Analysis of the protein-protein interactions involved in the activity of colicin in *Escherichia coli*

RIHAF ALFARAJ

A thesis submitted in partial fulfilment of the requirements of the University of Brighton for the degree of Doctor of Philosophy

August 2021

Declaration

I declare the research contained in this thesis, unless otherwise formally indicated within the text, is the original work of the author. The thesis has not been previously submitted to this or any other university for a degree, and does not incorporate any material already submitted for a degree.

Signed

Dated

This work is dedicated to the memory of my loving mother, you will always be with me.

Acknowledgments

I would like to express my deepest appreciation to my supervisors Dr. Ian Cooper, Dr. Lara Barnes and Dr. Lucas Bowler for their guidance through this journey, for their supervision, support and help in regard to research without their persistent help this thesis would not have been possible. I would like also to thank my supervisor at KSU Dr. Fadilah Aleanizy for her guidance, endless support, her encouragement and her passion, thank you for believing in me, thank you for going through the journey with me. Thank you Dr. Fulwah Alqahtani for being my teacher, my mentor, thank you for your endless care and support.

I would like to extend my sincere thanks to Dr. Simon Jeffs for his supervision at the beginning of this journey, Dr. Jon Salvage for his help with the confocal microscope and for Dr. Fergus Guppy and Dr. Sian Williams for their help with the statistics and to all administrative and technical staff at University of Brighton.

I would also want to say thank you to every who one has cooperated in this project. Thanks also to Dr. Matt Cliff at Manchester institute of Biotechnology for imparting his expertise in the study of NMR. Thanks to Dr. Chris Penfold at University of Nottingham for his cooperation. I also would like to thank Dr. Ahmad Bari and Dr. Awwad at King Saud University for their help with the NMR study.

I am highly indebted to King Saud University and External Joint Supervision Programme for their supervision and financial support.

I would like to express my special gratitude to my husband Waleed Alsalem for his enduring love for believing in me and for sharing my wish to reach the goal for completing this project.

To my father for his endless love, for his caring, for his sense of humor, for always telling me that everything is going to be ok. I love you so much.

Thanks to my beloved family for always supporting, helping, praying and standing by me, thank you my dearest brothers Raeed and Abdullah, Thank you my loving and supportive sister Reham and to my little sister and daughter Toona. Thank you Alaa just for being part of my family, and welcome to the world my baby nephew Khalid. Thank you, my little sweet cousin Kholoud, for your artistic gift.

Thank you, Fai Alkathiri for everything. I would like to thank my friend Sarah Bukhari and her family. And finally, to my dearest friends through my life's journey, Hind, Haifa and Shams Albdoor, I am grateful that you are in my life.

Abstract

Colicins are a type of bacteriocin produced by some bacterial species in response to stress. Colicins are cytotoxic to competing cells. They are classified into two groups: group A, also called “Tol-dependant colicins” and group B called “Ton-dependant colicins”. To exert their cytotoxic effects, colicins first need to be translocated into the target cells via the periplasm; however, the exact mechanism by which colicins achieve this translocation remains unclear. For Group A colicins, translocation requires an interaction between the N-terminal domain of the colicin and a series of membrane-bound and periplasmic proteins known as the Tol system. The translocation also includes interactions with more complicated assemblies. Recent developments in confocal laser scanning microscopy and in-cell nuclear magnetic resonance (CLSM and NMR, respectively) have prompted more studies of protein–protein interactions in living cells. CLSM and in-cell NMR spectroscopy were used to study the interaction between an overexpressed colicin translocation domain and periplasmic machinery. Cloning the full-length of the translocation domain, and the TolA, and TolB box sub-regions was accomplished. Recombinant protein expression was optimised to meet the requirements for CLSM and in-cell NMR analysis. Several factors, including culture medium and high cell density, were considered to have produced a successful and high level of recombinant proteins. The antimicrobial activity of colicin E9 on different *E. coli* cells expressing either the translocation domain or the TolA or TolB box was examined, and it was found that overexpressing the translocation domain, or part of it, provided some protection against colicin. Moreover, CLSM results indicate that TolA plays a key role in occupying the periplasm and preventing translocation of external colicin E9 to its target, which was found based on the high level of protection against ColE9. The same cells along with a positive control, protein GB1, were also used to develop a method to detect proteins of interest using in-cell NMR. Several parameters were investigated, including a labelling technique, culture medium, and the concentration of the in-cell sample. Analysis of 60% slurry (v/v) showed better NMR spectra, indicating the presence of an unfolded protein or the unstructured part of the translocation domain or TolB box. Furthermore, the data suggest that the expressed proteins, especially the TolA box, are involved in many interactions that can lead to the formation of large complexes responsible for occupying the periplasm and thus prevent translocation and consequently, activity of externally applied colicin.

Table of contents

List of Figures

List of tables

List of Abbreviations

µL	**microliter**
^{13}C	**Isotopic labelled carbon**
^{15}N	**Isotopic labelled nitrogen**
^{19}F	**Isotopic labelled fluorine**
1D	**One- dimensional**
^{1}H	**Isotopic labelled hydrogen**
2D	**Two- dimensional**
3D	**3- dimensional**
AB-3	**Acrylamide- bisacrylamide**
Arg	**Arginine**
APS	**Ammonium persulphate**
Asp	**Aspartate**
ATP	**Adenosine triphosphate**
BiFC	**Bimolecular fluorescence complementation**
bp	**Base pair**
BtuB	**Vitamin B12**
C- domain	**Cytotoxic domain**
CFU	**Colony forming unit**
CLSM	**Confocal laser scanning microscopy**
ColA	**Colicin A**
ColE9	**Colicin E9**
COSY	**Correlated spectoscopy**
D_2O	**Deuterium oxide**
DAPI	**4,6-diamino-2-phenylindole**
DNA	**Deoxyribonucleotide**

dH_2O	**Distilled water**
dNTPs	**Nucleotide triphosphate**
DTT	**Dithiothreitol**
df	**Degree of freedom**
E. coli	***Escherichia coli***
FRET	**Forser resonance energy transfer**
GFB	**Green fluorescence protein**
Glu	**Glutamine**
h	**hour**
HSQC	**Heteronuclear single quantum coherence spectroscopy**
IM	**Inner membrane**
Imm9	**Immunity protein of colE9**
IPTG	**Isopropyl β-D-1-thiogalactopyranoside**
IR	**infrared**
kB	**kilo**
kDa	**Kilodalton**
min	**minute**
ml	**milliliter**
mM	**Milli mol**
MWCO	**Molecular weight cut off**
ng	**nanogram**
nM	**Nano mol**
NMR	**Nuclear magnetic resonance**
OBS1	**OmpF binding site 1**
OD	**Optical density**
OM	**Outer membrane**

OmpF	**Outer membrane porin F**
p value	**Probability value**
Pal	**Peptidoglycan-associated lipoprotein**
PBS	**Phosphate buffer saline**
PCR	**Polymerase chain reaction**
PG	**peptidoglycan**
PI	**Propidium iodide**
PMF	**Proton motive force**
PPI	**Protein- protein interaction**
ppm	**Parts per million**
PVDF	**Polyvinylidene difluoride**
R- domain	**Receptor- binding domain**
rpm	**Rate per minute**
s	**Second**
SD	**Standard deviation**
SDS	**Sodium dodecyl sulphate**
Ser	**Serine**
SPR	**Surface plasma resonance**
SOC	**Super optimal broth**
TA	**Translocation domain of Colicin A**
TBE	**TolB binding epitope**
TBS	**Tris buffer saline**
TBST	**Tris buffer saline+ tween**
TCI	**Triple resonance cryoprobe**
TEMED	**Tetramethylethylendiamine**
Tyr	**Tyrosine**
Try	**Tryptophane**

UV	**Ultraviolet**
V	**Volt**

Chapter 1

Introduction

1.1 Introduction

Escherichia coli (*E. coli*) is a Gram-negative, non-sporulating, facultative anaerobic microorganism primarily found in the human gastrointestinal tract. It is a prokaryotic, single-celled organism able to cope with changes in the chemistry of the environment and genetic modifications (Huang *et al.*, 2008). *E. coli* is commonly used as a model organism in the molecular biology lab, chiefly because it multiplies rapidly and easily cultured. *E. coli* species range from pathogenic strains, which can cause outbreaks of severe diseases, to non-pathogenic isolates, which form part of the normal intestinal microbiota in humans. The pathogenicity of an *E. coli* strain is determined by the possession of genes coding for virulence factors such as adhesins, invasins, toxins, and capsules (Kuhnert, 2000). Some strains, both virulent and toxigenic—for example, extra-intestinal pathogenic *E. coli*—cause a various gastrointestinal disease leading to hundreds of thousands of cases of diarrhoea and even severe kidney damage each year (Rasko *et al.*, 2008).

Along with other members of the *Enterobacteriaceae* family, *E. coli* is known to produce toxins referred to as bacteriocins (Sharp *et al.*, 2017). Discovered in 1925 by André Gratia (Gaillard-Gendron, 2000), bacteriocins are generated by a diverse group of microorganisms, including yeast, protozoa, and pathogenic bacteria. *E. coli*, are bioactive, antimicrobial peptides or active proteins that can cause cellular damage to other microbial species, despite having no effects on their human hosts (Lie *et al.*, 2019). They are reported to have a molecular size of up to 80 kDa and a narrow spectrum of activity, with their bioavailability being defined in the context of their therapeutic potential (Lobmann *et al.*, 2019). Within this group of chemical actives, bacteriocins differ in their mechanism of cytotoxic action, and their relative size, structure, and release process, the last of which

varies depending on the species producing them (Hol *et al*., 2014). Bacteriocins differ from classic antibiotics because they are synthesised by ribosomes and deactivated by proteases. They are classified into two types: microcin, with a molecular weight of 10 kDa or less produced by Gram-negative bacteria, such as *Enterobacter*, and colicins. Colicins are plasmid-encoded, vary in molecular weight from 25 to 80 kDa, and are produced by *E. coli* and other members of *Enterobacteriaceae* (Ghazaryan *et al*., 2014a).

Although their mechanisms of action and microbial targets vary, bacteriocins are generally lethal to other species of bacteria, which compete for the same limited resources and space. They are closely related to the bacteriocins produced by other strains of the host species. The production of bacteriocins by Gram-negative bacteria such as *E. coli* occurs during stress or exposure to specific chemical triggers (Riley, 1993). These can include DNA-damaging agents, cold shock, nutrition depletion, and mutagenic triggers, to which bacteriocins interact through a regulatory mechanism (Ghazaryan *et al.*, 2014a). The producing strains of bacteriocins protect themselves from toxin-induced death by binding to a cognate immunity (Imm) protein co-expressed or naturally present inside the cells (Matsumoto-Nakano & Kuramitsu, 2006).

The apparent purpose of bacteriocin's production is to provide the producing strains with an advantage in the competition for resources and space between bacteriocin-producing bacteria and susceptible bacteria. The level of activity and competition depends on the bacteriocin's specific structure, the availability of nutrients, and the population cell density within a specific environment (Ghazaryan *et al*., 2019). The producing strains are not totally immune to the effects of the bacteriocins. However, an advantage of such competition is that it will lead to the equal distribution of nutrients that benefit the entire

population of bacteriocin-producing and bacteriocin-susceptible strains of the same species. This because bacteriocins will affect the producing strains once they reach the limit at which their production cannot be compensated for by binding with an Imm protein to confer immunity on that cell (Ghazaryan *et al*., 2019). For example, colicin-producing cells are capable of resistance to the effects of the colicins that they produce because they produce an Imm protein. The Imm protein can bind to the colicin to neutralize its cytotoxic effect (Majeed *et al*., 2013). It must be stated that the entire mechanism of colicin production and secretion into the environment is not fully understood. One feature worthy of investigation is the secretion mechanism itself. Because colicin-producing strains lack a colicin secreting system, which should be able to transfer colicin from the cytoplasm to the environment, the release of colicin requires > 3% of colicin producing strains to die by lysis (Inglis *et al*., 2009).

Bacteriocins thus play two roles: a defensive one, which involves preventing the attack on their population, and an offensive one, as observable in colicin-producing strains that give bacteriocin-sensitive cells a competitive edge within a mixed population involving another species that is susceptible to the bacteriocins. However, the exact competition mechanism between bacteriocin-producing strains is still unknown (Ghazaryan *et al*., 2019).

Durrett & Levin (1997) studied the effect of a small population of colicin-producing *E. coli* strains against sensitive *E. coli* strains. The researchers observed that in heterogeneous environments, both producing and sensitive strains were able to co-exist. Nevertheless, in environments such as agar plates or within the colons of mice, an established colony of either strain could resist invasion by the other.

Majeed and colleagues (2011) studied the competitive interactions between colicin-producing strains, as well as a cytotoxic effect; the production of a colicin by one strain can induce the production of colicin by another strain. Further research (Majeed *et al*., 2013) explored the capacity at which different strains were able to produce colicin. When a strain produced a colicin at a faster rate (also known as stronger rate) than another, the slower (also known as weaker) strain would eventually outcompete the faster one. If exposure to a colicin from one strain stimulates the production of colicin by another strain, it is theorized that the faster strains produced too much of their own colicin to compensate for cross induction with the weaker strain, as well as producing their own colicin.

In April 2019, a report from the United Nations revealed that antibiotic-resistant infection causes the death of approximately 7 million people worldwide each year (Dadgostar, 2019). That and other reports demonstrating a rise in cases of nosocomial infections in the past decade have shifted the attention of scientists to multidrug, antibiotic resistant bacteria, especially Gram-negative bacteria, which possess an outer membrane (OM) that impedes the entry of many antibiotics (Behrens *et al.*, 2017). Regardless of how bacteria are transmitted, the number of species demonstrating resistance to various antibiotics has risen, which has prompted researchers to investigate bacteriocins as a possible replacement for antibiotics. *In vitro* studies have shown that some species of bacteria, even though resistant to some antibiotics, remain sensitive to some bacteriocins (Behrens *et al.*, 2017). There is a substantial knowledge gap about the activity of bacteriocins *in vitro*, but especially *in vivo*. Recently, more protein bacteriocins have been identified, only approximately 100 have been studied *in vitro*, and only eight have been studied *in vivo*. The results of such

studies have nevertheless indicated promise for developments in the field and for discovering novel alternatives to antibiotics (Telke *et al.*, 2019).

Bacteriocins are currently in use in the food industry as antibacterial compounds. An example is nisin and lactacin. These have shown activity against vancomycin-resistant *Enterococci* and methicillin-resistant *Staphylococcus aureus* (Simon *et al.*, 2020). Another interesting finding is that bacteriocins also possess properties other than antimicrobial ones, including anticancer properties (Baindara *et al*, 2018). The cytotoxicity against cancer cells depends on the structural properties of the bacteriocins, their hydrophobicity, and the number of positively charged amino acids (Kaur, 2015). Bacteriocins target the negatively charged surface molecules on the cancer cells and induce cell membrane depolarization and subsequent cell death (Kaur, 2015). More recently, colicins have been shown to exert no adverse effects on humans. For example, no reports of allergies or hypersensitivity to colicins have been documented (Lobmann *et al.*, 2019). Meanwhile, other bacteriocins have yet to be fully characterized, and determining their mechanism of action has become an exciting objective of studies into their potential as therapeutic agents (Ghodhbane *et al*., 2015).

1.2 Colicins

According to the literature, approximately 54% of *E. coli* strains produce colicins with demonstrable cytotoxic activity against competing bacteria (Žgur-Bertok *et al*., 2012). Colicins are plasmid-encoded, each by a specific colicin gene (*cxa*). *cxa* is encoded in the first open reading frame in the colicin operon. In turn, the *cxa* gene is accompanied by another gene (*cxi*) that encodes an Imm protein, which is expressed constitutively by its promoter and protects the colicin-producing cells from its own colicin's cytotoxic effect by interacting with the cytotoxic domain (C) of the produced colicin (Kim *et al*, 2014). As described by Mavridou and colleagues (2018), each colicin operon has an associated lysis gene (*cxl)*, whose product is localized to the periplasmic space and is responsible for the release of colicin to the extracellular medium, called the bacteriocin release protein (BRP) (Casacales et al., 2007).

Based on their size and the translocation machinery they use to access target cells, colicins fall into two groups: group A and group B (Ghazaryan *et al*., 2014b). Colicins in group A are encoded by small, multi-copy plasmids of 6–10 Kb and include colicins A, E1 to E9, K, L, N, S4, U, and Y. Some of those colicins are co-expressed with (Imm) and secreted into the extracellular medium. Similarly, group B, such as colicins B, D, H, Ia, Ib, M, 5, and 10 — are also coupled with Imm proteins and secreted into the extracellular medium. However, they are encoded by large, mono-copy plasmids of 40 Kb. (Cao & Klebba, 2002; Cascales *et al.*, 2007). Colicins in both groups have the same structural organisation—which Cascales and colleagues (2007) have described, it comprises three functional domains, as shown in Figure 1.1.

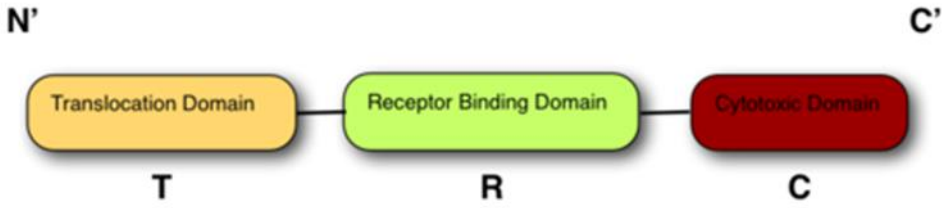


Figure 1.1 The structural organisation of colicins (Cascales *et al*., 2007). Three domains span the N-terminus (N') to the C-terminus (C'): the translocation (T) domain at N' and the receptor-binding (R) and cytotoxic (C) domains at C'.

Located between the C-terminus and N-terminus and with a hairpin-like structure, the receptor-binding (R) domain engages in receptor recognition. At one end, the first 83 residues of the (T) domain crystal structure are thought to be unstructured due to the lack of electron density in the region of the C-terminus (Soelaiman *et al*., 2001). At the other terminus is the (C) domain which is responsible for the cytotoxic activity of colicins (Cascales *et al*., 2007).

The therapeutic potential for colicins to act as alternatives to antibiotics has also been explored. Colicins have been shown to play a role in *E. coli* colonisation of the gastrointestinal tract due to their role in the inter-species competition. Different *E. coli* strains produce up to 25 different colicins, and nearly all of those types have three plasmid-encoded domains; the exception is colicin JS, which has no domains, and it is discovered recently (Micenková, 2019). Colicins have been tested for the treatment of Crohn's disease, which is the chronic form of inflammatory bowel disease characterised by intestinal dysbiosis (Brown *et al*., 2015). The causing phenotype of *E. coli* in Crohn's disease is adherent–invasive *E. coli* (AIEC). Such strains of *E. coli* can form biofilms adhering to and lyse the epithelial layer. These are themselves a protective factor in the immune system of

host humans (Lee *et al*., 2019), as the biofilms stimulate the production of cytokines (Buisson *et al.,* 2019). These *E. coli* strains were isolated from the ileal mucosa of individuals with Crohn's disease and treated with colicin. Research showed that the colicin demonstrated potency against ileal Crohn's disease mucosa associated with AIEC strains and did not show any toxicity to the macrophages or even stimulate the production of cytokines (Brown *et al*., 2015). The efficacy of colicin was also compared to the efficacy of antibiotics such as metronidazole and ciprofloxacin, which revealed the greater bactericidal activity of colicin against bacterial biofilm. Those results suggest that colicin-producing bacteria isolated from the human gastrointestinal tract may be beneficial in addressing the antibacterial resistance of biofilm-associated cells.

In other research, a novel colicin, type Z, was isolated from extra-intestinal *E. coli* from the anorectal abscess of a 17-year-old man, and its spectrum of activity was tested (Micenková, 2019). Colicin Z has shown a selective potency against entero-invasive *E. coli* and *Shigella* strains. Its mechanism of action also has been investigated, and it was found to be targeting the peptidoglycan (PG) layer of bacterial cells, leading to PG degradation (Micenková, 2019).

Other examples illustrating the effectiveness of other colicins that have been explored are colicins M and E7, used against enterohaemorrhagic *E. coli* (Schulz and colleagues, 2015). Colicin M, which targets the PG layer, showed high potency against five strains of enterohaemorrhagic *E. coli* and gave a synergistic effect when it was mixed with colicin E7 that has a nuclease effect on sensitive bacterial cells.

As well as colicin's potency has been studied, its contribution as a toxin was also explored. Abd El-Baky and colleagues (2019) to studied the relationship between virulence factors and resistance genes of extra-intestinal pathogenic *E. coli* were isolated from urine and blood samples. Different colicin genes were tested, and they showed a positive correlation between colicin production and how they contribute to the research of *E. coli* pathogenicity. They concluded how these genes need to be controlled to reduce the severity of infections by further investigations.

1.2.1 SOS response and the release of colicins

It has been determined that DNA damage that regulates the production of colicins also induces the bacterial SOS response (Gillor *et al*., 2008; Budič *et al*., 2011). The SOS response, induced to respond to DNA damage (Maslowska *et al*., 2019), typically involves LexA and RecA proteins stimulated by single-stranded DNA interactions (ssDNA) (Cascales *et al*., 2007). The primary function of LexA during normal growth conditions is to inhibit the expression of genes involved in the SOS response, and induce DNA bending upon interaction with the operator of the SOS inducible genes such as colicin (Mavridou *et al*., 2017). This is achieved by binding to the SOS box on the associated promoter. However, when ssDNA becomes exposed during cellular stress or DNA damage, RecA is upregulated and binds to the ssDNA, forming a heteroduplex complex. Next, LexA dissociates from the SOS genes, allowing for the transcription and expression of the SOS response and allowing DNA repair to occur and then repression of the SOS induced genes (Žgur-Bertok, 2012). This process is summarised in Figure 1.2.

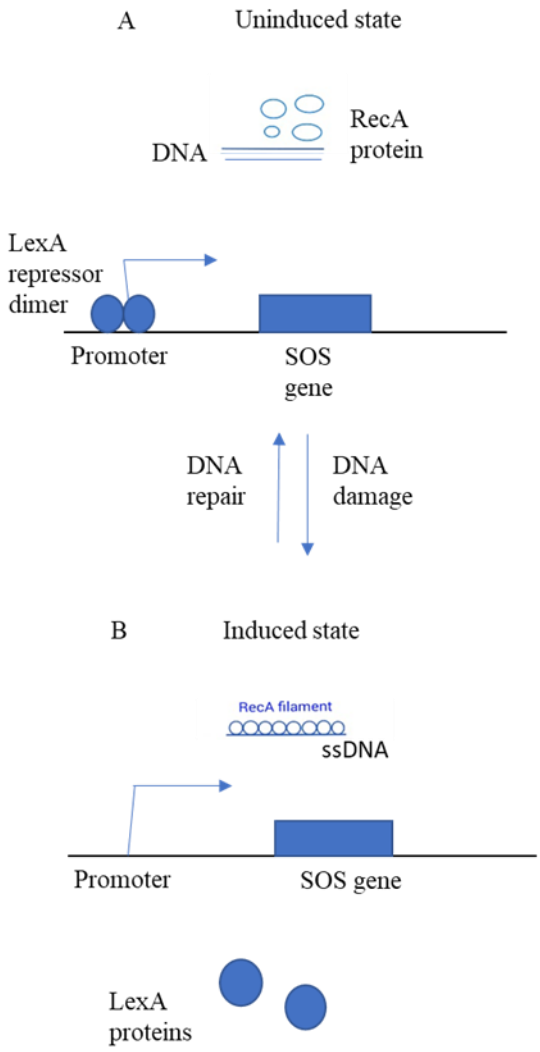


Figure 1.2 The role of LexA and RecA proteins in the SOS response in bacterial cells (adapted from Michel, 2005). A) in bacterial cells under normal conditions (i.e., uninduced state), SOS proteins such as LexA and RecA are downregulated. LexA binds to SOS-associated genes, resulting in the downregulation of DNA repair mechanisms. B) during conditions of DNA damage, SOS response is stimulated (i.e., induced state). DNA damage upregulates RecA, which binds to damaged single-stranded DNA (ssDNA), thereby forming a RecA heteroduplex complex (i.e., nucleoprotein filament). RecA/ssDNA nucleoprotein filament acts as co-protease leading to the dissociation of LexA from the SOS-associated genes, thereby allowing SOS gene expression to initiate DNA repair, the cells will return to uninduced state where the normal replication continues.

Once the SOS response is induced, several key processes commence. Firstly, since colicins are expressed from operons under the control of SOS promoter, and its synthesis is regulated by the LexA protein, upon induction, colicin plasmid operon is expressed, producing colicin that accumulates in the cytoplasm. Next, the OM phospholipase A is activated, facilitating the accumulation of lysophospholipids and free fatty acids in the OM (Snijder *et al*., 2003). This affects the OM's permeability and, in time, allows the release of colicins (Alonso *et al.*, 2000; Lloubès *et al*., 2013). Once present in the external environment, colicins are able to exert a cytotoxic effect on competing cells. However, at the same time, upregulation of the lysis gene occurs, which ultimately leading to the death of the colicin-producing cells (Majeed *et al.*, 2011). Colicins that are present in the external medium bind to the OM of the competing cells and promote protein–protein interactions (PPIs) in the OM and periplasm that facilitate another mechanism: translocation.

1.2.2 Translocation of colicins

In order to deliver a cytotoxic effect on a susceptible bacterial cell, the colicins need to be translocated across the OM by several protein mediators to reach their intracellular targets. It has been shown that different colicins utilise different translocation pathways. This determines which specific interactions occur during its translocation and sub-classifies the colicins into one of two groups: nuclease (enzymatic) or pore-forming colicins (Cursino *et al*., 2002). However, it has been observed that sometimes colicins can struggle to enter bacterial cells, this is believed to be linked to mutations on the OM receptors that affect their receptor-binding capacity or an abnormality in their translocation systems (Budič *et al.*, 2011). As different strains of bacteria present with different susceptibilities to colicins,

susceptibility is reported to be a phenotypic marker due to mutations in the colicin translocation system, with resistance due to the absence of an appropriate receptor protein in the OM of Gram-negative bacteria (Smarda, 1992).

Significant research has been undertaken to clarify how colicins translocate across the OM, and this has been linked to either Tol- or Ton-dependent systems in the periplasm of bacterial cells directly linked to the OM. Colicins in group A are Tol-dependent, meaning that they use the Tol system for translocation, whereas ones in group B are Ton-dependent, meaning that they use the Ton system (Alonso *et al.*, 2000; Morales *et al*., 2015). Among colicins in group A, the *E. coli* K12 strain shows phenotypes similar to Tol mutants. Research has shown that Tol-associated genes influence the integrity of the OM and its invagination during cell division and the polar localisation of key chemoreceptors (Lloubès *et al*., 2001). They are also involved in the Tol system shared with peptidoglycan-associated lipoprotein (Pal) linked to the *E. coli* cell envelope. This consists of seven proteins: Pal and YbgC in the OM, TolB, and YbgF in the periplasmic space, and TolA, TolR, and TolQ linked to the inner membrane (IM). In the OM, Pal forms a complex with the periplasmic protein TolB, where the binding affinity constants with the energised protein TolA in the IM have been linked to the phenomenon of proton motive force (PMF) and the associated involvement of the IM (Lloubès *et al*., 2001).

In order to facilitate translocation, colicins in group A target the OM receptor *BtuB,* which is involved in the high-affinity transport of vitamin B_{12} and is present at adhesion sites in the IM and OM during the initial steps of translocation (Collins *et al.*, 2002). They also require a second receptor, usually an OM porin (e.g., OmpF); an exception is colicin E1, which uses the TolC protein. This works in conjunction with OmpC and PhoE, which are

trimeric OM porins, whose physiological function is to allow the passage of small, hydrophilic nutrient molecules through the OM.

Conversely, colicins in group B bind to specific Ton-dependent transporters in the OM and, unlike ones in group A, require only one OM receptor (either FepA or FhuA). In addition, they do not require a secondary protein, although colicins 5 and 10, which use two proteins to bind to Tsx and TolC as a transport channel, are exceptions (Cascales *et al.*, 2007). In *E. coli*, as suggested by Dimove and colleagues (2005), the TonB protein is likely required to facilitate the transfer of PMF between the IM and OM via siderophore transporters such as FhuA, FecA, and FepA. After the group B colicin binds to receptors in the OM, the TonB–ExbB–ExbD complex allows the transduction of energy from the IM to the OM for the energy-dependent transport of siderophores, vitamin B_{12} and also for group B colicins to translocate across the OM (Dimov *et al*., 2005). Figure 1.3, which was adapted from (Kleanthous, 2010) depicts the mode of action of Group A and group B colicins which both include; nuclease (i.e., enzymatic) colicins and also pore-forming colicins. The pore-forming colicins depolarise the IM, which prompts ion efflux and thereby causing the depletion of cytoplasmic adenosine triphosphate (ATP) and, in turn, cell lysis (Housden *et al.*, 2005).

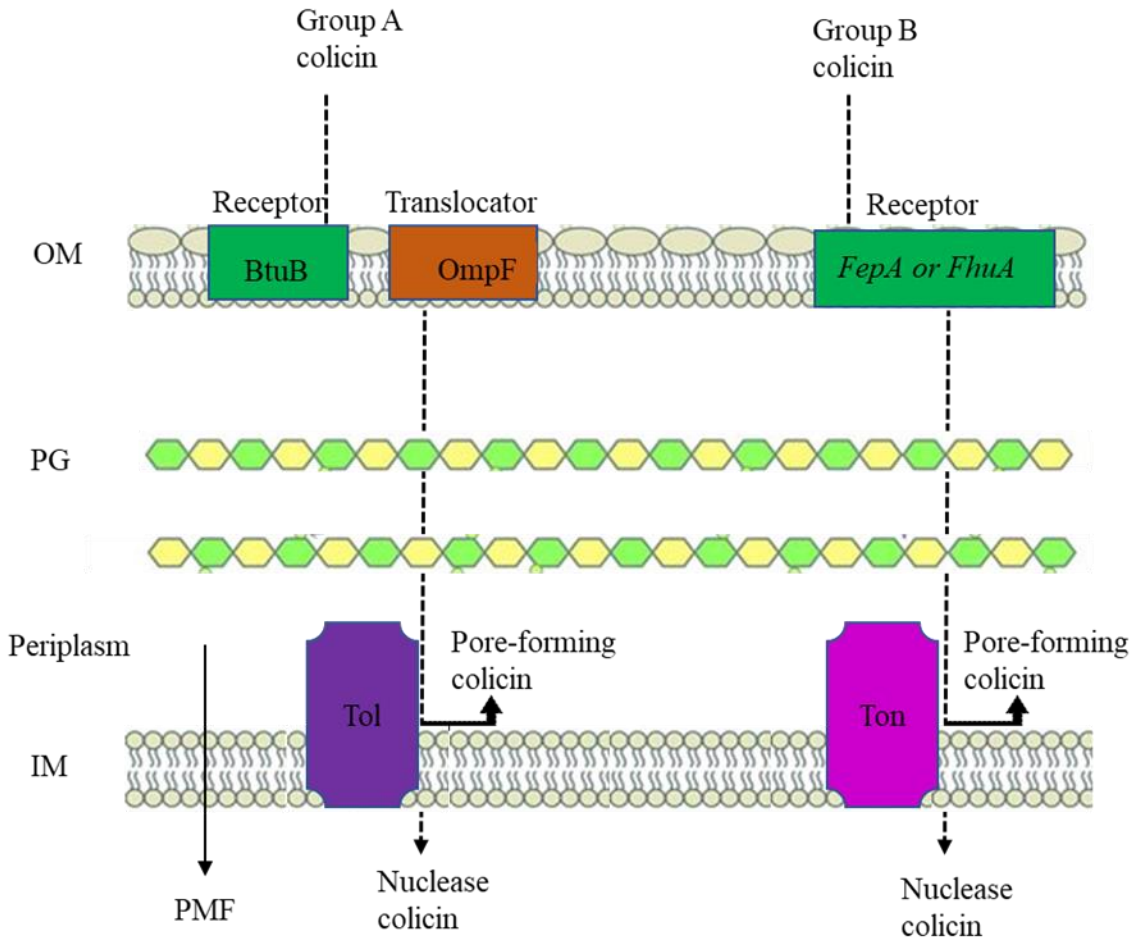


Figure 1.3 The mode of action of group A and group B colicins (adapted from Kleanthous, 2010). Both groups A & B colicins consist of nucleases that cause DNA or RNA cleavage and cell death. And pore-forming colicins, which depolarise the inner membrane (IM) and cause cell death. The difference is that group A colicins first bind to a receptor (i.e., BtuB) and translocator (i.e., OmpF) in the outer membrane (OM), after which they permeate the peptidoglycan (PG) layer, the periplasm, and the IM, where Tol proteins are recruited for translocation, a process enhanced by the presence of the proton motive force (PMF) in the IM. While group B colicins bind to one receptor, either *FepA* or *FhuA*, in the OM and translocate through the periplasm where they recruit Ton proteins.

1.2.2.1 Translocation of nuclease (group A) colicins

The nuclease (C) domain of colicins E2, E7, E8, and E9, acts to cleave the chromosomal DNA of colicin-sensitive *E. coli* cells. The four E colicins within this group share considerable sequence homology in their (R) and (T) domains, as well as exhibiting 80% homology in their (C) domains (Cascales *et al.*, 2007). To translocate across the OM, all nuclease colicins derive the required energy from the IM, which may aid not only protein unfolding upon entering the cell, but also the release of Imm proteins from the protein–colicin complexes (Vankemmelbeke *et al*., 2009; Kim *et al.*, 2014). Therefore, the multi-step import of colicins involves recognising and binding with receptors, recruiting Tol proteins if colicin is group A or the Ton proteins if colicin is in group B, and, in turn, traversing the cytotoxic domain (i.e., the (C) domain) to the site of toxicity (Figure 1.4).

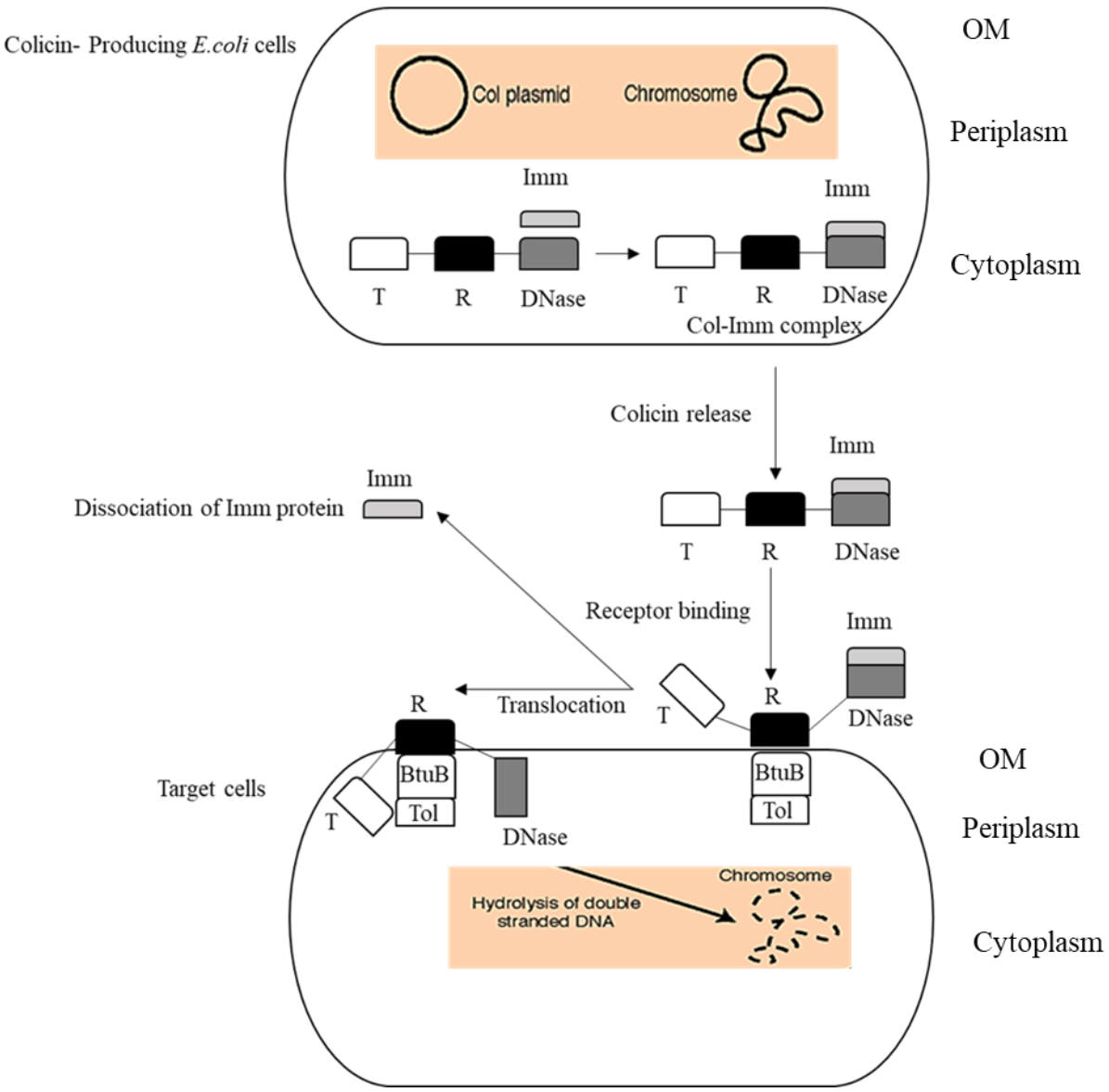


Figure 1.4 Diagrammatic representation of the translocation of nuclease colicins (adapted from Farrance *et al*., 2013). The functional domains of colicins consist of three domains: the translocation (T) domain, receptor binding (R) domain, and deoxyribonuclease (C) domain (i.e., DNase). Within the producing cell, they are bound with cognate immunity (Imm) proteins forming colicin–Imm protein complexes, which are secreted into the external environment. Once colicins bind to their receptor (i.e., BtuB) in the outer membrane (OM), the (T) domain translocate into the periplasm, allowing the dissociation of Imm protein from the (C) domain, which in turn, is internalised into the cytoplasm of the susceptible bacterial cell. Once internalised, the (C) domain is able to degrade the DNA of the susceptible cell, leading to cell death.

Imm proteins have a molecular weight of approximately 10 kDa and act to neutralise the cytotoxic activity of nuclease colicins by forming stable complexes. Once the group A colicins enter susceptible cells, the Imm proteins dissociate from those complexes, allowing the liberated (C) domain of the colicin to begin DNA cleavage, leading to cell death. The dissociated Imm proteins are detectable in the extracellular media after approximately 20 min after receptor-binding (Papadakos *et al.*, 2012).

1.2.2.2 Translocation of pore-forming (group A) colicins

The translocation of group A (pore-forming) colicins acts to disrupt the function of important intracellular components (e.g., ATP activity) and, as the pores form, transmembrane electrochemical gradient. The intracellular levels of potassium (K^+) and magnesium (Mg^{2+}) decrease, and by disrupting the voltage-dependent ion channels in the IM causes bacterial death (Alonso *et al.*, 2000), as shown in Figure 1.5.

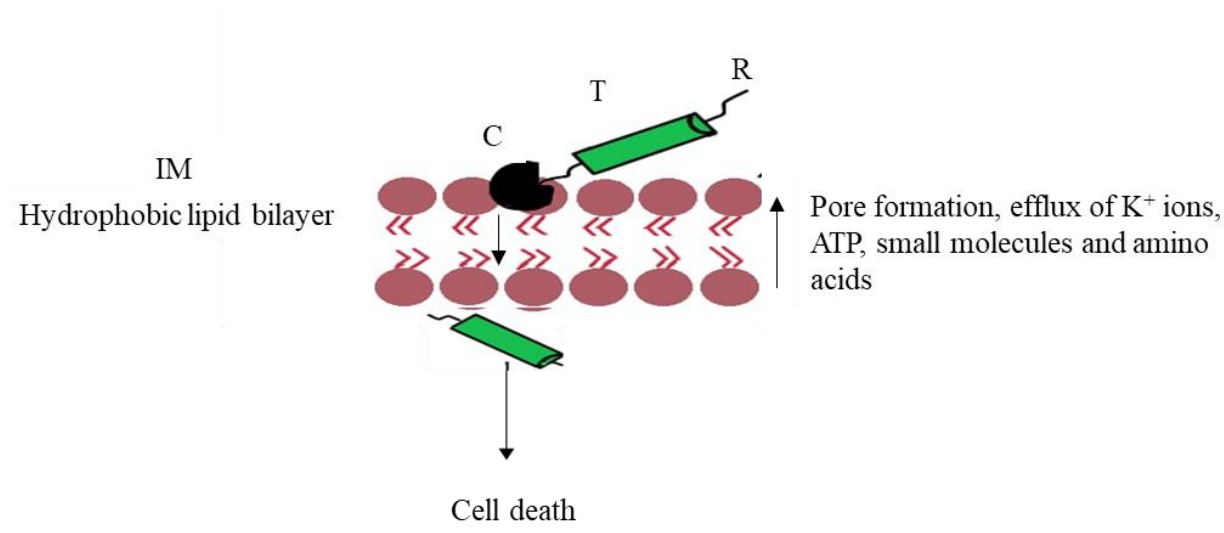


Figure 1.5 Pore formation by colicin (adapted from Bharti, 2015). While colicin and its structural domains (i.e., receptor binding (R), translocation (T), and cytotoxic (C) domains) enter the cells. Once the pore-forming cytotoxic (C) domain reaches the inner membrane (IM), here, the (C) domain acts to form a pore in the IM, leading to the loss of intracellular ions (K^+ and Mg^{2+}), as well as disrupting ATP activity, eventually leading to cell death.

Pore-forming colicins require an acidic pH to create pores in the IM. Zakharov and Cramer (2004) have observed that the initial interaction of the pore-forming domain (i.e. (C) domain) with the IM is electrostatic and that the (C) domain's orientation to the membrane surface is facilitated by a set of positive charges on the domain's surface. Once the (C) domain has bound to the IM, it unfolds, leading to the extension of the (C) domain's helices and its entry into the cytoplasm (Zakharov & Cramer, 2004).

Because pore-forming colicins are also cytotoxic to the bacteria that produce them, the co-synthesis of specific inhibitors (i.e., Imm proteins) is essential to protect the colicin producing strains against cell death. Indeed, a single molecule of colicin has been shown to kill producing cells (Johnson *et al.,* 2013). The Imm proteins of pore-forming colicins have been studied, and the gene encoded them to be located in the intergenic gap between the lysis gene and colicin on the plasmid (Metola *et al*., 2017).

A particular group A colicin, colicin M, has a mechanism of action unlike that of other colicins, one that kills sensitive *E. coli* cells in a process involving the hydrolysis of the PG lipid II intermediate. The primary structural component of the bacterial cell wall, PG, plays a role in protecting against external stress by maintaining the transport of nutrients across the OM of the bacteria and by preventing osmolysis. However, colicin M inhibits the polymerisation of PG, which results in cell lysis (Kamenšek & Žgur-Bertok, 2013).

1.3 Interaction of colicins with Tol proteins

The Tol–Pal translocation system consists of seven proteins. Five of these proteins (TolA, TolB, TolR, TolQ, and Pal) are involved in the translocation of group A colicins into the cytoplasm of a susceptible cell (Zakharov & Cramer, 2004). Of those five proteins, three are associated with the IM; TolA, TolQ, and TolR. These are linked together as a transmembrane unit, whereas TolB exists in the periplasmic space, and Pal is linked directly to the OM. This configuration is detailed in Figure 1.6.

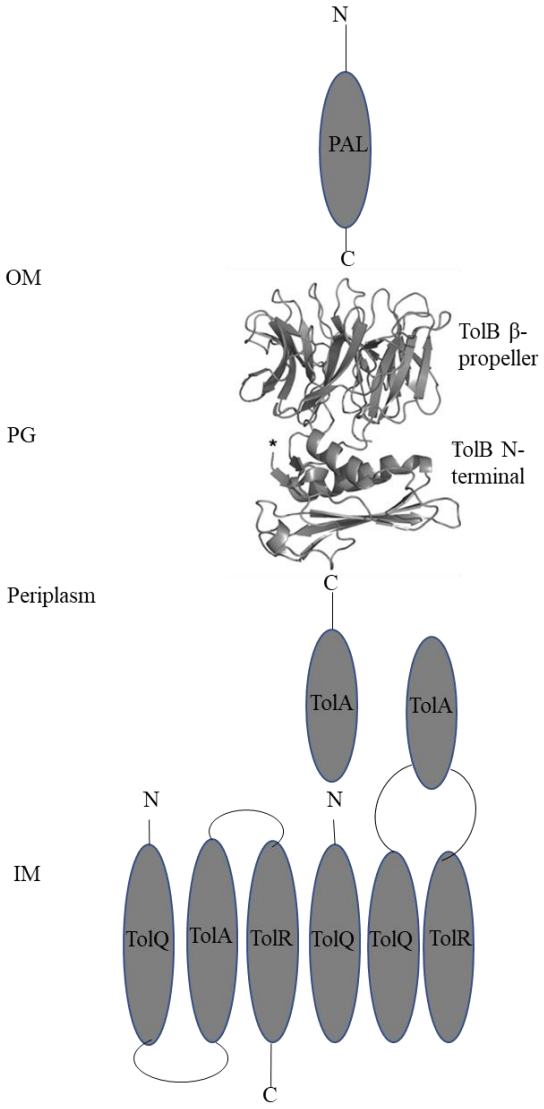


Figure 1.6 Tol–Pal system shown within Gram-negative bacterial cell wall (adapted from Lazzaroni *et al*., 2002). The location of Tol proteins with their domains (i.e., Pal, TolB (two domains, one bound to Pal where the other domain bound to TolA). TolA (three domains; one bound to TolB in the periplasm where the other two are bound to TolR and TolQ in the IM). TolR (two domains; both bound to TolA and TolQ) and TolQ (three domains; also bound to TolA and TolR)) are linked through the inner membrane (IM), directly through the periplasm, and extend into the outer membrane (OM). Upon binding with Tol- dependent colicin (i.e., group A), the Pal unit in the OM activates the β-propeller domain of the TolB unit in the peptidoglycan (PG), which in turn activates the TolB N-terminus unit within the periplasm, linking to TolA C- terminus, extending into the IM. This leads to binding with the remaining Tol proteins, thus providing the colicin with a physical link to translocate through the Gram-negative cell wall.

The functions of the five Tol proteins have been investigated by mutational analysis. This type of genetic manipulation designed to generate mutations within the Tol–Pal genes to determine their function. It has revealed that mutations in the Tol-Pal genes result in a defect in the integrity of the OM that, in turn, results in cellular hypersensitivity to detergents, drugs, and periplasmic proteins leaked into the external medium (Chan Li *et al*., 2012). Further research on TolA mutational analysis has revealed different uptake properties with colicin A (ColA), a pore-forming colicin, and E1, a nuclease colicin. Certain TolA point mutations have been isolated, and bacterial strains carrying these have demonstrated tolerance to ColA, but not to colicins E1, 2, 3, or K (Dubuisson *et al.*, 2002). Thus, the exact role of TolA remains unclear, with further research into the fundamental mechanisms still required (Zhou *et al.*, 2012; Rassam *et al*., 2018).

The TolA protein comprises three domains, combined with a molecular weight of 44 kDa (Egan., 2018), as illustrated in Figure 1.7. The first domain (Domain I) exists in the IM and is bound to other Tol proteins, namely TolQ and TolR, which interact at the N-terminus. Domain II exists in the periplasm and links Domain I and Domain III. Domain III exists in the periplasm as well and is responsible for PPIs with extracellular proteins.

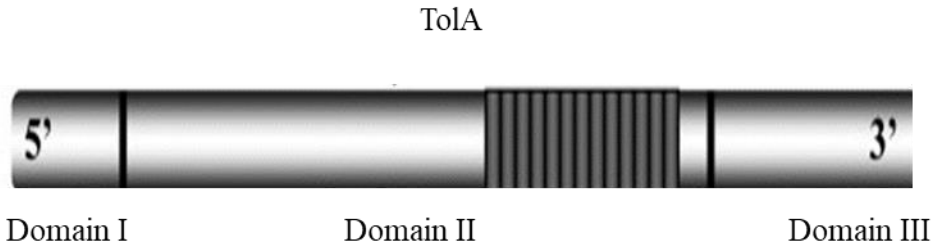


Figure 1.7 The structural organisation of TolA (Lazdunski et al, 1998). The figure depicts TolA, as well as its three Domains. Domain I exist within the inner membrane (IM); Domain II, III exist in the periplasm (the dark grey area is where the interactions with colicin occur), both domains are extended into the outer membrane (OM).

Research has demonstrated that TolA interacts with the TolB–Pal complex in the PG layer between the IM and OM and is directly linked to energy provided by PMF in the IM (Egan, 2018), as shown in Figure 1.8. The proteins in the IM (i.e., TolA, TolQ, and TolR) form a trimeric complex, which, in the presence of PMF, allows energy transfer through TolA to the Pal–TolB complex through the PG.

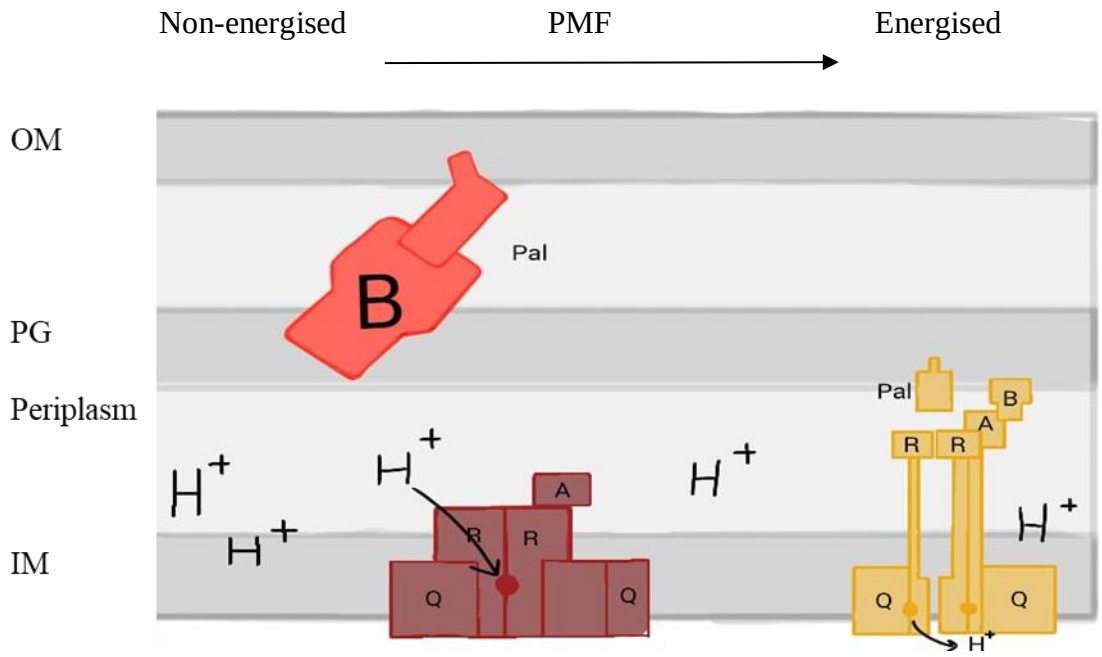


Figure 1.8 The interaction of TolA and the TolB–Pal complex in the Gram-negative bacterial cell wall, in association with hydrogen ion (H^+) movement during proton motive force (PMF) (adapted from Egan, 2018). The figure is showing (non-energised state), Tol proteins where TolB (B) and Pal in the outer membrane (OM) extending to the peptidoglycan (PG), where Tol (QAR) in the inner membrane (IM) and the presence H^+ of in the IM. In the presence of PMF, the accumulation of H^+ in the IM during PMF means that they are able to bind to the TolA units in the IM. By linking to the Pal-TolB units present in the PG, energy is able to be transferred to the OM (energised state).

However, in the absence of PMF, no interaction has been detected between Tol proteins in the OM and the IM. This subsequently inhibits the translocation of colicins across the membranes. Nevertheless, it has been detected that the association of other proteins in the IM, namely TolQ and TolR, enhance the interaction of TolA and the TolB–Pal complex (Lloubès *et al*., 2001).

In contrast to TolA, TolB has two domains, namely the β-propeller domain and the N-terminal domain, with a combined molecular weight of 47 kDa (Figure 1.9). The β-propeller is positioned where Pal binds to TolB C-terminus, whereas the N-terminal domain is positioned in the periplasm between the IM and OM (Bonsor *et al.*, 2009).

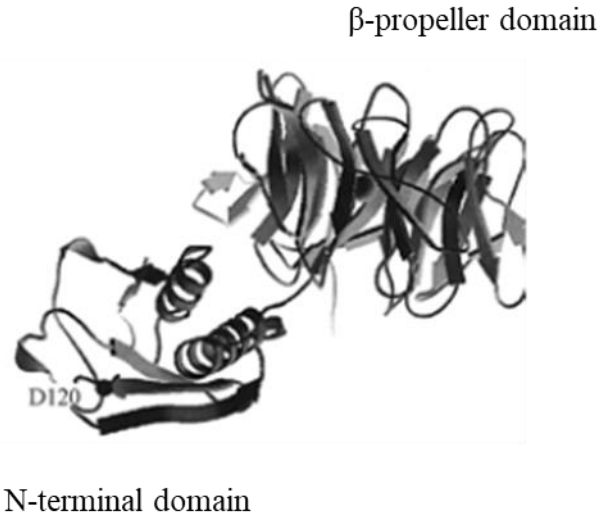


Figure 1.9 TolB structural organisation (Lazdunski *et al*, 1998). TolB structure includes the β-propeller domain existing in the OM and the N-terminal domain existing in the periplasm.

TolB has a total of 408 amino acid residues and exists in the periplasm, linking to the OM (Szczepaniak *et al*., 2020). The C-terminus, bound to Pal within the PG layer, consists of six β-propeller domains, and the N-terminus consists of α and β domains linking to the OM. Pal consists of four-stranded β and α sheets connected to the C-terminus of TolB to form the TolB–Pal complex. However, the binding is not rigid and can be flexible to allow TolB to interact with other proteins and for Pal to bind to the PG layer (Santos *et al.*, 2015).

Finally, TolQ and TolR. TolQ is a protein occurring in the IM with three transmembrane helices, and TolR attaches to the IM by a single helix and extends into the periplasm (Egan, 2018). TolQ and TolR form a complex with TolA in the IM, consisting of transmembrane helices linking through to the periplasm, allowing energy transfer from PMF interaction (Zhang *et al*., 2011). Using mutation analysis and study of two-hybrid systems which detect PPIs, Teleha and colleagues (2013) observed how TolQ and TolR engage in the PMF activation of TolA. This might support the interaction of TolA with the TolB–Pal complex in the OM, facilitating colicin translocation.

1.4 Colicin A (ColA)

ColA is a water-soluble colicin in group A with a molecular weight of 63 kDa. It acts to form pores in the IM of *E. coli* (Dunkel *et al*., 2015). X-ray crystallography has revealed that ColA has 10 amphiphilic α-helices, with a hydrophobic region between helices 8 and 9. A study on ColA crystal structure has suggested that after ColA binds to the receptor on the OM of *E. coli* (i.e., BtuB), helices 8 & 9 penetrate the OM, followed by the hairpin of helices 5 and 6. However, it must duly be noted that due to the instability of pore formation, investigating that process with nuclear magnetic resonance (NMR) or X-ray crystallography has proven to be difficult (Alonso *et al.*, 2000).

ColA has been studied using site-directed spin-labelling electron paramagnetic resonance (EPR) in order to observe the membrane proteins' structural properties (Jagannathan *et al.*, 2010). Pulagam & Steinhoff (2013) employed this technique to gain insight into its kinetics and binding affinities as a means to characterise its binding to *E. coli* OM. Such binding is an important factor in ColA translocation due to the involvement of the hydrophobic region of the molecule.

The translocation involves several stages. First, the (R) domain binds to the BtuB receptor, after which ColA unfolds. Second, when ColA binds to the surface receptor, the (T) domain enters into close proximity with OmpF. Next, the BtuB–OmpF translocon transfer the (T) and (C) domains into the periplasm. Third, the unfolded (T) domain threads through one of the channels within OmpF and enters into the periplasm. Next, the (T) domain binds to TolB protein in the periplasm using the TolB box, where ColA interacts with TolB. Lastly, TolA binds to the ColA (T) domain, and this binding site is referred to as the TolA box.

Bouveret *et al.* (2002) proposed that assuming the exposure of the N-terminus occurs on the OM periplasmic side, the (T) domain interacts with not only TolB at the N-terminus but also with TolA upstream from that terminus. If this is correct, then the (T) domain simultaneously interacts with both TolB and TolA. Building upon this research, Penfold and colleagues (2012) proposed that, if a Brownian motion interaction occurs, during which proteins are capable of converting chemical energy into mechanical motion during ColA translocation, then the TolB box will bind to TolB, and its TolA-box will bind to TolA via a high-affinity interaction. In turn, that interaction allows the (C) domain to enter the periplasm to reach the IM, where it can exert its cytotoxic effect via pore formation

and, in time, cause cell death (Zakharov & Cramer, 2004). This process is represented in Figure 1.10.

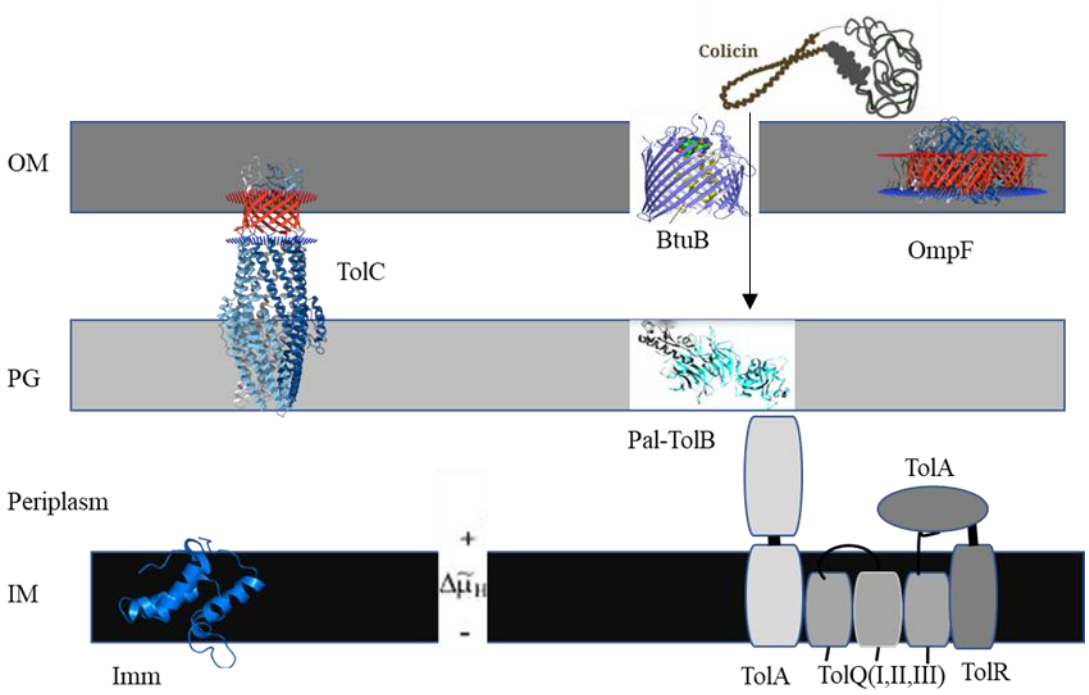


Figure 1.10 Diagrammatic representation of the components of the Tol–Pal cellular import pathway for the pore-forming colicin (e.g., ColA), including TolB, Pal, TolA, TolR and the immunity (Imm) protein in the inner membrane (IM) (adapted from Zakharov & Cramer, 2004). Pore-forming colicins are required to bind to BtuB and OmpF in the outer membrane (OM), noting that other types of colicins bind to TolC in the OM. By binding to BtuB and OmpF, ColA is subsequently able to bind to the Pal-TolB complex in the peptidoglycan (PG), Tol–ARQ complex in the IM, allowing for internalisation of ColA by providing energy ($^{\Delta\sim}\mu^{+}_{H}$) through the interaction of TolA with Pal-TolB complex in the periplasm. This will lead to pore formation and subsequent cell death.

Research suggests that an increase in the number of the pore-forming channels in the IM is directly linked to the presence of TolA and TolB aggregates in the cytoplasmic membrane and confirms their involvement in ColA translocation (Dubuisson *et al.*, 2002).

Research into the interaction between ColA, TolA, and TolB began with immune-blotting experiments that revealed ColA interaction with TolA and the formation of the TolB–Pal complex. Using surface plasmon resonance (SPR) studies on the binding affinity of ColA and TolA, it has been shown that ColA interacts with TolA (Lazdunski, 1998). Later research using mutational analysis focused on clarifying the involvement of the TolB–Pal complex in its interaction with ColA (Lazzaroni *et al.*, 2002). It has been shown that ColA binding to the surface receptors BtuB and OmpF facilitates the entry of the ColA (T) domain and its later interaction with TolB (Hands *et al.*, 2005).

Hecht and colleagues (2010) used *in vitro* NMR to confirm the direct interaction between ColA and both TolA and TolB in the (T) domain. The titration of an unlabelled ColA to a labelled TolA-III bound to unlabelled TolB has revealed the unfolding of TolA-III residue in the (C) domain, and the interaction of ColA binding to TolA and TolB separately to form a trimeric complex. They also reported the formation of the TolB–TolA-III complex. This research was important because, by the labelling of one protein, the level of background noise that accompanies NMR was reduced, and they were able to distinguish the protein of interest from other proteins. The authors suggested that the first unfolding of ColA occurs where the (R) domain binds to the OM receptor, followed by a second unfolding during the translocation where the N-terminus (i.e., (T) domain) interacts with TolA-III (i.e., C-terminus) (Hecht *et al.*, 2010). However, this contradicts with an earlier study which showed that ColA (T) domain formed two complexes (i.e., ColA–TolA–TolB, and ColA–

TolA–TolR) in different regions and that these interacted with TolA and TolB in different regions (Collins *et al*., 2002). Therefore, it must be stated that the entire process is not yet fully understood. Further research is needed for full elucidation of the mechanism.

It has also been demonstrated that in the absence of the Imm protein, ColA cannot reach the IM from the cytoplasm and, as a result, is not able to deliver a cytotoxic effect to colicin-producing cells (Fridd *et al*., 2002). In another study that same year, Duché (2002) investigated ColA interplay with the Imm protein to find that ColA (C) domain fuses with a single peptide on the cytoplasm to enter the cells and translocate through the periplasm until pore channels in the IM are formed. The colicin's cytotoxic effect was shown to be inhibited by the Imm protein, as confirmed by examining the structure of the ColA pore-forming domain (C) without the Imm protein. As a consequence, the colicins produced, upon being rendered unable to kill, remained in the cytoplasm of the producing cell due to the polarity of the transmembrane electrochemical potential. It must be noted that colicins that are external to the cells continued to be able to kill sensitive cells provided that the corresponding Imm proteins were absent (Duché, 2002).

Research on the ColA (C) domain has revealed that potassium ion efflux causes membrane depolarisation, ATP depletion, and, in turn, cell death. However, the pore-forming mechanism of ColA remains unclear. ColA pore-forming activity was first inferred from *in vitro* experiments and later from indirect *in vivo* observation. An interesting finding comes from Bénédetti and colleagues (1992), who showed that trypsin affected the efflux mechanism upon being added to ColA. Because trypsin cannot access the cytoplasmic membrane of bacterial cells, those authors suggested that after the translocation of ColA, part of it, notably its polypeptide chain, remains outside bacterial cells. This could explain

later the effect of trypsin on ColA outside cells as well as the closure of pores (Cascales *et al*., 2007). Upon being added to ColA, trypsin deactivated ColA and stopped the potassium efflux, likely due to its proteolytic activity. The extended effect of ColA unfolding suggests that part of ColA remained in contact with the external medium. However, another possibility is that ColA interaction with BtuB and OmpF caused it to unfold. Guihard and colleagues (1994) also analysed the kinetics of the efflux of cytoplasmic potassium induced by ColA in whole cells. Results indicated that the toxin (i.e. (C) domain) formed voltage-dependent channels in the IM. Although the mechanism of pore formation and the structure of the voltage-gated channels need clarification, later, Pulagam and Steinhoff (2013) examined ColA pore-forming function by measuring the lag time (i.e., the time required for a colicin's translocation through the cell envelope before the efflux of cytoplasmic potassium). The authors determined that the addition of urea to the system, which led to the denaturing of ColA, resulted in the lag time dropping from 30 to 10 s. However, after renaturation, the lag time returned to its normal value. Those findings indicate that ColA unfolded upon entry, which facilitated its translocation through the membrane system. However, it remains uncertain whether the entire ColA molecule is translocated into cells (Chang, 2018)., and further research is clearly needed for a full understanding of the interactions here.

1.5 Nuclease colicin (colicin E9)

Colicin E9 (ColE9) is a group A (nuclease) colicin. The (R) domain binds to the BtuB receptor in the bacterial OM, at which point the (T) domain crosses the membrane, facilitated by OmpF, and binds to TolB, thereby forming the translocon ColE9–BtuB–OmpF–TolB (Klein *et al.*, 2016). ColE9 translocation gains energy through association with TolA in the IM, as previously described. This causes the release of the Imm protein from the (C) domain of ColE9–Imm9 complex. Subsequently, TolB binds to TolA, which allows the (C) domain to enter the IM, where it can exert its cytotoxic activity by digesting the bacterial chromosomal DNA (Figure 1.11).

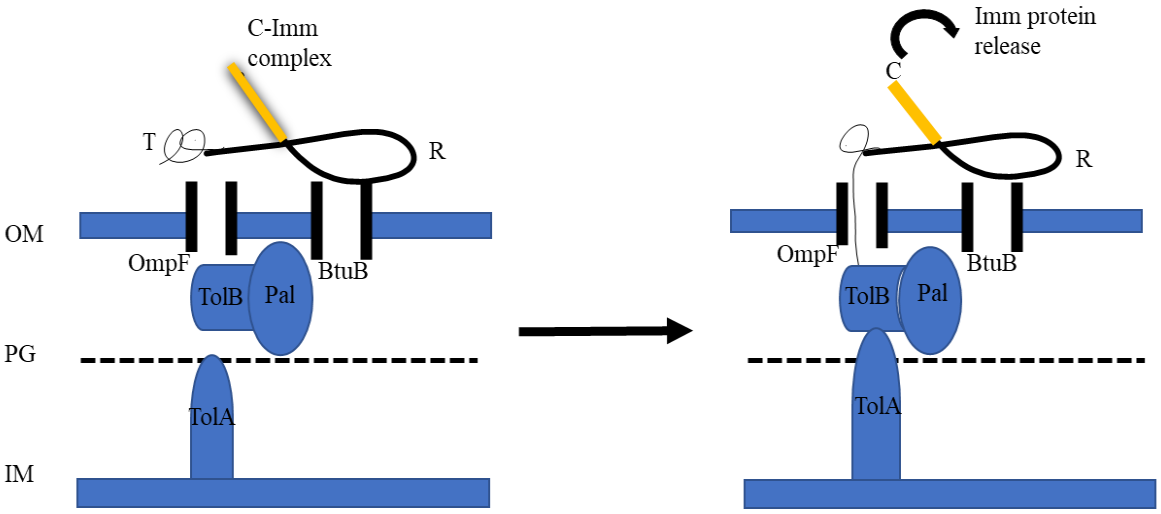


Figure 1.11 The mechanism of translocation for nuclease colicin ColE9 (adapted from Farrance *et al.*, 2013). The translocation of ColE9 begins when the receptor binding (R) domain binds to the BtuB receptor on the outer membrane (OM). Next, the (R) domain binds to OmpF to allow subsequent binding between the (T) domain and TolB-Pal complex; TolA interacts with the TolB–Pal complex to release the immunity (Imm) protein from the complex.

