.4 – 0.6). Alternatively, BHI agar plate (0.7% soft agar) was prepared and 1/1000 of PS88 was inoculated into the warm sterile agar solution and mixed well. After pour plate, the plate was then allowed to solidify in laminar flow. Subsequently, wells were then punched into the agar using sterile 1 mL pipette tips and 100 µg of endolysin were pipetted into the well and allowed to diffuse completely.

4.2.3.2 Turbidity reduction assays (TRA)

The experiment was adapted from Zhang et al. 2019, with some modification. Overnight cultured of PS88 was sub-cultured into fresh BHI broth and grown to mid-log phase (OD_{600}: 0.4 – 0.6) at 37 °C with shaking. The pellets were then harvested by centrifugation (4000 *x g*, r.t., 10 minutes), washed twice with PBS and finally resuspended in Tris buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5), and standardized to OD_{600}: 0.6 – 0.7 using 96-well plate and nanophotometer. The bacterial suspension was then pipetted into 96-well plate in aliquots of 180 µL , and 10 µg of endolysin (20 µL) was then mixed with the culture to a final volume of 200 µL per well. The assay was also repeated with varying amounts of endolysin (0.1 µg, 1 µg, 5µg and 10 µg). The OD decrease was measured every five minutes using the iMark Microplate Absorbance Reader (Bio-Rad). The lytic activity was quantified by measuring the percentage OD drop with the formula below (adapted from Ding et al. 2020)

$$\frac{(\text{OD600 of control} - \text{OD600 of the reacion mixture with endolysin})}{\text{Initial OD}} \text{ X } 100$$

4.2.4 Biophysical characterization of NXEndo88

Biophysical properties were determined by assessing how changes in environmental factors such as NaCl concentration, pH and temperature impacted on endolysin lytic activity. Different concentrations of NaCl were tested to assess the influence of ionic strength on the endolysin stability. The optimal pH range was determined by exposing the endolysin to different pH conditions. Additionally, the impact of temperature was examined by exposing the endolysin to different temperature prior to lytic assay.

4.2.4.1 Effect of Sodium chloride (NaCl) on lytic activity

Sterile 1M NaCl stock solution was prepared. Different concentration (50 mM, 100 mM, 150 mM, 300mM, 500 mM) of NaCl solutions were prepared by diluting the 1M NaCl stock with 50 mM Tris-HCl. Turbidity reduction assay was conducted as mentioned above (section 4.2.3.2), with slight changes. After the washing step, the mid-log phase cell pellet was resuspended with Tris buffer containing different concentration of NaCl. Then, the OD was standardised to 0.6 – 0.7 and proceeded with the assay using 10 µg of endolysin (adapted from Guo et al., 2017).

4.2.4.2 Effect of pH on lytic activity

To evaluate the effect of pH on lytic activity, TRA was conducted with slight changes on the buffer. The pH of 50 mM sodium acetate buffer was adjusted to pH 4.0, pH 5.0 and pH 6.0 and 50 mM Tris-HCl buffers which were adjusted to pH 7.0 – 11. These buffers were then used to resuspend the mid-log phase cell pellets and proceeded with TRA. Bacteria in each buffer without endolysin treatment were used as the negative control, to ensure that the bacteria was able to survive at the respective pHs before treatment with endolysin (adapted from Guo et al., 2017).

4.2.4.3 Effect of temperature on lytic activity

Turbidity reduction assay (TRA) was used to evaluate the effect of temperature on lytic activity, the steps were similar to section 4.2.3.2 with slight changes. Prior to treatment of endolyin on the bacteria, the endolysin was aliquoted into PCR tubes which were then incubated at different temperature (Table 4.8) (adapted from Ding et al., 2020; Obeso et al., 2008). The rationale for testing temperatures ranging from -20 °C to 100 °C was to evaluate the thermal stability of Endo88 under various conditions

relevant to storage and application. Room temperature (26 °C) and the optimal bacterial growth temperature (37 °C) were included, as 37 °C is also used for the lytic assay later. Gradual increases in temperature (45, 55, and 65 °C) allowed for the examination of enzyme stability at each increment. Low temperatures (-20 and 4 °C) were also used to assess the potential preservation of enzymatic activity during freezing/cold condition (Filatova et al., 2015), while high temperatures (100 °C) determined the protein's thermal denaturation threshold. Exposing the enzyme solution to 100 °C for more than 10 minutes resulted in precipitation, which affected the protein concentration to be used in subsequent analysis; thus, 10 minutes exposure was used to evaluate its stability. In contrast, longer incubation (60 minutes) allowed the protein to equilibrate for other selected temperatures, including freezing at -20 °C. However, further studies are required to evaluate long-term storage stability at -20 °C, including prolonged storage duration (few months or years) and the optimization of cryoprotectants (Cao et al., 2003; H. Park et al., 2021; Umbach et al., 2019).

Table 4.8: Effect of different temperature treatment on lytic activity

Temperature (°C)	Treatment times (minutes)
-20	60
4	60
26	60
37	30
45	30
55	30
65	30
100	10

4.2.5 Cloning (SOE-PCR) and expression of mutants

SOE-PCR was used to introduce the mutations designed in Chapter 3 into the wild-type Endo88 gene. The mutated gene was then cloned into pET28a using the same methodology in section 4.2.1. The mutants were generated by separating the gene into a few regions with primer-initiated complementary overhangs and the procedures are shown in Figure 4.1.

4.2.5.1 Construction of mutant gene (M, A1, A2)

The mutations were introduced into the SH3 domain by use of splicing by overlap PCR (SOE-PCR). SOE-PCR is a type of PCR which is used to generate point mutations at a specific place on a gene sequence (Higuchi et al., 1988; Ho et al., 1989; Zarghampoor et al., 2020). Additionally, it can be utilized to generate chimeric genes without the use of restriction sites (Horton et al., 1989). The SH3 domain gene sequence was separated into different regions by use of primers with overhang, where the overhang contained the mutation and the overhang for the first region is complementary to the overhang from the second region (Example shown in Figure 4.1A). Prior to joining the two regions, each regions were gel purified to prevent unspecific binding. The two regions were then allowed to anneal without primers (Table 4.9) and the ratio (1:1) amount of both regions were calculated using a ligation calculator (Qiagen webtool) (Figure 4.1 B). The annealed gene were then amplified from the reaction mixture using NXEndo88 primers (Table 4.2), and gel purified for downstream application.

Figure 4.1: Schema of mutant gene (A1) development through SOE-PCR. Red colour regions denote the overlapping region which allows annealing of region 1 and region 2. Green colours represents the original Endo88 sequence. **(A):** Two sets of primers were designed to amplify region 1 and region 2. **(B):** Each primer set generated sequences with flanking region that contain the mutations and overlapping region to facilitate the joining of the two fragments. The two regions were joined into one continuous DNA fragment successfully incorporating the mutation (Red region).

Table 4.9: Reagents for SOE-PCR, annealing condition for region joining.

Ingredients	Final amount in 25 µL	PCR conditions
dH_2O	Top up to 25 µL	98 °C 3 mins
5x Phusion HF buffer	1X	98 °C 10 s
dNTP Mix	0.2 mM	65 °C 30 s x15 cycles
Phusion High-Fidelity DNA polymerase	1.25 U	72 °C 30 s
Region 1	150 ng (6 ng/µL)	72 °C 5 mins
Region 2	14.56 ng (0.5824 ng/µL) (Mutant M) 28.2 ng (1.128 ng/µL) (Mutant A1) 21 ng (0.84 ng/µL) (Mutant A2)	4 °C -

The primers used for splicing by overlap extension PCR (SOE-PCR) were stated in Table 4.10. Mutant M was initially fused with pNZ8048 plasmid vector for an attempt to express in *Lactococcus lactis* NZ9000. However, the attempt was not successful and switched the expression system to *E. coli* BL21(DE3). The mutant M gene used in this study was directly amplified from recombinant pNZ8048(M) using NXEndo88-Forward and NXEndo88-Reverse primer (Section 4.2.1.4, Table 4.2). The PCR conditions and cloning methods were same as in section 4.2.3.

Table 4.10: SOE-PCR primer set for the mutants.

Mutant Gene	Primer	Primer sequence	Ta (°C)
M	Region1PstM-Fwd	5'- *CTGCAG* CC **ATGCAAGCAAAATTAACTAAAAAAG** -3'	60
	Region1PstM-Rev	5'- **AAGATGGTCATGTTTGGGTA** AGTTGGACAACAAGTGATGGC -3'	60
	Region2XbaIM-Fwd	5'- AGTTGGACAACAAGTGATGGC **CAACGTTATTACTTGCCTATAGA** -3'	60
	Region2XbaIM-Rev	5'- *TCTAGA* TTA CATCACCATCATCACCAC **ACTTATGGGGAGAAATCAGT** -3'	60
AVANA mutant – A1	Region1A1-Fwd	5'- CATG *CCATGG* CA **ATGCAAGCAAAATTAACTAAAAAAG** -3'	60
	Region1A1-Rev	5'- **TACTACATGGAAGAAAGTGCTAGA** TTTACCCCGAATACCGATATTATTACCCGCACCACCGGCCCG -3'	60
	Region2A1-Fwd	5'-TTTACCCCGAATACCGATATTATTACCCGCACCACCGGCCCG **TTCTTATCATGCCCGGTAGCTTA** -3'	60
	Region2A1-Rev	5'- CC *CTCGAG* GCC **AGGTGACTTATGGGGAGAAATCAGT**-3'	60
AVANA mutant – A2	Region1A2-Fwd	5'- CATG *CCATGG* CA **ATGCAAGCAAAATTAACTAAAAAAG** -3'	60
	Region1A2-Rev	5'- **CATTCTTATCATGCCCGGTA** AGCGGCGTGCTGAAAGCGGGCCAGACCATTCATTATGATGAAGTGATGAAA-3'	60
	Region2A2-Fwd	5'-AGCGGCGTGCTGAAAGCGGGCCAGACCATTCATTATGATGAAGTGATGAAA **CAAGATGGTCATGTTTGGGT** -3'	60
	Region2A2-Rev	5'- CC *CTCGAG* GCC **AGGTGACTTATGGGGAGAAATCAGT**-3'	60

Bold: Template's sequence; *Italic*: Restriction site; Red: SOE-PCR overlapping site. Note: (i) AVANA overlapping region was codon optimized using online tool-Codon usage Database (Nakamura,Y *et al*, 2000). (I) Mutant M was initially cloned into pNZ8048 for alternative study (unpublished), the mutant M gene used in this study was directly amplified from recombinant pNZ8048 + M using NXEndo88-Forward and NXEndo88-Reverse primer. The primers used for SOE-PCR in the case of mutant M were also provided in the table.

4.2.5.2 Expression of mutants and lytic host range assessments

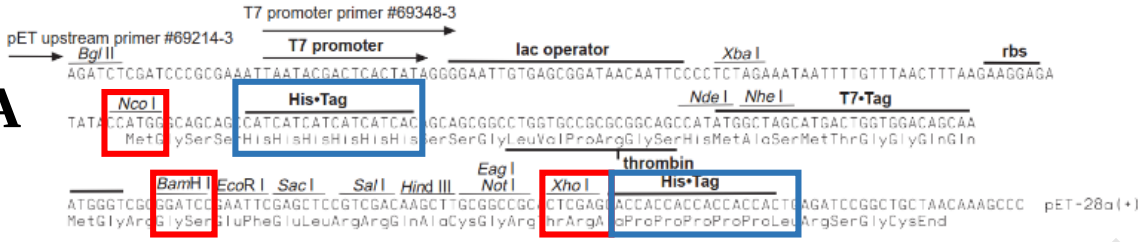
The mutants were expressed and purified with a similar condition to NXEndo88 (Section 4.2.2.2), with slight modifications in the buffer during buffer exchange. In brief, NaCl was not added in the buffer used for exchange, as the endolysin lysed better without the presence of NaCl (shown in 4.3.4.1). Subsequently, wild type endolysin and the mutants underwent host range analysis against various bacterial genera, including *Staphylococcus* (*S. aureus* PS88, *S. epidermidis*, *S. hominis*), *Enterococcus* (*E. faecalis*), *Streptococcus* (*S. pyogenes, S. agalactiae*), *Bacillus* (*Bacillus cereus, B. subtilis*), *Micrococcus* (*M. luteus*) and *Lactobacillus* (*L. plantarum*). Turbidity reduction assay was used to assess the ability of the endolysin to lyse different genus of bacteria.

4.3 Results & Discussion

4.3.1 Cloning strategy using different set of primers.

This section describes the effect of different cloning strategies in pET28a expression vector on solubility and functionality of Endo88, as previously published in our study (Tham et al., 2020). In this study, two different sets of primers were used to amplify *Endo88* gene: NXEndo88-Forward (Fwd), Reverse (Rev), and BXEndo88-Fwd, Rev (Table 4.2). Both primer pairs produced slightly differing sequences of the *Endo88*, with differences in the restriction enzyme (RE) sites and the location of the His-tag (Figure 4.2A). Initially, the *BXEndo88* primer sets were utilised, resulting in the expression of *BXEndo88*. This construct comprised of a His-tag, T7 tag and an additional 35 amino acids from the plasmid at the N-terminal, along with a 5x proline tail and 8 amino acids from the plasmid at the C-terminal (Figure 4.2B). Conversely, the *NXEndo88*-Fwd, Rev primers generated *NXEndo88* with the N-terminal His-tag and T7 tag being removed, and the His-tag was located at the C-terminal (Figure 4.2B).

A



B

>BXEndo88

MGSS*HHHHHH*SSGLVPRGSHMASMTGGQQMGRGSAMQAKLTKKEFIEWLKTSEGKQFNVDLWYGFQCFDY
ANAGWKVLFGLLLKGLGAKDIPFANNFDGLATVYQNTPDFLAQPGDMVVFGSNYGAGYGHVAWVIEATLDYII
VYEQNWLGGGWTDRIEQPGWGWEKVTRRQHAYDFPMWFIRPNFKSETAPRSIQSPTQASKKETAKPQPKAVE
LKIIKDVVKGYDLPKRGGNPKGIVIHNDAGSKGATAEAYRNGLVNAPSSRLEAGIAHSYVSGNTVWQALDESQVG
WHTANQLGNKYYYGIEVCQSMGADNATFLKNEQATFQECARLLKKWGLPANRNTIRLHNEFTSTSCPHRSSVLH
TGFDPVTRGLLPEDKQLQLKDYFIKQIRVYMDGKIPVATVSNESSASSNTVKPVASAWKRNKYGTYYMEESARFT
NGNQPITVRKIGPFLSCPVAYQFQPGGYCDYTEVMLQDGHVWVGYTWEGQRYYLPIRTWNGSAPPNQILGDL
WGEISPRAPPPPPLRSGC

>NXEndo88

MAMQAKLTKKEFIEWLKTSEGKQFNVDLWYGFQCFDYANAGWKVLFGLLLKGLGAKDIPFANNFDGLATVYQ
NTPDFLAQPGDMVVFGSNYGAGYGHVAWVIEATLDYIIVYEQNWLGGGWTDRIEQPGWGWEKVTRRQHAY
DFPMWFIRPNFKSETAPRSIQSPTQASKKETAKPQPKAVELKIIKDVVKGYDLPKRGGNPKGIVIHNDAGSKGATA
EAYRNGLVNAPSSRLEAGIAHSYVSGNTVWQALDESQVGWHTANQLGNKYYYGIEVCQSMGADNATFLKNEQ
ATFQECARLLKKWGLPANRNTIRLHNEFTSTSCPHRSSVLHTGFDPVTRGLLPEDKQLQLKDYFIKQIRVYMDGKI
PVATVSNESSASSNTVKPVASAWKRNKYGTYYMEESARFTNGNQPITVRKIGPFLSCPVAYQFQPGGYCDYTEV
MLQDGHVWVGYTWEGQRYYLPIRTWNGSAPPNQILGDLWGEISPLE*HHHHHH*

Bold: Methionine, *Italic*: His-tag, Underline: Endolysin amino acid sequence, Orange box: Extra amino acid from pET28a vector

Figure 4.2: Cloning strategy on pET28a by using different set of primers. (A) pET28a vector map adapted from Novagen. Red box: Restriction site used in this study; Blue box: Two his-tag located at different frames. (B) Amino acid sequence for NXEndo88 and BXEndo88. NXEndo88 consisted of the endolysin amino acid sequences with a C-terminal his-tag and BXEndo88 consisted of the endolysin's amino acid sequence, a N-terminal his-tag and other amino acid sequences from pET28a including a T7.Tag (Adapted from Tham et al., 2020).

4.3.1.1 Gradient PCR Optimisation of *BXEndo88* & *NXEndo88* using two sets of primer with different Restriction Enzyme site (RE).

Prior to conventional PCR, gradient PCR was employed to determine the optimal annealing temperature of the primer sets. Annealing temperatures ranging of 55°C, 60°C and 65°C were tested for each reaction. The DNA band sizes of the PCR products at different annealing temperature were visualized using agarose gel electrophoresis (Figure 4.3). *BXEndo88* (1464 bp) and *NXEndo88* (1464 bp) were successfully amplified across all tested annealing temperature given.

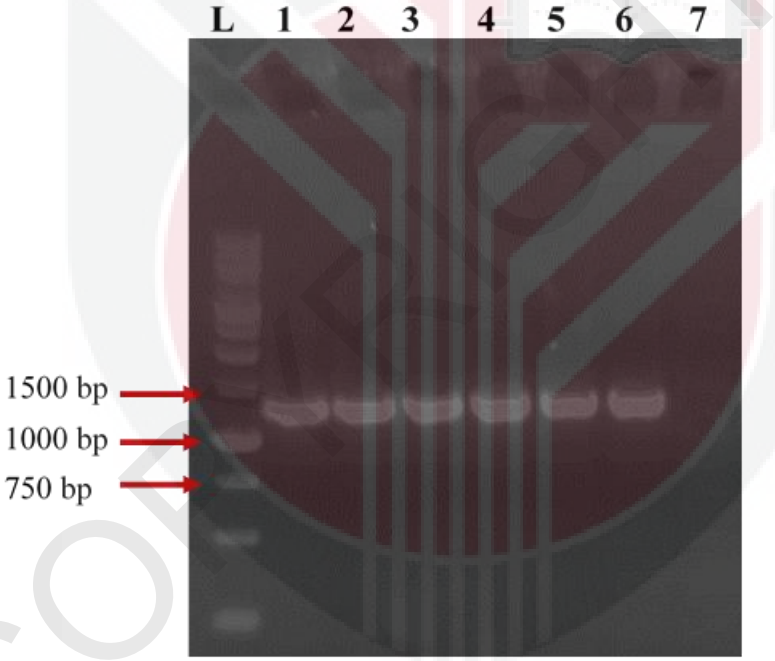


Figure 4.3: Gradient PCR product at different annealing temperature (55°C, 60°C & 65°C). Lane L: 1 kb DNA ladder (1st Base, Malaysia); Lane 1 to 3: PCR product of *BXEndo88* at 55°C, 60°C & 65°C; Lane 4 to 6: PCR product of *NXEndo88* at 55°C, 60°C & 65°C; Lane 7: Negative control (PCR Master mix without DNA template).

4.3.1.2 Cloning of *Endo88* gene (*BXEndo88* & *NXEndo88*) with pET28a expression vector into *E. coli* BL21(DE3) strain.

Recombinant plasmid pET28a harbouring *BXEndo88* or *NXEndo88* were generated using different pairs of restriction enzyme (RE). *Bam*HI & *Xho*I were used to cleave *BXEndo88*, while *Nco*I & *Xho*I were used to cleave *NXEndo88*. Concurrently, pET28a underwent double digestion with the corresponding RE. Double digestion of pET28a was expected to produce DNA band approximately 5323 bp (*Bam*HI & *Xho*I) and 5225 bp (*Nco*I & *Xho*I). Following electrophoresis, a single band was observed between 5000 bp and 6000 bp on the 1 kb ladder (Figure 4.4). Subsequently, the product of double digestion was then subjected to gel excision purification (section 4.2.1.7).

Both the purified insert (double digested *Endo88*) and vector (double digested plasmid) were then subjected to ligation (section 4.2.1.9) followed by transformation (section 4.2.1.11). The colonies were then picked with sterile pipette tips and subjected to colony PCR screening (section 4.2.1.12). Transformants harbouring the recombinant plasmid were expected to yield bands larger than the insert due to the design of the primers, which were positioned approximately 100bp away from both the upstream and downstream of the pET28a multiple cloning site. As depicted in Figure 4.5, transformants harboring the recombinant plasmid (lane 3, 4, 6, 7, 8) exhibited a band size slightly larger than the insert (above 1.5 kb band size in ladder), while those with empty plasmid (no insert) showed bands approximately 200 bp in size or no bands at all. Subsequent verification of the insert was done using DNA sequencing, which revealed that the sequence matched 100 % with the expected sequence (Appendix A).

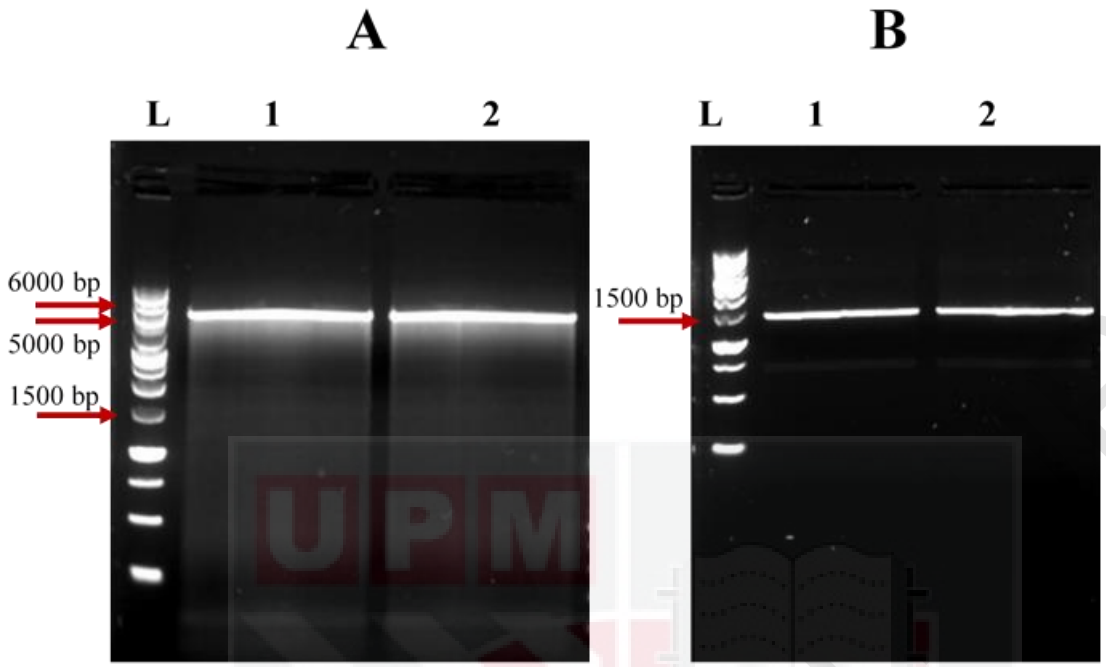


Figure 4.4: Double digestion of pET28a and PCR product (*BXEndo88* & *NXEndo88*) with two different pairs of RE. **(A)**: Lane L: 1 kb Ladder (1st Base, Malaysia), Lane 1: pET28 (~ 5.3 kb) double digested with *Bam*HI and *Xho*I; Lane 2: pET28a (~ 5.3 kb) double digested with *Nco*I and *Xho*I. **(B)**: Lane L: 1 kb Ladder (1st Base, Malaysia), Lane 1: *BXEndo88* (~1.5 kb) double digested with *Bam*HI and *Xho*I; Lane 2: *NXEndo88* (~1.5 kb) double digested with *Nco*I and *Xho*I.

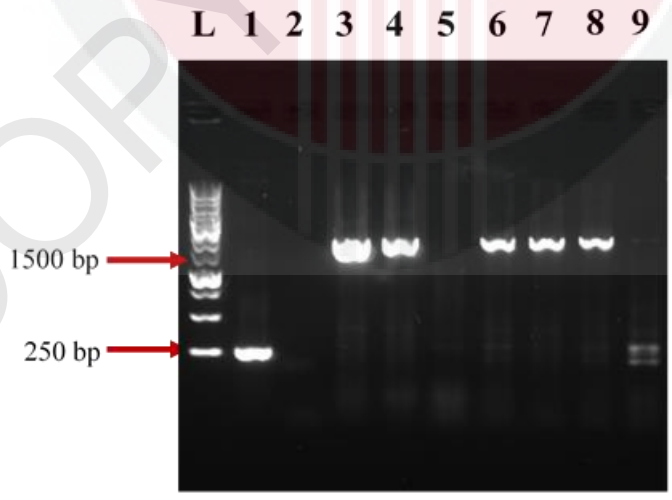


Figure 4.5: Colony PCR of recombinant *E. coli* harbouring pET28a (*BXEndo88*) and pET28a (NXEndo88). Lane L: 1 kb Ladder (1st Base, Malaysia); Lane 1 to 4: colony screening PCR of colonies acquiring pET28a (*BXEndo88*); Lane 5 to 8: colony screening PCR of colonies acquiring pET28a (*NXEndo88*). Lane 9: Negative control (Master mix without template).

4.3.2 Optimisation of protein expression and downstream processing of BXEndo88.

Prior to large scale protein expression, the induction of endolysin gene was optimized for optimum expression conditions. Both pellet (inclusion body) and supernatant (soluble fraction) were analysed via SDS-PAGE to determine the condition that yields the highest observable amount of soluble proteins (endolysin). Soluble proteins are found in the soluble fraction, while the proteins being separated in the pellet are usually non-functional aggregated protein caused by mis-folding which are known as the inclusion bodies (Bhatwa et al., 2021; Kopito, 2000; A. Singh et al., 2015). Expression of inclusion bodies is a common phenomenon in the over-expression of recombinant protein in *E. coli* expression systems (Baneyx, 1999; Sørensen & Mortensen, 2005; Upadhyay et al., 2012); it often caused low-yield of functional proteins during protein expression.

4.3.2.1 Optimisation of protein expression with different induction temperature and different concentration of IPTG.

One of the factors that contributed to inclusion body formation is the increase in temperature during gene induction. An increase in temperature may expose the cells to optimum growth rate which increases the metabolism during over expression, consequently causing the the folded protein intermediates to accumulate (Neudecker et al., 2012; Santucci et al., 2008; Sørensen & Mortensen, 2005) which is further exacerbated by insufficient chaperones which aids in folding (Sørensen & Mortensen, 2005). Moreover, depending on the protein being expressed, the induction temperature might not be the optimum temperature for protein stability (Hunke & Betton, 2003). Therefore, to counter this, expression at low temperature is thought to be a better solution (de Groot & Ventura, 2006; Hunke & Betton, 2003; Song et al., 2012). In this

study, *BXEndo88* was induced at two different temperatures (18 °C and 37 °C) to check the solubility of expressed endolysin. However, observation based on Figure 4.6 showed that induction using both temperateures were infavourable which resulted in the BXEndo88 was being expressed mostly in the pellets. On top of that, concentration of IPTG was also thought to indirectly affect the formation of inclusion body by increasing the expression rate (Silaban et al., 2018). Figure 4.7 showed different concentrations of IPTG used during induction. No distinct differences were observed as all proteins were expressed in the pellet fraction instead of the supernatant fraction.

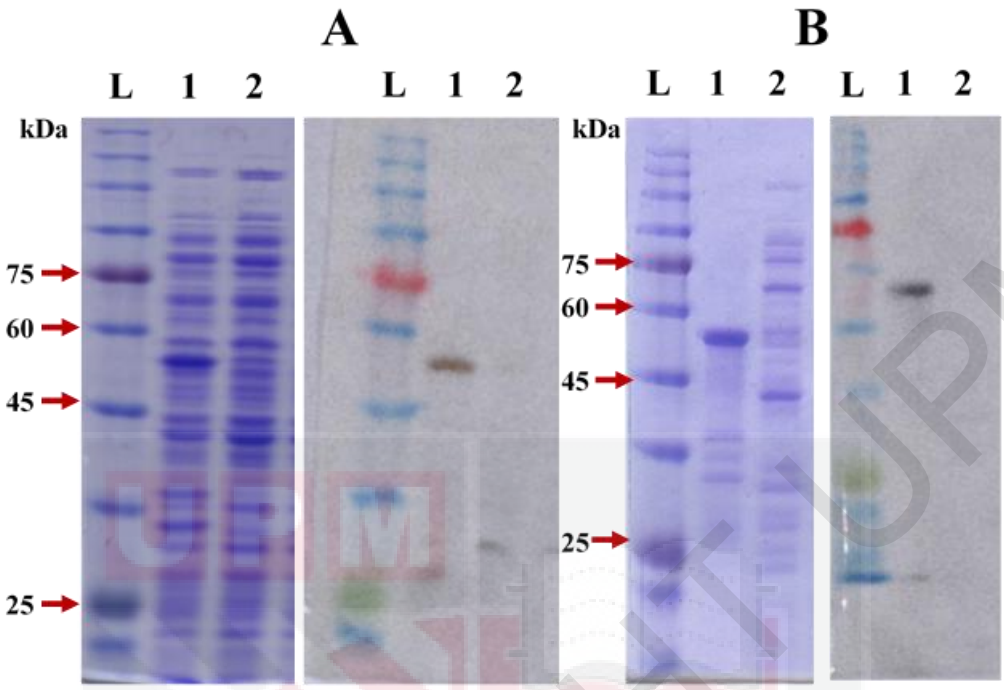


Figure 4.6: Expression of BXEndo88 analysed by SDS-PAGE (left panel) and Western blot (right panel). Red arrow: BXEndo88 was expressed in pellets (approximately 59 kDa) at both temperature (18 °C and 37 °C). Lane L: 5 – 245 kDa protein ladder; Lane 1: Pellet fraction; Lane 2: Soluble fraction. **(A)**: Expression at 18 °C; **(B)**: Expression at 37 °C. **(C):** 5 – 245 kDa protein ladder (1st Base, Malaysia).

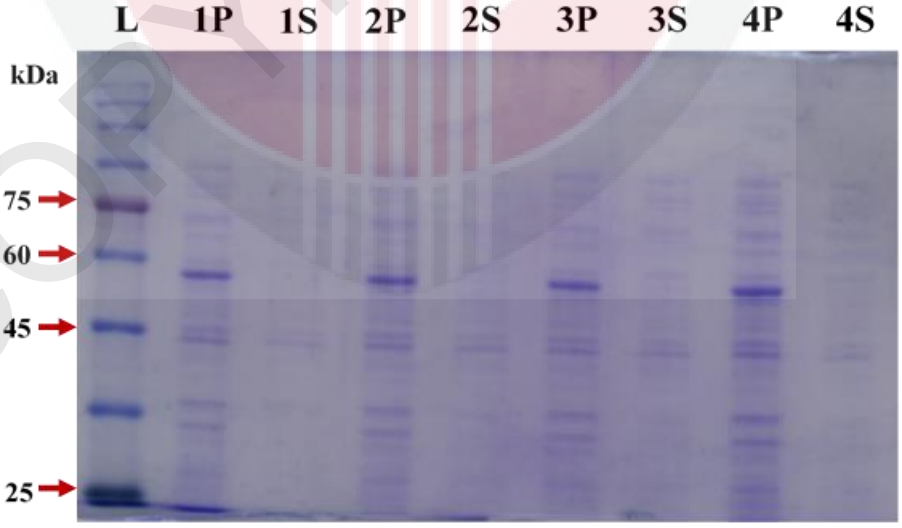


Figure 4.7: Effect of IPTG concentration during protein induction. No distinct differenced can be observed under different concentration of IPTG during induction. Red arrow: BXEndo88 approximately 59 kDa. P: Pellets fraction; S: Soluble fraction; Lane L: Ladder; Lane 1: 0.05 mM IPTG; Lane 2: 0.1 mM IPTG; Lane 3: 0.5 mM IPTG; Lane 4: 1 M IPTG.

4.3.2.2 Purification of BXEndo88 (Inclusion body) under denaturing condition.

Despite unsuccessful attempts on obtaining soluble protein during induction, the inclusion bodies were collected and subjected to purification under denaturing conditions (section 4.2.2.1). This process involved the uses of denaturing agent (8M Urea was used in this study) to solubilize the protein prior to his-tag purification, followed by washing step with decreasing amount of urea in the buffer to achieve protein tertiary structure re-folding, and elution with buffer without the denaturing agent (Pesarrodona et al., 2022). As depicted in Figure 4.8, BXEndo88 was successfully solubilised (Lane 1) and purified (Lane 7). The eluted protein fraction was subjected to plate lysis well-diffusion assay (section 4.3.3.3) to evaluate the functionality of the protein. However, as shown in Figure 4.9, no halo zones were observed indicating that the BXEndo88 was not functional even after solubilisation and purification (Figure 4.9A). This could be due to misfolded proteins during refolding (Kathir et al., 2005). SDS with 0.2% concentration was used as positive control, as it is commonly included in bio-surfactant to kill *S. aureus* and disrupt its biofilm. Typically, concentrations ranging from 0.2 % to 5 % are included in these formulations (Bezrodnykh et al., 2021; Díaz De Rienzo et al., 2016).

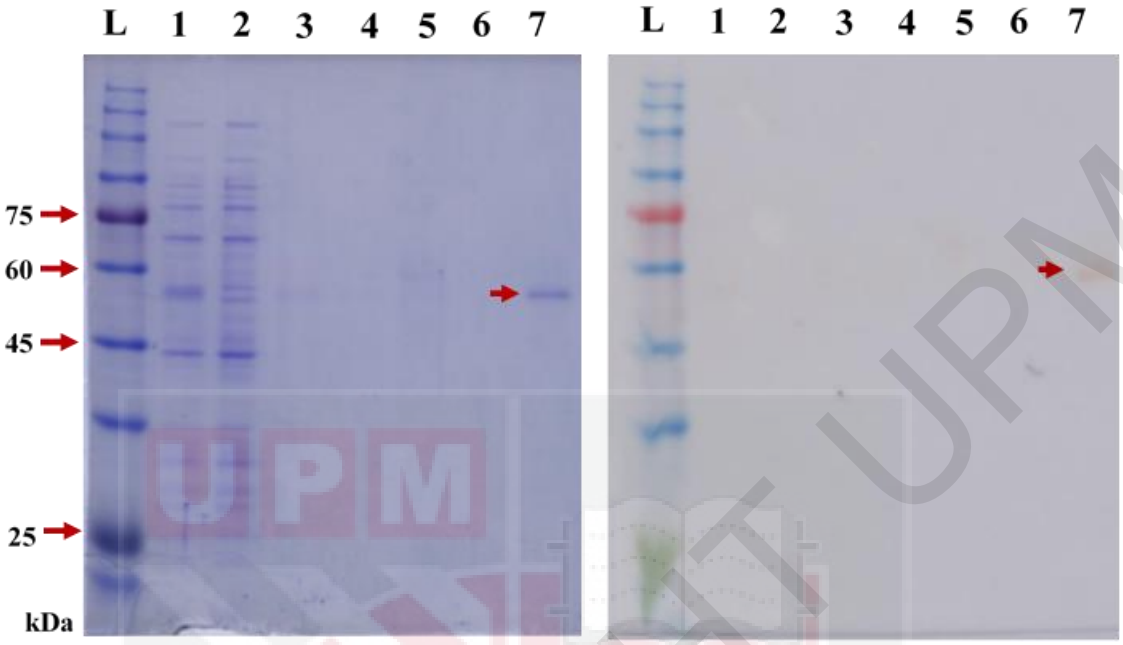


Figure 4.8: Purification of BXEndo88 (Inclusion body) under denaturing condition. Red arrow: The protein was successfully solubilised and purified under denaturing condition. Lane L: 5 – 245 kDa ladder; Lane 1: solubilised BXEndo88 treated with 8M urea; Lane 2: negative control (expression of empty plasmid - pET28a); Lane 3: flow-through; Lane 4: wash 1; Lane 5: wash 2; Lane 6: wash 3; Lane 7: elution.

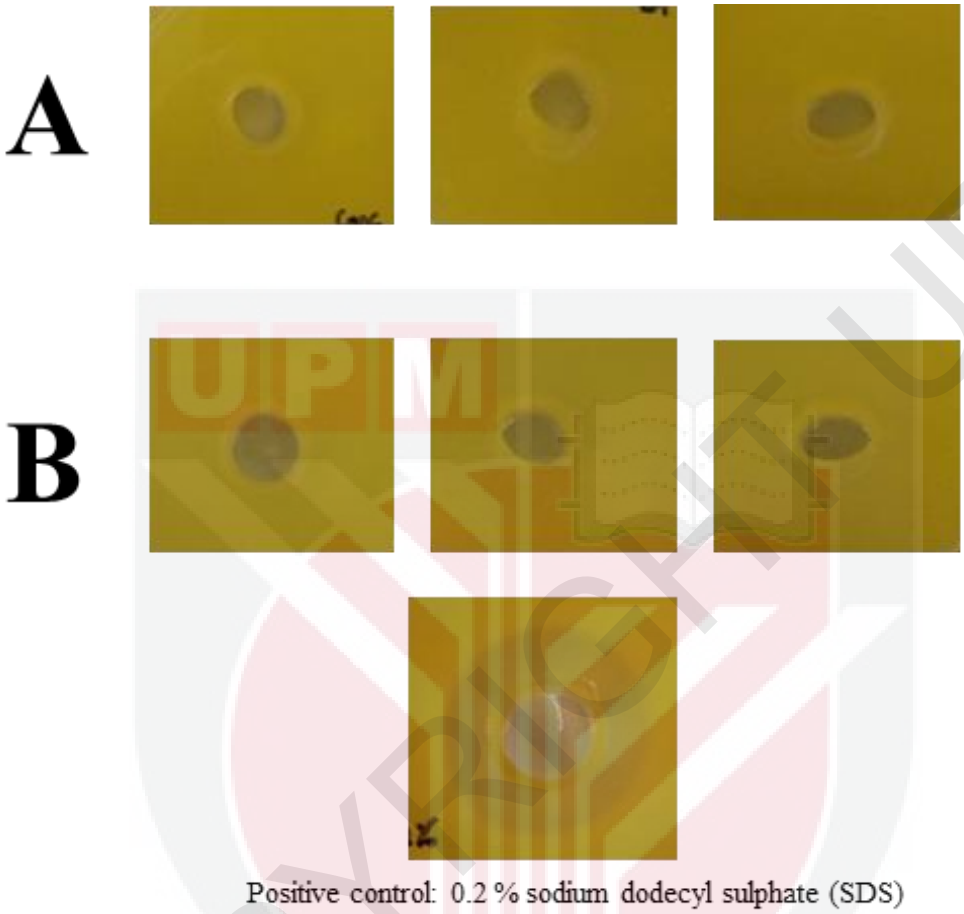


Figure 4.9: Plate lysis well-diffusion assay of BXEndo88 after refolding and purification. No halo zones were observed after treated with purified and refolded BXEndo88. **(A)**: Negative control (expression of empty plasmid - pET28a); **(B)**: Purified BXEndo88. The assay was done with three technical replicates. Positive control: 0.2 % sodium dodecyl sulphate (SDS).

4.3.2.3 Lysis buffer optimisation for protein stability

The choices of lysis buffer significantly impacts the solubility of proteins during extraction (Peach et al., 2015), specifically during the sonication step. Reactive oxygen species (ROS) generated during sonication could induce oxidative stress and damage the protein (Miljevic et al., 2014). Generally, the thiol side chains on two cysteine residues require oxidation to form a disulphide bond, which are is of the factor that is involved in stabilising the protein structure (Daniels et al., 2010; Haase-Pettingell et al., 2001; Nishino et al., 2022). The ROS generated during sonication could potentially affect the oxidation between the thiol side chain, resulted in mis-folding of the protein (Harris & Hansen, 2012; Miljevic et al., 2014; Stathopulos et al., 2004). There are six cysteines found in Endo88, which are potentially susceptible to the ROS. In light of that, different additives such as reducing agent, detergent and stabilizers which are added to buffers could aid in stabilising and releasing soluble protein from non-soluble component and preventing protein degradation (Leibly et al., 2012).

The reagents used in different buffers are stated in Table 4.7, section 4.2.2.3, all buffers were adjusted and maintained at pH 7.4 – 7.6, which were about two points below the isoelectric point of BXEndo88 – 9.02 (Predicted in ExPASy (Gasteiger et al., 2005)). In the buffer, the pH lower than the protein's isoelectric point ensures that the proper charge of the protein is maintained, facilitating the solubilisation of the protein molecules (Loell & Nanda, 2018; Tokmakov et al., 2021). The presence of NaCl was to increase the ionic strength and to stabilise the protein in the buffer (Leibly et al., 2012; Yong-Jin Mao et al., 2007). Beta-mercaptoethanol acts as a reducing agents to counter the oxidative stress (Son YJ et al., 2008) during sonication and

glycerol acts as a stabilizing agent in the buffer (Vagenende et al., 2009). Lastly, SDS solubilizes the protein (He & Ohnishi, 2017; Hjertén et al., 1988).

As depicted in Figure 4.10A, BXEndo88 was solubilized when Buffer B and D were applied, which indicated that the SDS alone was involved in the solubilization. The observations were expected as SDS is a detergent commonly used to solubilize proteins. SDS solubilize proteins by disrupting the hydrophobic interaction which leads to unfolding of the protein, depends on the condition (Hjertén et al., 1988; Winogradoff et al., 2020). Despite this, most of the protein remained in the pellet fraction as shown by the intense band in that fraction. To further evaluate the functionality of the solubilized protein, the soluble fraction from Buffer B was further purified and tested on plate lysis well-diffusion assay against *S. aureus* PS88.

As shown in Figure 4.10B, despite a faint halo zone was observed, it was ascertained that the lysis zone was caused by SDS carry over, since the negative control (Buffer B) also showed a clear halo zone. It was further validated by other reported study showing that Gram-positive bacteria are inhibited at 0.1 % SDS and above (Díaz De Rienzo et al., 2016; Flahaut et al., 1996). Similar to urea solubilization, this could be due to denaturation by SDS which resulted in mis-folding during sonication (Kathir et al., 2005). It suggested that the strong denaturation properties of SDS and urea that caused complete denaturation (Bennion & Daggett, 2003; Winogradoff et al., 2020) did not favor the re-folding into solubilized and functional proteins.

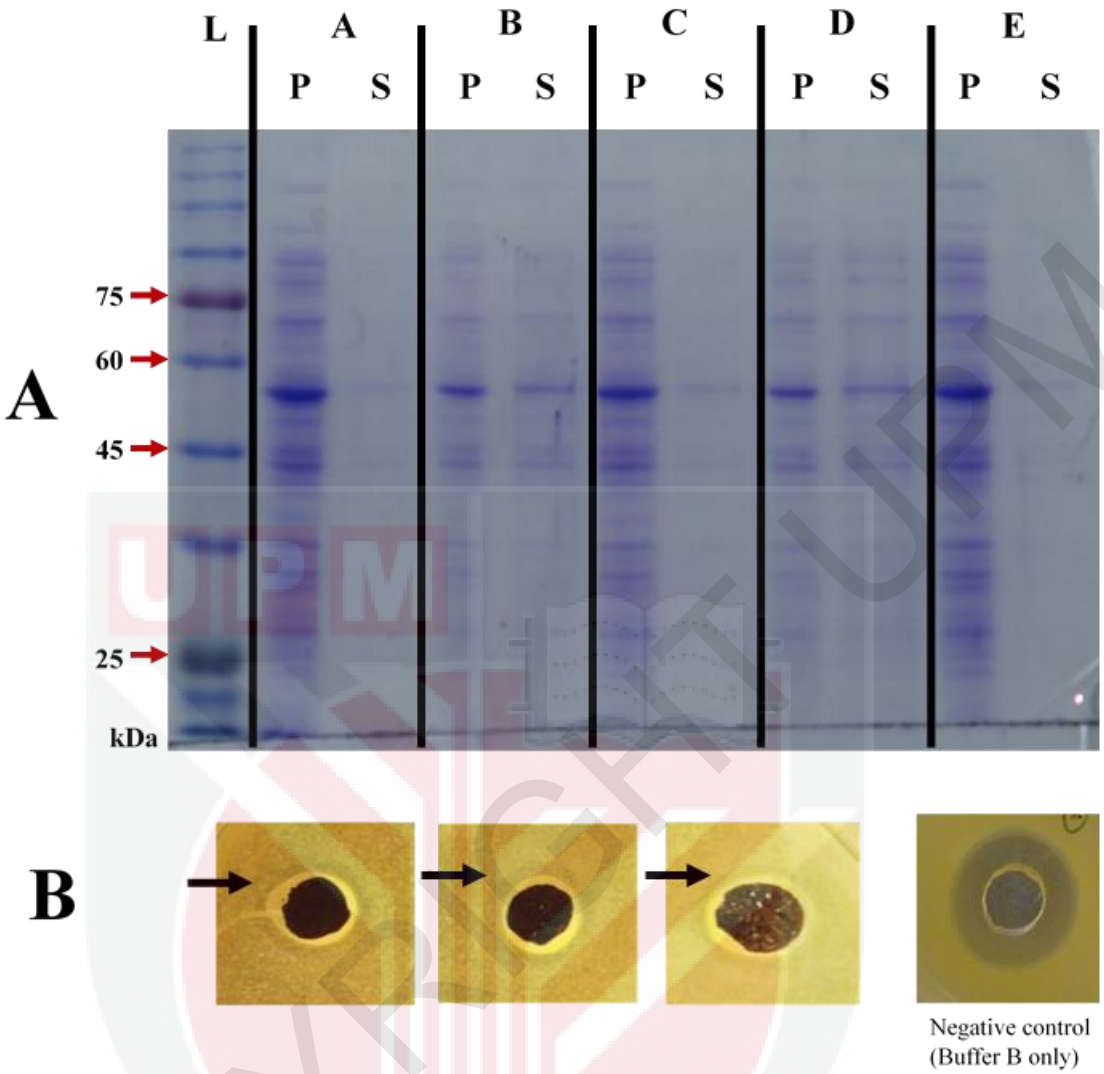


Figure 4.10: Optimisation of lysis buffer used during sonication. The solubility and functionality of BXEndo88 were evaluated using SDS-PAGE and PLA. P: Pellet fraction; S: Soluble fraction. Red arrow: BXEndo88 expressed at around ~ 59 kDa. **(A)**: SDS-PAGE showing the expression of BXEndo88 in pellet and soluble fraction after treatment of different lysis buffer. Lane L: 5 – 245 kDa ladder; Lane A: Buffer A; Lane B: Buffer B; Lane C: Buffer C; Lane D: Buffer D; Lane E: Buffer E. **(B)**: Plate lysis well diffusion assay tested with purified BXEndo88 in Buffer B. Black arrow: Observable faint halo zone. However, the negative control also showed a clear halo zone indicating that the SDS in the buffer was the cause of the halozone and not the solubilised endolysin.

4.3.2.4 Alternative approaches to obtain stable and solubilised endolysin.

The BL21(DE3) strain contains the *DE3* gene, which encodes T7 RNA polymerase, enabling high-level protein expression (Tabor, 1990). However, it does not specialize in facilitating disulphide bond formation between cysteine residues (Bhatwa et al., 2021), which are present in Endo88. Therefore, concurrent to buffer optimization, preliminary experiments were done on swapping the expression host to *E.coli* Origami 2(DE3) as an alternative expression host. This *E. coli* strain carries mutation in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) gene (Gordon et al., 2008; Hayat et al., 2018), which improve the formation of disulfide bond during protein expression (Hatahet et al., 2014; Xiong et al., 2005). This strains could facilitate the formation of disulfide bond between the 6 cysteine amino acid within Endo88. As depicted in Figure 4.11, the proteins were expressed in inclusion body (lane 1) and no observable band can be seen on the soluble fraction (lane 2). Therefore, attempts of obtaining soluble protein was not successful.

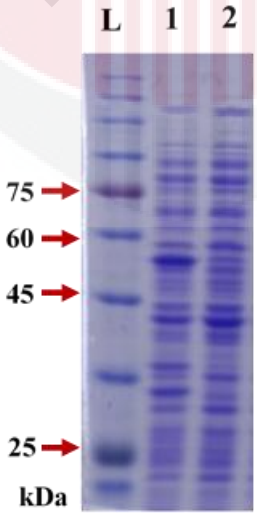


Figure 4.11: Expression of BXEndo88 in *E. coli* Origami 2(DE3). Band with size approximately 59 kDa was spotted in the pellet fraction, and no observable band can be observed in soluble fraction at the same size. Lane L: 5 – 245 kDa ladder; Lane 1: Pellet fraction; Lane 2: Soluble fraction.

4.3.3 Optimisation of protein expression and downstream processing of NXEndo88

Several studies have reported the successful expression of endolysins using the pET28a vector and cloning strategies similar to those employed for BXEndo88, with the proteins being expressed in the soluble fraction (Gerova et al., 2011; K. O. Lee et al., 2019; N. Lu et al., 2020). These proteins contained an N-terminal His-tag along with additional amino acids derived from the plasmid. However, unlike in BXEndo88, none resulted in proline tail at the C-terminus (Figure 4.2B). This difference appears to have a significant impact on protein solubility. Although proline is a small, uncharged amino acid, its unique pyrrolidine side chain can influence protein folding due to cis-trans isomerization and conformational rigidity (Cortes-Hernandez & Domínguez-Ramírez, 2017; S. C. Li et al., 1996; Osváth & Gruebele, 2003; Wedemeyer et al., 2002; Williamson, 1994; Zosel et al., 2018). Since all optimisation of BXEndo88 failed to yield soluble proteins, Endo88 was re-cloned using a different pair of primers with different set of restriction site (*Nco*I & *Xho*I), resulting in NXEndo88 (Figure 4.2B, section 4.3.1) which resulted in a frame shift that covered the proline tail to a C-terminal his-tag instead.

4.3.3.1 Induction of *NXEndo88* at different temperatures and concentrations of IPTG

NXEndo88 was induced at both cold temperature (18 °C) and warm temperature (37 °C) using different concentrations of IPTG (0.1 mM, 0.5 mM & 1 mM) at 18 °C (Figure 4.12), while only 0.5 mM IPTG was used at 37 °C. Higher IPTG concentration (e.g. 1 mM) at 37 °C has been correlated with inclusion body formation (Mühlmann et al., 2017; Soleymani & Mostafaie, 2019). Conversely, using too low IPTG

concentration might result in excessive cell growth with reduced protein expression. Similar to BXEndo88, most of the protein were being expressed in pellet fraction when induced with 37 °C (Lane 4, Figure 4.12). However, the protein was expressed in soluble fraction at 18 °C (Lane 1 – 3, Figure 4.12).

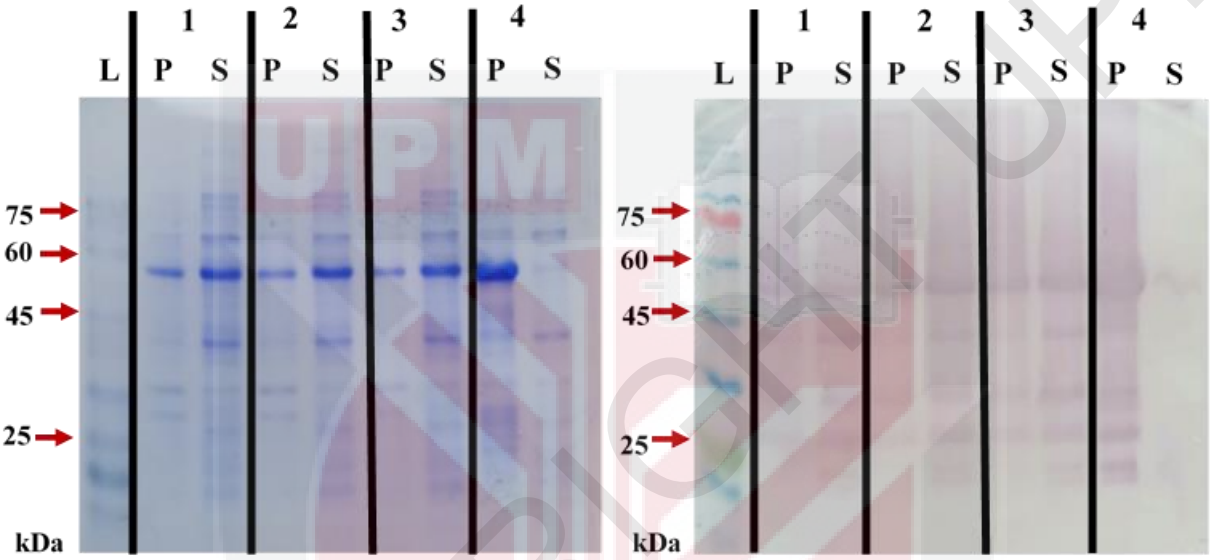


Figure 4.12: Optimisation of NXEndo88 expression in BL21(DE3) at different temperatures and IPTG concentrations. *NXEndo88* was induced in cold temperature (18 °C) and warm temperature (37 °C) with different concentration of IPTG (0.1 mM, 0.5 mM & 1 mM). Band with size approximately 55 kDa was spotted, indicated the presence of target protein. P: Pellets fraction, S: Soluble fraction. Lane L: 5 – 245 kDa ladder; Lane 1: 0.1 mM, 18 °C, overnight; Lane 2: 0.5 mM, 18 °C, overnight; Lane 3: 1.0 mM, 18 °C, overnight; Lane 4: 0.5 mM, 37 °C, 4 hours.

4.3.3.2 Purification of NXEndo88 under native conditions and screening of its functionality

Subsequently, the soluble fractions were pooled and subjected to purification under native condition (Section 4.2.2.2). As depicted in figure 4.13A, NXEndo88 was successfully purified and yielded good amounts as indicated by the thick band on lane 5 (eluted protein) and lane 6 (Buffer exchanged and concentrated protein) (Figure 4.13A). Both fractions from lane 5 and lane 6 showed clear halo zone formation on plate lysis well diffusion assay (Figure 4.13B). Additionally, although the soluble fraction (lane 1) appears to exhibit a band intensity similar to that of a purified fraction, the soluble fraction showed only faint halo zone formation in plate lysis assay. This is anticipated as 100 µg of soluble fraction contains crude proteins which include other cellular components, thus the concentration of endolysin in the soluble fraction is lower than using purified proteins.The Elution buffer (50 mM Tris-HCl, 150 mM NaCl, and 250 mM imidazole) without any other stabilizing additive, act as a negative control show no halo zone indicated that the lysis effect was from endolysin itself.

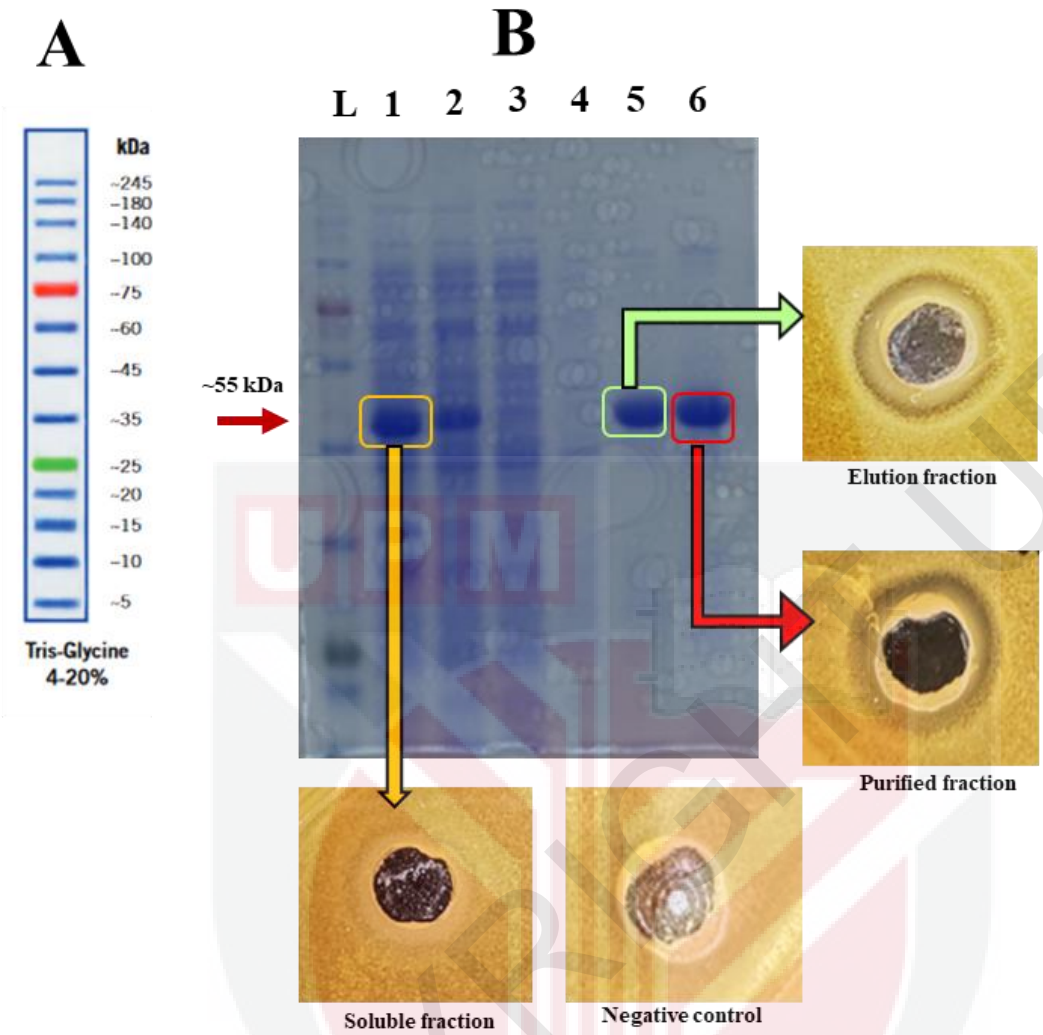


Figure 4.13: Purification and functional screening of different fractions of NXEndo88. Protein induction at 18 °C followed by purification under native conditions, the protein fractions were then subjected to plate lysis assay to evaluate the expression of soluble and functionality of NXEndo88. Bands with size approximately 55 kDa were observed, indicated the presence of target protein. **(A)**: 5 – 245 kDa Ladder. **(B)**: SDS-PAGE showing each protein fraction during purification under native condition. Lane L: 5 – 245 kDa Ladder, Lane 1: Soluble fraction; Lane 2: Pellet fraction; Lane 3: Flow through; Lane 4: Wash; Lane 5: Elution (with 250 mM imidazole); Lane 6: Purified and concentrated NXEndo88 (without imidazole). Elution buffer was used as negative control.

