THE UNIVERSITY
of EDINBURGH

Post-transcriptional regulation of a micro-injection system

Agata Barbara Wawszczyk

Doctor of Philosophy (PhD) in Infection and Immunity

Collage of Medicine and Veterinary Medicine

The University of Edinburgh

2022

Declaration

I declare that this this thesis was written and composed by me and the work presented is entirely my own. This work has not been submitted for any degree or professional qualification.

Agata Wawszczyk, 2022

Abstract

Enterohaemorrhagic *Escherichia coli* (EHEC) is a food-borne pathogen associated with outbreaks of bloody diarrhoea and haemolytic uraemic syndrome in humans. It originates from asymptomatic carriage in cattle and other ruminants. A critical step in EHEC infection of the human or ruminant host is attachment and effacement of the intestinal epithelium, which is dependent on production of a Type 3 Secretion System (T3SS). This is a molecular syringe encoded on the Locus of Enterocyte Effacement (LEE) composed of five operons LEE1,2,3,5 and 4. The expression of the T3SS is thought to be staged with a checkpoint after production of the membrane-embedded basal apparatus and before production of the LEE4-encoded translocon filament (EspABD) that contacts the host cell. The first protein encoded on the LEE4 operon is SepL and its expression is coupled to translocon proteins EspABD. SepL has been established to be essential for production of surface translocon filaments, although its mechanism of action is still unknown. A recent study by Wang et al. provided evidence that the conformation of the *sepL* mRNA determined its translation and the RNA-binding protein Hfq was shown to negatively regulate SepL expression. There is evidence for a LEE1-encoded factor required to enable LEE4 translation. Together, the *sepL* mRNA secondary structure, Hfq and LEE1-encoded components are hypothesised to form a regulatory interplay that determines SepL expression. The research described in this thesis aimed to further characterise the regulatory factors and unravel the mechanism behind the proposed post-transcriptional control of *sepL* mRNA.

Investigation into the LEE4 mRNA expression revealed a disconnect between the transcriptional activation and LEE4 transcript abundance. This was concluded to be due to rapid degradation of the LEE4 mRNA. The secondary structure of *sepL* mRNA was mapped

using a chemical probing technique called selective 2′-hydroxyl acylation analysed by primer extension (SHAPE). It revealed that the *sepL* transcript adopts a ‘closed’ conformation that likely obstructs the Shine-Dalgarno region from the ribosomal access. Previous research had shown that reporter fusions to truncated *sepL* mRNAs had high expression levels compared to fusions to the full length *sepL* mRNA. Those truncated *sepL* mRNA constructs were computationally predicted to form an ‘open’ secondary structure enabling ribosome binding. The ‘closed’ *sepL* mRNA conformation was linked to a higher rate of transcript decay in comparison to the *sepL* mRNA constructs predicted to form an ‘open’ secondary structure. I propose a model in which the ‘unfolding’ of the *sepL* mRNA leads to translation and protection from degradation. This was indicated to be EHEC background specific. While Hfq had a clear negative impact on *sepL* mRNA stability, deletion and complementation analyses demonstrated that the LEE1-encoded T3SS chaperone CesAB opposed the Hfq-mediated degradation of *sepL* mRNA and is a post-transcriptional regulator of SepL and other LEE4-encoded proteins. CesAB was shown to directly interact with the *sepL* transcript and the mRNA signatures involved in the interaction were investigated. The positive impact of CesAB on *sepL* mRNA maintenance was diverted by expression of CesAB’s protein cargo, the LEE4-encoded T3SS filament protein EspA. This indicates a hierarchical feedback loop controlling expression of the LEE4-encoded proteins. To our knowledge this is the first reported instance of a T3SS protein chaperone also regulating the expression of its subsequent protein cargo at the mRNA level. Therefore, it acts first to stabilise the LEE4 transcript followed by interaction with two of LEE4 mRNA-encoded proteins. The switch of CesAB from protein-RNA interaction to protein-protein interactions would naturally lead to degradation of the translocon mRNA. The duality of substrates may be a paradigm common to other substrates and this needs evaluation.

The conducted study aimed to widen the knowledge of the mechanisms that have evolved to stage the construction of complex bacterial organelles. In turn this can lead to a better understanding of the processes underlying potentially lethal infections and indicate new anti-virulence drug targets.

Lay Summary

Enterohaemorrhagic *Escherichia coli* (EHEC) is a bacterium associated with human and animal disease characterized by development of bloody diarrhoea. EHEC infections can lead to long lasting health problems and are potentially fatal. To infect and colonise its host EHEC attaches tightly to the cells that line the gastrointestinal tract. To do this, the bacterium injects the host cell with a cocktail of molecules called effector proteins, which bind to proteins within the cells and alter their function. This ultimately leads to alterations in the host cell's cytoskeleton, a network of filaments and microtubules which, just like bones, create an internal framework of the cell responsible for supporting cell's structure. The delivery of the effector proteins into the host cell is orchestrated by complex injection machinery called a Type 3 Secretion System (T3SS). The T3SS is a molecular syringe composed of the basal apparatus, embedded in the bacterial membrane, which extends into a 'injection filament' that projects outside of the bacterium. On the tip of the 'needle' a pore forms connecting the bacterium to the host cell. Assembly of the syringe was shown to be staged with the basal apparatus considered to be produced first, followed by formation of the 'needle'. One of the T3SS proteins, called SepL, was shown to control production of the 'needle' and without SepL, the 'needle' won't be produced. However little is known about factors involved in control of SepL production.

All proteins are produced from a line of code, called mRNA. The information encoded by the mRNA is translated into protein by the cellular machines called ribosomes. Those lines of code fold into different conformations and form 3D structures. The structure of the mRNA can either grant or inhibit ribosome access and thus regulate production of the protein. The mRNA molecules that are not translated by the ribosome are often unstable and prone to degradation. Some

proteins are able to attach to the mRNA and positively or negatively control its translation and maintenance.

My project indicated that the SepL-encoding mRNA is prone to degradation. A negative regulator of SepL production called Hfq was shown to be involved in the mRNA decay. T3SS protein called CesAB was shown to attach to *sepL* mRNA and oppose the Hfq-mediated degradation. The structure formed by *sepL* mRNA was determined to be 'closed' and likely to inhibit the ribosome access. Unfolding of the *sepL* mRNA is predicted to be necessary for its translation. The Hfq-CesAB interplay is thought to regulate the conformation of the *sepL* mRNA from a folded state that inhibits its translation and leads to degradation, to an unfolded state that allows ribosome binding and translation and protects the mRNA from decay. This mechanism underpins the production of the T3SS 'needle' and thus is crucial for EHEC infection. The primary finding of the work was the capacity of CesAB to firstly protect its mRNA target and then to be diverted from the mRNA by the translated products (injection filament components), thus providing a neat feedback loop to control production of the filament.

Studying the assembly of complex molecular machines, such as the Type 3 Secretion apparatus, allows us to fully understand the infection process and develop new ways to stop and treat these potentially lethal infections.

Acknowledgements

Throughout my research and the writing of this thesis I have received a huge amount of help and support from countless people. First I would like to thank my supervisors Prof. David Gally and Dr. Sander Granneman for giving me an opportunity to work on this project, be a part of their teams, all of the advice, patience and enthusiasm in guiding me through this academic experience. I want to specifically thank Prof. David Gally for teaching me to approach the grand ideas with creativity and passion. This attitude has shaped the scientist I am today.

I want to express my utmost gratitude to Sean McAteer for holding my hand through many experiments and whose support was simply invaluable. I cannot thank all of the members of the ZAP group enough for their advice and good times sharing the lab together.

Thank you to EASTBIO DTP for funding my project in the first place and allowing me to be a part of an amazing cohort. I also want to thank my kind and inspiring internship supervisor Dr. Katy Hayden at the Royal Botanic Gardens Edinburgh that introduced me to a fascinating world of *Phythopthora*.

I want to give massive thanks to my close ones: Elena for reading through this thesis from Texas; Izzy for a perfect mixture of meaningful and lighthearted support; Imogen for the shared T3SS enthusiasm; Wendy for being on my side; Alba for listening with kindness; Jimmy for constant encouragement; and Jean for making me smile throughout the writing process. I would not be able to do this without you.

Finally, I want to thank my family. My parents Jarosław and Barbara and my sister Alicja, who brought me up to be curious about the world, prioritised my education and keep on always believing in me. Kocham Was.

Contents

List of figures

List of Tables

List of abbreviations

A – adenine

ADP – adenosine diphosphate

A/E – attaching and effacing

APS – ammonium persulphate

asRNA – anti sense ribonucleic acid

ATP – adenosine triphosphate

C – cytosine

cDNA – complementary DNA

CDS – coding sequence

CFU – colony forming unit

CIMS - crosslink induced mutation sites

CLASH - cross-linking, ligation, and sequencing of hybrids

CLIP – cross-linking and immunoprecipitation

CLIP-seq - cross-linking and immunoprecipitation high-throughput sequencing

cm – centimetre

CRAC - called crosslinking and analysis of cDNAs

DAEC - diffusely adherent *Escherichia coli*

DMSO – dimethylsulphoxide

DNA – deoxyribonucleic acid

DTT - dithiothreitol

EAEC – enteroaggregative *Escherichia coli*

ECL - enhanced chemiluminescence

EDTA - Ethylenediamine tetraacetic acid

EF-G – elongation factor G

EHEC – enterohaemorrhagic *Escherichia coli*

EIEC – enteroinvasive *Escherichia coli*

EMSA - electrophoretic mobility shift assay

EPEC – enteropathogenic *Escherichia coli*

ETEC - enterotoxigenic *Escherichia coli*

G – guanine

g - gram

GFP – green fluorescent protein

GTC - guanidinium thiocyanate

GTP – guanosine triphosphate

HIV-1 – human immunodeficiency virus 1

H-NS - histone-like nucleoid-structuring protein

HRP – horseradish peroxidase

HUS – haemolytic uremic syndrome

IAA – isoamyl alcohol

IPTG - isopropyl-β-d-thiogalactopyranoside

kb – kilo bases

kDA – kilo Daltons

kV – kilo volts

LB - Luria-Bertani

LEE – Locus of Enterocyte Effacement

MEM-HEPES – minimal essential media HEPES

µF – micro Farad

µg – micro gram

µl – micro litre

M - mol

mg – milli gram

ml – milli litre

mm – milli meter

mM – milli molar

mRNA – messenger ribonucleic acid

ncRNA – non-coding ribonucleic acid

nm – nano meter

NMEC - neonatal meningitis *Escherichia coli*

OD – optical density

OMV – outer membrane vesicle

ONPG - *o*-nitrophenyl-β-galactoside

PAA – polyacrylamide

PARIS - psoralen analysis of RNA interactions and structures

PARS - parallel analysis of RNA structures

PBS – phosphate-buffered saline

PCR – polymerase chain reaction

PNK – polynucleotide kinase

qPCR – quantitative PCR

R-proteins – ribosomal proteins

RBP – ribonucleic acid binding protein

RBS – ribosomal binding site

RF – random forest

RIL-seq - RNA interaction by ligation and sequencing

RIP – ribonucleic acid immunoprecipitation

RNA – ribonucleic acid

RNA-seq – ribonucleic acid sequencing

RNAT – ribonucleic acid thermometer

RNP – ribonucleoprotein

rpm – rotations per minute

RT qPCR – reverse transcriptase quantitative polymerase chain reaction

SCV – *Salmonella* containing vacuole

SD – Shine-Dalgarno

SDM – site-directed mutagenesis

SDS-PAGE – sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SHAPE - selective 2'-hydroxyl acylation analysed by primer extension

SiM-KARTs - single molecule kinetic analysis of RNA transient structure

SL – stem loop

smFRET - single-molecule fluorescence fesonance energy transfer

SPI-1 - *Salmonella* Pathogenicity Island 1

sRNA – small ribonucleic acid

STMV – Satellite Tobacco Mosaic Virus

Stx – Shiga toxin

SVM – support vector machine

T- thymine

T1SS – Type 1 Secretion System

T2SS – Type 2 Secretion System

T3SS – Type 3 Secretion System

T4SS – Type 4 Secretion System

T6SS – Type 6 Secretion System

TEMED - Tetramethylethylenediamine

TIS - transposon insertion sequencing

TPP - thiamine pyrophosphate

tRNA – transfer ribonucleic acid

U – uracil

UPEC – uropathogenic *Escherichia coli*

USD – United States Dollar

UTR – untranslated region

UV – ultra violet

V - volts

w/v – weight per volume

Chapter 1 Introduction

1.1 Translational control in pathogenic bacteria

Most bacterial pathogens are extremely versatile microorganisms adapted to occupy different niches, overcome environmental pressures such as temperature, pH and osmolality change and utilise varying resources. In order to successfully infect the host, pathogens employ a multitude of virulence factors such as secretion systems or toxins, which aid in colonisation and immune-evasion (Mandell et al, 2015). Those tools are selectively deployed depending on changes in the environment and their expression is tightly regulated. This regulation can occur at any point along the continuum of gene expression: transcription of the DNA, translation of the messenger RNA (mRNA) or post-translationally via protein modification. Although transcriptional control plays a significant role in the regulatory cascades, a weak correlation between the transcript and protein abundance can be observed throughout the bacteria kingdom (Dressaire et al., 2010; Lu et al., 2007; Nie et al., 2006, Picard et al., 2013). This indicates apparent involvement of post-transcriptional regulation that was shown to govern all aspects of bacterial activity including house-keeping, metabolic and pathogenic processes (Bertram et al., 2014; Le Scornet et al., 2019; Romeo et al., 2013; Schiano et al., 2012). This part of the thesis describes the factors known to translationally control gene expression in Gram negative bacteria with a focus on mRNA secondary structure, regulatory noncoding RNAs and RNA binding proteins.

1.1.1 mRNA secondary structure

1.1.1.1 Global regulatory characteristics of mRNA conformation

mRNA carries information in a three-dimensional plane and in order to decode it, interpretation beyond reading of the nucleotide sequence is required. mRNA is by its nature prone to fold into higher-order secondary (hairpins and stem-loops) and tertiary (3D conformations stabilised by interactions between distantly located sequences) structural motifs. Those structures are highly dynamic and can adapt several different complex conformations in timescales measured in microseconds. Specific structures adopted by the transcript can perform biological functions and so transitions between the alternative structures are critical to the regulation of gene expression (Chemla et al., 2020; Ermolenko et al., 2020; Mauger et al., 2019; Meyer, 2017; Schroeder et al., 2004).

Comprehensive research by Del Campo et al., 2015 investigated global trends in *E. coli* mRNA structures and their effect on translational efficiency. The study established that unstructured conformation of the transcript correlates with increased translational initiation rates and protein expression. Openness of the mRNA conformation around the 5’ untranslated region (UTR) of the transcript was shown to determine the translational efficiency, more so than the complementarity of the mRNA Shine-Dalgarno (SD) sequence to the 30s subunit of the ribosome that is canonically used as a predictor of translational efficiency (Beck et al., 2016; Del Campo et al., 2015; Hockenberry et al., 2016; Jacob et al., 1987). It has been proposed that the unfolded 5’UTR is recognised by S1 ribosomal protein, known to have an affinity for mRNA binding upstream of the start codon. This interaction could act as primary recognition event and facilitate docking of the 30S subunit at the Shine-Dalgarno.

mRNA secondary structure adopted by the 5'UTR is often associated with transcript stability. Unlike ribosomal RNA and transfer RNA, mRNA is unstable and prone to decay (Cannistraro et al., 1985). RNase E is the main component of the RNA degradosome in *E. coli*, and is therefore essential for mRNA degradation and transcript processing. The RNase E binding motif consists of A/U-rich single-stranded mRNA regions. Several single gene studies, as well as a global analysis of *E. coli* genome, identified a structural signature, composed of an unpaired sequence preceded by a stem loop that is proposed to facilitate recognition by RNase E and cleavage of the transcript. Docking of the ribosome at the 5' UTR was shown to protect the mRNA from RNase E mediated cleavage (Braun et al., 1988; Choi et al., 2018) and inefficient ribosomal binding is typically associated with an increase in transcript degradation (Morris et al., 2021). This mutually exclusive interplay between translational machinery and the degradosome, utilising a specific structural signature as a recognition beacon, showcases the importance of sequence-independent mRNA structures in fine-tuned regulation of mRNA maintenance and translation initiation.

Apart from its importance in ribosomal recognition, unfolding of the mRNA is necessary for the next step of translation, elongation. Decoding of the mRNA, involves feeding of the mRNA strand through a narrow tunnel formed by S3, S4 and S5 ribosomal proteins. The architecture of this channel prevents structured motifs from entering, therefore the mRNA strand needs to be unfolded by approximately two codons in order to reach ribosomal decoding site. This change in mRNA structure is hypothesised to utilise the free energy released by EF-G-mediated GTP hydrolysis, during translocation. Complexity of the folding was shown to directly correlate with a decrease in elongation speed and ribosomal stalling (Choi et al., 2018). Additionally, mRNA structure-mediated pacing of translation is implicated in efficiency of recoding phenomenon, which refers to the production of alternative

proteins from the same nucleotide sequence. Recoding maximises density of encoded genetic information and can be utilised to regulate protein function (Chiaruttini et al., 2019).

mRNA structure was also shown to play a role in translation termination. Global analysis of secondary structures in the *E. coli* transcriptome, indicated a persistent but sequence unrelated structural signature upstream of the stop codon. This correlated with enrichment of ribosomes directly upstream of the structure, suggesting involvement of the motif in stalling of the ribosome. The ribosome pausing at the persistent structure was hypothesised to ensure accurate positioning of the ribosomal decoding site at the stop codon and thus mediate successful termination of translation (Del Campo et al., 2015).

1.1.1.2 Riboswitches and RNA thermometers

Apart from its general involvement in translation efficiency, mRNA structures are prominently utilised to act as sensory switches, called riboswitches or RNA thermometers (RNATs) (Abduljalil, 2018; Loh et al., 2010; Nerberhaus, 2010). Riboswitches are able to monitor environmental chemical or physical cues by binding to a specific ligand and, depending on the availability of the signal, inhibit or promote gene expression without obligate involvement of other factors. This control is dependent on conformational changes in the riboswitch or RNATs mRNA, which result in mutually exclusive mRNA structures, the ‘on’ anti-terminator and ‘off’ terminator mRNA states in terms of expression (Figure 1.1) (Pavlova et al., 2019). Riboswitches can act both on transcriptional and translational efficiency, however in Gram negative bacteria the inhibition of the translation initiation by sequestration of the Shine-Dalgarno sequence, is the most commonly observed mode of riboswitch-mediated expression regulation (Nudler et al., 2004; Garst et

al., 2011). Riboswitches and RNATs are noncoding stretches of RNA localised in the 5'UTR of the monocistronic or polycistronic mRNA for the genes under their control. A riboswitch is composed of two functional domains, the aptamer and the expression platform, linked by an overlapping region called the switching sequence. The aptamer region adopts a scaffolding-like fold that acts as a ligand-binding pocket that interacts with the cognate ligand (if required) with extreme specificity, discriminating between highly related compounds, such as adenine and guanine or one methylene group of lysine versus ornithine. Docking of the ligand at the aptamer results in reorganisation of the aptamer region and a conformational change in the expression platform structure, which interacts with translational machinery and determines progress in terms of repression or expression (Garst et al., 2011). Concentrations of various metabolites, tRNAs and ions, that reflect things like nutrient availability, temperature or pH change, were shown to elicit riboswitch-mediated regulation of metabolism and transport-related protein synthesis (Berdard et al., 2020). The architecture of riboswitches was shown to adhere to general patterns adopted by other RNAs, including loops, helices and bulges, which are then configured into more advanced structures such as GA3 tetraloop, pseudoknots, kink-turns, kissing-loop or T-loops (Garst et al., 2011).

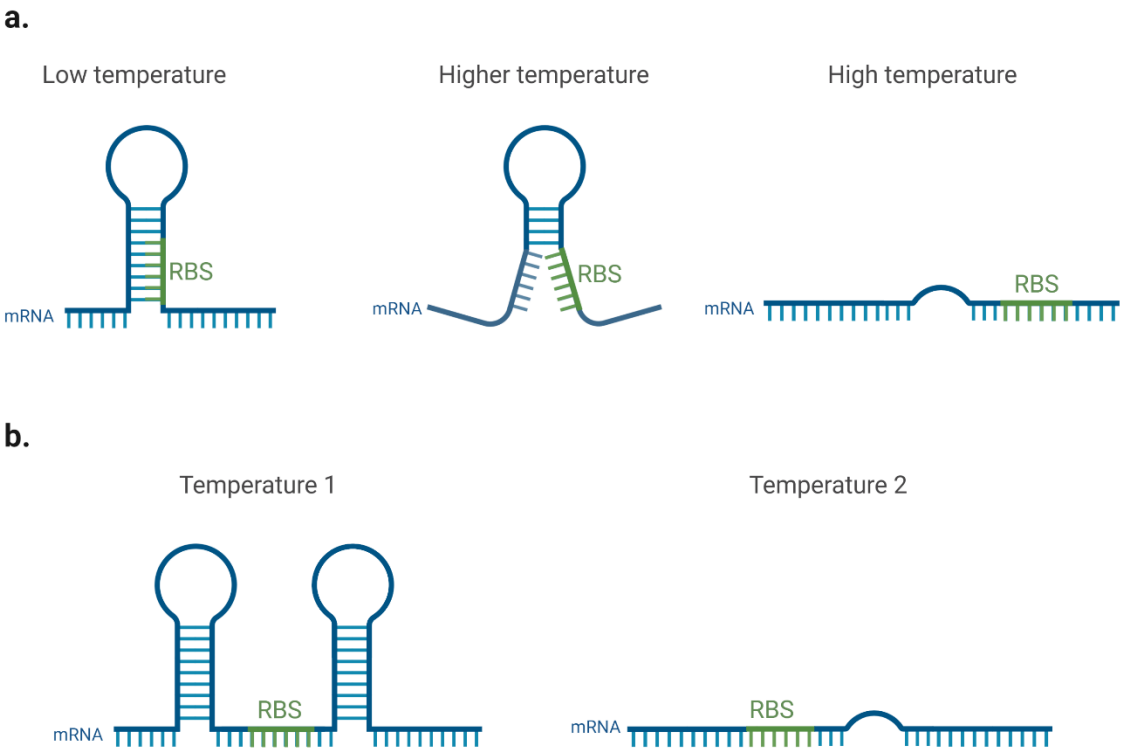


Figure 1.1. RNATs mRNA conformation rearangements leading to change in ribosomal binding site (RBS) accessibility. **a.** RNA zipper. Gradual change in temperature results in progressive 'melting' of the secondary structure. **b.** RNA switch. mRNA adopts two different, mutually exclusive conformations depending on change in temperature (Garst et al., 2011).

Due to the poor sequence conservation of riboswitches and RNATs, the computational prediction of such regulatory sequences based only on genomic information is particularly challenging (Chang et al., 2009). Sophisticated RNA structure probing techniques such as Single Molecule Kinetic Analysis of RNA Transient Structure (SiM-KARTS), Selective 2'-Hydroxyl Acylation analysed by Primer Extension (SHAPE), single-molecule Fluorescence Resonance Energy Transfer (smFRET), PARIS (psoralen analysis of RNA interactions and structures), and PARS (parallel analysis of RNA structures), as well as computational folding predictor software can be successfully utilised in order to determine potential regulatory structures within mRNA and indicate mechanistic aspects of their action (Chang et al., 2009; Chauvier et al., 2019). SiM-KARTS performed on the *Thermoanaerobacter tengcongensis* 7-aminomethyl-7-deazaguanine (preQ1)-sensing riboswitch, was employed to follow the alternate structure adaptation of the riboswitch dependent on the availability of the ligand. Those experiments indicated the conformational changes in the aptamer region upon ligand binding and correlated them with the opening of the Shine-Dalgarno region and expression of the gene of interest (Rinaldi et al., 2016). In addition, SHAPE analysis of the most studied bacterial riboswitch, the thiamine pyrophosphate (TPP), demonstrated that disruption of the folded structure interferes with formation of a functional riboswitch domain and renders the regulatory mechanism dysfunctional (Haller et al., 2013; Stoddard et al., 2008). Programmed pausing sites described above are additional mRNA structure-dependent regulatory elements often positioned within the expression platform of the riboswitches. Those structures slow down the decoding machinery and allow for correct folding of the aptamer, crucial for ligand detection and functionality of the switch. Further complexity of riboswitch-mediated regulation can be achieved by formation of another adjacent regulatory structure. Tandem architecture allows for increased sensitivity for the same ligand or more fine-tuned

regulation responsive to more than one cognate ligand (Sherlock et al., 2018; Sudarsan et al., 2006).

Riboswitches and RNATs are widely distributed among pathogenic bacteria. One of the better studied translational mRNA-encoded switches in Gram negative pathogens is the RNAT of *Yersinia pestis*, responsible for production of a transcriptional regulator LcrF that dictates production of the Yersinia type III secretion system and secreted *Yersinia* outer proteins (Yops); together these are critical virulence factors for infection and immune evasion. The *lcrF* RNAT occludes the Shine Dalgarno sequence and the AUG start codon, inhibiting the LcrF expression. At 37℃, indicative of the host environment, this structure melts and the Shine-Dalgarno and start codon are exposed which permits translation of the *lcrF* mRNA and subsequent expression of the virulence machinery (Hoe et al., 1993; Nuss et al., 2017). Similarly the RNATs of *Shigella dysenteriae* outer membrane heme receptor ShuA, important for acquisition of iron from complex molecules present in the host organism, depends on the conformational change of the mRNA structure that alters the access to the interaction sites that are critical for transcript decoding or processing. Alteration of transcript conformation is basis of many more riboswitches such as heat shock genes of *Salmonella typhimurium htrA* and σ^{32}, *rpoH* in *Escherichia coli* and adenine riboswitch of *Vibrio vulnificus*. RNase E cleavage-mediating *lycC* lysine switch and pH sensor *alx* in *Escherichia coli*, as well as *Salmonella* ATP and tRNA-sensing *mgt* riboswitch are also dependent on change in mRNA secondary structure (Barros et al., 2016; Chauvier et al., 2019; Kouse et al., 2013; Waldminghaus et al., 2007).

The examples described above highlight the layered nature of messages carried by various elements of the nascent mRNA molecule, with emphasis on the extent to which the mRNA structure can have an impact on translation initiation and efficiency. The regulatory attributes

of the mRNA structure can be further utilised in mRNA-RNA and mRNA-protein regulatory interactions as described below.

1.1.2 **Regulatory non-coding RNAs**

In the past decade non-coding RNAs (ncRNA) have taken gene expression regulation by storm. These micro regulators are abundant in all bacteria, managing processes such as adaptation to change in temperature and nutrient availability, adherence to host cells, immune-evasion or spread of the pathogen (Ahmed et al., 2016; Gottesman, 2005; Moodey et al., 2013; Perez-Reytor et al., 2017). The cellular mechanisms that were shown to undergo ncRNA-mediated regulation include transcriptional reprogramming, carbon storage, toxin and antitoxin systems, expression of entire virulence operons, iron homeostasis, envelope production, SOS response, enzyme modulation, dual action between sugar metabolism and virulence factor expression, transition phases and feedback regulation (Felden et al., 2021; Repoila et al., 2009; Perez-Reytor et al., 2017). It is likely that ncRNA control is standard for systems requiring responsive regulation. The mechanistic and physiological aspects of ncRNA-mediated regulation are described in more detail below.

ncRNAs can be divided into *cis*-encoded antisense RNAs (asRNAs), that are expressed from the complementary DNA strand to the RNA they act upon, and *trans*-encoded small RNAs (sRNAs) that control expression of genes encoded in different genomic locations (Saberi et al., 2016; Millar et al., 2021). *cis* asRNAs are most commonly associated with transcriptional interference via binding to their reciprocal mRNA sequences, however post-transcriptional regulatory mechanisms involving modulation of mRNA stability and translational control achieved by occlusion of the mRNA ribosomal binding sites, have been reported (Brantl, 2007; Georg et al., 2011). One of the examples of translational control mediated by asRNAs is the interaction between the asRNA and global RNA chaperone Hfq of *Legionella pneumophila*, a Gram-negative pathogen responsible for life-

threatening atypical pneumonia. Binding of the asRNA to *hfq* mRNA is mediated by the Hfq itself, resulting in a negative autoregulatory feedback loop, that inhibits Hfq expression in the outside-host environment (Oliva et al., 2017). Toxin-antitoxin systems are widely distributed in pathogenic bacteria. In Type I toxin-antitoxin systems, such as TisB/IstR-1 in *E. coli*, the asRNA is used as a neutralising agent that inhibits expression of the toxin, and thus represses cell killing or growth suppression in the organism and the daughter cells (Sarpong et al., 2021).

The first mode of action of *trans*-encoded sRNA is to affect gene expression directly by sRNA-mRNA interactions with or without aid of RNA chaperone proteins (Millar et al., 2021). Most reported instances of such mechanism involve base pairing between the complementary sRNA and the mRNA sequences around ribosomal binding sites, which occludes the access of the translational machinery and inhibits protein production (Desnoyers et al., 2013). An example of such interaction has been recently described in *Salmonella enterica* serovar Typhimurium in which the PinT sRNA regulates expression of two major transcriptional regulators of Salmonella Pathogenicity Island 1 (SPI1) *hilA* and *rtsA* (Kim et al., 2019). Kim et al., 2019 showed that binding of the PinT to the 5'UTR Shine-Dalgarno region of those transcripts leads to impaired ribosomal interactions and inhibition of translational activation.

Instances of translation activation mediated by sRNA binding to the transcript were also reported. In EHEC the expression of global transcriptional activator PchA is autoregulated by *pchA* coding sequence (CDS) pairing with the 5′ untranslated region (UTR) of the mRNA, which sequesters the ribosome binding site (RBS) and inhibits translation. This in turn negatively regulates expression of the EHEC Type 3 Secretion System (T3SS), a major colonisation factor (see later section 1.2.1.2). Under an oxygen limiting conditions that signal appropriate colonisation niche, production of a sRNA DicF that binds to the *pchA* CDS and disrupts the inhibitory *pchA* CDS-5'UTR interaction,

results in activation of *pchA* mRNA translation that subsequently allows for expression of T3SS (Melson et al., 2019).

Apart from direct competition of sRNAs with ribosomes, targeted degradation of the RNA duplexes by RNase E and RNase III is a widely utilised sRNA-mediated regulatory mechanism (De Lay et al., 2012; Desnoyers et al., 2012; Grutzner et al., 2021; Svensson et al., 2021). For example, RyhB, a sRNA involved in bacterial survival in iron-scarce conditions, was shown to directly mediate rapid degradation of at least 18 mRNAs which encode proteins which require iron for activity. The sRNA-mediated degradation of mRNA was shown to conclude translational repression, often employed before the processing of the transcript, deeming such gene silencing irreversible (Prevost et al., 2011). Two mechanisms are considered to play a role in sRNA-mediated mRNA degradation. Binding of the sRNA to the ribosomal binding sites results in loss of mRNA shielding by the translating ribosome and increased sensitivity of the transcript to the RNase activity (Sulthana et al., 2016). The *trans* sRNA-mediated control of mRNA decay is usually assisted by the RNA chaperone Hfq, that binds to ncRNAs with high affinity and stabilises the sRNA-mRNA-RNase E interaction (Morita et al., 2011; Santiago-Fargos et al., 2018). However, there is no requirement for the Hfq in formation of the RNA-RNA duplexes and it is considered to be more of a ‘helper’ molecule. *cis* sRNA-mediated mRNA degradation does not utilise the RNA chaperone and in some Gram-negative bacteria that were shown to employ sRNA-mediated regulation, Hfq is missing altogether (Gottesman et al., 2011). A more detailed description of Hfq and its translational control in bacteria, including the relevance of Hfq/ncRNA/RNase E complexes, can be found in section 1.1.3. Apart from their role in degradation-translation inhibition coupling, sRNA interactions were shown to actively contribute to transcript processing. RNase E is known to be involved in two mRNA decay pathways: a first and direct pathway allowing cleavage of triphosphorylated primary

transcripts (described above) and a 5′-end-dependent mRNA degradation pathway which requires prior conversion of the 5′-triphosphate extremity to a 5′-monophosphate by the RNA pyrophosphohydrolase RppH, in a manner similar to decapping of mRNAs in eukaryotes. RNase E favours binding on a monophosphorylated 5′-end (5′-P end), which can fit in a discrete pocket on the surface of RNase E, facilitating mRNA cleavage at a downstream location by the RNase E active site. Interaction of the 5′-P end of sRNA with a sensor pocket of RNase E aids closure of the catalytic domain to enhance nucleolytic activity (Lalaouna et al., 2013).

trans-encoded sRNAs can indirectly regulate mRNA translation by interfering with the regulatory proteins acting on the transcript. *Pseudomonas aeruginosa* posttranscriptional regulatory system Rsm (regulator of secondary metabolites) and its ortholog in *E. coli* Csr (carbon storage regulator), that are responsible for shift in expression from planktonic lifestyle-related factors like Type 4 Secretion System and Pel and Psl exopolysaccharides to colonisation virulence mechanisms such as pilus and Type 3 Secretion System, are among the best studied examples of regulatory protein-sRNA interplay. In this interaction two sRNAs CsrB/RsmY and CsrC/RsmZ sequester the regulatory protein CsrA/RsmA and prevent it from binding to its target mRNA, therefore indirectly controlling the mRNA translation (Janssen et al., 2018). It is an outstanding feature of many indirectly acting sRNAs that they modulate the multitasking regulatory proteins such as CsrA/RsmA and Hfq and thus target a multitude of products at once and facilitate widespread changes in virulence and metabolism that are often intertwined. For example, the sugar-phosphate stress and pathogenesis T3SS effector SopD of *Salmonella* are both regulated by the SgrS sRNA, and Hfq interacting TarA sRNA of *Vibrio* was shown to link glucose acquisition with expression of a major transcriptional activator of virulence, ToxT (Bobrovskyy et al., 2014; Richard et al., 2010).

It is worth noting that multiple sRNAs can elicit similar regulation with varying intensity and that the same sRNA can invoke different regulatory events depending on pairing compatibility with its target. Functional redundancy and functional overlap in sRNA repertoires are commonly observed and explain the difficulty in assigning a clear phenotype to a specific sRNA. Although effective base-pairing is crucial for sRNA ligand recognition, only a limited number of nucleotides are involved in mRNA-sRNA interactions. The secondary structures of the binding partners were shown to be essential for targeting of the sRNA cognate transcript elements. OxyS-*fhlA* sRNA-mRNA interaction in ***E. coli***, governing formate metabolism in oxidative stress conditions, was demonstrated to be based on two short loop-loop base-pairing complexes, instead of extended RNA duplex formation (Salim et al., 2010). Strength of the interaction between the quorum sensing-associated *Vibrio cholerae* sRNA Qrr3 and its target was shown to determine if the mRNA is going to be translated, sequestered in an inactive conformation or degraded alone or along with Qrr3 (Song et al., 2008).

Other interesting sRNA modes of action involve but are not limited to membrane localisation of the translation complex, sRNAs binding outside of the translation initiation region that was shown to mediate inhibition or activation of translation without alteration of mRNA secondary structure, and indirect regulation of expression via mRNA stabilisation (Aiba, 2007). It has been also proposed that 5' and 3' UTR cis-encoded elements including riboswitches (described in section 1.1.2) when cleaved off and released from the downstream mRNA, can act as a *trans* acting sRNAs with distinct regulatory function. For example, processing of the 5' UTR of a transposase encoding *tnpA* of *Salmonella typhimurium* leads to release of sRNA that sequesters the mRNA encoding the major transcriptional regulator of SPI-1 pathogenicity island, *invF*, and represses its expression, setting a threshold for activation of the SPI-1 virulence factors (Ellis et al., 2017).

In addition to the expanding work on known ncRNA systems, novel ncRNA classes including small CRISPR/Cas9-associated RNAs and antimicrobial eukaryotic long non-coding RNAs and new concepts such as cell-to-cell heterogeneity in ncRNA expression and extracellularly acting RNAs are gaining more research interest (Buettner et al., 2015; Hadjicharalambous et al., 2019; Lin et al., 2019). Global techniques such as dual RNA-seq, that provides information about infecting pathogen gene expression and the corresponding host response, are currently utilised to discover new sRNAs, asses their regulatory implications and mode of action. Another methodology the transposon insertion sequencing (TIS) can be applied to systematically evaluate on the fitness and virulence costs of mutations in informational sequences including those encoding sRNAs. Additionally, RNA-RNA and RNA-protein interaction detection methodologies such as UV cross-linking and immunoprecipitation high-throughput sequencing (CLIP-seq), cross-linking, ligation, and sequencing of hybrids (CLASH) and RNA interaction by ligation and sequencing (RIL-seq) are also applied to identify and study novel sRNAs. With the obvious scientific interest and the rapid progress in the ncRNA field, many more of specific ncRNA interactions as well as general mechanisms of action and ncRNA systems are predicted to be uncovered in the near future (Handzlik et al., 2020; Li et al., 2012).

1.1.3 **RNA binding proteins**

The most obvious and best conserved RNA-binding proteins that are present in all living organisms are ribosomal proteins (r-proteins), essential for ribosome assembly and supporting the complex translation process. However, in bacteria the r-proteins constitute only a third of the RBPs (Smirnov et al., 2016). In the last two decades the extensive research on regulation of bacterial translation identified heavy involvement of RBPs in modulation of protein expression. This section is going to focus on the role of RBPs in post-transcriptional processes, describe the general mechanisms behind the mRNA-protein interactions and supply the most prominent examples of the regulatory RBPs in Gram negative bacteria.

RBPs interactions range from binding of a single protein to a single RNA molecule, to creation of multi-factor complexes like those formed by the *E. coli* degradosome (Morita et al., 2005). Some of those interactions are stable, resulting in formation of ribonucleoproteins (RNPs), and some are transient, which allow for regulation of RNA function or processing. Regardless of the nature of the interaction, the selectivity of the target binding by the RBP is crucial, yet still not fully understood (Duss et al., 2019; Holmqvist et al., 2018). However, the biochemical underpinning of the RBP-RNA binding simmers down to the similarity of intramolecular forces acting on those molecules. Strength of the Van der Waals (VdW) interactions, hydrophobicity, **π** and stacking interactions between different combinations of protein residues and RNA interfaces were all shown to determine the affinity for RNA-RBP interactions and the stability of those complexes (Akopian et al., 2013; Hu et al., 2018; Teplova et al., 2011; Wilson et al., 2016). Analysis into patterns of amino acid sequences and nucleotides preferred in RBP-RNA binding as well as structural motifs of the interacting partners, allowed for identification of some conserved

protein signatures indicative of RNA-binding function; a comprehensive account of this can be found in the reviews by Van Assche et al., 2015 and Buckley et al., 2009. However, those motifs serve only as a clue and more direct methods such as binding assays are used to identify novel RBP-RNA interactions. Classical *in vitro* binding assays are based on immunoprecipitation of candidate RBP and subsequent identification of RNA target by global RNA-sequencing (RNA-seq), microarray or specific RT qPCR analysis. More sophisticated *in vivo* assays that couple UV crosslinking and immunoprecipitation (CLIP) allow for additional determination of specific protein-RNA binding sites. Another advantage of the *in vivo* assays is their consideration for the pleiotropic effects such as structural changes that occur in proteins and transcripts depending on environmental conditions and association with other interacting partners. Often the regulatory function of the RBP is discovered before the specific interaction is mapped. Candidate RBP knock out or overexpression backgrounds are assayed for changes in global or specific transcript abundance or proteins expressed by techniques such as RNA-seq, microarrays, reverse transcriptase quantification PCR (RT qPCR), proteomics or immunoblotting (Ramanathan et al., 2019). Results obtained from combinations of those techniques can help with determination of the RBP's mode of action. The most prominent post-transcriptional mechanisms mediated by the RBPs and the examples of the interactions are summarised in Figure 1.2.

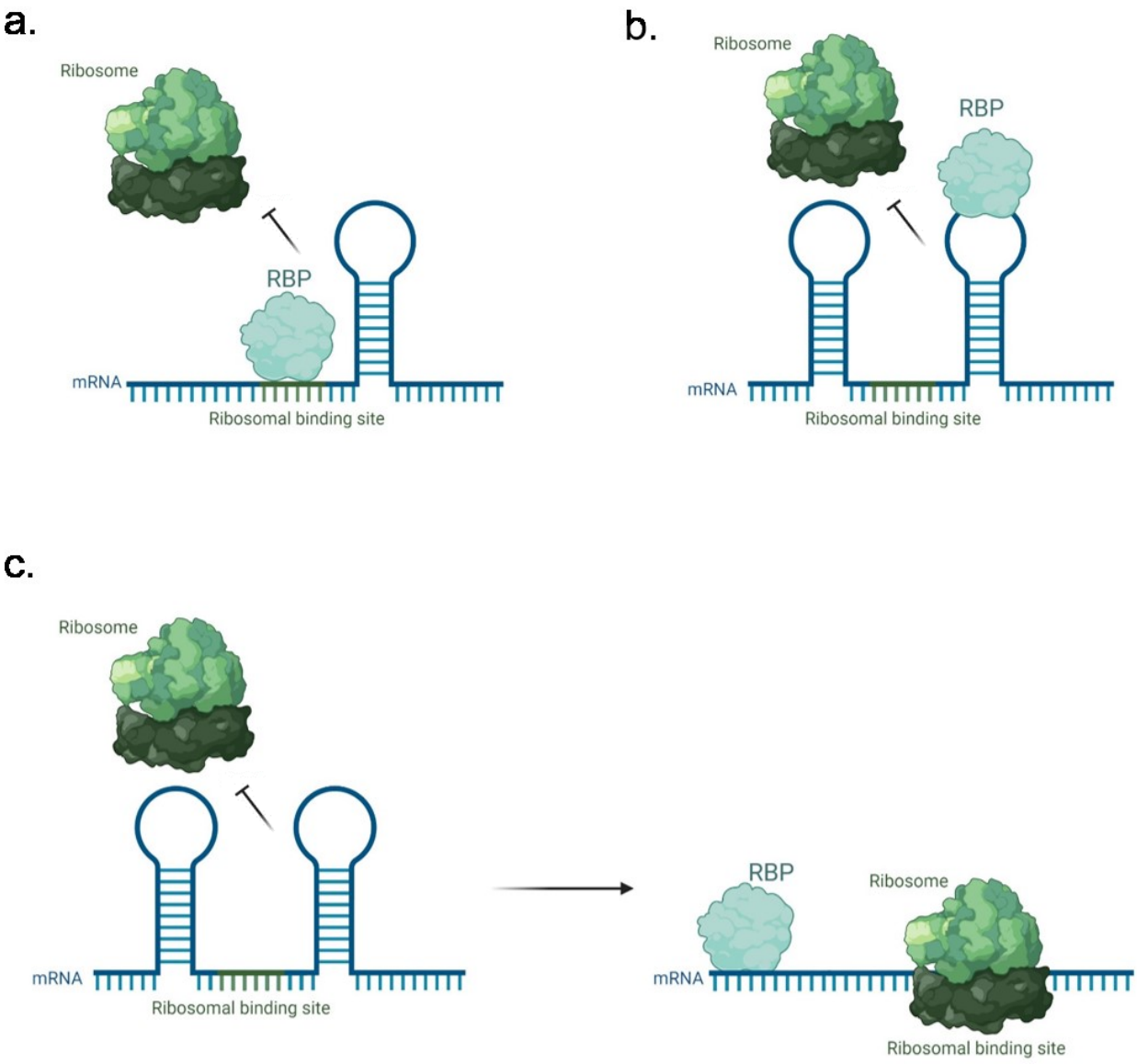

a.
Ribosome
RBP
mRNA
Ribosomal binding site
b.
Ribosome
RBP
mRNA
Ribosomal binding site
c.
Ribosome
mRNA
Ribosomal binding site
RBP
Ribosome
mRNA
Ribosomal binding site

Figure 1.2. Canonical modes of action of RNA binding proteins with focus on post-transcriptional regulation of mRNA. **a**.,**b.**&**c.** RBP binding to the transcript regulates ribosomal docking. **a.** Interaction of the RBP (turquoise) with the mRNA (blue) around the ribosomal binding site (green) obstructs ribosomal access and inhibits translation initiation (Baker et al., 2007). **b.** RBP-facilitated change in mRNA secondary structure results in burying of the ribosomal binding site and stops the ribosome from attaching to the transcript (Dubey et al., 2005). **c.** Binding of the RBP results in change of the transcript conformation from inhibitory (ribosomal binding site buried within the mRNA secondary structure) to open and permissive for ribosomal docking (Patterson-Fortin et al., 2013).

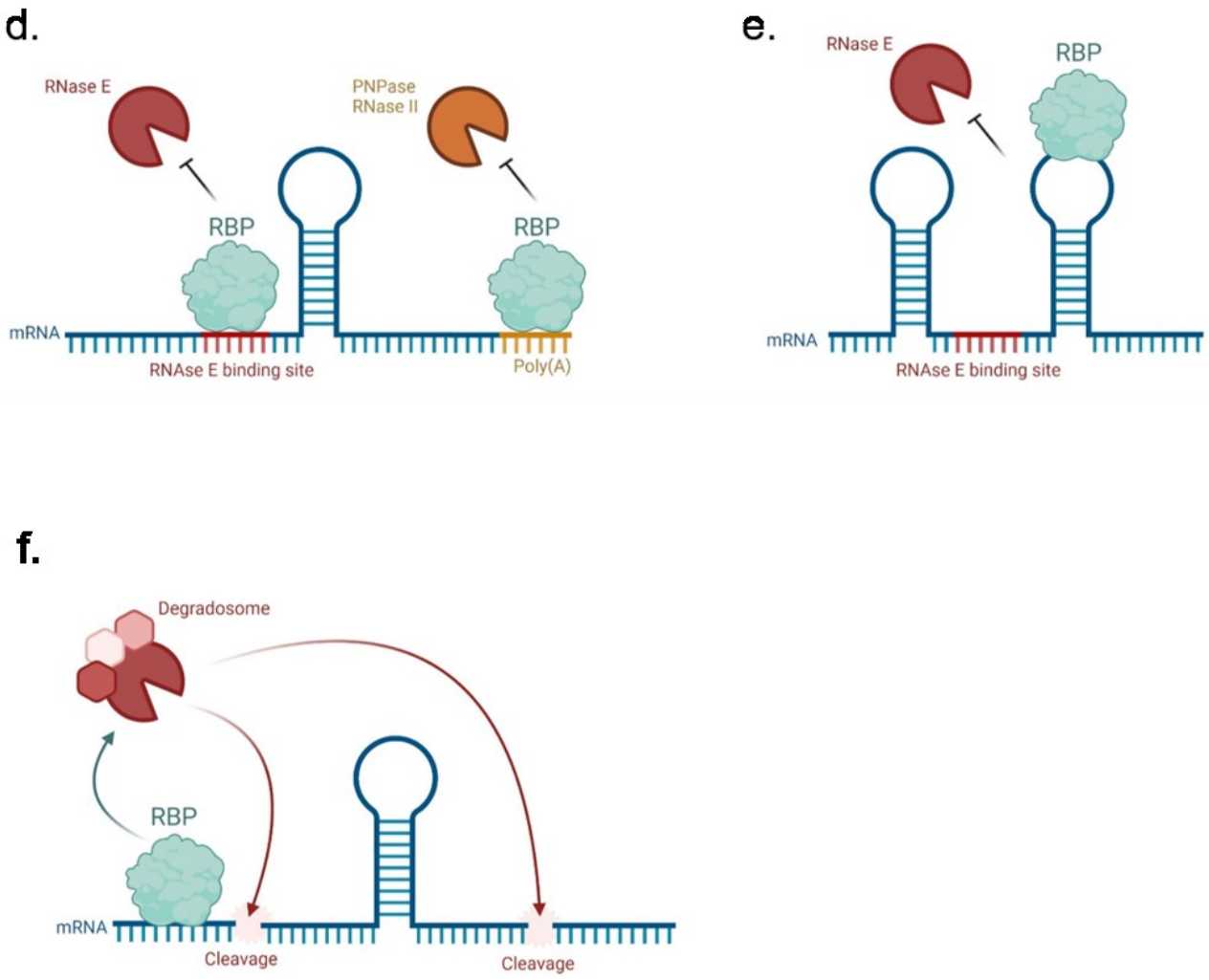


Figure 1.2.(continued). d.,e.& f. RBP interaction affects transcript susceptibility to ribonucleases. d. Binding of the RBP to the RNase E binding site (red) or the Poly(A)-tail recognised by PNPase and RNase II (orange) inhibits docking of the exonuclease (red) and the endonucleases (orange) and protects the transcript from degradation. e. RBP interaction alters the mRNA conformation, burying the ribonuclease interaction site and thus inhibits transcript cleavage. f. RBP binding signals for recruitment of the degradosome complex that facilitates cleavage of the transcript (Bruce et al., 2017; De Lay et al., 2013).

g.

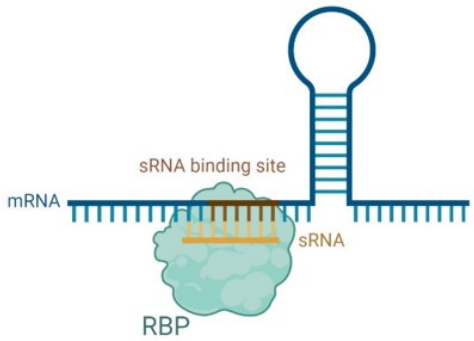


Figure 1.2.(continued). g. RBP acts as a chaperone for a regulatory sRNA. RBP facilitates base-pairing (often imperfect) between the transcript and the sRNA (yellow) which results in sRNA-facilitated regulatory processes described in section 1.1.2 of this thesis. Additionally, interaction of the RBP with the regulatory sRNA often stabilises the sRNA and thus indirectly affects mRNA translation or maintenance (Chao et al., 2012).

1.1.3.1 CsrA

One of the ways RBPs modulate post-transcriptional expression is by altering the access of the 30S ribosomal subunit to the ribosomal binding site (RBS), thus either inhibiting or activating translation initiation. Mechanistically this is achieved either by direct obstruction of the RBS or instigation of a change in the secondary structure of the target mRNA that either exposes or buries the RBS (Figure 1.2abc). This mode of action is widely observed in a well-characterised RBP carbon storage regulator A (CsrA) of *E. coli* and its homologues in other bacteria (Finn et al., 2014). CsrA is a global post-transcriptional regulatory factor that absence of which was shown to affect expression of up to 10% of genes in *Pseudomonas* and *Salmonella* species (Burrowes et al., 2006; Lawhon et al., 2003). The overarching characteristic of CsrA is its repression of stationary phase processes and enhancement of the factors involved in the exponential growth phase of bacterial proliferation (Dubey et al., 2003; Sabnis et al.,1995). CsrA was shown to play a role in crucial colonisation processes such as biofilm formation, quorum sensing, motility, adherence to host cells, carbon metabolism and stress responses (Dugar et al., 2016; Jackson et al., 2002; Pannuri et al., 2011; Sabnis et al., 1995; Yakhnin et al., 2011; Yakhnin et al., 2013). Analysis of the known CsrA targets and the binding sequences allowed for determination of a CsrA binding consensus, which was shown to be RUACARGGAUGU. Additionally, a hairpin conformation with the GGA codon at the loop of the mRNA structure was shown to act as another transcript signature targeted by CsrA (Dubey et al., 2005). CsrA functions as a homodimer with two identical RNA binding surfaces and dual-site binding of CsrA to its RNA targets was reported in multiple studies (Duss et al., 2014; Mercante et al., 2009).

The CsrA target consensus sequence closely resembles that of the ‘optimal’ Shine-Dalgarno, AGGAGG, which explains the preferential mode of CsrA-mediated regulation that focuses on occlusion of the RBS

(Jonas et al., 2008; Yakhnin et al., 2011). For example, the first reported case of CsrA-mediated control was the repression of translational activation of glucose-1-phosphate adenylytransferase-encoding *glgC* mRNA that was shown to be achieved by binding of the CsrA at four sites in the *glgC* transcript, one of which overlaps with the Shine-Dalgarno sequence (Mercante et al., 2009). As mentioned, multiple-site RNA binding is canonical for CsrA interactions, however CsrA was shown to modulate translational activation by binding to a single mRNA region around the Shine-Dalgarno. Interestingly, this is the case in CsrA-driven repression of translation of *hfq* mRNA, which encodes another regulatory RBP that often acts in opposition to CsrA (Baker et al., 2007). Another regulatory mode of CsrA involves alteration of mRNA secondary structure, facilitated by binding of the RBP. An example of such a mechanism is *Pseudomonas aeruginosa* RsmA translational repression of the *psl* mRNA that encodes a biofilm matrix component. Interaction with RsmA was shown to arrest the *psl* transcript in a conformation that sequesters the ribosomal binding site (Irie et al., 2010). Several studies reported that CsrA induces translation of mRNAs that encode factors governing bacterial pathogenesis and proliferation. For example Yakhnin et al., 2013 indicated that CsrA positively regulates the expression of the major transcriptional regulator of the flagellar T3SS in *E. coli*, FlhDC. Binding of CsrA to the *flhDC* mRNA was shown to protect it from the RNase E cleavage at the 5' end of the transcript (Yakhnin et al., 2013). Additionally, CsrA is able to positively regulate the mRNA translation by alteration of the mRNA secondary structure. In *E. coli* the expression of the *moaABCDE* operon that encodes the molybdenum cofactor, an important metallofactor for the enzymatic activity, was shown to be regulated by CsrA-mediated destabilisation of a repressive riboswitch structure (Patterson-Fortin et al., 2013). Furthermore the CsrA homologue in *Pseudomonas aeruginosa* was shown to regulate formation of the biofilm by modulating the expression of phenazine biosynthetic gene cluster *phz2*.

This is done by destabilisation of a hairpin structure around the Shine-Dalgarno region of a *phz2* mRNA that was shown to obstruct ribosomal access (Ren et al., 2014).

Several protein and sRNA antagonists have been reported with arguably the most notable ones being the CsrB and CsrC sRNAs. Homologues of those sRNAs are widely conserved throughout pathogenic Gram-negative bacteria (Baker et al., 2007; Fortune et al., 2006; Kay et al., 2005; Lenz et al., 2005; Parker et al., 2017; Sterzenbach et al., 2013). CsrBC sRNAs act as CsrA sponges sequestering the RBP and deviating it from its regulatory function on target transcripts (Romeo et al., 2013). Interestingly Hfq, with its most prominent function as an sRNA chaperone, was shown to stabilise CsrBC and enhance binding of the sRNAs to CsrA (Sonnleitner et al., 2006; Sorger-Domenigg et al., 2007).

1.1.3.2 Hfq

Some of the best characterised RBP-dependent regulatory events involve RBPs that act as chaperones for ncRNAs that localise the ncRNAs to the mRNA and facilitate RNA duplex formation that result in altered translation or stability of the transcripts (Figure 1.2g). In last couple of decades a nucleic-acid binding protein called Hfq has been extensively studied for its ability to bind imperfectly base-paired ncRNAs to the mRNA (Valentin-Hanses et al., 2004; Waters et al., 2009). Hfq was first documented in 1968 in *E. coli* as a host factor for bacteriophage Qβ RNA replication (Franze de Fernandez et al., 1968). Since then, it has been shown to be essential for bacterial proliferation in Gram negative bacteria as *hfq* deletion mutants exhibit growth impairment, sensitivity to environmental stresses and a dramatic decrease in virulence (Chao et al., 2010; Kulesus et al., 2008; Tsui et al., 1994). Studies into the structure of Hfq revealed its homohexameric nature together with witch its sequence classifies it as a member of

Sm/Lsm family of proteins that regulate RNA metabolism in eukaryotes. Today it is the RNA-chaperone and translational regulator function that Hfq is most well known for. The ring-like Hfq structure provides four solvent exposed regions: the proximal face, distal face, rim and C-terminal tail that were shown to facilitate Hfq-RNA interactions. The proximal face is associated with affinity for polyU sequences and the distal face binds A-rich sequences. Although of secondary importance, the rim and C-terminal tails were also attributed a function in Hfq-RNA binding.

Facilitation of the ncRNA-mRNA interaction by Hfq is multifaceted and was shown to involve changing the sRNA and mRNA secondary structures, targeting sRNAs to the transcript, neutralisation of negative charges at RNA pairing, stimulation of first base nucleation and stimulation of further annealing of the sequences (Fender et al., 2010; Holmqvist et al., 2010; Soper et al., 2010; Schu et al., 2015). The canonical Hfq-mediated post-transcriptional regulatory mode of action is guidance of the sRNA to the RBS and facilitation of mRNA-sRNA binding that inhibits ribosomal docking and translation initiation. For example three Hfq-dependent sRNAs, MgrR, RyhB and McaS were shown to bind to the 5'UTR region of the *glrA* and *glrR* mRNA that encode the respectively positive and negative regulators of Enteropathogenic *E. coli* (EPEC) Type 3 Secretion System (T3SS) micro-injection machinery, essential for bacterial adherence to the epithelial cells. The interactions lead to obstruction of the transcripts' Shine-Dalgarno sites, which prohibits ribosomal docking and translation initiation of one or the other T3SS regulator and thus dictate the production of the machinery and EPEC colonisation (Bhatt et al., 2017). Another Hfq-mediated regulatory mechanism involves alteration of the mRNA secondary structure upon sRNA-Hfq binding that leads to obstruction or exposing of the RBS. Production of the RpoS, a major regulator of *E. coli* virulence genes and enhancer of bacterial survival against host defences, was shown to be regulated by an Hfq-dependent

sRNA DsrA (Soper et al.,2009). This is mediated by a change in *rpoS* mRNA structure around the translation initiation region that was shown to be majorly dependent on Hfq (Peng et al., 2014; Soper et al., 2011). Restructuring of the transcript conformation facilitating a negative effect on mRNA translation by Hfq-sRNA binding was reported in regulation of biofilm formation factor, DgcM. Hfq-mediated structural change of the DgcM transcript allows for access of the Hfq-chaperoned sRNAs OmrA and OmrB to the early coding region of the transcript and repression of ribosomal binding (Hoekzema et al., 2019). Interactions with Hfq-sRNA complex were reported to facilitate the cleavage of the target mRNAs. Recent study by Lalaouna et al., 2021 showed that induction of rapid RNase E-dependent decay of *sodB* mRNA that encodes superoxide dismutase, which governs the iron-dependent metabolic processes in *E. coli,* is the primary function of Hfq in the interaction between the *sodB* transcript, Hfq and regulatory RyhB sRNA. Additionally, Hfq-mediated interaction between a MavR sRNA and the *eutR* transcript that encodes a transcription regulator of ethanolamine metabolism-related factors, that play a role in colonisation by Enterohaemorrhagic *E. coli*, was shown to stabilise the mRNA. This was determined to be through binding of the sRNA to the mRNA region targeted by the RNase E that inhibits RNase E interaction and transcript cleavage (Sauder et al., 2021). Additionally, Hfq has a stabilising function on the mRNA-regulating sRNAs that was shown to protect them from degradation and thus have an indirect effect on regulation of the transcripts affected by the stabilised sRNAs (Udekwu et al., 2005; Moon et al., 2011; Chao et al., 2012). Conversely, destabilisation of regulatory sRNAs by Hfq has also been reported (Han et al., 2019).

Hfq has been found to directly regulate transcript stability by promoting polyadenylation of the mRNA that targets the transcript for 3'-5' degradation. Direct binding of Hfq to 5'UTR of the transcript and obstruction of ribosomal access without a need for interacting sRNA was reported by Chen et al., 2017. Hfq was also recently shown to bind

to and process the pre-16S rRNAs and regulate the synthesis of mature 70S ribosomes and thus regulate mRNA translation by modulating the availability of functional ribosomes (Andrade et al., 2018). tRNAs were indicated as another Hfq substrate with processing affecting translational fidelity (Lee et al., 2008).

In Gram negative pathogenic bacteria Hfq was shown to govern production of crucial virulence factors such as toxin production in *Yersinia enterocolitica*, *Pseudomonas aeruginosa* and *Vibrio parahaemolyticus* (Nakao et al., 1995; Nakao et al., 2008; Sonnleitner et al., 2008). Quorum sensing that influences the timing of virulence gene expression was also shown to be regulated by Hfq and accompanying sRNAs in *Vibrio cholerae* and *Pseudomonas aeruginosa* (Bradill et al., 2011; Sonnleitner et al., 2006). Production of the Type 3 Secretion System (T3SS), a major colonisation and virulence factor, was shown to be regulated by the Hfq in *Salmonella*, *Escherichia*, *Shigella*, *Vibrio* and *Pseudomonas* species in a positive or negative manner depending on the bacterial background (Hansen et al., 2009; Mitobe et al., 2009; Shakhnovich et al., 2009; Wang et al., 2018). The Hfq-mediated regulation and other factors controlling T3SS in pathogenic *Escherichia coli* are covered in more detail in the following sections of this thesis.

1.2 Enterohaemorrhagic *Escherichia coli*

Escherichia coli is a Gram-negative, facultatively anaerobic, coliform bacterium from the Enterobacteriaceae family, primarily inhabiting the lower intestines of humans and other warm-blooded animals. It is bile-tolerant, lactose fermenting, nonfastidious organism that is easily cultured (Adamu et al., 2014; Blount et al., 2015). It was first described in 1885 by Dr. Theodor Escherich, an Austrian-German pediatrician that isolated the rod-shaped bacterium from infant stools, while studying the causes of neonatal dysentery (Escherich, 1885). Since then *E. coli* has become the most studied organism on the planet, it shaped bacterial genetics and helped give rise to an entire field of molecular microbiology (Blount et al., 2015). Although most *E. coli* are harmless and useful commensal organisms, acquisition of various virulence genes equipped the bacteria with a plethora of pathogenic attributes. This resulted in emergence of a number of commonly recognised *E. coli* pathotypes: enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC), enteroaggregative *E. coli* (EAEC), uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli* (NMEC) and the subject of this thesis, enterohaemorrhagic *E. coli* (EHEC) (Kaper et al., 2004).

EHEC is officially considered an emerging food-borne pathogen and a causative agent of bloody diarrhoea and hemolytic uremic syndrome (HUS) worldwide, with significant epicenters located in Northern America and certain parts of the United Kingdom (Goldwater et al., 2012). Although in many cases the disease caused by EHEC is self-limiting, approximately 8% of the cases will develop life-threatening HUS that results in 5% mortality and 25% chance of permanent renal damage (Garg et al., 2009). The main defining feature of this *E. coli* pathotype is production of a potent virulence factor, a phage-encoded cytotoxin called

Shiga toxin or Vero toxin named after its identity to the toxin produced by *Shigella dysenteriae* type 1A and the capacity of these toxins to kill kidney epithelial cells derived from the 'Vero' lineage (Croxen et al., 2010). In 1982, an outbreak of 43 severe bloody diarrhoea cases in United States, Michigan and Oregon, as well as 11-cases of HUS in 1983 (Midland UK) announced the arrival of pathogenic *E. coli* serotype O157 and EHEC on the list of emerging infectious diseases (Welinder-Olsson ety al., 2005). In the United States *E. coli* O157 is estimated to result in 63,000 illnesses and 2100 hospitalisations annually, costing the economy around 405 million USD (Fatima et al., 2021). In the United Kingdom the incidence of *E. coli* O157 human infections was shown to remain stable at 1029 and 1039 cases annually in 1995 and 2012, indicating persistence of the pathogen in the reservoir population (Pennington et al., 2014). According to Public Health England in 2018 1553 confirmed cases of Shiga-toxin producing *E. coli* including 607 O157 EHEC were confirmed in England and Wales alone (Public Health England, 2020). Although the *E. coli* O157 is the better known and the more studied EHEC serotype, non-O157 strains are a more common cause of haemorrhagic colitis and have the potential to cause large epidemics worldwide. Unfortunately, unlike the O157 serotype which can usually be distinguished from most *E. coli* strains by its inability to ferment sorbitol and lack of β-glucuronidase enzyme, many of the non-O157 shiga toxin positive strains that are a threat to human health have no defying phenotypic characteristics that can be reliably employed to isolate and identify those pathogens by standard screening (Welinder-Olsson et al., 2002, Valilis et al., 2018). This has hindered successful tracking of non-O157 related disease epidemiology and containment. A systematic review on non-O157 *E. coli* by Valilis et al., 2018, highlights the burden of 674 outbreaks reported since 1995, involving strains such as O26:H11, O45, O103:H25, O104:H4, O111:H8, O121, and O145:NM.

EHEC is a zoonotic pathogen that reaches its human host through consumption of contaminated goods or direct contact with infected

livestock, there is also the potential for human-to human spread, especially in nurseries and other childcare settings (Figure 1.3). Ruminants are a primary carrier and an asymptomatic reservoir of EHEC. In the European Union, 2-13.5% of cattle tested positive for EHEC between 2007 and 2011. Study by Henry et al., 2017, estimated that the prevalence of *E. coli* O157 in Scottish, English and Welsh cattle farms was 21-23% between 2014 and 2015. Transmission of EHEC is mediated by shedding of the pathogen in faeces. The fraction of EHEC carrying animals that shed more bacteria (over 10^3 CFU/g) than others are called 'super shedders'. Despite the low ratio of super shedders detected in cattle populations and a limited time period of increased shedding of the bacteria, up to 95% of excreted EHEC can be traced back to super shedders (Murphy et al., 2016; Omisakin et al., 2003; Chase-Topping et al., 2007). Dissemination of the pathogen to food products occurs through seeping of the contaminated manure or water to food crops and faecal contamination of animal produce during milking or slaughter. Widespread contamination of the environment with the pathogen and its ability to survive outside of the host for more than seven days, provides a continual source of EHEC and complicates the prevention of its spread (Wang et al., 1996). Although the majority of EHEC outbreaks are related to consumption of undercooked meat and dairy products, with around 75% of *E. coli* O157:H7 cases originating from contaminated ground beef, raw vegetables and salad plants have been increasingly implicated in human disease. In 1996, an outbreak in Japan had almost 8000 cases and originated from EHEC contaminated radish sprouts. Fresh lettuce and spinach were associated with spread of infection through the United States and Sweden in the early 2000s. Globalisation of food distribution exacerbates the risk of EHEC spreading to a large and geographically widespread proportion of the population, creating obstacles in management and tracking of outbreaks (Goldwater et al., 2012).

