Bioengineering of *Escherichia coli* Flagellar Type III Secretion system (FT3SS) for improved efficiency of protein export

By:

Nitin Sampatrao Kamble

A thesis submitted in partial fulfilment of the requirement for the degree of

Doctor of Philosophy

Supervisors:

Dr Graham Stafford

Dr Jagroop Pandhal

Prof. Phillip Wright

School of Clinical Dentistry

School of Chemical and Biological Engineering (ChELSI)

The University of Sheffield, (U.K.)

December 2018

(This page is left blank intentionally)

Summary

In order to produce high quality and yield of recombinant proteins one goal of the biotechnology industry is to reduce production costs as a large proportion (up to 80 %) goes into downstream processing. These processing costs could be reduced by direct export of protein product in to the medium. This in turn could be achieved by re-engineering of bacterial flagella type III secretion system (FT3SS) by converting it into a high efficiency protein export nano-machine. The flagella are comprised of motor and filament structures that provides the cells with motility but is also assembled de novo *via* protein exports of ~1700 subunit min^{-1} through the export apparatus. In terms of industrial biotechnology (IB), the FT3SS may enable a high efficiency, one-step export route from cytoplasm to the extracellular medium where the secretion rate is significantly higher than the range of IB relevant secretion systems.

The main aim of this research project was to fully explore and engineer the FT3SS in order to convert *Escherichia coli* into a 'Super-Secretor' strain, that would export proteins directly in to the medium. This was achieved by using a secretion construct which harboured Cutinase cargo along with an element of native flagellin (untranslated region and late secretion signal) and purifications tags. Following the optimisation of a high throughput assay, the most efficient secretion signal and engineered strain combination was investigated with a systems biology-proteomics approach (iTRAQ) to highlight protein regulation existed during secretion. In parallel, a genomic screening approach was set-up to highlight novel genes to improve protein export. Finally, the genes encoding the highlighted proteins along with potential regulatory small non-coding RNA (snRNA) were further tested in order to investigate whether they would improve protein secretion, however, the engineered Δ*clpX* Δ*arcZ* strain was obtained which significantly improved the protein export *via* the modified FT3SS.

These approaches comprehensively allowed secretion yields to be improved significantly, alongside the wealth of knowledge gained in this thesis is completely explored it suggest that the FT3SS platform might outcompete the current commercial protein secretion systems utilised in the IB industry.

Acknowledgement

I am grateful to my lovely Akka-Dada & brothers for their unconditional love and trust in me. Actually, it is their hard work and the sacrifices they made that I could reach this great height today! I love you guys; I'm indebted for life!

My special thanks and appreciation go to the very kind being, the most intelligent, my supervisor Dr Graham Stafford for his continuous support, encouragement and motivation, right from rescuing me from Chesterfield the day I landed in U.K. till the day of completion of this work. I have been greatly benefited from his advice, guidance, expertise and unfaltering confidence in me, without which this work would not have been possible. I have been privileged to seek continuous guidance and support throughout my project and initial struggling time from him. It was an absolute honour working with him; I cannot thank him enough in words for developing an expert Synthetic Biologist in me. I would also like to thank my supporting supervisors Prof Phillip Wright (Newcastle University) and Dr Jags Pandhal (Chemical and Biological Engineering, UoS) for their academic support.

My sincere thanks go to Technical staff, Brenka McCabe and Jason Heath, who work tirelessly behind the scenes and without whom the lab would be absolute chaos. Thank you both for providing such a friendly and sound environment. I sincerely apologise for the trouble; I have had given you both during Lab-work and even out of hours too! I will miss you guys and the British sarcasm as well! Further, I would like to extend many thanks to my fellow, current and past colleagues in Dental School and CBE. My sincere thanks are due to Magda-Nied, Areeg, Dr Abigail Pinnock, Dr Ali Khurram, Dr Caroline Evans, Charlotte Green, Natalya, Marianne, Lakhan, Khalisi, Dr Raghu, Yuga, Regin, Charu, Rahul–Priyanka, Dr Mohammad Al-Zubidi, Andrew Frey, Kate Naylor, Emmanuel Amababe, Sadhavi, Zaki, Zul, Ashley, Khoa, Joy Mukherjee, Katherine, Beth, & Anita. Thank you, folks! I'll be happy to host you guys in India!

Finally, I am thankful to the High Commission of India in London (UK), Government of India for sponsoring my Research and The University of Sheffield for providing the amazing working environment. Much Love, will miss you all!

List of Abbreviations

5 ' UTR / 3 ' UTR	5' Untranslated Region / 3' Untranslated Region
AA or aa	Amino Acids (# refer appendix 5)
Abs.	Absorbance
Amb.	Ammonium Bicarbonate
Amp^R	Ampicillin resistance
APS	Ammonium Per Sulphate
ATP	Adenosine Tri Phosphate
AU	Arbitrary Units
Bp	Nucleotides
Bp or bp	Base Pairs
BSA	Bovine Serum Albumin
C – terminus	Carboxyl terminus
Cmp^R	Chloramphenicol resistance
Conc.	Concentration
D / W	Distilled water / deionised water
$_dH_20$	Double Distilled Water
DMSO	Dimethyl sulfoxide
DNA	De–oxy Ribose Nucleic Acid
$_d$NTPs	Deoxy Ribonucleotides
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylene diamine Tetra Acetic Acid
Et Br	Ethidium Bromide
FA	Formic Acid
FDR	Flase Discovery Rate
Fig.	Figure
Flag Tag	Flag Octapeptide Tag
FSB	Frozen Storage Buffer
fTTSS (fT3SS)	Flagellar Type III secretion system
$Gent^R$	Gentamycin resistance
GRM	Gate recognition Moti
GMP	Good manufacturing practices
hGH	Human Growth hormone
HOAC	Acetic Acid
HRP	Horse redish peroxidase
IPTG	Isopropyl β–D–1thiogalactopyranoside
iTRAQ	Isobaric Tag for Relative and Absloute Quantification
Kan^R	Kanamycin resistance
kDa	Kilo Dalton
KO	Knockout mutagenesis
KOAC	Potassium Acetate
LB	Luria – Bertani Broth / media
LFQ	Label-Free Quantification
Lib.	Library
MCS	Multiple Cloning sites
MeCN (ACN)	Acetonitrile
ml	Mili Litre
mM	Mili molar
MMTS	S – Methyl methanethiosilfonate
mRNA	Messenger RNA
MUB	Methyllumberyferryl Butyrate

N – terminus	Amide terminus
N/F water	Nuclease Free water
NaCl	Sodium Chloride
nm	Nano meter
O / n	Overnight
OD_{600}	Optical Density at 600 nm
ORF	Open Reading Frame
PBS	Phosphate Buffer Saline
PCA	Principal component analysis
PCB	Phosphate Citrate Buffer
PCR	Polymerase Chain Reaction
pJExpress	pJEXpress 404 plasmid
pNK	pNK_{15}TcLib plasmid
PTM	Post Translational Modification
Q-TOF	Quadrupole Time of Flight
R. E.	Restriction Enzymes / Restriction Endonucleases
RNA	Ribose Nucleic Acid
RPM / rpm	Revolution per minute
SCX	Strong cation exchange chromatography
SD	Standard Deviation
SDM	Site Directed Mutagenesis
SDS	Sodium Dodecyl Sulphate
SDS–PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEM	Standard Error of the mean
SN	Supernatant
SnRNA	small non–coding RNA
SS or ss	Signal Sequence (Signature Sequence)
Stop Buffer	Sodium phosphate Buffer (pH 10.5)
T3SS	Type III secretion system
T9SS or t9ss	Type 9 Secretion system
TAE Buffer	Tris Acetate EDTA buffer
TCA	Trichloro acetic acid
TCEP	Tris (2 – Carboxyethyl) phosphine hydrochloride
TEAB	Triethylammonium bicarbonate buffer
TEMED	Tetramethylethylenediamine
Tet^R	Tetracycline resistance
TEV	Tobacco Etch virus
TFA	Trifluoro acetic acid
TFB	Transformation Buffer
Trp	Tryptophan
TTSC (T3SC)	Type III secretion chaperones
U.V. Light	Ultraviolet Light
uHPLC	Ultra High-performance Liquid Chromatography
WT	Wild type
X – gal	5–bromo–4–chloro – 3–indolyl–β–D galactopyranoside
X g	Gravitational constant
µM	Micro Molar

Table of contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction:

One persistent goal of the biotechnology industry is the production of recombinant biotherapeutics and heterologous proteins to meet the global biological demand for protein biotherapeutics. In general, these high economic value products are often costly and difficult to produce due to GMP rules alongside the strict requirements for the safety of use in humans. Significant efforts have been taken to improve the yield of these products and reduce their production costs, for example a large amount of the total cost goes into downstream processing for protein purification to separate the protein from undesired cellular contaminants (Pina *et al.,* 2014). Downstream processing costs could be reduced by achieving the secretion of recombinant proteins from the bacterial cell cytoplasm directly into the medium. This objective has been shown to increase the quality and yield of the recombinant proteins (Georgiou and Segatori, 2005; Ni and Chen, 2009; Schlegel *et al.*, 2013).

An appealing model for increasing quality and yield of recombinant proteins, would be a high-throughput, one-step direct protein export process where protein expressed in the bacterial cell cytoplasm is exported to the medium, so that it can be easily separated and harvested from the medium using simple purification techniques such as ultra-filtration or centrifugation. This could open the possibility of continuous protein secretion by a stable cell culture, which could be utilised for more than one generation. Protein secretion using Gram-negative bacteria is a major biotechnology challenge, as it is low yielding. This aim could be achieved by re-engineering of existing biological protein secretion systems. Therefore, this thesis will focus on the re-engineering of the common bacterium employed in the biotechnology industry, *Escherichia coli (E. coli),* to directly secrete proteins to extracellular medium *via* the Flagellar Type III secretion system (FT3SS). While a prototype for this system has been established by the team in Sheffield and by others, a combination of fundamental microbiology, molecular biology, recombineering, proteomics, systems biology and synthetic biology methodologies were utilised in this thesis to enhance protein secretion *via* the modified FT3SS.

1.2 Synthetic Biology

Synthetic biology is a relatively modern stream of bioscience that deals with designing and assembling novel biological systems and re-design of existing biological systems for human welfare and benefit (Nandagopal and Elowitz, 2011). It also aims to modify the approaches previously utilised in existing biosciences such as genetic engineering. It is synchronised with biology at the interface of engineering. In an ideal world, the utilisation of synthetic biology approaches would eliminate the uncertainty and lack of transferability of the engineering of biological processes, as they are unavoidably cumbersome and complex to manage (Heinemann and Panke, 2006; Mukherji and Oudenaarden, 2009). Characterisation and construction of biological system are tedious, unreliable and spontaneous variation may occur during the process as evolution is continuous (Endy, 2005; Serrano, 2007).

Therefore, in order to overcome these challenges, the conceptualised principles routed in engineering could provide a framework for synthetic biology work *viz.* standardisation, decoupling and abstraction of this technology.

1.3 Industrial Biotechnology (IB)

Although the bioscience industry has only existed for a quarter of a century in a true sense, living organisms have been utilised for thousands of years to advance human life quality (Bollinger and Stover, 1999). Industrial biotechnology (IB) is one of the fastest growing contemporary industries (Kotzsch *et al.*, 2011; Baeshen *et al.*, 2015). It uses biological resources to manufacture materials, chemicals and energy. Currently, IB companies in UK employ ~14,000 people that contribute upto £1.2 billion in Gross Value added to the economy. It is also estimated that value of global IB market might reach £ 360 billion by the year 2025. While in 2003, U.S. Biotechnology companies employed, 200,000 people and generated $39.2 billion in revenue (Hevesi and Bleiwas, 2005; Chambers *et al.*, 2015). Biomanufacturing is based on the utilisation of biological resources such as Bacteria, Algae and Yeast, where eight out of ten most prescribed drugs are bio-manufactured (Walsh, 2006). As shown in Table 1.1, these include a range of biologicals such as hormones (Insulin, Human growth hormone), antibody fragments, interferon for cancer treatment, colony-

stimulating factor and vaccines that have been approved for therapeutic use in human were produced in *E. coli*. The biotech market is also predicted to grow in the near future (Lessard and Walsh, 1999; Walsh, 2006).

Table 1.1: Biotherapeutics approved for human use with the trade name.

Product	Trade name	Use
Blood clotting factor VIII/IX	Alphanate	Replacement of clotting factor missing in patients with haemophilia A / B
Calcitonin	Calcimar / Miacalcin	Treatment of rickets
Chronic gonadotropin	Profasi, Novarel, Pregnyl, Ovidrel	Treatment of infertility
Erythropoietin (EPO)	Epogen / Procrit	Treatment for anaemia
Factor VIII		Treatment for haemophilia
Hepatitis B vaccine	Rcombinax HB Engerix B	Prevention of viral hepatitis
Human Growth Hormone	Protropin/ Humatrope	Replacement of missing hormone in short stature people
Insulin	Humulin	Treatment of insulin-dependent diabetes
Interferon (alpha and gamma)	Intron A	Possible treatment for cancer and viral infections
Interleukins	Proleukin	Enhancement of action of the immune system
Platelet-derived growth factor	Becaplermin/ Regranex	Stimulation of wound healing
Taxol		Treatment for ovarian cancer
Tissue plasminogen activator (TPA)	Activase	Treatment for heart attack and some strokes
Vaccine	Various (Quadracel, Pediarix, Pentacel)	Prevention of infectious diseases such as hepatitis B, herpes, influenza, pertussis, meningitis

1.4 Biotechnology industry and expression organisms (Chassis)

A number of biological expression systems have been available for heterologous protein production which have unique advantages and disadvantages. In terms of the synthetic biology lexicon, one might consider the host expression organism as a biological chassis on which the parts that achieve actual expression or secretion are archieved. The type of chassis utilised depends upon the nature, features and the specifications of protein products. Some of the commonly used expression systems are mentioned here, as this project focuses on protein production. Historically, eukaryotic cells, such as plants, microalgae,

insects cells, fungi, yeasts and animal cells have been utilised for recombinant protein production, usually in the form of a cell line rather than a complete organism (with notable exceptions such as Spider silk production in goat milk) (Clark, 1998; Service, 2002).

As an overall rule, mammalian cells are preferred in cases where post-transalational modification (PTM) are important, due to their ability to mimic human PTMs relatively faithfully. For example, they can produce correctly glycosylated biologics compatible with humans and, hence are utilised to produce antibodies and hormones with yields reported up 10 gL^{-1} from Chinese Hamster Ovary (CHO) cells, but this yield is not always achievable and is unpredictable (Kim *et al.,* 2012). Another common chassis organism is the *Pichia pastoris* (Weinacker *et al.*, 2013); it has comparatively simple and cheaper growth requirements than CHO cells, and its PTM machinery allows efficient production of proteins into the medium, with some examples of cell engineering to produce human-like glycosylation. As an example *Pichia pastoris* secreted 1.5 g L^{-1} of Insulin in an 80 hr cell culture (Gurramkonda *et al.*, 2010; Weinacker *et al.*, 2013; Fidan and Zhan, 2015), while secreted yields of ~15 gL^{-1} was reported for a 21 kDa porton of rodent collagen (Weinacker *et al.*, 2013). Similarly, the yeast, *Saccharomyces cerevisiae* has also been used for protein production but only in limited cases with amylase production achieved at a yield of 2.5 mgL^{-1} (Rodríguez-Limas *et al.,* 2015). Other examples include Microalgae, such as *Chlamydomonas reinhardtii,* which produced 11.8 gkg^{-1} of milk amyloid dry weight (no secretion) (Gimpel *et al.*, 2015).

The production of non-glycosylated or post-translationally modified protein high value biotherapeutics is often favoured in prokaryotic organisms. With the most common being *E. coli,* which grows on the cheapest carbon sources and is simple to scale up (Ferrer-Miralles and Villaverde, 2013). In addition, its genetic makeup is well known, and it has a short generation (doubling) time and ease of manipulation (Villaverde and Carrió, 2003). *Bacillus* has also been utilised for proficient protein production, achieving yields of up to 25 g L^{-1} for certain enzymes in a 72 hr culture (Liu *et al.*, 2013), but it gives only 1 g L^{-1} for proinsulin

and up to 1 g L^{-1} for amylase (Olmos-Soto and Contreras-Flores, 2003; Heng *et al.*, 2005). *Pseudomonas* has been reported to secrete up to 1 g L^{-1} of various proteins (Retallack *et al.,* 2012). However, despite these, *E. coli* is the most favoured organism when it comes to recombinant protein production and is a 'Workhorse' organism for the biotechnology industry, making an expression platform for a range of proteins.

1.5 *Escherichia coli* as a chassis for synthetic biology and biotechnology

The goal of IB companies is to produce a high value biotherapeutics at minimum cost, maintaining high quality and yields of product, and increasingly in a sustainable manner. In many cases these proteins are produced *via* secretion from the cell cytoplasm into the *E. coli* periplasm (to allow disulphide bond formation in proteins) *via* the Sec or Tat-dependent systems but not into the extracellular medium (Mergulhão *et al.,* 2005; Branston *et al.*, 2012; Matos *et al.*, 2012). Despite the fact that the *E. coli* harbours endotoxin and does not carry out PTMs, a large number of recombinant proteins derived from *E. coli* have been licensed for human use and produced on an industrial scale (Table 1.1) (Walsh 2006; Warikoo *et al.*, 2012). These examples demonstrate the feasibility and amenability of *E. coli* as a cost-effective chassis organism for protein expression, apart from its suitability to industrial scale-up. Recombinant biotherapeutics have been licensed, which highlights the effectiveness of products without any adverse effects on host and hence well viewed by the regulatory authorities.

1.5.1 Enabling tool-technologies for enhancing biotechnology process

A number of technologies of molecular manipulation of genomic and plasmid DNA have contributed to the amenability of *E. coli* for biotechnological applications (Sørensen and Mortensen, 2005; Rosano *et al.,* 2014). The custom made, well-tailored tools such as molecular cloning and protein purification are available for protein overexpression. The post-genomic era has revolutionised these technologies leading to faster and cheaper gene synthesis and sequencing. For recombinant protein expression, commercial cloning systems such as blunt end cloning are easily available. The genes of interest can be assembled by PCR or commercially synthesised *de novo* by services like GenScript or GeneArt® synthesis (Invitrogen) with prices continuously decreasing (£0.085/bp

GenScript, 2018). Codon optimisation for *E. coli* is also offered by gene synthesis companies due to genetic code being degenerate and leads to higher protein expression, although it does not guarantee solubility and functionality of the protein product. However, as the specificity of tRNA for each codon varies in different organisms, availability of codon optimisation tools leads to more efficient translation (Ikemura, 1981; Rosano *et al.,* 2014). Traditionally cumbersome and inefficient genetic manipulations were carried out using DNA ligases and restriction enzymes (Warming *et al.*, 2005). Due to advancement in molecular engineering techniques generation of knockout (KO) mutants and insertion of recombinant DNA has been possible in to the genome of *E. coli* (Casali, 2003; Thomason, *et al.,* 2007).

In order to engineer the genome of *E. coli,* techniques such as recombineering and phage transduction are ideal for insertion (Knock-in) and removal (Knock-out) of the genes and gene loci. In short, recombineering allows transformation of *E. coli* with linear DNA fragments as it contains exonucleases (RecBCD encoded exonuclease V) that degrade linear DNA fragments (Datsenko and Wanner, 2000; Thomason *et al.,* 2007), but also allows recombination with high efficiency (Bieker and Silhavy, 1990). In practice, the Lambda (**λ**) Red plasmid-based method is commonly used where -the plasmid encodes the Red recombinase protein (Gamma, Exo and Beta proteins). The plasmid is inducible hence allows the control of recombineering events (Murphy, 1998; Baba *et al.*, 2006). Datsenko and Wanner *et. al.* (2000) simplified this protocol in their classic paper and it will be utilised in Chapter 2, 3 and 6 (Datsenko and Wanner, 2000).

Alternatively, other bacteriophage-based recombineering methods have been used for many years that provide simple, efficient means of introducing mutations. The best example being, P1 phage transduction (Adams *et al.,* 1992; Donath *et al.,* 2011). In short, the challenges associated with protein production in *E. coli* could be overcome by using genetic manipulations tools by removing, adding or altering loci in the chromosome to improve performance and *E. coli* presents an ideal host for this purpose. For example, overexpression of proteins often results in protein aggregates forming inclusion bodies, making protein

inactive that must then be re-solubilised to be functional- a fact that complicates downstream processing (Rudolph and Lilie, 1996; Sørensen and Mortensen, 2005). Again, *E. coli* presents a good chassis for this purpose. Apart from molecular tools, strains have also been optimised for high quality and yield of protein expression and modified for post-translational modifications (PTM). Commercially available mutant strains such as Origami™ (Novagen) expresses glutathione reductase and thioredoxin reductase that enhances disulphide bond formation in the cytoplasm, which results in improved folding of proteins (Prinz *et al.*, 1997). Similarly, for enhanced cytoplasmic disulphide bond formation, Shuffle® (NEB) strain is utilised that harbours a disulphide bond isomerase (Lobstein *et al.*, 2012). BL 21-codon plus (Agilent technologies) and Rosetta™ (Novagen) strains have also been used for increasing protein expression of the proteins which have rare *E. coli* codon sequences (Hunt, 2005; Sahdev *et al.,* 2008). The controlled expression of recombinant proteins and their effective purification would be an attribute of these enabling technologies to the organism as an excellent production chassis. Additionally, recombinant proteins produced are not correctly modified in prokaryotes due to lack of eukaryotic PTM machinery which includes glycosylation, phosphorylation, acylation, methylation etc. (Yue *et al.*, 2000; Wacker *et al.*, 2002; Sahdev *et al.,* 2008). Therefore, there is an urgent need of engineering of *E. coli* to improve and implement tailored PTM machinery for protein production; and necessary steps, with limited success have been taken and will be shortly discussed (Section 1.4.2).

1.5.2 Improving global understanding of *E. coli* cell behaviour during protein production: systems biology and genome screening as potential solutions

When producing heterologous recombinant proteins in *E. coli*, a considerable reduction in cell growth is often observed, which is referred to as the metabolic burden (Glick *et al.,* 1995). This reduction in growth indicates that a number of resources, raw materials and energy utilised in order to express the heterologous protein from the host metabolism might be limiting (Glick *et al.,* 1995; Sørensen and Mortensen, 2005). In order to identify where this metabolic burden lies, several approaches have been devised. One such approach is to

examine global protein expression under conditions of protein production or in response to growth conditions. One could then compare total protein expression profile of the hosts and use this information to infer physiological stresses and response in the cell- allowing one to suggest alterations to growth conditions or strain characteristics *i.e.* to direct host strain engineering. One example of a technique that allows global protein expression profiling is the mass-spectrometry based technique of isobaric tagging for relative and absolute quantitation (iTRAQ) (Zieske, 2006). This technique involves unbiased labelling of all proteins in the proteome followed by global comparison of their expression and is an excellent tool for discovery proteomics to identify cellular changes in response to growth, mutations, protein production or secretion, information that can then be used to direct host-modification (Chiverton *et al.*, 2016).

An alternative approach to observing the behaviour of the proteome (or transcriptome) would be to use an unbiased screen of genetic elements that might boost production of a desired product. This could be achieved with genetic screening by constructing a genomic library (See Chapter 5) (Glick, 1995; Jones *et al.,* 2000; Lee *et al.*, 2009). In this method, a plasmid-based library containing random segments of the genome is co-expressed in the presence of a screenable production phenotype -such as an enzyme activity or colorimetric response. In these cases, plasmid-based expression systems are preferred for recombinant protein production as they enable control over protein expression, (Flores *et al.*, 2004; Rosano and Ceccarelli, 2014), for example, Pandhal *et al*. reported glycosylation (Pandhal *et al.*, 2012).

1.5.3 Tools of genetic manipulation for recombinant protein expression

The optimum protein expression has also been achieved by modifying promoter and ribosome binding sites (RBS) of the designed biological circuits (Young and Alper, 2010). The biological circuit has been simplified, also characterised promoters and terminators are available in BioBricks: registry of biological parts (http://parts.igem.org/) and ribosomal binding sites in ribosome databases (http://rdp.cme.msu.edu/). These expression systems can be incorporated in association with enzymes, chaperones, hydrophilic proteins

for correct protein folding (Yang *et al.*, 2001). It can also be incorporated with post-translational modifications (PTM) machinery such as N-linked glycosylation system of *Campylobacter jejuni* or prolyl-4-hydroxylase for proline hydroxylation (Khokhlatchev *et al.*, 1997; Wacker *et al.*, 2002; Pinkas *et al.*, 2011; Ikeno and Haruyama, 2013). The availability of the combination of protein purification tags and their unique binding properties with choice of cleavage sites has allowed optimising the protein purification for each expression system (Hunt, 2005). Some protein tags and fusions could also help in protein folding or secretion of the proteins into the medium (Hannig and Makrides, 1998; Rosano *et al.,* 2014).

1.5.4 Cellular localisation and its effect on the protein product

As mentioned above, proteins expressed inside the bacterial cell cytoplasm are directed either to periplasm, cell membrane, exported out of cell body by secretion or remains in the cell cytoplasm (Feldman and Cornelis, 2003; Filloux, 2004). The preferred site of localisation depends on the heterologous protein product in question, with the cytoplasm often favoured as it leads to higher yields. However, cytoplasmic production has several disadvantages (Sahdev *et al.,* 2008). Chief among these is that, disulphide bond formation usually occurs in the periplasm, but if the protein remains in cytoplasm due to absence of periplasmic chaperones and folding enzymes, the protein can be misfolded forming aggregates and eventually resulting into inclusion bodies (Strandberg and Enfors, 1991; Sørensen and Mortensen, 2005; Baeshen *et al.*, 2015). While folding chaperones have also been observed in cytoplasm, it is the periplasm that has the reducing environment that in turn aids in disulphide bond formation and protein folding (Idicula *et al.,* 2007; Eser *et al.*, 2009). A range of strategies have been developed to reduce protein aggregation such as growing cultures at a low temperature, optimising protein expression with low-copy number plasmid and low-strength promoters or using mutant strains specifically designed to overexpress molecular chaperones to reduce protein aggregation (Glick, 1995; Jones *et al.,* 2000; Nishihara *et al.*, 2000). Another strategy that in some cases improves solubility but, as mentioned above is essential for folding of disulphide bonded protein is direction to the periplasm. The nature of periplasm directing secretion signal peptides is well known with their widespread use for

directed secretion into the periplasm in presence of the purification tags (Blattner *et al.*, 1997; Bendtsen *et al.*, 2004). While used in many cases, with success, the extracellular secretion of proteins may also avoid inclusion body formation and make purification easier (removing the need for cell lysis, removal of contaminating molecules) while also avoiding internal or periplasmic proteolysis (Wal *et al.*, 1998). The profit for biotech companies would then also be increased by reducing downstream processing costs (Choi and Lee, 2004). In summary, secretion of recombinant proteins into the extracellular medium would be important to the IB industry for increasing overall profit.

1.5.5 Role of signal sequences in protein secretion

The signal peptides are short amino-terminal parts of secreted precursor proteins which direct the proteins to the export system located in the cytoplasmic membrane (Freudl, 2018). The first protein secretion signal identified was the cleavable N-terminal signal peptides that targeted proteins to endoplasmic reticulum in eukaryotic organisms (Brooks, 2004). However, maltose and lactose transport and catabolism have provided a significant wealth of genetic information to research protein export in *E. coli*; which due to the powerful approaches designed using special properties of beta-galactosidase (LacZ) fusion (Maffei, *et al.* 2017). Several bioinformatics tools have been developed to predict the signal peptides which are based on either weight matrices, sequence alignment or machine learning algorithms. For example, for Sec signal peptides, Phobious (http://phobious.sbc.su.se/) or SignlP are the most popular analysis algorithms that predict the likelihood of a given amino acid sequence for being a Sec signal peptide or not. Also, for Tat signal identification TatP (http://www.cbs.dtu.dk/services/TatP/) , Tatfind of PRED-Tat are utilised. These signal peptide prediction programs are the valuable tools for scanning genome of an organism for signal peptides which can be subsequenctly tested with respect to their performance in the secretion of a desired heterologous target protein by a given bacterial expression host (Freudl, 2018). These signal peptides discriminate exported proteins from those remaining in the cytoplasm. It mediates the targeting and binding of secreted protein precursors to the respective protein translocases in the cytoplasmic membrane.

Additionally, systematic studies were reported on the effect of signal peptide variation on the secretory production of heterologous proteins where they used 173 predicted Sec signal peptides from *B. subtilis* and fused them in an identical manner to a Cutinase gene from the *Fusarium solani pisi* (Brockmeier *et al.*, 2006). Apart from changing the entire signal sequences, the modification of signal sequence by random or site directed mutagenesis approaches represents an alternative to improve the secretion. However, it has become clear that no universal signal peptide exists that promotes the best possible secretory production of any desired target proteins in any given bacterial expression hosts (Brockmeier *et al.*, 2006; Freudl, 2018). One of the promising methods to find the optimal signal peptide for a desired protein is to screen the large diversity of signal peptides either generated by signal peptide variation using signal libraries or by directed or random modification approaches such as high throughput screening strategies reported by Roche *et al*. using the automated microbiorector platforms (Rohe *et al.*, 2012; Freudl, 2018).

1.5.6 Post Translational Modification (PTM)

As mentioned above, PTM is a common consideration in recombinant protein production in the IB industry when it comes to production in *E. coli*, due to absence of post-translational modification (PTM) machinery. This can be compensated by introducing PTM machinery on a plasmid-based system, for example, although *E. coli* phosphorylate recombinant proteins occasionally using native phosphoryl kinases, the reliable phosphorylation of proteins can be achieved by coexpression of recombinant kinases, often fused with target protein (Murata *et al.*, 2008; Sahdev *et al.,* 2008). This strategy has been successfully used for protein glycosylation and acetylases, methylases and their substrates (Yue *et al.*, 2000; Acharya *et al.*, 2005; Pandhal *et al.*, 2012). Another major PTM in eukaryotic protein, that is key for efficacy in human medical usage is Glycosylation. This is a complex PTM process, common in eukaryotes but less so in prokaryotes. There are two types of glycosylation, N-Linked and O-Linked. N-Linked glycosylation occurs in eukaryotes by transfer of oligosacchryl intermediate (Glc 3 Man 9 Glc NAc 2) to asparagine (Asn) residues *via* the action of Oligosaccharyl transferase enzyme (Lizak *et al.*, 2011; Merino and Tomás,

2014). The structure is trimmed by the action of α-glucosidase I and II as a quality control mechanism and is further truncated and decorated with monosaccharides. However, O-linked glycosylation is much simpler. It involves the attachment of single GalNAc monosaccharide to a Serine (S) or Threonine (T) of the polypeptide backbone, followed by the addition of a Sialic acid (SA) residue that terminates the chain or addition of further monosaccharides subunits to form a longer linear or branched chain (Brooks, 2004; Pandhal *et al.*, 2012). Glycosylation mechanisms have also been discovered in bacteria, but their characteristics are significantly different from common core to the terminal glycans present in eukaryotic proteins. It would be problematic if glycosylation is the key for the efficacy and function of heterologous protein, as it would affect stability, solubility, immunogenicity and half-life of the protein (Baker *et al.,* 2013).

The pharmacokinetics and biophysical properties of the proteins can be improved by N-glycosylation or recognition of N-glycans by carbohydrate-binding proteins while lectins allow cell and tissue-specific targeting of proteins (Brooks, 2004; Lizak *et al.*, 2011). O-Linked glycosylation helps in protein stability and recognition including immune response (Steen *et al.*, 1998; Lubas and Hanover, 2000). Comprehensively, due to all these reasons correct glycosylation of therapeutic products is advantageous as proteins would be more characteristic of native protein in structure and function. These are then more likely to interact with other proteins and less likely to elicit an immune response. Non-enzymatic, N-glycosylation has also been observed for recombinant human interferon (ɣ) in *E. coli* (Schellekens, 2002; Mironova *et al.*, 2005). N-Linked glycosylation that occurs in *Campylobacter jejuni* can be transferred to *E. coli*, adding to the possibility of bacteria derived glycoproteins (Wacker *et al.*, 2002; Fisher *et al.*, 2011). The efficiency of N-Linked glycosylation was poor and occurs only in the periplasmic space, though the research was focused on enabling PTMs in *E. coli*. Recent work has demonstrated that it could be improved by codon optimisation of the oligosaccharyl transferases *pglB* that transfers *C. jejuni* hepta-saccharide to asparagine (N) residue by increasing the expression of glycosyltransferase WecA (Wacker *et al.*, 2002; Pandhal *et al.*, 2012). *E. coli* can

also be engineered with the expression of recombinant enzymes to confer O-linked glycosylation (Lubas and Hanover, 2000; Henderson *et al.,* 2011).

1.5.7 Immunogenicity

It is desired that proteins derived for therapeutic uses in humans must have low or at least predictable immunogenicity. Even a slight deviation in 3-D structure of protein to that of human protein, leads to elicitation of strong immune response by human antibodies (Schellekens, 2002). The immunogenic properties of proteins are likely to be higher if a non-human expression host system is used for the expression of heterologous proteins for human use, as the proteins may not harbour all essential human PTM machinery. Also, if the protein is not correctly folded, it may lead to exposure of different antigenic sites resulting in unwanted immune responses (Schellekens, 2002). However, a significant progress has been made to reduce the immunogenicity by implementing PTMs (Section 1.4.3). Successful production of biotherapeutics, which do not need PTM e.g. glycosylation has been encouraging and is advancing to clinical trials (Clark, 1998; Frenzel *et al.,* 2013). Storage of recombinant proteins may also result into the aggregation, oxidation-reduction (redox) or change in overall 3-D structure, eliciting an immune response. This could be nullified by shielding the antigenic sites with covalent attachment of polyethylene glycol (PEGylation) to the recombinant therapeutic thus escaping the immune system (Schellekens, 2002; DeFrees, 2006; Cong *et al.*, 2012). Interestingly, PEGylation protects the protein from enzymatic degradation and increases its circulatory half-life. It also eliminates *E. coli* derived contaminants (Schädlich *et al.*, 2009; Cong *et al.*, 2012). Finally, purifying proteins for human use, steps must be taken to ensure complete removal of contaminant as even a minute amount of endotoxin contaminant may have a strong immune response (Petsch and Anspach, 2000; Schellekens, 2002).

1.6 Protein secretion systems in *Escherichia coli (E. coli)*

Protein secretion is an important cell function of prokaryotic cells, in which proteins are transported from the cell cytoplasm to other compartments of the cell, environment, other bacteria or directly into eukaryotic cells acting as

specialised tools for microbes to interact with their host or environment (Basler *et al.,* 2013; Costa *et al.*, 2015; Green and Mecsas, 2015). It has many roles, being key to cell growth, organelle biogenesis, nutrient acquisition and virulence factor expression as well as playing a role in host-microbe crosstalk (Bhowmick and Tripathy, 2014). To achieve this feat, bacteria have evolved many systems to allow proteins to cross membranes, and it is this that is the subject of the next section of this thesis and which is outlined in Fig. 1.1.

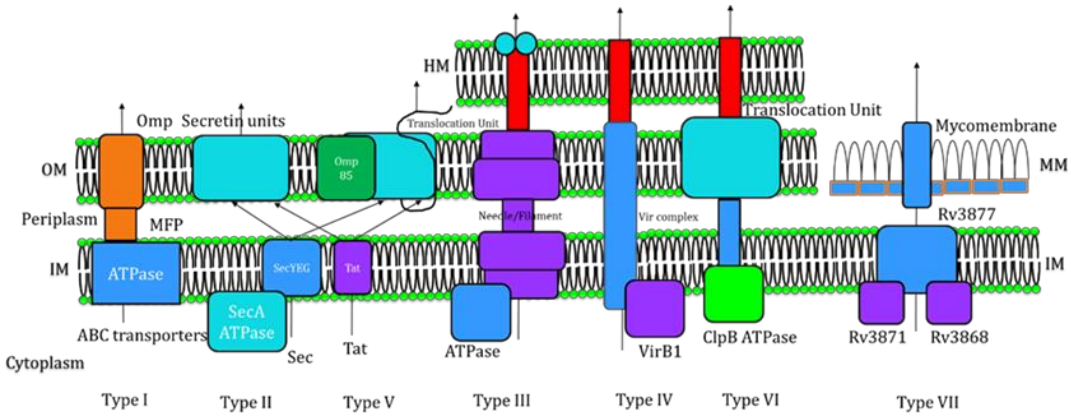


Figure 1.1: Secretion systems in Gram-negative bacteria.
Showing the (Type I-Type VII) and Sec-Tat secretion systems and their localisation in the inner membrane (IM), outer membrane (OM), and host cell membrane. Note: Type III injectisome is shown (Not Flagellar T3SS). Adapted from (Costa *et al.*, 2015).

1.6.1 Type I secretion system (T1SS)

The type I secretion system is an example of a one-step process in which protein is directly exported from the cell cytoplasm to the environment. It secretes a number of substrates often related with nutrient acquisition and virulence with their architecture closely resembling the Resistance-Nodulation-Division (RND) family of multidrug efflux pumps (Bhowmick, 2014; Costa *et al.*, 2015). These systems are made up of three components that utilise a double-membrane spanning channel comprised of the TolC membrane protein embedded in outer membrane, an ATP-binding cassette transporter (ABC) spanning the inner membrane while an inner membrane-anchored fusion protein (MFP) completes a tunnel, with TolC spanning the periplasm linking ABC and MFP (Chung *et al.*, 2009; Tseng *et al.,* 2009). TolC is a trimer which forms a helical barrel in the periplasm and β-barrel channel in OM. These systems include those secreting Hemolysin (HylA) and Colicin V (Andersen *et al.*, 2002;

Delepelaire, 2004). During secretion a twisting motion of the helix triggers the opening of the channel facilitated by ATP turn over by the ABC protein in the inner membrane (Andersen *et al.*, 2002). TolC associates with multiple MFP complexes while the ABC and MFP recognise the substrate for secretion. The protein is directed for export by a non-cleavable, C-terminal signal sequence by binding to the ABC cassette (Zhang *et al.*, 1995; Records, 2011). ATP hydrolysis energizes the protein secretion into the periplasmic cavity of the MFP, mediated by the C-terminus of ATP-binding cassette complex (Kanonenberg *et al.,* 2013; Thomas *et al.,* 2014). It is also observed that RND pumps use a proton gradient for secretion of the substrate, enabling the ABC-MFP complex to associate with TolC, triggering the opening of TolC that leads to release of substrate into the extracellular space (Thomas *et al.,* 2014; Costa *et al.*, 2015). Recombinant proteins have been secreted using the T1SS and HylA secretion signals with secretion efficiencies of 10-90 % and reported yields of ~2 mgL^{-1} (Fernández and Lorenzo, 2001; Ferna, 2004). In addition, recombinant lipases have been reported to be secreted *via* this system, however the lipases are the natural substrates in this case (Chung *et al.*, 2009). The cytoplasmic protein accumulation, dependence of growth phase and oxygen availability are the negative potential aspects of T1SS for recombinant protein production (Baneyx, 1999; Specht, 2003; Desvaux *et al.*, 2009).

1.6.2 Type II secretion system (T2SS)

The Type II secretion system spans both inner and outer membranes. It is comprised of a channel in the outer membrane referred to as secretin and a periplasmic pseudopilus anchored to an inner membrane platform, which is tightly associated with a cytoplasmic ATPase (Saier, 2006; McLaughlin *et al.,* 2012). First, the protein is exported from the cytoplasm to the periplasm *via* Sec or Tat pathway and then it is exported to environment from the periplasm (Mergulhão *et al.,* 2005). T2SS secretion substrates are transported to periplasm either as unfolded poly-peptides by the Sec translocon or as folded polypeptides by the Tat transporter and finally protein is exported outside in a folded state through the secretin complex (Francetic *et al.*, 2000; Nivaskumar and Francetic, 2014).

1.6.2.1 Sec secretion system

The Sec secretion machinery is the essential system that cell uses for ~ 95 % of its exported and membrane proteins (Jermy, 2012; Freudl, 2013). The proteins are targeted for secretion *via* the Sec translocon by two pathways SecB and SRP (Fig. 1.2). (1). The SecB recognises a cleavable hydrophobic N-terminal Ala-X-Ala signal sequence motif of protein signal sequences and directs them to periplasm *via* transferring it to the SecA protein, which drives its translocation through SecYEG channel with the help of ATP hydrolysis (Fig. 1.2 A) (Pallen *et al.,* 2003; Chatzi *et al.*, 2014; Denks *et al.*, 2014). (2). The signal recognition particle (SRP) pathway utilises a co-translational mechanism mainly to secrete inner membrane proteins. The SRP binds to the secretion signal of nascent proteins as they emerge from the ribosome. This compex is then directed to the inner membrane where the SRP recruits docking protein FtsY (Beckwith, 2013; Saraogi and Shan, 2014). The FtsY in turn binds to the SecY and delivers the ribosome-protein complex to the SecYEG channel, translocating nascent protein across the cytoplasmic membrane (Blight *et al.,* 1994; Chatzi *et al.*, 2013, 2014). Following secretion into the periplasm, the N-terminal signal peptide is cleaved, and the protein transverses into the periplasm. Although, current evidence of protein export to the extracellular space is limited, a prevalent idea is that in some cases, the protein binds to the tip of the pseudopilus and the periplasmic domain of secretin, stimulating ATPase activity with subunits being added to the growing pseudopilus which now functions as a piston; that pushes substrates through secretin channel. However, the detailed understanding of protein targeting to this complex is not clear and hence its use for biotechnology is not readily achievable (Sandkvist *et al.,* 2001; Nivaskumar and Francetic, 2014).

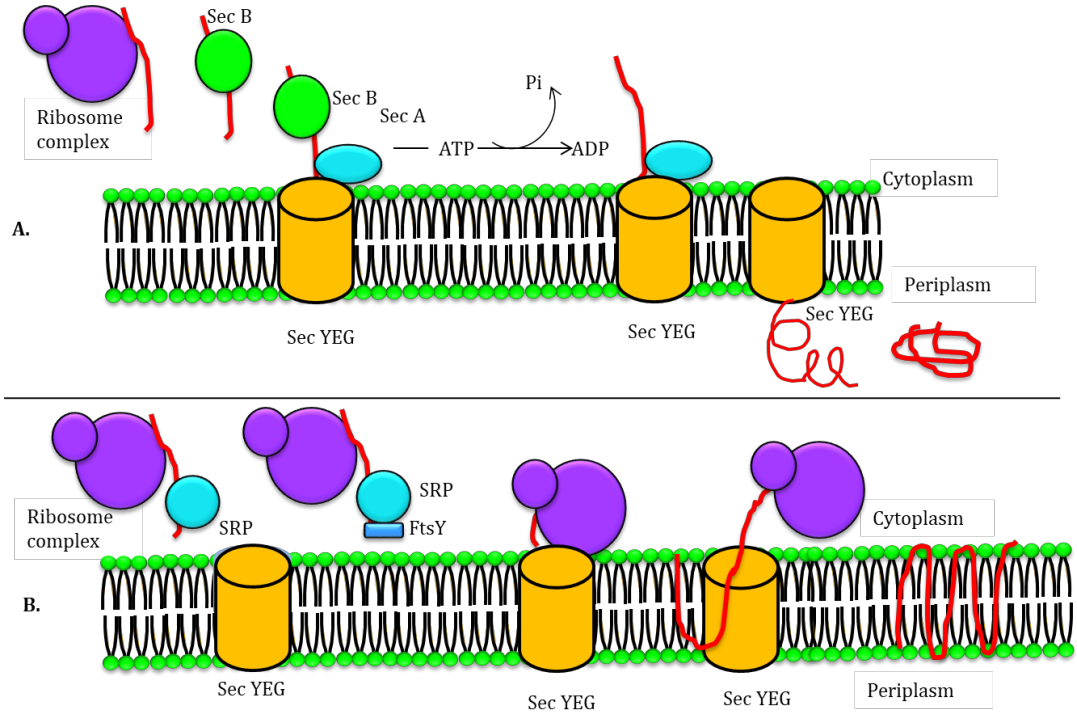


Figure 1.2: Sec secretion pathways.

The nascent polypeptides are translocated across the periplasm *via* SecB pathway (A) or remain embedded into the membrane as transmembrane protein (B) through Sec YEG complex, with the help of SecB-SecA proteins and SRP-FtsY complex (SRP pathway), respectively. Ribosome complex are shown in purple, SRP: signal recognition particle. Adapted from (Schaerlaekens *et al.*, 2004; Desvaux *et al.*, 2009).

1.6.2.2 Tat Secretion system

The structure of secretion substrates of both Sec and Tat pathways is similar, in that they harbour Ala-X-Ala motifs and the hydrophilic-hydrophobic regions. However, Tat signal contains a Twin-Arginine (SRRXFLK) residue motif and is less hydrophobic (Ulfig and Freudl, 2018). It also contains unique basic residues that are not found in the Sec signal that are thought to hinder interaction with the Sec secretion apparatus (Dijl *et al.*, 2002). The substrates are assembled at the inner membrane TatBC complex receptor and are recognised by TatC with transport of folded proteins across the inner membrane is facilitated by TatA (Fig. 1.3) which is driven by the proton motive force (PMF) (Alcock *et al.*, 2013). Following the cleavage of secretion peptides into the periplasm, the protein undergoes folding and PTMs. AB_5 chaperone recognises the tertiary and quaternary structures of these folded proteins and are directed for secretion through the secretin protein to the extracellular membrane (Nielsen and Engelbrecht, 1997; Sandkvist *et al.,* 2001).

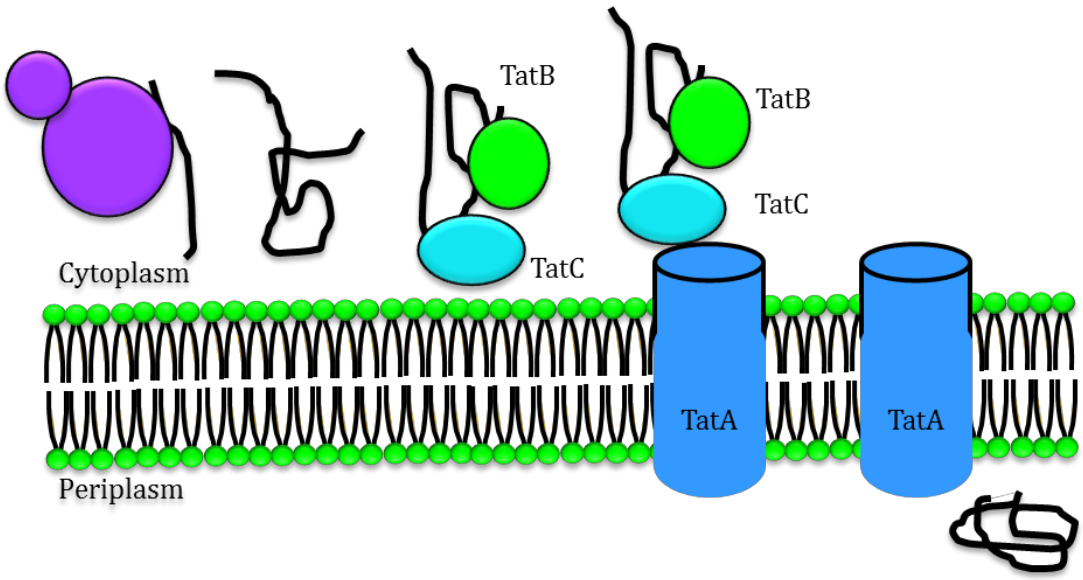


Figure 1.3: Secretion through the Tat pathway

The nascent peptide emerging from the ribosome docks at the cytoplasmic membrane with the help of TatB and TatC to the cytoplasmic TatA complex and translocated across the periplasm. Ribosome complex shown in purple. Adapted from (Schaerlaekens *et al.*, 2004; Green and Mecsas, 2016).

The Type II secretion system has several native secretion substrates. Their N-terminal secretion signals can be fused to recombinant protein and secreted *via* T2SS into the periplasm (Berlec and Štrukelj, 2013). The earliest evidence of clinically important recombinant proteins came through the availability of Human Insulin and Bovine growth hormone (Swartz, 2001). It took years of optimisation to yield ~ 4.34 g L^{-1} of Human insulin intracellularly and ~ 7.0 mg L^{-1} extracellularly. The later was achieved *via* the T2SS; Sec pathway although the possibility of cell lysis was not excluded (Baeshen et al 2014). In addition, TAT systems have been utilised for protein production, yielding ScFv (~4.4 mg L^{-1}), Green Fluorescent protein (~1 g L^{-1}), Interleukins and Human growth hormone (hGH) following periplasmic extraction *via* the TAT system (Choi and Lee, 2004; Branston *et al.*, 2012; McLaughlin *et al.,* 2012). Thioredoxin have been also secreted by fusion with a DsbA signal peptide using SRP pathway with ~60 mg L^{-1}, while GFP in small batch culture (~1 g L^{-1}) in fed-batch fermentation to an OD_{600} 150 (Matos *et al.*, 2012). Low recombinant protein translation is desired in this system, due to easy saturation of T2SS for protein transport from periplasm to the medium and hence yield is low (Mergulhão *et al.,* 2005). However, as described the major use of T2SS for recombinant protein

production focuses on protein secretion into the periplasm that enables desired properties to the protein such as protein folding, lack of proteolysis, however, the cells must be lysed to obtain the protein. These systems render limitation for industrial application if the protein needs to be secreted extracellularly.

1.6.3 Type III secretion system (T3SS)

The Type III secretion system (T3SS) is represented by two different systems, the bacterial flagellum and the T3SS Injectisome, as they share several components and many of their proteins have structural and functional homologies (Minamino and Macnab, 1999). There has been a heated debate around the evolutionary divergence of these two T3SS, with evidence of each being diversified from the other but sharing a common ancestor (Aizawa, 2001; Gophna *et al.,* 2003; Macnab, 2004; Abby and Rocha, 2012). Both systems span both Gram negative cellular membranes and both of these systems have the capacity to secrete a large number of protein subunits from the cytoplasm to the extracellular space or direct into the host membrane.

1.6.3.1 Injectisome

The injectisome T3SS is characteristically present in pathogenic bacteria including enteropathogenic *E. coli*, *Salmonella, Yersinia* and many plant pathogens like *Pseudomonas syringae* (Jarvis *et al.*, 1995; Dietsche *et al.*, 2016). Injectisomes share high structural and genetic homology with the parts of F_1F_0 ATP synthases and the flagellar apparatus (Portaliou *et al.*, 2016). The bacterium injects toxic proteins using the injectisome into the eukaryotic host in a process, which is essential for infection, survival and pathogenicity (Cornelis, 2000; Erhardt *et al.,* 2010). The injectisome is made up of ~25 proteins, assembled in a highly ordered and regulated manner. The central needle complex spans both inner and outer membranes and facilitates translocation of secreted effector proteins in one-step from the cytoplasm to the extracellular environment or host cell through a central channel of 28 Å diameter before they either polymerise at the tip or enter the host cell (Diepold and Wagner, 2014; Galán *et al.*, 2014; Dietsche *et al.*, 2017). In addition to the central needle complex, the injectisome consists of cytoplasmic proteins, rings and export apparatus (Galan and Wolf-Watz, 2006). The basal structure of the injectisome is formed of IM and OM rings

that encompass the inner rod of T3SS export apparatus situated within the IM concentric ring, which is made up of PrgK and PrgH proteins (Fig.1.4).