4.3.3.3 Characterisation of protein functionality by TRA and PLA

To quantitatively evaluate the inhibitory effect of NXEndo88 against *S. aureus* PS88, turbidity reduction assay was performed. As depicted in figure 4.14A, the fractions containing soluble NXEndo88 were able to lyse *S. aureus* PS88, as shown by the decrease in OD within 30 minutes at different rates. Remarkably, the elution fraction (purified NXEndo88 with 250 mM imizadole) reduced the OD slightly faster than the concentrated fraction (purified NXEndo88 without 250 mM imidazole as buffer has been exchanged with concentrator). This indicated that 250 mM imidazole did not halt the NXEndo88 activity, despite previous reports that imidazole affected the protein function (Chagas et al., 2023; Hamilton et al., 2003; J. J. Lee et al., 2007). The slight decrease of OD in the negative control (live cells only) and elution buffer is probably due to the nutrient depletion in the 96-well plate throughout the experiment, which resulted in natural cell death. NXEndo88 also showed a dose dependent manner, as shown in figure 4.14B. The highest concentration (10 µg) of NXEndo88 reached OD 0.1 within 20 minutes, while the lowest (0.5 µg) took almost 50 minutes to reach OD 0.1. The slight increase of OD during the first 5 minutes might be due to the low concentration of endolysin which was unable to overcome the growth of the *S. aureus* PS88 at the beginning, but become effective as time elapsed.

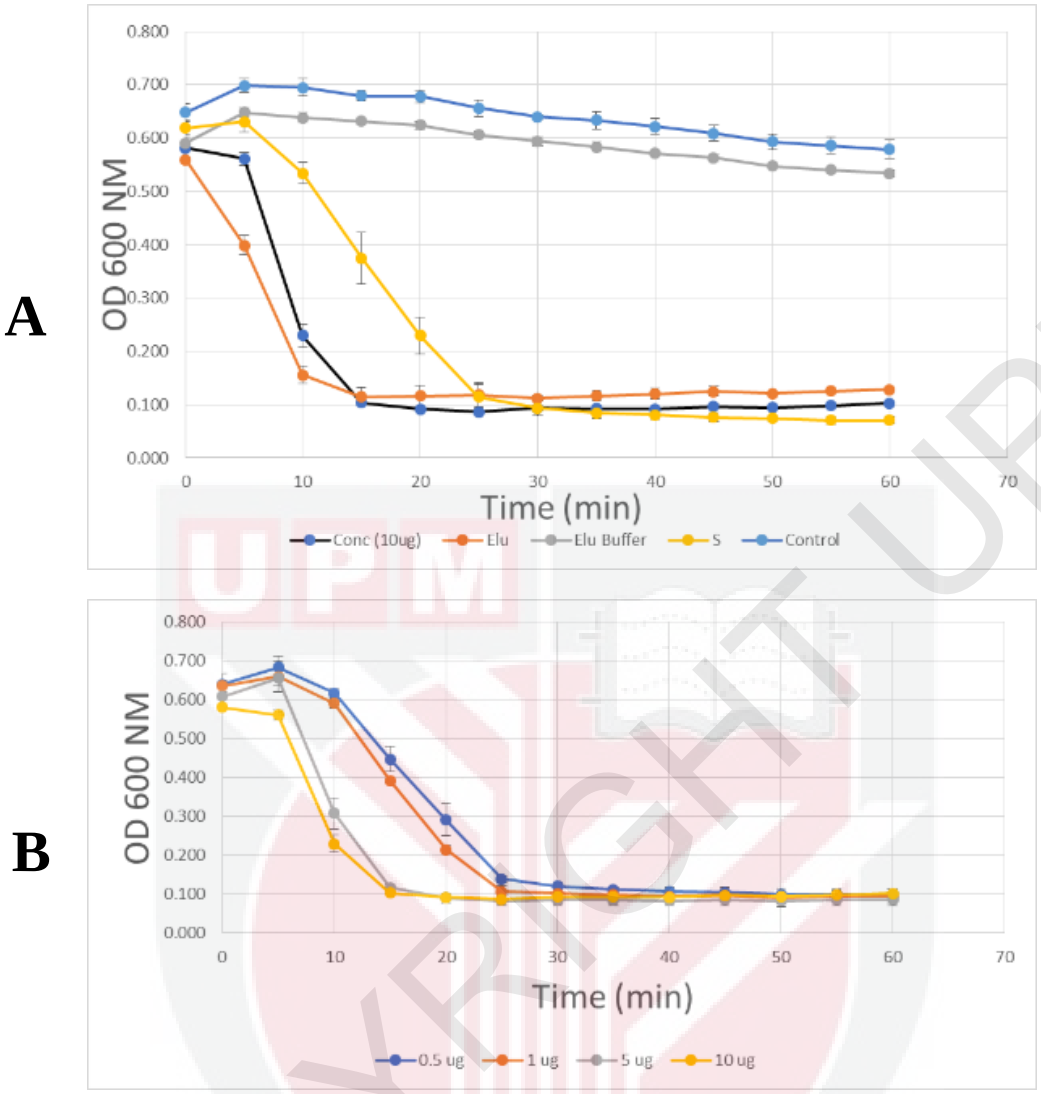


Figure 4.14: Turbidity reduction of NXEndo88. (A): Various fractions of NXEndo88 expressed at 18 °C. Conc: 10 µg purified endolysin (without imidazole); Elu: 10 µg NXEndo88 in elution fraction with 250 mM imidazole; S: 10 µg of soluble crude fraction; Elu buffer alone: Elution buffer (Tris-HCl, 150 mM NaCl, 250 mM imidazole); Control: untreated *S. aureus* PS88 as negative control. **(B)**: Turbidity reduction assay with different concentrations of NXEndo88.

4.3.4 Biophysical characterization of NXEndo88 (Denoted as Endo88).

The biophysical characterization of Endo88 involves studying the effects of temperature, NaCl concentration, and pH on enzyme activity and stability. These factors play crucial roles in determining the optimal conditions for enzyme activity.

4.3.4.1 Effect of NaCl on lytic activity

The effect of NaCl on enzyme activity can be complex and vary depending on enzyme and condition (Braham et al., 2021; Falk, 1918). This also apply on endolysin, which studies have suggested that the presence of NaCl can stabilize endolysins by affecting the ionic environment around the enzyme (Briers et al., 2014; Yong-Jin Mao et al., 2007), but diverse impacts on the lytic activity of endolysin has been reported (S. C. Becker et al., 2008; Etobayeva et al., 2018; Linden et al., 2015; Son et al., 2012). Different concentration (0 mM, 50 mM, 100 mM, 150 mM, 300mM, 500 mM)) of NaCl were included into the reaction buffer containing endolysin (Endo88) against *S. aureus* PS88. The percentage bacteria OD drop was calculated as mentioned in section 4.2.3.2. The highest OD drop (97.7 %) was observed on buffer without NaCl, while the lowest OD drop (6.2 %) was observed on 500 mM (Figure 4.15A). This indicated that the increase in NaCl concentration was inversely proportional to the effectiveness of Endo88 against *S. aureus* PS88. Similar results were observed with other endolysin, which prefer no salt/ low salt concentration in the buffer (Etobayeva et al., 2018; Filatova et al., 2015; Linden et al., 2015; Son et al., 2012). Additionally, Figure 4.15B showed that the decrease in OD was not caused by salt induced osmotic homeostasis as reports have shown that *S. aureus* acquire salt-tolerance in high salt condition (Feng et al., 2022; Parfentjev & Catelli, 1964; Trognon et al., 2023; Vaish et al., 2018). Thus,

this shows that the drop of OD in Figure 4.15A was caused by the endolysin and not due to the high salt concentration.

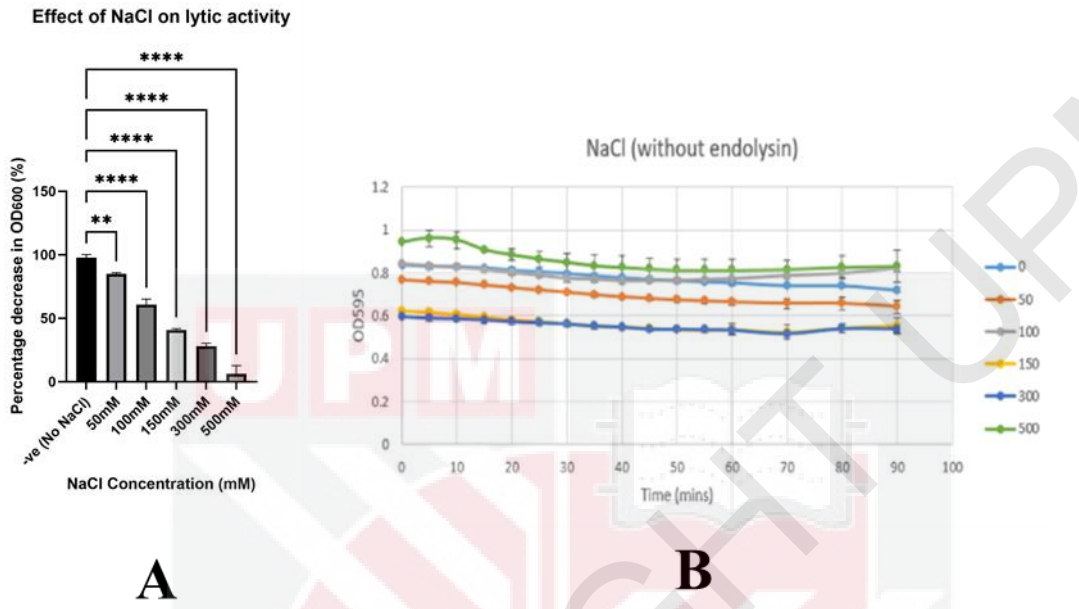


Figure 4.15: Effect of NaCl concentration on lytic activity. (A): Percentage of *S. aureus* PS88 OD dropped within 60 minutes after treated with different concentration of NaCl. Each column represents the mean of triplicate experiments, and error bars indicate the standard deviation. The asterisks indicate significant differences, "*": P<0.05; "**": P<0.01; "***": P<0.001; "****": P<0.0001. **(B)**: Negative control, *S. aureus* PS88 exposed to different concentration of NaCl without endolysin.

4.3.4.2 Effect of pH on lytic activity

The lytic activity of Endo88 was then tested in sodium acetate buffer and Tris-buffer with pH range from pH 4.0 – pH 11.0 to determine the optimum lytic conditions. Endo88 was active between pH 6.0 and pH 11.0, indicating a high alkaline tolerance, which was also found in other endolysin (range pH 8 – pH 10.5) (Y. Lu et al., 2021; Oliveira et al., 2014; Son et al., 2012) In contrast, many other endolysin exhibit different active range between pH 6.0 – pH 9.0 (Filatova et al., 2015; Obeso et al., 2008; H. Yang et al., 2017). Conversely, Endo88 lytic activity decrease significantly

at pH 5 and lower pH (Figure 4.16A). Since the active range was between pH 6 to pH 11, with no significant differences, pH 7.5 was selected for subsequent analysis since it is the most common pH used in other endolysin studies (S. C. Becker et al., 2008; Gu et al., 2014; Melo et al., 2018; Navarre et al., 1999; Schmelcher, Korobova, et al., 2012). Additionally, Figure 4.16B showed that the decrease in OD were not caused by the extreme pH as previous reports have shown that *S. aureus* can to adapt to extreme pH (Efthimiou et al., 2019; Vaish et al., 2018, 2019).

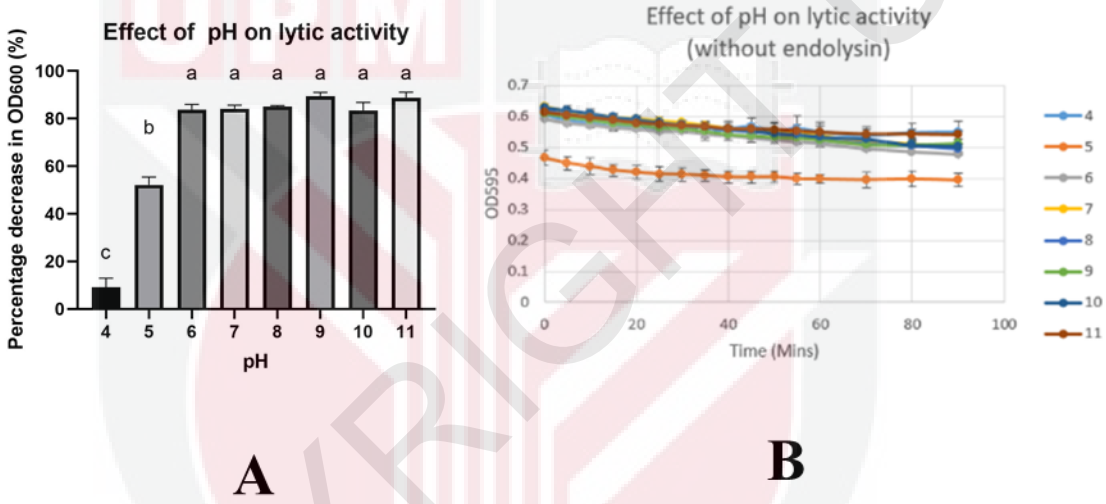


Figure 4.16: Effect of pH on lytic activity. (A): Percentage of *S. aureus* PS88 OD dropped within 60 minutes after treated with different pH. Each column represents the mean of triplicate experiments, and error bars indicate the standard deviation. Different alphabet (a, b & c) means they are significant to each other, same alphabet means not significant ($P<0.05$). **(B)**: Negative control, *S. aureus* PS88 exposed to different pH without endolysin.

4.3.4.3 Effect of temperature on lytic activity

The thermodynamic stability of Endo88 was investigated by exposing the endolysin to different temperatures prior to turbidity reduction assay. Endo88 exhibited high lytic activity (>90 %) after incubating at temperatures ranging from -20 °C to 37 °C (Figure 4.17). The lytic activity begins to abolish at 45 °C and stopped at 55 °C onward. This observation is consistent with other phage endolysin (Y. Chang et al., 2017; Linden et al., 2015; Obeso et al., 2008). Therefore, a temperature of 37 °C was used for subsequent analysis as it corresponds to the optimal growth conditions of many bacterial pathogens, including *S. aureus*. Since endolysin activity in this study was assessed against live bacterial cells, this temperature provides conditions that mimic natural bacterial environments.

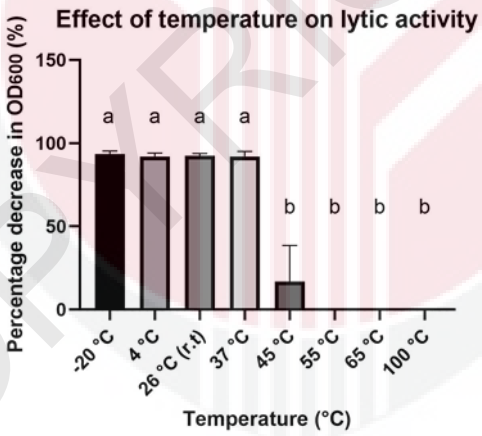


Figure 4.17: Effect of temperature on Endo88 prior turbidity reduction assay. Each column represents the mean of triplicate experiments, and error bars indicate the standard deviation. Different alphabet (a, b & c) means they are significant to each other, same alphabet means not significant ($P<0.05$).

4.3.5 Comparison of mutant genes (*M, A1* & *A2*) lytic host range with Endo88

4.3.5.1 Construction of mutant genes

Gradient PCR was employed to determine the optimal annealing temperature (55°C, 60°C and 65°C) for the primer sets (Table 4.10, section 4.2.5.1) amplifying each region's fragments. The expected bands were visualized through agarose gel electrophoresis, as depicted in Figure 4.18. The primer set for mutant M generated two region fragments with the expected band size (Region 1: 1370 bp & Region 2: 133 bp), as well as mutant A1 (Region 1: 1260 bp & Region 2: 237 bp), and A2 (Region 1: 1332 bp & Region 2: 183 bp). After gel purification, all regions were annealed together and amplified using NXEndo88 primers, except for Mutant M, for which a different primers set for pNZ8048 (Table 4.10) (Figure 4.19) was used. Notably at this stage, Mutant M was still amplified using the primer set designed for cloning into pNZ8048 plasmid vector. This was attempted to express in *Lactococcus lactis* NZ9000. However, the attempt was unsuccessful and was switched back to *E. coli* BL21(DE3) expression system, which uses the same primers (NXEndo88, Table 4.2). As depicted in Figure 4.20, the *mutant M* gene incorporated with *Nco*I and *Xho*I was successfully amplified.

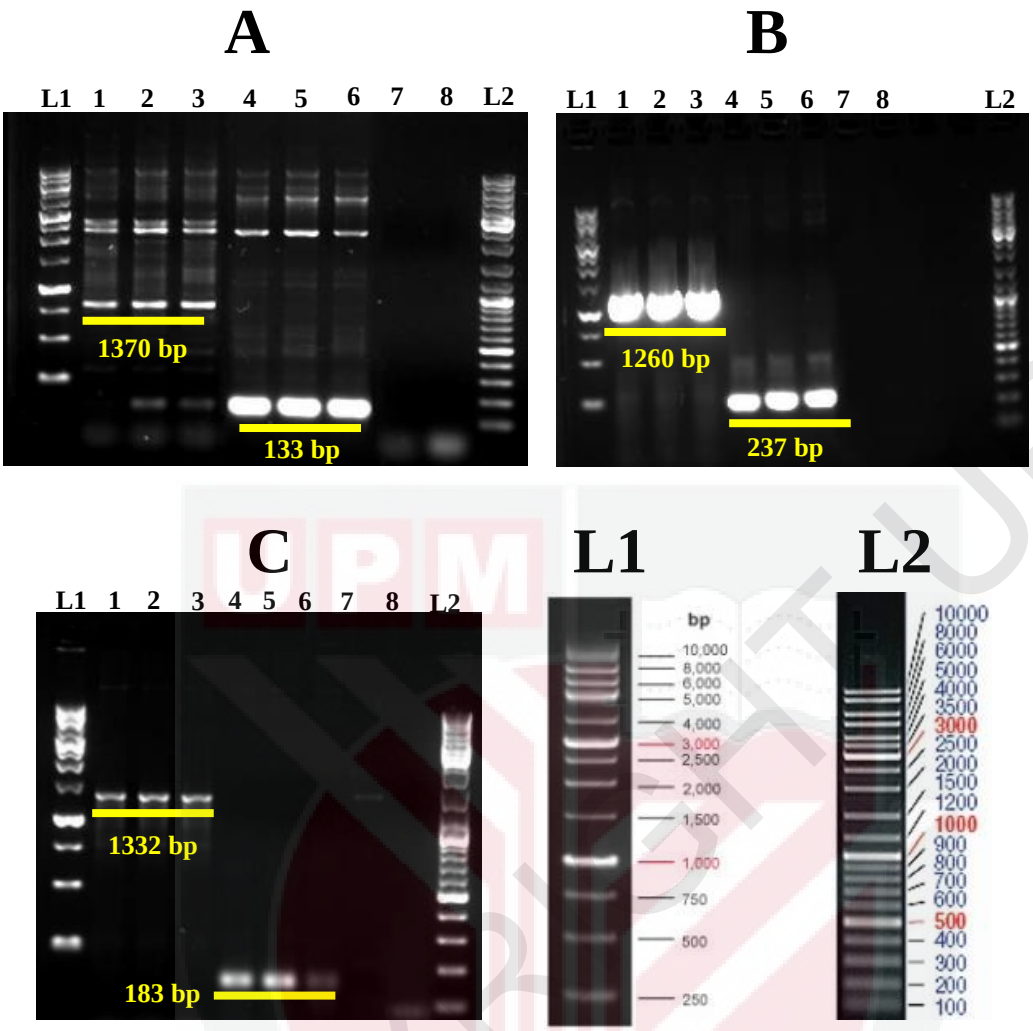


Figure 4.18: Gradient PCR to determine optimum annealing temperature for each primer sets. (A): Mutant M; **(B):** Mutant A1; **(C):** Mutant A2. Lane L1: 1 kb DNA ladder (1st Base, Malaysia); Lane L2: 100 bp ladder (GeneRuler, Thermo); Lane 1 to 3: PCR product of Region 1 at annealing temperature 55°C, 60°C & 65°C; Lane 4 to 6: PCR product of Region 2 at annealing temperature 55°C, 60°C & 65°C; Lane 7: Negative control of Region 1 PCR master mixture; Lane 8: Negative control of Region 2 PCR master mixture.

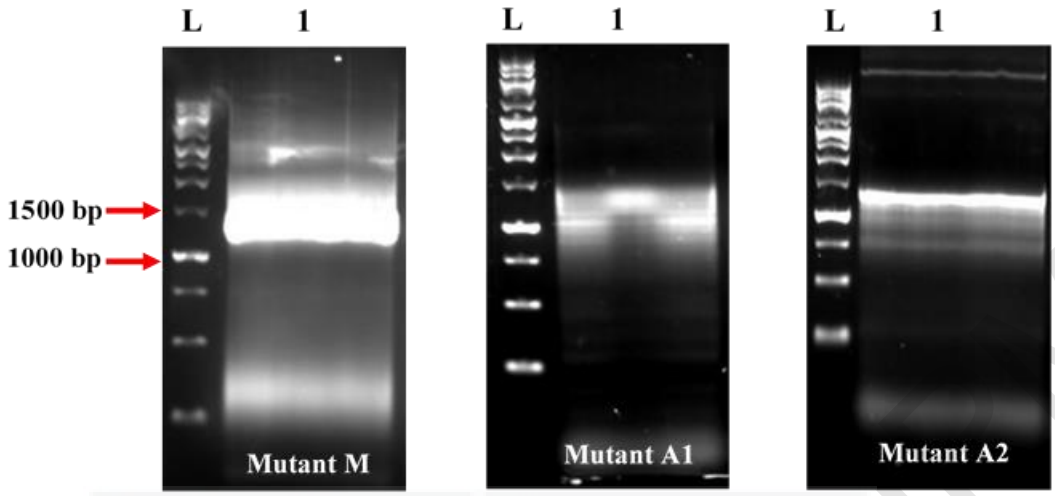


Figure 4.19: Conventional PCR to amplify the annealed region. Region 1 and region 2 of the mutants were annealed and then subjected to conventional PCR (Ta: 60 °C, section 4.2.1.5) to amplify the target mutant genes. Lane L: 1 kb DNA ladder (1st Base, Malaysia); Lane 1: Expected PCR product (~1466 bp).

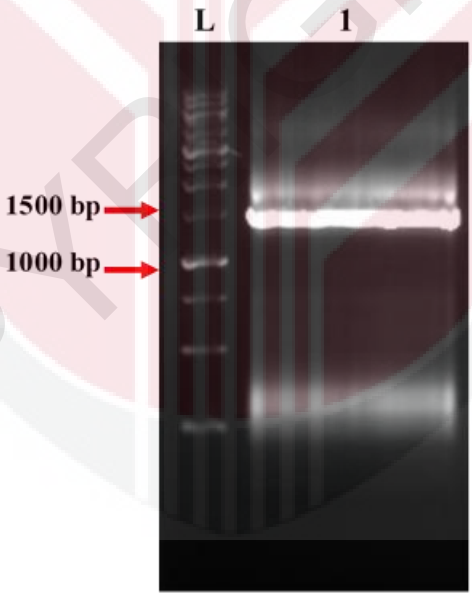


Figure 4.20: PCR product (incorporate with *Nco*I & *Xho*I) of mutant M amplified from pNZ8048 (*mutant M*). Lane L: 1 kb DNA ladder (1st Base, Malaysia); Lane 1: PCR product of Mutant M incorporate with *Nco*I & *Xho*I (~1466 bp).

The mutant genes (*M, A1, A2*) were subsequently cloned and transformed into *E.coli* BL21(DE3) using the same methodology as cloning of *NXEndo88* (Section 4.2.1.3). As depicted in Figure 4.21, colonies harbouring the expected recombinant plasmid displayed band size slightly above 1.5 kb band size on the ladder, while the empty plasmid showed band approximately 200 bp in size. Subsequent verification of the mutation was conducted through Sanger sequencing (Appendix A).

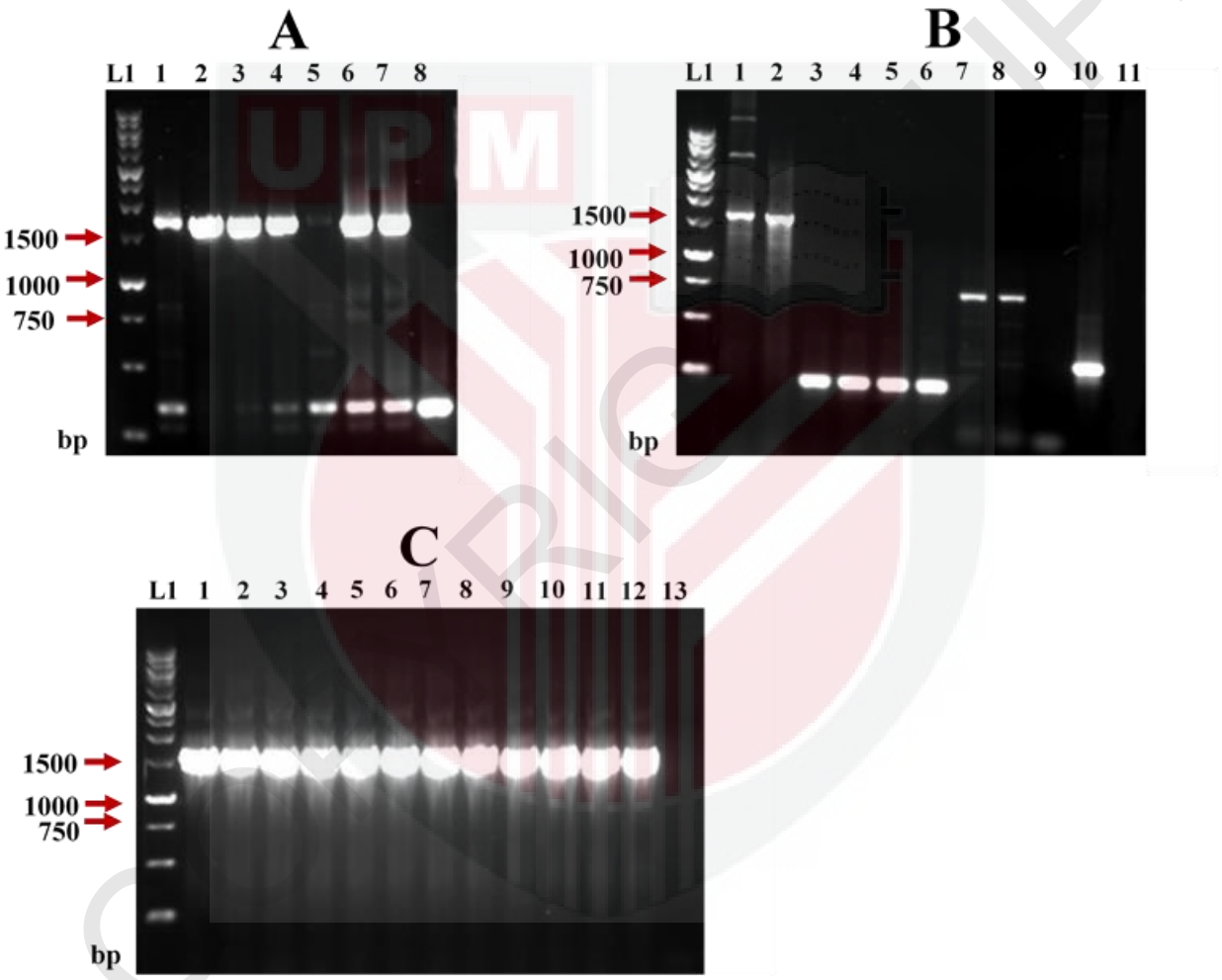


Figure 4.21: Colony PCR screening of recombinant plasmid (pET28a) harbouring mutant gene (*M, A1 & A2*). Lane L1: 1 kb DNA ladder (1st Base, Malaysia); Lane L2: 100 bp ladder (GeneRuler, Thermo). Red box: Expected band size (above 1.5 kb) indicated the insert was successfully inserted into pET28a. **(A) Colony PCR screening of pET28a (*M*)** Lane 1 – 7: colony screening PCR; Lane 8: Extracted and purified empty plasmid using the same primer sets (Negative control). **(B) Colony PCR screening of pET28a (*A1*):** Lane 1 – 8: colony screening PCR; Lane 10: Negative control. **(C) Colony PCR screening of pET28a (*A2*):** Lane 1 – 12: colony screening PCR.

4.3.5.2 Expression and purification of mutant protein (A1, A2 and M)

Mutants A2 and M were successfully induced and purified (Section 4.2.2.2) under the same condition as NXEndo88 (Figure 4.22), but mutant A1 was not expressed well (Figure 4.23). Most of the proteins were expressed as inclusion body even after optimization, as shown in (Figure 4.23). Therefore, A1 was excluded from subsequent analysis.

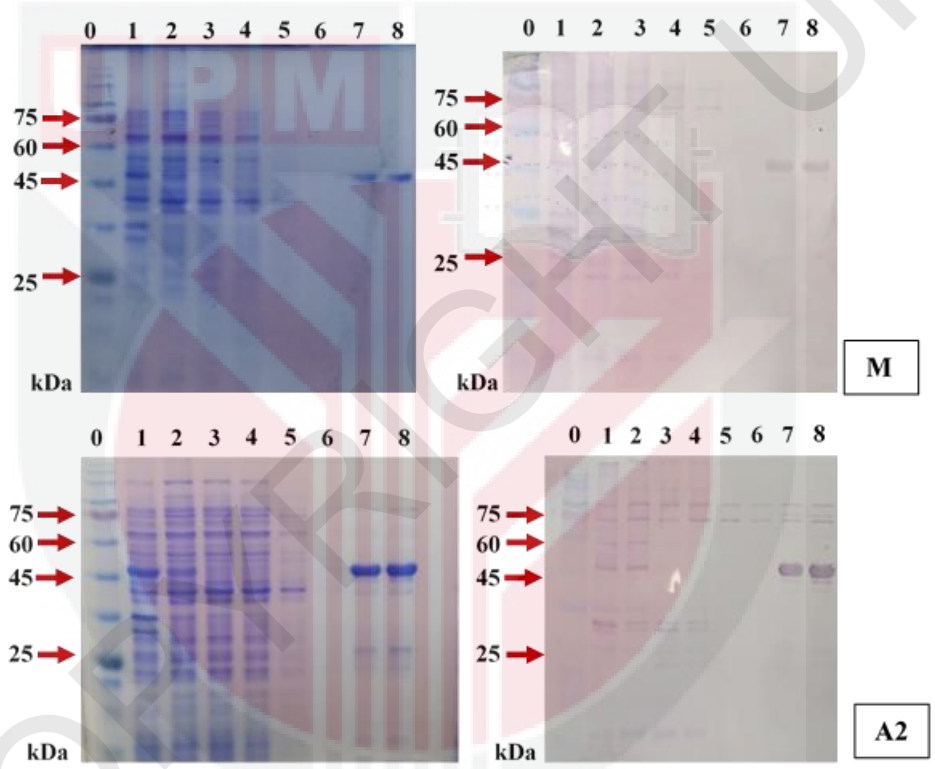


Figure 4.22: SDS-PAGE gel and Western Blot of mutants M and A2 purified under native condition (~55.4 kDa). Lane 0: 5 – 245 kDa protein ladder (1st Base, Malaysia); Lane 1: Pellets; Lane 2: Soluble fraction; Lane 3: Flow through; Lane 4 - 6: Wash 1 – 3; Lane 7: Elution; Lane 8: Concentrated and buffer exchanged protein.

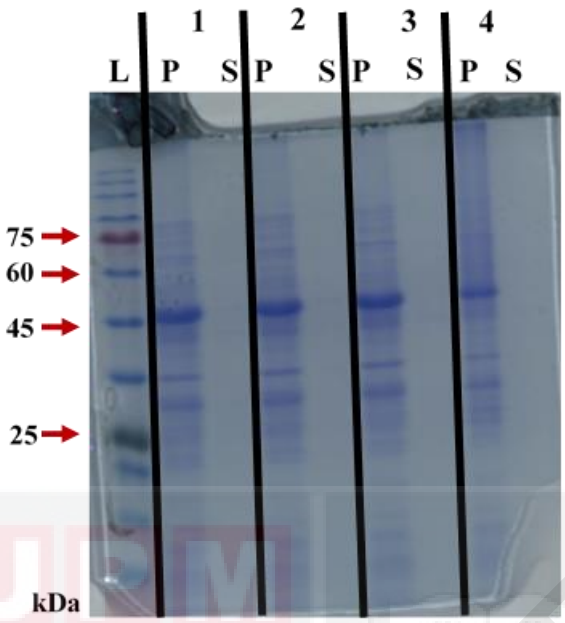


Figure 4.23: Mutant A1 expression and purification under native condition. Expression of Mutant A1 in cold temperature (18 °C) and warm temperature (37 °C) with different concentration of IPTG (0.1 mM, 0.5 mM & 1 mM). No bands were observed on soluble fraction even after optimisation. P: Pellets fraction (~55.4 kDa), S: Soluble fraction. Lane L: 5 – 245 kDa ladder; Lane 1: 0.1 mM, 18 °C, overnight; Lane 2: 0.5 mM, 18 °C, overnight; Lane 3: 1.0 mM, 18 °C, overnight; Lane 4: 0.5 mM, 37 °C, 4 hours.

4.3.5.3 Turbidity reduction assay (TRA)

The wild-type endolysin (Endo88) and the mutants (M and A2) were tested against different genera of bacteria (Figure 4.24). Endo88, A2 and M were able to lyse *S. aureus* PS88, *S. epidermidis, S. hominis and E. faecalis* with different degrees of lytic activities, but were incapable of lysing *S. pyogenes, S. agalactiae, B. subtilis, B. cereus, M. luteus* and *L. plantarum*.

Despite sharing a similar lytic host range, the mutants displayed lower lytic activity compared to Endo88. Specifically, A2 displayed significantly lower lytic activity, particularly in *S. aureus* PS88 (15.5% lesser), *S. epidermidis* (8% lesser) and *E. faecalis* (13.3% lesser). Mutant M exhibited the lowest lytic activity compared to Endo88 and A2, particularly in *S. aureus* PS88 (24.3% lesser), *S. epidermidis* (18.1%

lesser) and *E. faecalis* (26.8% lesser). However, for *S. hominis*, lytic activity was similar for wild-type, A2 and M with no significant differences. These observations indicate that the mutation on the SH3 domain influenced the lytic activity of the endolysin, despite the SH3 CBD having no lytic activity. This could be attributed to the folding of the endolysin despite the EAD and CBD regions being distinctively modular. Additionally, the 24.4 % decrease in OD observed in the negative control for *S. epidermidis* may be due to unfavourable buffer conditions and the absence of nutrients, which could have affected bacterial viability despite no treatment.

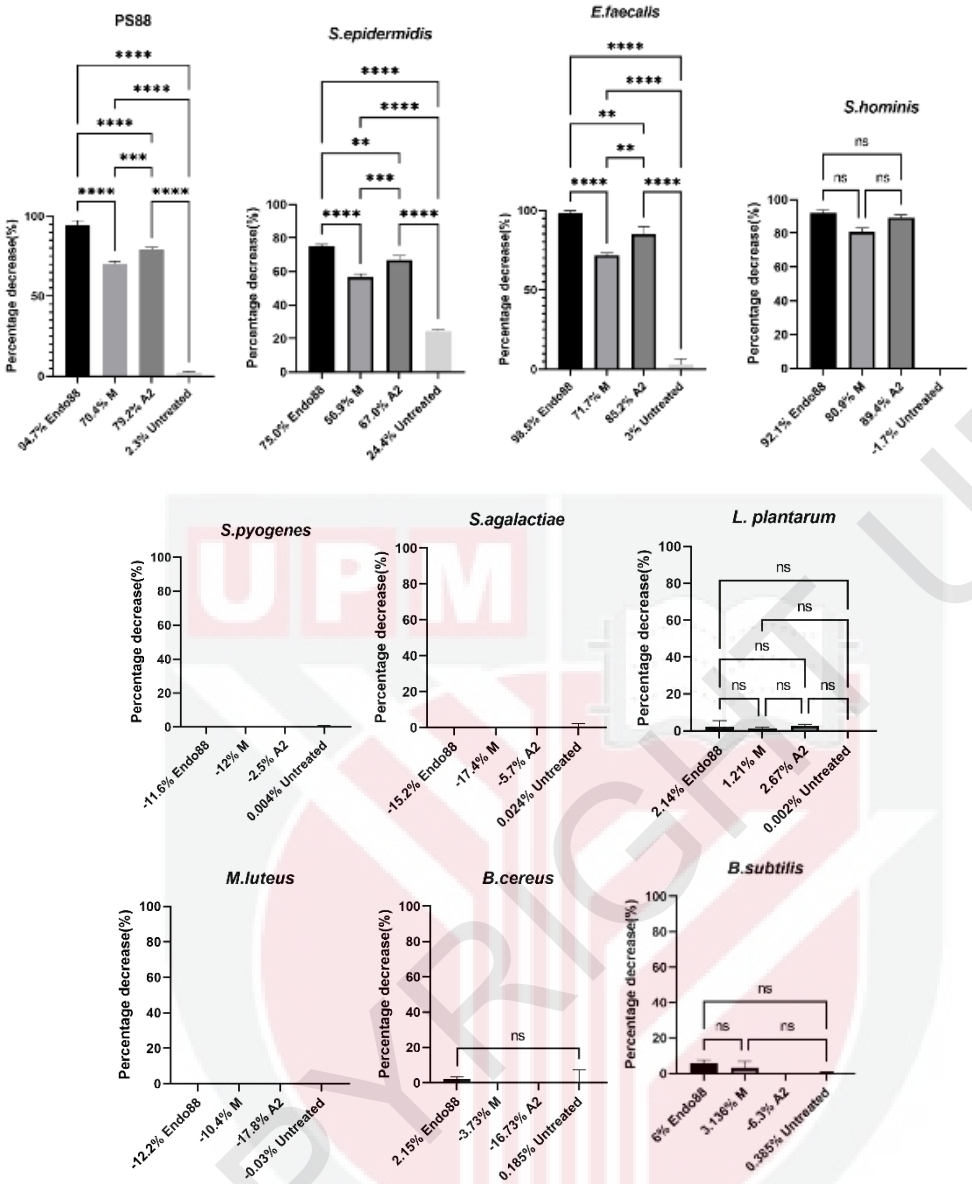


Figure 4.24: Turbidity reduction assay of wildtype Endo88 and mutants against various bacteria. The Endo88 and mutants (M and A2) were able to lysed *S. aureus* PS88, *S. epidermidis, S. hominis*, and *E. faecalis* with different degree of lytic activity. Untreated: Bacteria cell without endolysin treatment as negative control; "*" indicates significance with $p < 0.05$; "**": $p < 0.01$; "***": $p < 0.001$; $p < 0.0001$ and ns: not significant.

Both M and A2 were designed and anticipated to broaden their lytic host range compared to Endo88, but the mutants exhibited the same spectrum of lytic host range as Endo88. The unaltered host-range may be attributed to several reasons:

(I) The inaccurate prediction of the host range for the clades in the phylogeny tree.

Endo88 was initially predicted as a MH endolysin as it grouped in the MH clade (Figure 3.2, section 3.3.3.1) based on our hypothesis that SH3 sequences with high similarity may have similar host-range. However, the TRA revealed that the Endo88 could lyse across genera, as evidenced by its ability to lyse *E. faecalis.* This discrepancy may be caused by the limited number of sequences with defined host range in the Literature Database. Although the clades in the phylogenetic tree were clustered based on SH3 amino acid sequence similarity, the host range annotations were derived from endolysin sequences with experimentally proven host ranges. However, the host range of many endolysins in the Literature Database were ambiguous, as comprehensive testing across *Staphylococcus* sp is limitedly reported, subsequently leading to inaccurate designation of host range to the clades. For instance, the WH clade and NH clade each had only one experimentally proven endolysin to define their host ranges, and the MH clade (with Endo88) had only two (Figure 3.2). This limitation is compounded by the limited number of studies testing lytic activity across different genera as most studies on endolysin focusses on improving lytic activity rather than altering host-range, highlighting a research gap in endolysin studies.

Furthermore, the limited number of sequences in the database might led to increase in the background noise during calculation, potentially affecting the interpretation of the HCSS site. There are studies suggesting that at least 125 sequences are needed to reduce the background noise during AVANA analysis (Gloor et al., 2005; Martin et al., 2005).

(II) Differences in techniques used for quantifying lytic activity.

In the TRA, Endo88 was able to lyse *E. faecalis* (ATCC 29212) and *S. epidermidis* (ATCC 12228). However, an endolysin from the Literature Database, namely MV-L (BAF33253.1), which falls within the same clade with Endo88 was previously reported to be unable to lyse *E. faecalis* ENT1 (Isolate from Kochi Medical School Hospital) and *S. epidermidis* (ATCC 12228) (Rashel et al., 2007) despite sharing a 99% sequence identity with Endo88. The differences could be due to different techniques used to analyse lytic activity. In this study, TRA was used for both *E. faecalis* and *S. epidermidis* to determine the host range of Endo88, where the lysis reaction is performed in Tris buffer. This method is also used in many previous studies (Y. Chang & Ryu, 2017; Ding et al., 2020; Fujiki et al., 2018; Linden et al., 2015; Melo et al., 2018). In contrast, MV-L was tested against *S. epidermidis* but the protocol tested the lysis activity in culture media instead. Furthermore, another technique which is measurement of colony forming unit was used to measure MV-L lytic activity in *E. faecalis* (Rashel et al., 2007).

The inconsistencies of techniques to measure lytic activity of endolysins makes comparison between studies difficult. Notably, one study compared the effects of using buffer versus culture media in lysis reaction (Bhagwat et al., 2021). It was found that the endolysin exhibit weaker lytic activity in the culture media as compared to the buffer, and CBD binding assay also revealed weaker binding in culture media. The authors suggested that the components in the nutrient-rich media could interfere with the CBD binding, thus affecting the overall lytic activity of the endolysin.

Finally, the mutants A1, A2 and M were designed based on the hypothesis that SH3 sequences with high similarity may have similar host range as there may be certain conserved regions in the primary sequence which affects the host-range. However, the primary sequence of a CBD may not be a contributing factor to host-range at all and thus a different strategy based on the tertiary structure of endolysins and its binding site were then used to elucidate the factors that affects host-range in an endolysin (Chapter 5).

In summary, cloning strategies using primers with *Nco*I and *Xho*I restriction sites enabled the expression of soluble and functional Endo88 at 18 °C with 0.5 mM IPTG. biophysical characterization revealed that Endo88 lyse optimally in the absence of NaCl at 37 °C and pH 7.5, with retained activity at higher pH. While lysis was observed on *S. aureus*, *S. epidermidis*, *S. hominis* & *E. faecalis*, the mutations did not alter the host range, as all mutants (A2 and M) exhibited the same lytic host range as the wild-type Endo88. However, mutations in the SH3 domain affected lytic activity, confirming its role in modulating endolysin function. Since the previous mutation did not alter the host range, a different strategy was employed as described in Chapter 5.

CHAPTER 5

DEVELOPMENT OF MUTANTS BASED ON THE TERTIARY STRUCTURE

5.1 Introduction

The previous mutations based on primary sequences of the SH3 domain in Endo88 did not alter the host range as expected, but instead reduced its lytic activity. Therefore, a different approach was employed, focusing on the tertiary structure of the SH3 domain. This involved comparing the binding site between endolysins with different host range. Firstly, the binding site of SH3 domain in Endo88 was determined by: (1) Identification of ligand-interacting amino acids in the binding site of SH3 based on the crystal structure of a homologous SH3 domain (PDB: 6RK4) available online. (2) Binding site prediction using online tool (DoGsite3) to identify the binding site on SH3 domain of homologous endolysins (6RK4). Subsequently, structural alignment between 6RK4 and Endo88 was then carried out to identify the corresponding binding site on the Endo88 SH3 domain. Once the binding site was determined, the amino acids at the binding site of Endo88 was compared to an endolysin of narrower host range and one of wider host range than Endo88. Mutations were then introduced to Endo88 at these sites with expection of host range alteration. Given that the mutations were specifically targeted to the binding site, truncated SH3 domains (without the lytic domains) fused to Green Fluorescent Protein (GFP) were also generated to assess the binding activity of the SH3 domain and the mutants to bacteria cells.

Additionally, 3D structure of SH3 domain (Endo88 and mutants) were also generated through homology modelling. Given the lack of experimentally determined protein 3D

structure, homology modeling serves as a less labor-intensive and time-saving alternative for generating the 3D structure of proteins of interest (Haddad et al., 2020). High-quality 3D structures generated *in silico* can significantly reduce the time and cost needed for wet-lab 3D structure analysis (e.g. protein crystallization) (Cavasotto & Phatak, 2009). The process involves several key steps (Cavasotto & Phatak, 2009; Haddad et al., 2020; Sali & Blundell, 1993): First, homologous proteins with experimentally determines structures are retrieved from databases (e.g. PDB) to serve as templates. The target protein sequences is then aligned with the template squences to position structurally relevant regions to guide model generation. Multiple models are subsequently generated and evaluated for quality using structural validation tools such as the Ramachandran plot. Finally, model refinement including loop region refinements is performed to enhance structural accuracy and ensure proper folding.

This chapter also addresses the second and third objectives of the study which was to design and validate another batch of mutations designed based on their tertiary structure, subsequently studying the relationship between lytic and binding host range through binding assays (GFP-fused SH3 domain) and molecular docking (Objective four). The SH3 binding site of Endo88 was predicted using bioinformatics tools, and mutations were introduced at the predicted binding residues to assess their impact on lytic and binding host range. Molecular docking was used to provide insights into interactions between binding sites and varies peptide crosslinks, which representing different bacterial peptidoglycan.

5.2 Methodology

5.2.1 SH3 domain's 3D structure homology modelling and binding site prediction.

Protein sequence of Endo88 (Accession numbers: YP_240699) was submitted to the SWISS_MODEL online server (https://swissmodel.expasy.org) in December 2021 to retrieve suitable 3D templates for each domains. Suitable templates for each domain were retrieved, except for the looping region between the CHAP and AMD domains, which did not have an appropriate template. To address this, the AMD and SH3 domains were first joined. Subsequently, the AMD+SH3 construct was joined with the CHAP domain using MODELLER 10.1 (Fiser et al., 2000; Sali & Blundell, 1993).

The python script to operate the MODELLER is shown below:-

```
# Homology modeling by the automodel class
from modeller import * # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class
log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this model in
# directories for input atom files
env.io.atom_files_directory = ['.', '']
a = AutoModel(env,
alnfile = '4ols2mk5.pir', # alignment filename
knowns = ('4OLS','2MK5'), # codes of the templates
sequence = 'AMDSH3', # code of the target
assess_methods=(assess.DOPE, assess.GA341))
a.starting_model= 1 # index of the first model
a.ending_model = 20 # index of the last model
# (determines how many models to calculate)

a.make() # do the actual homology modeling
```

After that, the looping region was then refined using the python script shown below:-

```
# Loop refinement of an existing model
from modeller import *
from modeller.automodel import *
#from modeller import soap_loop

log.verbose()
env = Environ()

# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

# Create a new class based on 'LoopModel' so that we can redefine
# select_loop_atoms (necessary)
class MyLoop(LoopModel):
    # This routine picks the residues to be refined by loop modeling
    def select_loop_atoms(self):
        # One loop in chain A from residue 19 to 28 inclusive
        return Selection(self.residue_range('1:A', '25:A'))
        # Two loops simultaneously
        #return Selection(self.residue_range('19:A', '28:A'),
        #                 self.residue_range('38:A', '42:A'))

m = MyLoop(env,
           inimodel='NX88.pdb',    # initial model of the target
           sequence='NX88',                # code of the target
           loop_assess_methods=assess.DOPE) # assess loops with DOPE
#          loop_assess_methods=soap_loop.Scorer()) # assess with SOAP-Loop

m.loop.starting_model= 1           # index of the first loop model
m.loop.ending_model  = 20          # index of the last loop model
m.loop.md_level = refine.fast  # loop refinement method

m.make()
```

A total of twenty models were generated from each modelling during optimization, where three models with the lowest Discrete Optimized Protein Energy (DOPE) score and GA341 (nearest to 1) were chosen for further evaluation (Shen & Sali, 2006). Ramachandran plot generated by PROCHECK (Laskowski et al., 1993) was also used to evaluate the quality of the top 3 models. After the 3D structure was constructed, the structure was then viewed using PyMol 3 (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC).

5.2.1.1 Binding site prediction in SH3 domain of Endo88.

The crystal structure of a SH3 domain derived from another homologous protein (PDB ID: 6RK4) was used to identify the binding site. The crystal structure, co-crystallized

with a peptide crosslink/ligand, was previously deposited by other researchers (Gonzalez-Delgado et al., 2020). The crystal structure was then submitted to the Protein Ligand Interaction Profiler (PLIP) to identify the binding residues and various type of bonds between the ligand (peptide cross-link) and the binding residues in the crystal structure. Subsequently, the amino acid sequence of 6RK4 was then aligned with Endo88's SH3 using EXPRESSO pairwise structural alignment (https://tcoffee.crg.eu/apps/tcoffee/do:expresso), structural alignment score of 98 and a sequence identity of 54.8 % indicates that the proteins are homologous and share a highly similar three-dimentional structure despite moderate sequence divergence. The amino acids at the corresponding positions in Endo88 were predicted to be the binding residues in Endo88's SH3. Online tools – DoGsite3 (https://proteins.plus/#dogsite) (Graef et al., 2023) was used to further validate the binding site predicted in the previous section (5.2.1.1). After homology modelling of Endo88's SH3, the Program database (.PDB) file generated was submitted to DoGsite3 for binding site prediction.

5.2.2 Mutation on the interacting residue located within the SH3's binding groove.

Two endolysins with different host range than Endo88 were used as reference endolysins to design mutations based on the biding site. E-LM12 (AUV56903.1) (a narrower host range endolysin) (Costa et al., 2020; Melo et al., 2018) and LysWMY (BAD83402) (a wider host range endolysin) (Yokoi et al., 2005), were selected from the literature database as reference to design the mutations. These endolysins were seletected due to their well-defined host range, as demonstrated in their original studies. Pairwise sequence alignment was performed between the reference endolysin (E-LM12 & LysWMY) and Endo88 using EMBOSS Needle Wunsch pairwise

alignment tool. Additionally, extra five amino acids upstream of the SH3 domain (Endo88: NKYGT; E-LM12: NQFGT; LysWMY: NEYGT) were included in the alignment as they were predicted as binding residues. Through this alignment, the binding residues located on the SH3 domain of Endo88 were changed with residues of either E-LM12 or LysWMY at the corresponding amino acid position. The objective of this approach was to introduce the binding characteristics from the reference endolysins into Endo88, with the aim to modify its host range.

5.2.3 Construction of recombinant plasmid pET28a encoding for mutant gene (D1, D2)

SOE-PCR was used to introduce the mutation into the wild-type Endo88 gene using designed primers set (Table 5.1). After mutation, the gene was then cloned into pET28a using the same methodology in section 4.2.1. The SOE-PCR process for constructing mutant gene *D1* and *D2* was same as in section 4.2.5.1, with slight changes for annealing condition for region joining (Table 5.2). The first mutant – denoted D1, was generated by separating the Endo88 gene into 3 regions with primer-initiated complementary overhangs as shown in Figure 5.1A. The regions were assembled using SOE-PCR, region 2 (68 ng) was first combined with region 3 (150 ng), and then this combined sequence was ligated with region 1 (Figure 5.1B). For D2, initially it was divided into five regions by used of five different primer sets. However, region 2, 3, 4 and 5 failed to amplify in gradient PCR. To overcome this issues, sequences of region 2, 3, 4, 5 were combined into single region (denoted as Region 2) and synthesized (Integrated DNA Technologies_IDT) instead and joined with region 1. The synthesized Region 2 was then joined with Region 1 similar to Figure 5.1.

Table 5.1: SOE-PCR primer set for the mutants (*D1* & *D2*).

Mutant Gene	Primer	Primer sequence	Ta (°C)
D1	Region1_D1-Fwd	5'- CATG *CCATGG* CA **ATGCAAGCAAAATTAACTAAAAAAG** -3'	60
	Region1_D1-Rev	5'- **TTGCGAGTGCATGGAAACGT** AATCAGTTCGGTACTTACTACAT -3'	60
	Region2_D1-Fwd	5'- AATCAGTTCGGTACTTACTACAT**GGAAGA** -3'	60
	Region2_D1-Rev	5'- **CAACCAATCACTGTAAGAAAA** GGTGGACCATTCTTATCATGC -3'	60
	Region3_D1-Fwd	5'- GGTGGACCATTCTTATCATGC**CCG** -3'	60
	Region3_D1-Rev	5'- CC *CTCGAG* GCC **AGGTGACTTATGGGGAGAAATCAGT**-3'	60
D2	Region1_D2-Fwd	5'- CATG *CCATGG* CA **ATGCAAGCAAAATTAACTAAAAAAG** -3'	60
	Region1_D2-Rev	5'- **CGAGTGCATGGAAACGTAAT** GAATATGGTACTTACTGG -3'	60
	Region2A_D2-Fwd	5'- GAATATGGTACTTACTGG **ATGGAAGAAAGTGCTAGATTC** -3'	60
	Region2A_D2-Rev	5'- CC *CTCGAG* GCC **AGGTGACTTATGGGGAGAAATCAGT**-3'	60
D2 (Region 2, 3, 4, 5. Expected bands NOT detected)	Region2_D2-Fwd	5'- GAATATGGTACTTACTGG **ATGGAAGAAAGTGCTAGATTC** -3'	55
	Region2_D2-Rev	5'- **AACCAATCACTGTAAGAAAAACA** -3'	55
	Region3_D2-Fwd	5'- CCAATCACTGTAAGAAAAACA **GGACCATTCTTATCATGCC** -3'	55
	Region3_D2-Rev	5'- **GGTGGATATTGTGATTATACAACT** -3'	55

Table 5.1: Continued

Mutant Gene	Primer	Primer sequence	Ta (°C)
	Region4_D2-Fwd	**5'- GTGGATATTGTGATTATACAACT GTGATGTTACAAGATGGT -3'**	**55**
	Region4_D2-Rev	5'- **GAGGGGCAACGTTATTGG** -3'	55
	Region5_D2-Fwd	5'- GGGGCAACGTTATTGG **TTGCCTATTAGAACATGGAATGG** -3'	55
	Region5_D2-Rev	5'- CC *CTCGAG* GCC **AGGTGACTTATGGGGAGAAATCAGT** -3'	55
GFP (GFPSH3)	GFPa-Fwd	5'- CATG *CCATGG* CA **ATGGTGAGCAAGGGCGAGGA** -3'	60
	GFPa-Rev	5'- CGC *GGATCC* GCG **GGCATGGACGAGCTGTACAAG** -3'	60
SH3 domain (GFPSH3)	SH3a-Fwd	5'- CGC *GGATCC* GCG **GCGAGTGCATGGAAACGTAA** -3'	60
	SH3a-Rev	5'- CC *CTCGAG* GCG **ACTTATGGGGAGAAATCAGT** -3'	60
GFP (SH3GFP)	GFPb-Fwd	5'- CGC *GGATCC* GCG **ATGGTGAGCAAGGGCGAGGA** -3'	60
	GFPb-Rev	5'- CC *CTCGAG* GCC **GGCATGGACGAGCTGTACAAG** -3'	60
SH3 domain (SH3GFP)	SH3b-Fwd	5'- CATG *CCATGG* CA **GCGAGTGCATGGAAACGTAA** -3'	60
	SH3b-Rev	5'- CGC *GGATCC* GCG **ACTTATGGGGAGAAATCAGT** -3'	60

Bold: Template's sequence; *Italic*: Restriction site; Red: SOE-PCR overlapping site. Note: (i) the region 2 in D2 was synthesized (Integrated DNA Technologies_IDT) instead of generate using SOE-PCR. Therefore, primer Region2A_D2-Fwd and Region2A_D2-Rev were used to amplify region 2 of D2.

Table 5.2: Recipe for SOE-PCR, annealing condition for region joining.