Management of EHEC infections needs to address a multitude of life-threatening disease symptoms (Davis et al., 2013). Treatment of the disease includes general supportive approaches, use of probiotics and toxin neutralisers, intravenous fluid administration and in extreme cases dialysis and blood transfusion (Grisaru, 2014). Components of EHEC's major virulence factor the Type 3 Secretion System has shown potential as an effective target in vaccine studies in cattle (McNeilly et al., 2015). However, no commercial vaccine against EHEC is currently available and treatment with antibiotics is discouraged as conventionally used drugs were shown to increase the likelihood of HUS (Soysal et al., 2016). Therefore, development of novel anti-virulence treatments could be an effective strategy for future research targeting EHEC infections (Totsika, 2016).

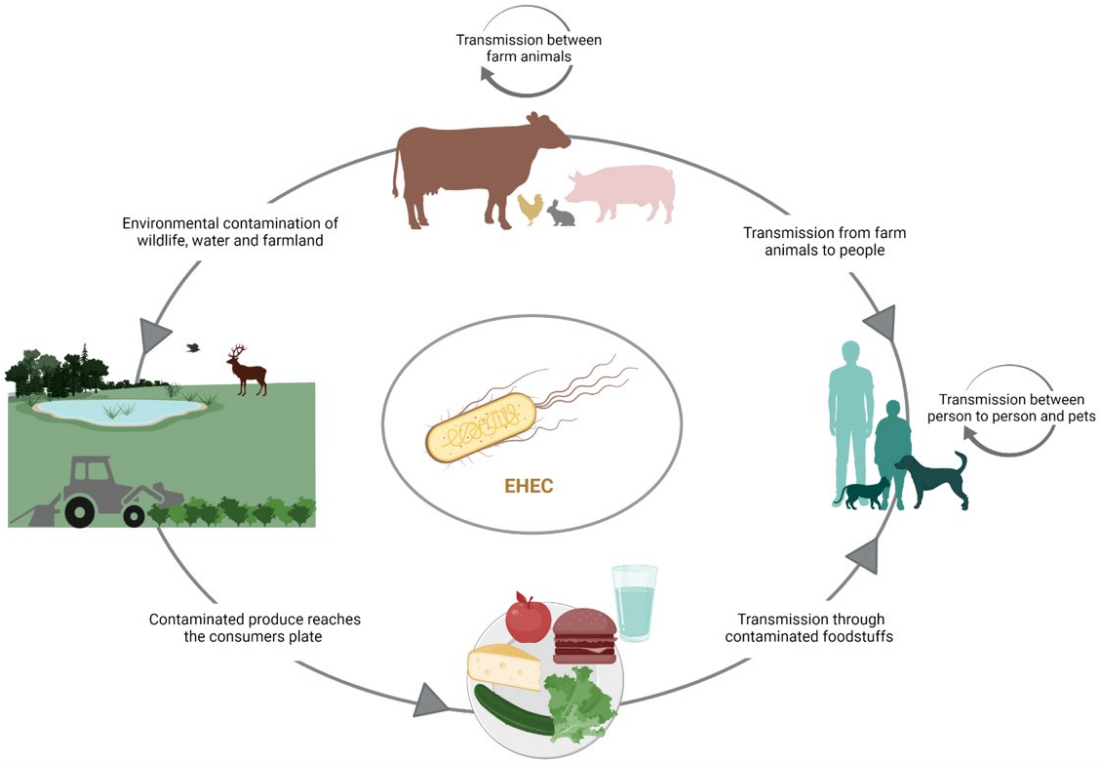


Figure 1.3. Farm-to-fork cycle of EHEC transmission. The main amplification of EHEC is through circulation of the pathogen in farm animals with cattle being the major asymptomatic reservoir. EHEC can reach the consumers plate directly through meat and dairy obtained from infected animals or through produce such as leafy greens that get contaminated with EHEC through contaminated farm waste that reaches the environment and water supplies used for agriculture. Spread of EHEC through the contaminated environment to wildlife can further amplify the circulation of the pathogen. Additionally contact with infected farm animals or contaminated animal waste can result in direct transmission of EHEC to a human host. Transmission from person to person or companion animals is also possible (Bolton et al., 2011; Quah et al., 2017).

1.2.1 **EHEC pathogenesis**

The infectious dose of EHEC is as low as 1 to 100 colony forming units (CFU). Apart from stress response mechanisms that manage survival of EHEC in extreme acidic environments, heat, cold and osmotic shock, there are a multitude of virulence factors that contribute to EHEC pathogenicity (Mellies et al., 2007). These virulence factors are primarily encoded on chromosomal pathogenicity islands, plasmids and prophage regions, and are instrumental in the ability of the pathogen to successfully colonise the host and/or cause disease. The presence of specific combinations of virulence factors may determine the risk of developing severe symptoms (Nguyen et al., 2012; Welinder-Olsson et al., 2002). The variability in virulence factors employed even by different EHEC strains is vast. The more strain specific virulence factors were shown to be encoded on the pO157 of *E. coli* O157 and the High-Pathogenicity Island present in many enterobacteria including certain non-O157 *E. coli* (Jiang et al., 2021). In *E. coli* O157:H7 the pO157 virulence plasmid encodes factors such as haemolysin, a bifunctional catalase-peroxidase, a Type 2 Secretion System, which was shown to promote EHEC adherence to gastrointestinal epithelium, and a secreted serine protease, EspP, that cleaves human coagulation factor V, metalloprotease StcE and a ToxB factor that were shown to promote colonisation and are thought to have an immunosuppressive function (Ho et al., 2008; Lim et al., 2009). Although the extent to which pO157 influences EHEC pathogenicity is still not fully understood, high prevalence of pO157 in clinical isolates was noted (Lim et al., 2010). Two absolutely crucial virulence strategies that actually define ‘traditional’ EHEC are production of Shiga toxin (Stx) and formation of attaching and effacing (AE) lesions requiring the micro-injection machinery, both of which are described in more detail below (Figure 1.4).

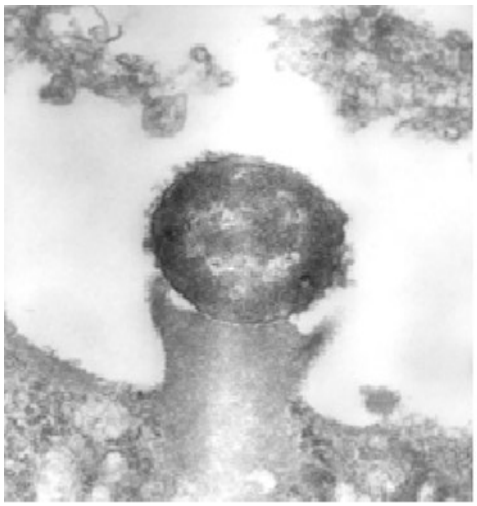


Figure 1.4. Electron microscopy image of EHEC attached to the enterocyte and with an induced attaching and effacing lesion (Naylor et al., 2003). Effector proteins delivered to the host cell through the T3SS facilitate intimate attachment between EHEC and the enterocyte by inducing host cell cytoskeleton rearrangements that result in the formation of an actin-dense cuplike pedestal and localised destruction of brush border microvilli (Gaytan et al., 2016).

Shiga toxin (Stx) is a main virulence factor and a defining feature of the EHEC pathotype. Production of the toxin in the colon results in local epithelium damage and possible spread of Stx through the bloodstream to the kidneys, where its action can result in development of HUS (Schuller et al., 2012). Stx is a family of prophage-encoded cytotoxins characterised by a high degree of diversity. Based on antigenic characterisation the Stx family members can be categorised into two groups Stx1 and Stx2, that are further divided into Stx variants, Stx1a, Stx1c and Stx1d and Stx2a, Stx2b, Stx2c, Stx2d, Stx2e, Stx2f, Stx2g (Melton-Celsa, 2014). The subgroups associated with the highest disease severity are Stx2a, Stx2c and Stx2d. This observation was confirmed by in vitro and in vivo studies showing higher potency of Stx2a and Stx2d specifically as compared to other Stx1 and 2 subtypes. Combinations of Stx variants from both of the major Stx groups can be found in EHEC isolates (Melton-Celsa et al., 2014). In the United Kingdom the most prominent strain of O157 EHEC encodes both Stx2a and Stx2c, this genotype is epidemiologically associated with super-shedding and has been linked to increased incidence of pathogen spread among livestock and zoonotic transmission. Recent transmission study by Fitzgerald et al., 2019 confirmed the association of Stx2a to the super shedder phenotype by showing an increase in pathogen transmission and significant increase in EHEC excretion in cattle infected with $Stx2a^{+}$ strain in lower dose, in exposure conditions closely mimicking natural spread of the pathogen as compared with an isogenic background.

The induction of Stx expression is generally coupled with induction of the bacteriophage by the SOS response. Classical cell lysis leading to bacterial death is the only conclusive toxin release system, although other delivery pathways such as packaging into outer membrane vesicles (OMVs) were proposed (Schüller, 2011). Stx mode of action involves inhibition of protein synthesis and cell apoptosis. Stx

polymeric structure consists of two major subunits catalytic subunit A bound to a pentamer of B subunits. B subunit (approximately 7.7 kDa) pentamer acts as a ligand for endothelial globotriaosylceramide-3 (Gb-3) cellular receptors expressed by Paneth cells in the intestinal mucosa and kidney epithelium (Melton-Celsa, 2011, Tarr et al., 2005). Distribution of Gb3 receptors dictates the localisation of toxin-induced pathology. Binding of the toxin to the cell surface allows for absorption of Stx and its systemic dissemination through the bloodstream. There is evidence suggesting that initial damage to the mucosal epithelium caused by A/E lesion formation aids in Stx endocytosis (Sandvig et al., 2010). The lack of vascular Gb3 receptors in the gastrointestinal tract of cattle provides an explanation for the tolerance of the infection and its asymptomatic outcome in the EHEC reservoir (Pruimboom-Brees et al., 2000).

The cytotoxic activity of Stx is mediated by subunit A (approximately 32 kDa), an RNA N-glycosidase enzyme, which cleaves the N-glycosidic bond at A-4324 in 28S rRNA and inactivates the eukaryotic ribosome, which in turn blocks protein synthesis and promotes host cell apoptosis. The aspects of inflammatory response against Stx are still debated. Few studies detected an IL-8 response to Stx release, however other research argues that this activation is suppressed by anti-inflammatory factors expressed by EHEC in an infection setting (Bellmeyer et al., 2009; Jandhyala et al., 2008). Cytotoxic action of Stx and the host response to endothelial cell damage can lead to destruction of the endothelial and vascular lining that is characteristic of EHEC haemorrhage.

The selective advantage of Stx expression in cattle is poorly understood, cell culture and mouse model studies linked expression of Stx with reorganisation of receptors recognised by the bacteria during initial adherence. Stx-mediated killing of grazing protozoa was proposed to aid in EHEC colonisation, however this relationship is not

conclusive as opposing results exist (Schmidt et al., 2016; Steinberg et al., 2007). A new phenotype observed by Fitzgerald et al. related Stx expression with changes in ileal proliferation. Expression of Stx2a and Stx2c prevented regeneration of epithelium and led to decreased budding of bovine ileal organoid culture, potentially leading to increase in persistence of the attached bacteria and thus an increase in the bacterial load observed in super-shedding cattle (Fitzgerald et al., 2019). Additionally other factors encoded on Stx and other prophages, such as sRNAs and regulatory proteins, were shown to impact EHEC colonisation, granting it with a growth advantage in bovine rectal mucus and orchestrating T3SS expression (Tree et al., 2014; Xu et al., 2012).

Due to the unarguable and critical involvement of Stx in EHEC pathogenesis, targeting the toxin via anti-virulence blocking compounds is an attractive avenue for treatment of EHEC inflicted disease (Huerta-Uribe et al., 2016).

Following sections addresses another hallmark of EHEC pathogensis the A/E lesions and specifically the machinery involved in their creaction, the Type 3 Secretion System.

1.3 Bacterial Secretion Systems

One of the major bacterial attributes is the ability to transport proteins and other molecules out of the cell into the milieu, into the outer membrane of the bacterium or inject them directly into the host cell or other bacterial cells. Secretion of these substrates is necessary for environmental sensing, efflux of harmful substances, colonisation and pathogenicity among others. More than one-third of the proteins synthesised are targeted outside of the bacterial cytoplasm, indicating the need for fine-tuned delivery structures, to carefully manage this flow of substrates (Orfanoudaki et al., 2014). Those complexes are elaborate macromolecular machineries called secretion systems. In Gram negative bacteria at least ten secretion systems have been recorded to date that can be classified into single-membrane-spanning and double-membrane-spanning systems. The double-membrane spanning type 1, 2, 3, 4 and 6 secretion systems are often extremaly complex machineries (Figure 1.5). Type 1 Secretion System (T1SS) resembles the ATP-binding cassette transporters and is mainly involved in Sec-independent secretion of small unfolded substrates such as adhesins, heme-binding factors, enzymes and toxins to the extracellular milieu (Symmons et al., 2009). Pathogenic bacteria can have multiple T1SSs dedicated to transport of one or few specific proteins. T1SS was shown to be important in virulence of multiple pathogenic bacteria, for example the common cause of nosocomial infection, *Serratia marcescens,* was shown to transport a haemophore, HasA, via the T1SS (Letoffe et al., 1996; Delepelaire et al., 2004). Type 2 Secretion Systems (T2SS) transport folded protein from the bacteria to the extracellular matrix. The initial step of the substrate transport includes Sec or Tat-dependent transport of the protein to the periplasm, where it is targeted to the T2S channel located in the outer membrane and subsequent extracellular secretion (Korotkov et al., 2012). Virulence factors transported through the

T2SS include cholera toxin, which facilitates watery diarrhoea, the hallmark of *Vibrio cholerae* infection (Sandkvist et al., 1997). Type 3 Secretion System (T3SS) is the core of the flagella involved in bacterial motility and injectisome machinery that connects bacteria to the host cell through formation of the pore at the eukaryotic cell membrane. Unfolded effector proteins are transported through the injectisome 'needle' in an ATP-dependent manner and once in the host cell modulate cellular functions and enable bacterial colonisation. *Salmonella enterica* possesses two injectisome T3SSs (T3SS-1 and T3SS-2) important for different stages of colonisation (Marcus et al., 2000). The effector proteins secreted through T3SS-1 enable *Salmonella* to initiate contact of bacteria with the host cell through secretion of effector proteins such as SipA, SipC, SopB, SopE, SopE2, and SptP that facilitate changes in the eukaryotic cell cytoskeleton that results in production of so-called ruffles and engulfment of Salmonella into the host cell in the *Salmonella* containing vacuoles (SCVs) (McGhie et al., 2001). T3SS-2 is active in the SCVs where it facilitates delivery of the effector proteins from the intracellular bacteria throughout the SCV membrane into the host cell cytoplasm where they inhibit the phagolysosome formation and manipulate SCV trafficking and maturation and promote *Salmonella* survival and replication (Bakowski et al., 2008). Type 4 Secretion System (T4SS) is related to bacterial DNA conjugation systems and is able to transport variety of complex substrates such as protein-protein and DNA-protein complexes. T4SSs can connect to the eukaryotic cell through formation of the pore and this was shown to govern processes such as conjugative transfer of DNA, DNA uptake and release and translocation of the complexes directly into host cell (Sgro et al., 2019). An interesting example of the T4SS-mediated process is with *Neisseria gonorrhoeae* and DNA uptake, which promotes acquisition of new virulence genes (Hamilton et al., 2005). Type 6 Secretion System (T6SS) is machinery primarily utilised for competition between

bacteria, however it can also be utilised to target eukaryotic host cells (Coulthurst, 2019; Hachani et al., 2016). T6SS uses a contraction mechanism closely related to that of bacteriophages that shoots a puncturing structure out of the attacking bacteria and through the opponent's membrane (Brackmann et al., 2017). The effector proteins are bound to the puncturing structure delivered to the rival cells and thus are translocated in a contact-dependent manner. The T6SS-delivered effectors were shown to govern processes such as outcompeting of commensal microflora, establishment of stable colonies in separate niches and alteration of host organism processes such as intestinal movement and immunity (Anderson et al., 2017; Chatzidaki-Livanis et al., 2016; Fu et al., 2018; Logan et al., 2018; Speare et al., 2018). Interestingly the attacking bacteria evolved a toxin-antitoxin immunity defence system in order to neutralise the harmful effect of the produced antimicrobial effector compounds. This also protects the genetically identical cells punctured with the T6SS from bactericidal outcomes. In addition, a recent study by Trunk et al., 2018 identified antifungal effectors delivered by Serratia marcescens T6SS that were shown to kill or inhibit *S. cerevisiae* and *Candida* species. T6SS was also shown to have a role in uptake of specific metal ions important for bacterial survival under conditions such as oxidative stress (Wang et al., 2015).

Secretion systems are especially prominent in pathogenic bacteria, which can express a combination of those nanomachines throughout their lifespan (Abrusci et al., 2014). Which secretion systems are going to be utilised is dependent on the surrounding conditions. For example in Enterohaemorrhagic *E. coli*, expression of flagella (flagellar Type 3 Secretion System) is employed in the free swimming stage of the infection and is switched to expression of Type 3 Secretion System injectisome, responsible for facilitation of adherence to epithelium, when host cell contact is established and colonisation occurs (Kirkpatrick et al., 2012). Due to the specificity of expression of some of

these secretion systems in bacterial pathogens, antimicrobials are being developed against these systems to augment our current repertoire of antibiotics.

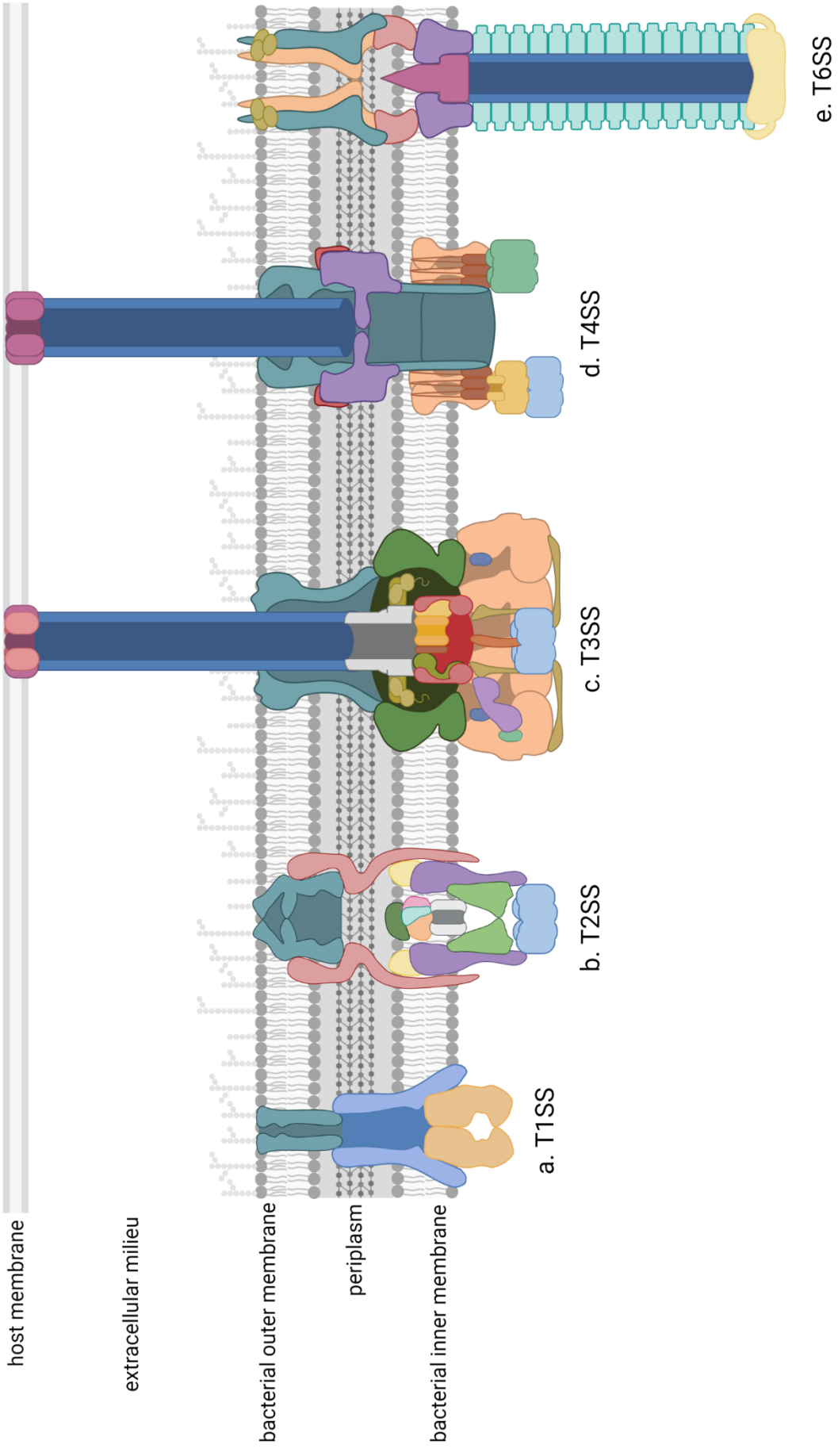

host membrane
extracellular milieu
bacterial outer membrane
periplasm
bacterial inner membrane
a. T1SS
b. T2SS
c. T3SS
d. T4SS
e. T6SS

Figure 1.5. Architecture of double-membrane spanning secretion systems in Gram-negative bacteria. **a.** Type 1 Secretion System (T1SS), **b.** Type 2 Secretion Systems (T2SS) **c.** Type 3 Secretion System (T3SS), **d.** Type 4 Secretion System (T4SS), **e.** Type 6 Secretion Sysem (T6SS). T1SS and T2SS secrete their substrates into the extracellular milieu. T3SS, T4SS and T6SS export the effector proteins directly to the host cell. T3SS and T4SS achieve it trough formation of a pore structure at the host membrane, while T6SS acts as a 'spear' that propelles itself from the bacterial cell and pierces the host membrane.

1.3.1 Type 3 Secretion System

This subsection focuses on Type 3 Secretion System in Gram negative bacterial pathogens with a specific focus on the EHEC T3SS. Type 3 Secretion Systems (T3SS) are at the core of, arguably, the most sophisticated nanomachines employed by EHEC, the flagellum and the injectisome (Figure 1.6). Bacterial flagella are primarily utilised as swimming and swarming propellers, which can spin clockwise or anti-clockwise at up to 10000 r.p.m.. Those rotary nanomotors are crucial for bacterial chemotaxis - migration of the cells towards attractant chemicals like nutrients and away from repellents such as toxins. Locomotion is only one of the flagella functions and the organelle was shown to also play a role in bacterial adhesion and environment sensing, biofilm formation, invasion and possibly host immune evasion (Kirov, 2003). Flagellum is considered to be an ancestor of the closely related injectisome machinery that assembly of is the main focus of this thesis. The injectisome delivery complex is commonly employed by EHEC and other Gram-negative pathogenic bacteria to directly translocate effector proteins into the host cell, where they can modulate the behaviour of the host in favour of the pathogen (Diepold et al., 2015). Expression of the injectisome is crucial for a plethora of bacterial processes such as: intimate attachment of enteropathogenic *E. coli* and enterohaemorrhagic *E. coli* to host enterocytes; uptake of *Salmonella* and *Shigella* into nonphagocytic cells and their intracellular survival and replication; modulation of immune response that hinder production of antibodies and induce the interleukin-10 response, which contributes to persistence of *Bordetella* in the respiratory tract; cytotoxic and enterotoxic pathology of *Pseudomonas* and *Vibrio parahaemolyticus* respectively; and maturation of *Chlamydia* from the metabolically inert elementary body state to active reticulate bodies, to name a few (Betts-Hampikian et al., 2010; Figueira et al., 2013; Gaytan et al., 2016; Gendrin et al., 2012; Ham et al., 2012; Nicholson et al., 2013). A more

extensive summary on the relation between injectosome and disease can be found in a comprehensive review by Coburn et al., 2007.

1.3.1.1 Environmental stimuli that govern T3SS expression

The stimulus for injectisome expression varies greatly between the bacterial species and is dependent on pathogen lifestyle. Environment-driven expression of the T3SS of EHEC allows it to establish a niche in the extremely diverse gastrointestinal environment and successfully colonise the host. In EHEC, quorum sensing of host-produced hormones, epinephrine and norepinephrine, by the histidine kinase response regulator systems QseBC and QseEF and subsequent activation of Lys-R type regulator QseA was indicated to transcriptionally activate T3SS expression in the large intestine (Carlson-Banning et al., 2018; Kendall et al., 2007; Lustri et al., 2017; Moreira et al., 2016). Dietary signals such as availability of mucin-derived carbohydrate N-acetyl-galactosamine and iron were also shown to alter expression of EHEC T3SS (Le Bihan et al., 2017; Sanchez et al., 2018). Apart from environmental signals that induce colonisation, there are factors which indicate niches that are not suitable for EHEC proliferation. D-serine, a metabolite prevalent in urinary tract and found in low concentrations in lower intestine, was shown to inhibit expression of the T3SS in EHEC (Connolly et al., 2016). Environmental stressors such as low pH and oxidative stress are known to repress T3SS production. Acid tolerance sensor GadE was shown to inhibit LEE transcription in a low pH environment. This is opposed by the GadE repressor of the global regulator of virulence A (GrvA), the expression of which depends on another regulator, RcsB (Morgan et al., 2015). Oxygen concentration-dependent interplay between global transcriptional regulator Cra and sugar sensors KdpE and FusR was shown to dictate expression of T3SS and determine the appropriate

location for aerobic EHEC colonisation in the low oxygen gut environment (Carlson-Banning et al., 2016).

a.

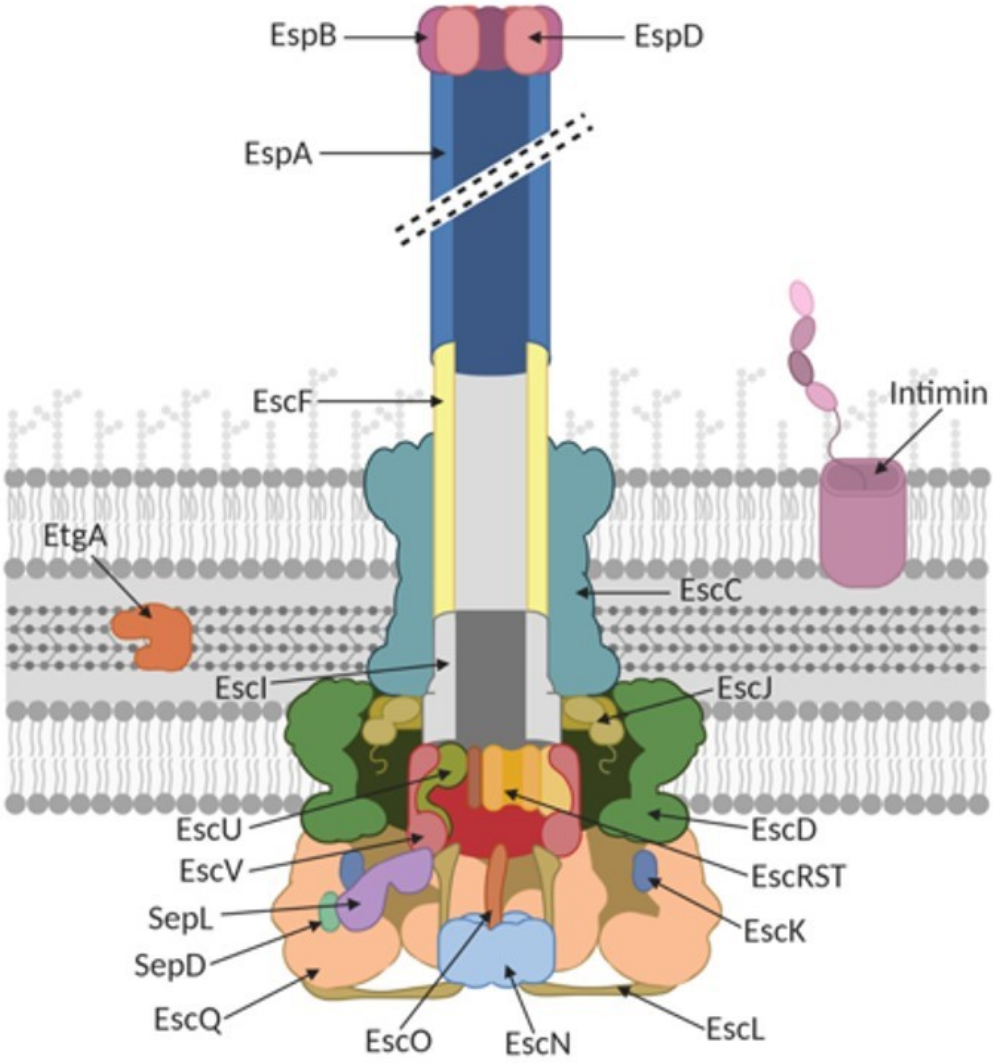


Figure 1.6. a. Schematic representation of Type 3 Secretion System. [Figure and figure legend continued on following pages]

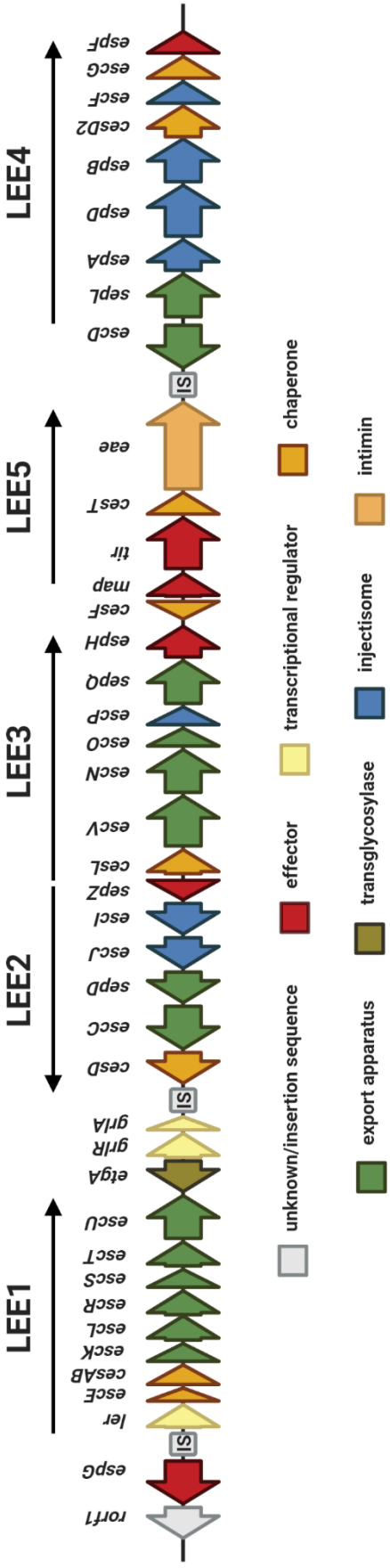

LEE1
LEE2
LEE3
LEE5
LEE4
orf1
espG
IS
ler
escF
cesAB
escK
escL
escR
escS
escT
escU
etgA
grlR
grlA
IS
cesD
escC
sepD
escJ
escI
sepZ
cesL
escV
escN
escO
escP
sepQ
espH
cesF
map
tir
cesT
eae
IS
escD
sepL
espA
espD
espB
cesD2
escF
escG
espF
unknown/insertion sequence
effector
transcriptional regulator
chaperone
export apparatus
transglycosylase
injectisome
intimin

Figure 1.6 (continued) b. Locus of Enterocyte Effacement (LEE) encoding the T3SS. LEE123-encoded proteins assemble to form an export/basal apparatus structure spanning the inner and outer bacterial membranes. Expression of the LEE123 is governed by the LEE1-encoded major transcriptional regulator, Ler, which is itself regulated by LEE-encoded GrlA and GrlR (Elliott et al., 2000; Russell et al., 2007). The translocon proteins of the injectisome machinery that connect the bacteria to the host cell are encoded on the LEE4 operon, alongside the gatekeeper protein SepL that together with its LEE2-encoded binding partner SepD governs hierarchical secretion of the translocon proteins and subsequent translocation of EHEC effector proteins (Deng et al., 2015; O'Connell et al., 2004). Additionally, LEE encodes the chaperone proteins that deliver the translocon and effector proteins to the gatekeeper complex and the ATPase component of the basal apparatus, the LEE3-encoded EscN (Chen et al., 2013; Creasey et al., 2003; Little et al., 2018; Neves et al., 2003). Some of the T3SS-secreted effector proteins are encoded on the LEE. The LEE5 operon encodes the major effector protein Tir that upon injections gets targeted to the host membrane and acts as a receptor for the LEE5 *eae*-encoded intimin that gets embedded in the bacterial outermembrane. Interaction between Tir and intimin facilitates strong bacterial adherence to the host cell (DeVinney et al., 1999; Hartland et al., 1999). Other effector proteins encoded on the LEE include *map, etgA, espG, espH, sepZ and espF*, the products of which are further involved in cytoskeletal remodelling, as well as inhibition of phagocytosis, mitochondrial and cellular trafficking disruption, cell apoptosis, disruption of ion channels and Rho GTPase modulation. A comprehensive review by Clements et al., describes the LEE and non-LEE encoded T3SS effector proteins involved in EHEC intimate attachment and formation of attaching and effacing lesions.

1.3.1.2 T3SS structure and function

Once in the appropriate environment EHEC T3SS facilitates adherence of the bacteria to the intestinal epithelium. Fully assembled EHEC T3SS is composed of a basal apparatus composed of 15 proteins that span through and across the inner and outer bacterial membranes. Then there are the translocon components that consist of an elongated hollow filament (EspA) that connects to the host cell via formation of a pore (EspBD) (Figure 1.6a). It is still unclear what the positioning and function of the smaller needle structure EscF is in EPEC and EHEC, it may help link the basal apparatus to the translocon. All of the structural T3SS proteins are essential for correct functioning of the apparatus (Burkinshaw et al., 2014; Puhar et al., 2014). Assembly of the T3SS is Sec-pathway dependent and is similar to that of flagella. However, export of the effector proteins is Sec-independent (Burkinshaw et al., 2014). Following the assembly of the basal apparatus, the translocon components EspABD are delivered to the basal apparatus with the help of LEE-encoded chaperone proteins CesAB, CesA2, EscA, CesD and CesD2. EspADB are then secreted through the basal apparatus machinery driven by the EspN ATPase (Creasey et al., 2003; Lodato et al., 2009; Su et al., 2008). Only after the translocon is formed are the other effector proteins exported. This hierarchical secretion of early translocon secretion substrates, and middle and late effector proteins is mediated via a sorting platform composed of SepL, SepD and CesL proteins (Díaz-Guerrero et al., 2021). The exported effectors include the translocated intimin receptor (Tir) that embeds itself in the host cell membrane and binds with intimin expressed on the bacterial surface. Tir-intimin binding and Tir clustering leads to a signalling cascade, which recruits the major regulatory N-WASP Arp2/3 complex leading to actin pedestal assembly. N-WASP activity requires other T3S secreted proteins, EspH involved in disruption of actin cytoskeleton and EspFu which forms a new nucleation core for actin polymerisation, in order to establish the actin

pedestal (Campellone et al., 2004). Changes in the host's actin cytoskeleton along with downregulation of filopodia and disruption of the host's microtubule network, mediated via an array of T3SS effectors, results in tight intimate attachment and disruption of intestinal microvilli, hence the term attaching & effacing (A/E) lesion. Activity of many T3SS proteins is shared among A/E pathogens and has been characterised in backgrounds closely resembling EHEC such as EPEC or *Citrobacter rodentium* (Dean et al., 2009). However notable differences in governing of T3SS expression in those species have been documented and the expression of a clone encoding the EHEC T3SS was shown to require an EHEC background (Elliott et al., 1999; Elliott et al., 2000; Hartland et al., 2013; Roe et al., 2003). T3SS is under finely-tuned, species and pathotype-specific regulation. Control of injectisome assembly and effector secretion in EHEC is further elaborated on below.

1.3.1.3 Regulation of T3SS assembly and effector secretion hierarchy

All of the T3SS structural proteins, as well as some regulators, chaperones and effector proteins are encoded on 35-kb chromosomal locus of enterocyte effacement (LEE) (Figure 1.6b). The LEE is composed of 5 major operons (LEE1-5) and several small transcriptional units, with at least 41 open reading frames (Franzin et al., 2015). The first gene of the LEE1 operon is *ler* which encodes the well-studied positive regulator of LEE expression, Ler (LEE-encoded regulator). Ler antagonizes the inhibition of LEE expression directed by the histone-like nucleoid-structuring protein (H-NS) by interfering with binding of H-NS to the promoter region of LEE operons (Shin et al., 2017; Wan et al., 2016). Apart from its crucial role in LEE expression, Ler is also known to be involved in regulation of non-LEE T3S associated factors. It is mainly the expression of Ler that is responsive

to the extracellular factors known to affect expression of the T3SS (Carlson-Banning et al., 2018; Donnenberg, 2013; Le Bihan et al., 2017; Sharp et al., 2007). Ler is itself known to be transcriptionally regulated by expression of multiple regulators including two other LEE-encoded proteins GrlA and GrlR, which bind to the LEE1 promoter and respectively activate or repress *ler* expression (Laaberaki et al., 2006; Russell et al., 2007). Multiple non-LEE factors such as RpoS were found to be indirectly involved in LEE transcription regulation via control of Ler expression (Laaberki et al., 2006). Other known transcriptional regulators of LEE expression include: the family of PcH transcriptional activators, that similarly to Ler compete with H-NS for binding to the LEE1 promoter region and antagonise other global transcriptional repressors such as StpA, Hha and YdgT; LEE3-encoded multiple point controller that represses positive Ler activity; as well as factors encoded on an inactive secondary T3SS of EHEC (T3SS2), EtrB that positively regulates LEE expression via transcriptional regulation of *ler* and *grlA,* and negative regulators EtrA and EivF (Luzader et al., 2016).

Apart from the Ler-mediated transcriptional control, EHEC T3SS is also thought to be regulated at the post-transcriptional level, managing processes such as assembly of the T3SS translocon and setting up the hierarchy between export of the translocator and effector proteins. Expression of a late non-LEE encoded T3SS effector protein NleA was shown to be repressed by a well-known post-transcriptional regulator CsrA. Interestingly NleA production was shown to be activated by CesT, a LEE5-encoded chaperone protein of the middle secretion substrate, Tir. This was shown to be an effect of CesT-CsrA interaction that ‘sponges’ the CsrA away from the *nleA* transcript (Ye et al., 2018; Katsowitch et al., 2017). Based on the studies in the closely related EPEC, CsrA is thought to control LEE expression in EHEC by binding to and repressing translation of the *grlA* transcript which results in a decrease in production of Ler (Bhatt et al., 2009). A recent study by Wang et al., 2018 indicated that CsrA and Hfq-mediated sRNA Spot42

post-transcriptionally regulate expression of the LEE4 operon. This interplay is described in more detail in subsection 1.3 of this chapter. In addition to its function as the chaperone of regulatory sRNAs, Hfq was indicated to repress LEE expression independently of associated sRNAs. This was shown to be achieved through direct binding of the Hfq distal face near the *grlA* and *ler* mRNA Shine-Dalgarno regions therefore obstructing ribosomal access and repressing activation of translation (Sudo et al., 2021).

Apart from Spot42, multiple non-coding sRNAs were indicated to control the assembly of T3SS in EHEC. The well-conserved GlmZ and GlmY sRNAs were shown to negatively regulate expression of the translocon through destabilisation of the LEE5-encoded transcripts and LEE4-encoded *espABD* mRNA (Gruber et al., 2014). GlmZ and GlmY were also shown to positively regulate production of non-LEE T3SS effector EspF. The GlmZ and GlmY-mediated regulation is thought to be important for correct ratio of structural translocon components and secreted effectors (Gruber et al., 2014). A study by Gruber et al., 2015 discovered that several EHEC-specific sRNAs namely sRNA56, sRNA103 and sRNA350 positively regulate expression of T3SS, either through regulating the *ler* mRNA or *espABD* transcript translation. There is also an account of a negative regulation of T3SS through destabilisation of LEE1 mRNA and inhibition of Ler protein synthesis elongation by a *cis* anti-sense sRNA called Arl that is encoded at the 3'UTR of *ler* (Tobe et al., 2014).

Additionally, the stability of *espADB* transcript and thus production of translocon proteins was shown to be controlled by RNase E that recognises a small six-codon mini-open reading frame at the 5'-end of the mRNA. This was shown to be inhibited by ribosomal-docking at the LEE4 Shine-Dalgarno and translation that interferes with RNase E interaction (Lodato et al., 2012; Lodato et al., 2017).

1.3.1.4 SepL - the ‘gatekeeper’ protein

One of the most crucial factors orchestrating assembly of the functional T3SS is a ‘gatekeeper’ protein SepL (Bhatt et al., 2011) (Figure 1.7). This 5.4 kD protein is encoded by the first gene on LEE4 operon, *sepL* and its homologues share more than 90% identity among A/E pathogens (Lodato et al., 2009). Multiple studies determined that without SepL, the EspA filament fails to be produced and the translocon is not assembled (Kresse et al., 2000, Deng et al., 2015; O’Connell et al. 2004; Wang et al., 2008). Cleavage of the LEE4 transcript by RNaseE, which occurs after translation of *sepL* mRNA, has been suggested to release *espABD* transcript and lead to translation of the translocon proteins only after SepL production (Lodato et al., 2009). Additionally, SepL along with its binding partner SepD and chaperone CesL form a molecular switch which controls the hierarchical transport of LEE-encoded effector proteins into the host cell (Younis et al., 2010). SepL was found to bind to the translocated intimin receptor (Tir), and delay effector release, allowing for prioritised secretion of translocon proteins EspABD (Deng et al., 2015; Tomalka et al., 2012; Tree et al., 2009, Wang et al., 2008). Studies conducted by Tomalka et al., 2012, indicated that interaction between the gatekeeper protein and translocator-specific signals located on EspABD determine the order of secretion, and prioritise export of the translocator proteins. In closely related EPEC, SepL interaction with EscV, a structural component of the basal apparatus, was shown to be crucial for targeting of translocon protein-chaperone complexes to the membrane (Portaliou et al., 2017). Expression of SepL itself is thought to be tightly regulated. Ler and SepD were shown to have a major impact on SepL production, with up to 10-fold decrease in its expression in *ler* deletion mutants (Wang, 2011). Despite its indubitable importance in control of T3SS assembly, the regulation of SepL production is still not well understood.

Characterisation of post-transcriptional control of SepL expression is the main topic of this thesis.

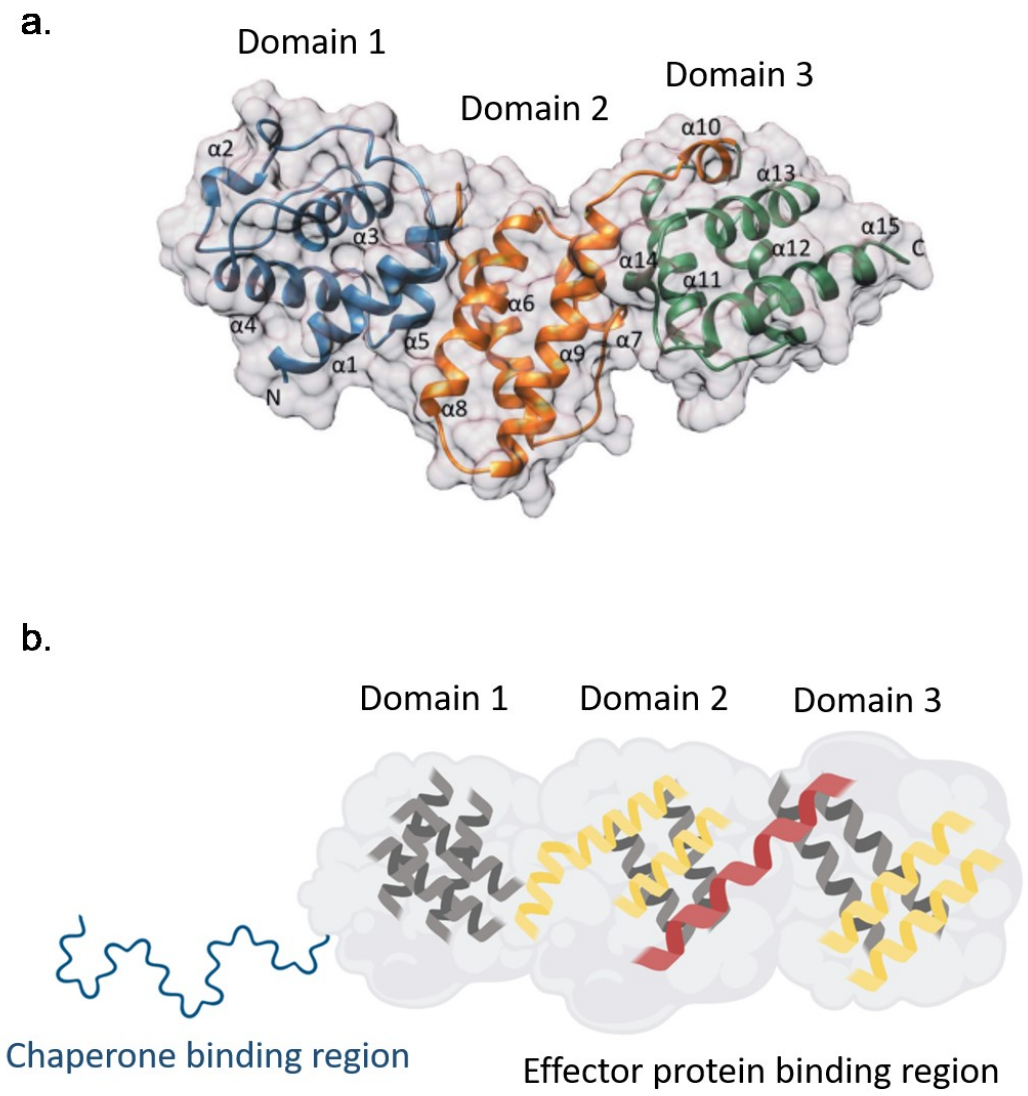


Figure 1.7. a. SepL gatekeeper structure in EPEC as solved by Burkinshaw et al., 2015. SepL has 3 active domains, domain 1 (blue), domain 2 (orange) and domain 3 (green) that are composed of 15 α-helices labelled 1-15. **b.** Schematic of SepL structure. Each of the SepL domains is thought to provide protein-protein interaction sites. Proposed effector and chaperone interaction regions have been labelled. Domain 3 region (around the yellow helices) was proven to interact with the effector protein Tir (Burkinshaw et al., 2015). [Adapted from Burkinshaw et al., 2015].

1.4 Previous studies leading to this project

Studies conducted previously in the laboratory indicated a heterogeneity of SepL expression in EHEC O157 backgrounds, which was directly correlated with heterogeneity of EspA filament production and was independent of Ler expression (Figure 1.8) (Roe et al., 2003; Wang et al., 2008). It was suggested that translational regulation of SepL expression might be involved in heterogeneity of T3SS filament expression manifested by EHEC. A series of sequentially truncated *sepL'*-GFP translational fusions were created in order to test the hypothesis. Loss of heterogeneity and a ten-fold increase in reporter GFP production was observed in bacteria carrying the truncated *sepL'*-GFP constructs without an obvious change in the *sepL* mRNA transcript levels produced (Wang et al., 2018; Wawszczyk, 2017) (Figure 1.9 b&c). The computational prediction of secondary structure of the full-length *sepL* mRNA and truncated *sepL* mRNA were shown to form respectively a 'closed' structure, with an obstructed ribosomal binding site (RBS), and more 'open' structure with an exposed RBS (Figure 1.9a). A recently published study by Wang et al., 2018 identified *sepL* mRNA conformation and mRNA-protein interactions as a determining factor in control of SepL expression. It was suggested that post-transcriptional changes are a driving factor behind regulation of SepL production and provide a possible lead in understanding of the observed heterogeneity (Wang et al., 2018). The study identified an interplay between the post-transcriptional regulator CsrA and the Hfq-mediated Spot42 sRNA which was shown to repress CsrA control of *sepL* mRNA translational activation. CsrA and Hfq-Spot42 were shown to bind to *sepL* transcript and were proposed to act on divergent and mutually exclusive mRNA structures (Wang et al., 2018). Accessibility to the CsrA/Hfq binding site determined by *sepL* mRNA conformation may play a role in the proposed interplay (Wang et al., 2018). However, the factors leading to the primary

switch from ‘OFF’ to ‘ON’ in terms of *sepL* mRNA translation are still to be determined (Wang, 2011; Wang et al., 2018).

There is some evidence suggesting that assembly of the T3SS basal body allows subsequent production of the translocon proteins, potentially via regulation of SepL expression (Diepold et al., 2014). Results obtained from studying the involvement of T3SS basal apparatus in SepL production determined a requirement for LEE1-encoded T3SS basal apparatus protein/s to be involved in SepL expression. Deletion of LEE1 operon resulted in a highly significant decrease in SepL expression but not *sepL* mRNA transcript production as measured using translational and transcriptional fusions. This indicates an involvement of LEE1-encoded proteins in post-transcriptional regulation of SepL production (Wawszczyk, 2017).

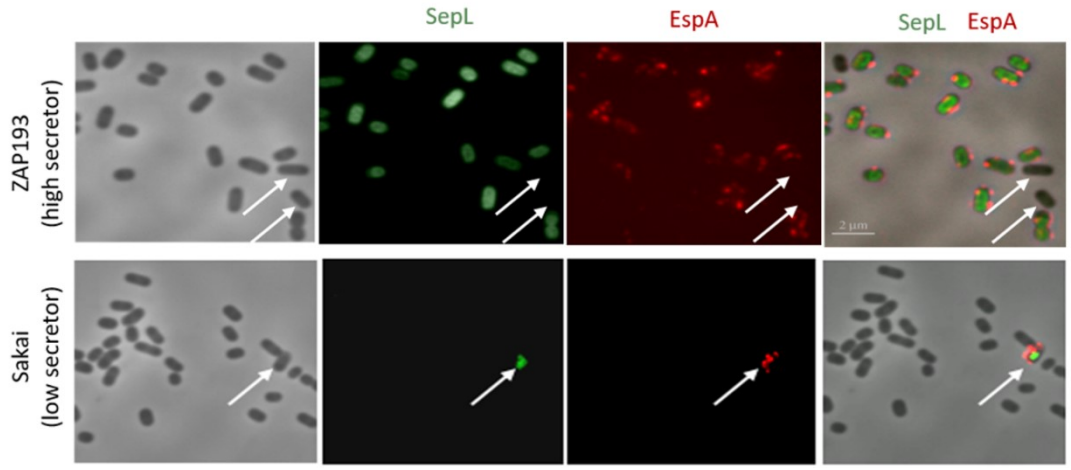


Figure 1.8. Heterogeneous expression of EHEC EspA filaments correlates with SepL expression. Production of EspA filament (mCherry-tagged in red) in low (Sakai) and high (ZAP193) secretor EHEC was shown to be heterogeneous in both strains with only a subset of bacteria in the colony producing the T3SS translocon proteins. This was shown to be correlated with expression of SepL gatekeeper protein (GFP-tagged in green) (Roe et al., 2003; Wang et al., 2008).

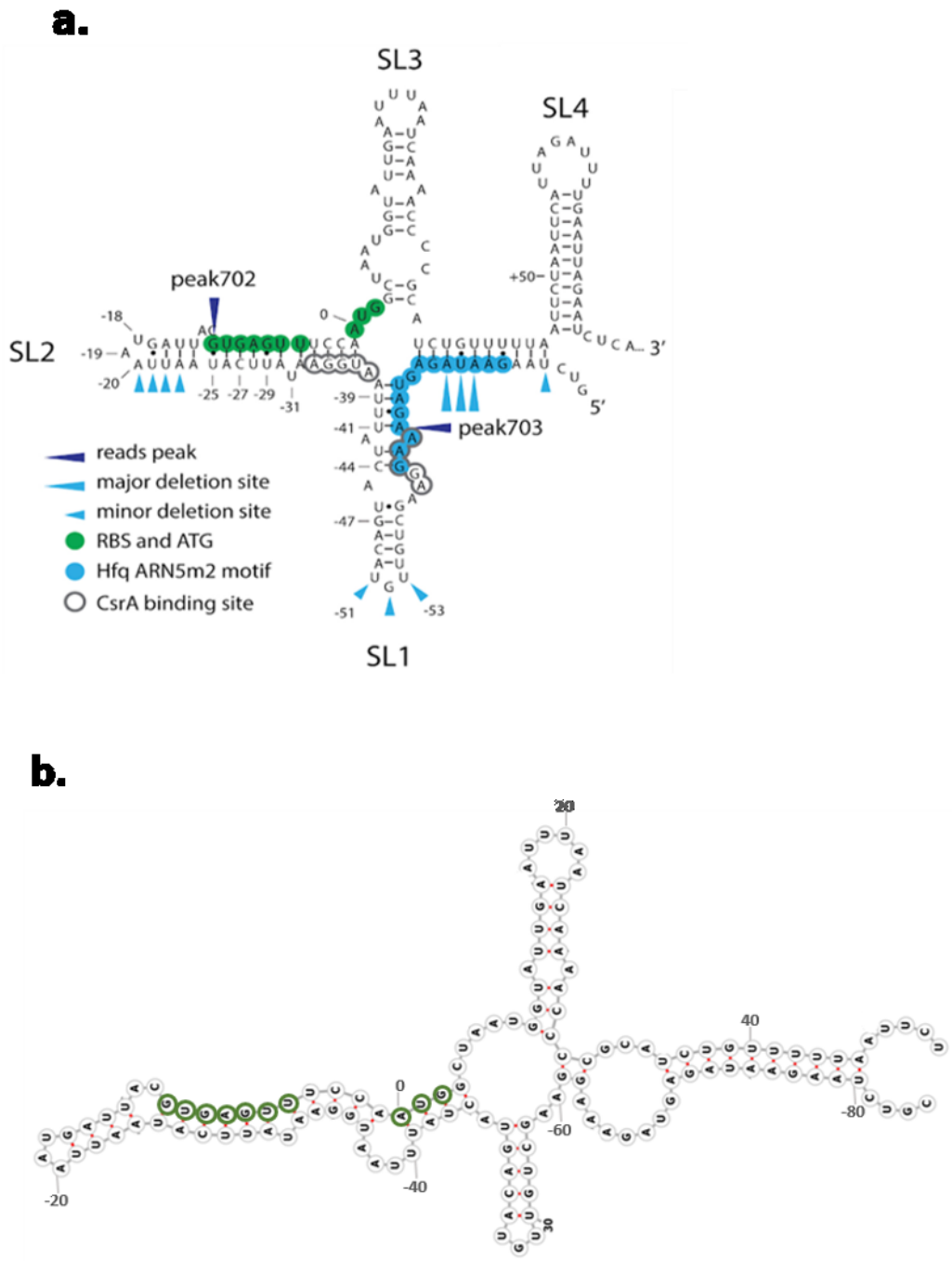


Figure 1.9. *sepL* mRNA characteristics and SepL expression. **a.** The computational prediction of *sepL* transcript secondary structure indicates a 'clover-leaf' conformation with the ribosomal binding site buried between the stem loop 2 (SL2) and stem loop 3 (SL3). **b.** Truncation of the sepL mRNA is predicted to expose the RBS.

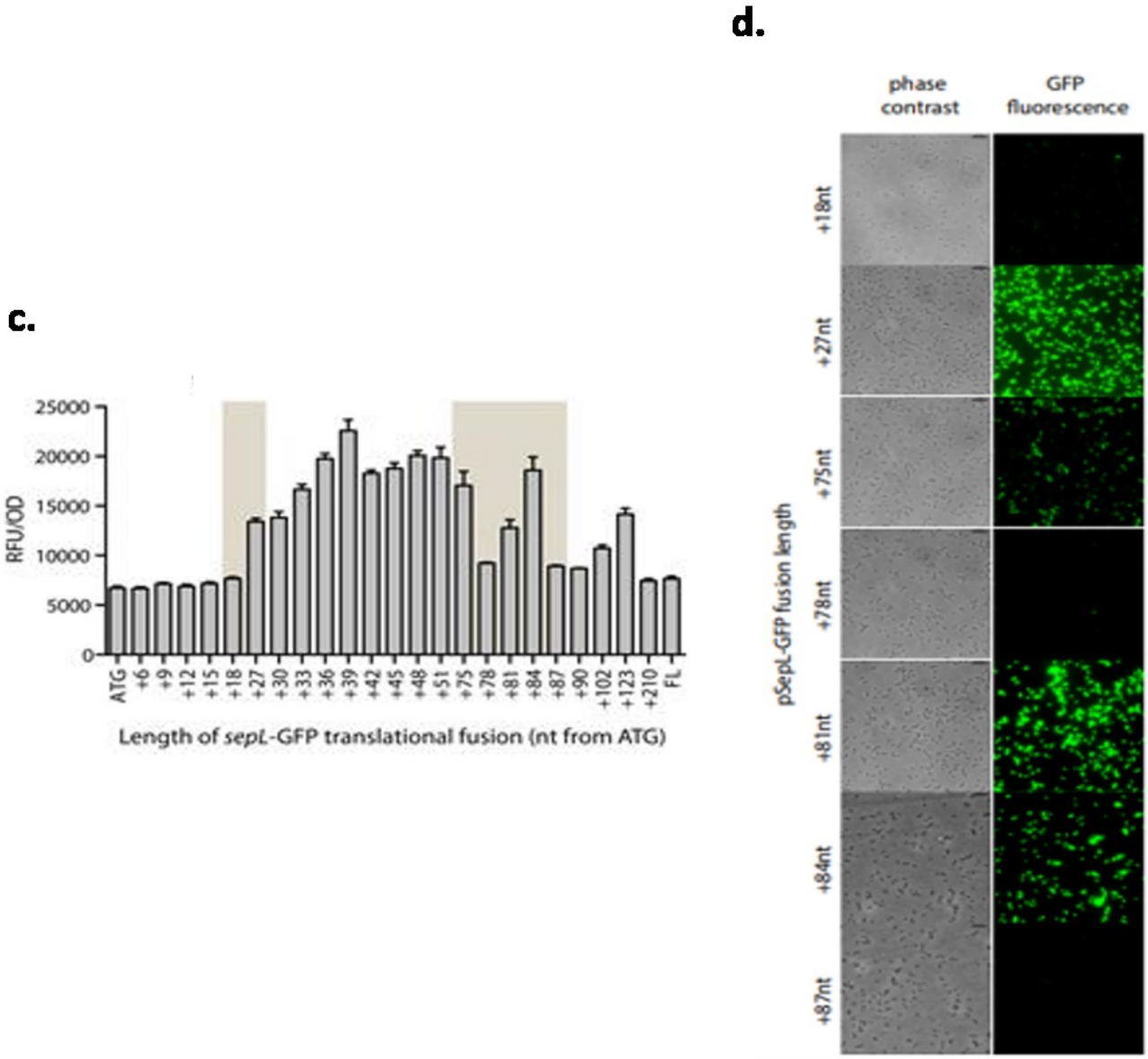


Figure 1.9 (continued). d.&e. Assessment of expression of the truncated *sepL* mRNA using translational reporter constructs indicates that *sepL* mRNA translation is dependent on the length of the transcript as **b.** visualised by fluorescence microscopy and **c.** quantified by fluorescence spectroscopy. This is thought to be due to alteration of the *sepL* mRNA conformation in the truncated fusions with different levels of exposure of the RBS (Wang et al., 2018).

1.5 Hypothesis and aims

The hypothesis of this project states that post-transcriptional regulation of SepL expression is to act as a checkpoint after production of the basal apparatus and before expression of the EspABD translocon. Taking into consideration previous research including the impact that *sepL* mRNA conformation has on SepL expression, we hypothesise that production of LEE1-encoded factors and assembly of the T3SS basal apparatus drives a change in *sepL* mRNA secondary structure and triggers an 'OFF' to 'ON' switch in terms of SepL mRNA translation (Wang et al., 2018). Therefore, we propose that LEE1-encoded factors play an essential role in post-transcriptional regulation of SepL production and thus formation of the EspABD translocon.

Using a combination of molecular biology techniques, chemical probing and RNA-protein binding assays, this project aimed to unravel new factors involved in translation of the *sepL* transcript and characterise the mechanism(s) underlying repression and activation of expression. The objectives of this project were as follows:

1. We aimed to investigate if the LEE4 regulation is distinct to that of other LEE operons by assessing the LEE4 transcription and mRNA abundance and comparing it with the other LEE transcripts.

2. We aimed to evaluate on the role of the *sepL* mRNA in SepL production via validation of the proposed computationally predicted mRNA secondary structure and determining its impact on mRNA abundance in different bacterial backgrounds.

3. The major objective of this project was to identify the LEE1-encoded factor/s responsible for modulation of the SepL expression, characterise its mode of action.

4. Subsequently we aimed to investigate the contact dependency of the proposed regulatory interaction by *sepL* mRNA-protein binding studies.

5. We aimed to clarify the involvement of the known SepL-regulatory factors the CsrA and the Hfq in the LEE1-mediated regulatory mechanism.

6. Additionally we aimed to investigate the impact of this mechanism on the finely-tuned hierarchy of LEE4-expression and the consequences of the translocon expression on the proposed *sepL* expression regulation.

The project ultimately aimed to better our understanding of the mechanisms underlying assembly of complex bacterial organelles such as T3SS.

Chapter 2 LEE4 transcription and maintenance

2.1 Background

LEE4 is the last operon on the Locus of Enterocyte Effacement (LEE). It encodes a translocon structure (EspADB) that connects bacteria to the host cell and allows for export of effector proteins and establishment of intimate attachment. The hallmark of EHEC pathogenesis, attaching/effacing lesions cannot be formed without a functional translocon. The first gene on the LEE4 operon is *sepL*, it encodes a 39.988 kDa protein SepL. SepL is a critical factor in orchestrating the assembly of the translocon and effector secretion. It functions as a 'gatekeeper' protein, which alongside its LEE2-encoded binding partner SepD and LEE3-encoded CesL forms a secretion switch that ensures formation of the translocon before effector release. SepL was found to bind to the translocated intimin receptor (Tir), and delay effector release, allowing for prioritised secretion of translocon proteins EspABD (Deng et al., 2015; Tomalka et al., 2012; Tree et al., 2009; Wang et al., 2008). Several studies indicated that *sepL* mutants are unable to secrete EspADB, which resulted in no filament formation and inability to produce A/E lesions. Additionally, a significant increase in effector proteins, such as LEE5-encoded Tir, was observed in *sepL* mutants, indicating loss of hierarchical control essential for T3SS function (Kresse et al., 2000, Deng et al., 2015; O'Connell et al. 2004; Wang et al., 2008).

SepL-mediated regulation of T3SS assembly is complex and a crucial process and so it is fair to assume expression of LEE4 and SepL itself has to be tightly regulated. There is evidence suggesting that assembly of the T3SS basal body ensures subsequent production of the translocon, potentially via regulation of SepL expression (Diepold et al.,

2014; Tree et al., 2009). Preliminary results obtained studying the involvement of T3SS basal apparatus in SepL production determined a requirement for LEE1-encoded T3SS basal apparatus protein/s in SepL production. Deletion of LEE1 operon but not LEE2 or LEE3 resulted in highly significant decrease in SepL expression (Wawszczyk, 2017).

Previous studies carried out in our laboratory indicated heterogeneity in SepL expression that correlated with heterogeneity of EspA filament formation at the single cell level. Changes in secondary structure of *sepL* mRNA were associated with alleviation of the heterogeneity and highly significant increase in SepL and EspA expression (Wang et al., 2008). Study by Wang et al. identified an interplay between post-transcriptional regulator CsrA and a RNA chaperon Hfq loaded with a Spot42 sRNA cargo, which were shown to competitively bind to *sepL* mRNA and lead to respectively translation activation and repression (Wang et al., 2018). Accessibility to the CsrA/Hfq binding site determined by *sepL* mRNA conformation was proposed to play a role in the interplay (Wang et al., 2018).

Taken together those observations point towards complex control of LEE4 expression involving the interplay of multiple factors, namely LEE1-encoded protein(s), *sepL* mRNA secondary structure, Hfq and CsrA. This section of my thesis aimed to characterise these dependencies in greater detail.

2.2 Results

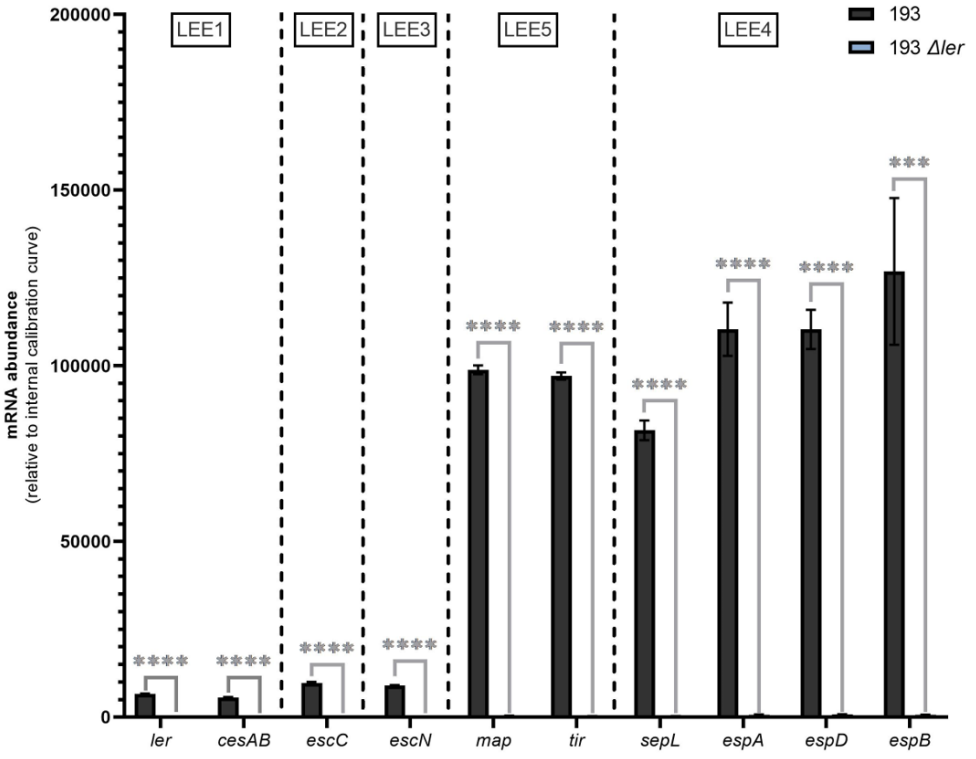
2.2.1 Role of LEE1-encoded factors in LEE4 transcript maintenance

Positive regulation of LEE transcripts by the LEE1-encoded major transcriptional activation factor Ler has been previously documented in multiple studies in EHEC and closely related EPEC (Bustamante et al., 2001; Elliott et al., 2000; Shin et al., 2017). It was demonstrated that Ler antagonises the repression of LEE transcription carried out by the histone-like nucleoid-structuring protein (H-NS). Ler acts by interfering with H-NS binding to the promoter region of several LEE operons, resulting in de-repression of transcription (Shin et al., 2017). However there is an inconsistent account of Ler regulation of the LEE4 transcript, with some studies reporting transcriptional dependency on Ler, while some showed no significant effect of Ler on LEE4 transcriptional activation. Several studies indicated that the absence of Ler results in highly significant decrease in detection of LEE4 transcripts (Elliott et al., 2000; Friedberg et al., 1999; Mellies et al., 1999; Sanchez San Martin et al., 2001; Sperandio et al., 2001). However Ler was shown to have no effect on expression of LEE4 transcript as measured by LEE4::*lacZ* transcriptional fusion in laboratory *E. coli* background K12 (Elliott et al., 2000). Additionally it is important to note that the main promoter of LEE4 is located upstream of *sepL* as suggested by Goldberg et al., 2001 and confirmed in our laboratory, and no direct activation of *sepL* transcription by Ler has been recorded (Wang et al., 2008).