It has been demonstrated through the isolation of stable ColE9–BtuB complex *in vitro* using gel filtration and chemical cross-linking that OmpF association with the BtuB–ColE9 complex, although necessary, is weak or transient, if not both (Law *et al.*, 2003). Further research has shown that the BtuB–ColE9–Imm9 complex recruits OmpF using an intrinsically disordered 83 amino acid N-terminal sequence (Collins *et al.*, 2002; Housden *et al.*, 2005). Additionally, Penfold and colleagues (2012) added that the interaction between ColE9 and TolB occurs after their interaction with OmpF, and importantly that no direct interaction with TolA occurs. More recently, TolA has been shown to function with ColE9 in releasing energy and removing Imm9 by interacting with TolB (Atanaskovic & Kleanthous, 2019).

To date, *in vitro,* NMR studies have involved examining ColE9 (R) and (T) domains, particularly to observe and confirm ColE9 interaction with TolB and the stability of that interaction. *In vitro* NMR studies indicate conformational changes occur resulting from TolB interaction with the ColE9 (T) domain. Furthermore, the unstructured sequence of the (T) domain suggests that other interactions are likely to be occurring with other Tol proteins. The interaction between ColE9 and TolB was additionally analysed after alanine mutations were identified in the TolB box that removed the biological activity of mutant ColE9 (James *et al*., 2002). The mutations were determined to be linked to three essential residues, D35, S37, and W39, of the TolB box pentapeptide sequence in the N-terminal (T) domain of nuclease E colicins. Through additional site-directed mutagenic research, this technique has determined that the TolB box sequence in ColE9 is larger than a pentapeptide and extends from residues 34 to 46 (Hands *et al*., 2005).

NMR has been conducted to investigate the interaction of ColE9 with Imm9. Research indicated that no changes in the chemical shifts had occurred. However, the resonance of the interaction region was affected, which indicates the presence of an interaction. Data suggests that Tol proteins, such as TolA, may indirectly facilitate translocation by interacting with TolB (Mosbahi *et al*., 2004). The interaction between ColE9 and TolB has also been studied using the two-hybrid system. Moreover, the data confirmed that their interaction is pivotal to ColE9 cytotoxic activity (James *et al*., 2002). However, a greater level of research in order to fully elucidate the mechanisms involved in these molecular systems is needed.

1.6 Live-cell imaging of proteins

Efforts to provide more stable conditions to study bacterial cells and monitor their growth have involved numerous modifications to cell-imaging systems in order to identify techniques that can furnish information about cell biology beyond what optical microscopy provides (Zucker, 2014). One such technique, live-cell imaging, is now quite common across research fields involving the study of cellular dynamics and cell function. In particular, live-cell imaging can be used in co-localisation studies, when two different fluorescent probes are employed to define, for example, intracellular compartments and to compare their functions, which provides information about the location of proteins within cells (Hoppe *et al.*, 2009).

The most common technique for live-cell imaging is light microscopy. In light microscopy, entire specimens are uniformly illuminated. However, this can result in out-of-focus and low-resolution imaging due to the three-dimension (3D) nature of biological samples, which usually range from 0.4-0.7 µm upwards in size. Therefore, conventional light microscopes are not suitable for *in vivo* studies of bacteria. Electron microscopes allow samples to be viewed in much higher resolution and in greater magnification. However, that technique also has limitations, including inapplicability on non-dehydrated samples (Singh & Gopinathan, 1998), meaning that living samples cannot be visualised.

A variation on light microscopy is fluorescence imaging. This involves using fluorescent labels in conjunction with light microscopy and selecting different excitation wavelengths, including those in the ultraviolet (UV) and infrared (IR) range, to visualise the different labels (Hoppe *et al.*, 2009). Although an enduring problem with using cell-imaging techniques has been long exposure times that can affect the viability of cells, the development of bright fluorescent dyes has allowed for a reduction in the time of exposure. The most common fluorescence imaging technique used to investigate PPIs entails using green fluorescent protein (GFP) as a visual marker (Rosochacki & Matejczyk, 2002). GFP is widely used because its folding properties facilitating the detection of PPIs via biomolecular fluorescence complementation (BiFC). BiFC works by dividing GFP into two fragments, and only in the case of PPI, the two fragments become fluorescent (Meyer & Dworkin, 2007). Another fluorescence imaging technique applied to study PPIs is fluorescence resonance energy transfer (FRET). When using FRET, a donor molecule needs to be excited by a photon of a specific wavelength, which transfers energy to other molecules by a non-radiative mechanism (Xing *et al*., 2016).

Another way of investigating PPIs is by studying the movement of proteins via fluorescence recovery, which involves photobleaching proteins and measuring the time they need to recover after the bleach diffuses from photobleached areas. In using photobleaching to detect PPI, pulses generated from a laser on the bleaching wavelength are similar to the fluorescence excitation; several images are produced in the same focal plane, which shows the dynamics of the change in fluorescence within the bleached region (Meyer & Dworkin, 2007). However, when fluorescent microscopy is used, phototoxicity may occur during the excitation of fluorescent dyes, partly due to the occurrence of the

oxygen-dependent reactions of free radicals (Frigault *et al.*, 2009), causing cellular damage to the living sample leading to death (Icha *et al*., 2017).

Among the most advanced live-cell imaging techniques, confocal laser scanning microscopy (CLSM) affords an advantage over other microscopic techniques because it precludes out-of-focus blurring and scattered light. In CLSM, light is collected through a pinhole detector that moves to scan the whole sample, such that the pinhole's small diameter decreases the glare from the sample, typically reduced in size to a thin optical section. Because the sample is illuminated with a laser, one caveat of using CLSM is to determine the laser intensity and pinhole size in advance in order to control the background fluorescence and photobleaching (Nwaneshiudu *et al.*, 2012). It should be noted that CLSM can provide clear images of living and fixed cells because the resolution of confocal microscopes exceeds that of light microscopes, which may reach resolutions of 180 nm laterally and 500 nm axially (Fouquet *et al.*, 2015). This resolution power allows the technique to provide more detailed information about cellular structures and cellular localisation of macromolecules such as proteins, RNA, and DNA (Singh & Gopinathan, 1998).

The co-localisation analysis of two fluorescently labelled proteins, easily performed by CLSM, can be used to visualise bacterial cells (Chen & Periasamy, 2007) and locate proteins of interest. CLSM has also proven its ability to track single molecules (Oikawa *et al.*, 2012) and is currently considered to be a good observation technique for tracing specific cells and providing 3D images (Nikolić *et al*., 2018). CLSM provides results that can be easily mapped and quantified as well (Wu *et al*., 2010). It can also be used to monitor fluorescently stained bacterial cells grown within mixed populations, in which case it

allows detailed microscopic studies of the differential sensitivity to antimicrobial agents within mixed populations (Pamp *et al.*, 2009).

When stimulated, dyes (also referred to as fluorophores) used in CLSM can re-emit light upon light excitation, more than one of which can be studied by monitoring the excitation light at different wavelengths. To begin, an excitation light is provided by a laser at a wavelength that excites and stimulates the fluorophore. When the laser reaches the target, it produces extremely intense fluorescence at a specific focal point; the excitation beam and fluorescence emission pass through a dichroic mirror that reflects the incoming laser light with a shorter wavelength and higher energy, and the higher wavelength with lower-energy light passes through to the light detector (Furia *et al*., 2014). Partly for that reason, the appropriate dye to use depends on the target molecule, and dyes should have sufficient sensitivity and selectivity to emit a sharp signal. Also, the concentration of fluorescent dyes should be minimised to reduce any stress to living cells and to cause phototoxicity (Logg *et al*., 2009).

Processing the images produced from CLSM requires a powerful computational facility with image-processing procedures and specific hardware for stereoscopic image display. CLSM can produce images of sections of the sample. In this way, the computer can generate a series of sectional images that can be stored and created as 3D images (Nikolić *et al*., 2018). However, a major limitation of CLSM is its relatively slow image acquisition, which requires multiple scans to ensure high-quality images, and the minimal time per each image's acquisition should be at least 1 min (Paddock & Eliceiri, 2014).

1.7 Nuclear magnetic resonance (NMR)

A novel and challenging method of studying PPIs is NMR. NMR allows for the visualisation of the structure, dynamics, and interactions of biological macromolecules. Using NMR, researchers have been able to demonstrate the different molecular structures and associated chemical shifts for specific biological molecules and identify them (Sugiki *et al*., 2017), as well as demonstrating the effect of solvents on sample preparation and how aromatic as well as aliphatic atoms are represented in the NMR spectrum. Such findings marked the beginning of one-dimensional (1D) NMR, which was used to visualise specific atoms, including ^{1}H, ^{13}C, and ^{15}N, within a chemical complex and then developed to study biological samples (Takis *et al.,* 2020).

Taking advantage of the magnetic properties of specific atomic nuclei and the magnetic fields around the atoms, NMR involves changing the resonance frequency of the target atoms (Sugiki *et al*., 2017). Briefly put, NMR targets atomic nuclei with odd mass numbers, such as, ^{13}C, ^{15}N, and ^{1}H (Poulsen, 2002), which, produces a magnetic field by rotating around a given axis. When a molecule is exposed to an external magnetic field, the electrons within the molecule produce a local current that opposes the external magnetic field. That dynamic condition reduces the total effective magnetic field and generates a chemical shift, and any shielding from the external magnetic effect depends upon the chemical structure of the tested sample (Poulsen, 2002). An example NMR output is presented in Figure 1.12.

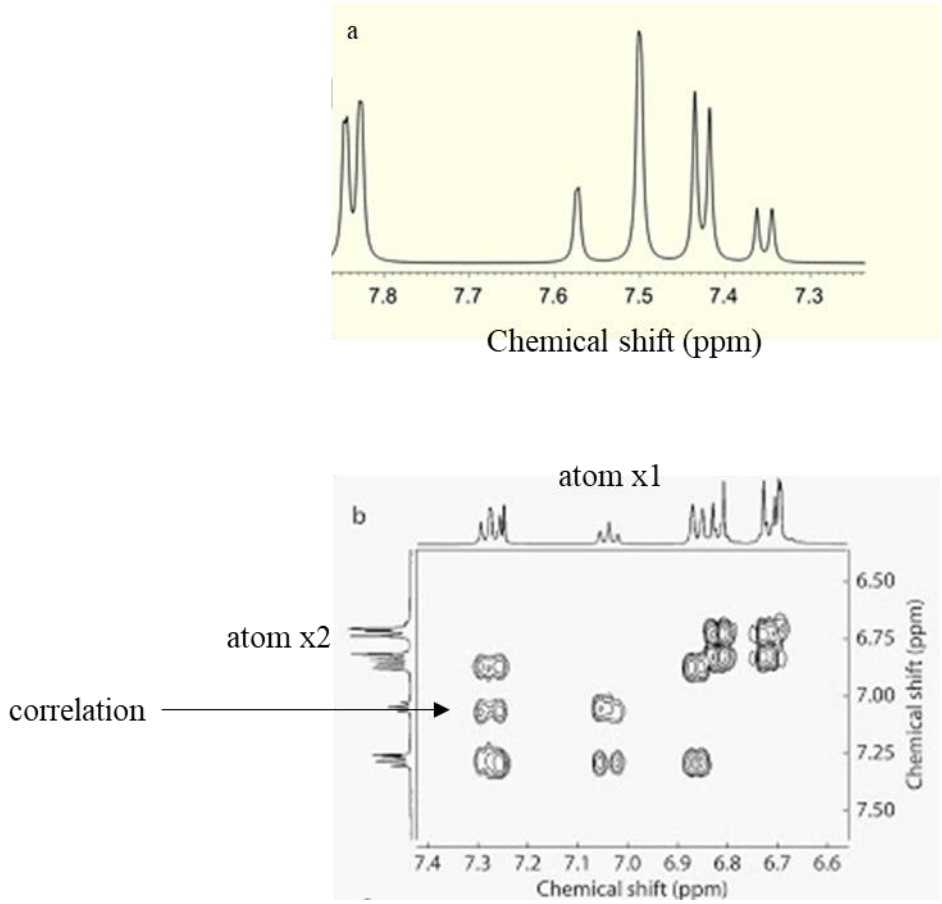


Figure 1.12 1D and 2D NMR spectra (adapted from Wüthrich,2003; WeiJin *et al.*, 2008). a) The 1D spectrum shows signals of one H atom that represents each peak. b) the 2D spectrum shows the correlation between two different H atoms x1 and x2.

Improvements have been to the basic NMR technology, and it is now possible to undertake two-dimensional (2D) NMR. In 2D experiments, emissions centre on a single frequency in order for correlated resonances to be observed. It is possible to identify the adjacent functional groups, and, in turn, the identification of associated resonances would be easier. An example of a 2D NMR experiment includes heteronuclear correlation experiments, in which emissions are centred on the peak of an individual nucleus. If the magnetic field correlates with another nucleus, then the method is referred to as homonuclear correlation

spectroscopy (COSY) sequencing. This can indicate which signals arise from neighbouring atoms by up to four bonds (Anaraki *et al*., 2019). By contrast, if the atoms are binding through intramolecular space, which means direct through-space interactions between magnetic dipoles, then the method used is nuclear overhauser effect (NOE) coupling, in which responses can be detected on the frequency of the correlated nucleus. Therefore, it can be shown that 2D NMR spectra provide more information about molecules than 1D NMR spectra especially in studying the structure of a molecule (Johnson *et al.*,2013). The most common 2D experiment for protein analysis is $^{1}H–^{15}N$ correlation heteronuclear single-quantum coherence (HSQC). This technique has allowed a host of protein structures and interactions to be identified.

1.7.1 In-cell NMR

With regards to the structure and function of the intracellular environment of living organisms, in-cell NMR is considered to be a novel tool for visualising biomolecules and monitoring their activity in living cells. This is an extremely complex situation as the potential for the disruption of normal physiological activity is high. At present, the technique still needs additional studies to fully develop the approach for prokaryotic and eukaryotic cells (Ito & Selenko, 2010). Concerning biomolecules, in-cell NMR has been developed out of the related technique, *in vivo* NMR, which is used in the analysis of small metabolites such as ATP. This has led to the development and successful application of in-cell NMR, which is also capable of analysing large macromolecules, including proteins, in addition to small molecules (Tochio, 2012). The chief difference between in-cell NMR and *in vivo* NMR concerns how the signals of the target molecule are differentiated from the

background signals, and this makes in-cell NMR requires a more complex labelling technique than the one required by *in vivo* NMR (Serber *et al.*, 2005). Over time, the use of in-cell NMR has expanded to include studies investigating protein–DNA interactions, the dynamics inside folded and unfolded proteins, and protein stability (Ito & Selenko, 2010). It must be noted that in-cell NMR has been developed to study also mammalian cells, for example, HeLa cells (Bekei *et al*, 2012), and it has been successfully used in drug-screening assays to observe delivery inside living cells and the binding affinity of the drug to its target (Burz & Shekhtman, 2009). To date, diverse studies on the applicability of in-cell NMR in a variety of cells have involved determining the structure of the TTHA1718 protein in prokaryotic cells. This study involved extraction of the putative metal-binding protein TTHA1718 from *Thermus thermophilus,* which was subsequently overexpressed in *E. coli* to be analysed by in-cell NMR (Sakakibara *et al.*, 2009; Ikeya *et al.*, 2019). Another successful in-cell NMR application was examining protein dynamics and PPIs with calmodulin (Serber *et al.*, 2004). Calmodulin (CaM) is a ubiquitous calcium-sensing protein among eukaryotes. Over the years, CaM was used as a tool to study calcium signalling in research. In-cell NMR analysis of calmodulin has been achieved by labelling its methyl group, and they were able to identify its structure (Grant & Marshall, 2019). In other research, Calmodulin signalling pathways have been also studied by in-cell NMR in how CaM interacts with and activates its targets. Among those targets, CaM has been shown to be an essential component of a calcium-sensing regulatory apparatus for several voltage-gated ion channels. In-cell NMR was used for structural characterization of CaM-peptide complexes, in particular for the study of IQ motifs, which binds CaM at that basal level of calcium in cells and thereby serve to localize CaM to its sites of action. The

application of a titration method of CaM with IQ motif peptides to determine the stoichiometry of the complex and to identify the residues at the binding interface was their approach (Damo *et al*.,2013). The resulted changes in amide chemical shifts were observed in the ^{15}N-^{1}H HSQC spectrum of ^{15}N-labeled CaM to dissect the stoichiometry of peptide binding, as well as to identify residues perturbed upon peptide binding as there were involved in the signalling pathway. The achievements were valuable for understanding the transduction of calcium signals by CaM and the activation of the intracellular signalling pathways (Damo *et al.*, 2013). The list of successful in-cell NMR applications extends to analysing protein folding, such as, human copper, zinc superoxide dismutase (SOD1) (Banci *et al.*, 2011). This protein is involved in the cadmium toxicity mechanism, and aggregation of the protein causes neurodegenerative diseases such as lateral sclerosis in human cells (Polykretis *et al*., 2019). SOD1 is an evolutionarily conserved antioxidant enzyme that is present in most tissues at relatively high concentration, particularly in neuronal cells, and is localized in the cytoplasm, nucleus, and mitochondria. In order to reach the enzymatically active form, SOD1 needs to dimerize, bind one zinc ion and one copper ion per monomer, and form an intramolecular disulphide bond. Analysis of the cell lysates by in-cell NMR revealed that SOD1 could not bind zinc even when zinc is supplemented after cell lysis forming irreversible aggregates which causing the disease, indicating that the misfolding state of the cells occurs before metal binding (Luchinat & Banci., 2018). In another study, SOD1 folding properties in living cells were investigated using in-cell NMR. The study confirmed the interaction between SOD1 and ebselen, which is an organo-selenium compound with broad antioxidant properties. In-cell NMR of

oxidation of SOD1 in living cells by ebselen showed how is the interaction affected its folding in the living cells (Grant & Marshall, 2019).

Although the use of in-cell NMR remains relatively uncommon, the method has proven its value by way of its ability to reveal the activity of proteins inside cells. However, it must be noted that this technique requires meticulous optimisation. Such modern in-cell NMR techniques have opened a new field of analysis of macromolecules, and understanding protein's mechanism of action with in-cell NMR has supported drug development and protein engineering (Burz & Shekhtman, 2012). Recently, in-cell NMR was used to study antituberculosis imidazopyridine amide (IPA) in living cells and observed drug binding to the cytochrome b in living cells (Kang, 2019). The study showed the engagement of an IPA anti-TB drug with its bacterial target, which is cytochrome b. Furthermore, in-cell NMR was used to characterize the hit−target binding affinity measurement mode by taking into account the dynamics of the association. Such interaction information can be very valuable for drug discovery programs (Bouvier *et al*., 2019).

1.7.2 The use of isotope labelling and protein overexpression in conjunction with in-cell NMR

The introduction of isotopes labelling can facilitate the selective visualisation of proteins of interest (Verardi *et al*., 2012). Target labelling not only aids the NMR-supported detection of macromolecules but also reduces background signals and thus provides more apparent results (Maldonado *et al*., 2011). However, the challenge with using the technique, is that efficiently producing isotopically labelled proteins inside cells can be achieved by two methods. Firstly, expressing the proteins in a labelled medium, where the

cells are able to uptake the isotopes and produce labelled proteins through normal metabolic functions. The second method is to inject labelled proteins directly into cells. However, microinjection is only applicable to large proteins (Burz & Shekhtman, 2009).

The often-time-consuming process of preparing samples for in-cell NMR begins with the overexpression of target protein within the cell system under study. Such overexpression is pivotal because it produces a high yield of proteins within the cellular environment without protein leakage (Maldonado *et al*., 2011). Many methods have been employed to achieve a high yield of recombinant proteins inside bacterial cells, particularly in *E. coli* (Mondal *et al*., 2013). The use of an inducible plasmid in bacterial cells has proven to be an effective method of producing isotopically labelled targets during protein overexpression. For example, bacterial T7 expression systems are capable of producing a high yield of proteins, thereby generating isotopically labelled targets rapidly (Freedberg & Selenko, 2014). Beyond that, using selective plasmids with specific induction time helps to control the high concentration of overexpressed proteins (Maldonado *et al*., 2011). During experiments, the induction time and temperature of the incubation of overexpressed proteins should be monitored to ensure cell viability as well. As an example of monitoring bacterial cell viability, Serber and colleagues (2005) managed to monitor the overexpression process of NmerA protein in *E. coli* cells using in-cell NMR. They obtained data every 2 h and determined that when protein signals started to form, they determined that the cells remained viable for 6 h, and this was one of the successful applications of in-cell NMR.

1.7.2.1 ^{15}N labelling

The majority of studies on efficient isotope labelling have used *E. coli* as a model organism due to the high level of protein overexpression obtainable from those cells (Rosano & Ceccarelli, 2014). Proteins may be isotopically labelled with ammonium chloride $^{15}NH_4Cl$, which incorporates the ^{15}N isotope nuclei into the protein's peptide backbone, which reduces the background signal from other proteins in bacterial cells (Maldonado *et al.*, 2011). A successful application of in-cell NMR with ^{15}N labelling, its applicability with *E. coli*, *Xenopus* oocytes, and HeLa host cells and, in such research, proteins have been isotopically labelled and identified, and 3D structures have been determined, as indicated by $^{15}N–^{1}H$ correlations (Burz & Shekhtman, 2012). However, it should be noted that using ^{15}N labelling in this manner ensures that every nitrogen molecule is labelled in the in-cell NMR sample, which may complicate investigating the protein of interest. To overcome that problem, some authors, such as the research of Burz & Shekhtman, 2012, have suggested that when using the overexpression technique, bacterial cells should be grown in an ordinary medium and then changed into a labelled medium only directly before inducing protein overexpression, which will predominantly direct labelling only to the induced protein of interest.

Another method of labelling the protein's backbone is to label specific amino acids that are incorporated in the newly synthesized protein, such as arginine, histidine, and lysine, which significantly reduces the number of residues in the overexpressed protein. However, such isotope labelling causes lower resolution because only labelled residues are visible in the in-cell NMR spectrum (Maldonado *et al.*, 2011). Also, isotope labelling them with ^{15}N may cause peaks to overlap due to large resonance line widths inherent in ^{15}N labelling, labelling

a specific type of amino acid with only ^{15}N has been suggested. Even so, that technique depends on many factors, including the role and the location of the targeted amino acid and the effect of isotope labelling on other amino acids in the in-cell NMR samples.

Further research piloted another protocol to reduce the background of other labelled proteins by adding an antibiotic (e.g., rifampicin) that inhibits protein biosynthesis by affecting bacterial RNA polymerase but not bacteriophage T7. The results indicated that because protein overexpression is under the T7 promoter, the production of proteins by other mechanisms will be suppressed leading, to reducing the background of the labelled molecule (Serber & Dötsch, 2001).

1.7.2.2 ^{13}C labelling

Another labelling protocol involves using ^{13}C as a source of carbon to label atomic nuclei in overexpressed proteins. However, ^{13}C labelling produces a high background of metabolites owing to the normal metabolism of the cell producing labelled products. To overcome that problem, ^{13}C labelling the amino acid side-chain residues, especially those in the methyl and methylene groups, reduces the background and clarifies the signals, as commonly done with large molecules by labelling the methyl groups (Buescher *et al*., 2015). In general, the application of ^{13}C labelling is limited to small molecules, because of the metabolic by-products. However, in the case of larger ones, the selective labelling of methyl groups in overexpressed proteins is required. A study involving the isotope labelling of methyl groups in proteins achieved a higher sensitivity with in-cell NMR than any other group, including the amide group. Such high sensitivity was achieved because the methyl group contains three ^{1}H protons coupled to ^{13}C, whereas the amide group

contains three bonds attached to the carboxyl or methyl group. Another advantage of methyl group labelling is that methyl protons do not undergo chemical reactions with water, which can cause the loss of signal, as in the case with the amide proton (Serber *et al*., 2006).

1.7.2.3 ^{19}F labelling

Fluorine (^{19}F) labelling has been found to be particularly effective with globular or disordered proteins. ^{15}N labelling of globular proteins tends to result in low signal intensity, primarily due to the high viscosity and weak interactions that present within the cytoplasm. Similar to ^{1}H labelling, ^{19}F labelling results in high sensitivity to the NMR with a range of chemical shifts more prominent than the range in protons. As natural proteins do not contain fluorine, labelling the amino acids with ^{19}F may result in low background signals during protein overexpression (Didenko *et al.,* 2013). Additionally, ^{19}F labelling is highly sensitive in that it does not require a relatively high concentration of proteins due to the short time needed for experiments. Further, it facilitates straightforward detection because the ^{19}F labelling of globular proteins normally produces sharp, well-distributed peaks. ^{19}F labelling is exceptionally useful in the site-selective labelling of fluorine-containing amino acids such as tryptophan and phenylalanine (Freedberg & Selenko, 2014).

1.7.3 Advantages of in-cell NMR

NMR is a non-invasive method applicable in both *in vitro* and *in vivo* studies. In the case of in-cell NMR, capturing the 2D and 3D structure of proteins would allow the study of proteins and how they function inside cells (Burz & Shekhtman, 2009). *In vitro* NMR has been applied to distinguish details of protein structure due to its well-resolved resonance, particularly when adding isotopic labelled molecules to be NMR-visible, which either disperse within the cellular membrane before their purification or are actively transported inside the cells. To date, the analysis of protein dynamics and PPIs *in vivo* has been performed successfully with a predominantly bacterial system, where it has been used as a novel technique that shows NMR-active nuclei within the 3D structure (Pastore, 2017).

In studies on PPIs, in-cell NMR potentially allows a researcher to determine the precise nature of the structural interactions through observation of the binding of the interacting proteins in different forms: one isotope-labelled, the other not. As a result, the method can capture the ability of molecules to adjust to the binding behaviour of a known interaction between two proteins (Ito & Selenko, 2010). Ultrafast methods for 2D and 3D NMR have been developed to support cell viability, which have reduced the time required for the 2D experiments to mere seconds and 3D experiments to mere hours. The recent ultrafast NMR methods are able to deliver 2D NMR spectra involving any kind of homo- or heteronuclear correlation in a single scan (Giraudeau & Frydman, 2014). Also, the ability of in-cell NMR analysis using the ultra-fast method has been extended to visualise folded and unfolded proteins because the distribution of signals over the spectrum is capable of identifying the difference between them (Poulsen, 2002).

One more advantage of in-cell NMR is that it not only allows visualisation of the location and dynamics of proteins and macromolecules but also shows changes in their structures and any interaction(s) with other cellular components by measuring chemical shifts and observing the position of resonance lines. A final set of advantages of in-cell NMR includes that each signal identified represents an atom, its chemical shift resolves its location, and the well-resolved spectra are always shown as narrow lines, which facilitates assigning each resonance, makes it easier for interpretation (Marion, 2013).

1.7.4 Disadvantages of in-cell NMR

As previously stated, a disadvantage of in-cell NMR is the high concentration of proteins required for the procedure. In-cell NMR experiments require cells to survive in generally unfavourable conditions in NMR tubes with no nutrients or oxygen. Needless to say, prolonged exposure to those conditions will undoubtedly affect the cells. To address that problem, in-cell NMR experiments need to be conducted rapidly using only a few scans, which is applicable only in high concentrations of proteins. At the same time, the overexpression of the cells should be controlled, which could cause unwanted effects, including that such toxicity might induce non-specific binding and protein leakage (Ito & Selenko, 2010). Beyond that, producing high concentrations of proteins may prompt protein aggregation, the formation of inclusion bodies, and high viscosity, all of which may make the sample difficult to analyse in in-cell NMR. At the same time, low levels of overexpression and long induction times can result in a high isotopic labelled background in the spectra and reduce cell viability during that analysis (Freedberg & Selenko, 2014).

Another disadvantage of in-cell NMR, as mentioned earlier, is that studying proteins under physiological conditions entails some limitations, including overcrowding in the native environment. This can cause non-specific interactions that obstruct detection of the protein of interest (Pielak *et al*., 2009).

As a consequence of those potential disadvantages, Robinson and colleagues (2012) have suggested two preliminary experiments need to be undertaken during in-cell NMR analysis as important quality controls. The first is an analysis of 'leakage' signals in the culture supernatants from the cells under study. The second is a viability test to assess the status of the cells after the in-cell NMR measurement has been completed (Li & Liu, 2013). Taken together, these two tests can be used to confirm whether the signals obtained by in-cell NMR are truly intracellular, as described in chapter 6.

1.8 Optimising conditions for in-cell NMR sampling

It is difficult to determine the ideal conditions for in-cell NMR studies and which environment the protein is best considered to be in its physiological optimum. With reference to Burz and Shekhtman's (2012) division of NMR techniques into three stages—sample preparation, analysis, and data processing. Several factors should be considered when preparing and running samples for in-cell NMR. First, the growth temperature of bacteria should be at 37 °C due to the cytosol's composition, and its solute concentration, and ionic strength because changes in these conditions might lead to peak variation, and in turn, the protein detection would be difficult. Second, the cell density of the sample is important because a very high concentration of cells causes deterioration of the proteins;

thus, the cells should form approximately 20% of the sample's volume. Third, protein degradation should be monitored, which is why the experiment should be run after cell collection. Finally, the cell viability should be monitored before and after the experiments using a colony plating test (Sharaf *et al*., 2010).

1.9 Aim of the thesis

Colicin could be a powerful, fast-acting alternative to antibiotics because of its cytotoxic activity mediated by cellular components disruptions, which are similar to most of the classic antibiotics (Behrens *et al*., 2017). The flexibility of different engineering types of colicins opens up the field to discover novel antibiotics to control bacterial infections (Jin *et al*., 2018). With the appreciation of the complexity of the intracellular environment when studying these colicins and how they are produced by bacteria as a general property of every living organism to proceed in this field, studying the details of colicin is required ((Bryant *et al*., 2005).

The gap of knowledge in colicin entry is that many questions remain unanswered about colicin import (Kleanthous, 2010). In order to fully understand the translocation process of colicin into the periplasm of *E. coli* cells, PPIs have to be investigated using CLSM and in-cell NMR. Through analysing the effect of ColE9 (nuclease colicin) on *E. coli* cells expressing the (T) domain of ColA. Is treating those cells with ColE9 result in protein-protein interactions?

This thesis aims to further the understanding of colicin internalisation within *E. coli* in order to elucidate the fundamental mechanisms that enable these molecules to deliver their cytotoxic effect. In order to achieve this aim, the following objectives will be addressed. The first objective is to clone the relevant genes and gene fragments (TA, TolA box, and TolB box) into appropriate vectors (e.g., pEt-15b and pBAD/gIIIc) to allow for both cytoplasmic or periplasmic expression. Secondly, to optimise overexpression and efficient labelling. Third, CLSM will be used to investigate the effect of colicin E9 on the viability of *E. coli* cells expressing TA, TolA box, and TolB box. Finally, in-cell NMR will be utilised to detect TA, TolA box, and TolB box domains within the cells. This will be a combination of experiments to optimise the condition suitable for in-cell NMR detection.

Chapter 2
Materials and methods

2.1 Materials and methods

2.1.1 Chemicals and Reagents

All chemicals and bacteriological media used were purchased from Sigma-Aldrich UK/USA, Oxoid (UK) or Thermo Fisher Scientific (USA) and Invitrogen unless stated otherwise. Custom oligonucleotides were synthesised by Eurofins Genomics (UK).

2.1.2 Equipment

Living cells images were acquired using the Leica Application Suite Advanced Fluorescence software 5 (Leica, UK), with an optical magnification of 40× using an oil-immersion objective. Images were sized to 1024 × 1024 pixels and recorded by scanning lasers over an area of 50 × 50 µm. An argon-based laser was employed for excitation at 488 nm, and an HeNe laser for excitation at 543 nm. The emission was set at 528 nm for SYTO 9 and 645 nm for propidium iodide (PI). Via sequential scanning, the images were obtained and processed in Image J software for quantification of bacterial cells.

The NMR research was initially conducted in the NMR facility at King Saud University in Riyadh, Saudi Arabia, NMR spectrometer -IE a Bruker Avance III (700 MHz) fitted with 5mm triple resonance TCI cryoprobe, after which it continued at the Manchester Institute of Biotechnology in Manchester, United Kingdom, using a Bruker Avance III (850 MHz) also fitted with a 5mm triple resonance TCI cryoprobe. All NMR spectra were obtained at 37 ℃ The spectra were acquired and processed using Bruker Topspin software v4.7.

2.1.3 Bacterial strains, growth conditions and media

Bacterial strains used are indicated in Table 2.1. *E. coli* DH5α was used as the host strain for plasmid stocks, whereas *E. coli* BL21(DE3) served as the host strain for plasmid expression. Bacterial cultures were grown at 37 ℃ overnight in Lysogeny broth (LB) medium supplemented with 100 µg/mL of ampicillin and stored as glycerol stocks at -80°C for further use.

Bacterial strain	Description	Source
***E. coli* DH5α**	F$^-$ *endA1 glnV44 thi1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ$^-$	Invitrogen (USA)
***E. coli* BL21 (DE3)**	F$^-$ *ompT gal dcm lon hsdS$_B$*($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	New England Biolab (UK)

Table 2.1. Bacterial strains used and their sources. Genotype description listed in the Appendix A.

2.1.4 Cloning and expression vectors

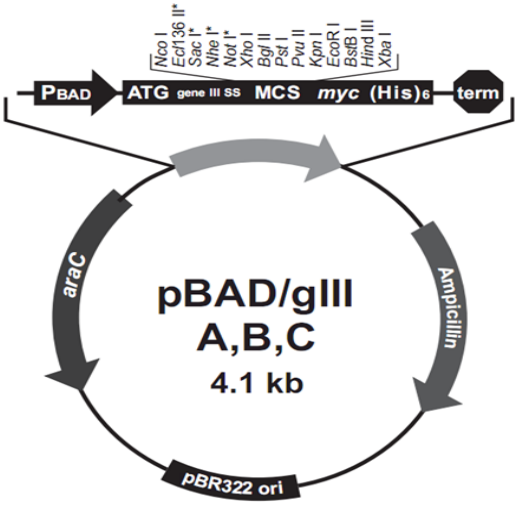
Two vectors were used for construction of recombinant clones (Table 2.2), both were cultured on LB agar supplemented with 100 µg/mL ampicillin. One is pBAD/gIIIc (4.1 Kb). Figure 2.1 shows the sequence map's multiple cloning sites for pBAD/gIIIc indicated. The other vector used is pET-15b (Novagen, USA) (5.7 Kb) its sequence map of its

multiple cloning sites is shown in Figure 2.2 similarly indicated. The restriction sites used in the cloning of the constructs is listed in Table 2.3. Recombinant pColA and pColE9 were a kind gift from Dr Chris Penfold (University of Nottingham) (see Appendix A).

Construct name	Description	Cloning vector
pTA/pTA1	TA (full) amino acid residues 1-172	pBAD/gIIIc / pET-15b
pTolA box	TA amino acid residues 52-172	pBAD/gIIIc
pTolB box	TA amino acid residues 1-52	pBAD/gIIIc

Table 2.2 Construct names and their descriptions.

A



B

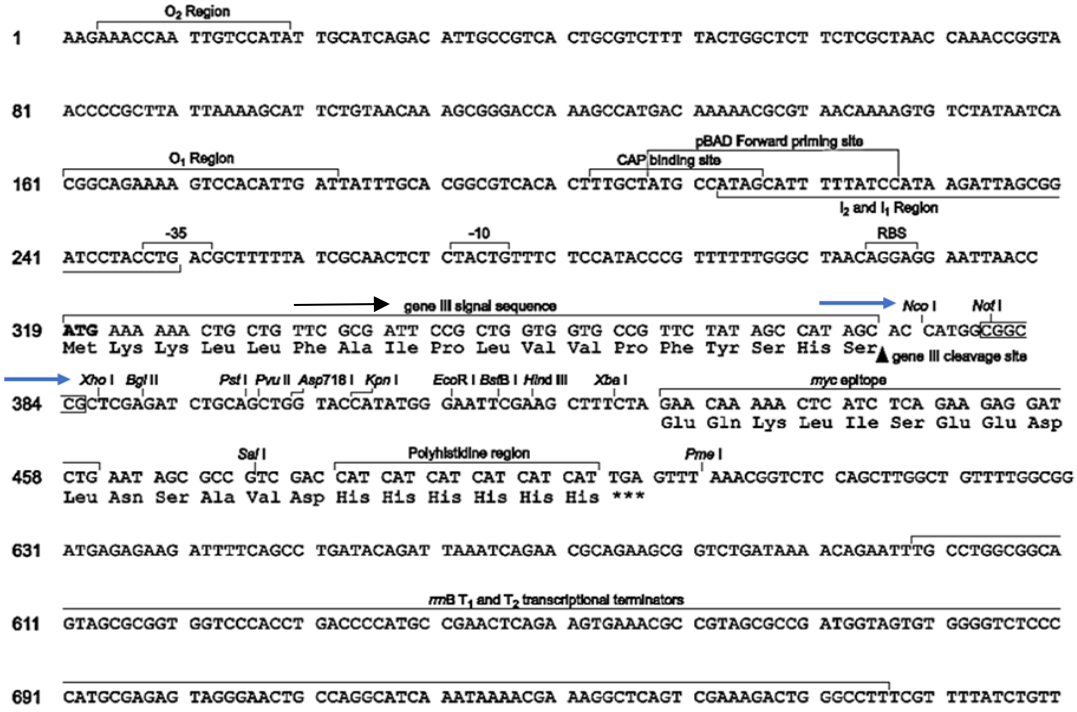


Figure 2.1 Map sequence of the pBAD/gIII vector (Thermofisher.com). A) Shows the vector's map of pBAD/gIII (A, B, C), derived from pBR322 encoding ampicillin resistance gene, using the araBAD promoter (PBAD), and the regulatory protein, AraC, that allows regulation of PBAD. B) Shows pBAD/gIIIc multiple cloning site sequence with restriction sites used for cloning, NcoI and XhoI, indicated by blue arrows. While black arrow indicates gene III sequence which utilized for secretion of recombinant protein into the periplasmic space.

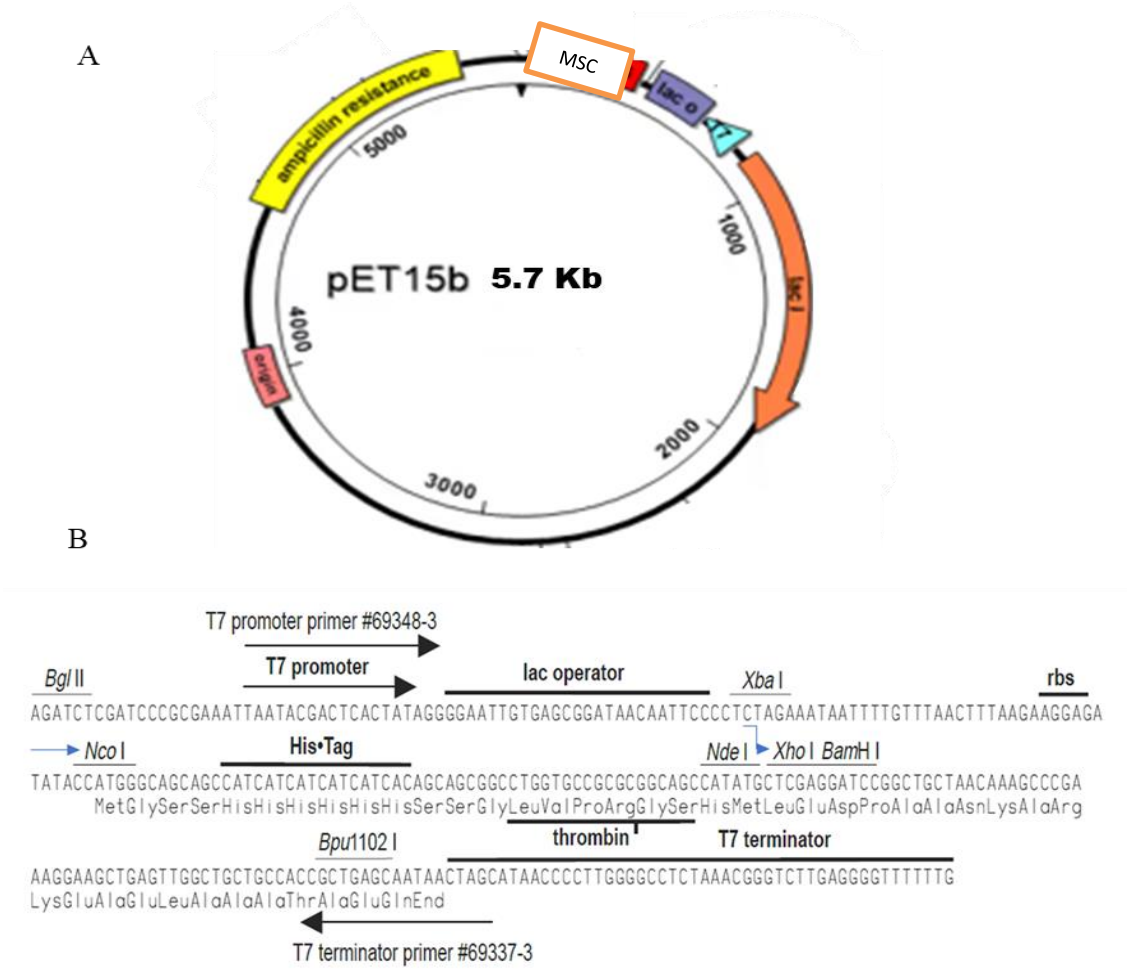


Figure 2.2 Map sequence of the pET-15b vector (Novagen.com). A) Shows the vector's map of pET-15b. derived from pBR322 encoding ampicillin resistance gene, using the T7 promoter and the lac operator for controlled expression. **B)** Shows vector's multiple cloning site sequence with restriction sites used for cloning, NcoI and XhoI, indicated by blue arrows.

2.1.5 Bacterial constructs and PCR primers

Figure 2.3 Shows the sequence of the TA region with the TolA and TolB box regions indicated in black and blue lettering respectively. The primer sequences used for the amplification of the TA regions are listed in Table 2.3.

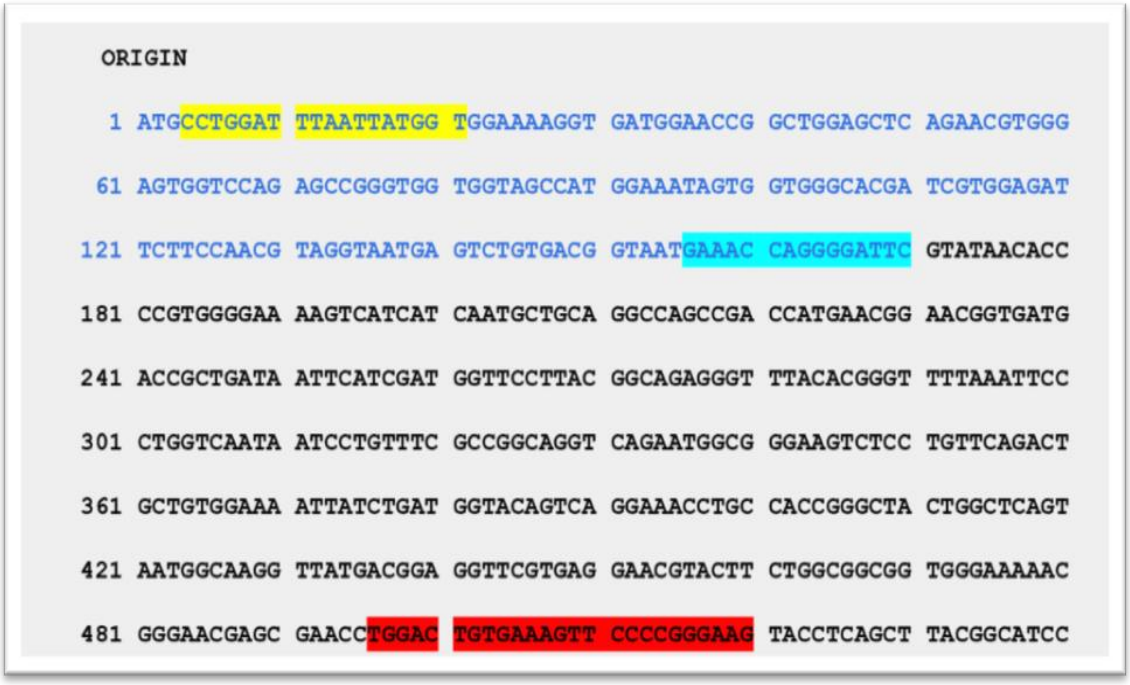


Figure 2.3 TA full sequence. The TolB box region is indicated in dark blue, the TolA box region in black. The sequences highlighted in yellow, light blue or red represent the priming regions used for PCR amplification and the introduction of the restriction sites required for cloning of the different TA regions into the vectors used (see also Table 2.3).

Primer	Oligonucleotide sequence 5`-3`	Description
RI1	GCCATGGCTGGATTTAATA TGGT	Forward primer to introduce the *Nco*I site in ColA TA1–172
RI2	GCTCGAGCCGGGGAACTTT CACAGTCCA	Reverse primer to introduce the *Xho*I site in ColA TA1–172
RI3	CGGCCATGGAACCAGGGG ATTCG	Forward primer to clone ColA TA52–172 and introduce an *Nco*I restriction site
RI4	CCCTCGAGCATTACCGTCA CAGA	Reverse primer to clone ColA TA1–52 and introduce an *Xho*I restriction site
pBAD Fwd	ATGCCATAGCATTTTTATC C	Forward primer for sequencing inserts in the pBAD/gIIIc vector.
pBAD Rev	GATTTAATCTGTATCAGG	Reverse primer for sequencing inserts in the pBAD/gIIIc vector.

Table 2.3 List of polymerase chain reaction (PCR) primers used for amplification of the different TA regions. The restriction endonuclease sites introduced to facilitate their cloning are indicated by the underlined sequences.

2.2 General methods

2.2.1 Chemical transformation of recombinant pColA using heat shock transformation into *E. coli* DH5α and BL21 (DE3)

First, 50 ng of pColA was added to 100 µL of chemically competent *E. coli* DH5α and to 50 µL of *E. coli* BL21 (DE3) as well for future expression. After incubation on ice for 30 min, the cells were heat shocked at 42 °C and again placed on ice for 5 min, all according to the manufacturer's instructions. The cells were incubated at 37 °C and shaking at 200 rpm for 1 h after the addition of 400 µL of pre-warmed super optimal broth with catabolite repression (SOC). 200 µL of the culture were plated on LB agar plates containing 100 µg/mL of ampicillin and grown for 18 h at 37 °C. Colonies were selected and recultured for further experiments.

2.2.2 Purification of pColA

A single colony of the transformed cells containing pColA was grown overnight at 37 °C with shaking at 200 rpm in 10 mL of LB supplemented with 100 µg/mL of ampicillin. The recombinant vector DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen, Germany) following the manufacturer's instructions. The miniprep was checked by agarose gel electrophoresis described in Section 2.2.3.2. With large-scale purification of pColA, a 500 mL of overnight culture (LB +amp) was conducted using a Plasmid Maxi Kit (Qiagen, Germany) in compliance to the manufacturer's instructions, and the purified plasmid was stored at -20 °C.

2.2.3 Cloning of recombinant TA region into pET-15b to produce pTA1 or into pBAD/gIIIc to produce pTA

2.2.3.1 Amplification of the DNA sequence using polymerase chain reaction (PCR)

First, 20 ng of recombinant plasmid was used amplify each DNA sequence (TA or TolA box or TolB box) by polymerase chain reaction (PCR) using a MyCycler™ PCR system (Biorad laboratories, USA). Typically, the PCR reaction mix comprised the components listed in Table 2.4, using the forward and reverse primer pairs as listed in Table 2.3. A negative control lacking template DNA was carried out to rule-out cross-contamination of reaction tubes and the various reagents.

Component	Volume
DNase- and RNase-free H_2O	23.3 µL
plasmid 20 ng	20 µL
Reverse primer (10 mM)	2.5 µL
Forward primer (10 mM)	2.5 µL
dNTPs (dATP, dTTP, dGTP, dCTP)2 mM	1.5 µL each
1 U Taq polymerase	0.2 µL
Total reaction volume	50 µL

Table 2.4. Components used for PCR amplification of target sequences.

Typical reactions consisted of a denaturing step performed at 95 ℃ for 1 min, followed by 30 cycles of 95 ℃ for 45 s, 55 ℃- 60 ℃ tm for 1 min and 72 ℃ for 2 min. A final incubation step performed at 72 ℃ for 30 min was used to ensure that full-length fragments were synthesised after all of the PCR cycles were completed**.** (tm= Primer melting temperature). PCR products were checked on agarose gel electrophoresis (see Section 2.2.3.2).

2.2.3.2 Agarose gel electrophoresis

Agarose gel electrophoresis using a GT™ Horizontal electrophoresis system, (Biorad, USA) was performed on all products of each step, following the method described by Sambrook and Russell (2001). Each sample is treated with DNA loading dye. A 10 µl sample was then loaded onto a 1% agarose gel containing ethidium bromide (10 mg/ml stock) to a 0.5 µg/ml final concentration was run at 110 V (constant voltage) for 45 min, and DNA was photographed under UV light using a Gel Doc™ EZ Gel documentation system (Biorad, USA). Fragments were sized by comparing them to the 1 kb DNA ladder.

2.2.3.3 Digestion of PCR product

Following the manufacturer's instructions and using the buffer provided, a double digestion of each PCR product (TA or TolA or TolB box) using XhoI and NcoI 1U/1 µg DNA (New England Bio-Labs, UK) was performed. The 50 µl mixture was incubated at 37 ℃ for 2 h, after which DNA loading dye was added to the product to be checked and visualized by agarose gel electrophoresis described in Section 2.2.3.2.

2.2.3.4 Ligation and chemical transformation

80 ng of each DNA insert (TA or TolA box or TolB box) was ligated into the either pBAD/gIIIc or pET-15b which also digested using same restriction enzymes (i.e., NcoI & XhoI) by T4 ligase enzyme (Roche, UK), following the manufacturer's instructions. After the reactions were incubated overnight at 16 °C, the ligation product was transformed into *E. coli* BL21 (DE3) by chemical transformation using heat shock as mentioned in Section (2.2.1).

2.2.4 DNA quantification

Genomic and plasmid DNA samples were quantified on a NanoDrop ND-2000 spectrophotometer by measuring absorbance at 260 nm. The purity of samples was estimated by the ratio of absorbance of nucleic acids (260 nm) to amino acids (280 nm).

Concentrations of nucleic acids can be directly calculated from their measured absorbance values at 260 nm, using the Beer-Lambert's equation:

$$c = \frac{A}{\varepsilon L}$$

Where, C= nucleic acid concentration in molar (M). A=UV absorbance in absorbance units (AU), ε=wavelength-dependent molar absorptivity coefficient (or extinction coefficient) in $M^{-1}cm^{-1}$.

L= light path in cm (cm).

2.2.5 DNA sequencing

Plasmid DNA was diluted to a final concentration of 100 ng/µL. A 15 µg sample was mixed with 2 µL sequencing primer (10 mM) was sent to Eurofin genomics, UK for sequencing. Recombinant vectors were sequenced using primers listed in Table 2.3.

2.2.6 Pilot expression of constructs based on the pBAD/gIIIc vector

The experiment was designed to test for and optimise the expression of recombinant proteins, using different concentrations of L-arabinose. A single colony from a stock plate of the *E. coli* BL21(DE3) containing a bacterial construct (pTA) that based on the pBAD/gIIIc was used to inoculate 5 mL of LB supplemented with 100 µg/mL of ampicillin at 37 °C overnight, after which five separate tubes were prepared in each of which 10 mL of LB containing 100 µg/mL ampicillin was added to 100 µL of the overnight culture and grown at 37°C with shaking (200 rpm) until an OD_{600} of 0.6 (log phase) was reached. At that point, the prepared L-arabinose solution was added to final concentrations of (20%, 2%, 0.2%, 0.02%, and 0.002%). Five 10-fold serial dilutions of 20% L-arabinose with sterile water were prepared, the details of which appear in the Appendix B. Grown at 37 °C with shaking for 3 h, the samples were centrifuged (Eppendorf 5804, USA) at 4 °C and 10,000 g for 12 min. Resulting cell pellets were stored at -20 °C overnight, they were resuspended in with loading buffer (see Appendix B) and analysed by Sodium dodecyl sulphate (SDS-PAGE) gel electrophoresis using a Mini Protean Tetra cell (Biorad, USA) as described in Section 2.2.9.

2.2.7 Recombinant protein overexpression of (TA, TolAbox and TolB box)

2.2.7.1 Recombinant protein overexpression

A single colony from a stock plate of transformants containing each bacterial construct was used to inoculate 5 mL of LB supplemented with 100 µg/mL of ampicillin at 37 °C overnight. A 100 µL of overnight culture was added to 10 mL of LB supplemented with 100 µg/mL of ampicillin, and the mixture was grown at 37 °C with shaking at 200 rpm until the OD_{600} of 0.6 (log phase). At that point, L-arabinose 0.002% (final concentration), was added to the pBAD/gIIIc-based plasmids, and 1 mM of isopropyl B-D-1-thiogalactopyranoside (IPTG), each added to the broth to induce overexpression. The culture was incubated for another 3 h at 37 °C with shaking at 200 rpm. The bacterial culture was then harvested, centrifuged at 10,000 g for 12 min at 4 ˚C in a benchtop centrifuge (Eppendorf 5804), the supernatant was discarded and the cell pellet was then stored at -20 ˚C for further use.

2.2.7.2 Isotope-labelled recombinant protein overexpression

A single colony taken from a stock plate of transformants containing each recombinant vector (pBAD/gIIIc or pET-15b) was used to inoculate either 5 mL of LB or M9 minimal medium supplemented with 100 µg/mL of ampicillin at 37 °C overnight. A 100 µL of overnight culture was added to 10 mL of LB supplemented with 100 µg/mL ampicillin, and the mixture was grown at 37 °C with shaking at 200 rpm until the OD_{600} of 0.6 (log phase) and in some conditions the OD_{600} of 1.0 (stationary phase) was reached. At that point, the culture was harvested by centrifugation at 800 g for 20 min at room temperature. The resulting cell pellet was resuspended in 10 mL M9 minimal medium+ glucose (see

Appendix B). ^{15}N isotope- labelled NH_4Cl (1g/L) and ^{13}C isotope- labelled glucose (1g/L) were added to the medium. In some experiments (described in chapter 4) ^{15}N isotope-labelled NH_4Cl was added alone. Then, L-arabinose 0.002% (final concentration), was added to the pBAD/gIIIc-based plasmids, and 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG), was added to pET-15b based plasmids to the broth to induce overexpression. The culture was incubated for another 3 h at 37 °C with shaking at 200 rpm. The bacterial culture then was harvested by centrifugation using (Eppendorf 5804, KSA) at 400 g for 10 min at room temperature. The resulting cell pellet was then stored at -20 ˚C for further use and the supernatant was discarded.

2.2.8 Protein concentration measurement (Bradford assay)

Total protein concentrations were determined using the Bradford assay (Bradford, 1976) using reagents obtained from Qiagen according to the manufacturer's instructions. The absorbance of proteins at 595 nm was measured using a Nano-Drop ND2000 spectrophotometer, according to the manufacturer's protocol. Bovine serum albumen was used to prepare standards of known protein concentration (typically the concentrations used were 1, 5, 7.5 and 10 µg protein /ml) to produce a calibration curve. The calibration curve was obtained by plotting the OD_{595} versus the concentration of the standards and the protein concentration of the unknown sample(s) then read from the curve. All protein assays were performed in triplicate

2.2.9 SDS-PAGE

Protein samples prepared as described in Section 2.2.7 were analysed by SDS-PAGE. Prior to SDS-PAGE analysis, cell pellets were resuspended in 50 µL lysis buffer (see Appendix B) and sonicated on ice using three 10-second bursts at high intensity using a Q2000 sonicator (Qsonica, USA). The samples were allowed to cool on ice between each round of sonication. The total protein concentration in each cell pellet was determined as described in 2.2.8., and accordingly the SDS loading buffer was added to give a final concentration 20 µg/ µl of each sample. Then, samples were boiled for 3 min, 10 µl of 20 µg total protein sample was loaded in each well. Then, proteins were resolved for SDS-PAGE as described by Laemmli (1970). 16% acrylamide resolving gels and 4% stacking gels (see Appendix B) were prepared, along with a pre-stained protein marker. After the separation of protein samples by SDS- PAGE at 150 V (constant voltage) for 1h using a Mini Protean Tetra cell (Biorad, USA), the gel was transferred to Coomassie stain (see Appendix B) for 1 h with shaking. Ultimately, the gel was transferred to a de-staining solution (see Appendix B) – overnight. Then the gel was visualised using a gel documentation system (Invitrogen iBright Imaging System, USA).

2.2.9.1 SDS-PAGE analysis of proteins with low molecular weight (3-30 KDa) (TolA box and TolB box)

Tricine-SDS-PAGE gels were prepared using a 16% resolving gel overlaid with a 4% stacking gel, then running the sample through gel electrophoresis using a Mini Protean Tetra cell (Biorad, USA), as detailed in the Appendix B (Schägger, 2006), using tricine lysis buffers (see Appendix B), The sample was incubated at 37 ℃ for 60 min before being sonicated on ice using three 10-second bursts at high intensity using a Q2000 sonicator (Qsonica, USA). The samples were allowed to cool on ice between each round of sonication for a 10 s. 20 µL of SDS loading buffer according to the protein sample concentration to 20 µg/µl final concentration. 10 µL of each sample was loaded into each well, the gel was run at 200 V (constant voltage) for 1.5 h. Gels were stained with Coomassie blue for 1 h with shaking. Ultimately, the gel was transferred to a de-staining solution for overnight and the size of the protein bands was determined by comparison with a pre-stained protein standard marker and the protein bands were visualised using an iBright Imaging System.

2.2.10 Immunoblotting technique

After analysis of samples by SDS-PAGE as described in section 2.2.9, the resolved proteins were transferred to a polyvinylidene fluoride (PVDF) membrane in transfer buffer (see Appendix B) using a Trans-Blot SD semidry transfer cell (Biorad, USA) at a constant current of 14 mA for 1 h. Following transfer with shaking, the membrane was incubated for 1 h at room temperature in blocking buffer comprising 5% bovine skimmed milk in Tris buffered-saline, tween 20% (TBST) (see Appendix B). Mouse anti-pentahistidine-

conjugated alkaline phosphate antibody which recognizes native or denatured, reduced forms of proteins tagged with 6X histidine, expressed in selected vector. The antibody is diluted 1:5000 in blocking buffer (5% skimmed milk and TBST) was then added and incubated overnight at 4 ℃ with shaking, after which the membrane was washed twice with TBST for 15 min, followed by incubation with a chemiluminescent alkaline phosphatase substrate (CDP-*Star*® Western Blot Chemiluminescence Reagent (PerkinElmer, USA)) which allows visualisation.

2.2.11 Recombinant protein purification under native conditions

The overexpressed protein ColE9-Imm9 in *E. coli* Bl21 (DE3) cells was purified using a Ni-NTA Spin Kit (Qiagen, Germany) following the manufacturer's instructions. The purity was checked by SDS as described in section 2.2.9. The protein sample was dialysed in dialysis tubing using cellulose membrane with molecular weight cut off (MWCO)=14,000, against 1L phosphate buffered saline (PBS) (Appendix B) pH 7.4 at 4 ℃ overnight (twice) and subsequently stored at -20 ℃ for further use.

2.2.12 Analysis of the antimicrobial activity of purified ColE9-Imm9 complex against different *E. coli* cells using a spot test

To 10 mL of molten 0.7% agarose cooled to 60 ℃, 100 µL of an overnight culture of different *E. coli* bacterial culture (mentioned in chapter 5) in LB medium was added, and the mixture was poured onto a single LB agar plate. Once the agar overlay had set, 10 µL of the purified ColE9-Im9 complex at three different concentrations (5 nM, 10 nM and 100

nM) was spotted on the agar. 10 µL PBS served as a control. The plate was incubated overnight at 37 °C.

2.2.13 Analysis of the antimicrobial activity of purified ColE9-Imm9 complex against *E. coli* BL21(DE3) in liquid culture.

A growth assay was performed to measure ColE9-Imm9 capacity to inhibit the growth of an *E. coli* BL21(DE3) in liquid culture. An overnight culture was grown in LB broth with shaking (200 rpm) at 37 ºC. 10 ml of the culture was diluted to 1/100 in M9 minimal medium+ glucose and incubated with shaking at 37 ºC until it reached an OD_{600} value of 0.1 (lag phase). 180-µL aliquots of this cell suspension were then added to each well of a flat-bottomed 96-well microtiter plate (Thermo Fisher Scientific, USA). Purified ColE9-Im9 was diluted in PBS to final concentrations of 10 and 100 nM and in some experiments 10 mM DTT was added, at which point a 20 µL of purified protein was added to each aliquot. The microtiter plate was then incubated in a microtiter plate reader (iEMS, Labsystems/Thermo Scientific, UK) at 37 °C for 12 h with shaking. OD_{640} readings were taken every 30 minutes with shaking for 5 s between each optical density reading. All proteins samples were tested in triplicate, and PBS was used as a negative control for antimicrobial activity.

2.2.14 Analysis of the antimicrobial activity of purified ColE9-Imm9 complex against *E. coli* BL21(DE3) expressing different TA regions in liquid culture

A single colony of each of the *E. coli* BL21(DE3) carrying recombinant pTA or pTolA ot TolB used to inoculate three separate 5 mL aliquots of LB supplemented with 100 µg/mL of ampicillin at 37 °C overnight. To 100 µL of each of the overnight cultures, 10 mL of

M9 minimal medium supplemented with 10 µg/ml ampicillin was then added, and the mixture was cultured at 37 °C with shaking (200 rpm) until an OD_{600} value of 0.6 (log phase) was reached. At that point, 100 µL of L-arabinose was added to a final concentration of 0.002% to induce overexpression, and the culture was incubated for a further 3 h at 37 °C with shaking (200 rpm), after which the culture was diluted to 1/100 and again incubated with shaking (200 rpm) at 37 ºC until it reached an OD_{600} value of 0.1 (lag phase). Next, 180-µL aliquots of each of the three different cell suspensions were added to the wells of three separate flat-bottomed 96-well microtiter plates (Thermo Fisher Scientific). Purified ColE9-Im9 was diluted in PBS to final concentrations of 10 and 100 nM and in some experiments 10 mM DTT was added, at which point a 20 µL of purified protein was added to each aliquot. The microtiter plate was then incubated in a microtiter plate reader (iEMS, Labsystems/Thermo Scientific, UK) at 37 °C for 24 h with shaking. OD_{640} readings were taken every 30 minutes with shaking for 5 s between each optical density reading. All proteins samples were tested in triplicate, and PBS was used as a negative control for antimicrobial activity.

2.2.15 Sample preparation for confocal laser scanning microscope (CLSM)

An overnight culture of each bacterial construct was diluted in M9 minimal medium to 1/100 and grown at 37 °C with shaking (200 rpm) until an OD_{600} value of 0.6 (log phase) was reached. For cells in which protein overexpression needed to be induced, L-arabinose 0.002% (final concentration) was added (as described in section 2.2.14) to the pBAD/gIIIc-based plasmids culture, and the culture was incubated for another 3 h at 37 °C with shaking (200 rpm). Following this, two 1 mL samples of the bacterial culture were taken, added to

100 mL of M9 minimal media and incubated at 37 °C with shaking (200 rpm) until an OD_{600} value of 0.1 was reached (lag phase). Next, 50 µL of the ColE9-Imm9 complex protein plus 10 mM of DTT in PBS was added to each bacterial culture, and the mixture was incubated at 37 °C with shaking for 3 h, then one of the two 1 ml sample was followed by centrifuging (Sigma 22003, UK) at 4000 g for 10 min at room temperature. Afterwards, the supernatant was removed, the obtained cell pellet was re-suspended in 0.85% NaCl; the sample was centrifuged again at 4000 g for 10 min, and the pellet was re-suspended in 1 mL of 0.85% NaCl, according the manufacturer's instructions for the Live/Dead BacLight™ (Qiagen) bacterial viability kit. The other 1-mL sample was incubated at 37 °C with shaking overnight, and the sample was processed after 18 h the same way.

2.2.16 Staining the bacterial sample for CLSM

Bacterial samples prepared as described in Section 2.2.15 were stained using the Live/Dead BacLight® bacterial viability kit, consisting of the dual fluorescent dyes SYTO 9 and propidium iodide (PI). A 1:1 mixture of SYTO 9 and PI was mixed in Safe-Lock® microcentrifuge tubes and centrifuged (Sigma 22003 centrifuge) for 1 minute at 5000 g at room temperature, after which 3 µL of the dye mixture was added to 1 mL of each of the bacterial samples in Safe-Lock® microcentrifuge tubes and incubated in the dark at room temperature for 15 min. Subsequently, 5 µL of the resulting stained culture was placed on a microscope slide and covered with a coverslip for analysis by CLSM.

2.2.17 Sample analysis by CLSM

Samples prepared as described in Section 2.2.16 were examined with a confocal laser scanning microscope (Leica TCS SP5, UK). Images were acquired using the Leica Application Suite Advanced Fluorescence software 5 (Leica, UK). Each experiment was conducted in triplicate with 20 fields of view counted each time, then the results achieved were statistically analysed as described in chapter 5.

2.2.18 Plate counts to determine the effect of ColE9-Imm9 on *E. coli* BL21(DE3)

Using the bacterial culture as prepared in Section 2.2.13, seven 10-fold serial dilutions in sterile normal saline were performed, after which 100 µL of each diluted culture was spread on LB agar plate and incubated at 37 °C overnight and colonies were counted using (Stuart digital colony counter, UK). The experiment was carried out in triplicate.

2.2.19 Sample preparation for in-cell NMR

The bacterial cell pellets obtained as described in Section 2.2.7.2 were each resuspended in three different media (Spent medium, LB and M9 minimal medium under different conditions) as described in Chapter 6 forming a bacterial slurry. A 10% D_2O (v/v) was added to bacterial slurry and transferred to 5-mm NMR tube. Along with a negative control which was culture media only.

2.2.20 ^{1}H-^{15}N correlation NMR of *E. coli* BL21(DE3) cytoplasmically expressing labelled TA in an autoinduction NMR medium

A single colony was selected from the agar plate of *E. coli* BL21(DE3) carrying pTA1 and inoculated into 3 mL of non-inducing medium from the auto-induction NMR medium kit (Novagen, USA). The culture was incubated at 37 °C with shaking at 300 rpm. After 7 h, The 3 mL culture was added to 100 mL of non-inducing medium pre-warmed to 37 °C in a 500 mL flask, placed in an incubator at 37 °C, and shaken at 250 rpm. After 18 h, 40 mL of the 100 mL culture was transferred to a 2 L polyethylene terephthalate bottle containing 460 mL of auto-induction medium and incubated at 37 °C with shaking. After 18 h, the cells were harvested by centrifugation for 20 min at 400 g at room temperature, re-suspended in unlabelled M9 medium containing 10% D_2O (v/v), and transferred to a 5-mm NMR tube.

2.2.21 Quality control test to detect protein leakage

Each representative NMR sample was centrifuged (Z206-A centrifuge, Hermle, UK) at 800 g for 10 minutes at room temperature, then the resulting pellet was re-suspended in unlabelled medium containing 10% D_2O, and transferred to a 5-mm NMR tube while the supernatant was transferred to other 5-mm NMR tube after adding 10% D_2O (v/v) for in-cell NMR analysis.

Chapter 3

Cloning of recombinant colicin A translocation domain and its reduntants

3.1 Introduction

This project aimed to study the interaction of colicin and Tol proteins. As previously discussed in Chapter 1, during the transfer of colicin A (ColA) protein across the periplasmic space, ColA interacts with TolA and TolB proteins in the bacterial periplasm at sites within the colicin structure called Tol boxes, leading to the formation of complexes.

ColA consists of three domains; 1–172 residue representing the translocation domain (TA) with TolA and TolB boxes (i.e., the region where ColA and Tol proteins bind), 173–388 residue spanning the receptor binding (R) domain, and 389–592 residue representing the cytotoxic (C) domain with 10 α-helices that form a voltage-gated channel in the cytoplasmic membrane (Penfold *et al*., 2012).

This chapter presents a method developed to clone the TA (residues 1–172), and the colicin TolB box (residues 1–52), and TolA box (residues 52–172) to enable the investigation of the interaction between them in the periplasm of *E. coli* cells.