Ingredients	Final amount in 25 µL	PCR conditions
dH_2O	Top up to 25 µL	98 °C 3 mins
5x Phusion HF buffer	1X	98 °C 10 s
dNTP Mix	0.2 mM	65 °C 30 s x15 cycles
Phusion High-Fidelity DNA polymerase	1.25 U	72 °C 30 s
Region 1	150 ng (6 ng/µL)	72 °C 5 mins
Region 2	36 ng (1.44 ng/µL) (Mutant D1, Region 2+3) 36.04 ng (1.44 ng/µL) ng (Mutant D2)	4 °C -

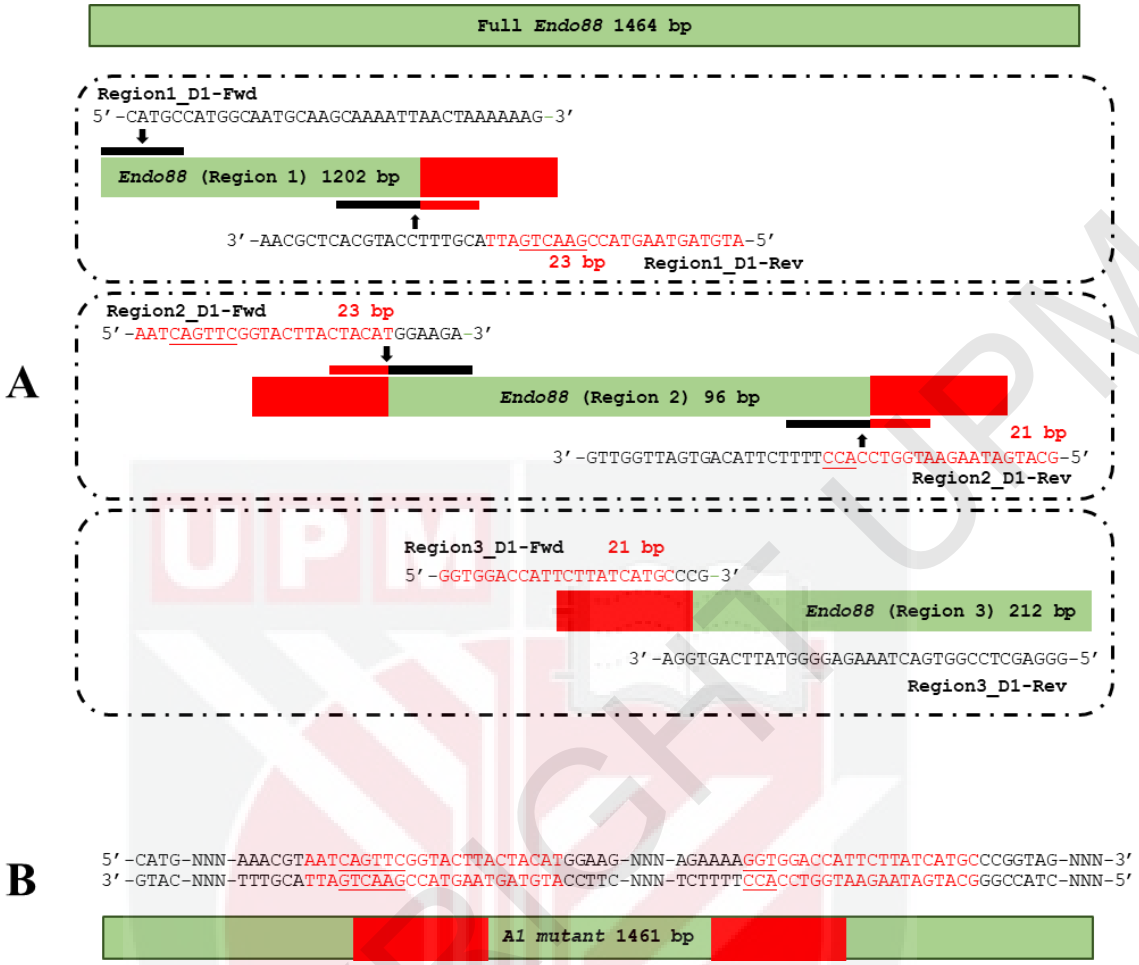


Figure 5.1: Schema of mutant gene (D1) development through SOE-PCR. Red color region denotes the overlapping region which allow annealing of all the regions. Underline represent the mutation to be introduces into Endo88 SH3 domain. **(A):** Primers were designed to amplify each regions. Each primer sets gereated sequences with flanking region (red) that contain the mutations and overlapping rgion to facilitate the joining of the fragments. **(B):** Regions were joined into one continuous DNA fragment incorporating the mutation (Underline).

5.2.3.1 Construction of GFP fused SH3 domain (SH3 of Endo88, D1 and D2)

For GFP-fused mutants, the SH3 gene of endolysins (Endo88, D1 and D2) and GFP were PCR amplified using two different primer sets (GFPSH3 & SH3GFP) listed in Table 5.1. The GFP gene was isolated from pOLTV5 (rAF-GFP) plasmid, which was gifted by Dr. Chia Suet Lin (UPM). The processes of amplifying the gene, including gradient PCR were similar in section 4.2.1.5. Initially, the GFP gene was first cloned into pET28a to generate pET28a (*GFP*) for later use. The GFP gene was fused upstream of SH3 gene at the N-terminal (Figure 5.2). However, it is noteworthy that the the putative binding site is located at the N-terminal of the SH3 domain (Gu et al., 2014; Hirakawa et al., 2009; J. Z. Lu et al., 2006; Mitkowski et al., 2019; Tossavainen et al., 2018), raising concern that the N-terminal GFP fusion might interfere with folding and affect the binding site. Therefore, the location of the GFP was taken into consideration.

Several other studies using GFP-fused SH3 domain were referred (Y. Chang & Ryu, 2017; Costa et al., 2020; Gu, Lu, et al., 2011; Kong et al., 2015) and the location of GFP varied among them. Therefore, an additional primer set (SH3GFP) was used to swap GFP to the C-terminal of SH3 (Figure 5.3), to determine the effect of GFP location on binding activity. The final products were then transformed into BL21(DE3) *E. coli* for expression.

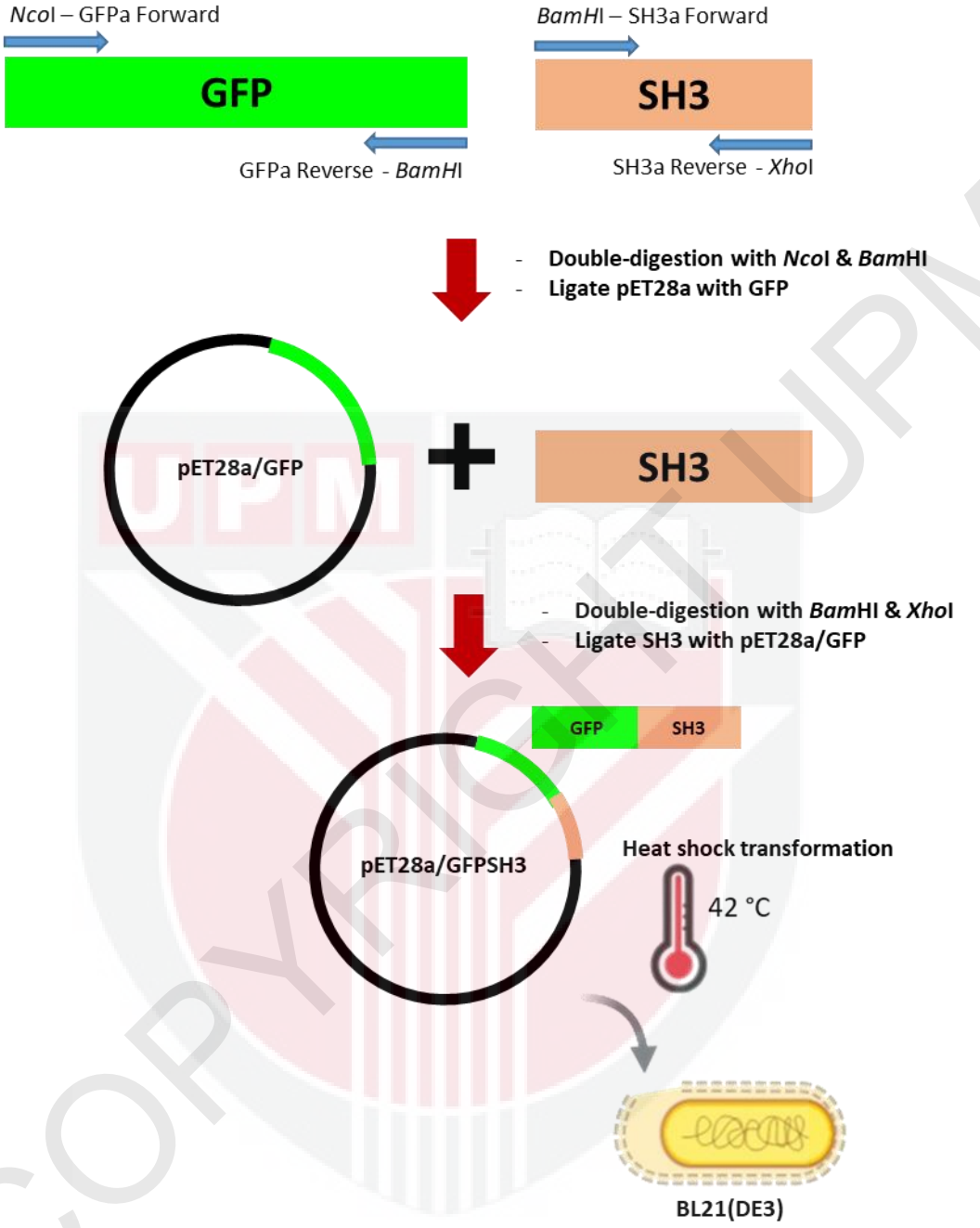


Figure 5.2: Schematic representation of GFP fused at N-terminal of SH3 with designed primer set. Initially, the GFP gene was first cloned into pET28a to generate pET28a (*GFP*) for later use. The GFP gene was fused upstream of SH3 gene at the N-terminal

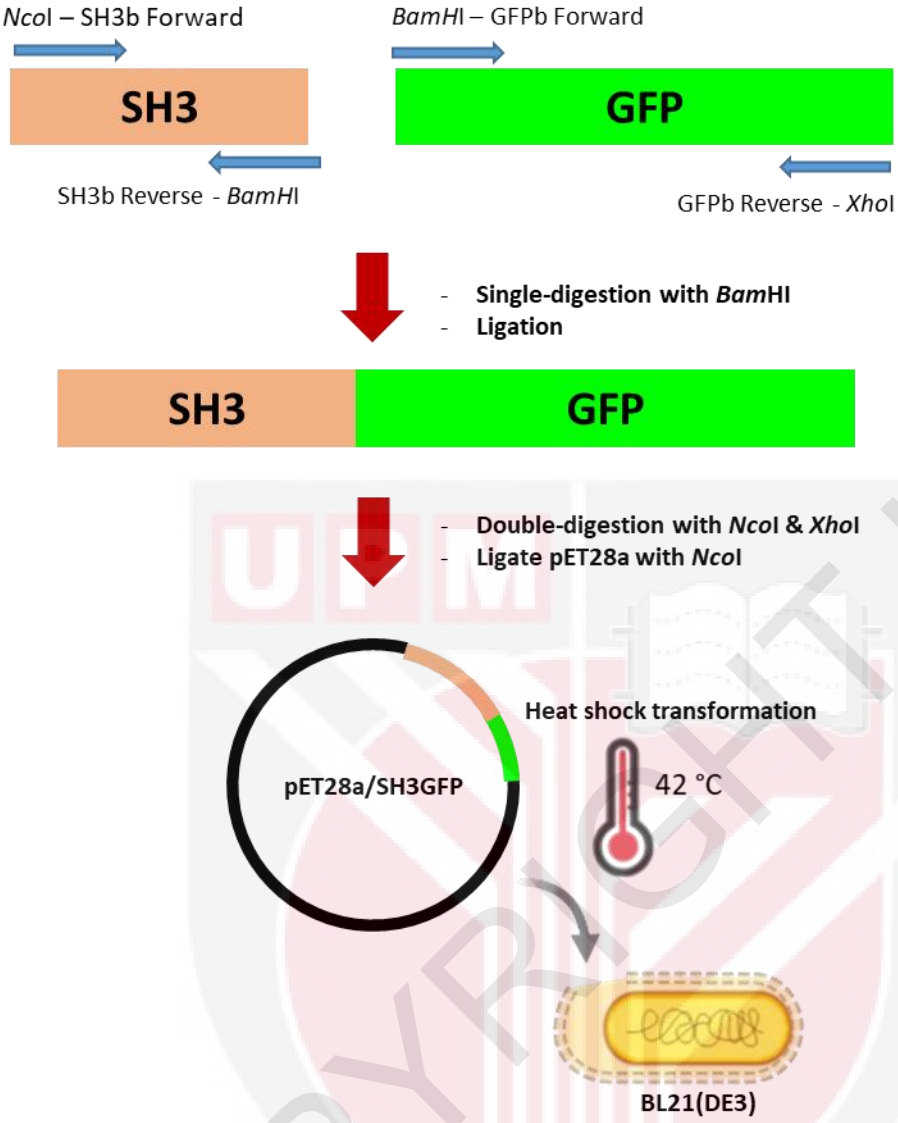


Figure 5.3: Schematic representation of GFP fused at C-terminal of SH3 with designed primer set. GFP was first fused at the C-terminal of the SH3 domain. Subsequently, the SH3_GFP was then cloned into pET28a.

5.2.3.2 Expression of mutants (D1 and D2) and lytic host range assessments

The expression and assay were conducted with same conditions as described in section 4.2.2.2.

5.2.4 Flow cytometry binding assay

The experiment was adapted from other studies (Y. Chang & Ryu, 2017; Costa et al., 2020; Gu, Lu, et al., 2011; Kong et al., 2015), with slight modifications. Bacterial cells were cultured overnight in their respective culture media. Subsequently, 200 µL of the overnight culture were subcultured into 20 mL of fresh culture medium and incubated until reaching OD600 0.6 – 0.7. Cell pellets were collected via centrifugation (4000 x*g*, 10mins, and 4 °C) and washed thrice with cold phosphate-buffered saline (PBS) buffer. After washing, 1 mL of culture was transferred into a 1.5 mL microcentrifuge tube, centrifuged (12, 000 xg, 4 °C, 5 mins) and the pellet was washed once with 500 µL of Tris buffer (50 mM Tris, pH 7.5) and then re-suspended with 100 µL of Tris buffer. Suspended pellets were subsequently incubated with GFPSH3/SH3GFP (10 µg) for 30 minutes at room temperature. After incubation, the mixture was centrifuged centrifuged (12, 000 xg, 4 °C, 5 mins) and washed thrice with cold PBS. The pellets were then re-suspended in cold fixing solution (2 % paraformaldehyde in PBS) and incubated for 10 minutes on ice. Subsequently, the mixture was washed thrice with 300 uL cold PBS and finally re-suspended in 500 uL PBS for flow cytometry analysis (NovoCyte Flow Cytometer, Agilent, USA). A total of 50,000 events with a sample flow rate of 10 uL min^{-1} were used for a single run, and fluorescence was detected with a FIT-C filter. Gram-negative bacteria *E. coli* served as a negative control in this assay.

5.2.5 Molecular docking

Two docking tools – HADDOCK and Autodock4 were used to compare their applicability on this study. These tools were validated by re-docking the protein-ligand crystal structure available in the PDB database (PDB: 6RK4), which consists of SH3 domain co-crystalised with a crosslink peptide ligand (Gonzalez-Delgado et al., 2020). The crosslink peptide ligand consists of a tetrapeptide stem (L-ala, D-isoGlu, L-Lys, D-ala) linked to a pentaglycine (Gly-Gly-Gly-Gly-Gly), representing the peptide stem between the *N*-acetylglucosamine (GlcNAc) and *N*-acetyl muramic acid (MurNAc) polymer (Figure 2.6) within the cell wall. The docking was justified based on several criteria: (i) the peptide crosslink fitting and binding into the binding groove. (ii) Root Mean Square Deviation (RMSD) comparison between the docking pose and the 6RK4 structure. (iii) The similarity of the binding residues involved with those in 6RK4.

5.2.5.1 Autodock4

Autodock 4.2.6 manual guide was referred to and the procedures were summarized in Figure 5.4. Prior to the docking process, an extended PDB format, denoted as PDBQT file, was used as the coordinate file. It contains the information required for AutoGrid and AutoDock software to generate the SH3 domains and peptide crosslink PDB file. These coordinate files consist of information including assignment of polar hydrogen atoms, partial charge (Kollman charges) and number of torsion. Polar hydrogen bonds and Kollman charges were assigned to the peptide crosslink and SH3 domain. Instead of rigid docking, flexible docking was performed**.** Various numbers of torsions were assigned on the peptide crosslink (ligand), and amide bonds were set to non-rotatable. This is to test the Autodock4's ability to handle flexibility in ligand, as bacterial

peptidoglycan are flexible in nature to adapt the osmotic changes in envorinment (Doyle & Marquis, 1994; Scheffers & Pinho, 2005; Vollmer et al., 2008)

Subsequently, an Autodock4 build-in tool- AutoGrid was used to create three-dimensional lattice grid map, which surrounded and centred on the region of interest within the SH3 domain. The space covered in the grid consists of calculated information which speed up the docking process. Two type of grids were set where: (i) A lattice grid covered the entire SH3 domain (Figure 5.5A) and (ii) a lattice grid covered only the binding groove, in accordance with residues predicted to be involved in binding (Y411, N405, E451, Y472, T429, Y407, K406) (Figure 5.5B). A total of 100 Genetic Algorithm (GA) runs were set for each docking. Each pose generated were then evaluated manually to check the binding location of the ligand and Root Mean Square Deviation (RMSD) differences with 6RK4 crystal structure.

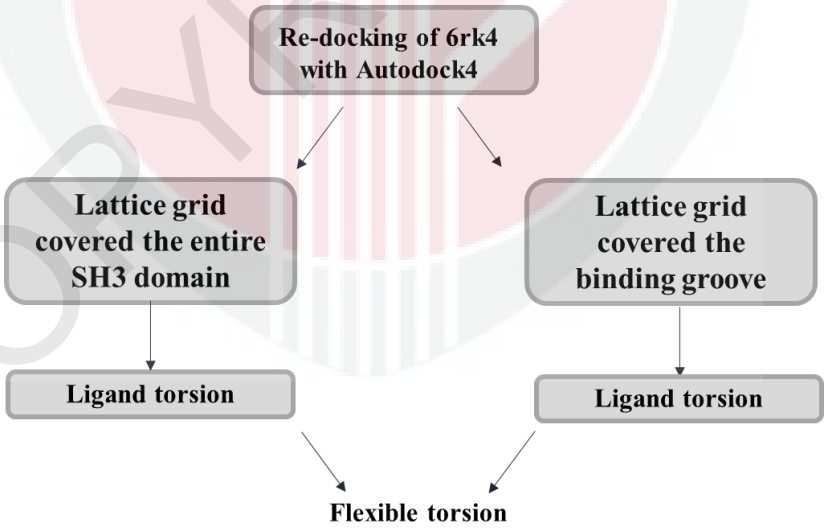
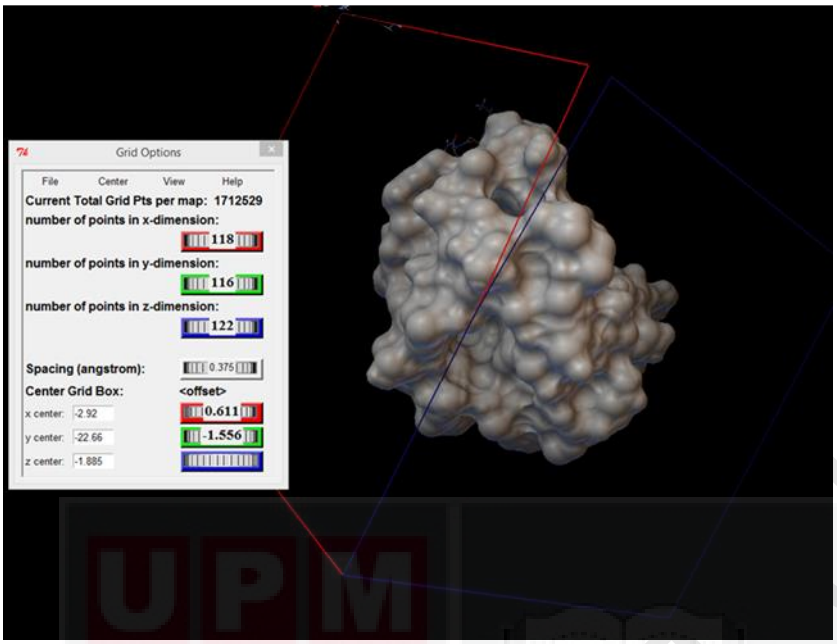


Figure 5.4: Overview of the workflow used to prepare the PDBQT coorfinate file for Autodock4. Two types of lattice grids were generated: (i) a lattice grid that encompassed the entire SH3 domain, which is known as the blind docking; (ii) a lattice grid that specifically targeted the binding groove.

A



B

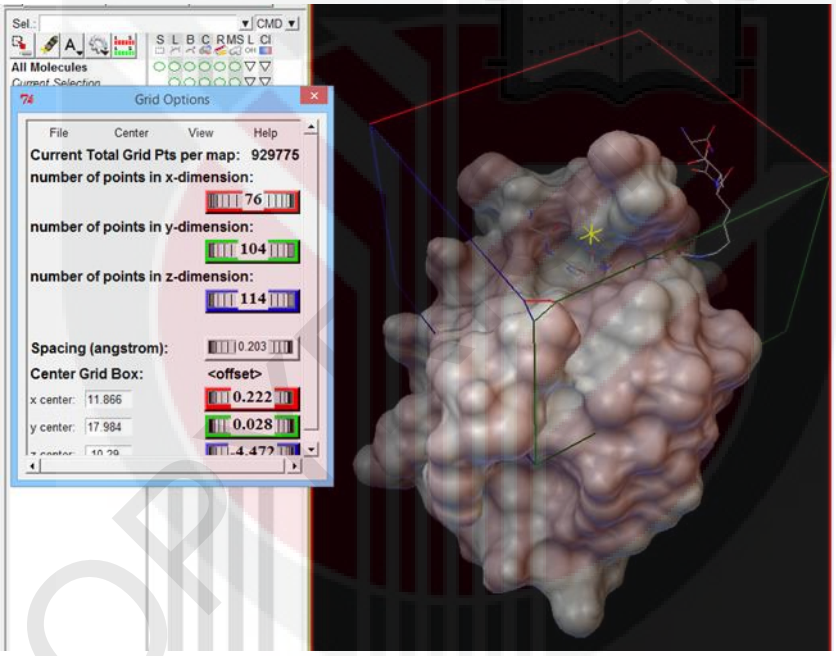


Figure 5.5: 6RK4-The crystal structure of Lysostaphin SH3b in complex with P4-G5 ligand. (A) Lattice grid was set encompass the entire SH3 domain. **(B)** Lattice grid was adjusted to specifically cover the binding groove.

5.2.5.2 HADDOCK 2.4

High Ambiguity Driven protein-protein DOCKing (HADDOCK) online tool was used for comparison with Autodock4. Although most settings were kept as default, a few modifications were made where: (i) the entire peptide crosslink were set as semi-flexible and (ii) docking with Ambiguous Interaction restraint (AIR) and without AIR were compared. AIR script was prepared to guide the peptide crosslink towards the binding site, akin to the functionality of the AutoGrid system in AutoDock. The distances in the AIR script were derived from the 6RK4 crystal structure, which was analysed by PLIP (Figure 5.6). The details of the AIR script are outlined below:

```
assign (resid 1 and segid A)
       (resid 405 and segid B) 2.6 1.0 1.0

assign (resid 1 and segid A)
       (resid 411 and segid B) 2.6 0.5 1.5

assign (resid 1 and segid A)
       (resid 451 and segid B) 2.6 0.5 0.5

assign (resid 1 and segid A)
       (resid 472 and segid B) 2.6 0.5 0.5

assign (resid 1 and segid A)
       (resid 429 and segid B) 2.6 0.5 0.5

assign (resid 1 and segid A)
       (resid 407 and segid B) 3.0 0.5 3.0

assign (resid 1 and segid A)
       (resid 406 and segid B) 2.6 0.5 1.0
```

```
assign (segid A and resid A)
       (segid B and resid B) Distance, lower-bound_correction, upper-bound correction
```

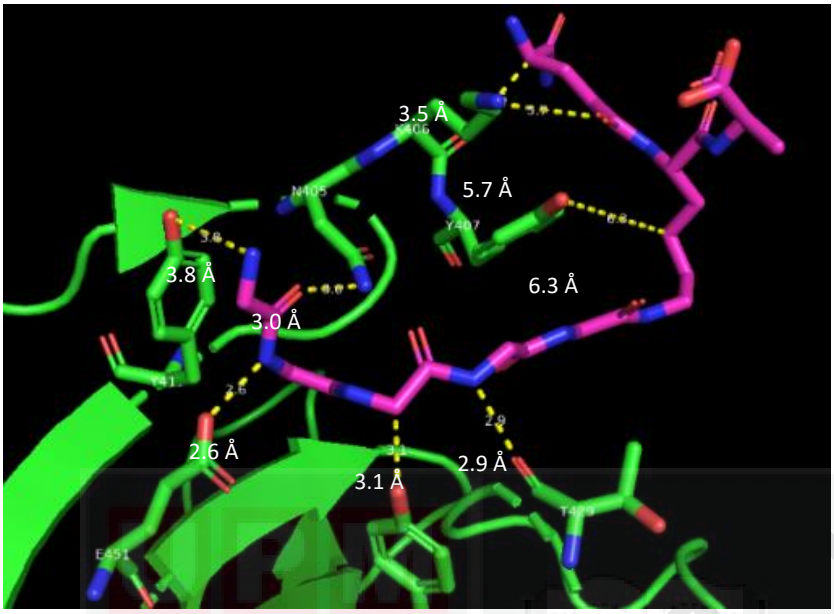


Figure 5.6: Crystal structure of LSS (6RK4)'s CBD analyzed by PLIP. Green: CBD; Purple: peptide crosslink. Pymol was used to measure the distance (Å) between the interacting amino acids and the ligand.

5.2.5.3 Molecular docking of SH3 domain with varies type of peptide crosslink.

Molecular docking was conducted to study the amino acid interactions between the SH3 domain and various peptide crosslink. First, the peptide crosslinks were generated using Avogadro 1.2.0 molecules editor and stabilized using UFF (Universal Force Field) and Steepest Descent algorithm. To ensure the reliability of the peptide crosslink generated from Avogadro, a peptide crosslink of *S. aureus* was generated and served as a control. The validation of the generated peptide crosslink was summarized in Figure 5.7. In brief, validation was performed by docking the generated *S. aureus* peptide crosslink to 6RK4 (ligand removed) and compared to the resulting amino acid interaction profile. Additionally, the peptide crosslink was docked into the modeled SH3 domain of Endo88 to create a control interaction profile for subsequent docking experiments.

HADDOCK (de Vries et al., 2010) and PRODIGY tools (Vangone & Bonvin, 2015; Xue et al., 2016) were used to evaluate the binding pose and predict the binding energy of the pose, respectively. The selection of the best binding pose was based on several criteria:

(i) Experimental validation from other studies, ensuring similarity between the amino acid interaction profile in the docking and experimental results (6RK4). The interacting amino acids were identified using Protein-Ligand Interaction Profiler (PLIP).

(ii) Consistency in the binding site, ensuring that the ligand is consistently bound to the same binding site across the docking poses generated.

(iii) The ability for the ligand to bind into the predicted binding groove.

(iv) The best binding score (HADDOCK score).

To accommodate the flexibility of the peptide crosslink, the docking was guided by Ambiguous restraint (AIR) and also Un-ambiguous restraint (UIR) in some circumstances to prevent uncommon binding and also enhance docking quality. The flexibility of the peptide crosslink increases the chances of peptide crosslink which could fit and accomodate to anywhere around the binding groove (Figure 5.8A, B, C), which reduce the accuracy of the docking. With the guidance from experimental result (PDB: 6RK4), the distance between atoms were used in AIR and UIR.

Subsequently, the docking was repeated with solvated docking and further refined with short molecular dynamics in explicit solvent (water). In the Input Data section, the peptide crosslink was set as “Ligand” instead of “Peptide” due to the presence of iso-

glutamic acid and unconventional peptide bonding, which were not recognized by the HADDOCK 2.4. Additionally, the peptide crosslink was set as full semi-flexible segments in the Input Parameters.

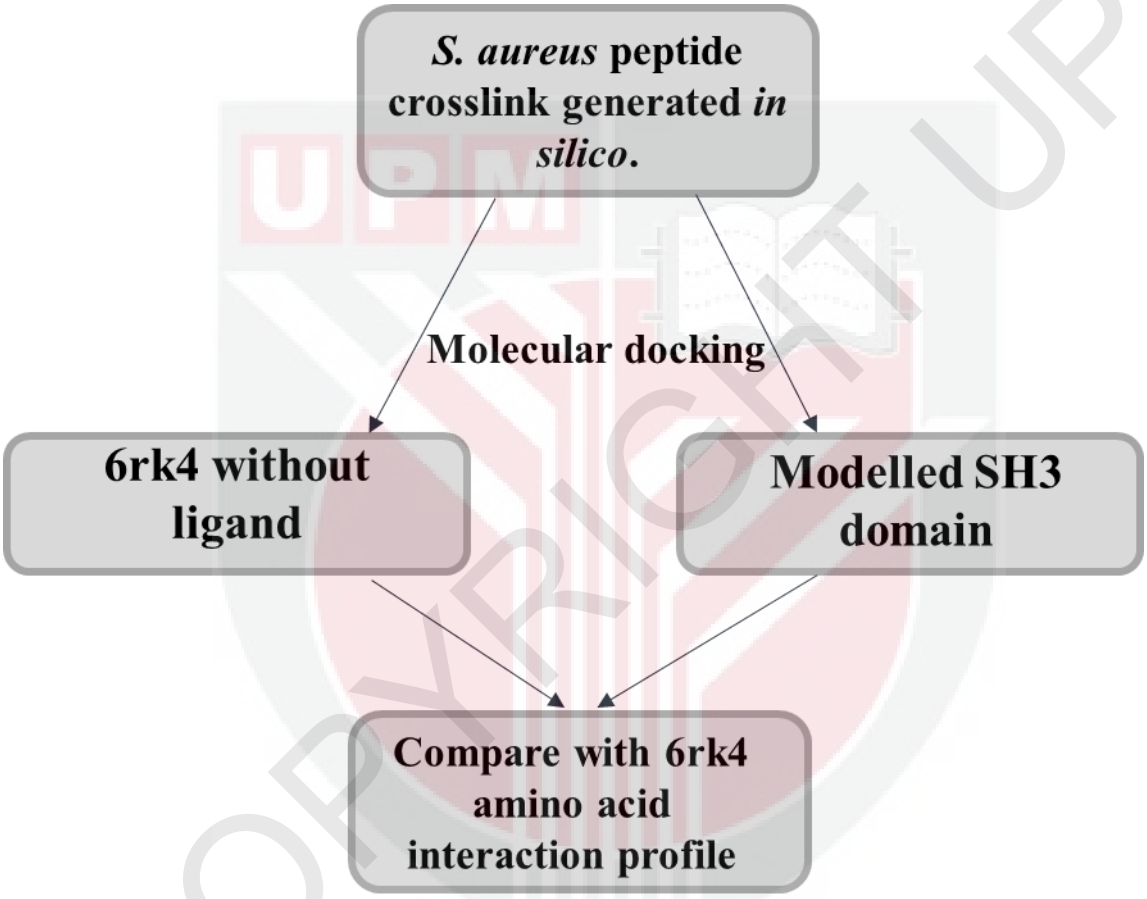


Figure 5.7: Validation steps for peptide crosslink generated *in silico*. The Avogadro 1.2.0 molecules editor was used to generate the peptide crosslink for molecular docking. The generated peptide crosslink was validated through the following steps: (i) Docking the peptide crosslink to 6RK4 (ligand removed); (ii) Docking the peptide crosslink to the modelled SH3 domain of Endo88. The amino acid interaction profiles from both docking experiments were then compared with the native 6RK4 structure (with ligand).

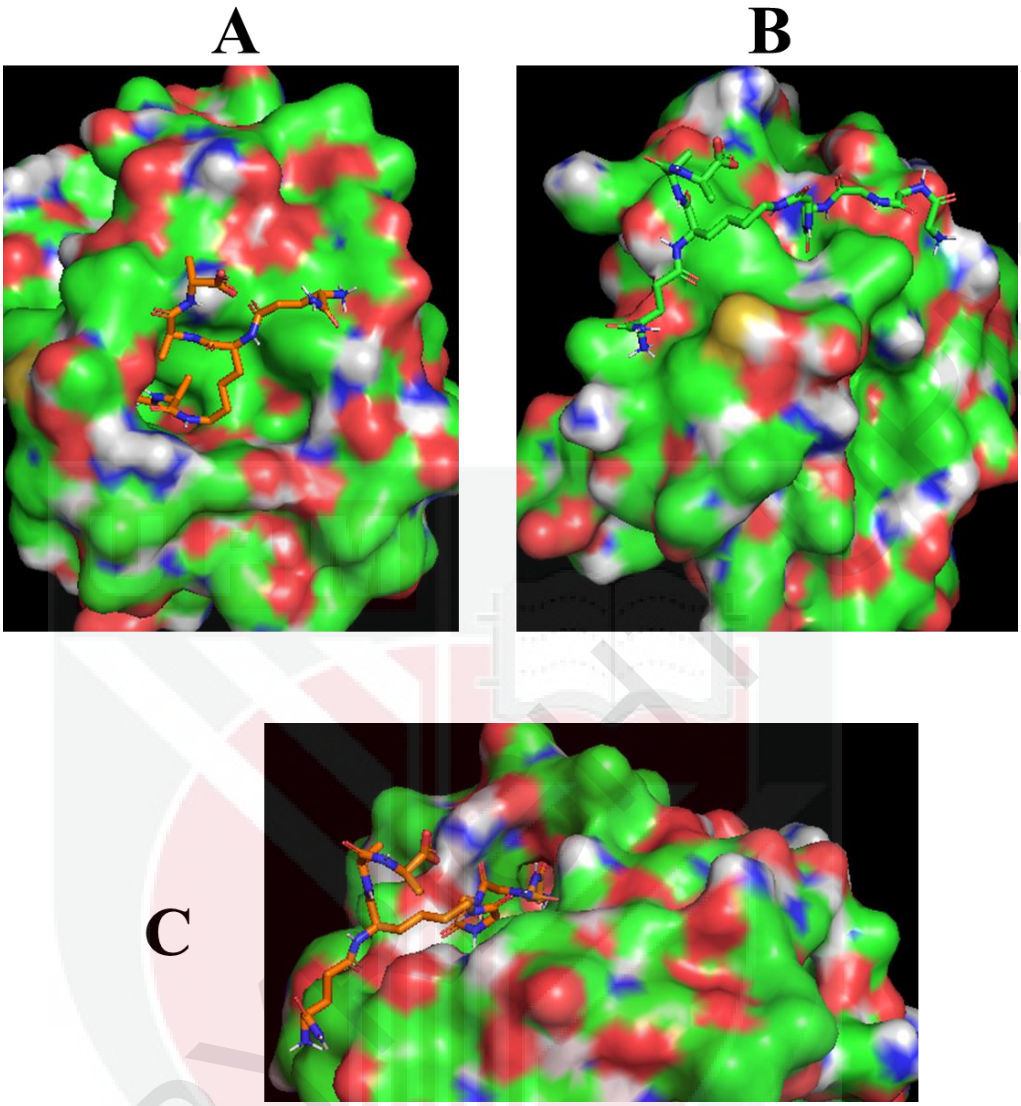


Figure 5.8: Example of docking pose with peptide crosslink bind at uncommon binding site with odd comformation. (A): One of the uncommon binding pose between peptide crosslink of *E. faecalis* and Endo88_SH3. The peptide crosslink was bound to a groove away from the binding groove, and the tail of the interpeptide bridge wedged into the binding groove. **(B):** One of the uncommon binding pose between peptide crosslink of *S. aureus* and Endo88_SH3. The peptide crosslink did not fit into any binding groove, but was bound and localized on the surface of the SH3 domain. **(C):** One of the uncommon binding pose between peptide crosslink of *S. aureus* and Endo88_SH3. The peptide crosslink was bound to a groove away from the binding groove, and the pentaglycine chain was wedged deep into the binding groove.

5.3 Result & Discussion

Homology modelling was conducted for Endo88 (WT) and its mutants D1 (predicted to have a narrower host range than WT) and D2 (predicted to have a wider host range than WT).

5.3.1 SH3 domain's 3D structure homologous modelling

Instead of generating the 3D structure of the SH3 domain alone, the entire endolysin structure was modeled. This approach is generally preferred as the interactions between domains within a protein can influence its overall structure and stability (Schauperl & Denny, 2022; Xia et al., 2023; Zhou et al., 2019). The amino acid sequence of Endo88 was submitted to the SWISS-Model database in December 2021. Three templates were retrieved (Table 5.3) and multiple sequence alignment was performed (Figure 5.9).

As shown in Figure 5.9, a loop region without structural information required loop refinement to resolve the missing structural data. Initially, the AMD and SH3 domains were first joined together and a total of 20 models were built.

Table 5.3: Templates for Endo88 homology modeling.

Template (PDB ID)	Corresponding domain	GMQE	Seq Identity	Coverage
6ist	CHAP	0.15	21.68	30%
4ols	AMD	0.28	51.22	42%
2mk5	SH3	0.18	45.08	25%

GMQE: Global Model Quality Estimation.GMQE score range between 0 and 1. The higher the number, the higher the reliability (Waterhouse et al., 2018).

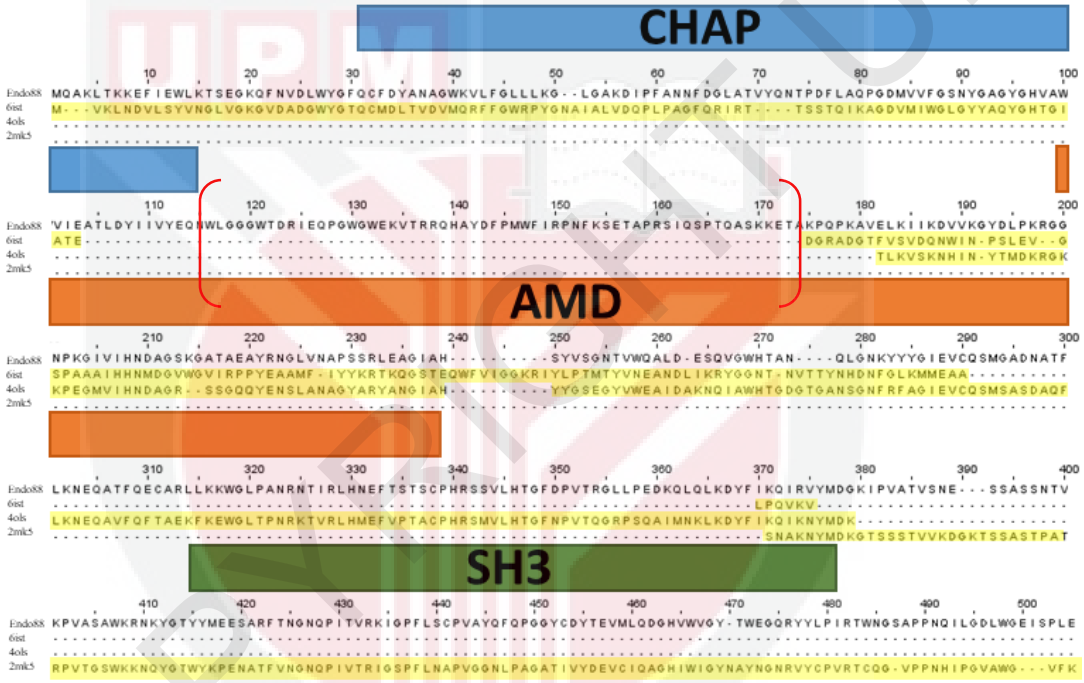


Figure 5.9: Multiple sequence alignment of Endo88 with the template (PDB: 6IST, 4OLS & 2MK5). Yellow highlights: Sequences containing structural information, which serve as the foundation for building the 3D model of each domain in Endo88. Red bracket: Loop region of Endo88 without any structural information.

The top three models were chosen based on the top three best DOPE and GA341 score obtained. A lower DOPE score indicates a better model, and a GA341 score closer to 1 is preferred. The best model was then submmited to PROCHECK (Laskowski et al., 1993) for Ramachandran plot analysis (Table 5.4). The Ramachandran plot provides insight into the stereochemical quality of the modeled structure by evaluating the distribution of backbone dihedral angles (ϕ and ψ) of amino acid residues. A higher percentage of residues in the favored and allowed regions suggests a well-defined model (Hollingsworth & Karplus, 2010; Lovell et al., 2003). While a value of >90 % in the favored region is generally considered good, models with >80 % in the favored region are still acceptable, especially given the lack of template during homology modeling. After the 3D structure of AMD+SH3 was confirmed, the process was repeated by joining AMD-SH3 to the CHAP domain. Loop refinement was performed on the complete 3D structure before submitting it to PROCHECK (Table 5.4). Finally, the top three best models were then submitted to PROCHECK to determine the best model (Figure 5.10).

Table 5.4: Models quality scoring generated by Modeller and Ramanchandran plot evaluation.

Models	DOPE	GA341	Ramanchandran	
			Most favoured regions (%)	Disallowed regions (%)
AMD+SH3	-30071.4	1	88.5	1.5
CHAP+(AMD+SH3)	-45100.1	1	83.8	1.8
Loop refinement	-2271.8	1	83.1	1.8

DOPE: Discrete Optimized Protein Energy

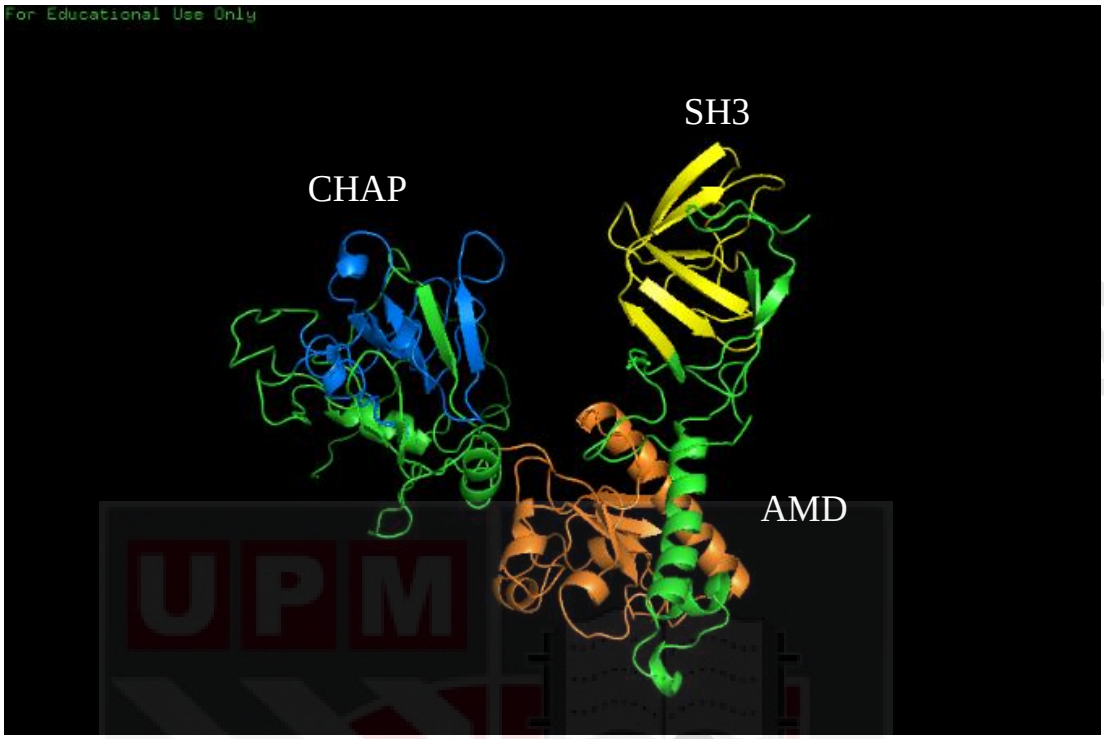


Figure 5.10: 3D structure of Endo88 generated by MODELLER and viewed in PyMol. Best model was chosen from PROCHECK analysis. Blue: CHAP domain; Orange: AMD domain; Yellow: SH3 domain.

5.3.2 Binding site prediction in SH3 domain of Endo88.

The binding site of SH3 domain in Endo88 was predicted using binding site information from other studies, in conjunction with a binding site prediction tool. The predicted binding residues were then used as mutation candidates for mutational studies.

6RK4 is an orthologous SH3 domain derived from bacterial lysostaphin (LSS) with experimentally validated binding sites. It served as a reference for Endo88 binding site prediction due to its unique co-crystallization with a peptide crosslink (Gonzalez-Delgado et al., 2020). This ortholog shares a sequence identity of 54.8 % and a similarity of 58.1 % with Endo88. Moreover, it achieved 98 out of 100 score in structural alignment (Figure 5.11A), indicating high structural similarity between them

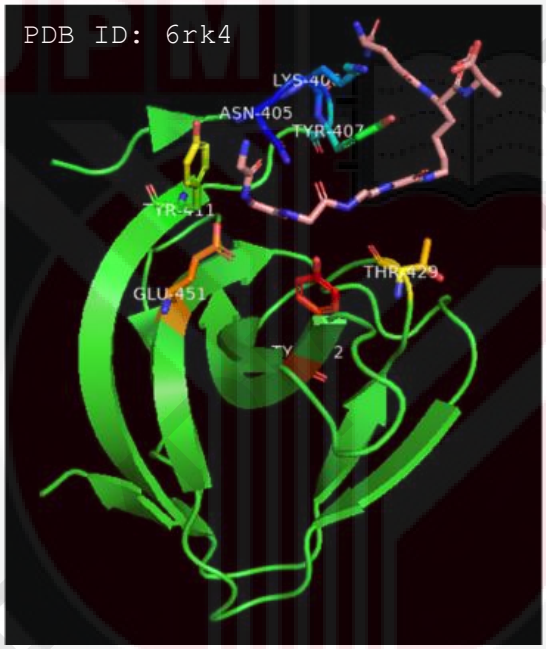
(C. Kim & Lee, 2007; Malmström et al., 2007; Zhou et al., 2019). The crystal structure was then submitted to PLIP to identify the interacting amino acid between the ligand and 6RK4 (Fig 5.11B). Through pairwise structural alignment, the amino acids at the corresponding position in Endo88 SH3 were predicted to be the interacting amino acid. These residues (N387, K388, Y389, Y393, E434, Y454, and I412) were then used as mutation candidates (Figure 5.12).

Online binding site prediction tools – DoGSite3 were utilized to further validate the previously predicted binding site. One of the detected binding grooves aligned with the earlier prediction (Figure 5.13), providing further validation. This binding groove was 45.026 $Å^3$, and consists of four Hydrogen (H) bond donors, four H bond acceptors and six aromatic atoms. This combination indicated that the binding mechanism likely involves non-covalent bonding (Adhav & Saikrishnan, 2023).

A

```
T-COFFEE, Version_11.00 (Version_11.00)
Cedric Notredame
SCORE=98
*
 BAD AVG GOOD
*
Endo88    :  98
LSS       :  98
cons      :   9

Endo88    NKYGTYYMEESARFTNGNQPITVRKIGPFLSCPVAYQFQPGGYCDYTEYMLQDGHVWVGYTW-EGQRYYLPI
LSS       NKYGTLYKSESASFTP-NTDIITRTTGPFRSMPQSGVLKAGQTIHYDEYMKQDGHVWVGYTGNSGQRIYLPV

cons      ***** * .*** **  *  * .*. *** * * :  ::.*   .* *** **********  .*** ***:


Endo88    RTWNGSAPPNQILGDLWGEIS
LSS       RTWNKS---TNTLGVLWGTIK

cons      **** *    .: ** *** *.
```

B



6rk4	Endo88
Y411	Y393
N405	N387
E451	E434
Y472	Y454
T429	I412
Y407	Y389
K406	K388

Figure 5.11: Binding site prediction in SH3 of Endo88 by the use of homologous protein reference. (A) Structural alignment of Endo88 and LSS_6RK4. Scores greater than 50 are considered reliable alignments, indicating structural similarity. Red box: 6RK4 amino acids involved in binding, as analysed by PLIP. Alignment was done to predict the binding amino acid in Endo88. **(B)** Crystal structure of 6RK4 illustrated using Pymol 3. Interacting amino acids (colored, stick) were identified by use of PLIP, pink color stick is the ligand.

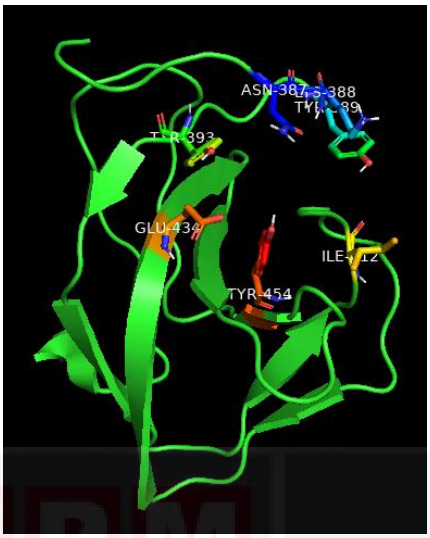


Figure 5.12: 3D model structure of Endo88's SH3. (A): The colored residues (N387, K388, Y389, Y393, E434, Y454, and I412) were the interacting amino acid predicted.

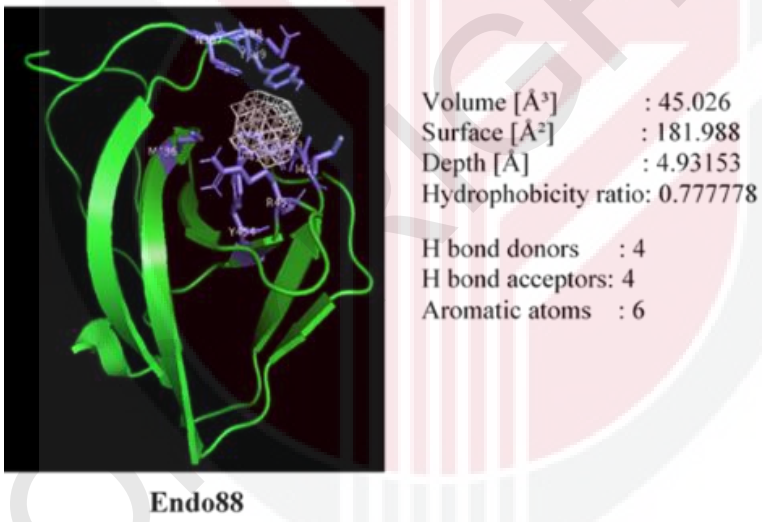


Figure 5.13: 3D structure analysis of the modeled SH3 domain. The size, shape and chemical features of the predicted binding groove (represented by the white mesh) were calculated using DoGSite3. Residues highlighted in blue indicate the predicted interacting amino acids in SH3 domain of Endo88.

5.3.3 Mutation on the binding residues located within the SH3's binding groove

Two reference endolysins – E-LM12 (AUV56903.1) and LysWMY (BAD83402) were used as the reference for mutation design due to their defined host ranges in this study. E-LM12 was categorized as a narrow host range endolysin, lysing only *S. aureus* species (Costa et al., 2020; Melo et al., 2018). LysWMY was defined as a wide host range endolysin, as it was able to lyse across genus (*Staphylococcus* sp, *Lactobacillus* sp, *Pediococcus* sp, *Micrococcus* sp) (Yokoi et al., 2005). Pairwise sequence alignment between the SH3 domains of the reference endolysins and Endo88 was performed to identify the binding residues on the reference endolysin.

E-LM12 shares 46.8% identity and 56.4% similarity with Endo88, while LysWMY shares 41.5% identity and 56.4% similarity. These identified binding residues were subsequently substituted into Endo88 (YP_240699) at the corresponding amino acid positions (**D1:** K391Q, Y392F, I415G; **D2:** K391E, Y396W, I415T, E437T, Y457W). (Figure 5.14)

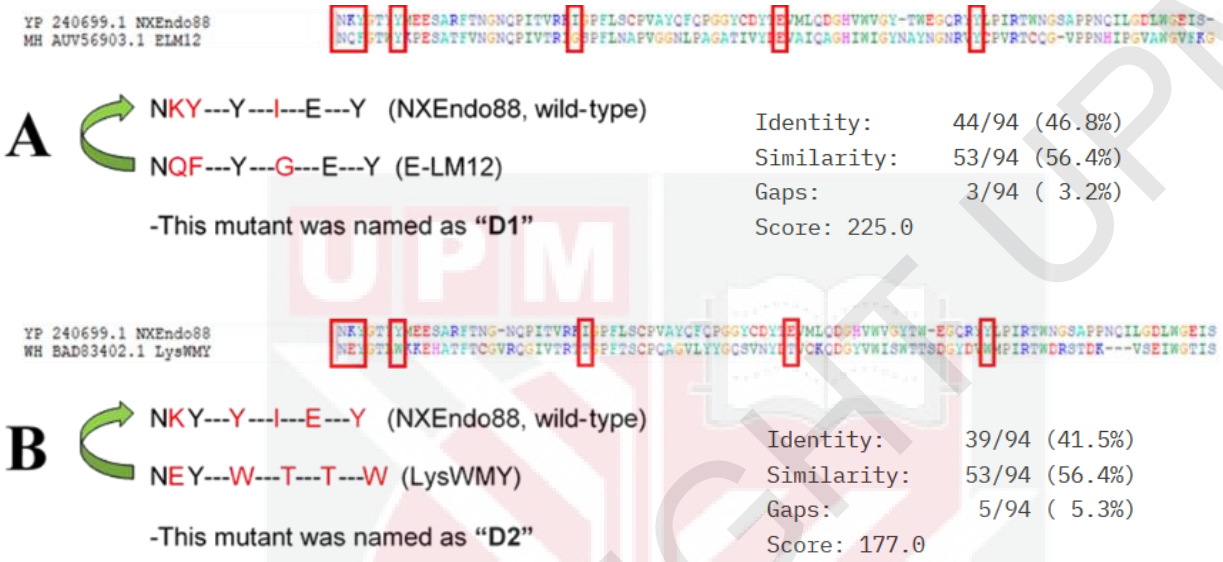


Figure 5.14: Needle-Wunsch pairwise alignment between Endo88 with E-LM12, as well as between Endo88 and LysWMY. The amino acid involved in binding were replaced in the protein sequence of Endo88 at the corresponding amino acid positions. Red box: Amino acids that are involves in binding. **(A)** Pairwise alignment between Endo88 and ELM12. **(B)** Pairwise alignment between Endo88 and LysWMY.

5.3.4 Construction of mutant gene

The optimal annealing temperature for the primer sets (Table 5.1) used to generate the mutants were determined using the method described in section 4.3.5.1. For mutant D1, the primer set produced three region fragments with expected sizes: Region 1: 1202 bp (lane 1 – 3), Region 2: 96 bp (lane 4 – 6), and Region 3: 212 bp (lane 7 – 9) (Figure 5.15). Lane 10 – 12 is the negative control for each region, primer dimers can be seen at the bottom of the gel. For D2, the initial approach involved separating it into five regions using five different primer sets. However, region 2 (75 bp), 3 (87 bp), 4 (83 bp), 5 (99 bp) failed to amplify in gradient PCR (Figure 5.16A). Consequently, the sequences of regions 2, 3, 4, 5 were combined and synthesized, denoted as region 2A (Figure 5.16B). The sequence of region 2A (283 bp) was then joined with region 1 (1200 bp). After gel purification of each region, all regions were annealed together and amplified using NXEndo88 primers (Figure 5.17).

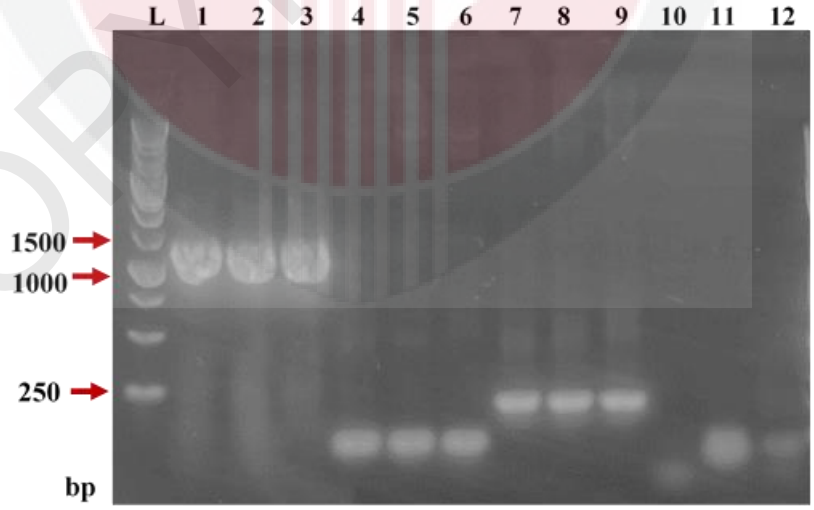


Figure 5.15: Gradient PCR to determine optimum annealing temperature for D1 mutant's primer sets. L: 1 kb DNA ladder (1st Base, Malaysia); Lane 1 to 3: PCR product of Region 1 (~1202 bp) at annealing temperature 55°C, 60°C & 65°C; Lane 4 to 6: PCR product of Region 2 (96 bp) at annealing temperature 55°C, 60°C & 65°C; Lane 7 to 9: PCR product of Region 3 (~212 bp) at annealing temperature 55°C, 60°C & 65°C; Lane 10 to 12: PCR master mix without template as negative control.