In order to characterise the regulatory role of Ler on LEE transcripts the mRNA abundance of genes representative of all five LEE operons was determined by RT qPCR in wild type and Ler deficient EHEC backgrounds (Figure 2.1). Quantification of the mRNA was

achieved using internal calibration curves produced by serially diluting quantified EHEC genomic DNA. ZAP193 and Sakai strains were assessed to ensure the obtained results were not strain specific. These strains do not produce Shiga toxin and so can be used in level 2 containment laboratories in the UK and Sakai strain has been fully sequenced and characterised (Hayashi, 2001). Additionally, ZAP193 exhibits a high secretor phenotype and Sakai is a low secretor of T3SS effectors, which facilitates detection of both decreases and increases in expression of LEE-encoded factors that might not be observable with only one secretion level. ZAP193 and Sakai strains were the EHEC strains used consistently throughout this project. In order to ensure optimal LEE expression the bacterial cultures were grown in minimal essential media (MEM)-HEPES, supplemented with 0.1% glucose and 250 nM $Fe(NO_3)_2$ that is the media previously shown to induce T3SS production (Roe et al., 2003). Supplemented MEM-HEPES was used to culture bacteria targeted for T3SS expression measurement experiments throughout this project unless stated otherwise. *Δler* ZAP193 (ZAP1004) and ΔLEE1 Sakai that were previously produced in the laboratory by Dr. Alison McIntosh and myself, were used as Ler deficient backgrounds (Low et al., 2006; Wawszczyk, 2017). There were either highly significant decreases or in some cases a lack of detectable mRNA for all of the tested LEE-encoded transcripts in these Ler negative backgrounds compared to their wild type Ler positive parent strains. It was concluded that the absence of Ler has a clear negative effect on abundance of all of the main LEE transcripts, LEE1,2,3,5 and 4. This was consistent with results obtained by Elliott et al., 2000. The transcript level as measured by RT qPCR does not necessarily give a conclusive account of transcriptional activity as some post-transcriptional processes such as transcript degradation can alter the levels of transcript detected.

a.


LEE1
LEE2
LEE3
LEE5
LEE4
193
193 Δler
mRNA abundance
(relative to internal calibration curve)
200000
150000
100000
50000
0
****
***
ler
cesAB
escC
escN
map
tir
sepL
espA
espD
espB

b.

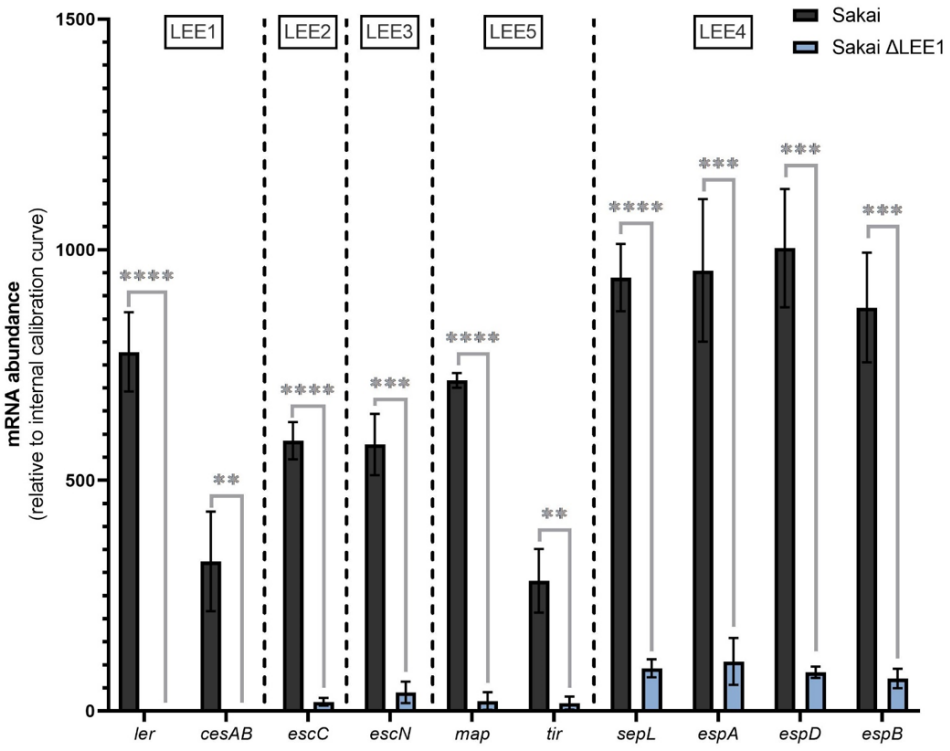

LEE1
LEE2
LEE3
LEE5
LEE4
Sakai
Sakai ΔLEE1
mRNA abundance
(relative to internal calibration curve)
1500
1000
500
0
****
***
**
ler
cesAB
escC
escN
map
tir
sepL
espA
espD
espB

Figure 2.1. Detection of LEE transcripts in Ler deficient **a.** ZAP193 high secretor and **b.** Sakai low secretor EHEC backgrounds. mRNA abundance of genes throughout the LEE (LEE1,2,3,5,4 operons) was probed in Δ*ler* ZAP193 and ΔLEE1 Sakai (blue) and compared to transcripts produced by wild type ZAP193 and Sakai (grey). Highly significant decreases in LEE transcripts abundance were noted in the absence of Ler in both high and low secretor EHEC (biological repeat n=3, technical repeat n≥9, unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; *** - $p<0.001$; **** - $p<0.0001$). It is worth noting that detectable amounts of transcripts (above Cq>35 threshold cut off) were documented in the absence of Ler throughout the LEE, with the exception of LEE1.

In order to determine if the observed Ler dependency is based on transcriptional regulation, as has been documented in studies focusing on Ler governed de-repression from transcriptional repressor H-NS, expression of plasmid-based LEE2::*gfp*, LEE3::*gfp*, LEE5::*gfp* and LEE4::*lacZ* transcriptional fusions (Flockhart et al., 2012) was measured in wild type EHEC and Ler deficient backgrounds with and without Ler complementation *in trans* (Figure 2.2). In transcriptional fusions, reporter genes such as *gfp* encoding for green fluorescent protein (GFP) or *lacZ* encoding for a β-galactosidase enzyme, are fused directly downstream of the promoter sequence of the gene or operon of interest. The intensity of the fluorescence emitted by the GFP was measured by fluorescence spectroscopy while β-galactosidase can be measured by its activity on *o*-nitrophenyl-β-galactoside (ONPG) that produces yellow coloured *o*-nitrophenol when cleaved. Activation of the promoter drives expression of the reporter proteins, which can be measured by assessing the fluorescence in the case of GFP or absorbance at 420nm for the β-galactosidase assay; both allow measurement of the level of transcriptional activation of the gene of interest from the plasmid construct. In my work, the LacZ reporter was used instead of GFP to measure LEE4 transcription because production of LEE4::*gfp* fusion was unsuccessful. High GFP levels were previously shown to be toxic to the bacterial cell, leading to growth arrest and cell lysis, which could be a potential reason for the difficulty in producing the LEE4::*gfp* transcriptional fusion (Feilmeier et al., 2000). *lacZ* is a native *E. coli* gene that can be expressed at high levels without detriment to the cell and a LEE4::*lacZ* construct was successfully produced in by Sean McAteer (Senior Lab Technician the the Gally group) for purposes of this project (Matthews et al., 2005). This enabled comparison of LEE4 transcriptional activation with transcript, protein and translational fusions allowing insights into different types of post-transcriptional regulation (Matthews et al., 2005). Transcription of LEE1,2,3 and 5 was determined by assessing the fluorescence intensity of the respective

gfp fusions, and was shown to be negatively affected by absence of Ler, which confirmed Ler's function as a positive transcriptional regulator of those operons, likely through inhibition of repressive activity of H-NS. Critically, the promoter activity of LEE4, as measured using LEE4::*lacZ* transcriptional fusion, was unchanged by absence of Ler (Figure 2.2). This suggests that transcription of LEE4 is not regulated by Ler, unlike other LEE operons. However results obtained from RT qPCR analysis (Figure 2.1) signify that Ler is necessary for detection of all of the LEE operon transcripts including LEE4. This observed disconnect between LEE4 transcription and mRNA abundance in *Δler* backgrounds indicates direct or indirect involvement of Ler in post-transcriptional regulation of LEE4 transcript maintenance or transcription completion.

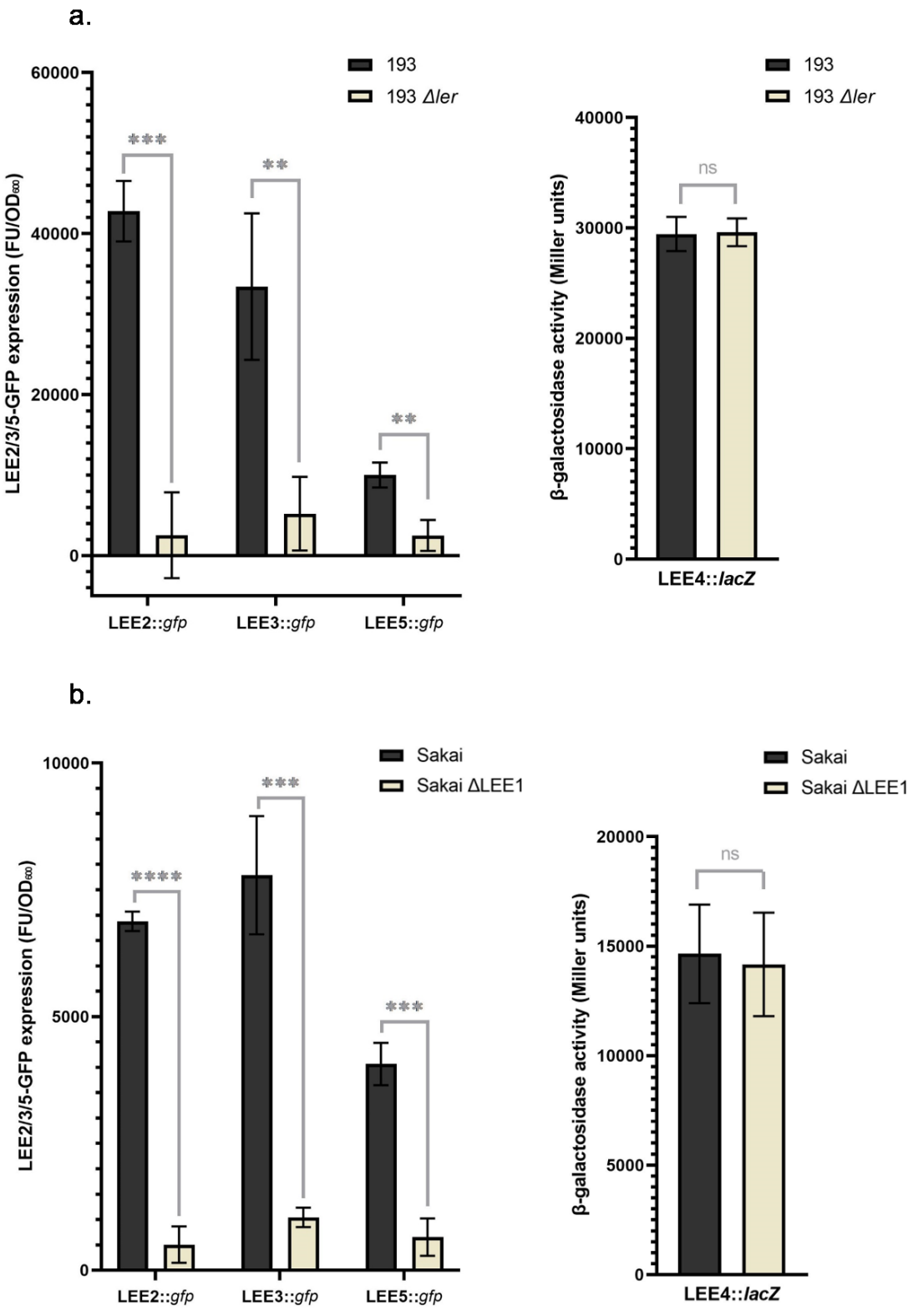
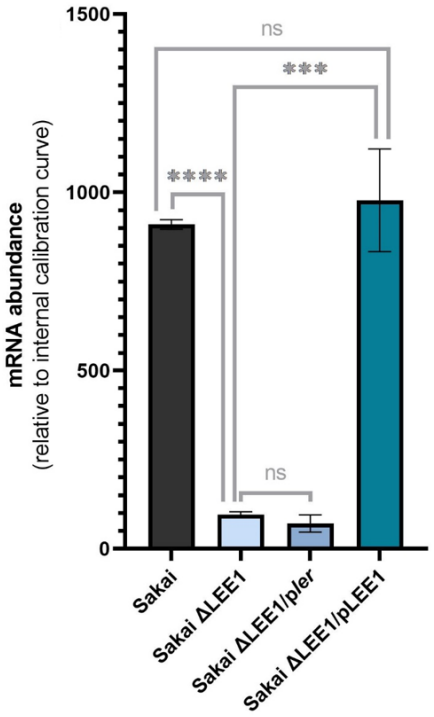

a.
193
193 Δler
LEE2/3/5-GFP expression (FU/OD600)
60000
40000
20000
0
***
**
**
LEE2::gfp
LEE3::gfp
LEE5::gfp
193
193 Δler
β-galactosidase activity (Miller units)
40000
30000
20000
10000
0
ns
LEE4::lacZ
b.
Sakai
Sakai ΔLEE1
LEE2/3/5-GFP expression (FU/OD600)
10000
5000
0
****
***
***
LEE2::gfp
LEE3::gfp
LEE5::gfp
Sakai
Sakai ΔLEE1
β-galactosidase activity (Miller units)
20000
15000
10000
5000
0
ns
LEE4::lacZ

Figure 2.2. Promoter activity of LEE operons in Ler deficient EHEC backgrounds. pLEE2-*gfp* (LEE2::*gfp*), pLEE3-*gfp* (LEE3::*gfp*), pLEE5-*gfp* (LEE5::*gfp*) and pLEE4-*lacZ* (LEE4::*lacZ*) transcriptional fusions were expressed in **a.** ZAP193 and **b.** Sakai. Ler was concluded to positively regulate transcriptional activation in LEE2,3 and 5 operons but not LEE4 in both high and low secretor EHEC strains (biological repeat n=3, technical repeat n≥6, unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; *** - $p<0.001$; **** - $p<0.0001$).

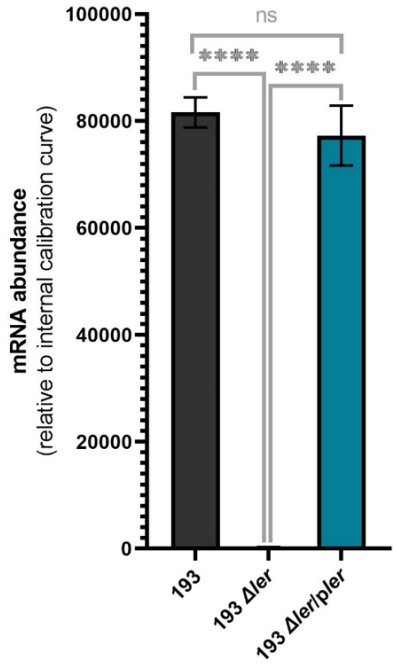
As demonstrated in Figure 2.2, Ler governs transcription of LEE1,2,3 and 5 operons. Deletion of the entire LEE1 operon but also deletion of LEE1-encoded basal apparatus proteins EscRSTU have been previously shown to result in significant decrease in production of the first protein encoded on LEE4 operon – SepL (Wagner et al., 2010; Wawszczyk, 2017; Yerushalmi et al., 2014). This was not observed in ΔLEE2 and ΔLEE3 backgrounds (Wawszczyk, 2017). Therefore it is possible that it is the lack of Ler-regulated LEE1-encoded factors and not direct Ler interaction that results in the decrease in LEE4 transcript abundance in Ler deficient backgrounds (Figure 2.1). In order to test this hypothesis: (1) the ΔLEE1 Sakai background was supplemented with plasmid-encoded *ler* clone (pWSK*ler*) previously produced in the laboratory by Dr. Xuefang Xu and the *sepL* mRNA abundance was measured by RT qPCR, (2) *sepL* mRNA abundance was also determined in the ΔLEE1 Sakai complemented with plasmid-based LEE1 clone (pWSKLEE1) produced by myself for purposes of this project and (3) the *Δler* ZAP193 (ZAP1004) strain was complemented with pWSK*ler* and *sepL* mRNA abundance determined (Figure 2.3). Measured *sepL* mRNA abundance was compared to that detected in wild type Sakai and ZAP193 and ΔLEE1 Sakai and *Δler* ZAP193 carrying empty pWSK29 vector used during construction of pWSKLEE1 and pWSK*ler*. Obtained results showed that supplementation of ΔLEE1 with Ler alone was insufficient to restore *sepL* mRNA abundance as compared to wild type Sakai to any significant level. However complementation of ΔLEE1 and *Δler* ZAP193 backgrounds with the LEE1 and *ler* clones, respectively, resulted in complete restoration of *sepL* transcript levels. β-galactosidase assay of LEE4::*lacZ* transcriptional fusion was carried out in all of the tested backgrounds to ensure that the observed differences were due to post-transcriptional regulation and not changes in transcriptional activity (Figure 2.3). Promoter activity was unchanged in all of the tested backgrounds as compared to wild type strains. Taken together it is evident from the data

that LEE1-encoded factors other than Ler are responsible for the post-transcriptional regulation of LEE4 transcript maintenance.

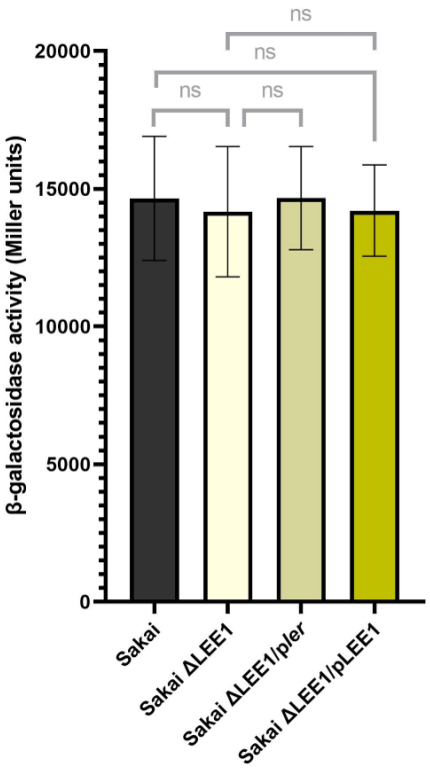
a.


mRNA abundance
(relative to internal calibration curve)
1500
1000
500
0
ns
***
****
ns
Sakai
Sakai ΔLEE1
Sakai ΔLEE1/pler
Sakai ΔLEE1/pLEE1

b.


mRNA abundance
(relative to internal calibration curve)
100000
80000
60000
40000
20000
0
ns
****
****
193
193 Δler
193 Δler/pler

c.


β-galactosidase activity (Miller units)
20000
15000
10000
5000
0
ns
ns
ns
ns
Sakai
Sakai ΔLEE1
Sakai ΔLEE1/pler
Sakai ΔLEE1/pLEE1

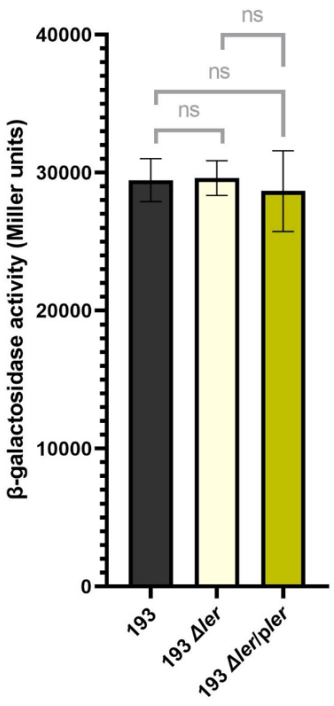

β-galactosidase activity (Miller units)
40000
30000
20000
10000
0
ns
ns
ns
193
193 Δler
193 Δler/pler

Figure 2.3. The impact of Ler and other LEE1-encoded factors on *sepL* transcript abundance. **a.** Supplementation of Sakai ΔLEE1 background with pWSK*ler* (p*ler*) did not result in significant change in *sepL* mRNA levels. Supplementation with pWSKLEE1 (pLEE1) did significantly increase the *sepL* transcript abundance to the levels comparable to wild type Sakai background (biological repeat n=3, technical repeat n=9). **b.** Supplementation of *Δler* ZAP193 background with pWSK*ler* (p*ler*) restored *sepL* mRNA to levels akin to wild type ZAP193 (biological repeat n=3, technical repeat n=9). **c.** Transcriptional activation of LEE4 promoter was shown to be unchanged by absence of Ler or LEE1-encoded factors as observed before and supplementation with pWSK*ler* (p*ler*) and pWSKLEE1 (pLEE1) did not affect the promoter activity either (biological repeat n=3, technical repeat n≥6). (unpaired t-test analysis: ns (not significant) – $p>0.05$; *** - $p<0.001$; **** - $p<0.0001$).

Additionally translation of SepL was determined in the same backgrounds utilising a SepL-GFP translational fusion (pDW6) previously produced in the laboratory by Dr. Dai Wang (Wang et al., 2018). Translational reporters fuse the reporter gene in-frame with the gene of interest, translation of the protein of interest leads to production of the reporter, concentration of which can be assessed by measuring fluorescence intensity by fluorescence spectroscopy, and give a quantifiable account of fused protein production. A significant decrease in SepL-GFP expression was observed in ΔLEE1 Sakai and *Δler* ZAP193 backgrounds compared to wild type backgrounds. Supplementation with pWSK*ler* resulted in restoration of protein expression only in the *Δler* ZAP193 with no change in ΔLEE1 Sakai strain. In the ΔLEE1 background complementation, with pWSKLEE1 was required to restore SepL-GFP production (Figure 2.4a&b). Taken together these results indicate that the LEE1-dependent changes in transcript abundance feed through to translation of a SepL-GFP reporter. Obtained results were additionally validated by Western blot analysis using serum raised against purified SepL (Figure 2.4 c&d).

Regulation of LEE4 transcript degradation is a potential post-transcriptional mechanism to account for the observed disconnect between LEE4 transcription and *sepL* mRNA levels. Stability of the LEE4 transcript was assayed in ZAP193 and Sakai backgrounds by inhibiting transcription through addition of rifampicin to the cell culture and a 10 minute time course of RT qPCR examined post-rifampicin treatment (Figure 2.5) (Mosaei et al., 2020). Rapid degradation of *sepL* mRNA was noted in both background with a higher rate of decline in Sakai background. This could be due to the lower starting levels of the *sepL* mRNA in the low secretor background or potentially enhanced processing/negative regulation of transcript stability in Sakai as compared to ZAP193. The following sections of this thesis assess the factors involved in maintenance of the *sepL* transcript.

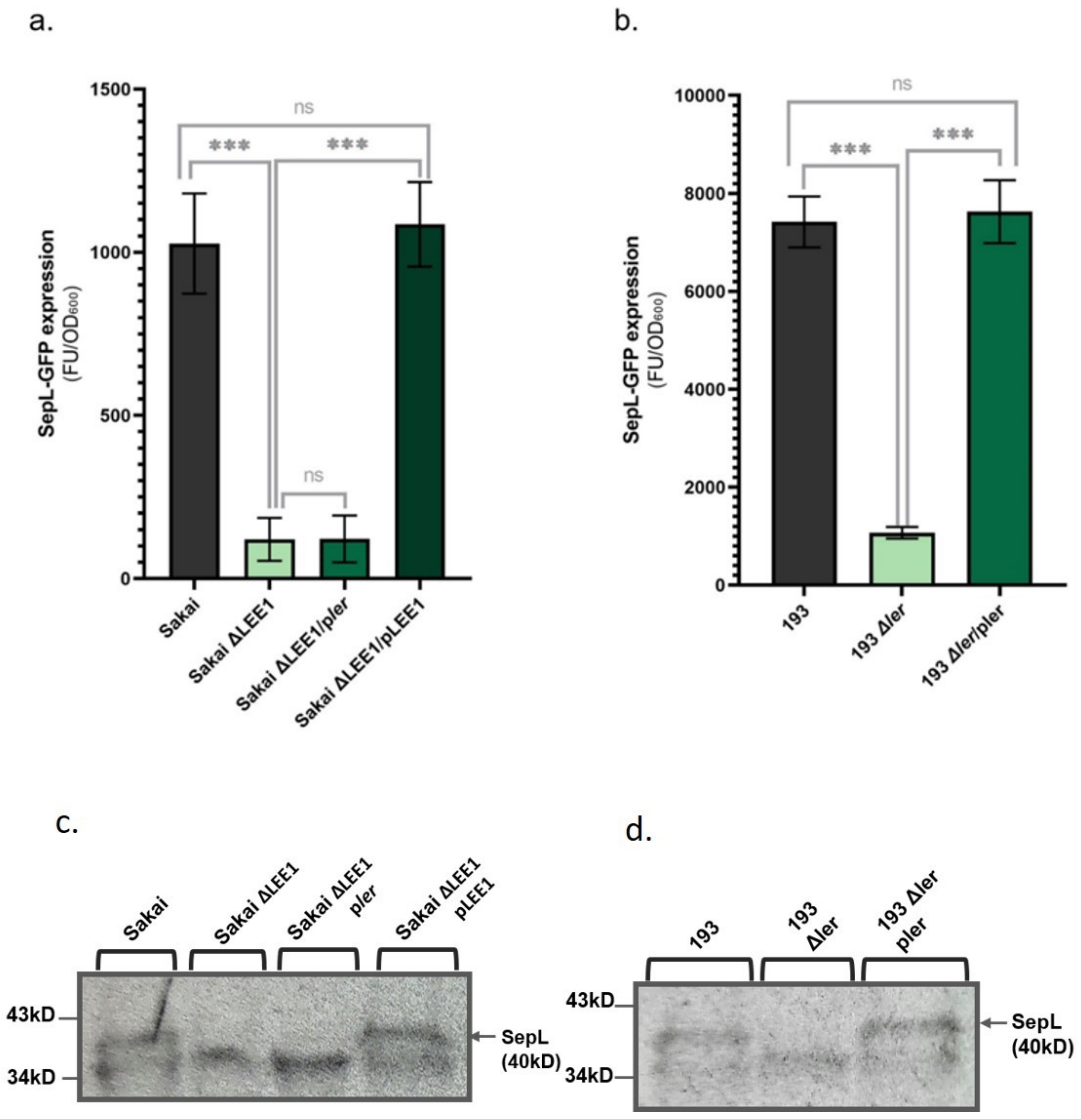

a.
b.
SepL-GFP expression (FU/OD600)
1500
1000
500
0
ns
***
***
ns
Sakai
Sakai ΔLEE1
Sakai ΔLEE1/pler
Sakai ΔLEE1/pLEE1
10000
8000
6000
4000
2000
0
193
193 Δler
193 Δler/pler
c.
d.
Sakai
Sakai ΔLEE1
Sakai ΔLEE1 pler
Sakai ΔLEE1 pLEE1
43kD
34kD
SepL (40kD)
193
193 Δler
193 Δler pler

Figure 2.4. The impact of Ler and other LEE1-encoded factors on SepL protein expression. SepL expression was measured by fluorescence spectroscopy detecting the SepL-GFP produced from a translational fusion construct (pDW6) in the **a.** low secretor Sakai and **b.** high secretor ZAP193 backgrounds. SepL-GFP was measured in background lacking Ler (Δ*ler* and ΔLEE1) and the Ler-complemented backgrounds. Significant decrease in SepL-GFP production was noted in both Ler negative backgrounds. SepL-GFP expression was only restored by Ler complementation (p*ler*) in the Δ*ler* ZAP193. ΔLEE1 Sakai required complementation with entire LEE1 clone (pLEE1) in order for SepL-GFP expression to be restored. (unpaired t-test analysis: ns (not significant) – $p>0.05$; * - $p<0.05$; *** - $p<0.001$; **** - $p<0.0001$). (biological repeat n=3, technical repeat n=9). Immunoblotting analysis in **c.** Sakai and **d.** ZAP193 backgrounds detected the 40kD band corresponding to SepL protein in p*ler* supplemented Δ*ler* ZAP193 and pLEE1 supplemented ΔLEE1 backgrounds but not the ΔLEE1 Sakai background supplemented with p*ler*. This validated SepL-GFP results. Δ*sepL* ZAP193 background acted as a negative control for SepL production (biological repeat n=3, technical repeat n≥3).

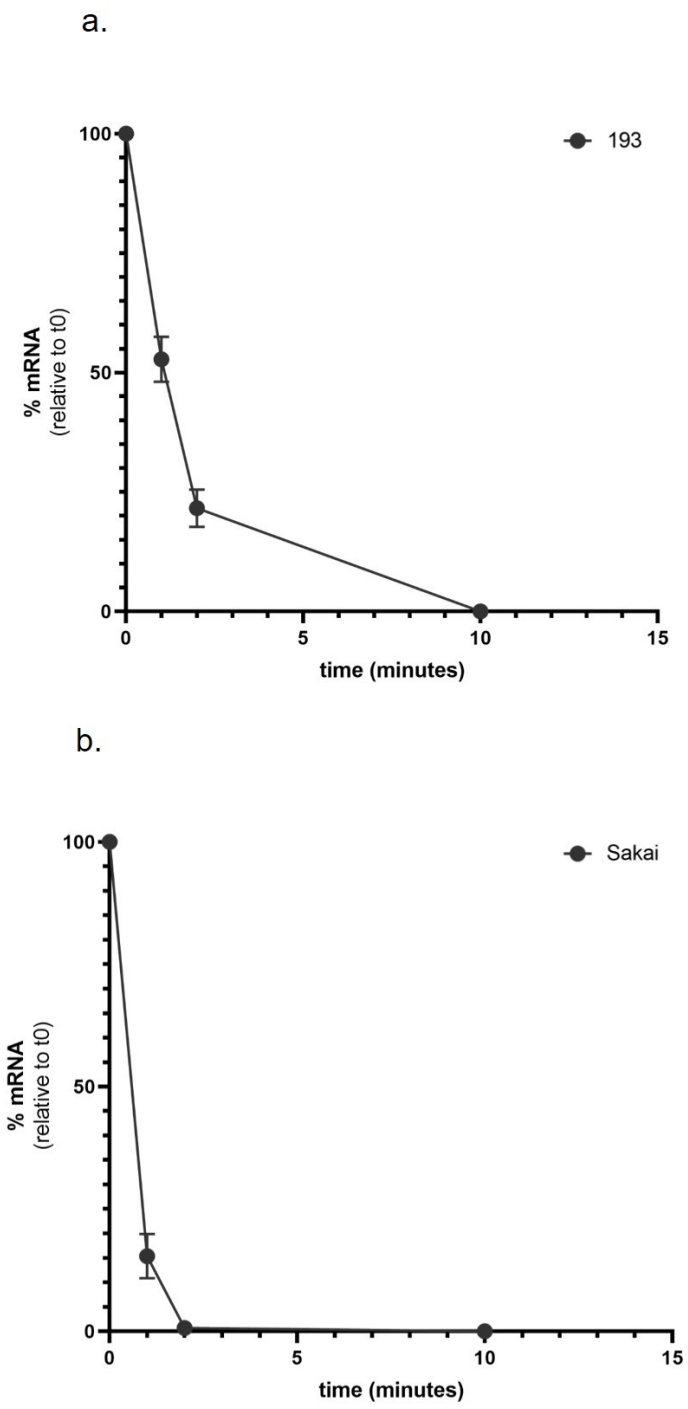


Figure 2.5. *sepL* transcript degradation in wild type **a.** ZAP193 and **b.** Sakai EHEC backgrounds. RNA samples were collected throughout 10 minute time course with time 0 indicating the starting quantity of the transcript before addition of transcription inhibitor rifampicin. Percentage of the transcript detected as compared to the starting quantity was plotted (biological repeat n=3, technical repeat n=9).

2.2.2 *sepL* mRNA structure affects LEE4 transcript stability

2.2.2.1 Validation of sepL mRNA secondary structure

The conformation adopted by mRNA is known to play a role in transcript maintenance, which refers to the stability of the transcript within the cellular environment and translation efficiency, including by determining access to ribosomal binding sites, ribosomal enhancers and binding and recruitment of interacting regulatory factors and processing complexes such as the degradosome (Chen et al., 2008; Del Campo et al., 2015). The computational prediction of *sepL* mRNA secondary structure indicated that its ribosomal binding site is likely to be obstructed and thus negatively regulate translation initiation (Wang et al., 2018). Truncation of the *sepL* transcript that is predicted to disrupt the folding of the mRNA, and expose the Shine-Dalgarno site, and this was shown to increase SepL expression. Thus further indicating that the conformation of the full-length *sepL* transcript is inhibitory to SepL expression (Wang et al., 2018; Wawszczyk 2017).

In order to validate the computational prediction, the secondary structure of the *sepL* transcript was mapped and the accessibility to the ribosomal binding site was determined using a chemical probing technique called selective 2'-hydroxyl acylation analysed by primer extension (SHAPE). SHAPE has been previously used to successfully map an entire genome of HIV-1 and STMV to a single-nucleotide resolution level as well as determine RNA-protein interactions and structural transitions (Archer et al., 2013; Grohman et al., 2013; McGinnis et al., 2012; Watts et al., 2019; Wilkinson et al., 2008). SHAPE utilises small hydroxyl-selective electrophilic reagents that acylate the 2'hydroxyl group of RNA nucleotides. Reactivity of the 2'-hydroxyl groups varies depending on the nucleotide flexibility, with more flexible residues, such as single stranded, loop and clip RNA regions, sampling

wide variety of conformation, some of them being more nucleophilic, which results in enhanced modification of the more dynamic RNA regions (McGinnis et al., 2012). The modifications can then be analysed by reverse transcription and primer extension of the treated sample (Figure 2.6) (Rice et al., 2014).

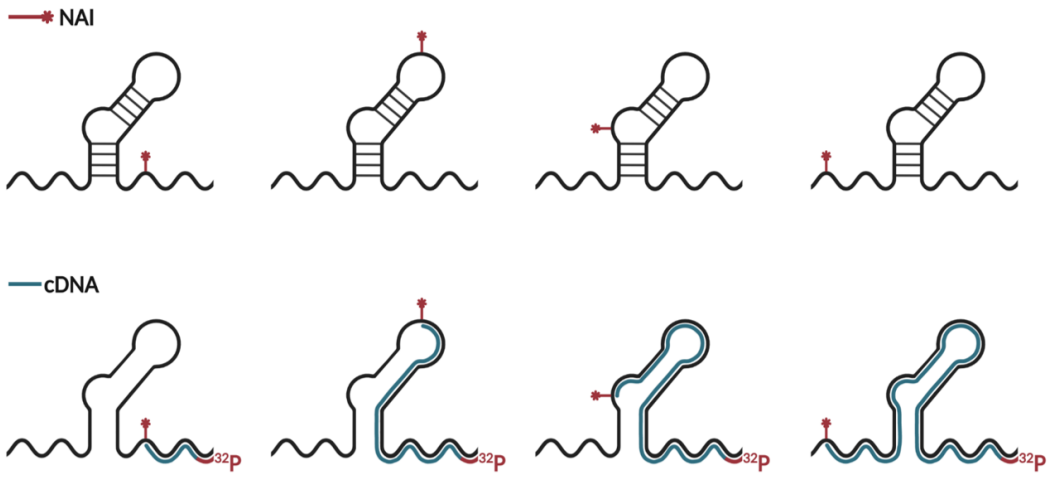


Figure 2.6. Selective 2'-hydroxyl acylation of probed RNA by acylation reagent NAI (1H-Imidazol-1-yl(2-methyl-3-pyridinyl)methanone) at the most flexible RNA regions (loops and single stranded RNA) followed by primer extension with radioactive marker ^{32}Phosphorus oligonucleotides and cDNA synthesis. Reverse transcriptase pauses at the nucleotides modified by NAI and thus the cDNA library produced is marked with reverse transcription stops at the sites of SHAPE modifications, which indirectly maps the RNA structure to a single nucleotide resolution. The modified regions can then be identified by running a gel of the RNA sequencing ladders (Rice et al., 2014).

SHAPE can be utilised to probe RNA structures both *in vitro* and *in vivo* (Watters et al., 2016). *sepL* mRNA was modified *in vivo* in order to ensure that the structure mapped is the closest to that naturally occurring in the bacterial cell. Several SHAPE acylation reagents can be used to modify the probed RNA. A comparative study by Lee et al. determined that acylimidazole reagent NAI produces the sharpest signal with the lowest background with an *in vivo* modification procedure (Lee et al., 2017). Additionally NAI was previously shown to successfully modify bacterial RNA, including transiently expressed *sepL* mRNA, therefore NAI was the reagent chosen to map sepL UTR and 5' region (Wawszczyk 2017). The radioactive marker ^{32}Phosphorus was used to tag the oligonucleotides for sequencing ladder production and primer extension of NAI-modified mRNA as opposed to newly employed fluorescent labels. ^{32}P is far more sensitive and stable than fluorescent tags, thus is more likely to detect the scarcely expressed *sepL* mRNA. Factors known to influence the efficiency of the probing such as optical density of the cells modified, urea PAA gel concentration, exposure time, and primer extension temperature and time were previously optimised (Wawszczyk, 2017). The samples were treated with exonuclease (ExoI) and ribonuclease (RNAse If) in order to degrade residual DNA and dinucleotide RNA bonds, which was shown to improve the detectability of the modifications. Overexpression of full-length SepL (pDW6), was shown to further enhance the detectability of the modifications due to an increase in the concentration of *sepL* mRNA (Wawszczyk, 2017). The Sakai background supplemented with pDW6 was used to probe *sepL* mRNA due to its low SepL expression profile, which under the working model at the time, was thought to increase detection of the 'folded' and 'untranslated' *sepL* mRNA.

Results obtained from probing of *sepL* mRNA secondary structure confirmed the computational prediction (Figure 2.7).

Modifications detected were representative of the predicted single stranded and more flexible regions of the mRNA. Analysis and quantification of the observed modifications indicate the presence of a ‘folded’, ‘cloverleaf’ secondary structure, with a major stem loop (SL2). The mapped conformation of the mRNA and especially the region proximal to the SL2 is likely to bury the ribosomal binding site and obstruct ribosomal access. Thus indicating that the observed naturally occurring *sepL* mRNA secondary structure is a potential limiting factor in mRNA translation and plays an important role in the switch from “OFF” to “ON” in terms of SepL production. This finding was published in Wang et al., 2018.

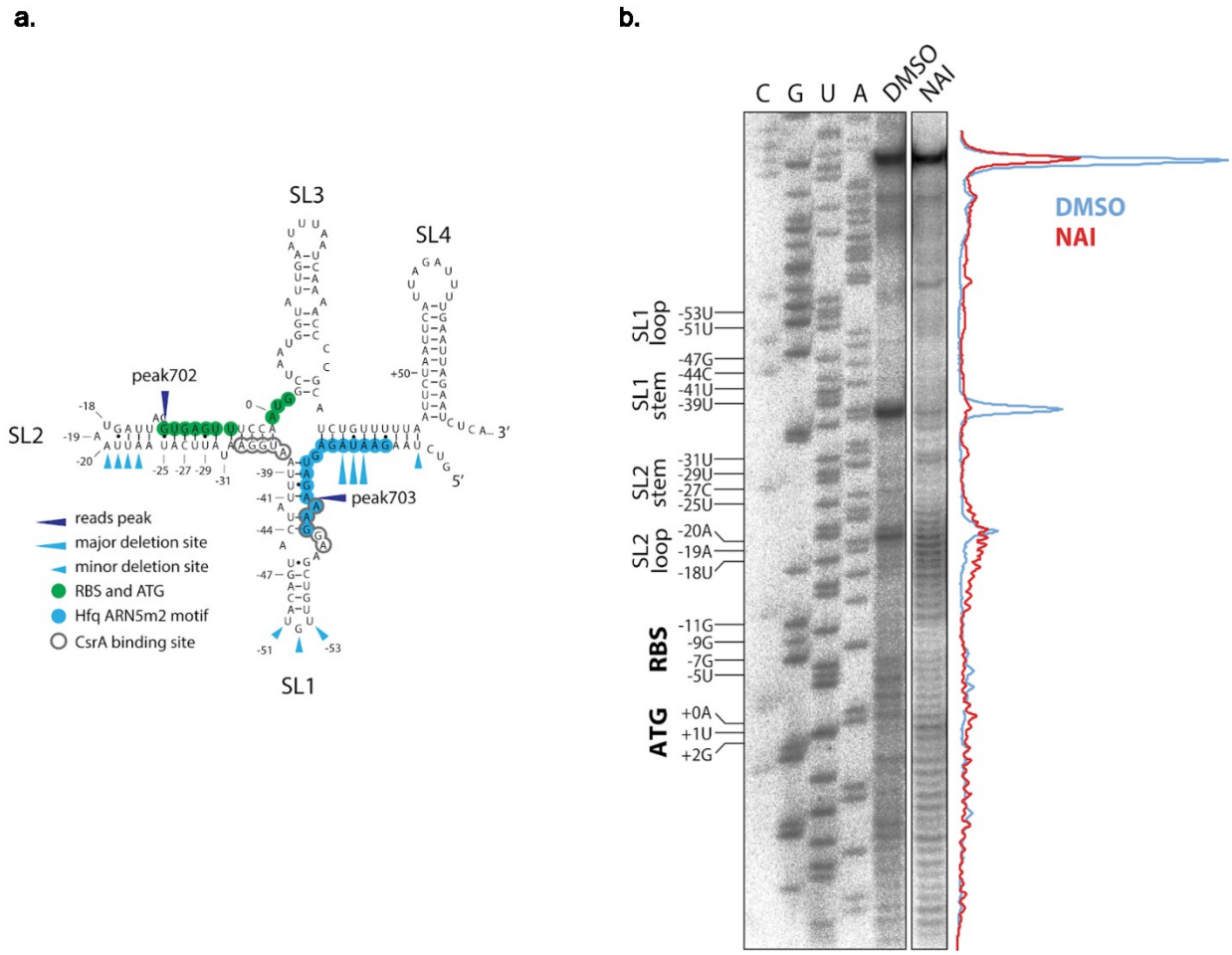


Figure 2.7. Predicted and structural analysis of the *sepL* 5'UTR and early coding region. **a.** RNAfold software prediction of sepL mRNA region from base -82 to +78, including ribosomal binding site and start codon (AUG). Four predicted stem loops were identified and are designated as SL1-4. RBS is predicted to be buried between SL2 and SL3. **b.** *in vivo* structural probing of the RBS and AUG *sepL* mRNA region (base -53 to +2) using SHAPE. pDW6 (SepL-GFP) supplemented Sakai culture was treated with SHAPE reagent NAI or control DMSO. The modified regions indicated by higher intensity radio-labelling bands can be identified by reading them to the produced sequencing ladder. Natural reverse transcriptase stops that can be identified in both DMSO control and NAI treated samples do not indicate the flexibility of the RNA regions mapped. The traces on the right handside indicate radio-labelling intensity. [As published in Wang et al., 2018]

2.2.2.2 Conformation of *sepL* mRNA affects transcript stability

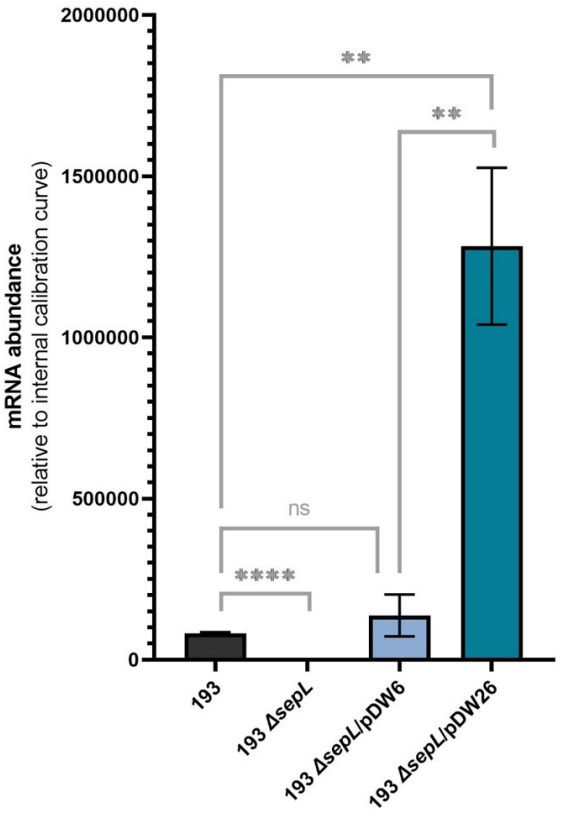
Wang et al., 2018 demonstrated that changes in the *sepL* mRNA structure result in differential SepL expression. It was proposed that a conformational change in the initially repressive *sepL* mRNA structure is necessary for the translation to occur (Wang et al., 2018). My work combined with previous published and unpublished studies lead to an updated working model of the regulation, with transcript being produced at high levels and the conformation of the *sepL* mRNA determines its stability and the initially formed 'folded' secondary structure is rapidly degraded. Fundamentally, I propose its translation is a competition between rapid degradation of the mapped structure and positive intervention of a factor to enable translation, most likely by opening up the structure to allow translation, possibly inhibiting degradation at the same time.

In order to test this, *sepL* transcript abundance was measured by RT qPCR in *ΔsepL* ZAP193 EHEC strain complemented with a full-length SepL-GFP expression construct (pDW6) predicted to form the inhibitory 'folded' mRNA structure, and a truncated SepL-GFP (pDW26) that was shown to be more readily expressed than the full-length SepL and is predicted to form a different, 'open' transcript conformation with exposed RBS (Wang et al., 2018). The measurements were compared to *sepL* mRNA abundance in wild type ZAP193 and *ΔsepL* ZAP193 backgrounds supplemented with pACYC184 cloning vector used in creation of pDW6 and pDW26 SepL-GFP expression fusions (Wang et al., 2018)

Despite a naturally high abundance of *sepL* mRNA in ZAP193, highly significant increase in the detected transcript was measured for the truncated SepL-GFP construct compared to the full-length SepL-

GFP fusion (Figure 2.8a.). Subsequently the stability of the full-length and truncated *sepL* mRNA was assayed as described in section 1.2.1. Percentages of the detectable mRNA as compared to the starting quantities at time 0 were plotted to allow for the comparison of degradation patterns between transcripts of different abundance levels. mRNA decay rate in the ZAP193 and *ΔsepL* ZAP193 supplemented with full-length SepL-GFP fusion was comparable, which provided confidence that the SepL-GFP fusion constructs are suitable tools for assessing the behaviour of the *sepL* transcript. In comparison the truncated 'open' *sepL* mRNA was shown to be degraded at a decreased rate and appeared more stable throughout the performed RT qPCR time course post-rifampicin treatment (Figure 2.8b.). Taken together obtained results indicate that the secondary structure of the *sepL* mRNA determines transcript stability and thus transcript availability for translation.

a.


2000000
1500000
1000000
500000
0
mRNA abundance
(relative to internal calibration curve)
**
**
ns
****
193
193 ΔsepL
193 ΔsepL/pDW6
193 ΔsepL/pDW26

b.

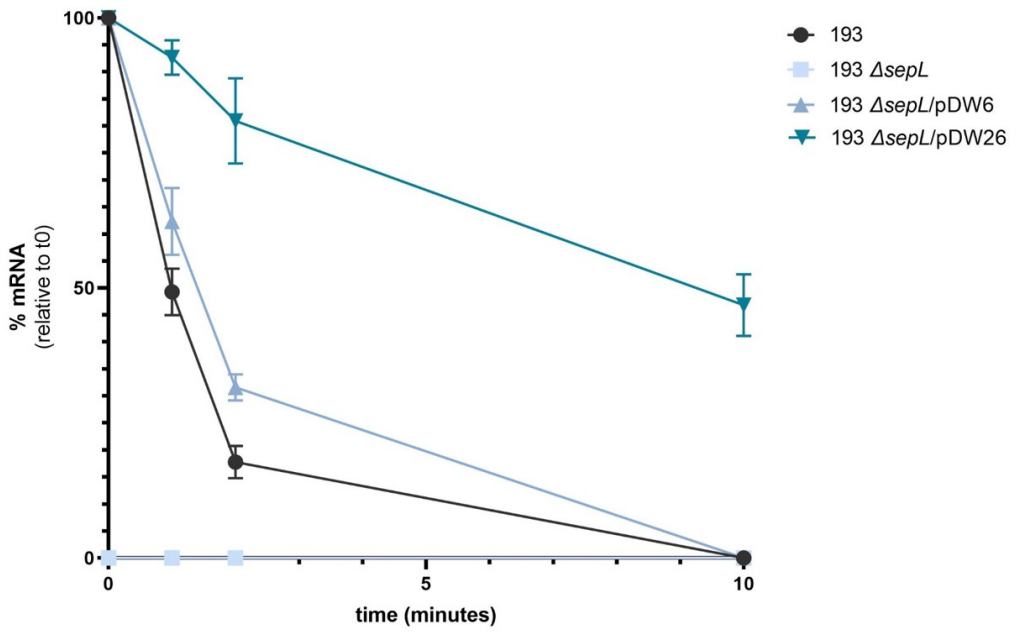

100
50
0
% mRNA
(relative to t0)
0
5
10
time (minutes)
193
193 ΔsepL
193 ΔsepL/pDW6
193 ΔsepL/pDW26

Figure 2.8. Maintenance of *sepL* transcripts predicted to form different secondary structures. **a.** *sepL* mRNA abundance in ZAP193, ZAP193 *ΔsepL* and ZAP193 *ΔsepL* supplemented with constructs producing full-length 'folded' *sepL* transcript (pDW6) and 'open' *sepL* transcript (pDW26). The 'open' sepL transcript (193 *ΔsepL*/pDW26) was shown to be detected at significantly higher levels than the 'folded' sepL mRNA (193 and 193 ΔsepL/pDW6) (biological repeat n=3, technical repeat n=9; unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; **** - $p<0.0001$). **b.** Degradation assay of sepL mRNA throughout a 10 minutes time course after transcription inhibitor addition. Folded full-length sepL mRNA from wild type ZAP193 and full-length SepL-GFP construct (193 and 193 *ΔsepL*/pDW6 respectively) was shown to have comparable degradation pattern. 'Open' *sepL* mRNA (193 *ΔsepL*/pDW26) exhibited higher stability with easily detectable transcript levels at the last time point. By contrast, the 'folded' sepL mRNA was not detectable at 10 minutes past transcription inhibitor treatment (biological repeat n=3, technical repeat n=9).

2.2.3 Rapid degradation of the LEE4 transcript is background specific

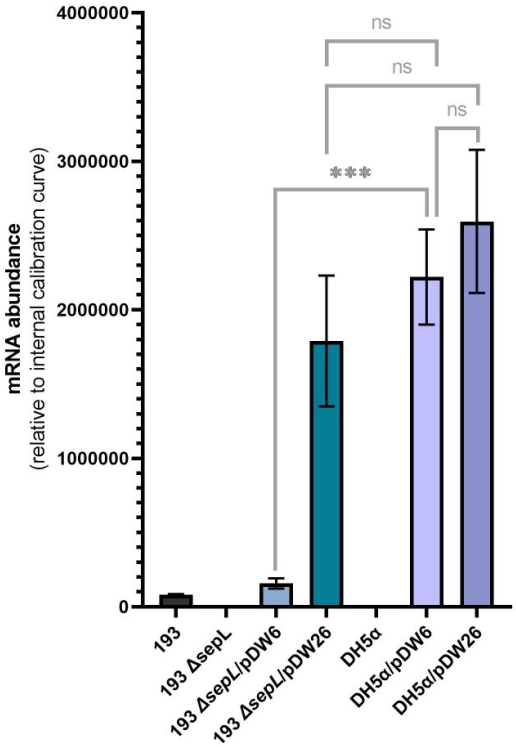
The extent of T3SS expression and its regulation is markedly different depending on the bacterial background the machinery is produced in. Multiple examples have been published including the comparative studies of closely related EHEC and EPEC *E. coli* pathotypes (Mellies et al., 2007; Roe et al., 2000; Spears et al., 2006; Turner et al., 2019). Recently Ruano-Gallego et al., demonstrated that functional injectisome can be produced by laboratory *E. coli* strain K12 (Ruano-Gallego et al., 2015). A study by Elliott et al. showed that a laboratory *E. coli* strain K12 is able to express LEE4 transcript despite lack of other LEE-encoded components, including Ler. However the study suggested that an specific regulatory factors that are not encoded on the core E. coli genome are necessary for production of SepL protein (Elliott et al., 2000).

In order to determine if the observed structurally-dependent stability of *sepL* mRNA is specific to an EHEC O157 background the degradation of full-length 'folded' (pDW6) and truncated 'open' (pDW26) SepL-GFP expression constructs were assayed in laboratory *E. coli* strain DH5α and *ΔsepL* ZAP193 (Figure 2.9a). *ΔsepL* ZAP193 was used to ensure the *sepL* mRNA produced was only from the plasmid constructs. This allowed us to compare the expression of the structurally different *sepL* transcripts in EHEC and nonEHEC *E. coli* backgrounds. Transcript abundance as measured by RT qPCR time course indicated that both full-length and truncated *sepL* mRNA can be expressed and were detected in similar quantities in the DH5α background. This is in contrast to the EHEC background in which the abundance of truncated 'open' *sepL* mRNA was significantly higher than that of full-length 'folded' *sepL* transcript. The degradation time course demonstrated that in the DH5α background the 'folded' and the

'open' *sepL* transcripts follow the same degradation pattern, comparable to that of the 'open' *sepL* transcript in EHEC background (Figure 2.9b). Taken together it is evident that the change in the secondary structure of the *sepL* mRNA does not affect the maintenance of the transcript to the same extent in the non-EHEC background (DH5α) and therefore there is a rapid degradation of the full-length, 'folded' *sepL* transcript in the EHEC O157 background.

Interestingly the high availability of the full-length *sepL* transcript in the laboratory *E. coli* background does not translate to production of the SepL protein. Expression of SepL-GFP full-length and truncated translational fusions (pDW6 and pDW26 respectively) in DH5α background were previously assayed by myself during my MRes and the experiments were repeated for the purposes of this thesis (Figure 2.10) (Wawszczyk, 2017). Despite comparable levels of *sepL* mRNA produced from the full-length and truncated constructs, only the 'open' truncated construct produced detectable SepL-GFP protein in the DH5α background, while none was detectable from the full-length SepL-GFP construct. This aligns with the working model in terms of additional SepL post-transcriptional regulation, in addition to the clear results concerning the transcript maintenance. This additional control is also dependent on the *sepL* mRNA structure and requires factors present in EHEC O157 and absent in DH5α.

a.


4000000
3000000
2000000
1000000
0
mRNA abundance
(relative to internal calibration curve)
ns
ns
ns
***
193
193 ΔsepL
193 ΔsepL/pDW6
193 ΔsepL/pDW26
DH5α
DH5α/pDW6
DH5α/pDW26

b.

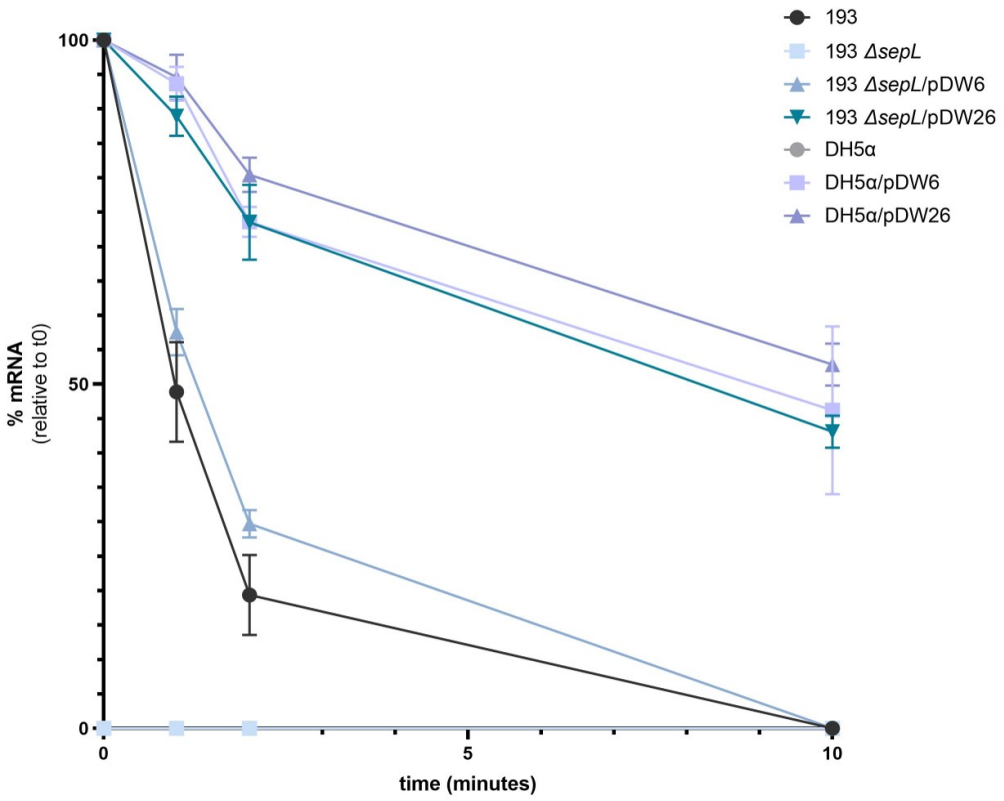

100
50
0
% mRNA
(relative to t0)
0
5
10
time (minutes)
193
193 ΔsepL
193 ΔsepL/pDW6
193 ΔsepL/pDW26
DH5α
DH5α/pDW6
DH5α/pDW26

Figure 2.9. *sepL* mRNA transcripts in nonT3SS *E. coli* (DH5α) background. **a.** Abundance of 'folded' (pDW6) and 'open' (pDW26) *sepL* mRNA in DH5α background compared to ZAP193. Significantly higher 'folded' *sepL* transcript levels were noted in the DH5α background and were comparable to those of the 'open' transcript in both ZAP193 and DH5α (biological repeat n=3, technical repeat n≥6; unpaired t-test analysis: ns (not significant) – $p>0.05$; *** - $p<0.001$). **b.** Stability of the 'folded' and 'open' *sepL* mRNA in the DH5α background as measured with a degradation time course assay. The degradation pattern of the 'folded' (pDW6) and 'open' (pDW26) transcript in DH5α background was akin to that of the 'open' *sepL* transcript in the EHEC ZAP193 background and was more stable than the 'folded' mRNA (biological repeat n=3, technical repeat n=6).

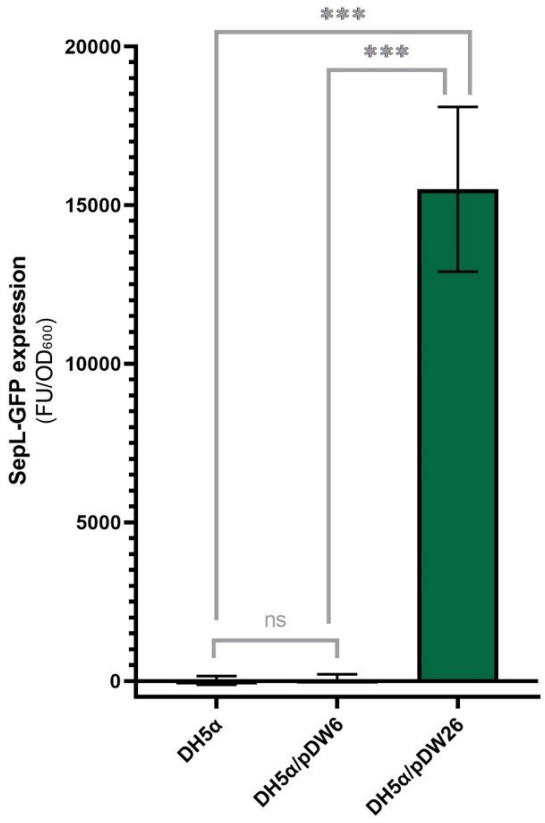


Figure 2.10. Expression of SepL-GFP translational fusions predicted to form 'folded' (pDW6) and 'open' (pDW26) mRNA conformations in nonT3SS *E. coli* background. SepL-GFP protein production was assayed using fluorescence spectroscopy at OD_{600}=0.8, that was shown to be the optimal growth stage for T3SS expression. Only *sepL* transcript that formed the 'open' conformation (pDW26) produced detectable levels of SepL-GFP (biological repeat n=3, technical repeat n=9; unpaired t-test analysis: ns (not significant) – $p>0.05$; *** - $p<0.001$)

2.2.4 **Hfq but not CsrA regulates LEE4 transcript maintenance**

In order to determine if the regulatory role of Hfq and CsrA on SepL expression involves manipulation of the *sepL* transcript maintenance, *sepL* mRNA abundance was determined using RT qPCR in ZAP193 and Sakai backgrounds supplemented with combinations of overexpressed *hfq* (pWSK-*hfq*) produced by Sean McAteer and *csrA* (pDW66) (Wang, 2011). The results obtained were compared to Sakai and ZAP193 backgrounds supplemented with empty pWSK29 and pACYC184 vectors used in creation of respectively pWSK-*hfq* and pDW66. Overexpression constructs of the regulators and not the *hfq* and *csrA* deletion backgrounds were utilised because viable *hfq* deletion mutants that did not harbour compensatory mutations was unable to be produced. Difficulty in generation of *hfq* mutants without compensatory mutations has been previously noted and is likely due to the absolutely crucial function of Hfq in a plethora of bacterial housekeeping processes (Ding et al., 2004; Shakhnovich et al., 2009; Sonnleitner et al., 2003; Tsui et al., 1994). The impact of Hfq and CsrA on *sepL* mRNA stability was additionally assessed by rifampicin degradation assay. Transcription of LEE4 was assessed in all of the backgrounds using the LEE4::*lacZ* fusion and showed no significant differences, thus ensuring that the obtained observations are due to post-transcriptional changes in *sepL* mRNA processing (Figure 2.11).

Overexpression of Hfq was shown to highly decrease *sepL* transcript abundance in both ZAP193 and Sakai, while overexpression of CsrA had no effect on the levels of *sepL* mRNA detected as measured by RT qPCR. Backgrounds that were supplemented with both p*hfq* and p*csrA*, produced *sepL* mRNA levels akin to that of backgrounds supplemented only with p*hfq* (Figure 2.12). The already rapid rate of *sepL* transcript degradation was shown to be further enhanced by Hfq overexpression with no detectable mRNA two minutes

after rifampicin-induced inhibition of transcription. p*csrA* supplementation did not alter *sepL* mRNA stability and overexpression of both Hfq and CsrA resulted in degradation pattern statistically indifferent to that of backgrounds supplemented with p*hfq* alone.