As previously discussed in Chapter 1, ColA is synthesized in the cytoplasm of the producing cells, and to attack sensitive cells, is exported and translocated into the periplasm of the sensitive cells (Crozel *et al*., 1984). ColA begins entering a sensitive *E. coli* cell by binding to an outer membrane (OM) receptor (i.e., vitamin B_{12} transporter BtuB) and to a porin translocator called OmpF (Di Masi *et al*., 1973), which acts as a co-receptor, as shown in Figure 3.1. The receptors facilitate the translocation of the unstructured region of the (T) domain of TA by recruiting the Tol proteins to facilitate the entry of the (C) domain through the OM to the inner membrane (IM) (Cao & Klebba, 2002; Housden *et al*., 2005).

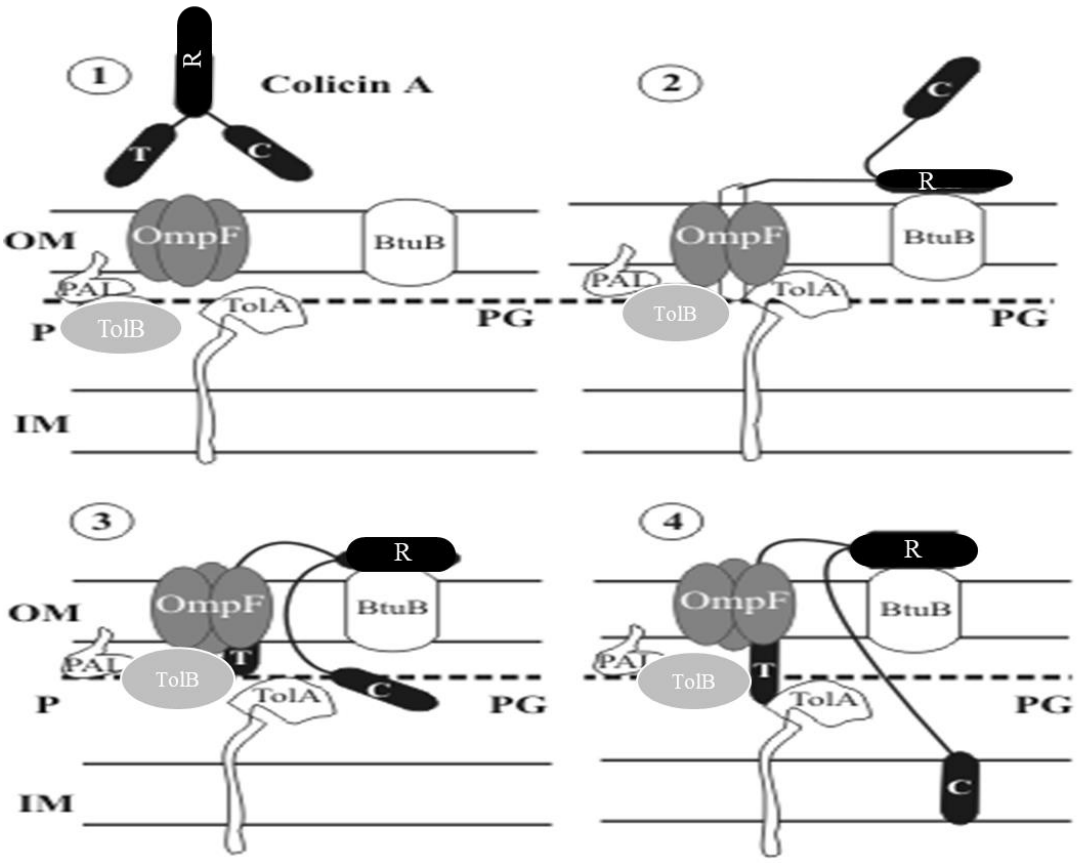


Figure 3.1 Proposed mechanism of ColA translocation (**adapted from Lazdunski *et al*., 1998).** 1) Shows the three colicin domains on the black structure: Translocation (T) domain, receptor binding (R) domain, and cytotoxic (C) domain of ColA. 2) At the outer membrane (OM), where (R) domain binds to BtuB. 3)The (T) domain translocate into the periplasm (P) through binding to OmpF, where TolB forms a complex with Pal and TolA, allowing the crossing of the (C) domain into the peptidoglycan layer (PG). 4) The (T) domain interacts with TolA and TolB, allowing the translocation of the (C) domain further into its target in the inner membrane (IM).

The direct interaction of ColA with TolA and TolB proteins is unique. In an extensive study on ColA entry into the periplasmic space and the importance of the TolA box in translocation, it was suggested that the TolA box might not be involved in the interaction between ColA and TolA (Bouveret *et al.*, 1998). However, in contrast, when their physical interaction was examined *in vitro* by Bouveret and colleagues (2002) using a series of deletion mutants in TolA, it has led them to the identification of regions involved in interactions between ColA and both TolA and TolB. They proposed that during ColA translocation, the TolB box first binds to TolB, after which its TolA box binds to TolA. Later, Penfold and colleagues (2012) concluded the same results, but the nature of those interactions remains unclear (Penfold *et al.*, 2012).

Based on such assumptions, the first objective of this project was to clone the ColA (T) domain (TA residues 1–172), and the TolB (TA residues 1–52) and TolA boxes (TA residues 52–172) that have been identified in the literature (Cascales *et al*., 2007).

This chapter describes the separate cloning of recombinant TA, the TolA box, and the TolB box into pBAD/gIIIc vector for periplasmic expression for live imaging studies and cloning of recombinant TA into pET-15b vector for cytoplasmic expression for in-cell NMR studies and transforming into *E. coli* cells.

3.2 Results

3.2.1 Cloning of recombinant TA region into pET-15b vector to produce pTA1

As described in Section 2.2.3.1, the TA domain in pColA was amplified by polymerase chain reaction (PCR), and the PCR product was analysed on a 1% agarose gel. The expected 500 bp band corresponding to the full length of TA domain is indicated in Figure 3.2.

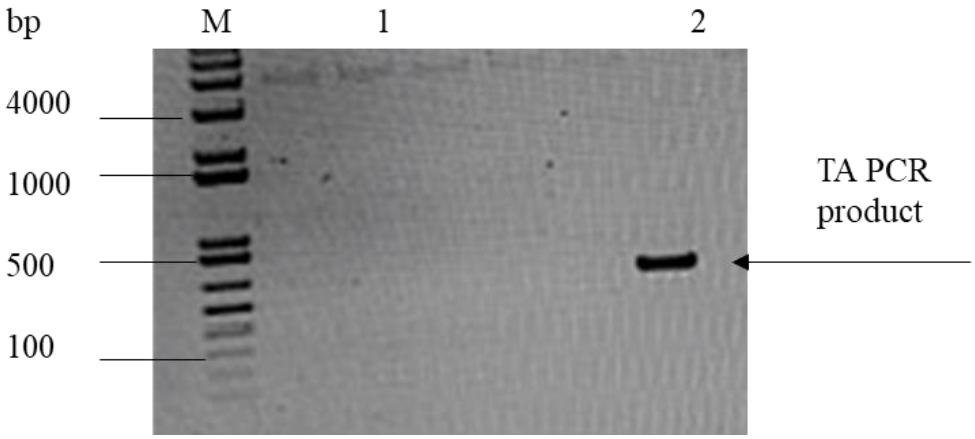


Figure 3.2 1% Agarose gel showing the PCR amplified TA region. Lane M: DNA ladder 1 kb. Lane 1: negative control (with no template). Lane 2: TA PCR product of (500 bp) at 55 °C.

Following a double digestion with NcoI and XhoI the 500 bp product corresponding to the TA domain was then cloned into pET-15b (as described in section 2.2.3.4) to create pTA1. Figure 3.3 shows the result of a double digestion of this construct by NcoI and XhoI enzymes. The TA insert at the expected size of 500 bp and the linearised pET-15b vector at 5.7 KB can be clearly seen.

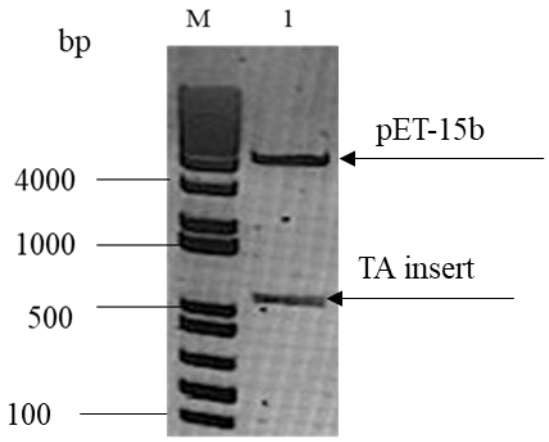


Figure 3.3 Double digestion of pTA1. Lane M: DNA ladder 1 Kb. Lane 1: The expected restriction fragments corresponding to pET-15b (with a linear size of 5.7 kb) and the TA insert (500 bp) are indicated.

3.2.2 Cloning of recombinant region into pBAD/gIIIc vector to produce pTA

As described in Section 2.2.3.1, the TA domain was amplified by PCR and the PCR product analysed on a 1% agarose gel. The expected 500 bp band corresponding to the full length of TA domain as indicated in Figure 3.4.

Following double digestion with NcoI and XhoI the 500bp product corresponding to the TA domain was then cloned into pBAD/gIIIc (as described in section 2.2.3.4) to create pTA. Figure 3.5 shows the result of a double digestion of this construct by NcoI and XhoI enzymes. The TA insert at the expected size of 500 bp and the linearised pBAD/gIIIc vector at 4.1 KB can be clearly seen.

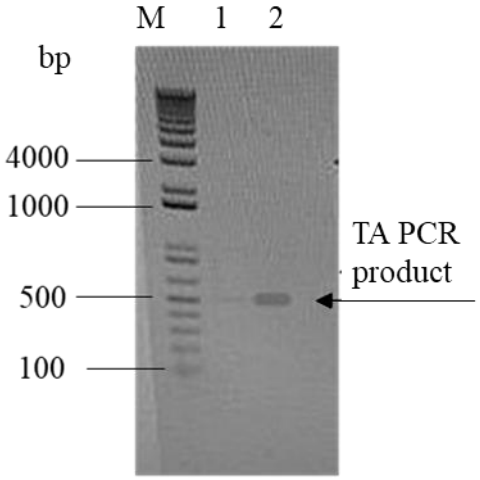


Figure 3.4 1% agarose gel showing the PCR amplified TA region. Lane M: DNA ladder 1 kb. Lane 1: negative control (no template). Lane 2: amplified TA fragment with linear size 500 bp.

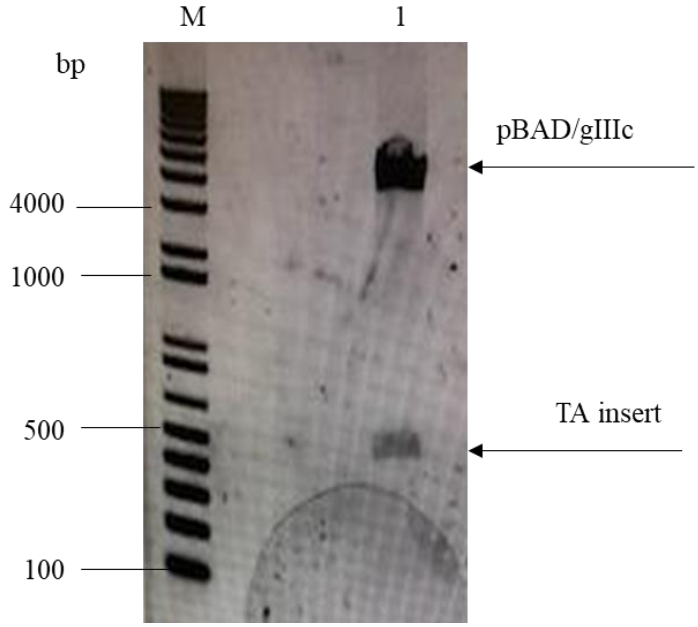


Figure 3.5 Double digestion of pTA. Lane M; DNA 1 kb ladder. Lane 1: The expected restriction fragments corresponding to pBAD/gIIIc (with a linear size of 4.1 Kb) and the TA insert (500 bp) are indicated**.**

3.2.3 Cloning of recombinant TolA box and TolB box into pBAD/gIIIc vector to produce pTolA box and pTolB box

As described in Section 2.2.3.1, the TolA box and TolB box regions were amplified by PCR and the PCR product analysed on a 1% agarose gel. Then following Section 2.2.3.4, they are both cloned into pBAD/gIIIc. The expected 363 bp band corresponding to the full length of TolA box domain is indicated in Figure 3.6. And the expected band for TolB box corresponding to 130 bp is shown in Figure 3.7.

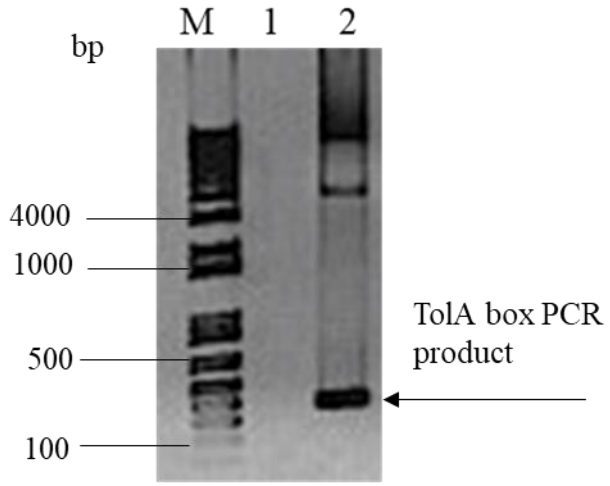


Figure 3.6 1% agarose gel showing the PCR amplified TolA box. Lane M: DNA ladder 1 kb. Lane 1: negative control (with no template). Lane 2: amplified TolA box fragment with linear size 363 bp.

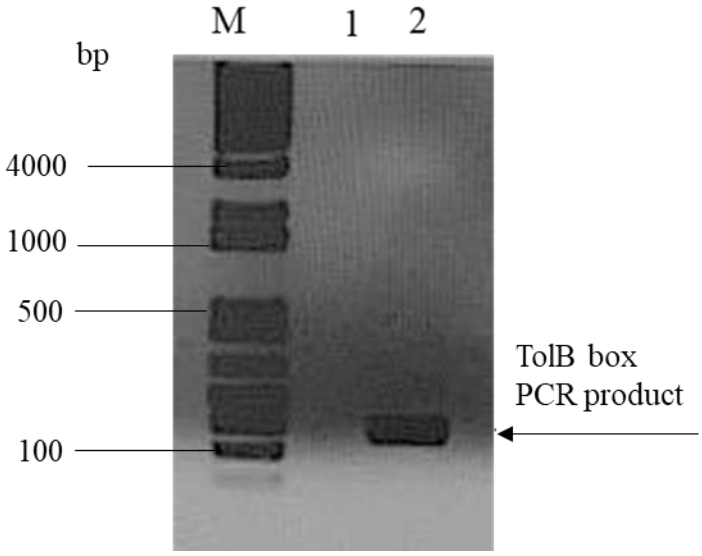


Figure 3.7 1% agarose gel showing the PCR amplified TolB box. Lane M: DNA ladder 1 kb. Lane 1: negative control (with no template). Lane 2: amplified TolB box fragment with linear size 130 bp.

3.3 Discussion

This chapter has focused on the procedures for cloning the TA, TolA and TolB boxes into two different vectors; pET-15b for cytoplasmic expression and pBAD/gIIIc for periplasmic expression. The recombinant vector pColA carrying the full length of TA region of ColA was a kind gift from Dr Chris Penfold from University of Nottingham, UK. pColA was transformed into *E. coli* DH5α for future use and into *E. coli* BL21(DE3) for future expression. The *E. coli* BL21(DE3) and its derivatives are the most used strains for protein expression; the λDE3 prophage was inserted in the chromosome of BL21 and contains the T7 RNAP gene under the *lac*UV5 promoter. In *E. coli* BL21(DE3), the expression of the gene encoding the recombinant protein is transcribed by the chromosomally encoded T7 RNA polymerase (T7 RNAP), which transcribes eight times faster than *E. coli* RNA polymerase (RNAP). *E. coli* BL21(DE3) is preferred host for protein expression because the higher the mRNA levels, the more recombinant protein can be produced (Zhang *et al*., 2015). *E. coli* BL21(DE3) strain was the chosen *E. coli* host for recombinant protein overexpression in this project.

DNA cloning technique is a common technique used to study the genes involved in biological processes (Lodish *et al*., 2000). Cloning technique using polymerase chain reaction (PCR) was used in this project for isolating and reproducing a large numbers of identical gene fragments (TA, TolA box and TolB box). On the basis of TA gene sequence (Figure 2.3), primers (Table 2.3) were designed to amplify TA gene which contained NcoI (CCATGG) and XhoI (CTCGAG) restriction sites for cloning into pET-15b to create pTA1 and into pBAD/gIIIc to create pTA. Following the PCR amplification method described in

Section 2.2.3.1, the TA PCR product was observed at 500 bp on agarose gel (Figure 3.2), after which the recombinant vector was digested with same restriction enzymes NcoI/XhoI as shown in Figure 3.3 and ligated into pEt-15b vector, where DNA ligase form phosphodiester bonds between the 3'hydroxyl and the 5' phosphate ends of the nucleic acid molecules. To be able to take up foreign DNA, the bacteria cells need to be made competent. This is often achieved by treating them with divalent cations under cold conditions. For *E. coli* transformation with plasmid the DNA needs assistance to pass through the cell membranes and to reach the site where it can be expressed and replicated. The chosen competent cells as mentioned before, are *E. coli* BL21(DE3). And the transformation involved heat shock of the solution to induce the cells to take up the recombinant plasmid pTA1, which can be used for future cytoplasmic expression. The reason behind choosing pET-15b is that the pET expression system is the most common promotor used in protein cytoplasmic expression. It features the T7 promoter (see Figure 2.2), which can make up to 50% of all cellular proteins. The strain BL21(DE3) is a T7 expression host under the control of the lac promoter; when cells are induced with isopropylyhio-β-galactoside (IPTG), the expression of T7 polymerase begins gene transcription that leads to the production of recombinant proteins (Boomershine *et al*., 2003). The highest level of expression from the lac promoter of *E. coli* only occurs when lactose is available (and glucose, the preferred energy source, is absent). When there is an absence of lactose, transcription of any associated recombinant protein from *lac* is blocked by a repressor protein. If both are present, then the lac promoter can be fully induced after the complete usage of both sugars. Consequently, using a low amount of glucose is a preferred solution for lac promoter induction. In the T7 promoter system utilised in the pET

vector, lactose, or its non-hydrolysable analogue, IPTG, is used for induction (Li *et al*., 2011).

Following the same amplification PCR method in Section 2.2.3.1, Figure 3.4 and Figure 3.5 show the results for cloning TA region into pBAD/gIIIc vector that was chosen future periplasmic expression, which can be seen at the expected linear size of 500 bp. The pBAD/gIIIc plasmids are pBR322-derived expression vectors designed for regulated (see Figure 2.1), secreted recombinant protein expression and purification in *E. coli*. The gene III signal sequence is utilized for secretion of the recombinant protein into the periplasmic space (Guzman, 1995)). The mechanism of araBAD promoter induction, which occurred in the pBAD vector, the Arac (the regulatory protein), had two functions. First, it represses translation by forming a protein–DNA complex that inhibits RNA polymerase from binding to the promoter. However, once L-arabinose is added, the second function starts, which involves the activation of transcription from the ara promoter.

The individual TolA and TolB boxes regions (i.e., where the interactions between colicins and Tol proteins are thought to occur) were both cloned into pBAD/gIIIc for secretion in the periplasm following the same technique. Other primers (Table 2.3) were used to amplify TolA box and TolB box genes which contained NcoI and XhoI restriction sites for cloning into pBAD/gIIIc. This will enable the study of each binding site separately in this project. Figure 3.6 shows the expected band of the pTolA box at 363 bp. Whereas Figure 3.7 shows the expected band of pTolB box at 130 bp. All constructs were verified by sequencing as described in Section 2.2.5 and checked using NCBI Blast (see Appendix A).

Chapter 4
Optimising protein overexpression

4.1 Introduction

Colicin A (ColA) relies on Tol proteins to be translocated through the periplasmic space to the cellular target of *E. coli*. The N- terminal (i.e., translocation domain) domain of ColA (TA) interacts with both TolA and TolB through sites called the TolA box and TolB box (Bouveret *et al.*, 1998). Cloning of both the TA region as well as the cloning of individual TolA and TolB boxes have been accomplished (as discussed in Chapter 3) to study their role in the translocation process. Isolation and analysis of the individual TolA and TolB boxes are essential to clarify the nature of the translocation process into the periplasmic space. For extensive analysis, those proteins need to be overexpressed and induced through an efficient expression system. This chapter discusses how the overexpression of certain proteins of ColA translocation domain (T) domain has been optimised for in-cell NMR analysis and imaging analysis to achieve the overall aim of the project.

Protein overexpression is a technique used to generate high levels of recombinant proteins for analysis. Although high yields of recombinant proteins have been obtained by various researchers, the level of production depends on many factors like the expression system and the induction conditions used as well as the nature of the protein of interest (Schlapschy *et al.*, 2006). *E. coli* is the most common bacterial species used for protein overexpression. However, cytoplasmic and periplasmic overexpression in that system continues to show limitations, including incorrect folding and aggregation of recombinant proteins. Nevertheless, protein overexpression in the periplasm of Gram-negative bacteria, especially *E. coli*, is widely used in many fields, in both prokaryotic and eukaryotic studies (Schlapschy *et al.*, 2006). Using *E. coli* as a host for protein expression began in the late 1960s, and since then, researchers have adjusted parameters like the structure of the

expression vector, incubation temperature, and induction temperatures used in order to improve production levels. Since the 1970s, researchers have made several advances in expressing isotope-labelled proteins promoter system as well (Mondal *et al*., 2013). Overall, *E. coli* has been shown to be a useful host bacterium for molecular biology experiments, as mentioned in Chapter 1 (Makino *et al*., 2011). Although high concentrations of proteins can also be obtained via chemical synthesis or by using cell-free protein expression systems, protein overexpression in *E. coli* remains one of the most effective methods. During protein overexpression, the protein of interest is first cloned using recombinant DNA cloning methods, and the protein is overexpressed using a suitable promoter system, and this is the approach used in this project for subsequent analysis.

Efficient expression of proteins in bacterial cells requires the adjustment of a number of specific cultivation parameters, including shaking for fast growth and aeration, which are essential during incubation. Shaking increases oxygenation, promoting aerobic growth and protein overexpression, but may also affect protein expression because as cell density increases, levels of oxygen decrease and may slow growth (Jia *et al*., 2016). Time of induction and incubation temperature are also important to avoid instability of protein by controlling proteolytic degradation (Mondal *et* al., 2013). LB medium is the most commonly used nutritionally rich medium. However, the required culture parameters can differ depending on the type of protein of interest (Jia *et al*., 2016). For instance, for nuclear magnetic resonance (NMR) techniques the use of an autoinduction medium is efficient for labelled cytoplasmic expression. The medium is cost-effective to obtain a high level of protein production and it was one of the chosen media for cytoplasmic expression in this project (see Chapter 6). In such a medium, using glycerol instead of glucose as the source

of carbon is preferred because glucose counteracts autoinduction by lactose (Mondal *et al*., 2013).

Another essential issue to consider when choosing an appropriate protein overexpression technique is protein solubility. Recombinant protein solubility can be enhanced by reducing the induction temperature to 15–18 °C. However, because the unstable pH of the culture medium due to cell metabolites can sometimes cause protein aggregation, an appropriate buffer should be added to the culture medium to maintain the desired pH level (Jia *et al*., 2016).

To investigate the involvement of TA in the translocation process, the overexpression of TA domain and the TolA and TolB boxes was optimised. As described in Chapter 3, target genes were cloned into the pET-15b vector under the control of strong bacteriophage T7 transcription and translation signals; expression was then induced by providing a source of T7 RNA polymerase in the host cell (Mierendorf *et al*., 1998). The other expression vector used, pBAD/gIIIc is designed for secretion of recombinant proteins into the periplasmic space using the gene III signal.

This chapter discusses how protein overexpression was optimised, how the best expression system was chosen, and the best conditions used for subsequent analysis.

4.2 Results

4.2.1 Pilot expression of constructs based on the pBAD/gIIIc vector

This pilot experiment was carried out in LB medium as described in section 2.2.6 in order to optimise the final concentration of L-arabinose required for maximal expression of TA protein from pBAD/gIIIc (pTA). Maximal expression was achieved using a final concentration of 0.002% L-arabinose (as illustrated in Figure 4.1) 0.002% L-arabinose was therefore the concentration of inducer used for all pBAD/gIIIc constructs used thereafter.

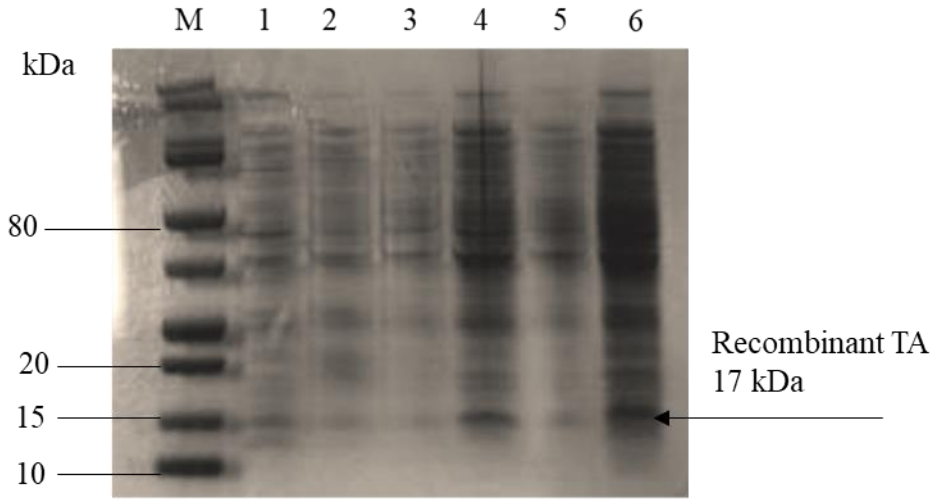


Figure 4.1 Coomassie blue stained SDS-PAGE analysis of pilot expression experiment using L-arabinose. Five 10-fold serial dilutions of L-arabinose. The induction of *E. coli* BL21(DE3) expressing recombinant TA at 17 kDa was the optimal with 0.002% L-arabinose. M: Dual pre-stained protein marker 10-250 kDa. Lane 1: non-induced cells; Lane 2: 0.2% L-arabinose induced cells expressing recombinant TA; Lane 3: 0.02% L-arabinose induced cells expressing recombinant TA; Lane 4: 0.002% L-arabinose induced cells expressing recombinant TA; Lane 5: 0.0002% L-arabinose induced cells expressing recombinant TA; Lane 6: 0.00002% L-arabinose induced cells expressing recombinant TA.

4.2.2 Overexpression of recombinant TA cloned into pET-15b vector

The TA region of *E. coli* BL21 (DE3) cells was cloned into pET-15b vector and the expression was induced in LB medium as described in section 2.2.7.1. SDS-PAGE was performed as mentioned in section 2.2.9, SDS-PAGE, along with non- induced sample as negative control revealed the expected band of the recombinant cytoplasmic TA at 17 kDa.

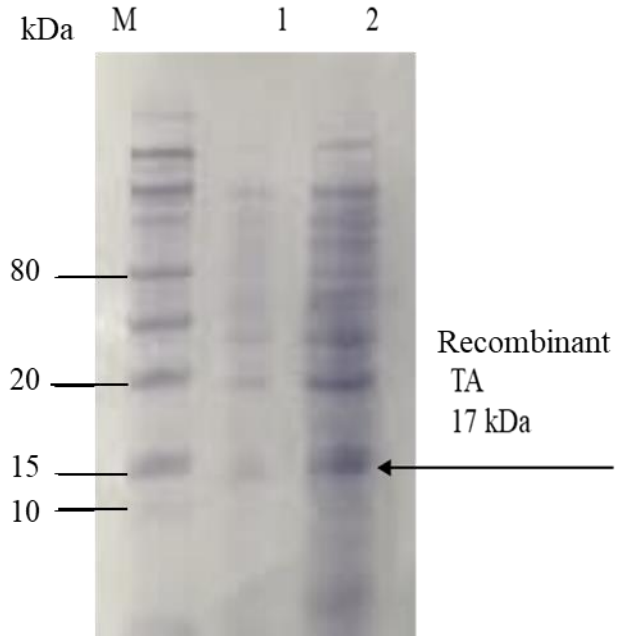


Figure 4.2 Coomassie blue stained SDS-PAGE analysis of cytoplasmic recombinant TA. Lane M: Pre-stained protein marker 10-250 kDa. Lane 1: non-induced cells as negative control. Lane 2: 1 mM IPTG induced cells expressing recombinant TA at the expected molecular weight of 17 kDa.

4.2.3 Overexpression of recombinant TA cloned into pBAD/gIIIc vector

Periplasmic protein expression of TA was achieved in LB with 0.002% L-arabinose, as described in section 2.2.7.1. SDS-PAGE of the sample, along with non- induced sample as negative control revealed the expected band for recombinant TA at 17 kDa as illustrated in Figure 4.3.

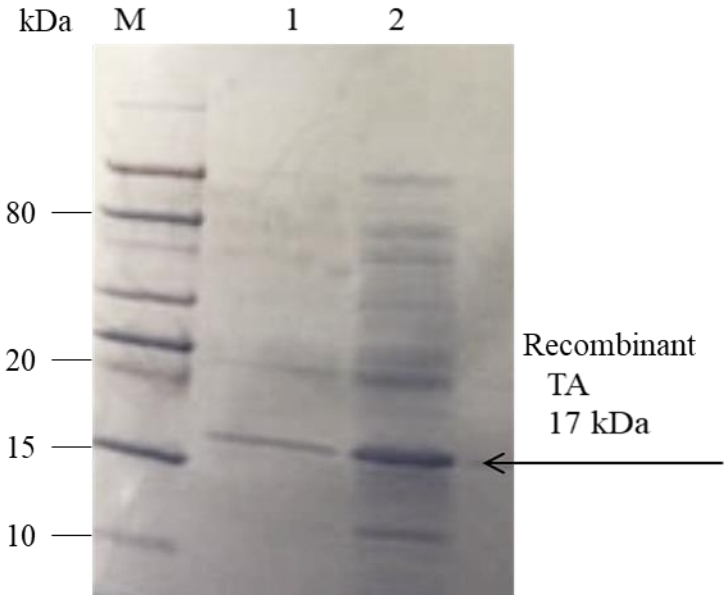


Figure 4.3 Coomassie blue stained SDS-PAGE analysis of periplasmic recombinant TA. Lane M: Dual pre-stained protein marker 10-250 kDa; Lane 1: non- induced cells; Lane 2: 0.002% L-arabinose induced cells expressing recombinant TA at the expected molecular weight of 17 kDa.

4.2.4 Overexpression of isotopically labelled recombinant TA in pBAD/gIIIc and pET-15b vectors in M9 minimal media

Protein overexpression from both constructs (i.e., pTA/pTA1) was carried out as mentioned in Section 2.2.7.2. Both starter culture (i.e., overnight culture) was M9 minimal medium, and at the time of induction, $^{15}NH_4Cl$ and ^{13}glucose were added both at same time to same culture. Figure 4.4 shows the result of poor induction that resulted weak level of expression at the expected molecular weight 17 kDa.

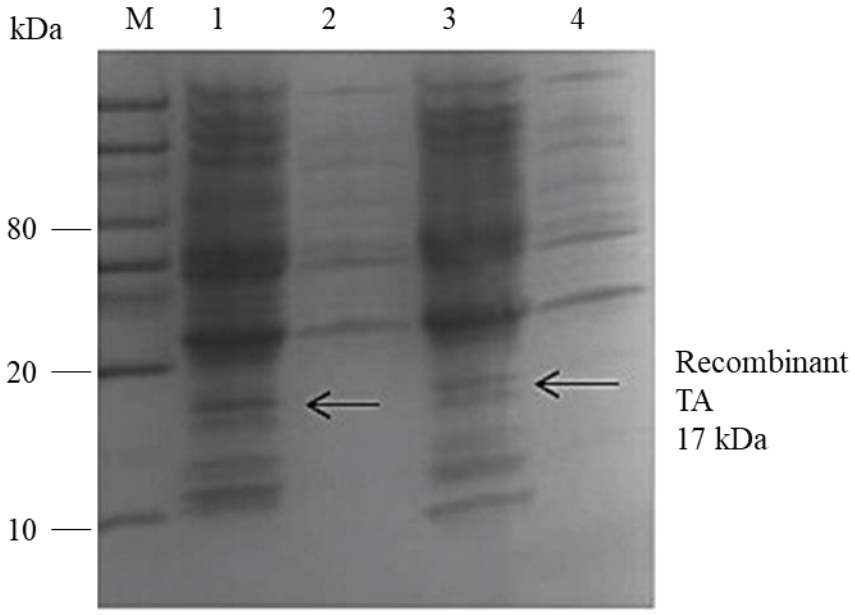


Figure 4.4 Coomassie blue stained SDS-PAGE analysis of overexpression of isotopically labelled recombinant TA in pBAD/gIIIc and pET-15b vectors in M9 minimal media. M: Dual pre-stained protein marker 10-250 kDa. Lane 1: 1 mM IPTG induced cells expressing isotopically labelled recombinant TA in the cytoplasm (pET-15b expression). Lane 2 & 4: non-induced cells as negative controls; Lane 3: 0.002% L-arabinose induced cells expressing isotopically labelled recombinant TA in the periplasm (pBAD/gIIIc expression). The expected recombinant TA protein should be observed at a molecular weight of 17 kDa.

4.2.5 Overexpression of isotopically labelled recombinant TA in pBAD/gIIIc and pET-15b vectors in LB then in M9 minimal media

The method of overexpression was modified in an attempt to increase the level of recombinant protein expression over that described in Section 4.2.4. The starter culture (i.e., overnight culture) was grown in rich medium (LB) instead of M9 minimal medium, then switched to isotopically labelled M9 minimal medium just prior to time of induction when the $O.D_{600}$ reaches 0.6. Nevertheless, the total protein concentration achieved which was calculated by Bradford assay was broadly comparable to that obtained using the methodology described in Section 4.2.4, and still low compared to the amount of protein product yielded from the approach used in Section 4.2.2 and 4.2.3. As illustrated in Figure 4.5, the result was not acceptable. The gel indicates only low-level expression at the expected size of 17 kDa.

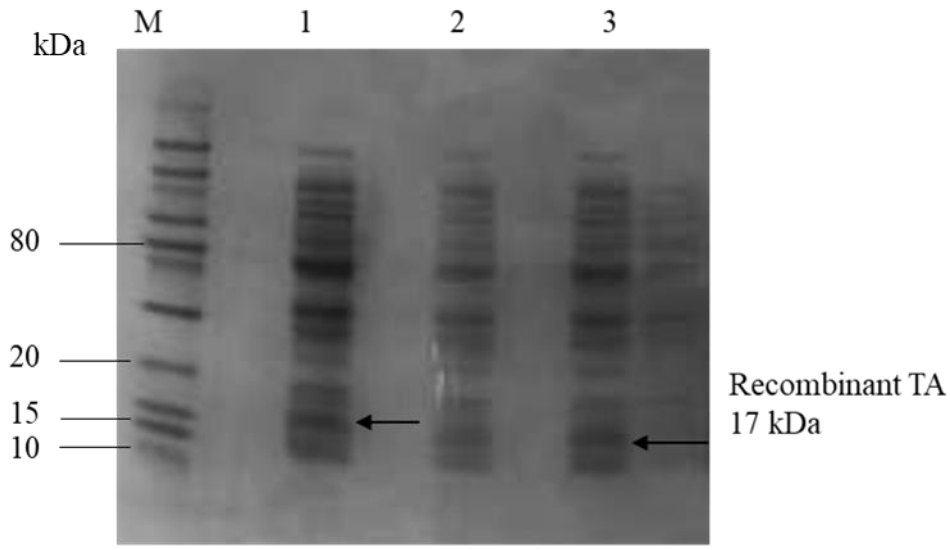


Figure 4.5 Coomassie blue stained SDS- PAGE analysis of isotopically labelled recombinant TA in LB then in M9 minimal medium. M: Dual pre-stained protein marker 10-250 kDa. Lane 1: 1mM IPTG induced cells expressing the isotopically labelled recombinant TA in the cytoplasm (pET-15b expression) at the expected molecular weight of 17 kDa. Lane 2: non-induced cells as negative control; Lane 3: 0.002% L-arabinose induced cells expressing isotopically labelled recombinant TA in the periplasm (pBAD/gIIIc expression), at the expected molecular weight of 17 kDa.

4.2.6 Overexpression of isotopically labelled recombinant TA cloned in both pET-15b and pBAD/gIIIc for overnight at low temperature.

Attempts to optimise protein overexpression, changes in conditions were applied to the method in section 4.2.5. Via overnight induction instead of 3 h to yield higher concentration of protein, the induction was carried out at a lower temperature of 25 °C instead of 37 °C (i.e., the normal induction temperature). However, it produced very similar results to those obtained in previous experiments (in Sections 4.2.4 and 4.2.5) i.e., the expression level of isotopically labelled TA can be seen to be relatively low as illustrated in Figure 4.6.

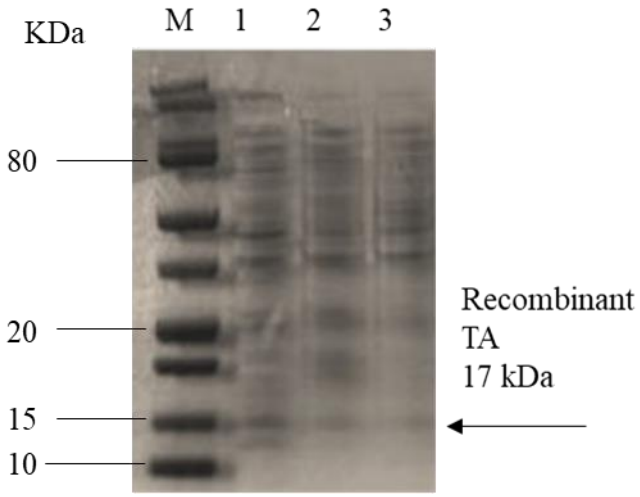


Figure 4.6 Coomassie blue stained SDS-PAGE analysis of an overnight expression of isotopically labelled recombinant TA at low temperature. M: Dual pre-stained protein marker 10-250 kDa. Lane 1: 0.002% L-arabinose induced cells expressing isotopically labelled recombinant TA in the periplasm (pBAD/gIIIc expression). Lane 2: 1 mM IPTG induced cells expressing isotopically labelled recombinant TA in the cytoplasm (pET-15b expression). Lane 3: non-induced cells as negative control. All showing very low level of expression at the expected molecular weight of 17 kDa.

4.2.7 Overexpression of isotopically labelled recombinant TA cloned in both pET-15b and pBAD/gIIIc; Induction at high cell density.

Using essentially the same method described in Section 4.2.5, but the bacterial culture to be induced was instead grown to an $O.D_{600}$ of 1.0 before switching to isotopically labelled M9 minimal medium. This gave the optimal results for the protein overexpression, SDS-PAGE of sample along with non- induced sample as negative control, showed the expected band at 17 kDa as illustrated in Figure 4.7.

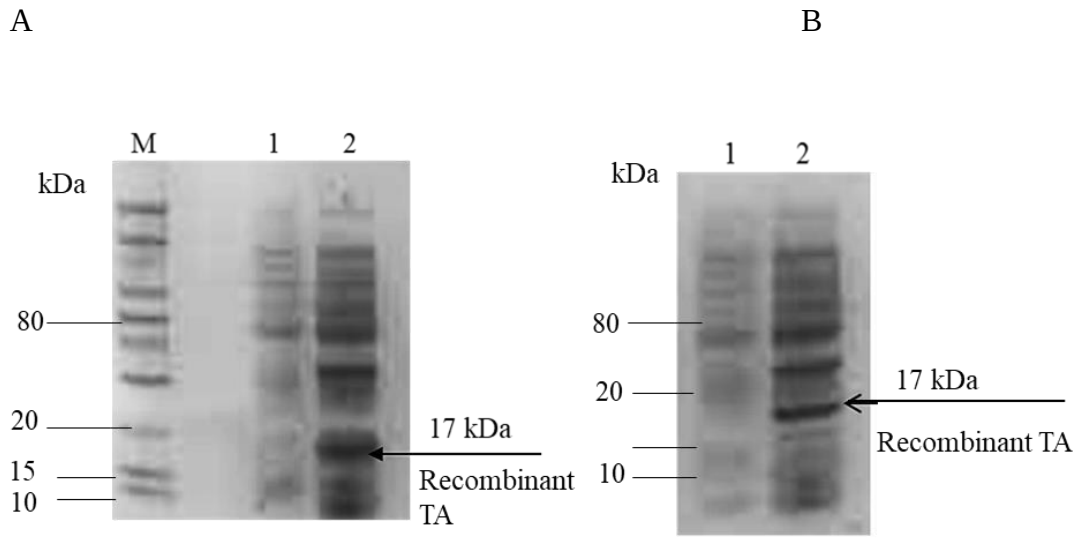


Figure 4.7 Coomassie blue stained SDS-PAGE analysis of isotopically labelled recombinant TA protein; Induction at high cell density. Lane M: Dual pre-stained protein marker 10-250 kDa. A) SDS-PAGE of isotopically labelled recombinant TA in the cytoplasm (pET-15b expression) at the expected molecular weight of 17 kDa. Lane 1: non-induced cells as negative control; Lane 2: 1 mM IPTG induced cells expressing isotopically labelled recombinant TA at the expected molecular weight of 17 kDa. B) SDS-PAGE of isotopically labelled recombinant TA in the periplasm (pBAD/gIIIc expression) at the expected molecular weight of 17 kDa. Lane 1: non-induced cells as negative control; Lane 2: 0.002% L-arabinose induced cells expressing isotopically recombinant TA at the expected molecular weight of 17 kDa.

4.2.8 Detection of recombinant TA protein using immunoblotting

Following the experiment in section 2.2.10, The TA protein domain cloned into either pET-15b (pTA1) or pBAD/gIIIc (pTA) was detected using immuno-blotting to reveal the expected band at 17 kDa (Figure 4.8) by analysing along the cells expressing recombinant TA in LB medium as positive control for the isotopic labelled proteins, both cytoplasmic and periplasmic recombinant TA were identified by SDS-PAGE, then they were detected at the same expected molecular weight as the positive controls at 17 kDa.

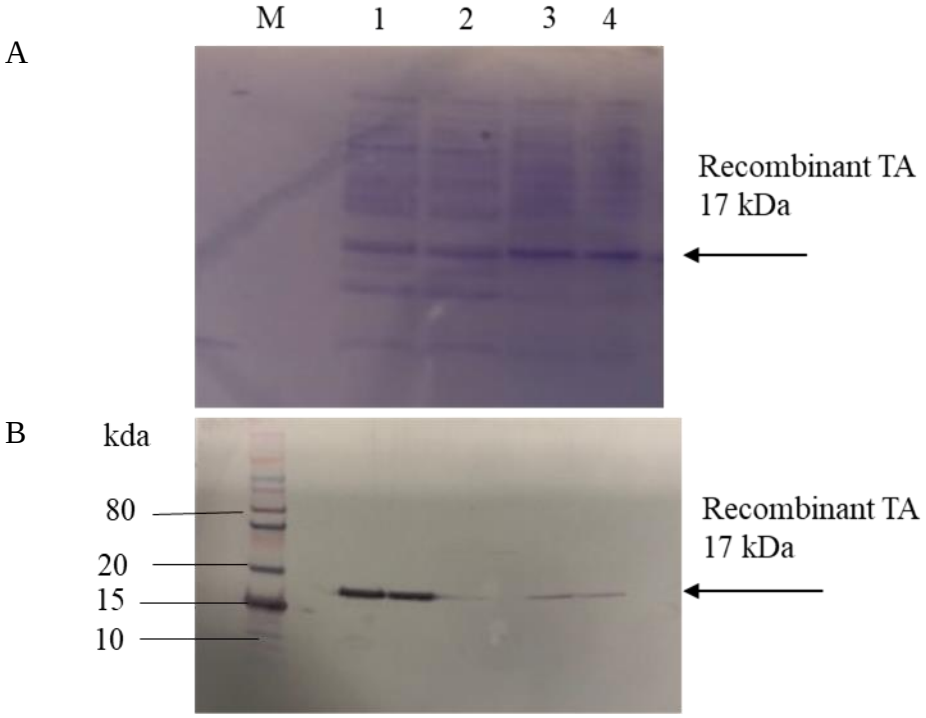


Figure 4.8 Coomassie blue stained SDS-PAGE analysis and immunodetection of isotopically labelled recombinant TA protein. A) protein immunoblotting. B) Coomassie blue stained SDS-PAGE**.** Lane M: Dual pre-stained protein marker 10-250 kDa. Lane 1: 1 mM IPTG induced cells expressing the recombinant TA in the cytoplasm (pET-15b expression. Lane 2: 1 mM IPTG induced cells expressing the isotopically labelled recombinant TA in the cytoplasm (pET-15b expression. Lane 3: 0.002% L-arabinose induced cells expressing recombinant TA in the periplasm (pBAD/gIIIc expression). Lane 4: 0.002% L-arabinose induced cells expressing isotopically labelled recombinant TA in the periplasm (pBAD/gIIIc expression), all at the expected molecular weight of 17 kDa.

4.2.9 Overexpression of isotopically labelled recombinant TolA box and TolB box in pBAD/gIIIc

Protein overexpression was carried out as described in Section 2.2.7.1. with the same conditions mentioned in Section 4.2.7. Because the molecular weights of the expressed proteins are low, it was necessary to use a tricine-based electrophoretic buffering system (see Section 2.2.9.1) which revealed the expected bands for the TolB and TolA boxes of 5 kDa and 12 kDa as illustrated in Figures 4.9 and 4.10, respectively.

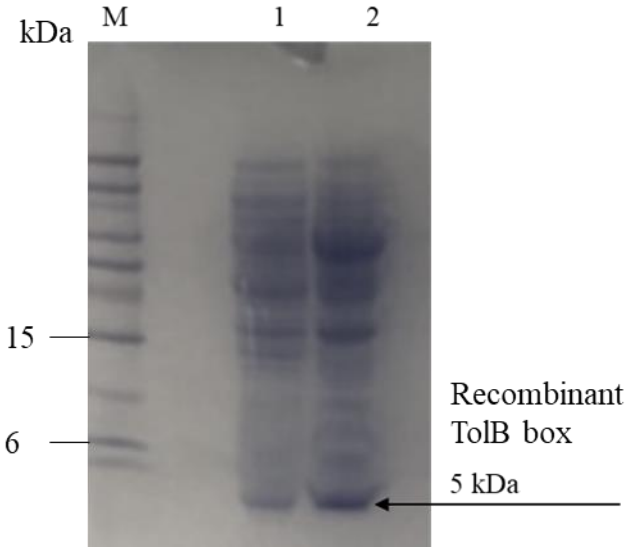


Figure 4.9 Tricine-SDS-PAGE analysis of isotopically labelled TolB box. M: Dual pre-stained protein marker (3.5-100 kDa); Lane 1: non-induced cells as negative control; Lane 2: 0.002% L-arabinose induced cells expressing isotopically labelled recombinant TolB box which showed the expected molecular weight of 5 kDa.

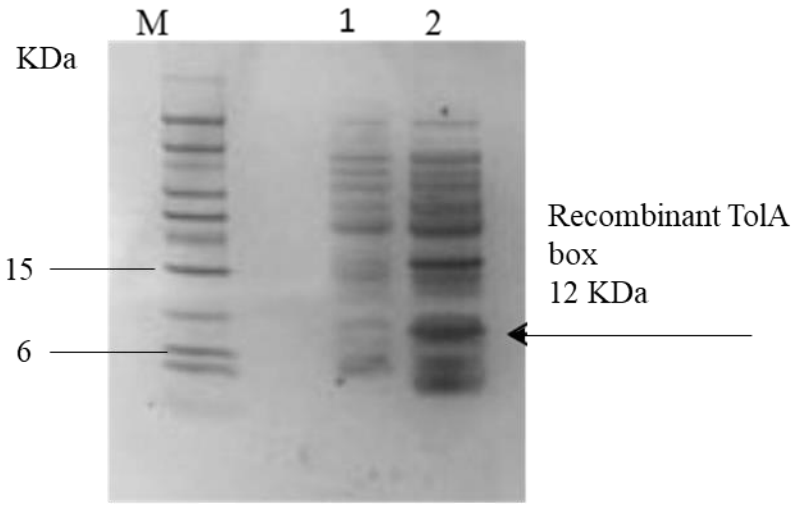


Figure 4.10 Tricine-SDS-PAGE analysis of isotopically labelled TolA box. M: Dual pre-stained protein marker 3.5-100 kDa. Lane 1: non-induced cells as negative control; Lane 2: 0.002 % L-arabinose induced cells expressing isotopically labelled recombinant TolA box which showed the expected molecular weight of 12 kDa.

4.3 Discussion

Two of the most popular techniques employed to study proteins are living cell imaging and in-cell nuclear magnetic resonance (NMR), both of which can provide rich information about macromolecules and their interactions. Both techniques require high intracellular levels of the protein(s) of interest and thus typically require the overexpression of recombinant proteins in *E. coli*. NMR-based studies also necessitate the use of isotope-labelled protein. The aim of the experiments described here was therefore, the optimisation of high-level recombinant protein overexpression, both cytoplasmically and periplasmically. To this end, this chapter describes the techniques used for optimising the overexpression of recombinant TA protein using two different expression systems, pET-15b and pBAD/gIIIc, and overexpression of the TolA and TolB boxes in pBAD/gIIIc.

One of the aims of this project to use in-cell NMR to visualise protein (s) of interest. The other aim of this project is to investigate protein- protein interactions (PPIs) in the periplasm, TolA and TolB boxes' overexpression were carried out in the periplasm only. Along with the investigation, overexpressing TA domain in both the cytoplasm and the periplasm was used to compare the results. In most of the previously conducted studies, researchers have investigated PPIs *in vitro* and *in vivo* in the cytoplasm. However, improper folding of many target proteins may occur during this process despite successful examples. In comparison, studies involving periplasmic localisation remained limited (Latifi *et al*, 2015). Given that the main subject of this study, the interactions between colicin and Tol proteins occur within the periplasm, it was therefore vital that a high-level expression of the proteins of interest in this environment is obtained.

Understanding the nature of the periplasmic space of Gram-negative bacteria helps in this optimisation process as in comparison to cytoplasmic expression, the periplasmic space provides a very different environment for protein production. The other reason behind using different vectors is to test which condition is best for in-cell NMR. The question is would molecular crowding affects in-cell NMR detection or not. In-cell NMR has been proved for its ability to detect folded proteins in the cytoplasm as proved by Pielak and colleagues (2009) and because testing TA by in-cell NMR approach has not been done before, this project attempts to visualize TA in the cytoplasm by this method. The periplasmic expression also enables to detect any changes to TA domain due to its known interactions with Tol system, components of which are present in this space.

Bacterial cytoplasm consists 25% proteins of the total volume. The complexity of the cytoplasm affects the structural and the dynamic properties of the proteins, which cause an increase in the thermodynamic activity, and proteins tend to fold (Smith *et al*., 2015). Cytoplasmic production also involved several advantages. There is no need for outer-membrane disruption to recover target proteins and avoid intracellular proteolysis by periplasmic proteases, leading to higher recombinant proteins production (Latifi *et al*, 2015), which worth to be visualised by in-cell NMR.

While the periplasmic secretion of *E. coli* contains other advantages such as enhanced product solubility and the ease of protein purification, although it is not needed in this study, for instance, the periplasmic space contains fewer proteins than the cytoplasm, and it is easily purified to conduct PPI analysis in the periplasm (Baneyx & Mujacic, 2004).

Additionally, potential substrates for PPI (e.g., colicins) are able to cross the outer membrane, unlike the cytoplasmic membrane (Bloois, 2012). However, periplasmic expression can be a challenging environment as a result of low secretion of recombinant proteins. Also, the level of recombinant proteins can be affected by proteolytic degradation (Latifi *et al.*, 2015).

In general, several pertinent factors influencing both cytoplasmic and periplasmic protein overexpression should be considered, including culture medium, growth phase and optical density at the time of induction, protein's molecular weight, and induction conditions. Although many researchers have observed that periplasmic protein expression is not significantly affected by the size of the protein, they have found that the composition of the culture medium can play a role. For example, protein expression in M9 minimal medium has yielded optimal periplasmic expression compared to rich medium (Bloois, 2012). Accordingly, M9 minimal medium is one of the chosen media for protein overexpression in this project.

The use of an appropriate promoter system for protein overexpression is another important factor for optimising protein production. For high-level protein synthesis, the promoter should allow the production of recombinant protein at more than 30% of the total cellular proteins. The promoter should easily induce and allow basal transcriptional activity (Makrides. 1996).

Other factors do however, need to be considered when attempting to optimise protein overexpression. For example, the nature of the starter culture (i.e., overnight culture) used. Using *E. coli* as the host expression strain takes advantage of its fast growth kinetics resulting in a 1/100 dilution of the starter culture reaching the stationary phase and a high final cell density in a few hours. The most common protocols for starting protein expression experiments involve using a diluted overnight culture typically, using a dilution factor of 1/100 into a large volume of fresh medium. After which, the inducer can be added at mid-log phase or at an early stationary phase, depending on the protein of interest (Berrow, 2006). And this protocol was applied in this project as explained in Section 2.2.7.

Protein expression in this project started with periplasmic expression using the pBAD/gIIIc vector. It is one of the most common commercially available vectors used for periplasmic protein expression. It is tightly controlled compared to other periplasmic expression systems, such as some of the relevant pET-series vectors. The control of expression levels, based on catabolite repression, makes the pBAD system ideal for producing high protein levels without leakage. Also, Guzman *et al*. (1995) identified that the use of *ara*BAD (PBAD) as a promoter using an optimised concentration of L-arabinose for induction could modulate protein overexpression over 2000-fold above the level of basal expression, greatly in excess of that typically achieved, for example, using the lacZYA operon in some other vector systems (Sommer *et al*., 2010). As the pBAD system is modulated by the amount of L-arabinose present in the culture, if a high level of recombinant protein is needed, then more L-arabinose can be added. Thus, the yield of the desired protein can be increased according to the amount of L-arabinose added to the culture (Schleif, 2010; Marschall *et al.*, 2017). Therefore, as described in Section 2.2.6, serial dilutions of 20% L-

arabinose were used (see Appendix B) to attempt to maximise protein overexpression. SDS-PAGE gel analysis showed the level of expression for each culture; Samples induced with a 0.002% final concentration of L-arabinose gave an optimal level of expression, with the protein of interest (recombinant TA) seen at the expected molecular weight of 17 kDa as shown in Figure 4.1, which is in agreement with the work of Benedetti and colleagues (1998). This amount of L-arabinose (i.e., 0.002% final concentration) was subsequently used for all periplasmic expression experiments using the pBAD/gIIIc vector described in this chapter. In general, the expression of the gene of interest is at very low basal levels in the absence of L-arabinose and is repressed in the presence of glucose (Guzman, 1995). However, this was not the case in this project; even with glucose added to the media (see Chapter 2), some expression was still seen with non-induced cells on SDS-PAGE.

As described in Chapter 1, one of the aims of the project is the production of high level of recombinant proteins. The first approach to achieve this aim is conducting protein overexpression of TA in both chosen vectors (i.e., pBAD/gIIIc & pET-15b) in LB medium. LB is the most common medium for culturing *E. coli*. It is easy to make, it has rich nutrient contents, and its osmolarity is optimal for growth at the early log phase. It is ideal for protein production (Sezonov *et al*., 2007).

The experiments were conducted as described in Chapter 2 (Section 2.2.7.1). After standardising the gel loading of the samples after the total protein concentration was determined as described in section 2.2.8. SDS- PAGE analysis was then used to resolve the samples. Bands were observed at the expected molecular weight of 17 kDa for TA protein, indicating the expression of the target protein as shown in Figure 4.2 for cytoplasmic recombinant TA, and in Figure 4.3 for periplasmic recombinant TA.

Because efficient labelling is required for in-cell NMR, there is a need for isotopic enrichment media. M9 minimal medium was chosen because a minimal medium, such as M9, + glucose (see Appendix B), is required to tightly regulate expression under reduced growth rate conditions. It is a defined growth medium that consists of a number of essential salts but excludes a carbon source. The buffering components in M9 salts maintain the pH at optimal for *E. coli* growth. Traces of iron, in addition to amino acids, such as casamino acid, were also added to the medium as suggested by Mondal and colleagues (2013) as this can increase the growth rate as well as final cell density. The reduced growth rate in M9 minimal medium is thought by some to researchers be more advantageous for production, for correct folding and most importantly, for incorporating specific residues into the protein (s) of interest (e.g., ^{13}C amino acids) for structural studies or in-cell NMR (Cai *et al*., 2016). Tighter control of expression can be achieved by the addition of 0.2–1% w/v glucose to the medium as rich media prepared with tryptone or peptone may contain the inducer lactose (Rosano & ceccarelli, 2014). The addition of glucose to the growth medium can suppress basal expression due to a reduction in cellular cAMP levels. In a glucose-free medium, such as M9, cAMP levels are high, and a cAMP-CRP (catabolite activator protein) complex bind to the pBAD promoter. This association is required for promoter activity, so the addition of glucose will robustly repress expression of the gene of interest (Guzman, 1995).

Initial experiments were conducted in M9 minimal medium supplemented with glucose and produced low levels of total protein (1 mg/mL, as determined by Bradford assay). The cultures were also slow growing, the desired optical density being achieved only after overnight incubation. SDS- PAGE analysis revealed low level of expression at the expected molecular weight 17 kDa for TA protein, as shown in Figure 4.4. The low growth rate

might be due to basal expression of the recombinant protein prior to induction, even in the presence of glucose. In order to overcome this issue, the culture medium selection was reconsidered. Sivashanmugam and colleagues (2009) suggested that growing the culture in a medium that provides low growth rates, such as M9 minimal medium, can result in weak expression unless it is used as an induction medium only, and the starter culture is subsequently transferred to a rich medium for induction. Sivashanmugam and colleagues (2009) suggested the reason for this is that the minimal medium only supports a low growth rate. Therefore, during induction, the labelled ^{13}glucose and $^{15}NH_4Cl$ might have been used more for cellular reproduction rather than labelling of the protein(s) of interest. This can reduce the efficiency of incorporation of isotopic labelled atoms (e.g., ^{13}C and ^{15}N). Accordingly, the next attempt at high level expression of TA in both the periplasm and the cytoplasm was as described in Section 4.2.5. In this experiment the host cells were cultured in LB medium with the aim of increasing the cell density and consequently the yield of the target protein. The culture was then switched to the labelled M9 minimal medium at the time of induction. Figure 4.5 shows the SDS- PAGE gel results. The expression level was slightly improved and the total protein concentration was 15 mg/mL, which is still relatively low level of expression that could be due to many reasons. First, the translocation of recombinant proteins into the *E. coli* periplasmic compartment is limited by the size of the periplasmic space and by the export capacity of the cell (The periplasmic compartment accounts for less than 20% of the total cell volume).

Furthermore, it is not surprising that an excessive *E. coli* stress response induced by protein overexpression could generate increased demand for protein folding and result in an increased metabolic burden on the cells (Sandomenico *et al.*, 2020). As for cytoplasmic expression, normally, proteins are able to fold correctly spontaneously. However, aggregation of overexpressed recombinant proteins in bacterial cells could result from the accumulation of high concentrations of folding intermediates or from inefficient processing by molecular chaperones (Sørensen & Mortensen, 2005). This similarity in the results between the cytoplasmic and periplasmic expression of isotopic labelled TA might be due to the fact that TA is an intrinsically disordered protein, and steric factors could result in incomplete folding, as suggested by Dunker and colleagues 2002, especially if that protein is involved in critical cellular mechanism as TA is likely to be.

To overcome the potential issue of increased metabolic stress that could be placed on the cells, another dynamic that merits attention for optimization of protein overexpression is using a low temperature during the induction for a long period, which can help to generate a high yield of proteins (Mondal *et al*., 2013). Slower rates of protein production give newly transcribed recombinant proteins more time to fold properly. Moreover, the most commonly used way to lower the rate of protein synthesis by reducing incubation temperature ((Rosano & Ceccarelli. 2014)). Low temperature reduces the risk of culture overgrowth, plasmid degradation, and antibiotic inactivation. Accordingly, the following experiment for TA expression in both the periplasm and the cytoplasm was carried out essentially as described in Section 4.2.5, but with modified induction conditions as mentioned in Section 4.2.6. Induction was carried out overnight at 25°C, but this still did not improve the yield of total protein and SDS- PAGE analysis revealed the same low level

of protein expression as illustrated in Figure 4.6. This outcome may be due to working at the lower end of the temperature range, as slower growth and reduced synthesis rates can result in lower protein yields, which negates the potential benefits in terms of improved protein folding.

Another factor considered was the optical density of the cell culture to be induced. The yield of the desired protein can be increased by growing the culture to high densities. In general, the time of induction is usually when the bacterial cell culture reaches the exponential phase (i.e., OD_{600} 0.6-0.8). However, when the cells expressing misfolded proteins inclusion bodies could be formed, which might compromise cell function or viability. Also, if the induction is conducted too soon after the inoculation and thus at low cell densities, overexpressed recombinant proteins could inhibit bacterial growth thereby resulting in low yield (Rosano & Ceccareli., 2014). The experiment described in Section 4.2.5 was subsequently modified by growing cultures to an OD_{600} of 1.0 before induction, and this was identified as the best approach for optimising protein overexpression. SDS-PAGE analysis of TA expression from constructs using both pET-15b and pBAD/gIIIc revealed a band at the expected molecular weight of 17 kDa for TA protein and indicated a higher level of expression. Figure 4.7 shows that the use of a higher OD at time of induction is superior as cells in stationary phase have reached a higher bacterial cell densities that result in higher level of protein's overexpression. This condition was optimal and was used in the project.

Immunoblotting to detect recombinant TA and to validate the results was carried out along with the expressed TA in LB medium to confirm identity as shown in Figure 4.8. The C-terminal poly-histidine regions for both vectors (i.e., pBAD/gIIIc & pET-15b) are illustrated in Chapter 2. They form metal-binding site for affinity purification of recombinant fusion protein on metal-chelating resins and also allow the detection of the recombinant protein with Anti-His (C-term) antibody, as described in Section 2.2.10.

Conditions for protein overexpression as described in section 4.2.7, using an OD_{600} of 1.0 at point of induction was similarly used on cells carrying the TolA box and Tol B box plasmids. A different buffer system for SDS-PAGE analysis of protein overexpression was used because of their small molecular weights (Schagger, 2006) has demonstrated a tricine-SDS-PAGE protocol that improves the separation and resolution of small proteins. The method used for tricine-SDS-PAGE electrophoresis is given in Section 2.2.9.1, which proved valuable for the analysis of proteins smaller than 30 kDa. The low concentrations of acrylamide used in this protocol had the additional advantage of enhancing protein blotting and staining. Tricine-SDS-PAGE analysis revealed the presence of the TolB box protein at the expected molecular weight of 5 kDa, and TolA box protein at approximately 12 kDa with an acceptable level of expression, as shown in Figures 4.9 and 4.10, respectively.

Chapter 5

Analysis of colicin E9 effect on cells expressing residues of colicin A translocation domain using confocal laser scanning microscopy

5.1 Introduction

The aims of this chapter are as follows; to investigate the effect of Colicin E9 (ColE9) against *E. coli*, specifically the ColE9-sensitive strain BL21(DE3). Secondly, to analyse the effect of ColE9 on cells which overexpress TA protein—that is, the full-length translocation domain, and the effect on cells overexpress TolA and TolB box domains (i.e., sub-regions of TA) as detailed in Chapters 3 and 4, regarding how they were cloned and how protein overexpression was optimised.

The overarching goal of the two aims is to determine whether the presence of those domains within the periplasm of *E. coli* BL21(DE3) cells provides any protection against ColE9, whether by interacting with cellular Tol proteins or, by competitively inhibiting interaction. To achieve those aims, live-cell imaging with confocal laser scanning microscopy (CLSM) was performed.

ColE9 is co-expressed with immunity protein 9 (Imm9), forming ColE9-Imm9 complex, which binds to the C-terminal region of colicins (Section 1.5). In extensive studies on the ColE9 N-terminal region, Housden and colleagues (2010, 2013, 2018) have observed that the region is intrinsically disordered i.e., that it lacks an ordered three-dimensional (3D) structure, and that it consists of two outer membrane protein F (OmpF)-binding sites: OmpF-binding site 1 (OBS1), where ColE9 binds to OmpF porins in the bacterial OM; and OmpF-binding site 2 (OBS2), where it binds to, or interacts with, a periplasmic TolB protein (Housden *et al*., 2010). The organisation of those binding sites is illustrated in Figure 5.1.

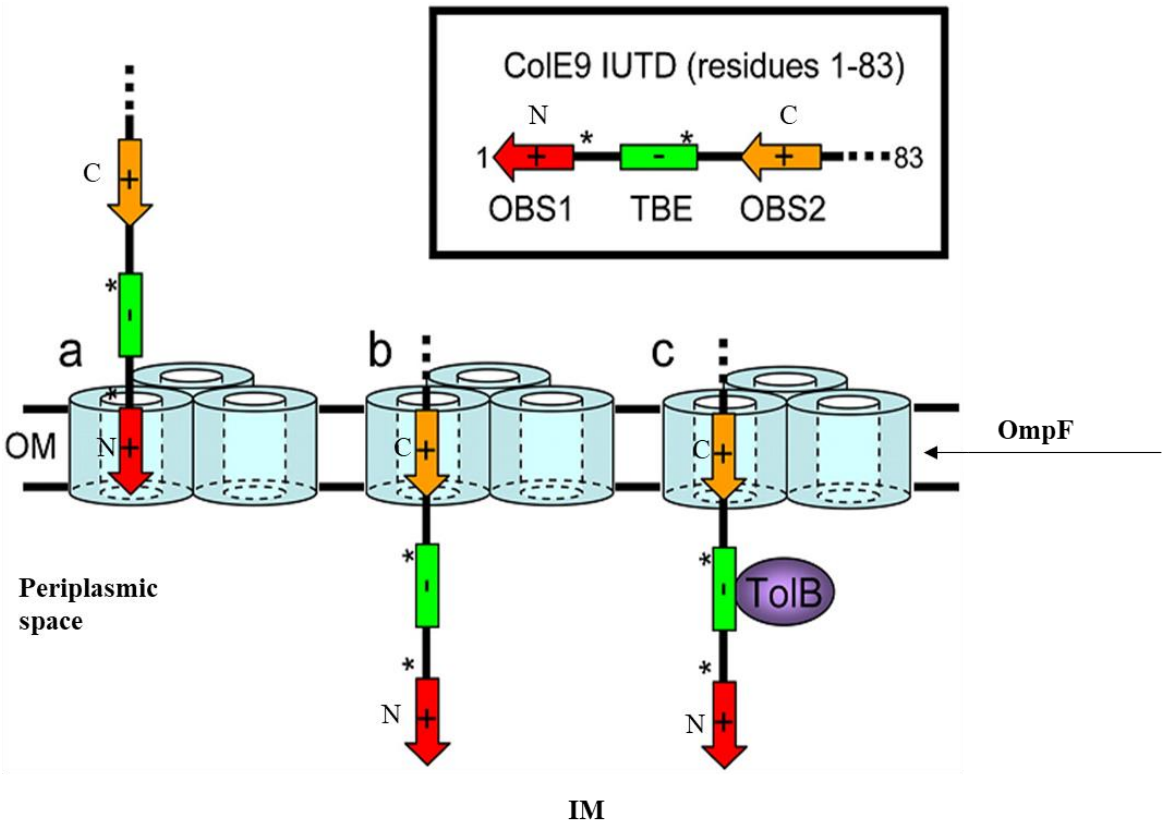


Figure 5.1 Porin binding of the intrinsically unstructured (T) domain (IUTD) of ColE9 (adapted from Housden *et al.*, 2010). a) ColE9 N-terminal residue binds to the extracellular side of OmpF (outer membrane porin F) at the positively charged OBS1 (OmpF binding site 1). b) whereas ColE9 C-terminal binds to the periplasmic side of the positively charged OmpF at OBS2 (OmpF binding site 2) followed by the diffusion of the ColE9–OmpF complex is directed through the pore to the periplasmic space. c) the negatively charged ColE9–OmpF complex interacts with TolB at the TBE (TolB binding epitope) binding site and, in turn, translocate into the IM (inner membrane).