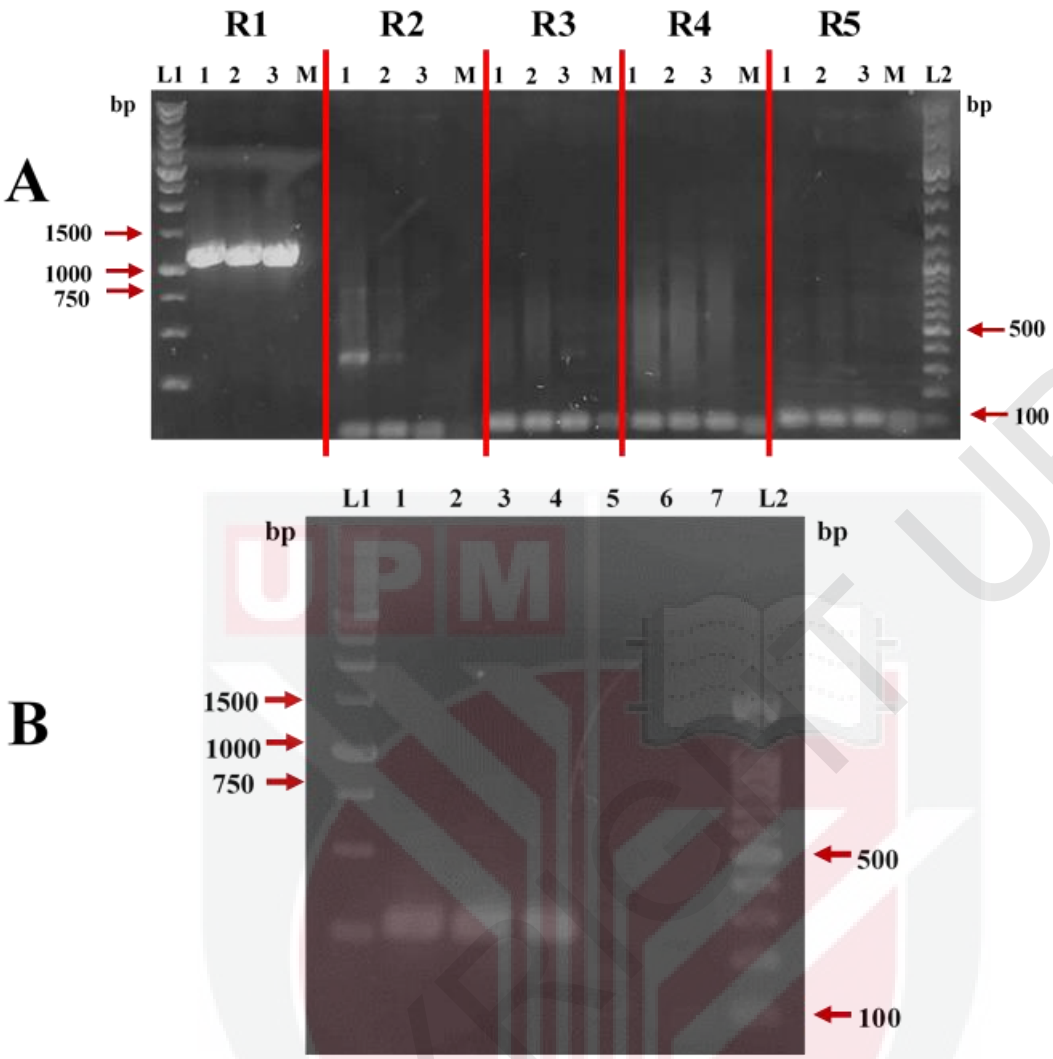


Figure 5.16: Gradient PCR to determine optimum annealing temperature for D2 mutant primer sets. (A): Region 1 (R1, 1200 bp) was successfully amplified at three different annealing temperatures, while no expected bands were observed for R2, R3, R4, and R5. Lane L1: 1 kb DNA ladder (1st Base, Malaysia); Lane L2: 100 bp ladder (GeneRuler, Thermo); Lane 1 to 3: PCR products at annealing temperature of 55°C, 60°C & 65°C; Lane M: Negative control. **(B):** Gradient PCR of region 2A (283 bp) at different annealing temperature. Lane 1 to 3: PCR products at annealing temperatures of 55°C, 60°C & 65°C; Lane 6: Negative control; Lane L1: 1 kb DNA ladder (1st Base, Malaysia); Lane L2: 100 bp ladder (GeneRuler, Thermo).

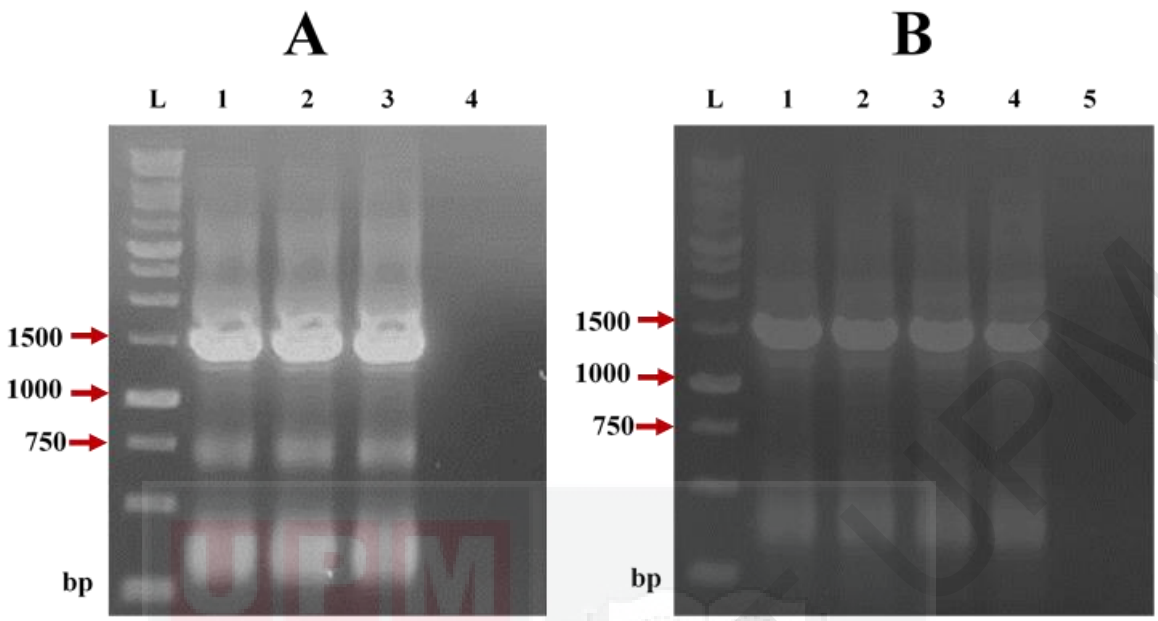


Figure 5.17: PCR product (incorporated with *Nco*I & *Xho*I) of the mutants (*D1* & *D2*) after annealing of each regions. (A): Mutant D1 (~1466 bp). Lane L: 1 kb DNA ladder (1st Base, Malaysia); Lane 1 – 3: D1 PCR product; Lane 4: Negative control. **(B):** Mutant D2 (~1466 bp). Lane L: 1 kb DNA ladder (1st Base, Malaysia); Lane 1 – 4: D2 PCR product; Lane 5: Negative control.

The mutant genes (D1 and D2) were then cloned and transformed into *E. coli* BL21(DE3), following the procedure outlined for cloning *NXEndo88* (Section 4.2.1.3). As shown in Figure 5.18, colonies containing the expected recombinant plasmid exhibited a band size slightly above 1.5 kb on the ladder. The mutations were subsequently verified using Sanger DNA sequencing (Appendix A).

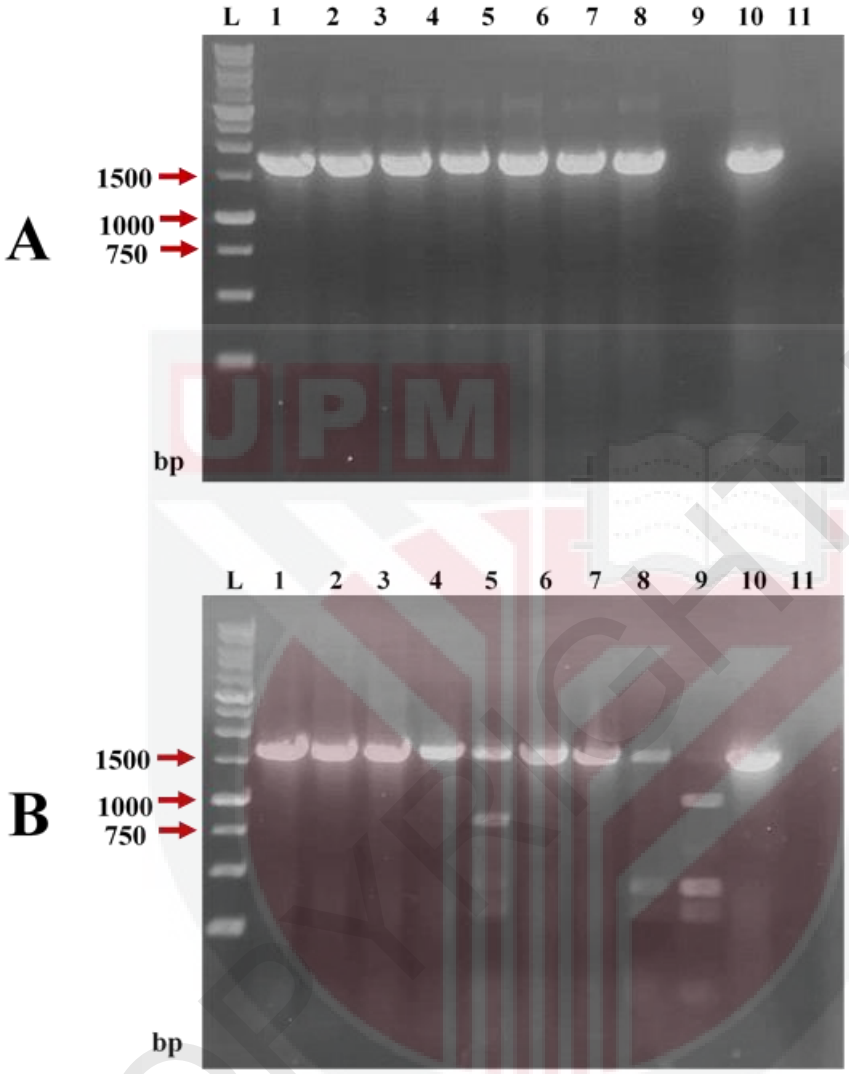


Figure 5.18: Colony PCR screening of recombinant plasmid (pET28a) harbouring mutant gene (*D1* & *D2*). Lane L: 1 kb DNA ladder (1st Base, Malaysia); Lane 1 – 10: Colony PCR screening. **(A):** pET28a (*D1*) (~1648 bp), **(B):** pET28a (*D2*) (~1648 bp). Transformants harboring the recombinant plasmid exhibit band size slightly larger than the insert (above 1.5 kb band size in ladder).

5.3.5 Construction of GFP fused SH3 domain

Two different primer sets were used to generate two types of GFP-fused SH3: *GFP-SH3*, with GFP fused at the N-terminal, and *SH3-GFP* with GFP fused at C-terminal. An annealing temperature of 60 °C was determined for both primer sets by use of gradient PCR (Figure 5.19).

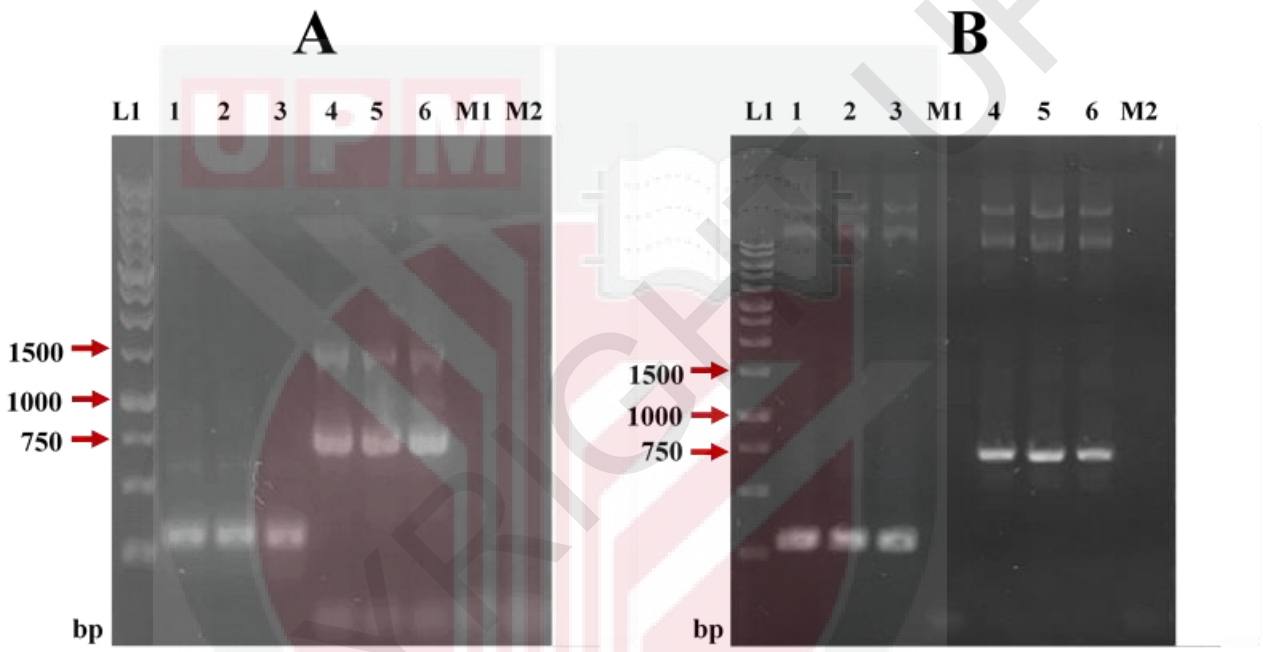


Figure 5.19: Gradient PCR to determine optimum annealing temperature for each primer sets. Primer sets incorporated with different RE were used to amplify *SH3* and *GFP* gene. Gradient PCR was used to determine the optimum annealing temperature for each primer sets. Lane L1: 1 kb DNA ladder (1st Base, Malaysia); Lane M1: Negative control of SH3; Lane M2: Negative control of GFP. **(A):** Primer set for SH3GFP. Lane 1 – 3: PCR product (~318 bp) of SH3 at annealing temperature 55°C, 60°C & 65°C; Lane 4 to 6: PCR product (~740 bp) of GFP at annealing temperature 55°C, 60°C & 65°C. **(B):** Primer set for GFPSH3. Lane 1 – 3: PCR product (~317 bp) of SH3 at annealing temperature 55°C, 60°C & 65°C; Lane 4 to 6: PCR product (~741 bp) of GFP at annealing temperature 55°C, 60°C & 65°C.

5.3.5.1 Construction of recombinant plasmid pET28a encoding for *GFP-SH3*

The GFP gene was first cloned and ligated into pET28a using *Nco*I and *Bam*HI restriction sites and subsequently transformed into *E. coli* BL21(DE3). As depicted in Figure 5.20A, colonies harbouring the expected recombinant plasmid displayed a band size of approximately 962 bp. The GFP sequence in the plasmid was verified through Sanger DNA sequencing (Appendix A). Subsequently, pET28a (*GFP*) and SH3 were digested and ligated at the *Bam*HI and *Xho*I restriction sites, resulting in pET28a (*GFP-SH3*). This construct was then cloned into BL21(DE3), and the colonies harbouring the expected recombinant plasmid were confirmed with colony PCR screening (~ 1228 bp) (Figure 5.20B) and Sanger DNA sequencing (Appendix A).

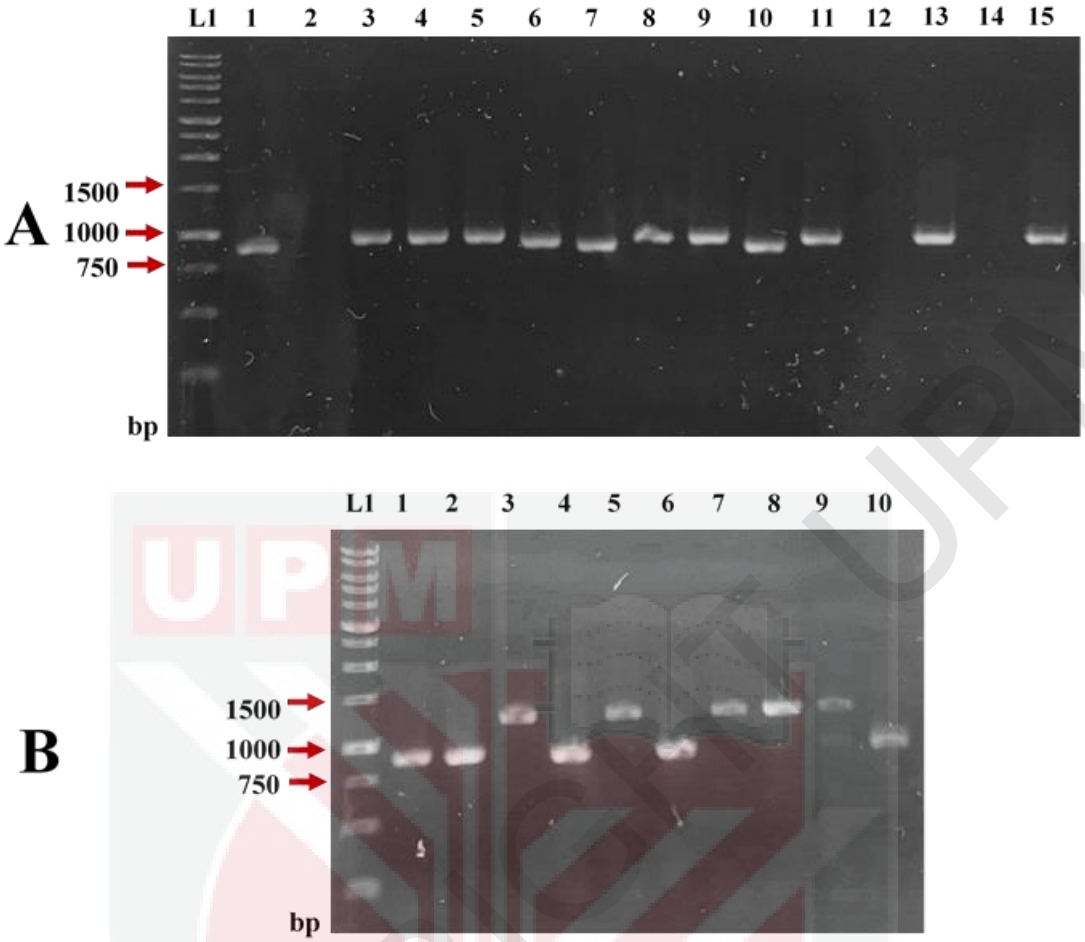


Figure 5.20: Colony PCR screening of recombinant plasmid. (A): pET28a + GFP. Lane L1: 1 kb DNA ladder (1st Base, Malaysia); Lane 1 – 15: colony PCR screening. Recombinant plasmid harboring the expected GFP gene displayed band size approximately 962 bp. **(B):** pET28a (*GFP-SH3*). Lane 1 – 10: colony PCR screening. Recombinant plasmid harboring the expected GFP gene displayed band size approximately 1228 bp.

5.3.5.2 Construction of recombinant plasmid pET28a encoding for *SH3-GFP*.

The DNA fragments – *Nco*I-*SH3*-*Bam*HI and *Bam*HI-*GFP*-*Xho*I generated in section 5.3.5 were double-digested and ligated, resulted in *SH3-GFP* (1046 bp). Conventional PCR was performed to amplify *SH3-GFP* (Figure 5.21A). The *SH3-GFP* was then cloned and ligated into pET28a using *Nco*I and *Xho*I, and subsequently transformed into *E. coli* BL21(DE3). As depicted in Figure 5.21B, colonies harbouring the expected

recombinant plasmid displayed band size of approximately 1228 bp. Subsequent verification was then conducted through Sanger DNA sequencing (Appendix A).

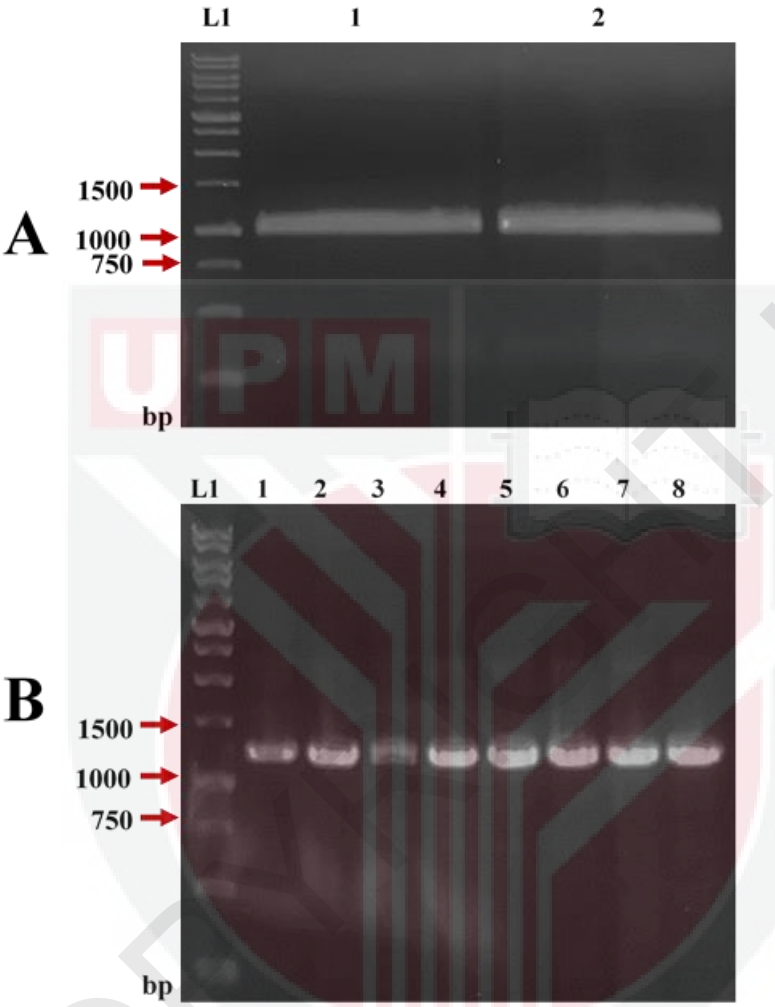


Figure 5.21: Cloning of *SH3GFP* into BL21(DE3). Lane L1: 1 kb DNA ladder (1st Base, Malaysia). **(A):** Lane 1 – 2: PCR of ligation product (*SH3-GFP*) (~1046 bp). **(B):** pET28a (*SH3-GFP*). Lane 1 – 10: colony PCR screening. Recombinant plasmid harboring the expected GFP gene displayed band size approximately 1228 bp.

5.3.5.3 Construction of recombinant plasmid pET28a encoding for *GFP-SH*3 with SH3 from the mutants D1 and D2

Flow cytometry binding assay showed no significant differences between SH3GFP and GFPSH3, and the results of these binding assays will be presented later in section 5.3.8.1. Therefore, the GFP-SH3 method (GFP fused at the N-terminal) was chosen for cloning the mutant SH3 domains. The SH3a-Fwd and SH3a-Rev primers (Table 5.1) was used to amplify the mutated SH3 domain from D1 and D2 endolysin, resulting in D1SH3 and D2SH3 (Figure 5.22A). Using the *GFP-SH3* method (section 5.3.5.1), the SH3 domains were then fused with pET28a and transformed into BL21(DE3). As depicted in Figure 5.22B (D1SH3 and D2SH3), colonies containing the expected recombinant plasmid exhibited a band size of approximately 1228 bp. The mutations were subsequently verified using Sanger DNA sequencing (Appendix A).

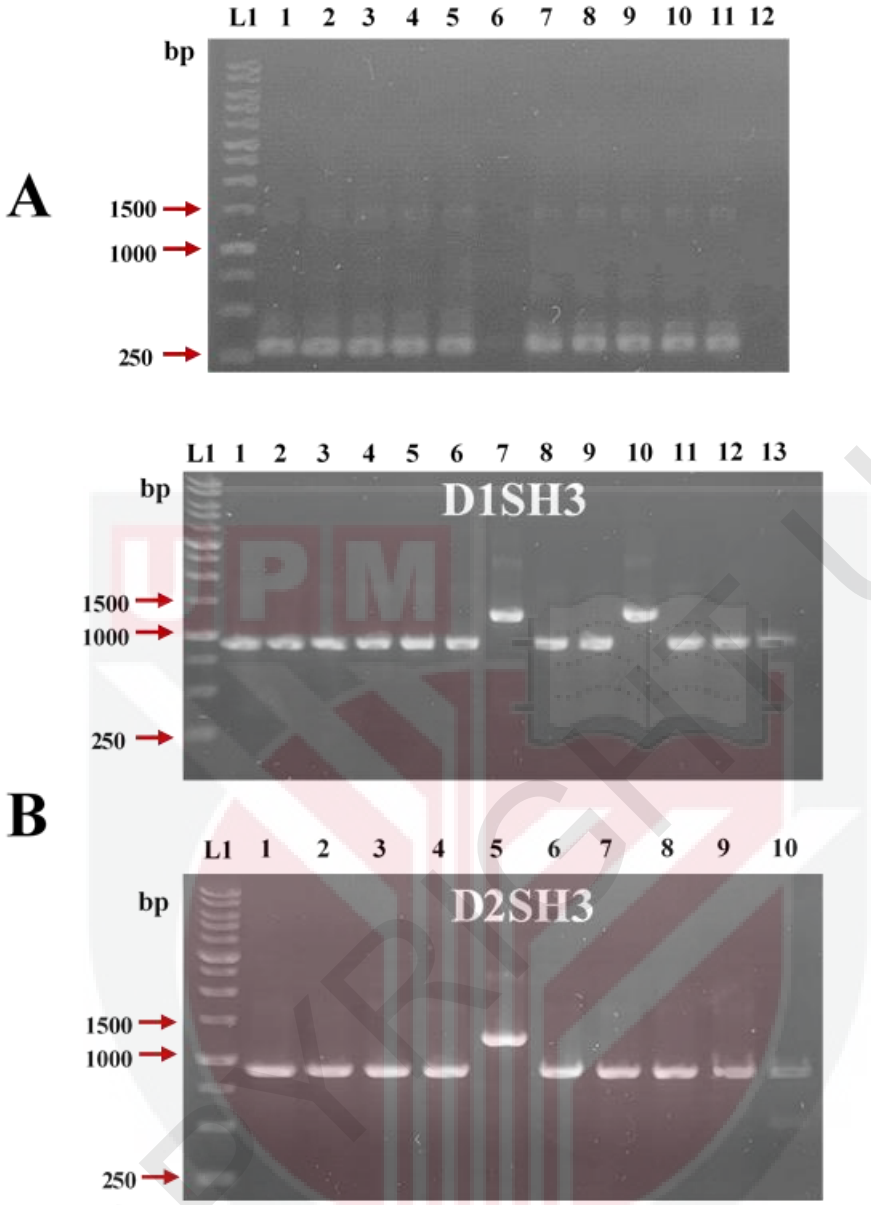


Figure 5.22: Cloning of GFP-fused SH3 mutants. Lane L1: 1 kb DNA ladder (1st Base, Malaysia). **(A):** Lane 1 – 5: PCR product of D1SH3 (~317 bp); Lane 6: Negative control (Master mix); Lane 7 – 11: PCR product of D2SH3 (~317 bp); Lane 12: Negative control (Master mix)**. (B):** pET28a (*GFP-SH3*). Recombinant plasmid harboring the expected GFP gene displayed band size approximately 1228 bp.

5.3.6 Expression and purification of mutant endolysins and SH3 domains

All the mutants were successfully induced and purified under same conditions as with Endo88 (section 4.2.2.2). As depicted in Figure 5.23, D1 and D2 endolysins were successfully expressed in the soluble fraction and purified using his-tag purification. Similarily, the GFPSH3 and the mutants were induced and purified under the same condition (Figure 5.24).

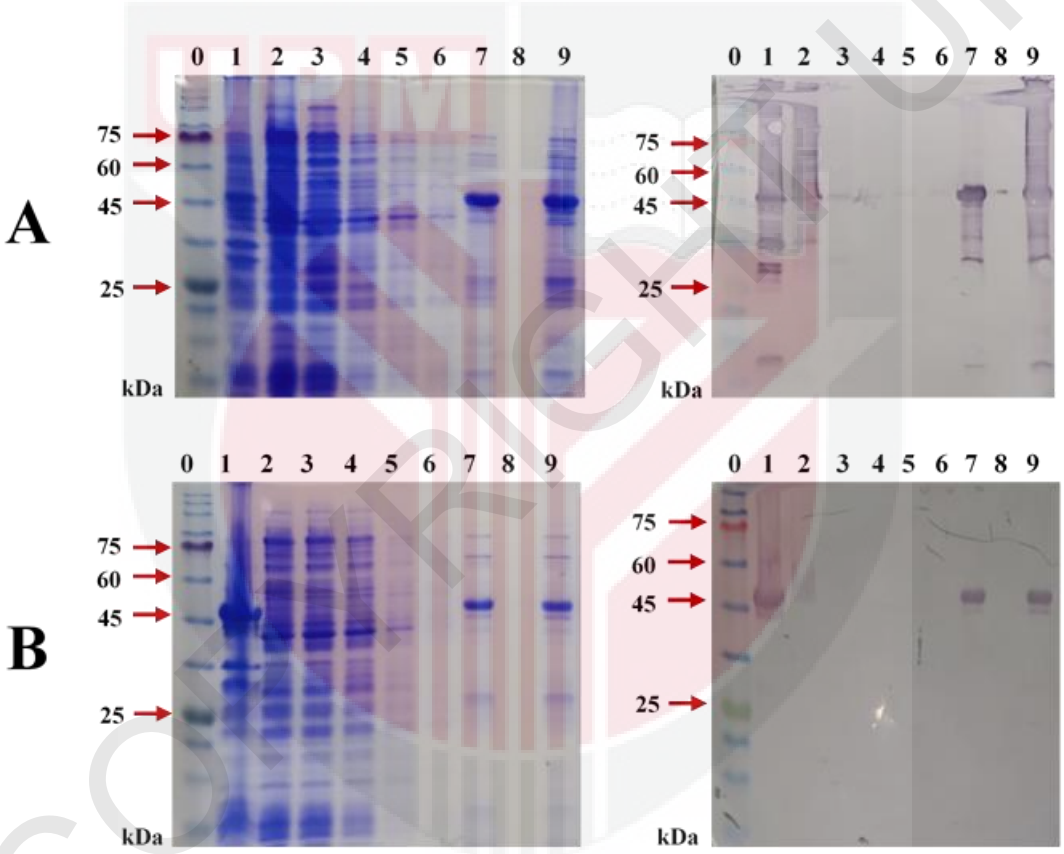


Figure 5.23: SDS-PAGE gel and Western Blot of mutants D1 and D2 purified under native conditions. (A): Expression of Mutant D1 (~55.4 kDa) and purified under native condition. **(B):** Expression of Mutant D2 (~55.4 kDa) and purified under native condition. Lane 0: 5 – 245 kDa protein ladder (1st Base, Malaysia); Lane 1: Pellets; Lane 2: Soluble fraction; Lane 3: Flow through; Lane 4 - 6: Wash 1 – 3; Lane 7: Elution; Lane 9: Concentrated and buffer exchanged protein.

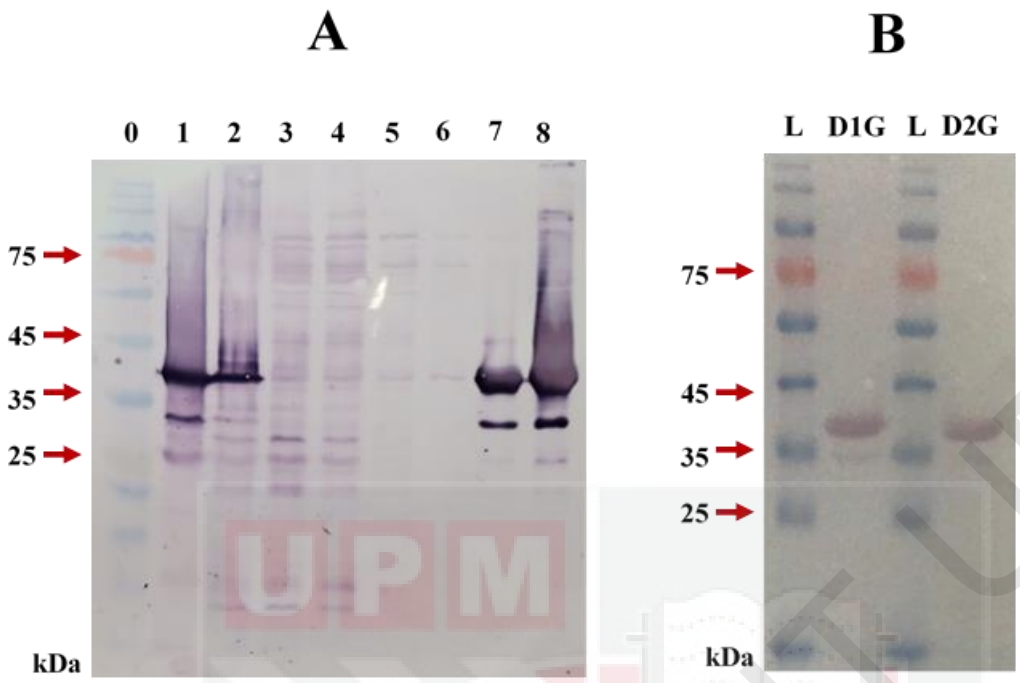


Figure 5.24: Western Blot of Endo88_GFP_SH3, D1_GFP_SH3, and D2_GFP_SH3. GFP-fused SH3 domain with mutants were expressed and purified under native condition (~39.1 kDa). **(A):** Endo88_GFP_SH3 purification under native condition. 0: 5 – 245 kDa protein ladder (1st Base, Malaysia); 1: Pellets; 2: Soluble fraction; 3: Flow through; 4: Wash 1; 5: Wash 2; 6: Wash 3; 7: Elution; 8: Concentrated protein. **(B):** Western Blot of purified and concentrated protein (~39.1 kDa). L: 5 – 245 kDa protein ladder; D1G: D1_GFP_SH3; D2G: D2_GFP_SH3.

5.3.7 Assessment of mutant (D1 & D2) lytic host range

Mutants D1 and D2 were tested against the same set of bacteria previously tested with Endo88 and other mutants (M and A2). Despite the anticipation for a change in host range, the mutants did not exhibit any alteration. However, these mutations similarly resulted in decreased lytic activity (Figure 5.25). Both mutatns showed significantly lower lytic activity compared to the wild-type, except for D1 against *S. hominis*, which showed no significant difference from the wild-type Endo88. Notably, D2 exhibited the lowest lytic activity against all bacteria tested.

LysWMY, which served as the mutation reference for D2, had its lytic host range experimentally tested in its original study (Yokoi et al., 2005). Lytic activity was observed on various bacterial genera, including *S. warneri, S.xylosus, S. simulans, S. epidermidis, S. hycus, S. haemolyticus, Lactobacillus gasseri, L. plantarum, Pediococcus pentosaceus*, and *Micrococcus luteus* (Yokoi et al., 2005). While D2 was similarly able to lyse all *Staphylococcus* spp. tested in the current study, it was unable to lyse *L. plantarum* & *M. luteus*. On the other hand, E-LM12, which served as a narrow host range reference for D1, was previously reported to be unable to exhibit lytic activity against *S. epidermidis, E. faecalis* and *S. hominis* (Costa et al., 2020), which were also tested in the current study. In contrast to E-LM12, D1 was still able to lyse these bacteria, similar to the wild-type Endo88. These observations indicated that mutations in the native CBD mainly affect the lytic activity of the endolysin without significantly altering its host range.

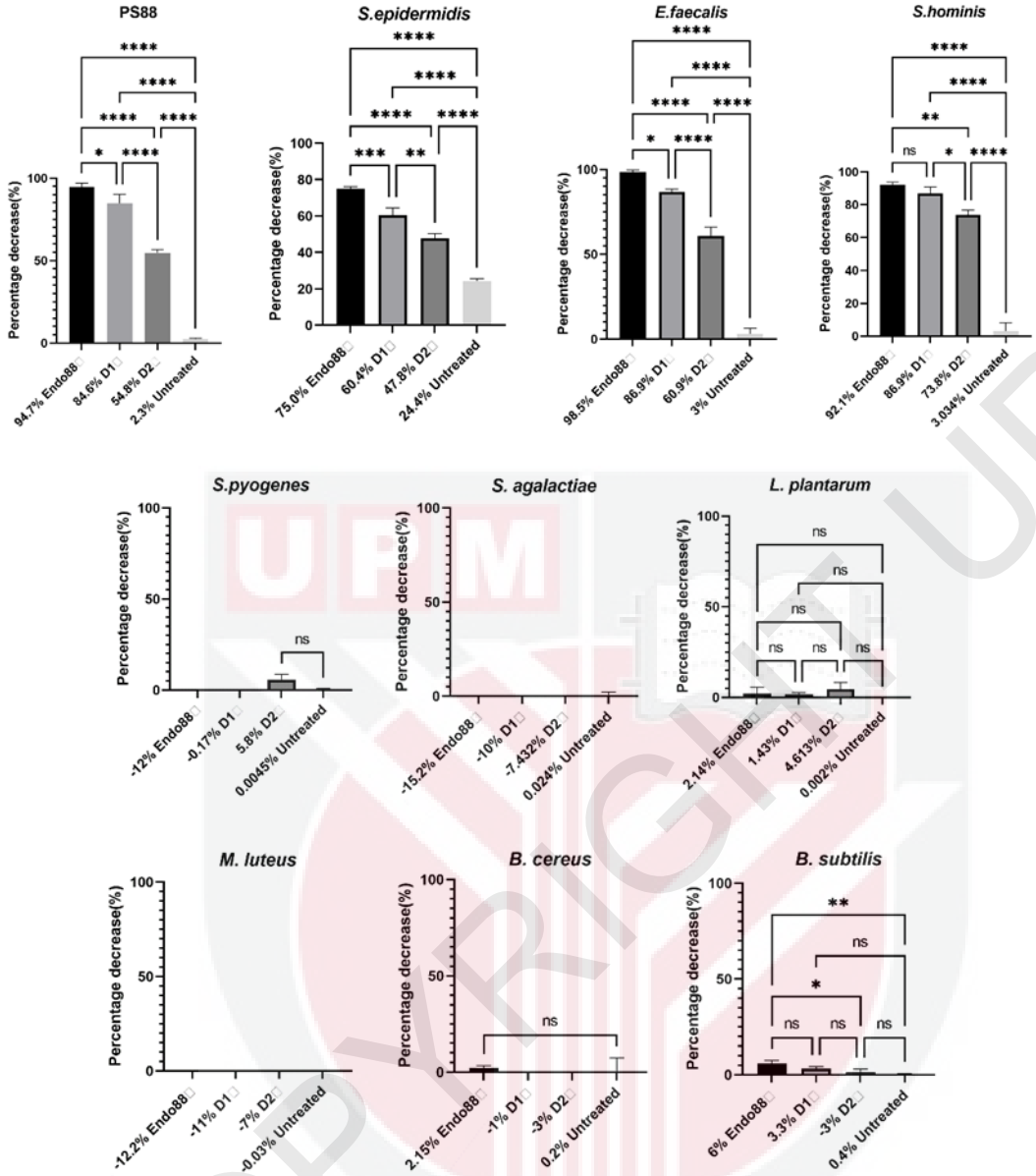


Figure 5.25: Turbidity reduction assay of wildtype Endo88 and mutants against various bacteria. The Endo88 and mutants (D1 and D2) were able to lysed *S. aureus* PS88, *S. epidermidis*, *S. hominis*, and *E. faecalis* with different degree of lytic activity. Untreated: Bacteria cell without endolysin treatment as negative control; "*" indicates significance with $p < 0.05$; "**": $p < 0.01$; "***": $p < 0.001$; $p < 0.0001$ and ns: not significant.

5.3.8 Flow cytometry binding assay with GFP-fused SH3 domains

In previous studies, there have been contradictory statements about the role of the CBD in lytic activity. Some studies have reported increased lytic activity after CBD removal (Cheng & Fischetti, 2007; Gaeng et al., 2000; Low et al., 2005; Schmelcher, Korobova, et al., 2012; Yokoi et al., 2005), while others have observed decreased efficiency after CBD removal (Abaev et al., 2013; S. C. Becker et al., 2015; S. C. Becker, Foster-Frey, et al., 2009; Donovan, Lardeo, et al., 2006; Fujiki et al., 2018; Huang et al., 2015; Linden et al., 2015). The precise mechanism by which the CBD confers specificity to endolysins remains uncertain, but it definitely possesses evolutionary impotance as it is nearly ubiquitous in all Gram-positive endolysins (Oechslin et al., 2022). Since the mutants in the current study did not alter the host-range which was tested through its ability to lyse a certain bacteria, we next studied the binding activity independent from the lytic activity to determine if binding and lytic activity was correlated. To do this, the SH3 domains of Endo88, as well as D1 and D2 mutant endolysins were fused to GFP without the EAD domains.

5.3.8.1 Effect of location of GFP at N- & C-terminal of the SH3 domains

SH3GFP and GFPSH3 were incubated separately with *S. aureus* PS88 and *E. coli*. The binding of the GFP-fused SH3 domain to the bacteria was then detected using FIT-C filter in flow cytometry. Both SH3GFP and GFPSH3 achieved above 90% fluorescence detection, while no fluorescence was detected for *E. coli* (Figure 5.26). This result indicates that the SH3 domain successfully binds to PS88 and does not bind to *E. coli*. There was no significant difference between SH3GFP and GFPSH3 constructs. Consequently, the N-terminal GFP construct was chosen for the mutant SH3 domains.

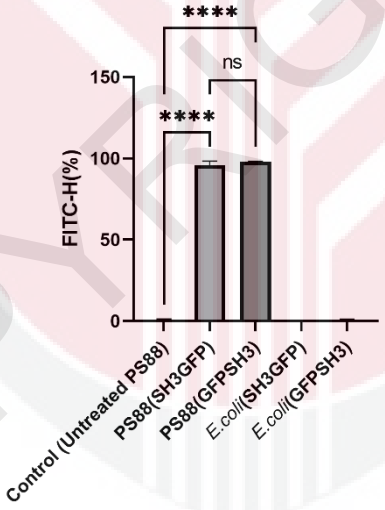


Figure 5.26: Comparison of C-terminal fused GFP construct (SH3GFP) and N-terminal fused GFP (GFPSH3). Percentage mean of fluorescence intensity over cell population were analyzed by flow cytometry. Untreated PS88 and *E. coli* BL21(DE3) were used as negative control. "*" indicates significance with $p < 0.05$; "**": $p < 0.01$; "***": $p < 0.001$; "****": $p < 0.0001$ and ns: not significant.

5.3.8.2 Binding assay of GFPSH3 and mutant SH3

To investigate the correlation between binding activity and lytic activity, the SH3 domains of Endo88 and its D1 and D2 mutant endolysins were fused to pET28a (*GFP*), with GFP fused at N-terminal. This resulted in Endo88_GFPSH3, D1_GFPSH3 and D2_GFPSH3. The same set of bacteria used in the turbidity reduction assay was employed to assess the binding activity of these GFP-fused SH3 domains.

Both Endo88_GFPSH3 and D1_GFPSH3 demonstrated high binding activity (>90%) on PS88 (Figure 5.27). Interestingly, despite the D2 full endolysin being able to lyse PS88 as shown previously in Figure 4.24 (section 4.3.5.3), albeit less effectively than Endo88 and D1, D2_GFPSH3 exhibited minimal binding activity (<5%) on PS88 (Figure 5.27).

An unexpected observation was made with *S. epidermidis*, where all the GFP-fused SH3 domains showed almost no binding activity (<2%). Nevertheless, Endo88, D1 and D2 full endolysins were able to lyse *S. epidermidis* effectively, as indicated by the turbidiy reduction assay. Interestingly, in one of our recent study (Krishnan et al., 2024) the truncation effect on Endo88, it was observed that SH3b reinforced the lytic activity of CHAP on *S. aureus*, but not on *S. epidermidis*. This indicated that the binding of SH3 may not be favoured on *S.epidermidis.* As per the current study, the SH3 domain of Endo88 was unable to bind *S. epidermidis* (Figure 5.27), and the lytic activity was reduced to below 80% (Figure 5.25, section 5.3.7) as compared to its lysis effect against *S. aureus.*

Conversely, the binding activity on *E. faecalis* and *S. hominis* showed similar patterns to the lytic activity observed in the turbidity reduction assay, with D2_GFPSH3 displaying the lowest binding activity among the SH3 domains, consistent with D2 exhibiting the lowest lytic activity. It is also noteworthy that D1 and D2 showed low fluorescence level (<20%) in *B. subtilis*, although no lytic activity was shown in the TRA (Figure 5.25, section 5.3.7).

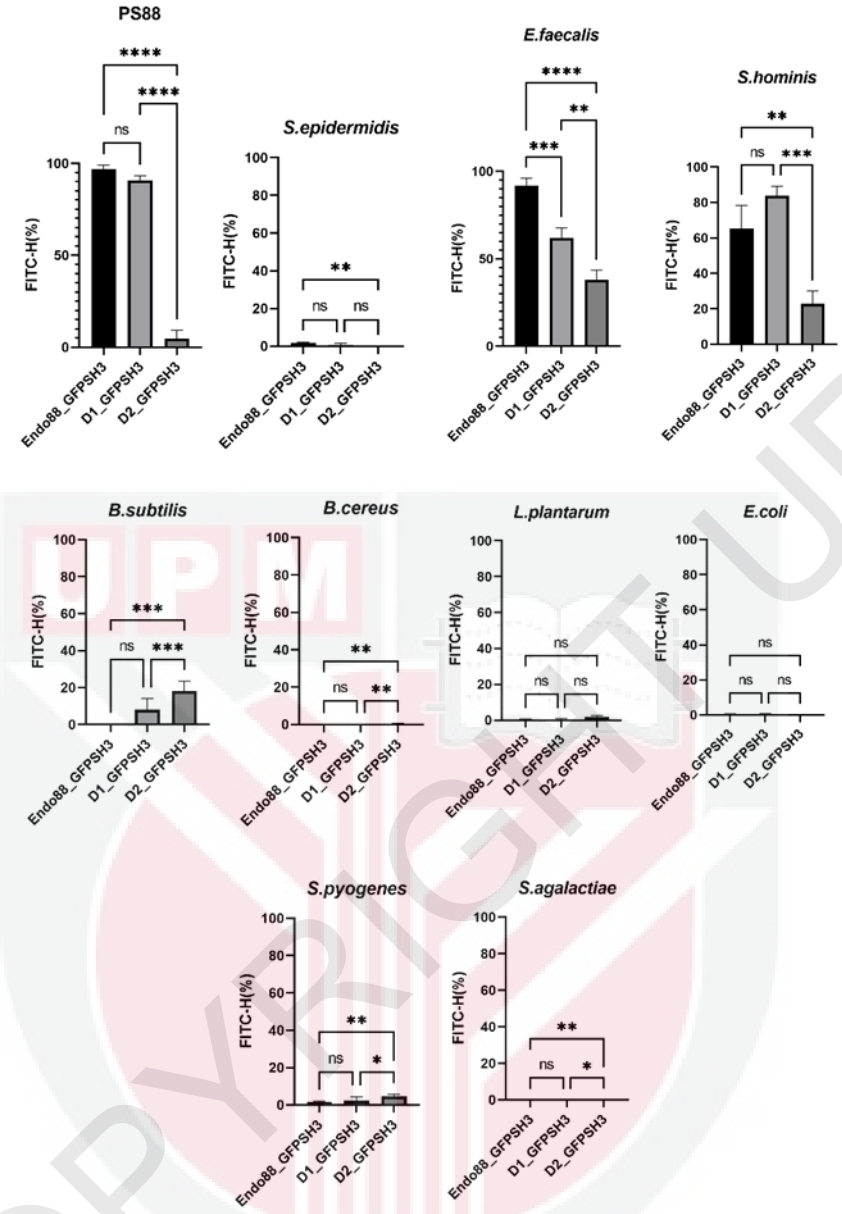


Figure 5.27: Percentage mean of fluorescence intensity over cell population analyzed by flow cytomtry. *E. coli* was used as negative control. The GFP-fused SH3 domains were incubated with different genus of bacteria and analyzed by flow cytometry. "*" indicates significance with $p < 0.05$; "**": $p < 0.01$; "***": $p < 0.001$; "****": $p < 0.0001$ and ns: not significant.

5.3.9 Molecular docking

Molecular docking tools are essential for studying the interactions between proteins and ligands. One of the key metrics for evaluating the compatibility of those tools is binding site consistency, which refers to the ability of the docking tool to reproduce a binding pose similar or identical to the experimentally proven binding pose (Agu et al., 2023; C et al., 2022). In this study, two molecular docking tools- Autodock4 and High Ambiguity Driven Protein-protein Docking (HADDOCK) were compared by re-docking ligands into the binding site of the experimentally solved crystal structure (PDB: 6RK4).

5.3.9.1 Comparison of docking tools - Autodock4

Two types of docking approaches were employed in this study. Firstly, blind docking was performed where the three-dimentional lattice grid covered the entire SH3 domain. A total of 100 runs were clustered into 49 clusters, and the top 10 poses with the lowest mean binding energy (ranging from -1.00 to -2.90 kCal/mole) were chosen and superimposed with 6RK4 crystal structure (Figure 5.28). However, none of the clusters had poses similar to 6RK4.

The docking was then repeated with the three-dimension lattice grid covering only the binding groove. A total of 52 clusters were generated, with only one pose from cluster 39 being closest to the native form. Despite this, although the pentaglycine region (interpeptide bridge) was bound into the groove, the peptide stem region was bent downwards (Figure 5.29A), resulting in a 1.942 Å difference from 6RK4. This discrepancy might be due to the high flexibility of lysine, caused by its three flexibile methylene group side chains (Azevedo & Saiardi, 2016; Clark et al., 2019; Kono,

2013). Additionally, manually fixing the torsion at the methylene side chain caused the ligand to deviate from the binding site, resulted in 1.389 Å difference from 6RK4 (Figure 5.29B). Concurrently, HADDOCK was used to compare the docking compatibility with Autodock4.

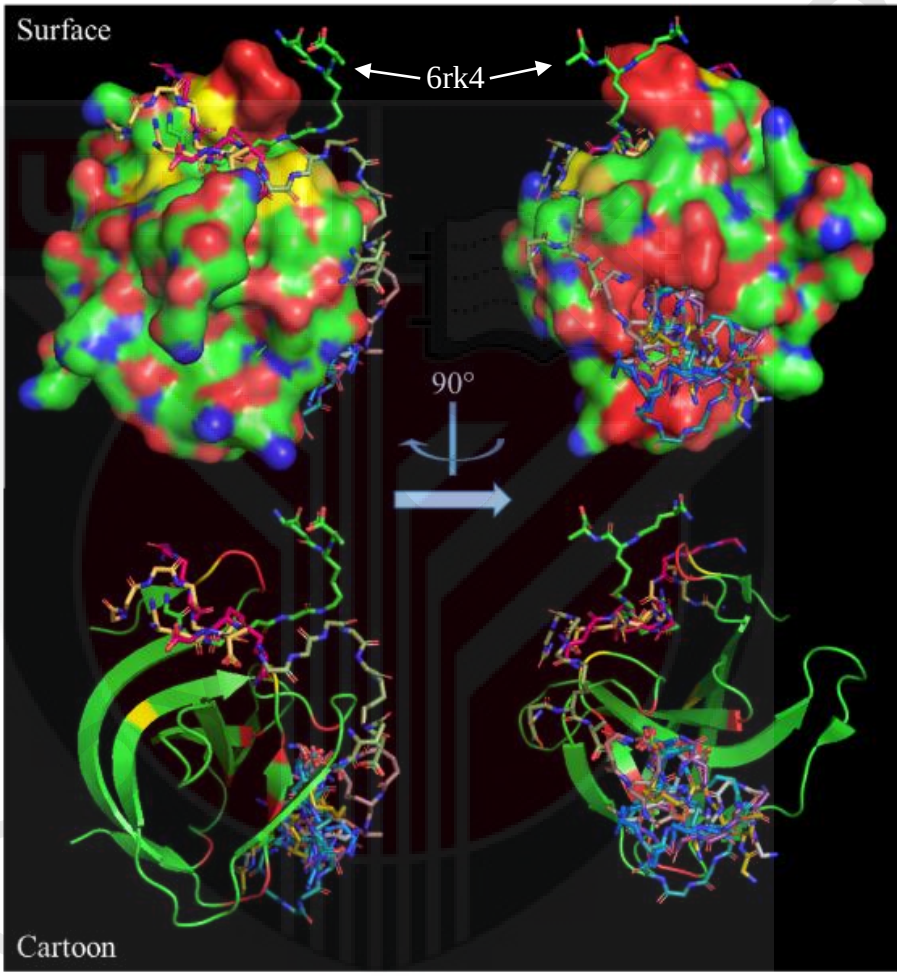


Figure 5.28: Cartoon and Surface representation of 3D docking structure. Top 10 poses were superimposed with the 6RK4 crystal structure (Green), none of the poses were similar to the ligand in 6RK4. Green: 6RK4.

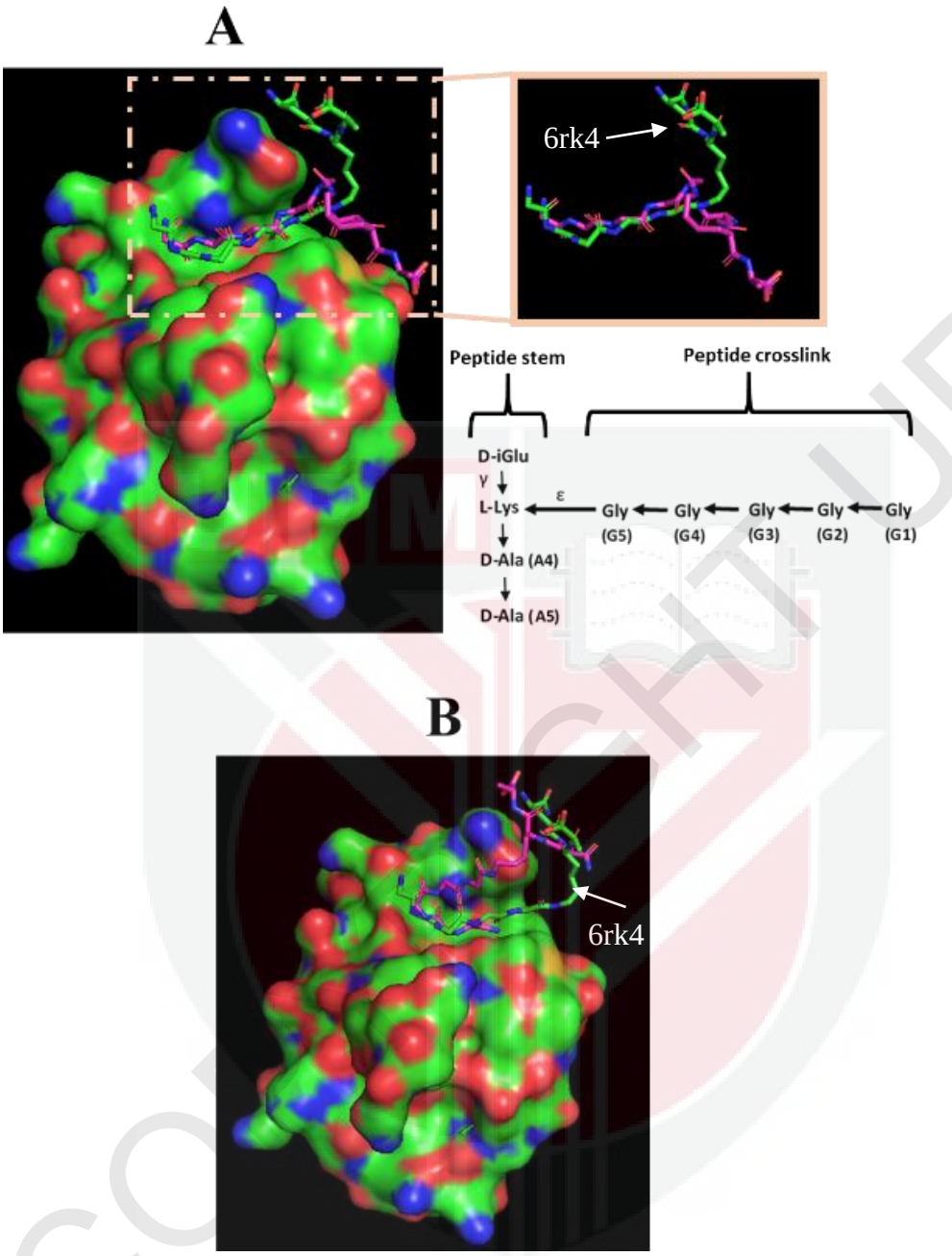


Figure 5.29: Illustration of docking pose superimposed with 6RK4. Green: 6RK4; **(A):** Smaller docking grid caused the ligand to successfully bind into the binding groove, but the peptide stem deviated downwards. **(B):** Fixed torsion at the methylene group of the lysine caused the ligand to deviate from the binding groove.