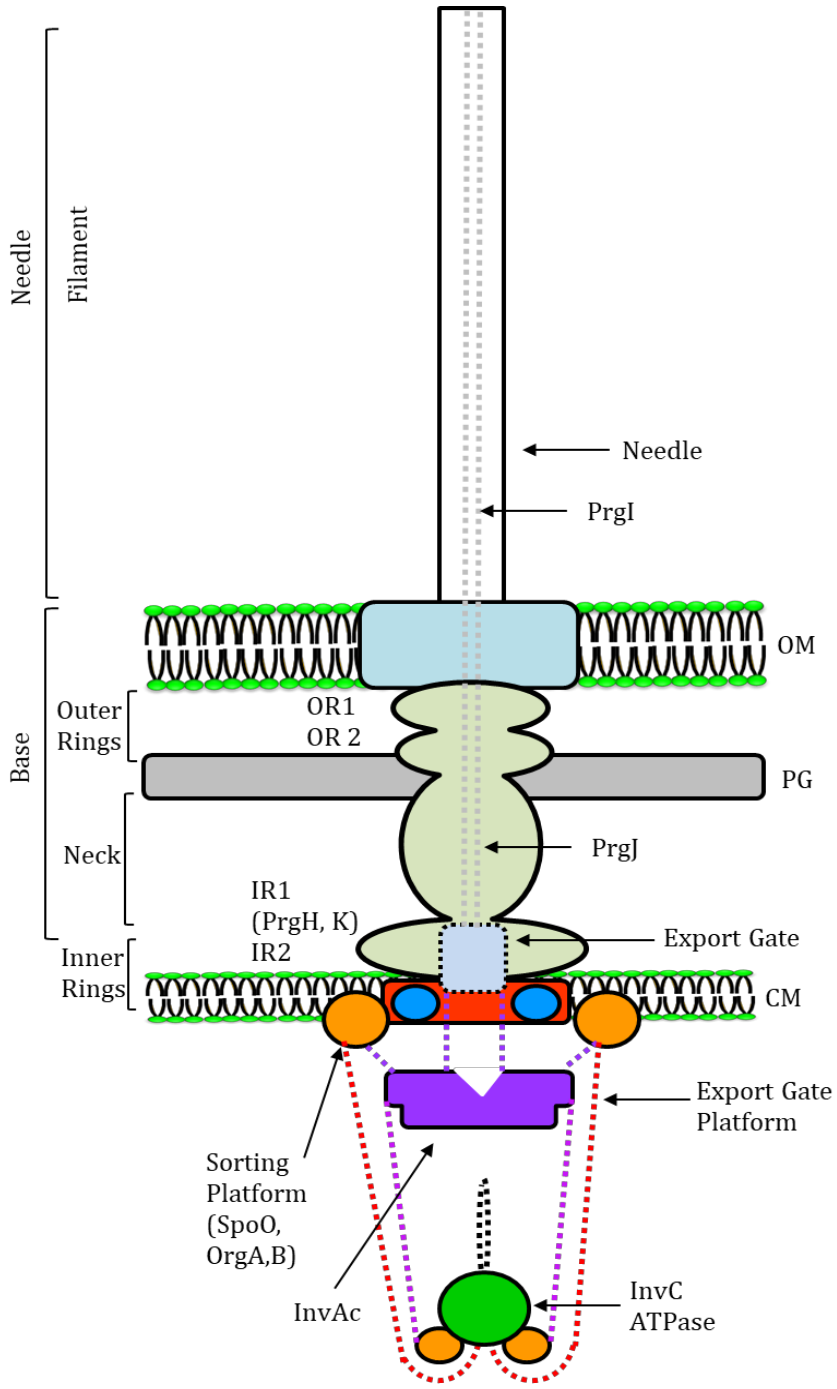


Figure 1.4: Schematic representation of the structure of the Injectisome.
The injectisome consists of base and needle complex, made up of rings, OR: Outer membrane ring, IR: Inner membrane ring, OM: Outer membrane, PG: Peptidoglycan layer, CM Cytoplasmic membrane. Adapted from (Kawamoto *et al.*, 2013).

The inner membrane hosts the secretion apparatus by forming a sorting platform for substrates. An ATPase complex mediates interaction with the export

apparatus exhibiting substrate recognition and unfolding substrates prior to export (Hu *et al.,* 2011). The export apparatus recognises a common N-terminal signal sequence which is also a binding site of T3SS chaperones for effector proteins such as SopB and SigE (Lloyd and Buckman, 1991; Anderson *et al.,* 2002; Munera *et al.*, 2010). The rod (PrgJ) assembles on IM ring, the InvG protein of the secretin family forms the OM ring and the Pilotin (InvH) protein helps escort the OM into place (Koster *et al.*, 1997; Crago and Koronakis, 1998). Next, the neck region links the OM ring to the IM ring by spanning the periplasm and then the needle structure assembles on the rod structure. The rod is made up of ~ 100 (10 kDa) polymerised Prgl proteins. For all these substrates, the InvC ATPase complex couples ATP hydrolysis to folding and release of effector and needle proteins from their chaperones, while the PMF also helps in secretion through the T3SS (Akeda and Galán, 2005; Kosarewicz *et al.,* 2012). Once the needle is complete, a translocon pore (IpaB and SipB) assembles at the tip of the needle that is opened only upon contact with the host membrane leading to the formation of translocon channel through which effector protein are delivered from bacterial cytoplasm directly to eukaryotic cells (Cornelis, 2006; Costa *et al.*, 2015).

1.6.3.2 Flagella

The flagellar T3SS (FT3SS) facilitates the building of a functional flagella that provides motility to the cell. The FT3SS shares mechanism of assembly and structural morphology with the injectisome. Both systems assemble in a similar ordered manner from the proximal to the distal end. The progression steps are controlled by regulatory checkpoints, pathways and contains several conserved proteins in the secretion apparatus (Chilcott and Hughes, 1998; Aizawa, 2001). The structure comprises a basal body and a cylindrical protein that extends from the cytoplasm and protrudes from the bacterial cell body. Assembly occurs at the tip of the growing structure. This mechanism of assembly starts at the base of the rod when its assembly in initiated on to the MS-ring foundation (Aldridge and Hughes, 2002; Macnab, 2003). The FT3SS will be discussed in more details in Section 1.7.

1.6.4 Type IV secretion system (T4SS)

The Type IV secretion system is ubiquitous in nature and used to transport nucleic acids, toxins and effector proteins from bacteria to host cell or environment (Fronzes *et al.,* 2009). It is an ATP dependent transport process in which the T4SS spans both membranes of Gram-negative bacteria (Lawley *et al.*, 2003; Zechner *et al.,* 2012). The translocation complex is made up of six scaffolding proteins, where two types of subunits form the pilus that extends into the extracellular space. VirB1 is required for pilus biogenesis, which has periplasmic lytic transglycosylase activity and degrades the peptidoglycan layer. The ATPase is located in the IM and powers the system, provides energy for pilus biogenesis and substrate translocation (Fig. 1.1) (Cascales and Christie, 2004; Fronzes *et al.*, 2009; Costa *et al.*, 2015). The T4SS pilus extends into the extracellular space and when it encounters a host cell, switches on substrate translocation of the T4SS secretion substrates. Proteins are directed for export by means of a C-terminal signal sequence that contains unstructured clusters of hydrophobic, positively charged residues, while nucleic acids are exported by binding of contact-dependent ATPase coupling protein (VirD 4) to a -GGGG- nucleotide structure (Darbari and Waksman, 2015; Fronzes *et al.*, 2009). The system also requires chaperone and adapter proteins for efficient secretion, similar to other systems reported (Fronzes *et al.,* 2009; Zechner *et al.,* 2012). The requirement of chaperones in protein secretion represents a means of prevention of premature protein misfolding and aggregation as most of the proteins are ultimately destined for polymerisation.

1.6.5 Type V Secretion system (T5SS)

The T5SS are employed to secrete the proteins across the Gram-negative OM, referred to as auto-transporters family and contains distinctly characterised 5 classes (a-e), while a-autotransporter (Va) and b-the two-partner secretion pathway (Vb) represents the classes most studied in depth (Fig. 1.1) (Henderson *et al.*, 2004; Ulsen *et al.*, 2014). Proteins secreted through the T5SS contains a signal sequence, a passenger domain and a β-domain (Desvaux *et al.,* 2004; Henderson *et al.*, 2004). During protein secretion *via* auto-transporters, the proteins are first recognised *via* their N-terminal signal sequences by the Sec or

SRP pathways and directed to the periplasm, where the signal sequences are cleaved, and the β-domains are assembled into β-barrel structures in the cell membrane. The passenger domain is then translocated and released across the membrane or remains anchored in the cell envelope for surface display (Desvaux *et al.,* 2004). In contrast to Va auto-transporters, in the Vb system, the β-domain and passenger domains are translated into two peptides, also referred to as two-partner secretion (Tps) system (Pallen *et al.,* 2003; Desvaux *et al.,* 2004; Henderson *et al.*, 2004). The Vc systems are most complex auto-transporters, the Vd systems are similar to Va systems. While the Ve systems are called inverse auto-transporters, based on the fact that the domain order is reversed in which the C-terminal part comprises a passenger domain while the N-terminal part forms the translocon pore (Henderson *et al.*, 2004; Oberhettinger *et al.*, 2012).

The type V, auto-transporter secretion system has potential application in recombinant protein secretion, especially for cell surface expression rather than complete secretion in the medium. The passenger domain of an autotransporter protein can be replaced with a recombinant protein, a construct that results in surface display on the outside of the bacterial cell and has great importance in vaccine development (Amsterdam *et al.*, 2014). For example, recombinant ScFv, lipases and toxins have been reported to be displayed on the surface of *E. coli* after secretion through the T5SS (Ulsen *et al.*, 2014). There is no evidence available for the use of the Type Vb-d for biotechnological applications, but the intimins (type Ve) have been used for biotech purposes. Autotransporters-mediated secretion or surface display has been used in various biotech applications that includes live vaccine development, whole cell biocatalysis, biosorbent-biosensor development and protein library screening (Ulsen *et al.*, 2014). Recently, heterologously expressed AT module from the *E. coli* protein Pet to secrete mCherry and antigens such as *Bordetella pertussis* Pertacin, *Y. pestris* YapA, and *Mycobacterium tuberculosis* ESAT6 at titers up to 5.4 mg/L in *E. coli* BL21, which have been proved to secrete folded and active heterologous proteins. However, this system could be limited by complexity and proteins containing disulphide bonds may not be secreted (Burdette *et al.*, 2018).

1.6.6 Type VI secretion system (T6SS)

The T6SS is an injectisome like structure that is structurally and mechanistically analogous to an intracellular membrane-attached contractile tail of T_4 bacteriophage (Ho *et al.,* 2014). It has a tail complex equivalent that resembles the contractile tail of phage, tail sheath, an inner tube on the base plate; which is anchored to cell membrane complex by extending from inner to outer membrane (Cascales, 2008; Costa *et al.*, 2015). The conformational change induced by the extracellular signal of the base-plate drives a contraction in the sheath that ejects the effectors out of the cell through the inner tube (Fig. 1.5) (Costa *et al.*, 2015). The T6SS spans the bacterial IM and OM and transfers toxic effector proteins from the cytoplasm into the eukaryotic cells or bacterial target cells in one-step, propelling the structure and effector proteins into a bacterial or mammalian host cell using ATP as energy source (Coulthurst, 2013; Ho *et al.,* 2014).

The T6SS acts as a hybrid system since it has components derived from the T4SS and the bacteriophage (Records, 2011). Its mechanism of action gives an opportunity of researching evolution, in terms of how bacteria have hijacked the viral system or virus has hijacked the bacterial system (Basler *et al.,* 2013). The Type VI secretion system is a versatile weapon deployed by many bacterial species to target the host or rival bacteria (Coulthurst, *et.al* 2016). It delivers multiple effector proteins directly into the target cells using dynamic 'firing' mechanism which is most likely followed by recycling of the machinery and are related to contractile bacteriophage tails (Coulthurst, *et al.* 2016). However, the mechanism of firing and where and when bacterium dictates firing still needs to be fully elucidated. Recent findings regarding mode of action of firing and translocation of effector proteins have provided a significant step forward and its discovery of mechanism that T6SS can deploy to efficiently secrete wide range of effectors has provided significant insight. However, numerous questions of anticipated utilisation of Type VI secretion system in biotech purposes remain unresolvd (Coulthurst, *et al,* 2016).

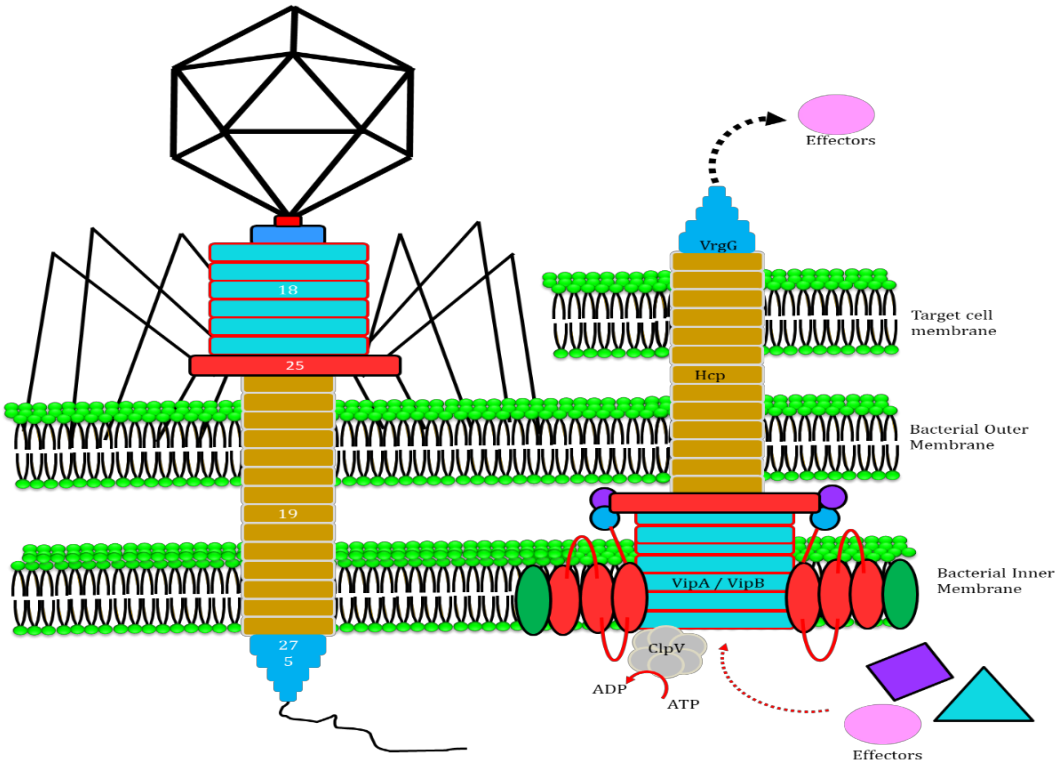


Figure 1.5: T4 Phage and T6SS injection machines.

Bacteriophage T4 molecular architecture (left) has been studied in detail, components are labelled with their gene product as numbers, while the T6SS model is based on the predicted subcellular localisation of proteins and the homology between T6SS and Phage protein sequences. The homologous proteins are coloured with the same colour. Adapted from (Records, 2011; Basler *et al.,* 2013).

1.6.7 Type VII secretion system (T7SS)

In Gram-positive organisms, including species of the *Mycobacterium* and *Corynebacterium,* the cell wall layer is heavily incorporated with lipids which form a dense, waxy, hydrophobic layer on the outer surface of bacterial membrane. It makes protein transport difficult in these bacteria and is often called the cytomembrane. It serves as an effective barrier against environmental and antimicrobial therapies (Freudl, 2013; Ates *et al.,* 2016). Therefore, these bacteria utilise a specialised mechanism of protein transport across their inner and myco/cyto-membrane referred to as the T7SS (Freudl, 2013; Ates *et al.,* 2016). The core structural components of the T7SS are encoded in a linked cluster, made up of EccB, EccC, EccD, EccE and MycP membrane proteins which were first reported in *Mycobacterium tuberculosis*, referred to as the ESX system. Except for EccD, all other proteins are hydrophobic and hence interact with multiple membranes and cytoplasmic chaperones. The four Ecc proteins form

ESx-5 that in turn forms a large inner membrane complex, which presumably contains a channel through which proteins transverse, while MycP is a myosin or subtilisin-like protease that plays an important role in protein regulation, while its role in protein translocation remains unknown (Freudl, 2013; Ates *et al.,* 2016).

Despite being rigorously studied, how a substrate navigates through the mycomembrane and or the mechanism of protein translocation through the T7SS remains unknown (Ates *et al.,* 2016). However, as these systems are essential for virulence and viability, the understanding of T7SS could direct research to improve the vaccines and develop novel drugs in IB industry (Ates, *et al.* 2016).

1.6.8 Type VIII secretion system (T8SS)

The T8SS is also referred to as the extracellular nucleation-precipitation pathway in Gram-negative bacteria. It is responsible for secretion and assembly of pre-pilins for fimbriae biogenesis, which is the prototypical curli fibre (Desvaux *et al.*, 2009). Notably the T8SS differs from the other secretion systems involved in assembly of surface appendages (e.g. flagella, pilli), in that fibre-growth of the curli fibres occurs extracellularly (Desvaux *et al.*, 2009). These curli fibres are also called Thin aggregative fimbriae (Tafi) and were first identified in *Salmonella* as a part of the *age* operon as key drivers of biofilm formation, while its subsequent homologs identified in *E. coli* are called Csg. One of the three putative curli assembly factors CsgF acts as a nucleation protein for the other proteins to assemble on the surface but understanding of the system is still in its infancy (Desvaux *et al.*, 2009).

The *E. coli* T8SS has been engineered for protein secretion and it can be overexpressed using a plasmid. However, there is no evidence for the comparison of secretion *via* a plasmid borne T8SS to the secretion by the genomic T8SS. For example, active single domain antibodies (sdAbs), antimicrobial peptides, mussel foot proteins and lanthanides binding domains have been secreted as C-terminal fusions to the curli subunit CsgA (Burdette *et al.*, 2018). Secretion *via* the T8SS was reported to be limited which might due to size fusions of mCherry, Bla, and alkaline phosphatase to CsgA were not secreted by the T8SS (Burdette *et al.*,

2018). Unlike the other bacterial secretion systems, T8SS engineering has been focused on engineering the secreted curli fibers to make nanofibrous meshes with a range of properties. CsgA has been secreted with C-terminal fusion to a wide range of peptide tags that provided route for functionalising the amyloid fibre network from enzyme scaffolding to nanoparticle patterning, suggesting the robustness of the machinery (Burdette *et al.*, 2018). Additionally, a simple method to purify genetically modified CsgA fibres *via* filtration with yields of hundreds of milligrams per litre have been reported (Courchesne *et al.*, 2017).

1.6.9 Type IX secretion system (T9SS)

The T9SS is a complex translocon found in *Flavobacterium johnsoniae* and *Porphyromonas gingivalis,* where it is used to assemble the bacterial gliding motility system and act to translocate important virulence factors across the outer membrane, respectively (Lasica *et al.*, 2017). This secretion system is characterised by an unusual mechanism. After translocation across the IM *via* Sec-dependent mechanism, the protein substrate recognised by the T9SS using a C-terminal domain (CTD) are directed through the OM translocon. Following translocation across the OM, the CTD is removed by a protease with sortase-like activity and an anionic LPS is attached to the newly formed C-terminus, resulting in secretion of cargo protein into the extracellular milieu or being covalently attached to the bacterial surface (Lasica *et al.*, 2017). The T9SS is regulated by a two-component system, however, the precise environmental signal which triggers its assembly and secretion remain unidentified (Abby and Rocha, 2017; Lasica *et al.*, 2017).

In many respects, T9SS is the most robust secretion system and it facilitates secretion of upto 35 cargo proteins bearing the CTD. Many of these are implicated in pathogenicity in *P. gingivalis* (Lasica *et al.*, 2017). Secretion of chitinase and cellulase requires T9SS and it functions as a non-invasive tool for food acquisition and movement (Lasica *et al.*, 2017). A thorough study with CTDs from various *P. gingivalis* T9SS cargo proteins (e.g. RgpB) fused with GFP found that GFP was secreted and PTM modified by *P. gingivalis* in the same way as the native T9SS cargos (Lasica *et al.*, 2017). These reports suggest that T9SS could

also be utilised for heterologous protein secretion (Abby and Rocha, 2017; Lasica *et al.*, 2017).

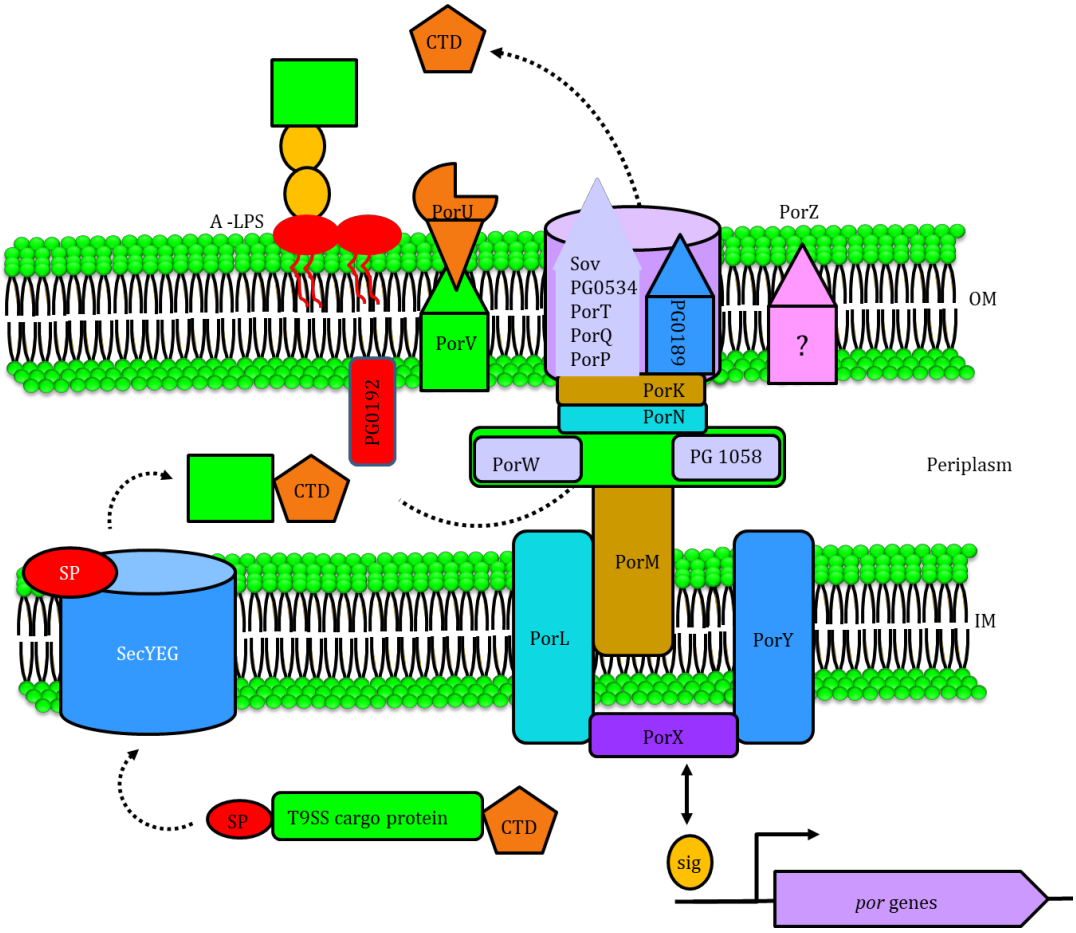


Figure 1.6: Schematic diagram of the Type IX secretion system (T9SS).

The T9SS cargo protein is equipped with two sorting signals. N-terminal signal peptide (SP) directs the proteins to the general secretion system (SecYEG) and conserved C-terminal domain (CTD) is recognised by T9SS. CTD is cleaved off by PorU sortase and a secreted protein is modified by attachment of A-LPS leading to anchorage of cargo protein to the cell surface. Adapted from (Abby and Rocha, 2017; Lasica *et al.*, 2017).

1.7 The bacterial flagellum

The bacterial flagellum is a cell surface organelle that allows cell motility in many Gram-negative and Gram-positive bacteria, as well as Spirochetes and some Archaebacteria (Macnab and Koshland, 1974; Moens and Vanderleyden, 1996; Wang *et al.*, 2012). The flagellum is made up of ~50 different proteins, each in multiple copies from a few to many thousands and extending from the cell cytoplasm across the membrane(s) and into the cell exterior (Macnab, 2003). As outlined above, it is a motor organelle that includes a protein secretion apparatus which enables self-assembly and is a member of the Type III family of bacterial

secretion systems, referred to as the FT3SS (Macnab, 2004). Each flagellum has a helical filament whose rotation is driven by ion flux, powered by a nanoscale electric motor (Berry, 2001; Berg, 2008). *E. coli* and *Salmonella* have an average 4-10 flagella that grow up to 20 μm in length (Martinez, 1966; Gerber *et al.,* 1973; Lino, 1977). The bacterial flagellum is the result of billions of years of evolution, comprised of different proteins that assemble in an ordered fashion from the proximal to the distal end and controlled by several well understood regulatory checkpoints (Chilcott and Hughes, 1998; Macnab, 2003). Most of the flagellar proteins in the outer structures are secreted as unfolded subunits through the existing central channel and assemble beneath FT3SS cap proteins (Yonekura *et al.,* 2002; Schuhmacher *et al.,* 2015). A sensory system drives the movement of flagella by dictating clockwise and anti-clockwise spin of motor (Smith and Koffler, 1971). If *E. coli* flagella rotate clockwise, it tumbles and moves ~30 diameter per second in one direction (Adler, 1966; Schade *et al.,* 1967). Chemical stimulants such as sugars, amino acids and chemical deterrents trigger the cell surface receptors leading to switching the direction of rotation of bacterial flagella to anti-clockwise, leading to assembly of flagella into a bundle allowing bacteria to move towards beneficial environments (Lino, 1977).

1.7.1 Structure and assembly of the Flagellum

The mechanism of biogenesis of flagella assembly will be explained in detail in this section, while their specific proteins and functions are listed in Appendix 1. Structurally, flagella are made up of a basal body, hook and filament and are generated in this order (Fig. 1.7) (Aldridge and Hughes, 2002). The basal body is a complex structure buried in the outer and inner membranes, while the flexible hook acts as a universal joint connecting the rotating rod and basal body to the extracellular structures filament (Macnab, 2003; Chevance and Hughes, 2008). The flagellar motor is a molecular machine of 11 MDa made up of 13 component proteins and ~25 further proteins required for its expression and assembly (Macnab *et al.,* 2004). The basal body acts as a motor and secretion apparatus while hook (FlgE) and filament (FliC) extend from the surface of the cell body (Macnab, 2004).

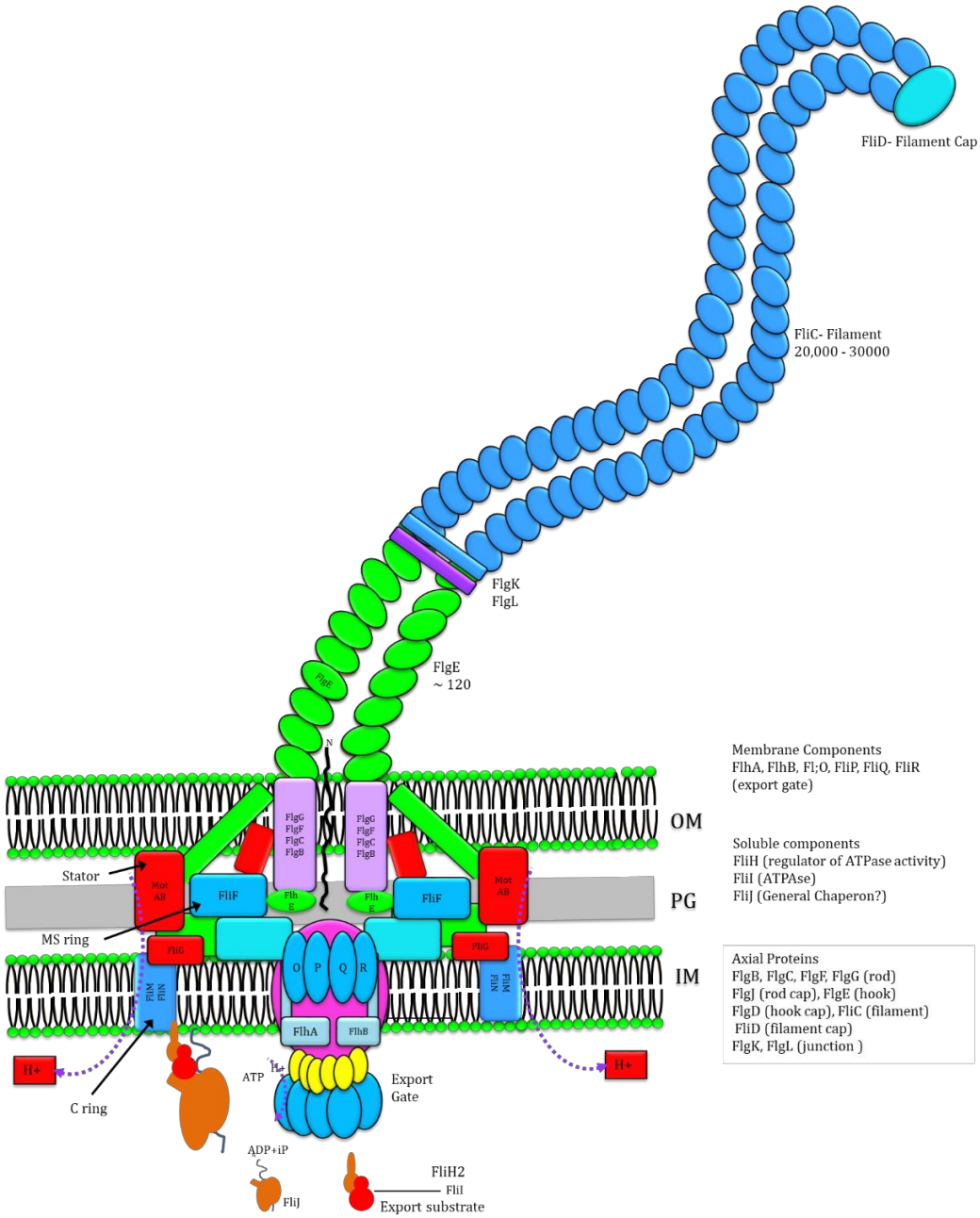


Figure 1.7: Bacterial flagella and Type III secretion export apparatus.

Showing the bacterial flagella and the distribution of proteins in the flagella Type III secretion system. Adapted from (Chevance and Hughes, 2008b; Evans *et al.*, 2013; Minamino *et al.*, 2014; Fabiani *et al.*, 2017).

1.7.1.1 The hook-basal body of the bacterial flagella

Flagellar assembly initiates at the MS ring, made up of FliF subunits that are incorporated into the bacterial cytoplasmic membrane after secretion *via* the Sec-dependent export mechanism and forms into an 8 MDa structure that acts as a platform for the rest of the motor to build upon (Desvaux *et al.*, 2009). This MS ring is integral to the inner membrane and also houses the FT3SS export

apparatus of the bacterial flagella. Upon the MS ring base, the internal C-ring (made up of FliGMN) forms, making a drum-shaped structure surrounded by the motor (Mot proteins) (McMurry *et al.*, 2015). This structure also houses the export apparatus (See Section 1.5.1.2 below), which enables secretion and assembly of the subsequent axial structures.

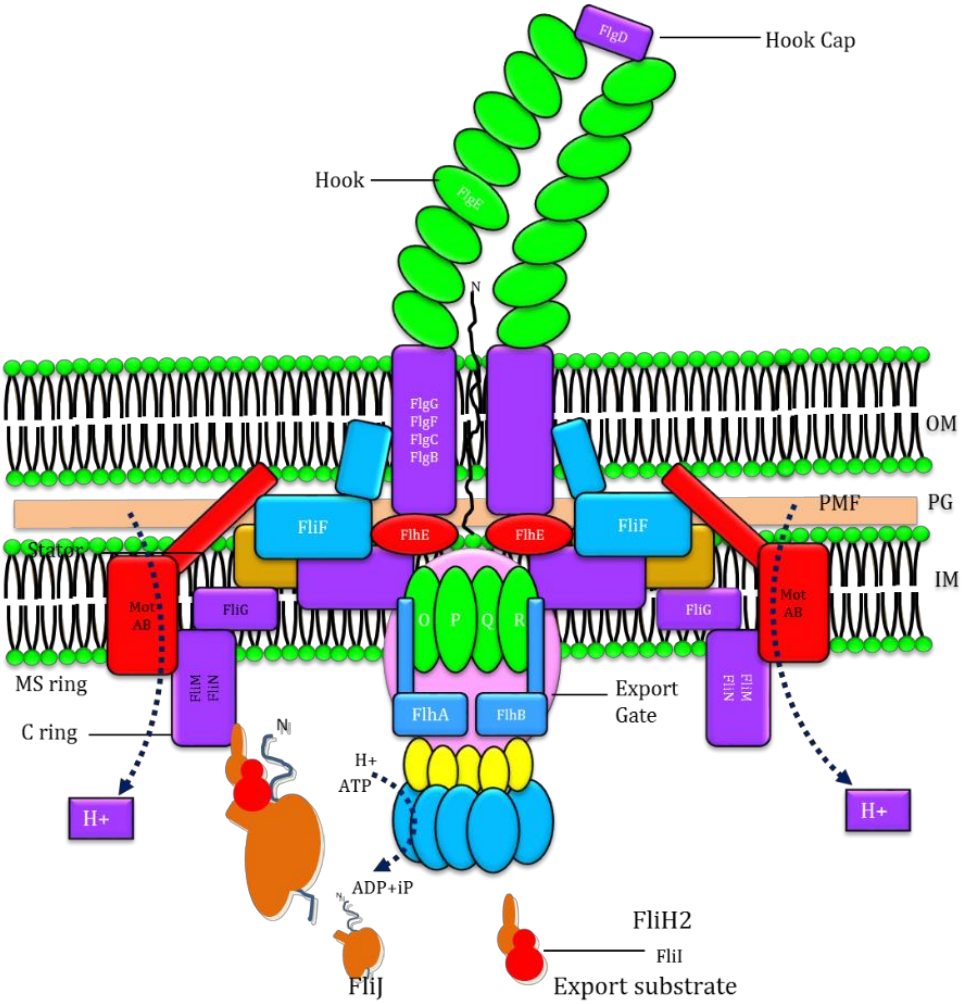


Figure 1.8: Flagella hook-basal body and the export apparatus.

Showing the bacterial flagella hook-basal body and the distribution of the proteins involved in its assembly. FlgD: hook cap protein, FlgE: hook subunits, OM: outer membrane, PG: peptidoglycan layer, IM: inner membrane. Adapted from (Aldridge and Hughes, 2002; Macnab, 2004; Samatey *et al.*, 2004; Matsunami *et al.*, 2016).

Once the MS ring, C-ring and the export apparatus is in place and functional, the proximal-distal rod proteins (FliE, FlgB, FlgC, FlgF, and FlgG) are secreted in order through the T3SS rod, with the FlgG protein assembling beneath its dedicated rod-cap foldase- FlgJ, a dual function protein that also

contains a β-N-acetylglucosaminidase activity (Fujii *et al.*, 2017). It enables degradation of the peptidoglycan layer and allows the rod to grow through the periplasmic space (Fig. 1.8) (Saijo-Hamano *et al.*, 2004; Herlihey *et al.,* 2014). Once the rod proteins are in the place, they allow the formation of P and L-ring protein assemblies in the peptidoglycan and outer membrane layers (composed of proteins FlgI and FlgH), respectively. These rings surround the rod proteins and provide a channel for both assembly and rotation, where P and L rings do not undergo signal peptide cleavage during the export process and therefore are not secreted by the general secretory pathway (Ohnishi *et al.*, 1994; Macnab, 2003). Once the rod is complete the FlgJ-cap is ejected and replaced by the hook foldase-FlgD, under which the hook subunits- FlgE polymerise to produce a curved hook that acts as a universal joint to convert the rotary movement of the straight rigid rod to the filament (Fujii *et al.*, 2017). The rod (FlgG) and hook (FLgE) protein subunits with 39% amino acid sequence identity are directly connected to each other but show distinct mechanical properties; as the rod is straight and rigid as a drive shaft whereas the hook is flexible in bending as a universal joint (Fujii *et al.*, 2017). The curved hook is formed using ~120 monomers of FlgE proteins which are assembled beneath the hook-cap protein (FlgD) and reaches upto ~55 ± 6 nm under the control of hook length regulator protein (FliK) (Fujii *et al.*, 2017). Once completed the hook cap is ejected and the flagella filament formed; however, this will be discussed further in Section 1.6.1.2, once we have examined the flagellar motor (next section) and export apparatus itself (Section 1.7.1.2).

1.7.1.2 The bacterial flagellar motor

Like any other rotary motor, the bacterial flagellar motor comprises of a rotor and a stator. The rotor is attached to the helical filament by a universal joint whereas the stator is anchored to the cell wall (Berg, 2008; Sowa and Berry, 2008). The C-ring proteins form the rotor/switch complex that contributes to control the direction of rotation while the motor proteins MotA and MotB form a $MotA_4$–$MotB_2$ complex attached to the C-ring *via* MotB, leading to the formation of a stator and proton channel in H^+ driven flagellar motors such as *E. coli* and *Salmonella typhimurium*, while the Na^+ driven motors of *Vibrio alginolyticus* and *Vibrio cholerae*, the stator is made up of PomA and PomB (Terashima *et al.,* 2008;

Sowa and Berry, 2008; Kojima, 2015). MotX and MotY are also present in Na^+ driven motors, but their function is unknown (McCarter, 2006). The unbound MotB blocks proton flow through the proton channel, but its collision with basal body motor drives a conformational change that removes the block initiating the proton flow (Hosking *et al.*, 2006). This allows proton transport down the electrochemical potential gradient through the cell membrane providing proton motive force (PMF) that drives secretion of flagella substrates and rotation of the flagella motor (Hosking *et al.*, 2006). The sodium motive force has also been recently shown to drive secretion (Terashima *et al.,* 2008; Minamino *et al.*, 2016). Torque generation of the flagella and protein secretion through the FT3SS is powered by ATP, proton and sodium motive forces (Fig. 1.9) under tight control and regulation, which is costly for the cell (Kojima and Blair, 2004).

1.7.1.3 The export apparatus of the Bacterial Flagella

The export apparatus in a T3SS is required for the translocation of most of the flagellar proteins that localise outside the cell membrane as well as the internal rod proteins and associated non-structural cap proteins and one regulatory protein (Minamino and Macnab, 1999). The flagellar export apparatus comprises of six transmembrane proteins (FlhA, FlhB, FliO, FliP, FliQ, FliR) and three soluble, membrane associated proteins (FliH, FliI, FliJ) (Macnab, 2003; Paul *et al.*, 2008). Within this structure, the transmembrane proteins are highly conserved and essential for export apparatus function (Paul *et al.*, 2008). Specifically, FlhA (75 kDa) is an integral membrane protein and forms a close contact with FlhB (42 kDa), which has a 19 kDa C-terminal cytoplasmic domain that functions in export substrate docking (Minamino, 2014). Key to the whole export apparatus function is the hexameric FliI ATPase (49 kDa) which forms a ring complex at the export gate together with FliH (26 kDa) and FliJ (17 kDa) and plays an important role in substrate recognition (Minamino, 2014). However, the functions of FliO (13 kDa), FliP (25 KDa), FliQ (10 kDa) and FliR (29 kDa) are not clear, except that they seem key for gating function. FliP has a N-terminal signal peptide required for its insertion into the cytoplasmic membrane where it is cleaved to produce mature FliP, it co-purifies along with FliR and the basal body, highlighting it as a key part of the complex (Minamino and Macnab, 1999; Hirano

et al., 2009). The MS ring and transmembrane proteins together form the export gate while soluble proteins associate with the MS ring, anchoring the export gate in the central pore of the MS ring by forming the channel through which remaining flagella proteins are exported (Yonekura *et al.,* 2003). In terms of secretion mechanism, the cytoplasmic domain of $FlhA_c$ forms the binding sites for chaperones, ATPase and the export substrates (Bange *et al.*, 2010). FlhB located in the export gate has two domains- membrane and cytoplasmic, which are connected with a flexible linker (Fig.1.9) (Williams *et al.*, 1996; Aldridge and Hughes, 2002; Hughes, 2014).

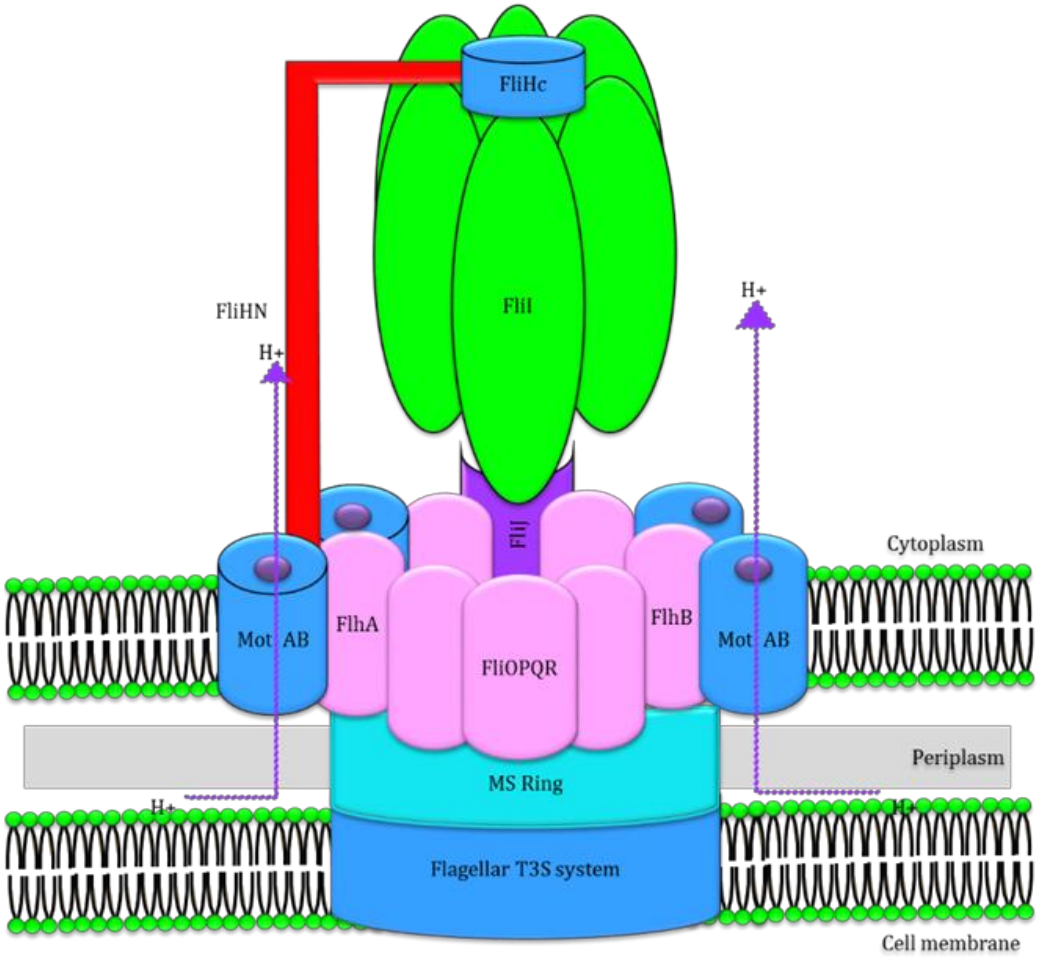


Figure 1.9: The export machinery of the flagellar type III secretion system. Showing the Flagellar Type III secretion apparatus and the distribution of the proteins involved. Transmembrane proteins FlhAB, FliOPQR and soluble FliI, FliH and FliJ are shown. For detailed molecular function of proteins, refer to section 1.5.1.3. Adapted from (Macnab, 2003b; Kojima and Blair, 2004; Minamino *et al.*, 2014).

Recently it has been observed that FlhA acts as a channel for the sodium motive force, but it is activated in absence of an ATPase complex, suggesting this dual power motor is effective in presence and absence of an ATPase complex (Minamino *et al.*, 2016). As previously stated, FliH inhibits the premature expenditure of ATP by the ATPase complex, while MotB blocks proton flow through the proton channel (Minamino *et al.*, 2003; Hosking *et al.*, 2006). The ATPase complex is not required for protein export; however it facilitates substrate association with the export gate and acts to unfold the proteins during this process (Minamino *et al.,* 2008).

1.7.1.4 Selective recognition of early and late substrates

A remarkable feature of flagellar-specific T3SS is not only its ability to secrete large quantities of proteins through a specialised export apparatus (see previous section), but also that this occurs in an ordered fashion with several regulatory mechanisms to modulate the process. One feature that ensures this is the selective recognition of proteins of different parts of the structure at different points in the assembly process by the export apparatus (Singer *et al.,* 2014). As part of this mechanism it is worth noting that the flagella protein substrate is divided into two specificity classes, named early substrates secreted before hook-basal body (HBB) completion and late secretion substrates secreted after HBB completion (Stafford *et al.*, 2007). On completion of hook assembly, hook length control protein (FliK) a molecular ruler sends a signal to membrane export component FlhB, resulting in a conformational change that switches export substrate specificity from early rod-hook substrates to late substrates, allowing secretion of the hook-filament junction and filament proteins (Williams *et al.*, 1996; Fraser *et al.*, 2003; Thomas *et al.,* 2004). The change is brought about by FlhB auto-cleavage that opens a region comprised of a basic stretch formed by α-helix and β-sheets adjacent to the NPTH loop. It is speculated that an acidic loop of FliK may 'wedge' into the cleft formed by the basic patch driving the substrate specificity switch (Chilcott and Hughes, 2000; Stafford *et al.*, 2007; Lorenz and Büttner, 2011; Mizuno *et al.*, 2011). The prevention of FlhB auto-cleavage maintains the secretion apparatus in an early locked state, which suggests that FlhB is necessary for the recognition of secretion substrate and secretion of late

substrates (Evans *et al.*, 2013). Following completion of the hook, the hook-filament junction protein FlgK is secreted through the hook displacing the hook-cap protein from the hook tip and permitting the assembly of the hook-filament junction. At the same time, secretion of the anti σ^{28} (FliA) factor FlgM is also facilitated. This releases free, uninhibited FliA in the cytoplasm (Fig. 1.10) (Hughes, 2014). Before this point in the assembly process, the FliA and FlgM are expressed in the cytoplasm with FlgM binding σ^{28} (FliA) and preventing its interaction with RNAP and thus repressing FliA-dependent transcription (Chadsey *et al.,* 1998; Erhardt and Hughes, 2010). These FliA dependent gene loci include all late flagellar proteins belonging to what is termed Class III flagellar loci, including (FliC) flagellin (Fig. 1.11) (Saito *et al.*, 1998). This intersection of FlgM with FliA also protects FliA from proteolysis (Barembruch and Hengge, 2007). This coupling of transcription and assembly shows an efficient ''just-in-time'' expression kinetics and allows the cell to respond quickly to hook completion and is discussed further in Section 1.7.2 (Fitzgerald *et al.,* 2014).

1.7.1.5 Secretion of early - late flagellar secretion substrates *via* the FT3SS

As outlined above, flagella assemble in a sequential order from early to late subunits. The early unchaperoned subunits interact with regulatory FliJ and ATPase FliI (Auvray *et al.*, 2001; Thomas *et al.,* 2004). The substrates are sorted for early and late export by means of their N-terminal secretion signals by docking at FliI ATPase where FliI partially unfolds the substrates (Stafford *et al.*, 2007). Once the hook is formed, hook-filament junction proteins (FlgK-FlgL) are exported replacing hook cap (FlgD). This junction is made up of ~13 molecules each in a matured flagellum with two proteins, FlgK (HAP1) and FlgL (HAP3) and their mutational studies suggest that the junction acts as a buffering structure and connects two filamentous structures together (hook and filament) with distinct mechanical properties (Ikeda *et al.,* 1996).

Finally, flagellin monomers are exported from the cytoplasm through the central channel and assembled at the distal end forming the filament cap protein FliD (Ohnishi *et al.*, 1994). However, unlike early substrate subunits secretion of the late subunits is chaperone dependent, where chaperones bind and present

the late substrates to the export gate and are recycled *i.e*. FlgN to FlgK and FlgL, FliT to FliD and FliS to FliC (Minamino, 2014). Specific cleavable signal peptides have not been observed for most of the secretion substrates where recognition by the export apparatus is believed to be achieved *via* a disordered N-terminal region that serves as secretion signal for protein export (Homma *et al.*, 1990). Deletion mutant studies have identified residues 26-47 within the N-terminal region of FliC-flagellin to be required for directed secretion, which was also reported for directed homologous protein secretion by means of fusion of the signal residues to the recombinant protein (Muskotál *et al.*, 2006). The FliD cap rotates as flagella monomers arrive at the tip aiding their polymerisation. About ~1000 monomers min^{-1} of flagellin are exported through flagella channel, where the filament is made up ~20,000-30,000 monomers (Erhardt *et al.,* 2010; Renault *et al.*, 2017). The export gate with protruding FlhAc domain binds chaperones, secretion substrates and ATPase forming 'export cage', where efficient secretion through FT3SS is accomplished (Minamino, 2014; Sajó *et al.*, 2014). The presence of FlhB domain in an export cage has two flexible domains that allow secretion (Fig. 1.10) (Fraser *et al.*, 2003; Hirano *et al.*, 2003; Meshcheryakov *et al.*, 2013). The ATPase complex (FliH, FliI, FliJ) is located beneath export cage initiates subunit export following the constitution of secretion structure in place with the help of FliJ bridging the gap between FliH-FliI and the FlhA (Abrusci *et al.*, 2013).