ColE9-Imm9 uses its disordered structure to engage the Tol–Pal system in the periplasm, yet requires the proton-motive force (PMF) to provide energy for transport across the periplasm via OmpF in the OM to the inner membrane (IM) to reach the target. In general, the ColE9 translocation mechanism becomes activated when ColE9 enters the OM of Gram-negative bacteria. Despite providing a strong protective barrier against hydrophilic and hydrophobic compounds alike, the OM contains small porins that facilitate the exchange of metabolites and the transport of nutrients required by bacteria. Although bacterial cells cannot uptake

molecules larger than (600-800 Da), colicins that target sensitive bacteria can nevertheless penetrate the OM with the help of Tol proteins (Nikaido, 1994). As the ColE9 is co-expressed with Imm9, the most important factor for ColE9 translocation is its interaction with Tol proteins, which disrupts the membrane permeability, and creates energy used to dissociate the ColE9–Imm9 complex. According to Housden and colleagues (2018), the dissociation of the ColE9–Imm9 complex allows ColE9 to enter the OM and travel through the periplasmic space to the IM, where it exerts its activity, confirming the results achieved by James and colleagues (2002) shown in Figure 5.2.

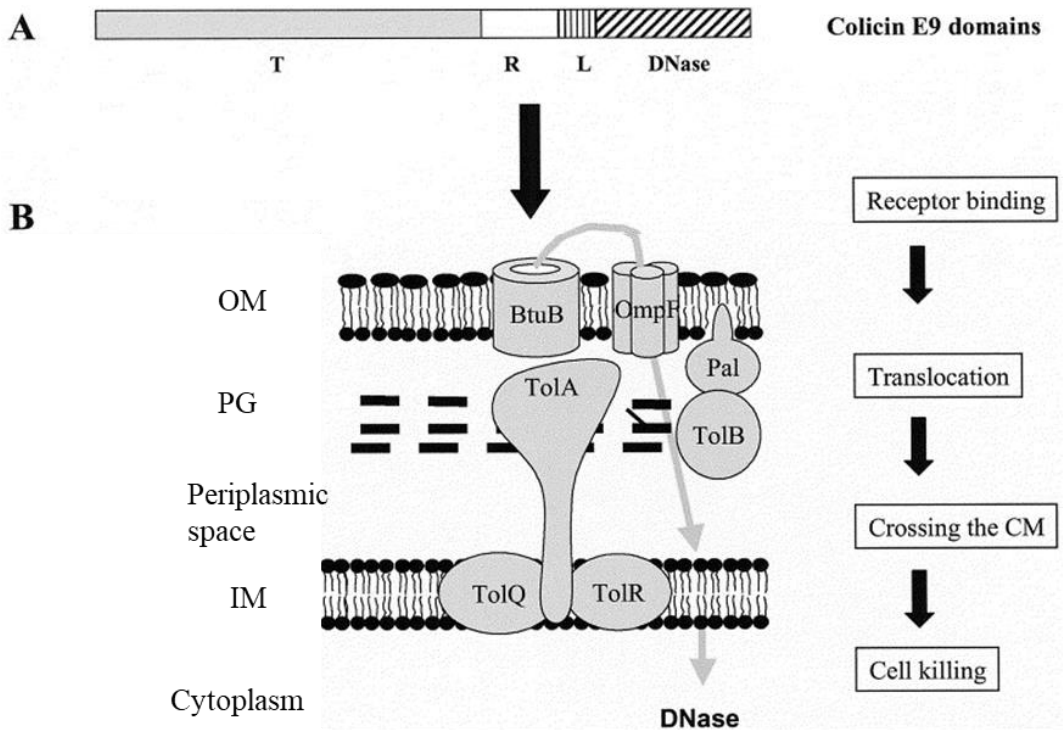


Figure 5.2 ColE9 mechanism of action (James *et al.*, 2002). A) ColE9 has three structural domains: The translocation (T) domain, the receptor binding (R) domain is connected to the deoxyribonuclease (i.e., C) domain by a linker region (L). B) Along with the binding of ColE9 (R) domain to BtuB and OmpF porin, the (T) domain enters the periplasmic space through binding to TolB-Pal complex in the peptidoglycan. The presence of Tol–AQR in the inner membrane (IM) facilitates the energy transfer to the OM (outer membrane) and the activation of the deoxyribonuclease domain's cytotoxic activity against DNA.

After identifying the mechanism of ColE9 translocation through the cells, the role of Tol proteins has been studied. Vanemmelbeke and colleagues (2009) used a deletion mutation study to show that in bacterial strains such as *E. coli* that lack cytoplasmic Tol proteins, especially TolA, the release of Imm9 from ColE9-Imm9 complex is compromised, which confirmed the proposed indirect interaction between TolA and ColE9 and the TolA role in the dissociation of the ColE9–Imm9 complex before reaching the target. They additionally proposed that the energy created by the TolA–TolB interaction also allows the release of Imm9 before the translocation of ColE9, and that proves how TolA is involved in the translocation process of ColE9 (Vankemmelbeke *et al*., 2009; Garinot-Schneider *et al*., 1997).

Other researchers have also investigated the displacement of the cognate Imm9 from ColE9. Collins and colleagues (2002), for instance, who examined the dissociation of the ColE9–Imm9 complex, observed that the energy transduced from the IM is crucial for ColE9 translocation in sensitive cells and, in turn, its cytotoxic effect. The authors added that because ColE9 is an energy-dependent colicin, with energy produced in the IM and transduced to the OM, TolA could be the driving force for Imm9 release by being active without any direct interaction with ColE9. Also, they demonstrated that when ColE9 C-terminal domain translocate into the periplasm, the (R) and (T) domains remain bound to OM and periplasmic proteins to complete the translocation process (Collins *et al*., 2002).

In addition to TolA, the other IM Tols (i.e., TolQ and TolR) have been studied recently. Rassam and colleagues (2018) have demonstrated that ColE9 entry into cells and binding to Tol–Pal proteins affects the mobility of Tol proteins in the IM and causes an indirect interaction between the IM Tol proteins, TolQ, and TolR, and the Tol–Pal–ColE9 complex as well. That binding process clarifies not only ColE9 translocation but also the way in which protein complexes in the IM and OM become involved in the interaction process, which produces energy that facilitates translocation of ColE9. The authors suggested that proteins

in the IM reflect the roles of those in the OM. And because the mobility of proteins in the OM is restricted to the OM, whereas those in the IM diffuse throughout the cytoplasm, visualising colicin's entry and translocation as well as the conformational changes that occur upon its entry until reaching the target can elucidate the colicin's mechanism of action and its potential use in discovering novel antibiotics (Rassam *et al*., 2018).

Unfortunately, there are some gaps in the knowledge about the functionality of the cellular Tol proteins with overproduced colicin domain residues in the periplasm when the cell is treated with an external colicin. Bouveret and colleagues (2002) investigated the interactions of colicins with Tol proteins *in vivo*, namely by overexpressing the (T) domain of ColA (TA) fused to G3p in the periplasm, in order to determine whether their interaction disrupts the function of the Tol–Pal system, as discussed in Chapter 1. The overexpression of TA has been shown to cause cells to resist ColA activity. The possible explanation for the resistance is that colicins' domains which are produced in the periplasm, should competitively inhibit external colicin's entry into the cells with the aid of Tol- Pal system. However, the role of the Tol–Pal system in colicin's translocation, in general, has not been fully elucidated. Although Bouveret and colleagues (2002) observed that any mutation in the Tol-Pal system affects the OM integrity and that the system aids colicin in reaching its target, it remains uncertain whether overproduced periplasmic proteins can generate the same effect as the (T) domain in terms of binding to Tol proteins, and, if this is the case, could this effect be measured. A further question is whether the presence of colicin domains in the periplasm of *E. coli* cells provides any protection against ColE9. Examining the effect of ColE9 cytotoxic effect on *E. coli* cells is essential, particularly whether TA domain in the periplasm protects the cells against ColE9.

Indeed, the results obtained from live-cell imaging with CLSM could help to quantify the effect of ColE9 and to determine the level of protection against it. This will be investigated in this chapter, helping to close gaps in how the translocation of colicins is currently understood.

As discussed in Chapter 1, live-cell imaging can serve as a rich source of information that does not require disrupting basic cellular functions. Initially developed from basic light microscopy, live-cell imaging later led to the creation of staining techniques used in florescence microscopy and CLSM (Schneider & Basler, 2016). Of those two methods, CLSM eliminates out-of-focus light by using a pinhole that illuminates a small volume of the tested specimen called a "confocal spot". CLSM produces light at narrow-wavelength bands, most commonly by an argon laser, as used in the research for this thesis, which causes excitation at 488 nm (Takeuchi & Frank, 2001). Generally, the choice of laser depends upon the excitation and emission wavelengths of the dyes used (Sanderson *et al*., 2014). After scanning the confocal spot with a laser for illumination, the image can be relayed to a computer for analysis and for the quantification of bacterial cells (Raarup & Nyengaard, 2001), and this is the approach for this chapter.

5.2 ColE9-Imm9 antimicrobial activity on *E. coli* BL21(DE3) cells expressing recombinant TA and TolA and TolB boxes

To test the cytotoxic effect of ColE9 (nuclease colicin) against *E. coli* BL21(DE3) cells, a ColE9–Imm9 mutant with the same cytotoxic activity as the wild type was provided by Dr Chris Penfold at the University of Nottingham. According to Dr Penfold, the construct (pColE9) was engineered by his team to ease the overexpression and purification of ColE9–Imm9 (see Appendix A). In this project, starting with the overexpression and the purification of ColE9-Imm9, its antimicrobial activity against colicin-sensitive cells *E. coli* BL21(DE3) was tested, and plate counting as well as bacterial growth monitoring were performed after treatment of *E. coli* BL21(DE3) cells with ColE9–Imm9. Bacterial cell viability was determined by using CLSM with a Live/Dead® BacLight™ viability kit consisting of a universal stain, SYTO 9, which internalises within both living and dead cells, and propidium iodide (PI), which penetrates only dead cells and displaces SYTO 9. Using two contrasting colours—SYTO 9 green for living cells, and PI red for dead ones—helped to distinguish and count cells (Rosenberg *et al*., 2019). The dual staining results visualised by using CLSM were quantified in terms of live versus dead cells and the results were compared.

5.3 Results

In this chapter, the results were demonstrated testing ColE9-Imm9 in the presence or absence of dithiothreitol (DTT) against different cells. DTT reduces the disulphide linkage of the ColE9-Imm9, and thus its effect was tested. To test all of the hypotheses, First, the data obtained in this chapter were tested for normality by Shapiro Wilk test. The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05, and because the data are not normally distributed. Secondly, testing the samples if they have equal variances, is called homogeneity of variance. Levene test was used to determine if the variances are equal for all samples, which they are not equal. Accordingly, Kruskal Wallis test was used for non-

parametric data to determine the difference significance (p-value). All results and descriptive statistical data are detailed in Appendix C.

5.3.1 Overexpression of ColE9–Imm9 in *E. coli* BL21(DE3) cells

ColE9–Imm9 was overexpressed in *E. coli* BL21(DE3) cells as explained in section 2.2.7, then the sample was analysed by SDS-PAGE as described in Section 2.2.9. Figure 5.3 shows the proteins expressed by IPTG-induced *E. coli* BL21(DE3) cells illustrating the band at 60 kDa expected for ColE9 and the band at 10 kDa expected for Imm9 on Lane 1. In comparison, Lane 2 indicates low level of expression from non-induced cells.

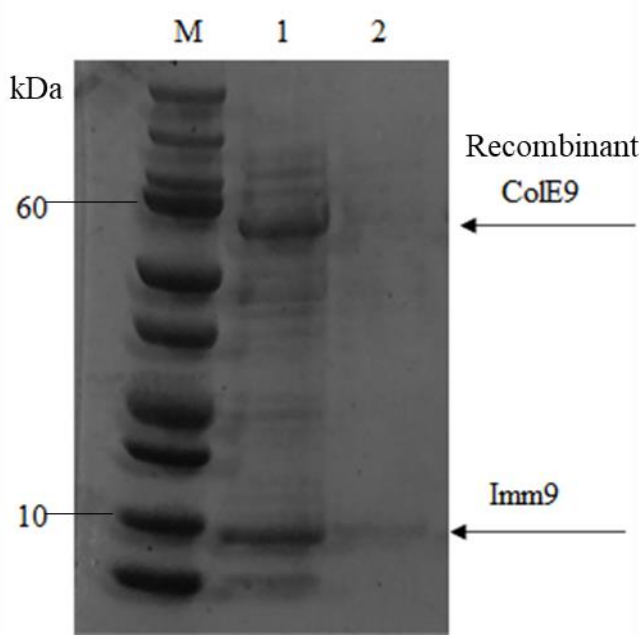


Figure 5.3 Coomassie blue stained SDS-PAGE analysis of recombinant ColE9–Imm9. M: Pre-stained protein marker 3.5-100 kDa. Lane 1: recombinant ColE9–Imm9 in *E. coli* BL21(DE3) cells, at the expected molecular weights: ColE9 at 60 kDa and Imm9 at 10 kDa. Lane 2: non-induced cells which shows basal level of expression as the negative control for the over-expression.

5.3.2 Recombinant protein purification of ColE9–Imm9 complex under native condition

As described in Section 2.2.11, the ColE9-Imm9 was purified and dialysed overnight, and samples were collected for analysis with SDS-PAGE, as described in Section 2.2.9. Figure 5.4 illustrates the expected band for ColE9 at 60 kDa and the expected band for Imm9 at 10 kDa for those samples obtained under different preparation conditions; following IPTG-induction, after purification, after overnight dialysis and following concentration.

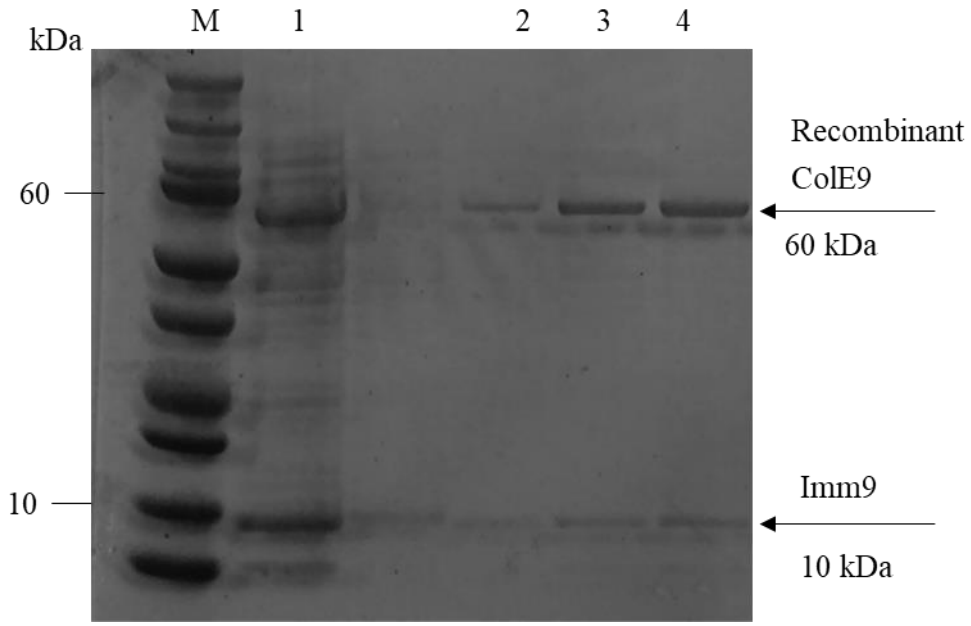


Figure 5.4 Coomassie blue stained SDS-PAGE analysis of recombinant ColE9–Imm9 complex after purification. M: Pre-stained protein marker 3.5-100 kDa. Lane 1: recombinant ColE9–Imm9 in *E. coli* BL21(DE3) cells. Lane 2: recombinant ColE9–Imm9 after purification under native conditions. Lane 3: recombinant ColE9–Imm9 after purification followed by overnight dialysis. Lane 4: recombinant ColE9–Imm9 after purification followed by protein concentration using (Pierce Slide-A-Lyzer™ Dialysis Cassettes® 3.5K MWCO). Both parts of the complex showed the expected bands: ColE9 at 60 kDa and Imm9 at 10 kDa in all lanes.

5.3.3 Analysis of the antimicrobial activity of the purified ColE9–Imm9 complex against *E. coli* BL21(DE3)/ *E. coli* BL21(DE3) cells expressing TA/ TolA box /TolB box using spot test

A spot analysis of the ColE9–Imm9 complex against *E. coli* BL21(DE3) cells was performed as described in Section 2.2.12. The results revealed zones of inhibition formed within the bacterial lawn due to the action of the ColE9-Imm9 complex produced under different conditions, as illustrated in Figure 5.5. No inhibition of the bacterial cells was observed with the negative control (PBS). While the other constructs pTA, pTolA and pTolB resulted no inhibition of growth in the agar plate.

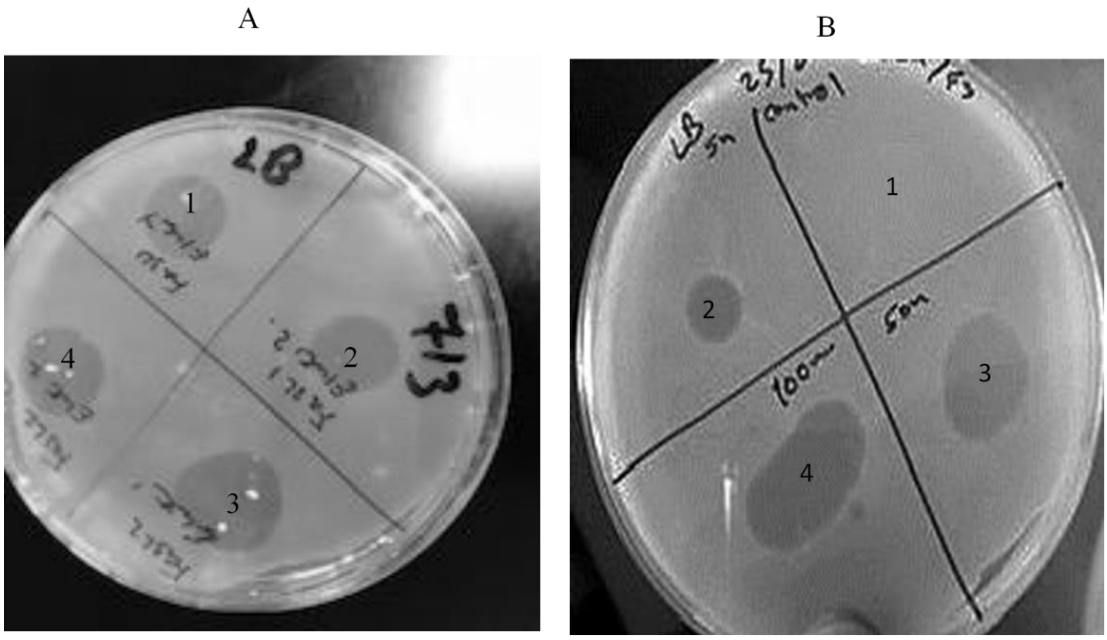


Figure 5.5 Spot analysis of ColE9–Imm9 complex prepared by various means (10 µL) added onto a lawn of *E. coli* BL21(DE3) cells grown on LB agar. A) 1 and 2- are duplicates which show zones of inhibition resulting from spotting recombinant ColE9-Imm9; 3 and 4- are duplicates which show zones of inhibition resulting from spotting recombinant ColE9–Imm9 after purification followed by overnight dialysis and protein concentration. B) zones of inhibition resulting from spotting recombinant ColE9–Imm9 after purification followed by overnight dialysis at different concentrations; 1 - negative control (PBS, no ColE9-Imm9); 2, 3 and 4 – show zones of inhibition resulting from spotting recombinant ColE9–Imm9 at 5, 50 and 100 nM, respectively.

5.3.4 Analysis of antimicrobial activity of purified ColE9–Imm9 complex against *E. coli* BL21(DE3) cells in liquid culture

The colicin-sensitive BL21(DE3) strain was assayed in triplicate to test the antimicrobial activity of ColE9-Imm9 in liquid culture, as determined using optical density measurements, as described in Section 2.2.13. Figure 5.6 shows a graph plotting the optical density (OD) of *E. coli* BL21(DE3) cells treated with ColE9–Imm9 (10 or 100 nM) over time, compared to untreated *E. coli* BL21(DE3) cells (i.e., negative control, bacterial cells only). All treatments showed approximately the same increase in the optical density over 12 h.

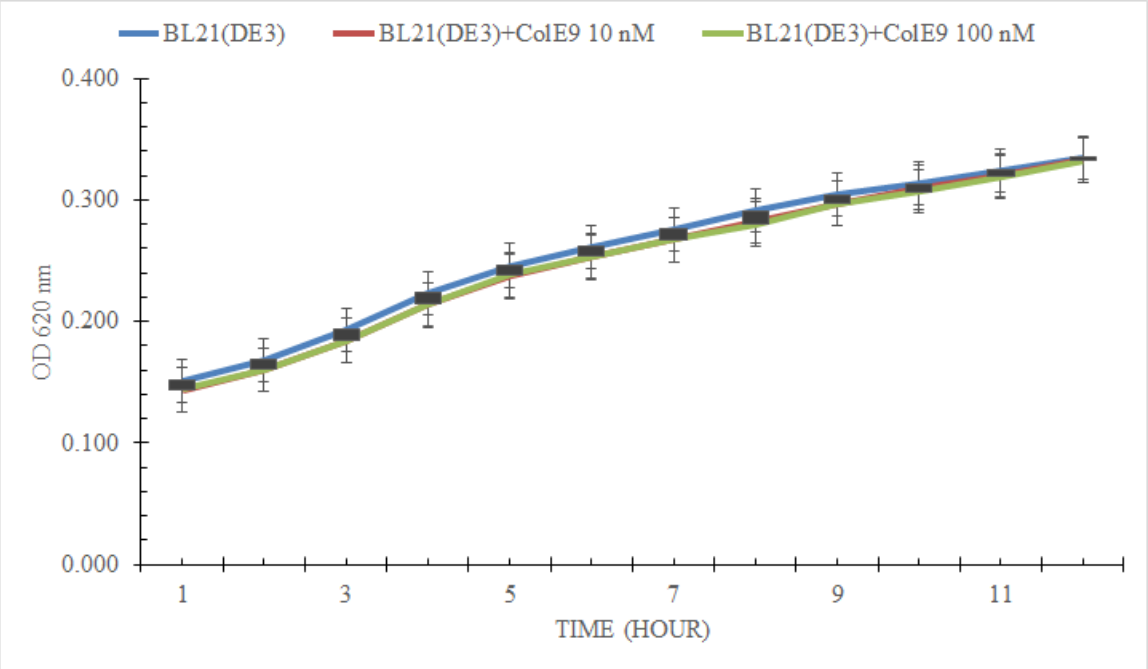


Figure 5.6 The antimicrobial activity of ColE9–Imm9 complex on the growth of *E. coli* BL21(DE3) cells in liquid culture. The graph shows no change in the optical density of *E. coli* BL21(DE3) cells when they are treated with 10 or100 nM of ColE9 compared to the optical density of untreated *E. coli* BL21(DE3) cells (n = 3; error bars represent the standard error of the mean).

5.3.5 Analysis of the antimicrobial activity of purified ColE9–Imm9 complex against *E. coli* BL21(DE3) cells in the presence of DTT in liquid culture

In order to reduce the disulphide linkage of ColE9-Imm9, the assay was performed as described in Section 2.2.13, with the following modifications. Immediately before treating the *E. coli* BL21(DE3) cells with ColE9–Imm9 (10 or100 nM) at room temperature, 1 M of DTT was added to the ColE9-Imm9 sample resulting in a final concentration of 10 mM of DTT. Figure 5.7 presents a graph of the mean optical density for *E. coli* BL21(DE3) cells treated with ColE9–Imm9, with and without DTT. The control group consisted of untreated *E. coli* BL21(DE3) cells, BL21(DE3) cells treated with 10 nM ColE9–Imm9 (BL21(DE3)/10) and 100 nM ColE9-Imm9 (BL21(DE3)/100), whereas the experimental group consisted of *E.*

coli BL21(DE3) cells treated with ColE9-Imm9 the same as the control plus DTT. In the control group, the highest mean optical density (OD_{620} nm) was 0.301, which occurred with the untreated *E. coli* BL21(DE3) cells, followed by 0.297 in the BL21(DE3)/10 group and 0.297 in the BL21(DE3)/100 group. After the addition of DTT, the highest mean optical density was 0.161, which occurred with the *E. coli* BL21(DE3) cells with DTT but no ColE9-Imm9 complex, followed by 0.145 in the BL21(DE3)/10 group with DTT and 0.122 in the BL21(DE3)/100 group with DTT. The reduction in the final optical density observed demonstrates that cell growth was affected by ColE9–Imm9 plus DTT.

Although the significant difference shown with the mean values, taking into account small sample (n=3) and violation of assumptions was tested with Levene test. It revealed ($p<<0.001$) which variances are not equal. The Shapiro-Wilk test was employed to determine whether the data are normally distributed, which revealed ($p<<0.001$) that the data are not normally distributed. As such, a non-parametric Kruskal-Wallis test was used to test for significant differences which revealed that there is no statistically significant difference between mean rank values of BL21(DE3) cells groups ($T(2) = 4.511$, $p=0.105$).

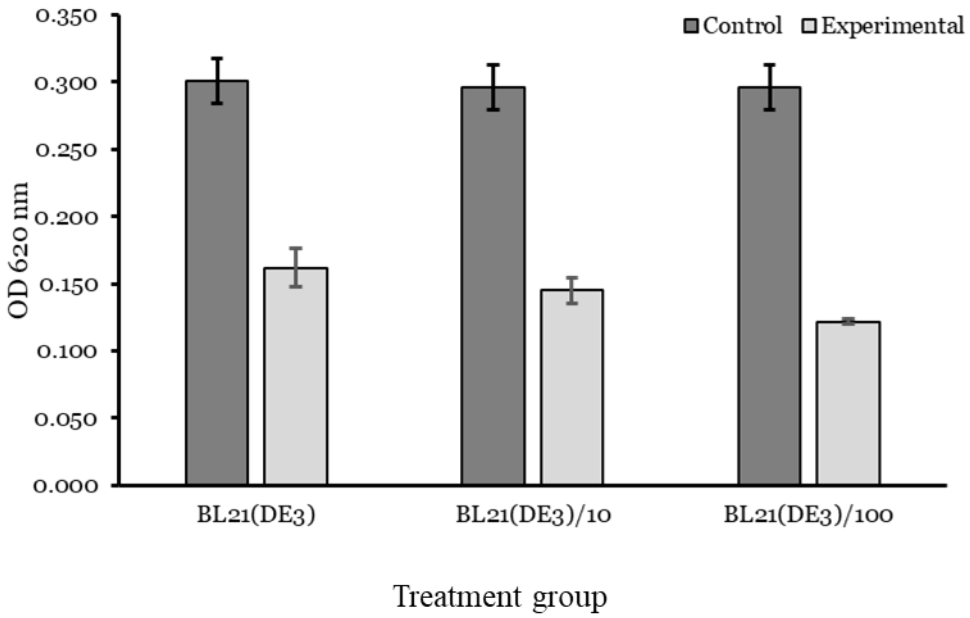


Figure 5.7 The antimicrobial activity of ColE9–Imm9 (control), and ColE9-Imm9 plus DTT (experimental) on the growth of *E. coli* BL21(DE3) cells in liquid culture. The graph shows the mean optical densities for the control group of *E. coli* BL21(DE3) cells treated with different concentrations of ColE9–Imm9; 10 nM (BL21(DE3)/10) and 100 nM (BL21(DE3)/100), and the experimental group of BL21(DE3) cells treated with different concentrations of ColE9–Imm9 (10/100 nM) plus DTT. The experimental group shows that treating *E. coli* BL21(DE3) cells with ColE9–Imm9 plus DTT at different concentrations resulted in no significant decrease in the final optical density of BL21(DE3)/10 and BL21(DE3)/100 as well as untreated cells of the experimental BL21(DE3) plus DTT compared to the control group. (n = 3; error bars represent the standard error of the mean).

5.3.6 Analysis of the antimicrobial activity of ColE9–Imm9 complex plus DTT against *E. coli* BL21(DE3) cells expressing TA, TolA box and TolB box in liquid culture

The antimicrobial activity of ColE9–Imm9 against *E. coli* BL21(DE3) expressing TA, TolA box and TolB box in the presence of DTT in liquid medium was tested as described in Section 2.2.14. Results presented in Figure 5.8 pinpointed the effect of ColE9–Imm9 plus DTT against *E. coli* BL21(DE3) cells expressing the TA, TolA box and TolB box compared to *E. coli* BL21(DE3) (D/C) treated cells with ColE9-Imm9 plus DTT as a positive control for the

antimicrobial activity of ColE9-Imm9 plus DTT. Figure 5.8 illustrates the findings for the TA group (*E. coli* BL21(DE3) cells expressing TA). The mean optical cell density (OD_{620} nm) was 0.334 for TA/C cells (i.e., untreated TA), 0.331 for the TA/10 cells (i.e., TA cells treated with 10 nM ColE9-Imm9 plus DTT) and 0.351 for the TA/100 cells (i.e., TA cells treated with 100 nM ColE9-Imm9 plus DTT) the last of which was higher than the control. When compared to *E. coli* BL21(DE3) cells treated with ColE9-Imm9 plus DTT(D/C), after testing the normality and homogeneity of variances (see Appendix C), the non-parametric Kruskal-Wallis test revealed here is a significant difference ($p<<0.001$) between the rank values of TA cells groups and the control cells (D/C).

For the TolA group (*E. coli* BL21(DE3) expressing TolA box), the mean optical density was 0.344 in the TolA/C group (i.e., untreated TolA cells), 0.342 in the TolA/10 group (i.e., TolA cells treated with 10 nM ColE9-Imm9 plus DTT) and 0.337 in the TolA/100 group (i.e., TolA cells treated with 100 nM ColE9-Imm9 plus DTT), the last of which was slightly lower than both the TolA/C and TolA/10 groups. After testing the normality and homogeneity of variances (see Appendix C), the non-parametric Kruskal-Wallis test revealed there is a significant difference ($p<<0.001$) between rank mean values for TolA groups and D/C control cells.

Lastly, for the TolB group, the optical density was 0.345 in the TolB/C group (i.e., untreated TolB cells), 0.360 in the TolB/10 group (i.e., TolB cells treated with 10 nM ColE9-Imm9 plus DTT) and 0.373 in the TolB/100 group (i.e., TolB cells treated with 100 nM ColE9-Imm9 plus DTT). After testing the normality and homogeneity of variances (see Appendix C), the non-parametric Kruskal-Wallis test revealed there is a significant difference ($p<0.001$) between rank mean values of ToIB groups and D/C control cells.

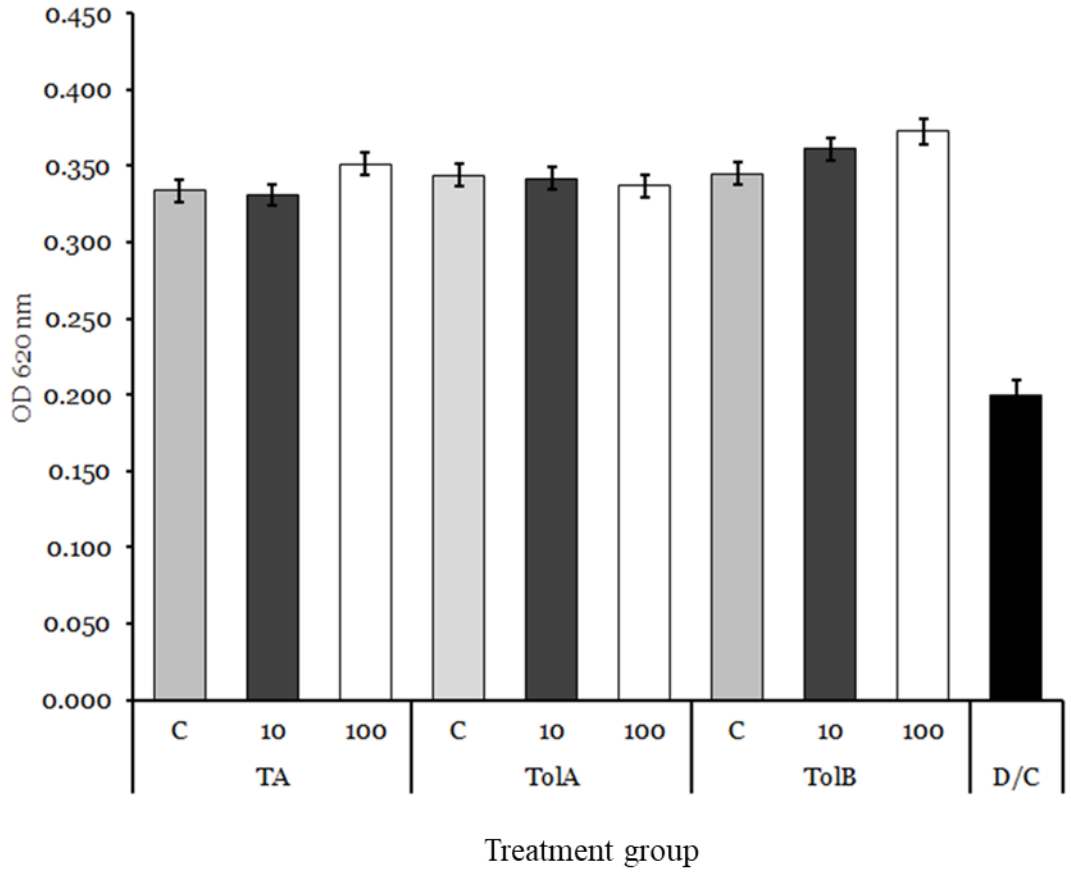
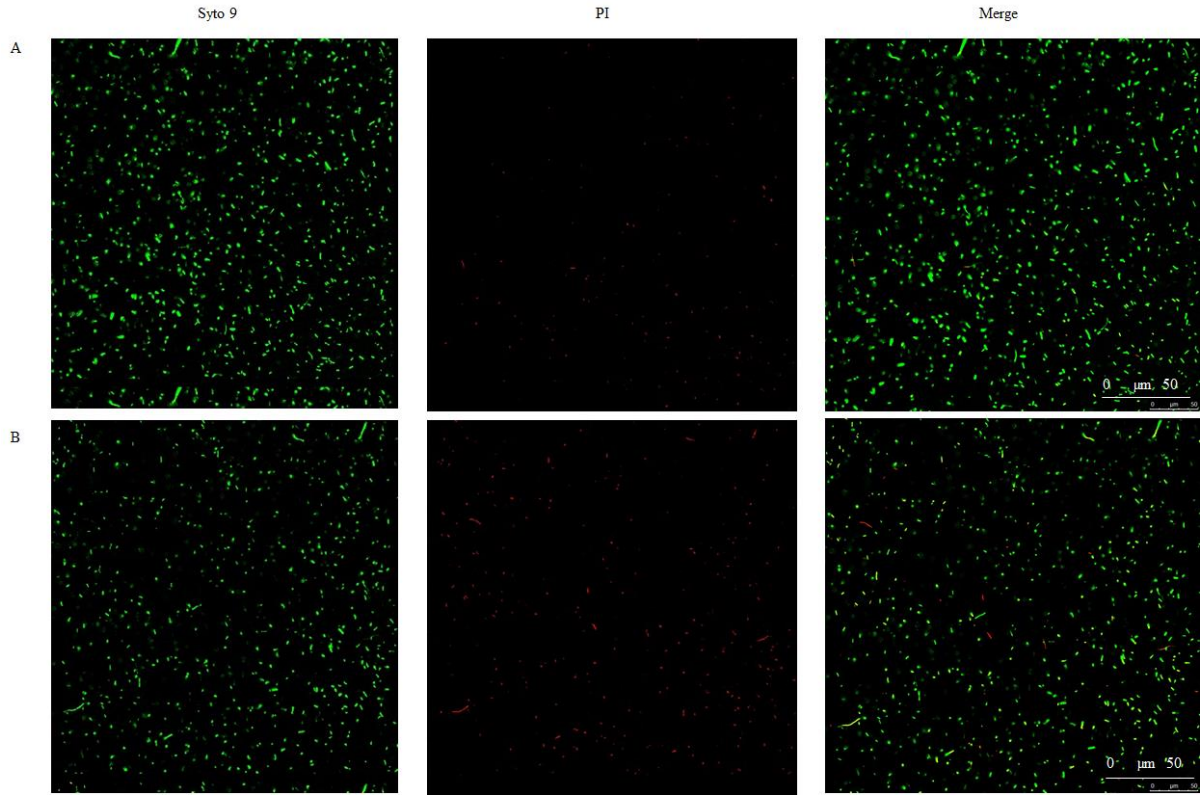


Figure 5.8 Graph of mean optical densities of *E. coli* BL21(DE3) constructs with and without ColE9–Imm9 at concentrations of 10 nM and 100 nM plus DTT. *E. coli* BL21(DE3) cells expressing TA (TA), *E. coli* BL21(DE3) cells expressing TolA box (TolA) and *E. coli* BL21(DE3) cells expressing TolB box (TolB) were treated with different concentrations of ColE9–Imm9 (i.e., 10 nM and 100 nM) plus DTT and compared to untreated cells of each construct type (C; negative controls with DTT), and *E. coli* BL21(DE3) cells plus 100 nM ColE9-Imm9 plus DTT (D/C) as a positive control for the antimicrobial activity of ColE9-Imm9. ColE9–Imm9 plus DTT exerted no inhibitory effect on the bacterial constructs except the *E. coli* BL21(DE3) cells (D/C). (n = 3; error bars represent the standard error of the mean).

5.3.7 Analysis of the antimicrobial activity of ColE9-Imm9 on *E. coli* BL21(DE3) cells using Live/Dead staining in conjunction with CLSM

5.3.7.1 CLSM imaging of E. coli *BL21(DE3) cells*

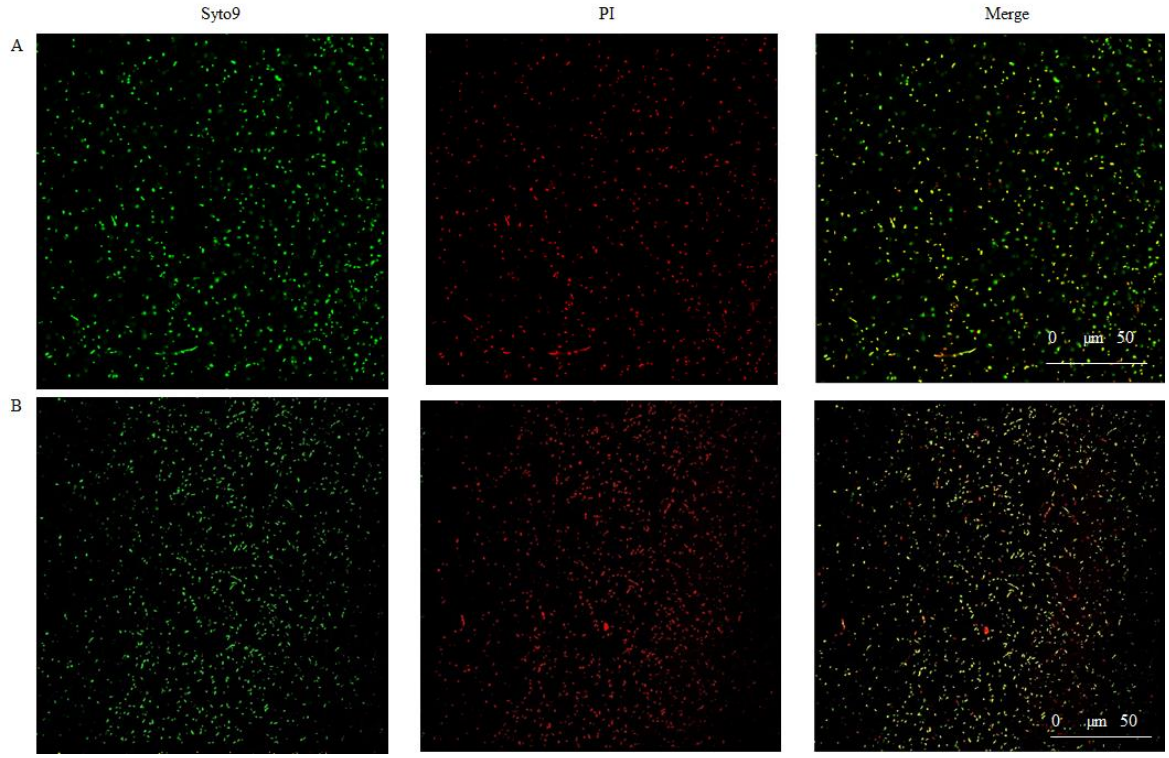
The effect of ColE9–Imm9 in conjunction with DTT on the viability of *E. coli* BL21(DE3) cells was investigated using LIVE/DEAD® BacLight™ stain in conjunction with CLSM as described in Section 2.2.15. *E. coli* BL21(DE3) cells were treated with either 10 nM of ColE9–Imm9 plus DTT (BL21(DE3)/10) or 100 nM of ColE9–Imm9 plus DTT (BL21(DE3)/100) for 3 h or 18 h and were compared with an untreated control group (BL21(DE3)). Post-treatment bacterial samples were stained, as described in Section 2.2.16, and visualised using CLSM. Figure 5.9 shows the results for untreated BL21(DE3) cells (control), where the vast majority of cells are stained green indicating viability, with few cells stained red (red being indicative of membrane damage). There was no visible difference between the samples incubated for 3 h and 18 h, in terms of the relative proportion of green and red cells.



Green= live / Red= dead

Figure 5.9 CLSM images of untreated *E. coli* BL21(DE3) cells with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI, and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells after 3 h of incubation at 37 ℃. B) *E. coli* BL21(DE3) cells after 18 h of incubation at 37 ℃. There was no visible change in the number of live/dead cells between the two incubation periods.

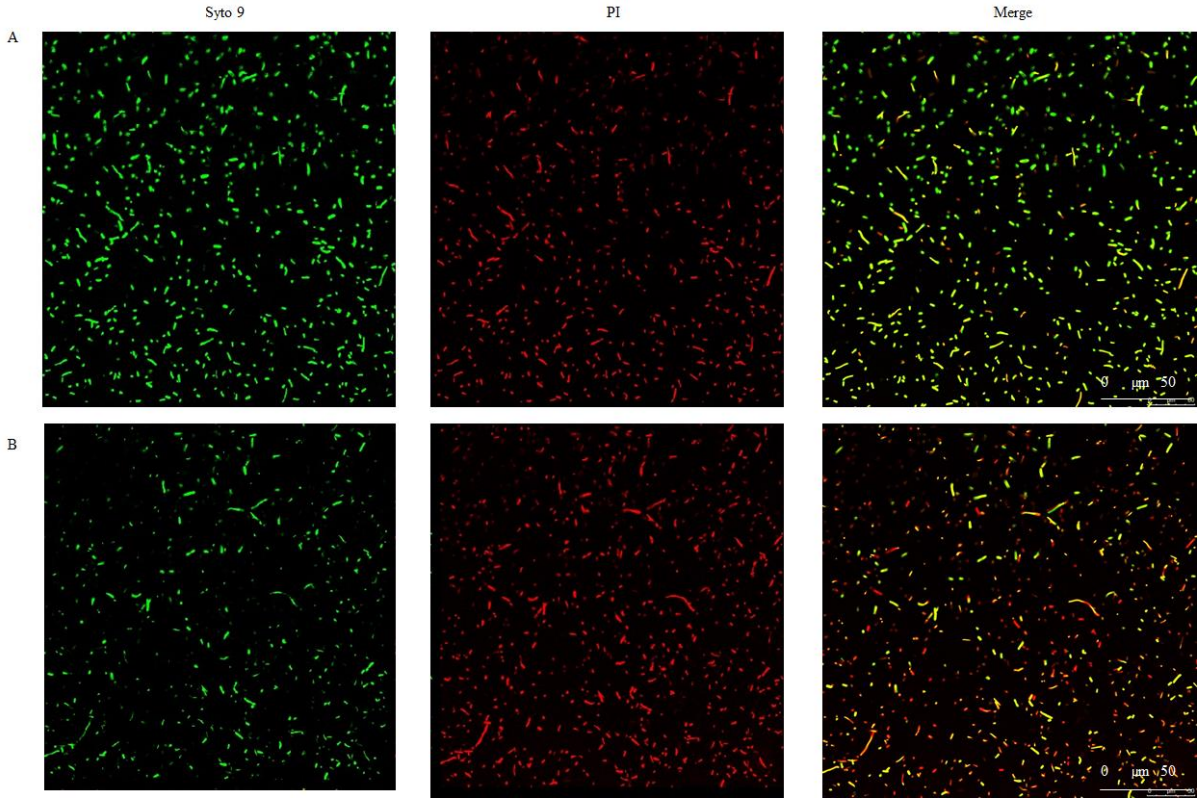
Across 60 fields of view, an average of 59% of BL21(DE3)/10 cells were found to be living cells (i.e., green), and 31% were found to be dead (i.e., red) after 3 h of incubation. While after 18 h, dead cells (i.e., red) represented 36%, whereas living cells (i.e., green) represented 64% of the total cells as shown in Figure 5.10.



Green= live / Red= dead

Figure 5.10 CLSM images of *E. coli* BL21(DE3) cells treated with 10 nM of ColE9–Imm9 plus DTT (BL21(DE3)/10) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells treated with 10 nM of ColE9–Imm9 plus DTT after 3 h incubation at 37 °C. B) *E. coli* BL21(DE3) cells treated with 10 nM of ColE9–Imm9 plus DTT after 18 h incubation. The viability of *E. coli* BL21(DE3) cells visibly began to be affected by the presence of ColE9–Imm9 plus DTT, given the slight increase of red cells (i.e., dead cells) during the two-incubation periods with the appearance of yellow cells that are considered to be damaged cells.

Across 60 fields of view, an average of 46% of BL21(DE3)/100 cells were found to be living cells (i.e., green), and 54% were found to be dead (i.e., red) after 3 h of incubation. While after 18 h, an increase in dead cells (i.e., red/yellow) at 66% was observed, whereas living cells (i.e., green) represented only 33% of the total cells as shown in Figure 5.11.



Green= live / Red= dead

Figure 5.11 CLSM images of *E. coli* BL21(DE3) cells treated with 100 nM of ColE9–Imm9 plus DTT (BL21(DE3)/100) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells treated with 100 nM of ColE9–Imm9 plus DTT after 3 h of incubation at 37 °C, most of the cells were yellow and thus damaged. B) *E. coli* BL21(DE3) cells treated with 100 nM of ColE9–Imm9 plus DTT after 18 h of incubation at 37 °C, most of the cells appeared red indicating death due to the high concentration of cytotoxic ColE9–Imm9 added.

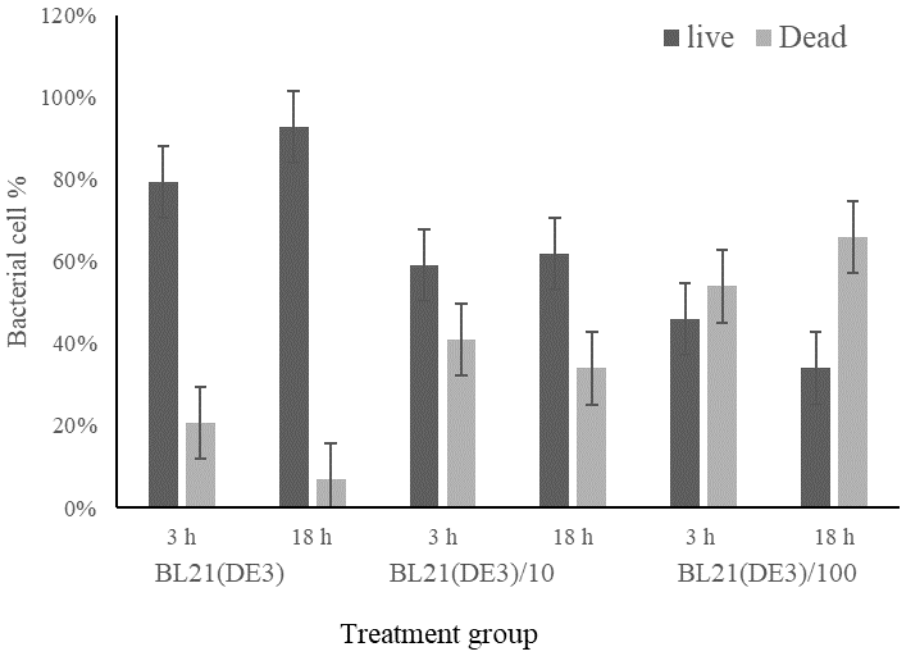


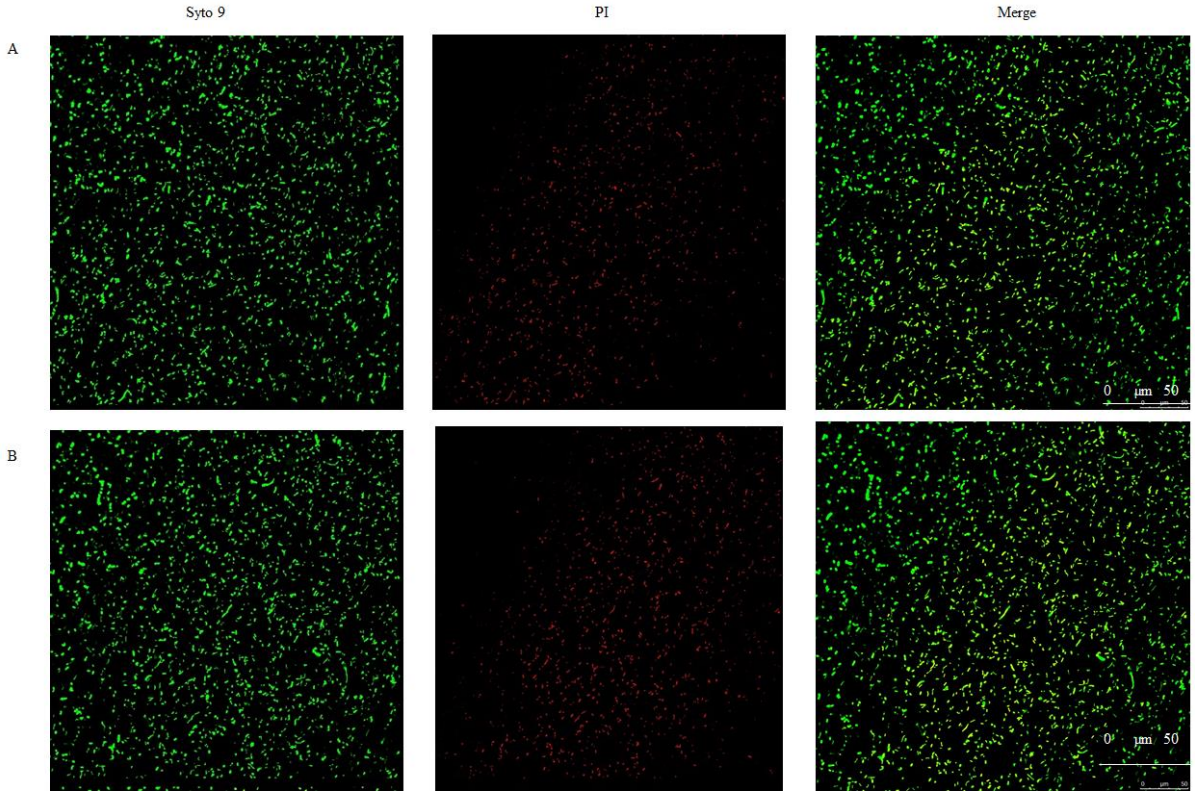
Figure 5.12 The difference in percentage of live versus dead bacterial cells from 60 randomly selected fields of view of *E. coli* BL21(DE3) cells with and without ColE9–Imm9 plus DTT following 3 or 18 h of treatment. The graph shows the highest percentage of live cells in untreated *E. coli* BL21(DE3) cells during the two incubation periods. For the BL21(DE3)/10 sample, the percentage of dead cells rose during the two incubation periods due to the addition of 10 nM ColE9-Imm9 plus DTT but didn't exceed the live cells. For the BL21(DE3)/100 sample the percentage of dead cells increased further leaving live cells low in number after the addition of 100 nM ColE9-Imm9. (n= 3, error bars represent the standard error of the mean).

According to the data shown in Figure 5.12, after testing the normality and homogeneity of variances, the non-parametric Kruskal-Wallis test revealed that after 3 h of incubation, there is no significant difference ($p=0.203$) between live cells with and without the treatment with ColE9-Imm9 (10 nM, 100nM), whereas, there is a statistical difference ($p<0.001$) between the dead cells. While, after 18 h, both live and dead cells revealed a significant difference ($p<0.001$) between them due to the effect of ColE9-Imm9 plus DTT as they exceed the number of living cells.

5.3.8 Analysis of the antimicrobial activity of ColE9-Imm9 on *E. coli* BL21(DE3) cells expressing TA using Live/Dead staining in conjunction with CLSM

5.3.8.1 CLSM imaging of E. coli *BL21(DE3) cells expressing TA*

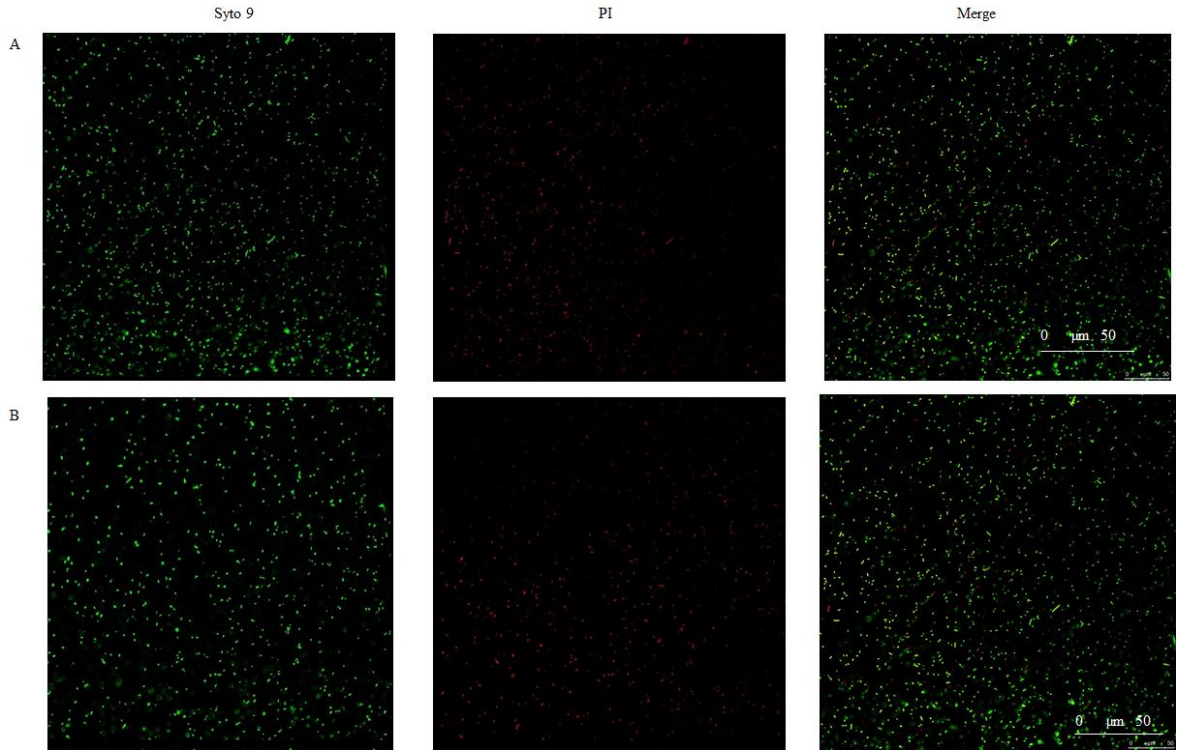
To examine *E. coli* BL21(DE3) cells exposed to different conditions using CLSM, bacterial samples were prepared as described in Sections 2.2.15 & 2.2.16, and visualised as described in Section 2.2.17. The experiment was conducted on cells expressing TA (i.e., control), TA/10 (i.e., cells expressing TA treated with 10 nM of ColE9–Imm9 plus DTT) and TA/100 (i.e., cells expressing TA treated with 100 nM of ColE9–Imm9 plus DTT) after two incubation periods;3 h and 18 h. Figure 5.13 shows the results for untreated TA cells (i.e., control), where the vast majority of cells are stained green indicating viability, with few cells-stained red which indicating dead cells. There was no visible difference between the samples incubated for 3 h and 18 h, in terms of the relative proportion of green and red cells.



Green= live / Red= dead

Figure 5.13 CLSM images of untreated *E. coli* BL21(DE3) cells expressing TA (TA cells) an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) expressing TA cells after 3 h of incubation at 37 °C. B) *E. coli* BL21(DE3) cells after 18 h of incubation at 37 °C. There was no visible change in the number of live/dead cells between the two incubation periods.

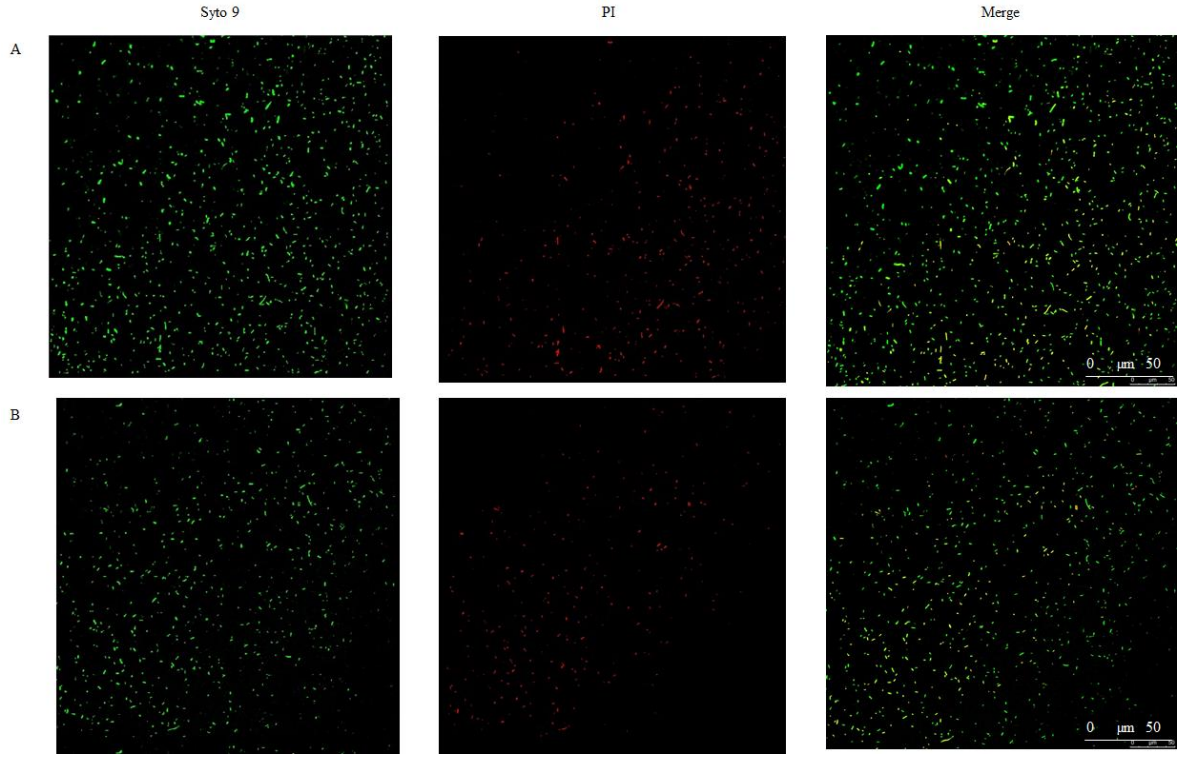
Across 60 fields of view, 76% of cells were found to be living cells (i.e., in green), and 23% cells were found to be dead (i.e., red) after 3 h incubation. While after 18 h, a slight increase in dead cells (i.e., in red) at 31% was observed, whereas living cells (i.e., green) represented 69% of the total cells as shown in Figure 5.14.



Green= live / Red= dead

Figure 5.14 CLSM images of *E. coli* BL21(DE3) cells expressing TA after the addition of 10 nM of ColE9–Imm9 plus DTT (TA/10 cells) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells expressing TA after the addition of 10 nM of ColE9–Imm9 plus DTT after 3 h incubation at 37 °C. B) *E. coli* BL21(DE3) cells expressing TA after the addition of 10 nM of ColE9–Imm9 plus DTT after 18 h incubation at 37 °C. The viability of TA/10 cells has not been visibly affected by the presence of ColE9–Imm9 plus DTT, given the high number of green cells (i.e., live cells) during the two-incubation periods, although the red cells (i.e., dead cells) have increased slightly in numbers.

Across 60 field of view, an average of 60% of TA/100 cells were found to be living cells (i.e green), and 40% were found to be dead (i.e., red) after 3 h of incubation. While after 18 h, the dead cells (i.e., red) were found to be 41%, whereas living cells (i.e., green) represented 59% of the total cells as shown in Figure 5.15.



Green= live / Red= dead

Figure 5.15 CLSM images of *E. coli* BL21(DE3) cells expressing TA after the addition of 100 nM of ColE9–Imm9 plus DTT (TA/100) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells expressing TA after the addition of 100 nM of ColE9–Imm9 plus DTT after 3 h incubation at 37 °C. B) *E. coli* BL21(DE3) cells expressing TA after the addition of 100 nM of ColE9–Imm9 plus DTT after 18 h incubation at 37 °C. The viability of TA/100 cells has not been visibly affected by the presence of ColE9–Imm9 plus DTT, given the high number of green cells (i.e., live cells) during the 3 h and 18 h incubation periods with slightly similar number of red cells (i.e., dead cells).

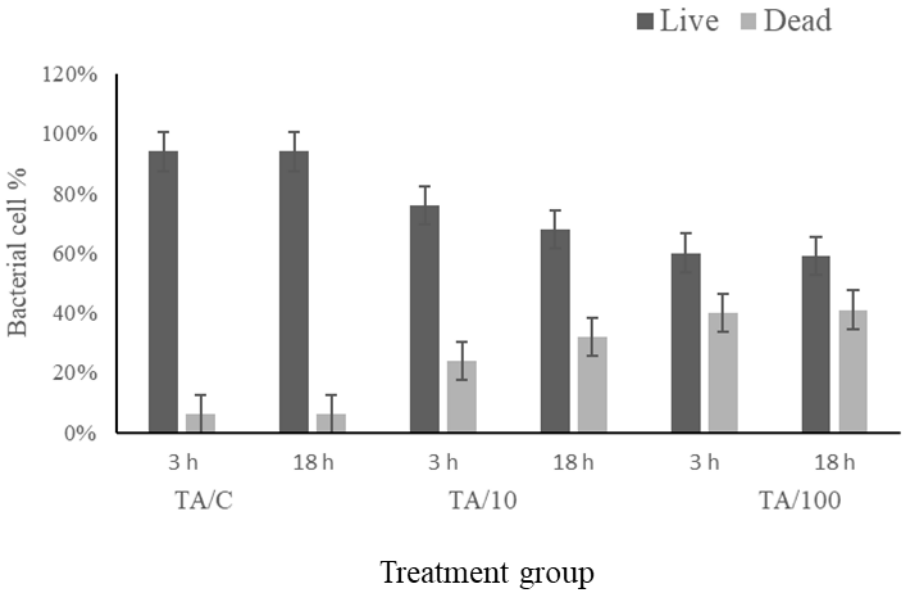


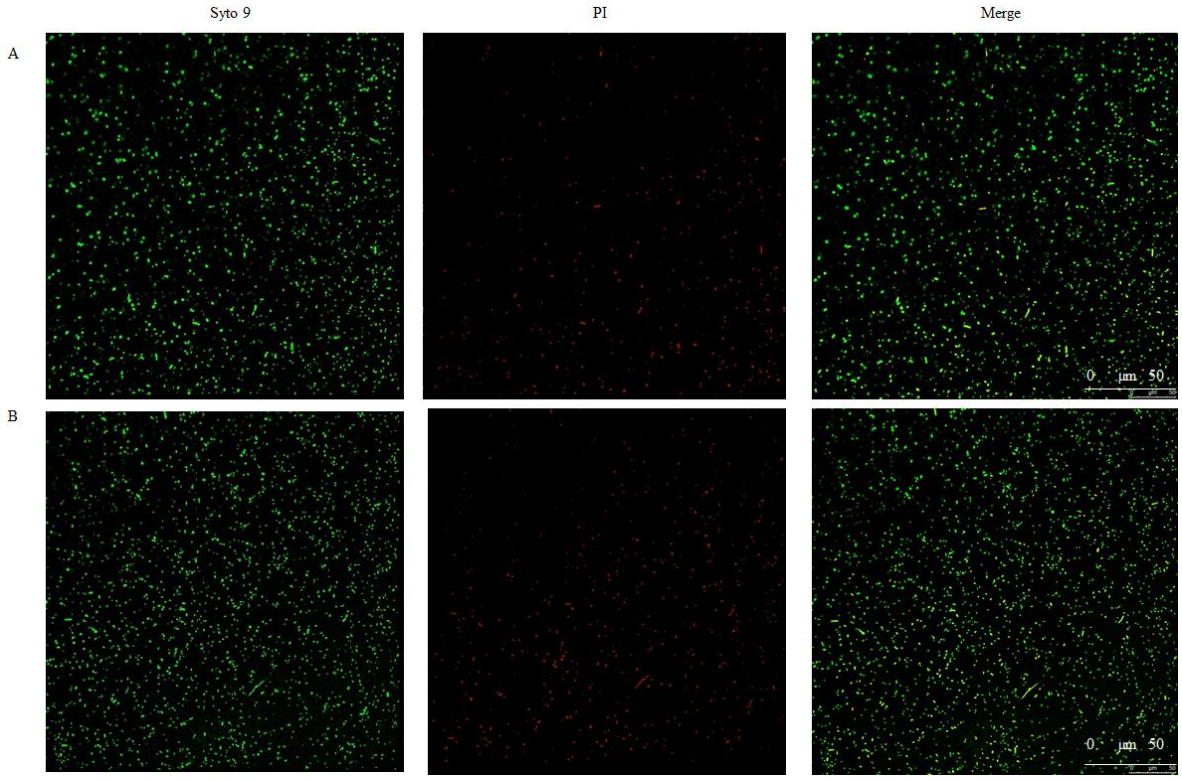
Figure 5.16 The difference in percentage of live versus dead bacterial cells from 60 randomly selected fields of view of *E. coli* BL21(DE3) expressing TA with and without ColE9–Imm9 plus DTT following 3 or 18 h of treatment. The graph shows a high number of live cells in the control TA sample during the two incubation periods. With TA/10 the number of dead cells increased during the two incubation periods but didn't exceed the number of live cells. The number of dead cells has increased but didn't exceed live cell numbers on TA/100 (n= 3, error bars= Standard error of the mean)

According to the data in Figure 5.16 unlike the data presented in Figure 5.12, the number of dead cells did not exceed the number of living cells even when the cells were treated with ColE9-Imm9 plus DTT. Although the non-parametric Kruskal-Wallis test revealed that there is a significant difference ($p<0.001$) between both live and dead TA groups after 3 h of incubation. After 18 h there is no significant difference ($p=0.306$) between live TA groups, but there is significant difference ($p<0.001$) between dead TA groups. However, the percentages of live cells still remain high among all TA cells unlike colicin sensitive *E. coli* BL21(DE3) cells in Section 5.3.7.1

5.3.9 Analysis of the antimicrobial activity of ColE9-Imm9 on *E. coli* BL21(DE3) cells expressing TolA box in conjunction with CLSM

5.3.9.1 CLSM imaging of E. coli *BL21(DE3) cells expressing TolA box*

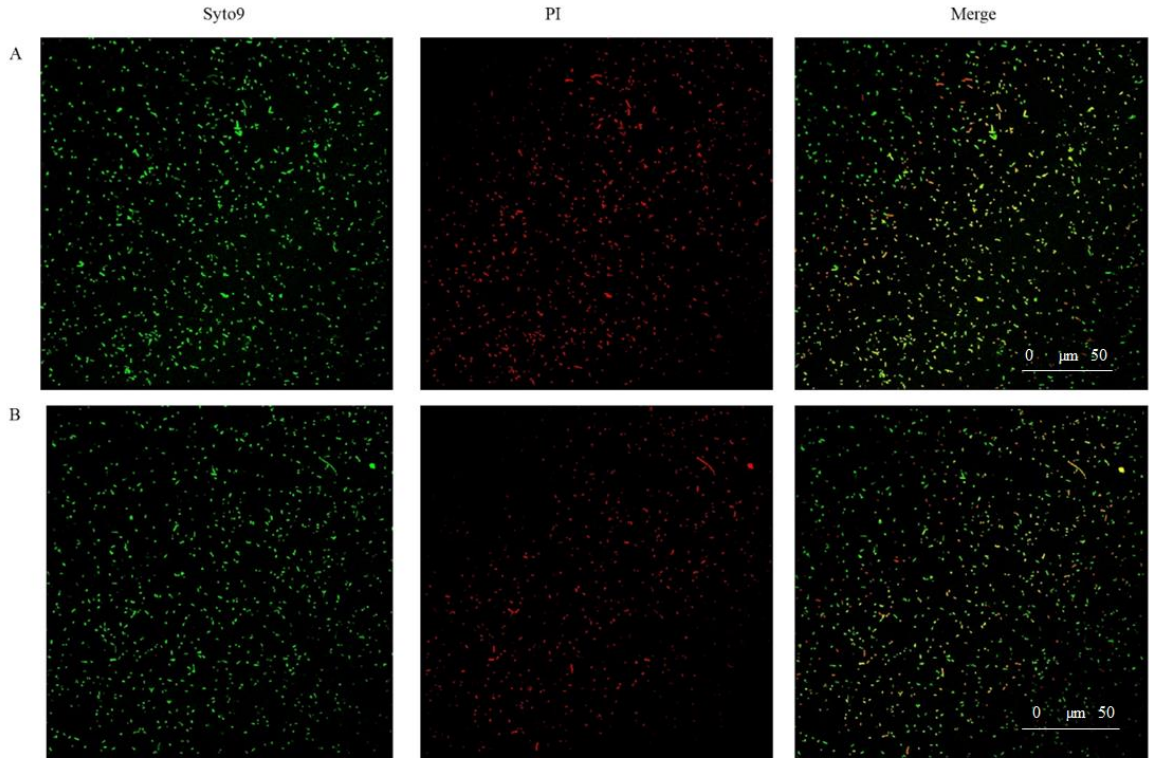
As described in Section 2.2.17, to investigate *E. coli* BL21(DE3) cells expressing TolA box in different conditions with CLSM, bacterial samples were prepared as described in Section 2.2.15 & 2.2.16, and visualised as described in Section 2.2.17. *E. coli* BL21(DE3) cells expressing TolA box were divided into three groups: control cells (i.e., TolA); cells treated with 10 nM of ColE9–Imm9 plus DTT (i.e., TolA/10); and cells treated with 100 nM of ColE9–Imm9 plus DTT (i.e., Tol/100). Figure 5.17 shows the results of untreated TolA cells (control), where the vast majority of cells are stained green indicating viability, with few cells stained red indicating dead cells. There was no visible difference between the samples incubated for 3 h and 18 h in terms of the proportion of green and red cells.