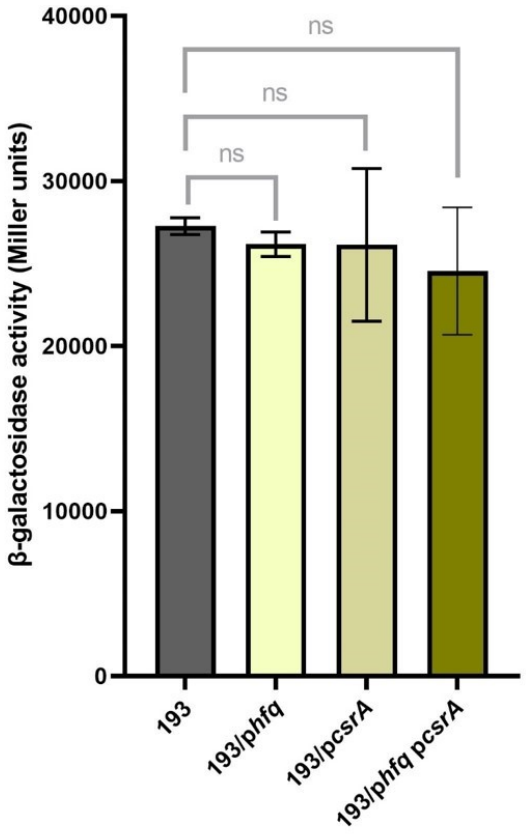
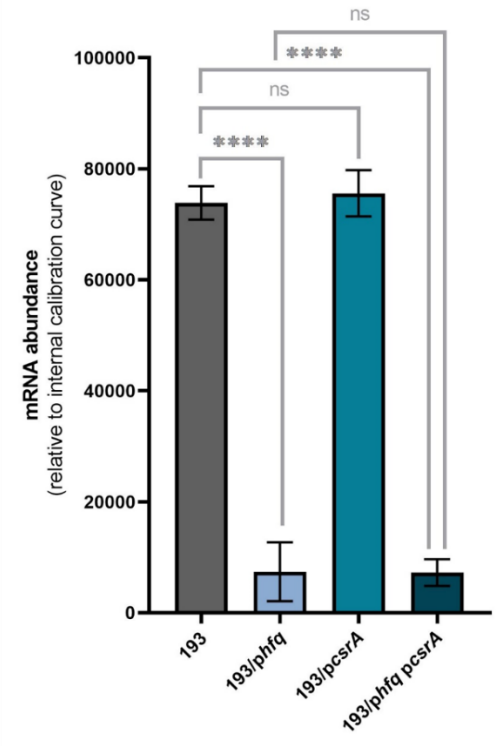


Figure 2.11. Impact of Hfq and CsrA on LEE4 transcription. Transcriptional activation of LEE4 promoter as measured by LEE4::*lacZ* transcriptional fusion indicated no significant differences in LEE4 transcription in EHEC ZAP193 background supplemented with combinations of pWSK*hfq* construct expressing Hfq (p*hfq*) and pDW66 construct expressing CsrA (p*csrA*) (biological repeat n=3, technical repeat n=6; unpaired t-test analysis: ns (not significant) – p>0.05)

a.


ns
****
ns
****
100000
80000
60000
40000
20000
0
mRNA abundance
(relative to internal calibration curve)
193
193/phfq
193/pcsrA
193/phfq pcsrA

b.

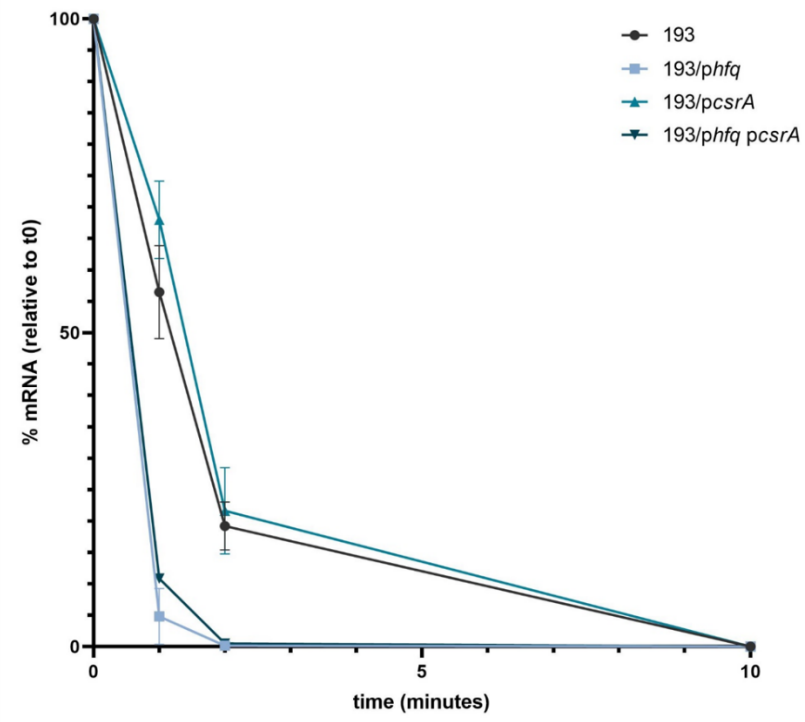

100
50
0
% mRNA (relative to t0)
0
5
10
time (minutes)
193
193/phfq
193/pcsrA
193/phfq pcsrA

Figure 2.12. Role of known SepL expression regulators Hfq and CsrA in *sepL* transcript maintenance. **a.** *sepL* mRNA abundance in overexpression of Hfq from a plasmid-based *hfq* clone pWSK*hfq* (p*hfq*) and CsrA from the pDW66 (p*csrA*) construct. Significant decrease in *sepL* mRNA detected was noted in Hfq-overexpression (193/p*hfq*). Supplementation with extra CsrA did not result in a change in *sepL* mRNA abundance (193/p*csrA*) and did not restore the sepL mRNA in the Hfq overexpression background (193/p*hfq* p*csrA*). (biological repeat n=3, technical repeat n≥6; unpaired t-test analysis: ns (not significant) – p>0.05; *** - p<0.001). **b.** Stability of the *sepL* mRNA in backgrounds supplemented with pWSK*hfq* (p*hfq*) and pDW66 (p*csrA*) as measured with a degradation time course assay. Overexpression of Hfq (193/p*hfq*) resulted in change in *sepL* mRNA decay pattern. *sepL* mRNA stability in EHEC O157 background overexpressing CsrA (193/p*csrA*) was akin to that of wild type strain (193). Supplementation with pDW66 (p*csra*) did not result in restoration of the *sepL* mRNA stability in the *hfq* overexpression background (193/p*hfq* p*csrA*) (biological repeat n=3, technical repeat n=6).

A negative effect of Hfq on *sepL* transcript abundance has been previously noted but was attributed to Hfq-driven negative regulation of Ler expression and hypothesised effect of Ler on LEE4 transcription (Shakhnovich et al., 2009). However the involvement of Ler or Hfq on transcription of LEE4 was not assessed in the Shakhnovich study. As indicated in subsection 2.2.1 of this thesis transcription of LEE4 in EHEC is distinct to that of other LEE operons and was shown to not be dependent on Ler and overexpression of Hfq was shown to have no effect on LEE4 promoter activity (Figure 2.11). However Ler-dependent LEE1-encoded factors were confirmed to be implicated in positive post-transcriptional regulation of *sepL* transcript maintenance (Figure 2.12). It is possible that Hfq-driven downregulation of Ler expression and thus downregulation of LEE1-encoded factors that stabilise the *sepL* mRNA is the mechanism behind the observed negative effect of Hfq on *sepL* transcript abundance. Alternatively a direct interaction of Hfq and potential Hfq-bound sRNA cargo with the LEE4 transcript leading to *sepL* mRNA degradation can be at play. A direct role of Hfq in transcript degradation is well documented and involves multiple modes of action such as translation inhibition, alteration of Poly(A) tails formation and RNase E access and aid in PNPase-driven decay, that were described in more detail in the subsection 1.1.3. (Cole et al., 1986; Hajnsdorf et al, 2000; Lalaouna et al., 2021; Masse et al., 2003; Park et al., 2021; Sledjeski et al., 2001; Vytvytska *et al.,* 2000). A study by Tree et al., 2014, documented multiple Hfq-*sepL* mRNA binding sites across the LEE mRNAs including two main sites in the UTR end of *sepL* and other throughout the LEE4 transcript. Many of these interactions have yet to be further characterised and could potentially be involved in control of *sepL* transcript degradation. Based on the results obtained in this thesis, which demonstrate highly significant decrease in *sepL* mRNA abundance but not LEE4 transcription we propose a previously unreported function of Hfq as a negative regulator of *sepL* mRNA

maintenance. However the specific mechanism underlying this control is yet to be revealed.

Despite its known positive impact on SepL expression overexpression of CsrA did not impact *sepL* mRNA transcription, abundance or stability and was unable to counter the Hfq-driven decrease in *sepL* transcript maintenance. Thus it was concluded that CsrA mode of action centres around translational regulation and not a positive modulation of *sepL* transcript stability and is likely a subsequent step in complex regulatory network controlling SepL expression. My results indicate that other regulatory players that counter the negative effect of Hfq have an important role in *sepL* transcript maintenance and in turn govern EspABD translocon production.

2.3 Discussion

Investigation of LEE4 expression has unravelled a discrepancy between transcriptional activation of LEE4 and other LEE operons. In contrast to LEE1,2,3 and 5, the LEE4 transcript was shown to be readily produced without requirement for the major transcriptional regulator Ler. High LEE4 promoter activity was also observed in DH5α, an *E. coli* background that does not encode a T3SS. This itself is an intriguing observation that indicates a need for the LEE4 transcript to be readily available prior to expression of the T3SS basal apparatus. It is possible that LEE4-encoded substrates are involved in processes other than secretion through the T3SS basal apparatus and translocon formation. There is evidence suggesting that T3SS pore forming proteins of *Salmonella enterica* SipB and SipC (homologues of LEE4-encoded EspDB) can be transported to the host cell via outer membrane vesicles (OMVs) without establishment of bacterial attachment. Translocon proteins of multiple pathogenic bacteria including EHEC have been previously shown to elicit changes in host immune response and rearrangements in the architecture of the host cytoskeleton (Chatterjee et al., 2015; Dewoody et al., 2013). It is possible that the LEE4-encoded proteins delivered to the host cell via OMVs might have alternative functions to their well characterised T3SS translocator role. However my work shows the requirement of LEE1-encoded factors for LEE4 transcript maintenance, which suggests that without the T3SS (LEE1) the LEE4 proteins cannot be produced. It would be interesting to determine if other factors are able to elicit LEE4 transcript maintenance and translocon protein expression and thereby allow T3SS-independent use of LEE4 effectors.

Another potential reason for independent transcription of LEE4 involves the tightly regulated timing of the hierarchical expression of the injectisome. My results indicated that maintenance of the LEE4

transcript and expression of the first LEE4-encoded protein SepL is dependent on a LEE1-encoded factor, or factors, other than Ler. Unpublished results produced in our laboratory show that LEE2-encoded CesD - a chaperone of LEE4-encoded EspD, negatively regulates SepL expression. Availability of the LEE4 transcript prior to production of its LEE1 and LEE2-encoded regulators might be necessary for the correct order of those LEE-dependent regulatory interactions. However this has not been determined and more research into the timing of T3SS expression and affinity of the *sepL* mRNA for its interacting partners is necessary to explore this hypothesis.

As described at length in section 1.1.1, the secondary structure of mRNA is a major player in determining transcript maintenance and translation. Modulation of the accessibility to the RBS being one of the most notable regulatory features. Ribosomal docking is a well-documented enhancer of transcript stability. Binding of the translation machinery is thought to competitively inhibit interaction with factors that mark the mRNA for degradation and the degradosome itself. Mapping of the *sepL* mRNA conformation using SHAPE methodology confirmed the *in silico* prediction proposing that the full-length *sepL* mRNA forms a 'folded' secondary structure with RBS buried between two stem loops. Although in order to obtain a fully mapped *sepL* mRNA structure, probing of further fragments of the LEE4 transcript is necessary. Experimental confirmation of the *in silico* prediction of the full-length *sepL* mRNA conformation gave us confidence that other *in silico* generated structures of *sepL* constructs that were predicted to disrupt the *sepL* mRNA secondary structure were likely correct. One such construct - pDW26 produces a SepL-GFP translational fusion of *sepL* mRNA truncated at base 51' and was predicted to form an 'open' conformation with an exposed RBS. Both the transcript maintenance and SepL-GFP protein production of the 'open' construct were shown to be markedly increased compared to the full-length SepL-GFP fusion. This indicated that the full-length *sepL* transcript 'folded' conformation

is likely inhibitory to translation and requires ‘opening’ to allow for translation (likely through enhancement in ribosomal binding) and without a change in the conformation the ‘folded’ transcript will be targeted for degradation. However it is also possible that the truncation of the *sepL* mRNA removed parts of the mRNA sequence or a structural signature, other to that affecting ribosomal access, that is necessary for interaction with a negative regulator of *sepL* mRNA stability and thus resulting in an increase in transcript maintenance.

The latter hypothesis is supported by the results obtained from study of *sepL* mRNA maintenance in different *E. coli* backgrounds. The inhibitory feature of the *sepL* mRNA secondary structure was noted in EHEC backgrounds but not the laboratory *E. coli* strain DH5α that was able to produce and maintain full-length *sepL* mRNA in quantities comparable to that of the ‘open’ transcript. However despite the high transcript abundance no SepL-GFP protein was produced in DH5α background from the ‘folded’ mRNA. This shows that folded *sepL* mRNA can be maintained without translational activation and thus without ribosomal docking-driven transcript stabilisation. This implicates some EHEC-encoded factors, that are absent in DH5α, as responsible for rapid degradation of ‘folded’ *sepL* mRNA in the EHEC O157 background. Prophage-encoded sRNAs unique to EHEC are plausible candidates for the degradation promoting function. Multiple pathways involving prophage-encoded sRNAs were shown to post-transcriptionally regulate bacterial mRNA abundance by altering its susceptibility to degradation machinery (Bandyra et al., 2012). It is possible that apart from the obstruction of the RBS by the ‘folded’ conformation of the *sepL* mRNA, *sepL* mRNA conformation-dependent binding of an EHEC-encoded factor antagonises translation initiation and results in loss of the ribosome docking-driven stabilisation of the transcript. This mode of action of prophage-encoded sRNA has been recently documented in study by Waters et al., that assayed RNA-RNA interactions on the scaffold component of mRNA degradation

machinery - RNase E, and revealed that a prophage encoded sRNA Esr41 binds to RBS of multiple distinct mRNAs, inhibits their translation and targets the bound transcript for RNase E-mediated cleavage (Waters et al., 2017). In order to determine if *sepL* transcript maintenance is under such sRNA-driven control it would be interesting to assay *sepL* mRNA-sRNA interactome via methodologies such as cross-linking, ligation, and sequencing of hybrids (CLASH), select the EHEC-specific sRNA hits and characterise them for their impact on *sepL* mRNA maintenance (Liu et al., 2017).

Taken together results obtained from studying *sepL* transcript expression and maintenance in a non-EHEC O157 background suggest that dependence of *sepL* transcript stability on the adopted conformation is an EHEC-specific trait. However, it is worth noting that the 'folded' *sepL* mRNA was unable to be translated in the nonEHEC background, while the 'open' *sepL* mRNA construct produced high levels of SepL protein. This suggests an additional positive regulation of *sepL* mRNA translation that is also governed by EHEC-specific factors.

The hypothesised sRNA-driven alteration of transcript stability could be facilitated by the estabilished RNA chaperone Hfq. Binding of Hfq to the *sepL* transcript was demonstrated by Tree et al. and Wang et al. (Tree et al., 2014, Wang et al., 2018). Two principal and one minor Hfq-*sepL* mRNA binding sites have been mapped at the 5'UTR of *sepL* transcript. The study by Wang et al. proposed that Hfq-recruited sRNA Spot42 acts as an off switch in *sepL* mRNA translation. However that interaction was indicated to take place after initial translational activation by a positive regulator of SepL expression - CsrA. Wang et al. suggested that the 'folded' conformation of *sepL* mRNA is the primary factor responsible for translational repression. My study confirms that statement and elaborates on it by indicating that the 'folded' *sepL* mRNA structure is targeted for degradation unless LEE1-

encoded factors are expressed. Interestingly investigation of *sepL* mRNA stability in Hfq overexpression backgrounds indicated that the observed degradation process is Hfq-dependent and was shown to involve factors encoded by EHEC but not DH5α background. This indicates that apart from the previously documented Hfq-Spot42-dependent mechanism of *sepL* mRNA translational repression, Hfq is also involved in the initial post-transcriptional regulation of *sepL* mRNA maintenance that requires the 'folded' transcript conformation. Crossover assessment of Hfq-sRNA interactions, previously mapped by Tree et al., 2014, that overlap with the proposed EHEC background-specific sRNA-*sepL* mRNA interactome study, could further narrow down the pool of potential sRNAs involved in *sepL* mRNA degradation.

A general yet major take away from studying LEE4 expression is the necessity to consider all stages of the transcript life before concluding the nature of regulatory interactions acting upon it. Transcriptional activation does not always translate into transcript availability and protein production often doesn't mirror transcript abundance. Assaying promoter activity is a powerful tool in determining changes in gene expression dependent on different environmental conditions, stages in the life cycle or bacterial background. It also allows for identification of transcriptional regulators such as Ler. However as demonstrated in this chapter and in many other studies measuring of transcriptional activation using tools such as transcriptional fusions, gene expression does not necessarily give an accurate account of transcript abundance. I consider that the debate over the impact of Ler on LEE4 expression stemmed from such an assumption. Additionally it is worth highlighting the importance of distinguishing between the post-transcriptional and the translational control. This is crucial for determining the complexity of expression control. Results obtained in this study suggest that LEE4 mRNA is under layered regulation involving alteration of transcript stability and translational control.

The following chapters investigate the potential involvement of LEE1-encoded factors in the proposed mechanism controlling *sepL* transcript degradation and address the characteristics of the *sepL* mRNA that are utilised in the regulatory interplay.

Chapter 3 CesAB-mediated post-transcriptional regulation of SepL expression

3.1 Background

Investigation into regulation of LEE4 transcript stability indicated a necessity for LEE1-encoded factor/s in *sepL* mRNA maintenance. As demonstrated in Chapter 2, the transcriptional regulator Ler is not directly involved in post-transcriptional control of the LEE4 transcript. Apart from Ler, the LEE1 operon encodes several structural proteins of the basal apparatus, the EscE T3SS needle chaperone, and notably CesAB a chaperone of T3SS translocon components (Figure 3.1a). CesAB is a non-structural cytoplasmic protein that associates with LEE4-encoded EspA filament and EspD pore-forming proteins as well as SepL. Given this association, it was considered to be the most likely LEE1-encoded candidate for an intrinsic regulatory function on the LEE4 transcript (Figure 3.1b).

CesAB is a 12.298 kDa protein that belongs to the family of T3SS translocator chaperones, which mediate secretion and stabilisation of one or more translocon proteins (Menard et al., 1994; Neyt et al., 1999; Tucker et al., 2000). Most of the research concerning CesAB has been conducted in an EPEC background, E2348/69. Secretion of the LEE4-encoded filament forming protein EspA and pore-forming EspB were shown to be CesAB-dependent. Both EspA and EspB require other chaperones to be secreted - CesA2 and EscA for EspA and CesD for EspB. CesAB is considered to be a primary chaperone for EspA but was concluded to be auxiliary for EspB translocation (Creasey et al., 2003; Portaliou et al., 2017). Studies on CesAB-dependent translocation of EspA filaments revealed that binding of EspA to CesAB

results in CesAB conformational changes that expose the CesAB C-tail (Chen et al., 2013). The LEE4-encoded SepL gatekeeper protein and LEE3-encoded EscV were shown to form a bipartite receptor for CesAB-bound EspA (Portaliou et al., 2017). The interaction between the C-tail of EspA-bound CesAB and EscV-associated SepL is necessary for EspA targeting to the basal apparatus. Recognition of the SepL-EscV receptor precedes the CesAB-dependent docking of the CesAB-EspA complex to the T3SS translocase - EscN ATPase, which is followed by dissociation of the translocon protein from the chaperone and finally EspA secretion (Portaliou et al., 2017). Additionally CesAB plays a role in inhibition of aggregation and premature degradation of EspA filament protein (Chen et al., 2013). However apart from its post-translational stabilising action on EspA, no account of CesAB-mediated regulation of protein expression has been documented to date.

This thesis chapter examines the involvement of CesAB in control of LEE4 mRNA stability in EHEC and contemplates the chaperone's duality as a translational regulator of the transcript that encodes the CesAB's protein cargo.

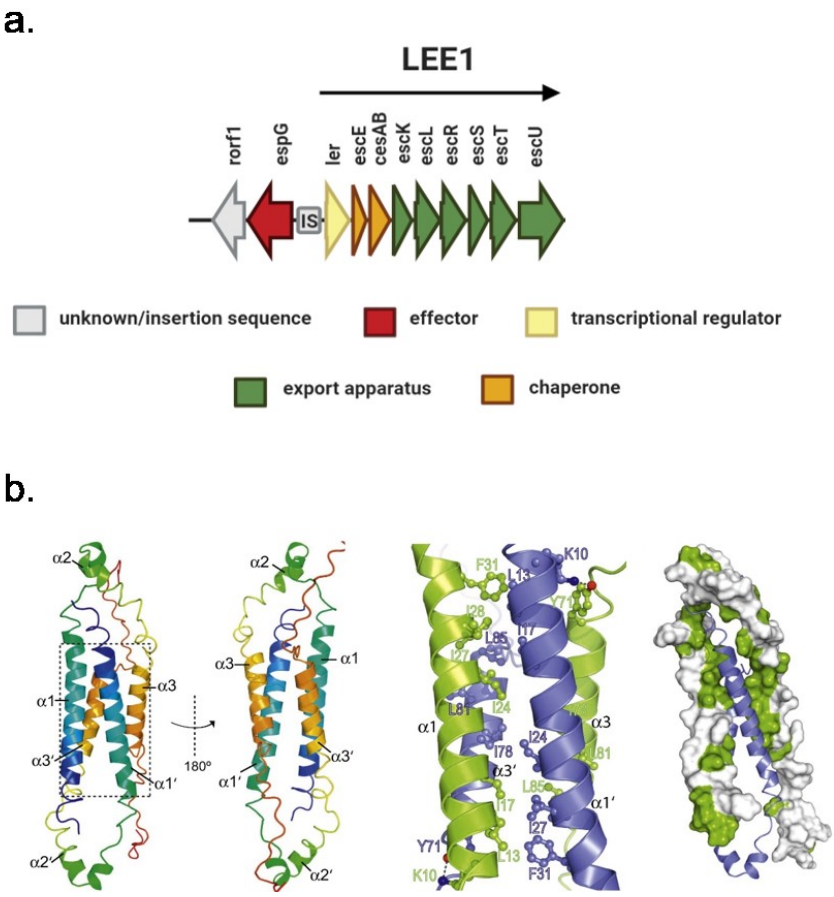


Figure 3.1. a. Schematic of the LEE1 operon arrangement. Third gene encoded on LEE1 is 314 base pair *cesAB*. *cesAB* encodes a chaperone of T3SS translocon components EspA and EspB (Creasey et al., 2003). **b.** Schematic of CesAB from EPEC. CesAB is composed of 3 α-helixes and forms a dimer. The dimer interface is populated by hydrophobic residues (green) that project into the centre of the dimer to form a hydrophobic core. The CesAB dimer monomerises once in contact with its binding partner, EspA (Chen et al., 2013). [Adapted from Chen et al., 2011].

3.2 Results

3.2.1 CesAB promotes *sepL* mRNA maintenance

In order to determine if CesAB is the LEE1-encoded factor involved in *sepL* transcript maintenance, a plasmid encoding *cesAB* (pWSK*cesAB*) was produced using the NEBuilder HiFi DNA assembly kit the ΔLEE1 Sakai as well as *Δler* ZAP193 low and high secretor EHEC strains respectively were transformed with the produced construct. This experiment was intended to investigate if this CesAB complementation is able to increase SepL expression, potentially by stabilising the *sepL* transcript. The pWSK expression vector was chosen due to its low copy number and that isopropyl-β-d-thiogalactopyranoside (IPTG) can be used to induce expression of the cloned genes. LEE4 transcript maintenance was determined by assessing *sepL* mRNA abundance by RT qPCR in these backgrounds with and without supplementation with pWSK*cesAB* (Figure 3.2). The backgrounds that were not supplemented with pWSK*cesAB* were transformed with empty pWSK29 vector to ensure any observed differences were solely due to complementation with *cesAB*. Significant increases in *sepL* mRNA levels were noted in both backgrounds supplemented with pWSK*cesAB*. However the restoration was not full when compared to the original wild type Sakai and ZAP193 background. The incomplete complementation of *sepL* mRNA in ΔLEE1 and *Δler* backgrounds might be due to the impact of those deletions on additional regulatory pathways controlling *sepL* mRNA expression or RNA stability. Additionally other factors such as absence of T3SS basal apparatus that localisation to might be important for CesAB-*sepL* mRNA interaction can potentially result in lack of full complementation of the *sepL* mRNA abundance in ΔLEE1 and *Δler* backgrounds.

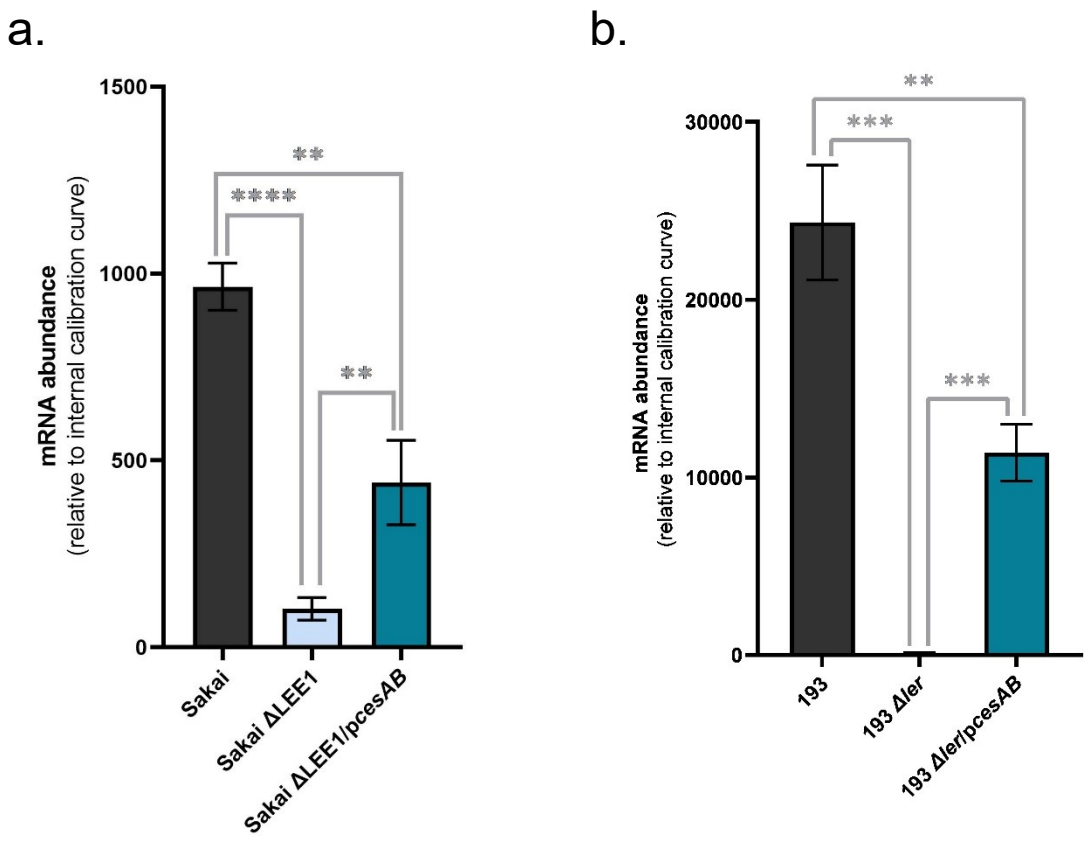


Figure 3.2. Detection of *sepL* mRNA in Ler deficient **a.** ΔLEE1 Sakai and **b.** *Δler* ZAP193 backgrounds supplemented with pWSK*cesAB* (p*cesAB*) CesAB overexpression construct. Supplementation with CesAB alone was sufficient to significantly increase the *sepL* transcript abundance in both Ler deficient EHEC backgrounds (biological repeat n=3, technical repeat n≥9). (unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; *** - $p<0.001$; **** - $p<0.0001$).

In order to determine if the observed changes in *sepL* transcript abundance are due to hypothesised post-transcriptional changes in *sepL* mRNA maintenance or the transcriptional regulation, LEE4*::lacZ* transcriptional activation with and without pWSK*cesAB* supplementation was measured (Figure 3.3). LEE4 transcription was shown to be comparable in all of the tested backgrounds and thus the observed differences in *sepL* mRNA abundance in ΔLEE1/Δ*ler* EHEC supplemented with pWSK*cesAB* are likely due to the impact of CesAB on *sepL* mRNA stability. This is consistent with findings by Creasey et al., 2003, which concluded that CesAB does not transcriptionally regulate the LEE4 operon.

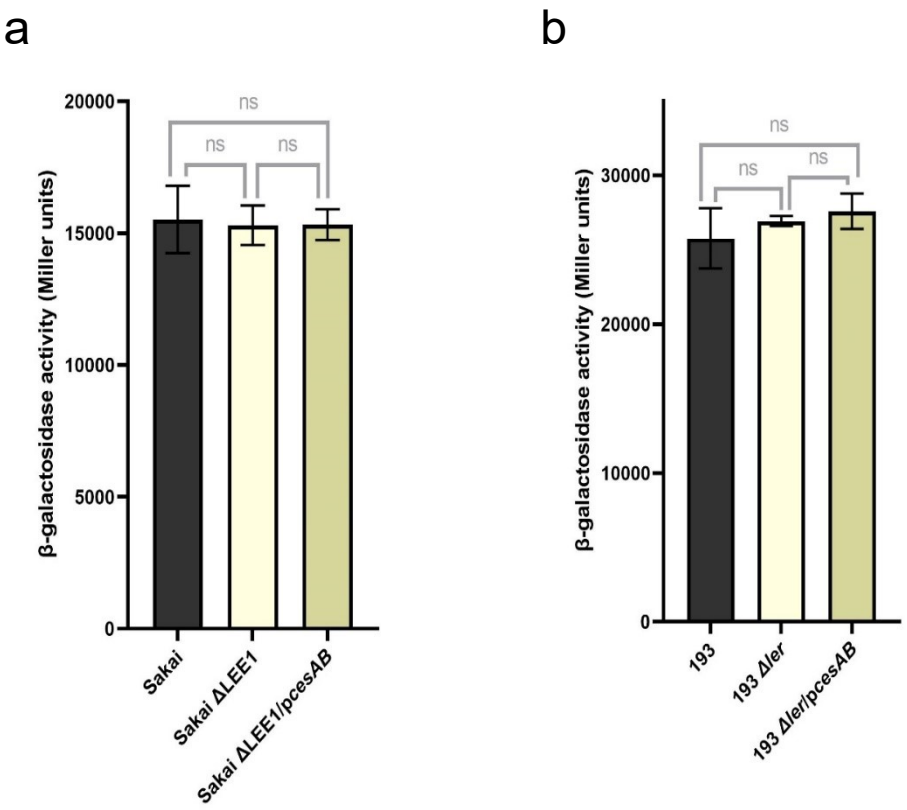


Figure 3.3. Transcriptional activation of LEE4 promoter in Ler defincient a. Sakai and b. ZAP193 EHEC backgrounds supplemented with pWSK*cesAB* (ΔLEE1/p*cesAB* and *Δler*/p*cesAB*). The LEE4*::lacZ* promoter activity was concluded to be unchanged by LEE1 and *ler* deletion with or without pWSK*cesAB* supplementation (biological repeat n=3, technical repeat n≥6). (unpaired t-test analysis: ns (not significant) – p>0.05)

In order to determine the necessity for CesAB in *sepL* transcript maintenance ZAP193 and Sakai *cesAB* knock-out deletion mutants (*ΔcesAB*) were produced by allelic exchange as originally described by Link et al., and optimised by Merlin et al.. This recombineering method makes use of the naturally occurring allelic exchange between homologous sequences of the region targeted for deletion (*cesAB*) and regions flanking the deletion sequence that were cloned into a supplemented replacement vector (pTOF25) carrying chloramphenicol resistance (*cat*), temperature sensitivity (*repA101*^{TS}) and sucrose sensitivity (*sacB*) markers. Tetracycline resistance cassette (*tet*) that was inserted in between the flanking regions, as well as chloramphenicol resistance marker (*cat*) on the pTOF25 vector allow for selection of the initial transformants. The incorporation into chromosome protects the recombinants from temperature sensitive phenotype. Therefore the bacteria undergoing first homologeous recombination were selected by incubation at 42 ºC using the temperature sensitivity marker carried on pTOF25 (*repA101*^{TS}). Secondary recombinants were selected by incubation at 42 ºC in presence of sucrose and tetracycline. The colonies obtained were screened for chloramphenicol sensitivity and tetracycline resistance indicative of the excision of the pTOF25 plasmid and incorporation of the tetracycline cassette in place of the sequence targeted for the deletion. The tetracycline cassette can be excissed using flippase and the homologeous flippase recognistion target sequences and subsequent screening of the colonies for tetracycline resistance (Figure 3.4). A tetracycline resistance cassette (*tet*) was used instead of the kanamycin resistance marker originally suggested by Merlin et al., because of the kanamycin resistance in Sakai strain, present due to original modification used to make the strain Stx negative. After careful selection and confirmation of the obtained recombinants by PCR targeting internal *cesAB* sequence and sequences flanking the *cesAB* region, as well as subsequent sequencing of the produced

backgrounds, the described approach yielded *ΔcesAB* ZAP193 and Sakai backgrounds.

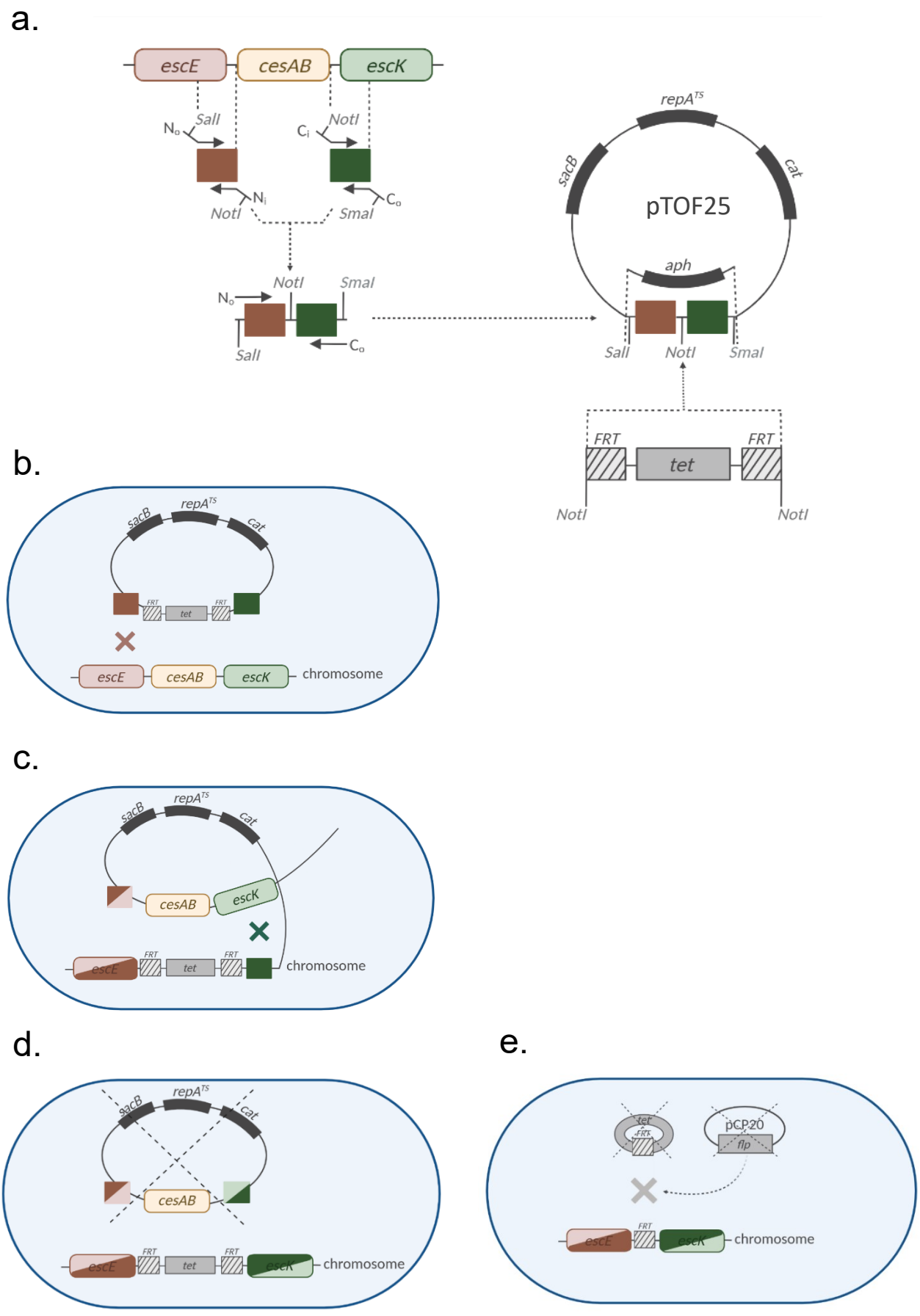
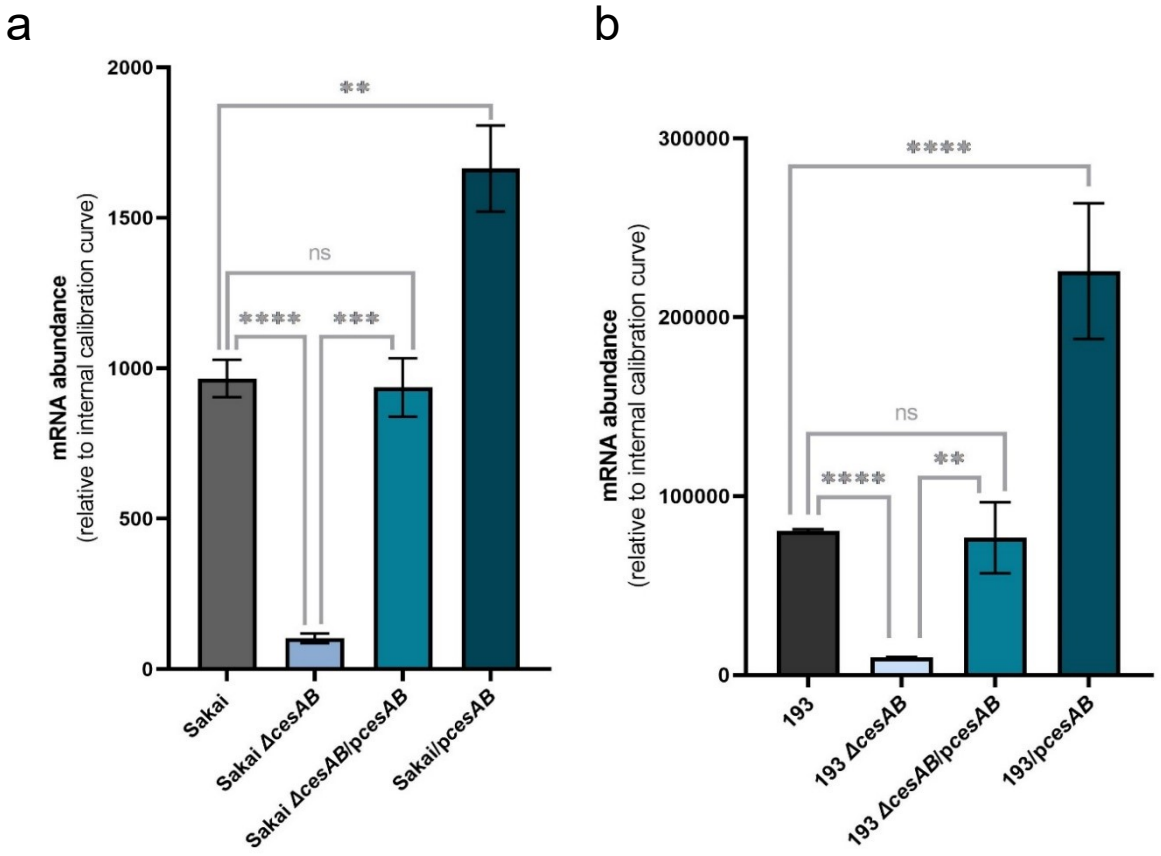

a.
escE
cesAB
escK
SalI
NotI
SmaI
repA^TS
sacB
cat
pTOF25
aph
FRT
tet
b.
chromosome
c.
d.
e.
pCP20
flp

Figure 3.4. Schematic of the plasmidic allelic exchange between genomic *cesAB* and replacement vector construct. **a.** amplification of DNA flanking *cesAB* (fragments of *escE* and *escK* genes) from N and C terminals with primers containing sequences of NotI, SmaI and SalI restriction sites (Ni – N-terminus internal and No – N-terminus outer, Ci – C-terminus outer and Ci – C-terminus internal). Subsequent crossover amplification of the amplification products with No and Co primers to join the *cesAB* flanking fragments containing SmaI and SalI restrictions sites and NotI restriction site in between the N-terminus and C-terminus fragments. Followed by cloning of the NoCo product into pTOF25 vector using the SalI and SmaI restriction sites, replacing the pTOF25 ampicillin resistance gene (amp). Insertion of the tetracycline cassette with FRT flanked with NotI restriction site sequence (from pTOF1) in between the cesAB flanking fragments. The markers on the pTOF25 replacement vector allow for selection of **b.** replacement vector transformants (*tet* and *cat*); **c.** first homologous recombinants with the entire replacement vector incorporated into the genome, which prevents temperature sensitivity ($repA101^{TS}$); **d.** secondary homologous recombinants, that no longer contain the replacement vector and thus are tetracycline resistant (*tet*), chloramphenicol sensitive and are not affected by sucrose or higher temperatures. **e.** the tetracycline resistance cassette can then be excised using FLP-FRT system [Adapted from Merlin et al., 2010].

sepL mRNA abundance in *ΔcesAB* backgrounds as measured by RT qPCR was shown to significantly decrease in levels similar to those reported in *Δler* ZAP193 and ΔLEE1 Sakai. This was fully reversible by complementation with the CesAB expression construct (pWSK*cesAB*) (Figure 3.5a&b). Additionally overexpression of CesAB (pWSK*cesAB*) resulted in significant increase in *sepL* mRNA abundance in both backgrounds, including ZAP193, despite having already high native ZAP193 *sepL* mRNA levels. In order to determine if the CesAB-dependent post-transcriptional regulation of *sepL* mRNA is due to a change in mRNA degradation pattern*, sepL* transcript was probed in *ΔcesAB* and pWSK*cesAB* supplemented backgrounds by RT qPCR over a time course after addition of the transcriptional antagonist rifampicin (Figure 3.5c). Deletion of *cesAB* was shown to result in more rapid degradation of *sepL* transcript while CesAB overexpression prolonged *sepL* mRNA maintenance in both EHEC strains. All of the tested backgrounds exhibited the same LEE4 promoter activity as tested using LEE4*::lacZ* transcriptional fusion, which again ensured that observed differences in *sepL* mRNA abundance and stability are due to CesAB-dependent post-transcriptional regulation of *sepL* mRNA (Figure 3.6).


a
b
mRNA abundance
(relative to internal calibration curve)
2000
1500
1000
500
0
**
ns
****
***
Sakai
Sakai ΔcesAB
Sakai ΔcesAB/pcesAB
Sakai/pcesAB
300000
200000
100000
0
****
ns
****
**
193
193 ΔcesAB
193 ΔcesAB/pcesAB
193/pcesAB

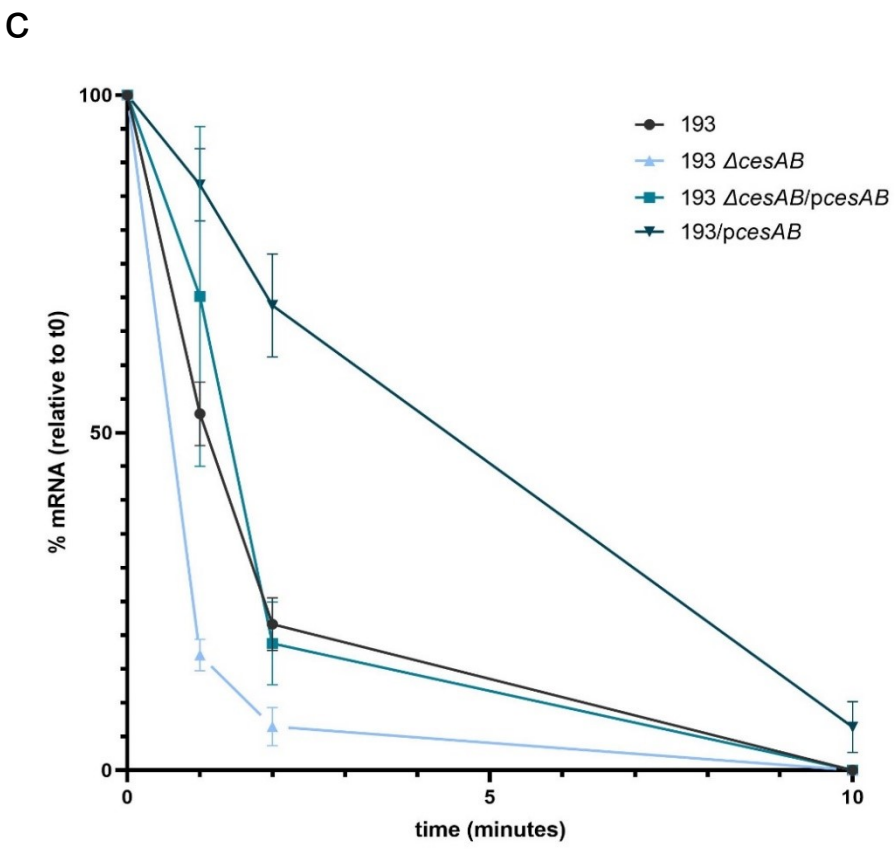

c
% mRNA (relative to t0)
100
50
0
0
5
10
time (minutes)
193
193 ΔcesAB
193 ΔcesAB/pcesAB
193/pcesAB

Figure 3.5. *sepL* mRNA abundance in **a.** low secretor Sakai and **b.** high secretor ZAP193 EHEC backgrounds with knock-out mutations of *cesAB* (*ΔcesAB*), cesAB complementation with pWSK*cesAB* (*ΔcesAB* /p*cesAB*) and CesAB overexpression (p*cesAB*) from pWSK*cesAB* construct. As compared to wild type *sepL* mRNA levels detected in Sakai and ZAP193. Significant decreases in *sepL* mRNA levels was measured in *ΔcesAB* backgrounds and the levels were restored by pWSK*cesAB* supplementation. Overexpression of CesAB (pcesAB) enhanced the *sepL* mRNA levels in both low and high secretor EHEC backgrounds. (unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; *** - $p<0.001$; **** - $p<0.0001$). (biological repeat n=3, technical repeat n≥9) **c.** stability of *sepL* mRNA in *ΔcesAB* ZAP193, *ΔcesAB* ZAP193 supplemented with pWSK*cesAB* (*ΔcesAB*/p*cesAB*) and CesAB overexpression background (p*cesAB)*. Increased rate of *sepL* transcript decay was noted in *ΔcesAB*. Supplementation of *ΔcesAB* with pWSK*cesAB* (*ΔcesAB*/p*cesAB*) reverted the pattern of *sepL* mRNA degradation to that documented in wild type ZAP193. Overexpression of CesAB (p*cesAB*) resulted in enhanced maintenance of the *sepL* mRNA (biological repeat n=3, technical repeat n≥6).

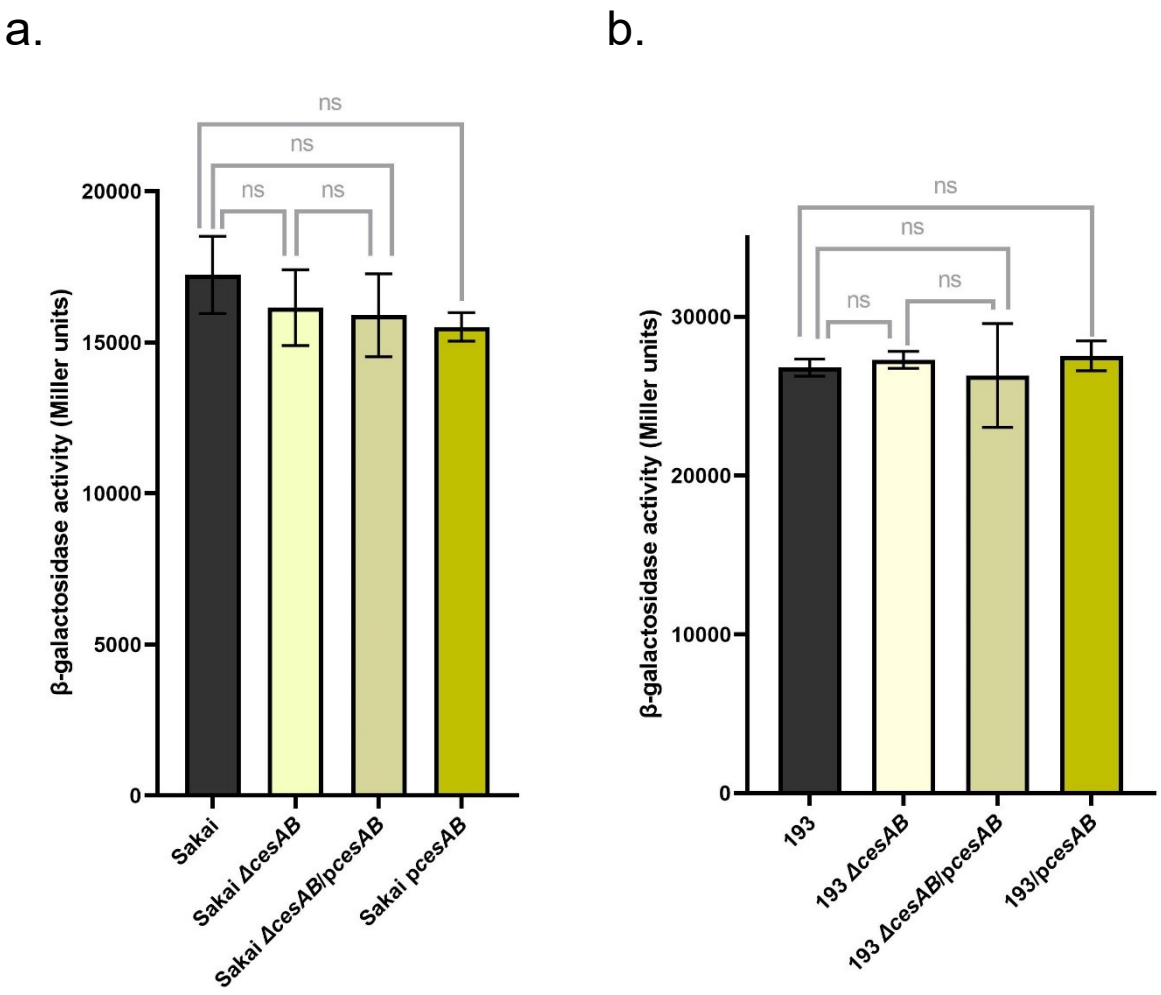
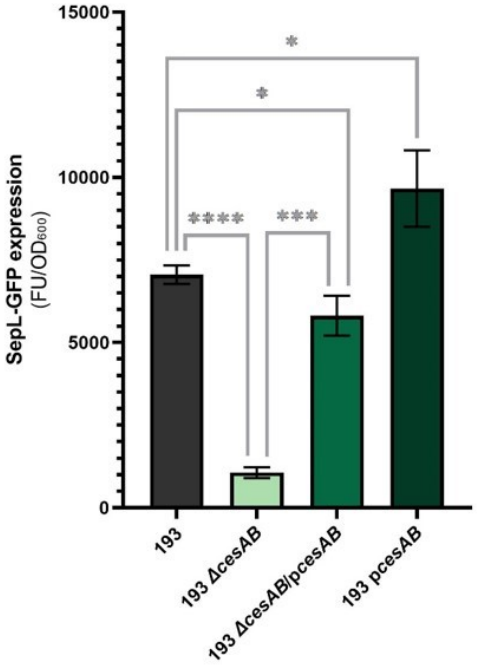


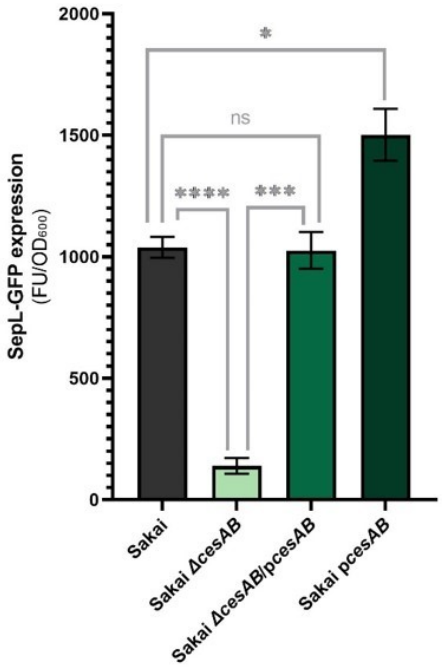
Figure 3.6. LEE4 promoter activity in *ΔcesAB* a. Sakai and b. ZAP193 EHEC backgrounds supplemented with pWSK*cesAB* (*ΔcesAB*/p*cesAB*) and under CesAB overexpression (p*cesAB*). The LEE4*::lacZ* transcription was concluded to be unchanged in the *cesAB* deletion backgrounds, or with *ΔcesAB* complementation or with CesAB overexpression (p*cesAB*) (biological repeat n=3, technical repeat n≥6). (unpaired t-test analysis: ns (not significant) – p>0.05)

In order to determine if CesAB-induced changes in *sepL* transcript abundance translate into expression of functional SepL protein, production of SepL-GFP translational fusion (pDW6) as well as expression of native SepL protein, measured by immunoblotting, were assayed in *ΔcesAB* and pWSK*cesAB* supplemented backgrounds. SepL-GFP expression was shown to mirror the decrease in *sepL* mRNA abundance in *ΔcesAB* that can be entirely (Sakai) or almost entirely (ZAP193) restored by complementation with pWSK*cesAB*. Overexpression of CesAB (pWSK*cesAB*) resulted in significant increases in SepL-GFP levels detected as compared to wild type for the ZAP193 and Sakai backgrounds (Figure 3.7a&b). This was validated by immunoblotting of SepL protein in all of the tested wild type background (no GFP reporter) using anti-SepL polyclonal rabbit serum antibody previously produced in the laboratory (Figure 3.7c&d). *ΔsepL* ZAP193 is a useful negative control in immune-detection of SepL. Due to the polyclonal nature of the SepL antibody, many unspecific bands appear on the western blot, some in close proximity to SepL and therefore the *sepL* deletion control allows for correct localisation of the SepL band in other tested backgrounds. SepL expression in the low secretor Sakai background is problematic to detect due to very low expression levels and so the blotting required extensive optimisation including changes in blocking time, primary antibody concentration, amount of protein loaded and exposure time. Typically for the purpose of detection of SepL via immunoblotting the expression of SepL in low secretor background is enhanced by supplementation with a Ler overexpression construct. However as of our understanding the positive effect of Ler overexpression on production of LEE4-encoded proteins is due to a Ler-mediated increase in CesAB production. Overexpression of Ler would result in detectable EspA from wild type Sakai background, however the *ΔcesAB* would not be affected by the supplementation. This would create an additional variable in the comparison between the wild type and *ΔcesAB* backgrounds that could not be accounted for.

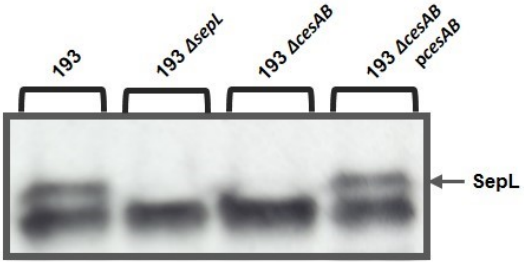
a.


SepL-GFP expression
(FU/OD600)
15000
10000
5000
0
*
*
****
***
193
193 ΔcesAB
193 ΔcesAB/pcesAB
193 pcesAB

b.


SepL-GFP expression
(FU/OD600)
2000
1500
1000
500
0
*
ns
****
***
Sakai
Sakai ΔcesAB
Sakai ΔcesAB/pcesAB
Sakai pcesAB

c.


193
193 ΔsepL
193 ΔcesAB
193 ΔcesAB pcesAB
SepL

d.

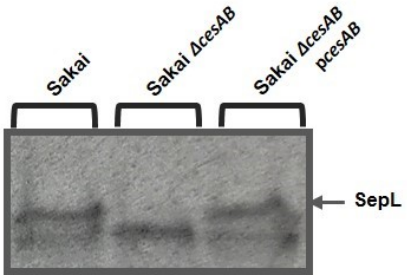

Sakai
Sakai ΔcesAB
Sakai ΔcesAB pcesAB
SepL

Figure 3.7. SepL protein production in absence and overexpression of CesAB. SepL expression was measured by fluorescence spectroscopy detecting the SepL-GFP translational fusion (pDW6) in the **a.** high secretor ZAP193 and **b.** low secretor Sakai backgrounds. SepL-GFP was measured in *cesAB* knock-out mutants (*ΔcesAB*) and the knock-out backgrounds complemented with pWSKcesAB expression construct (*ΔcesAB*/p*cesAB*). Significant decrease in SepL-GFP production was noted that could be restored by pcesAB complementation. Additionally significant increase in SepL-GFP production in CesAB overexpression backgrounds (p*cesAB*) was documented (unpaired t-test analysis: ns (not significant) – $p>0.05$; * - $p<0.05$; *** - $p<0.001$; **** - $p<0.0001$). (biological repeat n=3, technical repeat n=9). Immunoblotting analysis in **c.** ZAP193 and **d.** Sakai backgrounds detected a ~40kD band corresponding to SepL in the pWSK*cesAB* supplemented background, which validated SepL-GFP results. *ΔsepL* ZAP193 background acted as a negative control for SepL production (biological repeat n=3, technical repeat n≥3).

Additionally the impact of CesAB supplementation (pWSK*cesAB*) on SepL expression in *Δler* ZAP193, ΔLEE1 Sakai and DH5α backgrounds that were previously shown to be unable to express SepL protein was investigated. Supplementation with pWSK*cesAB* was shown to significantly increase SepL-GFP translational fusion production in *Δler*/ΔLEE1 EHEC strains (Figure 3.8a&b). Interestingly despite only partial restoration in *sepL* mRNA abundance in ΔLEE1 Sakai p*cesAB* background, expression of SepL-GFP was fully restored upon CesAB supplementation. This could be due to proximity to the detection threshold of the SepL-GFP levels produced by the low secretor profile of Sakai strain, which can decrease the accuracy of the read-outs. However the immunoblotting of SepL protein in both *Δler* ZAP193 and ΔLEE1 Sakai backgrounds showed restoration of SepL production upon supplementation with p*cesAB* (Figure 3.8c&d). Strikingly a laboratory *E. coli* background DH5α that does not encode for any other T3SS elements was able to produce detectable levels of SepL-GFP when supplemented with pWSK*cesAB* (Figure 3.9a). Supplementation of DH5α with the whole LEE1 operon resulted in an even higher increase in SepL-GFP expression (Figure 3.9a). After extensive optimisation necessary due to the low levels of the SepL protein produced this was validated by anti-SepL immunoblotting (Figure 3.9b). As shown previously full-length *sepL* mRNA is highly abundant but not translated in DH5 background (Figure 2.9a and 2.10 respectively). Expression of SepL-GFP upon supplementation with pWSK*cesAB* in DH5α background suggests potential involvement of CesAB in translational activation on top of its demonstrated role in regulation of *sepL* transcript stability.

Taken together the observed results indicate that CesAB is a positive regulator of *sepL* mRNA maintenance and potentially translation initiation. Following sections of this chapter investigate the implications of this control on other known regulatory pathways and its impact on the translocon expression.

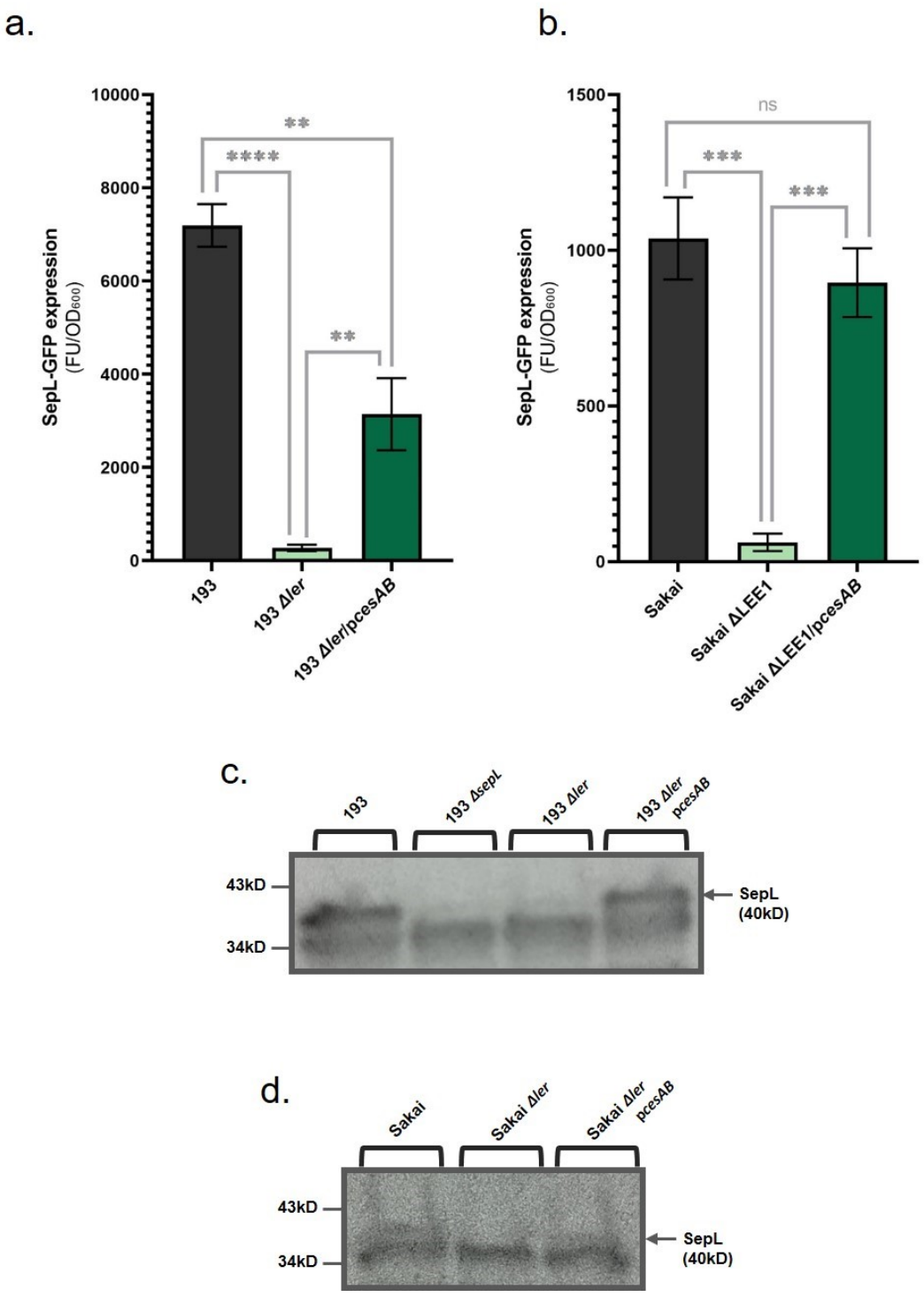
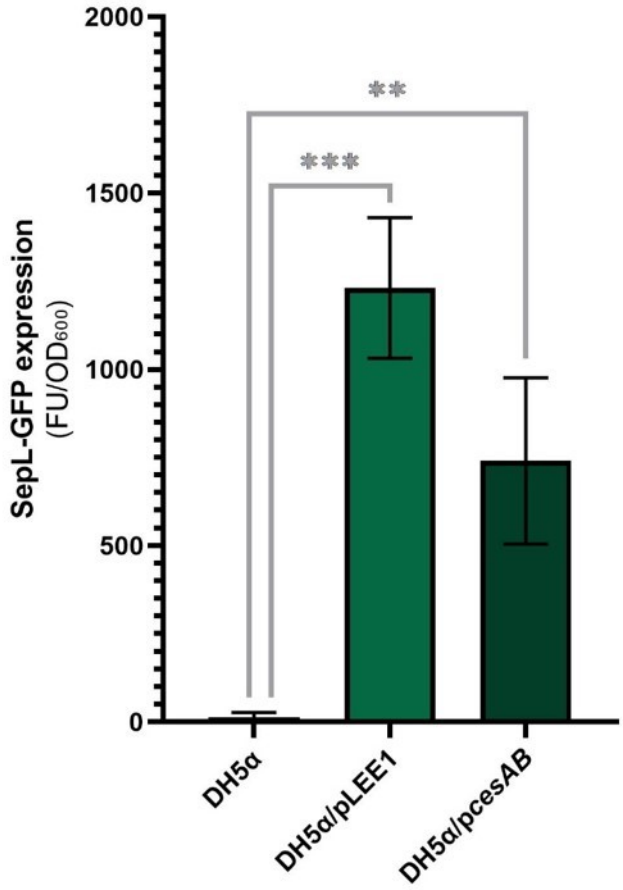

a.
SepL-GFP expression
(FU/OD600)
10000
8000
6000
4000
2000
0
**
****
**
193
193 Δler
193 Δler/pcesAB
b.
SepL-GFP expression
(FU/OD600)
1500
1000
500
0
ns
***
***
Sakai
Sakai ΔLEE1
Sakai ΔLEE1/pcesAB
c.
193
193 ΔsepL
193 Δler
193 Δler pcesAB
43kD
34kD
SepL
(40kD)
d.
Sakai
Sakai Δler
Sakai Δler pcesAB
43kD
34kD
SepL
(40kD)

Figure 3.8. SepL protein production in Ler deficient backgrounds supplemented with CesAB. SepL expression was measured by fluorescence spectroscopy detecting the SepL-GFP translational fusion (pDW6) in the **a.** high secretor ZAP193 and **b.** low secretor Sakai backgrounds. SepL-GFP was measured in *Δler* ZAP193 and ΔLEE1 Sakai and the backgrounds compleented with pWSKcesAB expression construct (*Δler*/p*cesAB* and ΔLEE1/p*cesAB*). Significant increase in SepL-GFP production was documented upon supplementation with pWSK*cesAB* in both backgrounds (unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; *** - $p<0.001$; **** - $p<0.0001$). (biological repeat n=3, technical repeat n=9). Immunoblotting analysis in **c.** ZAP193 and **d.** Sakai backgrounds detected 40kD band corresponding to SepL protein in pWSK*cesAB* supplemented background, which validated SepL-GFP results. *ΔsepL* ZAP193 background acted as a negative control for SepL production (biological repeat n=3, technical repeat n=3).

a.


2000
1500
1000
500
0
SepL-GFP expression
(FU/OD600)
**
***
DH5α
DH5α/pLEE1
DH5α/pcesAB

b.