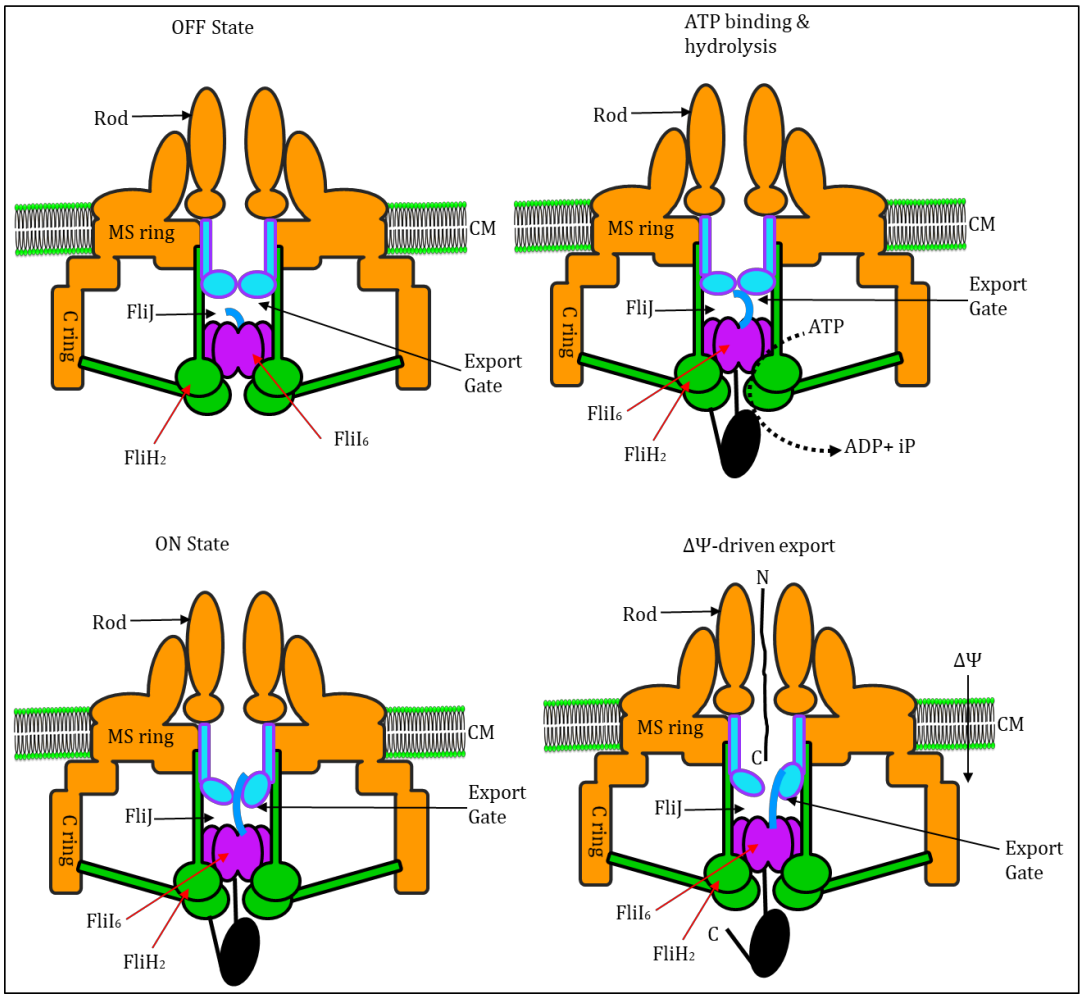


Figure 1.10: The export machinery of FT3SS in ON and OFF state.
Showing the Flagellar Type III secretion apparatus and the distribution of the proteins involved, in ON and OFF state, during ATP binding and hydrolysis. (See the text for explanation) Adapted from (Macnab, 2003; Minamino *et al.*, 2014).

1.7.1.6 Theories of Flagella-filament biogenesis

1.7.1.6.1 Chain mechanism of Flagellar elongation

It was observed that the N-terminal polypeptides of flagellar proteins retain the ability of secretion *via* flagellar export pathway into culture medium. A post-translational occurrence of flagellar protein export has also been reported, which indicates that the N-terminal amino acid sequences of substrates function as export signals and are recognised by FT3SS export apparatus (Minamino, *et al.,* 2014). The disordered N-terminal secretion signal and the mRNA secretion signal within 5'untranslated region (5'UTR) were reported for flagella assembly, however for late substrates, piloting proteins known as Type III chaperones were also required (Singer *et al.,* 2014). Once subunits are

secreted, they must travel through existing structure before they are crystallised at its tip. As the structure elongates, rate of protein export remains constant and filament can reach up to 20 μm in length (Turner *et al.,* 2012). It was thought that the constant rate of secretion is driven by head to tail linkage of FliC (Flagellin) and FlgE (Hook) protein monomers in a chain (Evans *et al.*, 2013). The chain is formed by linkages of N-and C-terminals of adjacent subunits enabling the transition of subunits through the existing flagella channel by a pulling or capillary force generated as the N-terminus of the foremost subunit shortens during polymerisation and crystallises at the flagella tip (Aldridge and Hughes, 2001; 2002; Minamino *et al.*, 2014). This leads to the pulling of subunits from the export apparatus into the existing flagella structure (Evans *et al.*, 2013).

1.7.1.6.2 Flagella filament elongation with diffusion mechanism

Recently a competing theory was put forward by Renault *et al.,* 2017. They reported the diffusion mechanism of flagella biogenesis using fluorescent labelling technique (Renault *et al.*, 2017). The reported initial rate of flagellum growth ~1,700 amino acids per second decreased with increasing length of filament and hence previously proposed chain mechanism does not contribute to filament elongation dynamics (Evans *et al.*, 2013; Renault *et al.*, 2017). It was reported that the inhibition of the proton motive force dependent export apparatus revealed a contribution of substrate injection/diffusion in driving filament elongation. The combination of experimental and mathematical evidence demonstrated the control of bacterial flagella growth outside the cell by simple, injection diffusion mechanism (Renault *et al.*, 2017).

1.7.2 The Flagellar gene hierarchy

As outlined above, flagellar biogenesis is a tightly regulated, energetically expensive process, in which flagellar genes are placed in an operon and assembled in transcriptional hierarchy, which imposes a huge burden on the cell (Chilcott and Hughes, 2000; Heel *et al.*, 2013). The hierarchy comprises three classes, with single Class I operon of *flhDC* sitting at the top of the hierarchy (Fig. 1.12). This atypical transcription factor forms a $FlhD_4C_2$ heterocomplex and acts as the master regulator of the flagella gene hierarchy (Soutourina and Bertin,

2003). It directs σ^{70}-RNA polymerase to transcribe from Class II promoters by binding to the upstream regulatory elements of Class II genes (Chilcott and Hughes, 2000). It results in expression of the Hook-Basal body (HBB) and regulatory proteins including sigma factor σ^{28} (FliA) and anti-sigma factor σ^{28} FlgM (Section 1.7.1) (Minamino and Macnab, 1999). As outlined above the expression of Class III genes is linked to the completion of the HBB and FliA promotes transcription of Class III genes, encoding filament, motor and chemotaxis proteins (Kutsukake and Doi, 1994; Claret and Hughes, 2000; Aldridge and Hughes, 2002).

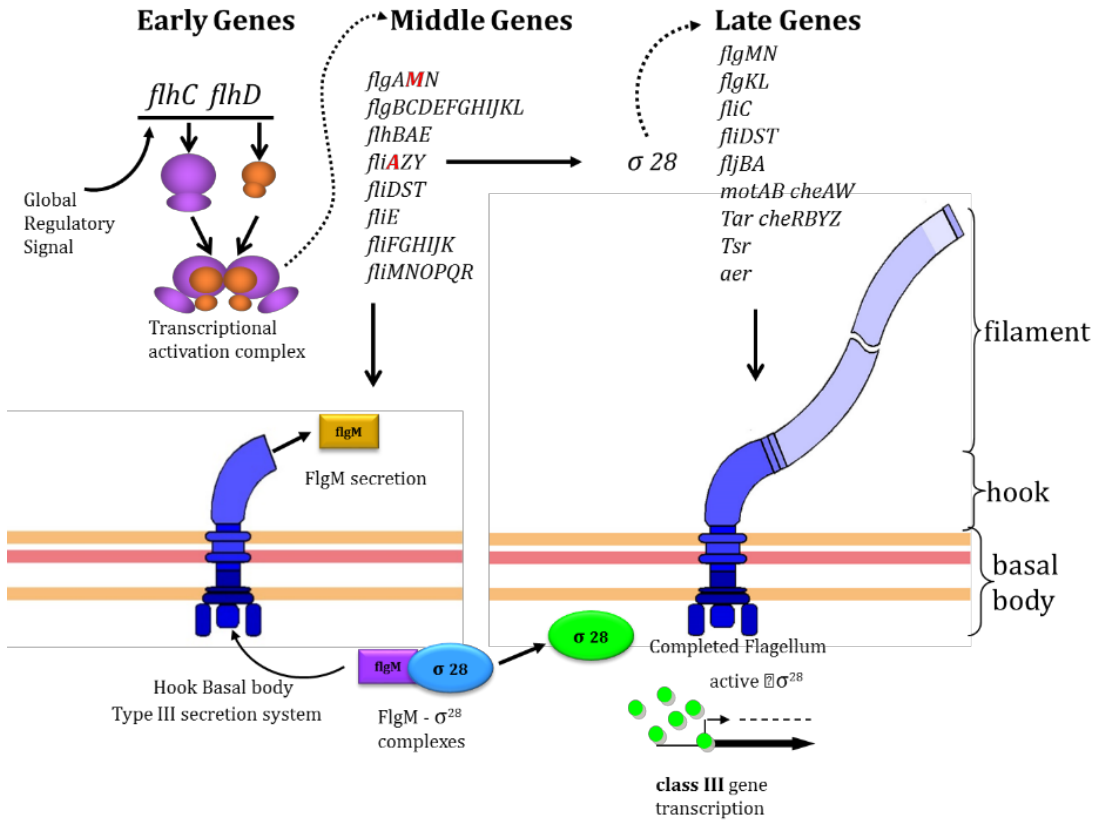


Figure 1.11: Flagella gene hierarchy and their assembly.
Showing the three classes of flagella genes and their corresponding morphological structures in flagella assembly. (Macnab and Parkinson, 1991; Namba and Vonderviszt, 1997; Aldridge and Hughes, 2002; Berg, 2008; Chevance and Hughes, 2008).

As might be expected, the expression of the master regulator, *flhDC* is regulated at a number of levels, allowing it to serve as an integrator of environmental, nutritional and growth phase signals by various proteins

expressed in response to environmental cues such as pH, temperature, nutrient availability and osmotic pressure (Soutourina and Bertin, 2003). FlhDC is an extensively studied flagella master regulator found in the members of phyla Proteobacteria *Salmonella typhimurium* and *E. coli* (Zhao *et al.*, 2013). The expression of *flhDC* is controlled by several other regulators (Fig. 1.13), including heat shock proteins DnaK, DnaJ and GrpE, which respond to changes in temperature (Li *et al.*, 1993); H-NS, proteins respond to changes in pH, OmpR in response to osmolarity, and cAMP-CAP in response to the availability of carbon sources (Kutsukake 1977).

It has been also reported that small non-coding RNA (snRNA) such as McaS and MicA positively regulate FlhDC expression, while ArcZ and OmrA act negatively (Lay and Gottesman, 2012; Jørgensen *et al.*, 2013). These regulatory proteins and snRNA may directly or indirectly affect transcriptional regulation of *flhDC* or may directly act on $FlhD_4C_2$ complex by inhibiting it from activation of gene expression of Class II promoter or may actively degrade the master regulator complex (Surette and Bassler, 1998; Habdas *et al.*, 2010; Lay and Gottesman, 2012; Jørgensen *et al.*, 2013). The master regulator transcriptional regulation is mediated through multiple sites in the large upstream region of *flhDC* promoter (Fahrner and Berg, 2015).

The FlhDC encoding mRNA has a conserved 5'-untranslated (5'UTR) sequence of ~198 nucleotides (Wei *et al.*, 2001). There are seven commonly accepted FlhDC-dependent operons (*flgAMN, flgBCDEFGHIJ, flhBAE, fliAZY, fliE, fliFGHIJK, fliLM-NOPQR*) that encode important regulatory factors and structural components of membrane-spanning basal body and associated export apparatus (Macnab, 1992). The feedback inhibition and activation of flagellar gene expression can be observed in which proteins produced due to *flhDC* expression in turn activate or repress master regulator or Class II or Class III protein expression (Fig. 1.14). These proteins are the protein chaperones FlgN, FliS, and FliT, where, FlgN regulate FlgM affecting Class III gene expression (Karlinsey *et al.*, 2000). The FliS-FliC complex binds to FlhA by the interaction of FliS with C-terminal cytoplasmic domain of FlhA (FlhAc), but neither interact with FliH, FliI

nor FliJ (Evans *et al.*, 2013). FliS also negatively regulates FlgM export as it binds to FlgM in the cytoplasm and suppresses its secretion following hook-basal body assembly (Yokoseki *et al.,* 1996; Delalez *et al.*, 2010), while FliT negatively regulates Class II gene expression *via* binding to $FlhD_4C_2$ master regulator (Yamamoto and Kutsukake, 2006).

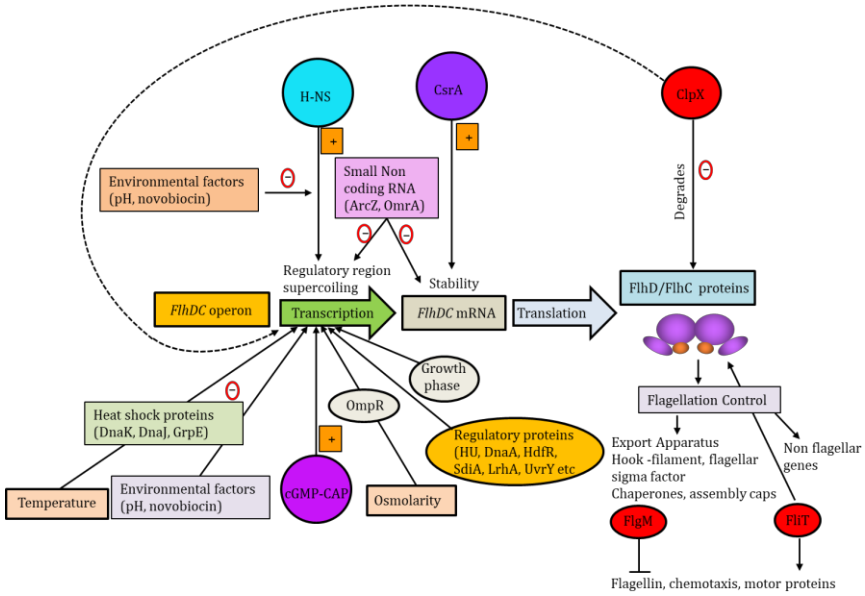


Figure 1.12: Multiple regulations of the *flhDC* flagellar master operon.
The main regulatory levels and effectors that are involved in the control of FlhDC synthesis are indicated: +: positive effect, –: negative effect, no indication, positive or negative effect. Adapted from (Soutourina and Bertin, 2003; Lay and Gottesman, 2012).

Alternatively, some proteins can directly digest the $FlhD_4C_2$ complex and these are listed in Fig. 1.13 (also see Appendix 2). Apart from Flagella biogenesis, $FlhD_4C_2$ complex, also regulates promoters of a wider regulon, as it binds to a multitude of promoter regions (Stafford *et al.,* 2005). Nevertheless, it has been shown to repress cell division, up-regulate galactose transport and downregulate aerobic respiration (Pru and Matsumura, 1996; Pruss *et al.*, 2001). The prime target of investigation into improved FT3SS performance could be the master regulator, FlhDC which forms a point of control over FT3SS. FlhDC could be investigated with respect to transcriptional regulators or proteases that respond to environmental cues and negative feedback from within the flagella regulatory hierarchy, e.g. *clpP* and *fliA* genes regulate flagellar gene expression, where ClpXP protease regulates proteolytic turnover of flagellar regulon master $FlhD_4$-$FlhC_2$ complex (Fig. 1.13) (Tomoyasu *et al.*, 2003).

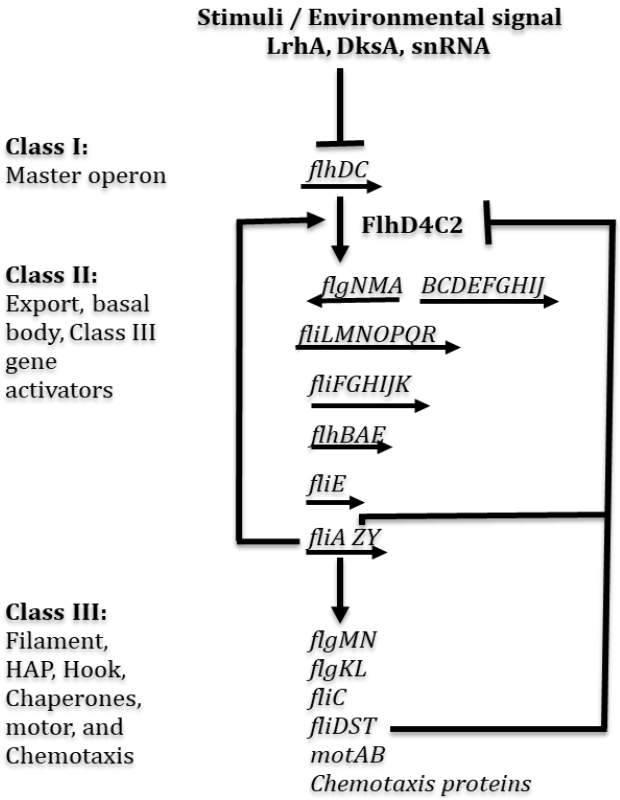


Figure 1.13: Flagella genes operons within regulatory classes.
Feedback inhibition is denoted by blunt-ended arrows. Adapted from (Karlinsey *et al.*, 2000; Aldridge and Hughes, 2002).

1.7.2.1 FliA dependent transcription

FliA triggers transcription of six commonly accepted Class 3 operons (*flgKL, fliDST, flgMN, fliC, tar-tap-cheRBYZ (meche)* and *motAB-cheAW (mocha*). In addition, it can also initiate transcription of its own operon *fliAZY* (Tanabe *et al.*, 2011; Fitzgerald *et al.,* 2014). FliA not only drives transcription of genes involved in chemotaxis (*trg*) and (*tsr*), aerotaxis (*aer*) and c-di-GMP regulation of motility (*yhjH* and *ycgR*) it also drives transcription of Class 2, *fliLMNOPQR* operon. FliA-dependent transcription of other Class 2 operons has also been suggested (Fitzgerald *et al.,* 2014). It also initiates transcription of Class III based genes that includes hook associated proteins (HAPs) FlgK, FlgL, FliD and 52 kDa FliC-Flagellin protein, components of motor and chemotaxis-related regulatory proteins (Daniell *et al.*, 2003; Fitzgerald *et al.,* 2014). Additionally, several well-characterized FliA-dependent promoters have been reported such as *modA* and *flxA* but no obvious function in flagellar synthesis or motility was observed. Bioinformatics studies have predicted FliA-dependent promoters based on

sequence identity but those with non-flagellar functions have not been tested experimentally (Huerta and Collado-Vides, 2003; Aldridge *et al.*, 2006).

1.8 Current scenario of recombinant protein secretion

As outlined above, around 80 % of the costs for production of recombinant proteins in the biomanufacturing industry are estimated to go into downstream processing (Hellwig *et al.*, 2004; Walsh *et al.,* 2006). However, if the recombinant proteins were secreted from the bacterial cell into the medium, downstream processing would be much easier, simple and cheaper, as the cells remain in culture medium without disruption, and proteins can be purified *via* simplified pipelines directly. In addition, since the cell continues to function the cellular factory producing the protein can continue to do so, unlike in a situation where cell lysis is required *i.e.* when proteins are produced in the cytoplasm or periplasm. Protein secretion doesn't allow inclusion body formation, reduce protein degradation and may aids in correct folding leading to structurally stable, robust and biologically active protein products (Baneyx, 1999; Mergulhão *et al.,* 2005). Secreted protein has N-terminal authenticity and as the secretion signal is cleaved following secretion, N-terminal methionine is lost, which affect biological properties of proteins, such as immunogenicity can be avoided (Baneyx, 1999).

Currently, in the biotechnology industry, secretion of recombinant proteins into the periplasm is often preferre due to the benefits arising from non-cytoplasmic protein expression such as confined small space of periplasm aids in initial isolation of protein. However, the protein must be retrieved and purified from the additional periplasmic proteins. The extracellular secretion of proteins is desired due to several advantages, such as improved yield and the quality of the proteins, easy downstream processing and cost-effectiveness. Current knowledge of secreted recombinant proteins relies on the presence of the secreted proteins rather than route of their secretion (Ni *et al.,* 2009). Arguably, protein secretion in *E. coli* is a complex process due to issues of incomplete secretion, incompatibility of secretion, low capacity of secretion, proteolysis and incapability of secretion of large proteins due to limited capacity of secretion system (Baneyx, 1999; Huang *et al.,* 2012). The protein expression and secretion

should be evenly balanced as if protein expression exceeds the secretion demand, intracellular accumulation of proteins occurs leading to inclusion body formation (Strandberg and Enfors, 1991; Schlegel *et al.,* 2013).

This thesis focuses on engineering T3SS of bacterial flagella for improving protein export. It has been previously attempted to modify it for protein secretion but with only a limited success. A recombinant lipase PelB (35 kDa) was reported to secrete through a flagella CAPless (Δ*fliCD*) mutant of *E. coli*, following 32 hr culture but without any evidence of how much cell lysis contributed to over-all 'secretion' (Narayanan *et al.,* 2010). Majander *et al.*, (2012) has reported the secretion of 12 mg L^{-1} for a 28 kDa protein Peb1 and the largest protein secreted was a 43 kDa Enolase. In another study, 25 amino acid long, sea snail venom fused with natively secreted protein and secreted *via Salmonella enterica* FT3SS (Singer *et al.*, 2012). Spider silk proteins have also been secreted through *Salmonella* injectisome by fusing with N-terminal secretion tag and a chaperone with the largest protein yield of 75 kDa and 6.9 nmol L^{-1} hr^{-1} (Widmaier *et al.*, 2009). A range of recombinant proteins with partial secretion tags have been successfully secreted such as human complement protein, GFP and highest yield reported was 1.8 $mgL^{-1}hr^{-1}$ (Muskotál *et al.*, 2006; Végh *et al.*, 2006). The details of recombinant protein production *via* FT3SS in terms of engineering of FT3SS and the utilisation of secretion signal will be discussed (Section 1.8.2) focusing on FT3SS regulation. The next section will emphasise why FT3SS was selected to form the basis of secretion platform with reference to recombinant protein secretion in this thesis.

1.8.1 Aspects of flagella T3SS secretion relating to biotechnological manipulation for secretion of protein

The first biological explanation suggesting that secretion of heterologous proteins and in fact recombinant protein secretion was the observation that broken flagella regrow after physical breakage- releasing misdirected monomers into the medium in the meantime (Stafford *et al.*, 2007). In addition, during the course of mechanistic studies, several mutations have resulted in the secretion of monomeric flagella proteins into the media. These mutations could be utilised for

further engineering of the FT3SS for improving protein secretion if used in concert with knowledge of secretion signals (See Section 1.7.1.5). For example, it has been shown that the flagella FT3SS gene deletions reported with respect to their functions, e.g. in *Salmonella,* were redundant among Gram-negative genera and hence could successfully be extrapolated to *E. coli*. The deletion mutations of flagella genes such as flagella cap, hook-filament protein or entire flagella, results in immotile mutants due to failure to polymerise functional flagella. Flagellar cap protein FliD helps in polymerisation of flagellar filament, where *fliC* monomers assemble under the pentameric cap. The *fliD* mutant does not assemble functional flagella in absence of flagella filament cap (Fig. 1.14) and hence flagellin monomers leak into the medium. It could still assemble the hook, as the hook cap protein was present, but it remained immotile (Yokoseki *et al.,* 1996). Motility could be restored to some extent by supplementing an exogenous source of flagellin into the medium, in which flagella filaments partially assemble upon the hook structures. It has been reported that *E. coli,* in absence of flagellin *fliC*, does not assemble filaments, and therefore remains immotile (Li *et al.*, 1993). The dramatic elevation in Class II promoter activity and gene expression can be observed in the filament cap mutant (Δ*fliD*) as *fliD* and *fliT* are in same operon, where *fliT* is the repressor of gene expression of master regulator. Due to absence of *fliT*, the FlhDC master regulator remains 'on', leading to increased Class II gene expression. Following double deletion of filament and filament cap proteins (Δ*fliCD*), mutants increased number of flagella basal bodies. It was reported due to absence of FlgM secretion competition through flagella channel (Singer *et al.*, 2012). The flagella filament junction protein mutants (Δ*flgKL*) do not assemble functional flagella filaments; instead the unpolymerized FliC monomers are secreted *via* flagella channel to the medium (Homma *et al.,* 1985). Additionally, increased levels of homologous proteins such as FlgM and FliD-filament cap were observed in Δ*flgKL* than wild type strain, suggesting that free FliD is not incorporated in Δ*flgKL* mutant flagella (Brown *et al.*, 2008).

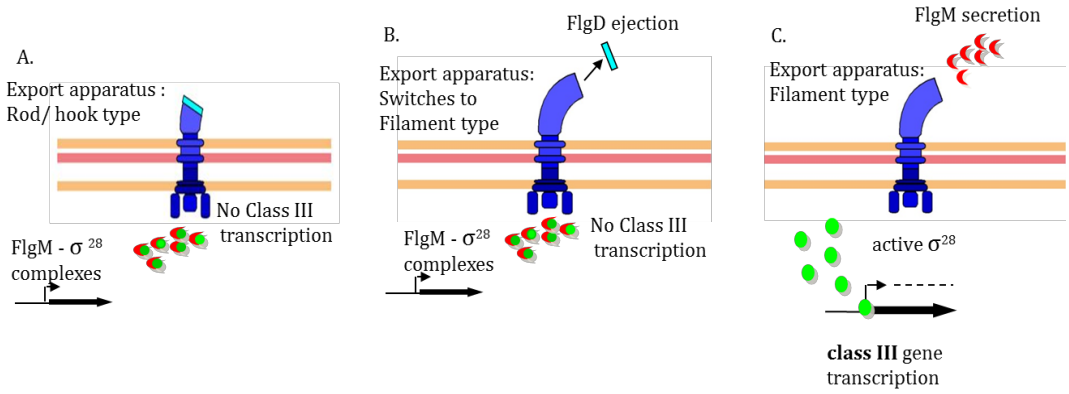


Figure 1.14: The switch from Class II to Class III flagella gene expression.

The flagella hook grows under the FliD cap. FlgM binds to σ^{28}, inhibiting it from initiation of transcription of class III genes (A.). Following completion of the hook, the hook cap FliD is ejected (B). FlgM is secreted, leaving σ^{28} free, to initiate the transcription of class III genes (C). Adapted from (Aldridge and Hughes, 2002; Hughes, 2017).

1.8.2 FT3SS and recombinant protein secretion

The engineering of FT3SS is an appealing model for protein production as FT3SS provides a high-efficiency, direct route of protein export from cytoplasm to the extracellular space. Specifically, ~30,000 monomers of flagellin (fliC) are exported from the cytoplasm to build the flagellum (Erhardt *et al.,* 2010). Considering 30 min generation time of *E. coli* and an export rate of ~1000 subunits min^{-1} (Renault *et al.*, 2017), it would equate to ~0.5 g L^{-1} hr^{-1} of protein. It could potentially outcompete the yields achieved by similar systems in yeast (*Pichia pastoris*) and *Bacillus* genera (Chen *et al.*, 2012; Love *et al.*, 2012). Therefore, this high efficiency secretion system could be utilised for protein secretion into the medium serving as an ideal model for several reasons; potentially providing protection from cellular proteases, high purity and simplified downstream processing (Singer *et al.*, 2012). The well-understood natural regulatory system for flagellar gene expression that exists provides control over flagellar manipulations. The control is rendered directly or indirectly by $FlhD_4C_2$ complex, which in turn is transcriptionally regulated by environmental cues. Efforts have been taken to convert FT3SS in to a protein secreting nanomachine, which mostly relies on the fact that -removal of filament and cap proteins result in a modified FT3SS capable of secreting heterologous proteins with considerable efficiency (Majander *et al.,* 2005; Narayanan *et al.,* 2010; Singer *et al.*, 2012). As mentioned, this thesis aims to investigate the

optimisation of an *E. coli* strain for directed secretion of proteins from cytoplasm to the extracellular medium by means of modification of FT3SS. Therefore, FT3SS will be engineered in such a way that it would serve as a platform secretion machinery for the export of a range of economically valuable biotherapeutics. A range of secretion signals were tested for directed secretion of proteins *via* the FT3SS in various knockout backgrounds, while in parallel, strain optimisation was carried out aiming to improve protein export output by controlling regulatory gene expression and reducing the metabolic burden by avoiding unnecessary gene expression. The secretion output was quantified by developing a fluorescence-based high-throughput assay and was validated by western immunoassay. Following establishment of these assays, protein secretion was tested by means of secretion construct harbouring cargo protein, comprised of secretion signal in N-terminus followed by cargo protein, detection and purification tags. The modular secretion construct could be modified for improving secretion, as it would allow alteration different secretion signals and cargo proteins. The modified machinery could be utilised for secretion of various proteins. The enzyme-based assay developed, would enable a high throughput screening of secretion by the allowing efficient protein export.

A range of proteins were secreted into the medium using Δ*fliCD* mutant, including Enolase, GFP, Peb1 adhesion, PalB lipase by fusing the target gene into 5' and 3' untranslated (5'UTR-3'UTR) region *fliC* (Majander *et al.*, 2005; Narayanan *et al.,* 2010). Protein secretion was improved by using FlgM secretion signal in a Δ*fliCD* mutant (Singer *et al.*, 2012). N-terminal FliC 47 amino acid secretion signal without 5'UTR was tested for protein export using Δ*fliC* mutant and it was reported that the first 47aa FliC signal is sufficient and it doesn't need the 5'UTR for protein secretion. Furthermore, it was reported that only residues 26-47aa of FliC contains the secretion signal and this was sufficient for directed protein export *via* the FT3SS of *Salmonella* (Végh *et al.*, 2006; Ve, 2010). Literature has been reported mostly on the use of late secretion signals for protein secretion, however secretion has been achieved to a certain extent but mostly using the intact hook. This project highlights the possibility of bypassing the intact hook by creating hook mutants and linking to protein secretion with

flagella early secretion signals such as FlgD-Hook cap and FlgE-hook subunit proteins discontinuing control rendered by FlgM on protein secretion and substrate switching. Previously, secretion has been reported by fusing proteins to early secretion signals-FlgD and FlgE in Δ*flgD* and Δ*flgE* mutants (Singer *et al.,* 2014). This will be considered and tested for protein secretion in different flagella mutant strains. Best secretion signal and strain combination will be further researched. Proteomics analysis (iTRAQ) will be performed for the identification of differential protein expression. In a parallel approach, a genomic library will be constructed and a plasmid library will be designed for the characterisation of regulatory genes that influence protein secretion *via* the FT3SS. A thorough literature search will be performed for the identification of non-coding RNAs (snRNA), which will be tested for protein secretion using an enzyme-based immunoassay.

1.8.3 Master regulator and its effect on protein secretion *via* the FT3SS

The FT3SS will be modified for the efficient protein production by means of engineering the basal body for protein secretion directly into the culture medium. The negative regulators of flagella gene expression will be removed for increasing the number of flagella. Abundance of $FlhD_4C_2$ would be increased by removing proteases of $FlhD_4C_2$ or by inhibiting Class II gene expression. Increased abundance of master regulator complex was expected to directly or indirectly promote expression of downstream flagella genes in flagella assembly. Gene deletions which resulted in increased *flhDC* expression or decreased $FlhD_4C_2$ proteolysis were reported in literature (Appendix 2). Some of these genes were studied previously in the group, hoping for improvement in protein secretion *via* the FT3SS. These gene deletions will be utilised which shows a great potential for modification of FT3SS for secretion. This secretion system is high throughput and well controlled using secretion signal. Genetic regulation, structure and function of FT3SS are well understood and as the gene regulatory network is well known- it is well suited for genetic manipulations. Previously, protein secretion has been achieved by the route of FT3SS modification with varying degrees of success. It encourages further engineering of FT3SS for improved protein secretion.

1.9 Aims and hypotheses of the study

In order to achieve the persistent goal of biotechnology industry a novel route of direct secretion of protein product from the cell cytoplasm to extracellular space was investigated by means of building on the re-engineered bacterial flagellar type III secretion (FT3SS) of *E. coli,* which secretes ~1700 subunits per minute of flagella proteins and renders the bacteria with motility. The aim of this research project was to create a 'Super-secretor' strain by further re-engineering FT3SS for improved protein secretion. However, the specific aims of the projects are reported as below,

Chapter 3: Investigation of the use of early, late or RNA-based signals (FliC 5'-UTR/ 3'-UTR) for directing secretion, through the Flagellar Type III secretion system (FT3SS)

Investigate secretion efficiency of early and late secretion signals in Hookless and substrate switched HAPless strains.

Optimisation of secretion assay and western immunoassay for the estimation of protein secretion.

Chapter 4: Mass spectrometry-based proteomics analysis of the strains for improving protein secretion *via* FT3SS with iTRAQ analysis

Investigation of protein regulation existed during protein secretion with systems biology approach, in order to understand the FT3SS for improving protein secretion.

Chapter 5: Construction of genomic library for the identification and screening of genes influencing protein secretion *via* the FT3SS.

Investigation of FT3SS with genomic screening to highlight the novel genes for improving protein secretion *via* FT3SS by constructing a genomic library.

Optimisation and refining of the secretion assay for library screening.

Chapter 6: Investigation of role of proteins and genes identified by systems biology, genome screening and preliminary investigation of SnRNA

Validation of systems biology and genomic screening findings highlighted in the thesis with recombineering.

Investigation of role of snRNA in improving protein secretion *via* FT3SS.

CHAPTER 2

MATERIAL AND METHODS

2 Material and methods

2.1 Strain

Table 2.1: Bacterial strains

Strains	Source
MC1000: F– Δ*(araA–leu)7697 [araD139] B/r*, Δ*(codB–lacI)3 galK16 galE15*(GalS) *λ– e14 relA1 rpsL150*(strR) *spoT1 mcrB1* (CGSC)	G. Stafford, UoS (Casadaban M.J., Cohen SN, 1980)
MG1655: F– LAM– rph–1 (CGSC)	Dr G. Stafford, UoS
MG1655 Δ*clpX:FRT – Km – FRT*	Prof. Jeff Green, UoS
(DH5α) NEB 5–alpha Competent *E. coli*: *fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi–1 hsdR17*	New England Biolabs
BL21 (DE3): *F–ompT gal dcm lon hsdSB (rB – mB–) λ (DE3 [lacI lacUV5–T7 gene 1 ind1 sam7 nin5])*	New England Biolabs
MC1000 Δ*flhDC::*FRT–*Km*–FRT	G P Stafford, UoS
MC1000 Δ*fliC::*FRT–*Km*–FRT	M. Hicks, GS lab
MC1000 Δ*fliCD::*FRT–*Km*–FRT	C.A. Green, GS lab
MC1000 Δ*fliC* Δ*flgKL* Δ*fliCD::*FRT–*Km*–FRT	Green *et al*, 2019
MC1000 Δ*fliC* Δ*flgKL::*FRT–*Km*–FRT	Green *et al*, 2019
MC1000 Δ*fliC* Δ*flgKL*	Green *et al*, 2019
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX::Km*	Green *et al*, 2019
MC1000 Δ*fliC* Δ*flgKL* Δ*motAB::*FRT–*Km*–FRT	Green *et al*, 2019
MC1000 Δ*fliC* Δ*flgKL* Δ*motAB::*FRT–*Cm*–FRT Δ*flgMN::*FRT–*Km*–FRT	Green *et al*, 2019
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX:: Km* Δ*motAB::*FRT–*Cm*–FRT	This study
MC1000 Δ*fliC* Δ*flgKL* Δ*flgDE::*FRT–*Km*–FRT	This study
MC1000 Δ*fliC* Δ*flgKL* ΔflgDE Δ*clpX::Km*	This study
MC1000 Δ*fliC* Δ*flgKL* ΔflgDE Δ*trg::Km*	This study
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX::Km:: ΔarcZ:: FRT – Cm – FRT*	This study
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX::Km:: ΔomrA:: FRT – Cm – FRT*	This study
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX::Km:: Δcrl:: FRT – Cm – FRT*	This study
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX::Km:: ΔclpB:: FRT – Cm – FRT*	This study
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX::Km:: ΔychR:: FRT – Cm – FRT*	This study
MC1000 Δ*fliC* Δ*flgKL* Δ*clpX::Km:: ΔgspL:: FRT – Cm – FRT*	This study

2.2 Plasmids:

Table 2.2: Plasmids used in this study.

Plasmids were purchased commercially, donated by The University of Sheffield (UoS) staff, previous colleagues in the group or generated during this study.

Plasmid	Description/ Features	Antibiotic Resistance	source

pTrc99a–	pTrc$_{99}$a with NdeI site in Multiple cloning regions. Empty Vector for protein expression.	Amp^r	Gillian Fraser (Oshima et al. 2006)
pTrc99a–E2	Referred to as pTrc$_{99}$a–E2 in text *fliC* Δ191–280 inserted between NdeI and BamHI of pTrc$_{99}$a–F	Amp^r	M. Hicks, GS lab
pJexpress–404–1–47 *fliC*	F*liC* 5'UTR– 1–47 secretion signal residues– XhoI–TEV EcoR1– PstI–TEV–FLAG tag– Strep tag– XbaI–HindIII– *fliC* 3' UTR, inserted between NdeI and BamHI of pJexpress–404.	Amp^r	M. Hicks, GS lab
pJexpress–404–1–47 C_H2	1–47 *fliC* C_H2 inserted between EcoRI and PstI of pJexpress–404–1–47 *fliC*	Amp^r	C_H2 insert amplified from plasmid from Dr Jagroop Pandhal, UoS
pJexpress–404–1–47 *fliC* C_H2 (no 5'UTR)	PCR amplified NdeI– 1–47 secretion signal residues– XhoI– TEV– EcoR1–C_H2– PstI TEV– FLAG tag– Strep tag XbaI–HindIII– *fliC* 3' UTR–inserted between *NdeI* and BamHI of pJexpress–404–1–47 *fliC*.	Amp^r	C A Green, Dr Graham Stafford, UoS
pJexpress–404–C_H2 (no 5'UTR)	PCR amplified NdeI–XhoI– TEV EcoR1–C_H2– PstI–TEV–FLAG tag–Step tag–XbaI– HindIII– fliC 3' UTR– BamHI inserted between NdeI and BamHI of pJexpress–404.	Amp^r	This Study
pJexpress–404–1–47 *fliC* Cutinase pSC O	Cutinase inserted between EcoRI and PstI of pJexpress– 404–1–47 fliC	Amp^r	This Study
pJexpress–404–1–47 *fliC* Cutinase (no 3' UTR) pSC 1	PCR amplified NdeI– *fliC* 5'UTR 1–47 of FliC secretion signal residues–XhoI–TEV–EcoR1–Cutinase –PstI–TEV–FLAG tag– Strep tag–XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404.	Amp^r	This Study
pJexpress–404–1–47 *fliC* Cutinase (no 5'UTR)	PCR amplified NdeI–1–47 secretion signal residues– XhoI–TEV–EcoR1–Cutinase – PstI–TEV–FLAG tag–Step tag–XbaI– HindIII– fliC 3' UTRBamHI inserted between NdeI and BamHI of pJexpress–404.	Amp^r	This Study

pJexpress–404–1–47 *fliC* Cutinase (no 5' or 3' UTR)	PCR amplified NdeI– 1–47 secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Strep tag–XbaI–HindIII–BamHI inserted between NdeI and BamHI of pJexpress–404.	Amp^r	Green *et al*, 2019
pJexpress–404–26–47 *fliC*Cutinase (no 5'UTR)	PCR amplified NdeI– 26–47 secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Strep tag– XbaI–HindIII– *fliC* 3' UTR BamHI inserted between NdeI and BamHI of pJexpress–404– 1–47 fliC	Amp^r	Green *et al*, 2019
pJexpress–404–26–47 *fliC*Cutinase (no 5' or 3' UTR)	PCR amplified NdeI– 26–47 secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Strep tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pJexpress–404–*fliC*–Cutinase (no 5'UTR)	PCR amplified NdeI– XhoI TEV–EcoR1– Cutinase – PstI TEV–FLAG tag– Strep tag–XbaI HindIII–fliC 3' UTR– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pJexpress–404– *fliC* Cutinase (no 5' or 3' UTR)	PCR amplified NdeI– XhoITEV–EcoR1– Cutinase – PstITEV–FLAG tag– Step tag–XbaI HindIII–BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pKD46	Expression of λ–RED recombinase system recombineering	Cmp^r	Barry Wanner, USA
pCP20	Expression of FLP recombinase for excision of FRT–flanked markers	Km^r	Barry Wanner, USA
pKD3	Template for overlap PCR and cassette construction	Km^r	Barry Wanner, USA
pKD4	Template for overlap PCR and cassette construction	Cmp^r	Barry Wanner, USA
pKD13	Template for overlap PCR and	Amp^r	

	cassette construction		Barry Wanner, USA
pKD32	Template for overlap PCR and Cassette construction		Barry Wanner, USA
pJExpress–404–1–100 FlgD Cutinase 3'UTR pSC 5	PCR amplified NdeI– *flgD*$_{100}$*aa* Secretion signal residues– XhoI–TEV–*Eco*R1–Cutinase – PstI– TEV– FLAG tag– Step tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pJExpress–404–1–50 FlgD Cutinase 3'UTR pSC 7	PCR amplified *Nde*I– *flgD*$_{50}$*aa* secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Step tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pJExpress–404–1–100 FlgE Cutinase 3'UTR pSC 9	PCR amplified *Nde*I– *flgE*$_{100}$*aa* secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Strep tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pJExpress–404–1–50 FlgE Cutinase 3'UTR pSC 11	PCR amplified NdeI– *flgE*$_{50}$*aa* secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Strep tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pJExpress–404–1–100 FlgD Cutinase No 3'UTR pSC 6	PCR amplified NdeI– *flgD*$_{100}$*aa* secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Strep tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study
pJExpress–404–1–50 FlgD Cutinase No 3'UTR pSC 8	PCR amplified NdeI– *flgD*$_{50}$*aa* secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Step tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJexpress–404–1–47 *fliC*	Amp^r	This Study

pJExpress–404–1–100 FlgE Cutinase No 3'UTR pSC 10	PCR amplified NdeI– *flgE*$_{100}$*aa* secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Step tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJExpress–404–1–47 *fliC*	Amp^r	This Study
pJExpress–404–1–50 FlgE Cutinase No 3'UTR pSC 12	PCR amplified NdeI– *flgE*$_{50}$*aa* secretion signal residues– XhoI– TEV– EcoR1– Cutinase – PstI– TEV– FLAG tag– Step tag– XbaI– HindIII– BamHI inserted between NdeI and BamHI of pJExpress–404–1–47 *fliC*	Amp^r	This Study
pACYC 184	pA15 ori,	Ter^r	Dr Mark Thomas, UoS
pNK15TcLib	p15AOri, MCS	Cmp^r,Tet^r	This Study
pNK15TcLib (Plasmids 1–3200)	P15A Ori, MCS, random Genomic DNA fragments of wild type of *Escherichia coli* MG 1655, terminator from BioBrick	Tet^r	This Study

2.3 Chemical reagent and buffers

All chemical reagents and buffers were purchased from Sigma-Aldrich and Fisher Scientific, unless otherwise stated in the text. All reagents used were of analytical grade and were purchased from the suppliers listed in (Table 2.3, Appendix 7). Buffers were prepared in distilled water and pH adjusted with HCl or NaOH. Solutions were sterilised by autoclave or sterile filtration (0.22 µm). The molecular biology kits were used according to manufacturers' instructions.

Table 2.3: Manufacturers and suppliers of chemicals kits and equipment.

Supplier	Location
BDH Laboratory Supplies	Poole, UK
Beckman Coulter	High Wycombe, UK
Bioline	London, UK
Bio Rad Laboratories	Hertfordshire, UK
BMG Labtech	Ortenberg, Germany
Calbiochem (Merck Millipore)	Watford, UK
Cell Signalling Technology	Danvers, Massachusetts, USA
Eppendorf	Hamburg, Germany
Expedeon	Swavesey, Cambridgeshire, UK
Fisher Scientific	Loughborough, UK
FlowGen Biosciences (SLS Life Sciences)	Hessle, East Riding of Yorkshire, UK

GATC Biotech	Konstanz, Germany
GE Healthcare Life Science	Buckinghamshire, UK
IBA Lifesciences	Goettingen, Germany
Life Technologies	Carlsbad, California, USA
Merck Millipore	Watford, UK
New England Biolabs (NEB)	Hitchin, Hertfordshire, UK
Norgen Promega	Southampton, UK
Sigma–Aldrich	Poole, UK
Statens Serum Institut	Copenhagen, Denmark
Syngene	Cambridge, UK
Tecan Group Ltd.	Männedorf, Switzerland
Thermo Scientific	Leicestershire, UK
Waring Laboratory Science	Torrington, Connecticut, USA
Xograph	Gloucestershire, UK

2.4 Molecular biology methods

2.4.1 DNA methods

2.4.1.1 Buffers and reagents

10 x TAE buffer

TAE buffer was used for running DNA Agarose gels. 10 X TAE buffer was prepared by dissolving, 48.4 g Tris base, 20 ml 0.5 M EDTA, pH 8.0 and 11.44 ml glacial acetic acid in up to 1 litre of dH_2O. 1 x TAE buffer was prepared in dH_2O for running the agarose gel.

6 x DNA loading buffer

6 x DNA loading buffer were used for preparing DNA sample for gel electrophoresis. It was prepared by using 60 ml glycerol, 6 ml Tris, pH 8.0, 1.2 ml 0.5 M EDTA, pH 8.0 and 60 mg Bromophenol blue in 100 ml in dH_2O.

2.4.1.2 DNA agarose gel electrophoresis

DNA analysis was carried out on 1 % (w / v) TAE agarose gels, unless otherwise stated. Agarose powder was dissolved in 1 x TAE buffer, added with traces of ethidium bromide (EtBr (0.3 µl)). DNA gels were set in casting trays. 6 x DNA loading buffer was added to DNA samples and run along with GeneRuler 1 kb DNA ladder (Thermo Scientific) or 100 bp DNA ladder (Bioline). DNA gels were run at 120 V until adequate separation of DNA bands was achieved. DNA was observed under UV light using G: BOX (Syngene Technologies). If required bands were excised using a scalpel. DNA sample was purified using ISOLATE II PCR and Gel Kit (Bioline) according to manufacturer's instructions, stored at –20 ^{O}C.

2.4.1.3 Bacterial plasmid DNA extraction

Plasmid DNA was isolated from liquid cell culture with Isolate II Plasmid Mini Kit (Bioline) according to manufacturer's instructions.

2.4.1.4 Rapid isolation of Miniprep DNA for double strand sequencing

2 ml of overnight cultures were resuspended in 130 µl of Solution I (50 mM Tris, 10 mM EDTA) and 260 µl of solution II (0.2 M NaOH, 1 % SDS). Eppendorf tubes were gently inverted to ensure complete lysis. 200 µl of Solution III (3 M KOAC, 2 M HOAC) was added, rocked (2-3 sec) and centrifuged for 2 min at 11,000 x g. 500 µl supernatant was collected, precipitated the nucleic acids with 1 ml of Isopropanol, mixed and centrifuged for 5 min at 11,000 x g. Pellet was washed with 70 % ethanol, centrifuged briefly to remove the residual ethanol; air dried for 2 min and resuspended in dH_2O. Incubated overnight and quantified by using Nanodrop 1000 spectrophotometer, stored at – $4^{O}C$ (Saunders and Burke, 1990).

2.4.1.5 Bacterial chromosomal DNA extraction

Chromosomal DNA was isolated from liquid cell culture with Wizard® Genomic DNA Purification Kit (Promega), according to manufacturer's instructions.

2.4.1.6 Method for the concentration of DNA samples

Dilute DNA samples, if required were concentrated by using 1: 10 volume of 3 M sodium acetate (pH 5.2) and 2-5 volumes of ice cold 100 % ethanol. Incubated at –20 ^{O}C for 1 hr to precipitate the DNA sample and recovered by centrifugation at 13000 x g in a microfuge for 15 min. Residual ethanol removed and pellet was washed with 70 % ethanol, air-dried DNA pellet resuspended in suitable volume of TE buffer, quantified with Nanodrop Spectrophotometer (Thermo Scientific).

2.4.1.7 Quantification of DNA concentration

Following calibration with nuclease free water, DNA concentration (ng μl^{-1}) in 1.0 µl isolated DNA solution measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

2.4.1.8 Polymerase chain reaction

Polymerase chain reaction (PCR) was carried out to amplify existing DNA samples. Primers were designed and synthesised by Sigma-Aldrich. Upon arrival

nuclease free water was added to dry oligonucleotides to result in 100 µM solution. Primers were stored at –20 ^{O}C. Primers used are listed (Appendix 3).

2.4.1.8.1 High-fidelity DNA polymerase PCR

High-Fidelity Phusion® DNA polymerase (NEB) was used to ensure the accuracy of DNA amplification for chromosomal mutagenesis, plasmid DNA cloning and DNA sequencing. PCR reactions were set up as shown in Table 2.4 and Table 2.5.

Table 2.4: Reagents for PCR with Phusion® High–Fidelity DNA polymerase

PCR reaction mixes were prepared in 50 µl. If multiple reactions carried out, for convenience master mixes were prepared using common reaction components.

Reagent	**µl**
N/F H_2O	9.7
5 x Phusion® High Fidelity Buffer	10
DMSO	0.1 (Optional)
DNA Template (10 µM)	0.1
Primer F (100 µM)	0.5
Primer R (100 µM)	0.5
dNTPS (100 µM)	0.5
Phusion® High Fidelity DNA polymerase	0.5

Table 2.5: Thermocycling for High–Fidelity DNA polymerase PCR reaction.

Different annealing temperature and duration were used depending on primers used. If no product was obtained following an annealing setting at 55 ^{O}C for 1.5 min, reactions were repeated with lower annealing setting while increasing time to 2 min. 30-35 cycles of the middle three steps were repeated. Extension time varied based on size of PCR product (30 sec / Kb).

Step	**Temperature (^{O}C)**	**Time (min: sec)**
Initial Denaturing	98	5: 0
Denaturing	98	0: 30
Annealing	A. 55 or B. 36	A. 1: 00 or B. 2: 0
Extension	72	0: 30 / kb product
Final Extension	72	7 – 10: 0

2.4.1.8.2 Overlap-extension PCR

Overlap -extension PCR was used to join DNA templates or introduce single point substitutions. High replication fidelity was achieved by using Phusion® High–Fidelity DNA Polymerase, as the product was incorporated into genomic DNA for transcription. Reactions were set up as described in Table 2.4; however, DNA

from multiple templates was added. The identical thermocycling conditions were used as mentioned (Table 2.5.).

2.4.1.8.3 Colony PCR

For the PCR based confirmation of chromosomal mutagenesis or vector DNA ligation (cloning) in individual *E. coli* colonies, colony PCR was utilised. The fidelity was not essential; hence, DreamTaq DNA Polymerase (Thermo Scientific) was used. PCR reactions which were set up as specified (Iulia *et al.*, 2013; Sheu, Wang and Lee, 2017) in Table 2.6.

Table 2.6: Reagents for PCR with Dream–Taq DNA polymerase.

PCRs were prepared in 20 µl PCR reaction mixes. The DreamTaq PCR Master Mix contains buffer, polymerase and dNTPs. If multiple reactions were carried out, master mixes were prepared using common components, for convenience. Instead of isolating DNA, a small scoop of DNA from a single colony was directly added to the PCR tube.

Reagent	**µL**
N / F H_2O	9.7
2 x DreamTaq PCR master mix	10
DMSO	0.1 (Optional)
DNA Template	N/A
Primer F (100 µM)	0.1
Primer R (100 µM)	0.1

Table 2.7: Settings for the PCR with DreamTaq DNA polymerase.

Based on the primer's features, annealing temperature and duration were set. If no product was obtained at an annealing setting of 55 ^{O}C for 30 sec, the reaction was repeated with a lower annealing setting while time increased to 1 min. The extension time varied based on the size of the PCR product. For colony PCR the initial denaturation time was extended to 5 min, to ensure cell lysis. 25-30 cycles of the middle three steps were repeated.