Green= live / Red= dead

Figure 5.17 CLSM images of *E. coli* BL21(DE3) cells expressing TolA box (TolA cells) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) expressing TolA box after 3 h of incubation at 37 °C. B) *E. coli* BL21(DE3) expressing TolA box after 18 h of incubation at 37 °C. There was no visible change in the number of live/dead cells between the two incubation periods.

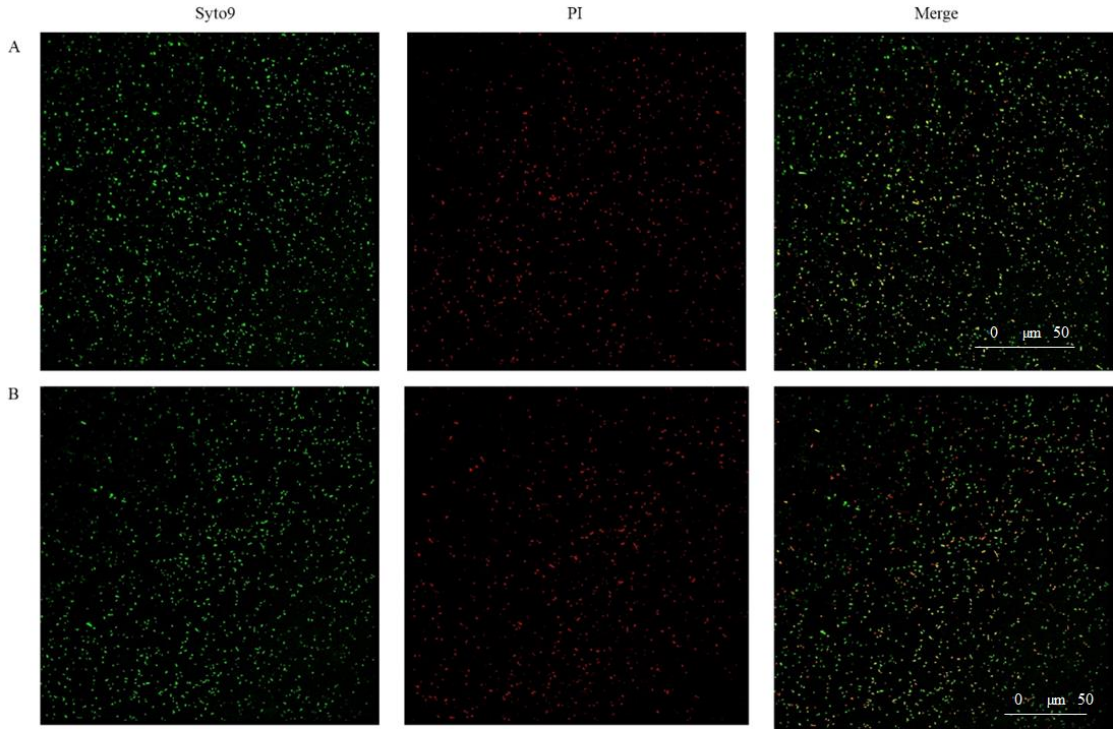
Across 60 fields of view, 66% of TolA/10 cells were found to be living cells (i.e., in green), and 43% cells were found to be dead (i.e., red) after 3 h incubation. While after 18 h, a slight increase in dead cells (i.e., in red) at 36% was observed, whereas living cells (i.e., green) represented 68% of the total cells as shown in Figure 5.18.



Green= live / Red= dead

Figure 5.18 CLSM images of *E. coli* BL21(DE3) cells expressing TolA box after the addition of 10 nm of ColE9–Imm9 plus DTT (TolA/10 cells) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells expressing TolA box after the addition of 10 nm of ColE9–Imm9 plus DTT after 3 h incubation at 37 °C. B) *E. coli* BL21(DE3) cells expressing TolA box after the addition of 10 nm of ColE9–Imm9 plus DTT after 18 h incubation at 37 °C. The viability of TolA/10 cells has not visibly been affected by the presence of ColE9–Imm9 plus DTT, given the high number of green cells (i.e., live cells) during the two incubation periods. Although the number of red cells (i.e., dead cells) has slightly increased.

Across 60 fields of view, an average of 58% of TolA/100 cells were found to be living cells (i.e., green), and 42% were found to be dead (i.e., red) after 3 h of incubation. While after 18 h, a decline in dead cells (i.e., red) at 27% was observed, whereas living cells (i.e., green) represented 73% of the total cells as shown in Figure 5.19.



Green= live / Red= dead

Figure 5.19 CLSM images of *E. coli* BL21(DE3) cells expressing TolA box after the addition of 100 nm of ColE9–Imm9 plus DTT (TolA/100) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells expressing TolA box after the addition of 100 nm of ColE9–Imm9 plus DTT after 3 h incubation at 37 °C. B) *E. coli* BL21(DE3) cells expressing TolA box after the addition of 100 nm of ColE9–Imm9 plus DTT after 18 h incubation at 37 °C. The viability of TolA/100 cells has not been visibly affected by the presence of ColE9–Imm9 plus DTT, given the high number of green cells (i.e., live cells) during the two-incubation period. Also, the number of red cells (i.e., dead cells) has declined after 18 h.

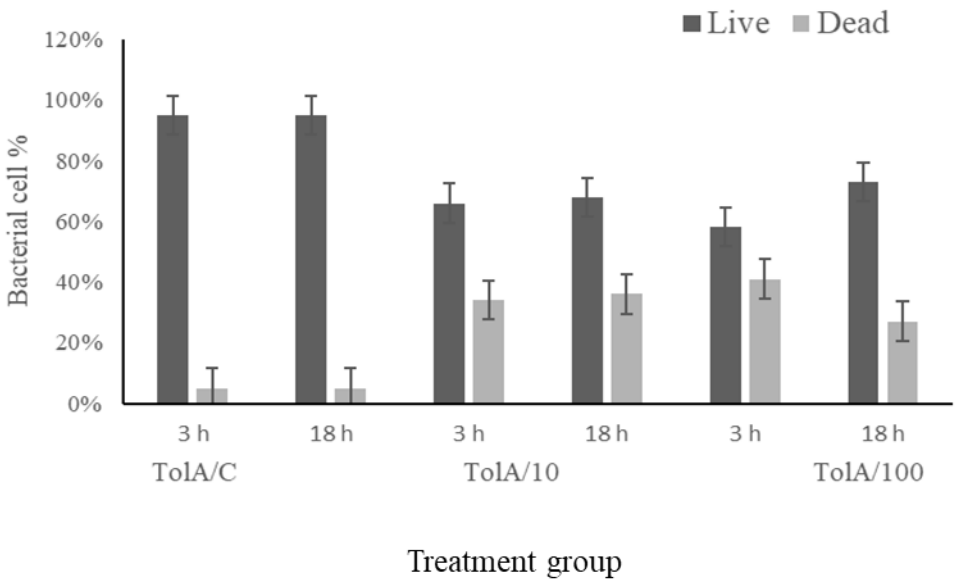


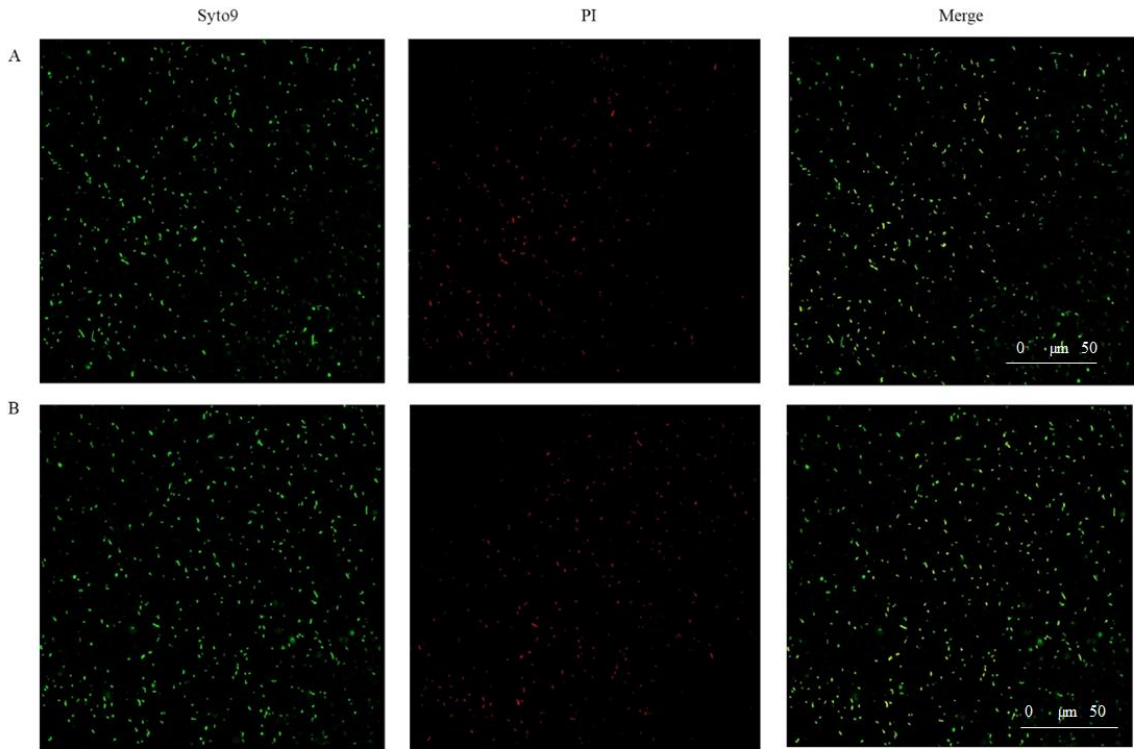
Figure 5.20 The difference in bacterial cell percentage of live versus dead cells from 60 randomly selected fields of view of *E. coli* BL21(DE3) expressing TolA box with and without ColE9–Imm9 plus DTT during two incubation periods. The graph shows high number of live cells in control TolA cells during the two incubation periods. While with TolA/10 the number of dead cells rose during the two incubation periods but didn't exceed the live cells number. The number of dead cells has declined in TolA/100, giving the high number of living cells after the addition of 100 nM ColE9-Imm9 plus DTT. (n= 3, Error bars= Standard error of the mean).

According to the data in Figure 5.20 unlike the data presented in Figure 5.12, the number of dead cells did not exceed the number of living cells even when the cells were treated with ColE9-Imm9 plus DTT. Although the non-parametric Kruskal-Wallis test revealed that there is a significant difference ($p<0.001$) between both live and dead of TolA groups after both incubation period (3 h & 18 h). However, the percentages of live cells still remain high among all TolA cells unlike colicin sensitive *E. coli* BL21(DE3) cells in Section 5.3.7.1

5.3.10 Analysis of the antimicrobial activity of ColE9-Imm9 on *E. coli* BL21(DE3) cells expressing TolB box in conjunction with CLSM

5.3.10.1 CLSM imaging of E. coli *BL21(DE3) cells expressing TolB box*

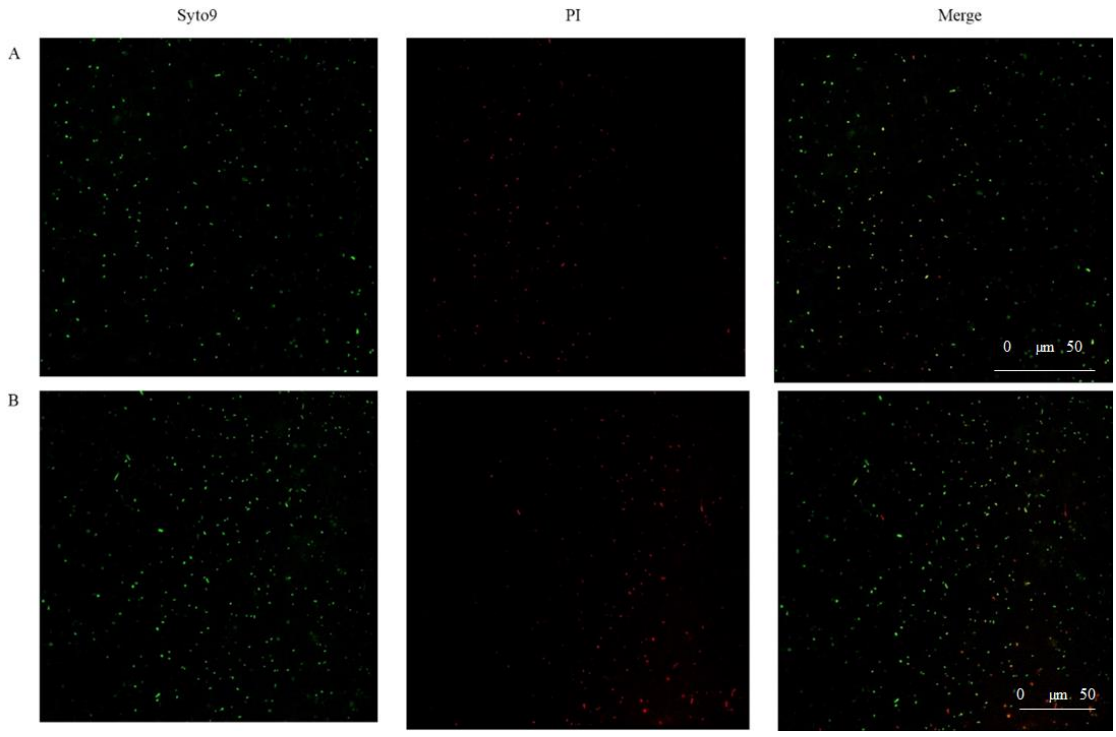
To examine *E. coli* BL21(DE3) cells expressing TolB box in different conditions under CLSM, as described in Section 2.2.15, bacterial samples were prepared as described in Section 2.2.16 and visualised as described in Section 2.2.17. *E. coli* BL21(DE3) cells expressing TolB box (TolB cells) were divided into three groups: untreated control cells (i.e., TolB); cells treated with 10 nM of ColE9–Imm9 plus DTT (i.e., TolB/10); and cells treated with 100 nM of ColE9–Imm9 plus DTT (i.e., TolB/100 group). Figure 5.21 shows the results for untreated control TolB cells, where the vast majority of cells are stained green indicating viability, with few cells stained red indicating dead cells. There was no visible difference between samples incubated for 3 h and 18 h, in terms of the relative proportion of green and red cells.



Green= live / Red= dead

Figure 5.21 CLSM images of *E. coli* BL21(DE3) cells expressing TolB box (TolB cells) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells expressing TolB box after 3 h of incubation at 37 °C. B) *E. coli* BL21(DE3) cells expressing TolB box after 18 h of incubation at 37 °C. There was no visible change in the number of live/dead cells between the two incubation periods.

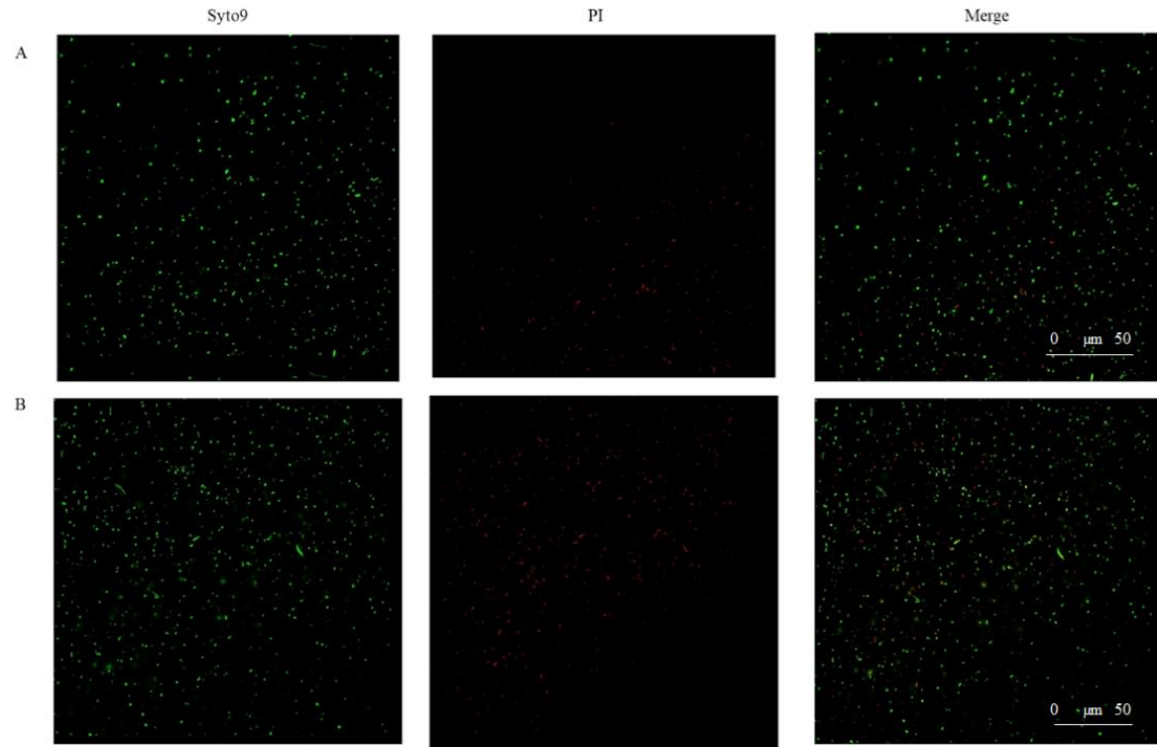
Across 60 fields of view, an average of 60% of TolB/10 cells were found to be living cells (i.e., green), and 40% were found to be dead (i.e., red) after 3 h of incubation. While after 18 h, a decline in dead cells (i.e., red) at 30% was observed, whereas living cells (i.e., green) represented 70% of the total cells as shown in Figure 5.22.



Green= live / Red= dead

Figure 5.22 CLSM images of *E. coli* BL21(DE3) cells expressing TolB box after the addition of 10 nm of ColE9–Imm9 and 10 mM of DTT (TolB/100) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells expressing TolB box after the addition of 10 nm of ColE9–Imm9 and 10 mM of DTT after 3 h incubation at 37 °C. B) *E. coli* BL21(DE3) cells expressing TolB box after the addition of 10 nM of ColE9–Imm9 and 10 mM of DTT after 18 h incubation at 37°C. The viability of TolB/10 cells has not been visibly affected by the presence of ColE9–Imm9 plus DTT, given the high number of green cells (i.e., live cells) during the two-incubation period. Also, the number of red cells (i.e., dead cells) has slightly declined.

Across 60 fields of view, an average of 72% of TolB/100 were found to be living cells (i.e., green), and 28% were found to be dead (i.e., red) after 3 h of incubation. While after 18 h, an increase in dead cells (i.e., red) at 46% was observed, whereas living cells (i.e., green) represented 54% of the total cells as shown in Figure 5.23.



Green= live / Red= dead

Figure 5.23 CLSM images of *E. coli* BL21(DE3) cells expressing TolB box after the addition of 100 nM of ColE9–Imm9 plus DTT (TolB/100) with an optical magnification of 40× using an oil-immersion objective. The figure shows green cells stained with SYTO 9, red cells stained with PI and the merged cells (live/dead). A) *E. coli* BL21(DE3) cells expressing TolB box after the addition of 100 nM of ColE9–Imm9 plus DTT after 3 h incubation at 37 °C. B) *E. coli* BL21(DE3) cells expressing TolB box after the addition of 100 nM of ColE9–Imm9 plus DTT after 18 h incubation at 37 °C. The viability of TolB/100 cells has not been visibly affected by the presence of ColE9–Imm9 plus DTT, given the high number of green cells (i.e., live cells) during the two-incubation period. Although the number of red cells (i.e., dead cells) has increased.

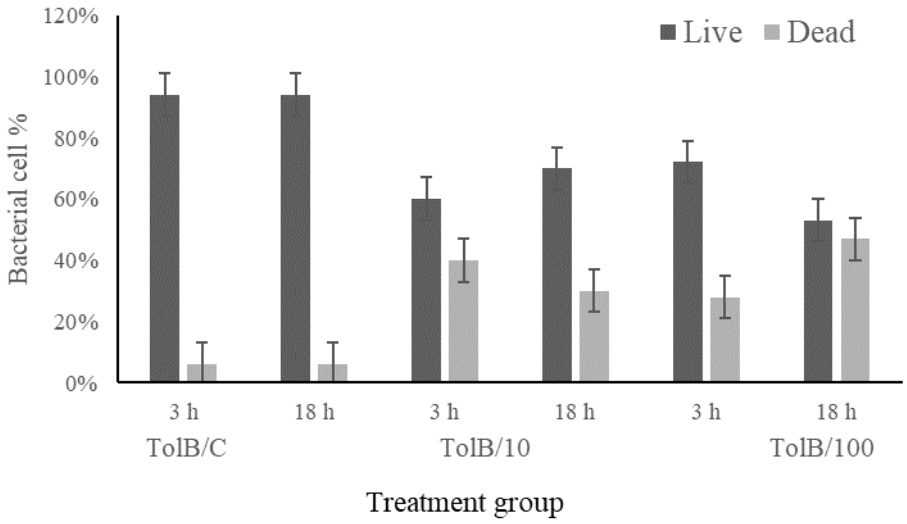


Figure 5.24 The difference in bacterial cell percentage of live versus dead cells from 60 randomly selected visual fields of *E. coli* BL21(DE3) expressing TolB box with and without ColE9–Imm9 plus DTT during two incubation periods. The graph shows a high proportion of live cells in control TolB (TolB/C) cells during the two incubation periods. While with TolB/10 the number of dead cells rose during the two incubation periods but didn't exceed the live cells number. The number of dead cells has increased in TolB/100 compared to TolB/10 but didn't exceed live cells number after the addition of 100 nM ColE9-Imm9 plus DTT. (n=3, error bars= standard error of the mean).

According to the data in Figure 5.24 unlike the data presented in Figure 5.12, the number of dead cells did not exceed the number of living cells even when the cells were treated with ColE9-Imm9 plus DTT. Although the non-parametric Kruskal-Wallis test revealed that there is a significant difference ($p<0.001$) between both live and dead TA groups after both incubation period (3 h & 18 h). However, the percentages of live cells still remain high among all TA cells unlike colicin sensitive *E. coli* BL21(DE3) cells in Section 5.3.7.1

5.3.11 Plate counts to determine the antimicrobial activity of ColE9–Imm9 plus DTT on bacterial constructs

To supplement the results obtained from CLSM, spread plate counts were performed on colicin-sensitive *E. coli* BL21(DE3) cells following challenge with 10 nM ColE9-Imm9 plus DTT, as described in Section 2.2.18. Figure 5.25(A) shows the *E. coli* BL21(DE3)-sensitive cells on an agar plate. Figure 5.25(B) shows the colonies obtained after the cells were challenged with100 nM of ColE9–Imm9 plus DTT. After calculating the colony forming units (CFUs) for *E. coli* BL21(DE3) before and after adding ColE9–Imm9, the average count of *E. coli* BL21(DE3) cells after adding 10 nm of ColE9-Imm9 plus DTT equalled 2.6×10^9 CFU/mL. While the average count of *E. coli* BL21(DE3) cells after adding 100 nm of ColE9 equalled 1.8×10^9 CFU/mL.

For untreated *E. coli* BL21(DE3) and other constructs expressing TA and the TolA and TolB boxes of ColA, all revealed how the colonies were too many to count on a 10^{-7} plate with estimated minimum of 300 colonies (i.e., the maximum countable number) even after ColE9–Imm9 plus DTT were added.

Assuming their CFU/ml is > 3.2 X 10^9, and because of the small sample size the data were tested by Kruskal Wallis to reveal ($T(2) = 7.261$, $p = .027$) that there is a statistically significant difference between three groups (i.e., untreated BL21(DE3), and 10 nM ColE9-Imm9 plus DTT treated cells, and 100 nM ColE9-Imm9 plus DTT treated cells (see Appendix C).

A B

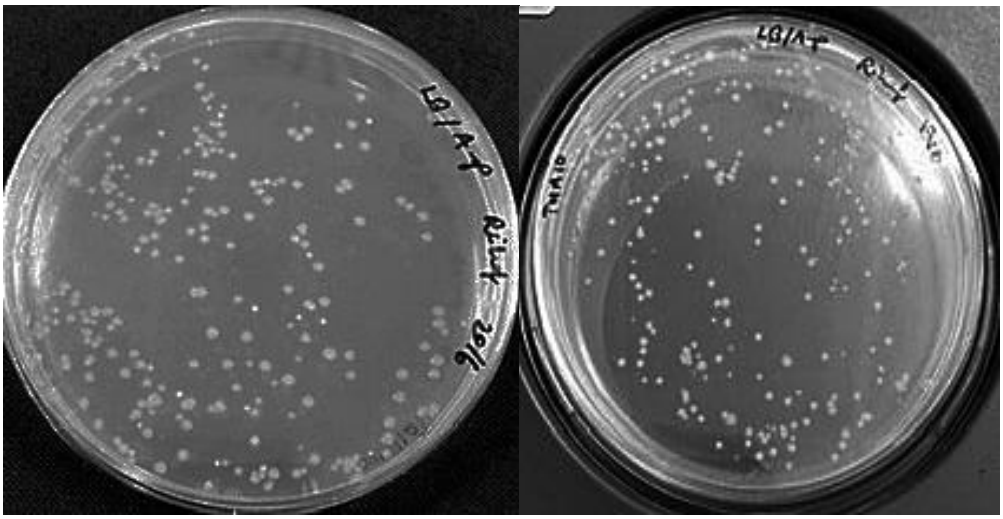


Figure 5.25 *E. coli* BL21(DE3) cells with ColE9–Imm9 plus DTT. A) Colonies grown on an agar plate after incubation with 10 nM of ColE9–Imm9 plus DTT at 37 °C overnight. (B) Colonies grown on the plate after incubation with 100 nM of ColE9–Imm9 plus DTT at 37°C overnight, during which time the number of colonies decreased.

5.6 Discussion

The aim of expressing and secreting colicin domains into the cell periplasm was to observe the interaction between the Tol system and externally added colicins. Because that interaction can be monitored by cellular patterns of sensitivity and resistance to externally added colicins, this research involved investigating whether the expression of ColA translocation domains protects sensitive cells, including *E. coli* BL21(DE3), from externally added ColE9. This hypothesis derives from the fact that interactions between the translocation domain and cellular Tol proteins preclude all subsequent interactions of Tol proteins with the (T) domains of externally added colicins. Beyond that, the competitive inhibition of TolA affects cell resistance more significantly than the inhibition of TolB, the examination of which may provide new insights into the role of different Tol proteins in the translocation of different colicins.

This chapter presents a novel approach for investigating protein–protein interaction and the effects of the cytotoxic activity of the ColE9–Imm9 complex against various *E. coli* BL21(DE3) cells. ColE9–Imm9 was tested against *E. coli* BL21(DE3) cells expressing TA, TolA box and TolB box using several techniques, including testing ColE9–Imm9 antimicrobial activity in the absence and presence of dithiothreitol (DTT) in liquid culture, live-cell imaging using CLMS, and plate counting. Penfold and colleagues (2004) previously constructed a ColE9–Imm9 mutation by introducing two cysteine residues at position 496 in the receptor-binding domain, or (R) domain, which caused a disulphide bond to form in the centre of the domain. Subsequently, the mutation facilitated ColE9–Imm9 purification as a means to produce recombinant ColE9–Imm9. In their study, once the mutant was constructed, the resulting plasmid was sequenced to confirm the presence of cysteine mutations in the (R) domain. Their recombinant plasmid pColE9 encoded ColE9 containing the Y324C–L447C mutations, as well as a poly-histidine tag on Imm9 for the purification of the (C) domain of

the 10-kDa in the ColE9–Imm9 complex when both proteins were co-expressed in T7 expression vectors. The authors introduced an *Xho*I site in the place of the immunity gene's stop codon in the plasmid pColE9 during a PCR. The 2012-bp *Nde*I–*Xho*I PCR fragment was cloned into plasmid pET-2la restricted with the same enzymes, and the resulting plasmid was transformed into *E. coli* BL21(DE3) cells (Penfold *et al.*, 2004).

In this research, the ColE9–Imm9 mutant was the chosen cytotoxic protein for two major reasons. Firstly, *E. coli* BL21(DE3) cells are sensitive to colicin (Ranjan *et al.*, 2018), and, secondly, ColE9–Imm9 is easy to purify. Although the ColE9–Imm9 mutant's cytotoxic activity could constitute a limitation because of the mutations, other researchers have examined the disulphide form of ColE9–Imm9 instead of the wild type owing to the flexibility of the former's (R) domain. That domain undergoes conformational changes that occur due to the entry of colicins, an event that ensures that the disulphide bonds do not alter the ColE9 activity or its affinity to Imm9 (Klein *et al.*, 2016). Those findings guided this research approach in testing the ColE9–Imm9 mutant. In the research, the overexpression and purification of ColE9–Imm9 in the cells was achieved in small- and large-scale bacterial cultures, as described in Section 2.2.7. As a result, the expected bands for 60 kDa of ColE9 and 10 kDa of Imm9 confirmed the co-expression of the immunity protein with ColE9 as they formed a complex as illustrated in Figures 5.3 & 5.4.

This chapter aims to address some fundamental questions regarding the translocation of ColE9–Imm9. For one, how can the effect of ColE9–Imm9 on sensitive cells be measured? Also, if the periplasm of the ColE9-Imm9 treated cells is occupied by periplasmic proteins, does it affect the translocation of ColE9 from the OM to the periplasm. If so, then is it due to the competitive inhibition of protein–protein interactions in the periplasm or the compromised dissociation of Imm9? To date, progress towards answering those questions has been limited

by the lack of a sensitive, rapid assay for ColE9–Imm9 translocation other than by monitoring the resulting inhibition of growth.

To determine the cytotoxic concentration of ColE9-Imm9 against the general host strain, *E. coli* BL21(DE3)—, a spot test assay was performed, a traditional cell-killing assay in which different concentrations of ColE9-Imm9 are spotted onto a lawn of sensitive cells to test their antimicrobial activity. The results confirmed the sensitivity of *E. coli* BL21(DE3) cells to ColE9–Imm9, which resulted in a zone of growth inhibition compared with a negative control as shown in Figure 5.5 with minimum inhibitory concentration of 5 nM. By contrast, the *E. coli* BL21(DE3) cells expressing TA, TolA box and TolB box showed complete growth and no zone of inhibition, demonstrating that the expression of the proteins protected the cells against the cytotoxic activity of ColE9-Imm9. This outcome supports the hypothesis that the cells can be protected against ColE9–Imm9 due to the expressed domains (TA, TolA box and TolB box) in the periplasm. Spotting low (i.e 5 nM) and high (i.e., 100 nM) concentrations of ColE9–Imm9 yielded a concentration-dependent effect, as expected, because the used concentrations were above the lethal concentration of ColE9-Imm9 suggested by Li and colleagues (2004). They also tested the sensitivity on another strain of *E. coli* (i.e., DH5α) to ColE9 activity using a spot test assay and determined that 4 nM of ColE9 was a minimum lethal concentration suitable for use in conventional assays.

Performing techniques other than the traditional cell-killing assay was necessary to improve sensitivity and for ease of data analysis. To confirm the results of the spot test, i.e., that ColE9–Imm9 affected the growth of *E. coli* BL21(DE3) but not the growth of the *E. coli* BL21(DE3) expressing TA, TolA box and TolB box, a liquid growth assay was performed whereby *E. coli* BL21(DE3) cells expressing TA, TolA box and TolB box were treated with ColE9–Imm9. As shown in Figure 5.6, ColE9–Imm9, once added to *E. coli* BL21(DE3) cells, did not affect the growth of the cells, even when added in higher concentrations than the

suggested lethal concentration (i.e., 10 & 100 nM), and the optical density achieved by the treated cells was similar to that of the untreated ones. Those results, despite contradicting the results of the spot test which showed inhibition of growth, they aligned with the findings of Penfold and colleagues (2004), who observed that the double cysteine mutation performed in ColE9 (R) domain, along with the formation of a disulphide bond, inhibited the activity of ColE9 without significantly affecting BtuB binding or binding to the cellular TolB. They also tested the mutant protein on *E. coli* DH5α, a colicin-sensitive strain, in a liquid growth inhibition experiment, the results of which confirmed the difference between the results of the plate assay and liquid growth assay in terms of inhibitory activity. To investigate the possible reasons for the loss of ColE9-Imm9 cytotoxic effect in liquid assay, they also compared the interaction between the (T) domain of ColE9 and TolB with the (T) domain of the ColE9 mutant and TolB using two-hybrid screening. The presence of disulphide in the mutant protein had no inhibitory effect on the interaction between their TolB boxes and TolB, although it did somewhat affect the activity of their nuclease activity, unless the bond was reduced with a reducing agent. Adding DTT to the oxidised proteins containing disulphide bonds restored the activity of ColE9, as previously demonstrated by Vankemmelbeke and colleagues (2005), who investigated Imm9 release from the mutant ColE9–Imm9 complex by conducting a fluorescence assay. Using DTT to reduce the disulphide linkage, which restored the activity of ColE9–Imm9, they observed that Imm9 release occurred immediately after the DTT-induced reduction, thereby restoring both flexibility in the (R) domain and the interaction of the (T) domain with the Tol proteins.

In light of those findings, in this research 10 mM (final conc.) of DTT was added to the ColE9–Imm9 complex before treating the *E. coli* cells. The results showed that ColE9–Imm9 activity was restored as expected. The results further indicated that the antimicrobial effect of ColE9–Imm9 with DTT on *E. coli* BL21(DE3) cells is concentration-dependent, as illustrated

in Figure 5.7. However, this difference is not significant ($p > 0.01$) observed with Kruskal Wallis test between ones treated with ColE9–Imm9 only (i.e., control), that showed no difference in growth, and the ones treated with ColE9–Imm9 plus DTT (i.e., experimental). Bacterial growth (i.e., optical density) decreased according to the concentration of ColE9–Imm9 plus DTT. The *E. coli* BL21(DE3) cells treated with the higher concentration of 100 nM of ColE9 plus DTT showed a marginally lower optical density (employed as an indicator of bacterial growth) than the *E. coli* BL21(DE3) cells treated with 0 nM and 10 nM of ColE9-Imm9 plus DTT. Those results were expected and support the findings of Vankemmelbeke and colleagues (2005) who challenged a colicin-sensitive *E. coli* strain, DPD1718, with a range of concentrations of the ColE9–Imm9 complex, from 0.4 to 40 nM and observed dose-dependent outcomes.

The hypothesis that the expression of ColA translocation domains could protect sensitive cells such as *E. coli* BL21(DE3) from externally added ColE9 was tested in a liquid growth inhibition assay, namely by monitoring the optical density of the growing cells after treating them with ColE9–Imm9 plus DTT. The optical density was not expected to be affected by ColE9–Imm9 plus DTT, even when cells were exposed to a high concentration of ColE9–Imm9 (100 nM), when compared to the untreated cells. As anticipated, the results confirmed the hypothesis as shown in Figure 5.8. Also, the results revealed that there is a significant difference ($p<0.001$) between the optical cell density of colicin sensitive strain *E. coli* BL21(DE3) (i.e., D/C) and the other constructs that gained protection against the antimicrobial activity of ColE9-Imm9 plus DTT. The same hypothesis was proposed by Bouveret and colleagues (2002), who observed in their study that the Tol–Pal system can be disturbed by the periplasmic production of the G3p protein, a protein produced in the inner membrane of *E. coli.* Among other results, they found that the interaction of the overproduced G3p with the Tol proteins inhibited their normal function—that is, the translocation of colicin.

Those findings prompted the question that guided this research: ‘‘What if the domains produced in the periplasm can be derived from colicins as well, instead of testing G3p we test the translocation domain of colicin’’?

The results of spot testing and liquid growth assays support the hypothesis that the expressed TA, TolA box and TolB box of ColA in the periplasm provide a certain degree of protection to the cells against ColE9–Imm9 plus DTT. Quantifying that degree of protection requires determining whether the level of protection afforded is the same for all examined cells needs to be applied. The implications of this statement raise a few important questions. Firstly, does the fact that ColE9 interacts only with TolB only for its translocation and cytotoxicity provide the greatest level of protection? Secondly, how can the expressed TolA box protect cells, especially since ColE9 does not appear to interact with TolA for its translocation, or, at least, not directly? The exact mechanism of ColE9 translocation needs to be more thoroughly investigated in order to pinpoint the exact role of TolA, which may be accomplished by focusing on cells that express the TolA box of ColA and by studying their status upon being treated with ColE9–Imm9 plus DTT.

Given current knowledge about the mechanism of ColE9 translocation, it can be expected that TA overproduction in the periplasm not only causes the interaction between cellular Tol proteins and TA but also causes TA to occupy the binding sites with ColE9, making them unavailable for ColE9 translocation in the periplasm. As explained in Section 5.1, ColE9 binds to TolB in the TolB box for its translocation into the cells. By extension, our hypothesis was that the TolA box’s overproduction does not afford the same level of protection as the overproduced TA and TolB box, because simply ColE9 doesn’t interact with TolA directly for its translocation through the periplasm. However, the results of the liquid growth inhibition assay indicated otherwise, and raised the question about the contribution of TolA to ColE9 translocation when separated from the TolB box. To measure the level of protection

resulting from the overproduction of the TolA box and compare it to both TA and TolB box, a novel approach was applied: live cell imaging.

The BacLight™ kit is a differential stain comprising SYTO 9, which is able to enter all cells staining them green, and propidium iodide (PI) which enters those cells with a compromised membrane and displaces SYTO 9 staining the dead cells red. Viable cells remain green. In treating the bacterial cells with the ColE9–Imm9 protein, which has a cytotoxic activity, if the bacteria were sensitive, then their membrane would be damaged, thereby leading to eventual cell death, and PI would displace SYTO 9 and stain the cells red. A yellow fluorescence, meaning damaged cells, is commonly observed with BacLight™ kits, and research has shown that PI cannot displace SYTO 9 completely, which probably accounts for the evolution of yellow fluorescence (Stiefel, 2015). Brown and colleagues (2015) used live cell imaging to investigate whether culturing colicin-sensitive bacterial cells with an antimicrobial such as gentamicin would provide any protection against ColE9. Although that hypothesis differed from the one presented in this thesis, relying on such a technique to quantify the level of protection against ColE9 supported the choice of this methodology.

The first experiment involved testing *E. coli* BL21(DE3) cells (See Figures 5.9, 5.10 and 5.11), which were sensitive to ColE9–Imm9 plus DTT in the antimicrobial activity assay in liquid culture. The findings supported the results of the liquid growth assay, that the effect of ColE9–Imm9 plus DTT on the bacterial cell depends upon the concentration. Compared with the control *E. coli* BL21(DE3) cells, a significant ($p < 0.001$) decrease in the percentage of living cells occurred when treated with 10 nM ColE9–Imm9 plus DTT reached 64 % after 18 h of incubation, which decreased even further with higher concentration of ColE9-Imm9 plus DTT (i.e., 100 nM) to become only 33 %. However, against expectations, no significant difference arose at the beginning between the treated and untreated living cells after 3 h of incubation, but it significantly a decrease in the percentage of living cells after both incubation

periods, as shown in Figure 5.12, which emphasises how the concentration of added ColE9-Imm9 plus DTT concentration impacts the decrease of living cells.

The same method was applied to evaluate the other *E. coli* BL21(DE3) cells carrying plasmids for TA, TolA box and TolB box. After expressing TA and achieving the induction of cells with L-arabinose, the cells were prepared for CLSM in the same way as the colicin-sensitive *E. coli* BL21(DE3) cells (see Figures 5.13, 5.14 and 5.15). Because the major function of TA is ColE9 translocation, it is reasonable to assume that expressed TA would provide the greatest level of protection represented in high percentage of living cells to the producing cells against ColE9–Imm9 plus DTT. The results included an unexpected difference in the percentage of living cells when comparing to the expressed TolA and TolB boxes. Figure 5.16 shows the difference in the difference in the percentage of live versus dead of cells expressed TA, the living cells were found to be 69 % and 59 % after treatment with 10 and 100 nM ColE9-Imm9 plus DTT. This is completely different from the results achieved with sensitive strain (*E. coli* BL21(DE3) in Section 5.3.7 where the percentage of living cells were only 33 % due to the antimicrobial activity of ColE9-Imm9 plus DTT. Also, the variability shown in Figure 5.16 clearly does not relate to ColE9–Imm9 cytotoxic effect, unless, possibly, if Tol proteins other than TolA and TolB are responsible for ColE9 translocation, which would explain why the level of protection was not greater. Even with the high concentration of ColE9–Imm9 plus DTT treated cells, the slight difference between the values was not significant ($p > 0.001$). Even so, the findings support the hypothesis, because the number of living cells among treated cells was relatively high indicated that TA overproduction in the periplasm provided some protection against ColE9, but not as great a level as was expected. As mentioned in Chapter 1, of the Tol proteins—for example, TolA, TolB, TolQ and TolR—the involvement of TolQ and TolR in ColE9 translocation has not been fully studied. However, Kim and colleagues (2014) were able to identify the contribution

of TolR in the translocation of ColE3 (shared 80% identity with ColE9), and suggested that TolR interacts with TolA by forming a complex that produces energy for ColE3–Imm3 dissociation and eventually aids the translocation of ColE3 into cells. If the same is true with ColE9–Imm9, it could explain why the protection provided by TA (i.e., TolA box and TolB box) was not as high as expected, namely due to TolR involvement. Another study suggesting the same theory was conducted by Journet and colleagues (1999), who showed that the overproduction of TolR central domain in the periplasm worked to protect the colicin-producing cells against ColA. A similar result was later achieved by Bouveret and colleagues (2001) while investigating how TolR overproduction affected ColE3. Their explanation was that the possible interaction of the overproduced TolR with cellular TolR blocked ColE3 translocation. This raised another question, however. if TolR box is overproduced in the periplasm, would it provide the same level of protection as TA?

Tol boxes were investigated separately. In particular, because ColE9 interacts with TolB in TolB box for translocation and requires energy from TolA for Imm9 release, how the overproduction of TolA and TolB boxes affects ColE9 action in cell viability was assessed using CLSM. Because the interaction between ColE9 and TolA has not yet been confirmed, and because their energy involvement via indirect interaction, the protection provided by the overproduction of the TolA box in *E. coli* BL21(DE3) box (TolA) was expected to see a substantial number of dead cells via live/dead staining due to the cytotoxic action of ColE9. Surprisingly, the results were similar to the ones achieved by TA overproduction (see Figures 5.17, 5.18, 5.19). Despite a significant difference ($p < 0.001$) in bacterial growth of all tested cells as shown in Figure 5.20, it is an expected variability in bacterial growth. However, live cell counts exceeded dead cell counts and remained close to rates in the control group and never exceeded the range of living cells (i.e., 60-70%), which is unexpected considering that generally the status of bacterial growth after 18 h (late stationary phase) means that the

numbers of live and dead cells are beginning to converge. Moreover, a decline in the number of dead cells has been observed, and this is because during incubation period the bacterial cells are still dividing, causing an increase in the proportion of live versus dead cells overtime. ColE9 is known to lack TolA box (Carr *et al.*, 2000), and the significant level of protection against ColE9 conferred by expressing TolA box in the periplasm, which occurred in 70% of living cells, indicates a major role for TolA in ColE9 translocation. It may be that the energy transferred to a TolB-bound ColE9, probably via a conformational change caused by TolA–TolB interaction, prompted Imm9 release and the translocation of the cytotoxic domain to its target (Vankemmelbeke *et al.*, 2009), which would explain the protection. Several colicin–TolA interactions have been investigated using agar overlay technique, surface plasmon resonance and two-hybrid screening for colicins A and E1 (Lazdunski *et al.* 2002). *In vivo* cross-linking and co-immuno-precipitation have revealed a complex of TA, TolA and TolB, the formation of which may contribute to translocation, in which the overproduced TolA box in *E. coli* BL21(DE3) cells would occupy Tol and prevent both the formation of the TA–TolA–TolB complex and consequent translocation of ColE9. Considering that possibility, scholars have classified TolA as a *hub protein*, a class of proteins involved in numerous interactions with other proteins and that are critical to maintaining the stability and function of the interaction network. Hub proteins are classified into two groups: *transient hubs*, which are engaged in only one interaction at a time, and *obligate hubs*, which can be involved in multiple interactions at once. Other names and descriptions classifying hub proteins have been introduced, including “date” and “party” as well as “sociable” and “non-sociable”, according to their number of binding sites, their size and their structures (Carro *et al.*, 2018). Researchers continue to investigate whether hub proteins have specific structural properties to facilitate their participation in various interactions with other proteins, as well as to determine whether they have any structural flexibility for folding and changing conformation.

Because hub proteins can be recognised by their disordered regions, they are large unfolded proteins with no secondary or tertiary structure (Patil *et al.*, 2010). Penfold and colleagues (2012) characterised TolA as a hub protein after discovering its binding site, or TolA box, where it interacts with colicin. However, because no direct interaction occurs between ColE9 and TolA, the interaction between ColE9 and TolB causes a conformational change in TolB, which consequently causes an interaction between ColE9 and TolA, resulting in the provision of the energy required for Imm9 release and ColE9 translocation. TolA also has structural flexibility between the OM and IM and participates in most protein–protein interactions, including those with other cytoplasmic Tol proteins (i.e., TolQ and TolR). To pinpoint the mechanism of the protein–protein interaction, an *in vivo* technique such as within-cell NMR needs to be developed, which justifies the second aim of this research: to visualise overproduced domains and determine their structural conformational changes before and after challenging the cells with ColE9.

Because the TolB box pentapeptide is essential for ColE9 killing activity (Garinot-Schneider et al., 1997), it is reasonable to speculate that if the TolB box is overproduced in the periplasm, it will block the interaction between ColE9 and cellular TolB. Thus, the level of protection was anticipated to be higher than for TolA and TA. Unexpectedly, however, the results obtained (see Figure 5.21, 5.22, 5.23) were similar to those of the expressed TA cells. The percentage of living cells was 54 % and only slightly less than that of living cells from treated cells with expressed TA (i.e., 59%). Again, a significant difference between the bacterial counts emerged, as shown in Figure 5.24, which is clearly related to normal variability in bacterial growth. Even so, the percentage of living cells still higher than dead cells (54%). However, the finding did not relate to ColE9–Imm9 cytotoxic effect upon concentration as seen with colicin sensitive strain *E. coli* BL21(DE3) as demonstrated in Section 5.3.7.

Overall, percentages of living *E. coli* BL21(DE3) cells expressing TA, TolA box and TolB box ranged from 50% to 70% following treatment with ColE9-Imm9 plus DTT, the highest being with *E. coli* BL21(DE3) cells expressing TolA box. That finding suggests that TolA has an important function in ColE9 translocation and proves its function as a hub protein. The lower percentage of living cells expressing the TolB box of ColA presumably related to differences in the affinity of the TolB boxes of ColE9 and ColA, approximately 1 μM (Loftus *et al.*, 2006) ColA and 10 μM respectively, were found to bind with TolB, and the low binding affinity could explain the unexpected result. The finding could also agree with the suggestion of Hands and colleagues (2005) that residues outside the TolB box region have a high affinity for periplasmic TolB, which could also explain the level of protection that we observed.

Using a liquid growth assay, CLSM and plate counting, this research revealed the significant difference ($p<0.01$) between cells due to the effect of ColE9-Imm9 cytotoxic activity after its reduction with DTT on *E. coli* BL21(DE3)-sensitive cells, and how that effect depends on the concentration of ColE9–Imm9 as was also demonstrated by plate counting as illustrated in Figure 5.25. Testing the same protein on *E. coli* BL21(DE3) expressing TA, TolA and TolB revealed that the overproduced TA, TolA and TolB provided some protection to cells against ColE9 activity. The presence of those proteins in the periplasmic space blocked ColE9 translocation due to other interactions between the overproduced proteins (TA, TolA box and TolB box) and the cellular TolA and TolB of the host cells rendering them unavailable to interact with ColE9, as shown using three methods. The statistical findings reported herein indicate, however, that the level of protection is not cumulative. Because ColE9 interacts directly with TolB, overexpressing TolB box and blocking the interaction with ColE9 should have afforded a high level of protection, as represented in a higher live cell count. However, this was not the case. Even considering TolA indirect role in Imm9 dissociation from the complex, cells that express the whole TA should show a high level of protection. Those

outcomes indicate that a series of interactions involving the Tol system and colicins have yet to be discovered, and identifying them will involve grasping just how large and structurally disordered the translocation domain is. To that end, in-cell NMR visualisation could allow confirming the novel findings in this chapter.

Chapter 6
Development of methodologies to study protein-protein interactions of overexpressed translocation domain of colicin A using in-cell nuclear magnetic resonance

6.1 Introduction

One of the common methods in studying protein-protein interactions (PPIs) is nuclear magnetic resonance (NMR). As mentioned in Chapter 1, the use of NMR has proven its ability to reveal protein status inside the cells (Ito & Selenko, 2010). This chapter discusses the applicability of in-cell NMR spectroscopy to the analysis of PPIs of the overexpressed translocation domain of Colicin A residues 1-172 (TA), TolA box (residues 52-172), and TolB box (residues 1-52) in *E. coli* BL21(DE3) cells. In order to provide optimal results, the most important parameters that should be monitored during in-cell NMR are chemical shifts, signal intensities, and linewidths. The chemical shift is the resonant frequency of particular nuclei relative to a standard in a magnetic field and it represents the position on the δ scale (in ppm) where the peak forms in the NMR spectrum (Williamson., 2013). Chemical shifts result from differences in the local environment around particular nuclei, for example, from changes in the electron density as the result of the application of an external magnetic field that induces motion in the electron cloud surrounding the nucleus. These changes cause electrons to generate their own magnetic fields, which contribute to the shifts observed. The presence of hydrogen bonds and charged molecule can also similarly result in and/or contribute to such resonance shifts (Balci, 2005). Signal intensities refer to the magnitude of NMR signals which are proportional to the concentration of the sample. Thus, a small or dilute sample will give a weak signal, while a concentrated sample increases the signal strength proportionally and will influence the sharpness of the formed peak; The stronger the signal, the easier for NMR detection. Lastly, the linewidths describe the width of the peak; narrower peaks are more distinctive and easier to identify (Wang *et al*., 2015).

Kumar *et al*. (2013) have found significant benefits in the use of in-cell NMR in studying proteins. As mentioned in Chapter 1, to analyse a protein inside cells, two-dimensional (2D) $^{1}H–^{15}N$ correlation is the most widely used method, which can allow the characterisation of two NMR sensitive nuclei, like ^{1}H and ^{15}N, such that each $^{1}H–^{15}N$ correlated peak represents an amino acid in the backbone of the protein. That method is easily applied and requires only the preparation of ^{15}N-labelled protein samples to allow the visualisation of any variations in the chemical shifts of the peaks (Kumar *et al*., 2013). The principal challenge of this method is when it produces overlapping cross-peaks which is a common problem in in-cell NMR analysis, which makes it difficult to identify every individual peak. Therefore, this requires additional higher dimensional experiments, which will distribute those peaks over a larger space facilitating their identification and thereby helping to assign the protein structure (Felli *et al*., 2014). However, higher dimensional NMR is limited and not applicable when studying large proteins because of the likelihood of overlapping cross-peaks (i.e., >30 kDa) or ones that bind to large structures (Burz *et al*., 2009). Similarly, other researchers have shown that small proteins (i.e., 6-14 kDa) are easier to analyse via in-cell NMR than large ones (Pielak *et al*., 2009). Accordingly, the initial aim of the work described in this chapter, was to visualise the expressed TA using NMR. Although the size of TA (17 kDa) is considered to be small and thus should be detected easily by in-cell NMR, the approach chosen was to simplify the protein sample by separating the TolA box domain (12 kDa) and TolB box domains (5 kDa) of the intact TA region in order to further facilitate the analysis as according to reported studies, this is likely to reduce the complexity of the resulting peaks (Selenko *et al*., 2007).

The translocation domain of ColA (TA) is an intrinsically disordered region. *In vitro* NMR reveals that when other protein(s) interact with TA, conformation changes occur to TA itself as well as the other interacting protein(s). This was proven by showing the conformational change in TolA when it interacts with TA using this technique. This demonstrates interaction occurring between TA and TolA and raises the question of what occurs after translocation of TA into the periplasm inside the bacterial cells (Hecht *et al*., 2009). To date, researchers in the field continue to seek the best way to simplify the analysis of target proteins inside bacterial cells and to improve their capacity to provide valuable information (Briendel *et al*., 2019). And because in-cell NMR is powerful tool to visualise the structure and the dynamics of intrinsically disordered proteins such as TA. The contribution of this project in this area is an attempt to develop a methodology with appropriate conditions for visualising the expressed TA inside *E. coli* cells using in-cell NMR, and to subsequently determine its structure.

This chapter's aims were firstly to provide the appropriate conditions necessary to provide high-level recombinant protein expression, maintain protein stability and permit efficient labelling of the expressed proteins. Achieving this aim requires the development of a novel in-cell NMR methodology for the analysis or visualisation of TA interactions within *E. coli* cells. ^{1}H NMR (Figure 6.1) is the fastest 1D NMR technique, that detects hydrogen atoms. Every region of the ^{1}H spectrum is a defined region for a functional group. It is also helpful in determining the conformational status of the protein. If the observed NMR peaks are broad and narrowly dispersed (i.e., the peaks cover only a small area of the spectrum), this means that the protein is either partially folded or completely unfolded (Oktaviani, 2014). While well-dispersed narrow peaks means that the protein is folded. This is because in folded proteins each proton has a distinct conformation which depends on its position in the tertiary structure of the protein (Page *et al.*, 2005). This is shown in Figure 6.2.

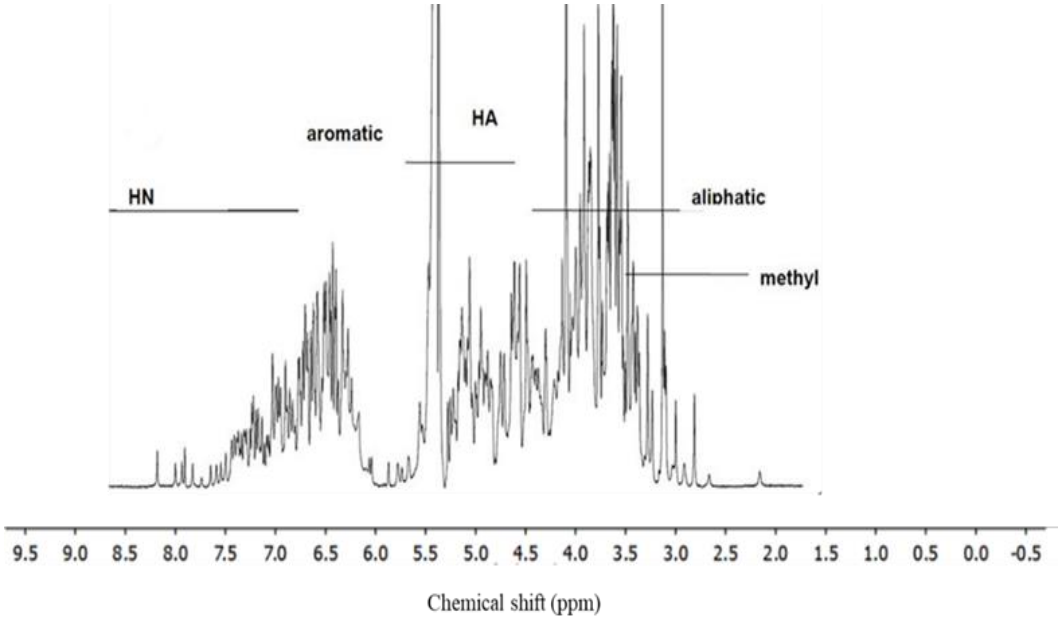


Figure 6.1 General 1H NMR spectrum of protein (Chou, 2008). The chemical shifts of H atoms in methyl side chain (δ= 0-2 ppm); H atoms of the aliphatic side chain (δ= 2-4 ppm); the HA atom of exchangeable amide (δ= 5-5.5 ppm); H atoms of the aromatic groups (δ= 6-8 ppm); the HN atom of the amide group (δ= 7-10 ppm).

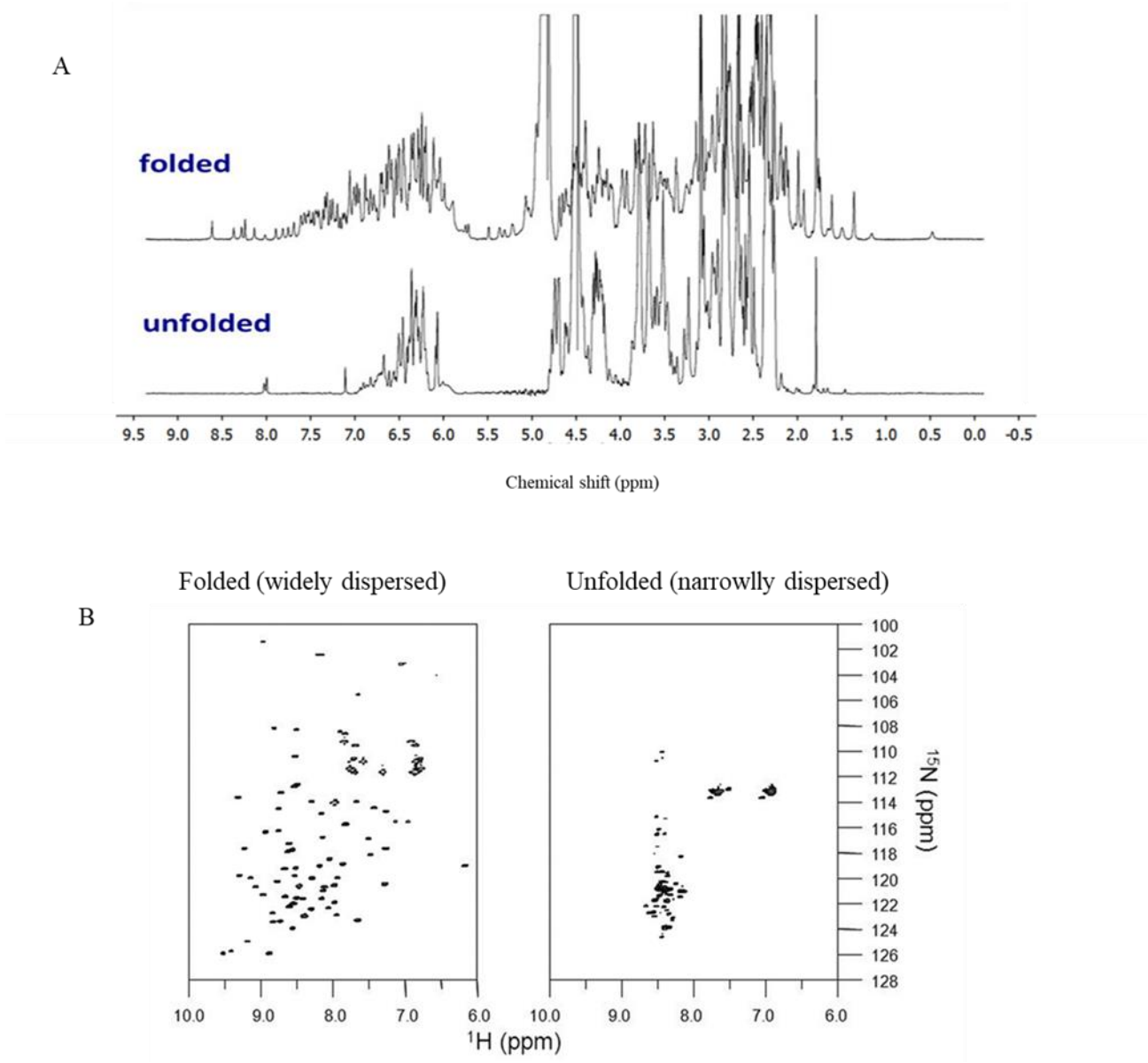


Figure 6.2 General 1D and 2D NMR spectrum of a folded and unfolded protein (Rehm *et al*., 2002). A) 1D NMR peak of the folded and unfolded protein spectrum. The peaks are more broadly dispersed over the spectrum, especially between chemical shifts δ= 7-10 ppm; The peaks are narrower and sharper than in the unfolded protein spectrum. B) 2D NMR of folded and unfolded proteins. The signal in the folded protein is more dispersed and distinctive while the peaks are overlapping in the unfolded protein.

Determination of protein conformation is crucial for understanding its functions and interactions. Dispersion of the ^{13}C and ^{15}N nuclei which are sensitive to amino acid sequence helps with the resonance assignment of unfolded and partly folded proteins. The intrinsic chemical shift dispersion for ^{15}N, ^{1}H, and ^{13}C, has been evaluated using full resonance assignment data for unfolded apomyoglobin, a well characterized globular protein that has been extensively employed as a model system for protein folding and stability studies, with two other unfolded proteins and five folded proteins. The dispersion of ^{13}C, ^{1}H resonances for the unfolded proteins is poor, whereas the dispersion of ^{15}N, ^{1}H-^{15}N and ^{13}CO is much greater, because of the sensitivity of these nuclei to the nature of the adjacent amino acid in the primary sequence. By contrast, the dispersion of the ^{13}C, ^{1}H nuclei are much greater in the folded proteins, because of the well-characterised environments around these nuclei and the known effects on secondary and tertiary structure. These differences in chemical shift dispersion necessitating the use of different strategies for accurate resonance assignment in unfolded proteins compared to those most commonly used for folded proteins (Yao, *et al.*, 1997).

6.2. Overview of methodologies

The main challenge presented in this study was the development of the in-cell NMR technique itself. Optimisation of the methodology required careful consideration of the key parameters, in particular the biological system under study, and the labelling technique utilised. In this subsection, molecular and biological techniques, labelling techniques, and experimental design are discussed.

6.2.1 Molecular and biological systems techniques

Analysis of macromolecules at the cellular level is facilitated by the use of a biological system that is easy to manipulate, like *E. coli* as discussed in Chapter 1. The first in-cell NMR study described by Serber *et al*. (2001) was performed in *E. coli* cells using a T7-dependent overexpression system. *E. coli* has previously been used as a model organism in the in-cell NMR field for analysis of both eukaryotic and prokaryotic systems. In this study, *E. coli* BL21(DE3) was the host used for expression of the target proteins and was used to determine the suitability of different overexpression systems, a T7 promotor-based system (pET-15b) for cytoplasmic protein expression, and an *ara*BAD system (pBAD/gIIIc) for protein expression and translocation into the periplasm. Methods for the periplasmic expression of the target proteins (TA, TolA box and TolB box) inside the cells have been discussed previously in Chapters 3 and 4.

6.2.2 Isotope labelling technique

Isotopic labelling of macromolecules can be carried out uniformly or selectively as described in Chapter 1. Also, labelling can be achieved using ^{15}N or ^{13}C or both, depending on the nature of the experiment to be performed and the information required. The labelling technique does however require careful optimisation in order to obtain the best results. This chapter will now discuss the different approaches that were taken to label the target proteins.

6.2.3 Assigning experimental controls

Using controls is necessary at this stage to prove that the technique is working. In this study GB1 was employed as a control to demonstrate the potential to detect protein molecules using in-cell NMR. GB1 is the B1 immunoglobulin-binding domain of streptococcal protein G (6.24 kDa). This protein has been fully characterized by in-cell and *in vitro* NMR because of its stability, solubility, and its small size (Gronenborn & Clore., 1993). GB1 is a globular protein that gives well-resolved in-cell ^{1}H-^{15}N correlation spectra, and because it is not intrinsic to *E. coli*, there is a low chance of any interaction with the cytoplasmic components in the host cells (Selenko *et al*., 2006). Rather than studying the functional role of protein GB1, the intent is to detect its structure in the cellular environment. However, it is not considered to be a positive control for the target proteins in this project because they are different in their properties, they do not localize in similar compartments, and their secretion, distribution and their binding properties with other protein inside the cells are different (Xing *et al*., 2016).

6.4 Results

6.4.1 ^{1}H and ^{13}C NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA

Following the experiment in Section 2.2.7.2, ^{15}N isotope labelled NH_4Cl and ^{13}C isotope-labelled glucose are both added at the time of induction. In order to induce the expression of recombinant labelled protein. The 20% (v/v) in-cell NMR slurry prepared in a labelled M9 minimal medium: D_2O (90:10 ratio). Figure 6.3 shows the similarity between the ^{1}H NMR spectrum derived from labelled M9 minimal medium (i.e., uninoculated medium-only control) and the spectrum from *E. coli* BL21(DE3) cells expressing isotopically labelled TA in the periplasm in the same medium. They both demonstrate sharp signals associated with the labelled M9 minimal medium itself. Absence of characteristic peaks at the chemical shift for amide functional groups (NH) (i.e., 8-9 ppm) revealed no signs for the TA target protein.