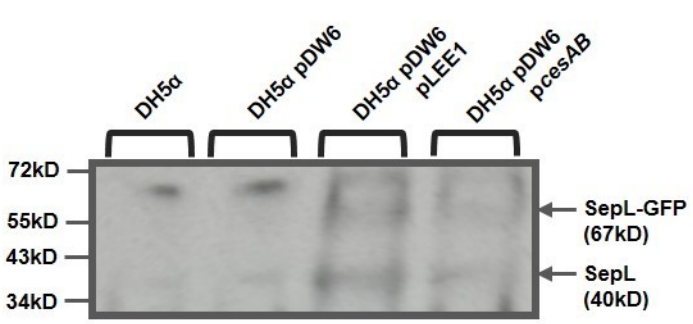

DH5α
DH5α pDW6
DH5α pDW6 pLEE1
DH5α pDW6 pcesAB
72kD
55kD
43kD
34kD
SepL-GFP (67kD)
SepL (40kD)

Figure 3.9. SepL protein production in nonT3SS *E. coli* background DH5α supplemented with LEE1 and CesAB. **a.** SepL expression was measured by fluorescence spectroscopy detecting the SepL-GFP translational fusion (pDW6) in the DH5α background with or without pWSKLEE1 (pLEE1) or pWSK*cesAB* (p*cesAB*). Significant increases in SepL-GFP were detected in the pWSKLEE1 and pWSKcesAB supplemented DH5α. (unpaired t-test analysis: ns (not significant) – $p>0.05$; * - $p<0.05$; *** - $p<0.001$; **** - $p<0.0001$). (biological repeat n=3, technical repeat n=9). **b.** Immunoblotting of the SepL protein documented two bands appearing in the supplemented backgrounds, the sizes of which correspond to SepL-GFP fusion (67 kD) and SepL (40 kD). SepL-GFP fusion was previously reported to be unstable and result in production of separate SepL and GFP proteins. (biological repeat n=2, technical repeat n=2).

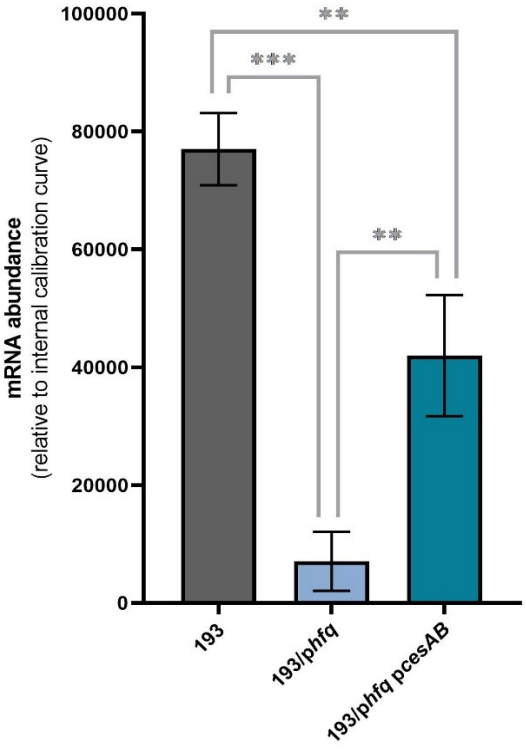
3.2.2 CesAB counters Hfq-induced degradation of *sepL* transcripts

Results obtained studying the mode of action underlying the Hfq-driven regulation of SepL expression uncovered a role for Hfq in promoting *sepL* transcript degradation (section 2.2.4). As stated above CesAB has a positive effect on *sepL* mRNA stability. Therefore it was interesting to determine if those two regulatory interactions are intertwined. In order to test this, CesAB-driven antagonism of Hfq activity was investigated. Expression and stability of *sepL* transcripts in EHEC background ZAP193 with different combinations of CesAB and Hfq overexpression were assayed by end-point RT qPCR and RT qPCR over a time-course after inhibition of transcription as described in previous sections. The ZAP193 background was chosen due to its high secretor profile that enables identification of Hfq-driven reductions in *sepL* mRNA levels and potentially the hypothesised restoration following supplementation with extra CesAB. Previously constructed CesAB and Hfq expression vectors were utilised in the pWSK29 backbone, which uses ampicillin (Ap^R) resistance as a selection marker. In order to select for successful transformation with both *hfq* and *cesAB* clones, a different vector, that harbours chloramphenicol resistance (*cat*), which was previously shown to be compatible with pWSK29-based vectors was used to produce a new CesAB expression construct (pACYC*cesAB*).

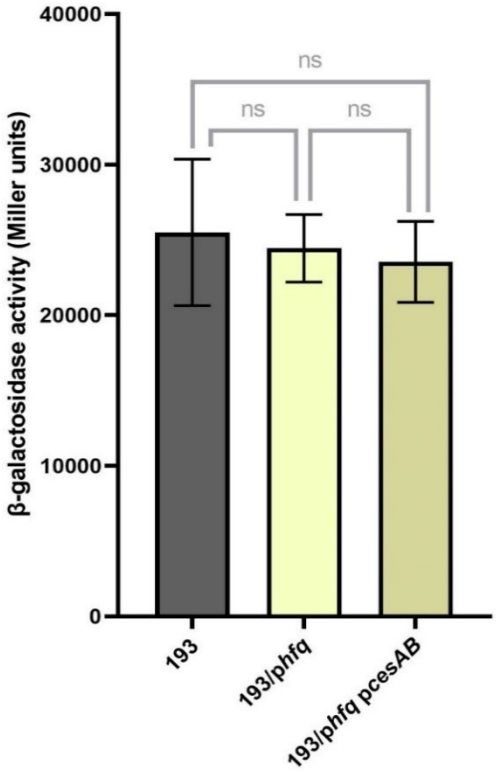
Results obtained showed that overexpression of CesAB (pACYC*cesAB*) in the EHEC ZAP193 background supplemented with pWSK*hfq* significantly increased *sepL* mRNA levels detected, that were shown to be reduced in backgrounds with only Hfq overexpression (Figure 3.10a). Additionally stability of *sepL* mRNA as measured by the RT qPCR time-course post rifampicin treatment indicated that supplementation with pACYC*cesAB* countered the negative effects of

Hfq overexpression on *sepL* transcript maintenance (Figure 3.10b). This control was again shown to be strictly post-transcriptional by measuring LEE4 promoter activity using the LEE4*::lacZ* transcriptional fusion, which remained unchanged in all pWSK*hfq* and pACYC*cesAB* supplemented backgrounds (Figure 3.10c). Taken together it can be concluded that Hfq-mediated degradation of *sepL* mRNA can be counteracted by CesAB and that antagonism of Hfq is a likely mode of action behind the CesAB-mediated positive post-transcriptional regulation of SepL expression. This could be achieved by CesAB in multiple ways: competitive binding to the Hfq-targeted *sepL* mRNA sequence; facilitation of a change in *sepL* transcript structure that leads to obstruction of Hfq-binding site or 'sponging' of Hfq away from its action on the *sepL* mRNA by direct protein-protein interaction. The specifics of the proposed CesAB-Hfq interplay could be further clarified by *sepL* mRNA-CesAB and CesAB-Hfq interaction studies, some of which were undertaken and are described in the following chapter of this thesis.

a.


mRNA abundance
(relative to internal calibration curve)
100000
80000
60000
40000
20000
0
**
***
**
193
193/phfq
193/phfq pcesAB

c.


β-galactosidase activity (Miller units)
40000
30000
20000
10000
0
ns
ns
ns
193
193/phfq
193/phfq pcesAB

b.

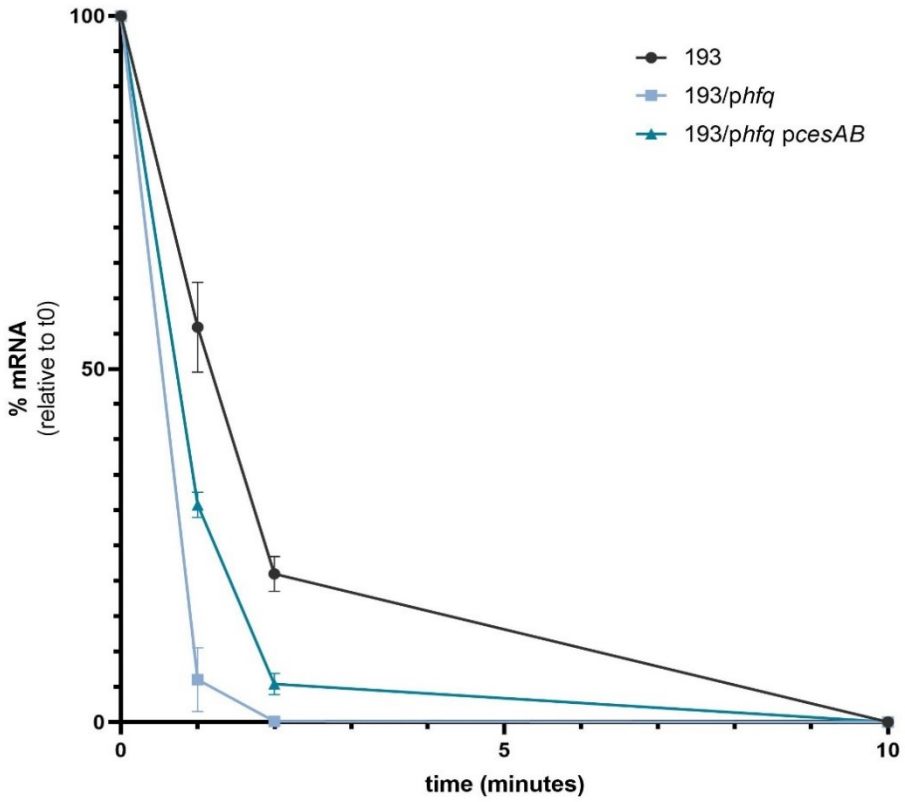

% mRNA
(relative to t0)
100
50
0
0
5
10
time (minutes)
193
193/phfq
193/phfq pcesAB

Figure 3.10. a. *sepL* mRNA abundance in ZAP193 EHEC background supplemented with pWSK*hfq* and overexpressing the negative post-transcriptional regulator Hfq with or without supplementation with pACYC*cesAB*. Expression of extra CesAB resulted in significant increase in *sepL* mRNA levels (193 p*hfq* vs. 193/p*hfq* p*cesAB*). **b.** the pattern of *sepL* mRNA degradation in backgrounds overexpressing Hfq (phfq) with or without CesAB overexpression (p*hfq* p*cesAB*). *sepL* mRNA maintenance in Hfq overexpression background was prolonged by supplementation with extra CesAB. **c.** LEE4 promoter activity in backgrounds overexpressing Hfq (193/p*hfq* and 193/p*hfq* p*cesAB*) and CesAB (193/p*hfq* p*cesAB*). The LEE4*::lacZ* transcription was unchanged *by* Hfq or CesAB overexpression. (unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; *** - $p<0.001$). (biological repeat n=3, technical repeat n≥6).

Additionally evidence for Hfq and CesAB interplay comes from unpublished results obtained by Sean McAteer that showed an increase in SepL-GFP translational fusion production (pDW6) in *Δhfq* K12 *E. coli* background that was comparable to that detected in K12 background supplemented with CesAB overexpression vector (pWSK*cesAB*). Overexpression of CesAB in *Δhfq* K12 *E. coli* did not result in further increase in SepL-GFP production. The lack of synergistic effect of *Δhfq* and CesAB overexpression suggests that the positive effect of CesAB on SepL-GFP expression is through action involving Hfq. Similar dependency was observed by Katsowitch et al. 2017. In EPEC CsrA is a negative regulator of T3SS effector NleA and CesT a T3SS chaperone positively regulates NleA expression. The study showed that in absence of CsrA, the CesT is no longer needed for NleA production (Katsowitch et al., 2017).

As previously indicated in DH5a background, *sepL* mRNA levels are high and not readily degraded as it is the case in EHEC backgrounds. However the mRNA is not translated. Detectable levels of SepL-GFP in *Δhfq* and pWSKcesAB supplemented backgrounds would suggest translational activation is achieved. This could be explained by Hfq 'locking' the *sepL* mRNA secondary structure in untranslatable state that can be derepressed by CesAB. In EHEC backgrounds there is the additional step of *sepL* mRNA degradation if the *sepL* mRNA is locked in the untranslatable secondary structure. However this is only a speculation and more in depth examination of Hfq and CesAB effects on *sepL* mRNA structure are necessary to investigate the hypothesis.

3.2.3 CesAB-dependent regulation of *sepL* mRNA impacts expression of translocon components

Apart from SepL, the LEE4 operon encodes the T3SS translocon proteins EspABD. The only confirmed promoter at which initiation of transcription of the entire LEE4 mRNA is thought to take place is situated at position -5 to -36 upstream of the *sepL* coding region (Kresse et al., 2000). However, there is an evidence for processing of the LEE4 transcript with RNase E cleavage in the 3' end of the *sepL* mRNA and before the *espA* transcript. Expression of the T3SS filament building block, EspA, was previously shown to mirror the unique heterogeneous expression of the SepL protein in EHEC backgrounds and this is thought to be controlled at a post-transcriptional level (Roe et al., 2003). In order to determine if the CesAB-induced changes in *sepL* mRNA maintenance result in a knock-on effect on EspA expression, *espA* mRNA abundance and EspA-GFP translational fusion expression and EspA protein expression as measured by anti-EspA immunoblotting, were determined in the absence and with supplementation of CesAB. Again immunoblotting of EspA was carried out only in the ZAP193 high secretor strain due to the low production and difficulty with detection of EspA in the low secretor Sakai strain.

RT qPCR of *espA* mRNA in *ΔcesAB* ZAP193 and Sakai backgrounds indicated a significant decrease in *espA* transcript abundance, that was proportionally akin to that of *sepL* mRNA. This was reversible by complementation of the *ΔcesAB* backgrounds with pWSK*cesAB* (Figure 3.11). As demonstrated previously the LEE4 promoter activity remains unchanged by deletion of *cesAB* (Figure 3.6), and thus the observed decline in *espA* transcript measured is concluded to be an effect of post-transcriptional changes. In order to determine if the observed decline in *espA* mRNA abundance in absence of CesAB has an effect on EspA protein production, expression of EspA-GFP

translational fusion (pAJR-EspA) that has been previously produced in the laboratory by Sean McAteer was measured (Figure 3.12a&b). Abolishment of EspA-GFP expression in *ΔcesAB* backgrounds was fully restored by complementation with pWSKcesAB in both high secretor ZAP193 and low secretor Sakai strains. Additionally anti-EspA immunoblotting using rabbit polyclonal anti-EspA antibody produced as described in McNeilly et al., 2010 was carried out in *ΔcesAB* ZAP193 with and without pWSKcesAB complementation. Obtained results confirmed that CesAB is necessary for EspA protein expression (Figure 3.12c). This is likely predominately due to the observed positive effect of CesAB on the availability of the LEE4 transcript.

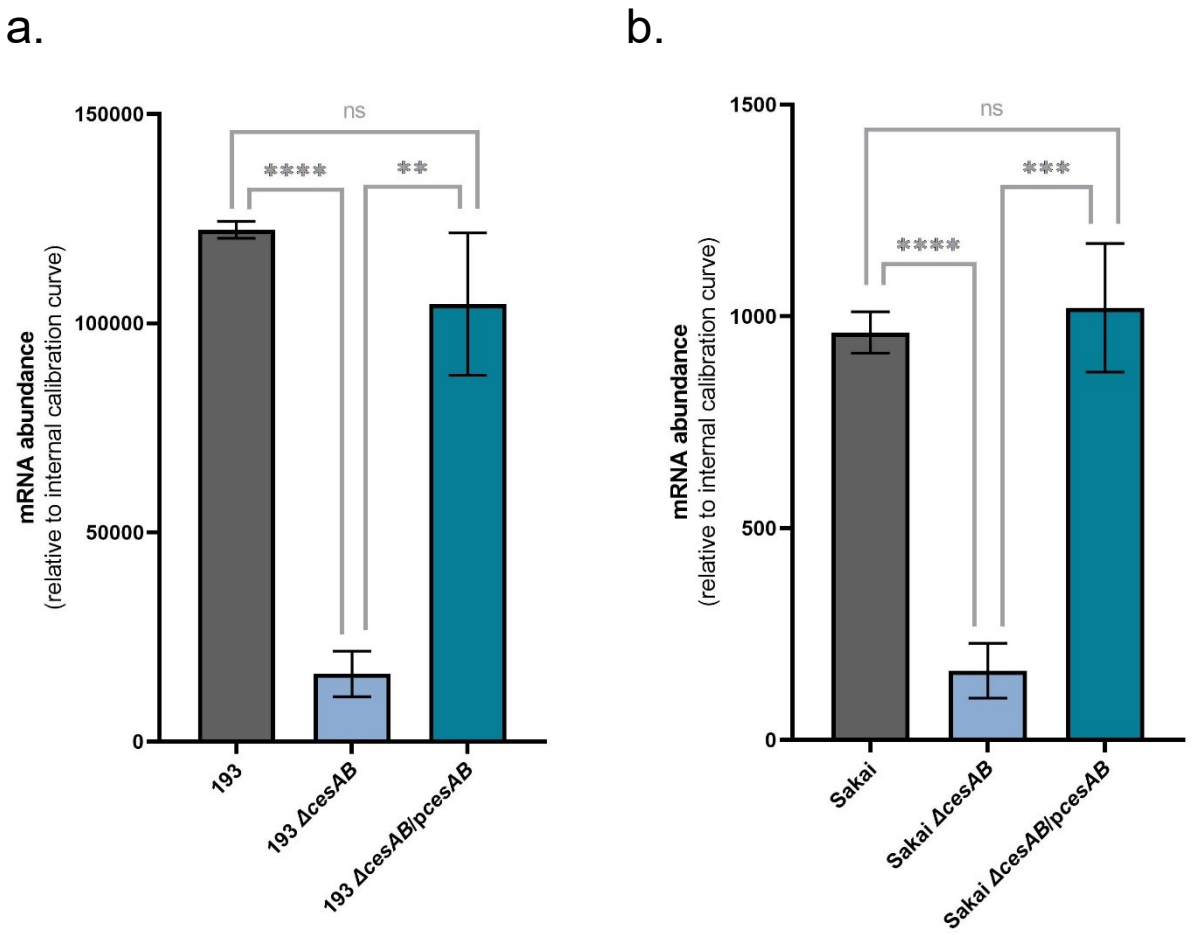


Figure 3.11. *espA* mRNA abundance in **a.** high secretor ZAP193 EHEC and **b.** low secretor Sakai backgrounds with a *cesAB* knock-out mutation with and without *cesAB* complementation. Deletion of *cesAB* (*ΔcesAB*) resulted in highly significant decrease in *espA* mRNA detected which was fully restored by supplementation with pWSKcesAB (*ΔcesAB*/p*cesAB*) (unpaired t-test analysis: ns (not significant) – p>0.05; ** - p<0.01; *** - p<0.001; **** - p<0.0001). (biological repeat n=3, technical repeat n=9).

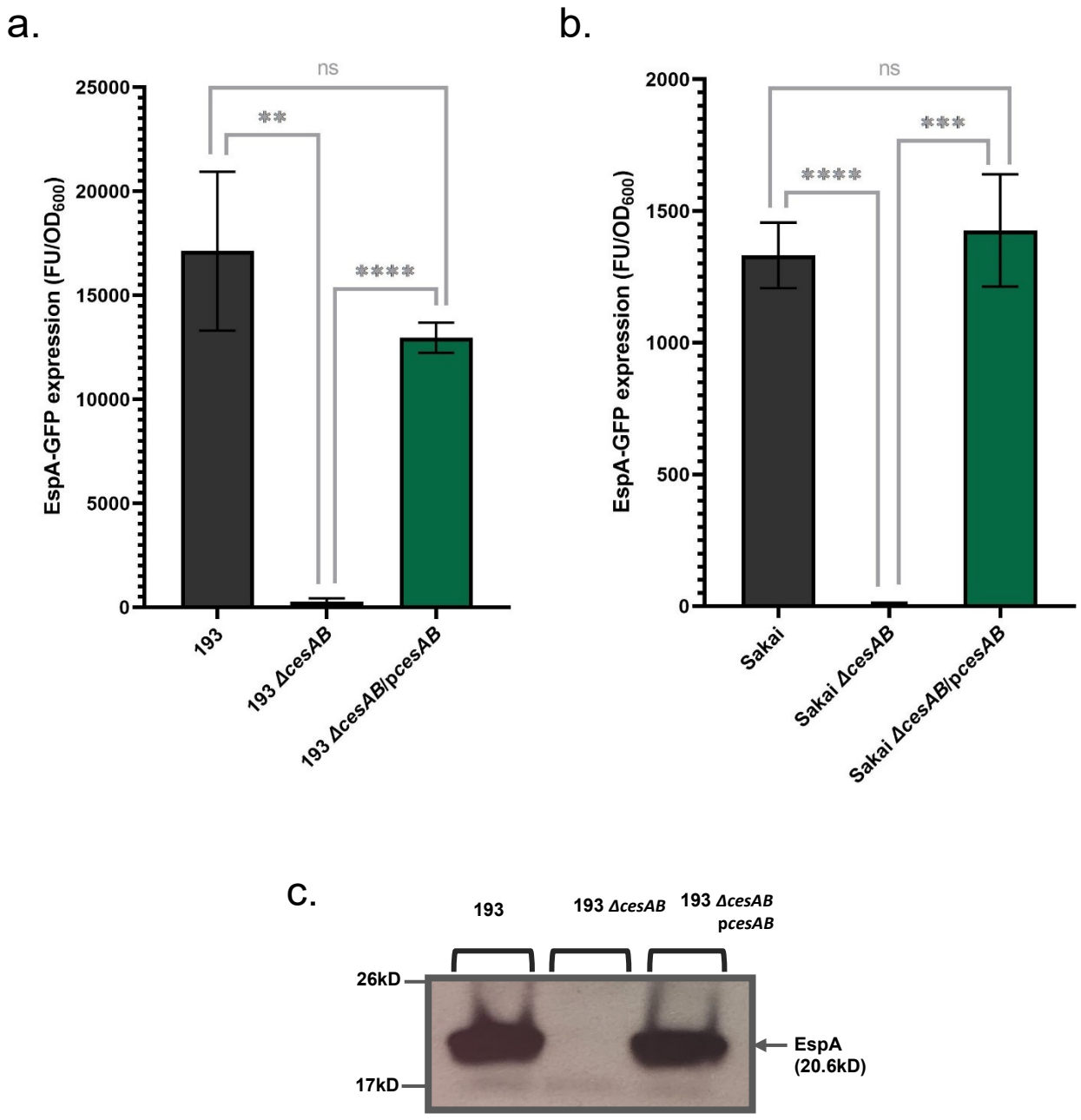

a.
b.
c.
ns
**
****
***
EspA-GFP expression (FU/OD600)
25000
20000
15000
10000
5000
0
2000
1500
1000
500
193
193 ΔcesAB
193 ΔcesAB/pcesAB
Sakai
Sakai ΔcesAB
Sakai ΔcesAB/pcesAB
193 ΔcesAB pcesAB
26kD
17kD
EspA
(20.6kD)

Figure 3.12. EspA-GFP (pAJR-EspA) translational fusion production in **a.** ZAP193 high secretor and **b.** Sakai low secretor EHEC *cesAB* knock-out backgrounds with and without CesAB complementation (pWSK*cesAB*) measured by fluorescence spectroscopy. Abolishment of EspA-GFP production was recorded in the *cesAB* deletion backgrounds (*ΔcesAB*) and was fully restored by complementation with pWSK*cesAB* (*ΔcesAB*/pcesAB). (unpaired t-test analysis: ns (not significant) – $p>0.05$; ** - $p<0.01$; *** - $p<0.001$; **** - $p<0.0001$). (biological repeat n=3, technical repeat n=6). **c.** Immunoblotting of the EspA protein documented loss of a band corresponding to the 20.6kD EspA protein in *ΔcesAB* ZAP193 backgrounds and restoration of the band and thus EspA protein production in *ΔcesAB* ZAP193 complemented with pWSKcesAB (193 *ΔcesAB*/p*cesAB)* (biological repeat n=2, technical repeat n=2).

However, CesAB is a well-documented protein-protein interacting partner of EspA that was shown to positively influence EspA protein stability and thus the observed decrease in EspA protein detection in *ΔcesAB* backgrounds cannot be solely attributed to the role of CesAB in the LEE4 transcript maintenance (Creasey et al., 2003). Therefore it was interesting to determine if expression of other LEE4-encoded components is affected by the chaperone. EspD is a pore-forming protein of the T3SS translocon that does not have any known links to CesAB. *espD* transcript abundance and EspD-GFP translational fusion (pAJR-EspD, produced by Sean McAteer) expression was determined in *ΔcesAB* ZAP193 and Sakai backgrounds with and without CesAB complementation (pWSK*cesAB*). Effect of absence and restoration of CesAB on *espD* mRNA abundance mirrored that of *sepL* and *espA* (Figure 3.13a&b). Although significantly reduced, production of EspD-GFP was not entirely abolished in *ΔcesAB* backgrounds as it was the case for EspA-GFP (Figure 3.13c&d). This indicated additional CesAB-mediated regulation of EspA protein abundance, likely through previously documented effect on EspA stability (Creasey et al., 2003).

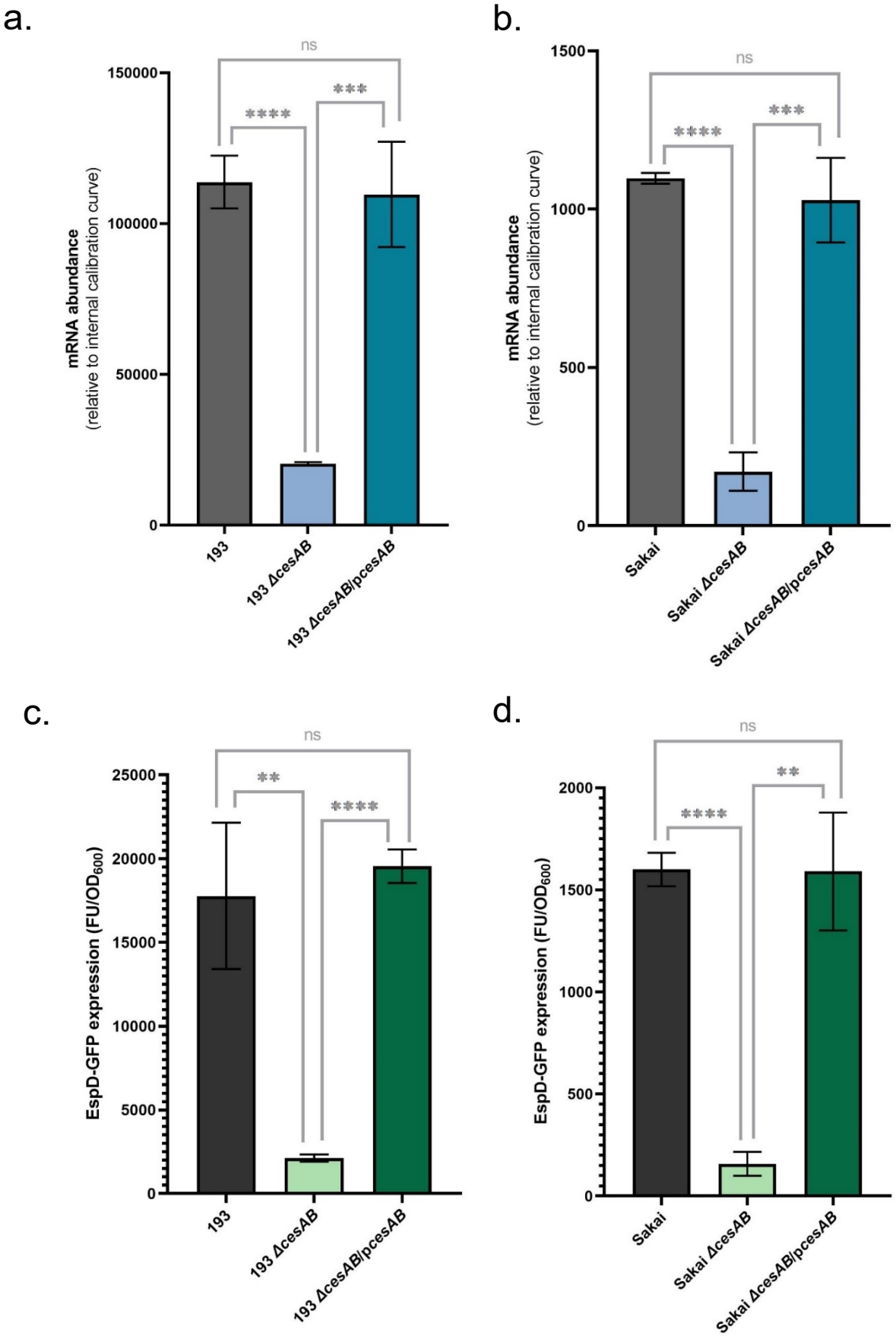

a.
b.
c.
d.
ns
****
***
**
mRNA abundance
(relative to internal calibration curve)
EspD-GFP expression (FU/OD600)
150000
100000
50000
0
1500
1000
500
25000
20000
15000
10000
5000
2000
193
193 ΔcesAB
193 ΔcesAB/pcesAB
Sakai
Sakai ΔcesAB
Sakai ΔcesAB/pcesAB

Figure 3.13. a-b. *espD* mRNA abundance in high secretor ZAP193 EHEC and low secretor Sakai backgrounds with a *cesAB* knock-out mutation with and without *cesAB* complementation. Deletion of *cesAB* (*ΔcesAB*) resulted in highly significant decrease in *espA* mRNA, which was fully restored by supplementation with pWSKcesAB (*ΔcesAB*/p*cesAB*). **c-d.** EspD-GFP (pAJR-EspD) translational fusion production in ZAP193 and Sakai *cesAB* knock-out backgrounds with and without CesAB complementation (pWSK*cesAB*) measured by fluorescence spectroscopy. Significant decrease in EspD-GFP detected was documented in *cesAB* deletion backgrounds (*ΔcesAB*) and was fully restored by complementation with pWSK*cesAB* (*ΔcesAB*/pcesAB). (unpaired t-test analysis: ns (not significant) – $p>0.05$; * - $p<0.05$; *** - $p<0.001$; **** - $p<0.0001$). (biological repeat n=3, technical repeat n≥6).

The possibility that CesAB regulates *espA and espD* transcripts stability directly and independently and not as a knock-on effect of antagonism of *sepL* mRNA degradation cannot be excluded. CRAC UV-crosslinking study by Tree et al., deposited at NCBI GEO under the accession number GSE46118, indicated multiple Hfq binding sites throughout the *espADB* region of the LEE4 transcript (Tree et al., 2014). It is possible that those sites are involved in Hfq-mediated transcript degradation that is opposed by CesAB, as is the case in the demonstrated regulatory interplay of Hfq and CesAB controlling *sepL* transcript maintenance (see section 2.2.4). However this is only a speculation and further studies concerning maintenance of *espAD* mRNA and other LEE4 transcripts are necessary to elaborate on this hypothesis.

Taken together it is apparent that expression of CesAB is necessary for maintenance of different regions of the LEE4 transcript (*sepL*, *espA* and *espD*) and that this mRNA-centred regulation has a significant impact on final concentration of T3SS translocon proteins. However if the observed regulation stems from CesAB-mediated regulation of the *sepL* mRNA or individual interaction of CesAB with different parts of LEE4 transcript is yet to be determined. Probing of potential interactions between CesAB and specific regions of LEE4 mRNA, with or without other interacting partners like Hfq, would be a clear next step in order to characterise this CesAB-mediated regulation.

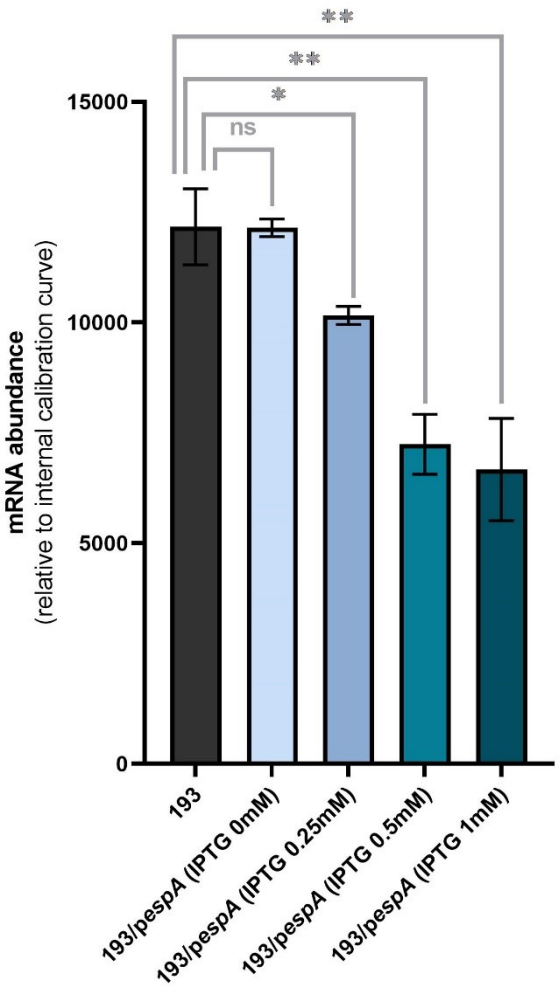
3.2.4 **EspA competes with *sepL* mRNA for CesAB interaction**

CesAB is a well-known chaperon of EspA protein that binds to EspA with high affinity (Portaliou et al., 2017). Therefore it was interesting to determine if the CesAB-EspA interaction is able to divert CesAB from its role as a positive regulator of *sepL* mRNA maintenance. In order to test this hypothesis an IPTG inducible EspA overexpression vector (pWSK*espA*) produced by Sean McAteer was expressed in ZAP193 EHEC background and *sepL* mRNA abundance measured. The effect of EspA overexpression on degradation of *sepL* mRNA transcript was additionally determined by RT qPCR time-course after transcription inhibition. In order to assure observed changes in *sepL* mRNA and SepL levels detected are due to post-transcriptional regulation LEE4 promoter activity was again assayed using LEE4*::lacZ* transcriptional fusion. SepL protein expression was measured by determining SepL-GFP (pDW6) expression by fluorescence spectroscopy and was validated by SepL immunoblotting. The work again focused on the ZAP193 strain due to its high secretor profile which allows for observations of decreases in SepL expression. The *sepL* mRNA abundance measurements were carried out over an IPTG induction gradient ranging from 0 mM to 1 mM. An IPTG induction gradient was not conducted while measuring SepL-GFP expression in EspA overexpression backgrounds as the pDW6 SepL-GFP translational fusion construct is itself IPTG inducible. Therefore decreases in IPTG concentration would result in less EspA to divert CesAB but also less SepL-GFP. This additional variable would make it impossible to observe changes in SepL-GFP expression proportional to EspA overexpression.

Results obtained by measuring expression of *sepL* mRNA abundance under EspA overexpression conditions indicated that EspA has a negative effect on *sepL* mRNA abundance (Figure 3.14a&b).

Increase in IPTG concentration and thus EspA expressed from the pWSK*espA* construct resulted in proportionally greater decrease in *sepL* transcript detected. This was concluded to be due to post-transcriptional changes in *sepL* mRNA stability (Figure 3.14c). This effect of EspA overexpression on *sepL* mRNA maintenance was confirmed to result in decrease in SepL-GFP protein production and SepL protein detected by western blot analysis (Figure 3.15).

a.


**
**
*
ns
15000
10000
5000
0
mRNA abundance
(relative to internal calibration curve)
193
193/pespA (IPTG 0mM)
193/pespA (IPTG 0.25mM)
193/pespA (IPTG 0.5mM)
193/pespA (IPTG 1mM)

b.

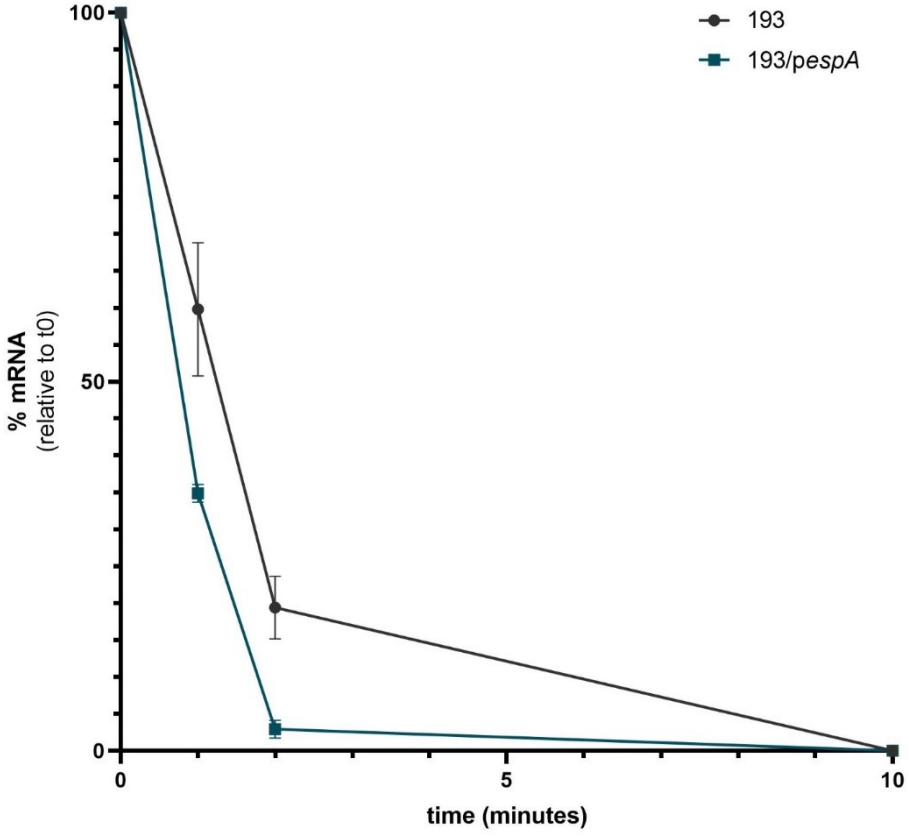

100
50
0
% mRNA
(relative to t0)
0
5
10
time (minutes)
193
193/pespA

Figure 3.14. a. *sepL* mRNA abundance in high secretor ZAP193 EHEC EspA overexpression background. *sepL* transcript was quantified with different EspA concentrations achieved by using a gradient of IPTG concentration (0 mM, 0.25 mM, 0.5 mM and 1 mM) to induce pWSK*espA* (p*espA*) expression. Increase in pWSK*espA* induction and thus EspA production resulted in proportional and significant decreases in *sepL* mRNA detected via RT qPCR. **b.** *sepL* mRNA maintenance time-course post rifampicin treatment with and without pWSK*espA* supplementation (IPTG induction 0.5 mM). Overexpression of EspA resulted in enhanced decay of *sepL* transcript. (unpaired t-test analysis: ns (not significant) – $p>0.05$; * - $p<0.05$; ** - $p<0.01$). (biological repeat n=3, technical repeat n≥6).

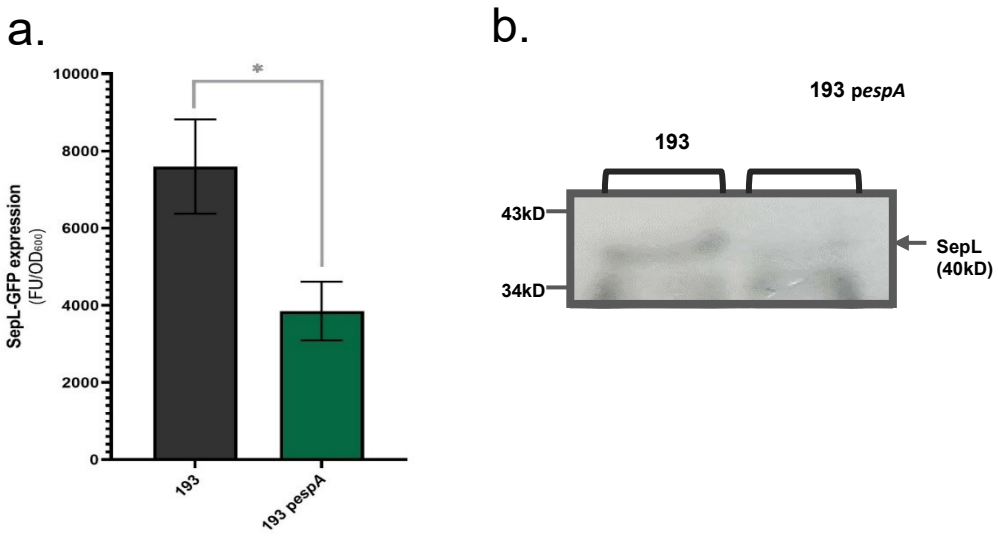


Figure 3.15. a. SepL-GFP (pDW6) translational fusion production in ZAP193 high secretor EHEC background overexpressing measured by fluorescence spectroscopy. Significant decrease in SepL-GFP signal was detected in background supplemented with pWSK*espA* (p*espA*). (unpaired t-test analysis: * - $p<0.05$). (biological repeat n=3, technical repeat n=6). **b.** Immunoblotting of the SepL protein documented reduced intensity of a band corresponding to the 40 kD SepL protein in ZAP193 overexpressing EspA (p*espA*) as compared to the wild type ZAP193 background (biological repeat n=2, technical repeat n=2).

However these experiments did not conclude if the observed decrease in *sepL* mRNA maintenance is due to hypothesised diversion of positive regulator CesAB by EspA or a direct effect of EspA on *sepL* mRNA stability. In order to investigate that EspA overexpression backgrounds were additionally supplemented with CesAB overexpression construct (pACYC*cesAB*). This showed an increase in CesAB available to act on *sepL* mRNA resulted in partial restoration of *sepL* transcript abundance irregardless of EspA overexpression (Figure 3.16). This result indicates that it is likely to be the competition of EspA with the *sepL* transcript for CesAB interaction that results in a decrease in *sepL* mRNA stability. As indicated in section 3.2.1 and 3.2.3 levels of the LEE4 transcript are substantially reduced if CesAB is not present. Diversion of this CesAB activity for LEE4 transcript maintenance by EspA can act as a negative regulatory feedback loop for LEE4 proteins expression and will play a role in orchestration of hierarchical production of T3SS translocon proteins.

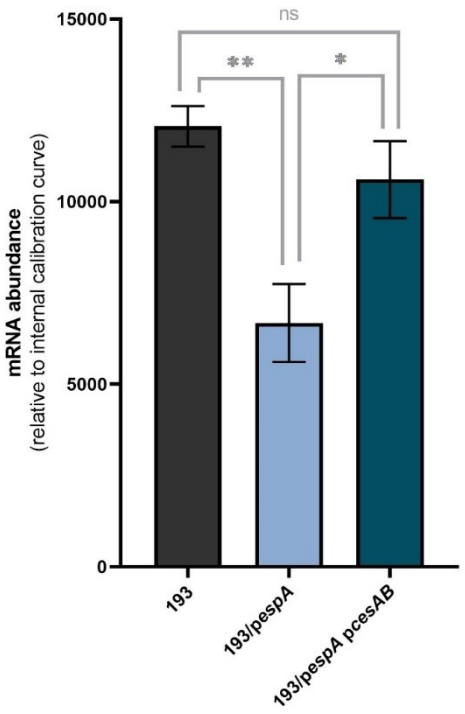


Figure 3.16. *sepL* mRNA abundance in high secretor ZAP193 EHEC overexpressing EspA and CesAB. Supplementation with pWSK*espA* (p*espA*) and pACYC*cesAB* (p*cesAB*) resulted in significant increase in *sepL* mRNA levels that were shown to be reduced by supplementation with pWSKespA alone (unpaired t-test analysis: ns (not significant) – $p>0.05$; * - $p<0.05$; ** - $p<0.01$). (biological repeat n=3, technical repeat n=9)

3.3 Discussion

In the last couple of decades the dual function of T3SS chaperons as secretion pilots and stabilisers of T3SS effectors has been documented in multiple studies. There is also evidence of their activity as regulators with such 'moonlighting' action extensively investigated for the LcrH chaperone of YopD (pore-forming translocon component, EHEC EspB homologue) in *Yersinia* species. A protein complex formed by the LcrH chaperone, its binding partner YopD, and the LcrQ protein involved in the low-calcium response was shown to post-transcriptionally regulate expression of T3SS effector proteins (Fei et al., 2021; Francis et al., 2001; Rimpiläinen et al., 1992). The complex was shown to bind to the mRNA of a major transcriptional regulator (LcrF) and facilitate recruitment of RNase E, leading to *lcrF* transcript decay (Fei et al., 2021).

Another T3SS chaperone FlgN, involved in flagellar assembly in *Salmonella typhimurium* was shown to couple FlgM effector secretion and *flgM* mRNA translation from the flagellar T3SS but not injectisome T3SSs (Aldridge et al., 2003). Furthermore multiple studies including those investigating EHEC CesD chaperone homologues, the *Shigella* IpgC and *Salmonella* SicA, documented little to no translocon substrate in the absence of the cognate chaperones, which was attributed to chaperone-mediated translational control, although no mechanism was provided (Tucker et al., 2000; Menard et al., 1994).

A study by Yi et al. in *Edwardsiella tarda*, documented a decrease in abundance of transcripts encoding for EseB and EseC (EspA and EspD EHEC homologues) translocon components in the absence of the EseE protein, which acts as the *Edwardsiella* EscC chaperone of a chaperone that governs EseB and EseC secretion. The authors concluded this is due to EseE-mediated transcriptional

regulation of the translocon-encoding operon, however the methodology used did not assess transcriptional activation and rather measured the transcript abundance. Thus the noted decrease in mRNA abundance in the absence of the EseE chaperon could potentially be an effect of post-transcriptional control of mRNA decay (Yi et al., 2016).

Another example of finely tuned chaperone-mediated coupling of translocon expression and secretion switching was documented in *Salmonella* species (Takaya et al., 2019). *Salmonella* pathogenicity island 2 (SPI2)-encoded chaperone proteins SsaH and SsaE were shown to form a heterodimer and govern secretion of SsaG (T3SS needle, EHEC EscI). SsaE was also identified as a chaperone for an early secretion substrate SseB (EHEC EspA filament homologue). The switch in needle to translocon secretion was hypothesised to be determined by different affinities of SsaE-SsaH-SsaG and SsaE-SsaB for T3SS secretion sorting platform. Interestingly SsaE was shown to regulate the cellular levels of the SsaH chaperone as well as the SsaB filament, which was concluded to play an important role in hierarchical secretion of the bound substrates. However the nature of this regulation was not fully explained (Takaya et al., 2019).

Another prominent way T3SS chaperones have been documented to control expression of LEE-encoded proteins is by antagonism of post-transcriptional regulators such as CsrA. Recent studies by Katsowich et al., 2017, Ye et al., 2018, and Elbaz et al., 2019, indicated a post-transcriptional feedback loop that involved CesT antagonism of CsrA, which positively regulate *tir* mRNA translation. The CsrA-CesT anti-Tir interaction, along with *tir* mRNA translation–dependent maintenance of LEE5 mRNA and Tir protein stabilising function of CesT ensures stable CesT chaperone to Tir substrate ratio essential for finely tuned hierarchical effector secretion (Elbaz et al., 2019).

It is apparent that T3SS chaperones have the capacity for post-transcriptional control of LEE expression, however the mechanisms behind those interactions are still predominately unknown. Additionally many chaperones are yet to be assayed for their potential function in the regulation of translation. Investigation of such mechanisms can lead to a clearer understanding of finely-tuned injectisome assembly and hierarchical secretion of T3SS effectors.

Results obtained in this chapter document a previously unreported function of CesAB as a post-transcriptional regulator of the LEE4 transcript maintenance and translation. This finding adds to the repertoire of T3SS effectors chaperones that elicit control over LEE expression. The mechanism behind CesAB-mediated regulation of LEE4 was characterised and was shown to involve stabilisation of the LEE4 transcript. The LEE4-encoded *sepL* transcript was rapidly degraded in the absence of CesAB and the maintenance of the measured mRNA was restored by CesAB complementation and enhanced by CesAB overexpression. Post-transcriptional regulation of *sepL* mRNA had a knock-on effect on EspA production and without stable *sepL* mRNA the *espA* transcript was not maintained. This control is suggested to be achieved through antagonism of Hfq-related degradation of *sepL* mRNA. Supplementation of extra CesAB partially restored *sepL* mRNA abundance and reduced the rate of transcript decay in EHEC background with increased expression of Hfq. My results align with Elbaz et al. hypothesis stating that effector chaperone can govern stable chaperone-effector ratio by repression of regulatory RNA-binding factors (Elbaz et al., 2019). Determination of direct RNA-protein and protein-protein interactions between *sepL* mRNA, CesAB and Hfq is a clear next step for further characterisation of the mechanism. This would potentially define the mRNA signatures such as secondary structure or specific binding sequence that are involved in the CesAB-Hfq-*sepL* mRNA regulatory circuit. The following chapter aims to address this by determining if there is evidence for a direct

physical interaction between CesAB and *sepL* mRNA by RNA-protein crosslinking and subsequent protein pull down assays.

Additionally it would be interesting to determine if CesAB-mediated repression of Hfq requires interaction with the *sepL* transcript only, or is it the 'sponging' or 'blocking' of Hfq by CesAB that diverts Hfq from targeting *sepL* mRNA for decay. In order to investigate this the expression of a known Hfq-regulated factor such as Shiga toxin (*stx*$_{2AB}$) could be assessed in backgrounds that overexpress CesAB but lack *sepL* mRNA (Kendall et al., 2011). If it is only the ability of CesAB to 'sponge' the Hfq through protein-protein interaction, overexpression of CesAB would lead to increase in expression of other factors that are negatively regulated by Hfq such as Shiga toxin (Kendall et al., 2011). However if that does not happen, it leaves the competitive interaction of CesAB and Hfq on the *sepL* transcript as the main way to control *sepL* mRNA decay.

CesAB antagonism of Hfq-mediated degradation of LEE4 transcript allows for translation of LEE4 mRNA-encoded factors, including CesAB substrates EspA and EspB. Portaliou et al. indicated that binding of the EspA-CesAB to SepL protein docked at the EscV basal apparatus subunit is essential for EspA secretion (Portaliou et al., 2017). My findings obtained through EspA competition assay suggest that expression of EspA diverts CesAB from its positive regulatory function for LEE4 mRNA maintenance. We propose that CesAB-binding partners EspA and SepL act as repressors of LEE4 expression, which is therefore an evolved negative feedback loop that aids the switch in production and secretion from early T3SS substrates (translocon proteins) to middle and late effectors (eg. Tir). This hypothesis is supported by findings of Portaliou et al., who stated that SepL increases the affinity of CesAB-EspA to the T3SS translocase and reduces the targeting of middle and late substrate effector complexes. This explains the phenotype of *ΔsepL* backgrounds that oversecrete late secretion

substrates prematurely. However correctly timed cessation of SepL production by the proposed negative feedback loop would allow for hierarchical translocation of the late effector proteins.

Additionally expression of LEE1-encoded CesAB can act as an indicator of basal apparatus assembly (LEE1-3-encoded, regulated by LEE1-encoded transcriptional regulator Ler), thus ensuring that other T3SS interacting partners such as EscV are present at the correct time for SepL and EspA interactions. However this is only a hypothesis and further investigation into the hierarchy of LEE operon expression is necessary to fully appreciate the importance of such regulation. Methodologies like real-time imaging of specific mRNA transcripts at a single cell level would allow for tracking of transient changes of translation dynamics throughout a time-course and could be used to determine the importance of timing of LEE-encoded factor expression for T3SS assembly (Wang et al., 2016). Additionally detailed mRNA co-expression patterns throughout a time-course could be assayed by novel ‘omic’ analysis approaches such as DynOmics designed by Straube et al., that take into consideration dynamic processes of mRNA degradation and translation (Straube et al., 2017).

Strikingly CesAB alone was shown to be sufficient to translationally activate SepL-GFP (pDW6) production in a non-T3SS *E. coli* background, DH5α. High level *sepL* mRNA was previously detected in DH5α background, however it was shown not to be translated. Although the primary regulatory function of CesAB appears to be repression of *sepL* transcript decay it is possible that interaction with CesAB leads to structural changes in *sepL* mRNA that enhance ribosomal docking. However this would require further research addressing the effects of CesAB interaction on *sepL* mRNA conformation and translational activation.

Unpublished data obtained in our laboratory shows that CesD, a LEE2-encoded chaperone of EspD translocon protein, has a negative

impact on SepL expression, while the LEE2-encoded T3SS structural protein EscC positively regulates SepL production. Interestingly deletion of the entire LEE2 operon doesn't affect SepL expression (Wawszczyk, 2017). This suggests that an interplay between CesD and EscC negates the CesD-mediated repression of SepL expression. In the light of the new findings concerning the CesAB regulation of *sepL* mRNA maintenance, it would be interesting to determine if CesD has a role in CesAB antagonism. This could be prior to T3SS basal apparatus production, when EscC is not at the correct position to interact with CesD, or after the translocon protein secretion, when both CesAB and CesD have lost their cargo. Unravelling this potential regulatory mechanism would add another layer to the complex control of T3SS assembly.

Chapter 4 Characterisation of CesAB-*sepL* mRNA interaction

4.1 Background

Study by Tree et al., 2014, showed that *sepL* mRNA directly interacts with the well-known RBP Hfq. Previous chapters of this thesis demonstrated that Hfq-mediated degradation of *sepL* mRNA is opposed by CesAB, expression of which stabilises the *sepL* transcript. The potential mechanisms behind this interplay are as follows: competitive binding of Hfq and CesAB to the same region of *sepL* transcript; CesAB-mediated change in *sepL* mRNA structure that antagonises Hfq interaction; sequestration of Hfq by CesAB; or CesAB-mediated control of factors that antagonise Hfq. Preliminary *in silico* predictions of CesAB-*sepL* mRNA and CesAB-Hfq interactions were generated during this project and aimed to indicate the most likely interactions underpinning the CesAB-mediated regulation of *sepL* mRNA. RNA-protein binding prediction software RPISeq, created by Mupiralla et al., 2011, was used to computationally assess the probability of CesAB-*sepL* mRNA interaction. The RPISeq machine learning approach uses Random Forest (RF) and Support Vector Machine (SVM) classifiers that range in values from 0 to 1 and values above 0.5 being considered as 'positive' for the predicted interaction. RF analysis indicated 0.75 chance of *sepL* mRNA interaction with CesAB and the SVM analysis predicted that the likelihood of the interaction is even higher at 0.88. Prediction of Hfq-CesAB interaction was assessed by protein to protein binding prediction tool PSOPIA (Murakami et al., 2014). This tool indicates the probability of protein to protein interactions by assessment of proteins sequence compatibility, propensity of domain-domain interactions and sum of edge weights along the shortest path between homologous proteins in a protein-protein interaction network and gives a score ranging from 0-1 for each of those qualities and the combined score taking into

consideration all of the assessed characteristics (Murakami et al., 2014). PSOPIA indicated the probability of Hfq-CesAB binding as low at 0.3537 for all of the features of the proteins tested. Therefore the next logical step in characterisation of this regulatory interplay was to attempt to validate the *in silico* prediction of a direct *sepL* mRNA-CesAB interaction.

RNA immunoprecipitation (RNA-IP or RIP) is a methodology commonly used to determine RNA molecules that interact with RNA binding proteins (RBPs). The pivotal steps of RIP involve initial step to 'fix' RNA interactions, then lysis of the cells, followed by precipitation of the RNA binding protein of interest with its associating RNAs'. These are then extracted and then identified by RT qPCR, microarray analysis (RIP-chip) or sequencing (RIP-seq). RIP purifies RNA-protein complexes either under native conditions or after formaldehyde crosslinking. Native RIP identifies only the RNAs directly bound to the protein of interest. However native RIP requires low stringency conditions that result in low specificity of the captured interactions. Formaldehyde treatment before the lysis of the cells, creates covalent bonds between RNA and the RBP that allow for high stringency conditions to be used during the purification, this in turn results in better specificity of captured interactions. Crosslinks created by formaldehyde can be reversed by heat treatment. Chemical crosslinking catalyses RNA-protein, DNA-protein and protein-protein crosslinks and thus selects for direct but also indirect RNA targets of the protein of interest. This characteristic makes formaldehyde RIP a useful methodology to identify interacting multi-molecular complexes such as degradosome (Balcerak et al., 2019).

Optimisation of RIP to account for described lack of specificity of native RIP and inability of formaldehyde crosslinking to distinguish between direct and indirect interactions, has led to development of crosslinking and immunoprecipitation (CLIP) methodology that introduces an ultra violet (UV) irradiation step to the RIP protocol. UV crosslinking creates irreversible covalent bonds between the protein and the interacting RNAs and allows for use of high stringency conditions, improving the specificity of the captured

interactions. However unlike the formaldehyde treatment, UV irradiation does not result in protein to protein and RNA to DNA crosslinking, thus only the RNA-protein interactions are captured by CLIP. This characteristic is especially useful while studying multifactorial regulatory RNA-protein mechanisms that involve multiple acting proteins. Use of UV crosslinking ensures that the detected RNAs bind directly to the probed protein and not to other molecules involved in the complex. UV crosslinking cannot be reversed. After UV crosslinking and cell lysis, CLIP employs a partial fragmentation of captured RNAs, which prevents co-purification of other proteins that directly bind the RNAs. An additionally this step provides insight into the specific RNA-protein binding sites, since the obtained RNA fragments are protected from RNase digestion by the interacting proteins. Post-precipitation of RBP-RNA complexes, CLIP makes use of ligation of RNA adapters to allow for high-throughput RNA sequencing. Additional quality control step employed by CLIP involves separation of the protein-RNA complexes on SDS-PAGE and extraction of the RNAs only from the excised band that correspond to the size of the RBP of interest. Subsequent digestion of bound protein by proteinase K treatment releases the RNAs that can then be reverse transcribed using primers complementary to the sequence of the ligated adapter. cDNAs are truncated at the site of the crosslink. Variants of CLIP protocol such as iCLIP were established to utilise this characteristic to map the specific RNA binding sites. Recent review by Lee et al., 2018 provides a detailed account of variations of CLIP protocol including iCLIP.

Immunoprecipitation is usually employed in order to purify the RBP-RNA complexes. However if a specific antibody is not available, the protein of interest can be tagged with and purified using affinity or epitope tags such as 6xHis-tag or FLAG (Einhauer et al., 2001; Bizzard et al., 2000). In fact a methodology derived from CLIP, called crosslinking and analysis of cDNAs (CRAC) employs affinity tag(s) to allow for fully denaturing conditions during protein purification, which increases the stringency and ensures full dissociation of multifactorial RNA-binding complexes (Granneman et al., 2009).

Research described in this chapter utilised a combination of described crosslinking and precipitation methodologies to determine the nature of proposed CesAB-*sepL* mRNA interaction. The potential CesAB-*sepL* mRNA binding site was also probed by deletion studies. The importance of the secondary structure of the *sepL* transcript in facilitation of the proposed interaction was also investigated.

4.2 Results

4.2.1 **CesAB-6xHis construction**

In the absence of a specific, usually monoclonal antibody, affinity tags are a useful tool in purification of specific proteins. The DNA sequence encoding a string of six (or more) histidine molecules is commonly used to produce recombinant proteins with the poly-His-tag fused to the N- or C-terminus. His-tagged proteins can be purified utilising the capacity of histidine to chelate several metals such as nickel, cobalt, zinc and copper and their expression can be detected by anti-His immunoblotting (Brasili et al., 2016). A 6xHis-tag was used to produce N and C-terminus CesAB-6xHis-tagged recombinant proteins cloned into a pWSK29 expression vector that was used for *in vivo* CesAB-RNA interaction studies described in detail in the following subsection. Tagging both N- and C-terminus increases the chances of producing a recombinant protein with an exposed tag, as on some occasions the tag can be buried and thus unable to be detected by the antibody or metals used for recombinant protein purification. The pWSK29 expression vector was chosen again as it is IPTG inducible and thus allows for the levels of protein produced to be adjusted, which can be useful in protein purification as high protein concentrations can lead to protein aggregation, insolubility and instability (Singh et al., 2015).

Originally site directed mutagenesis (SDM) methodology was planned to be used to introduce 6xHis-tag at the N- and C-CesAB terminii of pWSK*cesAB* clone. pWSK*cesAB* was previously shown to express functional CesAB that complements *ΔcesAB* backgrounds, therefore it was utilised as a scaffold for recombinant protein production. SDM allows for introduction of specific mutations in double stranded DNA plasmids by amplification of the entire plasmid DNA with primers designed to include desired modifications such as nucleotide substitutions, deletions and small or large sequence

insertions. Two sets of back-to-back oligonucleotides targeting the N- and C-terminii of *cesAB* clone (pWSK*cesAB*) with 3x-His sequence flanking both the forward and reverse primers were designed. Designed primers successfully amplified the pWSK*cesAB* plasmid sequence that was subsequently circularised and the template DNA was removed using kinase, ligase and DpnI (KLD) enzyme mix. *E. coli* DH5α (cloning strain) was then transformed with the produced constructs and the transformants were selected using ampicillin. However once checked by immunoblotting using anti-His antibody the obtained cultures were shown not to express a detectable His-tag. The SDM procedure was repeated several times but the required 6xHis-CesAB was not obtained using this approach.

Ultimately another strategy, T4-based cloning of synthetic DNA corresponding to sequences of *cesAB* flanked with N- or C- terminal 6xHis into the pWSK29 expression vector, was employed in order to obtain recombinant CesAB-6xHis protein (Figure 4.1a). The anti-His immunoblotting of crude cell extract from both DH5α pWSK*cesAB*-6xHis-N and pWSK*cesAB*-6xHis-C transformant cultures detected a protein corresponding to His-tagged CesAB (13.1kD) (Figure 4.1b). Interestingly the immunoblotting detected an additional band at approximately 26 kD mark. CesAB is known to dimerise and it is possible that the CesAB dimer that was not denatured was detected on the blot (Chen et al., 2013).

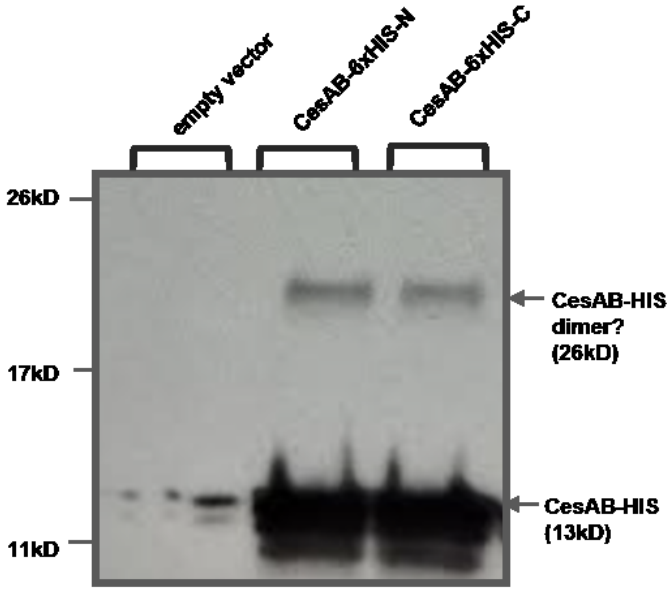
Although the expression of both N- and C-terminal 6xHis-tagged CesAB was confirmed, the functionality of the recombinant protein needed to be assessed. In order to do that the *ΔcesAB* ZAP193 background was supplemented with pWSK*cesAB*-6xHis-N and pWSK*cesAB*-6xHis-C constructs. Deletion of *cesAB* results in rapid decay of *sepL* mRNA and decrease in SepL production. Thus restoration of SepL protein expression by the recombinant CesA-6xHis proteins would indicate functionality of the constructs. Expression of SepL-GFP (pDW6) in *ΔcesAB* ZAP193 with and without pWSK*cesAB*-6xHis-N or pWSK*cesAB*-6xHis-C supplementation was assessed by fluorescence spectroscopy (Figure 4.1c). Obtained results

indicated that both N- and C-terminal 6xHis-tagged CesAB recombinant proteins complemented the *ΔcesAB* mutation and restored SepL production. This confirmed that the recombinant proteins maintained the characteristics that enable them to carry out the specific *sepL* regulatory function studied in this thesis. This observation was crucial as the following purpose of the created constructs was to assess if the CesAB-mediated regulation of SepL expression requires the direct interaction of CesAB with *sepL* mRNA.

a.


N-terminal 6xHis-tag
Start codon
5’ GTGCATCATCACCACCACCATNNN…TGA 3’
C-terminal 6xHis-tag
5’ GTGNNN…CATCATCACCACCACCATTGA 3’

b.


empty vector
CesAB-6xHIS-N
CesAB-6xHIS-C
26kD
17kD
11kD
CesAB-HIS dimer? (26kD)
CesAB-HIS (13kD)

c.