Step	**Temperature (^{O}C)**	**Time (min: sec)**
Initial Denaturing	95	5: 0
Denaturing	95	0: 30
Annealing	A. 55 or B. B. 36	A. 0: 30 B. 1: 0
Extension	72	1: 0
Final Extension	72	7 – 10: 0

2.4.1.9 Blunt end cloning

A storage vector CloneJET PCR Cloning Kit (Thermo Scientific) was used for blunt end cloning of linear fragments of DNA such as, synthetic genes or PCR products derived following overlap PCR. Efficient ligation of linear DNA into the vector was ensured by the induction of a suicide gene in any re-circularised vector. Blunt end cloning and the presence of proofreading enzymes ensured that the entire DNA fragment including ends was correctly incorporated into the CloneJet vector. Its transformation into suitable *E. coli* cells served to store DNA and as a simple means of DNA amplification by cell division. Insert DNA was recovered from the vector by either restriction digestion or PCR amplification.

2.4.1.10 DNA ligation

DNA fragments were cloned in vectors by restriction digestion followed by ligation. The insert DNA (donor) and cloning vectors (acceptors) were double digested with appropriate restriction enzymes in an appropriate buffer. Insert DNA was obtained either from a PCR or from plasmid DNA. Different phosphatases were used in restriction digest reactions to acceptor vector DNA to prevent self–ligation. The digestion reactions were set up as mentioned (Table 2.8). Reactions were incubated for 2 hr at 37 ^{O}C. Following incubation, 6 µl 10 x Thermo Trap Phosphatase reaction buffer and 2 µl ThermoTrap Phosphatase enzymes were added to acceptor vector DNA digest reactions and the mix was incubated for 1 hr at 37 ^{O}C. Following incubation, 5 µl of reaction mix was run on a 1 x TAE agarose DNA gel along with undigested donor DNA to compare digestion. Restriction enzymes in the vector DNA reaction were heat inactivated by incubating at 75 ^{O}C for 15 min. If required in exceptional cases, for example BamHI, inactivation was carried by clean up with the ISOLATE II PCR and Gel Kit (Bioline). The whole insert DNA reaction mixes were loaded on to a 1 x TAE agarose DNA gel, and DNA product of the expected size was excised and purified with the ISOLATE II PCR and Gel Kit (Bioline).

Table 2.8: Reagents and their quantities of restriction enzyme digest.

Reaction mixtures were set up in an Eppendorf tubes. The appropriate buffer for the two restriction enzymes was selected using the Double Digest Finder (NEB).

Reagent	**µl**
N / F H_2O	Up to a total of 50
Cutsmart buffer 10 X	5
R.E.1	1
R.E. 2	1
DNA polymerase enzyme	1 µg

DNA 1 µg total – for example 20 µl of 50 ng μl^{-1}

Following confirmation, vector and insert DNA were ligated using T_4 DNA ligase (NEB). Reaction mixes were set up as listed in Table 2.9 and incubated overnight at RT.

Table 2.9: Reagents and their quantities for DNA ligation reaction.

Reactions were set up in 1.5 ml Eppendorf tubes. The desired 1: 1 ratio of vector to insert sticky ends was achieved by adding them in 1: 3 ratios (NEBio Calculator), since vector DNA is bigger (Kbp) than insert DNA (bp).

Reagent	**µl**
N / F H_20	Up to 20
10 x T_4 DNA Ligase buffer	2
Vector DNA	1 part (as per requirement)
Insert DNA	3 parts (as per requirement)
T_4 DNA Ligase	1

3.0-3.5 µl ligation mixes were transformed into DH5-α competent *E. coli* cells (NEB) and screened by plating on LB agar plates supplemented with the relevant antibiotic. Colony PCR was used for the screening of successful transformants. Resultant positive PCR colonies were grown for plasmid extraction to undergo confirmatory restriction digest and sequencing.

2.4.1.11 DNA sequencing

The chromosomal DNA and plasmid DNA were sequenced using the LIGHTRUN™ sequencing service (GATC Biotech). The chromatogram of resulting nucleotide sequences was visualised using Finch TV/Snapgene (Fig. 2.1). Sequence coverage was achieved for around 1.0-1.4 Kb of DNA.

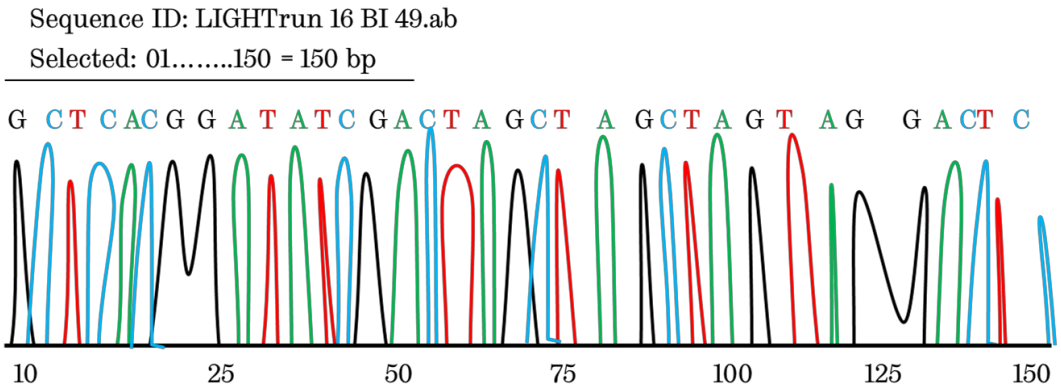


Figure 2.1: Sanger sequencing from GATC Biotech, Germany.
High Fidelity PCR, plasmids, cloning of DNA inserts etc. were verified by Sanger sequencing from GATC Biotech®, Germany. Peak color corresponds to the nucleotide. (Red: T, Green: A, Black: G, Blue: C). Clean sequence indicates the successful construction of the plasmid and the sequence was verified for mutations/error, manually by using SnapGene, online software program.

2.4.1.12 Gene synthesis

Synthetic genes were designed using online databases and sent for DNA String Fragments synthesis by GeneArt® synthesis, (Life Technologies). Upon arrival, dried DNA was resuspended in 50 µl nuclease free water. The blunt ended DNA fragments were cloned into a storage vector using the CloneJET PCR Cloning Kit (Thermo Scientific). The ligation mix was transformed into DH5-α, competent *E. coli* cells (New England Biosciences) and screened by plating on an appropriate antibiotic supplemented LB agar plate. Transformants were confirmed by colony PCR, restriction digestion and DNA sequencing.

2.5 Methods used for Genomic DNA Library preparation

2.5.1 Optimisation of Genomic DNA digestion

The *E. coli* MG 1655 genomic DNA was isolated from an overnight culture by using genomic DNA isolation as kit (Promega, Wizard®) according to manufacturer's instructions. The DNA Concentration was quantified by Nanodrop Spectrophotometer (Thermo Fisher) and if necessary was adjusted to 1200 ng μl^{-1}, by DNA concentration method (Section 2.4.1.6). The 100 µg of genomic DNA was digested using 1 µl (100 Unit) of 4-cutter enzymes. The reactions were incubated at 37 ^{O}C for 7 min, heat inactivated at 80 ^{O}C for 20 min, unless otherwise stated. The reactions were run on 0.85 % agarose gel, containing traces of EtBr, prepared in a 1 x TAE buffer and run at 45 V for 5 hr.

The continuous smear of DNA was visualised on G-box (SynGene Technologies) and desired bands in the range of 1.0 Kb-4.00 Kb were excised and purified by using Bioline PCR and Gel clean kit (Bioline). The genomic DNA digest would produce overhangs complementary to the ones created by 6-cutter restriction enzymes, designed into the multiple cloning sites (MCS) of $pNK_{15}TcLib$ plasmid.

2.5.2 Cloning of the random fragments of DNA into $pNK_{15}TcLib$ plasmid

The purified genomic DNA bands of expected size were excised; cleaned using PCR and Gel clean up kit (Bioline). The $pNK_{15}TcLib$ plasmids were digested with 6-cutter enzymes prior to cloning with cleaned fragments of genomic DNA. Genomic DNA fragments were ligated into the $pNK_{15}TcLib$, in 4: 1 ratio (NEB, Ligation Calculator) as shown in Section 2.4.1.10. The ligation reactions were incubated at 16 ^{O}C overnight. 3.5 µl ligation reaction was transformed in to DH5-$\propto$ cells and competent cells of the secretion strain, prepared by the Hahahan Method (Section 2.8). The transformation was carried out at 42 ^{O}C for 45 sec, recovered at 37 ^{O}C for 1 hr and plated on appropriate antibiotic selection plates. The colonies for the presence of the insert (random fragment of Genomic DNA) was checked and verified by Colony PCR and Digestion reactions.

2.5.2.1 Confirmation of clones in pNK plasmid by Colony PCR

The successful clones were verified by using colony PCR and Digestion. The PCR reactions and thermocycling condition were set as in Section 2.4.1.8.3.

2.5.2.2 Sanger sequencing of unknown nucleotide sequence

The positive colonies from colony PCR were Sanger sequenced by GATC Biotech, as mentioned before. The larger fragments were sequenced from either end of the clone by using appropriate forward and reverse primer sets using specific for the $pNK_{15}TcLib$ vector (Section 2.4.1.11).

2.5.3 Statistical Analysis

Each experiment was carried out in triplicate with three biological repeats unless otherwise stated. Statistical analysis used either student's t-test, One-way ANOVA for multiple comparisons, once normal distribution was determined and is detailed in the figure legend.

2.6 Methods used for non-coding RNA and recombineering Chapter 6.

2.6.1 Chromosomal DNA Mutagenesis

2.6.1.1 Chromosomal mutagenesis by Lambda (λ) Red recombination.

Chromosomal mutagenesis was achieved by a method reported by Datsenko & Wanner (2000). Chromosomal disruptions were carried out by Lambda Red recombinase induced homologous recombination, which include genomic insertions and knockout mutagenesis. For knockout mutagenesis, PCR products harboured FRT (FLP recombinase recognition site) flanked antibiotic resistance cassettes derived from template DNA of pKD3, pKD4, pKD13 or pKD32 plasmids and a ~ 35 bp homologous nucleotides to the regions adjacent to the gene, which was inactivated, were generated. If genomic insertions required, additional DNA (a gene or a promoter region) was also incorporated into the PCR product, adjacent to the antibiotic resistance cassette. Mutagenesis occured following electroporation of the PCR product into arabinose induced pKD46 lambda Red recombinase plasmid containing cells.

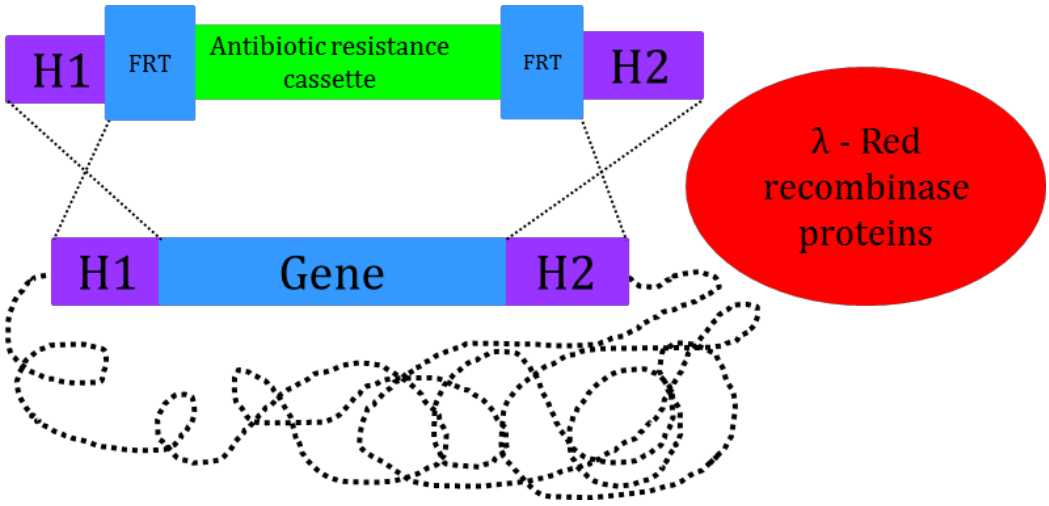


Figure 2.3: Chromosomal mutagenesis by with Recombineering.
Schematic of a cell harbouring plasmid derived Lambda Red recombinase proteins, PCR product template (top) and chromosomal DNA (bottom). The isolated PCR product construct contains the antibiotic resistance cassette, flanked with FLP recombinase target (FRT) sites, and homologous regions (H1 and H2) to chromosomal DNA adjoining the gene, which will be removed.

Bacterial cultures expressing the pKD46 lambda red recombinase plasmid were supplemented with 1 mM L-arabinose. Cultures were grown at 30 ^{O}C with agitation. Cells were transformed by electroporation with PCR derived linear FRT

flanked DNA and incubated for 1 hr at 30 ^{O}C for recovery and spread on LB agar plates supplemented with antibiotics. Following successful mutagenesis, the antibiotic resistance site could be removed by a FLP recombinase carrying helper plasmid (pCP20), which acts on the FRT sites. The red and helper plasmids were cured by growing cells at 43 ^{O}C, so that no further recombination events could occur. The antibiotic resistance gene flanked by FRT sites was removed from the chromosome of cells expressing the pCP20 plasmid by incubation at 43 ^{O}C.

2.6.1.2 Chromosomal mutagenesis by P1 phage transduction

To move a region of chromosomal DNA from one *E. coli* genome to another, P1 phage transduction was used, which utilises P1*vir* phage. P1*vir* is capable of packaging up to 90,000 kb DNA (Thomason *et al.,* 2007). It can be harnessed to transduce genes linked to selectable genetic markers (e.g. antibiotic resistance cassettes).

2.6.1.2.1 Phage lysate preparation

An overnight bacterial culture of the donor strain was diluted 1 in 100 in LB broth with 5 mM $CaCl_2$ and 0.2 % (w / v) glucose and incubated at 37 ^{O}C, 150 rpm for 1 hr. 100 µl P1 phage lysate was added to the culture and incubated for 1-3 hr or until the media was clear. Few drops of chloroform were added to the culture, vortexed and centrifuged at 13,000 x g for 2 min. The supernatant was transferred to a fresh tube and a few drops of chloroform added. The lysate was stored at -4 ^{O}C.

2.6.1.2.2 Phage transduction

An overnight culture of the recipient strain was centrifuged, and the pellet was resuspended in half the volume of LB broth with 100 mM $MgSO_4$ and 5 mM $CaCl_2$. 100 µl aliquots of culture were transferred into 1.5 ml eppendorf tubes. The eppendorfs contained (1) donor strain P1 lysate, (2) 50 µl donor strain P1 lysate while the negative controls contained (3) 100 µl recipient cells or (4) 100 µl donor strain P1 lysate only. Aliquots were incubated in stationary incubator at 30 ^{O}C for 30 min. Cells were recovered by adding 1 ml LB and incubation at 37 ^{O}C for 1 hr. Cells were spread on LB agar plates containing antibiotics and 10 mM sodium citrate, to chelate the calcium and inhibit reinfection (Murphy, 1998).

Successful transformants were passaged (3 x) on LB agar with appropriate antibiotic and sodium citrate to ensure eradication of P1 phage.

2.6.2 Strain storage

Following the confirmation of a novel strain or plasmid, cells were stored at –80 ^{O}C. Single colonies were picked from a LB agar plate and grown overnight in LB broth with relevant antibiotics. 1 ml medium was transferred to a cryotube with 0.5 ml 50 % (v/v) glycerol LB medium and stored at –80 ^{O}C.

2.6.3 Plate reader growth curve

For 96 well plate growth curve experiments, cells were incubated in the plate reader, Tecan (Magellan, Infinite® 200 PRO) at 37 ^{O}C, with 3 mm orbital agitation, O.D.600 readings were taken every 30 min.

2.6.4 Protein methods

2.6.4.1 Buffers and reagents

SDS-PAGE: Lower Tris buffer

For the preparation of SDS–PAGE resolving gel, 6.06 g Tris Base and 0.4 g Sodium dodecyl sulphate (SDS) were dissolved in 100 ml dH_2O, pH was adjusted to 6.8.

SDS-PAGE: Upper Tris buffer

For the preparation of SDS-PAGE stacking gel, 18.17 g Tris base and 0.4 g Sodium dodecyl sulphate (SDS) were dissolved in 100 ml dH_2O, pH was adjusted to 8.8.

10 x SDS-PAGE running buffer

SDS-PAGE running buffer was used for running SDS-PAGE gels using specific tanks. 30 g Tris base, 144.0 g glycine, 10.0 g SDS were dissolved up to a volume of 1000 ml with dH_2O. 1 x SDS-PAGE running buffer was prepared using dH_2O.

2 x SDS loading buffer

2 x SDS loading buffer were used for the preparation of protein samples for SDS–PAGE. It was prepared by using 100 mM Tris-Cl (pH 6.8), 4 % (w / v) SDS, 0.2 % (w / v) Bromophenol blue. Prior to use DTT was added to a final concentration of 200 mM from a 1 M stock solution.

Semi dry transfer buffer

Semi dry transfer buffer was used for protein transfer during Western blot. It was prepared with 2.9 g Tris base, 1.45 g glycine, 1.85 ml 10 % (w / v) SDS and 100 ml methanol and made up to 500 ml with H_2O.

10 x TBS buffer

TBS was used for the incubation of nitrocellulose membranes, during blocking and antibody exposure stages of Western blot. 24.0 g Tris base and 88.0 g NaCl was dissolved in dH_2O to a final volume of 1 litre. pH was adjusted to 7.6. 1 x solution of TBS was prepared prior to use in dH_2O.

2.6.4.2 Bacterial protein precipitation

Protein analysis was carried out using SDS-PAGE Western blot analysis. Bacterial cell cultures were centrifuged, and cell pellets were resuspended in 200 µl, 2 x SDS loading buffer per OD Unit of cells (1 ml cell culture at OD_{600} 1.0 is equivalent to 1 OD Unit). Samples were heated at 95 ^{O}C for 10 min. The supernatant from secretion assay experiments for total protein expression profile was precipitated with tri-chloro-acetic acid (TCA). 10 % (v / v) TCA was added to supernatant and incubated at 4 ^{O}C, overnight. The precipitate was recovered by centrifugation (4 ^{O}C) in a pre-calibrated microfuge while the supernatant was discarded. The pellet was resuspended in same volume of ice-cold acetone, vortexed and centrifuged. Pellet was collected and air dried. The pellet was suspended in 50 µl, 2 x SDS loading buffer/OD unit and heated at 95 ^{O}C for 10 min. The samples obtained from protein purification, 1: 1 ratio of 2 x SDS loading buffer was added and heated for 5 min at 95 ^{O}C.

2.6.4.3 SDS-PAGE gel

SDS-PAGE gels were prepared using casting kit of Mini-PROTEAN® (Bio-Rad). Long and short plates with 1.5 mM spacer were assembled on casting module. Resolving gel were prepared as listed in Table 2.10; and poured between the plates. An appropriate percent cross-linking was selected depending on size of the protein. Isopropanol was layered on top to level of the gel to avoid air bubbles and protect it from the air interface. Once the gel was set, isopropanol was removed and stacking gel (0.975 ml acrylamide, 2.1 ml 0.5 M Tris pH 6.8 (upper Tris), 4.725 ml dH_2O, 17 µl TEMED, 0.1 ml 10 % (w/v) ammonium persulphate)

was poured between the plates and comb inserted. Samples were loaded onto gels along with EZ-Run™ Prestained Rec protein ladder (Fisher Scientific) and run in a Mini-PROTEAN® Tetra Vertical Electrophoresis cell with 1 x SDS running buffer (Section 2.4.2.1). Gels were used for Coomassie staining (Section 2.6.4.4) or transferred on to Nitrocellulose paper for western blot analysis (Section 2.6.4.5).

Table 2.10: Reagents used for the preparation of SDS-PAGE gel.

Varying percent acrylamide was used in the resolving gel depending on the size of the protein of interest. 5 µl TEMED and 0.35 ml 10 % (w / v) ammonium persulphate (APS) were added to the reagents listed to initiate crosslinking.

% cross-linking	Acrylamide (ml)	Lower Tris (1.5 M, pH 8.8) (ml)	H_20 (ml)
10	2.475	2.5	4.825
12	3	2.5	4.3
15	3.75	2.5	3.55

2.6.4.4 Coomassie staining

SDS-PAGE protein gels were washed in dH_2O, 3 x times and strained with Instant Blue™ (Expedon), according to manufacturer's instructions. Gels were washed 3 x times in dH_2O before imaging on a scanner (Amersham).

2.6.4.5 Western blot analysis

SDS-PAGE gel was placed on a nitrocellulose membrane (GE Healthcare Life Science) and sandwiched between 3 layers of chromatography paper (GE Healthcare Life Science). The entire stack was saturated with 1 x semi-dry transfer buffer and assembled on a BioRad Trans-Blot Semi Dry Transfer Cell (Bio Rad Laboratories) and run at 20 V for 30 min. The nitrocellulose membrane was incubated in 50 ml 3 % (w / v) BSA in 1 x TBS / 0.1 % Tween or 5 % milk in 1 x TBS / 0.1 % Tween for 2 hr at RT or overnight at 4 ^{O}C. The membrane was washed with 1 x TBS / 0.1 % Tween, 3 x for 15 min. 20 ml 1 x TBS / 0.1 % Tween containing the primary antibody was added to the membrane for 1 hr at RT. Wash steps were repeated 3 x 15 min. The antibody incubation and wash steps were repeated if secondary antibody used. Antibodies were diluted in 1 x TBS / 0.1 %

Tween. The antibodies used are listed in Table 2.11. The membrane was soaked in Pierce® ECL chemiluminescent western blotting dark and light substrates (Thermo Scientific). Blots were developed on C-Digit, *LI-COR* western developing machine and analysed using *LI-COR,* Image studio (*Lite ver 5.2*). The blots were developed with scanning at high resolution on machine for 12 min.; alternatively, the blots were placed in cling film and taken into an X-Ray cassette folder (Kodak). Sheets of X-ray films, CL-Xposure™ (Thermo Scientific) were placed in the cassette on nitrocellulose membrane for varying time intervals (30 sec-10 min). Films were developed using a Compact X 4, X-ray Film Processor (Xograph). If required, membranes were stripped of antibodies and probed with alternative antibodies. Membranes were washed in 1 x TBS / 0.1 % Tween, 3 x 15 min and incubated in 10 ml Restore™ western blot stripping buffer (Thermo Scientific) for 15 min. Wash steps were repeated and the process could then be resumed at blocking stage.

Table 2.11: Antibodies used for western blot analysis.

The antibody used, manufacturer and concentration of antibody used in 1 x TBS / 0.1 % Tween are listed in Table.2.11

Antibody	**Manufacturer**	**Concentration**
H 48 Monospecific H Rabbit Antiserum, (Anti-FliC) *E. coli*	Statens Serum Institute	1 in 1000
Anti-rabbit HRP (horseradish peroxidase).	Cell Signalling Technology	1 in 3000
Monoclonal ANTI–FLAG® M2 antibody produced in mouse	Sigma-Aldrich	1 in 1000
Anti-mouse HRP (horseradish peroxidase)	Cell Signalling Technology	1 in 3000

2.6.5 Bacterial protein quantification by densitometry

A protein standard was run on SDS-PAGE gel to quantify relative concentrations of unknown protein samples. This was calculated by densitometry of western blot bands using ImageJ (http://rsbweb.nih.gov/ij/) software or *LI-COR* (Image studio *Lit ver* 5.2).

2.7 Microbiological culture methods

2.7.1 Bacterial culture

Luria Broth (LB) agar supplemented with necessary antibiotics (ampicillin: 100 μg ml^{-1} LB, kanamycin: 50 μg ml^{-1} LB, chloramphenicol: 25 μg ml^{-1} LB, tetracycline 100 μg ml^{-1}. All filter sterilised) were inoculated with *E. coli* either from cryo-vials or cell pellets following liquid culture. Cells were incubated at 37 ^{O}C overnight. The inoculated LB agar plates were stored at 4 ^{O}C for up to a month.

2.7.2 Bacterial cell cultures for secretion assay experiments

Individual *E. coli* colonies (3 colonies) for a secretion assay experiment were selected from LB agar plates and grown in 5 ml LB broth in a 20 ml tube, overnight at 37 ^{O}C. LB was supplemented with relevant antibiotics. Overnight starter cultures were used to inoculate fresh LB, supplemented with relevant antibiotics. Starter cultures were diluted 1 in 100 in 10 ml LB broth in a 50 ml Falcon tube. Exceptions to this include the inoculation of 100 ml of LB broth in 250 ml Erlenmeyer Flask for proteomics experiments, 500 ml LB broth in a 2.5 L conical flask for protein overexpression and purification, or 200 μl LB broth per well for 96 well plate for growth curve experiments. If necessary, additional supplements were added to initiate promoter induction of plasmids *viz.* included Isopropyl-β-d-galactopyranoside (IPTG) (Calbiochem) and L-arabinose (both filters sterilised 0.22 μm). Cells were cultured at 37 ^{O}C unless otherwise stated, with 180 rpm agitation (140 rpm for conical flask cultures). Optical density of cells was measured in a 1 ml cuvette at OD_{600} using a spectrophotometer (Jenway, 7315 UV / Visible).

2.8 The Hanahan method for the preparation and transformation of competent *E. coli* cells: High–efficiency transformation

The high efficiency of transformation was achieved by Hanahan method (Hanahan, 1983; Promega, 2012). Following buffers were used for making competent cells: 1. For instant use, and 2. For long term use. TFB was used when preparing competent cells for immediate use, while FSB was used to prepare stock of competent cells for long term use and stored at –80 ^{O}C.

2.8.1 Standard transformation buffer (TFB)

1 M MES was prepared by dissolving 19.52 g of MES in 80 ml dH_2O, pH was adjusted to 6.3 with 5 M KOH and dH_2O was added to bring it to final volume of 100 ml. The solution was sterilised by filtration through a disposable pre-rinsed Nalgene Filter (0.45 µm pore size) and divided into 10 ml aliquots and stored at – 20 ^{O}C. TFB was prepared by dissolving all the solute listed below in 500 ml of dH_2O and added 10 ml of 1 M MES (pH 6.3). Final volume of TFB to 1 L was adjusted with dH_2O.

Table 2.12: Solutes added to prepare TFB.

Reagent	Amount per Litre	Final concentration
1 M MES (pH 6.3)	10 ml	10 mM
$MnCl_2.4\ H_2O$	8.91 g	45 mM
$CaCl_2.2\ H_2O$	1.47 g	10 mM
KCl	7.46 g	100 mM
Hexamminecobalt chloride	0.80 g	3 mM
H_2O	To 1 litre	

The TFB was filter sterilised by using disposable 0.45 µm pore size filter. The solution was divided into 40 ml aliquots and stored at –4 ^{O}C.

2.8.2 Frozen storage buffer (FSB)

9.82 g of 1 M potassium acetate was dissolved in to 90 ml dH_2O, pH 7.5 was adjusted using 2 M acetic acid and final volume was adjusted to 100 ml using dH_2O and stored at –20 ^{O}C. FSB was prepared by dissolving all the solute listed below in 500 ml of dH_2O, pH 6.4 was adjusted using 0.1 N HCl. Final volume of FSB was adjusted to 1 L with dH_2O.

Table 2.13: Solutes added to prepare FSB.

Reagent	Amount per Litre	Final conc.
1 M potassium acetate (pH 7.5)	10 ml	10 mM
$MnCl_{2.4}H_2O$	8.91 g	45 mM
$CaCl_{2.2}H_2O$	1.47 g	10 mM
KCl	7.46 g	100 mM
Hexamminecobalt chloride	0.80 g	3 mM
Glycerol	100 ml	10 % (v/v)
H_2O	To 1 litre	

The FSB was filter sterilised by using disposable 0.45 µm pore size filter. The solution was divided into 40 ml aliquots and stored at –4 ^{O}C. Overnight cultures were transferred into sterile, disposable, ice cold 50 ml polypropylene

tubes and cooled to 0 ^{O}C, by keeping on ice for approx. 10 min. The cells were recovered by centrifugation at 27000 x g (4100 rpm) for 10 min at 4 ^{O}C. Supernatant was discarded and cell pellet was air dried. The pellet was resuspended by gentle swirling and vortexing in 20 ml (per 50 ml tube) of ice-cold TFB or FSB transformation buffer and kept on ice for 10 min. The cells were recovered by centrifugation at 2700 x g for 10 min at 4 ^{O}C. Supernatant was discarded and cells were resuspended by swirling or gentle vortexing in 4 ml of ice-cold TFB or FSB and aliquoted in 100 µl in eppendorfs.

2.9 Methods of transformation

2.9.1 Transformation of competent cells by heat shock

Heat shock method was used for the plasmid transformation into DH5-α cells or competent *E. coli* cells. Fresh cultures in LB broth were grown as described (Section 2.7.1) to O.D.$_{600}$ 0.4-0.6, in 10 ml culture and incubated on ice for 20 min. The cultures were centrifuged at 4 ^{O}C for 15 min in pre-equilibrated centrifuge, at 3500 rpm. The pellet was resuspended in same initial volume (10 ml) of 0.1 M ice-cold $CaCl_2$ and incubated on ice for 10 min. The centrifugation step was repeated with half the volume of $CaCl_2$ and finally the pellet was resuspended in 1 ml 0.1 M $CaCl_2$. 3.5 µl plasmid DNA (20-80 ng μL^{-1}) was transferred into 100 µl cell aliquots in 1.5 ml eppendorfs, incubated on ice for 30 min and heat shocked at 42 ^{O}C in a pre-set water bath for 45 sec and incubated on ice for 5 min. 1 ml warm LB or SOC medium was added and cells were incubated to recover at 37 ^{O}C for 1 hr. Cells were plated out on LB agar plates supplemented with antibiotics. Alternatively, for high efficiency of transformation, at the final resuspension step, cell pellet was resuspended in 1 ml 0.1 M $CaCl_2$, 10 % (v/v) glycerol. 100 µl cell aliquots in eppendorfs were stored at –80 ^{O}C. The protocol was resumed following defrost of the aliquots on ice for 10 min.

2.9.2 Transformation of competent cells by electroporation

Depending on the requirement, electroporation was used for plasmid transformation into *E. coli*. Fresh cultures of *E. coli* were grown (Section 2.5.1) to O.D.$_{600}$ 0.4–0.6, in 10 ml LB and incubated on ice for 20 min. Cultures were centrifuged in a pre-chilled (4 ^{O}C) centrifuge for 15 min at 3500 rpm. Cell pellet

was resuspended in 10 ml ice cold dH_2O. This step was repeated three times with half volume of dH_2O. Following the final centrifugation, cells were resuspended in up to 1000 µl dH_2O and aliquoted in 1.5 ml eppendorf tubes. 3.5 µl purified PCR product or plasmid DNA (20-80 ng μL^{-1}) was added to 100 µl cell aliquots and transferred to a pre-chilled electroporation cuvette (Flowgen Biosciences). Cells were electroporated with a BIO-RAD MicroPulser series 411 BR (BioRad Laboratories) using the settings (EC 1 setting, PL 5, potential difference: 2.5 kV, resistance: 200 Ω and capacitance: 25 µF). Cells were allowed to recover in 1 ml LB for 1 hr at 37 ^{O}C and plated out on LB agar plates supplemented with relevant antibiotics. For the storage purpose of the cells, at the final resuspension step, cell pellet was resuspended in 10 % (v/v) glycerol. 100 µl cell aliquots were transferred to eppendorfs and stored at -80 ^{O}C. The protocol was resumed following defrost of the aliquots on ice for 10 min.

2.9.2.1 Cell lysis

Cell lysis was required for the purification of intracellular proteins. Cell cultures were centrifuged, supernatant was discarded, and cell pellet was resuspended in the appropriate buffer for subsequent protein purification steps. Cells overexpressed with streptavidin II tagged proteins were disrupted (3 x) by french pressure at 1000 psi (French Press G-M®), incubated on ice for 1 min between each application. For the preparation of intracellular protein fractions for proteomics (iTRAQ) studies, cells were lysed by 10 x 10 sec repetitions of sonication with 30 sec incubation on ice between each sonication steps. Glass bead (Sigma Aldrich) beating technique was also used for cell disruption (Scientific Industries™ Genie Disruptor).

2.9.2.2 Concentration of protein

Either prior to or following protein purification, Amicon® Ultra 0.5 ml Filters (Merck Millipore) were used for the concentration of proteins by size exclusion. Protein sample (up to 500 µl) was applied to the column, centrifuged at 13,000 rpm, 4 ^{O}C until the desired volume of sample was achieved, followed by the addition of an alternative buffer and further concentration allowed buffer exchange using this system. Vivaspin 20 Protein Concentrator Spin Columns,

MWCO 10000 (GE Healthcare Life Science) were utilised, for larger volumes of protein, allowing the concentration of volumes up to 20 ml.

2.9.2.3 Protein purification

Protein was purified for further use as a protein standard for western blot analysis. For the purification of Streptavidin II tagged protein from the secreted fraction of cell culture, protein was suspended in 100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA pH 8.0 and purified using *Strep*-Tactin® Spin Column (IBA Life sciences) and eluted with 2 mM D-Biotin in the appropriate buffer. The corresponding intracellular protein was also purified using this method. The intracellular fraction of Streptavidin II tagged protein was purified on a 1 ml StrepTrap HP column (GE Healthcare) according to manufacturer's instructions. Protein was suspended in 100 mM Tris-HCl, 150 nM NaCl, 1 mM EDTA, pH 8.0 and applied to the column. The column was coupled to a P-1 peristaltic pump (GE Healthcare Life Science). Protein was eluted using 2.5 mM desthiobiotin. Eluted protein sample was dialysed overnight to yield purified protein in 100 mM Tris-HCl, pH 8.0. For SDS-PAGE protein standards, protein was prepared in 2 x SDS loading buffer (Section 2.6.4.2), aliquoted and stored at –20 ^{O}C.

2.9.2.4 Quantification of protein concentration

Concentration of purified protein was quantified using either Bradford assay (Sigma Aldrich) or Pierce™ BCA Protein Assay Kit (Thermo Scientific) in a 96-well plate according to manufactureres instructions. Absorbance of the reaction was measured using plate reader (Tecan Magellan, Infinite® 200 PRO). Results were plotted on a scatter graph.

2.9.2.4.1 Protein quantification using BCA Assay

Standard concentrations of Bovine Serum Albumin (BSA) were prepared from the 2000 µg ml^{-1} stock of BSA (Sigma Aldrich), ranging from 25 µg ml^{-1} to 2000 µg ml^{-1} in TEAB buffer (500 mM TEAB, 1: 20 (v / v) of 2 % SDS). The working BSA reagent was prepared by using 50: 1 (v / v) of reagent A: B. 25 µl of standards and unknown samples were incubated with 200 µl of BSA working reagent and incubated at 37 ^{O}C for 30 min. Plate was cooled to RT and read on Plate reader, (Tecan Magellan, Infinite® 200 PRO) at 562 nm.

2.9.2.4.2 Bradford Assay

Bradford assay for the estimation protein concentration was performed by plotting standard curve using BSA. Stock solution of 14 mg ml^{-1} of BSA was used for the preparation of standard dilutions, 0.25 µg $µl^{-1}$, 0.50 µg $µl^{-1}$, 0.75 µg $µl^{-1}$, 1.0 µg $µl^{-1}$, 1.25 µg $µl^{-1}$, and 1.40 µg $µl^{-1}$. 1 ml each of Bradford reagent was added to 1.5 ml eppendorf containing 30 µl of standard dilution, 0.25 µg $µl^{-1}$ to 1.40 µg $µl^{-1}$, respectively. Blank was prepared by using only 30 µl of TEAB buffer. Samples were diluted prior, 5 x times using TEAB buffer. 6 µl of concentrated protein and 24 µl of TEAB buffer were added to 1 ml of Bradford reagent. All the reactions were incubated at RT for 20 min. Absorbance was read at 595 nm using spectrophotometer (Jenway 7315 UV/Visible).

2.10 Assays to quantify the capacity of FT3SS for protein secretion.

2.10.1 Protein secretion assay

Bacterial culture (Section 2.7.1) prepared in LB medium was inoculated with overnight cell culture and supplemented with antibiotics and IPTG, OD_{600} was measured, cell and supernatant were separated by centrifugation at 3500 x g for 15 min, prior to sample preparation for SDS-PAGE and western blot analysis. The cells were directly added with 2 x SDS-loading buffer, while supernatants were precipitated with 10 % Trichloroacetic acid (TCA) (Sigma-Aldrich) and the precipitates were washed with chilled acetone, air-dried and added with 2 x SDS-loading buffer, prior to boiling at 95 ^{O}C for 15 min and western blot analysis carried out as described in Section 2.6.4.5.

2.10.2 Motility assay

Motility of bacterial cell was determined by measuring the zone of bacterial swimming that arises on a motility agar plate following inoculation of an individual colony. An isolated colony was picked from LB agar plate using sterile pipette tip and inoculated into a motility assay agar plate (0.85 g LB agar in 100 ml dH_2O). Plates were incubated at 30 ^{O}C in a stationary incubator.

2.10.3 MUB assay for Cutinase activity

Previously developed 4-MUB assay was further optimised to measure the activity of secreted Cutinase. 10 ml cell cultures of bacterial cells were induced

with 0.05 mM IPTG, grown to OD_{600} ~ 1.0 and supernatant collected with centrifugation at 4 ^{O}C, 13,000 rpm for 15 min. 900 µl supernatant was removed for assaying. (N.B. a sterile filtration technique had been tested by colleagues in the group, but it was not more effective at removing cells than centrifugation). LB broth was prepared in tandem to facilitate the removal of background interference by subtraction. Supernatants were normalised to OD_{600} with the addition of sterile LB broth. Solutions of 4-MUB substrate were prepared fresh. 250 mM MUB substrate was dissolved in dimethylformamide (DMF) with 1 % Triton X-100. Mixture was diluted in 0.05 M phosphate citrate buffer (pH 5.0) to a concentration of 100 µM MUB. 100 µM solution of 4-MU was also prepared similarly, to serve as a positive control. 25 µl prepared cell supernatant was added to 96-well plate and reaction was initiated with addition of 160 µl, 100 µM MUB and 15 µl of dH_2O. Additionally, following controls were used (1) LB and substrate, (2) substrate only, (3) LB and phosphate citrate buffer (4) phosphate citrate buffer only and (5) 4-MU to aid the removal of background interference through subtraction and to serve as a positive control. Plates were gently rocked for 1 min to mix solutions and incubated at 37 ^{O}C for 30 min. Fluorescence was measured, either by imaging 96 well plates under UV light (G: BOX, Syngene) or plate reader (Tecan Magellan, Infinite® 200 PRO). Following excitation and emission scans fluorescence were recorded at 302, 446 nm.

2.10.4 Assay development for the screening pNK genomic Library

A modified MUB assay was developed for the screening of pNK_{15}TcLib library. Genomic DNA of *E. coli* MG 1655 was randomly digested with 4-cutter enzymes and cloned in to pNK plasmid (Section 2.5.2) to create library of pNK plasmid that contained random fragments ranging from 1.0-4.00 Kb. Library of individual plasmid (1-3200) were transformed and co-expressed into the secretion strain (harbouring pSC 1). Secretor strain *E. coli* MC 1000 Δ*CKL* Δ*clpX*, harbouring pSC 1, expresses Cutinase from T_5 promoter which is exported in medium *via* FTTSS. Secretion strain co-expressing two different plasmids and proteins from different promoters were selected on Ampicillin and Tetracycline (100 µg $µl^{-1}$) plates. Cells were grown for assay in 96-well plate, overnight at 40 rpm at 37 ^{O}C, in 200 µL LB broth, supplemented with antibiotics. Expression cultures were prepared by

inoculating 200 µl of fresh LB broth with 10 µl of overnight culture and incubated for 4 hr in low shake at 37 ^{O}C. Their supernatants for MUB assay were collected by centrifugation (4 ^{O}C) at 3500 rpm using 96-well plate (Greiner) (Section 2.10.3). Assay plates were read at 302, 446 nm. Significantly changed strains were selected using original master plate and grown in 10 ml liquid broth for secretion assay (Section 2.7) and MUB assay was performed to cross verify the secretion. Their plasmids were isolated using Plasmid II miniprep kit (Bioline) (Section 2.4.1.4). pNK$_{15}$TcLib were Sanger sequenced (Section 2.4.1.11) to identify unknown genes incorporated using nucleotide BLAST searches.

2.11 Methods used for proteomics analysis

2.11.1 iTRAQ analysis pipeline

2.11.1.1 Isobaric tag for relative and absolute quantification (iTRAQ)

For iTRAQ studies, starter cultures were grown overnight in 10 ml culture, supplemented with antibiotic at 37 ^{O}C in shaking incubator at 180 rpm. Following day, three sets (3 x 3) of 250 ml Erlenmeyer flask containing 100 ml of autoclaved LB media were inoculated with 1 ml of overnight culture and 100 mg ml^{-1} antibiotic, induced with 0.05 mM IPTG, and incubated for 4.00 hr or until O.D.$_{600}$ reached 0.9-1.0 at 180 rpm. The supernatant was separated by centrifugation at 3500 x g (Beckman Avanti 25.50) in 50 ml Falcon tube, cells were collected in 2 ml eppendorf tubes (Sigma Aldrich) and stored at –20 ^{O}C.

2.11.1.2 Protein extraction

For protein extraction from whole bacterial cells, 500 mM tri-ethyl ammonium bicarbonate buffer (TEAB) pH 8.5 ± 0.1, containing 1: 20 of 2 % SDS and protease inhibitor tablet was added to low bind eppendorf tubes and equal volume of glass geads (v/v) (200-300 µm, Sigma Aldrich) were added to protein samples. Samples were agitated using a cell disruptor (Scientific Industries™ Genie Disruptor) for 10 times with 45 sec for each disruption and 3 min incubation on ice. Samples were then sonicated in an ultrasonic bath on ice, 3 min for 3 x times, with 2 min gap between each application. The unbroken cells and insoluble material including glass beads were removed by centrifugation at 17,000 x g for 15 min. Pellet was discarded, supernatant collected in 1.5 ml low

bind eppendorfs and stored at –20 ^{O}C. Soluble cell protein extracted was quantified with Bradford Assay (Section 2.9.2.4).

2.11.1.3 Protein digestion and labelling with iTRAQ reagents

100 µg protein samples were dissolved in 1 M triethylammonium bicarbonate, pH 8.5. Total protein concentrations were then reassessed using Bradford quantification assay. Samples were digested with 1: 50 ratio of sequencing grade Trypsin (Promega), incubated at 37 ^{O}C for 16 hr. Samples were then aliquoted in 8-identical fractions for 8-plex labelling. Labelling was performed according to manufacturers protocol (Applied Biosystems).

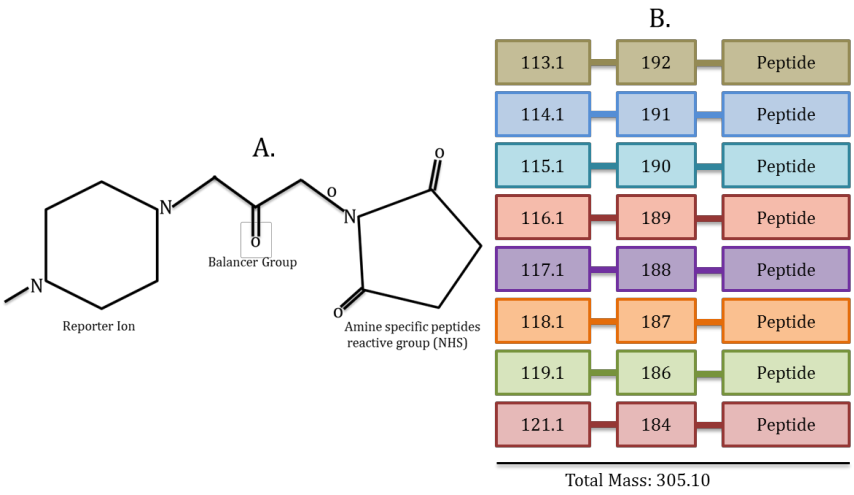


Figure 2.7: Structures of iTRAQ reagents

A. Showing the chemical composition of iTRAQ reagent. B. The 8–plex iTRAQ labels and their balancer group that show the equivalent masses, attached to the unique peptides *via* amide specific reactive group of iTRAQ reagent.

2.11.1.4 Protein fractionation with Hypercab

The dried labelled peptides (for iTRAQ) were resuspended in buffer A (3 % acetonitrile (ACN) and 0.1 % TFA) prior to fractionation with a 4.6 x 200 mm polyhydroxyethyl A column (5 µm, 200 Å, Hichrom Limited, UK), coupled with an HPLC 3000 system (Dionex, Germany), which was operated at a flow rate of 0.2 ml/min with UV detector used to monitor peptide abundance at a wavelength of 214 nm using a gradient with buffer B, containing 97 % ACN with 0.1 % TFA. The gradient was performed as follows: 10 min of buffer B before ramping up to 20 % of buffer B for 5 min, then up to 60 % of buffer B for 50 min, then ramped up

to 100 % of buffer B for 10 min and kept for 10 min, and finally 0 % of buffer B for 5 min. All the eluted peptides were collected every 3 min concentrated.

2.11.1.5 MS analysis

The selected peptide fractions were re-dissolved in 20 µl of buffer A consisting of 0.1 % formic acid (FA) in 3 % ACN and combined into 4, 8 and 12 different fractions prior to 3 µl of sample withdrawn and submitted on to a Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer coupled with a nano uHPLC 3000 system (Dionex, UK), operated at a flow rate of 0.3 µl / min. Peptides were separated using a C 18 column with a 105 min gradient of buffer B (0.1 % FA in 97 % ACN) as follows: 3% for 5 min, then ramped up to 10 % for 5 min, 50 % for 75 min, 90 % for 1 min, and then kept at 90 % for 4 min before ramped back to 3 % buffer B for 1 min and then finally maintained at 3 % for 14 min. The MS was operated in positive mode with resolutions of full MS and MS2 set at 60,000 and 15,000 respectively. AGC targets were set at 3.106 and 5.104 for full MS and ddMS 2, respectively; maximum IT times were set at 100 and 20 ms for full MS and ddMS2 respectively; a full mass scan ranging from 375 to 1500 m z^{-1} was applied for MS while the mass scan of 100–1500 m z^{-1} was applied for ddMS 2; default charge state of ion was set at 2. All raw data files from MS analysis were submitted to MaxQuant for protein identification against *E. coli* database (consisting of 4267 entries) downloaded in August 2016 from Uniprot. For iTRAQ technique, modifications of iTRAQ 8-plex reagents (on N-terminal and Lysine residue) and MMTS were set as fixed modifications while methionine oxidation (M) was set as variable modification; Trypsin digestion used with max missed cleavages of 2; minimum peptide length of 6 and maximum peptides mass of 4600 Da were set; tolerance of 20 and 4.5 ppm were applied for MS and MS/ MS, respectively. A False Discovery Rate (FDR) of 0.01 was used for identification of both peptides and proteins; a minimum score of 40 was used for modified peptides. All peptides containing iTRAQ intensities (MaxQuant output) were submitted to an in-house proteomic pipeline for quantification of proteins and determination of regulated proteins (Pham *et al.*, 2010). Only proteins identified with ≥ 2 unique peptides were used for quantification with median and isobaric corrections applied and paired t-tests (α = 0.05) were performed to determine

regulated proteins for each phenotypic comparison. Quantified proteins with *p*-values from each t-test ≤ 0.05 were considered as regulated.

2.11.1.6 t-test comparisons between the reporter ion intensities

For the analysis of two conditions in comparison *i* and *j,* each one corresponding to a specific reporter ion and a given protein for which N distinct MS/MS scans were obtained. The corrected intensities of the k^{th} MS/MS scan are denoted by I^k_i205710332057 and I^k_j205710332057. It was tested whether the distribution of *p*-values log (I^k_j / I^k_j) 205710332057 had a mean different from zero. We report the p-values associated with the said distribution (one-sided test). It was performed for every protein and because of the multiple times each test was performed, the threshold (α = 5 %) used for significance had to be corrected for data mining using a Bonferroni correction (α/*p,* where *p* is the number of proteins). Since 2,3, and 3 replicates are available from each condition (iTRAQ reporters 113, 114 in comparison A, 115, 116, 117 in comparison B and 118, 119, 121 in comparison C), a change is reported only if it is significant regardless of which replicate is chosen to perform the analysis. In practical terms for a given protein, eight or nine *p*-values are computed, based on the following t-tests: 113:118, 113:119, 113:121, 114:118, 114:119, 114:121 or 118:115, 118:116, 118:117, 119:115, 119:116, 119:117, 121:115, 121:116, 121:117, respectively. It was required that all the *p*-values be less than $\alpha/p = 3.54 \times 10^{-5}$ for data mining or α = 0.05 for confirmatory tests and the probability scores were defined as $-\log_{10}(p)$ to measure significance.

2.11.1.7 Bioinformatics analysis of the genes and proteins highlighted

The following software were utilised for data analysis, Cytoscape (https://cytoscape.org/), Stepdb (http://stepdb.eu/), David (https://david.ncifcrf.gov/), Panther (http://www.pantherdb.org/), Ecocyc (https://ecocyc.org/), NCBI Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

CHAPTER 3

INVESTIGATION OF THE USE OF EARLY, LATE OR RNA-BASED SIGNALS (5'-UTR) FOR DIRECTING SECRETION, THROUGH THE MODIFIED FLAGELLAR TYPE 3 SECRETION SYSTEM (FT3SS)

3 Introduction

As mentioned in Section 1.5, the mature bacterial flagellum is composed of a hollow (~2 nm) proteinaceous channel through which flagella proteins are exported prior to assembly beneath the pentameric FliD cap (Moriya *et al.*, 2006; Inaba *et al.*, 2013). Similarly, as outlined in Chapter 1, the subject of this thesis is the understanding and continued improvement of a prototype FT3SS platform strain for secretion of heterologous proteins out into the medium (supernatant). The key to the adaptation of the FT3SS for secretion is the optimisation of several parameters of both the cell but also specific events within the secretion process. One aspect is the signals used by the cell to direct proteins for FT3SS secretion. In this context, it is established that the FT3SS substrate proteins contain a signal sequence within the N-terminal 100 amino acids (Singer *et al.,* 2014). This rule holds true for both early (rod, hook) and late substrates (filament, FlgK) signals (Stafford *et al.*, 2007). The late native substrates are also bound by chaperone proteins that improve the efficiency of secretion due to the stabilisation of substrates in a secretion competent form, but also in some cases play a role in modulation of regulation either directly on the master regulator FlhDC or indirectly *via* FlgM/FliA or RNA interactions (Thomas *et al.,* 2004; Akeda and Galán, 2005; Stafford *et al.*, 2007).

A previous student in the group had established the optimal signal parameters for FliC secretion (See Fig 3.6) and directed export of proteins in a prototype strain (Green, 2016). In that study, the FliC N-terminal 47 aa was fused to its natural 5'UTR followed by the protein of interest (cargo) along with the FliC-3'UTR, where the 3'UTR proved detrimental to the secretion (Green *et al.*, 2019). These signal fusions were expressed in a strain lacking the hook-adaptor proteins (FlgKL), that were termed HAPless with secretion measured by western blotting using an anti-Flag tag antibody or *via* an assay of fused and exported Cutinase activity in supernatants, which was further refined in this project. While having achieved secretion at 0.16 mg/L, any protein that is exported *via* this system needs to travel (~100-120 nm) to reach the point of secretion into the supernatant (Green *et al.*, 2019). However, it was hypothesised that the reduction

of the distance needed to travel simply by use of an early signal in an early secretion locked strain might be more efficient than using 'late' FliC as a signal in a substrate switched strain. Another factor in using the early substrate is that, the time and energy used by the cell may be reduced since the need for substrate specificity switching or assembly of the hook might be negated if an early signal were used in conjunction with a strain where the hook and hook-cap proteins were removed genetically to creare a 'hookless' strain (Fig. 3.1).