5.3.9.2 Comparison of docking tools – HADDOCK

HADDOCK was used to re-dock 6RK4 with or without the use of AIR guidance. The AIR are restraints designed to guide the docking process by incorporating experimental data of the protein interactions, specifying the active and passive residues that may or may not be involved in the interaction (Saponaro et al., 2020). Docking without AIR resulted in lesser binding site consistency, while both approaches resulted in the ligands binding in proximity to the binding groove:

(I) With AIR

A total of 309 docking poses were generated and clustered into 28 clusters. Top 10 clusters with the highest HADDOCK scores were chosen for detailed analysis. Among the 10 clusters, 9 clusters were successfully docked into the binding groove, (Figure 5.30A), including the best cluster, which showed a RMSD of 0.354 Å from the 6RK4 structure (Figure 5.30B). Although one cluster did not dock directly into the binding groove, the pentaglycine chain was positioned in proximity to the binding groove.

(II) Without AIR

A total of 247 poses were generated and clustered into 30 clusters. Top 10 clusters with the highest HADDOCK scores were chosen for analysis. The best pose showed a RMSD of 0.357 Å from the 6RK4 structure. Additionally, within the 10 clusters, 6 clusters were successfully docked into the binding groove including the best clusters (Figure 5.30B).

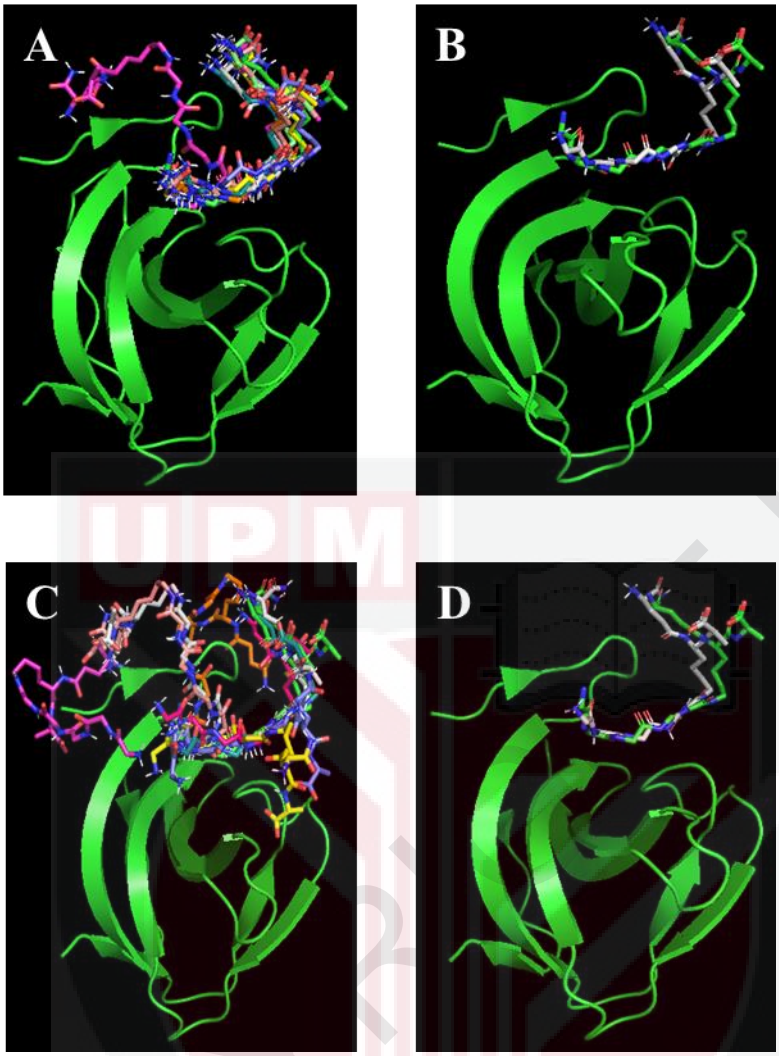


Figure 5.30: Superimposition of docking clusters with 6RK4. (A) Docking with AIR guidance: Top 10 best poses were superimposed with 6RK4; **(B) Docking with AIR:** Best docking pose was superimposed with 6RK4; **(C) Docking without AIR guidance:** Top 10 best poses were superimposed with 6RK4; **(D) Docking without AIR:** Best docking pose superimposed with 6RK4; Green: 6RK4 crystal structure.

5.3.9.3 Molecular docking of SH3 domain with varies type of peptide crosslink

Various peptide crosslink representing different bacterial peptidoglycan types were generated by use of Avogadro 1.2.0. This approach was employed to explore the binding interactions between these peptide crosslink and the SH3 domain, with the aim of correlating these interactions with the results obtained from the binding assays. The study aimed to understand the influence of the mutations on the SH3 domain developed in this study to various types of peptide crosslinks.

To further improve the accuracy of the docking, solvated docking with short molecular dynamic (MD) refinement was incorporated using HADDOCK (de Vries et al., 2010; Dominguez et al., 2003). This approach accounts for the effect of water molecules in the docking environment, which is crucial since protein-ligand interactions occur in aqueous conditions. Solvated docking enhances the reliability of predicted binding conformations in response to solvation effects. Subsequently, short MD refinement further optimizes the binding conformation by allowing side-chain rearrangements and relaxation of the docked complex into a more energetically favourable state. This step is crucial in generating a more physiologically relevant binding conformations, thereby complementing the results obtained from standard docking and binding assay (Bienstock, 2015; Genheden et al., 2011; Mahmoud et al., 2020; van Dijk & Bonvin, 2006).

AIR and UIR were also applied in the docking. AIR are restraints which part/molecules of two proteins might be interacting with some uncertainty, allowing some flexibility in the docking process. In contrast, UIR are more certain with specifically defined distance between atoms interacting with each other (Dominguez et al., 2003). Firstly,

the docking was run with only AIR (Figure 5.31A). If the docking poses generated a high number of docking poses with binding in uncommon binding sites, UIR will be applied to guide the docking. UIR was applied to docking of *S. aureus* (Figure 5.31B) and *E. faecalis* (Figure 5.31C) peptidoglycan.

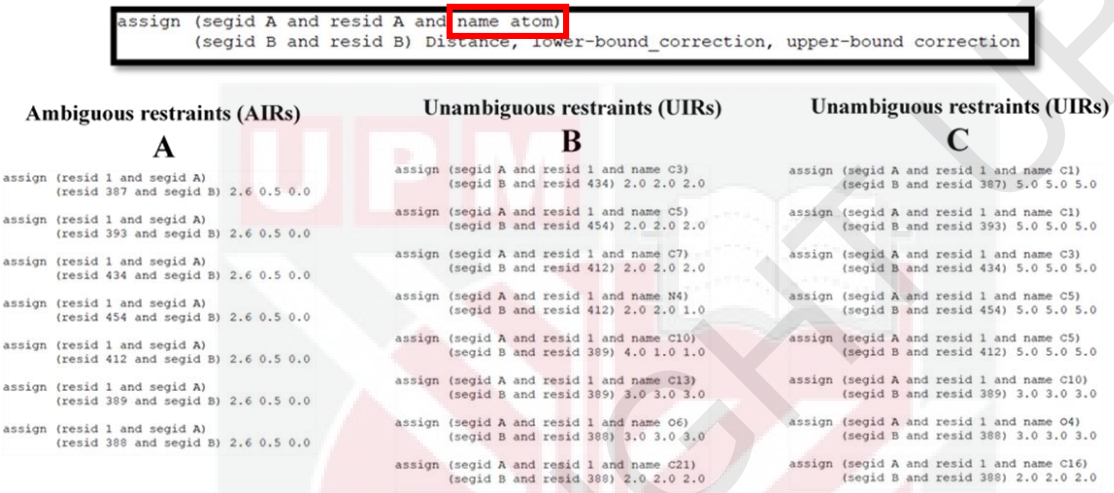


Figure 5.31: Ambiguous restraints (AIRs) and Unambiguous restraints (UIRs) were used for peptide crosslink to guide their movement in the proximity of the binding site. "Name atom" (red box) is for UIR to assign specific amino acid interaction. **(A):** AIRs was used for all peptide-crosslink to guide their movement in the proximity of the binding site. **(B)** UIRs was applied between *S. aureus* peptide cross-link and Endo88_SH3 for optimization, with the distance restraint referenced from the interaction between 6RK4 and its ligand (Figure 5.6). Additionally, the minimum length for hydrogen bond distance was considered (Dannenberg, 1998). **(C):** UIRs was applied between *E. faecalis* peptide cross-link and the SH3 domains.

Molecular docking between the SH3 domains and peptide crosslinks revealed several notable interactions:

(I) *S. aureus* peptide crosslink

As depicted in Figure 5.32, the peptide crosslink of *S. aureus* was generated, comprising a peptide stem (D-isoGlutamine, L-Lysine, D-Alanine, and D-Alanine) and a pentaglycine (interpeptide bridge). These components are typically found in *S. aureus* peptidoglycan (Schleifer & Kandler, 1972; Vollmer et al., 2008). The generated peptide crosslink was then docked into 6RK4 (ligand removed) for validation. The docking poses achieved a RMSD of 0.427 Å (Figure 5.33) when superimposed with 6RK4, and the amino acid interaction profile closely matched that of 6RK4 (Table 5.5). This indicated that the protocol used for generating the peptide crosslink is reliable.

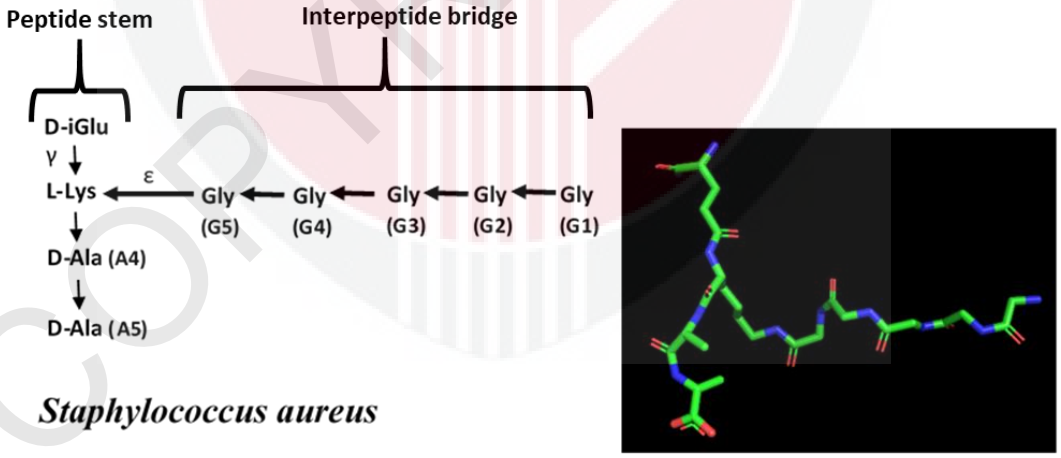


Figure 5.32: Peptide crosslink of *S. aureus* were generated using Avogadro 1.2.0. This structure were created based on references from existing studies (Sharif et al. 2009; Kim, Chang, and Singh 2015; Schleifer and Kandler 1972; Gonzalez-Delgado et al. 2020; Vollmer, Blanot, and de Pedro 2008).

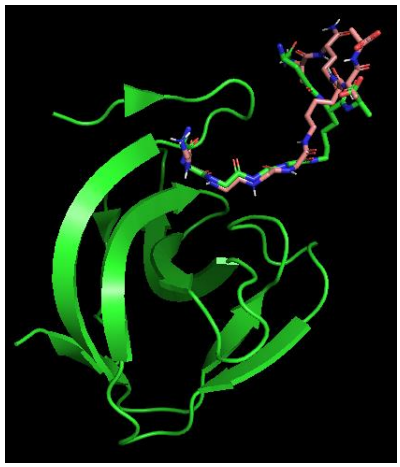


Figure 5.33: Superimposition of docking clusters with 6RK4. Peptide crosslink of *S. aureus* was generated *in silico* and docked with 6RK4 (ligand removed). The final docking pose was then superimposed with 6RK4. Green: 6RK4; Pink: Best docking pose.

Table 5.5: Amino acid interaction profile between the SH3 domain and the peptide crosslink. Bold font: Interaction profile between generated peptide crosslink and 6RK4 (Ligand removed); Red font: Interation profile of 6RK4.

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu **L-Lys** **D-Ala (A4)** **D-Ala (A5)**	Y407 (Y407)	(K406)	
G1		Y411, N405 (Y411), (Y405)	
G2		E451 (E451)	
G3		Y472 (Y472)	
G4		T429 (T429)	
G5			

The generated peptide crosslink was docked with the modeled SH3 domain with the guidance of AIR and UIR (Figure 5.31B), as well as with the mutant SH3 domains. Docking analysis of the Endo88 SH3 domain with the *S. aureus* peptide crosslink revealed that majority of hydrogen bonds formed at the pentaglycine region, supplemented by several hydrophobic interactions and one salt-bridge (Appendix C). Hydrogens bonds are commonly known for their crucial role in molecular recognition and binding affinity (Chen et al., 2016; Wade & Goodford, 1989), which were abundant in the interaction between the peptide crosslink and the SH3 domain. This observation is consistent with findings from other studies (Gu et al., 2014; J. Z. Lu et al., 2006; Mitkowski et al., 2019), suggesting that hydrogen bonds play a significant role in stabilizing the SH3 domain and potentially conferring its specificity.

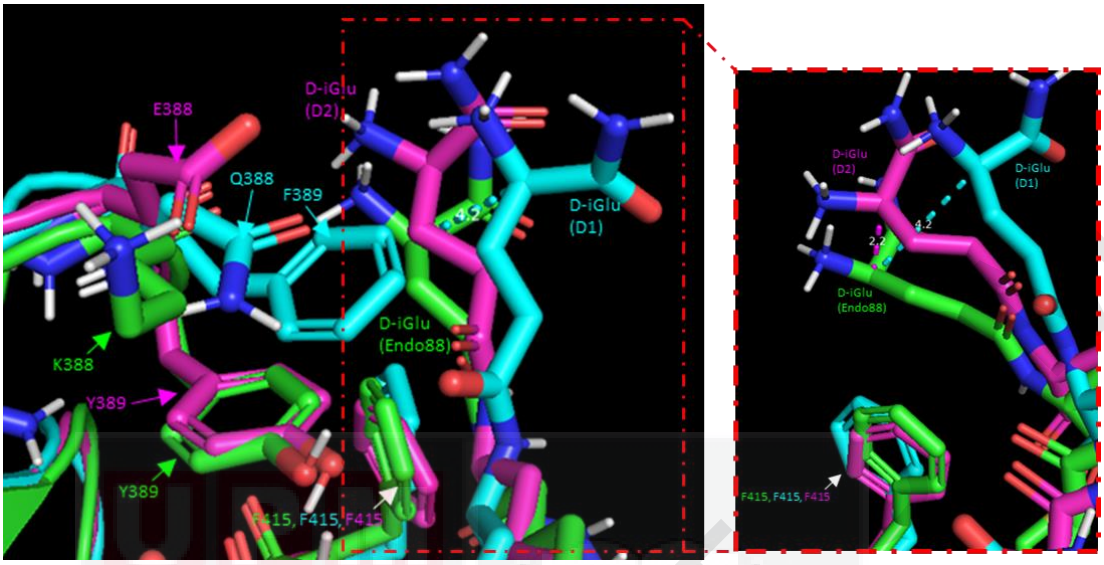
One obvious difference was observed in the docking poses between Endo88_SH3 and its mutants. In Endo88_SH3, the D-iGlu residue in the peptide crosslink formed a hydrophobic interaction with F415. However, mutations (K388Q and Y389F in D1 and K388E in D2) caused the D-iGlu to deviate approximately 2.2 Å for D2 and 4.2 Å for D1 from its original pose in Endo88 (Figure 5.34A). This deviation disrupted the hydrophobic interaction with F415 in both D1 and D2. Despite this noticeable deviation, the overall conformation of the peptide stem in D1 and D2 remained unchanged, the salt-bridge between K411 and D-Ala was preserved. Nevertheless, no binding activities were detected in the flow cytometry assay for D2. Despite having a high HADDOCK score and predicted energy (Table 5.6), D2 formed fewer hydrogen bonds at the pentaglycine region (2 hydrogen bonds) compared to Endo88 (5 hydrogen bonds) and D1 (3 hydrogen bonds) (Appendix C). Furthermore, after subjecting D2 to solvated docking with short MD refinement, no hydrogen bond interactions were

detected at the pentaglycine region, which may explain the lack of binding activity observed in flow cytometry Figure 5.34B.

Overall, the amino acid interaction profile revealed that the mutants exhibited fewer interactions with the *S. aureus* peptide crosslink (Appendix C), and demonstrated lower predicted binding energy in dry docking compared to the Endo88_SH3. These findings suggest that alterations in hydrogen bonds may compromise the stability of the binding interaction.

The decrease in electrostatic energy aligns with the observed pattern, indicating that electrostatic interactions play a role in changes in binding affinity. This reduction in electrostatic energy could be attributed to the substitution of the charged amino acid K388 (positive) in Endo88, with Q388 (polar) in D1 and E388 (negative) in D2. Additionally, the substitution of E434 (negative) with T434 (polar) in D2 may have further contributed to the decrease in electrostatic energy. The increase in molecular volume of the binding groove (**Endo88_SH3:** 45.026 Å; **D1_SH3:** 59.904 Å; **D2_SH3:** 72.704 Å) (Figure 5.35) might also lead to insufficient contacts between the peptide crosslink and the surrounding residues, resulting in decreased binding affinity and consequently reduced lytic activity (Ha et al., 2020; Pang et al., 2018).

A



B

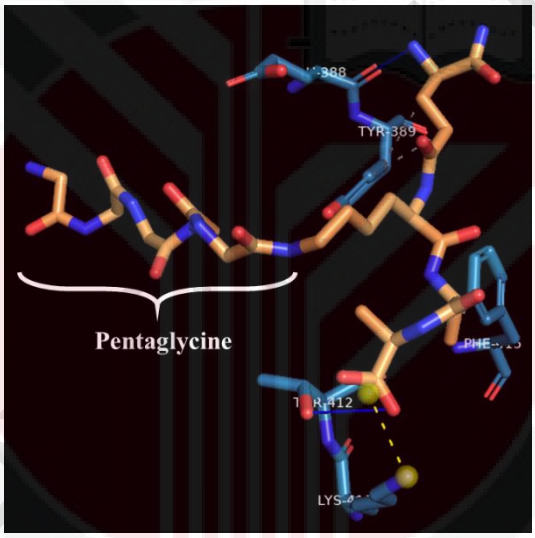
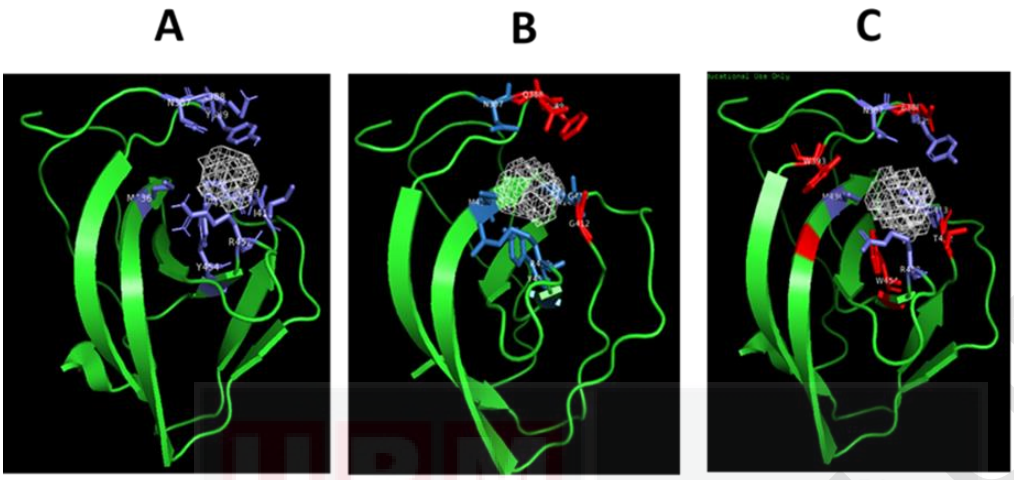


Figure 5.34: Illustration of molecular docking between the *S. aureus* peptide crosslink and the SH3 domains (Endo88, D1 and D2). (A): Superimposition of the best docking poses of the SH3 domains (Endo88, D1 and D2). Notable conformational changes were observed at the D-iGlu in the peptide crosslink. Green: Docking pose of Endo88_SH3 and the peptide crosslink; Magenta: Docking pose of D2_SH3 and the peptide crosslink; Cyan: Docking pose of D1_SH3 and the peptide crosslink. **(B):** Docking pose of D2_SH3 with the *S. aureus* peptide crosslink after subjected to solvated docking and refined with short molecular dynamic. No interactions were observed at the pentaglycine region.



Endo88	D1	D2
Volume [Å³] : 45.026	Volume [Å³] :59.904	Volume [Å³] : 72.704
Surface [Å²] : 181.988	Surface [Å²] : 239.311	Surface [Å²] : 264.759
Depth [Å] : 4.93153	Depth [Å] : 5.71314	Depth [Å] : 5.87878
Hydrophobicity ratio: 0.777778	Hydrophobicity ratio: 0.72	Hydrophobicity ratio: 0.727273
H bond donors : 4	H bond donors : 4	H bond donors : 6
H bond acceptors: 4	H bond acceptors: 3	H bond acceptors: 4
Aromatic atoms : 6	Aromatic atoms : 5	Aromatic atoms : 10

Figure 5.35: Physical properties of the SH3 domains analysed by DoGSite3. The volume of the binding groove increases after mutation in D1 and D2 SH3 domains.

Table 5.6: Molecular docking of the SH3 domain and different type of PG's peptide crosslink

		Dry docking			Solvated Docking + Short MD		
Bacteria PG's peptide crosslink	**SH3 domain**	**HADDOCK score**	**Predicted binding energy (Kcal/mol)**	**Electrostatic energy**	**HADDOCK score**	**Predicted binding energy (Kcal/mol)**	**Electrostatic energy**
PS88	**Endo88**	-47.6	-8.14	-146.73	-65.7	-8.86	-203.84
	D1	-34.4	-7.7	-126.99	-66.0	-7.49	-111.53
	D2	-39.1	-7.56	-103.71	-70.3	-7.72	-127.86
S. epidermidis	**Endo88**	-33.3	-6.99	-69.7	-	-	-
	D1	-37.9	-7.04	-71.11	-	-	-
	D2	-	-	-	-	-	-
E. faecalis	**Endo88**	-37.8	-8.24	-192.98	-45.9	-7.49	-95.72
	D1	-31.9	-7.32	-80.92	-40.9	-7.36	-115.33
	D2	-27.6	-6.96	-59.05	-50.6	-7.36	-85.86
M. luteus	**Endo88**	-45.9	-7.89	-91.03	-55.5	-7.15	-46.92
	D1	-38.8	-8	-95.06	-	-	-
	D2	-44.2	-8.59	-147.85	-47.4	-6.90	-31.81
L. plantarum	**Endo88**	-42.6	-7.61	-104.03	-47.7	-7.77	-148.77
	D1	-35.6	-7.21	-81.32	-40.1	-6.77	-72.72
	D2	-39.3	-7.23	-70.38	-50.4	-8.23	-191.33

(II) *S. epidermidis* peptide crosslink

The structure of peptide crosslink in *S. epidermidis* is similar to that of *S. aureus*, except that the peptide crosslink of *S. aureus* consists of little to no L-serine. Research have shown that serine replaces one of the glycine in the pentaglycine chain (Figure 5.36) (Schleifer & Kandler, 1972; Tipper, 1969; Willing et al., 2020).

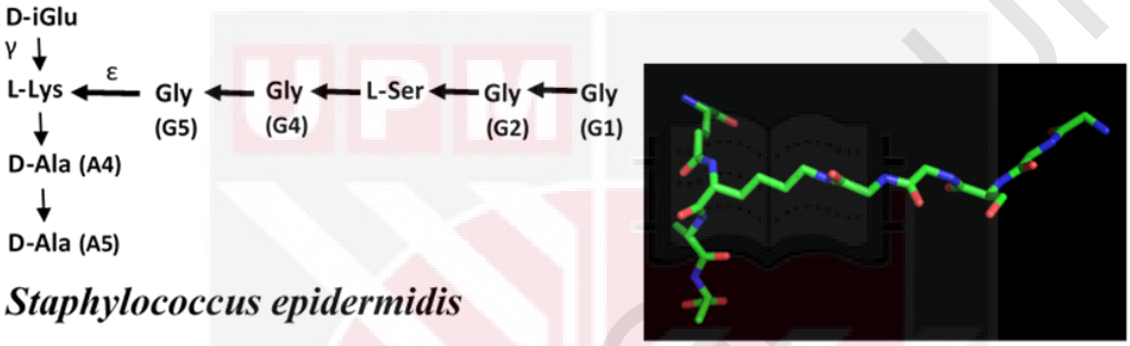


Figure 5.36: Peptide crosslink of *S. epidermidis* were generated using Avogadro 1.2.0. This structure was constructed based on references from existing studies (Sharif et al. 2009; Kim, Chang, and Singh 2015; Schleifer and Kandler 1972; Gonzalez-Delgado et al. 2020; Vollmer, Blanot, and de Pedro 2008).

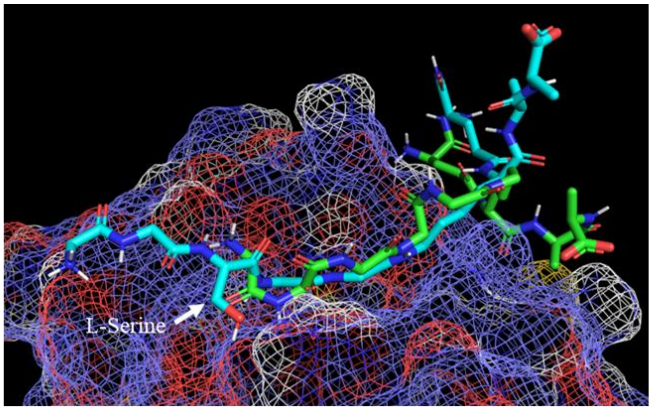
Previously, the binding assay for *S. epidermidis* exihibited almost no fluorescence detection in flow cytometry, accompanied by lower electrostatic energy and binding affinity compared to *S. aureus* in molecular docking (Table 5.6). Dry docking results revealed that the *S. aureus* peptide crosslink exihibited greater flexibility, enabling it to adapt more effectively to the shape of the binding groove. This flexibility resulted in higher binding coverage, as the *S. aureus* peptide crosslink interacted with a larger surface area within the binding groove compared to the *S. epidermidis* peptide crosslink, which appear to be straight and rigid (Figure 5.37A). For D2, the peptide crosslink did not bind to the binding groove during dry docking (Figure 5.37B). Additionally, the dissociation of the *S. epidermidis* peptide crosllink from the binding

site of SH3 domain was observed following solvated docking with short MD refinement for Endo88 and both mutants (Figure 5.37B).

It is noteworthy that in *S. epidermidis*, the glycine at position number 3 (denoted as G3) of pentaglycine is substituted with serine. This substitution may increase rigidity and potentially affect the binding of SH3 domain, resulting in almost no observed fluorescence in the binding assay (Abbas, 2017; Gray & Matthews, 1984; Mitkowski et al., 2019). Interestingly, lysostaphin also is unable to bind to some strains of *S. aureus* when the G3 is similarly substituted with serine, thus making these strains lysostaphin-resistant (Gonzalez-Delgado et al., 2020; Tossavainen et al., 2018). However, in constrast to lysostaphins, which are unable to bind or lyse these lysostaphin-resistant bacteria because the modified peptide crosslink served as both the binding and cleavage site, Endo88 and its mutants can still lyse the cells. This is likely because the CHAP's cleavage site is distinct from the pentaglycine chain.

The CHAP domain of phi_P24556 and MV-L, which share over 99% identity with the CHAP domain of Endo88, has been shown to possess D-alanyl-glycyl endopeptidase activity, cleaving the D-Ala-Gly bond (Figure 2.6) within the peptide crosslink (Navarre et al., 1999; Rashel et al., 2007). Given the high amino acid identity between these domains, it is plausible that the CHAP domain of Endo88 may also possess similar endopeptidase activity.

A



B

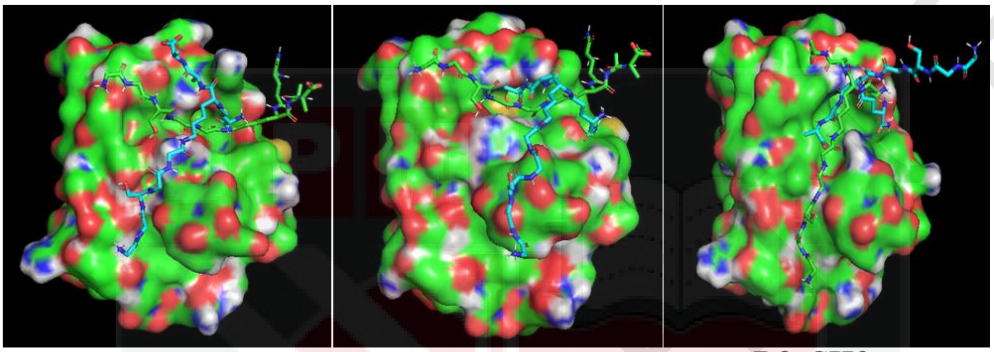


Figure 5.37: Superimposition of molecular docking. (A): Superimposition of docking between Endo_SH3 with *S. aureus* peptide crosslink AND Endo_SH3 with *S. epidermidis*. Green: Endo_SH3 with *S. aureus*; Cyan: Endo_SH3 with *S. epidermidis*. **(B):** Superimposition of dry docking between SH3 domains with *S. epidermidis* peptide crosslink AND after subjected to solvated docking with MD refinement. Green: Dry docking; Cyan: Solvated docking with MD refinement.

(III) *E. faecalis* peptide crosslink

The structure of the peptide crosslink in *E. faecalis* is similar to *S. aureus* at the peptide stem. However, in *E. faecalis* the peptide crosslink differs by incorporating two L-Alanine resideus instead of the typical interpeptide bridge found in *S. aureus* (Figure 5.38) (Bouhss et al., 2002; Schleifer & Kandler, 1972). The generated peptide crosslink was docked with the modeled SH3 domain with the guidance of AIR and UIR (Figure 5.31C)

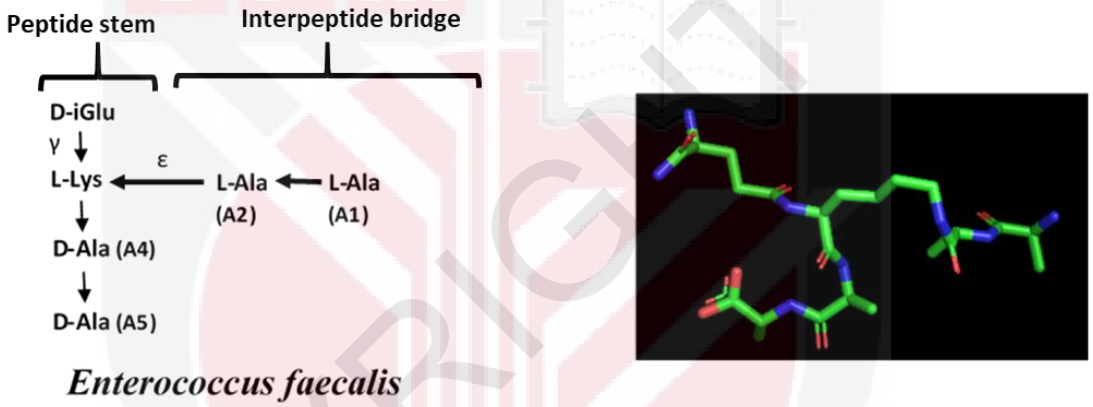


Figure 5.38: Peptide crosslink of *E. faecalis* were generated using Avogadro 1.2.0. This structure were created with reference to existing studies (Bouhss et al., 2002; Schleifer & Kandler, 1972).

The docking results aligned with the trends observed in both the Turbidity Reduction Assay (TRA) and binding assay. Specifically, D2 and D2_SH3 exhibited lytic and binding activity, respectively, with D1 showing higher activity than D2, although Endo88 was still the highest. Among the SH3 domains, Endo88_SH3 achieved the highest HADDOCK score and predicted binding affinity (HADDOCK score: -37.8; predicted binding energy: -8.24 kcal/mol), whereas D2 had the lowest (HADDOCK score: -27.6; predicted binding energy: -6.96 kcal/mol). This pattern persisted after solvated docking and refinement with short MD, where D2's HADDOCK score (-50.6) exceeded those of Endo88 (-45.9) and D1 (-40.9), with a predicted binding

energy same as D1 (-7.36 kcal/mol). The amino acid interaction profile for D2_SH3 revealed that only two residues (Y389 and T412) formed hydrogen bonds with the stem peptide, and no salt bridges were observed. The absence of salt bridges, which are present in Endo88_SH3 and D1_SH3, may account for the decreased in binding activity and lytic activities observed for D2.

The reduced number of interactions in the amino acid interaction profile of D2_SH3 raises concerns about potential inaccuracies in docking predictions. However, consistent results across multiple docking attempts indicate that the ligand consistently binds to the same site in all poses (Figure 5.39), albeit with a low number of interactions, suggesting the stability and reliability of the docking results. The discrepancy between the docking result and experimental result suggests that additional factors beyond the predicted docking interactions may influence the binding efficacy of SH3 domain in real-life situations.

As previously discussed, SH3 domains are commonly found in staphylococcal phage endolysin,s suggesting that these domains have evolved to specifically interact with the *Staphylococcus* peptidoglycan (Labrie et al., 2010; Oechslin et al., 2022; Vázquez et al., 2021). In contrast, *Enterococcus* phage endolysins are often abundant with LysM binding domains (Oliveira et al., 2013) instead of SH3 domains, which is interesting since Endo88 and all mutants generated in this study could both bind and lyse to *E. faecalis* in this study. As far as to our knowledge, our group is the first to report on a staphylococcal endolysin which can lyse *E. faecalis* (Krishnan et al., 2024). However, there have been reports of *Streptococcus* endolysins with SH3 domains which can lyse *Streptococcus*, *Staphylococcus* and *Enterococcus* (Gilmer et al., 2017; Huang et al.,

2015; Vander Elst et al., 2020). Due to a different binding domain found in *E. faecalis* compared to *S. aureus*, it is indicated that a different binding mechanism applies to the binding of the SH3 domain to *E. faecalis*. Research has shown that the LysM domains primarily target the N-acetylglucosamine (GlcNAc) within the glycan chain of peptidoglycan, with secondary interactions occurring with the peptide stems. This secondary interaction could be where SH3 binds. This specificity in target recognition influences both the binding affinity and specify of the LysM domains (Mesnage et al., 2014; Salamaga et al., 2023).

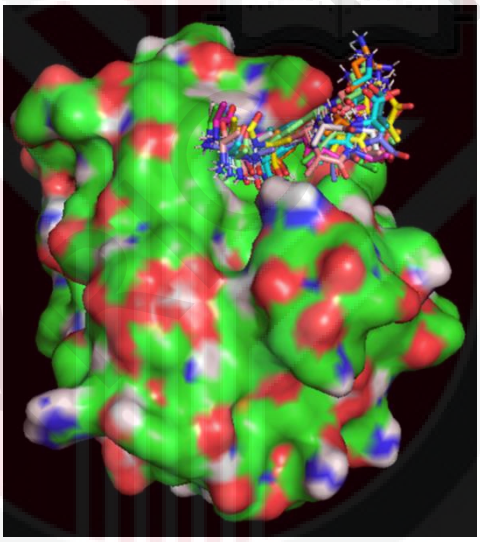


Figure 5.39: Molecular docking of D2_SH3 with the *E. faecalis* peptide crosslink. Docking poses were superimposed and illustrated consistent binding to the same site.

(IV) *M. luteus* and *L. plantarum* peptide crosslink

The peptide crosslinks in *M. luteus* and *L. plantarum* differ from the previous 3 peptide crosslink due to the absence of an interpeptide bridge. In these cases, the peptide crosslink are directly crosslinked by the peptide stem (Chapot-Chartier & Kulakauskas, 2014; Duchêne et al., 2019; Jensen & Campbell, 1976; KOCUR et al., 1972; Schleifer & Kandler, 1972). In this study, two stem peptides were joined and

generated in Avagadro for docking analysis. The peptide stem of *M. luteus* consists of a D-iGlu (side chain substituted with a glycine), L-lys and a D-Ala, with the L-Ala linking the peptide crosslink to the glycan chain (Figure 5.40). In *L. plantarum*, the side chain of D-iGlu is aminated (NH2), and the L-lys is replaced by *meso* - Diaminopimelic acid (m-Dpm), followed by a D – Ala (Figure 5.41).

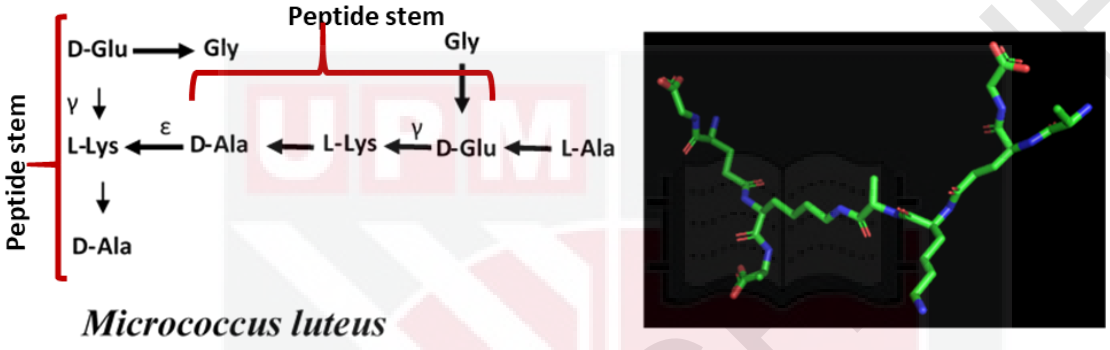


Figure 5.40: Peptide crosslink of *M. luteus* was generated using Avogadro 1.2.0. This structure was created with reference to existing studies (Jensen & Campbell, 1976; KOCUR et al., 1972; Schleifer & Kandler, 1972).

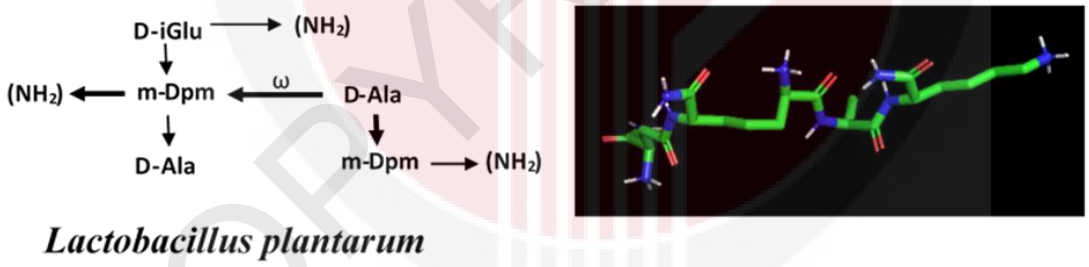


Figure 5.41 Peptide crosslink of *L. plantarum* was generated using Avogadro 1.2.0. This structure was created with reference to existing studies (Chapot-Chartier & Kulakauskas, 2014; Duchêne et al., 2019; Schleifer & Kandler, 1972)

For binding of *M. luteus* with Endo88, the docking results showed that the docking pose with the best HADDOCK score was positioned closed to the binding site, with other poses from other cluster being even more distant (Figure 5.42A). In the case of D1_SH3, all clusters conformations docked near the binding groove, yet only the highest HADDOCK score pose partially occupied the groove (Figure 5.42B).

However, post-solvated docking with short MD refinement revealed that these conformations eventually moved out of the groove. For D2, while initial docking showed all conformations were near the binding groove, none successfully fit into it (Figure 5.42C). Post-solvated docking with short MD refinement showed that only one pose managed to occupy the groove. However, this was accompanied by a significant drop in predicted binding energy and electrostatic energy.

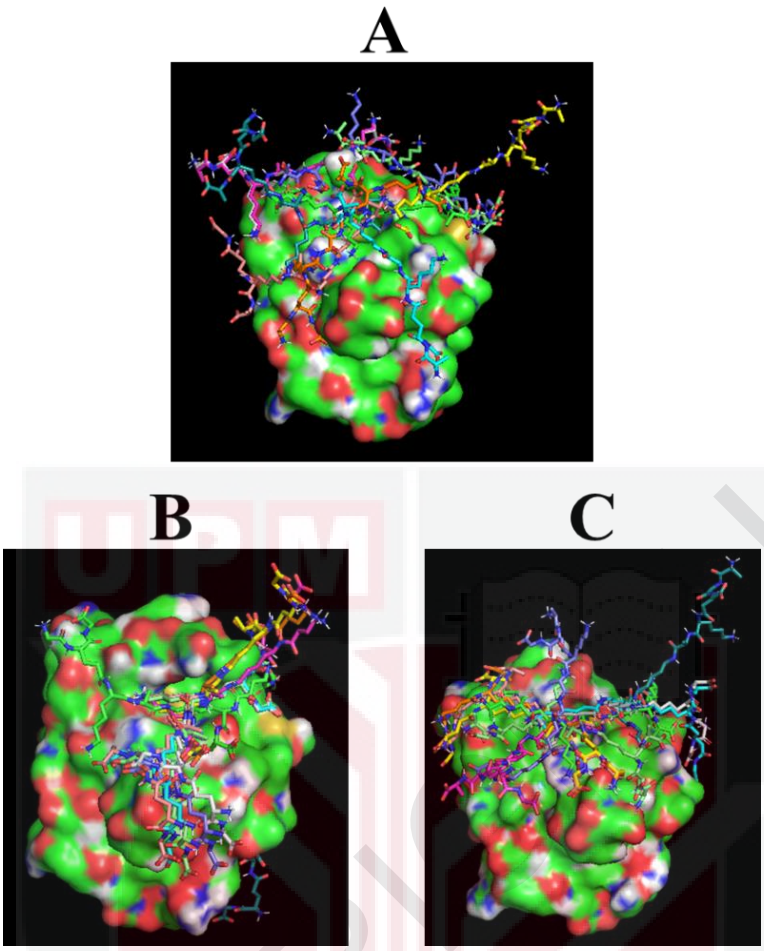


Figure 5.42: Superimpose of all the docking poses from the docking clusters. The binding interactions between SH3 domains and *M. luteus* peptide crosslink showed inconsistencies in conformation. **(A):** Endo88_SH3; **(B):** D1_SH3; **(C):** D2_SH3.

In summary, the docking results across all variants showed inconsistencies in the binding of the SH3 domain and the *M. luteus* peptide crosslink, complicating the identification of the most accurate binding conformation. Despite the limited understanding of the true interaction from other experimental data, the binding assay confirmed the absence of binding between the SH3 domain and the *M. luteus* peptide crosslink. While the peptide crosslink was positioned near the binding groove in the docking, it may not fully represent *in vivo* conditions, where other cell wall components such as teichuronic acid (Jensen & Campbell, 1976; Schleifer & Kandler,

1972; S. Yang et al., 2001) – analogous to teichoic acid and abundant in species like *M. luteus, B. subtilis,* and *B. licheniformis,* could influence the binding result.

Lactobacillus phage endolysin feature a diverse binding domain in their endolysin, including the SH3 domain (Figure 2.1). In contrast to *Staphylococcus* endolysins, which bind to the interpeptide bridge in the peptide crosslink, *Lactobacillus* does not have interpeptide bridge and the peptide stem was directly linked together at m-Dpm. It was postulated that the docking result may be similar to those observed for the *M. luteus* peptide crosslink.

For Endo88, 227 docking structures were clustered into 25 clusters. Among these, only 2 clusters, including the one with the best HADDOCK score, showed close fit into the binding groove. However, following solvated docking and short MD, the peptide crosslink left the binding groove (Figure 5.43A). For D1, 246 docking structures were clustered into 26 clusters. Of the top 10 clusters, 5 clusters displayed poses where the peptide crosslink fit into the binding groove. After solvated docking with short MD refinement, the peptide crosslink consistently remained within the binding groove across all clusters (Figure 5.43B). For D2, 293 docking structures were clustered into 29 clusters. The results were similar to D1, with 4 out of the top 10 clusters fitting into the binding groove. Even after solvated docking with short MD, the peptide crosslink consistently fit into the binding groove in all clusters (Figure 5.43C).

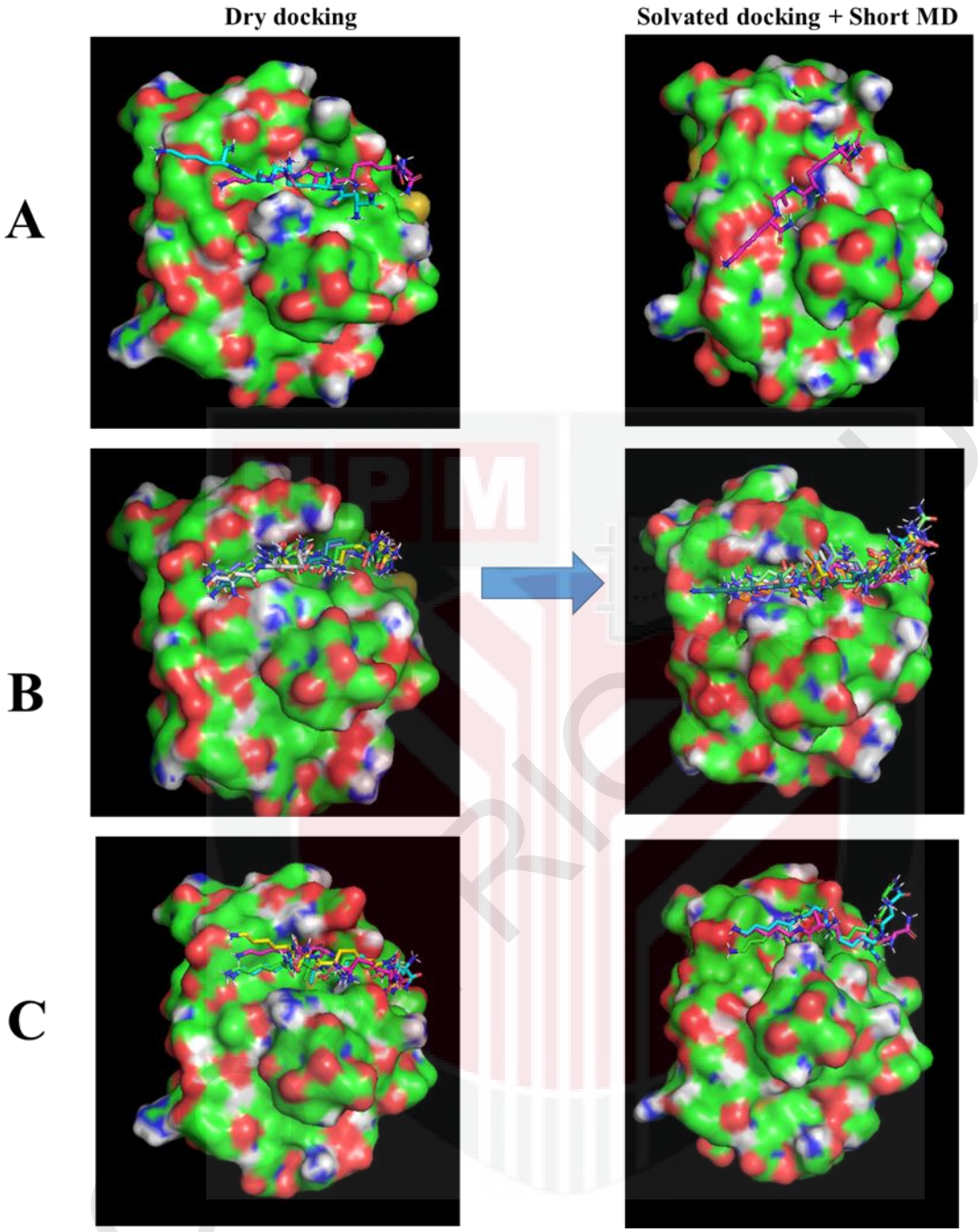


Figure 5.43: Superimposition of all the docking clusters close to the binding groove, both before and after solvated docking with short MD refinement. (A): Endo88_SH3 - Only one cluster was generated from the solvated docking, with the peptide crosslink dissociating from the binding groove. **(B): D1_SH3 -** Ten clusters were generated from the solvated docking, all remaining close to the binding groove. **(C): D2_SH3 -** Three clusters were generated from the solvated docking, all remaining close to the binding groove.

The docking result for the mutants suggest potential interactions between the *L. plantarum* peptide crosslink and D1_GFP_SH3 and D2_GFP_SH3. While D2_GFP_SH3 was designed to increase the host-range of Endo88 which can be seen from the molecular docking since the *L. plantarum* could dock into the binding groove, this was not correlated with the flow cytometry binding assay as no binding was detected. Even more baffling, the D1_GFP_sh3 docking results also showed docking of the *L.plantarum* peptide cross link with its binding site, although D1 was designed to be of narrower host-range than Endo88. This discrepancy raises concerns about the reliability of the docking predictions, as the true mechanism and binding conformation remain poorly defined. Two possible postulations for these observations can be considered:

(I) the observed binding in molecular docking might be a false positive. This could occur due to the docking of the peptide crosslink to the enlarged volume of binding groove on mutant SH3 domain (**Endo88_SH3:** 45.026 Å; **D1_SH3:** 59.904 Å; **D2_SH3:** 72.704 Å). Additionally, the *L. plantarum* peptide crosslink generated was relatively smaller compared to other peptide crosslink generated. The docking algorithm may then generate multiple poses for the peptide crosslink to fit within this larger space, even if these poses are not biologically relevant. Despite the use of solvated docking and short MD refinements, a longer duration of MD (e.g. >100 ns) might be required to observe the dissociation of peptide crosslink from the binding groove. Also, additional data such as a crystal structure that includes both the SH3 domain and the peptide crosslink would be essential for accuracy.

(II) The docking results is accurate but no binding was observed in the flow cytometry assay as docking primarily focuses on the interaction between the peptide crosslink and the SH3 domains while overlooking the presence of other cell wall components found in whole cells, which was used in the binding assay. In nature, the peptidoglycan layer is much more complex. A study by Dorosky et al., 2023 demonstrated that the removal of cell wall components (e.g. teichoic acid) from the *Lactobacillus* using trichloroacetic acid (TCA) led to an eight-fold increase in binding during flow cytometry. This indicated that components such as teichoic acid can interfere with the SH3 domain ability to bind optimally to the peptide crosslink, despite favorable docking predictions. Thus, while docking may accurately reflect the interaction between SH3 domain and peptide crosslink, the interference of other cell wall components must be considered in a biological context.

Moreover, the flow cytometry binding assay for D1_GFP_SH3 and D2_GFP_SH3 (Figure 5.27) against *B. subtilis* further highlights this interference. The peptide crosslink in *B. subtilis* shares high similarities with that of *L. planarum* (Angeles & Scheffers, 2021; Foster & Popham, 2001; Hayhurst et al., 2008; Schleifer & Kandler, 1972), where the m-Dpm residues directly link to each other and form the cross linkage in the peptidoglycan. Despite these structural similarities, weak binding of less than 20% was detected on *B. subtilis* during flow cytometry binding assay in comparison with *L. plantarum* which had no binding detected. This could suggest that *B. subtilis* has fewer interfering components on the whole cell wall compared to *L. plantarum*, resulting in lesser intereference of the SH3 domain binding.

In summary, the docking results revealed that the peptide crosslink was involved in the binding mechanism between SH3 domains and peptidoglycan. For the *S. aureus* peptide crosslink, the role of the interpeptide bridge (pentaglycine) as a host-range determining site for SH3 domains was aligned with other studies. This was further validated by the binding assays and molecular docking of *S. epidermidis*, where the replacement of one glycine with serine resulted in unsuccessful binding of the peptide crosslink to the SH3 domain. Although Endo88 was able to lyse *S. epidermidis* (Figure 5.25), the unsuccessful binding of the CBD could be one of the reasons the endolysins only achieved <80% OD reduction in turbidity reduction assay compared to *S. aureus* (Figure 5.25). Intriguingly, the SH3 domain from Endo88 exhibited a rather wide binding host range as it showed successful binding to *E. faecalis* in both binding assay and molecular docking. No binding was detected in *M. luteus* in binding assay. This was further validated by the molecular docking, which showed inconsistent binding pose. Albeit no binding was detected for *L. plantarum* in the binding assays, molecular docking revealed potential interactions between the peptide crosslink and the SH3 domains, suggesting that the binding could also be hindered by other components in the peptidoglycan.

5.3.9.4 Correlation between binding and lytic host range

Although numerous studies suggest that the SH3 domain confers specificity to the endolysins, the exact mechanism remains unclear to date. The observations in this study indicated that the presence of a SH3 domain in *Staphylococcus* phage endolysins is not essential for lytic activity, at least under laboratory conditions. This finding supports our study on Endo88 truncation and domain swapping study which was performed concurrently with the current study (Krishnan et al., 2024) where truncation of SH3 did not diminish lytic activity but reduced it significantly.

Host-range specificity of endolysins may not be dependent solely on the CBD but the EAD may also play an important role. The lytic ability of an endolysin depends on the specific bonds that the EAD can cleave. As aforementioned, in the current study, it was shown that Endo88, D1 and D2 were able to lyse *S. epidermidis* at different degrees although no binding was detected, possibly due to the fact that the EAD had a cleavage site different from the SH3 binding site. This was in contrast with lysostaphin-resistant *S. aureus* where substitution of a glycine in the peptide stem with serine caused lysostaphin to be unable to bind and also lyse the bacteria it since the binding and cleavage site is the same (Gonzalez-Delgado et al., 2020; Tossavainen et al., 2018). In another scenario, previous studies have also shown that swapping a CBD of a *Streptococcus* endolysin with a *Staphylococcus* CBD allowed the endolysin to lyse *Staphylococcus* (S. C. Becker, Foster-Frey, et al., 2009; Donovan, Dong, et al., 2006; Fernandes et al., 2012; Schmelcher, Korobova, et al., 2012). However, swapping the CBD of a Endo88 with a *Streptococcus* endolysin did not extend the host-range to *Streptococcus* and it was also postulated that this could be due to the fact that the EAD of the *Streptococcus* endolysin could lyse both *Streptococcus* and *Staphylococcus*, but

Endo88 EAD was unable to cleave *Streptococcus* peptidoglycans (Krishnan et al., 2024).

In addition, while our previous study (Krishnan et al., 2024) and various other studies have found that removal of CBD reduces lytic activity (Cheng & Fischetti, 2007; Gaeng et al., 2000; Low et al., 2005; Schmelcher, Korobova, et al., 2012; Yokoi et al., 2005), there were other studies that found that removal of the CBD did not affect the lytic activity (Cheng & Fischetti, 2007; Donovan, Foster-Frey, et al., 2006) or even enhances it (Abaev et al., 2013; Huang et al., 2015; Phothichaisri et al., 2022.). It has been suggested that the retained catalytic activity of the CHAP domain alone could be due to the overall positive charge on the CHAP domain (Low et al., 2011) of Gram-positive bacteria with overall negatively charged peptidoglycan which favours the binding of the EAD. However, this is not the sole factor contributing to lytic activity, as suggested by another study (Shang & Nelson, 2019). Both studies utilised site-directed mutagenesis to alter the overall charge of the CHAP domain to increase the positive charge. As compared to the full-length endolysin, Low et al was able to enhance the lytic activity of CHAP, while Shang & Nelson observed decrease in lytic activity. In both cases, the mutation sites were strategically introduced in non-interacting regions, regions without any secondary structure and surface accessible regions. The key differences were the distribution and location of these mutations: Shang & Nelson's mutations were evenly distributed on the surface, while Low et al was in the proximity of the active site. Furthermore, the active site in Shang and Nelson's mutants is located in a neutral groove, while in Low et al's mutants, it is positioned in an anionic pocket. Further studies are necessary to explore how these differences influence lytic activity of the CHAP domain.

On another note, one studies postulated that the retained/enhanced lytic activity of EADs without the CBD might be due to the purified and high concentration of EAD used in lytic assays (Loessner et al., 2002). In the studies of Phothichaisri et al., 2022, 100 μg of EAD was used in lytic assay, unlike 10 μg used for Endo88. In nature, the endolysin may be produced at relatively lower amounts within the host cell as compared to lab-based assay, making CBD necessary to facilitate binding to peptidoglycan when the EAD concentration is low (Loessner et al., 2002).

The above discussion shows that binding and lysis are two separate events which do not always correlate. Nonetheless, in previous reports where endolysin are tested against their original bacterial hosts, both lytic and binding activities are consistently observed together (Benešík et al., 2018; Etobayeva et al., 2018; Gu, Lu, et al., 2011; Linden et al., 2015; Loessner et al., 2002; Schmelcher et al., 2011; Vander Elst et al., 2020). Additionally, in the few studies where both lysis and binding assay were performed across different bacterial species or genera, it was also indicated that binding typically correlates with lysis (Y. Chang & Ryu, 2017; Kong et al., 2015; Loessner et al., 2002; Schmelcher et al., 2011). For example, the SH3 domain-containing endolysin, LysGH15 was able to lyse and bind to *S. aureus* (Gu et al., 2014; Gu, Lu, et al., 2011; Gu, Xu, et al., 2011). When tested against *S. pneumonia* and *B. subtilis*, both lysis and binding were not observed. An exception to this was reported for endolysin LysF1 (Benešík et al., 2018). In this study, binding and lysis were both observed in *S. aureus, S. epidermidis, S. rosti, S. capitis,* and *S. carnosus.* However, while binding was detected on both *S. sciuri* and *Streptococcus uberis*, no lysis was observed. It is noteworthy that the peptide crosslink of *S. sciuri* is substituted with one L-Ala in the glycine chain (L-Lys - L-Ala – Gly_{4-5}) (Schleifer & Kandler, 1972;

Schumann, 2011), which in the case of *S. epidermidis* and lysostaphin, seem to diminish binding when a substitution occur which is opposite from what was reported for LysF1. In addition, the peptide crosslink for *S. uberis* is similar to *E. faecalis* (L-Lys – L-Ala_2) (Figure 5.38), which showed both binding and lysis in this study, while in the case of LysF1 only binding was observed. As such, the correlation between CBD and EAD remains unclear due to conflicting reports and also due to a lack of studies which have conducted both lytic and binding assay on cross-species or –genus bacteria.