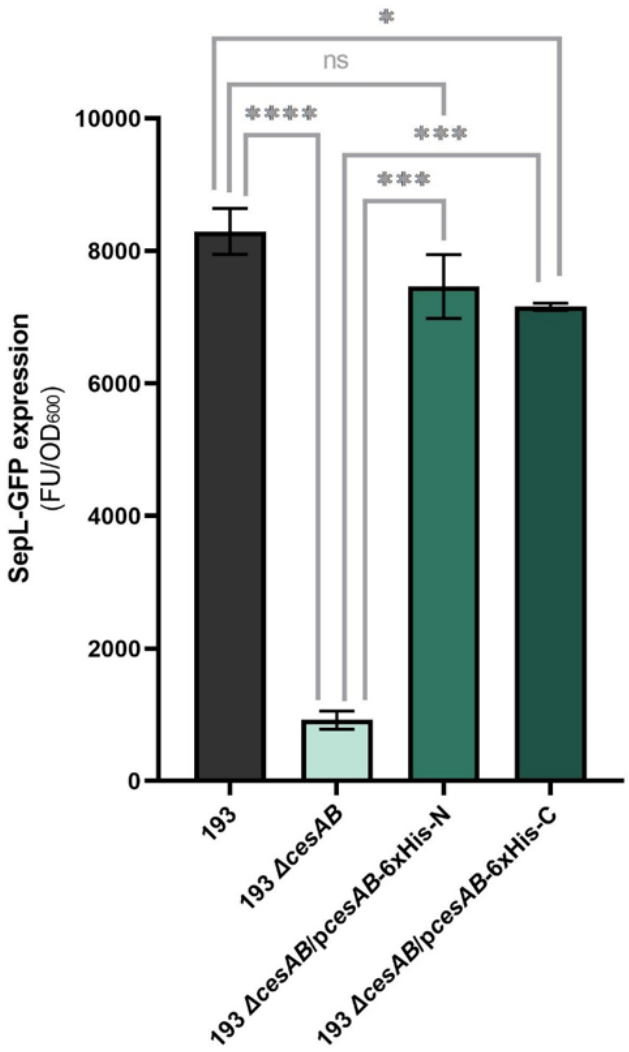

SepL-GFP expression (FU/OD600)
10000
8000
6000
4000
2000
0
*
ns
****
***
***
193
193 ΔcesAB
193 ΔcesAB/pcesAB-6xHis-N
193 ΔcesAB/pcesAB-6xHis-C

Figure 4.1. Construction of CesAB-6xHis recombinant protein. **a.** Design of synthesised 6xHis-tagged CesAB sequences cloned into pWSK29 expression vector. The N-terminus CesAB-6xHis contains 6xHis-tag sequence (pink) positioned directly upstream of the *cesAB* start codon (turquoise) followed by *cesAB* sequence including stop codon (grey). The C-terminus CesAB-6xHis contains 6xHis-tag sequence (pink) directly after *cesAB* sequence (grey) minus the stop codon. The stop codon (red) was added directly after the 6xHis sequence. **b.** anti-His immunoblotting of crude cell extracts of ZAP193 expressing the pWSK*cesAB*-6xHis-N (CesAB-6xHis-N) and pWSK*cesAB*-6xHis-C (CesAB-6xHis-C) constructs. Strong bands between the 11kD and 17kD protein ladder marker corresponds to CesAB-6xHis molecular weight (13kD), suggesting high yield of recombinant protein production. The bands observed between 17kD and 26kD markers are potential CesAB-6xHis dimers (26kD). **c.** Expression of SepL-GFP translational fusion (pDW6) in ZAP193 *cesAB* deletion background (193 *ΔcesAB*) supplemented with pWSK*cesAB*-6xHis-N (ZAP193 *ΔcesAB*/p*cesAB*-6xHis-N) and pWSK*cesAB*-6xHis-C (ZAP193 *ΔcesAB*/p*cesAB*-6xHis-C). Highly significant increase in SepL-GFP expression comparable to that of wild type ZAP193 (193) was observed.

4.2.2 **Investigation of *sepL* mRNA-CesAB-6xHis binding**

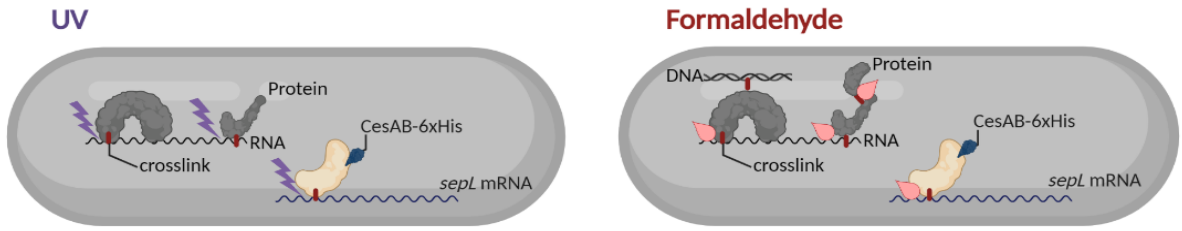
Formaldehyde crosslinking and UV crosslinking protocols that were heavily inspired by CLIP and CRAC methodologies were applied in this project (Figure 4.2). Use of both chemical and UV crosslinking allowed me to determine if CesAB-*sepL* mRNA interaction is direct. *sepL* mRNA read out from formaldehyde crosslinking and no *sepL* transcript detected from the UV crosslinking would suggest that the hypothesised interaction is indirect and involves creation of a complex with other molecule acting as a scaffold. No *sepL* mRNA detected from either of the generated samples would suggest that the CesAB's control of *sepL* transcript is an effect of CesAB-mediated regulation of a factor that regulates *sepL* mRNA maintenance.

Unlike global RNA-protein interaction studies, such as that by Holmqvist et al., which determined the CsrA and Hfq RNA binding sites in *Salmonella typhimurium* (Holmqvist et al., 2016), the question posed by my project concerned a specific RNA (*sepL* mRNA). Therefore RT qPCR using primers targeting *sepL* specifically, instead of transcriptome-wide microarray or adapter-based sequencing, was performed to investigate the binding of *sepL* mRNA to CesAB. This allowed me to omit several steps in CLIP methodology that are a requirement for subsequent sequencing. Ligation of RNA adapters was not performed and thus cDNA purification from the free adapters and RT primers was not necessary. This granted less processing of the captured nucleic acid sequences. Use of the primers targeting *sepL* sequence in cDNA amplification was thought to additionally increase the specificity of the assay. RNase digestion conditions recommended in CLIP protocols result in RNA sequences ranging from 30-200 bases. This is necessary for sequencing but was thought to significantly reduce the chances of annealing of the designed RT qPCR primers to the captured RNA and thus was omitted.

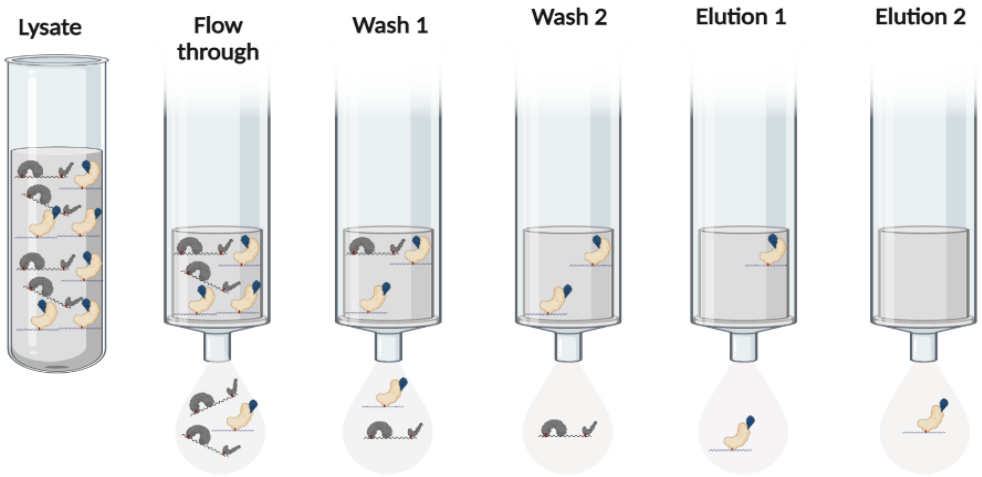
ZAP193 high secretor EHEC background transformed with the produced N and C terminal 6xHis-tagged CesAB constructs (pWSK*cesAB*-6xHis-N and

pWSK*cesAB*-6xHis-C) was used in both formaldehyde crosslinking and UV-crosslinking experiments. ZAP193 background supplemented with untagged pWSK*cesAB* was used as a negative control. The high secretor EHEC strain was used in order to maximise the amount of available *sepL* mRNA, abundance of which varies between strains as demonstrated in previous chapters of this thesis. Overexpression of CesAB from the pWSK*cesAB*-6xHis-N and -C constructs additionally increased the chances of capturing the CesAB-*sepL* mRNA interaction. However it needs to be noted that protein overexpression can change the governing of other cellular processes and thus might not represent the native conditions (Bolognesi et al., 2018).

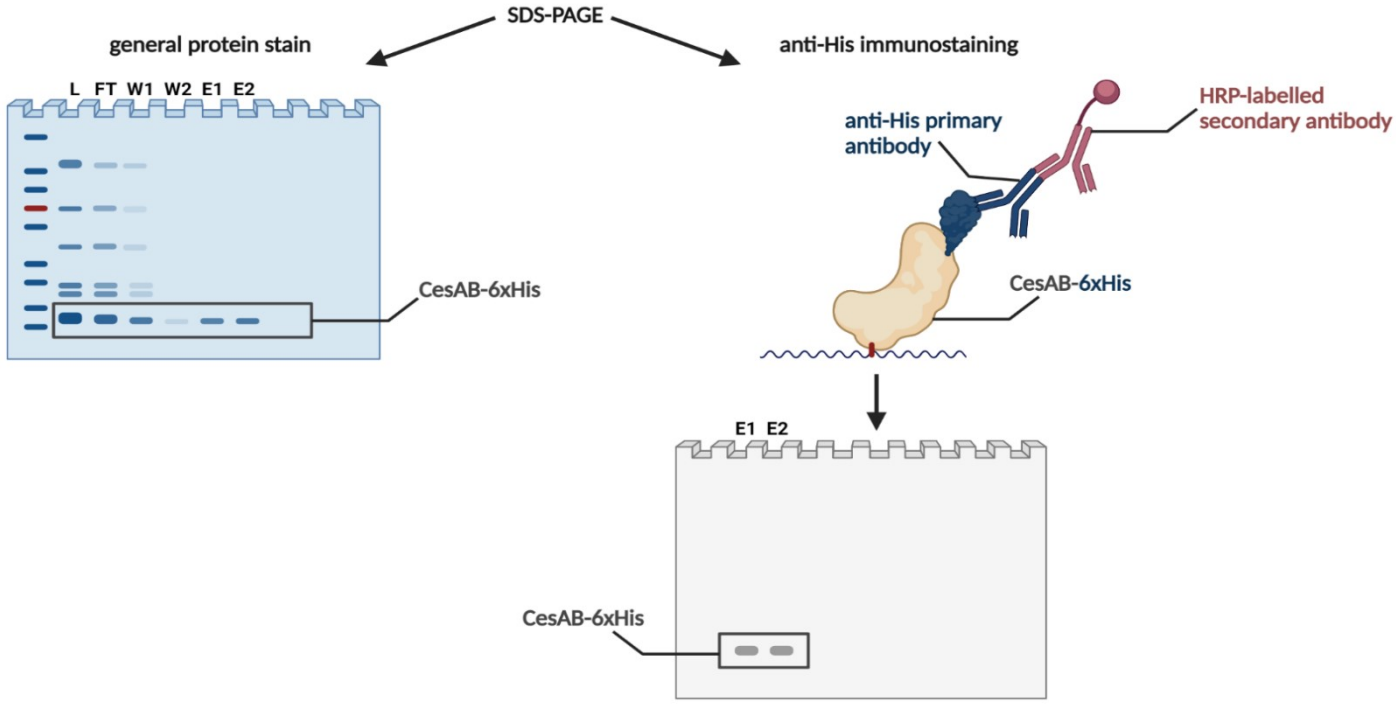
a. **Protein-RNA crosslinking**



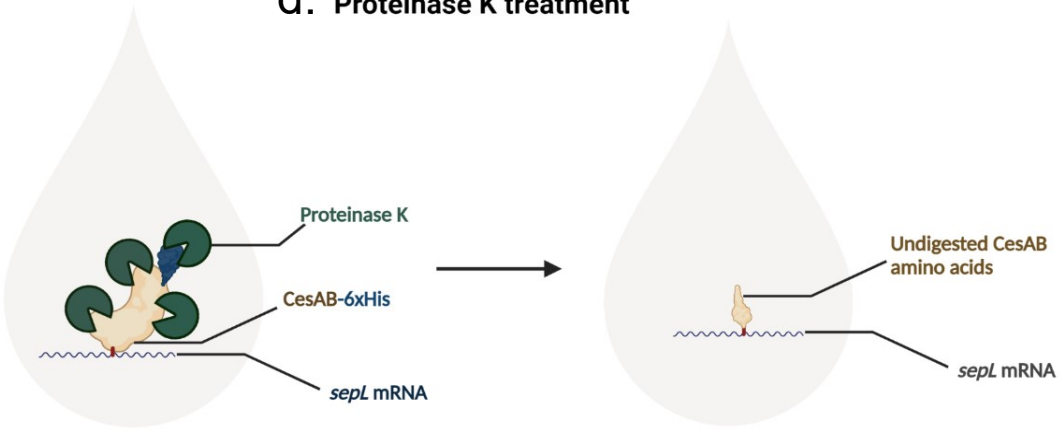
b. **Cell lysis and Ni-NTA His-tag affinity purification**



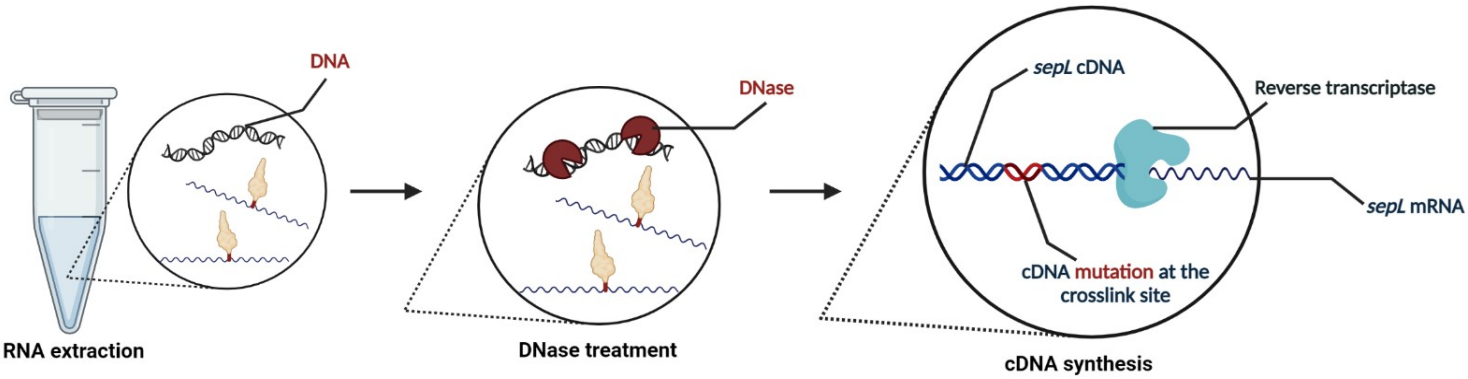
c. **Protein fraction analysis**



d. **Proteinase K treatment**



e. **mRNA purification and cDNA synthesis**



f. **qPCR**

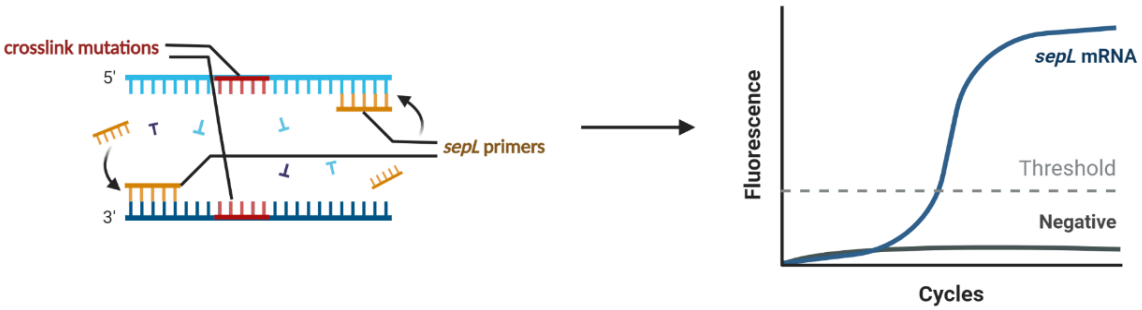


Figure 4.2. Schematic of workflow for detection of hypothesised CesAB-*sepL* mRNA interaction. **a.** the EHEC background expressing CesAB-6xHis recombinant protein or supplemented with untagged CesAB (negative control) were exposed to UV or formaldehyde treatment *in vivo*. The UV exposure created crosslinks between the proteins expressed by the EHEC and bound RNAs (hypothetically including *sepL* mRNA). The formaldehyde treatment crosslinked the proteins to bound RNAs but also other proteins and DNA. **b.** Obtained crosslinked cells were lysed and the His-tag affinity purification was performed using Ni-NTA Fast Start Kit manufactured by QIAGEN. The protein fractions were collected from the untreated lysate, Ni-NTA column flow through, after two rounds of washes that intended to remove proteins other than the CesAB-6xHis and two rounds of elutions of His-tagged CesAB bound to the Ni-NTA resin. The CesAB-untagged negative control samples were taken through the same process **c.** the obtained protein fractions were separated by SDS-PAGE and coomassie blue general protein stain was performed in order to determine protein fractions that contained only the CesAB-6xHis. The absence of a band at the elution fractions of the CesAB-untagged negative control samples indicated that CesAB-6xHis purification was successful. Western blot analysis using anti-His antibody and secondary HRP-labelled antibody followed by chemiluminescence analysis confirmed the presence of the CesAB-6xHis in the selected protein fractions. **d.** Proteinase K treatment was performed for subsequent detection of the CesAB-6xHis-bound *sepL* mRNA. UV crosslinking is irreversible and small CesAB fragments bound to the transcript remain undigested by the proteinase K, leaving the mark of the CesAB-mRNA interaction site. Formaldehyde treated samples were reverse crosslinked prior to Proteinase K treatment, that resulted in *sepL* mRNA without the undigested CesAB amino acids. **e.** Phenol-chloroform RNA extraction was performed. This was followed by DNase treatment in order to ensure only the CesAB-bound RNAs are being detected by subsequent analysis. This was specifically important for the formaldehyde treated samples since formaldehyde crosslinks proteins to DNA as well as RNA. Obtained RNA samples were reverse transcribed. The reverse transcription of UV-crosslinked samples produced cDNA with mutations at the site of the crosslinks, since the reverse transcriptase is unable to read through those sequences. **f.** qPCR

analysis using *sepL*-specific primers was performed to determine the presence of *sepL* mRNA in the crosslinked and CesAB-purified samples, which would suggest CesAB-*sepL* mRNA interaction. Negative control for this experiment was the detection of the mRNA of EHEC housekeeping gene *rpoA.* Absence of *rpoA* signal ensured that only CesAB-bound RNAs were present in the analysed samples.

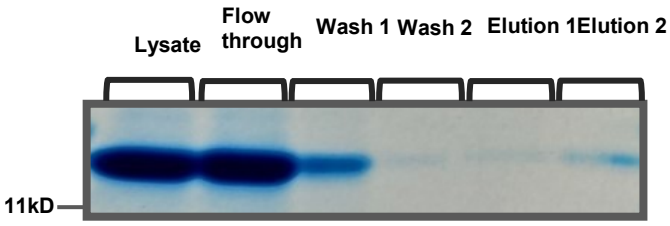
UV crosslinking procedure was carried out according to Tree et al., 2018, although a smaller volume of bacterial culture was used. Formaldehyde crosslinking was performed as described in the Materials and Methods chapter. Bacteria were cultured in MEM-HEPES supplemented with 0.1% glucose and 250 nM $Fe(NO_3)_3$ and the experiments were performed in exponential growth phase. Those conditions were previously shown to be optimal for T3SS expression (Tree et al., 2011). The cultures were supplemented with IPTG in order to induce CesAB-6xHis expression. Both experiments were performed in growing bacteria. This was crucial considering that secondary structures of RNAs are condition dependent and so can be very different *in vitro* vs. *in vivo* coupled to the importance of *sepL* mRNA secondary structure for the regulation of *sepL* transcript stability, as documented in previous chapters (Leamy et al., 2016). Additionally *in vivo* analysis ensured that other molecules that are potentially necessary for the *sepL* mRNA-CesAB interaction were present.

Ni-NTA Fast Start Kit manufactured by QIAGEN was used to pull down the CesAB-6xHis crosslinked to any putative interacting RNAs. Six fractions were collected from the column purification step and were analysed by SDS-PAGE and western blotting using anti-His antibody (Figure 4.3). This quality control step allowed for visualisation of purified CesAB-6xHis and ensured obtained samples do not include other proteins. A band corresponding to the size of CesAB-His was detected by commassie blue protein staining from CesAB-6xHis samples. The band gradually faded after washing of the column and reappeared slightly stronger at the elution fractions (E1 and E2) (Figure 4.3a). This suggested that although CesAB-6xHis was produced in high quantities, the binding of the His-tag to the column was not efficient. The elution fraction bands obtained from CesAB-6xHis-N (Figure 4.3a&b) appeared stronger than those produced by CesAB-6xHis-C (Figure 4.3c&d), the formaldehyde crosslinked CesAB-6xHis-C samples specifically produced very weak bands at the elution fractions. This could be due to potential burying of the His-tag in the structure adopted by C-terminal tagged construct. Thus only the N-terminal 6xHis-tagged elution franctions were chosen for further

analysis. As expected negative control expressing the untagged pWSK*cesAB* produced a strong band corresponding to CesAB protein molecular weight in the fractions before the elution step (F, C, W1, W2) (Figure 4.3e&f). However the band was not detectable in final elution fractions from those samples (E1 and E2). This confirmed that although low in yield the purification of His-tagged CesAB was successful. The elution fractions (E1 and E2) that were detectable in His-tagged samples but not untagged negative control were selected for further analysis. Presence of CesAB-6xHis-N in the elution samples was confirmed by the anti-His antibody immunostaining (Figure 4.4a&b). The low yield of purified CesAB-6xHis can be explained by an increase in insoluble protein aggregates that can form under protein overexpression conditions and hinder binding of the tagged protein to the Ni-NTA resin column (Zhang et al., 2004). This was accounted for reducing the amount of IPTG used for induction, when the crosslinking experiment was repeated. This marginally increased the yield of the protein detected by the anti-His (Figure 4.4b&c).

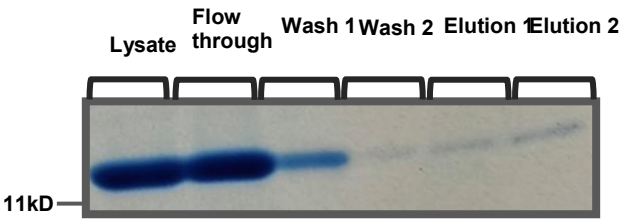
a.

CesAB-6xHis-N UV



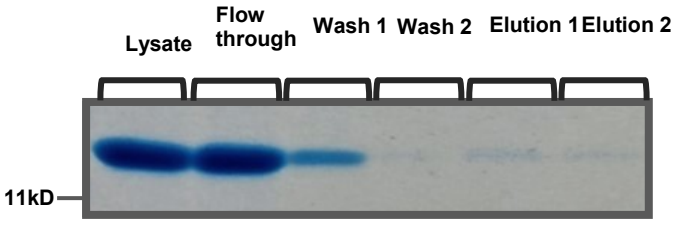
b.

CesAB-6xHis-N formaldehyde



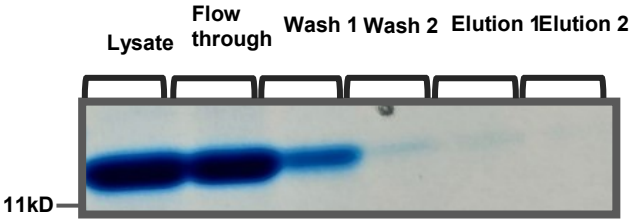
c.

CesAB-6xHis-C UV



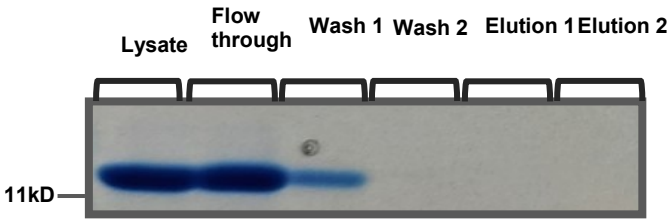
d.

CesAB-6xHis-C formaldehyde



e.

CesAB-untagged UV



f.

CesAB-untagged formaldehyde

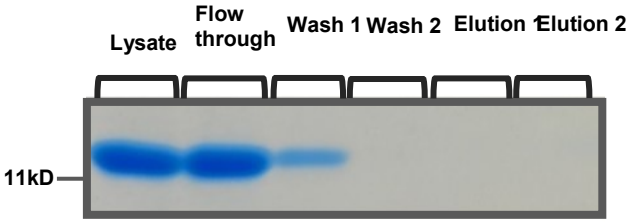


Figure 4.3. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) stained with coomassie blue of protein fractions collected from Ni-NTA Fast Start Kit (QIAGEN) His-tag purification of **a.,c.&e.** the UV crosslinked and **b.,d.&f.** the formaldehyde crosslinked ZAP193 backgrounds supplemented with the **a.&b.** pWSK*cesAB*-6xHis-N (CesAB-6xHis-N), **c.&d.** pWSK*cesAB*-6xHis-C (CesAB-6xHis-C) and **e.&f.** a negative control pWSK*cesAB* (CesAB-untagged) expression constructs. Lysate fractions (Lysate) were collected straight after cell lysis step. The samples were loaded onto Ni-NTA resin columns and the flow through fractions (Flow through) were collected. The columns were washed twice with manufacturer provided wash buffer and wash fractions (Wash 1 and Wash 2) were collected after each wash. Two elutions of the resin bound His-tagged recombinant proteins were performed with the manufacturer's elution buffer and two eluates were collected (Elution 1 and Elution 2).

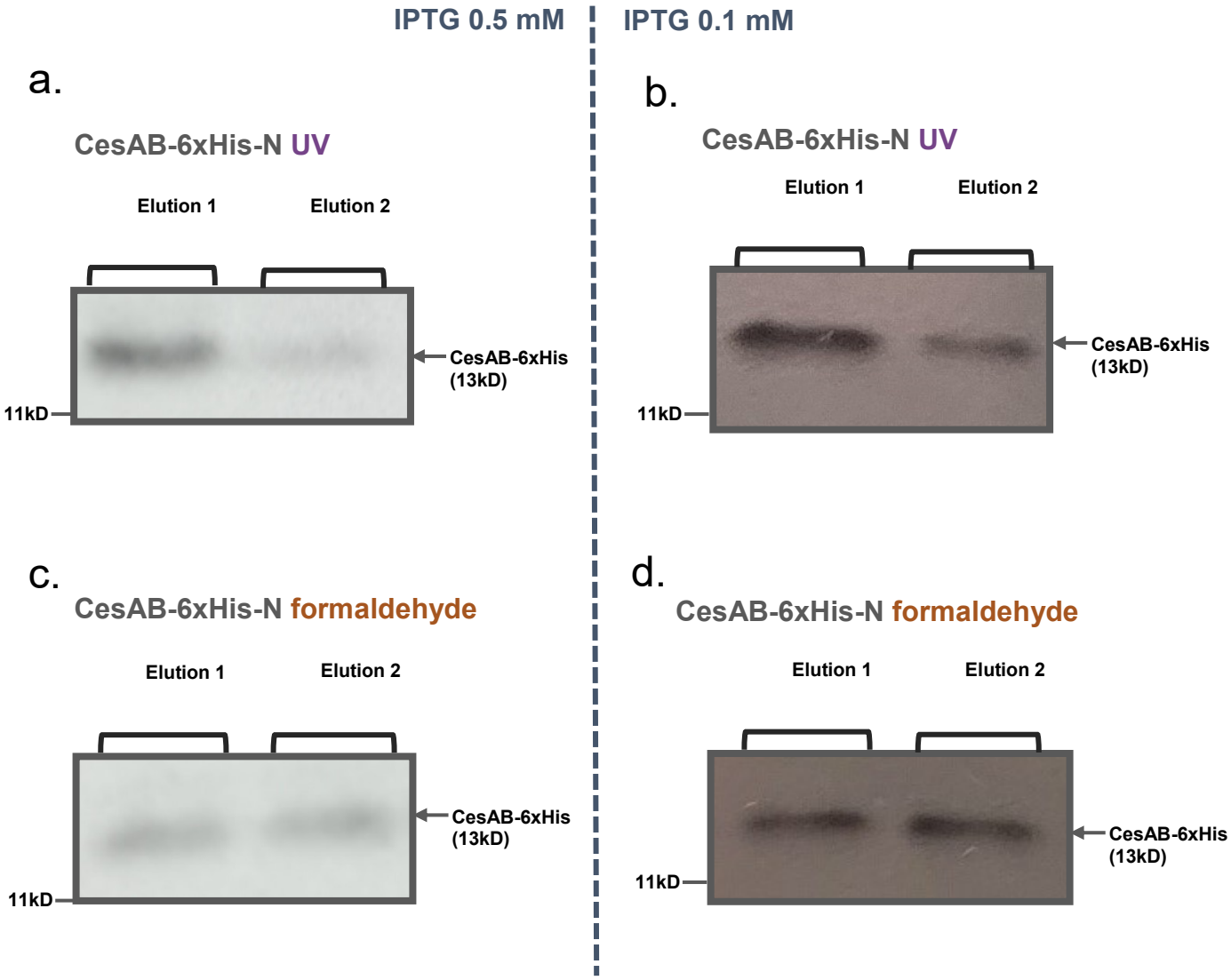


Figure 4.4. Western blot analysis of the **a&b.** UV and **c&d.** formaldehyde crosslinked CesAB-6xHis protein purification elution fractions. The extent of expression of CesAB-6xHis constructs is modulated by IPTG induction. CesAB-6xHis samples purified from cultures overexpressing CesAB-6xHis to **a&c.** higher (IPTG 0.5mM) or **b&d.** lower (IPTG 0.1mM) level were analysed in order to determine the more optimal conditions for higher yield CesAB-6xHis purification. Primary anti-His antibody and secondary HRP-labelled antibody immunostaining followed by chemoiluminescence detection was used to visualise the CesAB-6xHis-N. Bands corresponding to the molecular size of

CesAB-6xHis (13 kD) were detected in all of the samples. The bands detected from the samples with decreased IPTG induction (0.1 mM IPTG concentration) and therefore decreased CesAB-6xHis expression were stronger compared to those using higher levels of IPTG induction (0.5 mM IPTG). Potentially the overexpression of CesAB-6xHis results in poorer binding of CesAB-6xHis to the Ni NTA resin and lower yield of CesAB-6xHis purification.

Samples collected from both repetitions of the experiment were further processed with proteinase K treatment and reverse crosslinking was performed on the formaldehyde treated samples. Phenol-chloroform RNA extraction, followed by DNase treatment and cDNA synthesis were performed and the resultant samples were analysed by RT qPCR using primers targeting *sepL* sequences. The analysis detected *sepL* sequences in both UV and formaldehyde crosslinked samples from the CesAB-6xHis eluates but not from the negative control samples (Figure 4.5). The UV-crosslinking yielded more detectable *sepL* mRNA than formaldehyde crosslinking in both repeats of the experiments. This could be due to the loss of RNA through additional step of the crosslink reversal process for the formaldehyde treated samples or due to a better crosslinking efficiency of the UV. The samples were checked for the presence of transcripts encoding a ubiquitous housekeeping protein RpoA (Yin et al., 2011). No *rpoA* sequences were detected, thus increasing the confidence that the only RNAs detected in the tested sample were those bound to the CesAB-6xHis and not a contamination (Figure 4.5). In conclusion the obtained results indicated that CesAB binds to *sepL* mRNA transcript in a direct fashion.

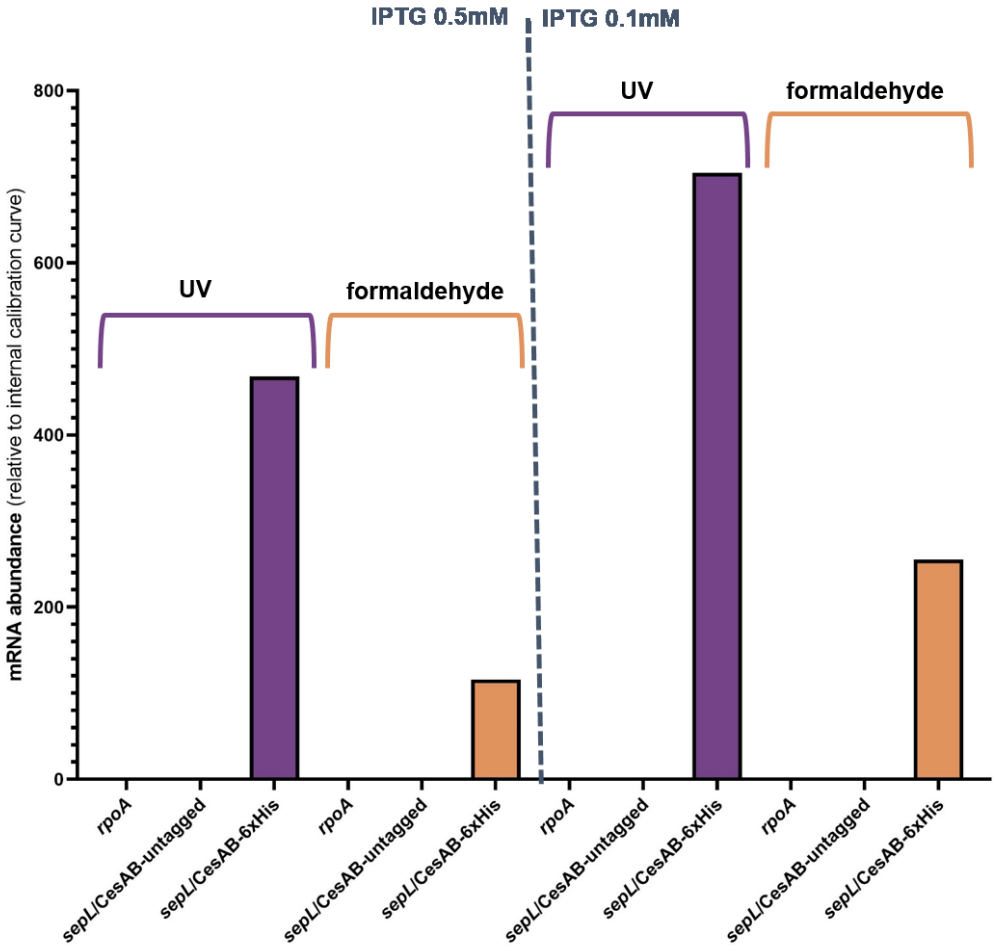


Figure 4.5. RT qPCR analysis of the proteinase K treated UV and formaldehyde crosslinked CesAB-6xHis and CesAB-untagged (negative control) protein purification elution samples. *sepL*-specific primers were used to determine the presence of *sepL* mRNA in the in CesAB-6xHis and CesAB-untagged cDNA samples. Determination of the *rpoA E. coli* housekeeping mRNA (using *rpoA*-specific primers) in the CesAB-6xHis elution fractions was used as a quality control check and an additional negative control. Absence of *rpoA* mRNA provided confidence that any RNAs pulled-down with the CesAB-6xHis were present in the samples. The RT qpCR analysis was repeated twice with higher (0.5mM) and lower (0.1mM) IPTG concentrations to induce CesAB-6xHis and CesAB-untagged expression. The mRNA levels were compared between the samples. The analysis indicated that *sepL* mRNA was present in all of the CesAB-6xHis samples but not the CesAB-untagged samples. *rpoA* mRNA was not

detected in the CesAB-6xHis samples. Decrease in IPTG induction resulted in significant increase in the detected *sepL* mRNA.

4.2.3 *sepL* transcript signatures involved in CesAB-mediated regulation

Previous studies conducted in our laboratory intended to determine the *sepL* mRNA sequence and structure characteristics involved in translational activation of *sepL* transcript. To test this Sean McAteer created specific short deletions in the *sepL* mRNA sequence (4-18nt, 19-27nt) comprising a stem loop structure (SL3) positioned close to the RBS and the start codon (Figure 4.6a). The deletions led to decrease the expression of SepL-GFP translational fusions compared to full-length SepL-GFP translational fusion (pDW6). I analysed expression from these constructs to determine if they might involve regions related to CesAB-*sepL* mRNA interactions (Figure 4.6b). The regions being scrutinised in the *sepL* transcript had not attributed to binding of any of the previously reported post-transcriptional regulators of SepL expression such as Hfq, Spot42 and CsrA (Tree et al., 2014; Wang et al., 2018).

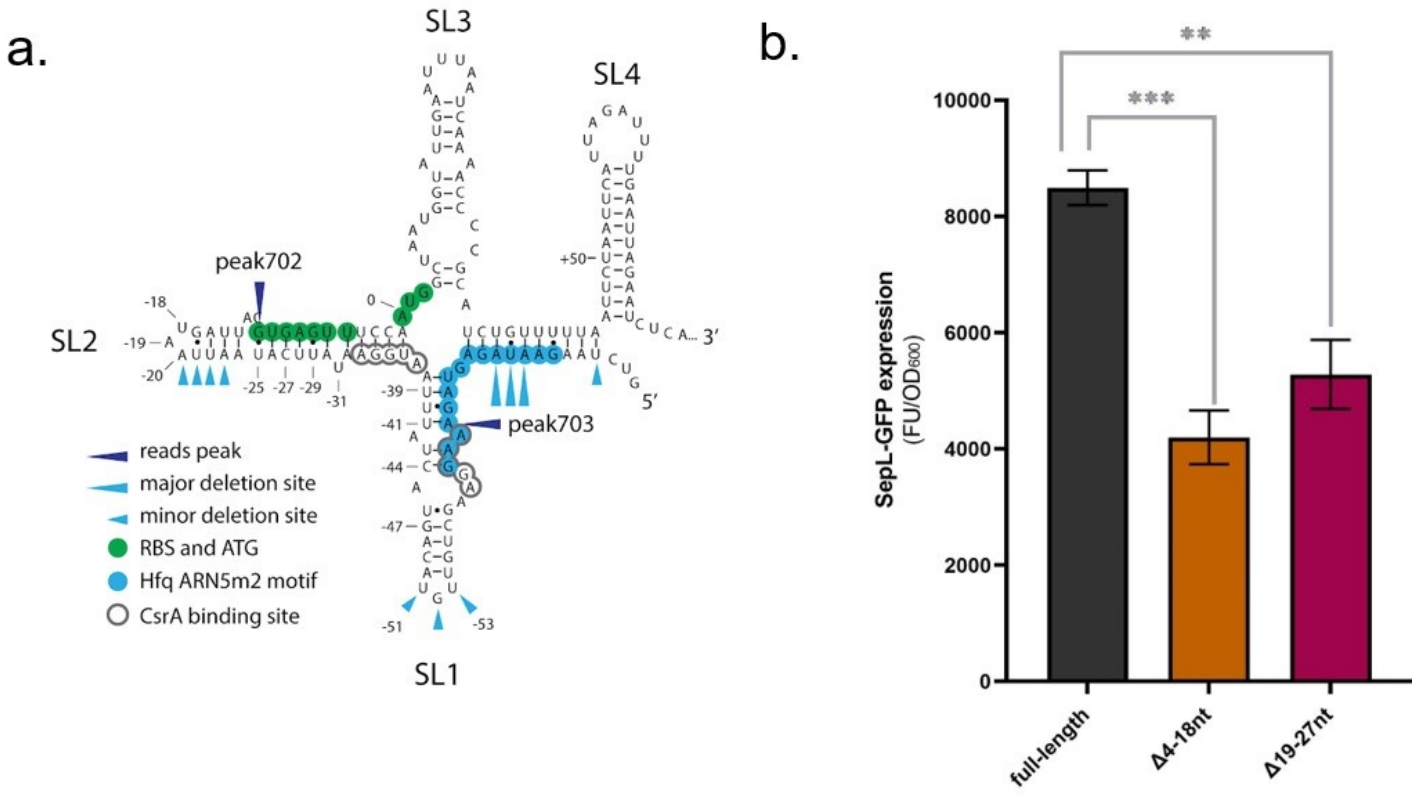


Figure 4.6. a. *sepL* mRNA secondary structure forming a stem loop (SL3) around the RBS and the start codon. **b.** SepL-GFP expression from the SDM constructs with deletions in *sepL* sequence regions 4-18nt and 19-27nt from the start codon, predicted to form a stem loop structure (SL3). The full-length SepL-GFP construct (pDW6) and the SDM constructs were expressed in the high secretor ZAP193 EHEC background. The deletions of the 4-18nt and 19-27nt resulted in significant decrease in SepL-GFP signal compared to the full-length construct.

As shown in the Chapter 2 of this thesis, SepL-GFP expression decreases significantly in ZAP193 *ΔcesAB*. Therefore it was interesting to investigate if any of the deleted regions are necessary for CesAB-*sepL* mRNA interaction. The Δ4-18nt Δ19-27nt SepL-GFP translational fusion constructs were measured in *cesAB* deletion background *ΔcesAB* ZAP193 and ZAP193 supplemented with the CesAB expression construct pWSK*cesAB*. The hypothesis behind this experiment was that if the deleted *sepL* mRNA region plays a role in CesAB-*sepL* mRNA interaction, changes in CesAB abundance would not affect SepL-GFP expression of the construct harbouring the specific deletion. In other words the construct would be insensitive to CesAB because it lacked the mRNA signature, such as specific structure necessary for CesAB-*sepL* mRNA interaction.

Obtained results indicated that the deletions did not result in decrease in SepL-GFP expression as drastic as that caused by deletion of *cesAB* (Figure 4.7). However the Δ19-27nt SepL-GFP construct was shown to be insensitive to CesAB in terms of SepL expression (Figure 4.7&Figure 4.8). Deletion of *cesAB* did not result in further decrease in SepL-GFP expression from the Δ19-27nt SepL-GFP construct (Figure 4.7). Overexpression of CesAB did not increase the SepL-GFP levels detected from the Δ19-27nt SepL-GFP translational fusion (Figure 4.8). However it is important to note that in is not possible to conclude that the 19-27bp sequence is directly involved in CesAB binding as the observed insensitivity to CesAB could be a result of disruption in the interaction of *sepL* transcript with the Hfq. It was shown in Chapter 3 of this thesis, that the negative regulation of SepL expression by Hfq was countered by CesAB. Impairment in Hfq binding might deem CesAB 'activation' redundant because its function as an antagonist of Hfq-directed regulation would no longer be applicable.

It is interesting to note that SepL-GFP expression from the Δ4-18nt SepL-GFP was shown to be affected by changes in CesAB abundance (Figure 4.7

and Figure 4.8). However this was not to the same extent as the wild type *sepL* mRNA sequence (pDW6). It is possible that the deletion has had a partial impact on CesAB activity that can be complemented by supplementation with extra CesAB (Figure 4.8). This indicated that 4-18nt *sepL* transcript region constructs is less likely to play a specific role in CesAB-mediated regulation of SepL expression. It is possible that the introduced deletion destabilises the SepL-GFP protein resulting in lower fluorescence read-out. A pulse-chase radiolabelled amino acid assay could be employed to investigate this possibility.

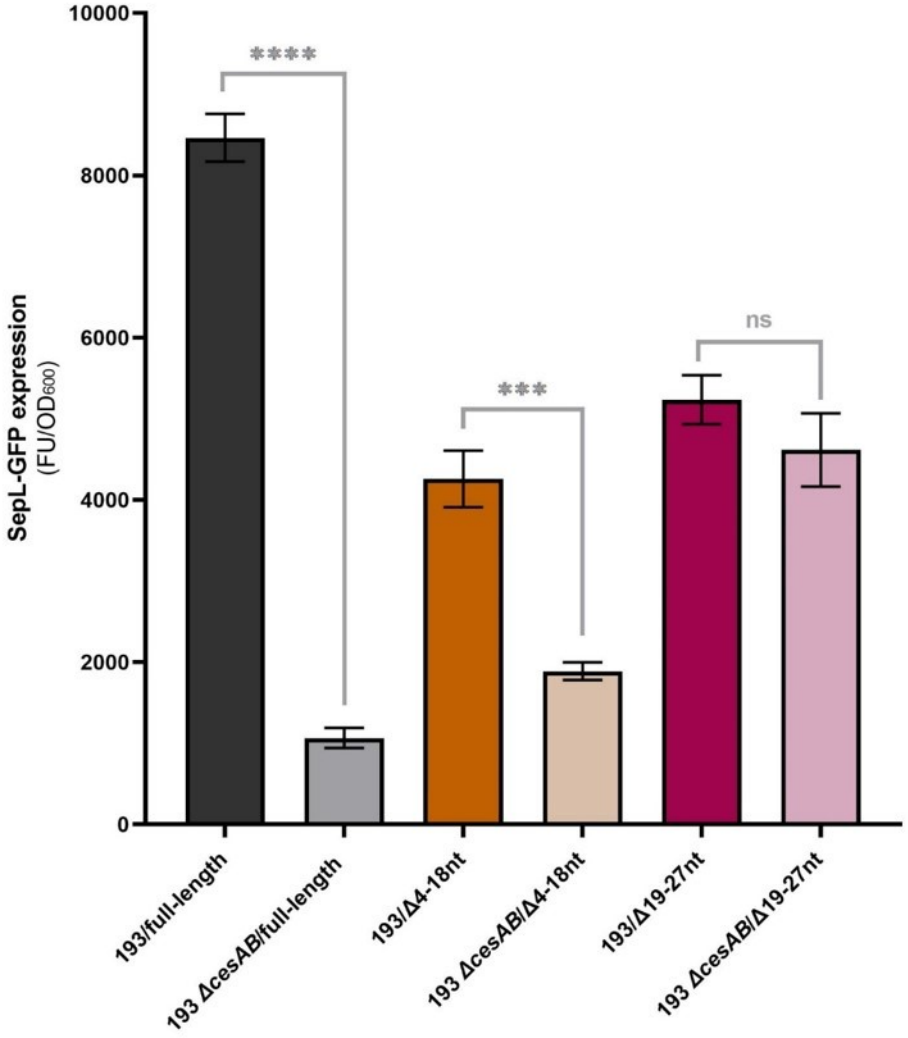


Figure 4.7. SepL-GFP expression from the full-length, Δ4-18nt and Δ19-27nt SepL-GFP constructs in *ΔcesAB* ZAP193 background. Deletion of *cesAB* resulted in significant decrease in SepL-GFP produced from the full-length and Δ4-18nt SepL-GFP as compared to the SepL-GFP produced from those constructs in the wild type EHEC background. No significant change in SepL-GFP expression was detected in the Δ19-27nt construct expressed in *ΔcesAB* ZAP193 background.

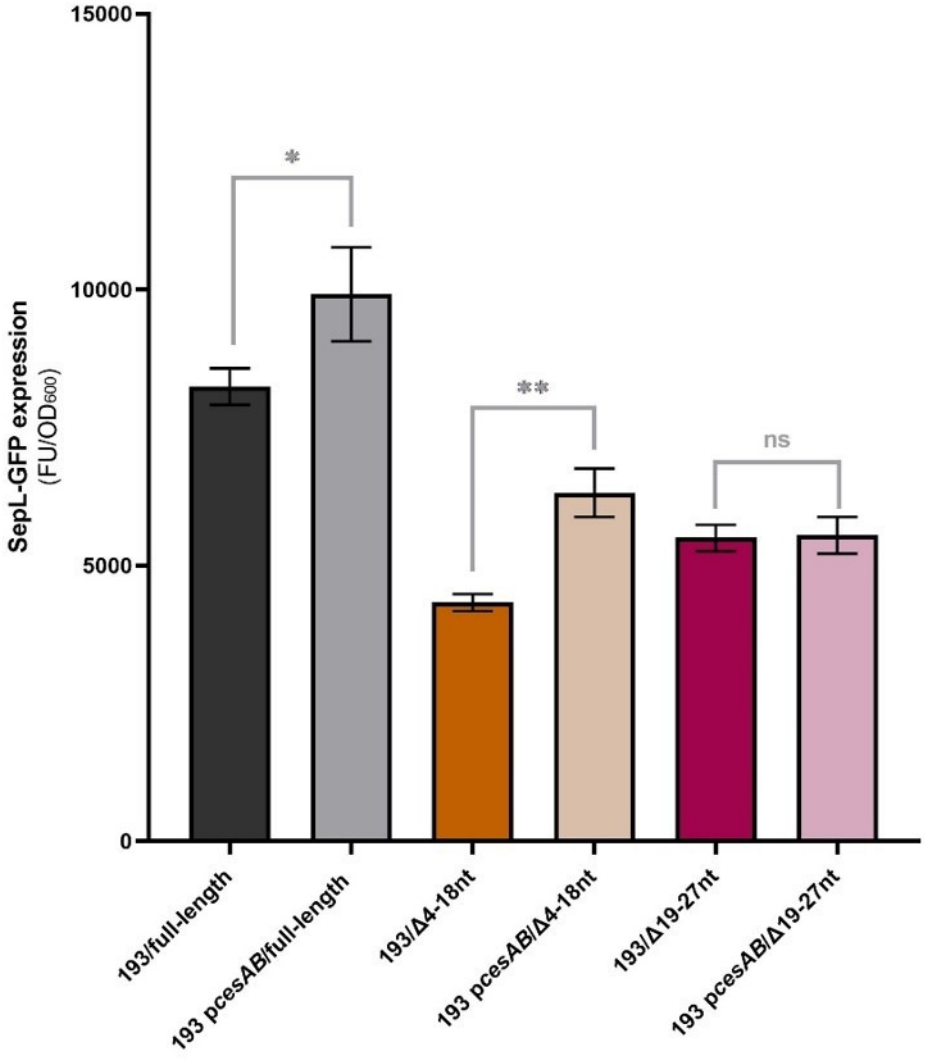


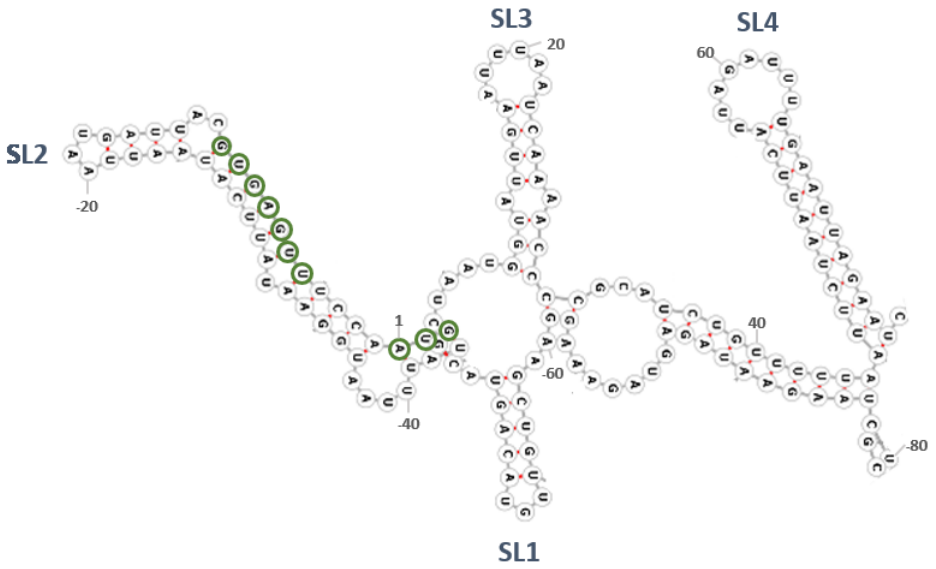
Figure 4.8. SepL-GFP expression from the full-length, Δ4-18nt and Δ19-27nt SepL-GFP constructs in ZAP193 background supplemented with CesAB expression plasmid pWSK*cesAB* (p*cesAB*). Plasmid-based complementation of CesAB resulted in significant increase in SepL-GFP produced from the full-length and Δ4-18nt SepL-GFP as compared to the SepL-GFP produced from those constructs in the wild type EHEC background. No significant change in SepL-GFP expression was detected in the Δ19-27nt construct expressed in ZAP193 pWSK*cesAB* background.

The *sepL* mRNA sequences deleted in the SDM constructs are predicted to form a stem loop structure at SL3 (spanning nucleotides +4 to 24) proximal to the RBS and start codon in the full-length *sepL* transcript. RNAfold software was previously used to map the secondary structure of *sepL* mRNA that was successfully validated by SHAPE experiments described in Chapter 2 of this thesis. Therefore RNAfold was applied to predict the secondary structure of the SL3 SDM constructs (Figure 4.9). This aimed to provide an indication as to whether the regulatory characteristic of the *sepL* transcript region between the 19th and the 27th bases is purely based on sequence complementarity or due to formation of a structural signature recognised by a regulatory factors CesAB and/or Hfq. The predictions indicated that *sepL* mRNA transcript with the 19-27nt deletion should not form the SL3 structure present in the full-length *sepL* transcript (Figure 4.9a&b). Notably the deletion of 19-27nt is predicted to also disrupt another loop formation formed by sequence -38nt to -69nt downstream of the AUG start codon (SL1) that has been previously indicated as a Hfq-binding site (Wang et al., 2018). The 4-18nt deletion construct was also predicted not to form SL3 at the 4-24nt region (Figure 4.9c). The predicted changes in the *sepL* mRNA structure and the recorded insensitivity of the 19-27nt construct to deletion and overexpression of CesAB indicate that the SL3 and/or SL1 are potential structural signatures that mediate the CesAB-*sepL* mRNA interaction.

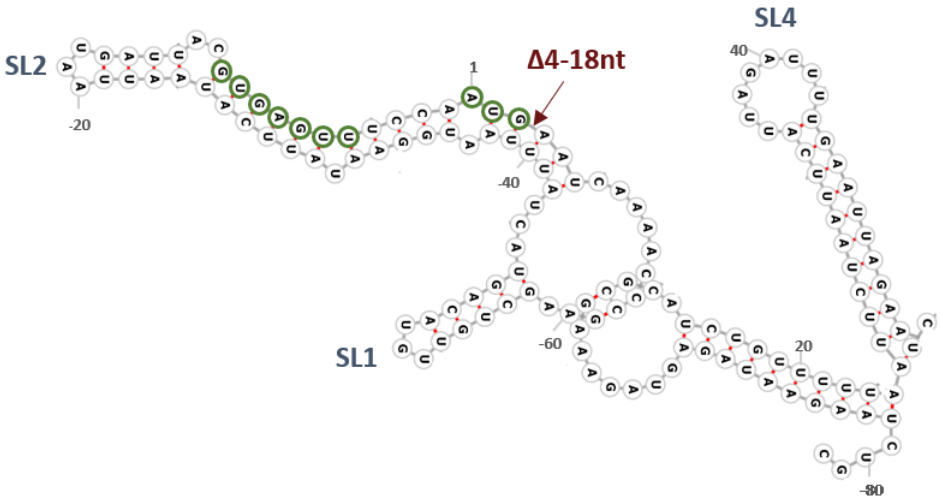
Additionally, the RNAfold predicted the Δ19-27nt mRNA conformation to be more ‘open’ around the RBS and the start codon than the ‘folded’ full-length *sepL* mRNA structure (Figure 4.9a&b). This characteristic could explain the lesser decrease in SepL-GFP expression from the Δ19-27nt construct as compared to that of a full-length SepL-GFP in *cesAB* deletion background. Although the positive effect on SepL expression mediated by CesAB is not exercised by the Δ19-27nt SepL-GFP construct, the negative effect of full-length ‘folded’ *sepL* mRNA structure on SepL expression is compromised in that mutant. This could potentially result in some translational activation of the

Δ19-27nt construct mRNA, bypassing Hfq-mediated degradation of the transcript.

a.


SL3
SL4
SL2
SL1
20
60
1
40
-20
-40
-60
-80

b.


SL4
SL2
Δ4-18nt
SL1
40
1
-20
-40
-60
20
-80

c.

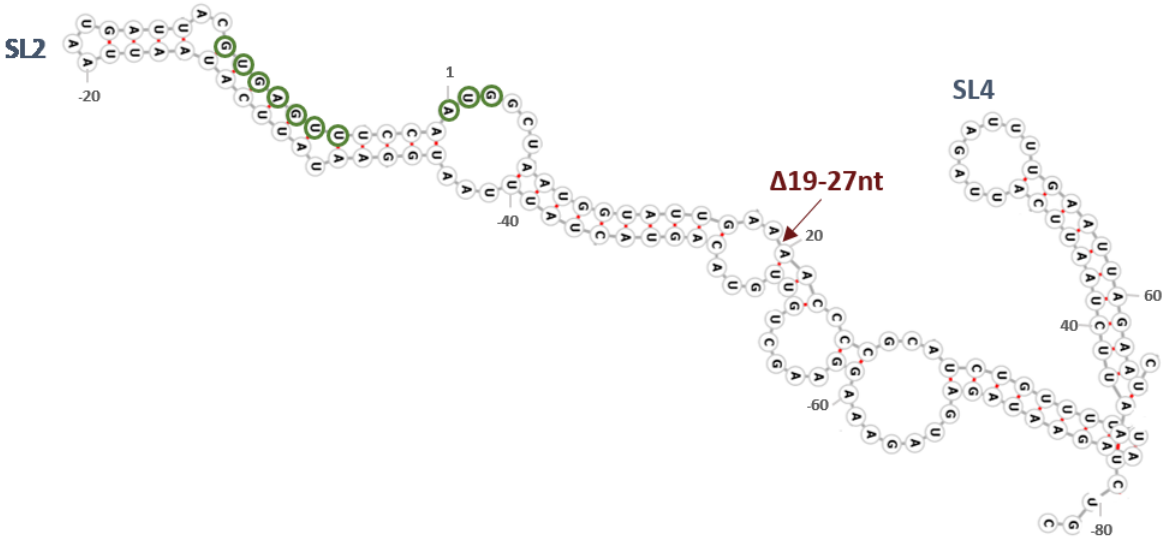

SL2
SL4
Δ19-27nt
1
-20
-40
20
-60
60
40
-80

Figure 4.9. RNAfold computational prediction of mRNA secondary structure around the Shine-Dalgarno site adopted by **a.** full length, **b.** Δ4-18nt and **c.** Δ19-27nt *sepL* transcripts. Shine-Dalgarno site and start codon sequences are highlighted in green and the sequence deletion SDM sites are pointed out in red. The sequences are numbered according to the distance from the start codon AUG. The 5'UTR sequences are assigned negative values decreasing with the distance from the start codon. Stem loops as predicted to form in the full-length *sepL* mRNA conformation are numbered 1-4 (SL1-4) and their formation is indicated in all of the assessed *sepL* mRNA sequences.

4.3 Discussion

The UV-crosslinking approach employed in this study allowed for identification of a direct CesAB-*sepL* mRNA interaction, in a relatively short time frame. This indicated that the CesAB-mediated control of *sepL* mRNA maintenance and translation activation described in Chapter 3, is likely to be facilitated by direct binding of CesAB to *sepL* transcript. However the crosslinking study came with a major shortcoming - it did not allow for the determination of specific CesAB-*sepL* mRNA binding sites. UV-crosslinking has been reported to induce mutations in the cDNA sequences at the sites of protein binding. This is thought to be a result of reverse transcriptase bypassing the crosslinked regions. Identifying the sequences with the highest mutation rates through sequencing has been acknowledged as a valid mean of mapping the specific RNA-protein binding sites (Zhang et al, 2011). Such crosslink induced mutation sites (CIMS) should be present in the cDNAs created from the UV-crosslinking experiments conducted for the purposes of this project. Therefore the produced cDNAs are planned to be sequenced by amplifying them with *sepL* specific primers containing linkers that will be then used to introduce the next generation sequencing-specific adapters at the second round of amplification (Fuellgrabe et al., 2015). The analysis of the sequencing results, with focus on the increase in mutation rate of the specific sequence regions is hoped to determine the *sepL* mRNA-CesAB interaction sites. iCLIP and CRAC that identify transcriptome wide RBP binding partners and RNA binding sites with a single nucleotide resolution can be employed to indicate the specific *sepL* mRNA-CesAB binding sites. Additionally, those methodologies would provide an information on other CesAB-bound RNAs. This could unravel other factors involved in CesAB-mediated post-transcriptional regulation of *sepL* mRNA and potentially other transcripts.

Alternatively to the protein purification and sequencing techniques the electrophoretic mobility shift assay (EMSA) could be applied to determine the proteins bound to the *sepL* mRNA. This methodology makes use of the slower

migration through the polyacrylamide or agarose gel of the protein-nucleic acid complexes as compared to free nucleic acids. In EMSA the nucleic acid of interest is labelled and visualised with radioisotope (canonically P^{32}) or a non-radioactive fluorescence tag such as Cyano dye Cy5 (Hellman et al., 2007; Ruscher et al., 2000). The proteins present in the detected complexes can be determined by subsequent western blot or mass spectrometry analysis. In comparison to RIP and CLIP, EMSA centres the detection of the protein-RNA binding on the RNA as opposed to the protein. This characteristic allows for mapping of the entire proteome interacting with the transcript of interest. However, EMSA is an *in vitro* analysis method. *in vitro* conditions might not be permissive for correct formation of *sepL* mRNA secondary structure, which is likely involved in CesAB-*sepL* mRNA interaction. To partially mitigate this, EMSA using the whole cell extract instead of the purified proteins can be performed. Employment of cell extract EMSA in the study of LEE4 transcript control could uncover other factors that regulate the mRNA prior, together with, or after the proposed CesAB-Hfq interplay and the previously reported Hfq-Spot42 and CsrA control (Wang et al., 2018).

In order to fully understand the mechanism underpinning regulation of LEE4 transcript maintenance, it is essential to consider the dynamics between all of the known acting components. Despite the low probability of Hfq-CesAB binding predicted by PSOPIA computational analysis, the possibility of this interaction cannot be discarded as of yet. Chemical crosslinking results in formation of RNA-protein but also protein-protein covalent bonds. Therefore, pull-down of formaldehyde treated CesAB-6xHis samples and subsequent western blot with commercially available anti-Hfq antibody or anti-FLAG antibody detecting Hfq-FLAG construct, previously produced by in our laboratory, or a mass spectrometry analysis can be used to determine if CesAB binds to the Hfq.

Although the CesAB was shown to directly bind *sepL* mRNA, that interaction might still require association of the CesAB-*sepL* mRNA complex with Hfq either through CesAB-Hfq or Hfq-*sepL* mRNA interaction. To

investigate this the crosslinking and pull-down protocol employed in this thesis can be performed in Δ*hfq E. coli* background or in *in vitro* conditions containing only the purified CesAB and *sepL* mRNA. If the *sepL* mRNA is still co-purified with CesAB-6xHis despite the absence of Hfq, it can be concluded that the formation of CesAB-*sepL* mRNA-Hfq complex is not required for binding of CesAB to *sepL* mRNA. Similarly EMSA with purified CesAB or whole cell extract from *Δhfq E. coli* can be employed. However those approaches cannot be performed in EHEC background. This is a major limitation given the importance of *sepL* mRNA conformation and the EHEC-specific factors acting on LEE4 transcript. Although proven to be difficult, production of *Δhfq* EHEC background would aid the investigation of the requirement for Hfq in CesAB-*sepL* mRNA interaction.

Deletion of 19-27nt of *sepL* mRNA that is predicted to result in misfolding of the transcript around the SL3 and SL1 region was shown to result in insensitivity of the *sepL* mRNA to the CesAB-mediated regulation. Thus suggesting the importance of those structures in CesAB-*sepL* mRNA interaction. However, conducted experiments do not conclude those region as a CesAB binding sites. Observed insensitivity of Δ19-27nt SepL-GFP construct to CesAB could be a result of hindered Hfq interaction. This is likely as the Hfq was shown to bind to the SepL mRNA close to the SL1 site that is predicted to be disrupted in Δ19-27nt *sepL* transcript. Lack of negative Hfq regulation of *sepL* mRNA would render the CesAB function as Hfq antagonist redundant and thus not detectable through SepL expression analysis. Apart from the identification of CIMS by sequencing or CRAC and iCLIP approaches described above, the proposed 19-27nt CesAB interaction site can be assessed by UV crosslinking and pull down of CesAB-6xHis in *ΔsepL* EHEC background expressing the Δ19-27nt construct. Inability to co-purify the Δ19-27nt *sepL* mRNA with CesAB-6xHis would conclude the 19-27nt region as crucial for CesAB binding. Additionally the importance of the 19-27nt *sepL* mRNA region for Hfq-mediated degradation can be assessed by determining the Δ19-27nt *sepL* mRNA transcript stability in overexpression of Hfq.

Determination of the Hfq-*sepL* mRNA binding site involved in Hfq-mediated degradation could further our understanding of the Hfq side of CesAB-Hfq-*sepL* mRNA regulatory interplay. Study by Tree et al., 2014 mapped the Hfq-LEE4 mRNA binding sites. As of yet some of the proposed interactions have not been assigned a function (Wang et al., 2018). Given the documented involvement of Hfq in *sepL* mRNA degradation it would be interesting to determine if disruption of the transcript sequences indicated by Tree et al., will result in increase in transcript maintenance.

If the 19-27nt region is crucial for CesAB interaction, the conditions that are unable to foster CesAB-*sepL* mRNA binding id est Δ19-27nt *sepL* mRNA and *ΔcesAB* EHEC backgrounds should exhibit similar levels of SepL expression. Although the expression of Δ19-27nt construct is reduced in comparison to full-length SepL in wild type EHEC background, the production of Δ19-27nt SepL-GFP is significantly higher than the expression of the full-length SepL in *ΔcesAB* EHEC. This could be explained by the predicted change in *sepL* mRNA secondary structure resulting in no SL3 formation in Δ19-27nt SepL-GFP construct. The Δ19-27nt *sepL* mRNA conformation is predicted to be less obstructive around the RBS and thus is potentially more permissive to translation despite absence of the CesAB antagonism of Hfq-mediated negative regulation. Alternatively apart from hindering the CesAB interaction the Δ19-27nt transcript conformation might also prohibit the Hfq binding, thus getting rid of the regulatory interplay. Additionally to the absence of Hfq-CesAB interplay on Δ19-27nt *sepL* mRNA, the misfolding of the Δ19-27nt transcript could lead to disruption of *sepL* mRNA-CsrA interaction that was shown to positively regulate SepL expression. Thus resulting in reduced Δ19-27nt SepL expression as compared to the full-length SepL.

Translational regulation of LEE by T3SS chaperons have been previously reported in multiple studies as described in subsection 3.1 of this thesis. However none of the studies have indicated direct binding of the chaperone to the regulated transcript. Up to my knowledge, the discovered CesAB-*sepL*

mRNA direct interaction is the first documented instance of binding between the T3SS chaperone and LEE transcript.