Aim: The specific aim of this chapter was to investigate secretion efficiencies of Early and Late proteinaceous signals in appropriate flagella mutant backgrounds while also considering whether the FliC-5'UTR alone could serve as an efficient signal.

Results

3.1 Construction of secretion platform strains

As outlined above, the first aim in this chapter was to test whether the FlgD/E early signals in 'hookless' strains might be more efficient signals for the export of proteins than the Late-5'UTR-FliC$_{47}$aa signal, established in our lab in a HAPless strain (Green *et al.*, 2019). The strains to be used for these studies are described in Fig. 3.1. It is also of note here that previous work in Dr. Stafford's lab established that the inactivation of the *clpX*P complex *via* deletion of *clpX* gene improved the secretion of the HAPless-FlgKL strain (Green *et al.*, 2019). As a result, comparisons in this chapter needed not only the constructions of a Hookless strain but one that also lacked the *clpX* gene. Therefore, in order to construct this strain, Δ*fliC* Δ*flgKL* Δ*flgDE* strain was used as a recipient in a P1 phage transduction experiment using *E. coli* MG1655 *clpX*:: Km strain, with Km^R selection marker (*donated by Prof J. Green*). As a final stage to verify the true genetic pedigree of all these strains and to ensure no undesired replacement of WT, *fliC, flgKL* or *FlgDE* genes had occurred; these alleles were established in comparison to the parental strains using colony PCR, as mentioned previously.

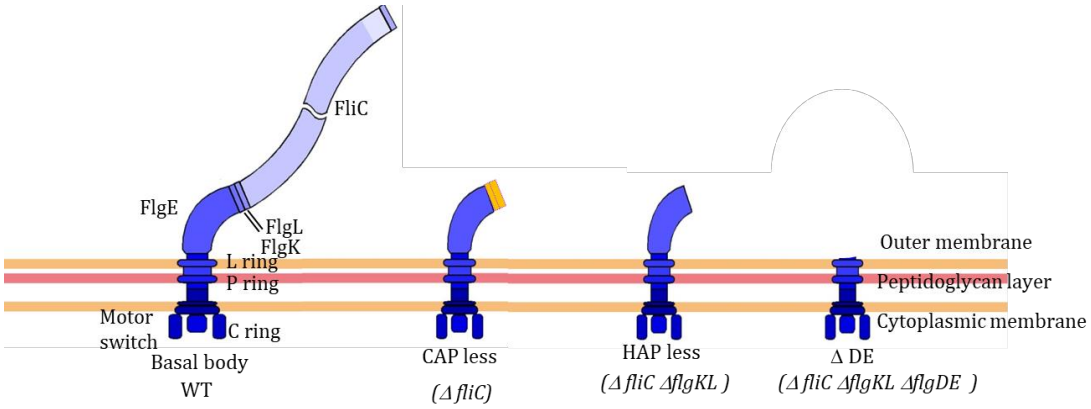


Figure 3.1: Structures of mutants used for protein secretion *via* the FT3SS.

Components of the flagella structures in different mutant strains. From left to right, these are Wild type (WT), CAP-less (Δ*fliC*), HAPless (Δ*flgKL*) and Hookless (Δ*flgKL* Δ*flgDE*).

The successful construction of these strains was verified by colony PCR using specific primer sets (Appendix 3: primers), for the *clpX* flanking regions, with a wild type *clpX* gene giving a product of (~1275 Kb) and a Km^R inserted allele at ~2.5 kb as shown in Fig. 3.2. The wild-type *E. coli* (MC 1000) was used as a control for the *clpX* allele while MC 1000 Δ*fliC* Δ*flgKL* Δ*clpX* was used as positive control for *clpX*::Km with the Δ*fli*C Δ*flgKL* Δ*flgDE* Δ*clpX* clearly indicating the presence of the *clpX*::Km allele.

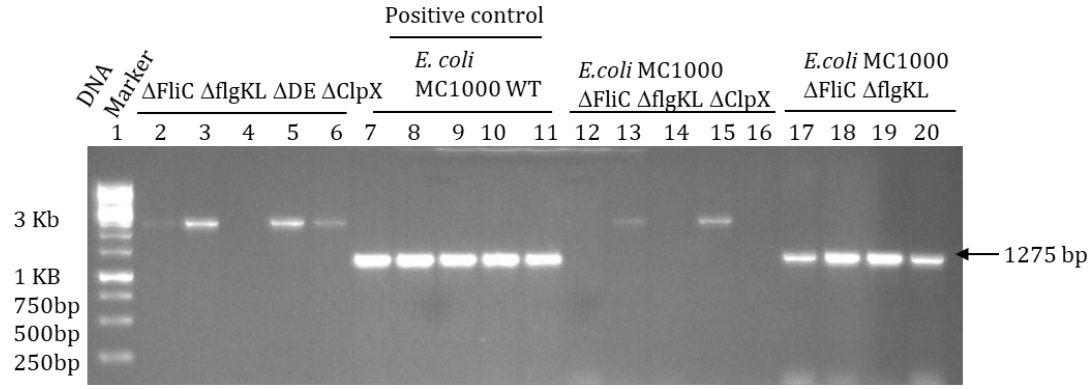


Figure 3.2: Confirmation of KO mutations by PCR.

KO mutants were confirmed with colony PCR (primers: Appendix 8.2), using *E. coli* WT strain as a positive control (Lane 7-11). Newly constructed Δ*flgDE* Δ*clpX* shown in (Lane 2-6), Δ*CKL* Δ*clpX* in (Lane 12-16), and *ΔCKL* in (17-20). Lane 1 was loaded with DNA marker. Reactions were run on an agarose gel in a 1 x TAE buffer and images were taken by exposing the gel to UV Light in G-box (SynGene Technologies).

3.2 Construction of secretion signal variant in pJET 404-based modular secretion plasmid

As mentioned above, a previous student in lab developed the prototype FT3SS system for secretion of proteins on which this project builds. Dr Graham Stafford and Dr CA Green (UoS) designed the prototype *in silico* as shown in Fig. 3.3. It contains a pUC *ori*-origin of replication (pMB1 derivative, up to 500-700 copy numbers, relaxed control), a native *fliC* ribosomal binding site (RBS), IPTG inducible Lac operon, T_5 promoter, convenient multiple cloning sites (MCS), terminator and a selection marker (Fig. 3.4) (Appendix 4: pJEXpress 404 Nucleotide and protein sequence).

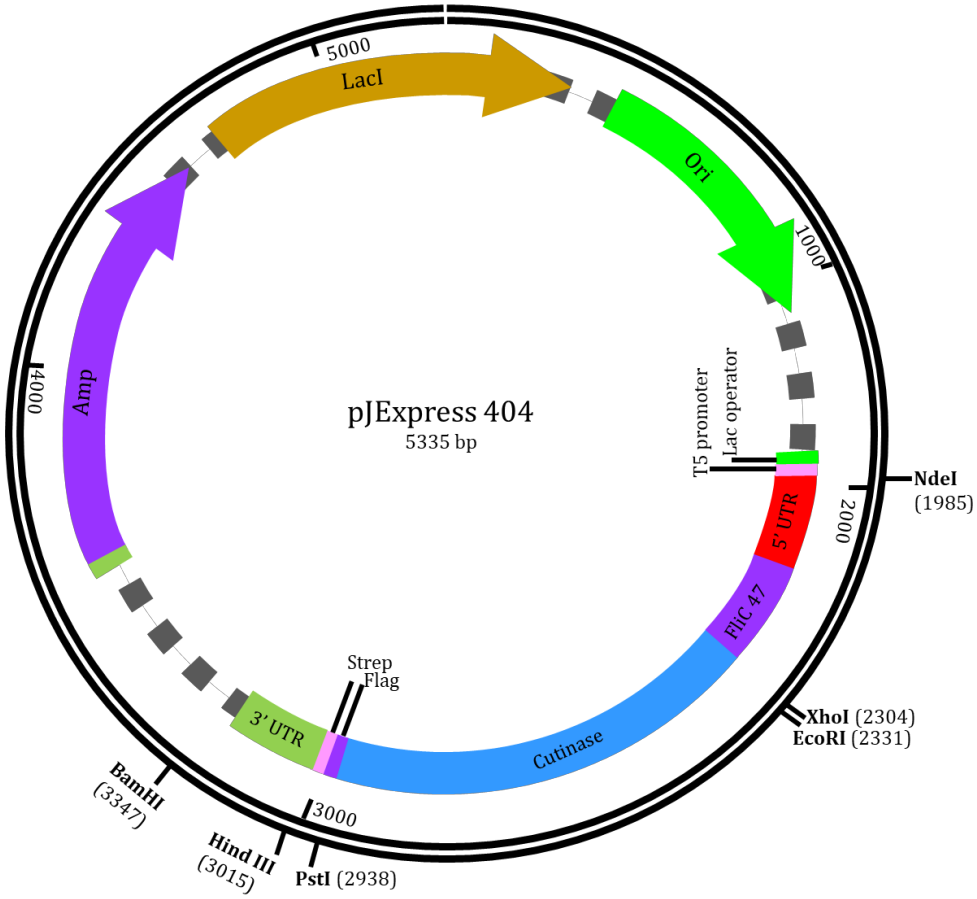


Figure 3.3: Graphical representation of the pJEXpress 404-FliC47-UTR-Cutinase plasmid.

The pJEXpress 404-FliC47-UTR-Cutinase contains the 5'-UTR of FliC (containing the natural FliC, σ^{28}-dependent promoter) downstream of the IPTG inducible T_5 promoter of pJEXpress 404 (DNA 2.0). This is followed by the first 47 amino acids of FliC (FliC47) fused in frame to the Cutinase gene (Appendix 8.4). Coding region is fused to Flag and Strep-tag and followed by the FliC-3'-UTR. Restriction enzyme cloning sites were incorporated throughout to allow modification of secretion constructs in a modular fashion. It also contains a pUC based origin of replication *ori*, ampicillin resistance Amp^R and MCS for loading cargo.

Other key features includes restriction sites to allow cloning of cargo in-frame with the N-terminal C47 amino acids signal sequence of flagellin (*fliC*), the native 5'UTR and promoter placed in between NdeI-XhoI restriction sites to allow both an IPTG inducible T_5 promoter induced and flagella-mediated gene expression (Class III gene *via* σ^{28} (FliA) and indirectly *via* the FlhDC). The C47 moiety is also followed by a TEV protease cleavage site (Glu, X, X, Tyr, X, Gln, Gly/Ser) to allow later removal of the signal if necessary. Importantly, the plasmid has EcoRI-PstI sites for loading cargo protein (shown above with Cutinase inserted), followed by another TEV protease cleavage site, Flag tag (DYKDDDDK, 1012 Da) for detection and Strep-tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) for purification and finally a TAA-termination (stop) codon, respectively. The original prototype also contained the 3'UTR of *fliC*-flagellin incorporated between BamHI-HindIII, which was latterly shown not to be beneficial to secretion (Green *et al.*, 2019). Finally, it has an ampicillin selection marker. Figure 3.3 shows the plasmid containing the Cutinase gene. This is a well-characterised 23.8 kDa enzyme from the fungus *Fusarium solani* that had been previously expressed intracellularly in *Pichia pastoris* and for which a well-established assay, based on its esterase activity is known (Carvalho *et al.,* 1998; Fojan *et al.*, 2000; Tammar *et al.*, 2016). As a part of a previous PhD, in which it was established that Cutinase could be secreted through the FT3SS (confirmed in this chapter) and based on this enzyme activity, a secretion assay based on assaying the activity of secreted enzyme was developed and improved. During the original cloning Cutinase was synthesised in a codon-optimised manner (GeneArt®) and cloned between EcoRI-PstI sites in pJEXpress 404. Resulting pJEXpress 404 with 1-47 FliC-Cutinase was predicted to yield a 30.6 kDa peptide from this construct. In order to assay its activity, Cutinase 's ability to cleave 4–methylumbelliferyl butyrate (MUB) to yield a fluorescent 4-methylumbelliferone (4-MU) molecule was utilised, in which the 4-MU can be measured fluorometrically to facilitate development of a simple fluorescence output secretion assay and it was optimised in Section 3.2.2.

3.3 Production of secretion signal variants fused to Cutinase

3.3.1 Optimisation of secretion signal with pJEXpress 404

Based on previous studies it was decided to test secretion using the first 50 and 100aa of early FlgD and FlgE proteins using an early secretion locked-Δ*flgDE* strain (Stafford *et al.*, 2007). Therefore, DNA encoding the first 50 and 100 aa of FlgD and FlgE were amplified by PCR using *E. coli* MC1000 genomic DNA as a template (Fig. 3.4, Primers: Appendix 3) and cloned into the secretion plasmid (pJExpress 404-Cutinase lacking or containing the FliC UTRs (5' or 3') as shown (Fig 3.6)), using NdeI-XhoI sites incorporated in the 5' and 3' primers, respectively.

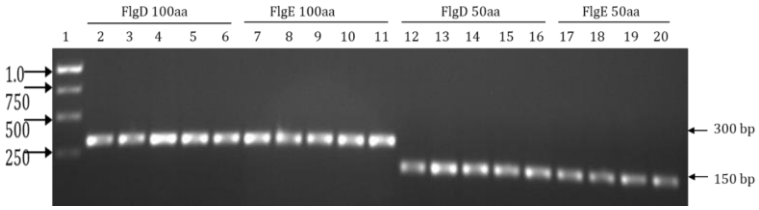


Figure 3.4: PCR for early secretion signals for cloning into pJExpress 404.
High Fidelity PCR amplification was performed for cloning of FlgD and FlgE genes (50 and 100_{aa}) using MC1000 DNA as template. Lane 1 was loaded with 1 Kb DNA ladder followed by $FlgD_{100}$aa (Lane 2-6), $FlgE_{100aa}$ (Lane 7-11), $FlgD_{50aa}$ (Lane 12-16) and $FlgE_{50aa}$ (Lane 17-20). Reactions were run on an agarose gel in a 1 xTAE buffer, image was taken by exposing the gel to UV Light in G-box (SynGene Technologies).

After ligation, successful transformants were then tested for the presence of signal inserts (300 bp for FlgD/FlgE and 150 bp for FlgD/FlgE) by colony PCR (eg. Fig. 3.5 A, B for FlgD and FlgE lacking 3' UTR) all constructs were verified by restriction digest (example Fig. 3.5. C-D) and Sanger sequencing, prior to secretion experiments.

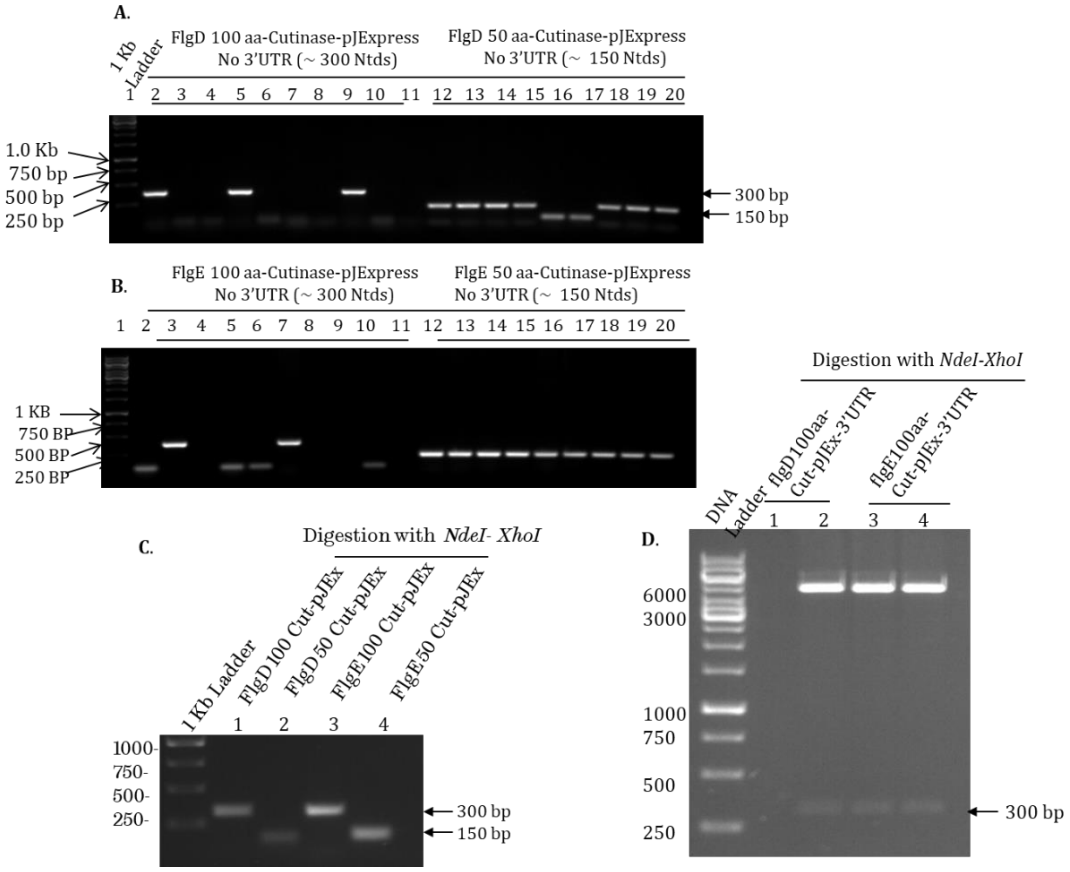
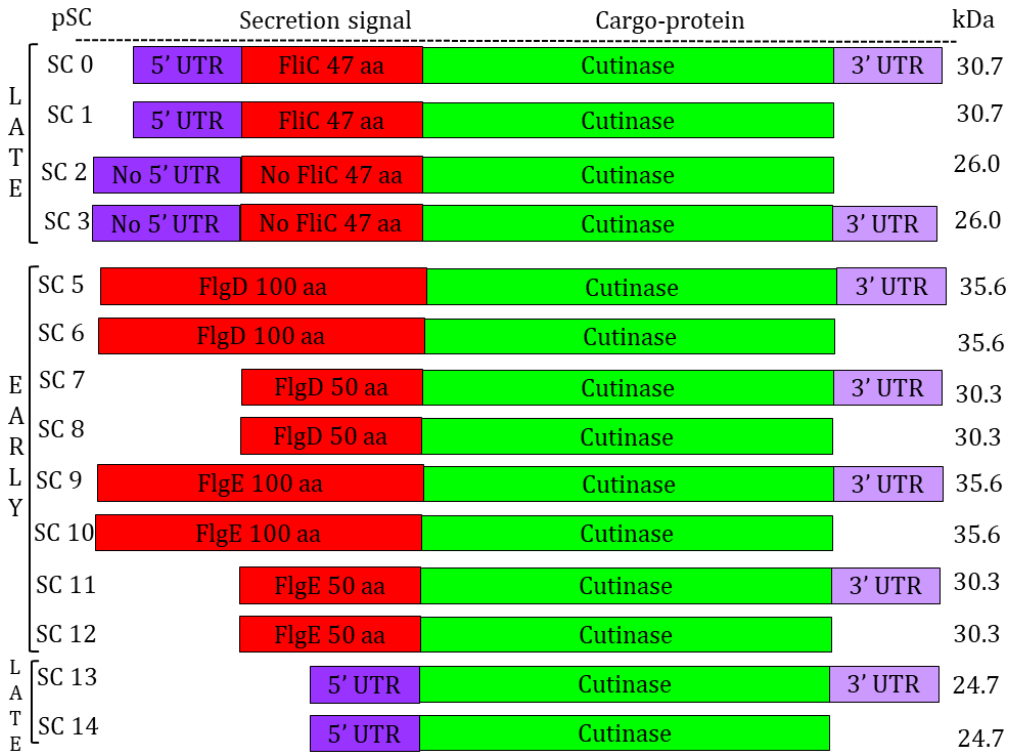


Figure 3.5: Confirmation of cloning by colony PCR and restriction digest.
FlgD and FlgE (300 and 150 bp) were cloned into pJEXpress 404 and confirmed with colony PCR. (A) Colony PCR of clones potentially containing $FlgD_{100aa}$ (Lane 2-11), $FlgD_{50aa}$ (12-20) and (B) $FlgE_{100aa}$ (Lane 2-11), $FlgE_{50aa}$ (12-20). Positive clones show bands at 300 bp for D/E100 and 150 bp for D/E 50 clones. Confirmation using restriction digest (NdeI-XhoI) (D)- only insert bands of positive clones are shown (C).

Both Early and Late signal containing Cutinase -pJEXpress 404 plasmids (with and without 3' UTR) were successfully constructed (Fig. 3.6) and further used for the determination of protein secretion-expression in Section 3.3.



Figures 3.6: pJEXpress 404-Cutinase with early and late secretion signals. Early and late secretion signals were cloned into Cutinase -pJEXpress 404 (with and without FliC-3'UTR), the resulting peptide sizes are shown. (kDa: Kilodalton, SC: Secretion construct, 5'UTR, 3'UTR: 5' and 3'Untranslated regions, aa: amino acids).

3.3.2 Improvement of throughput 4-methylumbelliferyl butyrate assay

In order to facilitate the assessment of the signal variants constructed in this chapter (Fig. 3.6) and subsequent chapters, the Cutinase based activity assay was first assessed and then improved to increase sensitivity and reliability. This assay was based upon the cleavage of 4-Methylumbelliferyl butyrate (4-MUB) into its constituent components, 4-Methylumbelliferrone (4-MU) and the butyrate. The 4-MU released, absorbs U.V. light at the specific excitation wavelength and fluoresces to emit the light at a wavelength, which can then be measured on a plate reader (Tecan, Magellan, Infinite® 200 PRO). The intensity of fluorescent light released is proportional to the concentration of active enzyme (Fig. 3.8), as previously established (Green *et al.*, 2019).

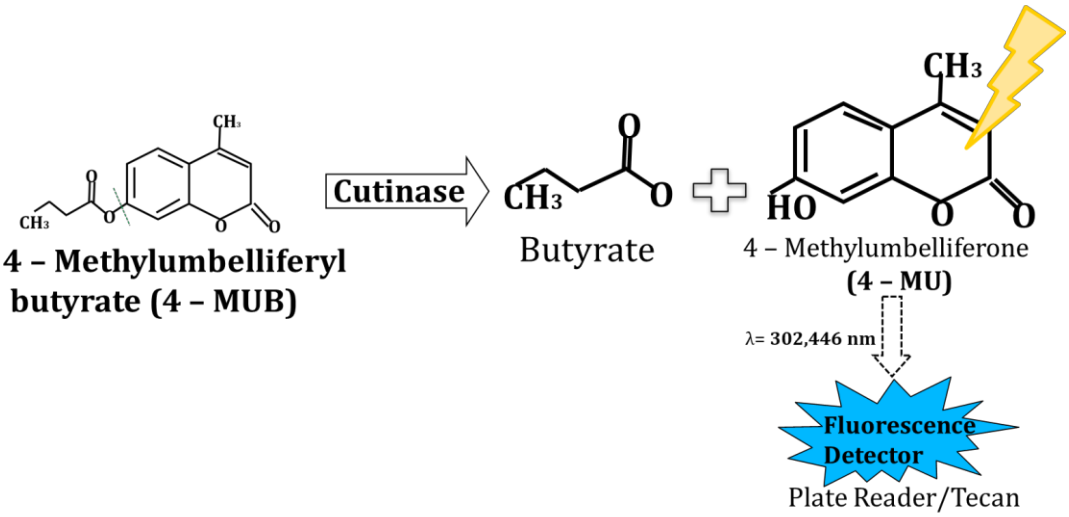


Figure 3.7: Schematic of the enzyme-substrate reaction for MUB assay.
4-MUB is cleaved by Cutinase to release 4-MU, which absorbs UV light at one specific wavelength and emits at others i.e fluorescence, which can be detected and quantified using Fluorescence detector such as Plate reader (Tecan Magellan, Infinite® 200 PRO), (4-MU: Methylumbelliferone, U.V.: Ultraviolet light).

As a first step, MUB assay conditions were optimised prior to protein detection with fluorescence output using emission and excitation scans, using 4-Methylumbelliferone (4-MU), which is the product of the cleavage reaction of MUB (Sigma Aldrich). Serial dilutions of 4-MU were prepared in a solution of Phosphate citrate buffer (PCB) and 200 µl aliquots were placed in 96-well plates (Greiner™) and assayed using a plate reader, with excitation scanning at 200-400 nm, and emission scanning at 340-600 nm. The highest excitation and emission wavelengths obtained were at 302 and 446 nm as shown (Fig. 3.8 a-b).

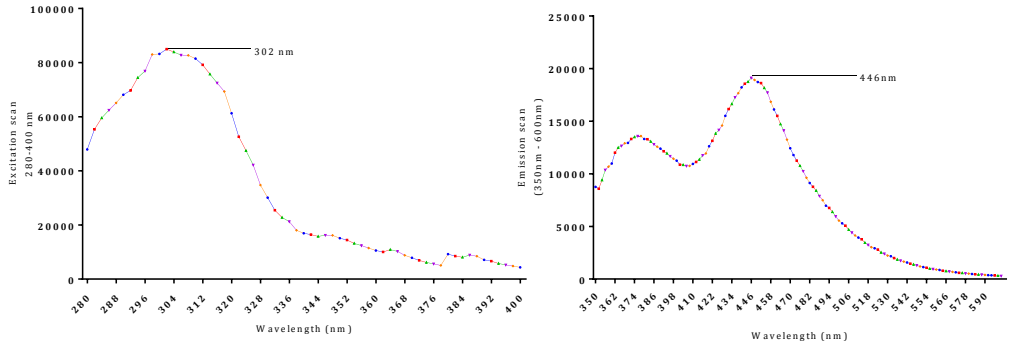


Figure 3.8: Excitation-Emission scans for the optimisation of MUB Assay.

The excitation (200-400 nm) (a: left) and emission (340- 600nm) (b: right) scans were performed on a Tecan Magellan, Infinite® 200 PRO Plate reader. Graphs were plotted using Graphpad prism 7.0.

In the second step, the assay conditions were optimised using culture supernatants (See Section 2.7.2). Supernatants from *E. coli* MC1000 Δ*fliC* Δ*flgKL* Δ*clpX* harbouring pSC 1 (Fig. 3.6) were induced with IPTG prior to assay using MUB substrate (100 µM), while Phosphate Citrate buffer (PCB), Luria-Bertani media (LB) and WT (supernatant from wild-type bacteria) were used as controls. In all cases, a reaction volume of 200 µl was maintained, which was made up of 160 µl MUB prepared in PCB (pH 5.0). The amount of IPTG needed for optimal response was established using different concentrations of Isopropyl β-D-1 thiogalactopyranoside (IPTG) from (0 to 0.5 mM) and using 40 µl of supernatant (Fig. 3.9.). These data show that 0.05 mM IPTG gave the maximum fluorescence output, and it was significantly different compared to all other concentrations of IPTG used for induction (Fig. 3.9). Henceforth, 0.05 mM IPTG was used for protein expression, secretion and protein purification.

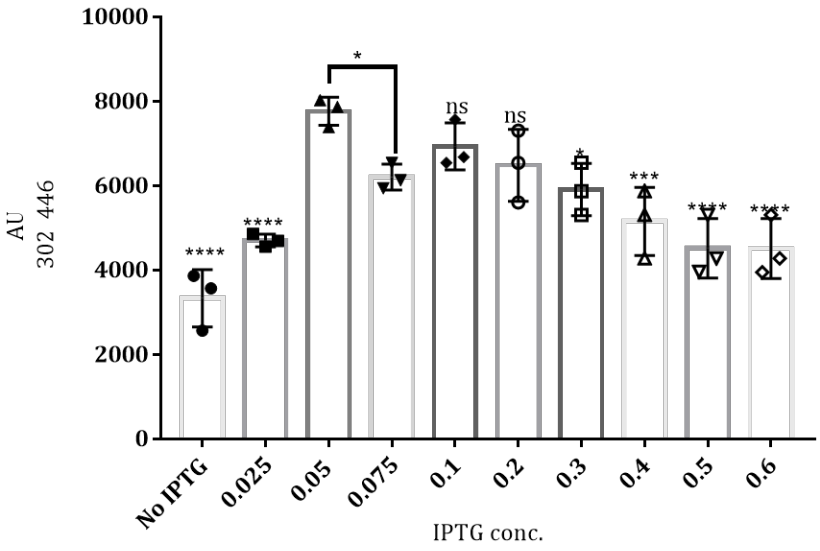


Figure 3.9: Optimisation of IPTG conc.for the induction of MUB assay.

The late signal harbouring plasmid was induced (0-0.6 mM IPTG) and MUB assay was performed as described (Section 3.3.2), data display 3 biological repeats with SEM and significance was determined by One-way ANOVA with *p*-value (**** $p<0.0001$).

Finally, the assay was optimised for the quantities of enzymes and substrates used in the reactions. As seen in Fig. 3.9, the cultures were induced with 0.05 mM IPTG and then in a 200 µl reaction 160 µl of 100 µM substrate and varying quantities of supernatant were used. As seen in Fig. 3.10 a-b, the optimal results (secretion output) were obtained with 20-30 µl with 25 µl taken as the

optimal volume for future screening. In all experiments for the MUB assay, induction with 0.05 mM IPTG and 25 µl supernatant were used in addition to controls using 200 µl each of PCB, MUB, LB broth, LB+MUB, MUB + WT (wild-type) supernatant and MUB + Empty pJEXpress while detection was carried out at 302 446 nm, throughput this thesis.

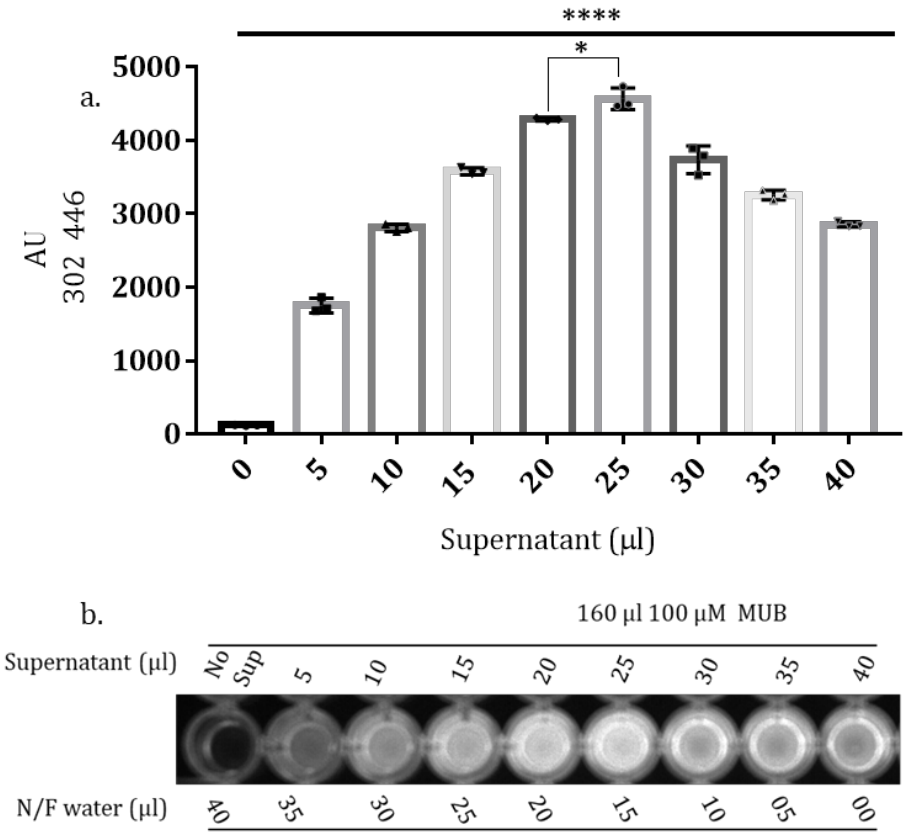


Figure 3.10: Optimisation of the MUB assay reaction.
(a) MUB assay enzyme-substrate reaction was optimised for the volume of enzymes (supernatant) in the reaction (0 to 40 µl), while 160 µl of substrate kept constant. (b) Graphs was plotted by using Graphpad Prism 7.0, significance determined using One-way ANOVA with *p*-value (**** $p<0.0000$), 3 technical repeats observed. Reactions were visualised by exposing to UV light in G-box (SynGene technologies).

3.4 Testing the efficiency of protein secretion with early (FlgD/E) and late (FliC) signal sequences

In order to compare the efficiency of protein secretion of the early secretion signal with the late FliC signal, a number of experiments were performed using FlgD and FlgE signal sequences with both the first, N-terminal 50 and 100 aa, cloned into pJEXpress 404 (Section 3.2.1) and shown in Fig. 3.6.

In short, the Δ*CKL* strain was used as a prototype secretion strain alongside the Δ*CKL* Δ*clpX* strain, which had previously been shown to increase secretion-probably due to increased flagella gene expression (Tomoyasu *et al.*, 2002; Green, 2016). In parallel, an early locked Δ*flgDE* deletion strain, with accompanying *clpX* mutation to produce an isogenic strain to Δ*CKL,* was also constructed. We had also previously shown that there was no benefit of deleting *flgM* in this strain as enough C_{47} construct was produced with IPTG in the media (Stafford *et al.*, 2007). In addition, in this section the effect of including the fliC-3'UTR on secretion using FlgD/FlgE signal sequences was tested by both methods, MUB assay (Fig. 3.11) and western blot analysis (Section 3.4).

These 4-strains (Fig. 3.1) were transformed with these early and late signals containing plasmids (Fig. 3.6) and grown in liquid broth culture and induced with IPTG as described (Section 2.7.2). The cells and supernatant were collected and MUB assay was performed (Section 3.2.2). The fluorescence output was plotted as shown in Fig. 3.11 and it clearly shows that Δ*DE* and Δ*DE* Δ*clpX*-early secretion locked strains do not allow secretion of either early (pSC 6, 7, 8, 9, 10, 11, 12) or late (pSC 1) secretion signals *via* the modified FT3SS.

However, protein secretion was observed through the substrate switched FT3SS of Δ*CKL* and Δ*CKL* Δ*clpX* strains, for both early (pSC 6: D_{100} and pSC 8: D_{50}) and late (pSC 1: C_{47}) secretion signals. It was also observed that the late secretion signal directed higher levels of protein secretion than early signal, i.e. Cutinase levels from pSC 1: C_{47} were 2.87 fold higher than pSC 6: D_{100} in Δ*CKL* and 2.23 fold higher in Δ*CKL* Δ*clpX* strain in this assay (Fig. 3.11 b.). It was also observed that pSC 1: C_{47} and pSC 6: D_{100} were secreted at higher levels (1.07 and 1.38-fold) in Δ*CKL* Δ*clpX* than Δ*CKL* (p<0.05) strain, respectively. Secretion was also observed for pSC 8: D_{50} in Δ*CKL* Δ*clpX,* while, pSC 1 was 2.4-fold more efficient than pSC 8 (Fig. 3.11 d).

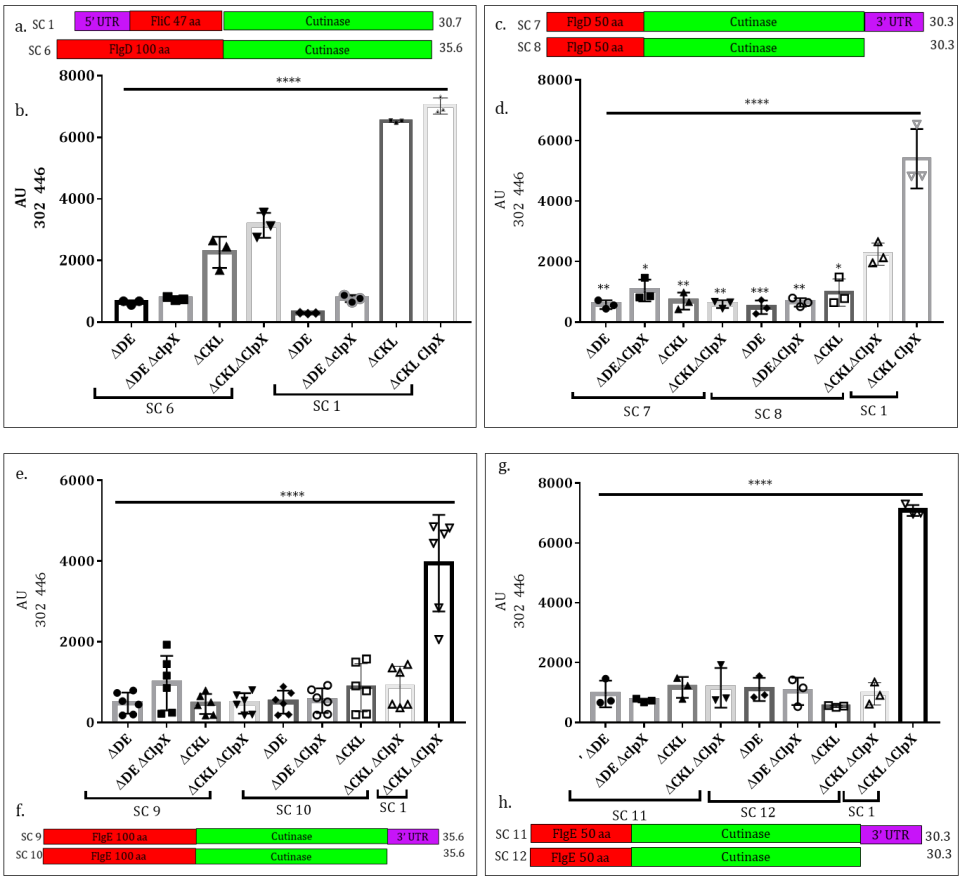


Figure 3.11: Comparison of protein secretion efficiency of early and late secretion signals *via* the FT3SS with MUB assay.
MUB assay was carried out for the comparison of Early (FlgD/FlgE 100aa and 50 aa) and Late (5'^{UTR}-FliC$_{47aa}$) secretion signals, cloned in to Cutinase -pJEXpress 404 (pSC 6, pSC 7, pSC 8, pSC 9, pSC 10, pSC 11 and pSC 12 *vs* pSC 1) in the strains shown (Fig. 3.1. One way ANOVA for multiple comparisons was carried out with pSC 1-Δ*CKL* Δ*clpX* (5'-C$_{47aa}$-Cut-pJEx) as a positive reference. Error bars (±) SEM, statistical significance was defined by ****p<0.0001. Assay was carried out in triplicate, with 3 biological repeats.

3.5 Testing the efficiency of protein secretion with early (FlgD/E) and Late (FliC) signal sequences : validation by western blot analysis

In order to fully investigate the MUB assay data for Cutinase and also to verify the validity of this approach *i.e.* to find correlation between Cutinase activity and western blot estimated secretion, parallel experiments using the detection of Flag-tag in western immunoblot assays were employed. These assays also allowed the estimation of protein concentration in the media, since a protein standard was also run on western immunoblot and a linear signal measurement system from LI-COR (*Image studio 5.2*) was used which is linear of 4-8 orders of magnitude and thus more quantitative than film-based western blots. In order to enable estimation of secreted protein fractions with western

blotting, the cultures were centrifuged as before and the cells collected to produce intracellular protein fractions, while supernatants were precipitated using 10 % Trichloroacetic acid (10 % v/v) prior to re-suspension with 2 x SDS-loading buffer (Section 2.6.4). Both samples were then run on SDS-PAGE gels and western blot analysis performed (Section 2.6.4.5) using the Anti-Flag tag antibody conjugated with HRP (Sigma).

These data once again showed that none of the FlgE containing constructs (pSC 9-pSC 12) secreted into the media, despite being adequately expressed intracellularly (Fig. 3.12) in agreement with MUB assay (Fig. 3.11). In contrast, the FlgD containing constructs were secreted into the medium with lower levels (pSC 5, pSC 6 and pSC 8). In the case of pSC 8 (D_{50aa}) secretion (for analysis see Fig. 3.13) expression across strains was again fairly consistent but secretion was only present in the Δ*CKL* Δ*clpX* strain, but strikingly the presence of the *fliC* 3'UTR abolished secretion, meaning that the signal seems to be detrimental to secretion regardless of whether it is attached to a late or early signal protein (Green, 2016). Of note here is that, in the case of $FlgE_{100}$ constructs (pSC 9-pSC 10) expression in the presence of the $fliC_{47}$-3'UTR was always significantly higher, however, this trend did not seem to be consistent as for the $FlgE_{50}$ constructs (pSC 11-pSC 12), there seemed to be no difference in intracellular expression.

It could be observed from, intracellular-extracellular protein secretion (Fig. 3.12a.) and densitometry (Fig.3.13a-b), that Cutinase was evenly expressed by all early and late pJEXpress 404, in all 4–flagella mutant strains. In agreement with previous MUB assay result (Fig. 3.11), Δ*DE* and Δ*DE* Δ*clpX*-early secretion locked strains do not allow secretion of either early ($FlgD_{100aa}$) pSC 6 or late pSC 1 (5'UTR-$FliC_{47}$aa) secretion signals. However, protein secretion was observed through substrate-switched Δ*CKL* and Δ*CKL* Δ*clpX* strains, for both early and late secretion signals. It can be observed from Fig. 3.12a that pSC 1 expressed more efficiently (2-fold, $p<0.001$) than the pSC 6, while it secreted (3-6 fold, $p<0.001$) more efficiently in Δ*CKL* and Δ*CKL* Δ*clpX* than the pSC 6 (Fig. 3.12a). An increase in protein secretion was also observed for early signals in Δ*CKL* Δ*clpX* than Δ*CKL* strains, however, it was not significant (* $p<0.05$).

It can be observed (Fig. 3.12b) that Cutinase protein was evenly expressed in all 4-strains for pSC 7-8. However, protein secretion was only observed in Δ*CKL* Δ*clpX* strain for pSC 8. Protein secretion does not occur through the FT3SS of the early locked strain for early or late signals, as previously mentioned, while protein secretion through the modified FT3SS of Δ*CKL* Δ*clpX* strain restricts the secretion of Early signal with 3'UTR signal, however, it allows the secretion of that lack the *fliC*-3'UTR. It was also observed (Fig. 12b.) that pSC 8 expressed more efficiently (2-fold, $p<0.001$) than the pSC 7 while protein secretion was only observed in Δ*CKL* Δ*clpX* where pSC 1 secreted (3-fold, $p<0.001$) more efficiently than pSC 8 in Δ*CKL* Δ*clpX* strain. It was also observed that Cutinase was evenly expressed by early signal (FlgE) pSC 9-10 and pSC 11-12, across all 4–mutant strains. The protein expression with pSC 9 was low with intact 3'UTR but it was increased with the exclusion of 3'UTR as shown by pSC 10 (Fig. 3.12c). It can also be observed that, pSC 1, expressed 2-fold more efficiently than the pSC 9 and pSC 10, respectively ($p<0.0001$) (Fig. 3.13e).

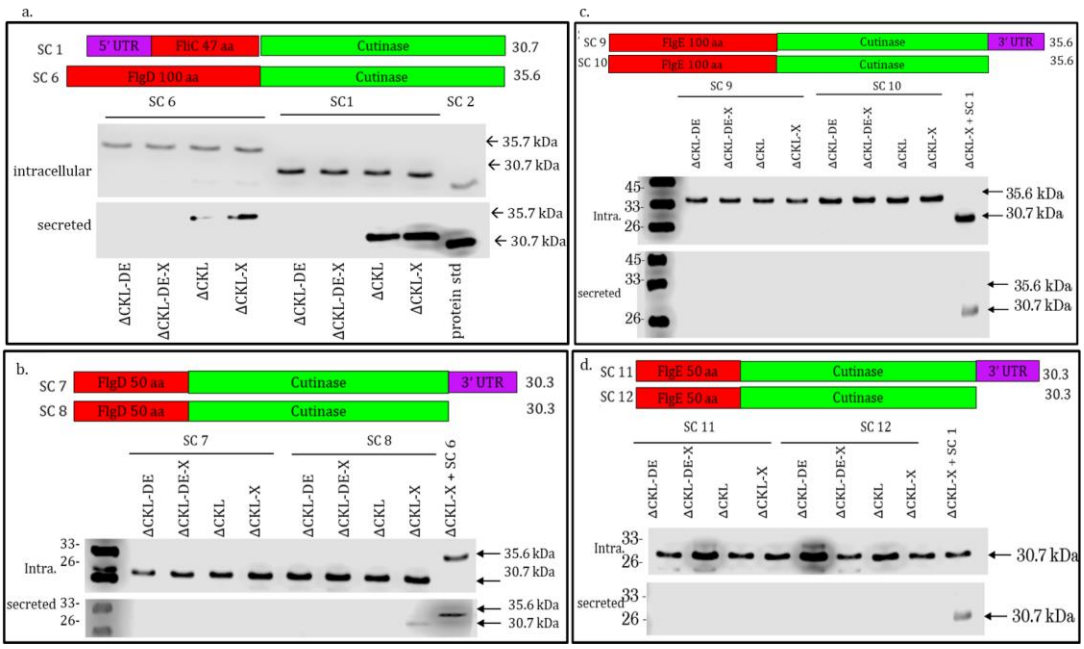


Figure 3.12: SDS-PAGE western blot analysis for protein detection.
Western blot analysis for the comparison of secretion efficiency of early secretion plasmids with intact 3'UTR and without 3'UTR in the pJEXpress 404 (pSC 5-12: a-h) (Fig. 3.6), by expressing in to flagella mutant strains Δ*DE*, Δ*DE* Δ*clpX,* Δ*CKL* and Δ*CKL* Δ*clpX,* respectively (Fig. 3.1). Protein expression and secretion are shown. pSC 1-Δ*CKL* Δ*clpX* was used as a control. Blots were developed using *LI-COR C–Digit*. Relative densities were calculated, results from (3x) biological repeats (Intra: intracellular).

Also, pSC 10 expressed more efficiently than pSC 9 and it was higher by 2-*fold* in Δ*CKL* and Δ*CKL* Δ*clpX*. It was also observed (Fig. 3.13f) that, there was a no significant difference (p<0.05) between the expression of pSC 11 and pSC 12 or pSC 1, across all the 4-strains tested. However, there was no protein secretion observed for either pSC 9-10 or pSC 11-12, the early signal pSC 9-pSC 12 (FlgE $_{50/100aa}$) pJEXpress 404, in agreement with the MUB assay results Fig. 3.11.

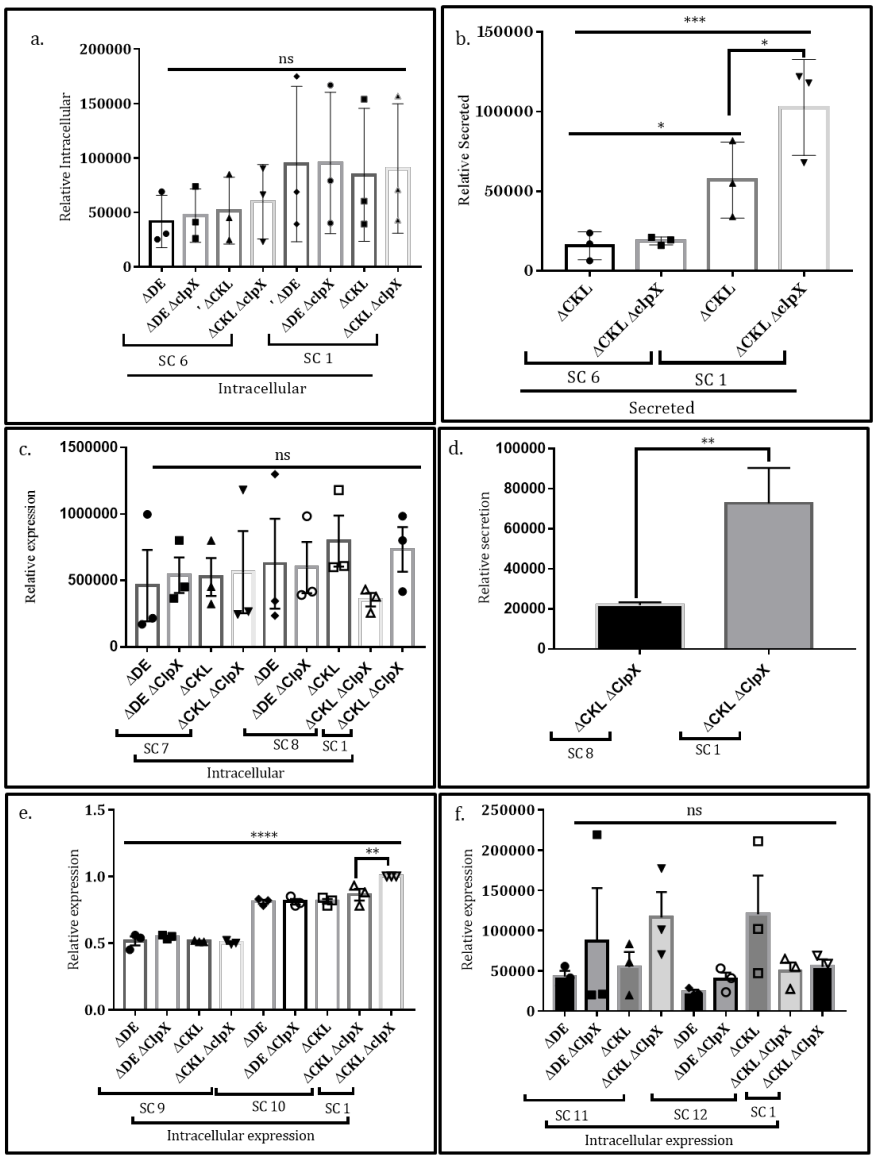


Figure 3:13: Comparison of protein secretion efficiency of pSC 5 and pSC 6. The efficiency of protein expression in 4-flagella mutant strains Δ*DE*, Δ*DE* Δ*clpX,* Δ*CKL* and Δ*CKL* Δ*clpX* (Fig 3.1) as shown (a, c, e, f), while of secretion (b and d.), using densitometry from Fig. 3.12. Images were analysed by LI-COR (Image Studio). Relative densities were calculated, results from (3 x) biological repeats, graphs plotted using Graphpad Prism 7.0. One way ANOVA for multiple comparisons was performed with Error bars (±) SEM, statistical significance was defined by **** p<0.0001.

(Note here, a human growth hormone (hGH) standard (92 µg µl^{-1}) was utilised for western blot, which in turn was purified by a colleague from FT3SS route of Δ*CKL* Δ*clpX* strain using 5'UTR-FliC$_{47aa}$-hGH-pJExpress 404, however it didn't reflect the true secretion of Cutinase *via* this route, hence it was not used for analysis and estimation of Cutinase secretion, as later differential secretion was reported for C$_H$2, Cutinase , hGH, E2 proteins (Green *et al.*, 2019)).

3.5.1 Effect of FliC-3'UTR on the secretion of Early FlgD secretion signal

As observed above, the *fliC*-3'UTR seemed to negatively influence secretion of proteins through the FT3SS -an observation that we had previously made for the secretion of proteins fused to the FliC-47 signal (Green *et al.*, 2019). Here the effect of the *fliC*-3'UTR was investigated further by comparing secretion of a FlgD100 (the most efficient early signal tested) in the presence (pSC 5) and absence of the *fliC* 3'UTR (pSC 6). The efficiency of secretion of pSC 5 and pSC 6 was compared with western immunoassay (Fig. 3.14).