Figure 6.4 shows the ^{13}C spectrum for the overexpressing cells and again there is no signal relating to the target protein. The characteristic crowd of peaks at the chemical shift regions that would represent the peptide chain of the TA target protein at δ =170-190, and 70-10 ppm are absent.

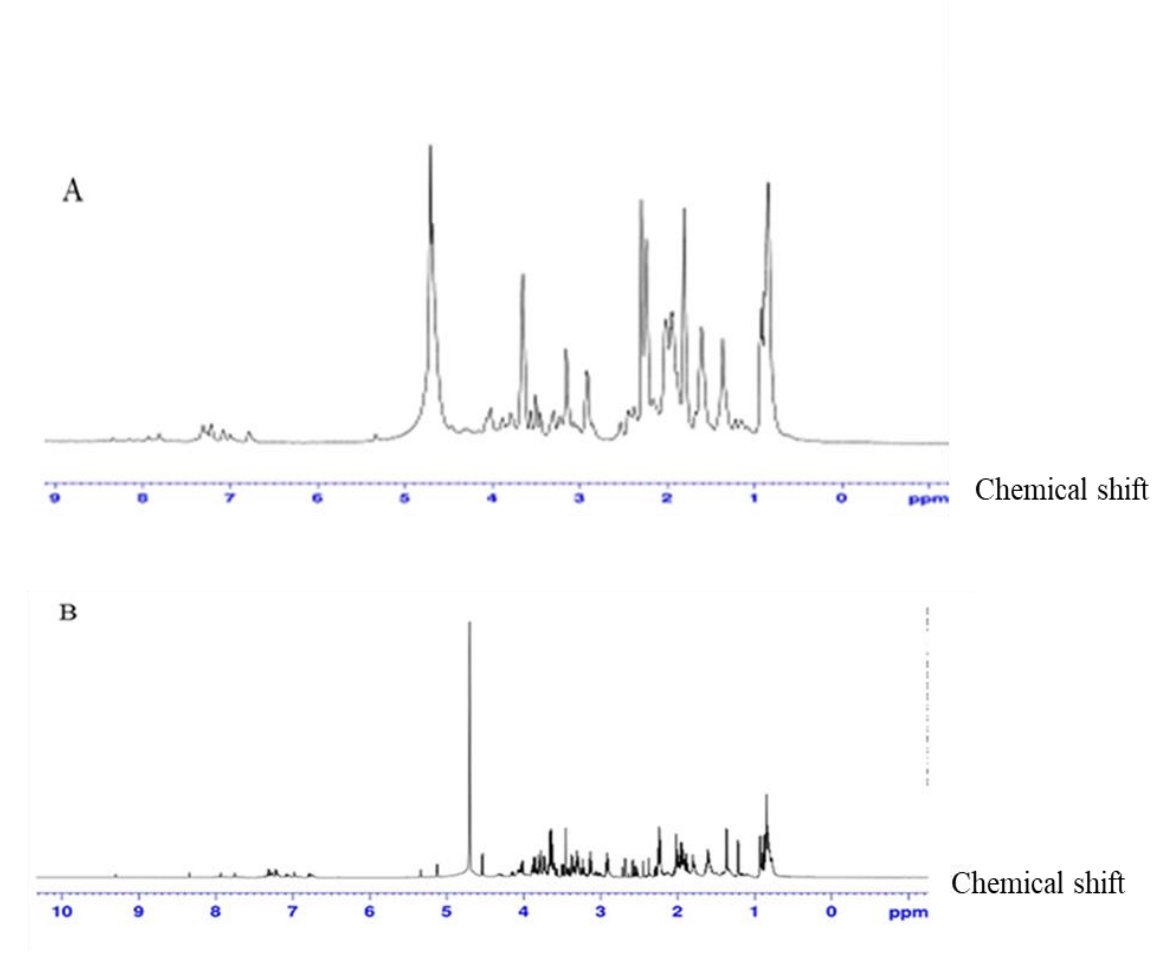


Figure 6.3 ^{1}H NMR spectrum of *E. coli* BL21(DE3) expressing ^{1}H-^{13}C isotopically labelled periplasmic TA. A) ^{1}H NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA. B) ^{1}H NMR spectrum of isotopically labelled M9 minimal medium (i.e., control), where 20% (v/v) in-cell NMR slurry was prepared in a labelled M9 minimal medium: D_2O (90:10 ratio). For both spectra the chemical shift region at δ= 0.7-2.4 ppm reveals sharp signals due to protons of casamino acids. Peaks at chemical shift δ= 3.5-5.5 ppm are due to glucose and thiamine. A strong peak at chemical shift δ= 4.9 ppm derives from H_2O. The small number of peaks at chemical shift region δ= 7-8 ppm represent proton exchange with D_2O.

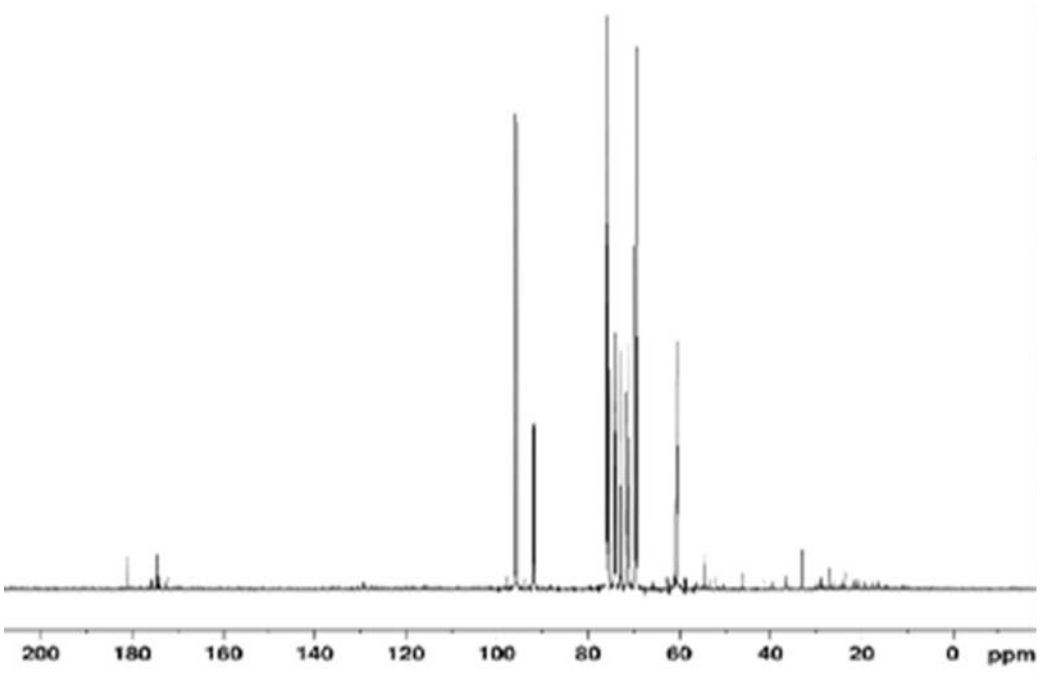


Figure 6.4 ^{13}C NMR spectrum of *E. coli* BL21(DE3) expressing ^{13}C-^{15}N isotopically labelled periplasmic TA. The 20% (v/v) in-cell NMR slurry prepared in a labelled M9 minimal medium: D_2O (90:10 ratio). The spectrum shows a few peaks relating to casamino acids at chemical shift δ=183-175, 60-15 ppm. Some peaks representing thiamine at chemical shift δ = 170, 130, 118, 65, 35-26 ppm. And characteristic peaks of glucose showing at δ = 100-60 ppm.

6.4.2 ^{1}H NMR spectrum of *E. coli* BL21(DE3) expressing ^{13}C-^{15}N isotopically labelled periplasmic TA (different solvent)

Following the same technique mentioned in section 2.2.7.2, instead of preparing the 20% (v/v) in-cell NMR slurry the spent medium, a fresh unlabelled M9 minimal medium: D_2O (90:10 ratio) was used. From the results shown in Section 6.4.1, both the sample and the negative control spectra show the same peaks that represent some of the components of the spent medium (the overexpression medium). As there is a possibility that these sharp peaks are masking the expected peaks for expressed TA protein, the media used for sample preparation was changed to unlabelled M9 minimal medium with 10 % D_2O. Figure 6.5 shows that less noise was present, but in the spectra for both the medium-only uninoculated control (A) and the sample containing expressed TA (B) no peaks were detected in the amide region, the presence of which would have indicated TA protein. Observed peaks were representative of components of the medium only, with the disappearance of the overlapping peaks in the spectrum as presented in Figure 6.5 (B).

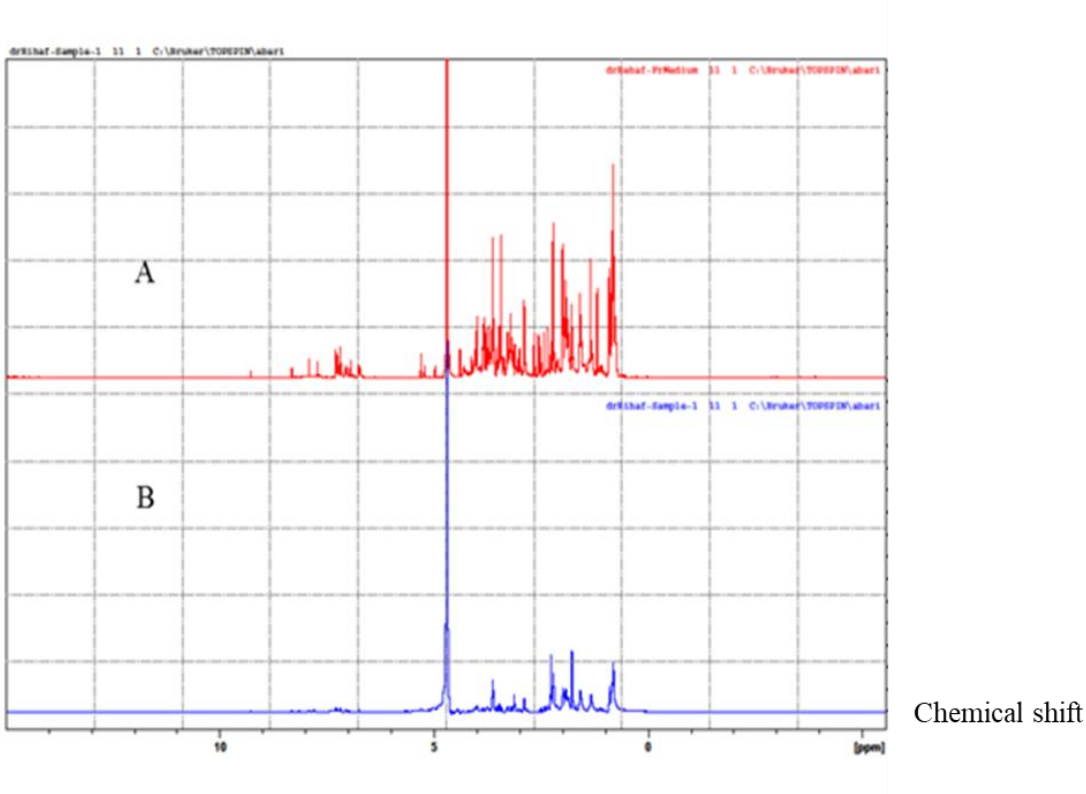


Figure 6.5 ^{1}H NMR spectrum of *E. coli* BL21(DE3) expressing ^{13}C-^{15}N isotopically labelled periplasmic TA (different solvent). A) ^{1}H NMR spectrum of the unlabelled M9 minimal medium (i.e., control). B) ^{1}H NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled TA, where 20% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10). Both are showing the chemical shift region at δ= 0.7-2.4 ppm reveals sharp signals due to protons of casamino acids. The peaks at chemical shift δ= 3.5-5.5 ppm are due to glucose and thiamine. And strong peak at chemical shift δ= 4.9 ppm for H_2O. Few peaks at chemical shift region δ= 7-8 ppm are representing proton exchange with D_2O.

6.4.3 ^{1}H-^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA (different labelling)

In order to simplify the peak detection by in-cell NMR, some modifications were employed to the experiment in Section 2.2.7.2. The overexpression performed in ^{15}N-labelled M9 minimal medium and the 20% (v/v) in-cell NMR slurry also, prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). ^{15}N labelled M9 minimal medium was used instead of ^{13}C-^{15}N labelled M9 minimal medium. Labelling the ^{15}N atoms only, instead of both ^{13}C and ^{15}N, gave better resolution of the peaks in the 1D ^{1}H spectrum. Those sharp peaks indicate methyl hydrogens in the side chains. No amide hydrogen in peptide backbone, that is related to target protein TA, was detected (Figure 6.6 (A)). Some new peaks were observed. in the 2D spectrum, narrowly dispersed over the spectrum with high background, most likely derived from an amide hydrogen backbone. These suggest the unfolded nature of a short peptide protein as shown in Figure 6.6 (B).

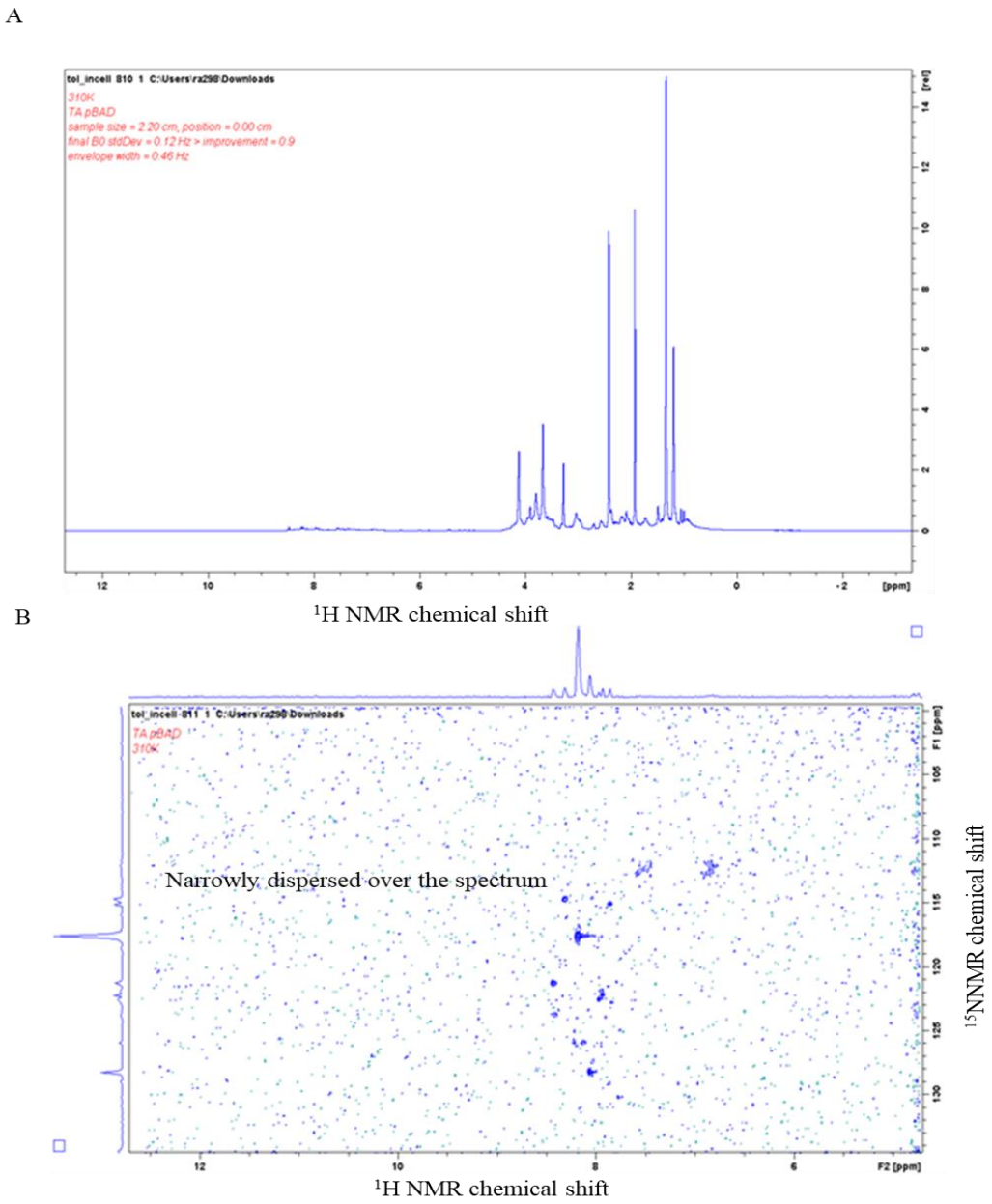


Figure 6.6 In-cell NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA (different labelling). The overexpression performed in ^{15}N-labelled M9 minimal medium and the 20% (v/v) in-cell NMR slurry prepared in unlabelled M9 medium: D_2O (90:10 ratio). A) ^{1}H NMR spectrum of expressed TA. Shows sharp peaks for side-chain methyl hydrogens at chemical shift δ= 0-2 ppm. Peaks representing Serine hydrogen atoms at chemical shift δ= 3 ppm, and peaks for hydrogen atoms in aliphatic side chains at chemical shift δ= 4 ppm can also be seen. B)^{1}H–^{15}N correlation spectrum of expressed TA. Shows the appearance of a few peaks lacking sharpness for Aspartate and Glutamine at ^{15}N chemical shift δ= 117 ppm; Phenylalanine, Tyrosine and Tryptophane residues at ^{15}N chemical shift δ= 120-130 ppm, that suggests the presence of the target protein comprising a short and unfolded peptide chain because of the high background.

These peaks were not observed in the non-induced *E. coli* BL21(DE3) cells (negative control), as sown in figure 6.7 A & B.

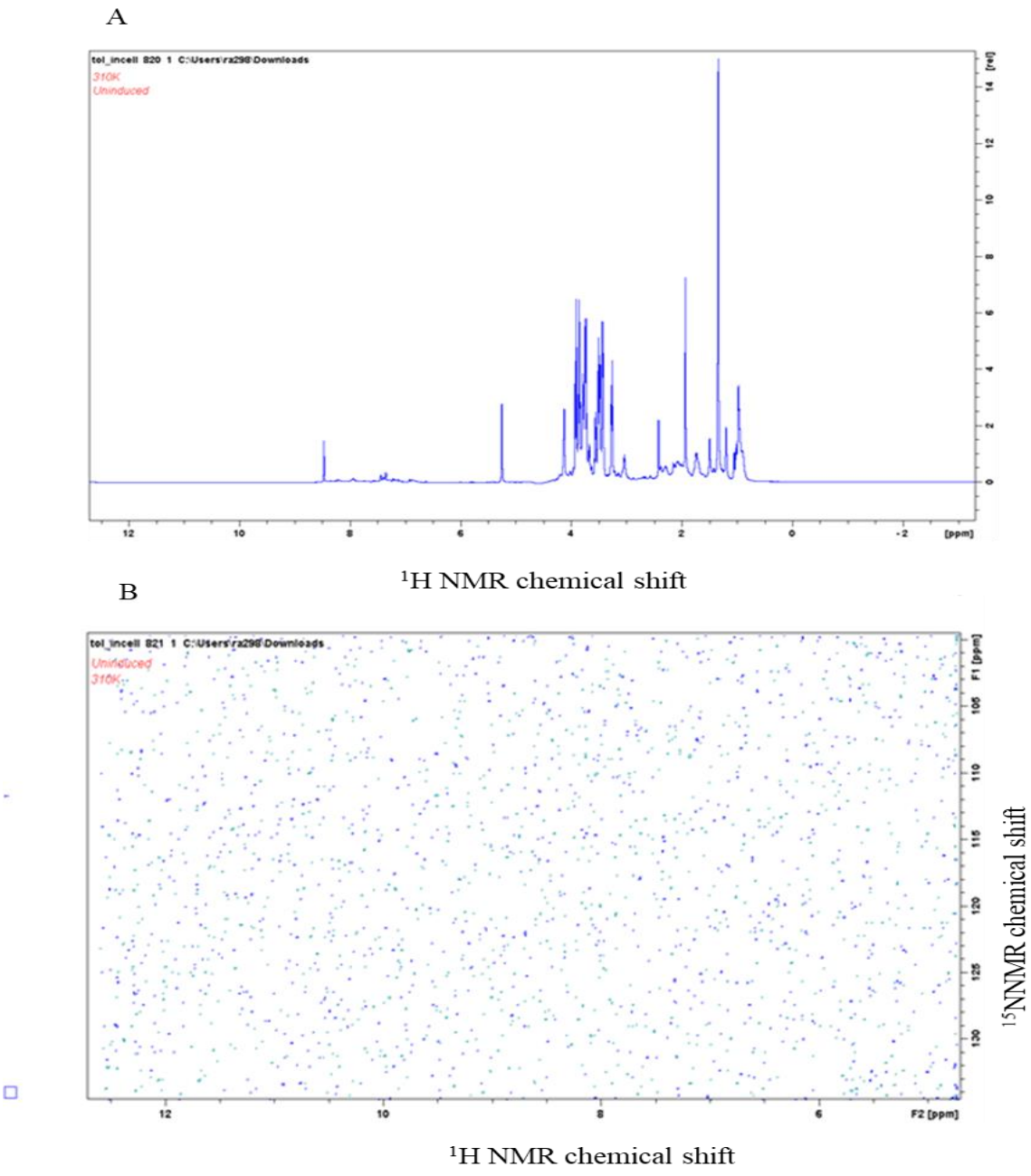


Figure 6.7 In- cell NMR spectrum of *E. coli* BL21(DE3) carrying pTA (non- induced cells). The 20% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). A) ^{1}H NMR spectrum of non- induced cells. Shows sharp peaks that represent the chemical shift region at δ= 0.7-2.4 ppm due to hydrogens of casamino acids present in the medium, while the peaks seen representing a chemical shift δ= 3.5-5.5 ppm are due to hydrogens in glucose and thiamine. B) ^{1}H–^{15}N correlation spectrum of non- induced cells. Shows absence of peaks and a high background level.

6.4.4 ^{1}H-^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA (different overexpression medium)

In order to optimise the detection of the target protein, other conditions were employed. Cultures of *E. coli* BL21(DE3) expressing TA were grown in a rich medium, LB, and were transferred to ^{15}N labelled M9 minimal medium immediately prior to induction, and the 20% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio)

Similar results were obtained to those described in section 6.4.3; a strong protein signal indicating structured TA was not detected. However, signals with high background level, indicating an unfolded short protein, were observed (Figure 6.8).

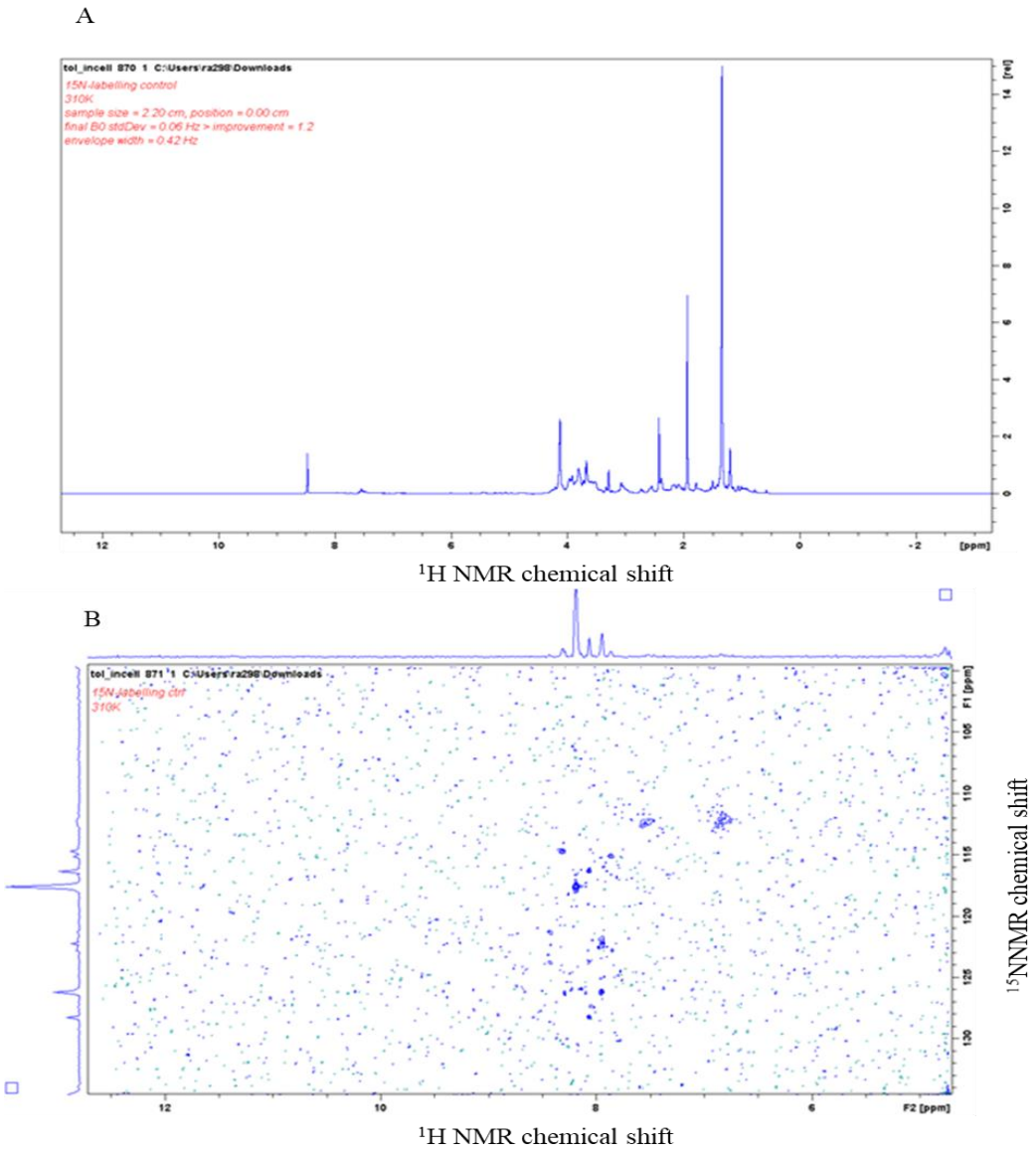


Figure 6.8 In-cell NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA (different over expression medium). Bacterial culture grown in LB medium then switched to ^{15}N-labelled M9 minimal medium at the time of induction and the 20% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). A) ^{1}H NMR spectrum of expressed TA. Shows sharp peaks for side chain methyl hydrogens at a chemical shift δ= 0-2 ppm. Peaks for hydrogens in Serine amino acid at chemical shift δ= 3 ppm. And peaks for hydrogens in aliphatic side chain at chemical shift δ= 4 ppm were also seen. B)^{1}H–^{15}N correlation spectrum of expressed TA. Few peaks are present other than a few poorly resolved peaks representing Aspartate and Glutamine at chemical shift δ= 117 ppm; and Phenylalanine, Tyrosine and Tryptophane residues at chemical shift δ= 120-130 ppm, that indicates the presence of the target protein as a short and unfolded peptide chain.

6.4.5 ^{1}H-^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled cytoplasmic GB1 protein

As described in section 6.2.3, GB1 has been extensively studied using in-cell NMR, and therefore it should be possible to detect it. Here, GB1 is employed as a positive control to confirm the validity of the technique being used. Continuing with the same experiment, where bacterial culture grown in LB medium then switched to ^{15}N-labelled M9 minimal medium at the time of induction. The 20% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). The 20% (v/v) slurry of *E. coli* BL21(DE3) cells expressing isotopically labelled GB1 protein generated weak signals of narrow peaks, representing low amounts of folded protein (Figure 6.9)**.** These results provided a reference to confirm the technique itself was working.

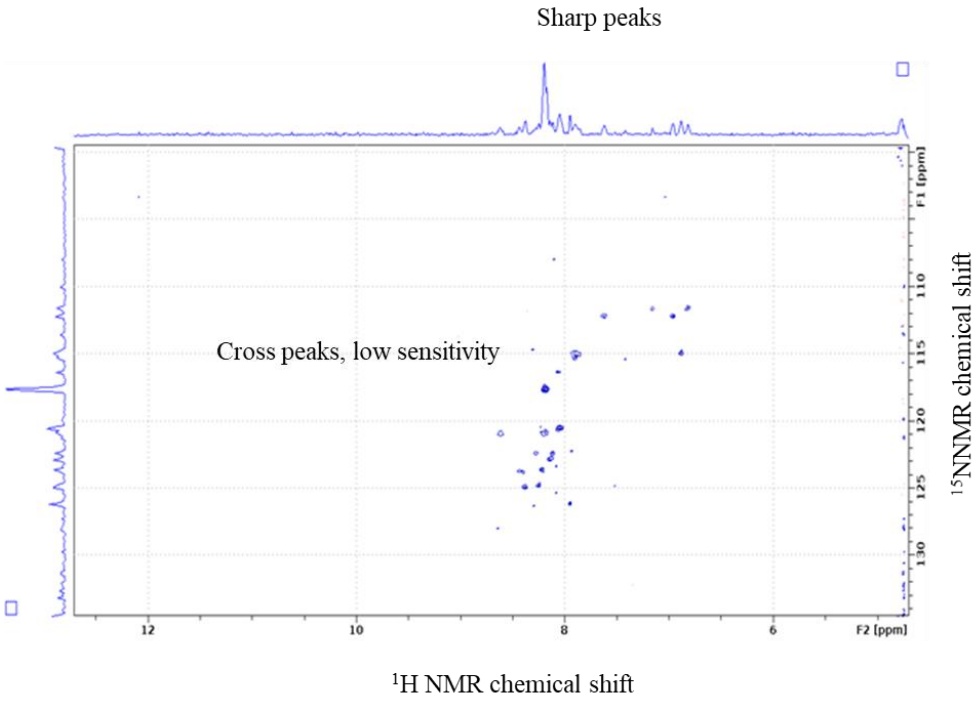


Figure 6.9 ^{1}H–^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled protein GB1. The bacterial culture grown in LB medium then switched to ^{15}N-labelled M9 minimal medium at the time of induction, the 20% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). Shows low intensity sharp peaks in the aromatic region δ= 7 ppm. Few low sensitive cross peaks in the amide regions of protein backbone at δ= 8-9 can be seen. The low background level indicates folded protein, while the weak signals indicate only small amounts of protein is present.

6.4.6 ^{1}H-^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled GB1 protein (high cell density).

The bacterial culture grown in LB medium then switched to ^{15}N-labelled M9 minimal medium at the time of induction. The culture of *E. coli* BL21(DE3) cells expressing GB1 protein was scaled up to a larger volume (100 mL) in order to make a more concentrated slurry (60% v/v) in an attempt to improve signal strength and thereby detection. The spectrum shown in Figure 6.10 (A) shows sharp peaks of high intensity relating to the GB1 protein backbone and side chain resembling those of the published 2D NMR of protein GB1 in solution *in vitro* as indicated in Figure 6.10 (B). This confirmed that detection of intracellular protein was possible using this in-cell NMR methodology.

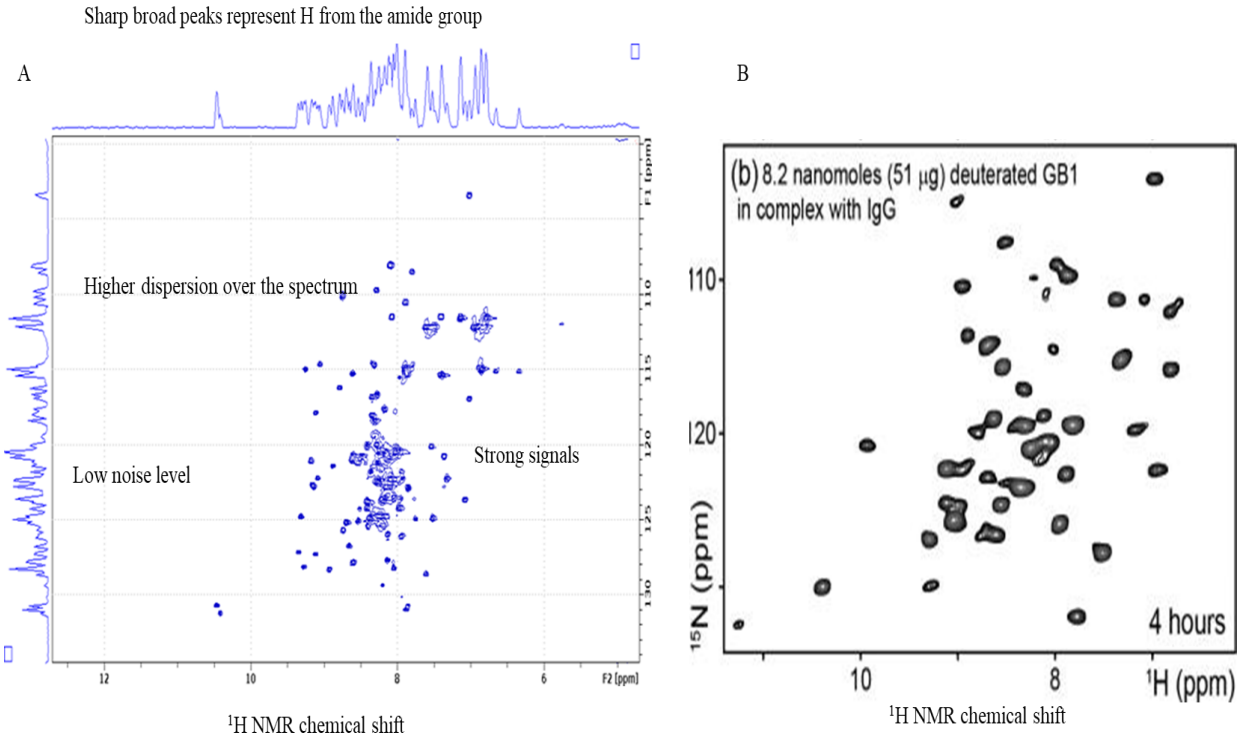


Figure 6.10 ^{1}H-^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled GB1 protein (high cell density). A) ^{1}H–^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing protein GB1. The bacterial culture grown in LB medium then switched to ^{15}N-labelled M9 minimal medium at the time of induction. The 60% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). B) Protein GB1 *in vitro* (Lamley et al., 2014). The results show similarity of sharp broad peaks in the amide region δ=8-10 ppm between the two spectra. The in-cell NMR method used produced strong signals that are highly dispersed across the spectrum with low background level, clearly representing GB1 protein as can be seen by comparison with the previously published data of Lamley and collegues (2014).

After in-cell NMR analysis, the bacterial cells were collected and the supernatant was checked by in-cell NMR, it indicated protein peaks in the extracellular medium, it is showing distinct and well-distributed cross peaks see Figure 6.11(A). while the collected cells showed protein peaks with the same cross peaks pattern as in (A) but some peaks are clustered together in the middle region of the spectrum in Figure 6.11(B). That indicates protein leakage outside the cells.

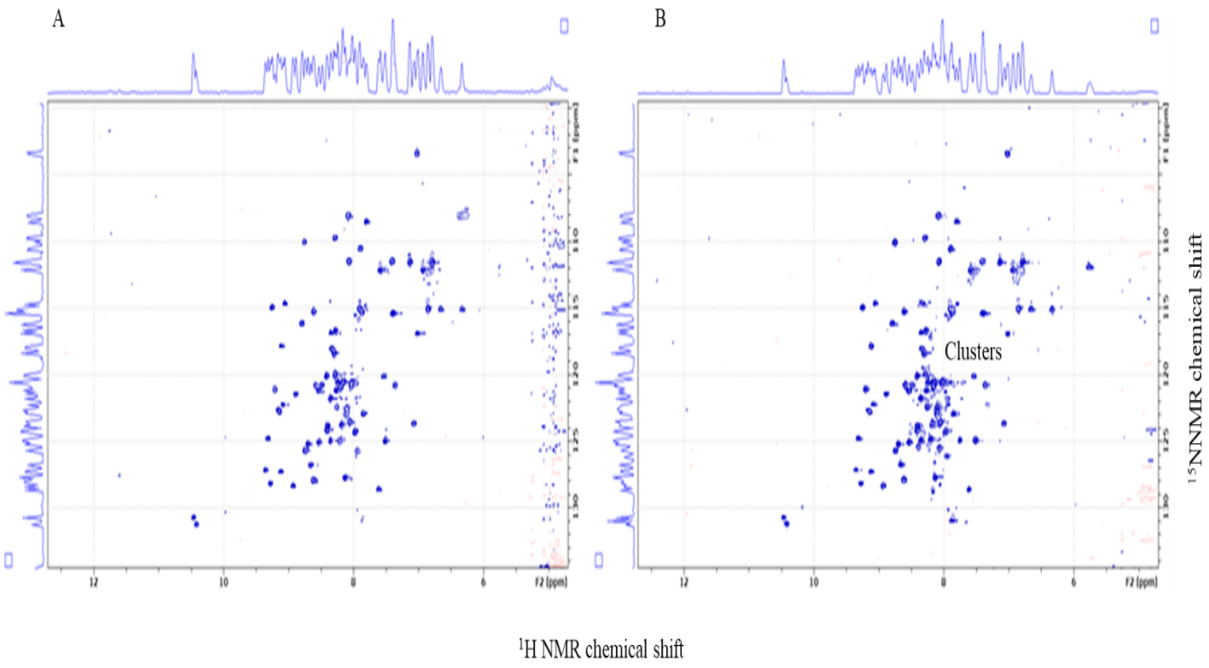


Figure 6.11 ^{1}H–^{15}N correlation NMR spectrum of GB1 protein after in-cell NMR analysis. A) ^{1}H-^{15}N correlation NMR spectrum of the collected cells. Shows the presence of sharp peaks of GB1protein inside the cells. B) ^{1}H-^{15}N correlation NMR spectrum of the supernatant. Shows the presence of GB1 protein with the appearance of clusters of some residues due to the sonication process. This indicates GB1 has leaked outside the cells during the expression.

6.4.7 ^{1}H-^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA (high cell density)

The same large scale (60% v/v) production conditions, which led to successful detection of expressed GB1 protein (Section 6.4.6), were applied to *E. coli* BL21(DE3) expressing TA. Figure 6.12 shows high intensity, narrowly dispersed broad peaks with a greater number of cross-peaks than seen in Figure 6.8 when 20% (v/v) slurry used at the same chemical shift regions. Although there is better resolution and more peaks are apparent, the presence of folded or structured protein was not detected. The partial distribution of those peaks indicates either partial association of TA with other protein or unfolding of TA. A follow up experiment was conducted to test if those peaks were derived from extracellular sources, indicative of protein leakage after cell collection. Figure 6.13 A & B reveal clear spectra of the supernatant (i.e., no protein leakage), while the spectrum derived from collected cells (Figure 6.14) shows very weak signals for disordered protein which may represent the unstructured part of TA.

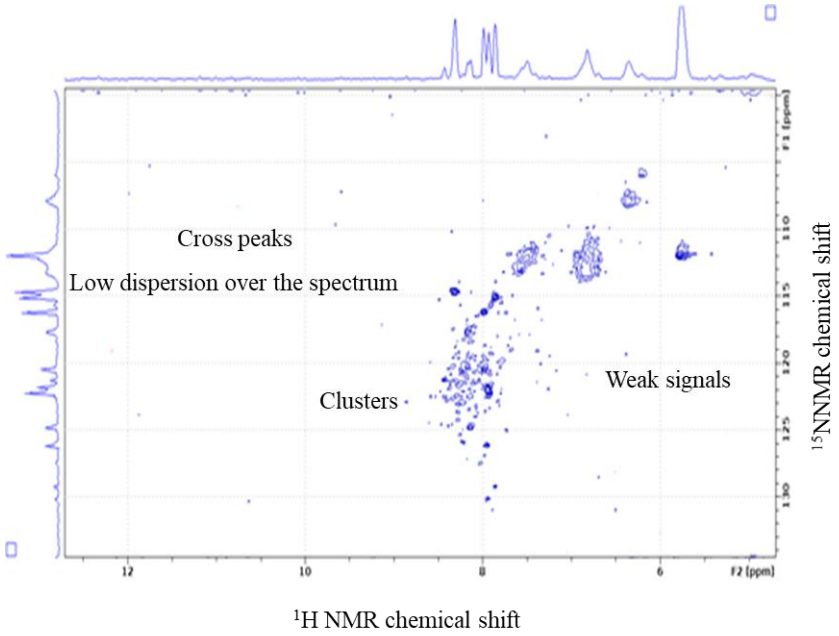


Figure 6.12 ^{1}H–^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled periplasmic TA (high cell density). The bacterial culture grown in LB medium then switched to ^{15}N-labelled M9 minimal medium at the time of induction and the 60% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). Shows sharp peaks for hydrogens in aromatic side chains at chemical shift δ= 7 ppm. Peaks for hydrogens in the amide region at ^{1}H chemical shift δ= 8-9 ppm. Cross peaks appeared for Aspartate and Glutamine at ^{15}N chemical shift δ= 117 ppm. Phenylalanine, Tyrosine and Tryptophane residues at ^{15}N chemical shift δ= 120-130 ppm. Cross-peaks that are narrowly dispersed over the spectrum. The presence of clusters of protein that indicates aggregation, the broadness of peaks represents the unfolded protein, the linewidth indicated the presence of a large molecule, the spectrum also showed a medium level of background. All indicate the presence of unstructured part of TA.

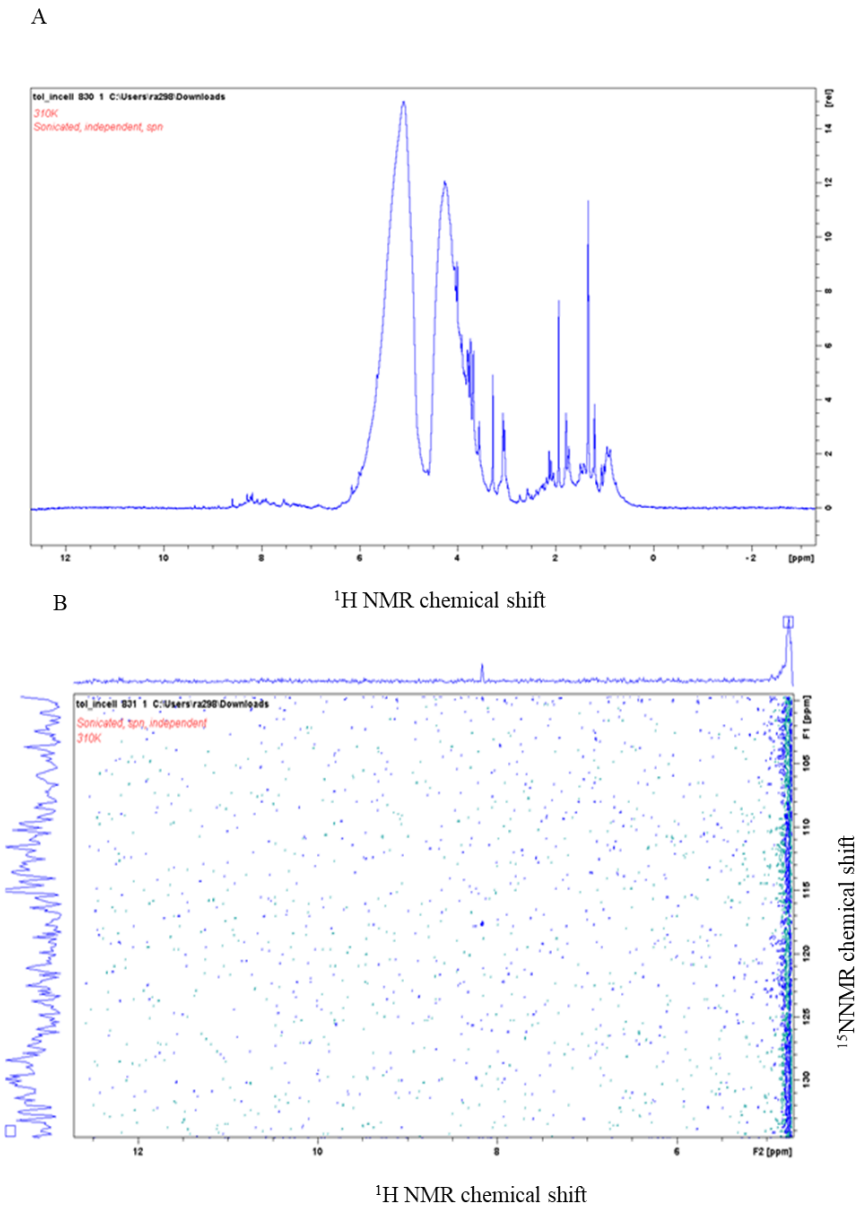


Figure 6.13 In- cell NMR spectrum of the supernatant of *E. coli* BL21(DE3) periplasmically labelled TA after cell collection. A)^{1}H NMR spectrum. Shows broad peaks in the aliphatic area for culture medium contaminants. B) ^{1}H–^{15}N correlation spectrum. There is no protein in the extracellular medium.

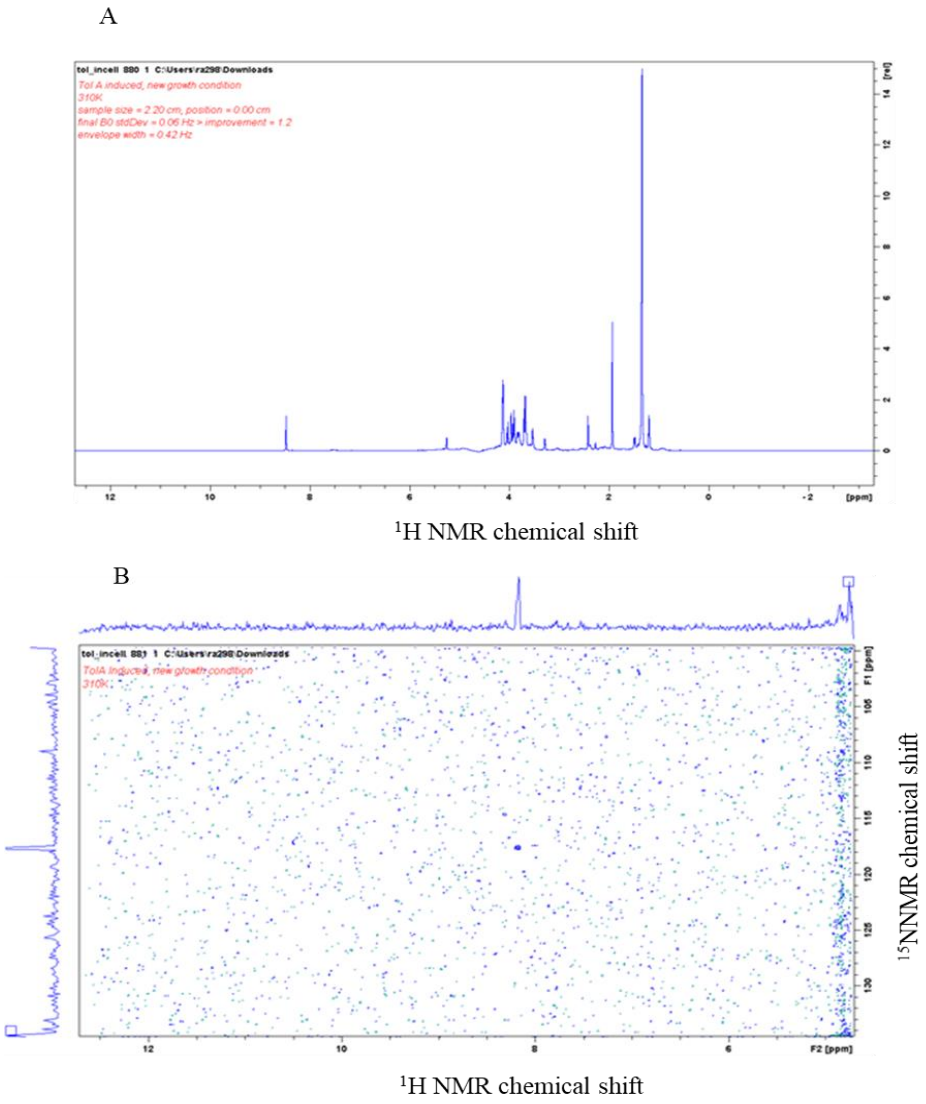


Figure 6.14 In- cell NMR spectrum of collected cells expressing TA. A) ^{1}H NMR spectrum. Shows sharp peaks for side-chain methyl hydrogens at chemical shift δ= 0-2 ppm. Peaks for hydrogens in Serine amino acid at chemical shift δ= 3 ppm. Peaks for hydrogens in aliphatic side chain at chemical shift δ= 4 ppm. B) ^{1}H–^{15}N correlation NMR spectrum. The spectrum shows a high background level which probably suggests the presence of unfolded protein which represents unstructured TA. The spectrum is consistent with a badly behaved protein.

6.4.8 $^{1}H–^{15}N$ correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled TolA (12 kDa) box

Continuing with the same method of in-cell NMR in Section 6.4.6, but on much smaller protein that is periplasmically expressed inside the cells as described in Chapter 4. The in-cell NMR spectrum obtained from non- induced *E. coli* BL21(DE3) cells carrying pTolA box is shown in Figure 6.15. There is high similarity with the spectrum obtained from the induced cells expressing TolA box, as seen in Figure 6.16. Both show no detectable signals attributable to the target protein, only peaks for components of the growth medium.

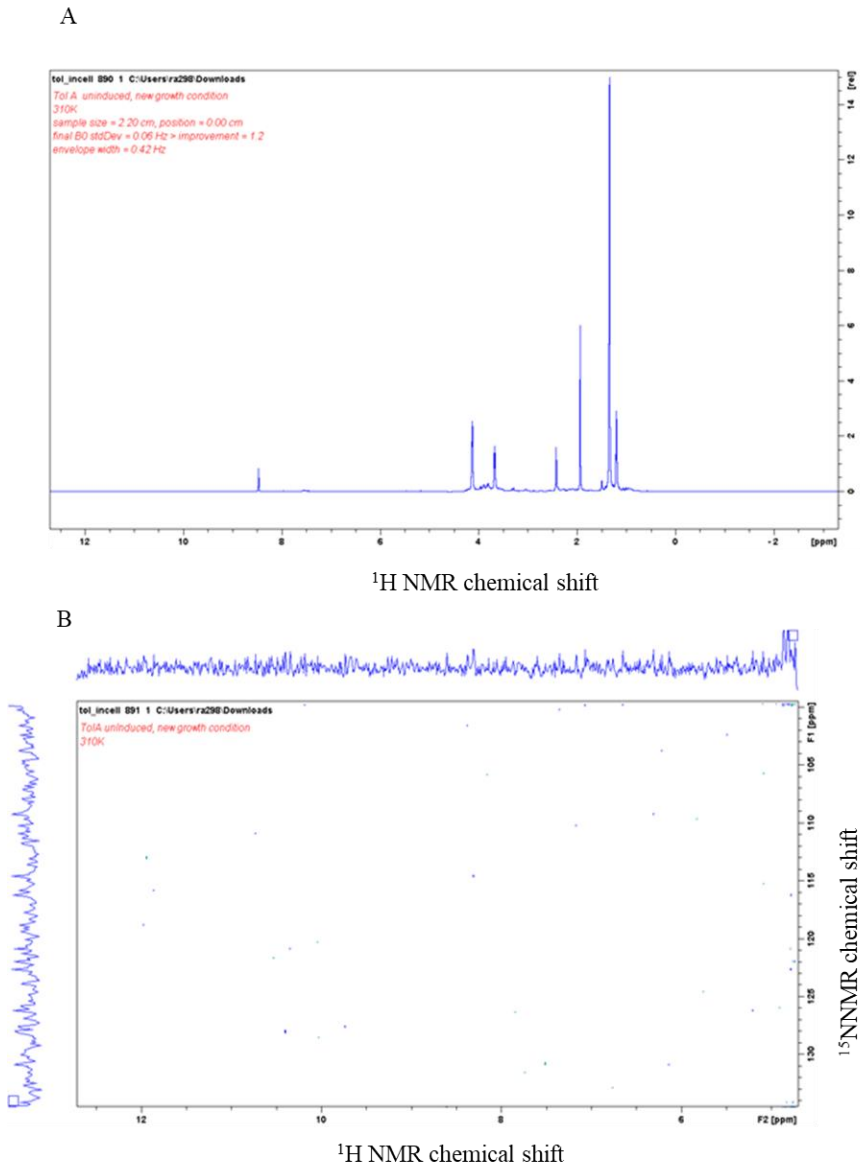


Figure 6.15 In-cell NMR spectrum of *E. coli* BL21(DE3) carrying pTolA (non-induced cells), A) ^{1}H NMR spectrum. Shows peaks at chemical shift δ= 3.5-5.5 ppm due to glucose and thiamine. Sharp peak for H_2O at chemical shift δ= 5 ppm. The strong signals are contaminants. B) ^{1}H–^{15}N NMR spectrum. Shows high background noise and absence of signals, indicating no protein is detected.

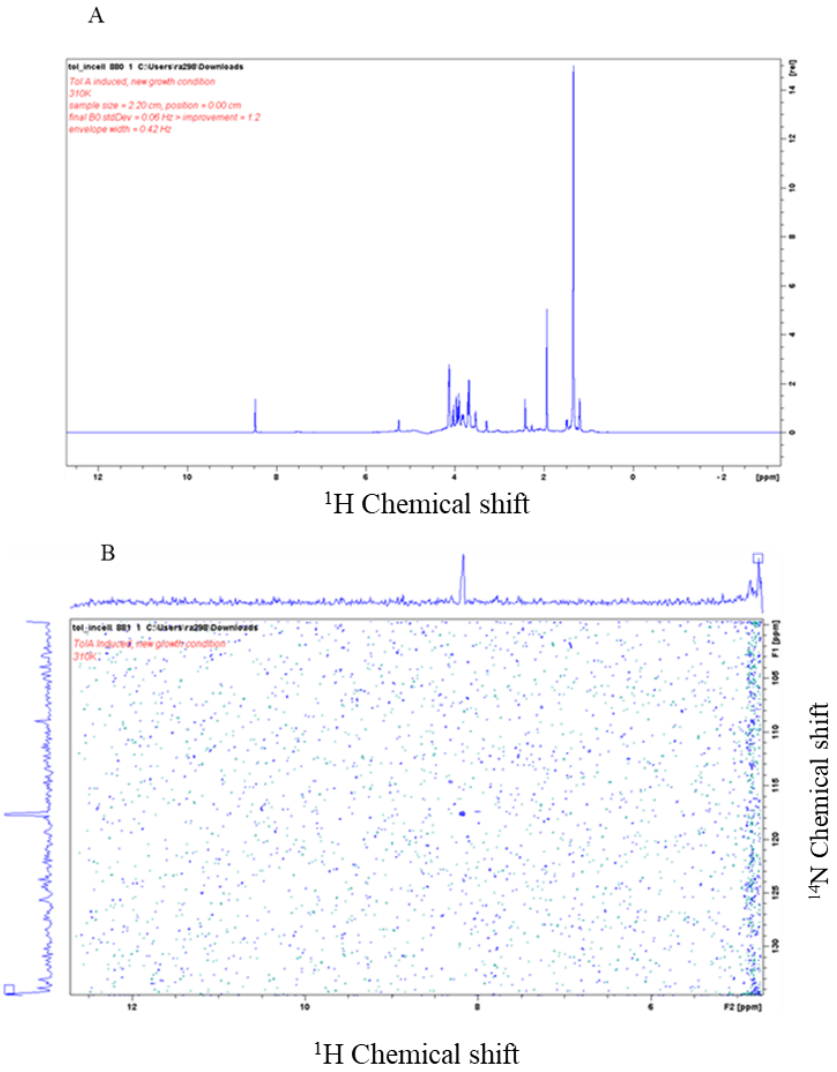


Figure 6.16 In-cell NMR of *E. coli* BL21(DE3) expressing isotopically labelled TolA box. The bacterial culture grown in LB medium then switched to ^{15}N-labelled M9 minimal medium at the time of induction and the 60% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). A) ^{1}H NMR spectrum. Shows peaks at chemical shift δ= 3.5-5.5 ppm due to glucose and thiamine. Sharp peak for H_2O at chemical shift δ= 5 ppm. The strong signals are culture medium contaminants. B) ^{1}H–^{15}N NMR spectrum. Shows high background level with very weak signals that are difficult to assign.

6.4.9 ^{1}H-^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled TolB box (6 kDa)

The experiment to detect the presence of TolB box (6 kDa) was carried out. Figure 6.17 (A) shows no well distributed peaks at the region of amide groups which would indicate the presence of protein. While Figure 6.17 (B) shows weak signals and high noise level which are characteristic for unfolded protein.

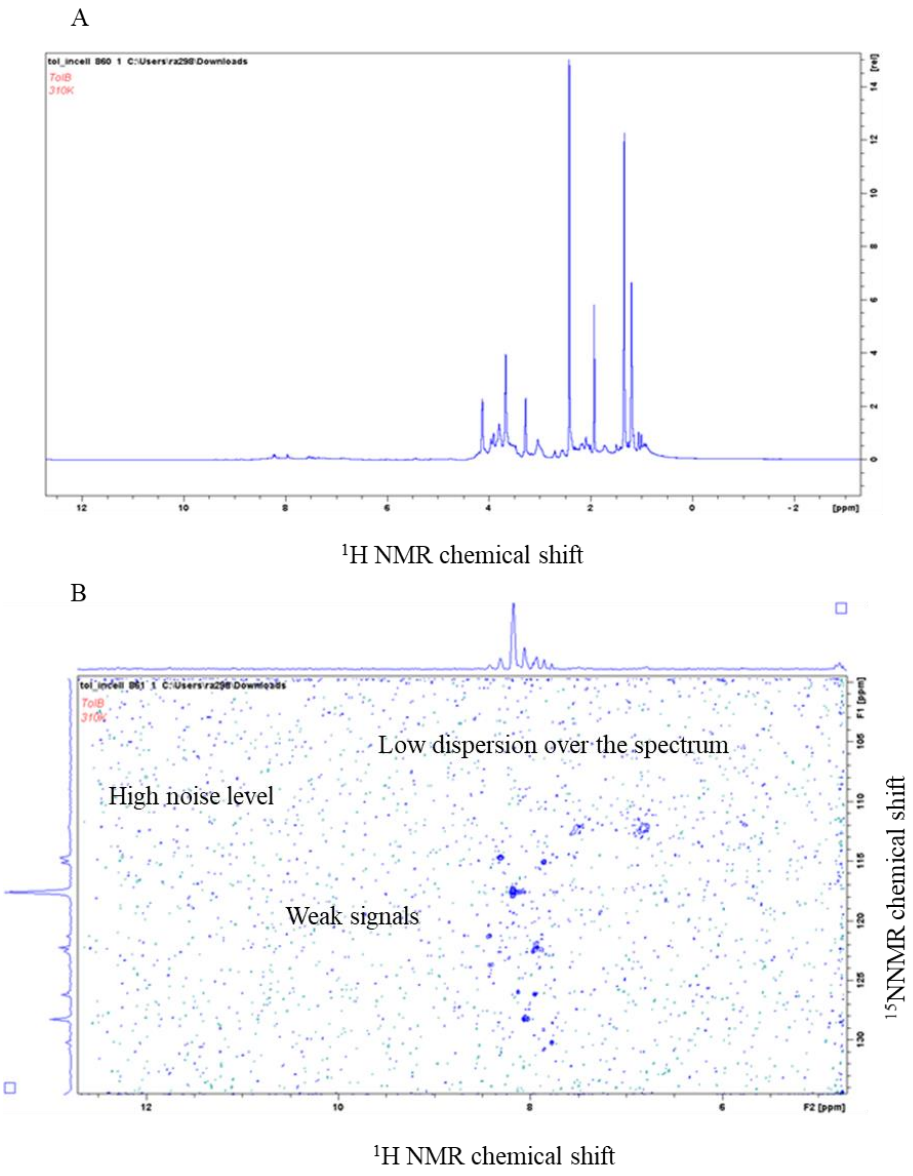


Figure 6.17 In- cell NMR of *E. coli* BL21(DE3) expressing isotopically labelled TolB box. The bacterial culture grown in LB medium then switched to 15N-labelled M9 minimal medium at the time of induction and the 60% (v/v) in-cell NMR slurry prepared in unlabelled M9 minimal medium: D_2O (90:10 ratio). A) ^{1}H NMR spectrum. Shows sharp peaks for side-chain methyl hydrogens at chemical shift δ= 0-2 ppm. Peaks representing for hydrogens in Serine amino acid at chemical shift δ= 3 ppm. Peaks for hydrogen atoms in aliphatic side chains at chemical shift δ= 4 ppm can also be seen. B)^{1}H- ^{15}N correlation NMR spectrum. Shows the appearance of weak signals in the amide region at ^{1}H chemical shift δ= 7-8 ppm. The appearance of few peaks lacking sharpness for Aspartate and Glutamine peaks at ^{15}N chemical shift δ= 117 ppm. Phenylalanine, Tyrosine and Tryptophane residues at ^{15}N chemical shift δ= 120-130 ppm**,** all peaks narrowly dispersed over the spectrum with high background level indicating the presence of short, unfolded protein.

6.4.10 ^{1}H-^{14}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled TA (17 kDa) using an autoinduction NMR medium

Due to un successful attempts to detect periplasmically expressed structured TA, the cytoplasmic expression of TA was investigated to see whether improved detection and spectral resolution of the structured region of this protein could be achieved. As can be seen in Figure 6.18 the observed peak signals appear stronger, more distinct and more narrowly distributed than tested sample for expressed periplasmic TA. In comparison with the spectrum obtained for the periplasmic TA, it shows more cross peaks of better quality, i.e., more distinct and better distributed cross peaks, indicative of higher protein content with more organised structure that indicates less inter- or intra-associated protein content.

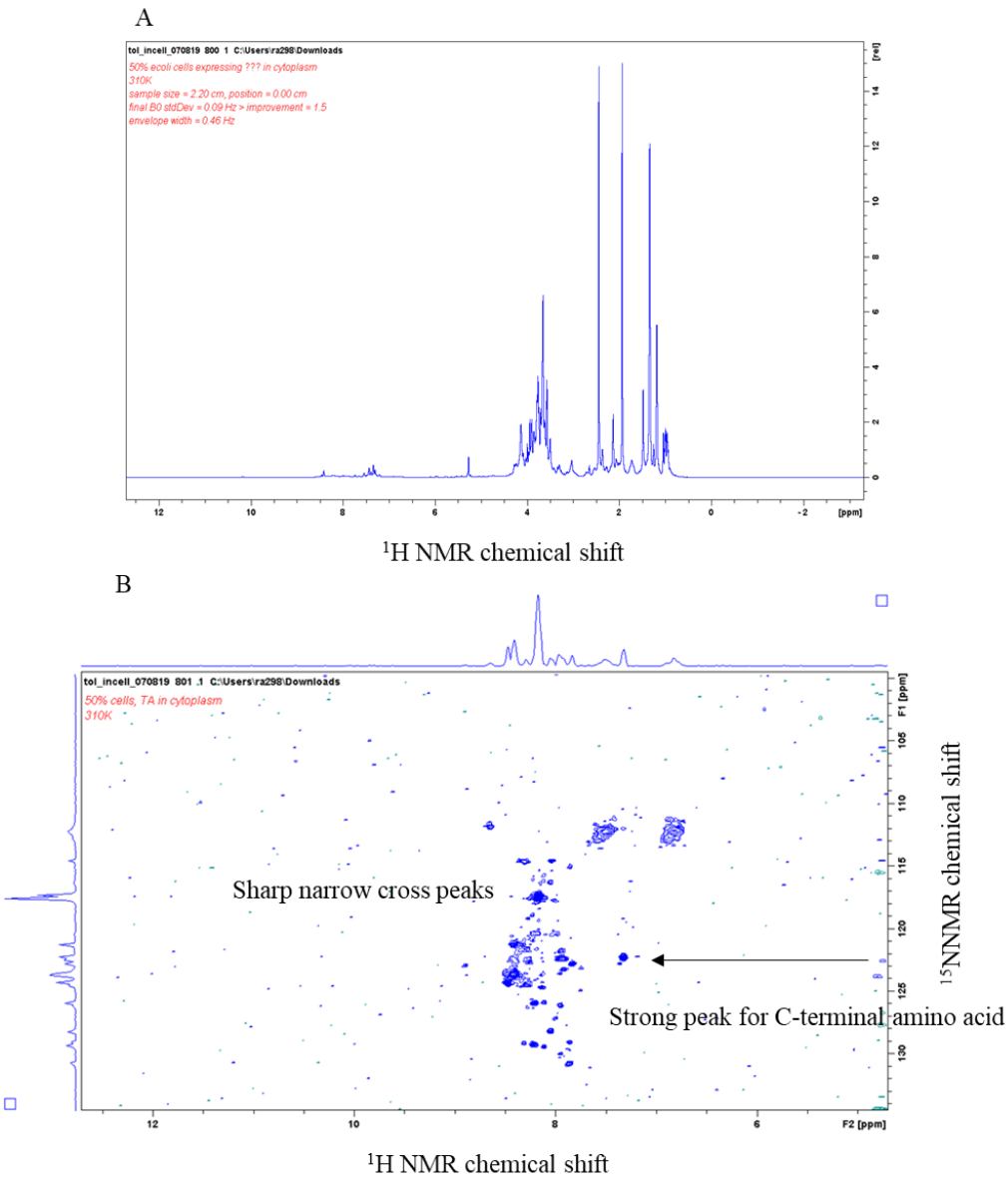


Figure 6.18 In- cell NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled TA using autoinduction medium. A) ^{1}H NMR spectrum. Shows sharp peaks for side-chain methyl hydrogens at chemical shift δ= 0-2 ppm. Peaks for hydrogens in Serine at chemical shift δ= 3 ppm. Shows weak signals in the amide region at chemical shift δ= 7-8 ppm. B) ^{1}H–^{15}N correlation NMR spectrum. Shows strong signals, distinct and sharp cross peaks. Sharp peaks for hydrogens of aromatic side chain at chemical shift δ= 7 ppm. Peaks for hydrogens amide region at ^{1}H chemical shift δ= 8-9 ppm. Cross peaks appeared for Aspartate and Glutamine at ^{15}N chemical shift δ= 117 ppm. Phenylalanine, Tyrosine and Tryptophane residues at ^{15}N chemical shift δ= 120-130 ppm. C-terminal peak at ^{15}N chemical shift δ= 126 ppm. The narrow peaks indicate the presence of the unstructured part of TA with intramolecular interactions.

6.4.11 ^{1}H-^{15}N correlation spectrum of *E. coli* BL21(DE3) expressing isotopically labelled TA using autoinduction medium to perform the induction within the NMR spectrometer

In order to monitor the appearance of NMR signals over time, and monitor the development of transient peaks which may evolve and disappear as protein-protein interactions initiate and conclude, the entire experimental process, from induction of expression, was performed in the NMR spectrometer and monitored over a period of 16 h. However, as can be seen in Figure 6.19, only very weak signals were obtained which were difficult to assign. The target protein was not detected.

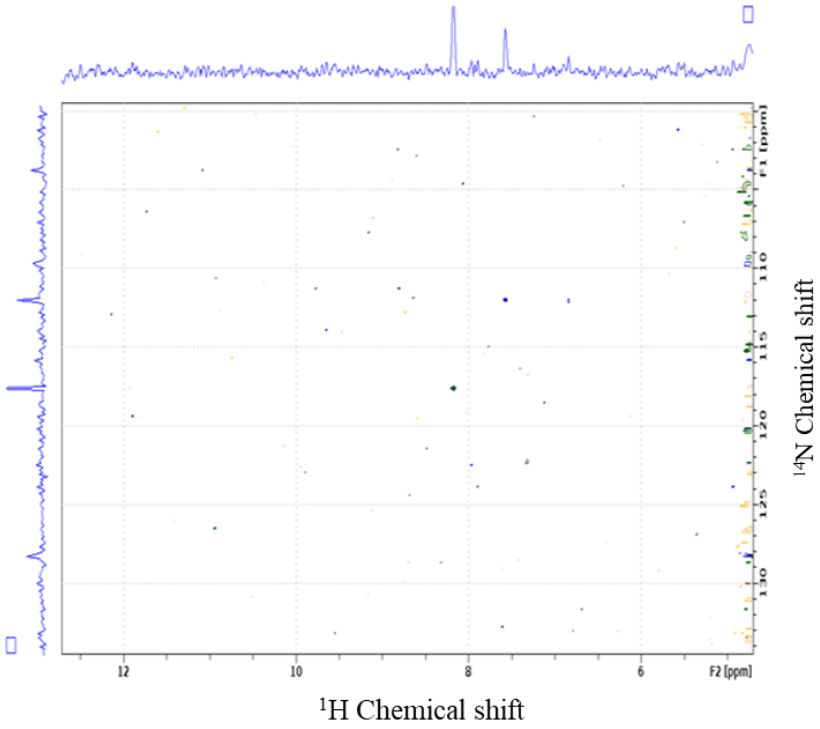


Figure 6.19 ^{1}H–^{15}N correlation NMR spectrum of *E. coli* BL21(DE3) expressing isotopically labelled cytoplasmic TA using auto induction medium within the NMR spectrometer. No proteins peaks were detected.