Finally, As aforementioned, the mutations leading to a broad host range in endolysins likely occur over longer periods of time as phage adapt to new hosts with different peptidoglycan components (Oechslin et al., 2022). This gradual process may be a protective measure to avoid loss of protein function due to excessive amino acid mutations at once (Strobel et al., 2024). This may explain why the mutations introduced in this study resulted in reduced lytic activity rather than an altered host range, as observed with the M, A2 (Figure 4.24), D1 and D2 mutants (Figure 5.25), and why the A1 mutant was expressed as non-functional inclusion bodies (Figure 4.23).

In summary, this study employed a structure-based approach to design mutations in the SH3 domain of Endo88, specifically targeting the predicted binding residue (NKY–Y–I–E–Y) identified through bioinformatics analysis. Binding activity was assessed using a GFP-fused SH3 domain. Similar to previous mutations (Chapter 4), the modifications did not alter the host range but reduced the lytic and binding activity of the endolysin. While wild-type Endo88 and its SH3 domain exhibit highest lytic and binding activity, respectively, D2 and its SH3 domain exhibit the lowest for both assay.

Notably, binding assays indicated that the binding activity did not always correlate with lytic activity, as shown by *S. epidermidis* lysis despite no binding were detected. Molecular docking revealed that hydrogen bonding plays an important role in binding interactions, particularly at the pentaglycine region. The Endo88 SH3 domain formed the highest number of hydrogen bonds with the peptide crosslink, whereas D2_SH3 exhibited fewer hydrogen bond with pentaglycine. No hydrogen bonds were observed even after solvated docking with short MD refinement, which may explain no binding activity was detected in the assay.

CHAPTER 6

CONCLUSION, LIMITATIONS, AND RECOMMENDATIONS

6.1 Conclusion

This study sought to elucidate the host specificity of the SH3 CBD of Endo88 as a model for *S. aureus* endolysins by assessing the effects of amino acid modifications within the domain. Five mutant endolysins were successfully designed, cloned and expressed, and their lytic and binding activity were compared with the wild type (Endo88). Initial attempt to widen the host range by replacing specific regions in Endo88 presumed to be exclusive in endolysins with a wide host range (resulting in mutants A1, A2 and M) failed to alter the host range. In addition, attempts to narrow or widen the host range based on mutations of the binding sites of endolysins (mutants D1 and D2, respectively) similarly did not alter the host range. Instead, both lytic and binding activity were decreased significantly, implying that the mutations in the SH3 domain affected both lytic and binding activity to a certain degree but did not alter host range as all mutants showed the same lytic host range as the wild-type comprising *S. aureus, S. hominis, S. epidermidis,* and *E. faecalis*. Finally, investigations using CBD-GFP fusion protein, indicated that the binding of the CBD and EAD lysis activities may not be closely correlated, as evidenced by lack of binding detection for *S. epidermidi,* despite positive lysis. This corroborates findings from other studies that some endolysins can still lyse bacteria without the CBD, implying that binding may not be a prerequisite for lysis, although most studies show that the CBD is needed for maximum lysis and that removal of the CBD results in major reduction of lytic activity in most cases. In the current study, the specific mutations in the CBD region

significantly reduced the overall lytic efficiency of the endolysin, highlighting the intricate relationship between the role of CBD and EAD, despite their modular structure. In addition, the molecular docking results obtained in this study suggests that molecular docking can offer a cheap, rapid and simpler approach for understanding the binding interactions between the CBD and various peptide crosslinks. However, this depends on having an adequate number of endolysin sequences with specific host ranges and that the host range data from these references are derived from a standardised method.

While advanced techniques such as molecular dynamics (MD) and NMR can provide detailed insights into molecular interactions, molecular docking offers a more rapid and simpler approach for understanding the binding interactions between the CBD and various peptide crosslinks. In future applications, by using substantial dataset of homologous endolysins as mutational references, molecular docking could efficiently screen through each candidate and identify the best mutants for subsequent analysis.

6.2 Limitations

The study faced challenges due to the limited availability of experimental data on *S. aureus* endolysins with well-defined host ranges. Although numerous sequences of *Staphylococcus* phage endolysin were retrieved from BLAST, most were the results of genomic sequencing without further analysis. Additionally, those that underwent experiemental characterisation mostly did not extend their host range studies across multiple genera. This lack of comprehensive data contributed to ambiguities in defining the host range of endolysin in this study. This is further complicated by the variations of protocols used to determine lysis activity and by extension, the host-range. This limitation may explain why the mutations introduced, based on the primary sequence of the endolysins, did no alter the host range, as dicussed in section 4.3.5.3.

Another limitation was the lack of understanding of the true binding mechanism between the CBD and its ligand. The interaction between most CBDs and theirs targets, including SH3, has not been fully elucidated. Despite some studies suggesting that the peptide crosslink is the binding site of the SH3 domain (from *Staphylococcus* phage endolysin), these conclusions were based on the analysis of peptidoglycan fragments rather than fully intact peptidoglycan, The complexity of the peptidoglycan, which includes components such as teichoic acids, membrane-bound proteins and compositions of peptidoglycan at different growth phase are likely to interfere with binding. It is also challenging to account for all these factors in a single analysis, further complicating the understanding of how these interactions influence the lytic host range.

6.3 Recommendations

To accurately determine and establish the host range of endolysins, based on the current study, a few recommendations are made for future studies as described below.

6.3.1 The choices of assay

When determining the host range of an endolysin, the choice of assay used must be standardized, as both the media and environmental conditions can influence both the bacteria (Oechslin et al., 2013; Vander Elst, 2024) and also the endolysin (Bhagwat et al., 2021). Broth-based assays support the growth of bacteria and maintain the "natural" condition of the bacteria cell, but the nutrient content can vary depending on the brand, consequently, affecting the structure or growth of the target bacteria. In contrast, buffer solutions provide a more controlled environment for assessing the lytic activity. Although real-life applications might encounter more complex conditions that could influence the lytic analysis, it is essential to first establish a baseline understanding in a standardized manner.

Moreover, the technique used can also influence host range determination for the same endolysin (Schmelcher et al., 2015; Vander Elst, 2024; Yokoi et al., 2005). Endolysins work synergistically between their EAD and CBD on the complex peptidoglycan structure. Therefore, it is recommended to use assay which preserve the integrity of the peptidoglycan overall structure (e.g. Turbidity reduction assay, time-kill assay and etc), rather than destructive methods (e.g. zymogram involving SDS, autoclaving of cells prior to analysis and etc). These destructive approaches may disrupt the natural interactions, leading to inaccurate host range results.

6.3.2 Separate evaluations of EAD, CBD and full-length endolysins

To thoroughly understand the contributions of each domains, experiments should include separate evaluations of the EAD, CBD and the full-length endolysin. This allows for better understanding of their individual roles and their synergistic effects on lytic activity agains different bacterial genera. For instance, in the current study and some others, binding host-range and lytic host-range are not the same. In addition, some studies have reported that their EAD lyse better than the full-length endolysin (Abaev et al., 2013; Huang et al., 2015; Phothichaisri et al., 2022.), which could lead to misinterpretation of the lytic host range if not assessed separately. The EAD that lyse better without CBD could become potential candidate for chimeric endolysin development, potentially leading to more effective antimicrobial agents.

6.3.3 Use of advanced technologies and computational methods

Given that peptidoglycan structures can vary across different genera (Garde et al., 2021; Khairulddin et al., 2004; S. J. Kim et al., 2015; R. D. Turner et al., 2014), these variations must be considered, as most endolysins are reported to target specific components on peptidoglycan. Even slight structural differences can significantly influence the lytic/binding activity, as seen in the case of Endo88 towards *S. epidermidis*。 Due to the many variations in peptidoglycan structure, it is difficult to predict the mechanism of action of endolysins on these peptidoglycan, thus proving it difficult to find mutations which can cause host-range alteration. However, achieving this understanding presents a challenge in maintaining the balance between mutating the amino acids in a particular type of CBD and preserving the tertiary structure's integrity amidst conformational changes induced by the introduced mutations.

To further understand this, more experimental data and computational analysis for different combinations of amino acids are needed. Advanced techniques such as molecular dynamic simulations, and nuclear magnetic resonance (NMR) and X-ray crystallography remains valuable for structural analysis (Sanz-Gaitero & van Raaij, 2021) and can provide a more comprehensive understanding of the binding mechanisms. In addition, emerging technologies such as cryo-electron microscopy (Cryo-EM) offer alternative methods for detailed structural visualization (Sanz-Gaitero & van Raaij, 2021). Additionally, AI-driven tools such as AlphaFold offer a cost-effective and time-efficient approach of predicting structure, thus facilitating rapid analysis (Nakonieczna et al., 2024; Oechslin et al., 2024). Machine learning may also prove useful in screening the numerous combinations of amino acid combinations with the different peptidoglycan structures whilst maintaining endolysin protein structure in order to come up with better mutations.

REFERENCES

Abaev, I., Foster-Frey, J., Korobova, O., Shishkova, N., Kiseleva, N., Kopylov, P., Pryamchuk, S., Schmelcher, M., Becker, S. C., & Donovan, D. M. (2013). Staphylococcal phage 2638A endolysin is lytic for Staphylococcus aureus and harbors an inter-lytic-domain secondary translational start site. *Applied Microbiology and Biotechnology*, *97*(8), 3449–3456. https://doi.org/10.1007/s00253-012-4252-4

Abbas, H. (2017). First-principles study of interaction of serine with nucleobases of DNA and RNA. *Journal of Biological Physics*, *43*(1), 105–111. https://doi.org/10.1007/s10867-016-9436-1

Abdelkader, K., Gerstmans, H., Saafan, A., Dishisha, T., & Briers, Y. (2019). The Preclinical and Clinical Progress of Bacteriophages and Their Lytic Enzymes: The Parts are Easier than the Whole. *Viruses*, *11*(2), 96. https://doi.org/10.3390/v11020096

Abedon, S. T. (2011). Lysis from without. *Bacteriophage*, *1*(1), 46–49. https://doi.org/10.4161/bact.1.1.13980

Abeysekera, G. S., Love, M. J., Manners, S. H., Billington, C., & Dobson, R. C. J. (2022). Bacteriophage-encoded lethal membrane disruptors: Advances in understanding and potential applications. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.1044143

Adhav, V. A., & Saikrishnan, K. (2023). The Realm of Unconventional Noncovalent Interactions in Proteins: Their Significance in Structure and Function. *ACS Omega*, *8*(25), 22268–22284. https://doi.org/10.1021/acsomega.3c00205

Agu, P. C., Afiukwa, C. A., Orji, O. U., Ezeh, E. M., Ofoke, I. H., Ogbu, C. O., Ugwuja, E. I., & Aja, P. M. (2023). Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Scientific Reports*, *13*(1), 13398. https://doi.org/10.1038/s41598-023-40160-2

Ahmed, A., Rushworth, J. V., Hirst, N. A., & Millner, P. A. (2014). Biosensors for Whole-Cell Bacterial Detection. *Clinical Microbiology Reviews*, *27*(3), 631–646. https://doi.org/10.1128/CMR.00120-13

Al-Hindi, R. R., Teklemariam, A. D., Alharbi, M. G., Alotibi, I., Azhari, S. A., Qadri, I., Alamri, T., Harakeh, S., Applegate, B. M., & Bhunia, A. K. (2022). Bacteriophage-Based Biosensors: A Platform for Detection of Foodborne Bacterial Pathogens from Food and Environment. *Biosensors*, *12*(10), 905. https://doi.org/10.3390/bios12100905

Anantharaman, V., & Aravind, L. (2003). Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biology*, *4*(2), R11. https://doi.org/10.1186/gb-2003-4-2-r11

Angeles, D. M., & Scheffers, D.-J. (2021). The Cell Wall of Bacillus subtilis. *Current Issues in Molecular Biology*, *41*(1), Article 1. https://doi.org/10.21775/cimb.041.539

Anthony C.Grabski. (2009). *Chapter 18 Advances in Preparation of Biological Extracts for Protein Purification—ScienceDirect* (2nd ed., Vol. 463). https://www.sciencedirect.com/science/article/pii/S0076687909630184

Arold, S., O'Brien, R., Franken, P., Strub, M. P., Hoh, F., Dumas, C., & Ladbury, J. E. (1998). RT loop flexibility enhances the specificity of Src family SH3 domains for HIV-1 Nef. *Biochemistry*, *37*(42), 14683–14691. https://doi.org/10.1021/bi980989q

Arshad, A. R., Ijaz, F., Siddiqui, M. S., Khalid, S., Fatima, A., & Aftab, R. K. (n.d.). COVID-19 pandemic and antimicrobial resistance in developing countries. *Discoveries*, *9*(2), e127. https://doi.org/10.15190/d.2021.6

Asbach, B., Ludwig, C., Saksela, K., & Wagner, R. (2012). Comprehensive analysis of interactions between the Src-associated protein in mitosis of 68 kDa and the human Src-homology 3 proteome. *PloS One*, *7*(6), e38540. https://doi.org/10.1371/journal.pone.0038540

Azevedo, C., & Saiardi, A. (2016). Why always lysine? The ongoing tale of one of the most modified amino acids. *Advances in Biological Regulation*, *60*, 144–150. https://doi.org/10.1016/j.jbior.2015.09.008

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, *10*(5), 411–421. https://doi.org/10.1016/S0958-1669(99)00003-8

Bastos, M. do C. de F., Coutinho, B. G., & Coelho, M. L. V. (2010). Lysostaphin: A Staphylococcal Bacteriolysin with Potential Clinical Applications. *Pharmaceuticals*, *3*(4), 1139–1161. https://doi.org/10.3390/ph3041139

Bateman, A., & Rawlings, N. D. (2003). The CHAP domain: A large family of amidases including GSP amidase and peptidoglycan hydrolases. *Trends in Biochemical Sciences*, *28*(5), 234–237. https://doi.org/10.1016/S0968-0004(03)00061-6

Becker, K., Heilmann, C., & Peters, G. (2014). Coagulase-negative staphylococci. *Clinical Microbiology Reviews*, *27*(4), 870–926. https://doi.org/10.1128/CMR.00109-13

Becker, S. C., Dong, S., Baker, J. R., Foster-Frey, J., Pritchard, D. G., & Donovan, D. M. (2009). LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells. *FEMS Microbiology Letters*, *294*(1), 52–60. https://doi.org/10.1111/j.1574-6968.2009.01541.x

Becker, S. C., Foster-Frey, J., & Donovan, D. M. (2008). The phage K lytic enzyme LysK and lysostaphin act synergistically to kill MRSA. *FEMS Microbiology Letters*, *287*(2), 185–191. https://doi.org/10.1111/j.1574-6968.2008.01308.x

Becker, S. C., Foster-Frey, J., Stodola, A. J., Anacker, D., & Donovan, D. M. (2009). Differentially conserved staphylococcal SH3b_5 cell wall binding domains confer increased staphylolytic and streptolytic activity to a streptococcal prophage endolysin domain. *Gene*, *443*(1–2), 32–41. https://doi.org/10.1016/j.gene.2009.04.023

Becker, S. C., Swift, S., Korobova, O., Schischkova, N., Kopylov, P., Donovan, D. M., & Abaev, I. (2015). Lytic activity of the staphylolytic Twort phage endolysin CHAP domain is enhanced by the SH3b cell wall binding domain. *FEMS Microbiology Letters*, *362*(1), 1–8. https://doi.org/10.1093/femsle/fnu019

Benešík, M., Nováček, J., Janda, L., Dopitová, R., Pernisová, M., Melková, K., Tišáková, L., Doškař, J., Žídek, L., Hejátko, J., & Pantůček, R. (2018). Role of SH3b binding domain in a natural deletion mutant of Kayvirus endolysin LysF1 with a broad range of lytic activity. *Virus Genes*, *54*(1), 130–139. https://doi.org/10.1007/s11262-017-1507-2

Bennion, B. J., & Daggett, V. (2003). The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5142–5147. https://doi.org/10.1073/pnas.0930122100

Bezrodnykh, E. A., Antonov, Y. A., Berezin, B. B., Kulikov, S. N., & Tikhonov, V. E. (2021). Molecular features of the interaction and antimicrobial activity of chitosan in a solution containing sodium dodecyl sulfate. *Carbohydrate Polymers*, *270*, 118352. https://doi.org/10.1016/j.carbpol.2021.118352

Bhagwat, A., Zhang, F., Collins, C. H., & Dordick, J. S. (2021). Influence of bacterial culture medium on peptidoglycan binding of cell wall lytic enzymes. *Journal of Biotechnology*, *330*, 27–34. https://doi.org/10.1016/j.jbiotec.2021.02.010

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X.-Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*. https://doi.org/10.3389/fbioe.2021.630551

Bienstock, R. J. (2015). Solvation methods for protein-ligand docking. *Methods in Molecular Biology (Clifton, N.J.)*, *1289*, 3–12. https://doi.org/10.1007/978-1-4939-2486-8_1

Bouhss, A., Josseaume, N., Severin, A., Tabei, K., Hugonnet, J.-E., Shlaes, D., Mengin-Lecreulx, D., Van Heijenoort, J., & Arthur, M. (2002). Synthesis of the L-alanyl-L-alanine cross-bridge of Enterococcus faecalis peptidoglycan. *The Journal of Biological Chemistry*, *277*(48), 45935–45941. https://doi.org/10.1074/jbc.M207449200

Braham, S. A., Siar, E.-H., Arana-Peña, S., Carballares, D., Morellon-Sterling, R., Bavandi, H., de Andrades, D., Kornecki, J. F., & Fernandez-Lafuente, R. (2021). Effect of Concentrated Salts Solutions on the Stability of Immobilized Enzymes: Influence of Inactivation Conditions and Immobilization Protocol. *Molecules*, *26*(4), 968. https://doi.org/10.3390/molecules26040968

Briers, Y., Walmagh, M., Puyenbroeck, V. V., Cornelissen, A., Cenens, W., Aertsen, A., Oliveira, H., Azeredo, J., Verween, G., Pirnay, J.-P., Miller, S., Volckaert, G., & Lavigne, R. (2014). Engineered Endolysin-Based "Artilysins" To Combat Multidrug-Resistant Gram-Negative Pathogens. *mBio*, *5*(4). https://doi.org/10.1128/mBio.01379-14

Broendum, S. S., Buckle, A. M., & McGowan, S. (2018). Catalytic diversity and cell wall binding repeats in the phage-encoded endolysins. *Molecular Microbiology*, *110*(6), 879–896. https://doi.org/10.1111/mmi.14134

Buey, R. M., Monterroso, B., Menéndez, M., Diakun, G., Chacón, P., Hermoso, J. A., & Díaz, J. F. (2007). Insights into Molecular Plasticity of Choline Binding Proteins (Pneumococcal Surface Proteins) by SAXS. *Journal of Molecular Biology*, *365*(2), 411–424. https://doi.org/10.1016/j.jmb.2006.09.091

Bustamante, N., Iglesias-Bexiga, M., Bernardo-García, N., Silva-Martín, N., García, G., Campanero-Rhodes, M. A., García, E., Usón, I., Buey, R. M., García, P., Hermoso, J. A., Bruix, M., & Menéndez, M. (2017). Deciphering how Cpl-7 cell wall-binding repeats recognize the bacterial peptidoglycan. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-16392-4

C, S., S., D. K., Ragunathan, V., Tiwari, P., A., S., & P, B. D. (2022). Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease. *Journal of Biomolecular Structure and Dynamics*, *40*(2), 585–611. https://doi.org/10.1080/07391102.2020.1815584

Cao, E., Chen, Y., Cui, Z., & Foster, P. R. (2003). Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. *Biotechnology and Bioengineering*, *82*(6), 684–690. https://doi.org/10.1002/bit.10612

Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery: Current trends and applications. *Drug Discovery Today*, *14*(13–14), 676–683. https://doi.org/10.1016/j.drudis.2009.04.006

Ceyssens, P.-J., Lavigne, R., Mattheus, W., Chibeu, A., Hertveldt, K., Mast, J., Robben, J., & Volckaert, G. (2006). Genomic analysis of Pseudomonas aeruginosa phages LKD16 and LKA1: Establishment of the phiKMV subgroup within the T7 supergroup. *Journal of Bacteriology*, *188*(19), 6924–6931. https://doi.org/10.1128/JB.00831-06

Chagas, R. S., Otsuka, F. A. M., Pineda, M. A. R., Salinas, R. K., & Marana, S. R. (2023). Mechanism of imidazole inhibition of a GH1 β-glucosidase. *FEBS Open Bio*, *13*(5), 912–925. https://doi.org/10.1002/2211-5463.13595

Chaibenjawong, P., & Foster, S. J. (2011). Desiccation tolerance in Staphylococcus aureus. *Archives of Microbiology*, *193*(2), 125–135. https://doi.org/10.1007/s00203-010-0653-x

Chan, W.-T., Verma, C. S., Lane, D. P., & Gan, S. K.-E. (2013). A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. *Bioscience Reports*, *33*(6), e00086. https://doi.org/10.1042/BSR20130098

Chandran, C., Tham, H. Y., Abdul Rahim, R., Lim, S. H. E., Yusoff, K., & Song, A. A.-L. (2022). Lactococcus lactis secreting phage lysins as a potential antimicrobial against multi-drug resistant Staphylococcus aureus. *PeerJ*, *10*, e12648. https://doi.org/10.7717/peerj.12648

Chang, C. Y., Nam, K., & Young, R. (1995). S gene expression and the timing of lysis by bacteriophage lambda. *Journal of Bacteriology*, *177*(11), 3283–3294. https://doi.org/10.1128/jb.177.11.3283-3294.1995

Chang, Y., Kim, M., & Ryu, S. (2017). Characterization of a novel endolysin LysSA11 and its utility as a potent biocontrol agent against Staphylococcus aureus on food and utensils. *Food Microbiology*, *68*, 112–120. https://doi.org/10.1016/j.fm.2017.07.004

Chang, Y., & Ryu, S. (2017). Characterization of a novel cell wall binding domain-containing Staphylococcus aureus endolysin LysSA97. *Applied Microbiology and Biotechnology*, *101*(1), 147–158. https://doi.org/10.1007/s00253-016-7747-6

Chapot-Chartier, M.-P., & Kulakauskas, S. (2014). Cell wall structure and function in lactic acid bacteria. *Microbial Cell Factories*, *13*(Suppl 1), S9. https://doi.org/10.1186/1475-2859-13-S1-S9

Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S. M., & Savidge, T. C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science Advances*, *2*(3), e1501240. https://doi.org/10.1126/sciadv.1501240

Cheng, Q., & Fischetti, V. A. (2007). Mutagenesis of a bacteriophage lytic enzyme PlyGBS significantly increases its antibacterial activity against group B streptococci. *Applied Microbiology and Biotechnology*, *74*(6), 1284–1291. https://doi.org/10.1007/s00253-006-0771-1

Chin, W. H., Kett, C., Cooper, O., Müseler, D., Zhang, Y., Bamert, R. S., Patwa, R., Woods, L. C., Devendran, C., Korneev, D., Tiralongo, J., Lithgow, T., McDonald, M. J., Neild, A., & Barr, J. J. (2022). Bacteriophages evolve enhanced persistence to a mucosal surface. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(27), e2116197119. https://doi.org/10.1073/pnas.2116197119

Cho, J.-H., Kwon, J.-G., O'Sullivan, D. J., Ryu, S., & Lee, J.-H. (2021). Development of an endolysin enzyme and its cell wall–binding domain protein and their applications for biocontrol and rapid detection of *Clostridium perfringens* in food. *Food Chemistry*, *345*, 128562. https://doi.org/10.1016/j.foodchem.2020.128562

Chong, L. C., & Khan, A. M. (2019). Identification of highly conserved, serotype-specific dengue virus sequences: Implications for vaccine design. *BMC Genomics*, *20*(9), 921. https://doi.org/10.1186/s12864-019-6311-z

Christie, G. E., Matthews, A. M., King, D. G., Lane, K. D., Olivarez, N. P., Tallent, S. M., Gill, S. R., & Novick, R. P. (2010). The complete genomes of Staphylococcus aureus bacteriophages 80 and 80α—Implications for the specificity of SaPI mobilization. *Virology*, *407*(2), 381–390. https://doi.org/10.1016/j.virol.2010.08.036

Clark, J. J., Benson, M. L., Smith, R. D., & Carlson, H. A. (2019). Inherent versus induced protein flexibility: Comparisons within and between apo and holo structures. *PLOS Computational Biology*, *15*(1), e1006705. https://doi.org/10.1371/journal.pcbi.1006705

Cortes-Hernandez, P., & Domínguez-Ramírez, L. (2017). Role of cis-trans proline isomerization in the function of pathogenic enterobacterial Periplasmic Binding Proteins. *PLoS ONE*, *12*(11). https://doi.org/10.1371/journal.pone.0188935

Costa, S. P., Dias, N. M., Melo, L. D. R., Azeredo, J., Santos, S. B., & Carvalho, C. M. (2020). A novel flow cytometry assay based on bacteriophage-derived proteins for Staphylococcus detection in blood. *Scientific Reports*, *10*, 6260. https://doi.org/10.1038/s41598-020-62533-7

Daniels, R., Mellroth, P., Bernsel, A., Neiers, F., Normark, S., von Heijne, G., & Henriques-Normark, B. (2010). Disulfide Bond Formation and Cysteine Exclusion in Gram-positive Bacteria♦. *The Journal of Biological Chemistry*, *285*(5), 3300–3309. https://doi.org/10.1074/jbc.M109.081398

Dannenberg, J. J. (1998). An Introduction to Hydrogen Bonding By George A. Jeffrey (University of Pittsburgh). Oxford University Press: New York and Oxford. 1997. ix + 303 pp. $60.00. ISBN 0-19-509549-9. *Journal of the American Chemical Society*, *120*(22), 5604–5604. https://doi.org/10.1021/ja9756331

D'Aquino, J. A., & Ringe, D. (2003). Determinants of the Src Homology Domain 3-Like Fold. *Journal of Bacteriology*, *185*(14), 4081–4086. https://doi.org/10.1128/JB.185.14.4081-4086.2003

de Groot, N. S., & Ventura, S. (2006). Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Letters*, *580*(27), 6471–6476. https://doi.org/10.1016/j.febslet.2006.10.071

de Vries, S. J., van Dijk, M., & Bonvin, A. M. J. J. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols*, *5*(5), 883–897. https://doi.org/10.1038/nprot.2010.32

DeHart, H. P., Heath, H. E., Heath, L. S., LeBlanc, P. A., & Sloan, G. L. (1995). The lysostaphin endopeptidase resistance gene (epr) specifies modification of peptidoglycan cross bridges in Staphylococcus simulans and Staphylococcus aureus. *Applied and Environmental Microbiology*, *61*(4), 1475–1479. https://doi.org/10.1128/aem.61.4.1475-1479.1995

Díaz De Rienzo, M. A., Stevenson, P., Marchant, R., & Banat, I. M. (2016). Antibacterial properties of biosurfactants against selected Gram-positive and -negative bacteria. *FEMS Microbiology Letters*, *363*(2), fnv224. https://doi.org/10.1093/femsle/fnv224

Dideberg, O., Charlier, P., Dive, G., Joris, B., Frère, J. M., & Ghuysen, J. M. (1982). Structure of a Zn2+-containing D-alanyl-D-alanine-cleaving carboxypeptidase at 2.5 A resolution. *Nature*, *299*(5882), 469–470. https://doi.org/10.1038/299469a0

Ding, Y., Zhang, Y., Huang, C., Wang, J., & Wang, X. (2020). An Endolysin LysSE24 by Bacteriophage LPSE1 Confers Specific Bactericidal Activity against Multidrug-Resistant Salmonella Strains. *Microorganisms*, *8*(5), 737. https://doi.org/10.3390/microorganisms8050737

Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, *125*(7), 1731–1737. https://doi.org/10.1021/ja026939x

Dong, Q., Wang, J., Yang, H., Wei, C., Yu, J., Zhang, Y., Huang, Y., Zhang, X.-E., & Wei, H. (2015). Construction of a chimeric lysin Ply187N-V12C with extended lytic activity against staphylococci and streptococci. *Microbial Biotechnology*, *8*(2), 210–220. https://doi.org/10.1111/1751-7915.12166

Donovan, D. M., Dong, S., Garrett, W., Rousseau, G. M., Moineau, S., & Pritchard, D. G. (2006). Peptidoglycan hydrolase fusions maintain their parental specificities. *Applied and Environmental Microbiology*, *72*(4), 2988–2996. https://doi.org/10.1128/AEM.72.4.2988-2996.2006

Donovan, D. M., & Foster-Frey, J. (2008). LambdaSa2 prophage endolysin requires Cpl-7-binding domains and amidase-5 domain for antimicrobial lysis of streptococci. *FEMS Microbiology Letters*, *287*(1), 22–33. https://doi.org/10.1111/j.1574-6968.2008.01287.x

Donovan, D. M., Foster-Frey, J., Dong, S., Rousseau, G. M., Moineau, S., & Pritchard, D. G. (2006). The Cell Lysis Activity of the Streptococcus agalactiae Bacteriophage B30 Endolysin Relies on the Cysteine, Histidine-Dependent Amidohydrolase/Peptidase Domain. *Applied and Environmental Microbiology*, *72*(7), 5108–5112. https://doi.org/10.1128/AEM.03065-05

Donovan, D. M., Lardeo, M., & Foster-Frey, J. (2006). Lysis of staphylococcal mastitis pathogens by bacteriophage phi11 endolysin. *FEMS Microbiology Letters*, *265*(1), 133–139. https://doi.org/10.1111/j.1574-6968.2006.00483.x

Dorosky, R. J., Lola, S. L., Brown, H. A., Schreier, J. E., Dreher-Lesnick, S. M., & Stibitz, S. (2023). Characterization of Lactobacilli Phage Endolysins and Their Functional Domains–Potential Live Biotherapeutic Testing Reagents. *Viruses*, *15*(10), 1986. https://doi.org/10.3390/v15101986

Doyle, R. J., & Marquis, R. E. (1994). Elastic, flexible peptidoglycan and bacterial cell wall properties. *Trends in Microbiology*, *2*(2), 57–60. https://doi.org/10.1016/0966-842x(94)90127-9

Duchêne, M.-C., Rolain, T., Knoops, A., Courtin, P., Chapot-Chartier, M.-P., Dufrêne, Y. F., Hallet, B. F., & Hols, P. (2019). Distinct and Specific Role of NlpC/P60 Endopeptidases LytA and LytB in Cell Elongation and Division of Lactobacillus plantarum. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.00713

Efthimiou, G., Tsiamis, G., Typas, M. A., & Pappas, K. M. (2019). Transcriptomic Adjustments of Staphylococcus aureus COL (MRSA) Forming Biofilms Under Acidic and Alkaline Conditions. *Frontiers in Microbiology*, *10*, 2393. https://doi.org/10.3389/fmicb.2019.02393

Etobayeva, I., Linden, S. B., Alem, F., Harb, L., Rizkalla, L., Mosier, P. D., Johnson, A. A., Temple, L., Hakami, R. M., & Nelson, D. C. (2018). Discovery and Biophysical characterization of PlyP56, PlyN74, and PlyTB40—Bacillus Specific Endolysins. *Viruses*, *10*(5), 276. https://doi.org/10.3390/v10050276

Falk, I. S. (1918). THE INFLUENCE OF CERTAIN SALTS ON ENZYME ACTION. *Journal of Biological Chemistry*, *36*(1), 229–247. https://doi.org/10.1016/S0021-9258(18)86433-X

Feng, Y., Ming, T., Zhou, J., Lu, C., Wang, R., & Su, X. (2022). The Response and Survival Mechanisms of Staphylococcus aureus under High Salinity Stress in Salted Foods. *Foods*, *11*(10), 1503. https://doi.org/10.3390/foods11101503

Fernandes, S., Proença, D., Cantante, C., Silva, F. A., Leandro, C., Lourenço, S., Milheiriço, C., de Lencastre, H., Cavaco-Silva, P., Pimentel, M., & São-José, C. (2012). Novel Chimerical Endolysins with Broad Antimicrobial Activity Against Methicillin-Resistant Staphylococcus aureus. *Microbial Drug Resistance*, *18*(3), 333–343. https://doi.org/10.1089/mdr.2012.0025

Filatova, L. Y., Donovan, D. M., Foster-Frey, J., Pugachev, V. G., Dmitrieva, N. F., Chubar, T. A., Klyachko, N. L., & Kabanov, A. V. (2015). Bacteriophage phi11 lysin: Physicochemical characterization and comparison with phage phi80α lysin. *Enzyme and Microbial Technology*, *73–74*, 51–58. https://doi.org/10.1016/j.enzmictec.2015.03.005

Fischetti, V. A. (2008). Bacteriophage lysins as effective antibacterials. *Current Opinion in Microbiology*, *11*(5), 393–400. https://doi.org/10.1016/j.mib.2008.09.012

Fiser, A., Do, R. K., & Sali, A. (2000). Modeling of loops in protein structures. *Protein Science : A Publication of the Protein Society*, *9*(9), 1753–1773.

Flahaut, S., Frere, J., Boutibonnes, P., & Auffray, Y. (1996). Comparison of the bile salts and sodium dodecyl sulfate stress responses in Enterococcus faecalis. *Applied and Environmental Microbiology*, *62*(7), 2416–2420.

Fokine, A., Miroshnikov, K. A., Shneider, M. M., Mesyanzhinov, V. V., & Rossmann, M. G. (2008). Structure of the bacteriophage phi KZ lytic transglycosylase gp144. *The Journal of Biological Chemistry*, *283*(11), 7242–7250. https://doi.org/10.1074/jbc.M709398200

Foster, S. J., & Popham, D. L. (2001). Structure and Synthesis of Cell Wall, Spore Cortex, Teichoic Acids, S-Layers, and Capsules. In *Bacillus subtilis and Its Closest Relatives* (pp. 21–41). John Wiley & Sons, Ltd. https://doi.org/10.1128/9781555817992.ch4

Francois, P., Pittet, D., Bento, M., Pepey, B., Vaudaux, P., Lew, D., & Schrenzel, J. (2003). Rapid Detection of Methicillin-Resistant Staphylococcus aureus Directly from Sterile or Nonsterile Clinical Samples by a New Molecular Assay. *Journal of Clinical Microbiology*, *41*(1), 254–260. https://doi.org/10.1128/JCM.41.1.254-260.2003

French, G. L. (2009). Methods for screening for methicillin-resistant Staphylococcus aureus carriage. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *15 Suppl 7*, 10–16. https://doi.org/10.1111/j.1469-0691.2009.03092.x

Fujiki, J., Nakamura, T., Furusawa, T., Ohno, H., Takahashi, H., Kitana, J., Usui, M., Higuchi, H., Tanji, Y., Tamura, Y., & Iwano, H. (2018). Characterization of the Lytic Capability of a LysK-Like Endolysin, Lys-phiSA012, Derived from a Polyvalent Staphylococcus aureus Bacteriophage. *Pharmaceuticals*, *11*(1). https://doi.org/10.3390/ph11010025

Gaeng, S., Scherer, S., Neve, H., & Loessner, M. J. (2000). Gene cloning and expression and secretion of Listeria monocytogenes bacteriophage-lytic enzymes in Lactococcus lactis. *Applied and Environmental Microbiology*, *66*(7), 2951–2958. https://doi.org/10.1128/AEM.66.7.2951-2958.2000 Gallagher, S. R., & Wiley, E. A. (2012). *Current Protocols Essential Laboratory Techniques*. John Wiley & Sons.

García, E., García, J. L., García, P., Arrarás, A., Sánchez-Puelles, J. M., & López, R. (1988). Molecular evolution of lytic enzymes of Streptococcus pneumoniae and its bacteriophages. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(3), 914–918.

Garde, S., Chodisetti, P. K., & Reddy, M. (2021). Peptidoglycan: Structure, Synthesis, and Regulation. *EcoSal Plus*, *9*(2). https://doi.org/10.1128/ecosalplus.ESP-0010-2020

Garg, S. K. (2021). Antibiotic misuse during COVID-19 Pandemic: A Recipe for Disaster. *Indian Journal of Critical Care Medicine : Peer-Reviewed, Official Publication of Indian Society of Critical Care Medicine*, *25*(6), 617–619. https://doi.org/10.5005/jp-journals-10071-23862

Garvey, K. J., Saedi, M. S., & Ito, J. (1986). Nucleotide sequence of Bacillus phage phi 29 genes 14 and 15: Homology of gene 15 with other phage lysozymes. *Nucleic Acids Research*, *14*(24), 10001–10008.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press. https://doi.org/10.1385/1-59259-890-0:571

Genheden, S., Mikulskis, P., Hu, L., Kongsted, J., Söderhjelm, P., & Ryde, U. (2011). Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *Journal of the American Chemical Society*, *133*(33), 13081–13092. https://doi.org/10.1021/ja202972m

Gerova, M., Halgasova, N., Ugorcakova, J., & Bukovska, G. (2011). Endolysin of bacteriophage BFK20: Evidence of a catalytic and a cell wall binding domain. *FEMS Microbiology Letters*, *321*(2), 83–91. https://doi.org/10.1111/j.1574-6968.2011.02312.x

Gibson, A. K., Baffoe-Bonnie, H., Penley, M. J., Lin, J., Owens, R., Khalid, A., & Morran, L. T. (2020). The evolution of parasite host range in heterogeneous host populations. *Journal of Evolutionary Biology*, *33*(6), 773–782. https://doi.org/10.1111/jeb.13608

Gill, J. J. (2014). Revised Genome Sequence of Staphylococcus aureus Bacteriophage K. *Genome Announcements*, *2*(1), e01173-13. https://doi.org/10.1128/genomeA.01173-13

Gilmer, D. B., Schmitz, J. E., Thandar, M., Euler, C. W., & Fischetti, V. A. (2017). The Phage Lysin PlySs2 Decolonizes Streptococcus suis from Murine Intranasal Mucosa. *PLoS ONE*, *12*(1), e0169180. https://doi.org/10.1371/journal.pone.0169180

Gloor, G. B., Martin, L. C., Wahl, L. M., & Dunn, S. D. (2005). Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. *Biochemistry*, *44*(19), 7156–7165. https://doi.org/10.1021/bi050293e

Gondil, V. S., Harjai, K., & Chhibber, S. (2020). Endolysins as emerging alternative therapeutic agents to counter drug-resistant infections. *International Journal of Antimicrobial Agents*, *55*(2), 105844. https://doi.org/10.1016/j.ijantimicag.2019.11.001

Gonzalez-Delgado, L. S., Walters-Morgan, H., Salamaga, B., Robertson, A. J., Hounslow, A. M., Jagielska, E., Sabała, I., Williamson, M. P., Lovering, A. L., & Mesnage, S. (2020). Two-site recognition of Staphylococcus aureus peptidoglycan by lysostaphin SH3b. *Nature Chemical Biology*, *16*(1), 24–30. https://doi.org/10.1038/s41589-019-0393-4

Gordon, E., Horsefield, R., Swarts, H. G. P., de Pont, J. J. H. H. M., Neutze, R., & Snijder, A. (2008). Effective high-throughput overproduction of membrane proteins in Escherichia coli. *Protein Expression and Purification*, *62*(1), 1–8. https://doi.org/10.1016/j.pep.2008.07.005

Graef, J., Ehrt, C., & Rarey, M. (2023). Binding Site Detection Remastered: Enabling Fast, Robust, and Reliable Binding Site Detection and Descriptor Calculation with DoGSite3. *Journal of Chemical Information and Modeling*, *63*(10), 3128–3137. https://doi.org/10.1021/acs.jcim.3c00336

Gray, T. M., & Matthews, B. W. (1984). Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins. *Journal of Molecular Biology*, *175*(1), 75–81. https://doi.org/10.1016/0022-2836(84)90446-7

Grisold, A. J., Leitner, E., Mühlbauer, G., Marth, E., & Kessler, H. H. (2002). Detection of methicillin-resistant Staphylococcus aureus and simultaneous confirmation by automated nucleic acid extraction and real-time PCR. *Journal of Clinical Microbiology*, *40*(7), 2392–2397. https://doi.org/10.1128/jcm.40.7.2392-2397.2002

Gründling, A., & Schneewind, O. (2006). Cross-Linked Peptidoglycan Mediates Lysostaphin Binding to the Cell Wall Envelope of Staphylococcus aureus. *Journal of Bacteriology*, *188*(7), 2463–2472. https://doi.org/10.1128/JB.188.7.2463-2472.2006

Gründling, A., Smith, D. L., Bläsi, U., & Young, R. (2000). Dimerization between the Holin and Holin Inhibitor of Phage λ. *Journal of Bacteriology*, *182*(21), 6075–6081.

Gu, J., Feng, Y., Feng, X., Sun, C., Lei, L., Ding, W., Niu, F., Jiao, L., Yang, M., Li, Y., Liu, X., Song, J., Cui, Z., Han, D., Du, C., Yang, Y., Ouyang, S., Liu, Z.-J., & Han, W. (2014). Structural and Biophysical characterization Reveals LysGH15 as an Unprecedented "EF-Hand-Like" Calcium-Binding Phage Lysin. *PLoS Pathogens*, *10*(5). https://doi.org/10.1371/journal.ppat.1004109

Gu, J., Lu, R., Liu, X., Han, W., Lei, L., Gao, Y., Zhao, H., Li, Y., & Diao, Y. (2011). LysGH15B, the SH3b domain of staphylococcal phage endolysin LysGH15, retains high affinity to staphylococci. *Current Microbiology*, *63*(6), 538–542. https://doi.org/10.1007/s00284-011-0018-y

Gu, J., Xu, W., Lei, L., Huang, J., Feng, X., Sun, C., Du, C., Zuo, J., Li, Y., Du, T., Li, L., & Han, W. (2011). LysGH15, a novel bacteriophage lysin, protects a murine bacteremia model efficiently against lethal methicillin-resistant Staphylococcus aureus infection. *Journal of Clinical Microbiology*, *49*(1), 111–117.https://doi.org/10.1128/JCM.01144-10

Guo, M., Feng, C., Ren, J., Zhuang, X., Zhang, Y., Zhu, Y., Dong, K., He, P., Guo, X., & Qin, J. (2017). A Novel Antimicrobial Endolysin, LysPA26, against Pseudomonas aeruginosa. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00293

Gutiérrez, D., Fernández, L., Rodríguez, A., & García, P. (2018). Are Phage Lytic Proteins the Secret Weapon To Kill Staphylococcus aureus? *mBio*, *9*(1). https://doi.org/10.1128/mBio.01923-17

Ha, V. L. T., Erlitzki, N., Farahat, A. A., Kumar, A., Boykin, D. W., & Poon, G. M. K. (2020). Dissecting Dynamic and Hydration Contributions to Sequence-Dependent DNA Minor Groove Recognition. *Biophysical Journal*, *119*(7), 1402–1415.https://doi.org/10.1016/j.bpj.2020.08.013

Haase-Pettingell, C., Betts, S., Raso, S. W., Stuart, L., Robinson, A., & King, J. (2001). Role for cysteine residues in the in vivo folding and assembly of the phage P22 tailspike. *Protein Science : A Publication of the Protein Society*, *10*(2), 397–410.

Haddad Kashani, H., Schmelcher, M., Sabzalipoor, H., Seyed Hosseini, E., & Moniri, R. (2017). Recombinant Endolysins as Potential Therapeutics against Antibiotic-Resistant Staphylococcus aureus: Current Status of Research and Novel Delivery Strategies. *Clinical Microbiology Reviews*, *31*(1). https://doi.org/10.1128/CMR.00071-17

Haddad, Y., Adam, V., & Heger, Z. (2020). Ten quick tips for homology modeling of high-resolution protein 3D structures. *PLoS Computational Biology*, *16*(4), e1007449.https://doi.org/10.1371/journal.pcbi.1007449

Hall, A. R., Scanlan, P. D., & Buckling, A. (2011). Bacteria-phage coevolution and the emergence of generalist pathogens. *The American Naturalist*, *177*(1), 44–53.https://doi.org/10.1086/657441

Hallak, G., Neuner, B., Schefold, J. C., Gorzelniak, K., Rapsch, B., Pfüller, R., Stengel, D., Wellmann, J., Ekkernkamp, A., & Walter, M. (2016). Preemptive Isolation Precautions of Patients at High Risk for Methicillin-Resistant Staphylococcus aureus in Combination With Ultrarapid Polymerase Chain Reaction Screening as an Effective Tool for Infection Control. *Infection Control and Hospital Epidemiology*, *37*(12), 1489–1491. https://doi.org/10.1017/ice.2016.204

Hamilton, S., Odili, J., Pacifico, M. D., Wilson, G. D., & Kupsch, J.-M. (2003). Effect of imidazole on the solubility of a his-tagged antibody fragment. *Hybridoma and Hybridomics*, *22*(6), 347–355. https://doi.org/10.1089/153685903771797048

Harkins, C. P., Pichon, B., Doumith, M., Parkhill, J., Westh, H., Tomasz, A., de Lencastre, H., Bentley, S. D., Kearns, A. M., & Holden, M. T. G. (2017). Methicillin-resistant Staphylococcus aureus emerged long before the introduction of methicillin into clinical practice. *Genome Biology*, *18*, 130. https://doi.org/10.1186/s13059-017-1252-9

Harris, C., & Hansen, J. M. (2012). Oxidative stress, thiols, and redox profiles. *Methods in Molecular Biology (Clifton, N.J.)*, *889*, 325–346. https://doi.org/10.1007/978-1-61779-867-2_21

Hassan, Z., Mustafa, S., Rahim, R. A., & Isa, N. M. (2016). Anti-breast cancer effects of live, heat-killed and cytoplasmic fractions of Enterococcus faecalis and Staphylococcus hominis isolated from human breast milk. *In Vitro Cellular & Developmental Biology. Animal*, *52*(3), 337–348. https://doi.org/10.1007/s11626-015-9978-8

Hatahet, F., Boyd, D., & Beckwith, J. (2014). Disulfide bond formation in prokaryotes: History, diversity and design. *Biochimica Et Biophysica Acta*, *1844*(8), 1402–1414. https://doi.org/10.1016/j.bbapap.2014.02.014

Hayat, S. M. G., Farahani, N., Golichenari, B., & Sahebkar, A. (2018). Recombinant Protein Expression in Escherichia coli (E.coli): What We Need to Know. *Current Pharmaceutical Design*, *24*(6), 718–725. https://doi.org/10.2174/1381612824666180131121940

Hayhurst, E. J., Kailas, L., Hobbs, J. K., & Foster, S. J. (2008). Cell wall peptidoglycan architecture in Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14603–14608. https://doi.org/10.1073/pnas.0804138105

He, C., & Ohnishi, K. (2017). Efficient renaturation of inclusion body proteins denatured by SDS. *Biochemical and Biophysical Research Communications*, *490*(4),1250–1253.https://doi.org/10.1016/j.bbrc.2017.07.003

Hennekinne, J.-A., De Buyser, M.-L., & Dragacci, S. (2012). Staphylococcus aureus and its food poisoning toxins: Characterization and outbreak investigation. *FEMS Microbiology Reviews*, *36*(4), 815–836. https://doi.org/10.1111/j.1574-6976.2011.00311.x

Hermoso, J. A., García, J. L., & García, P. (2007). Taking aim on bacterial pathogens: From phage therapy to enzybiotics. *Current Opinion in Microbiology*, *10*(5), 461–472.https://doi.org/10.1016/j.mib.2007.08.002

Hermoso, J. A., Monterroso, B., Albert, A., Galán, B., Ahrazem, O., García, P., Martínez-Ripoll, M., García, J. L., & Menéndez, M. (2003). Structural Basis for Selective Recognition of Pneumococcal Cell Wall by Modular Endolysin from Phage Cp-1. *Structure*, *11*(10), 1239–1249. https://doi.org/10.1016/j.str.2003.09.005

Heselpoth, R. D., Swift, S. M., Linden, S. B., Mitchell, M. S., & Nelson, D. C. (2017). Enzybiotics: Endolysins and Bacteriocins. In D. Harper, S. Abedon, B. Burrowes, & M. McConville (Eds.), *Bacteriophages: Biology, Technology, Therapy* (pp. 1–42). Springer International Publishing. https://doi.org/10.1007/978-3-319-40598-8_34-1

Higuchi, R., Krummel, B., & Saiki, R. K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: Study of protein and DNA interactions. *Nucleic Acids Research*, *16*(15), 7351–7367. https://doi.org/10.1093/nar/16.15.7351

Hirakawa, H., Akita, H., Fujiwara, T., Sugai, M., & Kuhara, S. (2009). Structural insight into the binding mode between the targeting domain of ALE-1 (92AA) and pentaglycine of peptidoglycan. *Protein Engineering, Design & Selection: PEDS*, *22*(7), 385–391. https://doi.org/10.1093/protein/gzp014

Hjertén, S., Sparrman, M., & Liao, J. (1988). Purification of membrane proteins in SDS and subsequent renaturation. *Biochimica Et Biophysica Acta*, *939*(3), 476–484.https://doi.org/10.1016/0005-2736(88)90094-6

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, *77*(1), 51–59. https://doi.org/10.1016/0378-1119(89)90358-2

Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular Concepts*, *1*(3–4), 271–283. https://doi.org/10.1515/BMC.2010.022

Horgan, M., O'Flynn, G., Garry, J., Cooney, J., Coffey, A., Fitzgerald, G. F., Ross, R. P., & McAuliffe, O. (2009). Phage lysin LysK can be truncated to its CHAP domain and retain lytic activity against live antibiotic-resistant staphylococci. *Applied and Environmental Microbiology*, *75*(3), 872–874. https://doi.org/10.1128/AEM.01831-08

Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., & Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. *Gene*, *77*(1), 61–68. https://doi.org/10.1016/0378-1119(89)90359-4

Huang, Y., Yang, H., Yu, J., & Wei, H. (2015). Molecular dissection of phage lysin PlySs2: Integrity of the catalytic and cell wall binding domains is essential for its broad lytic activity. *Virologica Sinica*, *30*(1), 45–51. https://doi.org/10.1007/s12250-014-3535-6

Hunke, S., & Betton, J.-M. (2003). Temperature effect on inclusion body formation and stress response in the periplasm of Escherichia coli. *Molecular Microbiology*, *50*(5), 1579–1589. https://doi.org/10.1046/j.1365-2958.2003.03785.x

Jacob, F., Fuerst, R., & Wollman, E. L. (1957). [Defective lysogenic bacteria. II. Physiological types resulting from prophage mutations]. *Annales De l'Institut Pasteur*, *93*(6), 724–753.

Jafri, A. K., Reisner, B. S., & Woods, G. L. (2000). Evaluation of a latex agglutination assay for rapid detection of oxacillin resistant Staphylococcus aureus. *Diagnostic Microbiology and Infectious Disease*, *36*(1), 57–59. https://doi.org/10.1016/s0732-8893(99)00105-4

Jalilsood, T., Baradaran, A., Song, A. A.-L., Foo, H. L., Mustafa, S., Saad, W. Z., Yusoff, K., & Rahim, R. A. (2015). Inhibition of pathogenic and spoilage bacteria by a novel biofilm-forming Lactobacillus isolate: A potential host for the expression of heterologous proteins. *Microbial Cell Factories*, *14*, 96. https://doi.org/10.1186/s12934-015-0283-8

Jarábková, V., Tišáková, L., & Godány, A. (2015). Phage Endolysin: A Way To Understand A Binding Function Of C-Terminal Domains A Mini Review. *Nova Biotechnologica et Chimica*, *14*(2). https://content.sciendo.com/view/journals/nbec/14/2/article-p117.xml

Jayaraman, V., Toledo-Patiño, S., Noda-García, L., & Laurino, P. (2022). Mechanisms of protein evolution. *Protein Science : A Publication of the Protein Society*, *31*(7), e4362. https://doi.org/10.1002/pro.4362

Jensen, S. E., & Campbell, J. N. (1976). Peptidoglycan biosynthesis in Micrococcus luteus (sodonensis): Transglycosidase and phosphodiesterase activities in membrane preparations. *Journal of Bacteriology*, *127*(1), 309–318.

Ji, W., Huang, Q., Sun, L., Wang, H., Yan, Y., & Sun, J. (2015). A novel endolysin disrupts Streptococcus suis with high efficiency. *FEMS Microbiology Letters*, *362*(24), fnv205. https://doi.org/10.1093/femsle/fnv205

Jonas, D., Speck, M., Daschner, F. D., & Grundmann, H. (2002). Rapid PCR-Based Identification of Methicillin-Resistant Staphylococcus aureus from Screening Swabs. *Journal of Clinical Microbiology*, *40*(5), 1821–1823. https://doi.org/10.1128/JCM.40.5.1821-1823.2002

Kaneko, T., Li, L., & Li, S. S.-C. (2008). The SH3 domain—A family of versatile peptide- and protein-recognition module. *Frontiers in Bioscience: A Journal and Virtual Library*, *13*, 4938–4952. https://doi.org/10.2741/3053

Kathir, K. M., Kumar, T. K. S., Rajalingam, D., & Yu, C. (2005). Time-dependent Changes in the Denatured State(s) Influence the Folding Mechanism of an All β-Sheet Protein *. *Journal of Biological Chemistry*, *280*(33), 29682–29688. https://doi.org/10.1074/jbc.M504389200

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Khairulddin, N., Bishop, L., Lamagni, T., Sharland, M., & Duckworth, G. (2004). Emergence of methicillin resistant Staphylococcus aureus (MRSA) bacteraemia among children in England and Wales, 1990–2001. *Archives of Disease in Childhood*, *89*(4), 378–379. https://doi.org/10.1136/adc.2003.028712

Kim, C., & Lee, B. (2007). Accuracy of structure-based sequence alignment of automatic methods. *BMC Bioinformatics*, *8*, 355. https://doi.org/10.1186/1471-2105-8-355

Kim, H., & Seo, J. (2023). A Novel Strategy to Identify Endolysins with Lytic Activity against Methicillin-Resistant Staphylococcus aureus. *International Journal of Molecular Sciences*, *24*(6), 5772. https://doi.org/10.3390/ijms24065772

Kim, S. J., Chang, J., & Singh, M. (2015). Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR. *Biochimica Et Biophysica Acta*, *1848*(1 Pt B), 350–362. https://doi.org/10.1016/j.bbamem.2014.05.031

KOCUR, M., PAČOVA, Z., & MARTINEC, T. (1972). Taxonomic Status of Micrococcus luteus (Schroeter 1872) Cohn 1872, and Designation of the Neotype Strain. *International Journal of Systematic and Evolutionary Microbiology*, *22*(4), 218–223. https://doi.org/10.1099/00207713-22-4-218

Kong, M., Shin, J. H., Heu, S., Park, J.-K., & Ryu, S. (2017). Lateral flow assay-based bacterial detection using engineered cell wall binding domains of a phage endolysin. *Biosensors and Bioelectronics*, *96*, 173–177. https://doi.org/10.1016/j.bios.2017.05.010

Kong, M., Sim, J., Kang, T., Nguyen, H. H., Park, H. K., Chung, B. H., & Ryu, S. (2015). A novel and highly specific phage endolysin cell wall binding domain for detection of Bacillus cereus. *European Biophysics Journal*, *44*(6), 437–446. https://doi.org/10.1007/s00249-015-1044-7

Kono, H. (2013). Two Lysine Residues Enhance the Flexibility of DNA Wrapping around CENP-A Containing Nucleosome Core Particle. *Biophysical Journal*, *104*(2),579a.https://doi.org/10.1016/j.bpj.2012.11.3213

Kopito, R. R. (2000). Aggresomes, inclusion bodies and protein aggregation. *Trends in Cell Biology*, *10*(12), 524–530. https://doi.org/10.1016/s0962-8924(00)01852-3

Korndörfer, I. P., Kanitz, A., Danzer, J., Zimmer, M., Loessner, M. J., & Skerra, A. (2008). Structural analysis of the L-alanoyl-D-glutamate endopeptidase domain of Listeria bacteriophage endolysin Ply500 reveals a new member of the LAS peptidase family. *Acta Crystallographica. Section D, Biological Crystallography*, *64*(Pt 6), 644–650. https://doi.org/10.1107/S0907444908007890

Kragh-Hansen, U., le Maire, M., & Møller, J. V. (1998). The mechanism of detergent solubilization of liposomes and protein-containing membranes. *Biophysical Journal*,*75*(6),2932–2946.