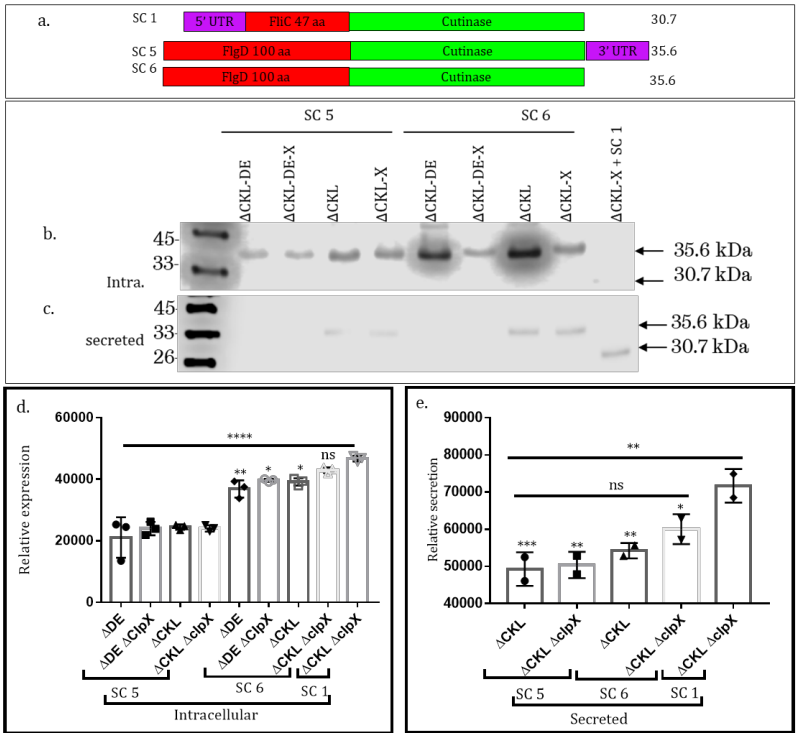


Figure 3.14: Protein secretion comparison of early signal with *fliC*-3'UTR
The efficiency of protein expression of pSC 5 and pSC 6 in Δ*DE*, Δ*DE* Δ*clpX*, Δ*CKL* and Δ*CKL* Δ*clpX* strains (Fig 3.1) as shown (b and d), while secretion in (c and e). Images analysed using LI-COR, Relative densities calculated, results from (3x) biological repeats, graphs plotted by using Graphpad Prism 7.0. One way ANOVA for multiple comparisons performed with Error bars (±) SEM, statistical significance defined by **** p<0.0001.

It can be observed from Fig.14b-d. that pSC 6 (no 3'UTR) expressed more efficiently than the 3'UTR-containing construct (pSC 5), where pSC 6 was 1.59-1.78x more highly expressed. However, there was no significant difference observed (p<0.05) in their secretion (Fig. 14 c-e.). The data shows that, all early and late secretion constructs evenly produced Cutinase (Fig. 3.12), however only early FlgD100 secreted the protein but not as efficiently as the Late secretion signal pSC 1 (5'UTR-FliC$_{47aa}$). Hence, the late signal (pSC 1) and Δ*CKL* Δ*clpX* was selected as the most efficient combination for the heterologous protein secretion *via* the modified FT3SS.

Therefore, in order to improve the protein secretion *via* the FT3SS, this most efficient combination was further investigated to check whether the FliC-5'UTR can direct protein secretion more efficiently (Végh *et al.*, 2006) than the intact late (5'UTR-FliC$_{47aa}$) signal and it is reported in Section 3.5.2.

3.5.2 Evaluation of the 5'UTR-FliC signal *via* the modified FT3SS

As shown in this chapter (and by a previous student) the most efficient secretion signal was that containing (5'UTR-FliC$_{47aa}$) in pSC 1 construct (Fig. 3.6). However, previous literature including one in which FT3SS secretion was characterised in a Δ*fliD* strain showed that the 5'UTR of *fliC* alone was able to direct secretion with several reports of co-translational secretion of FliC in *Salmonella typhimurium* (Singer *et al.,* 2014). Therefore, it was tested whether the 5'UTR alone is self-sufficient for exporting the target protein-Cutinase *via* the modified FT3SS of *E. coli*. The 5'UTR was therefore cloned in to pJEXpress 404 (with and without *fliC*-3'UTR (Section 3.2.1)) with placement to mean that its start codon-ATG was that of the Cutinase gene, but with no proteinaceous signal sequence (pSC 13 and pSC 14). As a first stage expression was again tested in the same 4-mutant strains (Fig. 3.1) by Western blot analysis (Section 3.4). The MUB assay was also performed on the supernatant (Section 3.2.2).

Firstly, by western blot, it was observed that Cutinase was evenly expressed in all 4-strains carrying the pSC 13 and pSC 14 and at a similar level to the pSC 1 construct (Fig. 3.15 d-f.). However, protein secretion was only observed

in the Δ*CKL*-Δ*clpX* strain for any of the constructs, with secretion using the 5'UTR signal able to direct secretion of Cutinase but at 20-40% of the levels of the 5'UTR-FliC47 (Fig. 3.15 c, e and g) in contrast to the proteinaceous signal, the inclusion of the *fliC*-3'UTR had no significant effect.

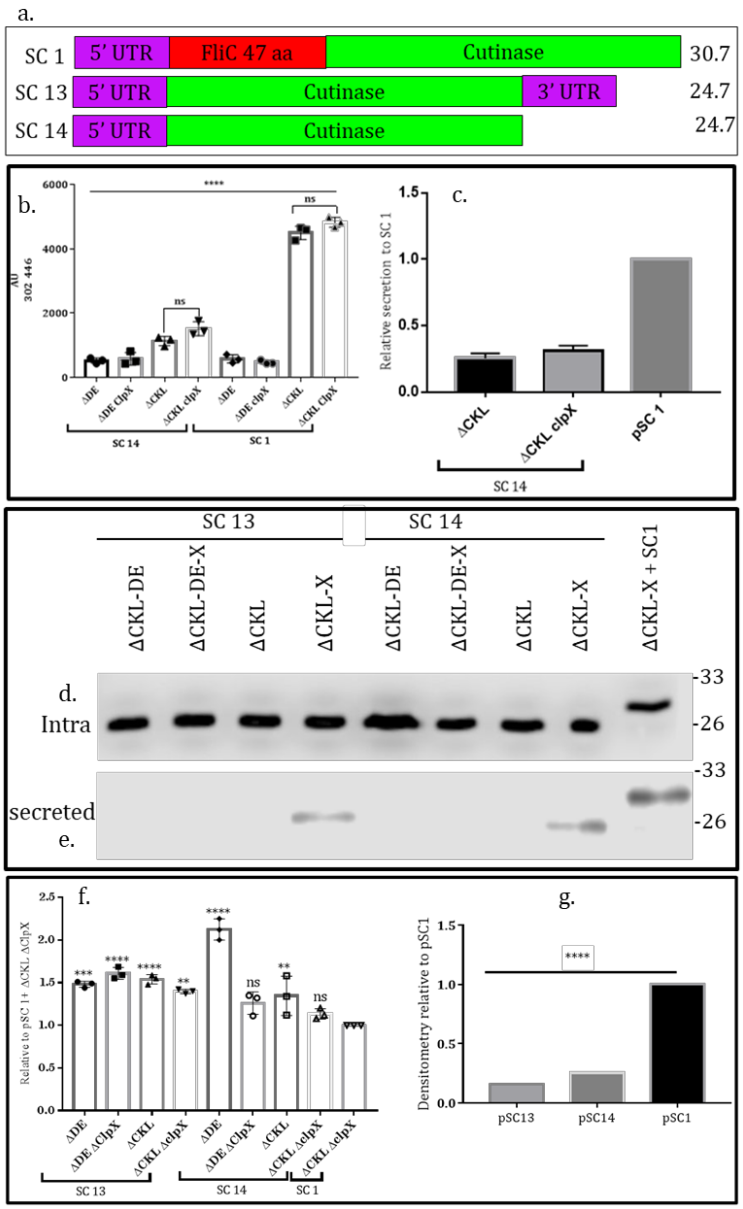


Figure 3.15: Late signal odification for protein secretion with SDS-PAGE. pSC 13 and pSC 14 (a) were expressed into Δ*DE,* Δ*DE* Δ*clpX,* Δ*CKL,* Δ*CKL* Δ*clpX* strains, respectively. MUB assay was performed as section 3.2.2 (b-c) and their intracellular proteins were run as shown (d.) with pSC 1-Δ*CKL* Δ*clpX* as control, while secreted fractions were run as (e.), their relative expression and secretion to control is shown in (f) and (g) respectively. Results from (3x) biological repeats. Relative densities were calculated using LI-COR (*Image Studio*), compared by One Way ANOVA for multiple comparisons, significance was determined by *p* values ($p<0.0001$****), Error bars (±) SEM, graphs were plotted by using Graphpad Prism 7.0.

Similar results were gained when assessed by MUB assay, i.e. no secretion in the Δ*DE* background strains and only limited secretion (pSC 14) in the Δ*CKL* based strains (Fig. 3.15 b-c.) at 4-6-fold less levels than pSC 1. However, it is curious that MUB assay did not detect secretion of pSC 13 strain indicating maybe this protein might have been produced in an unfolded and less active form to detect in MUB assay. Overall these data show that while the UTR may be able to direct secretion from FT3SS it is less efficient than the combined protein-RNA signal.

3.6 Discussion

In this chapter, several key parameters relating to heterologous protein export *via* the FT3SS were tested, namely whether an early or late signal is more effective at protein export and whether it might be feasible to utilise an RNA only (5'UTR-*fliC*) as an export signal. In addition, an improved secretion screening method based on the secretion of an active enzyme was established.

3.6.1 Establishing the streamlined Cutinase assay for secretion screening

Firstly, this chapter set out to establish a facile, streamlined, semi high-throughput assay, both to assess the signal constructs outlined but also as a prelude to later work presented in this thesis where gene libraries were screened for secretion boosting or attenuating genes (Chapter 5). Therefore, a well-established esterase activity based assay was adapted in which the well characterised 23.69 kDa fungal protein-Cutinase isolated from *Fusarium solani* was cloned in-frame with a range of flagella secretion signals (Zhiqiang *et al.*, 2009).

In order to assay its activity, the ability of Cutinase to cleave esterase bonds, assayed *via* use of the artificial substrate 4-methylumbelliferyl butyrate (MUB), cleaved to yield a fluorescent 4-methylumbelliferone (4-MU) (Fig. 3.7) which facilitated the development of a fluorometric secretion assay (Green, 2016). Therefore, Cutinase -pJExpress 404 secretion constructs harbouring early and late secretion signals were expressed in a range of flagella mutant strains (Fig. 3.1) and their cell-free supernatant was collected to perform the MUB assay. The assay was refined for visualisation and quantitation by performing excitation and emission scans following optimisation of IPTG induction of secretor strain

(harbouring secretion construct), the concentration of enzyme-substrate and the assay reaction volume in a plate reader-based assay. Therefore, it could also be qualitatively observed on a UV-transilluminator for quick visual inspection.

The specificity of the assay was tested and validated with western immunoblots of the same supernatants utilised for MUB assay and also on their intracellular fractions. A significant correlation between MUB assay and western blot was observed. The assay was able to detect a wide range of concentrations of secreted Cutinase in the medium. Therefore, this simple assay was utilised for screening of a larger number of strains, secretion signal variants and the genomic libraries (Chapter 5), which is otherwise cumbersome and not practical with traditional western blot-based analysis. The correlation between MUB and western immunoassay gives credence to the larger screening performed later for genomic library screening. However, it is not applicable to the intracellular protein estimation, unless the protein is re-solubilised into an appropriate buffer and MUB assay is performed in the same way as described for supernatant (Section 3.2.2). Thus, a high throughput enzyme secretion assay was developed, that enabled screening of a range of engineered strains and secretion signal harbouring constructs. Additionally, as the enzyme, activity requires correct folding of protein and therefore assay highlights that Cutinase was correctly folded in extracellular medium shown by the functionality of Cutinase. This is most desired considering IB applications of truncated FT3SS secretion platform, as correct folding was obtained (Section 3.2.2).

These findings demonstrate that for the truncated FT3SS following secretion through FT3SS in an unfolded state, the recombinant Cutinase is able to fold correctly into intended its 3D structure to allow functionality. This is very compelling considering IB application of truncated FT3SS platform. This could open the possibility of cell cultures that secrete proteins continuously, without need to sacrifice cell culture to retrieve the protein product (Rosano *et al.,* 2014). Additionally, this MUB assay could be utilised as an alternative to several assays used in genetic manipulations including the detection of protein secretion in supernatant, as it is quicker and more streamlined than lengthy western blot

assay. This assay might serve as an excellent alternative to, e.g, SptP phosphatase assay utilised for testing docking of early and late subunits to export gate, which did not work very-well (*GS: personal communication*) or the cumbersome green fluorescent protein assay utilised for genetic manipulations which do not secrete the protein (Matos *et al.*, 2012; Mahalik *et al.,* 2014). Therefore, the streamlined, optimised MUB assay might serve an excellent tool to these assays in terms of researching fundamental aspects of flagella gene regulations.

3.6.2 Establishment of 'late' secretion signal (*fliC*) over 'early' signals

In order to investigate the optimal secretion signal for efficient recombinant protein production *via* the modified FT3SS, secretion of early flagella substrate signal sequences, namely hook (FlgE) and hook-cap (FlgD) were compared to the secretion of the late FliC secretion signal. In both MUB and western blotting assay, the hook-protein FlgE did not direct secretion of the loaded Cutinase through the modified FT3SS of either early locked Δ*DE (*Δ*clpX)* or substrate switched Δ*CKL (*Δ*clpX)* backgrounds. However, the hook-cap protein FlgD did direct secretion successfully in the substrate specificity switched Δ*CKL (*Δ*clpX)* but not the early secretion locked Δ*CKL* Δ*DE* strain. These data indicate that for unknown reasons the switched ΔCKL strain was more efficient at secreting protein, and that protein secretion using the late secretion signal was significantly more efficient than early signals -upto 5-7 fold.

3.6.3 Role of UTRs in protein secretion *via* the modified FT3SS.

In this chapter, the role of the *fliC*-UTRs (5' and 3') was also investigated. For the late FliC signal, the 3'UTR was detrimental to secretion while the 5'UTR was of benefit (Fig.3.14, also seen in (Green *et al.*, 2019) to which this chapter contributed). It appears that the 3' UTR might produce an RNA based signal that decreases protein expression, however its mechanism of action on secretion *via* the FT3SS remains enigmatic. It could be investigated by checking expression levels of 3'UTR encoded RNA using qPCR, which might provide possible explanation.

In this chapter, and our previous work, the intact late secretion signal (5'UTR-FliC$_{47aa}$) was established as an efficient secretion signal for protein

production (Section 3.5.2), as had been found by others (Majander *et al.*, 2005; Singer *et al.*, 2012b). We then tested whether the 5'UTR alone could direct secretion, showing low but significant protein secretion (Fig. 3.15). These findings agreed, with some previous work, where the FliC-5'UTR (Majander *et al.*, 2005; Végh *et al.*, 2006), and 26–47 aa residue of FliC signal were utilised (Végh *et al.*, 2006; Ve, 2010) which alone directed secretion in *E. coli* and *Salmonella* respectively. However, it does not match with the evidence that 5'UTR is not necessary for secretion (Végh *et al.*, 2006) as secretion was abolished in absence of 5'UTR (Fig. 3.14) (Green, 2016). The requirement of intact 5'UTR and 47 aa for efficient protein secretion suggests that 5'UTR harbours a part of secretion signal which also suggests that the 5'UTR was not as efficient as intact late signal for protein secretion *via* the FT3SS (Majander *et al.*, 2005; Aldridge, Gnerer, *et al.*, 2006).

Some of the apparent contradictions in the literature and or results could be due to the potential variance in strains, the plasmid used, presence of T_5 and *fliC* promoters in the plasmid, the experimental conditions, reporter proteins used or the growth conditions used during the experiment. The improved secretion using intact late signal could be attributed to the protection of the mRNA transcript of late secretion signal with 5'UTR (Aldridge *et al.*, 2006). The independent secretion achieved using only 5'UTR (Majander *et al.*, 2005) and FliC 26-47aa (Végh *et al.*, 2006; Ve, 2010) may have an additive effect on secretion in an intact late signal, however it can be tested for this cargo protein as secretion was achieved with only 5'UTR but it needs to be checked in absence of 5'UTR, whether 47aa directs protein export by constructing 47aa-Cutinase-pJEXpress version of the secretion construct. It was also noted that pJEXpress 404 employed in these analyses contains two promoters, the T_5 promoter in the plasmid and the native flagella *fliC* promoter, contributing two different transcripts, comparatively longer from an upstream T_5 promoter and a shorter from the downstream native *fliC* promoter. It was unclear whether T_5 transcript was dominating or the native *fliC* during protein secretion. It was not clear whether the T_5 transcripts were recognised by export gate for secretion. The high-level intracellular expression may be suggesting the dominance of the T_5 transcripts

over the native *fliC* transcripts if the T_5 transcripts are not secreted. It may also suggest that T_5 promoter engages the cell machinery for transcription of longer transcripts, allowing a lower rate of transcription from *fliC* native promoter. However, in order to support these speculations, more robust experiments need to be designed with either removal of T_5 promoter from pJExpress or maybe without inducing the plasmid with IPTG and observing the basal secretion levels.

It was also observed that the inclusion of 3'UTR was not essential for secretion, irrespective of early (Fig. 3.11) or late signal (Green, 2016) or the 5'UTR only (Fig. 3.15). On the contrary, it was observed that the exclusion of 3'UTR increases the secretion, despite potentially reducing internal expression level (Fig. 3.11-3.15.). This may be due to the fact that 3'UTR influences the levels and stability of *fliC* mRNA, which in turn reduces the FlgM secretion alluding the improved secretion in absence of 3'UTR, as FlgM might be more readily secreted *via* the FT3SS (Guo *et al.*, 2014). However, there is no experimental evidence to support the improvement of protein secretion in absence of 3'UTR, unless 3'UTR interferes either with export gate or other proteins or may be the RNA transcript of 3'UTR downregulate the transcription from pJEXpress 404-T_5 or native FliC promoter by feedback inhibition and thus significantly decreases the rate of diffusion of substrate subunits through the export apparatus (Renault *et al.*, 2017). In addition to the discussion points made under each heading, general themes emerged for improving secretion and are reported in Chapter 7.

CHAPTER 4

MASS SPECTROMETRY-BASED PROTEOMICS ANALYSIS OF *E. coli* DURING PROTEIN SECRETION VIA FLAGELLAR TYPE III SECRETION SYSTEM (FT3SS)

4 Introduction

In Chapter 3, data was presented showing that the FT3SS was successfully engineered for heterologous protein-Cutinase export *via* the FT3SS, using enzymatic detection and western immunoassay. These data showed that a secretion signal comprising Late-5'UTR-FliC$_{47aa}$ (pSC 1) was more efficient for recombinant protein secretion through the FT3SS of a substrate-switched Δ*flgKL* strain containing an additional mutation in *clpX*. This later mutation likely relieved negative feedback on a range of FT3SS genes including *fliC, fliCD, motAB, flgKL, flgDE, flgMN, clpX, motche, tsr, trg, aer* (Gottesman *et al.*, 1993; Stafford *et al.*, 2007; Sato *et al.*, 2014; Green, 2016). The differential expression of proteins could be attributed to the strain phenotype. Therefore, in order to improve protein secretion *via* the FT3SS and further the understanding of Δ*CKL* and Δ*CKL* Δ*clpX* strains during differential protein expression, exploration of the fundamental biology of these strains was required to identify the potential regulatory proteins and pathways that might be altered in response to FT3SS action, which could be then modified to improve secretion (Ghosh *et al.,* 2012; Yadav *et al.*, 2012). In order to achieve this, a systems biology approach was utilised to understand the FT3SS and the effect of FT3SS secretion on the overall physiology of the organism using Mass Spectrometry-based proteomics. Analysis was performed by comparing Δ*CKL* and Δ*CKL* Δ*clpX* strains harbouring pSC 1 (5'UTR-FliC$_{47aa}$) and pSC 2 (No 5'UTR-FliC$_{47aa}$) secretion constructs as shown in Table 4.1. The reporter protein Cutinase (in pJExpress 404) was selected as it can be detected with a high throughput MUB assay (Section 3.2.2), as well as *via* a Flag-tag epitope in western immunoassay.

Table 4.1: Comparisons of iTRAQ analysis with secretion signal and strain.

Comparison 1 **Secretion signal: secretion strain**	**Comparison 2** **Secretion signal: secretion strain**
pSC 1: Δ*CKL* Δ*clpX*	pSC 2: Δ*CKL* Δ*clpX*
pSC 1: Δ*CKL* Δ*clpX*	pSC 1: Δ*CKL* Δ*clpX*
pSC 2: Δ*CKL* Δ*clpX*	pSC 1: Δ*CKL* Δ*clpX*

In this chapter, the assessment of the proteome was performed using an Isobaric tag for relative and absolute quantification (iTRAQ) of intracellular proteins. iTRAQ allows simultaneous identification and quantification of proteins

from multiple samples and it can expand the proteome coverage by labelling all the peptides (Ross *et al.*, 2004). It also increases the confidence in identification and quantification from the MS/MS spectra by tagging multiple peptides per protein (Zieske, 2006). Finally, it offers a simple workflow without sample fractionation for reduced-complexity of samples, by providing the flexibility to multiplex, up to eight different biological samples simultaneously, in a single experiment (Ross *et al.*, 2004; Gafken and Lampe, 2006).

Aim: The aim of this chapter was to investigate whether differential protein expression and biochemical pathways existed during protein secretion *via* the FT3SS, using high throughput mass spectrometry-based proteomics.

Results

4.1 Understanding of FT3SS secretion system using a Systems biology approach: experimental design for proteomics analysis

In order to elucidate the proteomic changes associated with the production and secretion of proteins through the FT3SS and also to interrogate the role of *clpX* in the improvement of secretion and its general influence on *E. coli* (as previously established (Chapter 3)) a series of conditions were set up and cultures were grown and induced identically. Firstly, the Δ*CKL* Δ*clpX* strain harbouring pSC 1 (5'UTR-FliC$_{47aa}$-Cutinase (Fig. 3.6)) was used as a maximally secreting strain and compared to the same strain containing a version of Cutinase that is retained in the cytoplasm (pSC 2 {no-sig}-Cutinase (Fig. 3.6)). Secondly, the Δ*CKL* strain (harbouring the pSC 1 construct) was compared to the Δ*CKL* Δ*clpX* strain (harbouring pSC 1) to establish both, the impact of the *clpX* mutation but also whether any of the changes conferred from the *clpX* deletion might be worth building upon for future iterations of strain construction. Finally, Δ*CKL* Δ*clpX* harbouring pSC 2 and Δ*CKL* harbouring pSC 1 were investigated in order to understand the response of these strains to the plasmid (see Table 4.1). In these comparisons, secretion and expression were verified by western blotting and Cutinase assay on the supernatant to verify qualitative consistency of secretion behaviour of the strains prior to protein extraction from whole cell pellets for

proteomics (iTRAQ) analysis. In these experiments, supernatant was also collected and precipitated for assay of secreted proteins (Section 2.6.4.2)

4.2 Verification of secretion strain behaviour for iTRAQ experiments

As a first verification measure, the growth parameters of these strains were established under secretion and induction conditions (LB, induction 0.05 mM IPTG) at 37 °C with shaking at OD_{600} using a plate reader (Tecan Magellan, Infinite® 200 PRO). As shown in Fig 4.1, there was no difference in growth curve shape or the final growth yield with any of the strains harbouring pJEXpress 404 (pSC 1 and pSC 2), with and without secretion signal or Δ*CKL vs* Δ*CKL* Δ*clpX* strains at any point in the growth curve.

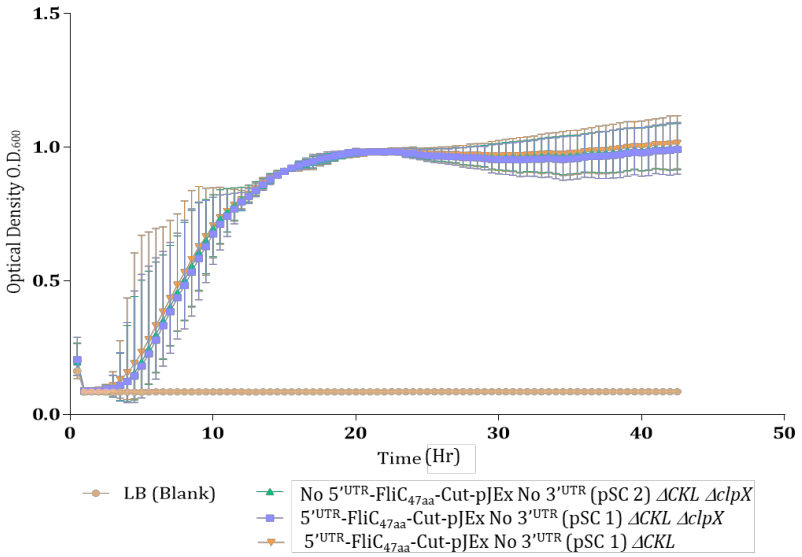


Figure 4.1: Monitoring growth curves of strains for proteomics analysis.

Growth curves were monitored for Δ*CKL* Δ*clpX* and Δ*CKL* flagella mutant strains. pSC 1 (5'-C_{47aa}-Cut-pJEx No 3') was expressed in both Δ*CKL* Δ*clpX* and Δ*CKL* cells, while pSC 2 (No 5'-C_{47aa}-Cut-pJEx No 3') into Δ*CKL* Δ*clpX* strain (Table 4.1), 3 separate transformants in triplicate were observed, OD_{600} was measured using plate reader (Tecan Magellan, Infinite® 200 PRO). Graph was plotted for mean and SEM using Graphpad Prism 7.0.

Once this had been established, we could have confidence that any changes that arose in the proteomics analysis were due to FT3SS dependent alterations rather than due to any other overall growth defects (Glick *et al.,*1995). The strains were then grown in a liquid broth culture in test conditions (10 ml cultures) for the assay of secretion *via* MUB assay. The MUB assay was performed on the supernatant (Section 3.2.2) and their secretion efficiencies were plotted

(Fig. 4.2). The assay reaction was visualised by exposing a reaction plate to UV light at 302 nm in a G-box (Syngene technologies) (Fig. 4.2a). As expected, it can be observed from Fig. 4.2a-b that the amount of Cutinase activity detected with pSC 1 was 1.27-fold higher for Δ*CKL* Δ*clpX* strain than Δ*CKL.* It can also be observed that pSC 2 supernatant showed no Cutinase activity for Δ*CKL* Δ*clpX* or (Δ*CKL*) strain due to the absence of the Type III secretion signal. (Note here, pSC 2 does not display MUB activity in supernatant of either Δ*flgKL* or Δ*flgKL* Δ*clpX,* however, only pSC 2-Δ*flgKL* Δ*clpX* was utilised for iTRAQ analysis).

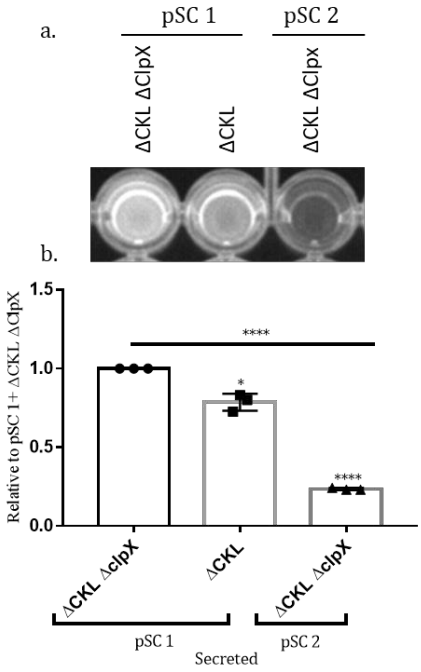


Figure 4.2: MUB assay on mutant strains subjected to proteomics analysis. Plasmid constructs were expressed as before and the MUB assay visualised on a UV transilluminator (a) and quantitatively using the previously developed MUB assay (b) (Chapter 3) to assess secretion relative to pSC 1-Δ*CKL* Δ*clpX* (triplicates with Error bars ± SEM, statistical significance defined by **** $p<0.0001$, determined by One-Way ANOVA.

Next, to confirm the data above, the supernatants were analysed by SDS-PAGE and western immunoassay (Fig. 4.3). In concordance with previous data, the data show that despite comparable expression of proteins intracellularly (Fig. 4.2), the secretion signal was obligatory for protein export through the modified FT3SS i.e no secretion was observed without the secretion signal. The efficient protein export *via* the FT3SS of Δ*CKL* and Δ*CKL* Δ*clpX* was also observed (Fig.

4.2a-c.), with a significant difference in protein export in Δ*CKL* Δ*clpX* strain over Δ*CKL* by ~25 fold (Fig. 4.3c.).

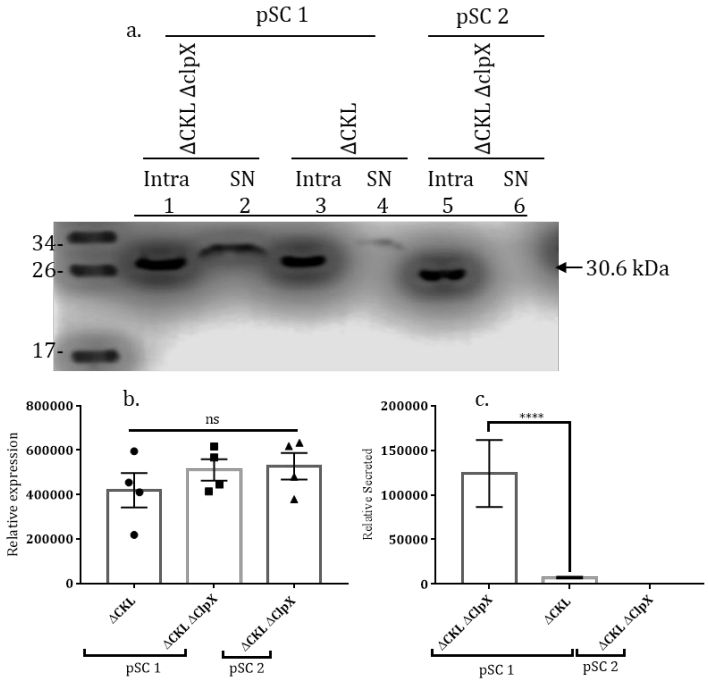


Figure 4.3: Confirmation of strains for proteomics using Western analysis. pSC 1 was expressed into Δ*CKL* and Δ*CKL* Δ*clpX* while pSC 2 into Δ*CKL* Δ*clpX.* Lanes 1, 3 and 5 were loaded with intracellular protein (intra) while lane 2, 4 and 6 with secreted proteins (SN) (a). Western blots were probed with anti-FLAG tag antibody, with visualisation on a C-digit (*LI-COR*), whole cell (b) secreted (c), data from 3-biological repeats, significance established with One-Way ANOVA for multiple comparisons, determined by *p* values ($p<0.0001$****), graphs plotted with Graphpad prism 7.0.

N.B. It could be noticed here that, there was a significant difference in the secretion of this construct in the Δ*CKL* Δ*clpX* relative to Δ*CKL,* reported in Fig. 3.12 *i.e.* secretion in this experiment was 25-fold higher in the ClpX strain, compared to 4-fold in Fig.3.12 and in Green *et al.,* 2019. This variation is in fact within the normal variation we see in secretion in these assays and in fact indicates that the Cutinase assay sometimes underestimates differences if the substrate becomes exhausted from the assay. The difference in the size of the intracellular exported protein can be observed which is mostly due to the difference in there loading volumes (2.5 µl *vs* 50 µl).

4.3 Total intracellular protein expression profile of the strains

Prior to iTRAQ analysis as a quality control (QC) and in order to examine the crude proteome of the cells, cultures were centrifuged and resuspended in 2

x SDS-lysis buffer and run on 15% SDS-PAGE gels (Section 2.6.4) and visualised with Coomassie staining (Instant Blue™, Expedeon) (Section 2.6.4.4). It can be observed (Fig. 4.4) that there was a good consistency in the intracellular protein expression profile of the replicates (3x)- a fact that provides evidence of consistency of the protein extraction protocol and serves as a quality control test for samples prior to iTRAQ analysis.

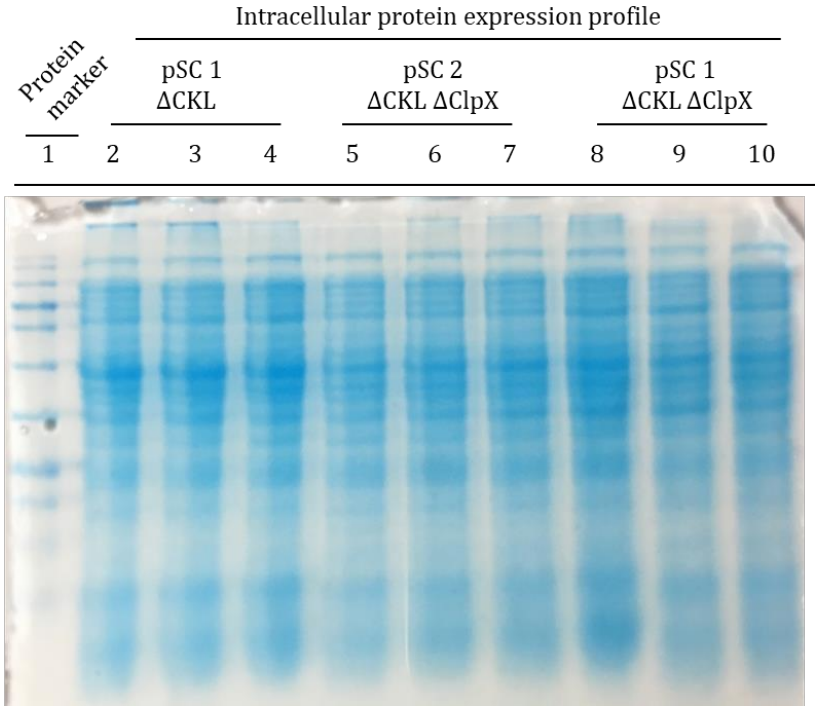


Figure 4.4: Protein expression profile of strains for proteomics analysis.
SDS-PAGE (15 %) of whole cell sample of pSC 1 and pSc 2 in the respective strains. Triplicate samples were run for each strain from biological repeat as shown. The gel was stained with Coomassie staining (Instant Blue™, Expedeon).

4.4 Isobaric tag for relative and absolute quantification (iTRAQ) analysis

As outlined in Section 4.1, the iTRAQ experiment was designed to characterise differential protein regulation in three experimental conditions. Figure 4.5 illustrates the strains used, their characteristics, and also the iTRAQ labels used for each condition and finally MS/MS analysis. As shown (Fig. 4.5), the strains and secretion constructs mentioned in Table 4.1 were expressed and grown in liquid broth culture in triplicate (3x) (Section 2.7.2) and proteins were extracted and recovered by centrifugation and digested with Trypsin (Section 2.11.1). Individual replicates were then labelled with 8-plex iTRAQ® reagents (Section 2.11.1.3 Fig. 2.7). Their secretion capacity is represented by a colour

code with high secretion (dark green), moderate secretion (pale green) and no secretion (red). (Note: pSC 1-Δ*CKL* Δ*clpX* and pSC 2-Δ*CKL* Δ*clpX* were used in triplicate while pSC 1-Δ*CKL* was used in duplicate due to limitation of 8-plex iTRAQ reagents).

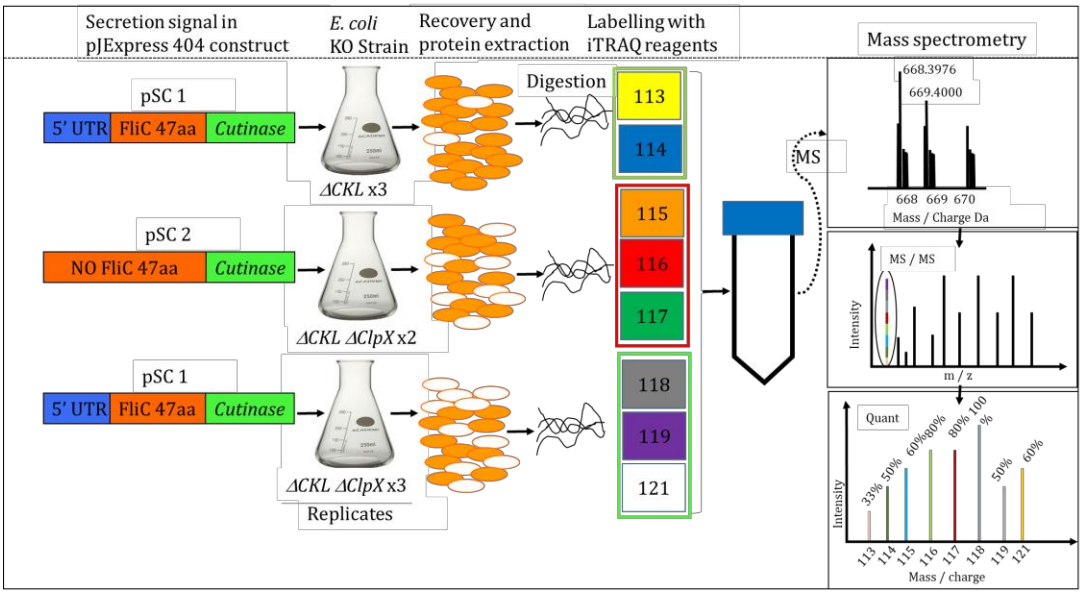


Figure 4.5: Overall schematic of iTRAQ® analysis.

Secretion constructs (pSC 1 and pSC 2) were transformed into Δ*CKL* and Δ*CKL* Δ*clpX* strains in triplicate (3 x), Total proteins were extracted, digested and labelled with iTRAQ label (113-121). Samples were subjected to MS analysis with database searching. Overall workflow, identification of MS and MS/MS analysis is shown by the bar diagram.

Samples labelled with 8-plex iTRAQ reagents were then fractionated using Ultra high performance liquid chromatography (uHPLC), collected at 2 min intervals, concentrated and combined prior to loading on to a Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific, Germany), coupled with a nano uHPLC 3000 system (Dionex, UK) tandem mass spectrometer (Section 2.11.1.4). The data obtained is represented in Section 4.6.

4.4.1 Quality control of data obtained from mass spectrometry analysis

The quality control of MS performance was monitored using a Hela standard (Pierce™, Thermo scientific) while the quality of biological replicates from iTRAQ data was analysed with a Dendrogram and Principal component analysis (PCA) using Perseus software (Tyanova *et al.*, 2016). The distribution of all the 3 replicates are plotted (Fig.4.6). Samples for pSC 1-Δ*CKL (*113.1-114.1*)*

(blue) cluster close to one another, as do the pSC 2-Δ*CKL* Δ*clpX* (115.1, 116.1, 117.1 (red)). In contrast, pSC 1-Δ*CKL* Δ*clpX* (118.1, 119.1, 121.1) samples are less well-clustered (green) but overall, all three samples appear to cluster in different parts of the graph space, indicating that three data sets are more distinct from each other than within their own replicates. Importantly, the blue samples that were only used in duplicate clustered well together, meaning that we can have confidence in the dataset.

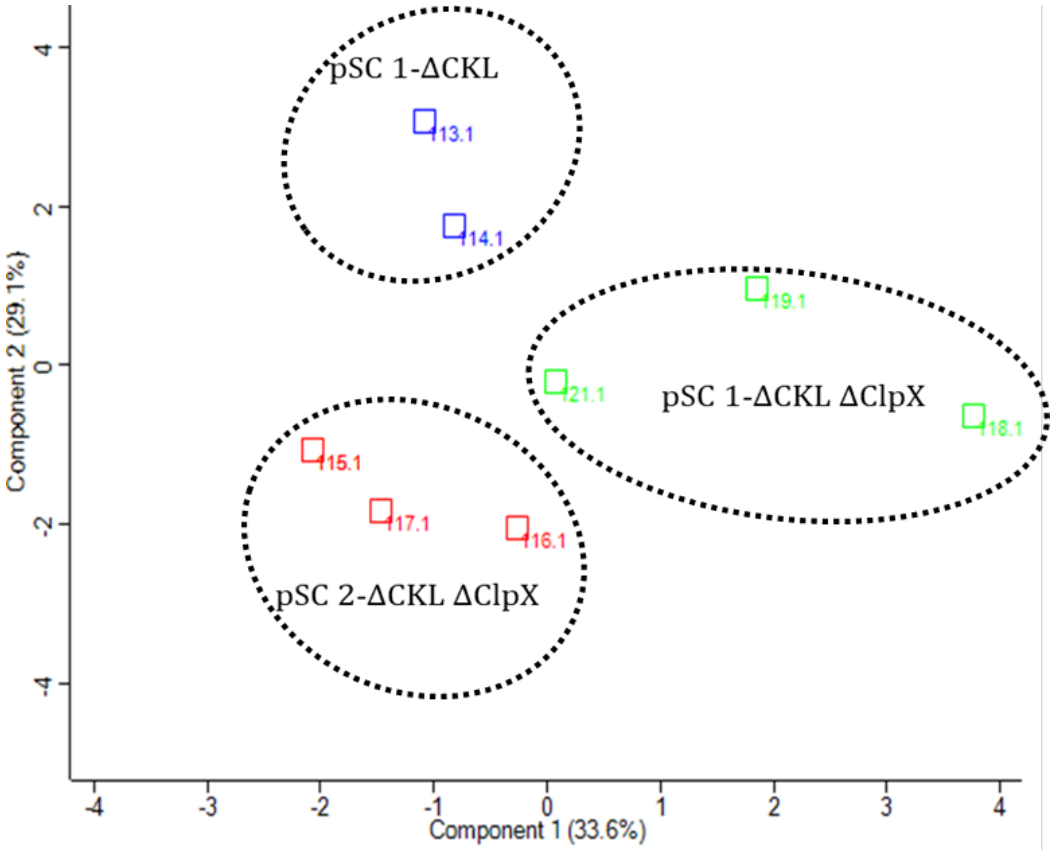


Figure 4.6: Distribution of protein expression in biological replicates with Principal component analysis (PCA).
Biological replicates of the iTRAQ samples were run on mass spectrometer, their consistency was verified by plotting scatter plot (generated by using *Perseus.exe*).

As a next step, the levels of each peptide from each protein were compared between each dataset to produce a measure of relative abundance based on the amount of each iTRAQ labelled peptide in each MS peak.

4.5 iTRAQ data analysis

Using median corrected data, each relative quantification was converted into a colour and represented in a heat map using the *perseus.exe*. Fig. 4.7 shows

these data alongside hierarchical clustering of the samples (horizontal-top) and peptide data (vertical-side). It can be observed that complete proteome expression of the individual replicates of pSC 1-Δ*CKL* (113.1-114.1), pSC 2-Δ*CKL* Δ*clpX* (115.1, 116.1, 117.1) and pSC 1-Δ*CKL* Δ*clpX* (118.1, 119.1, 121.1) all cluster together, as hoped with PCA analysis (Section 4.6), giving credence to more in-depth analysis presented in the remainder of this chapter. Following statistical analysis up and downregulated proteins during (1) pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX,* (2) pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* and (3) pSC 2-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparisons were obtained, shown in Table 4.2, 4.3 and 4.4, respectively.

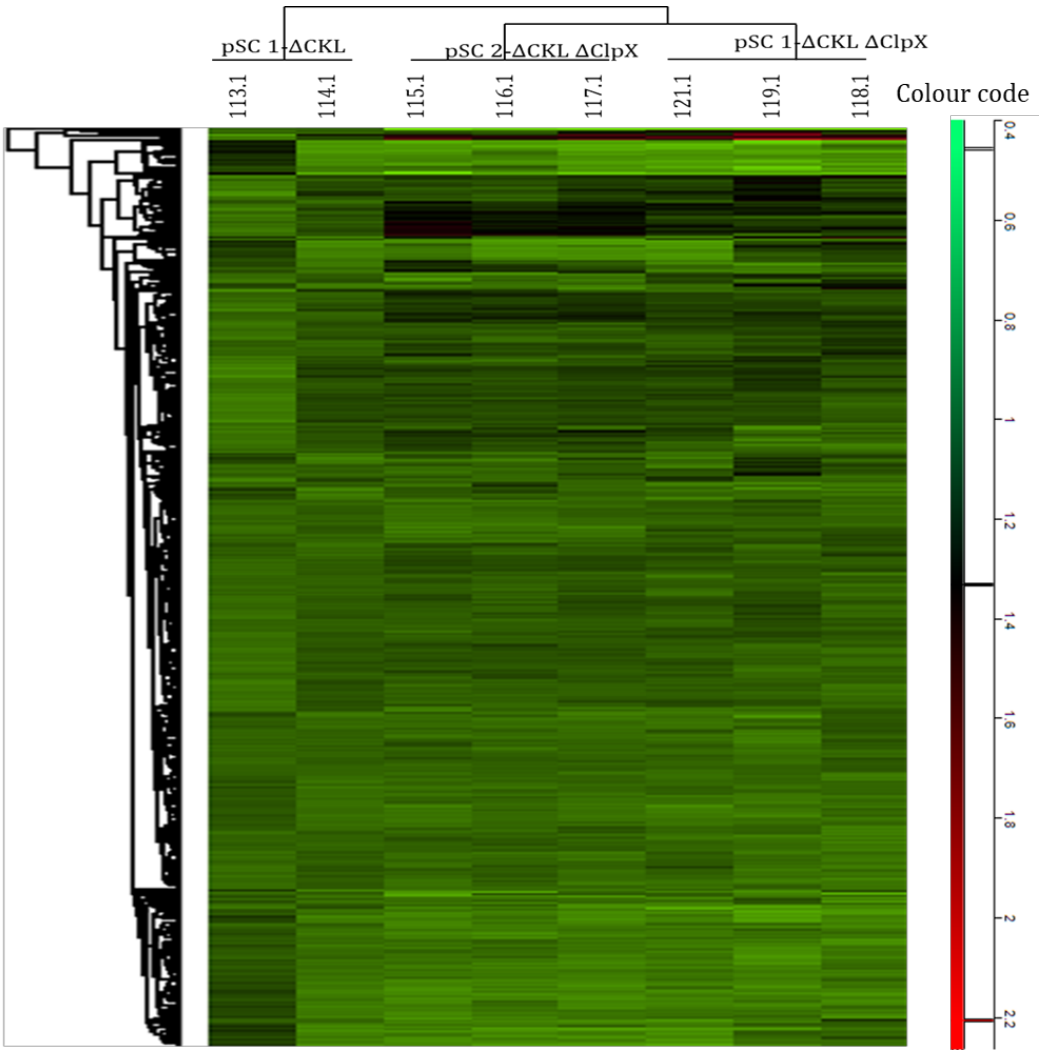


Figure 4.7: Clustering and proteome of the iTRAQ labelled peptide data.
Independent iTRAQ replicates were labelled as shown in Section 4.5 in duplicates (2x) and triplicates (3x). Clustering of the samples is shown, 113 114 (pSC 1-Δ*CKL*), 115 116 117 (pSC 2-Δ*CKL* Δ*clpX)* and 118 119 121 (pSC 1-Δ*CKL* Δ*clpX*) represents the iTRAQ labelling. Heat map generated by using *Perseus.exe* program, colour code shown on right.

4.6 Investigation of protein regulation highlighted in different comparisons of iTRAQ-proteomics analyses

4.6.1 iTRAQ data analysis pipeline

iTRAQ samples were fractionated and combined prior to analysis and a total of 14675 peptides were derived from 12 fraction combinations, corresponding to 1814 proteins. Data were analysed in order to determine which protein/peptides were significantly altered between experimental conditions tested. For these analyses, only proteins identified with ≥ 2 unique peptides were used for quantification with median and isobaric corrections applied and paired t-tests were then performed on geometric means of reporter intensities to determine regulated proteins (1411 proteins) for each phenotypic comparison. Quantified proteins with their *p*-values from each t-test, ≤ 0.05 were considered as regulated (Appendix: Table 8.4-8.9) while to establish statistically significantly regulated proteins, a Bonferroni correction was applied (based on the discovery of 1411 proteins), meaning only changes with a *p*-value ≤ 3.54×10^{-5} were considered as significant (Section 4.8- see methods also (Section 2.11)). Protein quantifications were obtained by computing the geometric means of reporter intensities. Median correction was applied to every reporter in each individual MS/MS scan.

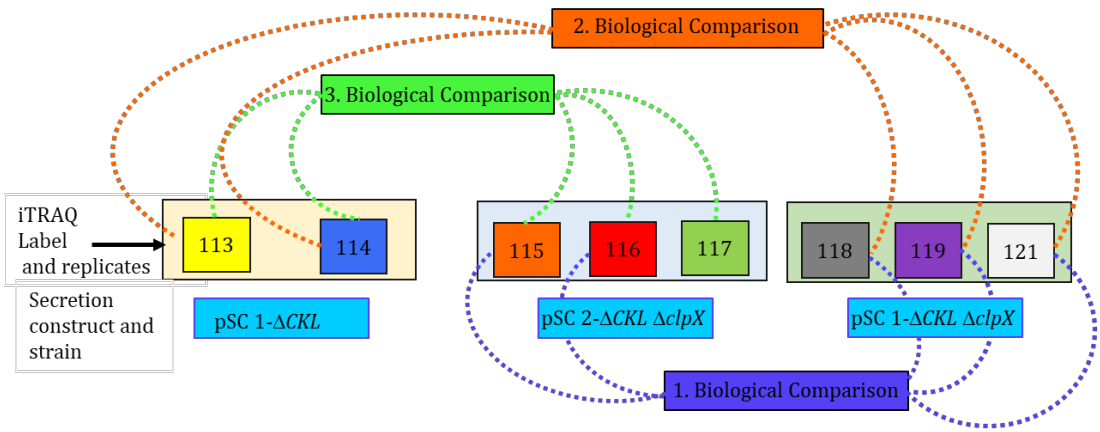


Figure 4.8: iTRAQ labelling and comparison of replicates.

Proteomics samples pSC 1-Δ*CKL*, pSC-Δ*CKL* Δ*ClpX* and pSC 1-Δ*CKL* Δ*ClpX* were labelled with 8-plex reagents and replicates were compared as shown with dotted lines. pSC-secretion construct, Δ*CKL* and Δ*CKL* Δ*ClpX* represents strains.

4.6.1 Regulation of proteins in Δ*CKL* Δ*clpX* harbouring pSC 1 and pSC 2

As outlined above, the first set of experiments compared the proteome of Δ*CKL* Δ*clpX* containing the secretion competent (pSC 1) and non-secreting (pSC 2) in 118 119 121 and 115 116 117 comparison. The iTRAQ data obtained was analysed (Section 2.11.1.5 and 2.11.1.6) based on their fold change and *p*-values. In this comparison, 38 regulated proteins were identified ($p<0.05$), but only 9 were considered significantly upregulated and 3 significantly downregulated using the Bonferroni correction method (Table 4.2).

Table 4.2: Regulated protein in pSC 1-Δ*CKL* Δ*clpX* and pSC 2-Δ*CKL* Δ*clpX* comparison.

(Note: Those genes indicated by * are just below threshold, but are notable exceptions highlighted for reasons explained in text).