6.5 Discussion

This chapter discusses whether in-cell NMR is an appropriate technique for investigating the conformational changes of TA, TolA box and TolB box that occur in the periplasm of *E. coli* BL21(DE3) before and after cells are treated with ColE9–Imm9 (i.e., The nuclease colicin E9 co-expressed with an immunity protein). Before making such a determination, however, the work involved identifying the most appropriate conditions for the technique. During this work, the technique was performed on a small (i.e., 6 kDa) model protein, Streptococcal protein GB1. GB1 is a thermodynamically stable, easily overexpressed protein that has been shown to be readily detectable by in-cell NMR and thus provides verification of the ability to detect small proteins using this technique. It therefore provides an ideal positive control. Although the primary focus of the work was the overexpression of ^{13}C–^{15}N-labelled target proteins, as described in Chapter 3, once in-cell NMR commenced, the focus shifted to finding the optimal conditions for the analysis.

The TA domain and the TolA and TolB boxes, the primary constituents of the translocation domain of ColA, were cloned, and the recombinant vectors were transformed into *E. coli* BL21(DE3) cells as described in Chapter 3. Subsequently, overexpression optimisation was performed, as described in Chapter 4, because in-cell NMR require the target protein to be at a higher concentration (i.e., millimolar) than other proteins. The results of CLSM analysis shown in Chapter 5 imply that the expressed TA, and the TolA and TolB boxes in *E. coli* BL21(DE3) interact with the cellular Tol proteins. Beyond that, the interactions provided the host cell with protection against the external colicin (i.e., ColE9–Imm9), as quantified by CLSM. The results obtained from CLSM also suggest that not only are TA and TolB involved in the interaction but also that TolA plays a major role in the interaction with ColE9, which could be further investigated by in-cell NMR. Such protection may be due to the interaction between the overexpressed periplasmic proteins and cellular Tol proteins, which might have

prompted the formation of large complexes that inhibited ColE9 from binding to TolB. If present, it should be possible to visualise the conformation of that large complex by in-cell NMR, which was one of the aims of this work. Added to that, the detection of a direct interaction between TolA and ColE9 would support the possibility that the interaction of the overexpressed domain and TolA also affected the energy transfer, which is responsible for the immunity protein's release from ColE9–Imm9 and the translocation of ColE9. That possibility also takes support from the findings presented in Chapter 5. A better understanding of the nature of those interactions and their mechanisms might answer several questions about the translocation of colicins, including what kind of conformational changes occur during their translocation and how TolA is involved in the process.

Accordingly, in-cell NMR was employed to analyse protein-protein interactions (PPIs) involving TA and the TolA and TolB boxes expressed in *E. coli* BL21(DE3) cells. However, producing efficiently labelled proteins suitable for analysis necessitates technical optimisation, especially in the case of double labelling with both carbon (^{13}C) and (^{15}N) nitrogen atoms. Because that initial approach proved difficult, a shift was made to (^{15}N) nitrogen labelling only.

The project's overarching goal was to develop a straightforward, effective technique for detecting the expressed protein by in-cell NMR. Various experimental methods were explored to determine which parameters were important in performing an effective in-cell NMR analysis on the proteins of interest in the *E. coli* system. The four parameters determined to be the most important to examine were the culture medium used to grow the bacterial cells, the expression system the proteins of interest were cloned into, the volume of the starting culture and the concentration of the sample slurry, and the size of the expressed protein (Sharaf *et al.*, 2010).

The TA domain, TolA and TolB boxes are mainly intrinsically disordered proteins hence their analysis should produce some broad peaks and strong signals with ^{1}H in-cell NMR because of the intrinsic flexibility of intrinsically disordered proteins. As the results achieved in Chapter 5 of this project suggest the presence of interactions of the expressed proteins (TA, TolA box and TolB box) with cellular proteins, the expected results from in-cell NMR analysis would be that these proteins would produce wider overlapping peaks indicating their involvement in protein-protein interactions (PPIs) inside *E. coli* cells which might causes loss of signals.

Experiments were initially conducted at the NMR facility at King Saud University in Riyadh, Saudi Arabia, beginning with 1 dimension (1D) ^{1}H and ^{13}C experiments. The backbone sequential assignment of the chemical shift of each spin preceding the amino acid residue was performed using a cryogenic probe, which in addition to the use of a static magnetic field, has been shown to improve the detection of ^{13}C and ^{1}H nuclei, resulting in an overall increase in the sensitivity and resolution of NMR (Emwas *et al*., 2019). In another important step, the 20% (v/v) in-cell NMR slurry was dissolved in 10% D_2O instead of H_2O to prevent the amide proton exchange that occurs when H_2O is added as an NMR solvent, which causes the formation of exceptionally sharp peaks that may overlap the resulting peaks in the NMR spectrum (Zhang *et al*, 1995) obscuring detail. By contrast, adding D_2O generally results in ^{1}H exchange with deuterium, which is undetectable by NMR. In general, the 1D ^{1}H NMR spectrum for proteins (> 5 kDa) is difficult for researchers to assign the formed peaks because of overlapping peaks, and the ^{1}H chemical shift of the protonation state in the spectrum is usually small (i.e., <1 ppm). Lastly, ^{1}H chemical shifts are more sensitive to charges other than the charges on the amino acid to which the proton is attached (Karplus *et al*., 1973). Therefore, using a 1D ^{13}C spectrum followed by a 2D ^{1}H–^{13}C correlation spectrum was

considered to be a key step in analysing the expressed proteins (i.e., TA, TolA box and TolB box).

The first attempt was performed with the sample for analysis prepared in the overexpression medium. As shown in Figure 6.3, both samples yielded highly similar results, with slightly broader signals derived from bacterial cells in which TA expression had been induced in comparison to the sharper peaks seen for the medium alone that represented only components of the medium. That outcome suggests that no signals were provided by the expressed TA. If the expressed TA protein had been detected, then peaks should have been observable between 8 and 9 ppm on the ^{1}H spectrum, a region of chemical shift that represents the presence of an amide group, indicative of labelled protein which did not appear in the experiment. The same sample was tested using ^{13}C NMR in an attempt to detect TA-derived signals. These results similarly showed no protein-derived signals between 10–70 ppm or 170–190 ppm (Figure 6.4). The absence of a protein chemical shift in both experiments (i.e., with ^{1}H and ^{13}C) could be due to several factors, including sub-optimal sample preparation and possibly an inefficient labelling technique. Considering that the sample in Section 6.4.1 was prepared in the spent medium (i.e., the overexpression medium), the essential nutrients in the medium had already been used by the cells. That circumstance suggests that although the cells are still a live, the growth might have been affected, which is indeed the primary requirement of in-cell NMR—that is, to have signals developed from living cells that have been tested to confirm they retain viability. Another potential source of the negative results was the presence of unincorporated labelled isotopes in the spent medium, which could have caused the sharp cross peaks' signals observed. Because of their labelled status, NMR detection was more sensitive; however, the formation of those peaks could have masked the formation of protein signals, if any were present.

In view of those results, further changes to the protocol were made and the final in-cell NMR sample was instead prepared in fresh unlabelled M9 minimal medium instead of spent medium (see Section 6.4.2). The medium was replaced with fresh, following the suggestion of Luchinat and colleagues (2015) with the expectation that fresh medium would reduce stress and promote growth by providing additional nutrients for the cells and additionally removing toxic metabolic by-products that may have accumulated, all of which may improve the cell's growth and NMR detection. Research has revealed that the short lifetime of cells during experiments such as the ones are conducted in this project, can be mitigated not only by preparing the final in-cell NMR sample in fresh medium, but also by using a bioreactor. Using a bioreactor, a constant flow of growth medium can be applied to the cells in order to provide a good supply of oxygen and nutrients. The bioreactor improves cell viability for long periods, thereby permitting relatively long NMR experiments, and allows investigating cellular processes in real time (Barbieri & Luchinat, 2021).

However, as shown in Figure 6.5, the results obtained with samples prepared in fresh medium were the same as those obtained with spent medium in Figure 6.4: no TA protein was detected. That finding led to the conclusion that the problem might relate to the isotope-labelling technique itself, rather than the sample preparation. Although ^{13}C NMR offers many advantages for metabolomics research—among others, a wide spectral dispersion, narrow singlets and a direct measure of the backbone structures of metabolites—using the ^{13}C isotopes tends to cause a relatively high background signal due to incorporation into many metabolites and other metabolic products of biochemical processes (Burz *et al.*, 2009). Moreover, the dispersion of ^{13}C resonances is generally poor compared with that of ^{15}N owing to the sensitivity of ^{13}C nuclei to the nature of the neighbouring amino acid in the primary sequence, especially if the protein of interest is unfolded. By contrast, the dispersion of ^{13}C nuclei is greater in folded proteins, given the well-known environments of those nuclei in

relation to secondary and tertiary structures (Yao *et al.*, 1997). Beyond that, because $^{1}H–^{13}C$ correlation experiments, one of the most common ways of analysing small proteins (i.e., especially <5 kDa) by NMR, for larger proteins such as TA (17 kDa) introducing a third dimension for the ^{15}N chemical shift should aid structural determination, according to Ferella and colleagues (2012). Above all, the primary advantage of using ^{15}N isotope labelling is that every ^{15}N position, whether in the protein backbone or side-chains, is separated from another ^{15}N atom by at least two bonds. Therefore, no $^{15}N–N$ couplings are present that could promote complicated, difficult-to-detect behaviour.

Thereafter, ^{15}N NMR experiments were performed at the Manchester Institute of Biotechnology (MIB) in the United Kingdom. The experiment described in Section 6.4.3 was conducted, albeit using a slightly different isotope-labelling medium, ^{15}N isotope labelling only, which was expected to simplify the experiment's preparation and reduce the background of the spectra such that the labelling isotope was ^{15}N only. The experiments conducted with ^{15}N-labelled M9 minimal medium were expected to produce clear spectra and easily assignable signals. The detection of nitrogen atoms ranks among the most important sources of information about protein backbones, because using nitrogen labelling (^{15}N) helps to reduce the protein's background noise and thus allow the detection of amino acids in the NMR spectrum, especially in 2D experiments with $^{1}H–^{15}N$ correlations. $^{1}H–^{15}N$ correlations, which represent coupling in the peptide bond, were the starting points in the analysis. Such bonds appear in every amino acid residue except that of proline. Normally, the spectrum reveals a ^{15}N-axis (i.e., x-axis) chemical shift and another axis for ^{1}H (i.e., y-axis), which partly explains the usefulness of the 2D spectrum (Poulsen, 2002). The technique has also proven applicable with *E. coli* cells, *Xenopus* oocytes and HeLa host cells, whose proteins have been successfully assigned and their 3D structures determined this way (Burz &

Shekhtman, 2012). Thus, at least in theory, shifting to uniformly ^{15}N labelling should increase the odds of detecting the TA protein with in-cell NMR.

Accordingly, $^{1}H–^{15}N$ correlation was performed with a small-scale (10 mL) culture of *E. coli* BL21(DE3) cells periplasmically expressing TA in labelled M9 minimal media to produce 20% (v/v) slurry as described in Section 6.4.3. In general, when the protein of interest (i.e., TA) is overexpressed within cells in labelled media, the isotopic labels are incorporated into all proteins and metabolites, and those labelled proteins should be easily detected. Although the results of 1D ^{1}H shown in Figure 6.6 (A) were sharp and included higher resolution peaks at 0–4 ppm, they indicate only the contents of the medium. By contrast, the 2D $^{1}H–^{15}N$ showed very weak signals that may represent amide hydrogens in the peptide backbone, as shown in Figure 6.6 (B), which may indicate the presence of protein. The appearance of a few weak peaks narrowly dispersed over the spectrum signifies the presence of unfolded protein or a short peptide chain that may relate to the TA protein. Meanwhile, spectra from non-induced *E. coli* BL21(DE3) carrying pTA (Figure 6.7) used as a negative control showed no peaks for proteins, as expected, which indicates that the short peptide chain detected in Figure 6.6 (B) is an in-cell protein with a low level of expression.

As a tentative conclusion, the weak formation of those peaks indicates the presence of a short-unfolded protein. That dynamic is expected because individual proteins may behave differently in different expression systems and because it is normally difficult to achieve the best expression conditions for a particular protein. Another reason, one investigated in Chapter 4 but with double labelling (i.e., ^{1}H and ^{13}C labelling), is because the starter culture medium was M9 minimal. Growing bacterial cells in M9 minimal medium results in a longer lag phase than growing cells in rich media, which could have affected the protein expression and the protein yield required for in-cell NMR. Researchers have managed to overcome that problem to some extent by increasing the buffering capacity of the medium which can

accelerate the growth rate, which ultimately leads to higher yields of expressed proteins (O'Brien *et al.*, 2018). By extension, the effectiveness of this approach was recently confirmed by Azatian and colleagues (2019), who suggested that doubling the concentration of buffering salts in minimal media facilitates protein production for an increased signal-to-noise ratio and decreases the duration of experiments. Although future studies may involve using that approach, it was not used here to overcome the low protein yield. Instead, the effect of cell growth conditions was explored by changing the initial culture medium to a rich medium, LB which would result in shorter lag phase.

The goal was simple: to increase the bacterial cell density such that the target protein was expressed at a sufficiently high level to enable detection using in-cell NMR. The method employed allowed for a significantly enhanced initial cell density at OD_{600} values of 1 before induction. Following the procedure of optimised expression discussed in Chapter 4, the experiment was conducted first in LB-rich medium in place of M9 minimal medium after which a switch was made to labelled M9 medium at the time of induction. Growing the bacterial culture in the LB-rich medium, as illustrated in Figure 6.8, revealed the appearance of a few cross-peak signals representative of amide hydrogens in peptide residues although these were of low intensity, which mirrored the results presented in Section 6.4.3 where the bacterial culture was grown in M9 minimal medium except for the high background level indicating the presence of unfolded protein. Obtaining the same results even with optimised conditions for protein expression implies that the unfolded protein could be part of TA, especially if it is accepted that the general structure of the (T) domain of all types of colicins is relatively the same. Kim and colleagues (2014), who investigated the structural conformation of two type A colicins, Colicin N (i.e., a pore-forming colicin) and Colicin E3 (i.e., nuclease), discovered that the (T) domains of both colicins are divided into two parts: an intrinsically unstructured T domain (IUTD), which could be what had been observed in

this work thus far, and a larger, structured (T) domain (STD) distal to the IUTD that we are aiming to detect by in- cell NMR.

A positive control, a small 6 KDa protein, GB1, known to be stable, was tested next. GB1 had been cloned into pET-21a at the MIB and transformed into *E. coli* BL21(DE3) for expression. Using the same conditions used to express TA in *E. coli* BL21(DE3), given the low sensitivity of ^{1}H NMR, only the 2D $^{1}H–^{15}N$ correlation was applied for NMR analysis of the positive control. The resulting NMR spectrum for GB1 shown in Figure 6.9 was not expected, however, for the peaks obtained implied a more complex folded protein than GB1. The weakness of signals allowing only poor, even invalid data with 20% slurry, could have been due to the low amount of protein in the slurry or to the protein's leakage into the medium. The same findings were reported by Ikeya and colleagues (2016), who performed in-cell NMR in eukaryotic cells to investigate both protein GB1 and the putative heavy-metal binding protein TTHA1718 from *Thermus thermophilus*. They found that a low concentration of GB1 in *E. coli* cells caused a much-reduced contrast between the NMR signals of GB1 and the background, whereas their successful in-cell NMR of eukaryotic cells revealed that TTHA1718 had been induced in a stable isotope-labelled medium (i.e., 100 ml) to high level. As a final development of sample preparation for in- cell NMR, a 60% v/v cell slurry was prepared for analysis. In previous attempts with in-cell NMR, the bacterial slurry used was relatively dilute (approximately 20% cells (v/v)). It is possible that such a low cell density caused the poor resolution obtained as well as the low intensity of the signals. Although researchers who have performed in-cell NMR have recommended that the NMR sample should contain at least 20% v/v cells in the slurry (Waudby *et al.*, 2013), others have used denser samples to prevent cell settling which, in turn, limits the ability of NMR to detect signals (Burz & Shekhtman, 2012). Currently, the majority of in-cell NMR studies involve using dense cell slurries (60% v/v) in standard NMR sample tubes (Siegal & Selenko, 2019).

Accordingly, protein GB1 was further investigated using the same experimental conditions described in Section 6.4.6, but with a larger starter volume of bacterial culture (100 ml) to enable production of a slurry of approximately 60% cells (v/v). Although the non-homogeneous and viscous sample might cause line broadening in the NMR spectrum that could affect the magnetic field (Franks *et al*., 2005), the signal broadening and the low intensity resulting from the reduced rate of tumbling (i.e., rotation of the molecule) resulting from the high viscosity of the intracellular medium and interactions with the macromolecular components of the cytosol (Pielak *et al*., 2009). Ultimately, using this approach, analysis of the GB1 positive control generated a spectrum closely resembling that of a published GB1 spectrum, as illustrated in Figure 6.10. The clear resolution of the spectrum and its close similarity to published spectral data clearly indicate that this approach, using a higher cell density of 60% v/v, is required for successful in-cell NMR.

A follow-up experiment was conducted to gauge the extent of the protein's leakage into the medium, which can occur due to cellular crowding and be detected by in-cell NMR (Pielak *et al*., 2009). As explained in Chapter 2, the follow-up experiment involved harvesting the NMR sample and testing only the supernatant to determine whether protein leakage had occurred, and to confirm that the protein was indeed inside the cells. As shown in Figure 6.11(A), protein GB1 was inside the cells; however, as shown in Figure 6.10(B), a large amount of extracellular protein GB1 also appeared, which indicates the partial unfolding or slight aggregation of the extracellular protein GB1 or release from damaged cells during centrifugation and sampling. Of course, that trend is quite common with protein GB1, which has been reported to leak easily into media during NMR experiments (Burz & Shekhtman, 2012).

Following the success of the above approach, *E. coli* BL21(DE3) periplasmically expressing TA was tested at a higher cell density (i.e., 60% slurry v/v). Although high cell densities pose disadvantages that can include plasmid loss from *E. coli* or reduced pH due to higher concentrations of cell metabolites, which could result in low protein production (Sivashanmugam *et al*., 2009), the result achieved by protein GB1 was encouraging. As depicted in Figure 6.12, the new resonance that appeared indicated the presence of some ^{15}N-labelled protein, with more cross peaks and clusters that could indicate protein aggregation or conformational flexibility. Again, the characteristics of the cross-peaks and the appearance of amide residues all indicated the presence of an unfolded protein, which could be the unstructured part of TA.

Resonance assignment is key to enabling the characterisation of a protein's conformational state. Although the linewidth of the peaks suggested the presence of a large molecule, the complexity of the peaks within the spectrum limited the assignment of protein's chemical shifts. The indication of a large molecule supports our hypothesis that the interaction of TA and cellular Tol causes the formation of a large complex. The fact that nitrogen forms only a third of the protein backbone and is only present in the side-chains in six of the twenty amino acids, means that ^{15}N NMR alone cannot provide full information about the conformational structure of a protein except for a few positions (e.g., Asn, Gln, His, Trp, Lys and Arg), as shown in Figure 6.12. Moreover, because the results indicate the presence of an unfolded protein, hydrophobic groups could have been exposed that bound to components in the cellular medium and caused the loss of signal, precisely because amide relaxation may be insensitive to motions in the hydrophobic core of the protein. On that topic, Watt and Loria (2010) have explained that certain motional modes of the protein backbone cannot be detected by monitoring relaxation at the amide positions.

The absence of signals in Figure 6.13 -shows that no sign of protein leakage manifested in the cellular medium. However, the spectrum shown in Figure 6.14, derived from the cells removed from the growth medium analysed in Figure 6.13 was consistent with a badly behaving protein, which might have been the unstructured part of TA. The numerous low-intensity peaks with broad linewidths seen may indicate the formation of large protein complexes due to protein-protein interactions, engagement with the cellular membrane and/or aggregation, as previously suggested and is typical of a 'badly behaved protein'.

Thus far, the findings indicated the success of the technique using a 60% slurry and the unstructured part of TA might have been detected. Reasons for the difficulty in detecting the structured part of TA could be that proteins differ in their folding pathways and that some proteins need pro-sequence or chaperones to fold correctly. That difference in the duration of the folding process should be considered when applying NMR with quench-flow hydrogen exchange method, to detect hydrogen-bonded amide groups (Wright & Dyson, 2004), These changes were made in the NMR settings, however, the TA structured part was still not detected.

Consideration of the most important parameters that would provide the best conditions for in-cell NMR led to considering the size of the target protein as well. According to the literature, although the largest protein successfully analysed with in-cell NMR is calmodulin at a size of 16.8 KDa, which was specifically labelled, the best results have typically been obtained with proteins less than 12 KDa (Burz & Shekhtman, 2012). In-cell NMR is far easier with smaller proteins because their tumbling rate (i.e., movement or rotation) is greater than that of larger proteins (Pielak *et al.*, 2009). Accordingly, following the results achieved with TA, the technique was applied to far smaller proteins, namely the TolA box domain (12 kDa) and TolB box domain (7 kDa), both of which are parts of the translocation domain. The results reported in Chapter 5 suggest that the overexpression of those proteins prompted an

interaction with cellular Tol proteins of *E. coli* BL21(DE3). Figure 6.15 (A) and 6.15 (B) shows the spectrum for the non-induced *E. coli* BL21(DE3) carrying pTolA box (i.e., negative control); as expected, no signals were detected. The reason behind testing the non- induced cells was the negative result obtained from the second attempt to analyse expressed TolA. As shown in Figure 6.16(A) and 6.16(B) show how the analysis of periplasmically expressed TolA box protein generated spectra with noise but no definite protein-related resonances, which aligns with our other findings. The absence of protein resonance may have been due to its major involvement in the interactions between TolA and cellular Tol (discussed in Chapter 5), which might have caused the loss of signals and indicate either a need for better resolution (i.e., higher dimension analysis) or more sequential experiments (repeated scans). Despite the higher sensitivity analysis achieved with the TolB box, high noise level was detected with low dispersion peaks (Figure 6.17), there was an indication of the presence of an unfolded protein, and the weak signals obtained could have also indicated the presence of PPIs, as suggested in Chapter 5. Such findings imply that the relatively large size of the target protein, TA, was not the cause of the complex spectra obtained, but potentially a characteristic of the protein itself.

The final experimental parameter investigated was the expression system used. Although an aim of the project was to study protein–protein interactions during the entry of colicins into the periplasm, using protein GB1 as a positive control for the technique, one that is expressed in the cytoplasm, influenced the method of testing the system. Researchers have shown that the bacterial cytoplasm is a good model for the analysis of eukaryotic proteins specially to study the influence of molecular crowding on protein folding and non-specific PPIs (Banci *et al*., 2011). Therefore, instead of testing the protein TA in the periplasm, where the levels achieved are typically limited, it was tested in the cytoplasm using a commercially available auto-induction NMR medium. Although using the rich LB medium has yielded tightly

controlled and high level of recombinant proteins expression, it is also time-consuming (Burz & Shekhtman, 2012) compared with using autoinduction, which was expected to save time and effort by involving less intervention and, in turn, imposing less stress on the cells as they work to maintain protein stability. The entire experiment depended on components of the medium that are metabolised differently to produce high-density growth and the automatic induction of protein expression from *lac*-based promoters (Grabiski *at al.,* 2005). The auto-induction medium contained glucose, lactose and other carbon sources. In general, when lactose is the carbon source, it is imported into the cell, supports the growth of the host cells and induces the expression of isotope-labelled recombinant proteins. The auto-induction medium does not require the monitoring of optical densities, or centrifugation steps, meaning less handling and potentially less human error (Tyler *et al.*, 2005).

Therefore, as described in Section 6.4.10, *E. coli* BL21(DE3) cytoplasmically expressing TA was tested. Figure 6.18 shows some of the characteristics of the spectrum of the periplasmic TA that was obtained. More peaks were detected with cytoplasmic expression, and they were sharper and more distinct than the peaks obtained from periplasmic expression, with a resolution representing the amino acid side-chains of Aln, Gln, Try and Phe. Despite the noise and poor quality of the spectrum, possibly because TA was cytoplasmically expressed, the resolution for the unstructured part of TA was far better than the that obtained with periplasmic expression. However, still no signals for the structured part of TA were detected. In general, the production of overly high concentrations of isotopic-labelled protein in the cytoplasm can have disadvantages. A study by Burz *et al*. (2012) suggested that very high concentrations of proteins can cause protein leakage into the medium or non-specific interactions that may complicate the detection of proteins with NMR, which could explain our results. Added to that, Fawzi and colleagues (2010) have reported weak protein–protein associations found by using NMR, including interactions between identical proteins (i.e., self-

association). They proposed a two-step process to explain this starting with the formation of an ensemble of transient, non-specific complexes dominated by electrostatic forces, followed by a rearrangement along the protein surface to form a final, well-defined complex stabilised by short-range hydrophobic interactions, hydrogen bonds and electrostatic forces. Johansson and colleagues (2014) later confirmed the non-specific interaction of proteins and suggested that the crowding of the cytoplasm as well as viscosity exert major effects on the development of such interactions.

To further investigate different condition, an overnight NMR experiment has been employed, that entailed monitoring the sample during in-cell NMR analysis by recording when the peaks formed and what changes happened to developed peaks in the spectrum. Majumder and colleagues (2014) previously reported being able to monitor changes over time in in-cell spectra. The results of their analysis showed the distribution of singular, distinct peaks followed by gradual changes in the spectra corresponding to specific interactions or the random binding of the target protein to components in the cytosol. However, this kind of monitoring is likely to affect the viability of cells due to the extended time that they must spend in the NMR tubes. But for the hope to witness the formation of protein peaks before it got lost. Therefore, attempt to investigate the developed signals of the spectrum, the same TA was induced in the NMR tube, left overnight and tested in the NMR spectrometer. However, the spectrum did not show any amide peaks, as shown in Figure 6.19, meaning that it was more likely the expression of TA was affected and caused a loss of signal.

The optimisation of conditions and environment for the successful in-cell NMR of proteins has proven difficult, especially as part of the TA target protein is intrinsically disordered. An essential requirement for using in-cell NMR is the high concentration of the protein needed; In this project, the crowding of the cytoplasm was not problematic, because the cells initially expressed the protein in the periplasm. Viscosity was not problematic, either, in the case of

the expressed protein GB1 but could have been for the expressed TA. Nevertheless, a more homogeneous, dilute sample was tested first, and the results were the same as those achieved with higher concentration samples: only signals of unfolded protein that might relate to the unstructured part of TA were detected.

The spectra yielded from in-cell NMR experiments in the project were predominantly characterised by broad signals. Barberiri *et al*. (2015) have explained that the broadening of signals in a protein NMR spectrum is typically due to weak, non-specific interactions of the protein with larger complexes or with cellular structures such as membranes, either of which can lower the tumbling rate, prevent NMR detection and cause signal loss. As suggested in Chapter 5, challenging the cells expressing the proteins of interest (i.e., TA, TolA box and TolB box) with ColE9–Imm9 afforded protection against the nuclease effect of ColE9, presumably due to the suggested interaction between the expressed proteins and cellular Tols in the periplasm. That interaction might have caused the results reported in this chapter, especially regarding the TolA box.

Another possible reason for the low-quality spectra obtained is the intrinsic behaviour of the proteins of interest themselves. Sample preparation that entails centrifugation and resuspension in different media is a stressful process that can result in the varied availability of oxygen, which might also affect the stability of proteins (Lippens *et al*., 2018). Although that outcome is entirely related to protein stability, it could have been the case with the colicins.

Chapter 7

General discussion

7.1 General discussion

Colicins are a type of toxin that are produced by certain species of bacteria, and can be cytotoxic to neighbouring bacterial cells which are competing for the same resources. Colicins are produced in response to stress or DNA damage. Both of these processes can activate the SOS response for DNA repair, and, in turn, upregulate the production of colicins within bacterial cells. Once colicins are released in the environment external to the producing cell (Zgur-Bertok, 2012), they are liable to exert a cytotoxic effect on neighbouring bacterial cells that do not produce colicins. However, it is important to note that they are not effective against bacterial cells which are producing colicins, because these cells are protected from the cytotoxic effect due to the co-expression of an immunity protein (Imm) alongside the colicin (Bouveret *et al.*, 2002).

To exert their cytotoxic effect within cells, colicins first have to be translocated into the cells via their translocation domain, or (T) domain, which is one of three functional domains within the toxin. The other two being a receptor-binding domain, or (R) domain, and a cytotoxic domain, or (C) domain (Cascales *et al.*, 2007). Depending on their mechanism of translocation, colicins can be divided into two groups. In Group A, colicins are Tol-dependent, meaning that they interact with Tol proteins in the periplasm. These group A toxins can be further classified into the pore-forming group (e.g., colicin A, or ColA), or the nuclease group (e.g., colicin E9, or ColE9). In Group B, by contrast, colicins are Ton-dependent, meaning that they interact with Ton proteins (Bonsor *et al.*, 2008).

To date, researchers have observed a series of events that occur during the translocation of different colicins, which is believed to influence their cytotoxic effect (Atanaskovic & Kleanthous, 2019). Research indicates that, generally, the (R) domain of colicins binds to an outer membrane (OM) receptor (i.e., BtuB) for *E. coli* cells, which consequently allows the

(T) domain to bind to an OM porin, OmpF. As a result, a translocon forms that can translocate the (T) domain through the OM. Once the (T) domain passes through the OM towards the periplasmic space, it encounters the Tol–Pal system, which comprises proteins distributed between the inner membrane (IM) and OM (Lloubes *et al*., 2001): Pal in the OM, TolB between the OM and periplasmic space and various Tols (i.e., TolA, TolQ and TolR) in the IM.

Next, after the (T) domain crosses the OM, the subsequent method of translocation via the periplasm to the IM depends upon the type of colicin involved (Houdsen *et al.*, 2012). Pore-forming colicins (e.g., ColA) interact with both TolA and TolB and their binding sites, namely TolA box and TolB box, respectively. Such interactions, when occurring simultaneously, form a TolA–TolB complex in the periplasm that interacts with ColA. Next, the (C) domain passes through to the IM, where it forms pores that will disrupt the IM, and ultimately kill the cells (Benedetti *et al.*, 1992; Cramer *et al.*, 2018). Colicins in the nuclease group (e.g., ColE9), however, translocate via a different mechanism. ColE9, for instance, is co-expressed with its immunity protein, Imm9, with which it forms a complex. Once the (R) and (T) domains bind to OM receptors, the (T) domain interacts with TolB and its binding site (i.e., TolB box). At that point, the Imm9 protein is released from the complex, thereby allowing the (C) domain to enter the periplasmic space, then to reach the IM and lead to DNA damage (Cascales, 2007).

Researchers who have studied Imm9 release prior to translocation have observed that TolA plays an important role in ColE9 translocation, even in the absence of direct interaction between ColE9 and TolA (Carr *et al.*, 2000). At the same time, when the researchers mutated TolA, the release of Imm9 became compromised, which suggested that an indirect interaction could indeed occur between ColE9 and TolA. Subsequent research revealed that the energy required for Imm9 release is generated by TolA activity in the IM and by other associated

Tols (Houdsen *et al.*, 2018). However, without that force, Imm9 cannot be released, and ColE9 cannot perform its cytotoxic activity.

Examining the properties of colicins, as powerful antimicrobial agents that are easy to generate, could pave the way for the development of novel antibiotics in response to multidrug resistance in populations worldwide (Behrens *et al*., 2017). Of particular necessity is research using new techniques that can fully illuminate the mechanism of colicin translocation. Given that combined techniques typically provide more detailed information, especially regarding proteins, the aim of this project was to investigate new techniques for studying protein–protein interactions involved in colicin's entry into *E. coli* cells. In this project, the colicins of interest were ColA and ColE9, two Tol-dependent colicins with different mechanisms of action. Col A, pore forming type colicin and ColE9 a nuclease type colicin.

Before studying protein–protein interactions relating to colicin activity in the periplasmic space of cells overexpressing the (T) domain or specific regions of that domain, as well as ColE9 effects on those cells, it was hypothesised that colicin-sensitive strains could become resistant to exogenous colicins if they overproduce the (T) domain or sub-regions of another colicin. To test that hypothesis, the project was designed to introduce the (T) domain of ColA for subsequent protein expression, and the (C) domain of ColE9 as an external antimicrobial colicin, and the colicin-sensitive *E. coli* strain BL21(DE3) as a host strain for expression. In those terms, the focal question was: what is the effect of ColE9-Imm9 on *E. coli* BL21(DE3) expressing (T) domain or sub-regions of TA. In answering that question, our novel contribution required live cell imaging, such that any effect possibly due to protein–protein interaction was examined via in-cell nuclear magnetic resonance (NMR) to investigate its applicability in detecting such interactions.

To that end, as described in Chapter 3, constructs expressing the full length of the TA region of ColA (i.e., Residues 1–172) in both the periplasm and the cytoplasm were engineered. Similarly, the TolA box (i.e., Residues 52–172) and TolB box domains (i.e., Residues 1–52), regions where interactions between colicins and Tol proteins occur were cloned into pBAD/gIIIc for secretion into the periplasm, which permitted the study of each binding site in isolation. Next, as described in Chapter 4, The overexpression of these constructs was optimised in *E. coli* BL21(DE3). Live-cell imaging and in-cell NMR required a high concentration of recombinant protein with efficient labelling. Although in-cell NMR has demonstrated its ability to detect folded proteins in cytoplasm (Pielak *et al.*, 2002), because TA has never been tested with in-cell NMR, we sought to detect TA in the cytoplasm. Meanwhile, periplasmic expression was performed to detect any changes in TA domain due to its known interactions with the Tol system, because of where they are located and the interaction occur.

To optimise conditions for both techniques, and thereby obtain high yields of recombinant proteins, certain parameters had to be investigated in terms of ensuring the most efficient implementation of the two expression systems, pET-15b and pBAD/gIIIc, and in the overexpression of the TolA and TolB boxes in pBAD/gIIIc. pBAD/gIIIc is one of the most commonly used vectors in periplasmic protein expression, with expression from the araBAD promoter more tightly controlled than that from equivalent promoters in other periplasmic expression systems (e.g., the pET series vectors). The control of expression levels, based on catabolite repression, makes the pBAD system ideal for producing high protein levels without leakage (Sommer *et al*., 2010). Moreover, expression using the pBAD system can be very effectively modulated by altering the amount of arabinose added to the culture (Sherif *et al*., 2010; Marschall *et al.*, 2017).

The next step was overexpressing isotope labelled TA. Despite being a highly sensitive technique, in-cell NMR cannot detect nuclei unless they are labelled. Although an ideal method to that end is introducing labelled active nuclei inside the cells by labelling newly expressed proteins, the challenge therein is producing a high concentration of the isotopically labelled proteins of interest. To that purpose, a number of approaches were taken to optimise the culture medium used. M9 minimal medium was chosen to induce expression under conditions promoting a reduced growth rate, as suggested by Mondal and colleagues (2013). This medium is also commonly used for incorporating specific residues into proteins of interest (e.g., ^{13}C amino acids) in structural studies or in-cell NMR (Cai *et al.*, 2016). However, as demonstrated in Chapter 4, the results obtained indicated that the level of expression was relatively low, and required improvement, possibly because the periplasmic production of protein is limited by the periplasm's size (Sandomenico *et al.*, 2020). Moreover, the attempt at cytoplasmic expression yielded a similar result see Section 4.2.5, both results possibly because TA is an intrinsically disordered protein, and steric factors would make completing the folding process impossible, as suggested by Dunker and colleagues (2002). That possibility is particularly strong if the protein concerned is involved in critical cellular mechanisms, which the TA undoubtedly is.

Mondal and colleagues (2013) suggested protein production over a long period at a low temperature as a viable approach. However, following that procedure resulted in poor expression of TA, likely because low temperatures reduce final cell density, plasmid degradation and antibiotic inactivation (Rosano & Ceccarelli, 2014). Growing the culture to an OD_{600} of 1.0 before induction was found to be the best approach to optimising protein overexpression and successfully yielded a high level of expression as shown in Section 4.2.7, the high cell density provided an efficient production of recombinant proteins. After achieving high level expression, confocal laser scanning microscopy (CLSM) was employed,

a method developed to detect protein–protein interactions and the co-localisation of proteins (Amos *et al.*, 2003). In CLSM, bacterial viability was tested as indirect measure of the protein interactions are detected using fluorescent dyes with specific excitation wavelengths that stain the cellular region or regions of interest and allow visualisation via microscopy. Thus, to study the nuclease effect of ColE9 on bacterial cell viability, a live/dead bacterial cell assessment was undertaken employing BacLight™ LIVE/DEAD stain in conjunction with CLSM. As reported in Chapter 5, after optimising the overexpression of the proteins, live cell imaging was performed using CLSM, chiefly to challenge the constructs that were engineered, as described in Chapter 3, and overexpressed, as described in Chapter 4, with ColE9–Imm9, whose nuclease-related activity leads to eventual cell death.

In studies on colicins seeking to establish whether transport across the OM is energy-dependent, a major problem has been the lack of appropriate measures of transport that are not dependent on cell death as an end point. To address that problem, we developed a live cell imaging platform using live/dead viability assessment to evaluate the effect of ColE9–Imm9 on *E. coli* cells via live cell imaging, specifically with CLSM. This technique was employed to elucidate the mechanism behind the translocation of colicins, which upon entering cells bind to receptors on the OM and recruit Tol proteins, thereby allowing the (C) domain to reach the IM and exert its killing activity (Majeed *et al.*, 2011). Knowing that the Tol system's chief function is to maintain the integrity of the cell membrane and that its interaction with colicins disturbs the cell membrane and compromises its function (Bernadac *et al.*, 1998), it was expected that such disturbances would allow dyes to pass through the cell membrane and stain the bacterial cells dye entry into damaged cells, but SYTO9 enters intact cells without membrane damage, PI enters those with compromised membrane. The techniques demonstrated in Chapter 5—namely, spot testing and liquid growth assay—confirmed our hypothesis about the protection gained by *E. coli* BL21(DE) cells expressing

TA, TolA box and TolB box, all overexpressed in the periplasm, when they are treated with ColE9–Imm9. The findings underscore the necessity of treating ColE9–Imm9 with dithiothreitol (DTT) before challenging cells in order to reduce the disulphide bond and, in turn, allow ColE9 to exert its activity, which corroborates the findings of Penfold and colleagues (2004).

Given current knowledge about the mechanism of ColE9 translocation, we expected that TA overproduction in the periplasm would not only cause interaction between cellular Tol proteins and TA but would also cause TA to occupy binding sites on ColE9, thereby rendering them unavailable for ColE9 translocation into the periplasm. As explained in Section 5.1, ColE9 binds to TolB in the TolB box region to achieve its translocation into cells. By extension, our hypothesis was that the TolA box's overproduction does not afford the same level of protection as the overproduced TA and TolB box. However, live-cell imaging revealed that *E. coli* BL21(DE3) cells expressing the TolA box showed the greatest protection against ColE9–Imm9 plus DTT, with a 20% difference between it and the results obtained from cells overexpressing the TolB box. A higher-than-expected level of protected cells expressing TA was also observed.

These findings cast doubt on a number of current theories concerning the ColE9 translocation mechanism. As previous work on the translocation mechanism of group A colicins (e.g. ColE9) has demonstrated, those colicins contact their periplasmic binding partners, typically TolB and/or TolA, via a direct epitope delivery mechanism after binding to the receptor on the cell's surface (Housden *et al.*, 2010; Jansen *et al.*, 2020). By extension, the first questionable fact is that ColE9 only interacts with TolB in the periplasm, specifically at the TolB box binding site for its translocation. If true, then our results would show that cells expressing the TolB box had considerable protection against ColE9. At the same time, the lower percentage of viable expressing the TolB box of ColA could presumably relate to

differences in the binding affinity of the TolB boxes of ColE9 and ColA—approximately 1 µM (Loftus *et al.*, 2006) ColA and 10 µM, respectively—to bind with TolB. In that case, their low binding affinity could explain the unexpected result. Even so, if the external colicin were ColA instead of ColE9, it would supposedly show a higher level of protection with the overproduced TolB box.

The second presumed fact that our findings question is the indirect interaction between ColE9 and TolA for translocation. Our study revealed that TolA is in fact likely to be more involved in this process due to the high level of protection against ColE9 activity gained from the overproduced TolA box. That finding supports the suggestion of Rassam and colleagues (2018) that TolA can fully extend through the periplasm via colicin-bound TolB at the OM, thereby displaying the TolA protein's unrestricted Brownian motion prior to its recruitment for translocation, which may explain the occupancy of TA binding sites and prevent the interaction between Tol and ColE9. In that case, however, the outcome would differ with overproduced TA (i.e., the combined TolA box and TolB box). Because the outcomes did not differ, we have to question whether other Tol proteins might also be involved in the translocation process, as previously suggested by Kim and colleagues (2014), who were able to identify TolR contribution to translocation. If the same were true with ColE9–Imm9, then it could explain why the protection provided by TA (i.e., a protein containing both the TolA and TolB boxes) was not as high as expected, namely due to TolR involvement., Szczepaniak and colleagues (2020) recently showed that the physiological role of the TolQ–TolR–TolA complex is to drive the dissociation of TolB–Pal complexes in the OM, thereby releasing Pal to bind the cell wall. As a consequence of that proton motive force (PMF)-driven cycle, TolB is thought to be translocated through the cell wall by TolA. This finding proves the need to investigate the other IM Tol proteins, especially TolR that might fill the gap of knowledge about colicin's translocation into the periplasm.

Despite the level of protection that might vary between the constructs, we believe that the occurrence of protein–protein interactions between overproduced proteins (i.e., TA, the TolA box and the TolB box) and cellular proteins in the periplasm may result in structural conformational changes, which it was hoped would be visualises with in-cell NMR (Burz *et al.*, 2006), if applied correctly. Using a highly sensitive spectrometer, in-cell NMR can reveal macromolecules in multidimensional states as well as their dynamics and conformational changes inside cells (Breindel *et al.*, 2019). Although in-cell NMR has some limitations that affect its use in biological systems, elucidating those limitations and identifying their factors helped us to develop a method that, in turn, clarified the nature of the (T) domain and its residues and produced some interesting findings in relation to our CLSM analysis, especially regarding the TolA box of ColA.

In-cell NMR experiments have revealed the method's limitations and challenges in this project, especially with intrinsically disordered proteins such as TA. In response, the project's overarching goal was to develop a straightforward, effective technique for detecting the expressed protein via in-cell NMR. We therefore investigated various experimental methods to determine which parameters were important in performing an effective in-cell NMR analysis of the proteins of interest in the *E. coli* system. The four parameters determined to be the most important were the culture medium used to grow the bacterial cells, the expression system the proteins of interest were cloned into, the volume of the initial culture to produce higher cell density and the size of the expressed protein of interest. Even after attempting to optimise conditions for in-cell NMR analysis, detecting the structured part of the TA protein remained difficult, the results achieved did nevertheless provide some support for the live cell imaging data obtained.

Because the TA domain and the TolA and TolB boxes are intrinsically disordered proteins, their intrinsic flexibility was expected to produce some broad peaks and strong signals with ^{1}H in in-cell NMR. As the results reported in Chapter 5 suggest, the presence of interactions of the expressed proteins (i.e., TA protein and the TolA and TolB boxes) with cellular proteins from in-cell NMR analysis should result in those proteins producing wider overlapping peaks indicating their involvement in protein–protein interactions inside *E. coli* cells, which might cause a loss of signals. Kim and colleagues (2014), who investigated the structural conformation of two type A colicins, ColN (i.e., a pore-forming colicin) and ColE3 (i.e., deoxyribonuclease), discovered that the (T) domains of both are divided into two parts: an intrinsically unstructured (T) domain (IUTD), which could be what was observed during these studies, and a larger structured (T) domain distal to the IUTD, which we aimed to detect with in-cell NMR, considering the high structural similarity between ColA and ColN.

The results achieved by analysing cells expressing TA with 1D ^{1}H and ^{13}C NMR did not reveal any protein signal. Meanwhile, 2D ^{1}H–^{15}N correlation NMR could detect IUTD only. Despite attempts to optimise conditions by several means: growing cells until the $O.D_{600}$ reached 1 before induction; and by growing the cells in LB medium first before switching them to labelled M9 minimal medium, preparing a 60% v/v slurry for analysis gave the best results and comes in agreement with Ikeya and colleagues research (2016). Some of those conditions were recently reviewed by Sugiki and colleagues (2020), who determined that the dependency of *E. coli*'s in-cell NMR spectra on the quality of the NMR spectra derived from the protein of interest was dependent on the growth phase of the host culture at time of expression target protein's stage along the growth stage arises from a yield of soluble recombinant proteins. For that reason, optical density is a sound tool for monitoring optimal conditions for heterologous protein overexpression.

The hypothesis concerning protein–protein interaction between overproduced proteins and cellular proteins suggests the inevitable formation of a complex molecule with a large molecular weight, which given the problems typically associated with NMR analysis of large proteins or protein complexes would consequently be likely to generate numerous low-intensity peaks with broad linewidths, as was seen in our analysis of the TA and the TolB box proteins. The increased molecular weight of the target protein reduces relaxation time, T2, which is the first step in any NMR experiment, and a delay or extension in this parameter will typically result in broader cross-peaks (Wuthrich, 1986). The resulting NMR signal from larger molecules also decays more rapidly and prompts line broadening (Burz *et al.*, 2019), which possibly happened in our study which provide some support to our hypothesis.

In light of poor results achieved with the TolA box (see Section 6.4.8), researchers have continued to investigate whether hub proteins have specific structural properties that facilitate their participation in various interactions with other proteins, as well as determining whether they have any structural flexibility for folding and changing conformation. Because hub proteins can be recognised by their disordered regions, they most are large, unfolded proteins with no secondary or tertiary structure (Patil *et al.*, 2010), it is reasonable to assume that the inherent nature of the TolA complex makes its detection difficult with in-cell NMR. However, the result supports our finding from live-cell imaging regarding a direct role for TolA in the translocation of ColE9. Studies to date have shown that no structures are available for intact TolA (Szczepaniak et al., 2020), while numerous laboratory experiments have demonstrated that TolA is a hub for protein–protein interaction, one able to form complexes with Tol proteins. Nevertheless, the proposed detail regarding most of those interactions have proven controversial, either due to a lack of corroborating biochemical data or because they are contradicted by other work (Szczepaniak *et al.*, 2020). Thus, findings suggest the involvement

of TolA in translocation process. However, the question remains: What is the exact role of TolA in the translocation of ColE9?

Altogether, the primary goal of this project was to determine whether the intramolecular binding properties for the TA of ColA are disrupted when interacting with the cellular Tol protein. As demonstrated in Chapter 5, the overexpressed TA interaction with cellular Tols provides protection against ColE9. At the same time, the results obtained here using in-cell NMR may have resulted from the intramolecular binding of TA and cellular Tol proteins and their interaction with Tol proteins, which could have generated larger complexes that are difficult to detect with in-cell NMR. The overproduction of the proteins of interest might have also caused the formation of large complexes, meaning that the target protein effectively behaved as a large protein and thus could not be readily detected by in-cell NMR (Gibbs *et al*., 2017). Another theory is that the overexpressed TA inside the cells needs be in an unfolded state in order to interact with other proteins in a similarly way to that for external colicins. As a case in point, when external colicins interact with cell membrane, and subsequently translocate through the membrane, they unfold to become able to span the periplasm and interact with the Tol proteins in the IM (Anderluh *et al*., 2003). The same dynamic could occur with the TolB box domain expressed inside the cells, hence its appearance in wide peaks but with a distribution similar to that of unfolded proteins.

Another difficulty faced in the project was that assigning the unstructured part of TA and the TolB box was difficult. Their unstructured nature requires repeated scanning through the ^{15}N planes of the 3D spectra in order to locate peaks at the desired chemical shifts, which is a highly time-consuming process and therefore threatens cell viability and causes loss of signals. In general, residue-type identification along the chain that requires removing ambiguities is a slow process that would greatly benefit from better set-ups and advanced equipment. Because the process is so slow, due to technical difficulties in analysing spectra

and complications in the interpretation of NMR data, the number of proteins investigated in their unfolded state to date has been far smaller than the number of folded ones (Luchinat *et al.*, 2020).

Despite progress in in-cell NMR methods, the presence of target molecules within cells with other macromolecules requires an efficient, ideally targeted, labelling technique in order to reduce the background of the spectra. Target proteins are liable to interact with other proteins and thereby form large complexes, which can lower signal sensitivity. The mechanism of action of colicins remains under investigation, and the results achieved thus far suggest that the translocation domain is engaged in many interactions that have yet to be described. Moreover, non-specific interactions may occur inside living cells and consequently reduce NMR signals as well. Hecht (2012) tested the translocation domain of ColN (TN), a pore-forming colicin, by *in vitro* NMR, and the results were the same as obtained in this work for the TA of ColA, which demonstrated the presence of a complex unstructured protein. However, when titrated with TolA, TN-TolA also showed more peaks than TN alone, which suggests a conformational change in the binding site and, in turn, intramolecular interactions between the ColN and TolA in the translocation domain. In view of such interactions, a conclusion made that the fluctuating clusters of side-chain interactions affect the (T) domain's intramolecular bindings. Another conclusion is that the closed conformation of the (T) domain is disrupted when it interacts with proteins in its target *E. coli* cell, which could also be the case with TA.

7.2 Future work

E. coli BL21(DE3) expressing TA, the TolA box and the TolB box were treated with ColE9–Imm9 and analysed by CLSM. The results of CLSM indicated that the overexpressed TA domain protects host cells against ColE9–Imm9, as well as the TolA and TolB boxes. The level of such protection within the same range suggests that the TolA box interacts with cellular TolA and that the occupied TolA blocks the activity of ColE9–Im9. To confirm that the responsible factor was the dissociation of Imm9 from the complex because TolA was no longer available to transfer the energy required for Imm9 release, ColE9 and Imm9 could be labelled with different fluorophores following their purification. With one fluorophore acting as a fluorescent receptor, and the other as a donor, each would then fluoresce when Imm9 was released from the complex. To expand upon those findings, a time-lapse video of the interaction could visualise the localisation and movement through the cells during CLSM and clarify whether Imm9 release is affected by the overexpression of the TolA box inside the cells.

TolR protein is believed to be involved in the translocation process of ColE9, applying the same analysis employed in this project on *E. coli* BL21(DE3) cells expressing TolR, would help answer the question about the unexpected level of protection that gained from cells expressing TA and TolB box against the cytotoxic effect of ColE9.

The applicability of using in-cell NMR on proteins in the (T) domain could also be clarified by determining whether those proteins are in unfolded states and could be subjected to a higher-dimensional experiment (e.g., 4D or 5D), which would reveal a wider spectrum of peaks and could facilitate the distribution of the signals. Increasing the dimensionality of the NMR spectra in future work might help by distributing the peaks across a larger spectral space and, in turn, reduce crowding (Grudziaz *et al.*, 2018). A recent advance in in-cell NMR

spectrometry is the development of multidimensional NMR with pulse sequences, including 5D NMR (Motáčková *et al*., 2010; Nováček *et al*., 2011). The technique has the capability of providing higher resolution but requires time-consuming measurements for each sample, which could affect the stability of unfolded proteins given the potential for aggregation as well as cell's viability (Okataviani, 2014). Also, performing that method with an NMR device with a magnet size exceeding 800 MHz would afford greater sensitivity. A new commercially available NMR spectrometer operating at 1.2 GHz has attracted keen interest in the research community, and the advantages of using the new instrument for NMR experiments have already been shown (Banci *et al.*, 2019), especially if provided with a bioreactor (i.e., a device to provide oxygen and nutrients to cells to maintain their viability) and for investigating biological processes with real-time in-cell NMR (Luchinat *et al.*, 2020). Such techniques can identify the location of all NMR signals in the slow exchange regime, where signals change shape or chemical shifts occur as a function of time.

Another suggested approach is specific labelling instead of uniform labelling, which can reduce the noise of the overlapping peaks, similar to labelling a specific amino acid. By following those suggestions, future work could provide more knowledge about and direct current understandings of the protein–protein interactions that occur inside bacterial cells.

Appendix A

A1 Genotype description

DH5 α genotype description

• *rec*A1 ensures increased insert stability and prevents unwanted recombination.

• *end*A1 improves the yield and quality of plasmid DNA prepared from minipreps.

• DH5α competent cells support replication of *M13mp* vectors but do not support plaque formation**.**

- *rec*A1 and *end*A1 mutations in DH5α cells increase insert stability and improve the quality of plasmid DNA prepared from minipreps.
- The mutations that the DH5-Alpha strain has are: dlacZ Delta M15 Delta(lacZYA-argF) U169 recA1 endA1 hsdR17(rK-mK+) supE44 thi-1 gyrA96 relA1. These mutations correspond to the distinct characteristics that make the DH5-Alpha strain excel in laboratory cloning procedures.

BL21(DE3) genotype description

- T7 RNA Polymerase: (*T7 gene1)* is encoded by the lambda *DE3* prophage present within the chromosome. T7 RNA polymerase is expressed from the *lacUV5* promoter, which is less sensitive to catabolite repression than the wt *lac* promoter.
- *DE3* strains may exhibit uninduced target protein expression. Although λ*DE3* is normally dormant in the host chromosome, the induction of the SOS cascade can occur as the result of expressing proteins that damage the *E. coli* chromosome, either directly or indirectly. This may lead to cell lysis.
- Protease Deficient (*[lon] ompT*): *E. coli* B strains are "naturally" deficient in the *lon* protease which in K-12 strains serves to degrade misfolded proteins and to prevent some cell cycle-specific proteins from accumulating. The OmpT protease resides at the surface of wild type *E. coli* in both K-12 and B strains, presumably helping the cells to derive amino acids from their external environment.
- T1 Phage Resistant (*fhuA2*): T1, an extremely virulent phage requires the *E. coli* ferric hydroxamate uptake receptor for infectivity. Deletion of this gene confers resistance to this type of phage, but does not significantly affect the transformation or growth characteristics of the cell.

A2 Vector maps

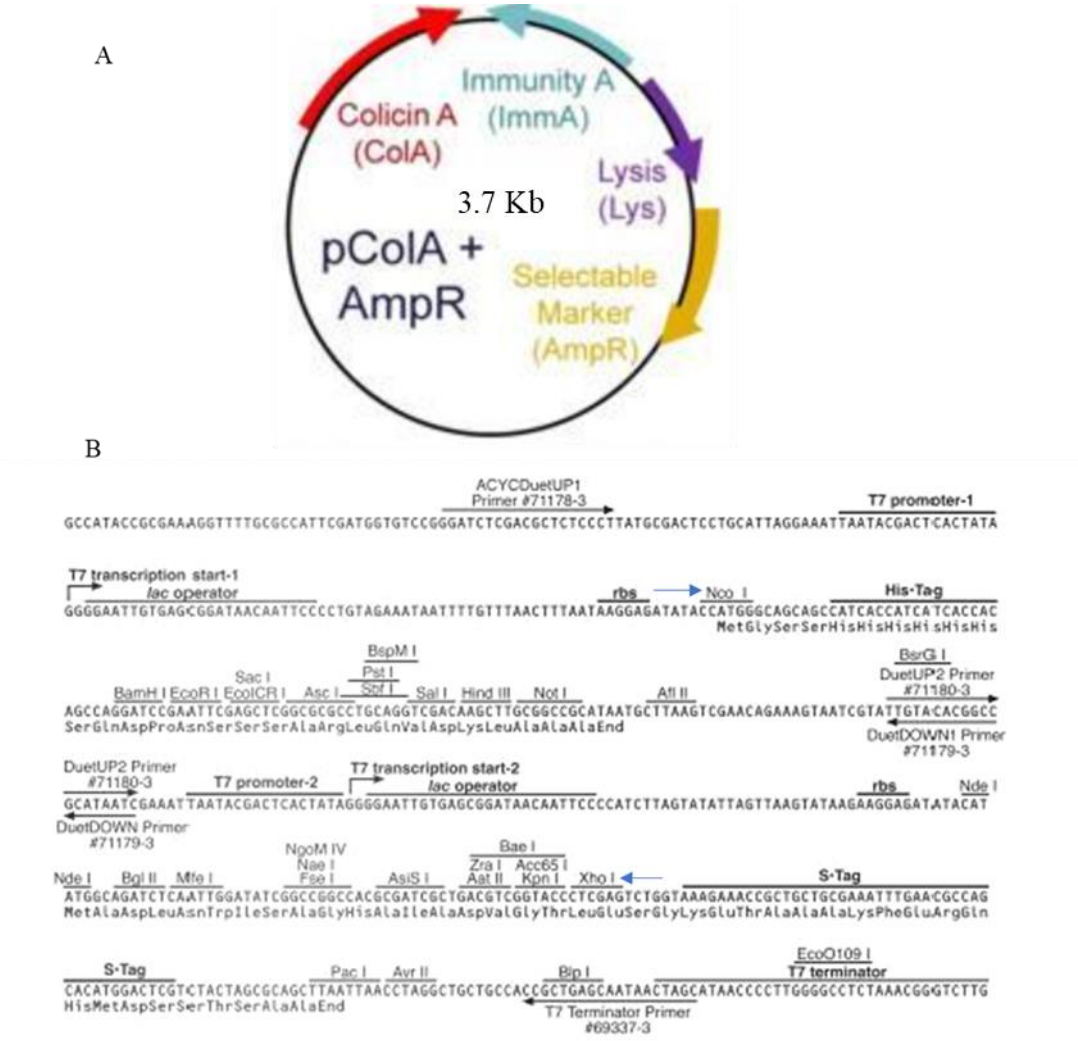


Figure A1 Map sequence of the pColA vector (Novagen.com). A) Shows the vector's map of pColA. The vector has the COLA replicon from ColA (1), encoding ampicillin resistance gene. B) Shows The vector encodes two multiple cloning sites (MCS) each of which is preceded by a T7 promoter, lac operator, and ribosome binding site (rbs). Multiple cloning site sequence with restriction sites used for cloning, NcoI and XhoI, indicated by blue arrows.

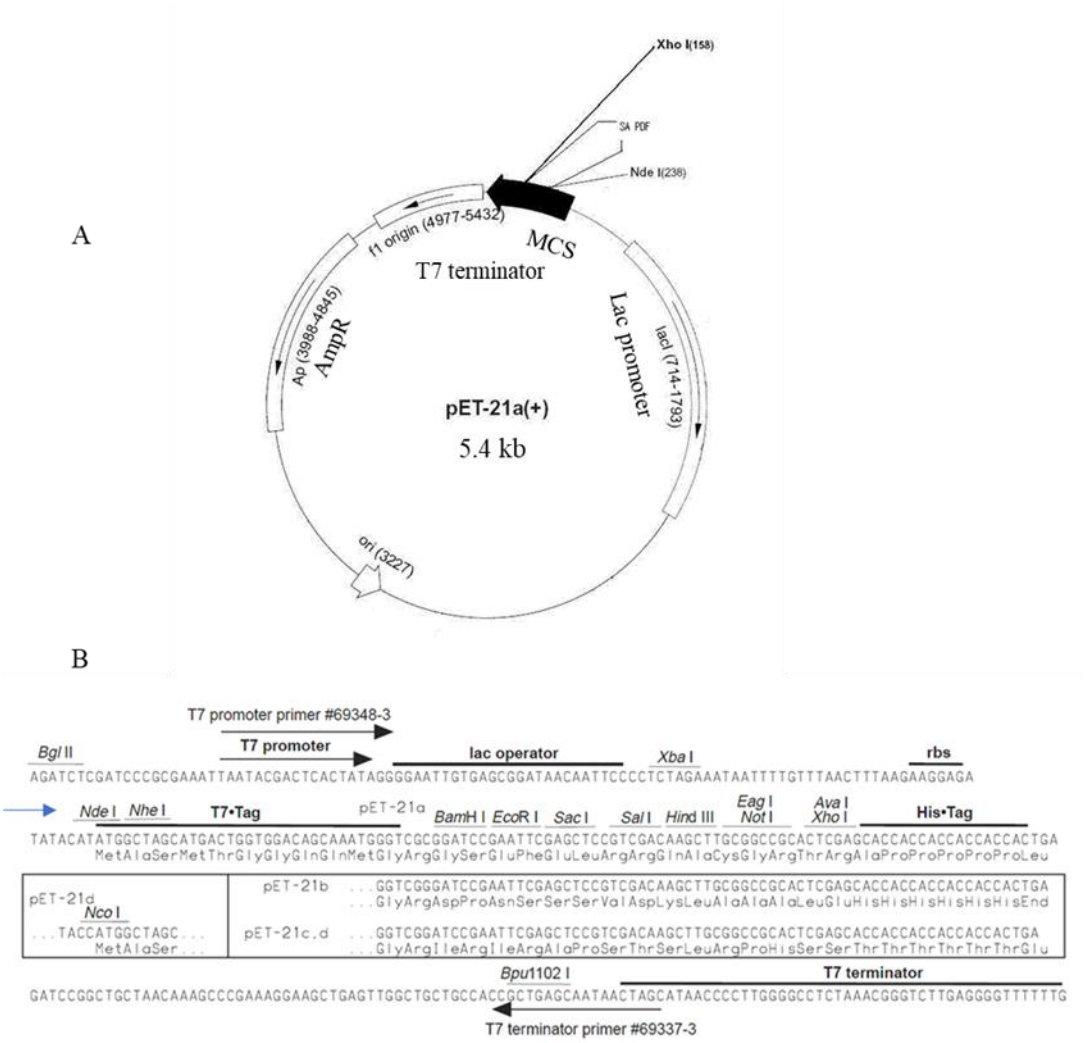


Figure A2 Map sequence of the pET-21a vector (Novagen.com). A) Shows the vector's map of pET-21a, carry an N-terminal T7•Tag sequence plus an optional C-terminal His•Tag sequence. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase, encoding ampicillin resistance gene. B) Shows multiple cloning site sequence with restriction sites used for cloning, NdeI and XhoI, indicated by blue arrows.

A 3 Gene sequence of TA, TolA box and TolB box of ColA

TA (ColA 1-172)

A

TACCTGAAGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTAACC
ATGAAAAAACTGCTGTTCGCGATTCCGCTGGTGGTGCCGTTCTATAGCCATAGCACCATGGCTGGATTTAAT
TATGGTGGAAAAGGTGATGGAACCGGCTGGAGCTCAGAACGTGGGAGTGGTCCAGAGCCGGGTGGTGGT
AGCCACGGAAATAGTGGTGGGCACGATCGTGGAGATTCTTCCAACGTAGGTAATGAGTCTGTGACGGTAAT
GAAACCAGGGGATTCGTATAACACCCCGTGGGGAAAAGTCATCATCAATGCTGCAGGCCAGCCGACCATGA
ACGGAACGGTGATGACCGCTGATAATTCATCGATGGTTCCTTACGGCAGAGGGTTTACACGGGTTTTAAATT
CCCTGGTCAATAATCCTGTTTCGCCGGCAGGTCAGAATGGCGGGAAGTCTCCTGTTCAGACTGCTGTGGAAA
ATTATCTGATGGTACAGTCAGGAAACCTGCCACCGGGCTACTGGCTCAGTAATGGCAAGGTTATGACGGAG
GTTCGTGAGGAACGTACTTCTGGCGGCGGTGGGAAAAACGGGAACGAGCGAACCTGGACTGTGAAAGTTC
CCCGGCTCGAGATCTGCAGCTGGTACCATATGGGAATTCGAAGCTTTCTAGAACAAAAACTCATCTCAGAAG
AGGATCTGAATAGCGCCGTCGACCATCATCATCATCATCATTGAGTTTAAACGGTCTCCAGCTTGGCTGTTTT
GGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGA
ATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCATGCCGAACTCAGAAGTGAAACGCCGTAGCG
CCGATGGTAGT

B

Sequence ID: Query_64317 Length: 485 Number of Matches: 1

Range 1: 1 to 485 Graphics ▼ Next Match ▲ Previous M

Score	Expect	Identities	Gaps	Strand
891 bits(482)	0.0	484/485(99%)	0/485(0%)	Plus/Plus

```
Query  32   ATGGAACCGGCTGGAGCTCAGAACGTGGGAGTGGTCCAGAGCCGGGTGGTGGTAGCCATG  91
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  1    ATGGAACCGGCTGGAGCTCAGAACGTGGGAGTGGTCCAGAGCCGGGTGGTGGTAGCCACG  60

Query  92   GAAATAGTGGTGGGCACGATCGTGGAGATTCTTCCAACGTAGGTAATGAGTCTGTGACGG  151
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GAAATAGTGGTGGGCACGATCGTGGAGATTCTTCCAACGTAGGTAATGAGTCTGTGACGG  120

Query  152  TAATGAAACCAGGGGATTCGTATAACACCCCGTGGGGAAAAGTCATCATCAATGCTGCAG  211
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  TAATGAAACCAGGGGATTCGTATAACACCCCGTGGGGAAAAGTCATCATCAATGCTGCAG  180

Query  212  GCCAGCCGACCATGAACGGAACGGTGATGACCGCTGATAATTCATCGATGGTTCCTTACG  271
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCCAGCCGACCATGAACGGAACGGTGATGACCGCTGATAATTCATCGATGGTTCCTTACG  240

Query  272  GCAGAGGGTTTACACGGGTTTTAAATTCCCTGGTCAATAATCCTGTTTCGCCGGCAGGTC  331
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAGGGTTTACACGGGTTTTAAATTCCCTGGTCAATAATCCTGTTTCGCCGGCAGGTC  300

Query  332  AGAATGGCGGGAAGTCTCCTGTTCAGACTGCTGTGGAAAATTATCTGATGGTACAGTCAG  391
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AGAATGGCGGGAAGTCTCCTGTTCAGACTGCTGTGGAAAATTATCTGATGGTACAGTCAG  360

Query  392  GAAACCTGCCACCGGGCTACTGGCTCAGTAATGGCAAGGTTATGACGGAGGTTCGTGAGG  451
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GAAACCTGCCACCGGGCTACTGGCTCAGTAATGGCAAGGTTATGACGGAGGTTCGTGAGG  420

Query  452  AACGTACTTCTGGCGGCGGTGGGAAAAACGGGAACGAGCGAACCTGGACTGTGAAAGTTC  511
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  AACGTACTTCTGGCGGCGGTGGGAAAAACGGGAACGAGCGAACCTGGACTGTGAAAGTTC  480

Query  512  CCCGG  516
            |||||
Sbjct  481  CCCGG  485
```

Figure A3 Gene sequence of colicin A translocation domain (TA). A) Gene sequence of Max ORF: 1-516, MW=17782. B) Sequence alignment' result using BLAST.