Krishnan, M., Tham, H. Y., Wan Nur Ismah, W. A. K., Yusoff, K., & Song, A. A.-L. (2024). Effect of Domain Manipulation in the Staphylococcal Phage Endolysin, Endo88, on Lytic Efficiency and Host Range. *Molecular Biotechnology*.https://doi.org/10.1007/s12033-024-01216-4

Krogh, S., Jørgensen, S. T., & Devine, K. M. (1998). Lysis genes of the Bacillus subtilis defective prophage PBSX. *Journal of Bacteriology*, *180*(8), 2110–2117.https://doi.org/10.1128/JB.180.8.2110-2117.1998

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kwan, T., Liu, J., DuBow, M., Gros, P., & Pelletier, J. (2005). The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(14), 5174–5179.https://doi.org/10.1073/pnas.0501140102

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews. Microbiology*, *8*(5), 317–327. https://doi.org/10.1038/nrmicro2315

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/10.1107/S0021889892009944

Layec, S., Gérard, J., Legué, V., Chapot-Chartier, M.-P., Courtin, P., Borges, F., Decaris, B., & Leblond-Bourget, N. (2009). The CHAP domain of Cse functions as an endopeptidase that acts at mature septa to promote Streptococcus thermophilus cell separation. *Molecular Microbiology*, *71*(5), 1205–1217.https://doi.org/10.1111/j.1365-2958.2009.06595.x

Lee, J. J., Bruley, D. F., & Kang, K. A. (2007). Effect of pH and imidazole on protein C purification from Cohn fraction IV-1 by IMAC. *Advances in Experimental Medicine and Biology*, *599*, 61–66. https://doi.org/10.1007/978-0-387-71764-7_9

Lee, K. O., Kong, M., Kim, I., Bai, J., Cha, S., Kim, B., Ryu, K.-S., Ryu, S., & Suh, J.-Y. (2019). Structural Basis for Cell-Wall Recognition by Bacteriophage PBC5 Endolysin. *Structure (London, England: 1993)*, *27*(9), 1355-1365.e4. https://doi.org/10.1016/j.str.2019.07.001

Leibly, D. J., Nguyen, T. N., Kao, L. T., Hewitt, S. N., Barrett, L. K., & Van Voorhis, W. C. (2012). Stabilizing Additives Added during Cell Lysis Aid in the Solubilization of Recombinant Proteins. *PLoS ONE*, *7*(12), e52482. https://doi.org/10.1371/journal.pone.0052482

Leprince, A., Nuytten, M., July, E., Tesseur, C., & Mahillon, J. (2022). Getting Outside the Cell: Versatile Holin Strategies Used by Distinct Phages to Leave Their Bacillus thuringiensis Host. *Journal of Virology*, *96*(14), e00696-22. https://doi.org/10.1128/jvi.00696-22

Li, S. C., Goto, N. K., Williams, K. A., & Deber, C. M. (1996). Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(13), 6676–6681. https://doi.org/10.1073/pnas.93.13.6676

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Lin, D. M., Koskella, B., & Lin, H. C. (2017). Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World Journal of Gastrointestinal Pharmacology and Therapeutics*, *8*(3), 162–173. https://doi.org/10.4292/wjgpt.v8.i3.162

Linden, S. B., Zhang, H., Heselpoth, R. D., Shen, Y., Schmelcher, M., Eichenseher, F., & Nelson, D. C. (2015). Biochemical and biophysical characterization of PlyGRCS, a bacteriophage endolysin active against methicillin-resistant Staphylococcus aureus. *Applied Microbiology and Biotechnology*, *99*(2), 741–752. https://doi.org/10.1007/s00253-014-5930-1

Liu, H., Hu, Z., Li, M., Yang, Y., Lu, S., & Rao, X. (2023). Therapeutic potential of bacteriophage endolysins for infections caused by Gram-positive bacteria. *Journal of Biomedical Science*, *30*(1), 29. https://doi.org/10.1186/s12929-023-00919-1

Loell, K., & Nanda, V. (2018). Marginal protein stability drives subcellular proteome isoelectric point. *Proceedings of the National Academy of Sciences*, *115*(46), 11778–11783. https://doi.org/10.1073/pnas.1809098115

Loessner, M. J. (2005). Bacteriophage endolysins—Current state of research and applications. *Current Opinion in Microbiology*, *8*(4), 480–487. https://doi.org/10.1016/j.mib.2005.06.002

Loessner, M. J., Gaeng, S., & Scherer, S. (1999). Evidence for a holin-like protein gene fully embedded out of frame in the endolysin gene of Staphylococcus aureus bacteriophage 187. *Journal of Bacteriology*, *181*(15), 4452–4460.

Loessner, M. J., Gaeng, S., Wendlinger, G., Maier, S. K., & Scherer, S. (1998). The two-component lysis system of Staphylococcus aureus bacteriophage Twort: A large TTG-start holin and an associated amidase endolysin. *FEMS Microbiology Letters*, *162*(2), 265–274.

Loessner, M. J., Kramer, K., Ebel, F., & Scherer, S. (2002). C-terminal domains of Listeria monocytogenes bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. *Molecular Microbiology*, *44*(2), 335–349.

Longshore-Neate, F., Ceravolo, C., Masuga, C., Tahti, E. F., Blount, J. M., Smith, S. N., & Amacher, J. F. (2024). The conformation of the nSrc specificity-determining loop in the Src SH3 domain is modulated by a WX conserved sequence motif found in SH3 domains. *Frontiers in Molecular Biosciences*, *11*.https://doi.org/10.3389/fmolb.2024.1487276

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. *Proteins*, *50*(3), 437–450. https://doi.org/10.1002/prot.10286

Low, L. Y., Yang, C., Perego, M., Osterman, A., & Liddington, R. (2011). Role of Net Charge on Catalytic Domain and Influence of Cell Wall Binding Domain on Bactericidal Activity, Specificity, and Host Range of Phage Lysins. *The Journal of Biological Chemistry*, *286*(39), 34391–34403. https://doi.org/10.1074/jbc.M111.244160

Low, L. Y., Yang, C., Perego, M., Osterman, A., & Liddington, R. C. (2005). Structure and lytic activity of a Bacillus anthracis prophage endolysin. *The Journal of Biological Chemistry*, *280*(42), 35433–35439. https://doi.org/10.1074/jbc.M502723200

Lu, J. Z., Fujiwara, T., Komatsuzawa, H., Sugai, M., & Sakon, J. (2006). Cell wall-targeting domain of glycylglycine endopeptidase distinguishes among peptidoglycan cross-bridges. *The Journal of Biological Chemistry*, *281*(1), 549–558.https://doi.org/10.1074/jbc.M509691200

Lu, N., Sun, Y., Wang, Q., Qiu, Y., Chen, Z., Wen, Y., Wang, S., & Song, Y. (2020). Cloning and characterization of endolysin and holin from Streptomyces avermitilis bacteriophage phiSASD1 as potential novel antibiotic candidates. *International Journal of Biological Macromolecules*, *147*, 980–989. https://doi.org/10.1016/j.ijbiomac.2019.10.065

Lu, Y., Wang, Y., Wang, J., Zhao, Y., Zhong, Q., Li, G., Fu, Z., & Lu, S. (2021). Phage Endolysin LysP108 Showed Promising Antibacterial Potential Against Methicillin-resistant Staphylococcus aureus. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.668430

Maciejewska, B., Olszak, T., & Drulis-Kawa, Z. (2018). Applications of bacteriophages versus phage enzymes to combat and cure bacterial infections: An ambitious and also a realistic application? *Applied Microbiology and Biotechnology*, *102*(6), 2563–2581. https://doi.org/10.1007/s00253-018-8811-1

Maervoet, V. E. T., & Briers, Y. (2016). Synthetic biology of modular proteins. *Bioengineered*,*8*(3),196–202.https://doi.org/10.1080/21655979.2016.1222993

Maestro, B., & Sanz, J. M. (2016). Choline Binding Proteins from Streptococcus pneumoniae: A Dual Role as Enzybiotics and Targets for the Design of New Antimicrobials. *Antibiotics (Basel, Switzerland)*, *5*(2). https://doi.org/10.3390/antibiotics5020021

Mahmoud, A. H., Masters, M. R., Yang, Y., & Lill, M. A. (2020). Elucidating the multiple roles of hydration for accurate protein-ligand binding prediction via deep learning. *Communications Chemistry*, *3*(1), 1–13. https://doi.org/10.1038/s42004-020-0261-x

Malmström, L., Riffle, M., Strauss, C. E. M., Chivian, D., Davis, T. N., Bonneau, R., & Baker, D. (2007). Superfamily Assignments for the Yeast Proteome through Integration of Structure Prediction with the Gene Ontology. *PLoS Biology*, *5*(4), e76. https://doi.org/10.1371/journal.pbio.0050076

Mao, J., Schmelcher, M., Harty, W. J., Foster-Frey, J., & Donovan, D. M. (2013). Chimeric Ply187 endolysin kills Staphylococcus aureus more effectively than the parental enzyme. *FEMS Microbiology Letters*, *342*(1), 30–36. https://doi.org/10.1111/1574-6968.12104

Martin, L. C., Gloor, G. B., Dunn, S. D., & Wahl, L. M. (2005). Using information theory to search for co-evolving residues in proteins. *Bioinformatics (Oxford, England)*, *21*(22), 4116–4124. https://doi.org/10.1093/bioinformatics/bti671

Mayer, B. J. (2001). SH3 domains: Complexity in moderation. *Journal of Cell Science*, *114*(Pt7),1253–1263.

Mayer, B. J., & Eck, M. J. (1995). SH3 Domains: Minding your p's and q's. *Current Biology*, *5*(4), 364–367. https://doi.org/10.1016/S0960-9822(95)00073-X

Mayer, B. J., Hamaguchi, M., & Hanafusa, H. (1988). Characterization of p47gag-crk, a novel oncogene product with sequence similarity to a putative modulatory domain of protein-tyrosine kinases and phospholipase C. *Cold Spring Harbor Symposia on Quantitative Biology*, *53 Pt 2*, 907–914. https://doi.org/10.1101/sqb.1988.053.01.104

McGowan, S., Buckle, A. M., Mitchell, M. S., Hoopes, J. T., Gallagher, D. T., Heselpoth, R. D., Shen, Y., Reboul, C. F., Law, R. H. P., Fischetti, V. A., Whisstock, J. C., & Nelson, D. C. (2012). X-ray crystal structure of the streptococcal specific phage lysin PlyC. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12752–12757. https://doi.org/10.1073/pnas.1208424109

Megat Mazhar Khair, M. H., Tee, A. N., Wahab, N. F., Othman, S. S., Goh, Y. M., Masarudin, M. J., Chong, C. M., In, L. L. A., Gan, H. M., & Song, A. A.-L. (2023). Comprehensive Characterization of a Streptococcus agalactiae Phage Isolated from a Tilapia Farm in Selangor, Malaysia, and Its Potential for Phage Therapy. *Pharmaceuticals (Basel, Switzerland)*, *16*(5), 698. https://doi.org/10.3390/ph16050698

Melo, L. D. R., Brandão, A., Akturk, E., Santos, S. B., & Azeredo, J. (2018). Characterization of a New Staphylococcus aureus Kayvirus Harboring a Lysin Active against Biofilms. *Viruses*, *10*(4), 182. https://doi.org/10.3390/v10040182

Mesnage, S., Dellarole, M., Baxter, N. J., Rouget, J.-B., Dimitrov, J. D., Wang, N., Fujimoto, Y., Hounslow, A. M., Lacroix-Desmazes, S., Fukase, K., Foster, S. J., & Williamson, M. P. (2014). Molecular basis for bacterial peptidoglycan recognition by LysM domains. *Nature Communications*, *5*, 4269. https://doi.org/10.1038/ncomms5269

Miljevic, B., Hedayat, F., Stevanovic, S., Fairfull-Smith, K. E., Bottle, S. E., & Ristovski, Z. D. (2014). To Sonicate or Not to Sonicate PM Filters: Reactive Oxygen Species Generation Upon Ultrasonic Irradiation. *Aerosol Science and Technology*, *48*(12), 1276–1284. https://doi.org/10.1080/02786826.2014.981330

Miotto, O., Heiny, A., Tan, T. W., August, J. T., & Brusic, V. (2008). Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC Bioinformatics*, *9 Suppl 1*(Suppl 1), S18. https://doi.org/10.1186/1471-2105-9-S1-S18

Mishra, A. K., Rawat, M., Viswas, K. N., Abhishek, Kumar, S., & Reddy, M. (2013). Expression and lytic efficacy assessment of the Staphylococcus aureus phage SA4 lysin gene. *Journal of Veterinary Science*, *14*(1), 37–43. https://doi.org/10.4142/jvs.2013.14.1.37

Mitkowski, P., Jagielska, E., Nowak, E., Bujnicki, J. M., Stefaniak, F., Niedziałek, D., Bochtler, M., & Sabała, I. (2019). Structural bases of peptidoglycan recognition by lysostaphin SH3b domain. *Scientific Reports*, *9*. https://doi.org/10.1038/s41598-019-42435-z

Mühlmann, M., Forsten, E., Noack, S., & Büchs, J. (2017). Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. *Microbial Cell Factories*, *16*, 220. https://doi.org/10.1186/s12934-017-0832-4

Murray, E., Draper, L. A., Ross, R. P., & Hill, C. (2021). The Advantages and Challenges of Using Endolysins in a Clinical Setting. *Viruses*, *13*(4), 680. https://doi.org/10.3390/v13040680

Nakamura, T., Kitana, J., Fujiki, J., Takase, M., Iyori, K., Simoike, K., & Iwano, H. (2020). Lytic Activity of Polyvalent Staphylococcal Bacteriophage PhiSA012 and Its Endolysin Lys-PhiSA012 Against Antibiotic-Resistant Staphylococcal Clinical Isolates From Canine Skin Infection Sites. *Frontiers in Medicine*, *7*, 234.https://doi.org/10.3389/fmed.2020.00234

Nakonieczna, A., Topolska-Woś, A., & Łobocka, M. (2024). New bacteriophage-derived lysins, LysJ and LysF, with the potential to control Bacillus anthracis. *Applied Microbiology and Biotechnology*, *108*(1), 76. https://doi.org/10.1007/s00253-023-12839-z

Navarre, W. W., Ton-That, H., Faull, K. F., & Schneewind, O. (1999). Multiple Enzymatic Activities of the Murein Hydrolase from Staphylococcal Phage φ11: IDENTIFICATION OF A d-ALANYL-GLYCINE ENDOPEPTIDASE ACTIVITY*. *Journal of Biological Chemistry*, *274*(22), 15847–15856. https://doi.org/10.1074/jbc.274.22.15847

Nelson, D., Schuch, R., Chahales, P., Zhu, S., & Fischetti, V. A. (2006). PlyC: A multimeric bacteriophage lysin. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(28), 10765–10770. https://doi.org/10.1073/pnas.0604521103

Neudecker, P., Robustelli, P., Cavalli, A., Walsh, P., Lundström, P., Zarrine-Afsar, A., Sharpe, S., Vendruscolo, M., & Kay, L. E. (2012). Structure of an Intermediate State in Protein Folding and Aggregation. *Science*, *336*(6079), 362–366. https://doi.org/10.1126/science.1214203

Nishino, H., Kitamura, M., Okada, S., Miyake, R., Okumura, M., & Muraoka, T. (2022). Cysteine-based protein folding modulators for trapping intermediates and misfolded forms. *RSC Advances*, *12*(41), 26658–26664. https://doi.org/10.1039/D2RA04044A

Obeso, J. M., Martínez, B., Rodríguez, A., & García, P. (2008). Lytic activity of the recombinant staphylococcal bacteriophage PhiH5 endolysin active against Staphylococcus aureus in milk. *International Journal of Food Microbiology*, *128*(2), 212–218. https://doi.org/10.1016/j.ijfoodmicro.2008.08.010

Oechslin, F., Daraspe, J., Giddey, M., Moreillon, P., & Resch, G. (2013). In vitro characterization of PlySK1249, a novel phage lysin, and assessment of its antibacterial activity in a mouse model of Streptococcus agalactiae bacteremia. *Antimicrobial Agents and Chemotherapy*, *57*(12), 6276–6283. https://doi.org/10.1128/AAC.01701-13

Oechslin, F., Zhu, X., Dion, M. B., Shi, R., & Moineau, S. (2022). Phage endolysins are adapted to specific hosts and are evolutionarily dynamic. *PLoS Biology*, *20*(8), e3001740. https://doi.org/10.1371/journal.pbio.3001740

Oechslin, F., Zhu, X., Morency, C., Somerville, V., Shi, R., & Moineau, S. (2024). Fermentation Practices Select for Thermostable Endolysins in Phages. *Molecular Biology and Evolution*, *41*(3), msae055. https://doi.org/10.1093/molbev/msae055

O'Flaherty, S., Coffey, A., Meaney, W., Fitzgerald, G. F., & Ross, R. P. (2005). The recombinant phage lysin LysK has a broad spectrum of lytic activity against clinically relevant staphylococci, including methicillin-resistant Staphylococcus aureus. *Journal of Bacteriology*, *187*(20), 7161–7164. https://doi.org/10.1128/JB.187.20.7161-7164.2005

Ohnuma, T., Onaga, S., Murata, K., Taira, T., & Katoh, E. (2008). LysM domains from Pteris ryukyuensis chitinase-A: A stability study and characterization of the chitin-binding site. *The Journal of Biological Chemistry*, *283*(8), 5178–5187.https://doi.org/10.1074/jbc.M707156200

Oliveira, H., Melo, L. D. R., Santos, S. B., Nóbrega, F. L., Ferreira, E. C., Cerca, N., Azeredo, J., & Kluskens, L. D. (2013). Molecular aspects and comparative genomics of bacteriophage endolysins. *Journal of Virology*, *87*(8), 4558–4570. https://doi.org/10.1128/JVI.03277-12

Oliveira, H., Sampaio, M., Melo, L. D. R., Dias, O., Pope, W. H., Hatfull, G. F., & Azeredo, J. (2019). Staphylococci phages display vast genomic diversity and evolutionary relationships. *BMC Genomics*, *20*, 357. https://doi.org/10.1186/s12864-019-5647-8

Oliveira, H., Thiagarajan, V., Walmagh, M., Sillankorva, S., Lavigne, R., Neves-Petersen, M. T., Kluskens, L. D., & Azeredo, J. (2014). A Thermostable Salmonella Phage Endolysin, Lys68, with Broad Bactericidal Properties against Gram-Negative Pathogens in Presence of Weak Acids. *PLOS ONE*, *9*(10), e108376. https://doi.org/10.1371/journal.pone.0108376

Osváth, S., & Gruebele, M. (2003). Proline Can Have Opposite Effects on Fast and Slow Protein Folding Phases. *Biophysical Journal*, *85*(2), 1215–1222.

Palmer, I., & Wingfield, P. T. (2004). Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia coli. *Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.]*, *CHAPTER*, Unit-6.3. https://doi.org/10.1002/0471140864.ps0603s38

Pandak, N., Al Sidairi, H., Al-Zakwani, I., Al Balushi, Z., Chhetri, S., Ba'Omar, M., Al Lawati, S., Al-Abri, S. S., & Khamis, F. (2023). The Outcome of Antibiotic Overuse before and during the COVID-19 Pandemic in a Tertiary Care Hospital in Oman. *Antibiotics*, *12*(12), Article 12. https://doi.org/10.3390/antibiotics12121665

Pang, Y.-P., Elsbernd, L. R., Block, M. S., & Markovic, S. N. (2018). Peptide-Binding Groove Contraction Linked to the Lack of T Cell Response: Using Complex Structure and Energy To Identify Neoantigens. *ImmunoHorizons*, *2*(7), 216–225.https://doi.org/10.4049/immunohorizons.1800048

Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Computation*,*15*(6),1191–1253.https://doi.org/10.1162/089976603321780272

Paradis-Bleau, C., Cloutier, I., Lemieux, L., Sanschagrin, F., Laroche, J., Auger, M., Garnier, A., & Levesque, R. C. (2007). Peptidoglycan lytic activity of the Pseudomonas aeruginosa phage phiKZ gp144 lytic transglycosylase. *FEMS Microbiology Letters*, *266*(2), 201–209. https://doi.org/10.1111/j.1574-6968.2006.00523.x

Parfentjev, I. A., & Catelli, A. R. (1964). TOLERANCE OF STAPHYLOCOCCUS AUREUS TO SODIUM CHLORIDE. *Journal of Bacteriology*, *88*(1), 1–3.

Park, H., Park, J.-Y., Park, K.-M., & Chang, P.-S. (2021). Effects of freezing rate on structural changes in l-lactate dehydrogenase during the freezing process. *Scientific Reports*, *11*, 13643. https://doi.org/10.1038/s41598-021-93127-6

Park, J., Yun, J., Lim, J.-A., Kang, D.-H., & Ryu, S. (2012). Characterization of an endolysin, LysBPS13, from a Bacillus cereus bacteriophage. *FEMS Microbiology Letters*, *332*(1), 76–83. https://doi.org/10.1111/j.1574-6968.2012.02578.x

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., Quail, M., Smith, F., Walker, D., Libberton, B., Fenton, A., Hall, N., & Brockhurst, M. A. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, *464*(7286), 275–278. https://doi.org/10.1038/nature08798

Peach, M., Marsh, N., Miskiewicz, E. I., & MacPhee, D. J. (2015). Solubilization of proteins: The importance of lysis buffer choice. *Methods in Molecular Biology (Clifton, N.J.)*, *1312*, 49–60. https://doi.org/10.1007/978-1-4939-2694-7_8

Pesarrodona, M., Latorre, I., & Salvatella, X. (2022). Intrinsically Disordered Proteins (IDP): Purification Under Denaturing Conditions. In E. Garcia Fruitós & A. Arís Giralt (Eds.), *Insoluble Proteins: Methods and Protocols* (pp. 359–370). SpringerUS.https://doi.org/10.1007/978-1-0716-1859-2_21

Phothichaisri, W., Chankhamhaengdecha, S., Janvilisri, T., Nuadthaisong, J., Phetruen, T., Fagan, R. P., & Chanarat, S. (n.d.). Potential Role of the Host-Derived Cell-Wall Binding Domain of Endolysin CD16/50L as a Molecular Anchor in Preservation of Uninfected Clostridioides difficile for New Rounds of Phage Infection. *Microbiology Spectrum*, *10*(2), e02361-21. https://doi.org/10.1128/spectrum.02361-21

Phothichaisri, W., Chankhamhaengdecha, S., Janvilisri, T., Nuadthaisong, J., Phetruen, T., Fagan, R. P., & Chanarat, S. (2022). Potential Role of the Host-Derived Cell-Wall Binding Domain of Endolysin CD16/50L as a Molecular Anchor in Preservation of Uninfected Clostridioides difficile for New Rounds of Phage Infection. *Microbiology Spectrum*, *10*(2), e0236121. https://doi.org/10.1128/spectrum.02361-21

Pritchard, D. G., Dong, S., Baker, J. R., & Engler, J. A. (2004). The bifunctional peptidoglycan lysin of Streptococcus agalactiae bacteriophage B30. *Microbiology (Reading, England)*, *150*(Pt 7), 2079–2087. https://doi.org/10.1099/mic.0.27063-0

Pritchard, D. G., Dong, S., Kirk, M. C., Cartee, R. T., & Baker, J. R. (2007). LambdaSa1 and LambdaSa2 prophage lysins of Streptococcus agalactiae. *Applied and Environmental Microbiology*, *73*(22), 7150–7154. https://doi.org/10.1128/AEM.01783-07

Proença, D., Fernandes, S., Leandro, C., Silva, F. A., Santos, S., Lopes, F., Mato, R., Cavaco-Silva, P., Pimentel, M., & São-José, C. (2012). Phage Endolysins with Broad Antimicrobial Activity Against Enterococcus faecalis Clinical Strains. *Microbial Drug Resistance*, *18*(3), 322–332. https://doi.org/10.1089/mdr.2012.0024

Rahman, M. ur, Wang, W., Sun, Q., Shah, J. A., Li, C., Sun, Y., Li, Y., Zhang, B., Chen, W., & Wang, S. (2021). Endolysin, a Promising Solution against Antimicrobial Resistance. *Antibiotics*, *10*(11). https://doi.org/10.3390/antibiotics10111277

Rashel, M., Uchiyama, J., Takemura, I., Hoshiba, H., Ujihara, T., Takatsuji, H., Honke, K., & Matsuzaki, S. (2008). Tail-associated structural protein gp61 of Staphylococcus aureus phage phi MR11 has bifunctional lytic activity. *FEMS Microbiology Letters*, *284*(1), 9–16. https://doi.org/10.1111/j.1574-6968.2008.01152.x

Rashel, M., Uchiyama, J., Ujihara, T., Uehara, Y., Kuramoto, S., Sugihara, S., Yagyu, K.-I., Muraoka, A., Sugai, M., Hiramatsu, K., Honke, K., & Matsuzaki, S. (2007). Efficient elimination of multidrug-resistant Staphylococcus aureus by cloned lysin derived from bacteriophage phi MR11. *The Journal of Infectious Diseases*, *196*(8), 1237–1247. https://doi.org/10.1086/521305

Recsei, P. A., Gruss, A. D., & Novick, R. P. (1987). Cloning, sequence, and expression of the lysostaphin gene from Staphylococcus simulans. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(5), 1127–1131. https://doi.org/10.1073/pnas.84.5.1127

Regulski, K., Courtin, P., Kulakauskas, S., & Chapot-Chartier, M.-P. (2013). A Novel Type of Peptidoglycan-binding Domain Highly Specific for Amidated d-Asp Cross-bridge, Identified in Lactobacillus casei Bacteriophage Endolysins. *The Journal of Biological Chemistry*, *288*(28), 20416–20426. https://doi.org/10.1074/jbc.M112.446344

Reischl, U., Linde, H.-J., Metz, M., Leppmeier, B., & Lehn, N. (2000). Rapid Identification of Methicillin-ResistantStaphylococcus aureus and Simultaneous Species Confirmation Using Real-Time Fluorescence PCR. *Journal of Clinical Microbiology*, *38*(6), 2429–2433. https://doi.org/10.1128/JCM.38.6.2429-2433.2000

Rodríguez-Rubio, L., Gutiérrez, D., Martínez, B., Rodríguez, A., & García, P. (2012). Lytic Activity of LysH5 Endolysin Secreted by Lactococcus lactis Using the Secretion Signal Sequence of Bacteriocin Lcn972. *Applied and Environmental Microbiology*, *78*(9), 3469–3472. https://doi.org/10.1128/AEM.00018-12

Romero, P., Bartual, S. G., Schmelcher, M., Glück, C., Hermoso, J. A., & Loessner, M. J. (2018). Structural insights into the binding and catalytic mechanisms of the Listeria monocytogenes bacteriophage glycosyl hydrolase PlyP40. *Molecular Microbiology*, *108*(2), 128–142. https://doi.org/10.1111/mmi.13922

Ryffel, C., Kayser, F. H., & Berger-Bächi, B. (1992). Correlation between regulation of mecA transcription and expression of methicillin resistance in staphylococci. *Antimicrobial Agents and Chemotherapy*, *36*(1), 25–31. https://doi.org/10.1128/AAC.36.1.25

Salamaga, B., Turner, R. D., Elsarmane, F., Galley, N. F., Kulakauskas, S., & Mesnage, S. (2023). A moonlighting role for LysM peptidoglycan binding domains underpins Enterococcus faecalis daughter cell separation. *Communications Biology*, *6*(1), 1–10. https://doi.org/10.1038/s42003-023-04808-z

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, *234*(3), 779–815. https://doi.org/10.1006/jmbi.1993.1626

Santucci, R., Sinibaldi, F., & Fiorucci, L. (2008). Protein folding, unfolding and misfolding: Role played by intermediate States. *Mini Reviews in Medicinal Chemistry*, *8*(1), 57–62. https://doi.org/10.2174/138955708783331522

Sanz-Gaitero, M., Keary, R., Garcia-Doval, C., Coffey, A., & van Raaij, M. J. (2014). Crystal structure of the lytic CHAPK domain of the endolysin LysK from Staphylococcus aureus bacteriophage K. *Virology Journal*, *11*, 133. https://doi.org/10.1186/1743-422X-11-133

Sanz-Gaitero, M., & van Raaij, M. J. (2021). Crystallographic Structure Determination of Bacteriophage Endolysins. *Current Issues in Molecular Biology*, *40*(1), Article 1. https://doi.org/10.21775/cimb.040.165

Saponaro, A., Maione, V., Bonvin, A. M. J. J., & Cantini, F. (2020). Understanding Docking Complexes of Macromolecules Using HADDOCK: The Synergy between Experimental Data and Computations. *Bio-Protocol*, *10*(20), e3793. https://doi.org/10.21769/BioProtoc.3793

Schauperl, M., & Denny, R. A. (2022). AI-Based Protein Structure Prediction in Drug Discovery: Impacts and Challenges. *Journal of Chemical Information and Modeling*, *62*(13), 3142–3156. https://doi.org/10.1021/acs.jcim.2c00026

Scheffers, D.-J., & Pinho, M. G. (2005). Bacterial Cell Wall Synthesis: New Insights from Localization Studies. *Microbiology and Molecular Biology Reviews*, *69*(4), 585–607. https://doi.org/10.1128/MMBR.69.4.585-607.2005

Schleifer, K. H., & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological Reviews*, *36*(4), 407–477.

Schmelcher, M., Donovan, D. M., & Loessner, M. J. (2012). Bacteriophage endolysins as novel antimicrobials. *Future Microbiology*, *7*(10), 1147–1171. https://doi.org/10.2217/fmb.12.97

Schmelcher, M., Korobova, O., Schischkova, N., Kiseleva, N., Kopylov, P., Pryamchuk, S., Donovan, D. M., & Abaev, I. (2012). Staphylococcus haemolyticus prophage ΦSH2 endolysin relies on cysteine, histidine-dependent amidohydrolases/peptidases activity for lysis 'from without'. *Journal of Biotechnology*, *162*(2–3), 289–298. https://doi.org/10.1016/j.jbiotec.2012.09.010

Schmelcher, M., Shen, Y., Nelson, D. C., Eugster, M. R., Eichenseher, F., Hanke, D. C., Loessner, M. J., Dong, S., Pritchard, D. G., Lee, J. C., Becker, S. C., Foster-Frey, J., & Donovan, D. M. (2015). Evolutionarily distinct bacteriophage endolysins featuring conserved peptidoglycan cleavage sites protect mice from MRSA infection. *Journal of Antimicrobial Chemotherapy*, *70*(5), 1453–1465. https://doi.org/10.1093/jac/dku552

Schmelcher, M., Tchang, V. S., & Loessner, M. J. (2011). Domain shuffling and module engineering of Listeria phage endolysins for enhanced lytic activity and binding affinity. *Microbial Biotechnology*, *4*(5), 651–662. https://doi.org/10.1111/j.1751-7915.2011.00263.x

Schumann, P. (2011). 5—Peptidoglycan Structure. In F. Rainey & A. Oren (Eds.), *Methods in Microbiology* (Vol. 38, pp. 101–129). Academic Press. https://doi.org/10.1016/B978-0-12-387730-7.00005-X

Shah, S., Das, R., Chavan, B., Bajpai, U., Hanif, S., & Ahmed, S. (2023). Beyond antibiotics: Phage-encoded lysins against Gram-negative pathogens. *Frontiers in Microbiology*, *14*, 1170418. https://doi.org/10.3389/fmicb.2023.1170418

Shang, X., & Nelson, D. C. (2019). Contributions of Net Charge on the PlyC Endolysin CHAP Domain. *Antibiotics*, *8*(2), 70. https://doi.org/10.3390/antibiotics8020070

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sharif, S., Singh, M., Kim, S. J., & Schaefer, J. (2009). Staphylococcus aureus Peptidoglycan Tertiary Structure from Carbon-13 Spin Diffusion. *Journal of the American Chemical Society*, *131*(20), 7023–7030. https://doi.org/10.1021/ja808971c

Shen, M., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science : A Publication of the Protein Society*, *15*(11), 2507–2524. https://doi.org/10.1110/ps.062416606

Silaban, S., Gaffar, S., Simorangkir, M., Maksum, I. P., & Subroto, T. (2018). Effect of IPTG Concentration on Recombinant Human Prethrombin-2 Expression in Escherichia coli BL21(DE3) ArcticExpress. *IOP Conference Series: Earth and Environmental Science*, *217*(1), 012039. https://doi.org/10.1088/1755-1315/217/1/012039

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories*, *14*(1), 41. https://doi.org/10.1186/s12934-015-0222-8

Singh, P. K., Donovan, D. M., & Kumar, A. (2014). Intravitreal injection of the chimeric phage endolysin Ply187 protects mice from Staphylococcus aureus endophthalmitis. *Antimicrobial Agents and Chemotherapy*, *58*(8), 4621–4629. https://doi.org/10.1128/AAC.00126-14

Singh, S. M., & Panda, A. K. (2005). Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering*, *99*(4), 303–310. https://doi.org/10.1263/jbb.99.303

Sipeki, S., Koprivanacz, K., Takács, T., Kurilla, A., László, L., Vas, V., & Buday, L. (2021). Novel Roles of SH2 and SH3 Domains in Lipid Binding. *Cells*, *10*(5), 1191.https://doi.org/10.3390/cells10051191

Soleymani, B., & Mostafaie, A. (2019). Analysis of Methods to Improve the Solubility of Recombinant Bovine Sex Determining Region Y Protein. *Reports of Biochemistry & Molecular Biology*, *8*(3), 227–235.

Son, B., Kong, M., Lee, Y., & Ryu, S. (2021). Development of a Novel Chimeric Endolysin, Lys109 With Enhanced Lytic Activity Against Staphylococcus aureus. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.615887

Son, B., Kong, M., & Ryu, S. (2018). The Auxiliary Role of the Amidase Domain in Cell Wall Binding and Exolytic Activity of Staphylococcal Phage Endolysins. *Viruses*,*10*(6),284.https://doi.org/10.3390/v10060284

Son, B., Yun, J., Lim, J.-A., Shin, H., Heu, S., & Ryu, S. (2012). Characterization of LysB4, an endolysin from the Bacillus cereus-infecting bacteriophage B4. *BMCMicrobiology*,*12*(1),33.https://doi.org/10.1186/1471-2180-12-33

Son YJ, Kim CK, Choi BT, Park YC, & Seo JH. (2008). *Effects of beta-mercaptoethanol and hydrogen peroxide on enzymatic conversion of human proinsulin to insulin. - PubMed—NCBI*. *18*(5), 983–989.

Song, J. M., An, Y. J., Kang, M. H., Lee, Y.-H., & Cha, S.-S. (2012). Cultivation at 6–10 °C is an effective strategy to overcome the insolubility of recombinant proteins in *Escherichia coli*. *Protein Expression and Purification*, *82*(2), 297–301. https://doi.org/10.1016/j.pep.2012.01.020

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

Sotomayor-Vivas, C., Hernández-Lemus, E., & Dorantes-Gilardi, R. (2022). Linking protein structural and functional change to mutation using amino acid networks. *PLoS ONE*, *17*(1), e0261829. https://doi.org/10.1371/journal.pone.0261829

Stathopulos, P. B., Scholz, G. A., Hwang, Y.-M., Rumfeldt, J. A. O., Lepock, J. R., & Meiering, E. M. (2004). Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Science*, *13*(11), 3017–3027. https://doi.org/10.1110/ps.04831804

Strobel, H. M., Labador, S. D., Basu, D., Sane, M., Corbett, K. D., & Meyer, J. R. (2024). Viral Receptor-Binding Protein Evolves New Function through Mutations That Cause Trimer Instability and Functional Heterogeneity. *Molecular Biology and Evolution*, *41*(4), msae056. https://doi.org/10.1093/molbev/msae056

Sullivan, L., Paredes, C. J., Papoutsakis, E. T., & Bennett, G. N. (2007). Analysis of the clostridial hydrophobic with a conserved tryptophan family (ChW) of proteins in *Clostridium acetobutylicum* with emphasis on ChW14 and ChW16/17. *Enzyme and Microbial Technology*, *42*(1), 29–43. https://doi.org/10.1016/j.enzmictec.2007.07.022

Suvorov, M., Lee, M., Hesek, D., Boggess, B., & Mobashery, S. (2008). Lytic Transglycosylase MltB of Escherichia coli and Its Role in Recycling of Peptidoglycan Strands of Bacterial Cell Wall. *Journal of the American Chemical Society*, *130*(36), 11878–11879. https://doi.org/10.1021/ja805482b

Tabor, S. (1990). Expression Using the T7 RNA Polymerase/Promoter System. *Current Protocols in Molecular Biology*, *11*(1), 16.2.1-16.2.11. https://doi.org/10.1002/0471142727.mb1602s11

Takeuchi, F., Watanabe, S., Baba, T., Yuzawa, H., Ito, T., Morimoto, Y., Kuroda, M., Cui, L., Takahashi, M., Ankai, A., Baba, S., Fukui, S., Lee, J. C., & Hiramatsu, K. (2005). Whole-genome sequencing of staphylococcus haemolyticus uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species. *Journal of Bacteriology*, *187*(21), 7292–7308. https://doi.org/10.1128/JB.187.21.7292-7308.2005

Tamai, E., Yoshida, H., Sekiya, H., Nariya, H., Miyata, S., Okabe, A., Kuwahara, T., Maki, J., & Kamitori, S. (2014). X-ray structure of a novel endolysin encoded by episomal phage phiSM101 of lostridium perfringens. *Molecular Microbiology*,*92*(2),326–337.https://doi.org/10.1111/mmi.12559

Tang, F., Bossers, A., Harders, F., Lu, C., & Smith, H. (2013). Comparative genomic analysis of twelve Streptococcus suis (pro)phages. *Genomics*, *101*(6), 336–344.https://doi.org/10.1016/j.ygeno.2013.04.005

Tham, H. Y., Song, A. A.-L., Yusoff, K., & Tan, G. H. (2020). Effect of different cloning strategies in pET-28a on solubility and functionality of a staphylococcal phage endolysin. *BioTechniques*, *69*(3), 161–170. https://doi.org/10.2144/btn-2020-0034

Tipper, D. J. (1969). Structures of the cell wall peptidoglycans of Staphylococcus epidermidis Texas 26 and Staphylococcus aureus Copenhagen. II. Structure of neutral and basic peptides from hydrolysis with the Myxobacter al-1 peptidase. *Biochemistry*, *8*(5), 2192–2202. https://doi.org/10.1021/bi00833a061

Tokmakov, A. A., Kurotani, A., & Sato, K.-I. (2021). Protein pI and Intracellular Localization. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/fmolb.2021.775736

Tolba, M., Ahmed, M. U., Tlili, C., Eichenseher, F., Loessner, M. J., & Zourob, M. (2012). A bacteriophage endolysin-based electrochemical impedance biosensor for the rapid detection of Listeria cells. *Analyst*, *137*(24), 5749–5756. https://doi.org/10.1039/C2AN35988J

Tossavainen, H., Raulinaitis, V., Kauppinen, L., Pentikäinen, U., Maaheimo, H., & Permi, P. (2018). Structural and Functional Insights Into Lysostaphin–Substrate Interaction. *Frontiers in Molecular Biosciences*, *5*. https://www.frontiersin.org/articles/10.3389/fmolb.2018.00060

Totté, J. E. E., van Doorn, M. B., & Pasmans, S. G. M. A. (2017). Successful Treatment of Chronic Staphylococcus aureus-Related Dermatoses with the Topical Endolysin Staphefekt SA.100: A Report of 3 Cases. *Case Reports in Dermatology*, *9*(2), 19–25. https://doi.org/10.1159/000473872

Trognon, J., Rima, M., Lajoie, B., Roques, C., & El Garah, F. (2023). NaCl-induced modulation of species distribution in a mixed P. aeruginosa / S. aureus /B.cepacia biofilm. *Biofilm*, *6*, 100153. https://doi.org/10.1016/j.bioflm.2023.100153

Turner, N. A., Sharma-Kuinkel, B. K., Maskarinec, S. A., Eichenberger, E. M., Shah, P. P., Carugati, M., Holland, T. L., & Fowler, V. G. (2019). Methicillin-resistant Staphylococcus aureus: An overview of basic and clinical research. *Nature Reviews Microbiology*, *17*(4), 203–218. https://doi.org/10.1038/s41579-018-0147-4

Turner, R. D., Vollmer, W., & Foster, S. J. (2014). Different walls for rods and balls: The diversity of peptidoglycan. *Molecular Microbiology*, *91*(5), 862–874. https://doi.org/10.1111/mmi.12513

Umbach, A.-K., Failing, K., Goericke-Pesch, S., & Wehrend, A. (2019). Freezing and storage effects on enzyme, electrolyte, and mineral compositions in canine prostatic fluid. *Veterinary Clinical Pathology*, *48*(4), 691–694. https://doi.org/10.1111/vcp.12803

Upadhyay, A. K., Murmu, A., Singh, A., & Panda, A. K. (2012). Kinetics of Inclusion Body Formation and Its Correlation with the Characteristics of Protein Aggregates in Escherichia coli. *PLOS ONE*, *7*(3), e33951. https://doi.org/10.1371/journal.pone.0033951

Vagenende, V., Yap, M. G. S., & Trout, B. L. (2009). Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry*, *48*(46), 11084–11096. https://doi.org/10.1021/bi900649t

Vaish, M., Jereen, A., Ali, A., & Krulwich, T. A. (2019). *The alkaliphilic side of Staphylococcus aureus* (p. 735191). bioRxiv. https://doi.org/10.1101/735191

Vaish, M., Price-Whelan, A., Reyes-Robles, T., Liu, J., Jereen, A., Christie, S., Alonzo, F., Benson, M. A., Torres, V. J., & Krulwich, T. A. (2018). Roles of Staphylococcus aureus Mnh1 and Mnh2 Antiporters in Salt Tolerance, Alkali Tolerance, and Pathogenesis. *Journal of Bacteriology*, *200*(5), 10.1128/jb.00611-17. https://doi.org/10.1128/jb.00611-17

van Dijk, A. D. J., & Bonvin, A. M. J. J. (2006). Solvated docking: Introducing water into the modelling of biomolecular complexes. *Bioinformatics*, *22*(19), 2340–2347. https://doi.org/10.1093/bioinformatics/btl395

Vander Elst, N. (2024). Bacteriophage-derived endolysins as innovative antimicrobials against bovine mastitis-causing streptococci and staphylococci: A state-of-the-art review. *Acta Veterinaria Scandinavica*, *66*(1), 20. https://doi.org/10.1186/s13028-024-00740-2

Vander Elst, N., Linden, S. B., Lavigne, R., Meyer, E., Briers, Y., & Nelson, D. C. (2020). Characterization of the Bacteriophage-Derived Endolysins PlySs2 and PlySs9 with In Vitro Lytic Activity against Bovine Mastitis Streptococcus uberis. *Antibiotics*, *9*(9), 621. https://doi.org/10.3390/antibiotics9090621

Vangone, A., & Bonvin, A. M. (2015). Contacts-based prediction of binding affinity in protein–protein complexes. *eLife*, *4*, e07454. https://doi.org/10.7554/eLife.07454

Vázquez, R., García, E., & García, P. (2021). Sequence-Function Relationships in Phage-Encoded Bacterial Cell Wall Lytic Enzymes and Their Implications for Phage-Derived Product Design. *Journal of Virology*, *95*(14), 10.1128/jvi.00321-21. https://doi.org/10.1128/jvi.00321-21

Vollmer, W. (2008). Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiology Reviews*, *32*(2), 287–306. https://doi.org/10.1111/j.1574-6976.2007.00088.x

Vollmer, W., Blanot, D., & de Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, *32*(2), 149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x

Vollmer, W., & Seligman, S. J. (2010). Architecture of peptidoglycan: More data and more models. *Trends in Microbiology*, *18*(2), 59–66. https://doi.org/10.1016/j.tim.2009.12.004

Vybiral, D., Takác, M., Loessner, M., Witte, A., von Ahsen, U., & Bläsi, U. (2003). Complete nucleotide sequence and molecular characterization of two lytic Staphylococcus aureus phages: 44AHJD and P68. *FEMS Microbiology Letters*, *219*(2), 275–283. https://doi.org/10.1016/S0378-1097(03)00028-4

Wade, R. C., & Goodford, P. J. (1989). The role of hydrogen-bonds in drug binding. *Progress in Clinical and Biological Research*, *289*, 433–444.

Wang, I. N., Smith, D. L., & Young, R. (2000). Holins: The protein clocks of bacteriophage infections. *Annual Review of Microbiology*, *54*, 799–825. https://doi.org/10.1146/annurev.micro.54.1.799

Wang, S., Li, X., Ma, J., Duan, X., Wang, H., Wang, L., Hu, D., Jiang, W., Li, X., & Qian, P. (2025). Structural and functional analysis reveals the catalytic mechanism and substrate binding mode of the broad-spectrum endolysin Ply2741. *Virulence*, *16*(1), 2449025. https://doi.org/10.1080/21505594.2024.2449025

Wang, X., Wilkinson, B. J., & Jayaswal, R. K. (1991). Sequence analysis of a Staphylococcus aureus gene encoding a peptidoglycan hydrolase activity. *Gene*, *102*(1), 105–109. https://doi.org/10.1016/0378-1119(91)90547-o

Wang, Y., Sun, J. H., & Lu, C. P. (2009). Purified recombinant phage lysin LySMP: An extensive spectrum of lytic activity for swine streptococci. *Current Microbiology*, *58*(6), 609–615. https://doi.org/10.1007/s00284-009-9379-x

Wassenberg, M. W. M., Kluytmans, J. a. J. W., Box, A. T. A., Bosboom, R. W., Buiting, A. G. M., van Elzakker, E. P. M., Melchers, W. J. G., van Rijen, M. M. L., Thijsen, S. F. T., Troelstra, A., Vandenbroucke-Grauls, C. M. J. E., Visser, C. E., Voss, A., Wolffs, P. F. G., Wulf, M. W. H., van Zwet, A. A., de Wit, G. A., & Bonten, M. J. M. (2010). Rapid screening of methicillin-resistant Staphylococcus aureus using PCR and chromogenic agar: A prospective study to evaluate costs and effects. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *16*(12), 1754–1761. https://doi.org/10.1111/j.1469-0691.2010.03210.x

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Wedemeyer, W. J., Welker, E., & Scheraga, H. A. (2002). Proline cis-trans isomerization and protein folding. *Biochemistry*, *41*(50), 14637–14644. https://doi.org/10.1021/bi020574b

Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. *The Biochemical Journal*, *297 (Pt 2)*, 249–260. https://doi.org/10.1042/bj2970249

Willing, S., Dyer, E., Schneewind, O., & Missiakas, D. (2020). FmhA and FmhC of Staphylococcus aureus incorporate serine residues into peptidoglycan cross-bridges. *The Journal of Biological Chemistry*, *295*(39), 13664–13676. https://doi.org/10.1074/jbc.RA120.014371

Winogradoff, D., John, S., & Aksimentiev, A. (2020). Protein unfolding by SDS: The microscopic mechanisms and the properties of the SDS-protein assembly. *Nanoscale*, *12*(9), 5422–5434. https://doi.org/10.1039/c9nr09135a

Xia, Y., Zhao, K., Liu, D., Zhou, X., & Zhang, G. (2023). Multi-domain and complex protein structure prediction using inter-domain interactions from deep learning. *Communications Biology*, *6*(1), 1–17. https://doi.org/10.1038/s42003-023-05610-7

Xiong, S., Wang, Y.-F., Ren, X.-R., Li, B., Zhang, M.-Y., Luo, Y., Zhang, L., Xie, Q.-L., & Su, K.-Y. (2005). Solubility of disulfide-bonded proteins in the cytoplasm of Escherichia coli and its "oxidizing" mutant. *World Journal of Gastroenterology : WJG*, *11*(7), 1077–1082. https://doi.org/10.3748/wjg.v11.i7.1077

Xu, Z., Hou, Y., Peters, B. M., Chen, D., Li, B., Li, L., & Shirtliff, M. E. (2016). Chromogenic media for MRSA diagnostics. *Molecular Biology Reports*, *43*(11),1205–1212.https://doi.org/10.1007/s11033-016-4062-3

Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: A web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, *32*(23), 3676–3678. https://doi.org/10.1093/bioinformatics/btw514

Yang, H., Yu, J., & Wei, H. (2014). Engineered bacteriophage lysins as novel anti-infectives. *Frontiers in Microbiology*, *5*, 542. https://doi.org/10.3389/fmicb.2014.00542

Yang, H., Zhang, H., Wang, J., Yu, J., & Wei, H. (2017). A novel chimeric lysin with robust antibacterial activity against planktonic and biofilm methicillin-resistant Staphylococcus aureus. *Scientific Reports*, *7*. https://doi.org/10.1038/srep40182

Yang, S., Sugawara, S., Monodane, T., Nishijima, M., Adachi, Y., Akashi, S., Miyake, K., Hase, S., & Takada, H. (2001). Micrococcus luteus Teichuronic Acids Activate Human and Murine Monocytic Cells in a CD14- and Toll-Like Receptor 4-Dependent Manner. *Infection and Immunity*, *69*(4), 2025–2030. https://doi.org/10.1128/IAI.69.4.2025-2030.2001

Yang, Z., Zhang, L., Zhang, Y., Zhang, T., Feng, Y., Lu, X., Lan, W., Wang, J., Wu, H., Cao, C., & Wang, X. (2011). Highly Efficient Production of Soluble Proteins from Insoluble Inclusion Bodies by a Two-Step-Denaturing and Refolding Method. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0022981

Yokoi, K.-J., Kawahigashi, N., Uchida, M., Sugahara, K., Shinohara, M., Kawasaki, K.-I., Nakamura, S., Taketo, A., & Kodaira, K.-I. (2005). The two-component cell lysis genes holWMY and lysWMY of the Staphylococcus warneri M phage varphiWMY: Cloning, sequencing, expression, and mutational analysis in Escherichia coli. *Gene*, *351*, 97–108. https://doi.org/10.1016/j.gene.2005.03.006

Yong-Jin Mao, Xiang-Rong Sheng, & Xian-Ming Pan. (2007). *The effects of NaCl concentration and pH on the stability of hyperthermophilic protein Ssh10b | BMC Biochemistry | Full Text*. *8*(28). https://bmcbiochem.biomedcentral.com/articles/10.1186/1471-2091-8-28

Yoong, P., Schuch, R., Nelson, D., & Fischetti, V. A. (2004). Identification of a Broadly Active Phage Lytic Enzyme with Lethal Activity against Antibiotic-Resistant Enterococcus faecalis and Enterococcus faecium. *Journal of Bacteriology*, *186*(14), 4808–4812. https://doi.org/10.1128/JB.186.14.4808-4812.2004

Young, R. (1992). Bacteriophage lysis: Mechanism and regulation. *Microbiological Reviews*, *56*(3), 430–481.

Yu, J., Zhang, Y., Zhang, Y., Li, H., Yang, H., & Wei, H. (2016). Sensitive and rapid detection of staphylococcus aureus in milk via cell binding domain of lysin. *Biosensors & Bioelectronics*, *77*, 366–371. https://doi.org/10.1016/j.bios.2015.09.058

Zarghampoor, F., Behzad-Behbahani, A., Azarpira, N., Khatami, S. R., Fanian, M., Hossein Aghdaie, M., & Rafiei Dehbidi, G. (2020). A Single Tube Overlap Extension PCR Method for Splicing of Multiple DNA Fragments. *Avicenna Journal of Medical Biotechnology*, *12*(1), 37–43.