Sr No	Protein ID	Gene name	Protein names	Fold change	*p*-Value
Upregulated proteins in pSC 1-Δ*CKL* Δ*clpX*					
1	P08622	*dnaJ*	Chaperone protein DnaJ	1.30	7.4E-05
2	P0A6F5	*groL*	60 kDa chaperonin	1.36	1.4E-09
3	P0A6Y8	*dnaK*	Chaperone protein DnaK	1.39	7.7E-14
4	POA6Z3	*htpG*	Chaperone protein HtpG	1.32	2.2E-09
5	P0A9M0	*Lon*	Lon protease	1.20	5.6E-05
6	P27298	*prlC*	Oligopeptidase A	1.19	2.0E-05*
7	P63284	*clpB*	Chaperone protein *ClpB*	1.32	1.0E-09
8	P69910	*potA*	Spermidine/putrescine import ATP-binding protein PotA	1.31	9.5E-06
9	P9999	*cutA*	Cutinase	1.57	4.4E-06
Downregulated proteins in pSC 1-Δ*CKL* Δ*clpX*					
10	P07363	*cheA*	Chemotaxis protein CheA	- 1.28	8.9E-06
11	P37329	*modA*	Molybdate-binding periplasmic protein	- 1.31	1.2E-05*
12	P75937	*flgE*	Flagellar hook protein FlgE	- 1.31	4.0E-05

Firstly, it can be observed (Table 4.2) from fold changes and *p*-values that during pSC 1-Δ*CKL* Δ*clpX* and pSC 2-Δ*CKL* Δ*clpX* comparison, the target protein-Cutinase was present significantly higher levels, ($p < 4.4E-06$).

The data also highlight upregulation of a number of proteins that might be considered as part of a stress response. These include upregulated proteins ATPase dependent folding chaperone DnaJ (HspJ40) (1.30 fold, $p<7.4E-05$) and 60 kDa chaperonin GroL (Hsp60) (1.36-fold, $p<1.4E-09$) and protein

disaggregating chaperone ClpB (Hsp100) (1.32 fold, p<1.0E-09), along with chaperone protein DnaK (Hsp70) (1.39 fold, p<7.7E-14) and Chaperone protein HtpG (Hsp90) (1.32 fold, p<2.2E-09). In this dataset one imagines most responses are due to the extra burden of secreted protein over it being held in the cytoplasm. However, some of these responses may have actually helped to boost flagellin synthesis *via* boosting FlhDC and σ^F dependent expression due to overexpression of DnaK/J. This idea suggested by the findings of Shi *et al.,* (1992), who reported a 10-20 fold decrease in flagella gene expression due to loss of DnaK/DnaJ and suggested, they may play a role in FlhDC stability and stability of other flagella proteins (Shi *et al.*, 1992; Li *et al.*, 1993).

Of note in relation to this dataset is that Tsao *et al.,* reported the coincident upregulation of the chaperone GroEL during enhanced protein production, where it was demonstrated that the increased protein production *via* GroEL was as a result of AI-2 (luxS) signalling and not due to stress (Tsao *et al.*, 2011). Similarly, Lon protease is an ATP dependent protease responding to stress (Higashitani *et al.*, 1997) and was also (1.20-fold, p < 5.6E-05) upregulated during high secretion. Lon protease is involved in the degradation of misfolded and regulatory proteins and prevention of the aggregation of proteins (Chung and Goldberg, 1981). Fitting in with these ideas, the significant upregulation in Lon protease was also reported in swarming cells where FlhDC levels are elevated-as they may be here as a result of DnaKJ upregulation (Claret and Hughes, 2000; Clemmer and Rather, 2008). Additionally, upregulation of DnaKJ may be indicate the increased stress, which might be in response to increased errors in transcription, and therefore DnaKJ may carry out RNA degradation (Kedzierska *et al.*, 1999). However, the downregulation of the Class III gene *cheA* and Class II gene *flgE* do not really makes sense in the context of potential upregulation of flagella genes, and in fact indicate that something else is causing downregulation of Class II and III regulation in these cells- although at this stage the mechanism for this is unclear.

4.6.2 Regulation of proteins in Δ*CKL* Δ*clpX* and Δ*CKL* strains harbouring the same secretion construct (pSC 1)

A second comparison of iTRAQ experiments compared expression of the same secretion construct, namely 5'UTR-FliC$_{47aa}$ (pSC 1) in both Δ*CKL* Δ*clpX* and Δ*CKL* strains (*i.e.* high and low secretion: labels 118 119 121 *vs* 113 114). As mentioned, Δ*CKL* Δ*clpX* secretes protein more efficiently than the Δ*CKL,* due to the exclusion of the *clpX* gene. ClpX protein is a negative regulator of $FlhD_4C_2$ complex (Tomoyasu *et al.*, 2002) while the presence of *clpX*, maybe downregulate the protein secretion in Δ*CKL* strain (Fig. 4.2). In this comparison, 495 regulated proteins were detected in which 50 proteins were significantly abundant (signifiacantly expressed proteins but below threshold). However, 6 proteins were significantly upregulated and 4 were downregulated. Their protein ID's, gene names, *p*-values and fold changes are reported in Table 4.3. Many of these changes in the protein regulation could be attributed to the *clpX* gene as both the strains expressed the same type III secretion signal harbouring (5'UTR-FliC$_{47aa}$) pSC 1. C*lpX* is an ATP dependent molecular chaperone that serves as a substrate-specifying adaptor for ClpP serine protease in ClpXP and ClpAxP protease complexes -which is a member of AAA$^+$ (ATPases associated with diverse cellular activities) family of ATPases (Baker and Sauer, 2012; Olivares *et al.,* 2015).

Table 4.3: Protein upregulation identified from pSC 1-Δ*CKL* Δ*clpX* and pSC 1-Δ*CKL* comparison of iTRAQ analysis.

Sr No	Protein ID	Gene name	Protein names	Fold change	*p*-Value
Protein Upregulation					
1	P0A6F5	*groL*	Cpn60 chaperonin GroEL, the large subunit of GroESL	1.192	2.5E-05*
2	P0A6Y8	*dnaK*	chaperone Hsp70, with co-chaperone DnaJ	1.239	7.6E-06
3	P0ABT2	*dps*	Fe-binding and storage protein; stress-inducible DNA-binding protein	1.577	9.7E-06
4	P22256	*gabT*	4-aminobutyrate aminotransferase, PLP-dependent	1.234	8.6E-05
5	P69910	*gadB*	glutamate decarboxylase B, PLP-dependent	1.575	4.4E-08
6	P9999	*cutA*	Cutinase	1.468	1.1E-05*
Protein Downregulation					

7	P0A6H1	*clpX*	ATPase and specificity subunit of *clpX*-ClpP ATP-dependent serine protease	- 2.388	7.5E-08
8	P0ABQ2	*garR*	tartronate semialdehyde reductase	- 1.333	3.4E-05*
9	P38104	*rspA*	bifunctional D-altronate/D-mannonate dehydratase	- 2.033	2.1E-07
10	P38105	*rspB*	putative Zn-dependent NAD(P)-binding oxidoreductase	- 2.498	7.0E-05

Firstly, despite not meeting the strict threshold of $p<3.5x10^{-5}$, it can be observed (Table 4.3) that the target protein Cutinase was significantly upregulated in the pSC 1-Δ*CKL* Δ*clpX* strain than pSC 1-Δ*CKL* even though expressed from identical promoters. This indicates that a ClpX dependent upregulation of the promoter in this plasmid may be occurring. Similarly, ClpX was detected at a vastly reduced level in the mutant, effectively zero given that iTRAQ underestimates differences.

As in the first comparison above (Section 4.6.1), the secretion of more protein in the Δ*CKL* Δ*clpX* strain, also results in increase in the levels of the stress protein DnaK (1.23-fold, $p<7.6E-06$). As well as the stress protein Dps (1.57-fold, 9.7E-06) while there are others that are regulated at a less stringent level of statistical analysis- such as GroEL (cpn60, $p=2.2x10^{-5}$), potentially due to higher protein production and secretion levels putting stress on the cell due to increased protein turn over as indicated by the upregulation of protein folding and disaggregation chaperones involved in RNA degradation pathway (Gupta *et al.*, 2009; Berlec and Štrukelj, 2013; Schlegel *et al.*, 2013). It may be also suggesting that *clpX* negatively regulates both ATPase dependent folding chaperones and protein disaggregation chaperones, as shown by the significant upregulation of DnaJ, GroEL, DnaK, and ClpB proteins, in the *clpX* deficient strain (Table 4.3). However, these chaperones were also significantly upregulated in pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* comparison (Section 4.1.2, Table 4.2), suggesting that FliC signal maybe the regulator of these chaperones, however, the mechanism remains enigmatic.

Of the other regulated proteins (but not meeting threshold) the metabolic pathways altered are of great interest. These include proteins in the butanoate and alanine, aspartate and glutamate metabolism pathways; many of which seem

to cluster in pathways and operons, indicating this is unlikely to be by chance despite not meeting the Bonferroni correction. Specifically, these are *gabA*, Succinate-semialdehyde dehydrogenase I, NADP dependent *gabD* (1.233 fold, $p<0.003$) (Table 8.6) and 4-aminobutyrate aminotransferase, PLP-dependent *gabT* (1.23 fold, $p<8.6E-05$) enzymes (Table 4.2) (Dover and Halpern, 1972; Cozzani *et al.*, 1980). These significantly abundant proteins leads to conversion of biochemical intermediates of different metabolic pathways, amino acids biosynthesis into the biochemical energy currency of the cells e.g. the conversion to succinate, which is a Citrate cycle intermediate, and leads to energy generation (Guest and Russell, 1992). This hypothesis could be supported by the upregulation of Pyruvate oxidase *poxB* (1.2 fold, $p<0.002$) (Appendix Table 8.6), which is a peripheral membrane enzyme that catalyses the oxidative decarboxylation of pyruvate to form acetate and CO_2, coupled with electron transport chain *via* ubiquinone (Abdel *et al.,* 2001; Vemuri *et al.*, 2005). Overall, suggesting the upregulation of central metabolism to meet the increasing demand of biochemical energy required for protein export in pSC 1-Δ*CKL* Δ*clpX* than pSC 1-Δ*CKL* comparison (Weber *et al.,* 2002).

The upregulation of the amino acid degradation pathway maybe indicates the increased demand for biochemical energy during protein production in the *clpX* deficient strain compared to the Δ*CKL* strain. It has been also reported that the TCA cycle has limited capacity and remains constant during increased demand for biochemical energy for protein secretion (Heyland *et al.,* 2011), therefore it appears that maybe the alternative energy sources are utilised with the upregulation of these alternate biochemical pathways. An engineered host for ATP and improved protein secretion have also been reported using the alternative energy sources for metabolism (Kim *et al.*, 2012). In addition, altered expression of more than one enzyme from these pathways gives confidence in the results, suggesting that this pathway (butanoate and alanine, aspartate and glutamate metabolism pathways) serves as an alternate either energy source or amino acid pool in high protein export condition (Cozzani *et al.*, 1980; Mahalik *et al.,* 2014). This might open the possibility that external provision of certain components of central metabolic pathways directly into the growth medium

might improve protein secretion, by the utilisation of biochemical energy for protein export rather than investing it into biochemical synthesis (Hoffmann and Rinas, 2001; Mahalik *et al.,* 2014).

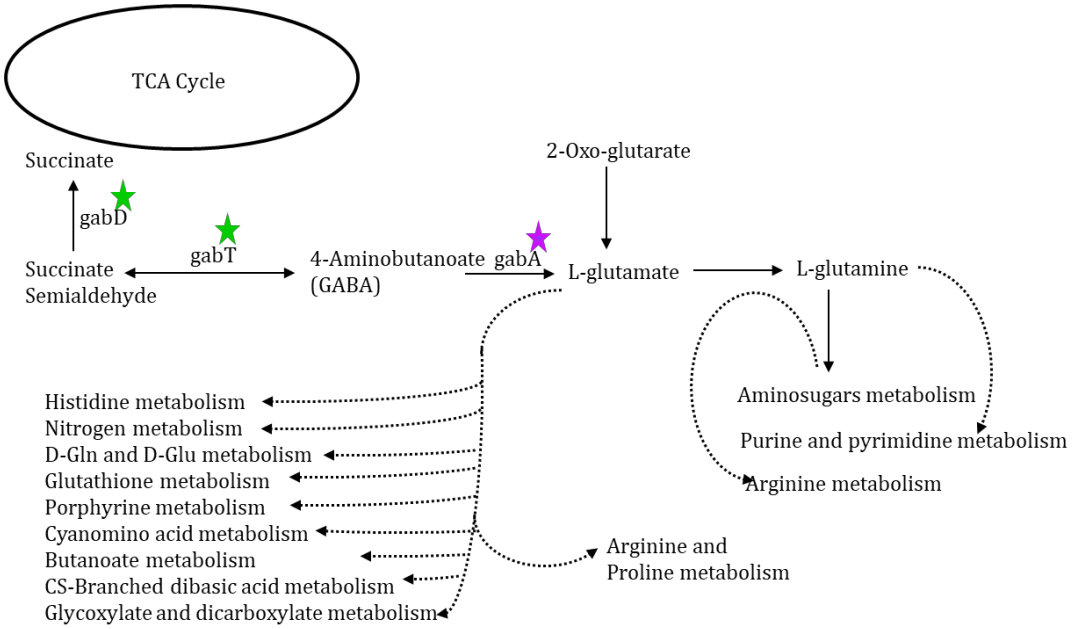


Figure 4.9: Upregulation of metabolic pathway.
The alanine, aspartate, and glutamate metabolism, the green sign represent significantly upregulated proteins, while purple sign shows abundant protein (proteins below threshold) (Appendix: Table 8.4-8.9). Adapted from (D. Huang *et al.,* 2009; Keseler *et al.*, 2017).

The tartarate semialdehyde reductase *garR* (-1.33-fold, p<3.4E-05) was also significantly downregulated. It catalyses the reduction of tartronate semialdehyde, which is one of two isozymes in *E. coli* to yield glycerate (Hubbard *et al.*, 1998). It could be inferred that during high protein secretion (pSC 1-Δ*CKL* Δ*clpX*), cells might experience increased demand for biochemical energy to carry out protein export and therefore may utilise biochemical energy for protein secretion. This might lead to the downregulation of anabolic pathways, such as the energy conversion into Glycerolipid metabolism, Inositol phosphate metabolism, Ascorbate and alternative metabolism or amino acid and nucleoside sugar metabolism (Fig. 4.9). In fact. this theory was supported by the significant downregulation of proteins such as regulatory stationary phase A, *rspA* that encodes a mandelate racemase/muconate lactonising enzyme family protein (-2.033 fold, p<3.4E-05) but which has been implicated in the starvation response

in *E. coli* (Huisman and Kolter, 1994) as well as recombinant protein production (Weikert *et al.*, 2000) and, *rspB* encodes a putative zinc-binding dehydrogenase (-2.498 fold, p<7.0E-05) (Table 4.2). Therefore, Δ*CKL* Δ*clpX* maybe utilises the conserved biological energy pool to direct protein export *via* the FT3SS in pSC 1-Δ*CKL* Δ*clpX*: high secretion than pSC 1-Δ*CKL*-moderate secretion comparison.

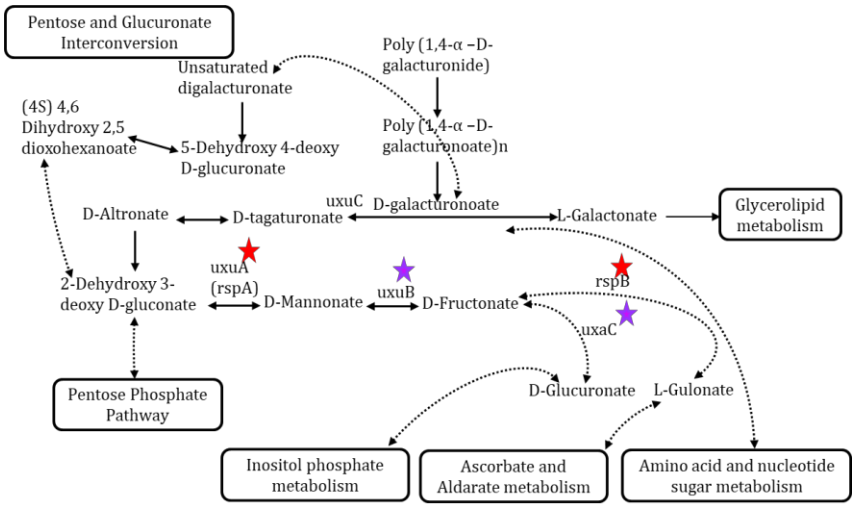


Figure 4.10: Downregulation of biochemical pathway.
Showing pentose glucuronate interconversion pathway, significantly down regulated proteins are shown in red, while abundant proteins are shown in purple sign and reported in Appendix (Table 8.7). Adapted from (Huang *et al.,* 2009; Keseler *et al.*, 2017).

4.6.3 The difference in regulation of proteins in strains Δ*CKL* Δ*clpX* harbouring pSC 2 and Δ*CKL*-harbouring pSC 1

In this experimental comparison (115, 116, 117 *vs* 113, 114), pSC 2 (no secretion signal) was expressed in the Δ*CKL* Δ*clpX* strain while pSC 1 (with secretion signal) was expressed in the Δ*CKL* strain. This comparison therefore allows us to look at the effects of ClpX again- where we would expect similar changes to the last comparison if these were solely due to ClpX, and any that differ might be due to secretion only. In this comparison, 604 proteins were identified in which 107 proteins were expressed in abundance. However, 11 proteins were significantly upregulated (gabT: abundant protein) and 7 were downregulated (dnaK: abundant protein), their protein ID's, gene names, p-values and fold changes are reported in Table 4.4.

Table 4.4: Regulated proteins in pSC 2-*ΔCKL* Δ*clpX vs* pSC 1-*ΔCKL* comparison

Sr No	Protein ID	Gene name	Protein names	Fold chain	P-Value
Protein Upregulation					
1	P07003	*poxB*	pyruvate dehydrogenase, thiamine triphosphate-binding, FAD-binding	1.35	7.7E-06
2	P07363	*cheA*	fused chemotactic sensory histidine kinase in a two-component regulatory system with CheB and CheY: sensory histidine kinase/signal sensing protein	1.43	1.25E-07
3	P0A8G6	*wrbA*	NAD(P)H: quinone oxidoreductase	1.35	6.80E-05
4	P0ABT2	*dps*	Fe-binding and storage protein; stress-inducible DNA-binding protein	1.40	1.26E-06
5	P0AD59	*ivy*	inhibitor of c-type lysozyme, periplasmic	1.32	3.92E-05
6	P0AEM9	*fliY*	cystine transporter subunit	1.25	7.50E-06
7	P0AEX9	*malE*	maltose transporter subunit	1.33	5.6E-06
8	P18843	*nadE*	NAD synthetase, NH3/glutamine-dependent	1.14	7.6E-05
9	P22256	*gabT*	4-aminobutyrate aminotransferase, PLP-dependent	1.23	1.04E-05*
10	P23847	*dppA*	dipeptide/heme ABC transporter periplasmic binding protein; dipeptide chemotaxis receptor	1.22	3.6E-06
11	P37329	*modA*	molybdate ABC transporter periplasmic binding protein; chlorate resistance protein	1.30	3.7E-06
12	P75937	*flgE*	flagellar hook protein	1.42	2.8E-07
Protein Downregulation					
13	P0A6F5	*groL*	Cpn60 chaperonin GroEL, large subunit of GroESL	- 1.13	6.7E-06
14	P0A6H1	*clpX*	ATPase and specificity subunit of *clpX*-ClpP ATP-dependent serine protease	- 2.18	8.6E-09
15	P0A6Y8	*dnaK*	chaperone Hsp70, with co-chaperone DnaJ	- 1.12	2.8E-05*
16	P0A6Z3	*htpG*	protein refolding molecular co-chaperone Hsp90, Hsp70-dependent; heat-shock protein; ATPase	- 1.20	3.6E-05
17	P0ABB4	*atpD*	F1 sector of membrane-bound ATP synthase, beta subunit	- 1.13	8.6E-06
18	P38104	*rspA*	bifunctional D-altronate/D-mannonate dehydratase	- 1.93	4.5E-08
19	P38105	*rspB*	putative Zn-dependent NAD(P)-binding oxidoreductase	- 2.28	4.4E-05
20	P63284	*clpB*	protein disaggregation chaperone	- 1.16	1.3E-07

In this dataset, FliY (1.25 fold, p<7.50E-06), *FlgE* (1.42 fold, p<2.8E-07), maltose ABC transporter periplasmic binding protein *MalE* (1.3 fold, p< 5.6E-06), NAD synthetase NH3-dependent enzyme *NadE* (1.14 fold, p<7.6E-05), dipeptide ABC transporter periplasmic binding protein *DppA* (1.22 fold, p<3.6E-06) were significantly upregulated. Upregulation of *FliY* was also observed in pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparison (Table 4.2). Upregulation of stress response proteins such as *NadA* and *DppA*, suggests that maybe nutrients were depleted in the medium and cells were experiencing stress in pSC 2-Δ*CKL* Δ*clpX*. On the contrary, in pSC 1-Δ*CKL* these stress proteins were downregulated. It could be further investigated whether external supply of nutrient-rich medium to the pSC 2-Δ*CKL* Δ*clpX* cultures would be beneficial. Upregulation of these proteins could also be attributed to the absence of *clpX* in pSC 2-Δ*CKL* Δ*clpX,* as *clpX* is a protease associated with diverse cellular activities (Gottesman *et al.*, 1993; Neuwald *et al.*, 1999). During pSC 2-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* (Table 4.3) it can be observed that *clpX* protein was significantly downregulated in Δ*CKL* Δ*clpX* than Δ*CKL* (-2.18, p<8.60E-09) as *clpX* gene is absent in pSC 2-Δ*CKL* Δ*clpX* but present in pSC 1-Δ*CKL.*

Additionally, the F1 sector of membrane-bound ATP synthase, alpha subunit *AtpA* (-1.10, p<0.002 (Appendix: Table 8.9)) and F1 sector of membrane-bound ATP synthase, beta subunit *AtpD* (-1.13 fold, p<8.6E-06) from oxidative phosphorylation were also significantly downregulated (Table 4.4). Similar to Fig. 4.10, a significant downregulation of pentose and glucoronate interconversion pathway genes such as regulatory stationary phase A, mandelate racemase/muconate lactozising enzyme family protein *RspA* (-1.93 fold, p<4.5E-08) and putative zinc-binding dehydrogenase, *RspB* (-2.28 fold, p<4.4E-05) (Table 4.4) were observed. Downregulation of chaperone protein HtpG (-1.20 fold, p<3.6E-05) which helps in folding of newly formed peptides was observed. It is also noted that *GroL* (-1.13 fold, p<6.7E-06), *DnaK* (-1.12 fold, p<2.8E-05 (Abundant proteins: Table 8.8)), *ClpB* (-1.16 fold, p<1.3E-07) proteins were significantly downregulated in this comparison, were significantly upregulated in pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* (Table 4.2) and pSC 1-Δ*CKL* Δ*clpX vs*

pSC 1-Δ*CKL* (Table 4.3). From these observations, it appears that, maybe these proteins are necessary for protein secretion. Hence, it might be worth investigating these proteins to engineer the host strain for efficient protein secretion *via* the FT3SS.

It can also be observed that chemotactic protein CheA involved in bacterial chemotaxis was significantly upregulated in the ClpX strain (1.43-fold, p<1.25E-07). Upregulation of chemosensory CheA and significant abundance of CheB, CheD, CheY, CheW (Appendix: Table 8.8) maybe suggests that in absence of secretion- a conserved pool of biochemical energy is utilised for chemotaxis. Upregulation of these methyl-accepting proteins (MCPs) CheA, CheB, CheD, Chey, CheW, CheZ has been reported during carbon deficiency in the medium that in turn upregulate *flhDC* levels, leading to more flagella gene expression (Adler, 1969; Springer *et al.,* 1977). This leads to gradual increase in motility in low-quality carbon supplies by spending a large amount of energy on flagella-filament synthesis flagella filament is made of 20-30,000 subunits of flagellin *fliC* (Aldridge *et al.*, 2006). It makes almost 8 % of total cell protein and spending ~2 % and 0.1 % of energy on flagellar synthesis and operation, respectively (Macnab and Koshland, 1974). This energy intensive flagella in poor nutrient environment triggers a high risk of exhausting the sole energy supply referred to as risk-prone foraging, in which bacteria take a risk and utilise the flagellar system to actively search for better conditions for growth and survival (Hastjarjo *et al.,* 1990). However, during pSC 2-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparison, these proteins were significantly upregulated in pSC 2-Δ*CKL* Δ*clpX* than pSC 1-Δ*CKL,* in which the former does not secrete proteins. This maybe suggests that biochemical energy is utilised for protein export in the later, pSC 1-Δ*CKL.* Also, it was observed that chemotactic pathway was downregulated in pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* (Fig. 4.5, Table 4.2), adding confidence to the multiple comparisons carried out using independent replicates.

Additionally, Pyruvate oxidase *PoxB*-a peripheral membrane enzyme catalyses oxidative decarboxylation of pyruvate to acetate and CO_2, coupled with electron transport chain was upregulated (1.35-fold, p<7.7E-06). PoxB in the

stationary phase induces main pathway of acetate production, which is inhibitory to normal cell growth as reported in Chapter 5 (Fig. 5.22). It was also observed that NAD(P)H: quinone oxidoreductase *WrbA* was upregulated (1.35 fold, p<6.80E-05), which has been reported to increase on acetate during stationary phase and it is dependent of RpoS and Crl (Yang *et al.,* 1993). Also, upregulation of *Dps* (1.40 fold, p<1.26E-06) which is a highly abundant protein in stationary phase required for long-term viability and involved in protection from multiple stresses such as oxidative stress maybe suggests the onset of stationary phase as periplasmic chaperone, inhibitor of vertebrate C-type lysozyme-*Ivy* was also upregulated (1.32 fold, p<3.92E-05), which is highly expressed during stationary phase (Alimi *et al.*, 2000).

4.7 Discussion: shared protein pathways and their implications

As reported in Chapter 3, a combination of an efficient secretion signal (5'UTR-FliC$_{47aa}$) and secretor strain (Δ*CKL* Δ*clpX*) was tested for protein export *via* the FT3SS. Sequential exclusion of a number of accessory flagella genes and proteases, such as *clpX,* were removed by mutagenesis to create a secretor strain with significantly improved efficiency of protein export. Encouraged with these results, a systems biology approach was utilised to find additional targets for improving secretion *via* the FT3SS. Therefore, the most efficient secretion signal pSC 1 (5'UTR-FliC$_{47aa}$) which gave differential protein secretion in Δ*CKL* and Δ*CKL* Δ*clpX* strains was investigated along with pSC 1, which secretes protein and pSC 2 which does not secrete proteins *via* FT3SS (due to absence of the secretion signal). Furthermore, as explained in Table 4.1 in order to investigate the effect of ClpX on the Δ*CKL* Δ*clpX* and Δ*CKL* strains harbouring pSC 1 were compared. Also, the protein up/downregulation in the Δ*CKL* Δ*clpX* strain harbouring pSC 1 or pSC 2 was investigated with proteomics analysis (Section 4.4).

4.7.1 iTRAQ analysis of strains for improving protein export *via* FT3SS

4.7.1.1 Biochemical regulation in the secretor strain (Δ*CKL* Δ*clpX*) independently harbouring pSC 1 and pSC 2 plasmids

In pSC 1-Δ*CKL* Δ*clpX* *vs* pSC 2-Δ*CKL* Δ*clpX*, where pJExpress 404 with (pSC 1) and without (pSC 2) secretion signal were expressed in the same strain, Δ*CKL* Δ*clpX*. pSC 1 efficiently secreted the heterologous protein-Cutinase *via* the FT3SS

of Δ*CKL* Δ*clpX* strain; while pSC 2, in absence of secretion signal, did not secrete Cutinase, which remained inside the cell (Table 4.2). It was confirmed with MUB assay and the western blot analysis (Fig. 4.2 and 4.3). Detection of Cutinase in the medium of pSC 1-Δ*CKL* Δ*clpX* and its complete absence in pSC 2-Δ*CKL* Δ*clpX,* could be due to the absence of 5'UTR-FliC$_{47aa}$ secretion signal. In order to investigate the role of secretion signal in Δ*CKL* Δ*clpX* and how it influences proteome expression during secretion and non-secretion, iTRAQ analysis was performed. It was observed that, the target protein Cutinase was significantly upregulated (1.57-fold, p<4.4E-06) during pSC 1-Δ*CKL* Δ*clpX* than pSC 2-Δ*CKL* Δ*clpX* comparison (Table 4.2). It can also be noted that Cutinase concentration was significantly higher in the supernatant than intracellularly which supports one of the aims of this study of establishment of continuous protein secreting culture without the need to sacrifice the cultures (Warikoo *et al.*, 2012; Peebo and Neubauer, 2018). Continuous protein secreting cultures is associated with several benefits such as lower running costs, reduced equipment size, process scalability and product quality (Walther *et al.*, 2015). However, further engineering of the strain might be required in order to establish the concept of continuous culture with respect to growth parameters, production medium, number of batches or the adverse effect of culturing on the strain and to mitigate the stability issues of culture tunable expression (Marschall *et al.,*2016). It could be inferred from these observations that the secretor strain might experience the stress for protein export as protein synthesis consumes two-thirds of total energy produced by the cell and in response the energy expensive biosynthetic pathways are downregulated to conserve the energy, while central metabolic pathways are upregulated to meet the increasing energy demand of the cell (Mahalik *et al.,* 2014). It could be compensated by growing the secretor strain in nutrient-rich medium or by directly providing precursors in the medium (Mahalik *et al.,* 2014). It could also be optimised for growth parameters such as pH, temperature, adequate aeration in order to optimise and improve secretion *via* the FT3SS (Joseph *et al.,* 2015). In this, pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* comparison, *GapA* was upregulated (1.22 fold, p<0.007) (Appendix: Table 8.4), which is an important enzyme of glycolysis (Sabnis *et al.,* 1995). This protein up-

regulation maybe suggests the involvement of glycolysis pathway, during protein export than no protein export comparison, alluding maybe to the increased demand for energy for protein export. Similarly, protein folding and disaggregation chaperones were upregulated in the independent analysis pSC 1-Δ*CKL* Δ*clpX* (secretion) than the pSC 2-Δ*CKL* Δ*clpX* (no secretion), *viz*. DnaJ (1.30 fold, p<7.4E-05), GroL (1.16 fold, p<1.4E-09) DnaK (1.39 fold, p<7.7E-14) and ClpB (1.32 fold, p<1.0E-09). Upregulation of chaperone increases with increased secretion which suggests a direct proportion between chaperones and protein secretion (Shi *et al.*, 1992; Martínez-Alonso *et al.*, 2008). It would be interesting to investigate these observations in order to further engineer the strain for improving protein secretion. It has also been reported that co-expression of DnaK, DnaJ or GroEL significantly improves protein production (Marco, 2004; Martínez-Alonso *et al.*, 2008: Jhamb and Sahoo, 2012) and, it could be tested by either cloning of these chaperones in secretion construct pSC 1 or co-expression with pNK$_{15}$TcLib (Chapter 5, Fig. 5.6) into the Δ*CKL* Δ*clpX* strain. It could also be tested by mutagenesis, as exclusion of these chaperones might decrease protein secretion (Shi *et al.*, 1992; Nishihara *et al.*, 1998). It might be worth investigating the effect of 5'UTR-FliC$_{47aa}$ signal on regulation of DnaK, DnaJ and GrpE chaperones, as Shi *et al*. (1992) had previously speculated direct relationship between DnaK, DnaJ and GrpE chaperones and FlhD and FliA regulation (Shi *et al.*, 1992). This could be established by independent cloning of these chaperones or with KO mutagenesis.

4.7.1.2 Biochemical regulation in Δ*CKL* Δ*clpX* and Δ*CKL* strains harbouring plasmid with same secretion signal (pSC 1)

In pSC 1-Δ*CKL* Δ*clpX* *vs* pSC 1-Δ*CKL* comparison, same pSC 1 (5'UTR-FliC$_{47aa}$) was expressed into both Δ*CKL* and Δ*CKL* Δ*clpX* strains. iTRAQ analysis on intracellular proteins showed upregulation of several proteins involved in glycolysis-gluconeogenesis, butanoate metabolism and downregulation of pentose-glucuronate interconversion pathway, in pSC 1-Δ*CKL* Δ*clpX* (high secretion) than in pSC 1-Δ*CKL* (moderate secretion). There was a significant upregulation of protein folding DnaJ (1.19-fold, p<0.003), GroEL (1.19-fold, P<2.5E-05), DnaK (1.23-fold, P<7.6E-06) and protein disaggregation chaperones

ClpB (1.13-fold, p <0.001) (Table 8.4-8.9). This significant change in upregulation in pSC 1-Δ*CKL* Δ*clpX* (improved secretion) than pSC 1-Δ*CKL* (moderate secretion) may be attributed to the absence of *clpX* gene in the former strain, as both strains were expressed with the same secretion plasmid pSC 1. This could also suggest that the rate of increased protein production in absence of *clpX* in pSC 1-Δ*CKL* Δ*clpX*, maybe *via flhDC* expression as reported (Tomoyasu *et al.*, 2002). It could also be noted that, protein chaperones DnaJ, DnaK, GroEL and ClpB were significantly upregulated in pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* (Table 4.3) than pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparison (Table 4.2). DnaJ, GroL, DnaK and ClpB were upregulated in pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* than pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparison. These observations might suggest that the rate of protein chaperone regulation increases with increase in protein secretion, suggesting direct proportion between protein secretion and regulation of chaperones (Shi *et al.*, 1992; Kedzierska *et al.*, 1999; Zolkiewski, 1999).

4.7.1.3 Differential biochemical regulation in flagella mutant strains Δ*CKL* Δ*clpX* (harbouring pSC 2) and Δ*CKL* (harbouring pSC 1)

In no-secretion *vs* moderate secretion -pSC 2- Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL*, where Δ*CKL clpX* and Δ*CKL* strains were expressed with plasmid without secretion signal (pSC 2) and harbouring secretion signal (pSC 1), respectively. It was observed that chemotaxis proteins were significantly upregulated which were significantly downregulated in independent pSC 1-Δ*CKL* Δ*clpX vs* 2-Δ*CKL* Δ*clpX* (Table 4.2) and pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparisons (Table 4.3). Also, in pSC 2-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparison, the energy generating pathways such as TCA cycle and oxidative phosphorylation were also significantly downregulated, which implies that these pathways were significantly upregulated in pSC 1-Δ*CKL*, where cell exports Cutinase (moderate secretion than pSC 1-Δ*CKL* Δ*clpX*). This indicates that protein-secreting cells might meet the increasing demand of biological energy by arranging upregulation of energy generating pathways with downregulation of non-desired biosynthetic pathways during secretion and instead may utilise biochemical energy for exporting proteins *via* the FT3SS. It can be observed that this

comparison could be utilised to support the findings of pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* and pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* comparisons.

4.8 Proteins highlighted from the iTRAQ comparisons

It was observed that Cpn 60 chaperonin GroEL- a large subunit of GroESL (*GroL*) and chaperone Hsp70, with co-chaperone DnaJ (*DnaK*) proteins were significantly regulated in all three experimental comparisons, suggesting increased cell activity by means of protein turnover as these are the ATP dependent folding chaperones (Gupta *et al.*, 2009; Berlec and Štrukelj, 2013; Schlegel *et al.*, 2013). The iTRAQ analysis revealed fundamental regulation of intracellular proteins during different comparisons of iTRAQ analysis (Sections 4.7). ClpB was significantly upregulated in pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* (1.32 fold, p<10E-09) and pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparison, while it was downregulated in pSC 2-Δ*CKL* Δ*clpX*- pSC 1-Δ*CKL,* maybe suggesting it has a potential role in responses to and may boost secretion (Appendix Table 8.5-8.9) and hence it was hypothesised that, exclusion of *ClpB* might improve the protein secretion. While, YcgR-flagellar brake protein was downregulated in pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* Δ*clpX* and upregulated during pSC 2-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* comparison (Appendix Table 8.8: though below the threshold limits, it was investigated as YcgR is directly related to flagella and motility) which might suggest that it was associated with decrease in protein secretion and therefore it was hypothesised that its removal might improve the protein secretion. Note that, due to time constraint only upregulated ClpB and downregulated YcgR were selected for further analysis (Chapter 6). However, in order to achieve the aim of this project of engineering the host strain for high-efficiency protein secretion, several of these proteins (Section 4.10) could be investigated to create super-secretor strain maybe by creating KO mutant library of these proteins (Chapter 7).

In addition to the summary mentioned here, general discussion themes emerged that integrate the significance of this work into the wider picture for improving protein export *via* the FT3SS and is reported in Chapter 7 (Discussion).

CHAPTER 5

A PRELIMINARY GENOMIC SCREENING STUDY TO IDENTIFY LOCI IMPACTING PROTEIN SECRETION *VIA* THE FLAGELLAR TYPE III SECRETION SYSTEM (FT3SS)

5 Introduction

The work in this thesis so far outlined assessment of optimal protein signals for secretion (Chapter 3) through the FT3SS and the proteomic changes observed both during FT3SS secretion as a result of *clpX* deletion (a mutation that improves FT3SS secretion: Chapter 4). In this chapter, a parallel set of experiments were carried out to take an unbiased approach to examining if the expression of extra copies of *E. coli* genes or loci might boost or reduce secretion, using a Cutinase based screen (developed and improved in Chapter 3). In order to achieve this, a version of a genetic screening approach was utilised, where a genomic library of *E. coli* was used to highlight the genes that might boost secretion *via* the FT3SS. Some investigators also refer to this as inverse metabolic engineering where desired phenotype is identified and used to screen the environmental or genetic conditions that would confer or improve the desired phenotype (Ghosh *et al.,* 2012). Once identified this information is then used to alter the phenotype of selected hosts by genetic engineering or other means (Delgado and Liao, 1997; Bailey *et al.*, 2002). This inverse metabolic engineering approach has been successfully utilised to investigate different biological processes such as glycosylation (Pandhal *et al.*, 2013), or over-expression of recombinant proteins (Ghosh *et al.,* 2012).

The principle of genomic library construction applied here, involved the collection of random gene fragments of *E. coli* in a suitable plasmid system. The library was then screened to highlight the effect of unknown gene(s) (and their encoded proteins) on recombinant protein secretion *via* the FT3SS. The clones affecting secretion would then be assessed for further investigation, using the previously optimised MUB assay, which was further refined in this chapter for higher throughput screening. Meanwhile, in order to create the genomic library, a suitable library plasmid was designed and constructed to accumulate the random genome fragments (genes) independently on the plasmid. The independent library plasmids, each harbouring a unique fragment of genomic DNA (genes), were then co-expressed in a secretion strain, along with the secretion construct (pSC 1: 5'UTR-FliC$_{47aa}$-Cutinase -pJExpress No 3'UTR (Fig.

3.4/6)) and screened using the MUB assay. The protein secretion affecting library plasmids were then further investigated.

Aim: The main aim of this chapter was to investigate the potential for developing a plasmid-based *E. coli* genomic library to screen for genes that might enhance the FT3SS function.

Results

5.1 Design and construction of the genomic library

To achieve the aim, it was decided to adopt an approach whereby the genomic library and secretion construct were co-expressed in a high-level secretion strain. In this approach, the genomic library needed to be inserted into a plasmid to enable co-expression with our secretion construct- pSC 1 (containing 5'UTR-FliC$_{47aa}$-Cutinase -pJExpress 404 -used in previous chapter).

As outlined in Chapter 3 (Section 3.2), pSC 1 plasmid has a pUC origin of replication, T_5 promoter and ampicillin resistance gene, multiple cloning sites (MCS) for cloning secretion signal and loading cargo protein. Finally, it has a Flag-tag and Strep-tag for easy detection and purification (Fig. 3.4/6). Therefore, in order to simply screen the effects of library gene expression on secretion *via* the pSC 1, a plasmid with alternative antibiotic selection, promoter and more importantly, a compatible origin of replication maybe with lower copy number was required. However, after searching the iGEM (http://igem.org), Addgene (https://www.addgene.org/) and the Biobrick (https://biobricks.org/) plasmid repositories of biological parts no convenient plasmids were present. Therefore, a bespoke plasmid for this purpose was designed *in silico* and produced *via* a combination of gene synthesis and ligation to produce the plasmid pNK$_{15}$TcLib (Fig. 5.6), which contains a p15A origin of replication, tetracycline resistance gene (Tet^R) and a tailored multiple cloning site (MCS) for library construction (see below).

In the first stage, pACYC184 plasmid, (Gene Bank Accession No #XO6403 (*donated by Prof. Mark Thomas, Dept. of Infection and Immunity, The University of*

Sheffield (Fig. 5.1))), was analysed and chosen as a backbone for new plasmid as it has a compatible p15A origin of replication with low copy number (~15) and a convenient ~2.2 kb SacII-AvaI fragment released upon digestion that contains both the origin of replication and Tet^R resistance gene. Crucially, the p15A origin allows it to co-exist in cells with plasmids of ColE1 compatibility group such as pUC19 (and therefore pJExpress 404-C47) (Walia *et al.*, 2007). The next part of the design phase required the consideration of the region of the plasmid to be used for library cloning and incorporation of transcriptional terminators at each end of a cloning site to prevent any effect from plasmid read-through on the inserts. For this purpose, a 129 bp long transcription terminator (BBa_K282001-iGEM repository (https://biobricks.org/) was used, and is shown below (Fig. 5.2),

BBa_282001_129_bp:ACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTG
TTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCG
TTTATATGGAAAGCGGGCAGTG.

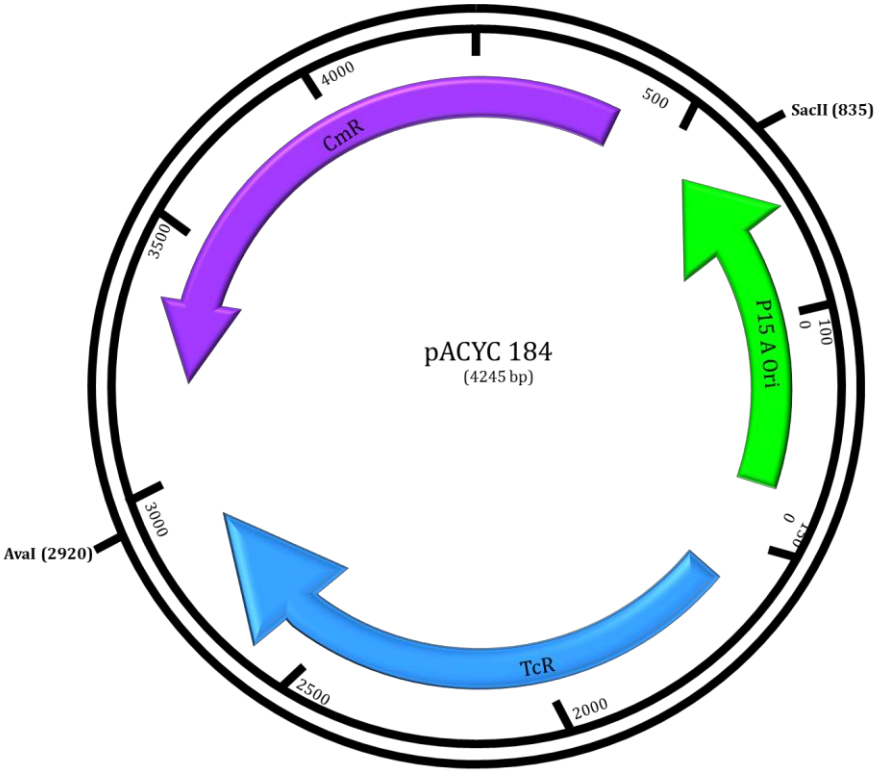


Figure 5.1: Graphical representation of the pACYC 184 plasmids.
Showing p15A Ori-Origin of replication, $Cm^{R:}$ Chloramphenicol resistance gene, Tet^R: Tetracycline resistance marker, R.E. Restriction Enzymes.

Secondly, one of the important characteristic features of this newly constructed library plasmid was the multiple cloning site (MCS) specifically designed for library construction. Several factors were considered while designing the MCS, such as the introduction of sites with compatibility with 'sticky-ended' 4-base cutter enzymes that were intended to be used in digesting the genomic DNA. As a result, the 6–cutter restriction enzymes used in the designing the MCS of the library plasmid were: MluI (ACGCGT), BglII (AGATCT), NcoI (CCATGG), SpeI (ACTAGT), BclI (TGATCA) and BssHII (GCGCGC). These 6-base cutter restriction enzymes upon digestion produce the same sticky overhangs as the corresponding 4-base cutter enzymes, AciI (GGCG), DpnII (GATC), CviaII (CATG), XspI (CTAG), DpnII (GATC) and SsiI (CCGC), respectively.

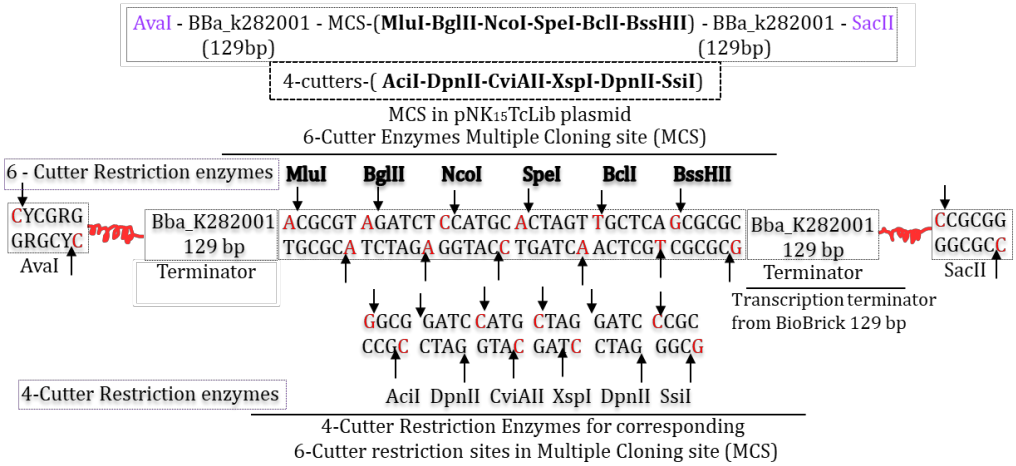


Figure 5.2: Schematic representation of MCS designed for Library plasmid. Multiple cloning sites (MCS) for 6-cutter restriction enzymes in the pNK$_{15}$TcLib plasmid were manually designed, corresponding sites for 4-cutters are shown, specific for genomic DNA digestion, transcriptional terminator sequence (BBa_K282001 129 bp) was accessed from Briobrick.org.

After carefully designing the MCS, it was assembled *in silico* and ordered from GeneArt (Thermo scientific) incorporating the transcription terminator sequences (BBa_K282001 129 bp) on either side of 36 bp MCS sequence. Also, on one side of the terminator sequence, AvaI (CYCGRG) and on the other side SacII (CCGCGG) restriction enzymes were introduced for cloning into the pACYC 184 backbone. To make the process more efficient for cloning, it was synthesised in GeneArtplasmid pMK-RQ (Km^{R}) (See Fig. 5.3), in which it was cloned by GeneArt as a 306 bp SfaI-SfaI fragment.

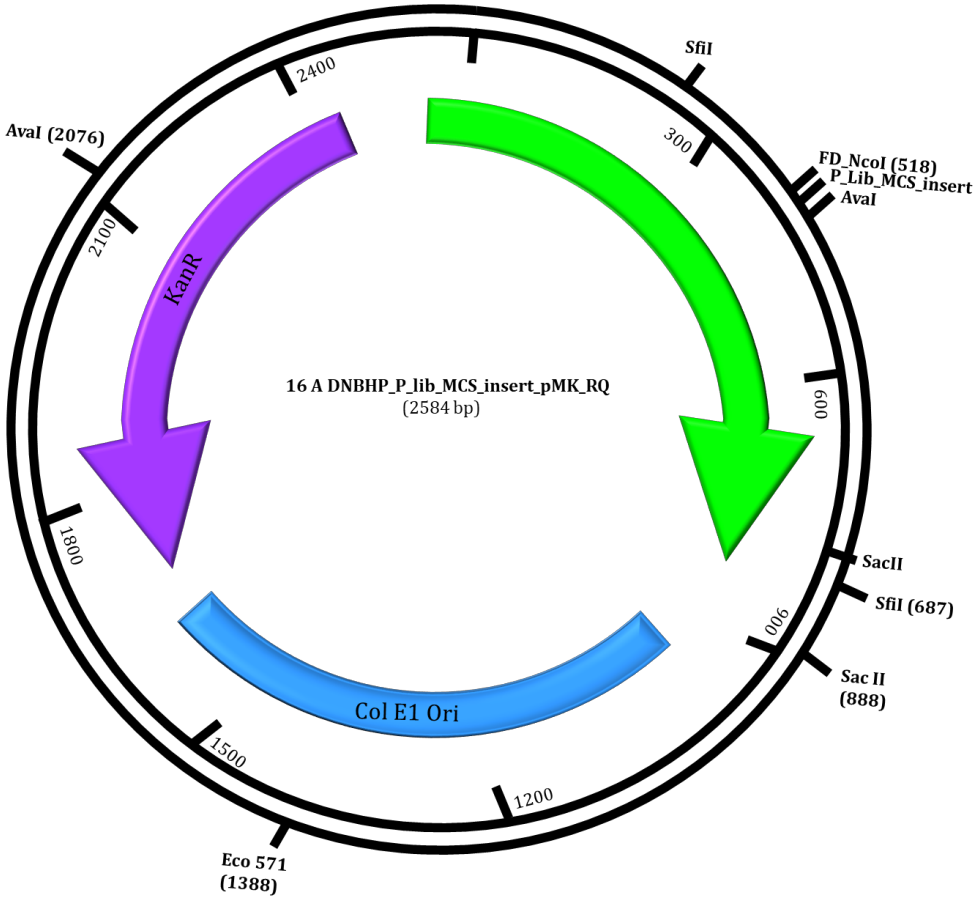


Figure 5.3: The pMK–RQ (Kan^R) plasmid containing the gene insert.
Showing p_lib_MCS_insert, ColE1 ori, Kan^R, gene string inserted between *SfiI–SfiI* sites, and convenient R.E. sites. (kan^R: kanamycin resistance gene, R.E.: restriction enzyme sites, ColE1: ori-origin of replication.

For cloning, the pMK-RQ (kan^R) plasmid (containing the desired 306 bp synthetic gene string) and pACYC184 were digested at 37 ^{O}C with AvaI and SacII and run on a 1% Agarose gel (with EtBr) as shown (Fig. 5.4). On the left, the pACYC184 digestion shows a band (s) at around 2.2 kb, note here that, the Tc^R / p15A fragment and the Cm^R fragments are nearly identical in size so this band will contain both, hence the selection relied on the Tc^R selection for producing correct clones (see later). Meanwhile, the 306 bp fragment containing the MCS is clearly visible from pMK-RQ (Km^R). Both the bands were excised and cleaned up using PCR and Gel extraction kit (Bioline) and 30 µl of each were eluted in buffer and ligated (Section 2.5).