TolA box (ColA 52-172)

A

ATGGAACCAGGGGATTCGTATAACACCCCGTGGGGAAAAGTCATCATCAATGCTGCAGGCCAGCCGACCAT
GAACGGAACGGTGATGACCGCTGATAATTCATCGATGGTTCCTTACGGCAGAGGGTTTACACGGGTTTTAA
ATTCCCTGGTCAATAATCCTGTTTCGCCGGCAGGTCAGAATGGCGGGAAGTCTCCTGTTCAGACTGCTGTGG
AAAATTATCTGATGGTACAGTCAGGAAACCTGCCACCGGGCTACTGGCTCAGTAATGGCAAGGTTATGACG
GAGGTTCGTGAGGAACGTACTTCTGGCGGCGGTGGGAAAAACGGGAACGAGCGAACCTGGACTGTGAAA
GTTCCCCGG

B

Sequence ID: Query_57183 Length: 363 Number of Matches: 1

Range 1: 1 to 363 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
665 bits(360)	0.0	362/363(99%)	0/363(0%)	Plus/Plus

```
Query  123  ATGAAACCAGGGGATTCGTATAACACCCCGTGGGGAAAAGTCATCATCAATGCTGCAGGC  182
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGGAACCAGGGGATTCGTATAACACCCCGTGGGGAAAAGTCATCATCAATGCTGCAGGC  60

Query  183  CAGCCGACCATGAACGGAACGGTGATGACCGCTGATAATTCATCGATGGTTCCTTACGGC  242
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CAGCCGACCATGAACGGAACGGTGATGACCGCTGATAATTCATCGATGGTTCCTTACGGC  120

Query  243  AGAGGGTTTACACGGGTTTTAAATTCCCTGGTCAATAATCCTGTTTCGCCGGCAGGTCAG  302
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AGAGGGTTTACACGGGTTTTAAATTCCCTGGTCAATAATCCTGTTTCGCCGGCAGGTCAG  180

Query  303  AATGGCGGGAAGTCTCCTGTTCAGACTGCTGTGGAAAATTATCTGATGGTACAGTCAGGA  362
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  AATGGCGGGAAGTCTCCTGTTCAGACTGCTGTGGAAAATTATCTGATGGTACAGTCAGGA  240

Query  363  AACCTGCCACCGGGCTACTGGCTCAGTAATGGCAAGGTTATGACGGAGGTTCGTGAGGAA  422
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  AACCTGCCACCGGGCTACTGGCTCAGTAATGGCAAGGTTATGACGGAGGTTCGTGAGGAA  300

Query  423  CGTACTTCTGGCGGCGGTGGGAAAAACGGGAACGAGCGAACCTGGACTGTGAAAGTTCCC  482
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CGTACTTCTGGCGGCGGTGGGAAAAACGGGAACGAGCGAACCTGGACTGTGAAAGTTCCC  360

Query  483  CGG  485
            |||
Sbjct  361  CGG  363
```

Figure A4 Gene sequence of colicin A TolA box (TA 52-172). A) Gene sequence of Max ORF: 1-363, MW=12882. B) Sequence alignment' result using BLAST.

TolB box (ColA 1-52)

A

ATGGAACCGGCTGGAGCTCAGAACGTGGGAGTGGTCCAGAGCCGGGTGGTGGTAGCCATGGAAATAGTG
GTGGGCACGATCGTGGAGATTCTTCCAACGTAGGTAATGAGTCTGTGACGGTAATGAAACCA

B

Sequence ID: M37402.1 Length: 6720 Number of Matches: 1

Range 1: 3327 to 3457 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
243 bits(131)	3e-60	131/131(100%)	0/131(0%)	Plus/Plus

```
Query  1     ATGGAACCGGCTGGAGCTCAGAACGTGGGAGTGGTCCAGAGCCGGGTGGTGGTAGCCATG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3327  ATGGAACCGGCTGGAGCTCAGAACGTGGGAGTGGTCCAGAGCCGGGTGGTGGTAGCCATG  3386

Query  61    GAAATAGTGGTGGGCACGATCGTGGAGATTCTTCCAACGTAGGTAATGAGTCTGTGACGG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3387  GAAATAGTGGTGGGCACGATCGTGGAGATTCTTCCAACGTAGGTAATGAGTCTGTGACGG  3446

Query  121   TAATGAAACCA  131
             ||||||||||
Sbjct  3447  TAATGAAACCA  3457
```

Figure A5 Gene sequence of colicin A TolB box (TA 1-52). A) Gene sequence of MaxORF:1-156, MW=5032. B) Sequence alignment' result using BLAST.

Appendix B

B1 Buffers and reagents

PBS

NaCl 8 g

KCl 200 mg

Na_2HPO_4 1.44 g

KH_2PO_4 245 mg

Made up 1L with dH_2O. The pH adjusted to 7.4

M9 salts

Na_2HPO_4 6 g/l

KH_2PO_4 3 g/l

NaCl 0.5 g/l

Dissolved with dH_2O. The pH was adjusted to 7.4

Minimal medium broth

Casamino acid 2% (w/v)

Glucose 0.2% (w/v)

M9 salts 1x

Biotin 15% (w/v)

Thiamine 15% (w/v)

0.1 M CaCl2 140 g

1 M MgSO4 124 g

Made up to 1L with dH_2O. The pH was adjusted to 7.4. then the broth is autoclaved.

Labelled M9 minimal medium

Minimal medium broth 1L

$^{15}NH_4Cl$ 1 g/l

^{13} Glucose 1 g/l

Coomassie stain

Acetic acid 10% (v/v)

G-250 Coomassie 0.025 % (w/v)

De-staining solution

Methanol 30% (v/v)

Acetic acid 10% (v/v)

Agarose gel

Agarose 1% (w/v)

TAE buffer 2 ml (5x)

Made up to 10 ml with dH_2O. Melt and add 5 µl ethidium bromide (10 mg/ml)

Sodium dodecyl sulfate (SDS- loading buffer 4x)

1 M Tris HCl, pH 6.5 2 ml

1 M DTT 4 ml

SDS 0.8 g

Bromophenol blue 40 mg

Glycerol 3.2 ml

Made up to 10 ml with dH_2O. make 500 ul aliquots and store at -20°C.

Sodium dodecyl sulfate (SDS- running buffer 10x)

0.25 M Tris base	30.3 g
2.5 M glycine	187.7 g
1% (w/v) SDS	10 g

Made up to 1L with dH_2O.

Resolving gel

1.5 M Tris-Cl (PH 8.8)	1.7 ml
30% (w/w) Acrylamide/Bis-acrylamide	2.33 ml
10% (w/v) Ammonium persulfate (APS)	30 µl
Tetramethyl ethylenediamine (TEMED)	30 µl
dH_2O	2.88 ml

Stacking gel

0.5 M Tris-Cl (pH 6.8)	1 ml
30% (w/w) Acrylamide/Bis-acrylamide	0.8 ml
10% (w/v) Ammonium persulfate (APS)	30 µl
Tetramethyl ethylenediamine (TEMED)	30 µl
dH_2O	2.18 ml

Lysis buffer

65 mM Tris	0.078 g
2% (w/v) SDS	0.2 g
10% (v/v) glycerol	1 ml

Made up 10 ml with dH_2O

Transfer buffer

Tris- base	5.8 g
Glycine	2.9 g
SDS	0.37 g
Methanol	200 ml

Made up 1L with dH_2O

Semi-dry blotting buffer

Tris- base	5.82 g
Glycine	2.93 g
10% (w/v) SDS	3.75 ml
Methanol	200 ml

Made up 1L with dH_2O

10x TBS buffer

Tris- base	24 g
NaCl	88 g

Made up 1L with dH_2O. the PH adjusted to 7.6

TBST buffer

10x TBS	100 ml
Tween 20	1 ml

Made up 1L with dH_2O

1 M stock Isopropyl-β -d-1 thiogalactopropanoside (IPTG)

IPTG	23 g
dH_2O	10 ml

Sterilized by filtration and stored at -20°C

Ampicillin stock

Ampicillin 1 g

dH_2O 10 ml

sterilized by filtration and stored at -20°C

Tube	Volume (mL)	Stock solution	Final concentration
1 (1 ml)	0.1	0.002%	0.00002%
2 (1 ml)	0.1	0.02%	0.0002%
3 (1 ml)	0.1	0.2%	0.002%
4 (1 ml)	0.1	2%	0.02%
5 (1 ml)	0.1	20%	0.2%

Table B1 Serial dilution of 20% L-arabinose.

	Anode buffer 10x	Cathode buffer 10x	Gel buffer 3x
Tris HCl M	1	1	3
Tricine M		1	
SDS		1	0.3
pH	8.9	8.25	8.45

Table B2 Tricine-SDS-PAGE buffers.

AB-3	0.5 mL	5 mL
Gel buffer 3x	**1.5 mL**	**5 mL**
Glycerol		3 mL
H_2O	**Up to 6 mL**	**Up to 15 mL**
APS 10%	45 µL	50 µL
Tetramethylethylenediamine (TEMED)	**4.5 µL**	**5 µL**

Table B3 Tricine-SDS-PAGE gel components.

Appendix C

Statistical data

Df= degree of freedom, SD= Standard deviation, Asymp. Sig= p-value based on chi-square approximation.

C1 Analysis of the antimicrobial activity of ColE9-Imm9 with and without DTT complex against *E. coli* BL21(DE3) in liquid culture

Bacterial strains	**Control**	**Experimental**
	Mean±SD	
Bl21DE3	**0.301±0.073**	**0.249±0.049**
Bl21/10	**0.296±0.076**	**0.145±0.032**
Bl21/100	**0.296±0.076**	**0.122±0.008**

Table C1 Descriptive statistics.

Levene statistics	Df1	Df2	significance
13.252	**2**	**44**	**0.001**

Table C2 Homogeneity of variance.

Shapiro-Wilk	**statistic**	**df**	**significance**
	0.867	36	0.000

Table C3 Normality test.

Chi-Square	**4.511**
df	2
Asymp. significance	0.105

Table C4 Non-parametric Kruskal Wallis test.

C2 Analysis of the antimicrobial activity of ColE9-Imm9 with and without DTT complex against *E. coli* BL21(DE3) expressing TA, TolA box and TolB box compared to *E. coli* BL21(DE3) in liquid culture

	TA	TolA	TolB
	Mean±SD		
Mean	0.338±0.056	0.340±0.058	0.359±0.060

Table C5 Descriptive statistics.

TA cells

Levene statistics	Df1	Df2	significance
14.353	**3**	**60**	**0.001**

Table C6 Homogeneity of variance.

Shapiro-Wilk	statistic	df	significance
	0.879	64	0.001

Table C7 Normality test.

Chi-Square	36.522
df	3
Asymp. significance	0.001

Table C8 Non-parametric Kruskal Wallis test

TolA

Levene statistics	Df1	Df2	significance
13.737	**3**	**60**	**0.001**

Table C9 Homogeneity of variance.

Shapiro-Wilk	**statistic**	**df**	**significance**
	0.866	64	0.001

Table C10 Test of normality.

Chi-Square	**35.740**
df	2
Asymp. significance	0.001

Table C11 Non-parametric Kruskal Wallis test.

TolB

Levene statistics	Df1	Df2	significance
13.133	**3**	**60**	**0.001**

Table C12 Homogeneity of variance.

Shapiro-Wilk	**statistic**	**df**	**significance**
	0.867	64	0.001

Table C13 Test of Normality

Chi-Square	**36.083**
df	3
Asymp. significance	0.001

Table C14 Non-parametric Kruskal Wallis test.

C3 Analysis of the antimicrobial activity of ColE9-Imm9 on E.coli using Live/Dead staining in conjunction with CLSM

Group		After 3 hrs	After 18 hrs
		Mean±SD	
Bl21	Live	314.17±344.48	541.78±232.28
	Dead	82.27±98.60	41.28±40.75
Bl21/10	Live	321.98±195.98	402.10±290.61
	Dead	218.97±172.57	219.43±176.36
Bl21/100	Live	311.65±294.25	311.65±190.50
	Dead	362.65±294.54	372.52±184.45

Table C15 Descriptive statistics.

	Levene statistics	Df1	Df2	significance
3h				
Live	**3.535**	**2**	**177**	**0.031**
Dead	**29.732**	**2**	**177**	**0.000**
18 h				
Live	**6.137**	**2**	**177**	**0.003**
Dead	**43.973**	**2**	**177**	**0.000**

Table C16 Homogeneity of variance.

Shapiro-Wilk	**statistic**	**df**	**significance**
3h			
Live	0.874	180	0.000
Dead	0.172	180	0.000
18 h			
Live	0.907	180	0.000
Dead	0.882	180	0.000

Table C17 Test of Normality

Kruskal Wallis 3h	Chi-square	df	Asymp.significance
Live	3.185	2	0.203
Dead	46.460	2	0.000
18 h			
Live	59.813	2	0.000
Dead	102.509	2	0.000

Table C18 Non-parametric Kruskal Wallis test.

TA

Group		After 3 hrs	After 18 hrs
		Mean±SD	
TA	Live	605.42±248.76	525.80±169.01
	Dead	38.08±35.68	37.72±46.25
TA 10	Live	686.93±342.39	481.27±225.65
	Dead	205.70±172.63	218.53±160.56
TA 100	Live	359.55±205.43	472.57±278.19
	Dead	229.05±172.06	293.85±213.08

Table C19 Descriptive statistics.

	Levene statistics	Df1	Df2	significance
3h				
Live	**4.478**	**2**	**177**	**0.013**
Dead	**28.129**	**2**	**177**	**0.000**
18 h				
Live	**8.546**	**2**	**177**	**0.000**
Dead	**43.426**	**2**	**177**	**0.000**

Table C20 Homogeneity of variance

Shapiro-Wilk	statistic	df	significance
3h			
Live	0.940	180	0.000
Dead	0.822	180	0.000
18 h			
Live	0.907	180	0.000
Dead	0.882	180	0.000

Table C21 Test of normality.

Kruskal Wilk	Chi-square	df	Asymp.significance
3h			
Live	38.479	2	0.000
Dead	78.096	2	0.000
18 h			
Live	2.366	2	0.306
Dead	86.414	2	0.000

Table C22 Non-parametric Kruskal Wallis test.

TolA

Group		After 3 hrs	After 18 hrs
		Mean±SD	
TolA	Live	747.72±204.52	510.35±177.05
	Dead	36.32±31.75	31.63±36.64
TolA 10	Live	587.72±365.36	641.10±306.87
	Dead	309.03±222.50	296.00±210.00
TolA 100	Live	597.90±503.28	639.20±319.46
	Dead	427.98±379.83	238.88±266.17

Table C23 Descriptive statistics.

	Levene statistics	Df1	Df2	significance
3h				
Live	**9.290**	**2**	**177**	**0.000**
Dead	**25.172**	**2**	**177**	**0.000**
18 h				
Live	**9.487**	**2**	**177**	**0.000**
Dead	**22.252**	**2**	**177**	**0.000**

Table C24 Homogeneity of variance.

Shapiro-Wilk	**statistic**	**df**	**significance**
3h			
Live	0.890	180	0.000
Dead	0.768	180	0.000
18 h			
Live	0.975	180	0.003
Dead	0.749	180	0.000

Table C25 Test of normality.

Kruskal Wallis	**Chi-square**	**df**	**Asymp.significance**
3h			
Live	18.606	2	0.000
Dead	94.641	2	0.000
18 h			
Live	7.780	2	0.020
Dead	92.371	2	0.000

Table C26 Non-parametric Kruskal Wallis test.

TolB

Group		After 3 hrs	After 18 hrs
		Mean±SD	
TolB	Live	597.03±242.69	472.37±219.92
	Dead	38.52±43.44	31.37±32.44
TolB/ 10	Live	364.50±266.05	872.17±533.54
	Dead	242.32±172.33	363.05±258.75
TolB /100	Live	528.08±280.40	755.20±417.32
	Dead	206.77±195.08	648.55±434.62

Table C27 Descriptive statistics.

	Levene statistics	Df1	Df2	significance
3h				
Live	**1.100**	**2**	**177**	**0.335**
Dead	**29.182**	**2**	**177**	**0.000**
18 h				
Live	**35.414**	**2**	**177**	**0.000**
Dead	**27.252**	**2**	**177**	**0.000**

Table C28 Homogeneity of variance.

Shapiro-Wilk	statistic	df	significance
3h			
Live	0.947	180	0.000
Dead	0.825	180	0.000
18 h			
Live	0.869	180	0.003
Dead	0.786	180	0.000

Table C29 Test of normality.

Kruskal Wallis	Chi-square	df	Asymp.significance
3h			
Live	18.606	2	0.000
Dead	66.990	2	0.000
18 h			
Live	16.540	2	0.000
Dead	103.409	2	0.000

Table C30 Non-parametric Kruskal Wallis test.

C4 Plate counts to determine the antimicrobial activity of ColE9-Imm9 plus DTT on *E. coli* BL21(DE3)

Chi-Square	7.261
df	2
Asymp. significance	0.027

Table C31 Non-parametric Kruskal Wallis test.

	SD±M
BL21 (DE3)	3.20E9 ± 1.000E8
BL21(DE3)/10	2.70E9 ± 1.000E8
BL21(DE3)/100	1.83E9 ± 5.774E7

Table C32 Descriptive statistics.

References

Abd El-Baky, R. M, Ibrahim, R. A, Mohamed, D. S., Ahmed, E. F., Hashem, Z. S. (2020). Prevalence of virulence genes and their association with antimicrobial resistance among pathogenic *E. coli* isolated from Egyptian patients with different clinical infections. *Infect Drug Resist, 13*, 1221-1236.

Alonso, G., Vilchez, G., & Rodriguez Lemoine, V. (2000). How bacteria protect themselves against channel-forming colicins. *Int Microbiol*, *3*(2), 81-88.

Amos, W. B., & White, J. G. (2003). How the confocal laser scanning microscope entered biological research. *Biology of the Cell*, *95*(6), 335-342.

Anaraki, T., M., Bermel, W., Dutta Majumdar, R., Soong, R., Simpson, M., Monnette, M., & Simpson, A. J. (2019). 1D spikelet projections from heteronuclear 2D NMR data-permitting 1D chemometrics while preserving 2D dispersion. *Metabolites*, *9*(1), 16.

Anderluh, G., Hong, Q., Boetzel, R., MacDonald, C., Moore, G. R., Virden, R., & Lakey, J. H. (2003). Concerted folding and binding of a flexible colicin domain to its periplasmic receptor TolA. *J Biol Chem*, *278*(24), 21860-21868.

Atanaskovic, I., & Kleanthous, C. (2019). Tools and Approaches for dissecting protein bacteriocin import in Gram-negative bacteria. *Frontiers in Microbiology*, *10*(646).

Azatian, S. B., Kaur, N., & Latham, M. P. (2019). Increasing the buffering capacity of minimal media leads to higher protein yield. *Journal of Biomolecular NMR, 73*(1-2), 11–17.

Baindara, P., Korpole, S. & Grover, V. (2018). Bacteriocins: Perspective for the development of novel anticancer drugs. *Appl Microbiol Biotechnol* 102, 10393-10408.

Baneyx, F., & Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol*, *22*(11), 1399-1408.

Banci L, Barbieri L, Bertini I, Cantini F, Luchinat E (2011) In-cell NMR in *E. coli* to monitor maturation steps of hSOD1. *PLoS ONE*. 6(8): e23561.

Banci, L., Barbieri, L., Calderone, V., Cantini, F., Cerofolini, L., Ciofi-Baffoni, S., Felli, I., Fragai, M., Lelli, M., Luchinat, C., Luchinat, E., Parigi, G., Piccioli, M., Pierattelli, R., Ravera, E., Rosato, A., Tenori, L., & Turano, P. (2020). Biomolecular NMR at 1.2 GHz. *arXiv: Chemical Physics*.

Metin Balci. (2005). Basic 1H- and 13C-NMR Spectroscopy. *Elsevier Science*. 25-85.

Barbieri, L., Luchinat, E. & Banci, L. (2015). Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci Rep* **5,** 14456.

Barbieri, L., Luchinat, E. (2021). Monitoring protein-ligand interactions in human cells by real-time quantitative in-cell NMR using a high cell density bioreactor. *J Vis Exp, 169,* e62323.

Bayramoglu, B., Toubiana, D., van Vliet, S., Inglis, R. F., Shnerb, N., & Gillor, O. (2017). Bet hedging in bacteriocin producing *Escherichia coli* populations: the single cell perspective. *Scientific Reports*, *7*, 42068.

Behrens, H. M., Six, A., Walker, D., & Kleanthous, C. (2017). The therapeutic potential of bacteriocins as protein antibiotics. *Emerging Topics in Life Sciences*, *1*(1), 65-74.

Bekei, B., Rose, H. M., Herzig, M., Dose, A., Schwarzer, D., Selenko, P. (2012). In-cell NMR in mammalian cells: part 1. *Methods Mol Biol,* 895:43-54. doi:10.1007/978-1-61779-927-3_4

Benedetti, H., Lloubes, R., Lazdunski, C., & Letellier, L. (1992). Colicin A unfolds during its translocation in *Escherichia coli* cells and spans the whole cell envelope when its pore has formed. *EMBO J, 11*(2), 441-447.

Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition. New York: W H Freeman; (2002). Section 9.3, Restriction enzymes: Performing highly specific DNA-cleavage reactions.

Bernadac, A., Gavioli, M., Lazzaroni, J.-C., Raina, S., & Lloubès, R. (1998). *Escherichia coli* tol-pal mutants form outer membrane vesicles. *Journal of bacteriology*, *180*(18), 4872-4878.

Berrow, N. S., Bussow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G. E., . . . Busso, D. (2006). Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallographica Section D*, *62*(10), 1218-1226.

Bharti, V., Mehta, A., Singh, S., Jain, N., Ahirwal, L., & Mehta, S. (2015). Bacteriocin: A novel approach for preservation of food. *International Journal of Pharmacy and Pharmaceutical Sciences*, *7*(9), 20-29.

Bloois, E.V. (2012). Biotechnological applications of periplasmic expression in *E. coli*.

Bonsor, D. A., Hecht, O., Vankemmelbeke, M., Sharma, A., Krachler, A. M., Housden, N. G., . . . Kleanthous, C. (2009). Allosteric β-propeller signalling in TolB and its manipulation by translocating colicins. *The EMBO Journal*, *28*(18), 2846-2857.

Bonsor, D. A., Meenan, N. A., & Kleanthous, C. (2008). Colicins exploit native disorder to gain cell entry: A hitchhiker's guide to translocation. *Biochem Soc Trans*, *36*(Pt 6), 1409-1413.

Boomershine, W. P., Raj, M. L., Gopalan, V., & Foster, M. P. (2003). Preparation of uniformly labeled NMR samples in *Escherichia coli* under the tight control of the araBAD promoter:

Expression of an archaeal homolog of the RNase P Rpp29 protein. *Protein Expr Purif*, *28*(2), 246-251.

Bouveret, E., Journet, L., Walburger, A., Cascales, E., Benedetti, H., & Lloubes, R. (2002). Analysis of the *Escherichia coli* Tol-Pal and TonB systems by periplasmic production of Tol, TonB, colicin, or phage capsid soluble domains. *Biochimie*, *84*(5-6), 413-421.

Bouveret, E., Rigal, A., Lazdunski, C., & Benedetti, H. (1998). Distinct regions of the colicin A translocation domain are involved in the interaction with TolA and TolB proteins upon import into *Escherichia coli*. *Mol Microbiol*, *27*(1), 143-157.

Bouvier, G., Simenel, C., Jang, J., Kalia, N. P., Choi, I., Nilges, M., Pethe, K., & Izadi-Pruneyre, N. (2019). Target engagement and binding mode of an antituberculosis drug to its bacterial target deciphered in whole living cells by NMR. *Biochemistry, 58*(6), 526–533.

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

Breindel, L., Burz, D. S., & Shekhtman, A. (2019). Interaction proteomics by using in-cell NMR spectroscopy. *Journal of Proteomics, 191*, 202-211.

Brown, C. L., Smith, K., Wall, D. M., & Walker, D. (2015). Activity of species-specific antibiotics against Crohn's disease–associated adherent-invasive *Escherichia coli*. *Inflammatory Bowel Diseases*, *21*(10), 2372-2382.

Bryant, J. E., Lecomte, J. T. J., Lee, A. L., Young, G. B., & Pielak, G. J. (2005). Protein dynamics in living cells. *Biochemistry*, *44*(26), 9275-9279.

Budič, M., Rijavec, M., Petkovšek, Z., & Zgur-Bertok, D. (2011). *Escherichia coli* bacteriocins: antimicrobial efficacy and prevalence among isolates from patients with bacteraemia. *PLoS One*, *6*(12), e28769-e28769.

Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. (2015).A roadmap for interpreting (13)C metabolite labeling patterns from cells. C*urr Opin Biotechnol.*

Buisson, A., Douadi, C., Ouchchane, L., Goutte, M., Hugot, J. P., Dubois, A., Minet-Quinard, R., Bouvier, D., Bommelaer, G., Vazeille, E., & Barnich, N. (2019). Macrophage's inability to mediate adherent-invasive *E. coli* replication is linked to autophagy in Crohn's disease patients. *Cells*, *8*(11), 1394.

Burz, A. S. a. D. S. (2012). *Protein NMR techniques*. New york.

Burz, D. S., & Shekhtman, A. (2009). Inside the living cell. *Nature*, *458*(7234), 37–38.

Burz, D. S., Breindel, L., & Shekhtman, A. (2019). Improved sensitivity and resolution of in-cell NMR spectra. *Methods in Enzymology*, *621*, 305–328.

Burz, D. S., Dutta, K., Cowburn, D., & Shekhtman, A. (2006). Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat Methods*, *3*(2), 91-93.

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

Cao, Z., & Klebba, P. E. (2002). Mechanisms of colicin binding and transport through outer membrane porins. *Biochimie*, *84*(5), 399-412.

Carr, S., Penfold, C. N., Bamford, V., James, R., & Hemmings, A. M. (2000). The structure of TolB, an essential component of the tol-dependent translocation system, and its protein-protein interaction with the translocation domain of colicin E9. *Structure*, *8*(1), 57-66.

Carro, L. (2018). Protein–protein interactions in bacteria: a promising and challenging avenue towards the discovery of new antibiotics. *Beilstein Journal of Organic Chemistry*, *14*, 2881-2896.

Cascales, E., Buchanan, S. K., Duché, D., Kleanthous, C., Lloubes, R., Postle, K., . . . Cavard, D. (2007). Colicin biology. *Microbiology and Molecular Biology Reviews*, *71*(1), 158-229.

Chang, J.-W., Sato, Y., Ogawa, T., Arakawa, T., Fukai, S., Fushinobu, S., & Masaki, H. (2018). Crystal structure of the central and the C-terminal RNase domains of colicin D implicated its translocation pathway through inner membrane of target cell. *The Journal of Biochemistry*, *164*(5), 329-339.

Chen, Y., & Periasamy, A. (2007). Localization of protein-protein interactions in live cells using confocal and spectral imaging FRET microscopy. *Indian J Exp Biol*, *45*(1), 48-57.

Chou J. (2008) Introduction to protein NMR spectroscopy. BCMP 201, Spring.

Collins, E. S., S. B. Whittaker, et al. (2002). Structural dynamics of the membrane translocation domain of colicin E9 and its interaction with TolB. *J Mol Biol*, *318*(3), 787-804.

Cramer, W. A., Sharma, O., & Zakharov, S. D. (2018). On mechanisms of colicin import: the outer membrane quandary. *Biochemical Journal*, *475*(23), 3903-3915.

Crozel, V., Lazdunski, C., Lloubes, R., & Cavard, D. (1984). A colicin A fragment containing the receptor binding domain can be directed to the periplasmic space in *E. coli* through gene fusion. *FEBS Lett*, *172*(2), 183-188.

Cursino, L., Šmarda, J., Chartone-Souza, E., & Nascimento, A. M. A. (2002). Recent updated aspects of colicins of *Enterobacteriaceae*. *Brazilian Journal of Microbiology*, *33*, 185-195.

Dadgostar P. (2019). Antimicrobial resistance: implications and costs. *Infection and drug resistance*, *12*, 3903–3910.

Steven Damo, Michael D Feldkamp, Benjamin Chagot, Walter J Chazin. NMR studies of the interaction of calmodulin with IQ motif peptides. *Methods in Molecular Biology*, Humana Press/Springer Imprint, 2013, Calcium-binding proteins and rage 963, pp.173-86.

de Zamaroczy, M., & Buckingham, R. H. (2002). Importation of nuclease colicins into *E. coli* cells: endoproteolytic cleavage and its prevention by the immunity protein. [Review]. *Biochimie*, *84*(5-6), 423-432.

Di Masi,D. R., J. C. White, et al. (1973). Transportofvitamin B12 in *Escherichia coli*: Common receptor sites for vitamin B12 and the E colicins on the outer membraneofthecellenvelope. *JBacteriol*, *115*(2), 506-513.

Didenko, T., Liu, J. J., Horst, R., Stevens, R. C., & Wüthrich, K. (2013). Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Current Opinion in Structural Biology*, *23*(5), 740–747.

Dimov, S. G., Ivanova, P. M., Harizanova, N. T., & Ivanova, I. V. (2005). Bioactive peptides used by bacteria in the concur-rence for the ecological niche: General classification and mode of action (Overview). *Biotechnology & Biotechnological Equipment*, *19*(2), 3-22.

Dubuisson, J. F., Vianney, A., & Lazzaroni, J. C. (2002). Mutational analysis of the TolA C-terminal domain of *Escherichia coli* and genetic evidence for an interaction between TolA and TolB. *Journal of Bacteriology*, *184*(16), 4620–4625.

Duché, D., Issouf, M., & Lloubès, R. (2009). Immunity protein protects colicin E2 from OmpT protease. *Journal of Biochemistry*, *145*(1), 95-101.

Dunkel, S., Pulagam, L. P., Steinhoff, H. J., & Klare, J. P. (2015). In vivo EPR on spin labeled colicin A reveals an oligomeric assembly of the pore-forming domain in E. coli membranes. *Physical chemistry chemical physics: PCCP*, *17*(7), 4875–4878.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochemistry*, *28*;41(21), 6573-82.

Durret, R., and Levine, S. (1997). Alleopathy in spatially distributed poulations. J. *Theor. Biol.* 185, 165-171.

Dyson H. J. &. Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev*, *104*, 3607−3622.

Egan, A. J. F. (2018). Bacterial outer membrane constriction. *Mol Microbiol*, *107*(6), 676-687.

Emwas, A. H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M., & Wishart, D. S. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, *9*(7), 123.

Farrance, O. E., Hann, E., Kaminska, R., Housden, N. G., Derrington, S. R., Kleanthous, C., . . . Brockwell, D. J. (2013). A Force-activated trip switch triggers rapid dissociation of a colicin from its immunity protein. *PLOS Biology*, *11*(2), e1001489.

Felli, Isabella & Gonnelli, Leonardo & Pierattelli, Roberta. (2014). In-cell C-13 NMR spectroscopy for the study of intrinsically disordered proteins. *Nature protocols*. 9. 2005-2016.

Ferella, L., Luchinat, C., Ravera, E. *et al.* (2013) SedNMR: a web tool for optimizing sedimentation of macromolecular solutes for SSNMR. *J Biomol NMR* **57,** 319–326.

Fouquet, C., Gilles, J. F., Heck, N., Dos Santos, M., Schwartzmann, R., Cannaya, V., Morel, M. P., Davidson, R. S., Trembleau, A., & Bolte, S. (2015). Improving axial resolution in confocal microscopy with new high refractive index mounting media. *PloS One*, *10*(3), e0121096.

Franks, W. T., Zhou, D. H., Wylie, B. J., Money, B. G., Graesser, D. T., Frericks, H. L., Sahota, G., & Rienstra C. M. (2005). TITLE. *Journal of the American Chemical Society*, *127*(35), 12291-12305.

Freedberg, D. I., & Selenko, P. (2014). Live cell NMR. *Annu Rev Biophys*, *43*, 171-192.

Fridd, S. L., Gokce, I., & Lakey, J. H. (2002). High level expression of His-tagged colicin pore-forming domains and reflections on the sites for pore formation in the inner membrane. *Biochimie*, *84*(5-6), 477-483.

Frigault, M. M., Lacoste, J., Swift, J. L., & Brown, C. M. (2009). Live-cell microscopy – tips and tools. *Journal of Cell Science*, *122*(6), 753-767.

Furia, L., Pelicci, P., & Faretta, M. (2014). Confocal microscopy for high-resolution and high-content analysis of the cell cycle. *Current Protocols in Cytometry*, *70*(1), 7.42.41-47.42.14.

Gaillard-Gendron, S., Vignon, D., Cottenceau, G., Graber, M., Zorn, N., van Dorsselaer, A., & Pons, A. M. (2000). Isolation, purification and partial amino acid sequence of a highly hydrophobic new microcin named microcin L produced by *Escherichia coli*. *FEMS Microbiol Lett*, *193*(1), 95-98.

Garinot-Schneider, C., Penfold, C. N., Moore, G. R., Kleanthous, C., & James, R. (1997). Identification of residues in the putative TolA box which are essential for the toxicity of the endonuclease toxin colicin E9. *Microbiology*, *143*(9), 2931-2938.

Ghazaryan, L., Soares, M. I., & Gillor, O. (2014). Auto-regulation of DNA degrading bacteriocins: molecular and ecological aspects. *Antonie Van Leeuwenhoek*, *105*(5), 823-834.

Ghazaryan, L., Tonoyan, L., Ashhab, A. A., Soares, M. I., & Gillor, O. (2014). The role of stress in colicin regulation. *Arch Microbiol*, *196*(11), 753-764.

Ghazaryan, L., Giladi, I., & Gillor, O. (2019). The effects of colicin production rates on allelopathic interactions in *Escherichia coli* populations. *Microorganisms*, *7*(11), 564.

Ghodhbane, H., Elaidi, S., Sabatier, J. M., Achour, S., Benhmida, J., & Regaya, I. (2015). Bacteriocins active against multi-resistant gram negative bacteria implicated in nosocomial infections. *Infect Disord Drug Targets*, *15*(1), 2-12.

Gibbs, E. B., Cook, E. C., & Showalter, S. A. (2017). Application of NMR to studies of intrinsically disordered proteins. *Archives of Biochemistry and Biophysics*, *628*, 57-70.

Gillor, O., Vriezen, J. A. C., & Riley, M. A. (2008). The role of SOS boxes in enteric bacteriocin regulation. *Microbiology*, *154*(Pt 6), 1783-1792.

Giraudeau, P., Frydman, L. (2014). Ultrafast 2D NMR: An emerging tool in analytical spectroscopy. *Annual Review of Analytical Chemistry*, *7*:129-161.

Grabski, A., Mehler, M. & Drott, D. (2005). The overnight express autoinduction system: High-density cell growth and protein expression while you sleep. *Nat Methods*, *2*, 233–235.

Grant, B. M., Marshall, C. B., Ikura. M. (2019). Expression and purification of calmodulin for NMR and other biophysical applications. *Methods In Molecular Biology*.

Gronenborn, A.M. and Clore, G.M., 1993. Structural studies of immunoglobulin-binding domains of streptococcal protein G. *Immunomethods*, *2*(1), 3-8.

Grudziąż, K., Zawadzka-Kazimierczuk, A., & Koźmiński, W. (2018). High-dimensional NMR methods for intrinsically disordered proteins studies. *Methods*, *148*, 81-87.

Guihard, G., Boulanger, P., Benedetti, H., Lloubes, R., Besnard, M., & Letellier, L. (1994). Colicin A and the Tol proteins involved in its translocation are preferentially located in the contact sites between the inner and outer membranes of *Escherichia coli* cells. *J Biol Chem*, *269*(8), 5874-5880.

Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, *177*(14), 4121–4130.

Hahn-Löbmann, S., Stephan, A., Schulz, S., Schneider, T., Shaverskyi, A., Tusé, D., . . . Gleba, Y. (2019). Colicins and Salmocins – new Classes of plant-made non-antibiotic food antibacterials. *Frontiers in Plant Science*, *10*(437).

Hands, S. L., Holland, L. E., Vankemmelbeke, M., Fraser, L., Macdonald, C. J., Moore, G. R., . . . Penfold, C. N. (2005). Interactions of TolB with the translocation domain of colicin E9 require an extended TolB box. *Journal of Bacteriology*, *187*(19), 6733-6741.

Hecht, O., Macdonald, C., & Moore, G. R. (2012). Intrinsically disordered proteins: lessons from colicins. *Biochem Soc Trans*, *40*(6), 1534-1538.

Hecht, O., Ridley, H., Lakey, J. H., & Moore, G. R. (2009). A common interaction for the entry of colicin N and filamentous phage into *Escherichia coli*. *J Mol Biol*, *388*(4), 880-893.

Hecht, O., Zhang, Y., Li, C., Penfold, C. N., James, R., & Moore, G. R. (2010). Characterisation of the interaction of colicin A with its co-receptor TolA. [Research Support, Non-U.S. Gov't]. *FEBS Lett*, *584*(11), 2249-2252.

Hol, F. J., Voges, M. J., Dekker, C., & Keymer, J. E. (2014). Nutrient-responsive regulation determines biodiversity in a colicin-mediated bacterial community. *BMC biology*, *12*, 68.

Hoppe, A. D., Seveau, S., & Swanson, J. A. (2009). Live cell fluorescence microscopy to study microbial pathogenesis. *Cell Microbiol*, *11*(4), 540-550.

Housden, N. G., Hopper, J. T. S., Lukoyanova, N., Rodriguez-Larrea, D., Wojdyla, J. A., Klein, A., . . . Kleanthous, C. (2013). Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science*, *340*(6140), 1570-1574.

Housden, N. G., Loftus, S. R., Moore, G. R., James, R., & Kleanthous, C. (2005). Cell entry mechanism of enzymatic bacterial colicins: porin recruitment and the thermodynamics of receptor binding. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 13849-13854.

Housden, N. G., Rassam, P., Lee, S., Samsudin, F., Kaminska, R., Sharp, C., . . . Kleanthous, C. (2018). Directional porin binding of intrinsically disordered protein sequences promotes colicin epitope display in the bacterial periplasm. *Biochemistry*, 57(29), 4374-4381.

Housden, N. G., Wojdyla, J. A., Korczynska, J., Grishkovskaya, I., Kirkpatrick, N., Brzozowski, A. M., & Kleanthous, C. (2010). Directed epitope delivery across the *Escherichia coli* outer membrane through the porin OmpF. *Proc Natl Acad Sci U S A*, *107*(50), 21412-21417.

Huang, K. C., Mukhopadhyay, R., Wen, B., Gitai, Z., & Wingreen, N. S. (2008). Cell shape and cell-wall organization in Gram-negative bacteria. *Proceedings of the National Academy of Sciences*, *105*(49), 19282-19287.

Icha, J., Weber, M., Waters, J.C., Norden, C. (2017). Phototoxicity in live fluorescence microscopy, and how to avoid it. Bioessays, *39*(8).

Ikeya, T., Güntert, P., & Ito, Y. (2019). Protein structure determination in living cells. *International Journal of Molecular Sciences*, *20*(10), 2442.

Ikeya, T., Hanashima, T., Hosoya, S. et al. (2016). Improved in-cell structure determination of proteins at near-physiological concentration. *Sci Rep*, *6*, 38312.

Inglis, R. F., Gardner, A., Cornelis, P., Buckling, A. (2009). Spite and virulence in the bacterium *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA, 106*:5703–5707.

Ito, Y., & Selenko, P. (2010). Cellular structural biology. *Current Opinion in Structural Biology*, *20*(5), 640-648.

James, R., Penfold, C. N., Moore, G. R., & Kleanthous, C. (2002). Killing of *E. coli* cells by E group nuclease colicins. *Biochimie*, *84*(5), 381-389.

Jansen, K. B., Inns, P. G., Housden, N. G., Hopper, J., Kaminska, R., Lee, S., Robinson, C. V., Bayley, H., & Kleanthous, C. (2020). Bifurcated binding of the OmpF receptor underpins import of the bacteriocin colicin N into *Escherichia coli*. *The Journal of biological chemistry*, *295*(27), 9147–9156.

Jia, B., & Jeon, C. O. (2016). High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. *Open Biol*, 6(8).

Jin, X., Kightlinger, W., Kwon, Y.-C., & Hong, S. H. (2018). Rapid production and characterization of antimicrobial colicins using *Escherichia coli*-based cell-free protein synthesis. *Synthetic Biology*, *3*(1).

Johansson, H., Jensen, M. R., Gesmar, H., Meier, S., Vinther, J. M., Keeler, C., Hodsdon, M. E. & Led, J. J. (2014). Specific and nonspecific interactions in ultraweak protein–protein associations revealed by solvent paramagnetic relaxation enhancements. *Journal of the American Chemical Society*, *136*(29), 10277-10286.

Johnson, C. L., Ridley, H., Pengelly, R. J., Salleh, M. Z., & Lakey, J. H. (2013). The unstructured domain of colicin N kills *Escherichia coli*. *Mol Microbiol*, *89*(1), 84-95.

Jonkman, J., & Brown, C. M. (2015). Any way you slice it-a comparison of confocal microscopy techniques. *J Biomol Tech*, *26*(2), 54-65.

Journet L, Bouveret E, Rigal A, Lloubès R, Lazdunski C, Bénédetti H. (1999). Import of colicins across the outer membrane of *Escherichia coli* involves multiple protein interactions in the periplasm. *Mol Microbiol.* 2001;42:331–344

Kamenšek, S., & Žgur-Bertok, D. (2013). Global transcriptional responses to the bacteriocin colicin M in *Escherichia coli*. *BMC Microbiology*, *13*(1), 42.

Kang, C. (2019). Applications of in-cell NMR in structural biology and drug discovery. *Int J Mol Sci*, *20*(1), 139.

Karplus, S., Snyder, G.H., and Sykes, B.D. (1973). Nuclear magnetic resonance study of bovine pancreatic trypsin inhibitor. Tyrosine titrations and backbone NH groups. *Biochemistry 12*, 1323–1329.

Kaur S. & Kaur S. (2015). Bacteriocins as potential anticancer agents. *Frontiers in pharmacology*, *6*, 272.

Kim, Y. C., Tarr, A. W., & Penfold, C. N. (2014). Colicin import into *E. coli* cells: A model system for insights into the import mechanisms of bacteriocins. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1843*(8), 1717-1731.

Kleanthous, C. (2010). Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat Rev Microbiol, 8*(12), 843-848.

Klein, A., Wojdyla, J. A., Joshi, A., Josts, I., McCaughey, L. C., Housden, N. G., . . . Kleanthous, C. (2016). Structural and biophysical analysis of nuclease protein antibiotics. *Biochem J, 473*(18), 2799-2812.

Kumar, T. K. S., Thurman, R., & Jayanthi, S. (2013). In-Cell NMR spectroscopy-*In vivo* monitoring of the structure, dynamics, folding, and interactions of proteins at atomic resolution. *Journal of Analytical & Bioanalytical Techniques, 4*(1), e112–.

Kuhnert P, Boerlin P, Frey J. Target genes for virulence assessment of *Escherichia coli* isolates from water, food and the environment. *FEMS Microbiol Rev*. 2000 Jan;24(1):107-17.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the headofbacteriophageT4. *Nature*, *227*(5259), 680-685.

Lamley, J. M., Iuga, D., Oster, C., Sass, H. J., Rogowski, M., Oss, A., . . . Lewandowski, J. R. (2014). Solid-state NMR of a protein in a precipitated complex with a full-length antibody. *J Am Chem Soc*, *136*(48), 16800-16806.

Latifi, A. M., Khajeh, K., Farnoosh, G., Hassanpour, K., & Khodi, S. (2015). The cytoplasmic and periplasmic expression levels and folding of organophosphorus hydrolase enzyme in *Escherichia coli*. *Jundishapur Journal of Microbiology, 8*(12), e17790-e17790.

Law, C. J., C. N. Penfold, et al. (2003). OmpF enhances the abilityof BtuB to protect susceptible *Escherichia coli* cells from colicin E9 cytotoxicity. *FEBS letters, 545*(2-3), 127-132.

Lazdunski, C. J., Bouveret, E., Rigal, A., Journet, L., Lloubès, R., & Bénédetti, H. (1998). Colicin Import into *Escherichia coli* Cells. *Journal of Bacteriology*, *180*(19), 4993-5002.

Lazzaroni, J.-C., Dubuisson, J.-F., & Vianney, A. (2002a). The Tol proteins of *Escherichia coli* and their involvement in the translocation of group A colicins. *Biochimie*, *84*(5-6), 391-397.

Lee, J. G., Han, D. S., Jo, S. V., Lee, A. R., Park, C. H., Eun, C. S., Lee Y. (2019). Characteristics and pathogenic role of adherent-invasive *Escherichia coli* in inflammatory bowel disease: Potential impact on clinical outcomes. *PloS one*, *14*(4), e0216165.

Lei, J., Sun, L., Huang, S., Zhu, C., Li, P., He, J., Mackey, V., Coy, D. H., & He, Q. (2019). The antimicrobial peptides and their potential clinical applications. *American journal of translational research*, *11*(7), 3919–3931.

Li, C. and M. Liu (2013). Protein dynamics in living cells studied by in-cell NMR spectroscopy. *FEBS Lett* 587(8), 1008-1011.

Li, C., Zhang, Y., Vankemmelbeke, M., Hecht, O., Aleanizy, F. S., Macdonald, C., … Penfold, C. N. (2012). Structural evidence that colicin a protein binds to a novel binding site of TolA protein in *Escherichia coli* periplasm. *The Journal of Biological Chemistry*, *287*(23), 19048–19057.

Li, Z., Kessler, W., van den Heuvel, J., & Rinas, U. (2011). Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems. *Applied Microbiology and Biotechnology*, *91*(4), 1203-1213.

Lippens, G., Cahoreau, E., Millard, P., Charlier, C., Lopez, J., Hanoulle, X., & Portais, J. C. (2018). In-cell NMR: from metabolites to macromolecules. *Analyst*, *143*(3), 620-629.

Lloubes, R., Bernadac, A., Houot, L., & Pommier, S. (2013). Non classical secretion systems. [Research Support, Non-U S Gov't Review]. *Res Microbiol*, *164*(6), 655-663.

Lloubès, R., Cascales, E., Walburger, A., Bouveret, E., Lazdunski, C., Bernadac, A., & Journet, L. (2001). The Tol-Pal proteins of the *Escherichia coli* cell envelope: an energized system required for outer membrane integrity. *Research in Microbiology*, *152*(6), 523-529.

Lodish H, Berk A, Zipursky SL, et al. (2000). *Molecular Cell Biology*. 4th edition. New York: W. H. Freeman; Section 7.1, DNA cloning with plasmid vectors.

Loftus SR, Walker D, Mate MJ, Bonsor DA, James R, Moore GR, Kleanthous C. (2006). Competitive recruitment of the periplasmic translocation portal TolB by a natively disordered domain of colicin E9. *Proc Natl Acad Sci USA*103:12353–12358

Logg, K.., Bodvard, K., Blomberg, A., Käll, M. (2009). Investigations on light-induced stress in fluorescence microscopy using nuclear localization of the transcription factor Msn2p as a reporter, *FEMS Yeast Research*, 9 ()6, 875–884,

Luchinat, E., Gianoncelli, A., Mello, T., Galli, A., & Banci, L. (2015). Combining in-cell NMR and X-ray fluorescence microscopy to reveal the intracellular maturation states of human superoxide dismutase 1. *Chemical communications (Cambridge, England)*, *51*(3), 584–587.

Luchinat, E., & Banci, L. (2018). New structural and functional insights from in-cell NMR. *Emerging Topics in Life Sciences*, ETLS20170136.

Luchinat, E., Barbieri, L., Campbell, T. F., Banci, L. (2020). Real-time quantitative in-cell NMR: Ligand binding and protein oxidation monitored in human cells using multivariate curve resolution. *Analytical Chemistry*, *92*(14), 9997-10006.

Luchinat, E., Barbieri, L., Cremonini, M. et al. (2021). Protein in-cell NMR spectroscopy at 1.2 GHz. *J Biomol NMR* 75, 97–107.

Majeed, H., Gillor, O., Kerr, B., & Riley, M. A. (2011). Competitive interactions in *Escherichia coli* populations: the role of bacteriocins. *The ISME journal*, *5*(1), 71–81.

Majeed, H., Lampert, A., Ghazaryan, L., Gillor O. (2013). The weak shall inherit: bacteriocin-mediated interactions in bacterial populations. *PloS one*, *8*(5), e63837.

Majumder, S., DeMott, C. M., Burz, D. S., & Shekhtman, A. (2014). Using singular value decomposition to characterize protein-protein interactions by in-cell NMR spectroscopy. *Chembiochem : a European journal of chemical biology*, *15*(7), 929–933.

Makino, T., Skretas, G., & Georgiou, G. (2011). Strain engineering for improved expression of recombinant proteins in bacteria. *Microb Cell Fact*, *10*, 32.

Makrides S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews*, *60*(3), 512–538.

Maldonado, A. Y., Burz, D. S., & Shekhtman, A. (2011). In-cell NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *59*(3), 197–212.

Marion, D. (2013). An Introduction to biological NMR spectroscopy. *Molecular & Cellular Proteomics*, *12*(11), 3006-3025.

Marschall, L., Sagmeister, P., & Herwig, C. (2017). Tunable recombinant protein expression in *E. coli*: promoter systems and genetic constraints. *Applied Microbiology and Biotechnology*, *101*(2), 501-512.

Maslowska, K. H., Makiela-Dzbenska, K., & Fijalkowska, I. J. (2019). The SOS system: A complex and tightly regulated response to DNA damage. *Environmental and molecular mutagenesis*, *60*(4), 368–384.

Matsumoto-Nakano, M., & Kuramitsu, H. K. (2006). Role of bacteriocin immunity proteins in the antimicrobial sensitivity of Streptococcus mutans. *Journal of Bacteriology*, *188*(23), 8095–81

Mavridou, D. A. I., Gonzalez, D., Kim, W., West, S. A., & Foster, K. R. (2018). Bacteria use collective behavior to generate diverse combat strategies. *Curr Biol*, *28*(3), 345-355.e344.

Metola, A., Bouchet, A. M., Alonso-Mariño, M., Diercks, T., Mäler, L., Goñi, F. M., & Viguera, A. R. (2017). Purification and characterization of the colicin A immunity protein in detergent micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1859*(11), 2181-2192.

Meyer, P., & Dworkin, J. (2007). Applications of fluorescence microscopy to single bacterial cells. *Res Microbiol*, *158*(3), 187-194.

Micenková, L., Bosák, J., Kucera, J., Hrala, M., Dolejšová, T., Šedo, O., Linke, D., Fišer, R., & Šmajs, D. (2019). Colicin Z, a structurally and functionally novel colicin type that selectively kills enteroinvasive Escherichia coli and Shigella strains. *Scientific reports*, *9*(1), 11127.

Michel, B. (2005). After 30 Years of study, the bacterial SOS response still surprises us. *PLOS Biology*, *3*(7), e255.

Mierendorf, R. C., Morris, B. B., Hammer, B., & Novy, R. E. (1998). Expression and purification of recombinant proteins using the pET system. *Methods in molecular medicine*, *13*, 257–292.

Motáčková V, Nováček J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Žídek L, Šanderová H, Krásný L, Koźmiński W, Sklenář V. Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J Biomol NMR*. 2010;48:169–177.

Mondal, S., Shet, D., Prasanna, C., & S. Atreya, H. (2013). High yield expression of proteins in E. coli for NMR studies. *Advances in Bioscience and Biotechnology*, 4, 751-767

Morales, M., Attai, H., Troy, K., & Bermudes, D. (2015). Accumulation of single-stranded DNA in *Escherichia coli* carrying the colicin plasmid pColE3-CA38. *Plasmid*, *77*, 7-16.

Mosbahi, K., Walker, D., Lea, E., Moore, G. R., James, R., & Kleanthous, C. (2004). Destabilization of the colicin E9 endonuclease domain by interaction with negatively

charged phospholipids: Implications for colicin translocation into bacteria. *Journal of Biological Chemistry*, *279*(21), 22145-22151.

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Šanderová H, Krásný L, Koźmiński W, Sklenář V. 5D ^{13}C detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR*. 2011;50:1–11.

Nicolas L. Fawzi, Michaeleen Doucleff, Jeong-Yong Suh, G. Marius Clore. (2010). Mechanistic details of a protein–protein association pathway revealed by paramagnetic relaxation enhancement titration measurements. *Proceedings of the National Academy of Sciences*, *107*(4), 1379-1384.

Nikaido, H. (1994). Porins and specific diffusion channels in bacterial outer membranes. *J Biol Chem*, *269*(6), 3905-3908.

Nikolic, M., Conrad, C., Zhang, J., & Scarcelli, G. (2018). Noninvasive imaging: Brillouin confocal microscopy. *Adv Exp Med Biol*, *1092*, 351-364.

Nwaneshiudu, A., Kuschal, C., Sakamoto, F. H., Anderson, R. R., Schwarzenberger, K., & Young, R. C. (2012). Introduction to confocal microscopy. *J Invest Dermatol*, *132*(12), e3.

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in endocrinology*, *9*, 402.

Oikawa, H., Suzuki, Y., Saito, M., Kamagata, K., Arai, M., & Takahashi, S. (2013). Microsecond dynamics of an unfolded protein by a line confocal tracking of single molecule fluorescence. *Scientific Reports*, *3*, 2151.

Oktaviani, N. (2014). NMR studies of folded and unfolded proteins: method developments and biological insight.

Paddock, S. W., & Eliceiri, K. W. (2014). Laser scanning confocal microscopy: History, applications, and related optical sectioning techniques. *Methods Mol Biol*, *1075*, 9-47.

Page, R., Peti, W., Wilson, I. A., Stevens, R. C., & Wuthrich, K. (2005). NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. *Proc Natl Acad Sci U S A*, *102*(6), 1901-1905.

Pamp, S. J., Sternberg, C., & Tolker-Nielsen, T. (2009). Insight into the microbial multicellular lifestyle via flow-cell technology and confocal microscopy. *Cytometry Part A*, 75A(2), 90-103.

Papadakos, G., Wojdyla, J., & Kleanthous, C. (2012). Nuclease colicins and their immunity proteins. *Quarterly Reviews of Biophysics*, *45*(1), 57-103.

Pastore, A., & Temussi, P. A. (2017). The Emperor's new clothes: Myths and truths of in-cell NMR. *Arch Biochem Biophys*, *628*, 114-122.

Patil, A., Kinoshita, K., & Nakamura, H. (2010). Hub promiscuity in protein-protein interaction networks. *Int J Mol Sci*, *11*(4), 1930-1943.

Penfold, C. N., B. Healy, et al. (2004). Flexibility in the receptor-binding domain of the enzymatic colicin E9 is required for toxicity against *Escherichia coli* cells. *Journal of Bacteriology*, *186*(14), 4520-4527.

Penfold, C. N., Li, C., Zhang, Y., Vankemmelbeke, M., & James, R. (2012). Colicin A binds to a novel binding site of TolA in the *Escherichia coli* periplasm. *Biochem Soc Trans*, *40*(6), 1469-1474.

Pielak, G. J., Li, C., Miklos, A. C., Schlesinger, A. P., Slade, K. M., Wang, G.-F., & Zigoneanu, I. G. (2009). Protein nuclear magnetic resonance under physiological conditions. *Biochemistry*, *48*(2), 226-234.

Polykretis, P., Cencetti, F., Donati, C., Luchinat, E., & Banci, L. (2019). Cadmium effects on superoxide dismutase 1 in human cells revealed by NMR. *Redox biology*, *21*, 101102.

Poulsen, F. M. (2002). A brief introduction to NMR spectroscopy of proteins.

Pulagam, L. P., & Steinhoff, H. J. (2013). Acidic pH-induced membrane insertion of colicin A into *E. coli* natural lipids probed by site-directed spin labeling. *J Mol Biol*, *425*(10), 1782-1794.

Raarup, M. K., & Nyengaard, J. R. (2011). Quantitative confocal laser scanning microscopy. *Image Anal Stereol* 2006;25:111-120

Ranjan, A., Scholz, J., Semmler, T., Wieler, L. H., Ewers, C., Muller, S., . . . Guenther, S. (2018). ESBL-plasmid carriage in *E. coli* enhances in vitro bacterial competition fitness and serum resistance in some strains of pandemic sequence types without overall fitness cost. *Gut Pathog*, *10*, 24.

Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., . . . Ravel, J. (2008). The Pangenome Structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology*, *190*(20), 6881-6893.

Rassam, P., Long, K.R., Kaminska, R. et al. (2018). Intermembrane crosstalk drives inner-membrane protein organization in *Escherichia coli*. *Nat Commun*, *9*, 1082.

Rehm, T., Huber, R., & Holak, T. A. (2002). Application of NMR in structural proteomics: Screening for proteins amenable to structural analysis. *Structure*, *10*(12), 1613-1618.

Riley, M. A. (1993). Molecular mechanisms of colicin evolution. *Molecular Biology and Evolution*, *10*(6), 1380-1395.

Robinson, K. E., Reardon, P. N., & Spicer, L. D. (2012). In-cell NMR spectroscopy in *Escherichia coli*. [Review]. *Methods Mol Biol*, *831*, 261-277

Rosano Germán L., Ceccarelli Eduardo A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5:172.

Rosenberg, M., Azevedo, N. F., & Ivask, A. (2019). Propidium iodide staining underestimates viability of adherent bacterial cells. *Scientific Reports*, *9*(1), 6483.

Rosochacki, S. J., Matejczyk, M. (2002). Green fluorescent protein as a molecular marker in microbiology. *Acta Microbiol Pol*, *51*(3), 205-216.

Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., . . . Ito, Y. (2009). Protein structure determination in living cells by in-cell NMR spectroscopy.. *Nature, 458*(7234), 102-105.

Sambrook, J., Fritsh, E. F. et al. (2001). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.

Sanderson, M. J., Smith, I., Parker, I., & Bootman, M. D. (2014). Fluorescence microscopy. *Cold Spring Harb Protoc*, *2014*(10), pdb.top071795.

Sandomenico, A., Sivaccumar, J. P., Ruvo, M. (2020). Evolution of *Escherichia coli* expression system in producing antibody recombinant fragments. *Int J Mol Sci*, *31*;21(17), 6324.

Santos, C. A., Janissen, R., Toledo, M. A., Beloti, L. L., Azzoni, A. R., Cotta, M. A., & Souza, A. P. (2015). Characterization of the TolB-Pal trans-envelope complex from Xylella fastidiosa reveals a dynamic and coordinated protein expression profile during the biofilm development process. *Biochim Biophys Acta, 1854*(10 Pt A), 1372-1381.

Abba J. Kastin. (2006). Colicins: Bacterial/antibiotic peptides, Handbook of biologically active peptides. *Academic Press.* 115-123,

Sharp C, Bray J, Housden NG, Maiden MCJ, Kleanthous C, (2017). Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using hidden Markov models. *PLoS Comput Biol* 13(7): e100565

Schägger, H. (2006). Tricine–SDS-PAGE. *Nature Protocols*, *1*, 16.

Schlapschy, M., et al. (2006). A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in *Escherichia coli*. *Protein Eng Des Sel* 19(8), 385-390.

Schleif, R. (2010). AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. [Review]. *FEMS Microbiol Rev*, *34*(5), 779-796.

Schneider, J. P., & Basler, M. (2016). Shedding light on biology of bacterial cells. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *371*(1707), 20150499.

Schulz, S., Stephan, A., Hahn, S., Bortesi, L., Jarczowski, F., Bettmann, U., . . . Gleba, Y. (2015). Broad and efficient control of major foodborne pathogenic strains of *Escherichia coli* by mixtures of plant-produced colicins. *Proc Natl Acad Sci U S A*, *112*(40), E5454-5460.

Selenko, P. and G. Wagner (2007). Looking into live cells with in-cell NMR spectroscopy. *Journal of Structural Biology*, *158*(2), 244-253.

Selenko, P., Serber, Z., Gadea, B., Ruderman, J. and Wagner, G., 2006. Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. *Proceedings of the National Academy of Sciences*, *103*(32), pp.11904-11909.

Serber, Z., Straub, W., Corsini, L., Nomura, A. M., Shimba, N., Craik, C. S., Ortiz de Montellano, P., & Dötsch, V. (2004). Methyl groups as probes for proteins and complexes in in-cell NMR experiments. *Journal of the American Chemical Society*, *126*(22), 7119–7125.

Serber, Z., & Dötsch, V. (2001). In-cell NMR spectroscopy. *Biochemistry*, *40*(48), 14317-14323.

Serber, Z., Corsini, L., Durst, F., & Dotsch, V. (2005). In-cell NMR spectroscopy. *Methods Enzymol*, *394*, 17-41.

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. *Journal of bacteriology*, *189*(23), 8746–8749.

Sharaf, N. G., Barnes, C. O., Charlton, L. M., Young, G. B., & Pielak, G. J. (2010). A bioreactor for in-cell protein NMR. *Journal of Magnetic Resonance*, *202*(2), 140-146.

Siegal, G., Selenko, P. Cells, drugs and NMR. (2019). *Journal of Magnetic Resonance*, 306

Simons, A., Alhanout, K., & Duval, R. E. (2020). Bacteriocins, Antimicrobial Peptides from Bacterial Origin: Overview of Their Biology and Their Impact against Multidrug-Resistant Bacteria. *Microorganisms*, *8*(5), 639.

Singh, A., & Gopinathan, K. P. (1998). Confocal microscopy: A powerful technique for biological research. *Current Science*, *74*(10), 841-851.

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. and Li, Q. (2009). Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science*, *18*:936-948.

Smarda J. (1992). Resistance and tolerance of bacteria to E colicins. In: James R., Lazdunski C., Pattus F. Bacteriocins, Microcins and Lantibiotics. *NATO ASI Series (Series H: Cell Biology)*, 65.

Smith, A. E., Zhang, Z., Pielak, G. J., & Li, C. (2015). NMR studies of protein folding and binding in cells and cell-like environments. *Current opinion in structural biology*, *30*, 7–16.

Snijder, H. J., Timmins, P. A., Kalk, K. H., & Dijkstra, B. W. (2003). Detergent organisation in crystals of monomeric outer membrane phospholipase A. *Journal of Structural Biology*, *141*(2), 122-131.

Sommer, B., Friehs, K., & Flaschel, E. (2010). Efficient production of extracellular proteins with *Escherichia coli* by means of optimized coexpression of bacteriocin release proteins. *J Biotechnol*, *145*(4), 350-358.

Soelaiman, S & Jakes, Karen & Wu, N & Li, Chunmin & Shoham, Menachem. (2001). Crystal structure of colicin E3: implications for cell entry and ribosome inactivation. *Molecular cell*. 8. 1053-62.

Sørensen, H.P., Mortensen, K.K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb Cell Fact*, *4*, 1.

Stiefel, P., Schmidt-Emrich, S., Maniura-Weber, K., & Ren, Q. (2015). Critical aspects of using bacterial cell viability assays with the fluorophores SYTO9 and propidium iodide. *BMC Microbiology*, *15*, 36-36.

Sugiki, T., Kobayashi, N., & Fujiwara, T. (2017). Modern technologies of solution nuclear magnetic resonance spectroscopy for three-dimensional structure determination of proteins open avenues for life scientists. *Computational and structural biotechnology journal*, *15*, 328–339.

Sugiki, T., Yamaguchi, Y., Fujiwara, T. et al. (2020) In-cell NMR as a sensitive tool to monitor physiological condition of *Escherichia coli*. *Sci Rep*, *10*, 2466.

Szczepaniak J., Press C., Kleanthous C. (2020a). The multifarious roles of Tol-Pal in Gram-negative bacteria, *FEMS Microbiology Reviews*, *44*(4), 490–506.

Szczepaniak, J., Holmes, P., Rajasekar, K. et al. (2020b). The lipoprotein Pal stabilises the bacterial outer membrane during constriction by a mobilisation-and-capture mechanism. *Nat Commun*, *11*, 1305.

Takeuchi, K., & Frank, J. F. (2001). Confocal microscopy and microbial viability detection for food research. *J Food Prot*, *64*(12), 2088-2102.

Takis, P. G., Jiménez, B., Sands, C. J., Chekmeneva E. & Lewis. M. R. (2020). SMolESY: an efficient and quantitative alternative to on-instrument macromolecular ^{1}H-NMR signal suppression.
*Chem. Sci.*11, 6000-6011

Teleha, M. A., Miller, A. C., & Larsen, R. A. (2013). Overexpression of the *Escherichia coli* TolQ protein leads to a null-FtsN-like division phenotype. *MicrobiologyOpen*, *2*(4), 618-632.

Telke A. A., Ovchinnikov K. V., Vuoristo K. S., Mathiesen G., Thorstensen T., Diep D. B. (2019). Over 2000-fold increased production of the leaderless bacteriocin garvicin KS by increasing gene dose and optimization of culture conditions. *Frontiers in Microbiology*, *10*:389-392.

Tochio, H. (2012). Watching protein structure at work in living cells using NMR spectroscopy. *Curr Opin Chem Biol*, *16*(5-6), 609-613.

Tyler, R. C., Sreenath, H. K., Singh, S., Aceti, D. J., Bingman, C. A., Markley, J. L., Fox, B. G. (2005). Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expression and Purification*, *40*(2), 268-278.

Vankemmelbeke, M., Zhang, Y., Moore, G. R., Kleanthous, C., Penfold, C. N., & James, R. (2009). Energy-dependent immunity protein release during tol-dependent nuclease colicin translocation. *Journal of Biological Chemistry*, *284*(28), 18932-18941.

Vankemmelbeke M, Healy B, Moore GR, Kleanthous C, Penfold CN, James R. Rapid detection of colicin E9-induced DNA damage using Escherichia coli cells carrying SOS promoter-lux fusions [published correction appears in J Bacteriol. 2005 Sep;187(18):6601]. *J Bacteriol*. 2005;187(14):4900-4907.

Verardi, R., Traaseth, N. J., Masterson, L. R., Vostrikov, V. V., & Veglia, G. (2012). Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Advances in Experimental Medicine and Biology*, *992*, 35–62.

Wang, B., Goodpaster, A. M., & Kennedy, M. A. (2013). Coefficient of variation, signal-to-noise ratio, and effects of normalization in validation of biomarkers from NMR-based metabonomics studies. *Chemometr Intell Lab Syst*, *128*, 9-16.

Watt, E D. & Loria J. P. Biomolecular NMR: Isotope labeling methods for protein dynamics studies. *ISOTEC® Stable Isotopes 2008 - 2010 Catalog*

Waudby CA, Camilloni C, Fitzpatrick AWP, Cabrita LD, Dobson CM, Vendruscolo M, et al. (2013). In-cell NMR characterization of the secondary structure populations of a disordered conformation of α-synuclein within *E. coli* cells. *PloS one, 8*(8), e72286.

WeiJin, Yang & YaWei, Wang & Zhou, Qunfang & Tang, Huiru. (2008). Analysis of human urine metabolites using SPE and NMR spectroscopy. *Science in China series b-chemistry.* 51. 218-225. 10.1007/s11426-008-0031-6.

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. [Review]. *Prog Nucl Magn Reson Spectrosc, 73*, 1-16.

Wu, Y., Eghbali, M., Ou, J., Lu, R., Toro, L., & Stefani, E. (2010). Quantitative determination of spatial protein-protein correlations in fluorescence confocal microscopy. *Biophys J, 98*(3), 493-504.

Kurt Wüthrich. (1986) NMR with Proteins and Nucleic Acids. *Europhysics News*. 7,11–13

Wüthrich, Kurt. (2003). NMR studies of structure and function of biological macromolecules.. Bioscience reports. 23. 119-68.

Xing, S., Wallmeroth, N., Berendzen, K. W., & Grefen, C. (2016). Techniques for the analysis of protein-protein interactions in vivo. *Plant physiology, 171*(2), 727-758.

Yao, J. Dyson, H. J. Wright, P. E. (1997). Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Letters, 419*, 285-289.

Zakharov, S. D., & Cramer, W. A. (2004). On the mechanism and pathway of colicin import across the *E. coli* outer membrane. *Front. Biosci, 9*, 1311-1317.

Zgur-Bertok, D. (2012). Regulating colicin synthesis to cope with stress and lethality of colicin production. *Biochem Soc Trans, 40*(6), 1507-1511.

Zhang, X. Y.-Z., Goemaere, E. L., Seddiki, N., Célia, H., Gavioli, M., Cascales, E., & Lloubes, R. (2011). Mapping the interactions between *Escherichia coli* TolQ transmembrane segments. *Journal of Biological Chemistry, 286*(13), 11756-11764.

Zhang, Z., Kuipers, G., Niemiec, Ł. et al. (2015). High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. *Microb Cell Fact, 14*, 142.

Zhang, Y. Z., Paterson, Y., & Roder, H. (1995). Rapid amide proton exchange rates in peptides and proteins measured by solvent quenching and two-dimensional NMR. *Protein science : a publication of the Protein Society, 4*(4), 804–814

Zhou, K., Michiels, C. W., & Aertsen, A. (2012). Variation of Intragenic tandem repeat tract of tolA modulates *Escherichia coli* stress tolerance. *PLOS ONE, 7*(10), e47766.

Zucker, R. M. (2014). Evaluating confocal microscopy system performance. *Methods Mol Biol, 1075*, 321-374.

Websites and softwares

http://www.novagen.com/

https://www.thermofisher.com/sa/en/home.html

https://blast.ncbi.nlm.nih.gov/

Image J with Java 1.8.0_172

SPSS Statistics 17.0

Bucker TopSpin 7.4