Zhang, H., Buttaro, B. A., Fouts, D. E., Sanjari, S., Evans, B. S., & Stevens, R. H. (2019). Bacteriophage φEf11 ORF28 Endolysin, a Multifunctional Lytic Enzyme with Properties Distinct from All Other Identified Enterococcus faecalis Phage Endolysins. *Applied and Environmental Microbiology*, *85*(13). https://doi.org/10.1128/AEM.00555-19

Zhou, X., Hu, J., Zhang, C., Zhang, G., & Zhang, Y. (2019). Assembling multidomain protein structures through analogous global structural alignments. *Proceedings of the National Academy of Sciences*, *116*(32), 15930–15938. https://doi.org/10.1073/pnas.1905068116

Zosel, F., Mercadante, D., Nettels, D., & Schuler, B. (2018). A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nature Communications*, *9*(1), 3332. https://doi.org/10.1038/s41467-018-05725-0

APPENDICES

APPENDIX A: Sequences of the wild-type and mutants

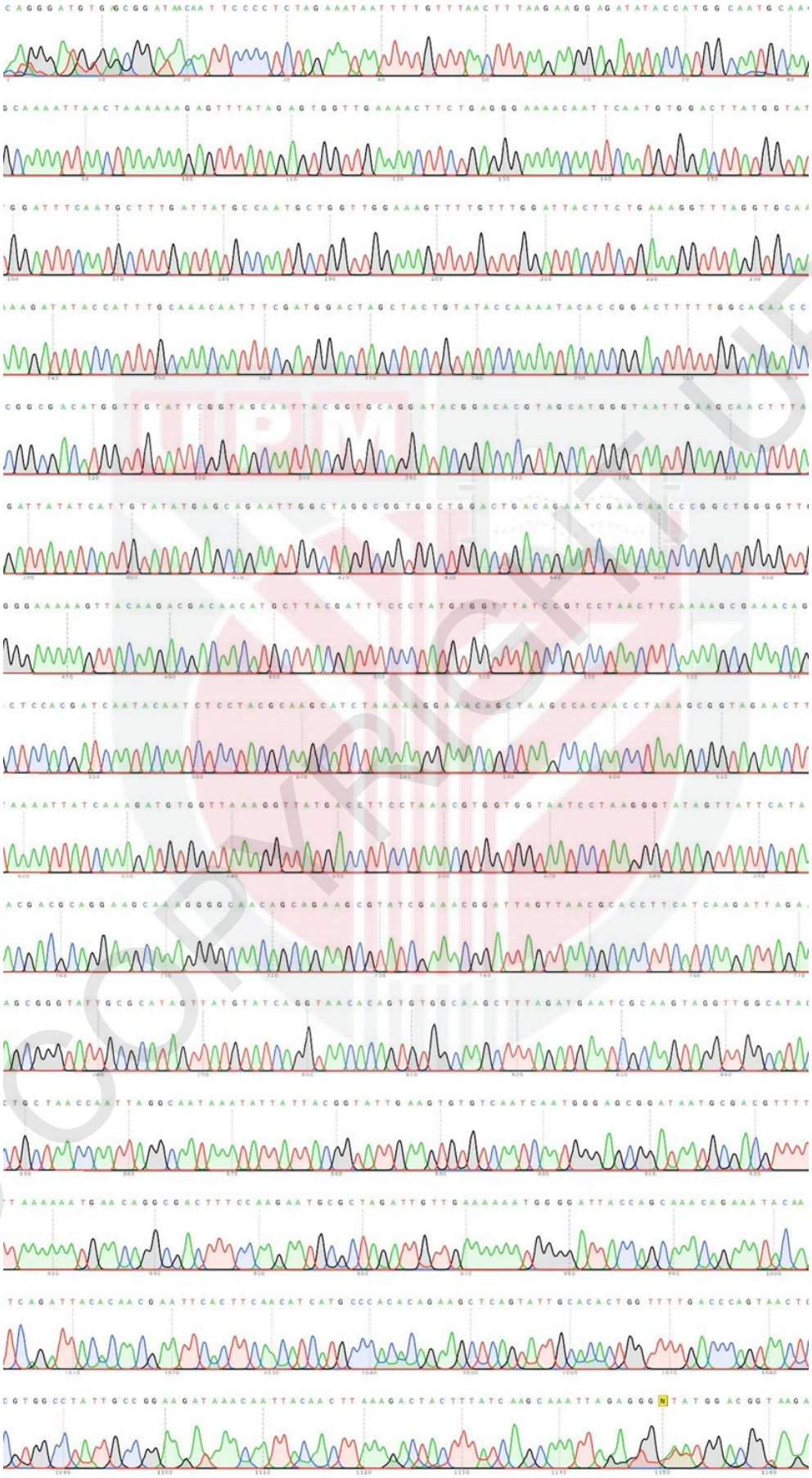
Expected Endo88 sequence (1466 bp)

5'-
CATGCCATGGCAATGCAAGCAAAATTAACTAAAAAAGAGTTTATAGAGTGGTTGAAAAC
TTCTGAGGGAAAACAATTCAATGTGGACTTATGGTATGGATTTCAATGCTTTGATTATGC
CAATGCTGGTTGGAAAGTTTTGTTTGGATTACTTCTGAAAGGTTTAGGTGCAAAAGATAT
ACCATTTGCAAACAATTTCGATGGACTAGCTACTGTATACCAAAATACACCGGACTTTTT
GGCACAACCCGGCGACATGGTTGTATTCGGTAGCAATTACGGTGCAGGATACGGACACG
TAGCATGGGTAATTGAAGCAACTTTAGATTATATCATTGTATATGAGCAGAATTGGCTAG
GCGGTGGCTGGACTGACAGAATCGAACAACCCGGCTGGGGTTGGGAAAAAGTTACAAGA
CGACAACATGCTTACGATTTCCCTATGTGGTTTATCCGTCCTAACTTCAAAAGCGAAACA
GCTCCACGATCAATACAATCTCCTACGCAAGCATCTAAAAAGGAAACAGCTAAGCCACA
ACCTAAAGCGGTAGAACTTAAAATTATCAAAGATGTGGTTAAAGGTTATGACCTTCCTAA
ACGTGGTGGTAATCCTAAGGGTATAGTTATTCATAACGACGCAGGAAGCAAAGGGGCAA
CAGCAGAAGCGTATCGAAACGGATTAGTTAACGCACCTTCATCAAGATTAGAAGCGGGT
ATTGCGCATAGTTATGTATCAGGTAACACAGTGTGGCAAGCTTTAGATGAATCGCAAGTA
GGTTGGCATACTGCTAACCAATTAGGCAATAAATATTATTACGGTATTGAAGTGTGTCAA
TCAATGGGAGCGGATAATGCGACGTTTTTAAAAAATGAACAGGCGACTTTCCAAGAATG
CGCTAGATTGTTGAAAAAATGGGGATTACCAGCAAACAGAAATACAATCAGATTACACA
ACGAATTCACTTCAACATCATGCCCACACAGAAGCTCAGTATTGCACACTGGTTTTGACC
CAGTAACTCGTGGCCTATTGCCGGAAGATAAACAATTACAACTTAAAGACTACTTTATCA
AGCAAATTAGAGTGTATATGGACGGTAAGATACCAGTTGCCACTGTCTCTAATGAGTCAA
GCGCTTCAAGTAATACAGTTAAACCAGTTGCGAGTGCATGGAAACGTAATAAATATGGT
ACTTACTACATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAG
AAAAATAGGACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTG
TGATTATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAGGATATACATGGGAGG
GGCAACGTTATTACTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATAT
TAGGTGACTTATGGGGAGAAATCAGTCCGCTCGAGCC -3'

Sequencing result using pETF2_Fwd

Expected Endo88 sequence (Top) vs sequencing result (Below)

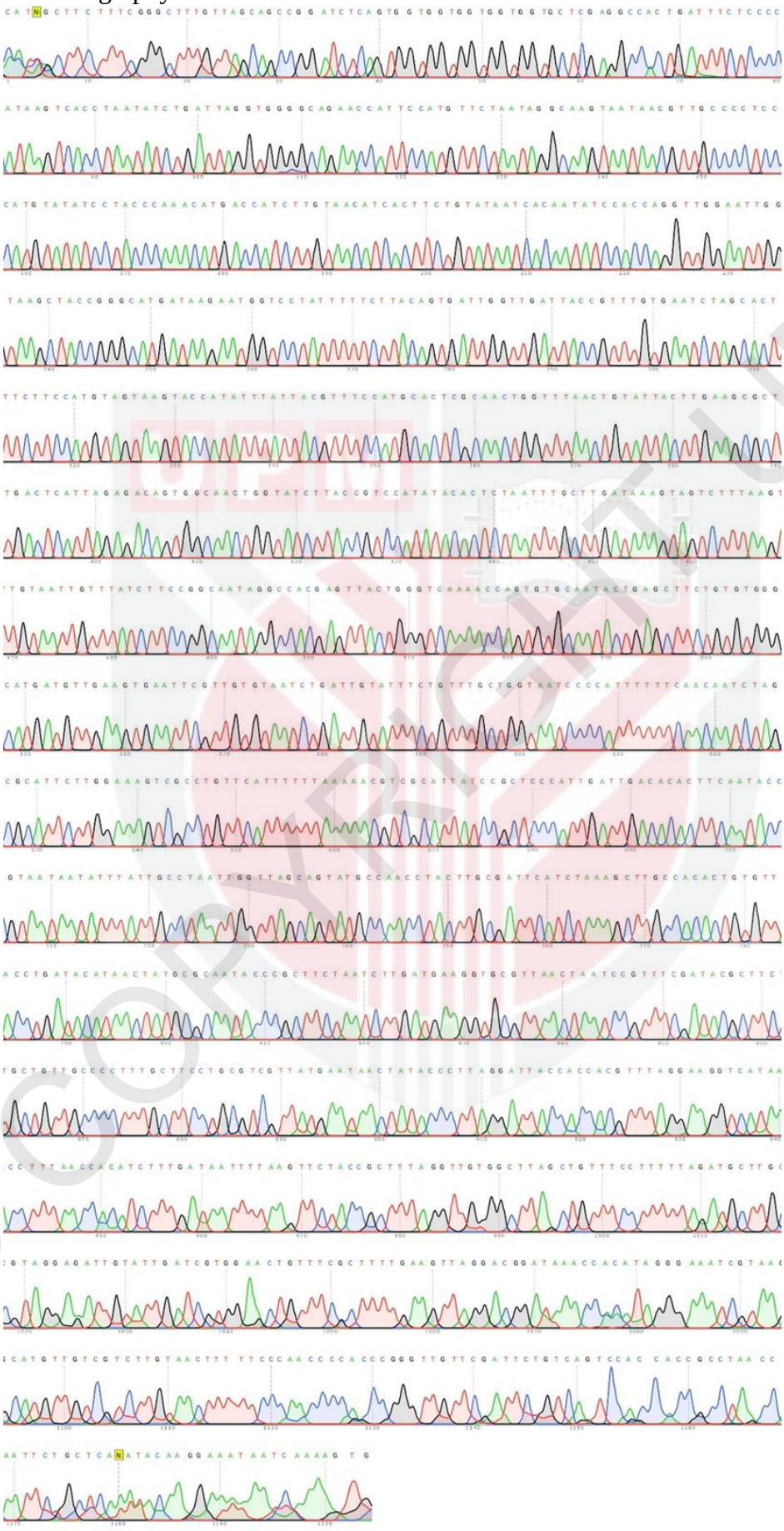
DNA chromatography:



Sequencing result using pETR_Rev

Expected Endo88 sequence (Top) vs sequencing result (Below)

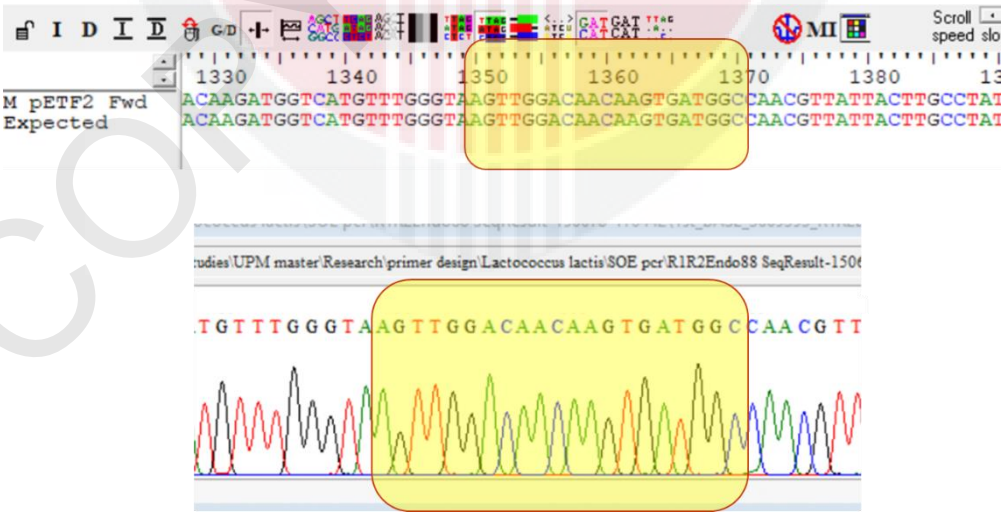
DNA chromatography:



Expected sequence of M (1469 bp), Red: mutation region

5'-
CATGCCATGGCAATGCAAGCAAAATTAACTAAAAAAGAGTTTATAGAGTGGTTGAAAAC
TTCTGAGGGAAAACAATTCAATGTGGACTTATGGTATGGATTTCAATGCTTTGATTATGC
CAATGCTGGTTGGAAAGTTTTGTTTGGATTACTTCTGAAAGGTTTAGGTGCAAAAGATAT
ACCATTTGCAAACAATTTCGATGGACTAGCTACTGTATACCAAAATACACCGGACTTTTT
GGCACAACCCGGCGACATGGTTGTATTCGGTAGCAATTACGGTGCAGGATACGGACACG
TAGCATGGGTAATTGAAGCAACTTTAGATTATATCATTGTATATGAGCAGAATTGGCTAG
GCGGTGGCTGGACTGACAGAATCGAACAACCCGGCTGGGGTTGGGAAAAAGTTACAAGA
CGACAACATGCTTACGATTTCCCTATGTGGTTTATCCGTCCTAACTTCAAAAGCGAAACA
GCTCCACGATCAATACAATCTCCTACGCAAGCATCTAAAAAGGAAACAGCTAAGCCACA
ACCTAAAGCGGTAGAACTTAAAATTATCAAAGATGTGGTTAAAGGTTATGACCTTCCTAA
ACGTGGTGGTAATCCTAAGGGTATAGTTATTCATAACGACGCAGGAAGCAAAGGGGCAA
CAGCAGAAGCGTATCGAAACGGATTAGTTAACGCACCTTCATCAAGATTAGAAGCGGGT
ATTGCGCATAGTTATGTATCAGGTAACACAGTGTGGCAAGCTTTAGATGAATCGCAAGTA
GGTTGGCATACTGCTAACCAATTAGGCAATAAATATTATTACGGTATTGAAGTGTGTCAA
TCAATGGGAGCGGATAATGCGACGTTTTTAAAAAATGAACAGGCGACTTTCCAAGAATG
CGCTAGATTGTTGAAAAAATGGGGATTACCAGCAAACAGAAATACAATCAGATTACACA
ACGAATTCACTTCAACATCATGCCCACACAGAAGCTCAGTATTGCACACTGGTTTTGACC
CAGTAACTCGTGGCCTATTGCCGGAAGATAAACAATTACAACTTAAAGACTACTTTATCA
AGCAAATTAGAGTGTATATGGACGGTAAGATACCAGTTGCCACTGTCTCTAATGAGTCAA
GCGCTTCAAGTAATACAGTTAAACCAGTTGCGAGTGCATGGAAACGTAATAAATATGGT
ACTTACTACATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAG
AAAAATAGGACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTG
TGATTATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAAGTTGGACAACAAGTGA
TGGCCAACGTTATTACTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGAT
ATTAGGTGACTTATGGGGAGAAATCAGTCCGCTCGAGCC -3'

Mutation part (yellow highlight) and DNA chromatography



Expected sequence of A1 (1463 bp), Red: mutation region

5'-
CATGCCATGGCAATGCAAGCAAAATTAACTAAAAAAGAGTTTATAGAGTGGTTGAAAAC
TTCTGAGGGAAAACAATTCAATGTGGACTTATGGTATGGATTTCAATGCTTTGATTATGC
CAATGCTGGTTGGAAAGTTTTGTTTGGATTACTTCTGAAAGGTTTAGGTGCAAAAGATAT
ACCATTTGCAAACAATTTCGATGGACTAGCTACTGTATACCAAAATACACCGGACTTTTT
GGCACAACCCGGCGACATGGTTGTATTCGGTAGCAATTACGGTGCAGGATACGGACACG
TAGCATGGGTAATTGAAGCAACTTTAGATTATATCATTGTATATGAGCAGAATTGGCTAG
GCGGTGGCTGGACTGACAGAATCGAACAACCCGGCTGGGGTTGGGAAAAAGTTACAAGA
CGACAACATGCTTACGATTTCCCTATGTGGTTTATCCGTCCTAACTTCAAAAGCGAAACA
GCTCCACGATCAATACAATCTCCTACGCAAGCATCTAAAAAGGAAACAGCTAAGCCACA
ACCTAAAGCGGTAGAACTTAAAATTATCAAAGATGTGGTTAAAGGTTATGACCTTCCTAA
ACGTGGTGGTAATCCTAAGGGTATAGTTATTCATAACGACGCAGGAAGCAAAGGGGCAA
CAGCAGAAGCGTATCGAAACGGATTAGTTAACGCACCTTCATCAAGATTAGAAGCGGGT
ATTGCGCATAGTTATGTATCAGGTAACACAGTGTGGCAAGCTTTAGATGAATCGCAAGTA
GGTTGGCATACTGCTAACCAATTAGGCAATAAATATTATTACGGTATTGAAGTGTGTCAA
TCAATGGGAGCGGATAATGCGACGTTTTTAAAAAATGAACAGGCGACTTTCCAAGAATG
CGCTAGATTGTTGAAAAAATGGGGATTACCAGCAAACAGAAATACAATCAGATTACACA
ACGAATTCACTTCAACATCATGCCCACACAGAAGCTCAGTATTGCACACTGGTTTTGACC
CAGTAACTCGTGGCCTATTGCCGGAAGATAAACAATTACAACTTAAAGACTACTTTATCA
AGCAAATTAGAGTGTATATGGACGGTAAGATACCAGTTGCCACTGTCTCTAATGAGTCAA
GCGCTTCAAGTAATACAGTTAAACCAGTTGCGAGTGCATGGAAACGTAATAAATATGGT
ACTTACTACATGGAAGAAAGTGCTAGATTTACCCCGAATACCGATATTATTACCCGCACC
ACCGGCCCGTTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTGTGATT
ATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAGGATATACATGGGAGGGGCAA
CGTTATTACTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATATTAGGT
GACTTATGGGGAGAAATCAGTGGCCTCGAGCC -3'

Mutation part (yellow highlight) and DNA chromatography (Rev-complement)

Expected sequence of A2 (1466 bp), Red: mutation region

5’-
CATGCCATGGCAATGCAAGCAAAATTAACTAAAAAAGAGTTTATAGAGTGGTTGAAAAC
TTCTGAGGGAAAACAATTCAATGTGGACTTATGGTATGGATTTCAATGCTTTGATTATGC
CAATGCTGGTTGGAAAGTTTTGTTTGGATTACTTCTGAAAGGTTTAGGTGCAAAAGATAT
ACCATTTGCAAACAATTTCGATGGACTAGCTACTGTATACCAAAATACACCGGACTTTTT
GGCACAACCCGGCGACATGGTTGTATTCGGTAGCAATTACGGTGCAGGATACGGACACG
TAGCATGGGTAATTGAAGCAACTTTAGATTATATCATTGTATATGAGCAGAATTGGCTAG
GCGGTGGCTGGACTGACAGAATCGAACAACCCGGCTGGGGTTGGGAAAAAGTTACAAGA
CGACAACATGCTTACGATTTCCCTATGTGGTTTATCCGTCCTAACTTCAAAAGCGAAACA
GCTCCACGATCAATACAATCTCCTACGCAAGCATCTAAAAAGGAAACAGCTAAGCCACA
ACCTAAAGCGGTAGAACTTAAAATTATCAAAGATGTGGTTAAAGGTTATGACCTTCCTAA
ACGTGGTGGTAATCCTAAGGGTATAGTTATTCATAACGACGCAGGAAGCAAAGGGGCAA
CAGCAGAAGCGTATCGAAACGGATTAGTTAACGCACCTTCATCAAGATTAGAAGCGGGT
ATTGCGCATAGTTATGTATCAGGTAACACAGTGTGGCAAGCTTTAGATGAATCGCAAGTA
GGTTGGCATACTGCTAACCAATTAGGCAATAAATATTATTACGGTATTGAAGTGTGTCAA
TCAATGGGAGCGGATAATGCGACGTTTTTAAAAAATGAACAGGCGACTTTCCAAGAATG
CGCTAGATTGTTGAAAAAATGGGGATTACCAGCAAACAGAAATACAATCAGATTACACA
ACGAATTCACTTCAACATCATGCCCACACAGAAGCTCAGTATTGCACACTGGTTTTGACC
CAGTAACTCGTGGCCTATTGCCGGAAGATAAACAATTACAACTTAAAGACTACTTTATCA
AGCAAATTAGAGTGTATATGGACGGTAAGATACCAGTTGCCACTGTCTCTAATGAGTCAA
GCGCTTCAAGTAATACAGTTAAACCAGTTGCGAGTGCATGGAAACGTAATAAATATGGT
ACTTACTACATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAG
AAAAATAGGACCATTCTTATCATGCCCGGTAAGCGGCGTGCTGAAAGCGGGCCAGACCA
TTCATTATGATGAAGTGATGAAACAAGATGGTCATGTTTGGGTAGGATATACATGGGAG
GGGCAACGTTATTACTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATA
TTAGGTGACTTATGGGGAGAAATCAGTGGCCTCGAGCC -3’

Mutation part (yellow highlight) and DNA chromatography (Rev-complement)

Expected sequence of D1 (1466 bp), Red: mutation region

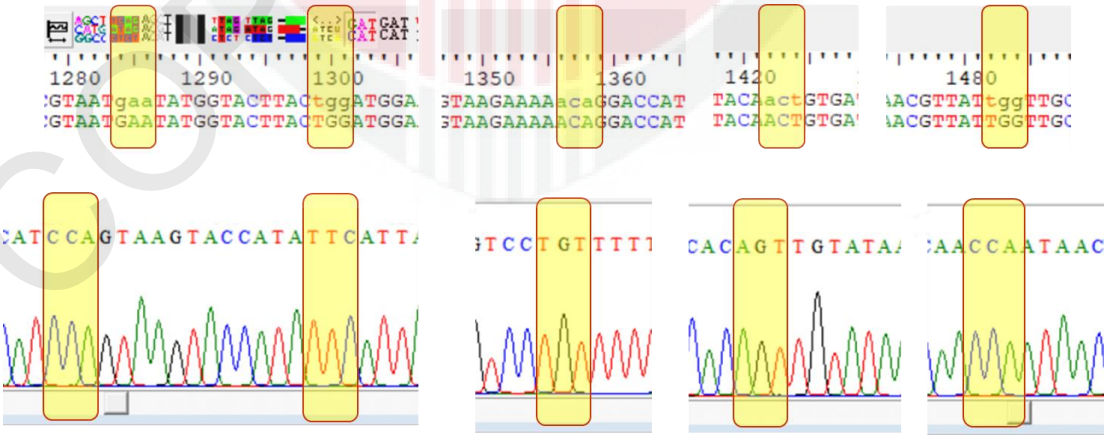
5’-
CATGCCATGGCAATGCAAGCAAAATTAACTAAAAAAGAGTTTATAGAGTGGTTGAAAAC
TTCTGAGGGAAAACAATTCAATGTGGACTTATGGTATGGATTTCAATGCTTTGATTATGC
CAATGCTGGTTGGAAAGTTTTGTTTGGATTACTTCTGAAAGGTTTAGGTGCAAAAGATAT
ACCATTTGCAAACAATTTCGATGGACTAGCTACTGTATACCAAAATACACCGGACTTTTT
GGCACAACCCGGCGACATGGTTGTATTCGGTAGCAATTACGGTGCAGGATACGGACACG
TAGCATGGGTAATTGAAGCAACTTTAGATTATATCATTGTATATGAGCAGAATTGGCTAG
GCGGTGGCTGGACTGACAGAATCGAACAACCCGGCTGGGGTTGGGAAAAAGTTACAAGA
CGACAACATGCTTACGATTTCCCTATGTGGTTTATCCGTCCTAACTTCAAAAGCGAAACA
GCTCCACGATCAATACAATCTCCTACGCAAGCATCTAAAAAGGAAACAGCTAAGCCACA
ACCTAAAGCGGTAGAACTTAAAATTATCAAAGATGTGGTTAAAGGTTATGACCTTCCTAA
ACGTGGTGGTAATCCTAAGGGTATAGTTATTCATAACGACGCAGGAAGCAAAGGGGCAA
CAGCAGAAGCGTATCGAAACGGATTAGTTAACGCACCTTCATCAAGATTAGAAGCGGGT
ATTGCGCATAGTTATGTATCAGGTAACACAGTGTGGCAAGCTTTAGATGAATCGCAAGTA
GGTTGGCATACTGCTAACCAATTAGGCAATAAATATTATTACGGTATTGAAGTGTGTCAA
TCAATGGGAGCGGATAATGCGACGTTTTTAAAAAATGAACAGGCGACTTTCCAAGAATG
CGCTAGATTGTTGAAAAAATGGGGATTACCAGCAAACAGAAATACAATCAGATTACACA
ACGAATTCACTTCAACATCATGCCCACACAGAAGCTCAGTATTGCACACTGGTTTTGACC
CAGTAACTCGTGGCCTATTGCCGGAAGATAAACAATTACAACTTAAAGACTACTTTATCA
AGCAAATTAGAGTGTATATGGACGGTAAGATACCAGTTGCCACTGTCTCTAATGAGTCAA
GCGCTTCAAGTAATACAGTTAAACCAGTTGCGAGTGCATGGAAACGTAATCAGTTCGGTA
CTTACTACATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAGA
AAAGGTGGACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTGT
GATTATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAGGATATACATGGGAGGG
GCAACGTTATTACTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATATT
AGGTGACTTATGGGGAGAAATCAGTGGCCTCGAGGG -3’

Mutation part (yellow highlight) and DNA chromatography (Rev-complement)

Expected sequence of D2 (1466 bp), Red: mutation region

5’-
CATGCCATGGCAATGCAAGCAAAATTAACTAAAAAAGAGTTTATAGAGTGGTTGAAAAC
TTCTGAGGGAAAACAATTCAATGTGGACTTATGGTATGGATTTCAATGCTTTGATTATGC
CAATGCTGGTTGGAAAGTTTTGTTTGGATTACTTCTGAAAGGTTTAGGTGCAAAAGATAT
ACCATTTGCAAACAATTTCGATGGACTAGCTACTGTATACCAAAATACACCGGACTTTTT
GGCACAACCCGGCGACATGGTTGTATTCGGTAGCAATTACGGTGCAGGATACGGACACG
TAGCATGGGTAATTGAAGCAACTTTAGATTATATCATTGTATATGAGCAGAATTGGCTAG
GCGGTGGCTGGACTGACAGAATCGAACAACCCGGCTGGGGTTGGGAAAAAGTTACAAGA
CGACAACATGCTTACGATTTCCCTATGTGGTTTATCCGTCCTAACTTCAAAAGCGAAACA
GCTCCACGATCAATACAATCTCCTACGCAAGCATCTAAAAAGGAAACAGCTAAGCCACA
ACCTAAAGCGGTAGAACTTAAAATTATCAAAGATGTGGTTAAAGGTTATGACCTTCCTAA
ACGTGGTGGTAATCCTAAGGGTATAGTTATTCATAACGACGCAGGAAGCAAAGGGGCAA
CAGCAGAAGCGTATCGAAACGGATTAGTTAACGCACCTTCATCAAGATTAGAAGCGGGT
ATTGCGCATAGTTATGTATCAGGTAACACAGTGTGGCAAGCTTTAGATGAATCGCAAGTA
GGTTGGCATACTGCTAACCAATTAGGCAATAAATATTATTACGGTATTGAAGTGTGTCAA
TCAATGGGAGCGGATAATGCGACGTTTTTAAAAAATGAACAGGCGACTTTCCAAGAATG
CGCTAGATTGTTGAAAAAATGGGGATTACCAGCAAACAGAAATACAATCAGATTACACA
ACGAATTCACTTCAACATCATGCCCACACAGAAGCTCAGTATTGCACACTGGTTTTGACC
CAGTAACTCGTGGCCTATTGCCGGAAGATAAACAATTACAACTTAAAGACTACTTTATCA
AGCAAATTAGAGTGTATATGGACGGTAAGATACCAGTTGCCACTGTCTCTAATGAGTCAA
GCGCTTCAAGTAATACAGTTAAACCAGTTGCGAGTGCATGGAAACGTAATGAATATGGT
ACTTACTGGATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAG
AAAAACAGGACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTG
TGATTATACAACTGTGATGTTACAAGATGGTCATGTTTGGGTAGGATATACATGGGAGGG
GCAACGTTATTGGTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATATT
AGGTGACTTATGGGGAGAAATCAGTGGCCTCGAGGG -3’

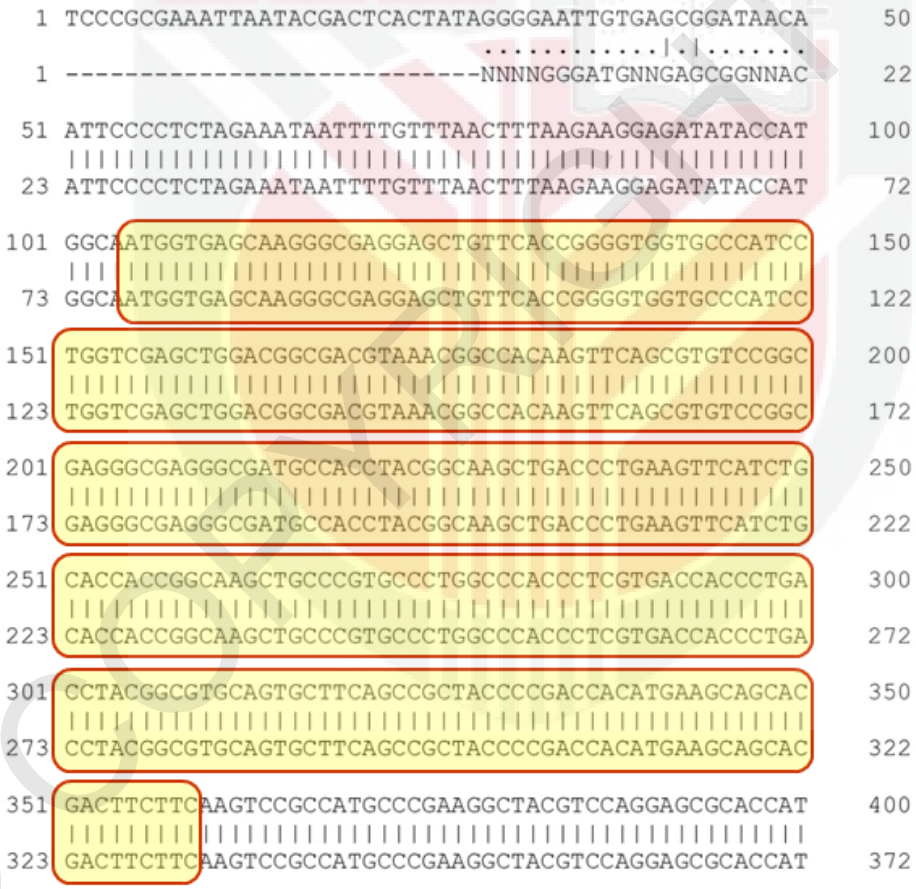
Mutation part (yellow highlight) and DNA chromatography (Rev-complement)



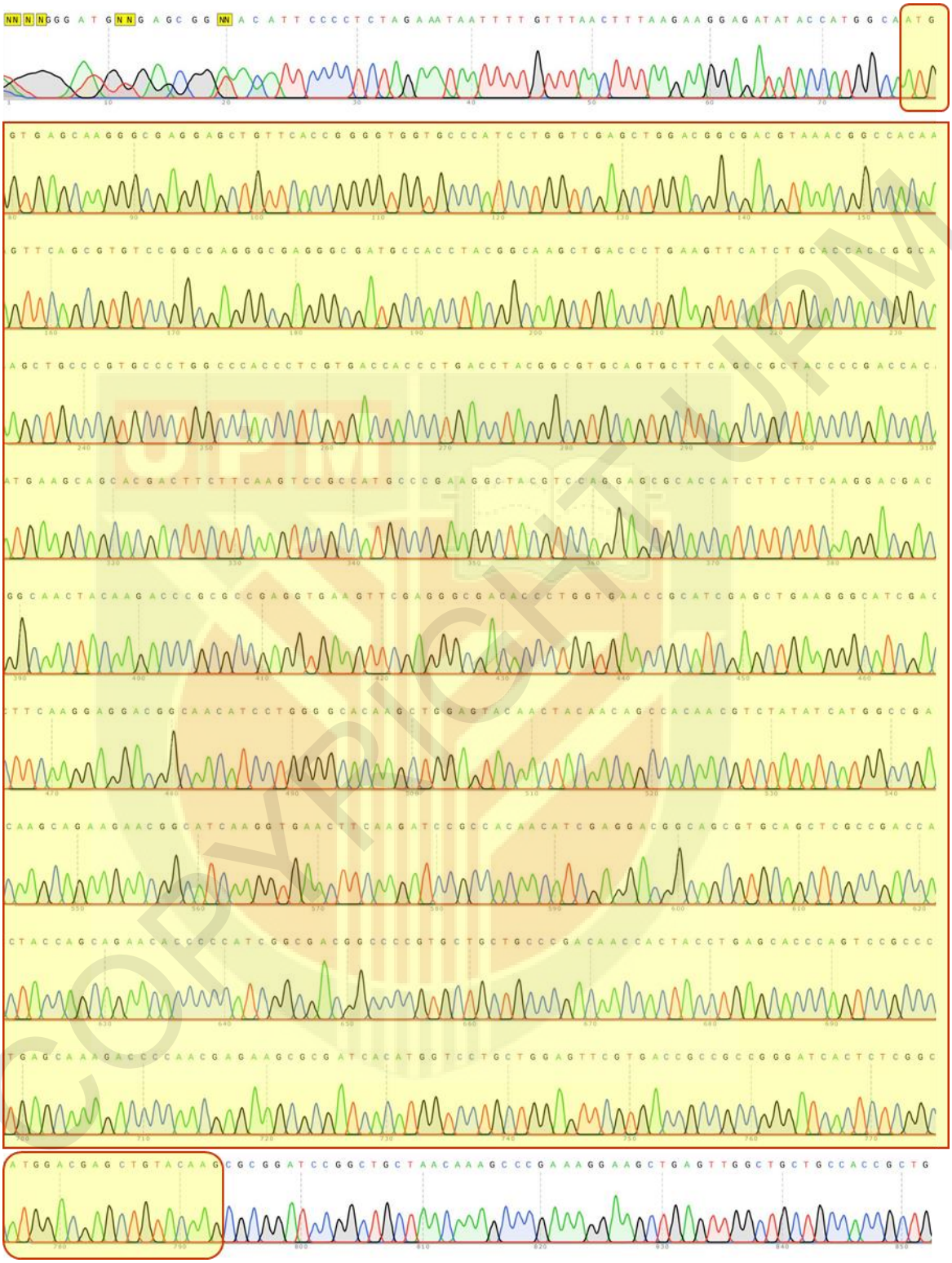
Expected sequence of GFP protein (720 bp)

5’-
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA
AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT
TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAG
CTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC
AACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCA
CATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
CAAG -3’

Expected GFP sequence (Top) vs sequencing result (Below)



DNA chromatography:



Expected sequence of Endo88_SH3 (294 bp)

5’-
GCGAGTGCATGGAAACGTAATAAATATGGTACTTACTACATGGAAGAAAGTGCTAGATT
CACAAACGGTAATCAACCAATCACTGTAAGAAAAATAGGACCATTCTTATCATGCCCGGT
AGCTTACCAATTCCAACCTGGTGGATATTGTGATTATACAGAAGTGATGTTACAAGATGG
TCATGTTTGGGTAGGATATACATGGGAGGGGCAACGTTATTACTTGCCTATTAGAACATG
GAATGGTTCTGCCCCACCTAATCAGATATTAGGTGACTTATGGGGAGAAATCAGT -3’

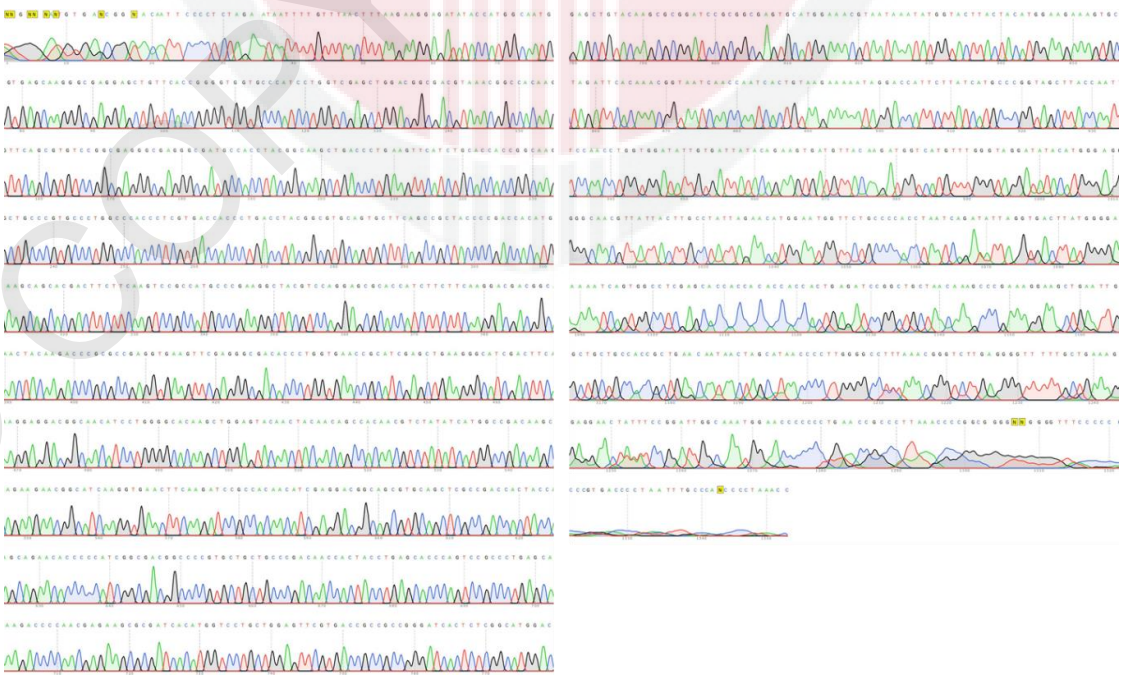
Expected sequence of Endo88_GFP_SH3 (1040 bp)

5’-
CCATGGCAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCC
GACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
CTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT
GGACGAGCTGTACAAGCGCGGATCCGCGGCGAGTGCATGGAAACGTAATAAATATGGTA
CTTACTACATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAGA
AAAATAGGACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTGT
GATTATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAGGATATACATGGGAGGG
GCAACGTTATTACTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATATT
AGGTGACTTATGGGGAGAAATCAGTGGCCTCGAG -3’

Expected GFPSH3 sequence (Top) vs sequencing result (Below)

```
  1 ----------------------------------------------------  0  635 CAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA  684
                                                               |||||||||||||||||||||||||||||||||||||||||||||||||||||
  1 NNGNNNANTGTGANCGGNACAATTCCCCTCTAGAATAATTTTGTTTAACT  50  701 CAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA  750

  1 ----------------CCATGGCAATGGTGAGCAAGGGCGAGGAGCTGTT  34  685 CCGCCGCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGCGCGGATCC  734
                    ||||||||||||||||||||||||||||||||||                ||||||||||||||||||||||||||||||||||||||||||||||||||||
 51 TTAAGAAGGAGATATACCATGGCAATGGTGAGCAAGGGCGAGGAGCTGTT  100  751 CCGCCGCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGCGCGGATCC  800

 35 CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCC  84  735 GCGGCGAGTGCATGGAAACGTAATAAATATGGTACTTACTACATGGAAGA  784
    ||||||||||||||||||||||||||||||||||||||||||||||||||||                ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCC  150  801 GCGGCGAGTGCATGGAAACGTAATAAATATGGTACTTACTACATGGAAGA  850

 85 ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG  134  785 AAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAGAAAAATAG  834
    ||||||||||||||||||||||||||||||||||||||||||||||||||||                ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG  200  851 AAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAGAAAAATAG  900

135 CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCC  184  835 GACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATAT  884
    ||||||||||||||||||||||||||||||||||||||||||||||||||||                ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCC  250  901 GACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATAT  950

185 CACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACC  234  885 TGTGATTATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAGGATA  934
    ||||||||||||||||||||||||||||||||||||||||||||||||||||                ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 CACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACC  300  951 TGTGATTATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAGGATA  1000

235 CCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC  284  935 TACATGGGAGGGGCAACGTTATTACTTGCCTATTAGAACATGGAATGGTT  984
    ||||||||||||||||||||||||||||||||||||||||||||||||||||                ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 CCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGC  350 1001 TACATGGGAGGGGCAACGTTATTACTTGCCTATTAGAACATGGAATGGTT  1050

285 TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC  334  985 CTGCCCCACCTAATCAGATATTAGGTGACTTATGGGGAGAAATCAGTGGC  1034
    ||||||||||||||||||||||||||||||||||||||||||||||||||||                ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC  400 1051 CTGCCCCACCTAATCAGATATTAGGTGACTTATGGGGAAAAATCAGTGGC  1100

335 CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC  384 1035 CTCGAG----------------------------------------------  1040
    ||||||||||||||||||||||||||||||||||||||||||||||||||||                ||||||
401 CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC  450 1101 CTCGAGCACCACCACACCACCACTGAGATCCGGCTGCTAACAAAGCCCGA  1150

385 TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG  434
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG  500

435 GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAA  484
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAA  550

485 GAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCA  534
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
551 GAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCA  600

535 GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC  584
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
601 GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC  650

585 CCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAG  634
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
651 CCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAG  700
```

DNA chromatography :

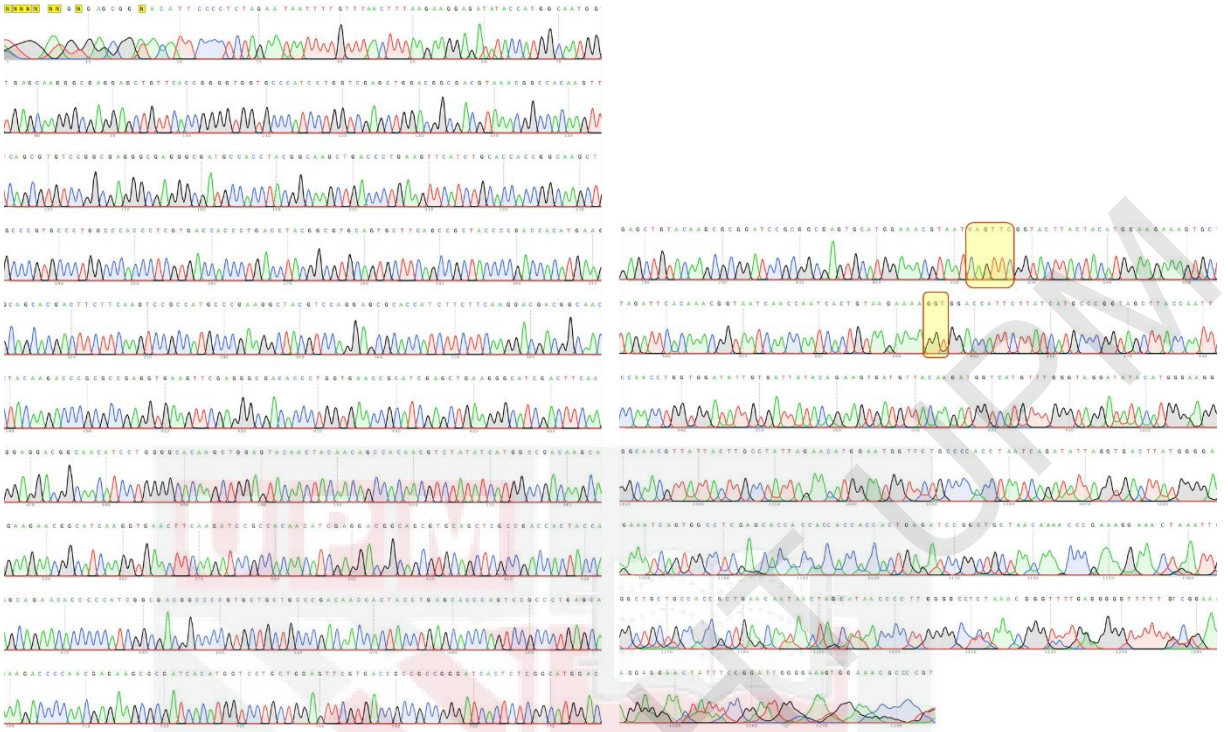


Expected sequence of D1_GFP_SH3 (1042 bp), Red: mutation region

5'-
CCATGGCAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCC
GACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
CTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT
GGACGAGCTGTACAAGCGCGGATCCGCGGCGAGTGCATGGAAACGTAATCAGTTCGGTA
CTTACTACATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAGA
AAAGGTGGACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTGT
GATTATACAGAAGTGATGTTACAAGATGGTCATGTTTGGGTAGGATATACATGGGAGGG
GCAACGTTATTACTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATATT
AGGTGACTTATGGGGAGAAATCAGTGGCCTCGAGGG -3'

Expected D1_GFPSH3 sequence (Top) vs sequencing result (Below)

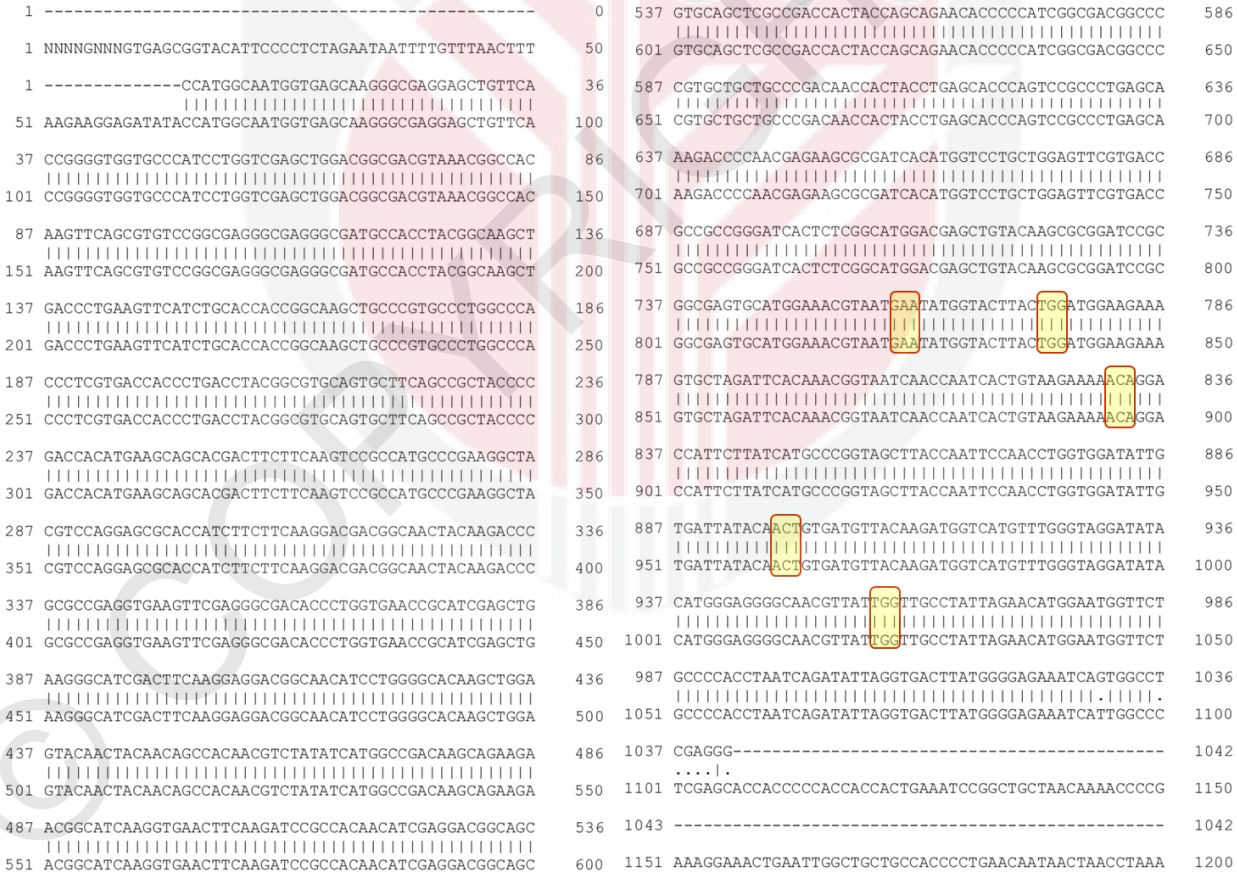
DNA chromatography:



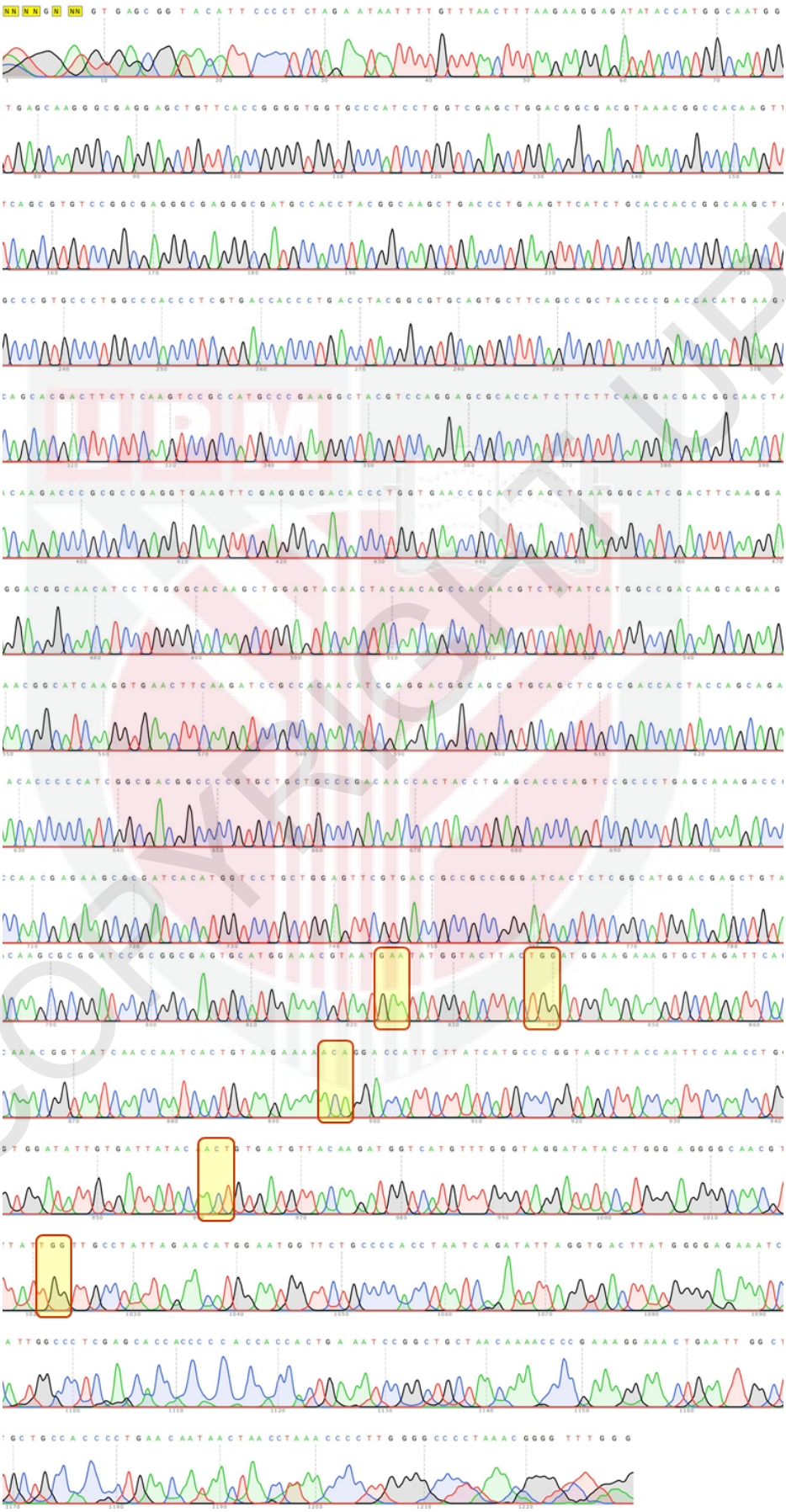
Expected sequence of D2_GFP_SH3 (1042 bp), Red: mutation region

5’-
CCATGGCAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCC
GACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
CTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT
GGACGAGCTGTACAAGCGCGGATCCGCGGCGAGTGCATGGAAACGTAATGAATATGGTA
CTTACTGGATGGAAGAAAGTGCTAGATTCACAAACGGTAATCAACCAATCACTGTAAGA
AAAACAGGACCATTCTTATCATGCCCGGTAGCTTACCAATTCCAACCTGGTGGATATTGT
GATTATACAACTGTGATGTTACAAGATGGTCATGTTTGGGTAGGATATACATGGGAGGGG
CAACGTTATTGGTTGCCTATTAGAACATGGAATGGTTCTGCCCCACCTAATCAGATATTA
GGTGACTTATGGGGAGAAATCAGTGGCCTCGAGGG -3’

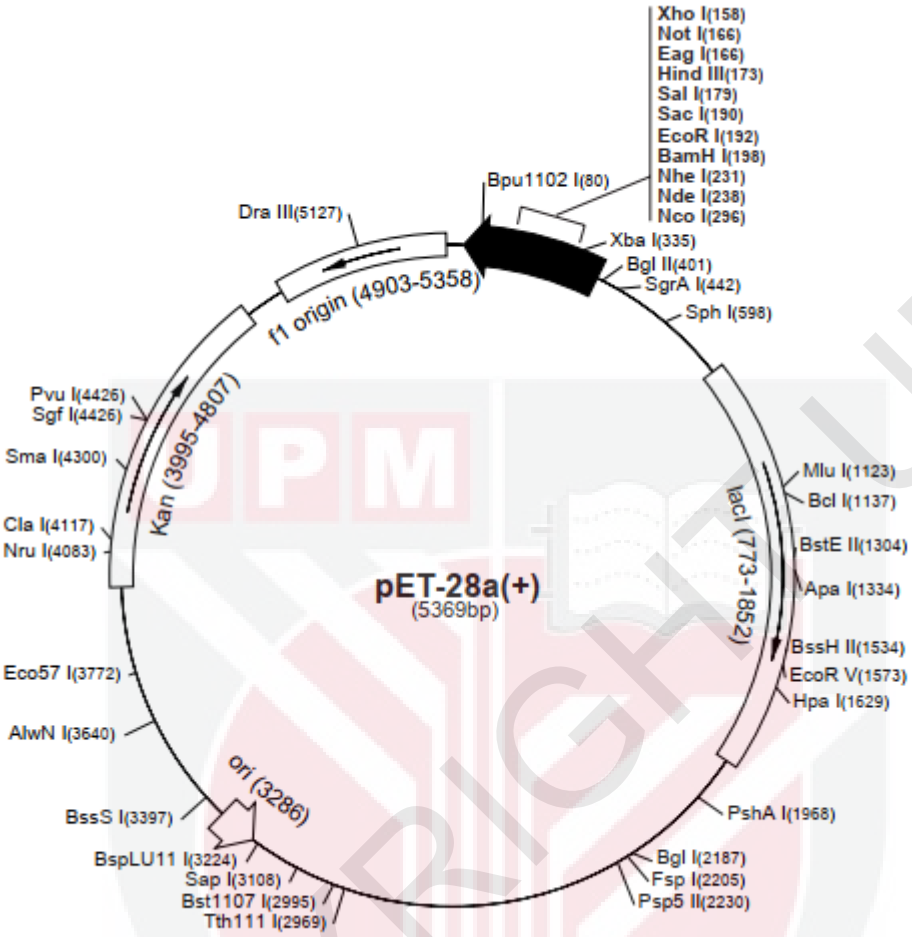
Expected D2_GFPSH3 sequence (Top) vs sequencing result (Below)



DNA chromatography:



Schematic Diagram of pET28a used in this study.



APPENDIX B:

Buffer composition for SDS-PAGE gel preparation and analysis

Composition	12 % Resolving Gel	4 % Stacking Gel
dH_2O	3.3 mL	2.1 mL
10% (w/v) SDS	0.1 mL	0.03 mL
Resolving/Stacking buffer	2.5 mL of Resolving buffer	0.38 mL of Stacking buffer
30 % Acrylamide solution	4 mL	0.5 mL
10 % Ammonium persulfate	0.1 mL	0.03 mL
TEMED	0.004 mL	0.003 mL

Note: Preparation for two gels

Buffers: -

Resolving buffer	
Tris 1.5 M	36.34 g of Tris base
Deionized dH_2O	Top up to 200 mL -adjust pH to 8.8 using HCl

Stacking buffer	
Tris 0.5 M	15.76 g of Tris-HCl
Deionized dH_2O	Top up to 200 mL -adjust pH to 6.8 using HCl

10X SDS running buffer	
Tris 0.277 M	60.4 g of Tris base
Glycine 2.13 M	288 g
SDS 0.0385 M	20 g
Deionized dH_2O	Top up to 2 L

4X Loading buffer	
Stacking buffer	5 mL
10 % SDS	8 mL
80 % Glycerol	4 mL
DTT	0.62 g
Bromophenol Blue	4 mg
Deionized dH_2O	Top up to 20 mL

Coomassie Blue Staining solution	
Coomassie R250	1 g
Glacial Acetic acid	100 mL
Methanol	400 mL
Deionized dH_2O	500 mL

De-staining solution

Methanol	200 mL
Glacial Acetic acid	100 mL
Deionized dH_2O	700 mL

10 X Western blot transfer buffer

Tris-base	30 g
Glycine	144 g
Deionized dH_2O	Top up to 1 L -Adjust pH to 8.3

10 X Tris-buffered Saline (TBS)

Tris base	25 g
NaCl	88 g
Deionized dH_2O	Top up to 1 L -Adjust pH to 7.6

1 X TBST

10 X TBS	100 mL
Tween 20	1 mL
Deionized dH_2O	Top up to 1 L

APPENDIX C: Molecular docking amino acid interaction profile table

Staphylococcus aureus

Endo88

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu	Y389, F415	K388, Y389	
L-Lys	I412		
D-Ala (A4)	P419, K411		
D-Ala (A5)			K411
G1		Y393	
G2		E434, R452	
G3		R452	
G4		I412	
G5			

Green: Stem peptide; Blue: Pentaglycine

D1

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu	F389	Q388	
L-Lys			
D-Ala (A4)	K411, P419		
D-Ala (A5)			K411
G1		E395	
G2			
G3		R452	
G4		Y454	
G5			

Green: Stem peptide; Blue: Pentaglycine

D2

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu	Y389	E388	
L-Lys			
D-Ala (A4)	P419		
D-Ala (A5)		T412	K411
G1		R452	
G2			
G3			
G4		Y389	
G5			

***Dry docking. Green: Stem peptide; Blue: Pentaglycine**

Staphylococcus epidermidis

Endo88

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu	Y389		
L-Lys	I412	Y389	
D-Ala (A4)			
D-Ala (A5)			
G1		E395	
G2		E395	
S		Y393, E434	
G4		E434, R452	
G5		R452, E434	

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

D1

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu	F415, F389		
L-Lys	F415, F389		
D-Ala (A4)			
D-Ala (A5)			
G1		E395, Y393	
G2			
S		Y393, E434, N387	
G4		E434, R452	
G5		Y454	

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

D2

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu			
L-Lys			
D-Ala (A4)			
D-Ala (A5)			
		-Not In Binding Groove-	
G1			
G2			
S			
G4			
G5			

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

Enterococcus faecalis

Endo88

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu	K388	K388, N387	
L-Lys		R452	
D-Ala (A4)			
D-Ala (A5)			K388
Ala-1	E434	E395, E434	
Ala-2		E434, R452	

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

D1

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu	F389	Q388	
L-Lys		G412, R452, E434	
D-Ala (A4)			
D-Ala (A5)		G412, Q451	K411
Ala-1		Y393, N387	
Ala-2			

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

D2

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu		Y389	
L-Lys		Y389	
D-Ala (A4)			
D-Ala (A5)		T412	
Ala-1			
Ala-2			

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

Micrococcus luteus

Endo88

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
Gly		L416, S417	
D-Glu	F415		
L-Lys	F415	Y389	
D-Ala			
D-Ala	Y389		
L-Lys		G450	
D-Glu			
Gly		Y393	K385
L-Ala			

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

D1

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
Gly			
D-Glu	F389		
L-Lys		G412	
D-Ala			K411
D-Ala		R452	
L-Lys		T433	
D-Glu		Y393	
Gly			K385
L-Ala			

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

D2

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
Gly		L416, S417	
D-Glu			
L-Lys	F415	T412, Y389	
D-Ala	K411		K411
D-Ala			
L-Lys		R452	
D-Glu		N387	
Gly			
L-Ala		W393, R386	K385

***Dry docking. Green: Stem peptide; Blue: Peptide cross-link**

Lactobacillus plantarum

Endo88

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu			
m-Dpm			
D-Ala	Y454		
m-Dpm	Y393	Y454, N387, R452 M394,E395	

***Dry docking. Green: Stem peptide; m-Dpm: meso-Diaminopimelic Acid**

D1

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu			
m-Dpm			
D-Ala		Q388, N387	
m-Dpm	Y393	R452, E434	

***Dry docking. Green: Stem peptide; m-Dpm: meso-Diaminopimelic Acid**

D2

Residues on peptide cross-link	Hydrophobic interaction	Hydrogen bond	Salt-bridges
D-iGlu		T412	
m-Dpm		T412	
D-Ala	W454	N387	
m-Dpm	W393	E395	

***Dry docking. Green: Stem peptide; m-Dpm: meso-Diaminopimelic Acid**

BIODATA OF STUDENT

Mr. Tham Hong Yun was born on 26th December 1995 in Kuala Lumpur. He completed his primary and secondary school education in SJK(C) Batu 11 Cheras and Seri Suria School. He pursued his A-levels at Sunway College before continuing his Bachelor's degree at the University of Nottingham, Malaysia Campus. Passionate about research, he has always aimed to contribute to the scientific community. After completing his degree, he continued his academic journey by enrolling in a Master's program at University Putra Malaysia, which he later converted to a PhD. In his free time, he enjoys playing computer games while remaining dedicated to advancing his research career.

LIST OF PUBLICATIONS

Chandran, C., Tham, H.Y., Abdul Rahim, R., Lim, S.H.E., Yusoff, K., Song, A.A.-L., 2022. Lactococcus lactis secreting phage lysins as a potential antimicrobial against multi-drug resistant Staphylococcus aureus. PeerJ 10, e12648. https://doi.org/10.7717/peerj.12648

Krishnan, M., Tham, H.Y., Wan Nur Ismah, W.A.K., Yusoff, K., Song, A.A.-L., 2024. Effect of Domain Manipulation in the Staphylococcal Phage Endolysin, Endo88, on Lytic Efficiency and Host Range. Mol Biotechnol. https://doi.org/10.1007/s12033-024-01216-4

Tham, H.Y., Song, A.A.-L., Yusoff, K., Tan, G.H., 2020. Effect of different cloning strategies in pET-28a on solubility and functionality of a staphylococcal phage endolysin. Biotechniques 69, 161–170. https://doi.org/10.2144/btn-2020-0034

Tham, H. Y., Chong, L. C., Krishnan, M., Khan, A. M., Choi, S. B., Tamura, T., Yusoff, K., Tan, G. H., & Song, A. A.-L. (2025). Characterization of the host specificity of the SH3 cell wall binding domain of the staphylococcal phage 88 endolysin. Archives of Microbiology, 207(2), 47. https://doi.org/10.1007/s00203-025-04242-1