Chapter 5 Conclusions

The T3SS is an essential factor for EHEC colonisation. Expression of the micro-injection machinery is tightly controlled and influenced by a multitude of environmental stimuli that notably revolves around activation or repression of the major transcriptional regulator Ler. However, results obtained during this project indicated that unlike the basal apparatus-encoding operons, the LEE4 operon which is composed of the genes encoding the translocon filament, pore-forming proteins and the gatekeeper SepL, is independent of direct Ler transcriptional control. This raises a question about why the production of the translocon-encoding transcript is independent to other components of the apparatus and which processes are orchestrating the evident regulation of LEE4 expression.

High levels of LEE4 transcriptional activation were reported in this study, however the mRNA was shown to undergo rapid degradation under conditions prohibitive to T3SS formation. The high rate of LEE4 transcript decay was shown to be specific to EHEC backgrounds as compared to a laboratory *E. coli* strain that exhibited high levels of detectable LEE4 mRNA with a slower rate of degradation.

Previous work by Wang et al., 2018 established the importance of *sepL* mRNA conformation with regard to SepL production. This study confirmed the computationally predicted *sepL* transcript secondary structure through chemical probing and highlighted the ‘buried’ nature of the RBS and start codon which is likely to be inhibitory to translation. Wang et al., 2018 proposed that the transcript is under negative regulation by the RNA binding protein Hfq and positive control elicited by CsrA. A study by Tree et al., 2014 indicated two Hfq-binding sites around the occluded *sepL* mRNA region. This study indicated that the reported degradation of the folded transcript is mediated by Hfq (Figure

5.1a). CsrA was shown not to be involved in *sepL* mRNA maintenance but is a critical factor to enable translation activation of an 'open form' *sepL* mRNA. The fact the rapid LEE4 transcript decay does not take place in the laboratory *E. coli* background, which also expresses the Hfq, indicates the potential requirement for an EHEC-specific Hfq cargo, sRNA(s), for the observed degradation process.

My research revealed the major role of the CesAB, the LEE1-encoded chaperone of EspA and EspB, to enable translation of both SepL and the translocon proteins, EspADB. The primary novel finding is that the initial function of the CesAB chaperone is to positively regulate the actual production of its own protein cargo. This was shown to be through positive modulation of mRNA stability in opposition to Hfq-mediated degradation. Protein-RNA interaction studies carried out for purposes of this project established the presence of a direct interaction between the *sepL* transcript and CesAB. Mechanistically, the binding of CesAB to *sepL* mRNA represses the degradation either through: (1) inhibition of Hfq activity; (2) stopping ribonuclease access to the transcript; (3) enhancement of ribosomal docking which protects the transcript from degradation (Figure 5.1b). Additionally, while there is some evidence from work in a non-background O157 that the presence of CesAB does trigger translational initiation, the main role for CesAB in in this process in *E. coli* O157 is to inhibit Hfq-mediated degradation. We propose that this is most likely through the CesAB-mediated change in *sepL* mRNA secondary structure that enhances ribosomal access and potentially the binding of other known positive regulators of SepL expression such as CsrA. This is supported by the results indicating that the *sepL* mRNA construct that is predicted to form an altered secondary structure around the defined Hfq-binding site is unresponsive to CesAB control.

Once the LEE4-encoded proteins are made, CesAB can then interact with its protein cargo and assist in SepL-mediated effector

export. CesAB activity on the LEE4 transcript was shown to be diverted by the expression of LEE4-encoded EspA, thus producing a negative regulatory feedback loop. We hypothesise that the *sepL* mRNA that is again unoccupied by CesAB is then subject to the Hfq-Spot42 inhibition of translation proposed by Wang et al., 2018 followed by the Hfq-mediated degradation identified in my work (Figure 5.1c).

Taken together, there is a hierarchical regulation loop in which CesAB switches the LEE4 transcript from a translationally “OFF” state, normally prone to Hfq-mediated degradation, to an “ON” state permissive for production of SepL and the translocon proteins. Subsequent interaction between CesAB and its LEE4-encoded substrates returns the transcript to its “OFF” state (Figure 5.1). To our knowledge this is the first reported instance of a T3SS chaperone also regulating the expression of its subsequent protein substrates at the mRNA level.

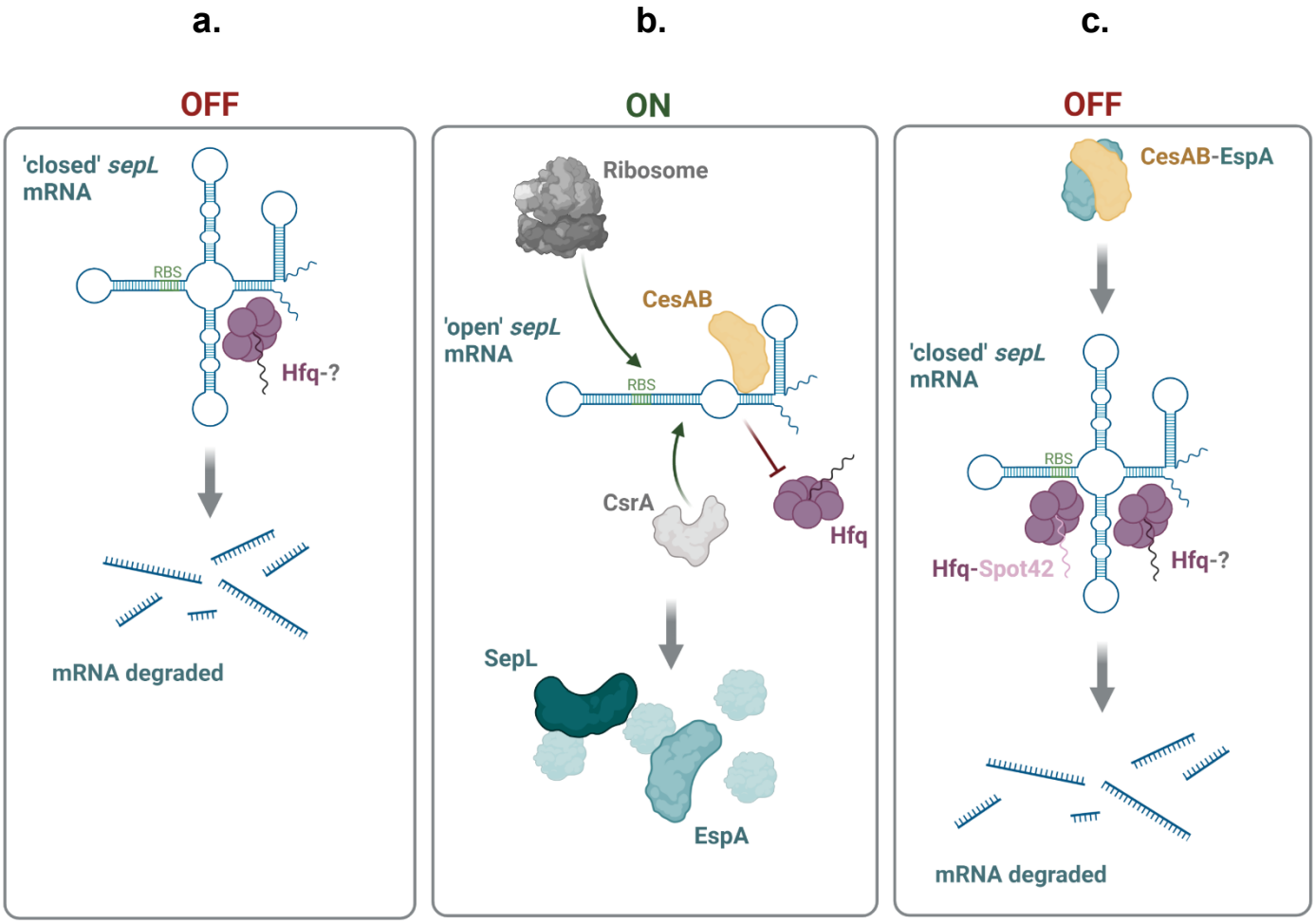


Figure 5.1. Model of CesAB-elicited positive control of *sepL* mRNA maintenance. **a.** *sepL* transcript is produced in a translationally "OFF" state with an obstructed RBS that is prone to degradation by Hfq and as of yet unidentified EHEC-specific factor. **b.** In conditions favourable for T3SS, the LEE1-encoded CesAB chaperone is produced and through a direct interaction with *sepL* mRNA, alters the conformation of the transcript. This achieves two things: (1) it switches the transcript to an "ON" state allowing for ribosomal/CsrA access; (2) It prevents Hfq access to the transcript stopping degradation. CesAB activity results in production of the SepL gatekeeper as well as other LEE4-encoded proteins including translocon filament EspA. **c.** Interaction between CesAB and its substrate EspA diverts CesAB away from the *sepL* transcript. This, along with Hfq-Spot42 repression of translation, returns the mRNA to the translationally "OFF" state that is again prone to Hfq-mediated degradation.

This duality can be a paradigm common to other chaperone proteins. The potential for chaperone-mediated regulation of their cargo production should be examined in order to fully untangle the complex hierarchical control of secretion system assembly. Additionally future studies directly related to this project should pursue the Hfq substrate involved in the EHEC background-specific degradation of the LEE4 transcript, through methods such as an sRNA screen study. Furthermore the evolutionary driver behind why the expression of LEE4 mRNA is independent to the other LEE transcripts should be investigated and the regulatory players behind it determined. It is possible that a transcriptional activator other than Ler controls the expression of LEE4 mRNA for purposes other than the T3SS, such as outer membrane vesicle transport (Kim et al., 2019; Yoon et al., 2011). This would also have to involve another nonT3SS translation activation factor that would oppose the Hfq-mediated degradation as CesAB would not be made if LEE1 is not expressed. Alternatively, the observed disconnect between transcription of LEE operons could related to tight control of the T3SS assembly timing. Binding affinity studies like those performed by Portaliou et al., 2017, could be applied throughout a T3SS expression time course to elaborate on this hypothesis. Additionally, it would be important to investigate if the *sepL* mRNA-CesAB interaction plays a role in localisation of the LEE4-encoded proteins to the basal apparatus. Imaging of the co-localisation of the substrates of interest with protein and RNA-imaging methods such as fluorescence-tag microscopy and MS2 system could be carried out (Dunn et al., 2011; Raj et al., 2008; Wang et al., 2008).

The major take away from this study is that in order to fully understand the regulatory intricacies of a system, the potential steps in expression control between transcript production up to post-translational modifications all need to be considered. A better understanding of the mechanisms orchestrating the assembly of complex bacterial machines, such as the T3SS, will broaden our

knowledge of pathogenesis, identifiy drugable targets and indicate new approaches to combating infection.

Chapter 6 Methods and Materials

6.1 Bacteriology

6.1.1 Bacterial strains, plasmids and culture conditions

The bacterial strains and plasmids used in the study are described in Tables 6.1. and 6.2. respectively.

Minimal essential media (MEM)-HEPES media (Sigma) supplemented with 250 nM $Fe(NO_3)_2$ and 0.1% glucose, was used throughout expression experiments. Luria-Bertani (LB) broth (Oxoid) was used to grow the inocula overnight. Antibiotics were included when required at the following concentrations: chloramphenicol (Cm) 15 µg/ml (resuspended in ethanol), tetracycline (Tet) 25 µg/ml (resuspended in 50/50 ratio of ethanol to water), Ampicillin (Amp) 50 µg/ml (resuspended in water), Rifampicin 100 mg/ml (resuspended in dimethylformamide). The cultures were incubated at 37 ºC at 200 rpm unless stated otherwise.

The pre-inocula of desired bacterial backgrounds were prepared the day before the experiment. A loopful of each strain colony was taken from an agar plate in triplicate (unless stated otherwise) and was inoculated separately in 3 ml of enriched (MEM)-HEPES supplemented with IPTG and appropriate antibiotics where required. The pre-inocula were grown overnight at 37 ºC and 200 rpm shaking unless stated otherwise. Optical density (OD) of the pre-inocula was measured using the spectrophotometer.

All strains were maintained on LB agar plates. The glycerol stocks of the cultures were made by inoculating a single colony into 5ml of LB

broth and grown overnight. 1 ml of the overnight cultures were then spun down and pellets were resuspended in a mixture of LB and 25 % final concentration of glycerol solution supplemented with appropriate antibiotics when required. The cultures were then stored at -80 ℃ indefinitely.

Strain	Description	Source
ZAP193	*Escherichia coli* O157:H7 *stx*⁻	NCTC 12900
RIMD 0509952	*Escherichia coli* O157:H7 Sakai *stx*⁻, Kan^R	Dahan et al., 2004
DH5α	*Escherichia coli* cloning K12-strain	New England Biolabs®
ΔsepL ZAP193	*Escherichia coli* O157:H7 ZAP193 stx⁻, unmarked *ΔsepL*	
ΔLEE1 RIMD 0509952	*Escherichia coli* O157:H7 Sakai stx⁻, Kan^R, ΔLEE1, (LEE1 replaced with tetracycline cassette)	Wawszczyk, 2017
ZAP1004	*Escherichia coli* O157:H7 ZAP193 *stx*⁻, unmarked *Δler*	Low et al., 2006
ΔcesAB RIMD 0509952	*Escherichia coli* O157:H7 Sakai stx⁻, Kan^R *ΔcesAB* (*cesAB* replaced with tetracycline cassette)	This project
ΔcesAB ZAP193	*Escherichia coli* O157:H7 ZAP193 stx⁻, Kan^R *ΔcesAB* (*cesAB* replaced with tetracycline cassette)	This project

Table 1. Bacterial strains used in the study.

Plasmid	Description	Source
pWSK29	Cloning vector, Amp^R	Wang et al., 1991
pACYC184	Low copy number cloning vector, Cm^R	New England Biolabs
pDW6	pAJR70 (Roe et al., 2003) digested with BamHI/KpnI; *sepL* with its own promoter amplified from ZAP193, cloned in frame 5' to GFP, Cm^R	Wang, 2011
pDW26	pACYC184 digested with BamHI/KpnI; 1-51bp *sepL* gene with its own promoter amplified from ZAP193, cloned in frame 5' to GFP and inserted, Cm^R	Wang, 2011
pXLS01 (pWSK*ler*)	pWSK29 digested with SalI and Pst with *ler* 567 bp with the promoter insert, Amp^R	Xu et al., 2012
pWSKLEE1	pWSK29 digested with SalI and SmaI; LEE1 operon with its own promoter amplified from ZAP193, Amp^R	This study
pWSK*cesAB*	pWSK29 digested with SalI and SmaI with *cesAB* insert amplified from ZAP193, Amp^R	This study
pACYC*cesAB*	pACYC184 digested with digested with SalI and SmaI with *cesAB* insert amplified from ZAP193, Cm^R	This study
pWSK*cesAB*-6xHis-N	pWSK29 digested with BamHI and SmaI; with synthesised *cesAB* sequence tagged with 6xHis sequence and the 5' end	This study
pWSK*cesAB*-6xHis-C	pWSK29 digested with BamHI and SmaI; with synthesised *cesAB* sequence tagged with 6xHis sequence and the 5' end	This study
pLEE4-*lacZ*	LEE4::*lacZ* transcriptional fusion, LEE4 promoter amplified from pDW6 then cloned into pRS551-SL	Lab collection (produced by Sean McAteer)
pLEE2-*gfp*	LEE2::*gfp* transcriptional fusion, upstream region of LEE2 cloned into GFP^+ reporter pKC26	Flockhart, 2012
pLEE3-*gfp*	LEE3::*gfp* transcriptional fusion, upstream region of LEE3 cloned into GFP^+ reporter pKC26	Flockhart, 2012
pLEE5-*gfp*	LEE5::*gfp* transcriptional fusion, LEE5 promoter amplified from pDWLEE5 cloned into GFP^+ reporter pKC26	Lab collection (produced by Sean McAteer)
pTOF25	pKO3 digested with BamHI-BamHI with 1,264 bp insert of aph from pUC4K, Cm^R, Km^R, T^S, Suc^S	Merlin et al., 2010
pTOF1-*tetR*	pUC18 digested with EcoRI and HindIII with 162 bp, NotI-FRT-SmaI-FRT-NotI insert and tetracycline cassette insert at SmaI site; Amp^R, Tet^R	Merlin et al., 2010

Table 2. Plasmids used in the study.

Plasmid	Description	Source
pDW6_Δ4-18nt	pDW6 with SDM-produced deletion at nucleotide 4-18 upstream of the *sepL* start codon	Lab collection (produced by Sean McAteer)
pDW6_Δ19-27nt	pDW6 with SDM-produced deletion at nucleotide 19-27 upstream of the *sepL* start codon	Lab collection (produced by Sean McAteer)
pWSK29-*hfq*	pWSK29 digested with BamHI and EcoRI; *hfq* gene amplified from ZAP193 and inserted	Lab collection (produced by Sean McAteer)
pDW66	pWSK29 digested with BamHI and EcoRI; *csrA* gene amplified from ZAP193 and inserted	Wang, 2011
pAJR-EspA	pAJR70 (Roe et al., 2003) digested with BamHI/KpnI; EspA with LEE4 promoter amplified from ZAP193, cloned in frame 5' to GFP, Cm^R	Lab collection (produced by Sean McAteer)
pAJR-EspD	pAJR70 (Roe et al., 2003) digested with BamHI/KpnI; EspD with LEE4 promoter amplified from ZAP193, cloned in frame 5' to GFP, Cm^R	Lab collection (produced by Sean McAteer)
pWSK*espA*	pWSK29 digested with SalI and SmaI with *cesAB* insert amplified from ZAP193, Amp^R	Lab collection (produced by Sean McAteer)

Table 2 (continued). Plasmids used in the study.

6.1.2 **Electrocompetent cell preparation and transformation**

Fresh 50ml of LB was inoculated with a grown pre-inoculum culture so that OD_{600} ~0.02. The mixture was then grown at 37 ºC (unless stated otherwise) and 200 rpm shaking until OD_{600} ~0.6. The culture was then subjected to centrifugation for 20 minutes at 4500 rpm and washed twice with ice-cold MilliQ water. The supernatant was discarded and pelleted cells were resuspended in 1ml of ice-cold MilliQ water. 50µl aliquots of the competent cells were kept on ice.

Plasmid extraction was carried out using a QIAGEN® QIAprep Spin Miniprep kit or E.Z.N.A.® Plasmid DNA Mini Kit I, according to manufacturer instructions.

50 µl of previously prepared competent cells were mixed with 2 µl of MilliQ-eluted plasmid or alternatively the ligation mixture. The mixture was incubated on ice for 30 minutes and then transferred into an ice-cold Bio-Rad 2 mm Gene Pulser®/Micropulser™ cuvette. Electroporation was performed at 12.5 kV/cm, 25 µF capacitance and 1000 Ω, using a Bio-Rad GenePulserXcell electroporator. 1000 µl of LB liquid media was immediately added to the mixture. The bacteria were set to recover at 37 ºC (unless stated otherwise) and 200 rpm for 1 hour. The mixture was then serially diluted in PBS (dilution series 10^{-1} to 10^{-8}) and the aliquots were plated onto LB agar plates containing appropriate antibiotics.

6.1.3 Chemically competent cells preparation and transformation

Plasmid extraction was carried out using a QIAGEN® QIAprep Spin Miniprep kit or E.Z.N.A.® Plasmid DNA Mini Kit I, according to manufacturer instructions.

50 µl aliquots of NEB® 5-alpha Competent *E. coli* (High Efficiency) (New England Biolabs®) were thawed on ice and mixed with 2 µl of extracted plasmid or ligation mixture. The mixture was then incubated on ice for 30 minutes and then heat-treated in a 42 °C water bath for 45 seconds. It was transferred to ice for 2 minutes. Subsequently 0.95 ml of preheated SOC medium (2 % w/v tryptone, 0.5 % w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose) or LB medium was added. The transformation mixture was then incubated at 37 °C for 1 hour with 200 rpm shaking. 150 µl of transformation mixture was plated on LB agar supplemented with antibiotics when appropriate and incubated at 37 °C overnight. Remaining transformation mixture was left at room temperature for a longer recovery time in case no transformants were obtained after initial plating. As a transformation control, pUC18 was used.

6.2 Molecular biology

6.2.1 DNA amplification and detection

DNA amplification was routinely performed by polymerase chain reaction (PCR). A loopful of a single colony was resuspended in 200 µl of MilliQ distilled water for each DNA sample to be tested. The samples were then incubated at 100 ºC for 13 minutes in a dry block. Incubated samples were thoroughly vortexed and subjected to centrifugation at 13000 rpm for 2 minutes. The lysed supernatant of the samples were used as a DNA template for PCR reactions. Alternatively, purified plasmid DNA was used as a DNA template during production of constructs. Primers used for the conventional PCR reactions are listed in Table 6.3.

Ready to use *Taq* 2X PCR mastermix (Biolabs) was used for PCR amplification. All of the primers used for the analysis are listed in Table 3. The volume of each of the reaction components was multiplied by the number of DNA samples to be analyzed plus the positive and negative controls. 22 µl of the mastermix aliquots were distributed into 0.2 ml eppendorfs and 3 µl of previously prepared DNA template was added. The eppendorfs were placed in a PCR machine and the samples were processed under the following conditions: 1 cycle of denaturalization - 5 minutes at 95 ºC, 30 cycles of PCR - denaturalization for 30 seconds at 95 ºC, alignment for 30 seconds at 50-58 ºC (depending on oligonucleotide melting temperature), elongation for 2 minutes at 72 ºC, 1 cycle of final elongation - 10 minutes at 72 ºC, hold at 4 ºC.

The PCR products and 100 bp or 1 kb DNA ladder (Promega) were loaded on previously prepared 1.2 % agarose (BioTools) gel stained with SYBR® Safe DNA stain. The DNA was then separated by

gel electrophoresis at 100 V for 40-60 minutes, depending on the size of the gel. DNA bands were then visualised using ultraviolet light in the G:BOX image capture system (Syngene).

Primer	Sequence (5'-3')	Application
Polymerase chain reaction (PCR)		
cesAB-6xHIS-N_SDM_F	CACCACCACAGCCAAACAAGAAATAAAG	Production of pWSK*cesAB*-6xHis-N and pWSK*cesAB*-6xHis-C constructs by SDM for CesAB-6xHis expression
cesAB-6xHIS-N_SDM_R	GTGGTGGTGCACAATACTCATCCTCTAT	
cesAB-6xHIS-C_SDM_F	CACCACCACTGACAATTTTTAATAAAATAGAC	
cesAB-6xHIS-N_SDM_R	GTGGTGGTGTACTATTTTTCTATTATTTC	
*cesAB*_KO_No	TTGCTGGTCTCGGTACCCGGGGTTGGTCCTTCCTGATAAGGTCGC	Production of *ΔcesAB* ZAP193 and *ΔcesAB* Sakai
*cesAB*_KO_Ni	CGCTCTTGCGGCCGCTTGGAACGGCTCATCCTCTATTTATTATTAATC	
*cesAB*_KO_Co	CCGTTCCAAGCGGCCGCAAGAGCGCCAATCATGATGGTTCATGATGAATG	
*cesAB*_KO_Ci	TCCCATTCGCCACCGGTCGACGAACATTTACTGAGGGCGAGGAGG	
LEE1clone_F	TTGATATCGAATTCCTGCAGCCCGGGTTAATAATCAAGGTCTATCGCAATACG	Production of pWSKLEE1
LEE1clone_R	GCGGCCGCTCTAGAACTAGTGGATCCTTACTTATTAGGGACAAATTTCGCGAG	
LEE1clone_seq_F	TTAATAATCAAGGTCTATCGCAATACG	LEE1 clone sequencing
LEE1clone_seq_R	TTACTTATTAGGGACAAATTTCGCGAG	
*cesAB*clone_F	TTGATATCGAATTCCTGCAGCCCGGGGGACTATAGAGATAATGCGAATCAGG	Production of pWSK*cesAB*
*cesAB*clone_R	GCGGCCGCTCTAGAACTAGTGGATCCCATTCATCATGAACCATCATGATTGG	
*cesAB*clone_F_pACYC	GGCGACCACACCCGTCCTGTCCCGGGGGACTATAGAGATAATGCGAATCAGG	Production of pACYC*cesAB*
*cesAB*clone_R_pACYC	CATGGCCTGCCCGGTTATTAGGATCCCATTCATCATGAACCATCATGATTGG	
*cesAB*clone_seq_F	GGACTATAGAGATAATGCGAATCAGG	*cesAB* clones sequencing
*cesAB*clone_seq_R	CATTCATCATGAACCATCATGATTGG	

Table 3. PCR primers used in the study.

Primer	Sequence (5'-3')	Application
Polymerase chain reaction (PCR)		
pDW6_1	CACCACCACAGCCAAACAAGAAATAAAG	SHAPE *sepL* mRNA structure probing
pDW6_2	GTGGTGGTGCACAATACTCATCCTCTAT	
*sepL*_F	GAATAGAGTAGAAAGGAAGC	Amplification of internal fragment of *sepL* sequence
*sepL*_R	ATTAGAAGAATTTTTTTGCG	
*cesAB*_F	TCGCCGAATTTGATGTCGTAA	Amplification of internal fragment of *cesAB* sequence
*cesAB*_R	CGCTCTTCACCATATGTGTACC	

Table 3(continued). PCR primers used in the study.

6.2.2 **Construction of *ΔcesAB* mutants**

Construction of the *cesAB* deletion mutants in EHEC strains Sakai and 193 was carried out as described previously (Emmerson et al., 2003 and Merlin et al., 2006.) using allelic exchange-based methodology. The deletion cassette was constructed using a two-step crossover PCR utilising *cesAB* flanking regions as described by Merlin et al., 2006. The cassette was cloned into a replacement vector pTOF25 (Cm^R, $repA^{ts}$, *sac*B), using SmaI-SmaII restriction enzymes (Thermo Scientific™). The chemically competent DH5α (laboratory strain) background was transformed with the exchange plasmid and the transformants recovered on LB-chloramphenicol agar plates at 30 °C. The tetracycline FRT cassette was cloned into the NotI restriction site in between fused PCR products and the integrants were obtained by transformation and selection on LB-chloramphenicol-tetracycline agar plates at 30 °C. The two target backgrounds (Sakai or 193) were then transfoemed with the exchange vector. The transformants were recovered at 30 °C and plated on LB-tetracycline-chloramphenicol plates. The overnight culture was grown from a single transformant colony in LB-tetracycline-chloramphenicol at 30 °C. Primary recombinants were selected by plating on LB-chloramphenicol-tetracycline agar plates at 42°C. One primary recombinant (big colony) was repurified on the LB-tetracycline plate at 42°C, and then cultured at 30°C in LB medium in order for the cointegrate to resolve utilising second homologous recombination. The secondary recombinants were selected by plating the serial dilutions of the overnight culture on LB-sucrose-tetracycline plates at 42 °C. The obtained colonies were tested by streaking on chloramphenicol and tetracycline LB agar plates. The tetracycline resistant and chloramphenicol sensitive colonies were then tested by PCR utilising primers for the tetracycline cassette and genes immediately upstream and downstream of *cesAB*.

6.2.3 **Construction of LEE1 and *cesAB* clones**

NEBuilder® HiFi DNA Assembly Cloning Kit was used in order to clone a LEE1 operon (8306 bp) and *cesAB* (314 bp) into the pWSK29 vector. *cesAB* was also cloned into a pACYC184 vector. Primers for insert amplification (LEE1clone_F, LEE1clone_R, *cesAB*clone_F, *cesAB*clone_R, *cesAB*clone_F_pACYC, *cesAB*clone_R_pACYC) were designed according to manufacturer instructions with BamHI and SmaI restriction sites and pWSK29/pACYC184 overlapping region. LongRange PCR KIT (QIAGEN) was used in order to amplify the insert. The PCR product of the right size was gel purified using QIAquick PCR Purification Kit (QIAGEN). pWSK29/pACYC184 plasmid was linearised using FastDigest enzymes BamHI and SmaI according to manufacturer's instructions (Thermo Scientific™). The assembly reaction was set up with NEBuilder HiFi DNA Assembly Cloning Kit reagents according to manufacturer's instructions. The electrocompetent DH5α background were then transformed with the reaction product and transformants selected on LB ampicillin plates. The transformants carrying pWSK29/pACYC184 plasmid with an insert were selected by restriction digestion screening and PCR screens.

6.3 Assessment of promoter activity and RNA detection

6.3.1 Beta-galactosidase assay

Transcriptional activity of LEE4 operon was assessed using the *lacZ* transcriptional fusion (LEE4*::lacZ*) expressed from the IPTG inducible pRS551-SL vector and was constructed by Sean McAteer. 50ml of enriched (MEM)-HEPES media with appropriate antibiotics and IPTG as required were inoculated with the prepared pre-inoculum culture so that OD_{600} ~0.02. The mixture was then cultured at 37 ºC and 200 rpm shaking until OD_{600} ~0.6. The negative control was that of IPTG uninduced cultures. The volume of the cell culture tested was optimised so that the OD_{420} after completion of the reaction was between 0.6 and 0.9. The culture was subjected to centrifugation at 10,000 rpm for 1 minute and the pellet suspended in the Z-buffer (0.1 M $NaPO_4$ pH7, 0.01 M KCl, 0.001 M $MgSO_4$, 0.05 M β-mercaptoethanol, 0.005 % SDS and 50 ml of distilled water) up to 1 ml total volume and 0.1 ml of chloroform. 1 ml of Z-buffer was used as a blank for the OD measurements. The samples were placed on ice and 0.2 ml of *o*-nitrophenyl-β-D-galactoside (ONPG; 4 mg/ml) was added. The samples were transferred to a 28 ºC water bath. The timer was started and the reaction was stopped with 0.5ml of 1M Na_2CO_3 after sufficient yellow colour has developed. The time the reaction was completed was noted. The OD_{420}, OD_{550} and OD_{600} were recorded for each sample. The units of activity (Miller units) representative of the transcriptional activity were calculated using the following equation: $100x\ [(OD_{420} - 1.75 \times OD_{550})]/(\text{time in minutes} \times \text{volume in ml} \times OD_{600})$.

6.3.2 **Reverse-transcriptase quantitative PCR (RT-qPCR)**

In order to detect or quantify the specific bacterial RNAs, reverse-transcriptase quantitative PCR (RT qPCR) was performed. RNaseZap™ RNase Decontamination Solution (Thermo Fisher Scientific) was used to decontaminate bench and equipment throughout all RNA-involving experiments. The bacterial cultures were prepared by inoculation of fresh enriched (MEM)-HEPES media supplemented with IPTG and appropriate antibiotics with pre-inocula of desired bacterial backgrounds so that OD_{600}~0.02 and grown until OD_{600}~0.6.

For assessing the rate of RNA decay, rifampicin was utilised to inhibit RNA synthesis (Mosaei et al., 2020). Once the cultures reached OD_{600}~0.6, rifampicin (final concentration of 100 µg/ml) was added to the culture and incubated for a further 10 minutes with samples being taken after a minute, two minutes and 10 minutes. The specific time-point samples were immediately processed as described below.

1 ml of the cultures were subjected to centrifugation at 10,000 rpm for 1 minute, the supernatants discarded and the pellets resuspended in 100 µl nuclease-free water. In order to preserve the RNA, 200 µl of RNAprotect Bacteria Reagent was added to the samples. The mixture was vortexed for 5 seconds and incubated at room temperature for 5 minutes. The samples were then subjected to centrifugation for 10 minutes at 5000 *g*. The pellets were decanted and kept on ice for further processing or stored at -80 °C if not used on the same day.

6.3.2.1 Column RNA extraction

RNA extractions of the crude cell extracts were carried out using RNeasy Mini kit (QIAGEN) unless specified otherwise. TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0) containing 50 mg/ml lysozyme was prepared fresh for lysis of the RNAprotect-treated bacterial pellets. The samples were resuspended in 100 µl of TE-lysozyme buffer, mixed by vortexing for 10 seconds and incubated at room temperature for 5 minutes on a shaker. 350 µl of manufacturer supplied RLT buffer was added to each sample and the mixture was vortexed vigorously. 250 µl of ethanol was then added and the samples were mixed by pipetting. The column extraction was performed as instructed by the manufacturer. The final elution was in 30-50 µl of MiliQ distilled water. The yield and quality of the extracted RNA was measured by Nanodrop. The samples were kept on ice for further processing or frozen down at -80 °C if not used on the same day.

6.3.2.2 DNAse treatment

The obtained RNA samples were treated with DNase in order to ensure no DNA contamination. The DNase treatment was performed using TURBO DNA-free™ kit (Thermo Fisher Scientific) according to manufacturer's specifications.

Briefly the RNA samples were standardised by dilution in RNAse and Nuclease free water so that 2 µg of RNA (as measured by Nanodrop) was added to each 50 µl reaction. 5 µl of manufacturer supplied buffer and 1 µl of DNase were added to the RNA samples and the mixtures were incubated in a 37 °C water bater bath for 30 minutes. The reaction was stopped by addition of 5 µl of DNase inactivation

reagent supplied by the manufacturer and incubation at room temperature for 5 minutes. The samples were subjected to centrifugation at 2000g for 5 minutes. The cleaned-up RNA-containing supernatant was transferred to a new Eppendorf tube. The success of the reaction was determined by PCR of the samples and gel electrophoresis. No DNA amplification was indicative of no DNA contamination. RNA quantity after the DNAse treatment was determined by Nanodrop measurement.

6.3.2.3 Synthesis of complementary DNA (cDNA)

The DNase treated RNA samples were then subject to first strand complementary DNA (cDNA) synthesis using random primers performed using the iScript™ cDNA Synthesis Kit (Bio-Rad). The RNA template concentration was standardised to 500 ng. All incubation steps were performed in the thermocycler. Reactions containing 4 µl 5x iScript reaction mix, 1 µl iScript reverse transcriptase and 500ng of RNA were made up to 20 µl total volume. The samples were incubated for 5 minutes at 25 °C, 20 minute at 46 °C and 1 minute at 95 °C. The 20 µl reactions were then diluted in 150 µl of nuclease-free distilled water. The obtained cDNAs were kept on ice or frozen at -20 °C until ready to use.

6.3.2.4 Quantitative PCR

Quantitative PCR was performed on three biological replicates of the samples with two or more technical replicates for each of the genes being analysed. Two water controls and three no template controls were included for contamination check.

Gene-specific oligonucleotides were designed using NCBI-Primer BLAST and were tested for specificity with conventional PCR using genomic EHEC DNA prior to the qPCR analysis (Table 6.4).

Serial dilutions (10^{-1} to 10^{-6}) of EHEC genomic DNA of known quantity gifted by Dr. Stephen Fitzgerald were used to produce an internal calibration curve. The internal calibration curve was used to assign standardised arbitrary quantity of the nucleic acid present in each of the tested samples (Fitzgerald S., 2012). Duplicates of qPCR reaction mixes with each serial dilution were loaded onto the 96-well plate for each of the gene tested.

qPCR was performed using SYBR ® GREEN (Bio-Rad) qPCR supermix according to manufacturer's specifications. 10 µl of the reaction mix, 8 µl of cDNA template and 0.4 µl of forward and reverse primers (final concentration 200nM) were diluted with MiliQ distilled water up to a final volume of 20 µl. The reaction set up was as follows: 105 °C hot start, 95 °C for 5 minutes followed by 40 cycles of 95 °C for 10 seconds, 60 °C for 1 minute. This was performed in the CFX96™ Real-Time System (Bio-Rad).

Bio-Rad CFX Manager software was used to analyse the data. Dissociation curves were checked whenever a new primer was used to ensure a single product was generated. The efficiency value was ensured to be between 90%-105%.

Primer	Sequence (5'-3')	Application
Quantitative polymerase chain reaction (qPCR)		
*ler*_RT_F	AGTCCATCATCAGGCACATTAG	*ler* RNA quantification through RT qPCR
*ler*_RT_R	ACTTTGAACTTCCTGCTCTCG	
*cesAB*_RT_F	TGCTGGACTCAGTGTCTCTAT	*cesAB* RNA quantification through RT qPCR
*cesAB*_RT_R	GCTCTTCACCATATGTGTACCC	
*escC*_RT_F	CTATGCGTCAGCGACAGATATAA	*escC* RNA quantification through RT qPCR
*escC*_RT_R	AATGGTTCACTTCCCAGTACC	
*escN*_RT_F	TCAGGCGCTATGTGAAGAAA	*escN* RNA quantification through RT qPCR
*escN*_RT_R	TACGCCTGCTTAGAGGCAAT	
*map*_RT_F	TGACAATGGCAGGCAGATC	*map* RNA quantification through RT qPCR
*map*_RT_R	GACTGCACGGGCATGAACCTC	
*tir*_RT_F	TTCAATTCCTCCTGCACCTC	*tir* RNA quantification through RT qPCR
*tir*_RT_R	CTTACAGGCGTAAATAGCGC	
*sepL*_RT_F	GGCGCCTCTTTACTTGACTGG	*sepL* RNA quantification through RT qPCR
*sepL*_RT_R	TCGAACGACAGCGCCCTAAT	
*espA*_RT_F	CGGTGTTTTTCAGGCTGCGA	*espA* RNA quantification through RT qPCR
*espA*_RT_R	GTGCCGTGGTTGACGCTTTA	
*espD*_RT_F	GGACAGAAAGTAGCCCTTTACC	*espD* RNA quantification through RT qPCR
*espD*_RT_R	CGACTTGTAACCTCACCACTAAT	
*espB*_RT_F	CGTACCTTCAGCAAGACTGGTC	*espB* RNA quantification through RT qPCR
*espB*_RT_R	CGAAGATCTTGCAGACGCCGCC	
*rpoA*_RT_F	GCGCTCATCTTCTTCCGAAT	*rpoA* DNA quantification as a negative control
*rpoA*_RT_R	CGCGGTCGTGGTTATGTG	

Table 4. qPCR primers used in the study.

6.4 Protein detection

6.4.1 **Measurement of GFP fluorescence**

The expression of translational protein-GFP fusions previously constructed in the laboratory was measured in order to compare levels of transcript production and translation in different bacterial backgrounds. Pre-inocula of strains transformed with desired GFP fusions were prepared the day before the experiment, along with pre-inocula of the control strain transformed with empty vector utilised in the formation of the fusion. Fresh enriched (MEM)-HEPES medium supplemented with IPTG and appropriate antibiotics was inoculated with grown pre-inocula so that OD_{600} ~0.02. Inocula were grown until OD_{600} ~0.6. 400 µl of inocula were placed into wells in a Thermo Scientific™ 96-well black plate. Each sample was added in triplicate. A POLARstar® Omega microplate reader spectrophotometer was used to measure sample fluorescence. The background fluorescence measurement (empty vector samples) was subtracted from the readings in order to standardise the data.

6.4.2 **SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)**

SDS-PAGE was performed to separate proteins prior to visualisation by Coomassie staining or immunostaining.

The cultures were grown under conditions specified in section 6.1. The cells were harvested by centrifugation at 4,000 rpm at 4°C for 20 min with two ice-cold PBS washing steps. The crude cell pellets were lysed by resuspension in 200 µl of Laemmli buffer (100 mM Tris pH 6.8, 4% w/v SDS, 0.2% w/v Bromophenol blue, 20% v/v glycerol and 2% β-

mercaptoethanol) (Sigma). The samples were boiled at 99 ˚C for 10 minutes. If the samples were not used immediately they were frozen at -20 ˚C and then thawed and boiled again before loading onto a gel. The gels used were pre-cast Mini-Protean TGX gels (Bio-Rad). 10-15 µl of the boiled samples were loaded onto the gel alongside Page Ruler™ (Fermentas) protein ladder. The proteins were separated by electrophoresis at 90 V for 90 minutes. The running buffer consisted of 25 mM Tris base, 192 mM glycine and w/v=0.1% SDS.

6.4.3 **Coomassie blue staining**

The protein gels were covered in 10-15 ml Coomassie Brilliant Blue G-250 (Bio-Rad) and incubated overnight on a 2D rocker. The gels were washed with water at 15-minute intervals until protein bands were clearly discernible.

6.4.4 **Western blotting**

All of the antibodies used for the western analysis are described in Table 6.5. The proteins were transferred from the protein gel to a nitrocellulose membrane. Transfer buffer consisted of 25 mM Tris,192 mM glycine and 20% v/v methanol.

After the transfer the nitrocellulose membrane was incubated in a blocking buffer composed of 5% w/v non-fat milk in PBS on a 2D rocker for 15 min at room temperature and transferred to stationary incubation at 4 ˚C overnight. After the blocking step the membrane was washed for 15 minutes in PBS two times. The primary antibody was diluted in the 5% w/v non-fat milk PBS solution and the membrane was incubated in the

antibody solution for a minimum of 1 hour at room temperature or overnight at 4 ˚C on the 2D rocker. The membrane was then washed with PBS for 15 min, the wash was repeated 4 times. The species specific HRP-labelled secondary antibody was diluted in the 5% w/v non-fat milk PBS solution and the membrane incubated in the solution for 1 hour at room temperature or at 4 ˚C overnight. The membrane was then washed with PBS for 15 min, the wash was repeated 4 times.

The immunostained proteins were visualised by chemiluminescence imaging. This is based on addition of enhanced chemiluminescence (ECL) solutions. The stocks of ECL solutions were made as follows, the ECL 1 was composed of 1 ml luminol, 440 µl coumaric acid solution and 10 ml 1M Tris-HCl at pH 8.5 diluted in distilled water to 100 ml final volume and ECL2 was composed of 64 µl hydrogen peroxide, 10 ml 1M Tris-HCl at pH 8.5 diluted in distilled water to 100 ml final volume. The membranes were flooded with equal volumes of ECL1 and ECL2 solutions and incubated for 5 min on a 2D rocker. The light emitted from the reaction, indicative of the immunostained protein was captured on X-ray film in a dark room.

Protein	Description	Additional information	Source
SepL	Rabbit polyclonal	1:10000 dilution	In house
EspA	Murine monoclonal	0.5µg/ml concentration	McNeilly et al., 2010, in house
EspD	Murine monoclonal	0.5µg/ml concentration	Santos et al., 2015, in house
6xHis-tag	Murine monoclonal	0.1µg/ml concentration, ID: 34660	QIAGEN
Anti-mouse IgG	Polyclonal Goat Anti-Mouse Immunoglobulins, HRP	1:10000 dilution, ID: ab6789	DAKO
Anti-rabbit IgG	Peroxidase-conjugated Swine Immunoglobulins, HRP	1:10000 dilution, ID: ab6721	DAKO

Table 5. Antibodies used in the study.

6.5 CesAB-*sepL* mRNA interactions

6.5.1 Construction of 6xHis-tagged CesAB

Originally the site directed mutagenesis (SDM) introducing a 6xHis-tag to the N-terminus and C-terminus was planned to be performed on the previously produced pWSK*cesAB* construct. This was unsuccessful and a T4 ligase-based cloning of the synthetically produced (Integrated DNA Technologies) *cesAB* sequences flanked with 6xHis-tag at 5' or 3'-end into a pWSK29 expression vector was performed and yielded the pWSK*cesAB*-6xHis-N and pWSK*cesAB*-6xHis-C CesAB-6xHis expression constructs.

6.5.1.1 Site directed mutagenesis (SDM)

Site directed mutagenesis was performed using Q5® Site-directed mutagenesis kit according to manufacturer's instructions.

Primers *cesAB*-6xHIS-N_SDM_F, *cesAB*-6xHIS-N_SDM_R, *cesAB*-6xHIS-C_SDM_F and *cesAB*-6xHIS-N_SDM_R were used to introduce the 6xHis-tag insert into the pWSK*cesAB* construct and were designed using the NEBaseChanger tool recommended by the manufacturer. The 6xHis-tag sequence was split between the forward and the reverse primers for the long fragment insertion methodology. The Q5 polymerase amplification step (conditions described in section 6.2.1) yielded the correct size products (5766 bp) that were gel purified using QIAquick PCR Purification Kit (QIAGEN). The obtained linear fragments were subsequently treated by the manufacturer provided KLD mix comprised of kinase, ligase and DpnI to remove the DpnI susceptible dam methylated parental DNA and circularise the plasmid

carrying the desired *cesAB*-6xHis sequence (Weinder et al., 1994). The chemically competent DH5α *E. coli* cloning background was transformed with the obtained pWSK*cesAB*-6xHis-N_SDM and pWSK*cesAB*-6xHis-C_SDM constructs and the transformants selected on ampicillin LB plates (as described in section 6.1.3). The colonies were checked for the carriage of the pWSK*cesAB*-6xHis-N_SDM and pWSK*cesAB*-6xHis-C_SDM constructs with PCR analysis using *cesAB* specific primers and sequenced (as described in section 6.2.1). The expression of the CesAB-6xHis was assessed by Western blot analysis of crude cell extracts using anti-His antibody (Biolegend) (as described in section 6.4.4).

6.5.1.2 Cloning of synthetic *cesAB*-6xHis fragments into pWSK29

cesAB sequence flanked by 6xHis sequence at either the 5' or 3'-end of the sequence and with the BamHI and SmaI restriction sites was synthesised by Integrated DNA Technologies. BamHI and SmaI restriction enzyme digestion using the Thermo Scientific™ FastDigest enzymes was performed according to the manufacturer's specifications for the synthesised insert sequences and the pWSK29 expression vector in order to produce compatible overhanging (sticky) ends. The cut pWSK29 was run on an agarose gel and both the cut vector and the inserts were purified using the QIAquick PCR Purification Kit. The obtained products were ligated using New England BioLabs® T4 ligase and T4 DNA ligase buffer according to manufacturer's instructions. The DH5α *E. coli* cloning background was transformed with the ligation mixture (as described in section 6.1.3). The expression of the CesAB-6xHis was assessed by Western blot analysis of crude cell extracts using anti-His antibody (Biolegend) (as described in section 6.4.4).

6.5.2 **CesAB crosslinking**

The electro-competent *E. coli* O157 ZAP193 were transformed with the pWSK*cesAB*-6xHis-N and pWSK*cesAB*-6xHis-C constructs as well as the negative control (untagged pWSK*cesAB*) (as described in section 6.1.2). 200 ml of 0.1% glucose and 250 nM $Fe(NO_3)_3$ supplemented (MEM)-HEPES media with ampicillin were inoculated with pWSKcesAB-6xHis-N ZAP193 cultures and were grown under IPTG induction until OD_{600}~0.6. The cultures were then chemically or UV crosslinked as described below.

6.5.2.1 Chemical crosslinking

5.42 ml of 37% formaldehyde was added to the 200 ml culture prepared as above. The culture was agitated for 30 min at 100 rpm. The reaction was stopped by addition of 13.7 ml of ice-cold 2M glycine. The crosslinked cells were harvested by centrifugation at 4,000 rpm at 4°C for 20 min with two ice-cold PBS washing steps and frozen at -20°C.

6.5.2.2 UV crosslinking

As recommended by Tree et al., 2018 the 200ml culture was exposed to 1800 mJ cm^{-2} of UV-C (254 nm) using the UV Stratalinker 1800 (Stratagene). The crosslinked cells were harvested by centrifugation at 4,000 rpm at 4°C for 20 minutes with two ice-cold PBS washing steps and frozen at -20°C.

6.5.3 **Protein purification and detection**

His-tagged protein purification was performed by resin column purification under denaturing conditions using the QIAExpress® Ni-NTA Fast Start kit according to the manufacturer's instructions. Protein fractions were collected at the initial lysate flow through, each with two wash and the elution steps. The proteins in the collected fractions were separated by SDS-PAGE and visualised by Coomassie Blue staining and western blotting analysis using anti-His antibody was performed to confirm the presence of His-tagged CesAB in the obtained fractions (as described in section 6.4).

6.5.4 **Reverse-crosslinking and proteinase K treatment**

In order to recover the CesAB-crosslinked RNA the chosen protein fractions were subsequently treated with 10 mg/ml proteinase K at 45 ˚C overnight. Additionally, the formaldehyde crosslinked samples were reverse-crosslinked by incubation with 0.3 M (final concentration) NaCl at 65 ˚C for 6 hours prior to the proteinase K treatment.

6.5.5 **Phase separation RNA extraction**

The phase separation using TRIzol reagent (Invitrogen) was performed to extract the total RNA. The procedure was performed on ice. 650 µl of TRIzol reagent was added to the samples that were vortexed for 3 minutes with 50 µl glass beads. The samples were then incubated at 65°C for 10 minutes and placed on ice for 10 minutes. 300 µl of chloroform/isoamyl alcohol (IAA) was added to the suspension and 80 µl

of 100 mM NaOAc pH5.2 added. Samples were vortexed and subjected to centrifugation for 5 minutes at 14000 rpm. The upper phase was added to 500 µl phenol:chloroform:IAA and spun for 2 minutes at 14000 rpm. The upper phase was removed and added to 500 µl of chloroform:IAA. It was phase separated for 2 minutes at 14000 rpm and the upper phase was removed and added to 1 ml ethanol and 20 µg glycogen. RNA was precipitated overnight at -20°C. It was then pelleted following centrifugation for 30 minutes at 14000 rpm at 4°C and the ethanol removed. Pellets were washed with 70% ethanol, air-dried and resuspended in 50 µl MilliQ water. The quantity and quality of the obtained RNA was measured by Nanodrop.

6.5.6 *sepL* mRNA detection

The RNA samples obtained from the crosslinking experiments were DNAse treated using the TURBO DNA-free™ kit (Thermo Fisher Scientific) and cDNAs for the samples were generated using the iScript™ cDNA Synthesis kit (Bio-Rad) as described in section 6.3. In order to determine the presence of *sepL* mRNA in the crosslinked samples, quantitative PCR was performed on the cDNA samples using *sepL* specific primers as described in section 6.3. Additionally, the samples were probed for presence of *rpoA* mRNA using *rpoA* specific primers. This served as a negative control to ensure only crosslinked RNAs were being detected.

6.6 mRNA structure probing

Selective 2-hydroxyl acylation by primer extension (SHAPE) methodology was utilised in order to map the *sepL* mRNA secondary structure. The protocol was adapted from that described by Mortimer et al., 2012.

6.6.1 *in vivo* NAI modification and RNA extraction

Modification of EHEC *in vivo* was carried out using SHAPE chemical NAI produced by Scott Cockroft's group using published protocols. Overnight culture of Sakai was set up at 37 °C. 1:100 dilution of the overnight culture was grown until OD_{600}= 0.6 at 37 °C. NAI was added to the 5 ml aliquots of the inoculum to the desired final concentration (25 mM, 50 mM, 100 mM). The mixture was then incubated for 10 minutes at 37 °C with shaking. It was then subjected to centrifugation for 5 minutes at 4000 rpm at 4 °C, the supernatant removed and the pellet flash-frozen in liquid nitrogen and stored at -80 °C or on ice if used immediately. The RNA extraction was performed as described in section 6.5.5 with the only difference being the use of GTC Phenol instead of TRIzol.

6.6.2 Oligonucleotide labelling

100µl of 10µM oligonucleotide (Table) solutions were spun through Illustra ProbeQuant G-50 Micro Column for 4 minutes at 3.4rpm in order to desalt the oligonucleotide. The labelling reaction of 2µl 10x PNK buffer, 2µl 100mM DTT, 5µl 10 µM oligonucleotide, 1µl PNK

enzyme, 3µl ^{32}P-gamma ATP and 7µl MiliQ water was incubated at 37°C for 1 hour. The reaction was diluted with 50µl MiliQ water and spun through ProbeQuant G-50 column. The oligonucleotides were stored on ice or at -20°C in the dedicated radiation laboratory.

6.6.3 **Sequencing ladder production**

Template plasmid was extracted using QIAGEM® QIAprep Spin Miniprep kit according to manufacturer instructions. Minimum of 5µg of DNA plasmid was mixed with 10 µl 2M NaOH and 10 µl 2 mM EDTA and diluted up to 100 µl with MiliQ water. The mixture was incubated for 30 minutes at 37 °C. 1/10 V 3 M NaAc pH 5.2 and 20 µg of glycogen were added and the mixture was then EtOH precipitated overnight at -20°C. The pellet was resuspended in 35 µl of MiliQ water. The annealing reaction composed of 7 µl of DNA, 1 µl of labelled oligonucleotide and 2 µl 5x Sequanase buffer was incubated for 10 minutes at 65°C and then cooled down to 10°C in a thermoblock. The labelling mix was made by adding 1µl 100 mM DTT, 2µl commercial labelling mix and 2µl diluted Sequenase to the annealing mix. 2.5µl of each dideoxynucleotide was preheated at 37°C for 3-5 minutes and 3.5µl of the labelling mix was added to each dideoxynucleotide. The mixtures were incubated for 5-10 minutes at 37°C. 4µl of stop solution was added to each mixture. The dideoxynucleotide mixes were stored on ice or at -20°C in radiation laboratory.

6.6.4 **Primer extension and PAGE electrophoresis**

1.75µl of minimum 10µg RNA and 1µl of labelled oligonucleotide were incubated for 3 minutes at 85°C and put on ice for 3-5 minutes. 0.5 µl 5 mM dNTP, 1 µl 5x RT buffer, 0.25 µl 100 mM DTT, 0.25 µl RNasin, 0.25µl Super Script III were mixed and added to RNA/oligonucleotide solution and incubated for 50 minutes at 50 °C. 0.5 µl of ExoI and 0.5 µl of RNAseIf were added to the reaction and incubated for 20-30 minutes at 37 °C. 5 µl of FA loading dye was added to the reaction. The reaction was denatured for 2 minutes at 95 °C along dideoxynucleotide ladder prepared previously and snap chilled on ice. 8 % PPA-Urea gel was prepared by adding 250 µl 10 % APS and 25 µl TEMED to 45 ml of 8 % PPA-Urea. The primer extension reaction was loaded on the sequencing gel and the samples were run at 25 W.

6.7 Software

GraphPad Prism 9.0 was used to perform all statistical analysis. The main analysis involved two-tailed, two-sample unequal variance t-test for which *p* values less then 0.05 were considered significant.

ApE software was used to design the oligonucleotides unless stated otherwise.

GraphPad Prism 9.0 and BioRender tools were used for generation of most of the figures.

PSOPIA protein-protein binding software was used to predict probability of Hfq-CesAB (Murakami et al., 2014). RNA-protein binding prediction software RPISeq (Mupiralla et al., 2011), was used to computationally assess the probability of CesAB-*sepL* mRNA interaction.

Predictions of *sepL* mRNA secondary structures were generated using RNAFold software.

Bibliography

Adamu, M. T., Shamsul, B. M. T., Desa, M. N., Khairani-Bejo, S. (2014). A Review on *Escherichia coli* O157: H7-The Super Pathogen. *Health*, 5(2), 118-134.

Abduljalil, J.M. (2018). Bacterial riboswitches and RNA thermometers: Nature and contributions to pathogenesis. *Non-coding RNA Research*, 3(2), pp.54–63.

Abrusci, P., McDowell, M. A., Lea, S. M., Johnson, S. (2014). Building a secreting nanomachine: a structural overview of the T3SS. *Current opinion in structural biology*, 25, 111-117

Ahmed, W., Zheng, K. and Liu, Z.-F. (2016). Small Non-Coding RNAs: New Insights in Modulation of Host Immune Response by Intracellular Bacterial Pathogens. *Frontiers in Immunology*, 7.

Aiba, H. (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. *Current Opinion in Microbiology*, 10(2), pp.134–139.

Akopian, D., Shen, K., Zhang, X. and Shan, S. (2013). Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annual Review of Biochemistry*, 82(1), pp.693–721.

Aldridge, P., Karlinsey, J. and Hughes, K.T. (2003). The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL. *Molecular Microbiology*, 49(5), pp.1333–1345.

Anderson, M.C., Vonaesch, P., Saffarian, A., Marteyn, B.S. and Sansonetti, P.J. (2017). *Shigella sonnei* Encodes a Functional T6SS Used for Interbacterial Competition and Niche Occupancy. *Cell Host & Microbe*, 21(6), pp.769-776.e3.

Andrade, J.M., Santos, R.F., Chelysheva, I., Ignatova, Z. and Arraiano, C.M. (2018). The RNA -binding protein Hfq is important for ribosome biogenesis and affects translation fidelity. *The EMBO Journal*, 37(11).

Archer E. J., Simpson M. A., Watts N. J., O'Kane R., Wang B., Erie D. A., McPherson A., Weeks K. M. (2013). Long-range architecture in a viral RNA genome. *Biochemistry 52*, 3182–3190.

Baker, C.S., EöryL.A., Yakhnin, H., Mercante, J., Romeo, T. and Babitzke, P. (2007). CsrA Inhibits Translation Initiation of *Escherichia coli hfq* by Binding to a Single Site Overlapping the Shine-Dalgarno Sequence. *Journal of Bacteriology*, 189(15), pp.5472–5481.

Bakowski, M.A., Braun, V. and Brumell, J.H. (2008). Salmonella-Containing Vacuoles: Directing Traffic and Nesting to Grow. *Traffic*, 9(12), pp.2022–2031.

Barros, S.A., Yoon, I. and Chenoweth, D.M. (2016). Modulation of the *E. coli rpoH* Temperature Sensor with Triptycene-Based Small Molecules. *Angewandte Chemie*, 128(29), pp.8398–8401.

Beck, H.J., Fleming, I.M.C. and Janssen, G.R. (2016). 5'-Terminal AUGs in *Escherichia coli* mRNAs with Shine-Dalgarno Sequences: Identification and Analysis of Their Roles in Non-Canonical Translation Initiation. *PLOS ONE*, 11(7), p.e0160144.

Bédard, A.-S.V., Hien, E.D.M. and Lafontaine, D.A. (2020). Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863(3), p.194501.

Bellmeyer, A., Cotton, C., Kanteti, R., Koutsouris, A., Viswanathan, V.K. and Hecht, G. (2009). Enterohemorrhagic *Escherichia coli* suppresses inflammatory response to cytokines and its own toxin. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 297(3), pp.G576–G581.

Bertram, R. and Schuster, C.F. (2014). Post-transcriptional regulation of gene expression in bacterial pathogens by toxin-antitoxin systems. *Frontiers in Cellular and Infection Microbiology*, 4.

Betts-Hampikian, H.J. and Fields, K.A. (2010). The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. *Frontiers in Microbiology*, 1.

Bhatt S., Edwards A.N., Nguyen H.T.T., Merlin D., Romeo T., Kalman D. (2009). The RNA binding protein CsrA is a pleiotropic regulator of the locus of enterocyte effacement pathogenicity island of enteropathogenic *Escherichia coli*. *Infect Immun*., 77, pp.3552–3568. https://doi.org/10.1128/IAI .00418-09.

Bhatt, S., Romeo, T., Kalman, D. (2011). Honing the message: post-transcriptional and post-translational control in attaching and effacing pathogens. *Trends in microbiology*, 19(5), 217-224.

Bill, A., Espinola, S., Guthy, D., Haling, J.R., Lanter, M., Lu, M., Marelli, A., Mendiola, A., Miraglia, L., Taylor, B.L., Vargas, L., Orth, A.P. and King, F.J. (2021). EndoBind detects endogenous protein-protein interactions in real time. *Communications Biology*, 4(1).

Blount, Z. D. (2015). The natural history of model organisms: The unexhausted potential of *E. coli*. *Elife*, 4, e05826.

Bobrovskyy, M. and Vanderpool, C.K. (2014). The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria. *Frontiers in Cellular and Infection Microbiology*, 4.

Bolognesi, B. and Lehner, B. (2018). Protein Overexpression: Reaching the limit. *eLife*, 7.

Bolton, D. J. (2011). Verocytotoxigenic (Shiga toxin–producing) *Escherichia coli*: virulence factors and pathogenicity in the farm to fork paradigm. *Foodborne pathogens and disease*, 8(3), 357-365.

Bandyra, K. J., Said, N., Pfeiffer, V., Górna, Maria W., Vogel, J. and Luisi, Ben F. (2012). The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *Molecular Cell*, 47(6), pp.943–953.

Bardill, J.P., Zhao, X. and Hammer, B.K. (2011). The *Vibrio cholerae* quorum sensing response is mediated by Hfq-dependent sRNA/mRNA base pairing interactions. *Molecular Microbiology*, 80(5), pp.1381–1394.

Brizzard, B. and Chubet, R. (2001). Epitope tagging of recombinant proteins. *Current Protocols in Neuroscience*, Chapter 5, p.Unit 5.8.

Brackmann, M., Nazarov, S., Wang, J. and Basler, M. (2017). Using Force to Punch Holes: Mechanics of Contractile Nanomachines. *Trends in Cell Biology*, 27(9), pp.623–632.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Current Opinion in Microbiology*, 10(2), pp.102–109.

Brasili, D., Watly, J., Simonovsky, E., Guerrini, R., A. Barbosa, N., Wieczorek, R., Remelli, M., Kozlowski, H. and Miller, Y. (2016). The unusual metal ion binding ability of histidyl tags and their mutated derivatives. *Dalton Transactions*, 45(13), pp.5629–5639.

Braun, F. (1998). Ribosomes inhibit an RNase E cleavage which induces the decay of the *rpsO* mRNA of *Escherichia coli*. *The EMBO Journal*, 17(16), pp.4790–4797.

Brown, L. and Elliott, T. (1996). Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires host factor I, an RNA-binding protein encoded by the *hfq* gene. *Journal of bacteriology*, 178(13), pp.3763–3770.

Bruce, H.A., Du, D., Matak-Vinkovic, D., Bandyra, K.J., Broadhurst, R.W., Martin, E., Sobott, F., Shkumatov, A.V. and Luisi, B.F. (2017). Analysis of the natively unstructured RNA/protein-recognition core in the *Escherichia coli* RNA

degradosome and its interactions with regulatory RNA/Hfq complexes. *Nucleic Acids Research*, 46(1), pp.387–402.

Buckley, P. T., Eberwine, J. (2009). RNA Binding Proteins Methods. Encyclopedia of Neuroscience, pp. 383-388.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C. and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology*, 33(2), pp.155–160.

Burkinshaw, B., Strynadka, N. (2014). Assembly and structure of the T3SS. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1843(8), 1649-1663.

Burrowes, E., Baysse, C., Adams, C., and O'Gara, F. (2006). Influence of the regulatory protein RsmA on cellular functions in *Pseudomonas aeruginosa* PAO1, as revealed by transcriptome analysis. *Microbiology*, 152, 405–418.

Bustamante, V.H., Santana, F.J., Calva, E. and Puente, J.L. (2001). Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Molecular Microbiology*, 39(3), pp.664–678.

Campellone, K. G., Robbins, D., Leong, J. M. (2004). EspF U is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. *Developmental cell*, 7(2), 217-228.

Cannistraro, V.J. and Kennell, D. (1985). Evidence that the 5' end of *lac* mRNA starts to decay as soon as it is synthesized. *Journal of Bacteriology*, 161(2), pp.820–822.

Carlson-Banning, K.M., Sperandio, V. (2016) Catabolite and oxygen regulation of enterohemorrhagic *Escherichia coli* virulence. *mBio*, 7, e01852-16

Carlson-Banning, K. M., Sperandio, V. (2018). Enterohemorrhagic *Escherichia coli* outwits hosts through sensing small molecules. *Curr. Opin. Microbiol*, 41, pp:83–88.

Chang, T. H., Huang, H. D., Wu, L. C., Yeh, C. T., Liu, B. J., Horng, J. T. (2009). Computational identification of riboswitches based on RNA conserved functional sequences and conformations. *RNA*, 15(7), pp.1426–1430.

Chase-Topping, M. E., McKendrick, I. J., Pearce, M. C., MacDonald, P., Matthews, L., Halliday, J., Allison, L., Fenlon, D., Low, L. C., Gunn, G., & Woolhouse, M. E. (2007). Risk factors for the presence of high-level shedders of *Escherichia coli* O157 on Scottish farms. *Journal of clinical microbiology*, 45(5), 1594-1603.

Chatterjee, A., Caballero-Franco, C., Bakker, D., Totten, S. and Jardim, A. (2015). Pore-forming Activity of the Escherichia coli Type III Secretion System Protein EspD. *Journal of Biological Chemistry*, 290(42), pp.25579–25594.

Chatzidaki-Livanis, M., Geva-Zatorsky, N. and Comstock, L.E. (2016). *Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut *Bacteroidales* species. *Proceedings of the National Academy of Sciences*, 113(13), pp.3627–3632.

Chao, Y. and Vogel, J. (2010). The role of Hfq in bacterial pathogens. *Current Opinion in Microbiology*, 13(1), pp.24–33.

Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M. and Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *The EMBO Journal*, 31(20), pp.4005–4019.

Chemla, Y., Peeri, M., Heltberg, M.L., Eichler, J., Jensen, M.H., Tuller, T. and Alfonta, L. (2020). A possible universal role for mRNA secondary structure in bacterial translation revealed using a synthetic operon. *Nature Communications*, 11(1).

Chen, L., Balabanidou, V., Remeta, David P., Minetti, Conceição A.S.A., Portaliou, Athina G., Economou, A. and Kalodimos, Charalampos G. (2011). Structural Instability Tuning as a Regulatory Mechanism in Protein-Protein Interactions. *Molecular Cell*, 44(5), pp.734–744.

Chen, S.J. (2008). RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics. *Annual Review of Biophysics*, 37(1), pp.197–214.

Chen, J. and Gottesman, S. (2017). Hfq links translation repression to stress-induced mutagenesis in *E. coli*. *Genes & Development*, 31(13), pp.1382–1395.

Chen, L., Ai, X., Portaliou, Athina G., Minetti, Conceicao A.S.A., Remeta, David P., Economou, A. and Kalodimos, Charalampos G. (2013). Substrate-Activated Conformational Switch on Chaperones Encodes a Targeting Signal in Type III Secretion. *Cell Reports*, 3(5), p.1754.

Chiaruttini, C. and Guillier, M. (2019). On the role of mRNA secondary structure in bacterial translation. *WIREs RNA*, 11(3).

Choi, J., Grosely, R., Prabhakar, A., Lapointe, C.P., Wang, J. and Puglisi, J.D. (2018). How Messenger RNA and Nascent Chain Sequences Regulate Translation Elongation. *Annual Review of Biochemistry*, 87(1), pp.421–449.

Cole, J.R. and Nomura, M. (1986). Changes in the half-life of ribosomal protein messenger RNA caused by translational repression. *Journal of Molecular Biology*, 188(3), pp.383–392.

Connolly, J.P.R., Gabrielsen, M., Goldstone, R.J., Grinter, R., Wang, D., Cogdell, R.J. et al. (2016). A highly conserved bacterial D-serine uptake system links host metabolism and virulence. *PLoS Pathog*, 12

Coulthurst, S. (2019). The Type VI secretion system: a versatile bacterial weapon. *Microbiology*, 165(5), pp.503–515.

Creasey, E.A., Friedberg, D., Shaw, R.K., Umanski, T., Knutton, S., Rosenshine, I. and Frankel, G. (2003). CesAB is an enteropathogenic

Escherichia coli chaperone for the type-III translocator proteins EspA and EspB. *Microbiology*, 149(12), pp.3639–3647.

Croxen, M. A., Finlay, B. B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature reviews. Microbiology*, 8(1), 26.

Davis, T. K., McKee, R., Schnadower, D., Tarr, P. I. (2013). Treatment of Shiga toxin–producing *Escherichia coli* infections*. Infectious Disease Clinics*, 27(3), 577-597.

De Lay, N. and Gottesman, S. (2012). RNase E Finds Some sRNAs Stimulating. *Molecular Cell*, 47(6), pp.825–826.

De Lay, N., Schu, D.J. and Gottesman, S. (2013). Bacterial Small RNA-based Negative Regulation: Hfq and Its Accomplices. *Journal of Biological Chemistry*, 288(12), pp.7996–8003.

Dean, P., Kenny, B. (2009). The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell. *Current opinion in microbiology*, 12(1), 101-109.

Del Campo, C., Bartholomäus, A., Fedyunin, I. and Ignatova, Z. (2015). Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLOS Genetics*, 11(10), p.e1005613.

Delepelaire, P. (2004). Type I secretion in gram-negative bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1694(1-3), pp.149–161.

Deng, W., Yu, H.B., Li, Y. and Finlay, B.B. (2015). SepD/SepL-Dependent Secretion Signals of the Type III Secretion System Translocator Proteins in Enteropathogenic *Escherichia coli*. *Journal of Bacteriology*, 197(7), pp.1263–1275.

Desnoyers, G., Bouchard, M.-P. and Massé, E. (2013). New insights into small RNA-dependent translational regulation in prokaryotes. *Trends in Genetics*, 29(2), pp.92–98.

Desnoyers, G. and Masse, E. (2012). Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes & Development*, 26(7), pp.726–739.

Dewoody, R.S., Merritt, P.M. and Marketon, M.M. (2013). Regulation of the *Yersinia* type III secretion system: traffic control. *Frontiers in Cellular and Infection Microbiology*, 3.

DeVinney, R., Stein, M., Reinscheid, D., Abe, A., Ruschkowski, S. and Finlay, B.B. (1999). Enterohemorrhagic *Escherichia coli* O157:H7 Produces Tir, Which Is Translocated to the Host Cell Membrane but Is Not Tyrosine Phosphorylated. *Infection and Immunity*, 67(5), pp.2389–2398.

Díaz-Guerrero, M., Gaytán, M.O., Soto, E., Espinosa, N., García-Gómez, E., Marcos-Vilchis, A., Andrade, A. and González-Pedrajo, B. (2021). CesL Regulates Type III Secretion Substrate Specificity of the Enteropathogenic *E. coli* Injectisome. *Microorganisms*, 9(5), p.1047.

Diepold, A. and Armitage, J.P. (2015). Type III secretion systems: the bacterial flagellum and the injectisome. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1679), p.20150020.

Diepold, A. and Wagner, S. (2014). Assembly of the bacterial type III secretion machinery. *FEMS Microbiology Reviews*, 38(4), pp.802–822.

Ding, Y., Davis, B.M. and Waldor, M.K. (2004). Hfq is essential for *Vibrio cholerae* virulence and downregulates σE expression. *Molecular Microbiology*, 53(1), pp.345–354.

Donnenberg, M. (2013). *Escherichia coli*: pathotypes and principles of pathogenesis. *Academic Press*.

Dressaire, C., Picard, F., Redon, E., Loubière, P., Queinnec, I., Girbal, L. and Cocaign-Bousquet, M. (2013). Role of mRNA Stability during Bacterial Adaptation. *PLoS ONE*, 8(3), p.e59059.

Dubey A.K. (2005). RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA*, 11(10), pp.1579–1587.

Dubey, A.K., Baker, C.S., Suzuki, K., Jones, A.D., Pandit, P., Romeo, T. and Babitzke, P. (2003). CsrA Regulates Translation of the Escherichia coli Carbon Starvation Gene, *cstA*, by Blocking Ribosome Access to the *cstA* Transcript. *Journal of Bacteriology*, 185(15), pp.4450–4460.

Dugar, G., Svensson, S.L., Bischler, T., Wäldchen, S., Reinhardt, R., Sauer, M. and Sharma, C.M. (2016). The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in *Campylobacter jejuni*. *Nature Communications*, 7(1).

Dunn, K.W., Kamocka, M.M. and McDonald, J.H. (2011). A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology*, 300(4), pp.C723–C742.

Duss, O., Michel, E., Yulikov, M., Schubert, M., Jeschke, G. and Allain, F.H.-T. (2014). Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature*, 509(7502), pp.588–592.

Duss, O., Stepanyuk, Dr.G.A., Puglisi, Dr.J.D. and Williamson, J.R. (2019). Transient Protein-RNA Interactions Guide Nascent Ribosomal RNA Folding. *SSRN Electronic Journal*.

Einhauer, A. and Jungbauer, A. (2001). The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of Biochemical and Biophysical Methods*, 49(1-3), pp.455–465.

Elbaz, N., Socol, Y., Katsowich, N. and Rosenshine, I. (2019). Control of Type III Secretion System Effector/Chaperone Ratio Fosters Pathogen Adaptation to Host-Adherent Lifestyle. *mBio*, 10(5).

Elliott, S.J., Sperandio, V., GirónJ.A., Shin, S., Mellies, J.L., Wainwright, L., Hutcheson, S.W., McDaniel, T.K. and Kaper, J.B. (2000). The Locus of Enterocyte Effacement (LEE)-Encoded Regulator Controls Expression of Both

LEE- and Non-LEE-Encoded Virulence Factors in Enteropathogenic and Enterohemorrhagic *Escherichia coli*. *Infection and Immunity*, 68(11), pp.6115–6126.

Elliott, S. J., Yu, J., & Kaper, J. B. (1999). The Cloned Locus of Enterocyte Effacement from Enterohemorrhagic *Escherichia coli* O157: H7 Is Unable To Confer the Attaching and Effacing Phenotype upon *E. coli* K-12. *Infection and immunity*, 67(8), 4260-4263.

Ellis, M.J., Carfrae, L.A., Macnair, C.R., Trussler, R.S., Brown, E.D. and Haniford, D.B. (2017). Silent but deadly: IS200 promotes pathogenicity in *Salmonella Typhimurium*. *RNA Biology*, 15(2), pp.176–181.

Ermolenko, D.N. and Mathews, D.H. (2020). Making ends meet: New functions of mRNA secondary structure. *WIREs RNA*, 12(2).

Escherich, T. (1885). Die darmbakterien des neugeborenen und sauglings. *Fortshr. Med.*, 3(5) 15-522, 547-554.

Fatima, R. and Aziz, M. (2019). Enterohemorrhagic *Escherichia Coli* (EHEC). Nih.gov. Available at: https://www.ncbi.nlm.nih.gov/books/NBK519509/.

Fei, K., Yan, H., Zeng, X., Huang, S., Tang, W., Francis, M.S., Chen, S. and Hu, Y. (2021). LcrQ Coordinates with the YopD-LcrH Complex To Repress *lcrF* Expression and Control Type III Secretion by *Yersinia pseudotuberculosis*. *mBio*, 12(3).

Feilmeier, B.J., Iseminger, G., Schroeder, D., Webber, H. and Phillips, G.J. (2000). Green Fluorescent Protein Functions as a Reporter for Protein Localization in *Escherichia coli. Journal of Bacteriology*, 182(14), pp.4068–4076.

Felden, B. and Augagneur, Y. (2021). Diversity and Versatility in Small RNA-Mediated Regulation in Bacterial Pathogens. *Frontiers in Microbiology*, 12.

Fender, A., Elf, J., Hampel, K., Zimmermann, B. and Wagner, E.G.H. (2010). RNAs actively cycle on the Sm-like protein Hfq. *Genes & Development*, 24(23), pp.2621–2626.

Figueira, R., Watson, K.G., Holden, D.W. and Helaine, S. (2013). Identification of *Salmonella* Pathogenicity Island-2 Type III Secretion System Effectors Involved in Intramacrophage Replication of *S. enterica* Serovar Typhimurium: Implications for Rational Vaccine Design. *mBio*, 4(2).

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J. and Punta, M. (2013). Pfam: the protein families database. *Nucleic Acids Research*, 42(D1), pp.D222–D230.

Fitzgerald, S., (2012). 'Expression of *hns* and *stpA* in *Salmonella enterica* servoar Typhimurium', [thesis], Trinity College (Dublin, Ireland). School of Genetics and Microbiology, pp 300

Fitzgerald, S.F., Beckett, A.E., Palarea-Albaladejo, J., McAteer, S., Shaaban, S., Morgan, J., Ahmad, N.I., Young, R., Mabbott, N.A., Morrison, L., Bono, J.L., Gally, D.L. and McNeilly, T.N. (2019). Shiga toxin sub-type 2a increases the efficiency of *Escherichia coli* O157 transmission between animals and restricts epithelial regeneration in bovine enteroids. *PLOS Pathogens*, 15(10), p.e1008003.

Flockhart, A.F., Tree, J.J., Xu, X., Karpiyevich, M., McAteer, S.P., Rosenblum, R., Shaw, D.J., Low, C.J., Best, A., Gannon, V., Laing, C., Murphy, K.C., Leong, J.M., Schneiders, T., La Ragione, R. and Gally, D.L. (2011). Identification of a novel prophage regulator in *Escherichia coli* controlling the expression of type III secretion. *Molecular Microbiology*, 83(1), pp.208–223.

Fortune, D.R., Suyemoto, M. and Altier, C. (2005). Identification of CsrC and Characterization of Its Role in Epithelial Cell Invasion in *Salmonella enterica* Serovar Typhimurium. *Infection and Immunity*, 74(1), pp.331–339.

Francis, M.S., Lloyd, S.A. and Wolf-Watz, H. (2001). The type III secretion chaperone LcrH co-operates with YopD to establish a negative, regulatory loop for control of Yop synthesis in *Yersinia pseudotuberculosis*. *Molecular Microbiology*, 42(4), pp.1075–1093.

Franzin, F. M., Sircili, M. P. (2015). Locus of enterocyte effacement: a pathogenicity island involved in the virulence of enteropathogenic and enterohemorragic *Escherichia coli* subjected to a complex network of gene regulation. *BioMed research international*.

Franze de Fernandez, M.T., Eoyang, L., and August, J.T. (1968). Factor fraction required for the synthesis of bacteriophage Qb-RNA. *Nature*, 219, pp.588–590

Friedberg, D., Umanski, T., Fang, Y. and Rosenshine, I. (1999). Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic *Escherichia coli*. *Molecular Microbiology*, 34(5), pp.941–952.

Fröhlich, K.S. and Vogel, J. (2009). Activation of gene expression by small RNA. *Current Opinion in Microbiology*, 12(6), pp.674–682.

Fuellgrabe, M.W., Herrmann, D., Knecht, H., Kuenzel, S., Kneba, M., Pott, C. and Brüggemann, M. (2015). High-Throughput, Amplicon-Based Sequencing of the CREBBP Gene as a Tool to Develop a Universal Platform-Independent Assay. *PLOS ONE*, 10(6), p.e0129195.

Garst, A.D., Edwards, A.L. and Batey, R.T. (2010). Riboswitches: Structures and Mechanisms. *Cold Spring Harbor Perspectives in Biology*, 3(6), pp.a003533–a003533.

Garg, A.X., Salvadori, M., Moist, L.M., Suri, R.S. and Clark, W.F. (2009). Renal prognosis of toxigenic *Escherichia coli* infection. *Kidney International*, 75, pp.S38–S41.

Gaytán, M.O., Martínez-Santos, V.I., Soto, E. and González-Pedrajo, B. (2016). Type Three Secretion System in Attaching and Effacing Pathogens. *Frontiers in Cellular and Infection Microbiology*, 6.

Gendrin, C., Contreras-Martel, C., Bouillot, S., Elsen, S., Lemaire, D., Skoufias, D.A., Huber, P., Attree, I. and Dessen, A. (2012). Structural Basis of Cytotoxicity Mediated by the Type III Secretion Toxin ExoU from *Pseudomonas aeruginosa. PLoS Pathogens*, 8(4), p.e1002637.

Georg, J. and Hess, W.R. (2011). cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiology and Molecular Biology Reviews*, 75(2), pp.286–300.

Goldberg, M.D., Johnson, M., Hinton, J.C.D. and Williams, P.H. (2001). Role of the nucleoid-associated protein Fis in the regulation of virulence properties of enteropathogenic *Escherichia coli*. *Molecular Microbiology*, 41(3), pp.549–559.

Goldwater, P. N., & Bettelheim, K. A. (2012). Treatment of enterohemorrhagic *Escherichia coli* (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC medicine*, 10(1), 12.

Gottesman, S. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends in Genetics*, 21(7), pp.399–404.

Granneman, S., Kudla, G., Petfalski, E. and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proceedings of the National Academy of Sciences*, 106(24), pp.9613–9618.

Grisaru, S. (2014). Management of hemolytic-uremic syndrome in children. *International journal of nephrology and renovascular disease*, 7, 231.

Grohman J. K., Gorelick R. J., Lickwar C. R., Lieb J. D., Bower B. D., Znosko B. M., Weeks K. M. (2013). A guanosine-centric mechanism for RNA chaperone function. *Science*, *340*, 190–195.

Gruber, C.C. and Sperandio, V. (2015). Global Analysis of Posttranscriptional Regulation by GlmY and GlmZ in Enterohemorrhagic *Escherichia coli* O157:H7. *Infection and Immunity*, 83(4), pp.1286–1295.

Grützner, J., Remes, B., Eisenhardt, K.M.H., Scheller, D., Kretz, J., Madhugiri, R., McIntosh, M. and Klug, G. (2021). sRNA-mediated RNA processing regulates bacterial cell division. *Nucleic Acids Research*, 49(12), pp.7035–7052.

Hachani, A., Wood, T.E. and Filloux, A. (2016). Type VI secretion and anti-host effectors. *Current Opinion in Microbiology*, 29, pp.81–93.

Hadjicharalambous, M.R, Lindsay M., A. (2019). Long Non-Coding RNAs and the Innate Immune Response. *Non-Coding RNA*, 5(2), p.34.

Hajnsdorf, E. and Regnier, P. (2000). Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proceedings of the National Academy of Sciences*, 97(4), pp.1501–1505.

Haller, A., Altman, R.B., Soulière, M.F., Blanchard, S.C. and Micura, R. (2013). Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proceedings of the National Academy of Sciences*, 110(11), pp.4188–4193.

Ham, H. and Orth, K. (2012). The role of type III secretion System 2 in *Vibrio parahaemolyticus* pathogenicity. *Journal of Microbiology*, 50(5), pp.719–725.

Hamilton, H.L. and Dillard, J.P. (2005). Natural transformation of *Neisseria gonorrhoeae*: from DNA donation to homologous recombination. *Molecular Microbiology*, 59(2), pp.376–385.

Han, Y., Chen, D., Yan, Y., Gao, X., Liu, Z., Xue, Y., Zhang, Y. and Yang, R. (2019). Hfq Globally Binds and Destabilizes sRNAs and mRNAs in *Yersinia pestis*. *mSystems*, 4(4).

Handzlik, J.E., Tastsoglou, S., Vlachos, I.S. and Hatzigeorgiou, A.G. (2020). Manatee: detection and quantification of small non-coding RNAs from next-generation sequencing data. *Scientific Reports*, 10(1).

Hansen, A.M. and Kaper, J.B. (2009). Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli. Molecular Microbiology*, 73(3), pp.446–465.

Hartland, E.L., Batchelor, M., Delahay, R.M., Hale, C., Matthews, S., Dougan, G., Knutton, S., Connerton, I. and Frankel, G. (1999). Binding of intimin from enteropathogenic *Escherichia coli* to Tir and to host cells. *Molecular Microbiology*, 32(1), pp.151–158.

Hartland, E.L. and Leong, J. (2013). Enteropathogenic and enterohemorrhagic E. coli: ecology, pathogenesis, and evolution. Frontiers in cellular and infection microbiology, 3, p.15.

Hayashi, T. (2001). Complete Genome Sequence of Enterohemorrhagic *Eschelichia coli* O157:H7 and Genomic Comparison with a Laboratory Strain K-12 (Supplement). *DNA Research*, 8(1), pp.47–52.

Hellman, L.M. and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols*, 2(8), pp.1849–1861.

Hockenberry, A.J., Pah, A.R., Jewett, M.C. and Amaral, L.A.N. (2017). Leveraging genome-wide datasets to quantify the functional role of the anti-Shine–Dalgarno sequence in regulating translation efficiency. *Open Biology*, 7(1), p.160239.

Hoe, N.P. and Goguen, J.D. (1993). Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. *Journal of Bacteriology*, 175(24), pp.7901–7909.

Hoekzema, M., Romilly, C., Holmqvist, E. and Wagner, E.G.H. (2019). Hfq-dependent mRNA unfolding promotes sRNA -based inhibition of translation. *The EMBO Journal*, 38(7).

Holmqvist, E., Reimegård, J., Sterk, M., Grantcharova, N., Römling, U. and Wagner, E.G.H. (2010). Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *The EMBO Journal*, 29(11), pp.1840–1850.

Holmqvist, E., Wright, P.R., Li, L., Bischler, T., Barquist, L., Reinhardt, R., Backofen, R. and Vogel, J. (2016). Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *The EMBO Journal*, 35(9), pp.991–1011.

Holmqvist, E. and Vogel, J. (2018). RNA-binding proteins in bacteria. *Nature Reviews Microbiology*, 16(10), pp.601–615.

Hu, W., Qin, L., Li, M., Pu, X. and Guo, Y. (2018). A structural dissection of protein–RNA interactions based on different RNA base areas of interfaces. *RSC Advances*, 8(19), pp.10582–10592.

Huerta-Uribe, A., Marjenberg, Z.R., Yamaguchi, N., Fitzgerald, S., Connolly, J.P.R., Carpena, N., Uvell, H., Douce, G., Elofsson, M., Byron, O., Marquez, R., Gally, D.L. and Roe, A.J. (2016). Identification and Characterization of Novel Compounds Blocking Shiga Toxin Expression in *Escherichia coli* O157:H7. *Frontiers in Microbiology*, 7.

Irie, Y., Starkey, M., Edwards, A.N., Wozniak, D.J., Romeo, T. and Parsek, M.R. (2010). *Pseudomonas aeruginosa* biofilm matrix polysaccharide Psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA. *Molecular Microbiology*, 78(1), pp.158-72

Jacob, W.F., Santer, M. and Dahlberg, A.E. (1987). A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation

of many proteins. *Proceedings of the National Academy of Sciences*, 84(14), pp.4757–4761.

Jackson, D.W., Suzuki, K., Oakford, L., Simecka, J.W., Hart, M.E. and Romeo, T. (2002). Biofilm Formation and Dispersal under the Influence of the Global Regulator CsrA of *Escherichia coli. Journal of Bacteriology*, 184(1), pp.290–301.

Jandhyala, D.M., Ahluwalia, A., Obrig, T. and Thorpe, C.M. (2008). ZAK: a MAP3Kinase that transduces Shiga toxin- and ricin-induced proinflammatory cytokine expression. *Cellular Microbiology*, 10(7), pp.1468–1477.

Janssen, K.H., Diaz, M.R., Golden, M., Graham, J.W., Sanders, W., Wolfgang, M.C. and Yahr, T.L. (2018). Functional Analyses of the RsmY and RsmZ Small Noncoding Regulatory RNAs in *Pseudomonas aeruginosa. Journal of Bacteriology*, 200(11).

Jiang, L., Yang, W., Jiang, X., Yao, T., Wang, L. and Yang, B. (2021). Virulence-related O islands in enterohemorrhagic *Escherichia coli* O157:H7. *Gut Microbes*, 13(1).

Jonas, K., Edwards, A. N., Simm, R., Romeo, T., Römling, U., and Melefors, O. (2008). The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Mol. Microbiol*. 70, 236– 257.

Kaper, J. B., Nataro, J. P., Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nature reviews. Microbiology*, 2(2), 123.

Katsowich, N., Elbaz, N., Pal, R.R., Mills, E., Kobi, S., Kahan, T. et al (2017). Host cell attachment elicits posttranscriptional regulation in infecting enteropathogenic bacteria. *Science*, 355, pp.735–739

Kay E, Dubuis C, Haas D., (2005). Three small RNAs jointly ensure secondary metabolism and biocontrol in *Pseudomonas fluorescens*. *Proc Natl Acad Sci U S A*, 102, pp.17136–17141.

Kendall, M.M., Gruber, C.C., Rasko, D.A., Hughes, D.T. and Sperandio, V. (2011). Hfq Virulence Regulation in Enterohemorrhagic *Escherichia coli* O157:H7 Strain 86-24. *Journal of Bacteriology*, 193(24), pp.6843–6851.

Kendall, M.M. and Sperandio, V. (2007). Quorum sensing by enteric pathogens. *Curr. Opin. Gastroenterol*, 23, pp.10–15

Kim, S.I., Kim, S., Kim, E., Hwang, S.Y. and Yoon, H. (2019). Corrigendum: Secretion of *Salmonella* Pathogenicity Island 1-Encoded Type III Secretion System Effectors by Outer Membrane Vesicles in *Salmonella enterica* Serovar Typhimurium. *Frontiers in Microbiology*, 10.

Kim, K., Palmer, A.D., Vanderpool, C.K. and Slauch, J.M. (2019). The Small RNA PinT Contributes to PhoP-Mediated Regulation of the *Salmonella* Pathogenicity Island 1 Type III Secretion System in *Salmonella enterica* Serovar Typhimurium. *Journal of Bacteriology*, 201(19).

Kirkpatrick, C.L. and Viollier, P.H. (2011). Reflections on a sticky situation: how surface contact pulls the trigger for bacterial adhesion. *Molecular Microbiology*, 83(1), pp.7–9.

Kirov, S.M. (2003). Bacteria that express lateral flagella enable dissection of the multifunctional roles of flagella in pathogenesis. *FEMS Microbiology Letters*, 224(2), pp.151–159.

Korotkov, K.V., Sandkvist, M. and Hol, W.G.J. (2012). The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology*, 10(5), pp.336–351.

Kouse, A.B., Righetti, F., Kortmann, J., Narberhaus, F. and Murphy, E.R. (2021). Correction draft: RNA-Mediated Thermoregulation of Iron-Acquisition Genes in Shigella dysenteriae and Pathogenic *Escherichia coli. PLOS ONE*, 16(6), p.e0252744.

Kresse, A. U., Beltrametti, F., Müller, A., Ebel, F., Guzmán, C. A. (2000). Characterization of SepL of Enterohemorrhagic *Escherichia coli. Journal of bacteriology*, 182(22), 6490-6498.

Kulesus, R.R., Diaz-Perez, K., Slechta, E.S., Eto, D.S. and Mulvey, M.A. (2008). Impact of the RNA Chaperone Hfq on the Fitness and Virulence Potential of Uropathogenic *Escherichia coli. Infection and Immunity*, 76(7), pp.3019–3026.

Laaberki M.H. (2006). Concert of regulators to switch on LEE expression in enterohemorrhagic *Escherichia coli* O157:H7: interplay between Ler, GrlA, HNS and RpoS. *Int. J. Med. Microbiol*., 296 (4–5), pp. 197-210

Lalaouna, D., Prévost, K., Park, S., Chénard, T., Bouchard, M.-P., Caron, M.-P., Vanderpool, C.K., Fei, J. and Massé, E. (2021). Binding of the RNA Chaperone Hfq on Target mRNAs Promotes the Small RNA RyhB-Induced Degradation in *Escherichia coli. Non-Coding RNA*, 7(4), p.64.

Lalaouna, D., Simoneau-Roy, M., Lafontaine, D. and Massé, E. (2013). Regulatory RNAs and target mRNA decay in prokaryotes*. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(6-7), pp.742–747.

Lawhon, S. D., Frye, J. G., Suyemoto, M., Porwollik, S.,McClelland, M., and Altier,C. (2003). Global regulation by CsrA in *Salmonella typhimurium. Mol. Microbiol.* 48, 1633–1645.

Le Bihan, G., Sicard, J.-F., Garneau, P., Bernalier-Donadille, A., Gobert, A.P., Garrivier, A. (2017). The NAG sensor NagC regulates LEE gene expression and contributes to gut colonization by *Escherichia coli* O157:H7. *Front. Cell Infect. Microbiol*, 7(134)

Le Scornet, A. and Redder, P. (2019). Post-transcriptional control of virulence gene expression in *Staphylococcus aureus. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1862(7), pp.734–741.

Leamy, K.A., Assmann, S.M., Mathews, D.H. and Bevilacqua, P.C. (2016). Bridging the gap between *in vitro* and *in vivo* RNA folding. *Quarterly Reviews of Biophysics*, 49.

Lee, T., Feig, A.L. (2008). The RNA binding protein Hfq interacts specifically with tRNAs. *RNA*, 14(3), pp.514–523.

Lee, B., Flynn, R.A., Kadina, A., Guo, J.K., Kool, E.T. and Chang, H.Y. (2016). Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA*, 23(2), pp.169–174.

Lenz, D.H., Miller, M.B., Zhu, J., Kulkarni, R.V. and Bassler, B.L. (2005). CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae. Molecular Microbiology*, 58(4), pp.1186–1202.

Lee, F.C.Y., Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular Cell*, 69(3), pp.354–369.

Létoffé, S., Delepelaire, P. and Wandersman, C. (1996). Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding. *The EMBO Journal,* 15(21), pp.5804–5811.

Li, W., Ying, X., Lu, Q. and Chen, L. (2012). Predicting sRNAs and Their Targets in Bacteria. *Genomics, Proteomics & Bioinformatics*, 10(5), pp.276–284.

Lim, J.Y., La, H.J., Sheng, H., Forney, L.J. and Hovde, C.J. (2009). Influence of Plasmid pO157 on *Escherichia coli* O157:H7 Sakai Biofilm Formation. *Applied and Environmental Microbiology*, 76(3), pp.963–966.

Lim, J.Y., Yoon, J. and Hovde, C.J. (2010). A Brief Overview of *Escherichia coli* O157:H7 and Its Plasmid O157. *Journal of Microbiology and Biotechnology*, 20(1), pp.5–14.

Little, D.J. and Coombes, B.K. (2018). Molecular basis for CesT recognition of type III secretion effectors in enteropathogenic *Escherichia coli. PLOS Pathogens,* 14(8), p.e1007224.

Liu, T., Zhang, K., Xu, S., Wang, Z., Fu, H., Tian, B., Zheng, X. and Li, W. (2017). Detecting RNA-RNA interactions in *E. coli* using a modified CLASH method. *BMC Genomics*, 18(1).

Lodato, P.B. and Kaper, J.B. (2009). Post-transcriptional processing of the LEE4 operon in enterohaemorrhagic *Escherichia coli. Molecular Microbiology*, 84(5), pp.990–990.

Lodato P.B., Hsieh P.K., Belasco J.G., Kaper J.B. (2012). The ribosome binding site of a mini-ORF protects a T3SS mRNA from degradation by RNase E. *Mol Microbiol*, 86(5), pp.1167–1182.

Lodato P.B., Thuraisamy T., Richards J., Belasco J.G. (2017). Effect of RNase E deficiency on translocon protein synthesis in an RNase E-inducible strain of enterohemorrhagic *Escherichia coli* O157:H7. *FEMS Microbiol Lett*, 364(13).

Logan, S.L., Thomas, J., Yan, J., Baker, R.P., Shields, D.S., Xavier, J.B., Hammer, B.K. and Parthasarathy, R. (2018). The *Vibrio cholerae* type VI secretion system can modulate host intestinal mechanics to displace gut bacterial symbionts. *Proceedings of the National Academy of Sciences*, 115(16), pp.E3779–E3787.

Loh, E., Righetti, F., Eichner, H., Twittenhoff, C. and Narberhaus, F. (2018). RNA Thermometers in Bacterial Pathogens. *Microbiology Spectrum*, 6(2).

Lu, P., Vogel, C., Wang, R., Yao, X. and Marcotte, E.M. (2006). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology*, 25(1), pp.117–124.

Lustri, B.C., Sperandio, V. and Moreira, C.G. (2017). Bacterial Chat: Intestinal Metabolites and Signals in Host-Microbiota-Pathogen Interactions. *Infection and Immunity*, 85(12).

Low, A.S., Holden, N., Rosser, T., Roe, A.J., Constantinidou, C., Hobman, J.L., Smith, D.G.E., Low, J.C. and Gally, D.L. (2006). Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. *Environmental Microbiology*, 8(6), pp.1033–1047.

Luzader, D.H., Willsey, G.G., Wargo, M.J. and Kendall, M.M. (2016). The type three secretion system 2-encoded regulator EtrB modulates enterohemorrhagic *Escherichia coli* virulence gene expression. *Infect. Immun.*, 84, pp.2555–2565

McNeilly, T.N., Mitchell, M.C., Rosser, T., McAteer, S., Low, J.C., Smith, D.G.E., Huntley, J.F., Mahajan, A. and Gally, D.L. (2010). Immunization of cattle with a combination of purified intimin-531, EspA and Tir significantly reduces shedding of *Escherichia coli* O157:H7 following oral challenge. *Vaccine*, 28(5), pp.1422–1428.

Mandell, G.L., Robert Gordon Douglas, John Eugene Bennett, Dolin, R. and Blaser, M.J. (2015). Mandell, Douglas, and Bennett's principles and practice of infectious diseases. Vol. 1 Philadelphia, Pa. Elsevier, Saunders C

Marcus, S.L., Brumell, J.H., Pfeifer, C.G. and Finlay, B. (2000). *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes and Infection*, 2(2), pp.145–156.

Masse, E. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes & Development*, 17(19), pp.2374–2383.

Matthews, B.W. (2005). The structure of *E. coli* β-galactosidase. *Comptes Rendus Biologies*, 328(6), pp.549–556.

McGhie, E.J. (2001). Cooperation between actin-binding proteins of invasive Salmonella: SipA potentiates SipC nucleation and bundling of actin. *The EMBO Journal*, 20(9), pp.2131–2139.

McGinnis, J.L., Dunkle, J.A., Cate, J.H.D. and Weeks, K.M. (2012). The Mechanisms of RNA SHAPE Chemistry. *Journal of the American Chemical Society*, 134(15), pp.6617–6624.

Mellies, J.L., Barron, A.M.S. and Carmona, A.M. (2007). Enteropathogenic and Enterohemorrhagic *Escherichia coli* Virulence Gene Regulation. *Infection and Immunity*, 75(9), pp.4199–4210.

Mellies, J.L., Elliott, S.J., Sperandio, V., Donnenberg, M.S. and Kaper, J.B. (1999). The Per regulon of enteropathogenic *Escherichia coli* : identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Molecular Microbiology*, 33(2), pp.296–306.

Melson, E.M. and Kendall, M.M. (2019). The sRNA DicF integrates oxygen sensing to enhance enterohemorrhagic *Escherichia coli* virulence via distinctive RNA control mechanisms. *Proceedings of the National Academy of Sciences*, 116(28), pp.14210–14215

Melton-Celsa, A.R. (2014). Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiology Spectrum*, 2(4).

Melton-Celsa, A., Mohawk, K., Teel, L., & O'Brien, A. (2011). Pathogenesis of Shiga-toxin producing *Escherichia coli*. In Ricin and Shiga Toxins (pp. 67-103). Springer Berlin Heidelberg.

Ménard, R., Sansonetti, P., Parsot, C. and Vasselon, T. (1994). Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of *S. flexneri*. *Cell*, 79(3), pp.515–525.

Mercante, J., Edwards, A.N., Dubey, A.K., Babitzke, P. and Romeo, T. (2009). Molecular Geometry of CsrA (RsmA) Binding to RNA and Its Implications for Regulated Expression. *Journal of Molecular Biology*, 392(2), pp.511–528.

Merlin, C., McAteer, S., Masters, M. (2002). Tools for characterization of *Escherichia coli* genes of unknown function. *Journal of bacteriology*, *184*(16), 4573-4581.

Meyer, M.M. (2016). The role of mRNA structure in bacterial translational regulation. *Wiley Interdisciplinary Reviews: RNA*, 8(1), p.e1370.

Millar, J.A. and Raghavan, R. (2021). Modulation of Bacterial Fitness and Virulence Through Antisense RNAs. *Frontiers in Cellular and Infection Microbiology*, 10.

Mitobe, J., Morita-Ishihara, T., Ishihama, A. and Watanabe, H. (2009). Involvement of RNA-binding protein Hfq in the osmotic-response regulation of invE gene expression in *Shigella sonnei. BMC Microbiology*, 9(1).

Moody, M.J., Young, R.A., Jones, S.E. and Elliot, M.A. (2013). Comparative analysis of non-coding RNAs in the antibiotic-producing *Streptomyces* bacteria. *BMC Genomics*, 14(1).

Moon, K. and Gottesman, S. (2011). Competition among Hfq-binding small RNAs in *Escherichia coli. Molecular Microbiology*, 82(6), pp.1545–1562.

Moreira, C.G., Russell, R., Mishra, A.A., Narayanan, S., Ritchie, J.M., Waldor, M.K., Curtis, M.M., Winter, S.E., Weinshenker, D. and Sperandio, V. (2016). Bacterial Adrenergic Sensors Regulate Virulence of Enteric Pathogens in the Gut. *mBio*, 7(3).

Morgan, J.K., Carroll, R.K., Harro, C.M., Vendura, K.W., Shaw, L.N. and Riordan, J.T. (2015). Global regulator of virulence A (GrvA) coordinates expression of discrete pathogenic mechanisms in enterohemorrhagic *Escherichia coli* through interactions with GadW-GadE. *J. Bacteriol*., 198, pp.394–409

Morita, T., Aiba, H. (2011). RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes & Development*, 25(4), pp.294–298.

Morita, T., Maki, K., Aiba, H. (2005). RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes & Development*, 19(18), pp.2176–2186.

Morris, C., Cluet, D. and Ricci, E.P. (2021). Ribosome dynamics and mRNA turnover, a complex relationship under constant cellular scrutiny. *WIREs RNA*, 12(6).

Mortimer, S.A., Trapnell, C., Aviran, S., Pachter, L. and Lucks, J.B. (2012). SHAPE–Seq: High-Throughput RNA Structure Analysis. *Current Protocols in Chemical Biology*, 4(4), pp.275–297.

Mosaei, H. and Zenkin, N. (2020). Inhibition of RNA Polymerase by Rifampicin and Rifamycin-Like Molecules. *EcoSal Plus*, 9(1).

Muffler, A., Fischer, D. and Hengge-Aronis, R. (1996). The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for *rpoS* translation in *Escherichia coli. Genes & Development*, 10(9), pp.1143–1151.

Muppirala, U.K., Honavar, V.G. and Dobbs, D. (2011). Predicting RNA-Protein Interactions Using Only Sequence Information. *BMC Bioinformatics*, 12(1).

Murakami, Y. and Mizuguchi, K. (2014). Homology-based prediction of interactions between proteins using Averaged One-Dependence Estimators. *BMC Bioinformatics*, 15(1).

Murphy, B. P., McCabe, E., Murphy, M., Buckley, J. F., Crowley, D., Fanning, S., & Duffy, G. (2016). Longitudinal Study of Two Irish Dairy Herds: Low Numbers of Shiga Toxin-Producing *Escherichia coli* O157 and O26 Super-Shedders Identified. *Frontiers in microbiology*, 7.

Nakao, H., Watanabe, H., Nakayama, S. and Takeda, T. (1995). *yst* gene expression in *Yersinia enterocolitica* is positively regulated by a chromosomal region that is highly homologous to *Escherichia coli* host factor 1 gene (*hfq*). *Molecular Microbiology*, 18(5), pp.859–865.

Nakano, M., Takahashi, A., Su, Z., Harada, N., Mawatari, K. and Nakaya, Y. (2008). Hfq regulates the expression of the thermostable direct hemolysin gene in *Vibrio parahaemolyticus. BMC Microbiology*, 8(1).

Narberhaus, F. (2010). Translational control of bacterial heat shock and virulence genes by temperature-sensing mRNAs. *RNA Biology*, 7(1), pp.84–89.

Neves, B.C., Mundy, R., Petrovska, L., Dougan, G., Knutton, S. and Frankel, G. (2003). CesD2 of Enteropathogenic *Escherichia coli* Is a Second Chaperone for the Type III Secretion Translocator Protein EspD. *Infection and Immunity*, 71(4), pp.2130–2141.

Neyt, C., Cornelis, G. R. (1999). Role of SycD, the chaperone of the *Yersinia* Yop translocators YopB and YopD. *Mol Microbiol*, 31, 143–156.

Nguyen, Y., Sperandio, V. (2012). Enterohemorrhagic *E. coli* (EHEC) pathogenesis. *Frontiers in cellular and infection microbiology*, 2.

Nicholson, T.L., Brockmeier, S.L., Loving, C.L., Register, K.B., Kehrli, M.E. and Shore, S.M. (2013). The *Bordetella bronchiseptica* Type III Secretion System Is Required for Persistence and Disease Severity but Not Transmission in Swine. *Infection and Immunity*, 82(3), pp.1092–1103.

Nie, M. and Htun, H. (2006). Different modes and potencies of translational repression by sequence-specific RNA–protein interaction at the 5′-UTR. *Nucleic Acids Research*, 34(19), pp.5528–5540.

Nuss, A.M., Heroven, A.K. and Dersch, P. (2017). RNA Regulators: Formidable Modulators of *Yersinia* Virulence. *Trends in Microbiology*, 25(1), pp.19–34.

O'Connell, C. B., Creasey, E. A., Knutton, S., Elliott, S., Crowther, L. J., Luo, W., Albert, M., Kaper, J., Frankel, G., Donnenberg, M. S. (2004). SepL, a protein required for enteropathogenic *Escherichia coli* type III translocation, interacts with secretion component SepD. *Molecular microbiology*, 52(6), 1613-1625.

Oliva, G., Sahr, T., Rolando, M., Knoth, M. and Buchrieser, C. (2017). A Unique cis -Encoded Small Noncoding RNA Is Regulating *Legionella pneumophila* Hfq Expression in a Life Cycle-Dependent Manner. *mBio*, 8(1).

Omisakin F., MacRae M., Ogden I. D., Strachan N. J. (2003). Concentration and prevalence of *Escherichia coli* O157 in cattle feces at slaughter. *Appl. Environ. Microbiol.* 69, 2444–2447

Orfanoudaki, G. and Economou, A. (2014). Proteome-wide Subcellular Topologies of *E. coli* Polypeptides Database (STEPdb). *Molecular & Cellular Proteomics*, 13(12), pp.3674–3687.

Pannuri, A., Yakhnin, H., Vakulskas, C.A., Edwards, A.N., Babitzke, P. and Romeo, T. (2011). Translational Repression of NhaR, a Novel Pathway for Multi-Tier Regulation of Biofilm Circuitry by CsrA. *Journal of Bacteriology*, 194(1), pp.79–89.

Park, S., Prévost, K., Heideman, E.M., Carrier, M.-C., Azam, M.S., Reyer, M.A., Liu, W., Massé, E. and Fei, J. (2021). Dynamic interactions between the RNA chaperone Hfq, small regulatory RNAs, and mRNAs in live bacterial cells. *eLife*, 10.

Parker, A., Cureoglu, S., De Lay, N., Majdalani, N. and Gottesman, S. (2017). Alternative pathways for *Escherichia coli* biofilm formation revealed by sRNA overproduction. *Molecular Microbiology*, 105(2), pp.309–325.

Patterson-Fortin, L.M., Vakulskas, C.A., Yakhnin, H., Babitzke, P., Romeo, T. (2013). Dual Posttranscriptional Regulation via a Cofactor-Responsive mRNA Leader. *Journal of Molecular Biology*, 425(19), pp.3662–3677.

Pavlova, N., Kaloudas, D. and Penchovsky, R. (2019). Riboswitch distribution, structure, and function in bacteria. *Gene*, 708, pp.38–48.

Peng, Y., Curtis, J.E., Fang, X. and Woodson, S.A. (2014). Structural model of an mRNA in complex with the bacterial chaperone Hfq. *Proceedings of the National Academy of Sciences,* 111(48), pp.17134–17139.

Pennington, H. (2014). *E. coli* O157 outbreaks in the United Kingdom: past, present, and future. *Infection and Drug Resistance*, p.211

Pérez-Reytor, D., Plaza, N., Espejo, R.T., Navarrete, P., Bastías, R. and Garcia, K. (2017). Role of Non-coding Regulatory RNA in the Virulence of Human Pathogenic Vibrios. *Frontiers in Microbiology*, 7.

Picard, F., Loubière, P., Girbal, L. and Cocaign-Bousquet, M. (2013). The significance of translation regulation in the stress response. *BMC Genomics*, 14, p.588.

Portaliou, A.G., Tsolis, K.C., Loos, M.S., Balabanidou, V., Rayo, J., Tsirigotaki, A., Crepin, V.F., Frankel, G., Kalodimos, C.G., Karamanou, S. and Economou, A. (2017). Hierarchical protein targeting and secretion is controlled by an affinity switch in the type III secretion system of enteropathogenic *Escherichia coli. The EMBO Journal*, 36(23), pp.3517–3531.

Prevost, K., Desnoyers, G., Jacques, J.F., Lavoie, F. and Masse, E. (2011). Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes & Development*, 25(4), pp.385–396.

Pruimboom-Brees, I.M., Morgan, T.W., Ackermann, M.R., Nystrom, E.D., Samuel, J.E., Cornick, N.A. and Moon, H.W. (2000). Cattle lack vascular receptors for *Escherichia coli* O157:H7 Shiga toxins. *Proceedings of the National Academy of Sciences*, 97(19), pp.10325–10329.

Public Health England (2020). *Escherichia coli* (*E. coli*) O157: annual totals. Available at: https://www.gov.uk/government/publications/escherichia-coli-e-coli-o157-annual-totals.

Puhar, A. and Sansonetti, P.J. (2014). Type III secretion system. *Current Biology*, 24(17), pp.R784–R791.

Quah, S.R. (2017). International encyclopedia of public health : volume 6, PRE-SOC. Oxford ; Waltham, Ma: Academic Press.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, 5(10), pp.877–879.

Ramanathan, M., Porter, D.F. and Khavari, P.A. (2019). Methods to study RNA-protein interactions. *Nature methods*, 16(3), pp.225–234.

Ren, B., Shen, H., Lu, Z.J., Liu, H. and Xu, Y. (2014). The phzA2-G2 Transcript Exhibits Direct RsmA-Mediated Activation in *Pseudomonas aeruginosa* M18. *PLoS ONE*, 9(2), p.e89653.

Repoila, F. and Darfeuille, F. (2009). Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biology of the Cell*, 101(2), pp.117–131.

Rice, G. M., Leonard, C. W., & Weeks, K. M. (2014). RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *Rna*, *20*(6), 846-854.

Richard, A.L., Withey, J.H., Beyhan, S., Yildiz, F. and DiRita, V.J. (2010). The *Vibrio cholerae* virulence regulatory cascade controls glucose uptake through activation of TarA, a small regulatory RNA. *Molecular Microbiology*, 78(5), pp.1171–1181.

Rimpiläinen, M., Forsberg, A. and Wolf-Watz, H. (1992). A novel protein, LcrQ, involved in the low-calcium response of *Yersinia pseudotuberculosis* shows extensive homology to YopH. *Journal of Bacteriology*, 174(10), pp.3355–3363.

Rinaldi, A.J., Lund, P.E., Blanco, M.R. and Walter, N.G. (2016). The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nature Communications*, 7(1).

Roe, A. J., Hoey, D. E., Gally, D. L. (2003). Regulation, secretion and activity of type III-secreted proteins of enterohaemorrhagic *Escherichia coli* O157. *Biochem Soc Trans*, 31, 98-103.

Roe, A. J., Tysall, L., Dransfield, T., Wang, D., Fraser-Pitt, D., Mahajan, A., Constandinou, C., Inglis, N., Downing, A., Talbot, R., Smith, D. G., Gally, D. (2007). Analysis of the expression, regulation and export of NleA–E in *Escherichia coli* O157: H7. *Microbiology*, 153(5), 1350-1360.

Romeo, T., Vakulskas, C.A. and Babitzke, P. (2013). Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environmental Microbiology*, 15(2), pp.313–324.

Ruano-Gallego, D., Álvarez, B. and Fernández, L.Á. (2015). Engineering the Controlled Assembly of Filamentous Injectisomes in *E. coli* K-12 for Protein Translocation into Mammalian Cells. *ACS Synthetic Biology*, 4(9), pp.1030–1041.

Ruscher, K., Reuter, M., Kupper, D., Trendelenburg, G., Dirnagl, U. and Meisel, A. (2000). A fluorescence based non-radioactive electrophoretic mobility shift assay. *Journal of Biotechnology*, 78(2), pp.163–170.

Russell, R.M., Sharp, F.C., Rasko, D.A. and Sperandio, V. (2007). QseA and GrlR/GrlA Regulation of the Locus of Enterocyte Effacement Genes in Enterohemorrhagic *Escherichia coli. Journal of Bacteriology*, 189(14), pp.5387–5392.

Saberi, F., Kamali, M., Najafi, A., Yazdanparast, A. and Moghaddam, M.M. (2016). Natural antisense RNAs as mRNA regulatory elements in bacteria: a review on function and applications. *Cellular & Molecular Biology Letters*, 21(1).

Sabnis, N.A., Yang, H. and Romeo, T. (1995). Pleiotropic Regulation of Central Carbohydrate Metabolism in *Escherichia coli* via the Gene *csrA*. *Journal of Biological Chemistry*, 270(49), pp.29096–29104.

Sanchez, K.K., Chen, G.Y., Schieber, A.M.P., Redford, S.E., Shokhirev, M.N., Leblanc, M. et al. (2018). Cooperative metabolic adaptations in the host can favor asymptomatic infection and select for attenuated virulence in an enteric pathogen. *Cell*, 175, pp.1–13

Sánchez-San Martín, C., Bustamante, V.H., Calva, E. and Puente, J.L. (2001). Transcriptional Regulation of the *orf19* Gene and the *tir-cesT-eae* Operon of Enteropathogenic *Escherichia coli*. *Journal of Bacteriology*, 183(9), pp.2823–2833.

Sandkvist, M., Michel, L.O., Hough, L.P., Morales, V.M., Bagdasarian, M., Koomey, M., DiRita, V.J. and Bagdasarian, M. (1997). General secretion pathway (*eps*) genes required for toxin secretion and outer membrane biogenesis in *Vibrio cholerae*. *Journal of bacteriology*, 179(22), pp.6994–7003.

Sandvig, K., Bergan, J., Dyve, A.-B., Skotland, T. and Torgersen, M.L. (2010). Endocytosis and retrograde transport of Shiga toxin. *Toxicon*, 56(7), pp.1181–1185.

Santiago-Frangos, A. and Woodson, S.A. (2018). Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdisciplinary Reviews: RNA*, 9(4), p.e1475.

Sarpong, D.D. and Murphy, E.R. (2021). RNA Regulated Toxin-Antitoxin Systems in Pathogenic Bacteria. *Frontiers in Cellular and Infection Microbiology*, 11.

Schiano, C.A. and Lathem, W.W. (2012). Post-Transcriptional Regulation of Gene Expression in *Yersinia* Species. *Frontiers in Cellular and Infection Microbiology*, 2.

Schmidt, C.E., Shringi, S. and Besser, T.E. (2016). Protozoan Predation of *Escherichia coli* O157:H7 Is Unaffected by the Carriage of Shiga Toxin-Encoding Bacteriophages. *PLOS ONE*, 11(1), p.e0147270.

Schroeder, R., Barta, A. and Semrad, K. (2004). Strategies for RNA folding and assembly. *Nature Reviews Molecular Cell Biology*, 5(11), pp.908–919.

Schu, D.J., Zhang, A., Gottesman, S. and Storz, G. (2015). Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *The EMBO Journal*, 34(20), pp.2557–2573.

Schüller, S. (2011). Shiga toxin interaction with human intestinal epithelium. *Toxins*, 3(6), 626-639.

Sgro, G.G., Oka, G.U., Souza, D.P., Cenens, W., Bayer-Santos, E., Matsuyama, B.Y., Bueno, N.F., dos Santos, T.R., Alvarez-Martinez, C.E., Salinas, R.K. and Farah, C.S. (2019). Bacteria-Killing Type IV Secretion Systems. *Frontiers in Microbiology*, 10, 1078.

Shakhnovich, E.A., Davis, B.M. and Waldor, M.K. (2009). Hfq negatively regulates type III secretion in EHEC and several other pathogens. *Molecular Microbiology*, 74(2), pp.347–363.

Sharma, A.K. and Payne, S.M. (2006). Induction of expression of *hfq* by DksA is essential for *Shigella flexneri* virulence. *Molecular Microbiology*, 62(5), pp.1498–1498.

Sharp, F. C., Sperandio, V. (2007). QseA directly activates transcription of LEE1 in enterohemorrhagic *Escherichia coli. Infection and immunity*, 75(5), 2432-2440.

Sherlock, M.E., Sudarsan, N., Stav, S. and Breaker, R.R. (2018). Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *eLife*, 7.

Shin, M. (2017). The mechanism underlying Ler-mediated alleviation of gene repression by H-NS. *Biochem. Biophys. Res. Commun*. 483, pp.392–396

Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M. and Panda, A.K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial Cell Factories*, 14(1).

Sledjeski, D.D., Whitman, C. and Zhang, A. (2001). Hfq Is Necessary for Regulation by the Untranslated RNA DsrA. *Journal of Bacteriology*, 183(6), pp.1997–2005.

Smirnov, A., Förstner, K.U., Holmqvist, E., Otto, A., Günster, R., Becher, D., Reinhardt, R. and Vogel, J. (2016). Grad-seq guides the discovery of ProQ as a major small RNA-binding protein*. Proceedings of the National Academy of Sciences*, 113(41), pp.11591–11596.

Song, T., Mika, F., Lindmark, B., Liu, Z., Schild, S., Bishop, A., Zhu, J., Camilli, A., Johansson, J., Vogel, J. and Wai, S.N. (2008). A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Molecular Microbiology*, 70(1), pp.100–111.

Sonnleitner, E., Hagens, S., Rosenau, F., Wilhelm, S., Habel, A., Jäger, K.-E. and Bläsi, U. (2003). Reduced virulence of a *hfq* mutant of *Pseudomonas aeruginosa* O1. *Microbial Pathogenesis*, 35(5), pp.217–228.

Sonnleitner, E., Schuster, M., Sorger-Domenigg, T., Greenberg, E.P. and Bläsi, U. (2006). Hfq-dependent alterations of the transcriptome profile and effects on quorum sensing in *Pseudomonas aeruginosa. Molecular Microbiology*, 59(5), pp.1542–1558.

Sonnleitner, E., Sorger-Domenigg, T., Madej, M.J., Findeiss, S., Hackermüller, J., Hüttenhofer, A., Stadler, P.F., Bläsi, U. and Moll, I. (2008). Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools. *Microbiology*, 154(10), pp.3175–3187.

Soper, T.J. and Woodson, S.A. (2008). The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA*, 14(9), pp.1907–1917.

Soper, T., Mandin, P., Majdalani, N., Gottesman, S. and Woodson, S.A. (2010). Positive regulation by small RNAs and the role of Hfq. *Proceedings of the National Academy of Sciences*, 107(21), pp.9602–9607.

Soper, T.J., Doxzen, K. and Woodson, S.A. (2011). Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *RNA*, 17(8), pp.1544–1550.

Sorger-Domenigg, T., Sonnleitner, E., Kaberdin, V.R. and Bläsi, U. (2007). Distinct and overlapping binding sites of *Pseudomonas aeruginosa* Hfq and RsmA proteins on the non-coding RNA RsmY. *Biochemical and Biophysical Research Communications*, 352(3), pp.769–773.

Soysal, N., Mariani-Kurkdjian, P., Smail, Y., Liguori, S., Gouali, M., Loukiadis, E., Fach, P., Bruyand, M., Blanco, J., Bidet, P., Bonacorsi, S. (2016). Enterohemorrhagic *Escherichia coli* hybrid pathotype O80: H2 as a new therapeutic challenge. *Emerging infectious diseases*, 22(9), 1604.

Speare, L., Cecere, A.G., Guckes, K. R., Smith, S., Wollenberg, M. S., Mandel, M.J., Miyashiro, T. and Septer, A.N. (2018). Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proceedings of the National Academy of Sciences*, 115(36), pp.E8528–E8537.

Spears, K. J., Roe, A. J., Gally, D. L. (2006). A comparison of enteropathogenic and Enterohaemorrhagic *Escherichia coli* pathogenesis. *FEMS Microbiology Letters*, 255 (2).

Sperandio, V., Torres, A.G., Giron, J.A. and Kaper, J.B. (2001). Quorum Sensing Is a Global Regulatory Mechanism in Enterohemorrhagic *Escherichia coli* O157:H7. *Journal of Bacteriology*, 183(17), pp.5187–5197.

Steinberg, K. and Levin, B.R. (2007). Grazing protozoa and the evolution of the *Escherichia coli* O157:H7 Shiga toxin-encoding prophage. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621), pp.1921–1929.

Su, M. S., Kao, H. C., Lin, C. N., and Syu, W. J. (2008). Gene *l0017* encodes a second chaperone for EspA of enterohaemorrhagic *Escherichia coli* O157: H7. *Microbiology*, 154(Pt 4), 1094–1103.

Sudarsan, N., Hammond, M.C., Block, K.F., WelzR., Barrick, J.E., Roth, A. and Breaker, R.R. (2006). Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. *Science*, 314(5797), pp.300–304.

Sudo, N., Lee, K., Sekine, Y., Ohnishi, M. and Iyoda, S. (2021). RNA-binding protein Hfq downregulates locus of enterocyte effacement-encoded regulators independent of small regulatory RNA in enterohemorrhagic *Escherichia coli. Molecular Microbiology.*

Sulthana, S., Basturea, G.N. and Deutscher, M.P. (2016). Elucidation of pathways of ribosomal RNA degradation: an essential role for RNase E. RNA, 22(8), pp.1163–1171.

Sterzenbach, T., Nguyen, K.T., Nuccio, S.-P., Winter, M.G., Vakulskas, C.A., Clegg, S., Romeo, T. and Bäumler, A.J. (2013). A novel CsrA titration mechanism regulates fimbrial gene expression in *Salmonella typhimurium*. *The EMBO Journal*, 32(21), pp.2872–2883.

Stoddard, C.D., Gilbert, S.D., Batey, R.T. (2008). Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA*, 14(4), pp.675–684.

Straube, J., Huang, B.E. and Cao, K.-A.L. (2017). DynOmics to identify delays and co-expression patterns across time course experiments. *Scientific Reports*, 7(1).

Svensson, S.L. and Sharma, C.M. (2021). RNase III-mediated processing of a trans-acting bacterial sRNA and its cis-encoded antagonist. *eLife*, 10.

Symmons, M.F., Bokma, E., Koronakis, E., Hughes, C. and Koronakis, V. (2009). The assembled structure of a complete tripartite bacterial multidrug efflux pump. *Proceedings of the National Academy of Sciences*, 106(17), pp.7173–7178.

Takaya, A., Takeda, H., Tashiro, S., Kawashima, H. and Yamamoto, T. (2019). Chaperone-mediated secretion switching from early to middle substrates in the type III secretion system encoded by *Salmonella* pathogenicity island 2. *Journal of Biological Chemistry*, 294(10), pp.3783–3793.

Tarr, P.I., Gordon, C.A. and Chandler, W.L. (2005). Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *The Lancet*, 365(9464), pp.1073–1086.

Teplova, M., Malinina, L., Darnell, Jennifer C., Song, J., Lu, M., Abagyan, R., Musunuru, K., Teplov, A., Burley, Stephen K., Darnell, Robert B. and Patel, Dinshaw J. (2011). Protein-RNA and Protein-Protein Recognition by Dual KH1/2 Domains of the Neuronal Splicing Factor Nova-1. *Structure*, 19(7), pp.930–944.

Tobe T., Yen H., Takahashi H., Kagayama Y., Ogasawara N., Oshima T., (2014). Antisense transcription regulates the expression of the enterohemorrhagic *Escherichia coli* virulence regulatory gene *ler* in response to the intracellular iron concentration. *PLoS One*, 9(7):e101582.

Tomalka, A. G., Stopford, C. M., Lee, P. C., Rietsch, A. (2012). A translocator-specific export signal establishes the translocator–effector secretion hierarchy that is important for type III secretion system function. *Molecular microbiology*, 86(6), 1464-1481.

Totsika, M. (2016). Benefits and challenges of antivirulence antimicrobials at the dawn of the post-antibiotic era. *Drug Delivery Letters*, 6(1), 30-37.

Tsui, H.-C.T., Leung, H.-C.E. and Winkler, M.E. (1994). Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. *Molecular Microbiology*, 13(1), pp.35–49.

Tree, J., Granneman, S., McAteer, Sean P., Tollervey, D. and Gally, David L. (2014). Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic *Escherichia coli*. *Molecular Cell*, 55(2), pp.199–213.

Tree, J.J., Roe, A.J., Flockhart, A., McAteer, S.P., Xu, X., Shaw, D., Mahajan, A., Beatson, S.A., Best, A., Lotz, S., Woodward, M.J., La Ragione, R., Murphy, K.C., Leong, J.M. and Gally, D.L. (2011). Transcriptional regulators of the GAD acid stress island are carried by effector protein-encoding prophages and indirectly control type III secretion in enterohemorrhagic *Escherichia coli* O157:H7. *Molecular Microbiology*, 80(5), pp.1349–1365.

Tree, J.J., Wolfson, E.B., Wang, D., Roe, A.J. and Gally, D.L. (2009). Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli. Trends in microbiology*, 17(8), pp.361-370.

Tucker, S. C., Galan, J. E. (2000). Complex function for SicA, a *Salmonella enterica* serovar Typhimurium type III secretion associated chaperone. *J Bacteriol*, 182, 2262–2268.

Turner, N.C.A., Connolly, J.P.R. and Roe, A.J. (2018). Control freaks—signals and cues governing the regulation of virulence in attaching and effacing pathogens. *Biochemical Society Transactions*, 47(1), pp.229–238.

Udekwu, K.I. (2005). Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes & Development*, 19(19), pp.2355–2366.

Valentin-Hansen, P., Eriksen, M. and Udesen, C. (2004). The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Molecular Microbiology*, 51(6), pp.1525–1533.

Valilis, E., Ramsey, A., Sidiq, S. and DuPont, H.L. (2018). Non-O157 Shiga toxin-producing *Escherichia coli*—A poorly appreciated enteric pathogen: Systematic review. *International Journal of Infectious Diseases*, 76, pp.82–87.

Van Assche, E., Van Puyvelde, S., Vanderleyden, J. and Steenackers, H.P. (2015). RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Frontiers in Microbiology*, 6.

Vytvytska, O., Moll, I., Kaberdin, V.R., von Gabain, A. and Bläsi, U. (2000). Hfq (HF1) stimulates *ompA* mRNA decay by interfering with ribosome binding. *Genes & Development*, 14(9), pp.1109–1118.

Wagner, S., Königsmaier, L., Lara-Tejero, M., Lefebre, M., Marlovits, T.C. and Galán, J.E. (2010). Organization and coordinated assembly of the type III secretion export apparatus. *Proceedings of the National Academy of Sciences*, 107(41), pp.17745–17750.

Waldminghaus, T., Heidrich, N., Brantl, S. and Narberhaus, F. (2007). FourU: a novel type of RNA thermometer in *Salmonella. Molecular Microbiology*, 65(2), pp.413–424.

Wan, B., Zhang, Q., Tao, J., Zhou, A., Yao, Y.F. and Ni, J. (2016). Global transcriptional regulation by H-NS and its biological influence on the virulence of enterohemorrhagic *Escherichia coli. Gene*, 588, pp.115–123

Wang D. (2011). A Study of Expression and Function of SepL, a Regulator of Type 3 Secretion in Enterohaemorrhagic *Escherichia coli* O157 (Doctoral dissertation). Retrieved from Royal (Dick) School of Veterinary Studies thesis and dissertation collection. URI: http://hdl.handle.net/1842/5560

Wang, C., Han, B., Zhou, R. and Zhuang, X. (2016). Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell*, 165(4), pp.990–1001.

Wang, D., Roe, A. J., McAteer, S., Shipston, M., Gally, D. L. (2008). Hierarchal type III secretion of translocators and effectors from *Escherichia coli* O157: H7

requires the carboxy terminus of SepL that binds to Tir. *Molecular microbiology*, 69(6), 1499-1512.

Wang, T., Si, M., Song, Y., Zhu, W., Gao, F., Wang, Y., Zhang, L., Zhang, W., Wei, G., Luo, Z.-Q. and Shen, X. (2015). Type VI Secretion System Transports Zn2+ to Combat Multiple Stresses and Host Immunity. *PLOS Pathogens*, 11(7), p.e1005020.

Wang, G., Zhao, T. and Doyle, M.P. (1996). Fate of enterohemorrhagic *Escherichia coli* O157:H7 in bovine feces. *Applied and Environmental Microbiology* 62, 2567–2570.

Waters, S.A., McAteer, S.P., Kudla, G., Pang, I., Deshpande, N.P., Amos, T.G., Leong, K.W., Wilkins, M.R., Strugnell, R., Gally, D.L., Tollervey, D. and Tree, J.J. (2016). Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *The EMBO Journal*, 36(3), pp.374–387.

Waters, L.S. and Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, 136(4), pp.615–628.

Watters, K.E., Yu, A.M., Strobel, E.J., Settle, A.H. and Lucks, J.B. (2016). Characterizing RNA structures in vitro and in vivo with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods*, 103, pp.34–48.

Watts, J. M., Dang K. K., Gorelick R. J., Leonard C. W., Bess J. W., Swanstrom R., Burch C. L., Weeks K. M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, *460*, 711–716.

Wawszczyk A., B. (2017). Building a type III secretion system: Does translation of the sepL mRNA require the basal apparatus? (Masters by Research dissertation). Royal (Dick) School of Veterinary Studies thesis and dissertation collection.

Weiner, M.P., Costa, G.L., Schoettlin, W., Cline, J., Mathur, E. and Bauer, J.C. (1994). Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene*, 151(1-2), pp.119–123.

Welinder-Olsson, C., Badenfors, M., Cheasty, T., Kjellin, E. and Kaijser, B. (2002). Genetic Profiling of Enterohemorrhagic *Escherichia coli* Strains in Relation to Clonality and Clinical Signs of Infection. *Journal of Clinical Microbiology*, 40(3), pp.959–964.

Welinder-Olsson, C., & Kaijser, B. (2005). Enterohemorrhagic *Escherichia coli* (EHEC). *Scandinavian journal of infectious diseases*, 37(6-7), 405-416.

Wilkinson K. A., Gorelick R. J., Vasa S. M., Guex N., Rein A., Mathews D. H., Giddings M. C., Weeks K. M. (2008). High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol*, *6*: e96.

Wilson, K.A., Holland, D.J. and Wetmore, S.D. (2016). Topology of RNA–protein nucleobase–amino acid π–π interactions and comparison to analogous DNA–protein π–π contacts. *RNA*, 22(5), pp.696–708.

Xu, X., McAteer, S.P., Tree, J.J., Shaw, D.J., Wolfson, E.B.K., Beatson, S.A., Roe, A.J., Allison, L.J., Chase-Topping, M.E., Mahajan, A., Tozzoli, R., Woolhouse, M.E.J., Morabito, S. and Gally, D.L. (2012). Lysogeny with Shiga Toxin 2-Encoding Bacteriophages Represses Type III Secretion in Enterohemorrhagic *Escherichia coli. PLoS Pathogens*, 8(5), p.e1002672.

Yakhnin, H., Baker, C.S., Berezin, I., Evangelista, M.A., Rassin, A., Romeo, T. and Babitzke, P. (2011). CsrA Represses Translation of *sdiA* , Which Encodes the N -Acylhomoserine- l -Lactone Receptor of *Escherichia coli*, by Binding Exclusively within the Coding Region of *sdiA* mRNA. *Journal of Bacteriology*, 193(22), pp.6162–6170.

Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T. and Babitzke, P. (2013). CsrA activates *flhDC* expression by protecting

flhDC mRNA from RNase E-mediated cleavage. *Molecular Microbiology*, 87(4), pp.851–866.

Ye, F., Yang, F., Yu, R., Lin, X., Qi, J., Chen, Z., Cao, Y., Wei, Y., Gao, G.F. and Lu, G. (2018). Molecular basis of binding between the global post-transcriptional regulator CsrA and the T3SS chaperone CesT. *Nature Communications*, 9(1).

Yerushalmi, G., Litvak, Y., Gur-Arie, L. and Rosenshine, I. (2014). Dynamics of Expression and Maturation of the Type III Secretion System of Enteropathogenic *Escherichia coli*. *Journal of Bacteriology*, 196(15), pp.2798–2806.

Yi, J., Xiao, S.B., Zeng, Z.X., Lu, J.F., Liu, L.Y., Laghari, Z.A., Nie, P., Yu, H.B. and Xie, H.X. (2016). EseE of *Edwardsiella tarda* Augments Secretion of Translocon Protein EseC and Expression of the *escC* - *eseE* Operon. *Infection and Immunity*, 84(8), pp.2336–2344.

Yin, X., Zhu, J., Feng, Y., Chambers, J.R., Gong, J. and Gyles, C.L. (2011). Differential Gene Expression and Adherence of *Escherichia coli* O157:H7 In Vitro and in Ligated Pig Intestines. *PLoS ONE*, 6(2), p.e17424.

Yoon, H., Ansong, C., Adkins, J.N. and Heffron, F. (2011). Discovery of Salmonella Virulence Factors Translocated via Outer Membrane Vesicles to Murine Macrophages. *Infection and Immunity*, 79(6), pp.2182–2192.

Younis, R., Bingle, L., E., Rollauer, S., Munera, D., Busby, S., J., Johnson, S., Dean, J., E., Lea, S., M., Frankel, G., Pallen, M. J. (2010). SepL resembles an aberrant effector in binding to a class 1 type III secretion chaperone and carrying an N-terminal secretion signal. *Journal of Bacteriology*, 192, pp.6093–6098

Zhang, C. and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Biotechnology*, 29(7), pp.607–614.