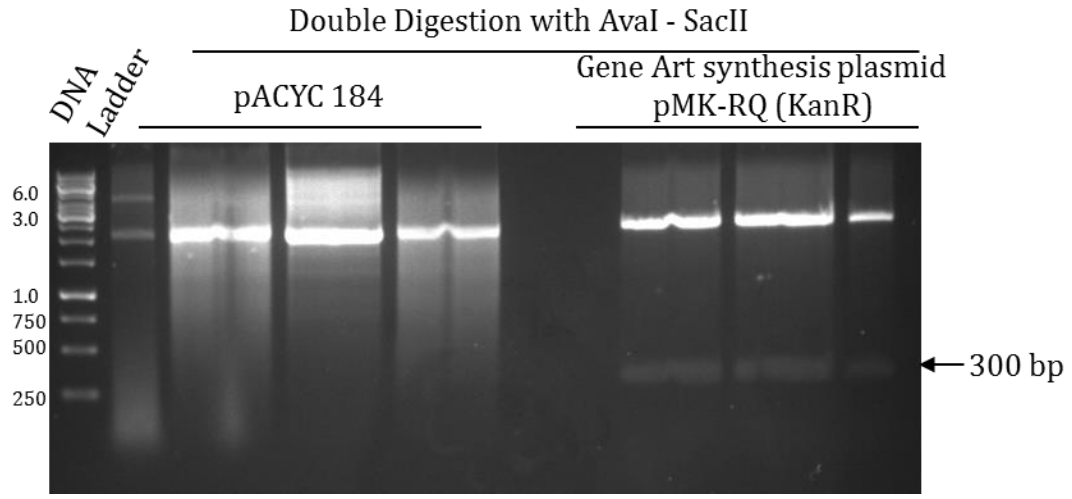


Figure 5.4: Digestion of pACYC184 and pMK-RQ for pNK$_{15}$TcLib synthesis.
pACYC 184 and synthetic gene string (received from Gene Art Syntheis™ in a plasmid containing c*olE1* Ori and kanamycin cassette) were digested with AvaI and SacII. pACYC 184 upon digestion yields identical bands of ~2.2 kb as shown in lane 1. Lower bands were selected and excised. Synthetic string of 306 bp was easily selected.

Successful cloning of the 306 bp synthetic fragment into a ~2.2 kb fragment of pACYC184 plasmid was confirmed with colony PCR of tetracycline resistance clones (Note: if the 'other' fragments of pACYC 184 ligated with MCS, it would lack both an origin and Tet^R genes) using following primer pair and restriction digestion, as shown in Fig. 5.5.

Forward primer: binds in Tet^R cassette: CTTACCAGCCTAACTTCGATCA

Reverse primer: bind in terminator sequence: ACTGCCCGCTTTCCATATAA

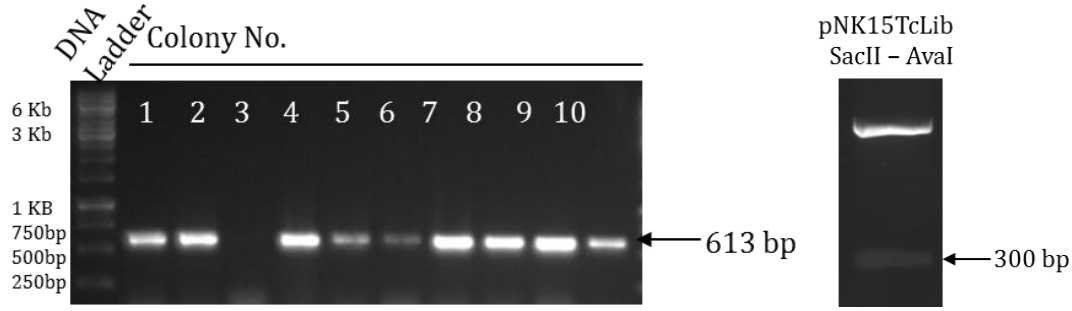


Figure 5.5: Confirmation of pNK$_{15}$TcLib plasmid synthesis with colony PCR (left) and restriction digestion (right).
Colony PCR was performed using primers (pNK_F and pNK_R, Table 8.3-see Fig. 5.6)-with ~613 bp indicating successful cloning (left). The inserion of the synthetic MCS region of 300 bp was confirmed by AvaI–SacII R.E digestion (right).

The final confirmation was obtained by Sanger Sequencing (GATC Biotech, Germany). The 2388 bp novel library plasmid was named pNK$_{15}$TcLib, where 15

represents p15A origin of replication, Tc: tetracycline resistance cassette and lib represents the library plasmid as shown in Fig. 5.6.

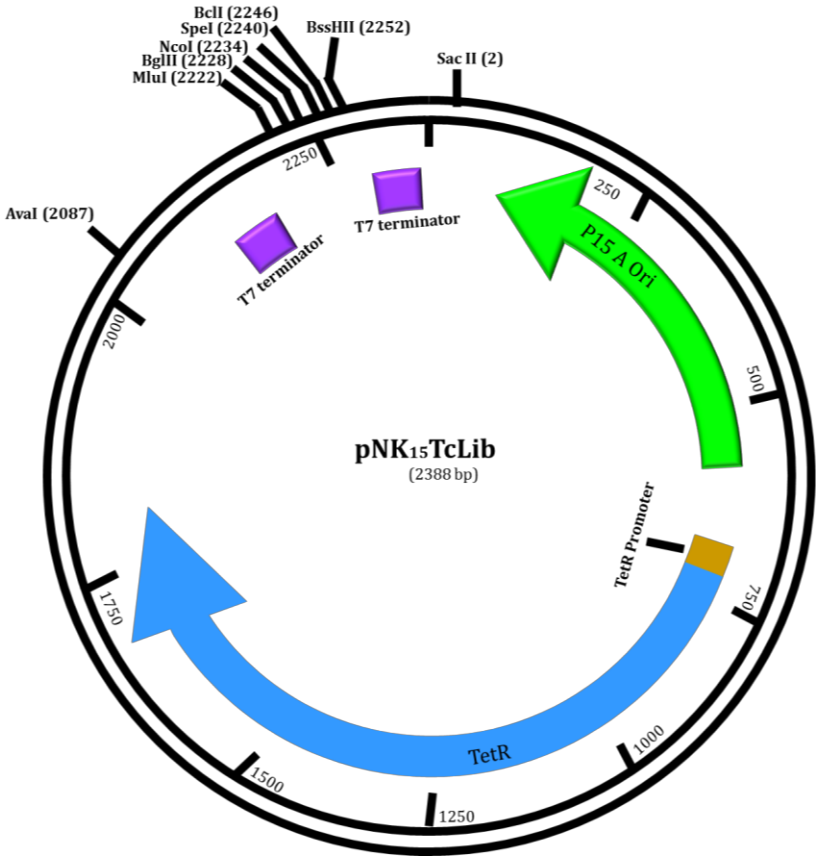


Figure 5.6: Graphical representation of the pNK$_{15}$TcLib (2388 bp) plasmid. Circular graphic map of plasmid pNK$_{15}$TcLib, it contains p15A origin of replication, Tet^R cassette and MCS containing six, 6-cutter R.E. sites and BBa_282001 transcription terminator on either side of MCS (BioBrick (https://biobricks.org/)) (MCS: multiple cloning sites, R.E.: Restriction sites, Tet^R: tetracycline resistant cassette, ter: Terminator).

5.2 Optimisation of genomic DNA digestion and cloning into library plasmid pNK$_{15}$TcLib

In order to achieve the deep coverage the genetically unmodified, wild type *E. coli* MG1655 was chosen over *E. coli* MC1000 which harbours several genomic modifications. The wild-type *E. coli* MG1655 genomic DNA was extracted with Genomic DNA Isolation kit (Promega Wizard®, Promega Inc.™) as per manufacturer's instructions and concentrated up to 1.200 µg µl^{-1} using 3M sodium acetate and 70% ethyl alcohol (Section 2.4.1.6). This genomic DNA was then digested with 4-cutter restriction enzymes which produce overhangs that can be cloned into the MCS of pNK$_{15}$TcLib (Fig. 5.2). The aim here was to produce a digest that contained a high amount of fragments in the range of 2-4 kb, with

the possibility of harbouring complete gene/operon (s) sequence with their native promoter. After several trials, it was found that AciI enzyme produced the most effective and even digests and hence this was used for genomic DNA digestion (optimisation process not shown).

To further improve the digests and produce the most fragments in the target region, AciI 4-cutter restriction enzyme (1000 Units μl^{-1}) was serially diluted and digestion reactions set up as follows: 30 µl genomic DNA (Conc. 1200.00 ng μl^{-1}) + 1 µl of 4-Cutter R.E. (NEB) + 5 µl Cutsmart buffer (NEB) + 14 µl dH_2O water = 50 µl. The reactions were incubated at 37 ^{O}C for 10 min as it was optimised from previous experiments and heat inactivated at 65 ^{O}C for 15 min (or 80 ^{O}C for 5 min) and run on 0.75 % agarose gels (Fig. 5.7).

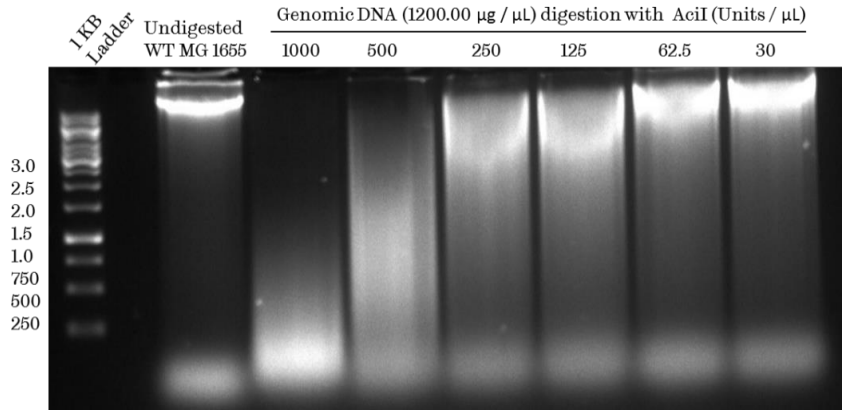


Figure 5.7: Optimisation of Genomic DNA digestion with 4-cutters.
Genomic DNA of *E. coli* MG1655 wild-type strain was digested with AciI using different concentrations of enzyme for 30 min at 37 ^{O}C and enzymes were heat inactivated at 65 ^{O}C for 15 min, digestion efficiency was checked by running undigested genomic DNA alongside. 1 kb ladder was run as shown in the left side of the gel. Picture was taken by exposing the gel to UV light at 302 nm in G-box (Syngene technologies)

After these experiments, samples digested using 125 and 250 U were selected as they contained a shift from uncut DNA at the top of the gel and a good range of fragments in 2-4 kb range. Bands were excised from agarose gel and cleaned up using a Gel extraction kit (Bioline). In parallel, $pNK_{15}TcLib$ was digested with the 6-cutter MluI (NEB) which produces sticky overhangs compatible with AciI digestion. It was similarly cleaned up using PCR and gel extraction kit (Bioline). Re-circularisation of these digested $pNK_{15}TcLib$ was circumvented by dephosphorylating open phosphate ends with FastTrap

alkaline phosphatase (Thermo scientific). Ligations were then incubated overnight at ~16 ^{O}C. 3.5 µl of ligation reaction was transformed into competent DH5∝ cells (NEB) (Section 2.9). A selection of positive colonies were used for colony PCR with more than 90% colonies positive for inserts and indicating presence of fragments of genomic DNA (1.5-4.0 kb inserted in pNK$_{15}$TcLib (data not shown)). However, it was surprising that fragments were not larger overall, but due to time constraints screens were conducted using this library (Fig. 5.8).

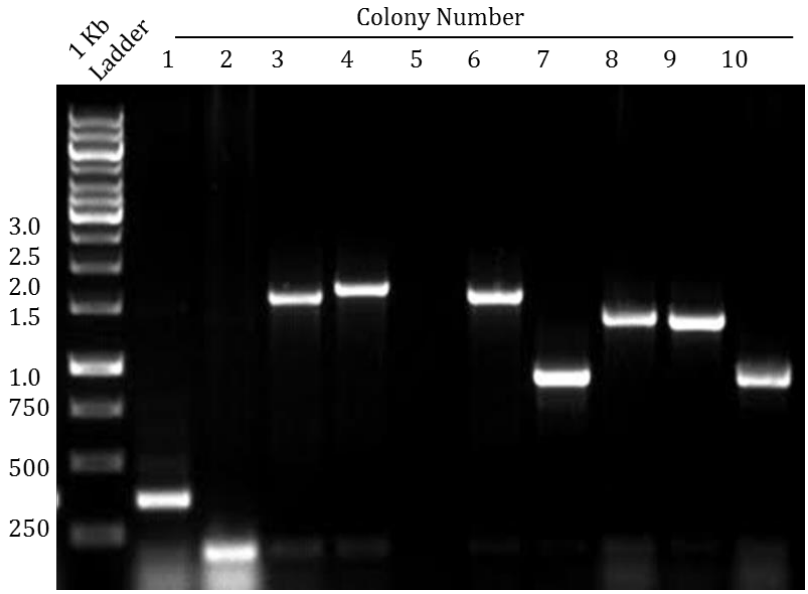


Figure 5.8: Colony PCR for the confirmation of inserts in the pNK15TcLib.
A representative colony PCR for the identification of insert into pNK$_{15}$TcLib plasmid, using primers that bind prior to MluI and after MluI site in MCS (primer set Appendix 3). Picture taken by exposing the gel to UV light at 302 nm in G-box (Syngene technologies).

5.3 Co-expression of the plasmids in secretor strain *ΔflgKL ΔclpX*

In order to investigate the effect of library genes on protein secretion *via* the FT3SS, the efficient secretion signal containing pJExpress 404 (pSC 1) and newly constructed pNK$_{15}$TcLib (harbouring random genomic DNA fragments) were co-transformed in the efficient secretor strain (Δ*flgKL* Δ*clpX*) (Fig. 5.9). The co-transformation as shown in Fig. 5.9 was carried out using chemical competent as well as electrocompetent (Δ*flgKL* Δ*clpX*) cells. However, the transformation of ligation directly into these competent, pSC 1 harbouring *ΔflgKL ΔclpX* cells did not work well. Therefore, the Hanahan method was utilised for making (Δ*flgKL* Δ*clpX*) chemical competent cells for co-transformation of pSC 1 and pNK plasmids (Hanahan, 1983; Chan *et al.*, 2013). Nevertheless, it was not as efficient

and practical as only 5-10 colonies were obtained on a dual antibiotic selection plate. Therefore, the ligation (3.5 µl) mixtures were directly transformed into DH5∝ cells (NEB) and ~75-92 colonies per plate were obtained. ~10% of these colonies from each plate were confirmed for the presence of insert in pNK clone, following their selection on antibiotic selection plates, once again revealing insert sizes in the range of ~2.0 kb. The remaining colonies on several plates were scraped off the plate and grown in liquid broth and their plasmids were isolated and co-transformed into Δ*flgKL* Δ*clpX* cells (harbouring pSC 1) and finally selected on dual-antibiotic selection plates containing ampicillin and tetracycline for pSC 1 and pNK, respectively. As previously mentioned, pSC 1 harbours the Cutinase enzyme (Section 3.2) and for which a well-established enzyme assay was optimised (Section 3.2.2). Therefore, for library screening, pSC 1 encoded Cutinase was used as a reporter protein to highlight the influence of library clones on protein secretion of pSC 1 *via* the FT3SS.

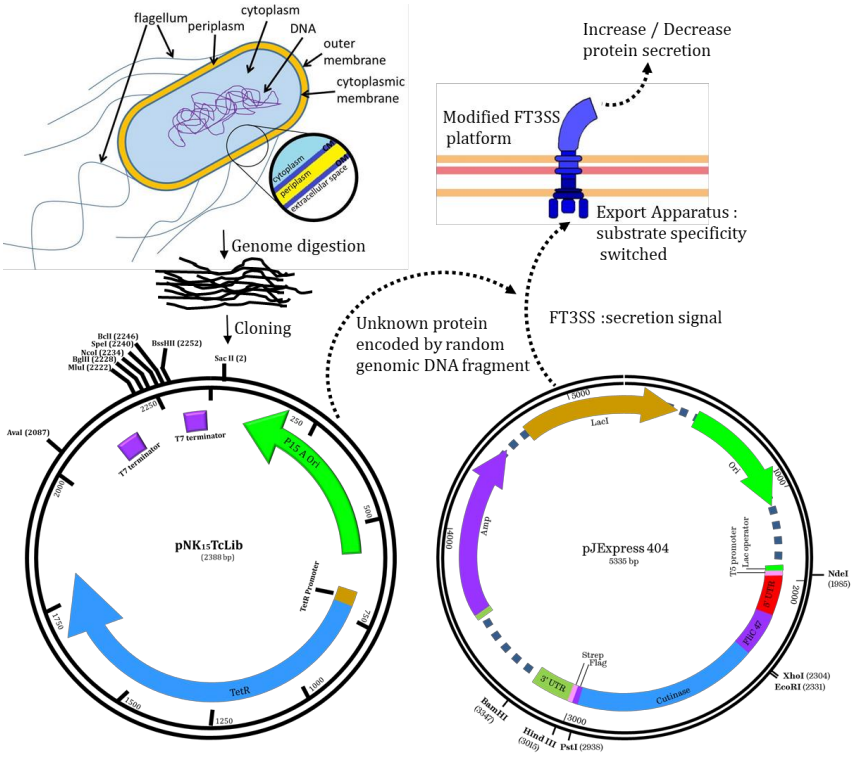


Figure 5.9: Co-expression of plasmids in the flagella mutant strain.
Graphical representations of the co-expression strategy for the expression of two plasmids in Δ*flgKL* Δ*clpX*, (1) secretion construct pSC 1 (5'C_{47aa}-Cut-pJEx No 3'), and (2) pNK_{15}TcLib (2388 bp) plasmid harbouring the random fragments of genomic DNA. The genomic DNA was randomly digested with AciI R.E. and cloned into the pNK plasmid, prior to co-expression with secretion construct pSC 1.

5.4 Investigation of the influence of pNK clones on the secretion strain Δ*flgKL* Δ*clpX* harbouring pSC 1 secretion construct

In order to establish the effect of co-transformation and co-expression of two plasmids and their cargo genes (and expressed proteins) on the growth parameters of *E. coli* by influencing metabolism, the Δ*flgKL* Δ*clpX* strain containing both pSC 1 and pNK (empty) plasmids were monitored and verified with growth curve analysis. The co-expression of pSC 1 and pNK (without insert) into Δ*flgKL* Δ*clpX* strain leads to protein secretion *via* the modified FT3SS, however, the proteins encoded by pNK (Tc^R resistance and replication proteins) didn't affect the protein secretion of pSC 1 *via* the FT3SS (Fig. 5.10). Therefore, the independent bacterial strain Δ*flgKL* Δ*clpX,* without any plasmids, with only pSC 1 and pNK and the one co-expressed with both pSC 1 and pNK (empty), were grown overnight in 5 ml of LB broth with appropriate antibiotics, before dilution 50-fold and inoculation of 200 µl into the sterile, transparent 96-well plate (Greiner). Growth was then followed in a plate reader (A 600 nm, Tecan, Magellan, Infinite® 200 PRO) at 37 ^{O}C with shaking every 30 min. It can be observed (Fig. 5.10) that there was no significant difference between the growth of cells harbouring either independent plasmids or co-transformed plasmids or cells without any plasmid. It may be inferred therefore that-harbouring two plasmids did not impose any extra metabolic burden on these strains, and that the genomic screening strategy was ready for initiation.

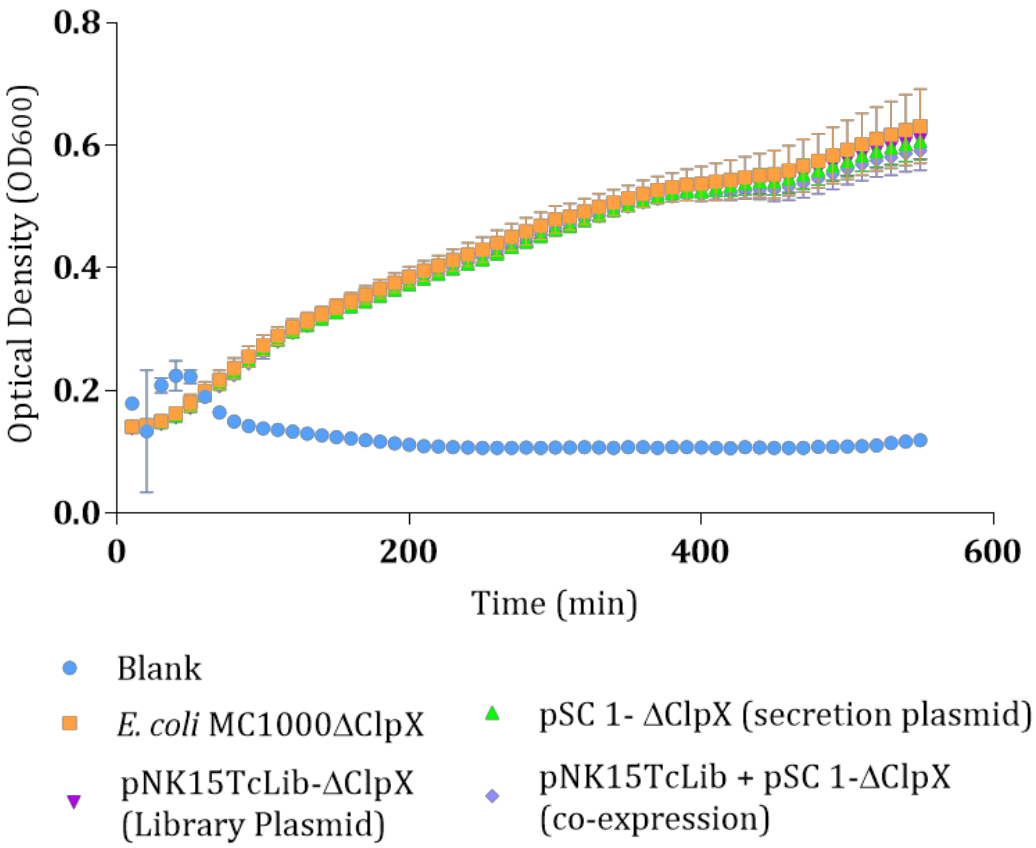


Figure 5.10: Growth curve to monitor metabolic burden of co-expression.

Growth curve to monitor the growth defect in secretor strain for carrying pNK and pSC 1 plasmids. Δ*CKL* Δ*clpX* strain without plasmid seves as a control. Independent and co-expressed plasmids in Δ*CKL* Δ*clpX* strain were grown, results from 3 biological repeats, error bars (±) SEM, and read using Plate reader (Tecan Magellan, Infinite® 200 PRO).

5.5 Screening of genomic library with high-throughput MUB assay

Positive colonies obtained on dual antibiotic selection plates (Amp^R and Tet^R) were grown in 96-well plates with appropriate antibiotics. The following day, these cultures were transferred into a fresh LB broth in a sterile 96-well plate (Greiner). Cultures were induced with 0.05 mM IPTG for 240 min, i.e. in mid exponential phase under these conditions (see Fig. 5.10 above). These cultures were then centrifuged at 3500 rpm (in a 96-well plates) at 4 ^{O}C for 15 min and the 4-methylumbelliferyl butyrate (MUB) assay was performed using 25 µl supernatant, 15 µl water and 160 µl 100 µM MUB. These were incubated for 30 min at 37 ^{O}C and fluorescence was measured at 302-446 nm. In these screens, 6 -wells were used for control (PCB, MUB, LB, LB+MUB, MUB+pNK (empty) and MUB +pSC 1 (positive control)).

Initially, the pNK libraries were grown under stationary conditions, to reduce chances of culture spill over and mixing in the 96-well plate format. Plates were incubated for 240 min at 37 ^{O}C and the improved assay was performed on supernatant (Section 5.6). It was observed (Fig. 5.11) that pNK$_{15}$TcLib plasmid 46 (pNK 46$^{Stat.}$) showed a significant (1.16 fold, p<0.05) increase in protein secretion over the control (pSC 1-Δ*CKL* Δ*clpX*), however, pNK$_{15}$TcLib 37$^{Stat.}$ 57$^{Stat.}$ 66$^{Stat.}$ and 89$^{Stat.}$ show a significant decrease in protein secretion *via* the modified FT3SS (p<0.05) (Note: The ($^{Stat.}$) represents the cells grown at stationary).

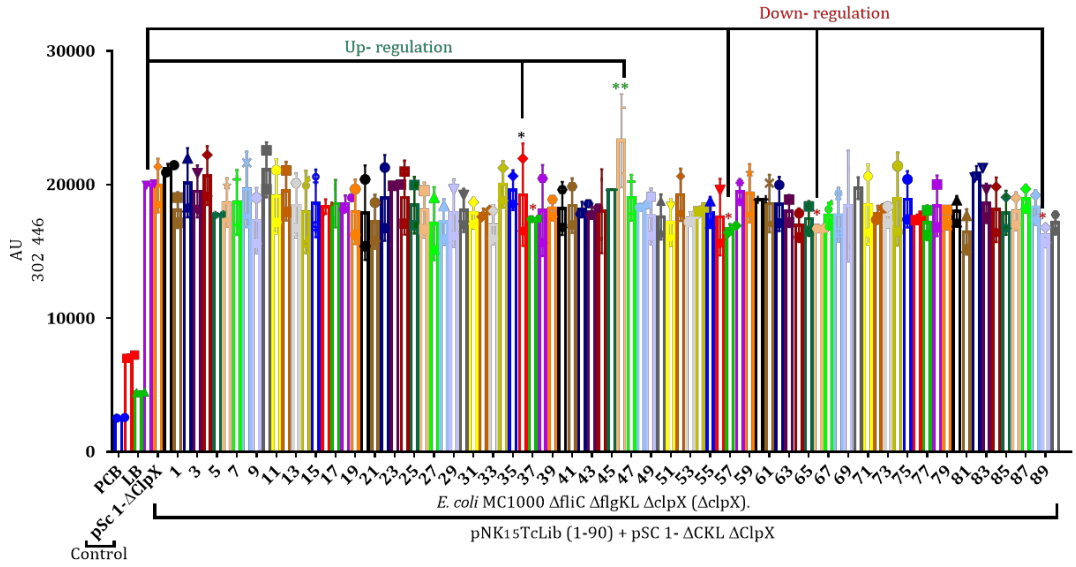


Figure 5.11: Screening of pNK clones from library grown at stationary
Screening of pNK$_{15}$TcLib library grown at stationary using the controls: PCB, MUB, LB, LB+MUB (Lane 1-4), pSC 1 and pSC 2 expressed into *ΔflgKL ΔclpX* serves as positive and negative controls (Lane 5-6), respectively. pNK$_{15}$TcLib clones screened with MUB assay (Lane 7 onwards) in a 96-well plate (Greiner). Fluorescence detected at 302, 446, results from 3 biological repeats, error bars (±) SEM, statistical significance defined by ****p<0.0001, determined by One way ANOVA for multiple comparisons, Graph plotted with Graphpad Prism 7 (Key: green sign * : upregulation, red sign * : downregulation).

Therefore, these potential five pNK$_{15}$TcLib clones were subjected to further investigation using the previously optimised, larger scale MUB assay (Section 3.2.2) by growing them in regular 10 ml cultures (Fig. 5.12) and by western immunoassay (Fig. 5.13). It could be observed (Fig. 5.12) that the data has a significant correlation with that observed in Fig. 5.11. pNK 46$^{Stat.}$ shows improved protein secretion as expected, however, it was not significantly different than control pSC 1-Δ*CKL* Δ*clpX* in 10 ml cultures, while pNK 37^{Stat}, 57^{Stat}, 66^{Stat} and 89^{Stat} showed significant decrease in protein secretion (p<0.0001),

suggesting that proteins encoded by these plasmids decrease the protein secretion of pSC 1 *via* the modified FT3SS. Finally, these were then confirmed with western immunoassay (Fig. 5.13).

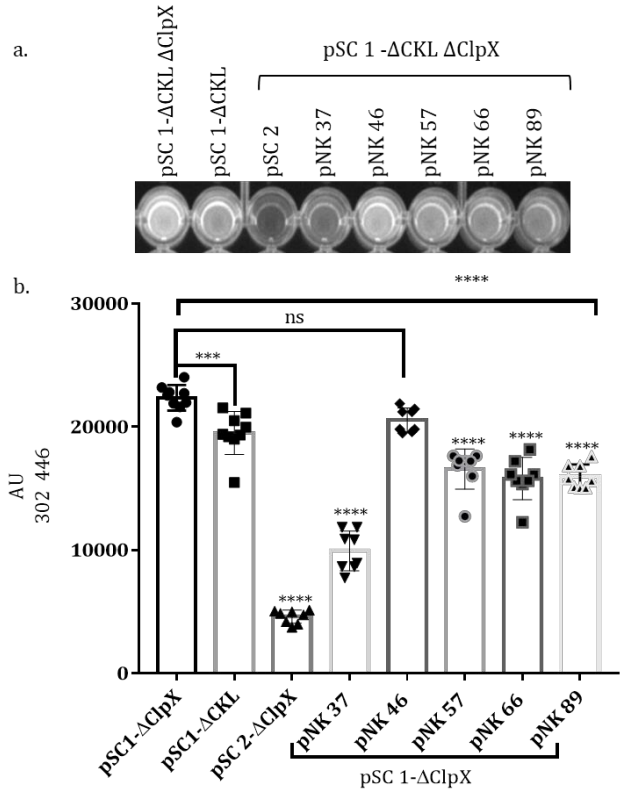


Figure 5.12: MUB assay on pNK clones highlighted from preliminary screening for protein secretion.
(a.) MUB assay reactions visualised by exposing the plate to U.V. Light at 302 nm in G-box (SynGene Technologies). (b.) pSC 1 and pSC 2 expressed into *ΔflgKL ΔclpX* were used as positive and negative controls, respectively. pNK$_{15}$TcLib 37$^{Stat.}$ 46$^{Stat.}$ 57$^{Stat.}$ 66$^{Stat.}$ and 89$^{Stat.}$ were tested, Fluorescence detected at 302, 446. Graph was plotted with Graphpad Prism 7.0, error bars (±) SEM, statistical significance was defined by **** p<0.0001, determined by One way ANOVA for multiple comparison, 3 biological repeats observed.

It can be observed (Fig. 5.13 a-c) that there was no significant difference in the intracellular expression of pSC 1 and pNK plasmids 37$^{Stat.}$, 46$^{Stat.}$, 57$^{Stat.}$, 66$^{Stat.}$ or 89$^{Stat.}$ as observed with statistical analysis with one-way ANOVA for multiple comparisons. However, there was a significant difference between their secretion. Previously pNK 46$^{Stat.}$ showed improved secretion (Fig. 5.11 and 5.12), however it showed a significant decrease in protein secretion in western immunoassay (Fig. 5.13b-d). However, in agreement with Fig. 5.11 and 5.12 the remaining pNK clones showed significant decrease in protein secretion (in the

range of 1.46 to 51.7-fold). These pNK plasmid encoded proteins (that decrease protein secretion of pSC 1 *via* the FT3SS) were then identified with sequencing and bioinformatics analysis (Section 5.6), (Note here, the pSC 1 plus pNK-empty had no effect on secretion (Fig. 5.11)).

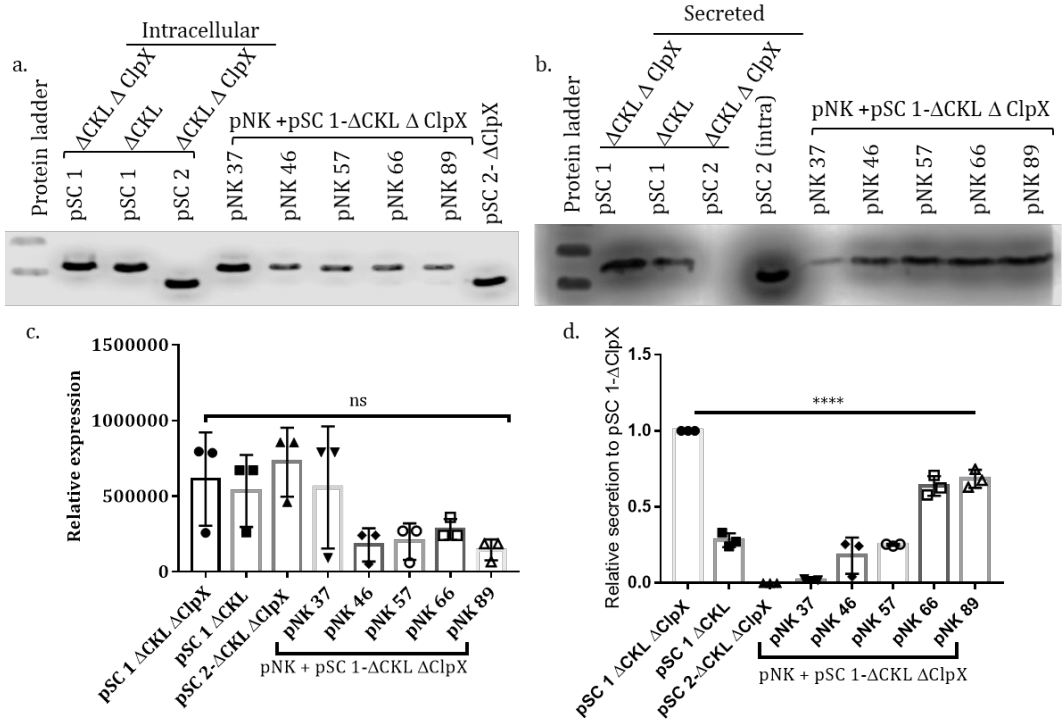


Figure 5.13: SDS-PAGE-western blot analysis on the protein highlighted from pNK$_{15}$TcLib screening at stationary.

pNK$_{15}$TcLib 37^{Stat,} 46^{Stat,} 57^{Stat,} 66^{Stat,} 89$^{Stat.}$ were coexpressed with pSC 1 in Δ*flgKL* Δ*clpX* strain. pSC 1-expressed in *ΔflgKL* and *ΔflgKL ΔclpX* and pSC 2 expressed into Δ*flgKL ΔclpX* were used as positive and negative control, respectively. Blots were developed using *LI-COR* C–Digit, graph plotted with Graphpad Prism 7, Error bars (±) SEM, 3 biological repeats, statistical significance was defined by **** p<0.0001, determined by One way ANOVA for multiple comparisons (a. intracellular, c. intracellular densitometry b. secreted d. secreted densitometry).

Note here that the number of pNK clones that significantly increased or decreased protein secretion were far lower than anticipated and perhaps reason could be attributed to the volume of the expression cultures or the aeration (shaking) required during growth (McDaniel *et al.,* 1965). Therefore, in the process of library screening-optimisation the remaining library was screened by growing cells with low shaking (section 5.5.2). As a preliminary screen and validation step to assess the approach of genome screening using this two-plasmid system it was decided to also try the assay with shaking culture. This

decision was taken based on increased aeration with shaking and therefore improved growth.

In further optimisation of the library screening, pNK libraries were grown in 96-well plate with low shaking (at 40 rpm) to avoid the possibility of spillage and mixing of the cultures (as previously speculated) for 240 min at 37 ^{O}C and the improved MUB assay was performed on culture supernatants as before. Data from one example plate is shown in Fig. 5.14, where a number of clones with increased or decreased secretion were observed -labelled with asterisks. The increased secretion clones were pNK (12^{RS}, 25^{RS}, 35^{RS}, 90^{RS})(Note: (^{RS}) represents the cells grown with regular shaking at 40 rpm) and those decreasing secretion pNK (13^{RS}, 36^{RS}, 46^{RS}, 60^{RS}) were selected. These were then grown overnight and inoculated in a fresh 10 ml broth for protein expression and assay was performed (section 3.2.1). These were then assayed for both the MUB activity and probed for Cutinase secretion using the western blot assay.

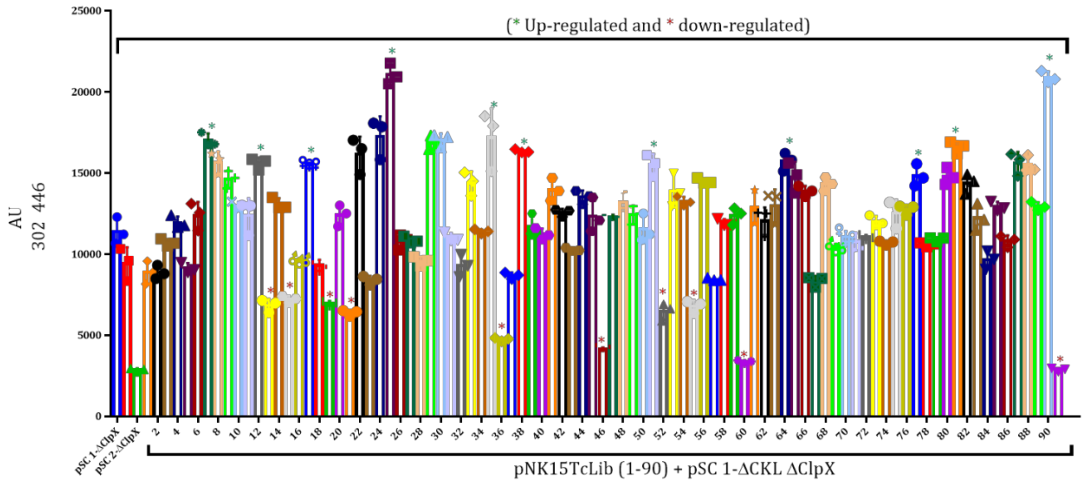
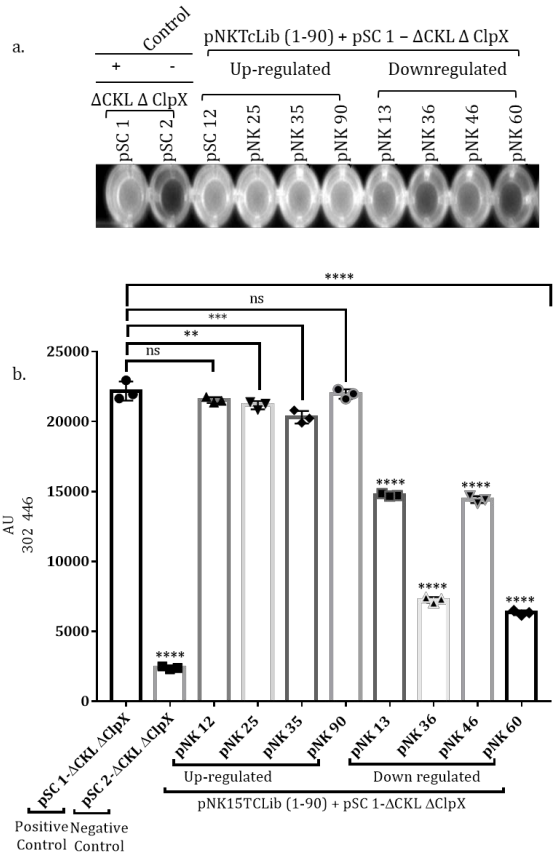


Figure 5.14: Screening of pNK15TcLib Genomic Library grown with regular shaking using the MUB assay.

Screening of pNK$_{15}$TcLib library grown with shaking was performed as described and clones with differential secretion from control (pSC 1-pNK (empty) indicated by *. Fluorescence was detected at 302, 446 (Tecan Magellan, Infinite® 200 PRO). Graph was plotted using Graphpad Prism 7.0, Error bars (±) SEM, statistical significance was defined by **** p<0.0001, determined by One way ANOVA for multiple comparisons, 3 biological repeats observed.

It can be observed from Fig. 5.15 that there was no significant difference between the secretion of original pSC 1-Δ*flgKL* Δ*clpX* and pNK 12^{RS} and 90^{RS}

clones (Fig. 5.14), while secretion of pNK 25^{RS} and 35^{RS} was decreased by less than 5% compared to control pSC 1-Δ*flgKL* Δ*clpX,* meaning these were not followed up. However, the downregulated pNK (13^{RS}, 36^{RS}, 46^{RS}, 60^{RS}) were significantly different than original pSC 1-Δ*flgKL* Δ*clpX* secretion in concordance with Fig. 5.14.



5.15: MUB assay on pNK clones highlighted to influence protein secretion.
MUB assay reactions were visualised by exposing the plate to U.V. Light at 302 nm in G-box (SynGene Technologies) (a.) and the quantitative plate assay (b). Cultures are in triplicate with error bars (±) SEM and statistical significance defined by **** p<0.0001, determined by One way ANOVA for multiple comparisons.

As before western blotting was used to validate the MUB assay findings. As seen in Fig. 5.16 A: i-iii, pNK 12^{RS} and 25^{RS} and pNK 60^{RS}, show a significant increase in intracellular protein expression of the pSC 1 protein to control (pSC 1-Δ*flgKL* Δ*clpX*). When secretion was assayed, it can also be observed (Fig. 5.16) that clones pNK 12, 25, 35, 90, 13, 36, and 60 affected pSC 1 derived Cutinase secretion significantly (p<0.05); while clone pNK 46 had no effect. Of those pNK 60 had the largest effect with almost complete loss of secretion. These

experiments validate this approach as having the potential to identify novel genes that might affect expression, however only a small number of clones were screened here due to time constraints.

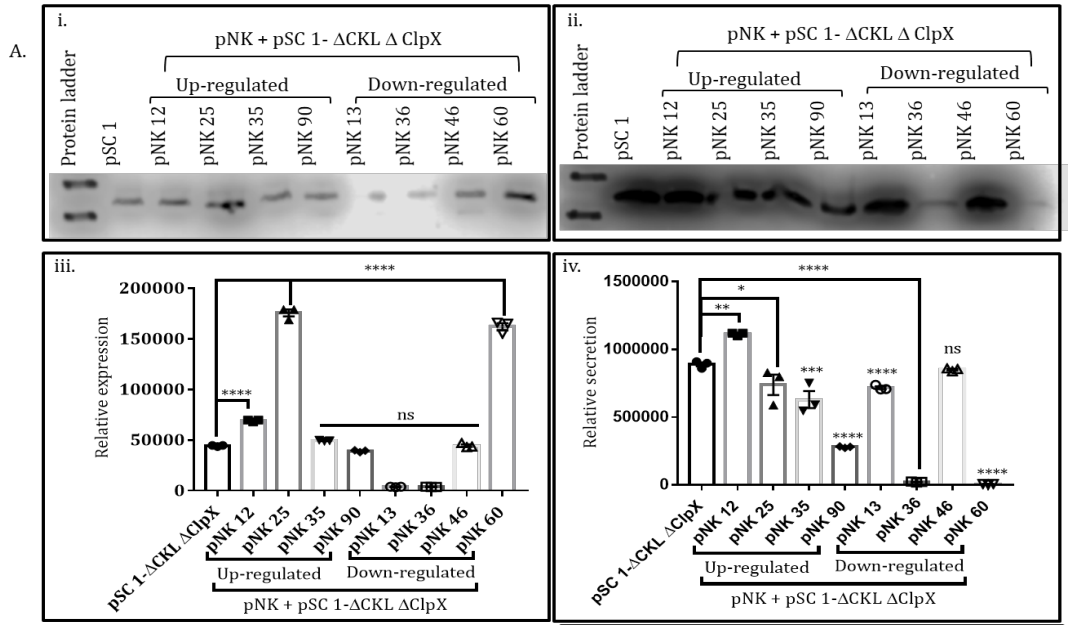


Figure 5.16: Western blot analysis on proteins identified from pNK screening with shaking.

pNK clones from genomic libraries, co-expressed with pSC 1 in Δ*flgKL* Δ*clpX* strain. pSC 1 expressed in Δ*flgKL* and Δ*flgKL* Δ*clpX* and pSC 2 into Δ*flgKL* Δ*clpX* were used as positive and negative controls, respectively. (i) and (iii) represents intracellular expression while their extracellular proteins (secreted) are shown in (ii-iv). Blots were developed using *LI-COR* C-Digit, 3 biological repeats observed, graphs plotted with Graphpad Prism 7, Error bars (±) SEM, statistical significance was defined by **** p<0.0001, determined by One way ANOVA for multiple comparisons.

5.6 Investigation of potential pNK clones highlighted from preliminary genomic library screening

As outlined above, potential pNK clones were identified following the final confirmation of library screening with MUB (Fig. 5.12 and 5.15) and western blot analysis (Fig. 5.13 and 5.16) that affected pSC 1 derived secretion. In this section, a small selection of the clones highlighted by the screens and confirmed by assays was sequenced in one direction using a primer in the 5' region of the MCS of pNK plasmid. All sequences were BLAST searched against the chromosome of MG1655 to map them on to the genome of this strain to see which genes were encoded in the clones. Pictures were either generated using SnapGene or taken from the Ecocyc database for MG1655 genome. Unfortunately, due to time

constraints and sequence quality only a selection of these were sequenced at high quality. There is also a short discussion of the possible mechanisms of potential regulation of secretion by these clones. Clones from the stationary grown library are shown with ($^{stat.}$) and those grown with regular shaking by ($^{Reg.shake\ or\ RS}$) signs, green upward (↑) arrow represents the clone that increased protein secretion of Cutinase (pSC 1) *via* the FT3SS while red downward (↓) arrow represents the clones that decreased, in both MUB and western immunoassays. A purple downward (↓) arrow represents the clones that showed increased protein secretion in MUB assay but in western immunoassay, it was either statistically non-significant or decreased compared to the control (pSC 1-Δ*CKL* Δ*clpX)*.

1. pNK$_{15}$TcLib 37$^{Stat.↓}$

Following Sanger sequencing and BLAST searches of this clone it was identified that, it contained a complete *crl* gene with native promoter. It also contains an incomplete Fermentation-respiration switch protein (*frsA*) upstream. However, only *crl* was considered significant for further analysis. The Crl is a polymerase holoenzyme assembly factor that changes the dynamics between σ^S and σ^{70} for binding to RNA polymerase core enzyme towards σ^S, resulting into increased RNA polymerase sigma S (E σ^S), which in turn allows more transcription from promoters within the σ^S regulon (Pratt and Silhavy, 1998; Kourennaia *et al.,* 2005; Typas *et al.*, 2007). It has been reported that, total cellular levels of σ factor produced in a cell exceeds that of RNA core polymerase, which results in competition between σ factor for binding to limiting amount of core polymerase (Farewell *et al.,* 1998; Ishihama, 2000). Additionally, it was reported that a phenotypically cryptic curli subunit gene *csgA* is transcriptionally activated by a multicopy plasmid harbouring *crl* gene (Arnqvist *et al.*, 1992). Therefore, as the over-expression of *crl* had significantly decreased protein secretion of pSC 1 (Cutinase) *via* the FT3SS of Δ*flgKL* Δ*clpX* strain, it appears that it might be due to the fact that the *crl* appears to utilise the RNA polymerase and changes the sigma factor dynamics towards curli formation; thus downregulating the flagella gene expression due to the lack of RNA polymerase (Farewell *et al.,* 1998; Jishage and Ishihama, 1999). However, it needs to be further investigated

in order to validate these observations and the genomic screening approach utilised in the genomic library construction. It could be investigated by either over-expressing the independent *crl* gene in the pSC 1 harbouring Δ*flgKL* Δ*clpX* strain using pNK plasmid or by removing it from the secretor Δ*flgKL* Δ*clpX* strain with knockout mutagenesis. It could also be investigated by the quantification of the RNA polymerase, *crl* and sigma factor expression levels with qPCR. Additionally, microscopic techniques might be worth for investigating Curli fibre formation.

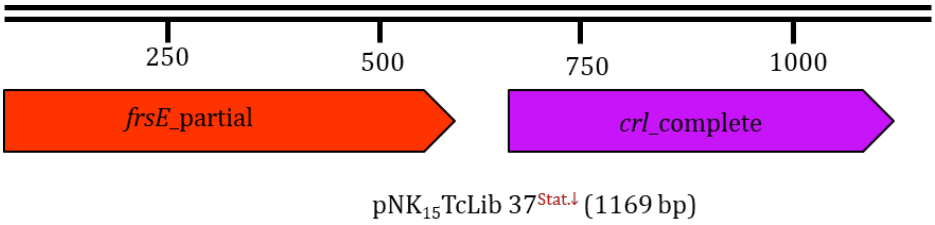


Figure 5.17: pNK$_{15}$TcLib37$^{stat↓}$ clone.

2. pNK$_{15}$TcLib 57$^{Stat.↓}$

The pNK 57$^{stat.↓}$ clone (Fig. 5.18) caused significantly downregulated protein secretion (Fig. 5.11-13). Bioinformatics analysis revealed that it contained a part of the non-essential gene *sbmC* partially. sbmC is a DNA gyrase inhibitor RecA-dependent SOS-regulon gene, in which its expression is induced at the onset of the stationary growth phase or by DNA damaging agent (Baquero *et al.*, 1995; Nakanishi *et al.*, 1998). Over-expression of SbmC has been reported to cause filamentous cell morphology and growth defects, as does antisense expression and its expression is positively regulated by H-NS and RpoS, that regulate SOS induction of the *sbmC* gene (Oh *et al.,* 2001).

However, the significant decrease in the protein secretion of pSC 1 (Cutinase) *via* the FT3SS is not well understood. It might be due to the onset of SOS response to increased SbmC protein in the cell. It could be due to the interaction of *sbmC* with the extrachromosomal DNA, maybe *via* the H-NS protein or it could be due to the action of RpoS protein. However, it could be further investigated with either qPCR or molecular cloning using pNK or knockout mutagenesis using the secretor strain, Δ*CKL* Δ*clpX* while checking the protein secretion *via* the FT3SS.

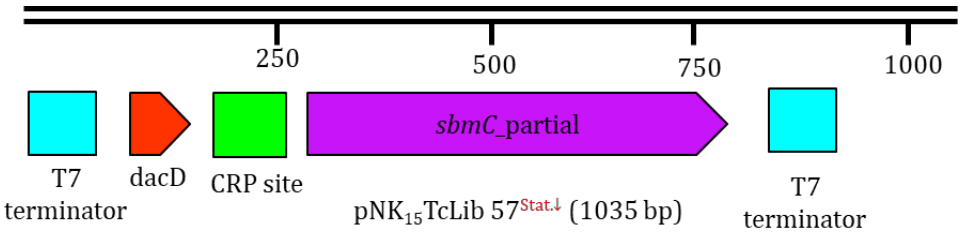


Figure 5.18: pNK$_{15}$TCLib 57$^{stat.↓}$ clone.

3. pNK$_{15}$TcLib 12$^{Reg.Shake↑}$

The pNK$_{15}$TcLib$^{Reg.shake↑}$ clone 12 (Fig. 5.19) caused significantly improved protein secretion of pSC 1 (Cutinase) harbouring Δ*flgKL* Δ*clpX* strain *via* the FT3SS (Fig. 5.14-16). Bioinformatics analysis following Sanger sequencing and BLAST searches revealed that it contains partial *gspK* and *gspL* gene sequences that might have encoded partial proteins. The *gspL* gene encodes the Type II secretion protein involved in a type II secretion system (T2SS-formerly general secretion pathway (GSP)) for the export of proteins (Korotkov *et al.,* 2012). It is a member of an operon of genes (*gspC-O*) that are generally not expressed, however are homologous to those encoding the secreton or type II secretion machinery in *Klebsiella oxytoca* (Pugsley *et al.*, 1997; Nivaskumar *et al.*, 2014).

The *GspL* is an inner membrane protein that interacts with GspC, GspE and GspM proteins in the formation of hetero-oligomer formation (Pugsley *et al.*, 1997; Py *et al.,* 2001; Sandkvist *et al.,* 2001). As the partial *gspK* and *gspL* had significantly improved protein secretion, therefore it needs to be further investigated in order to validate the genomic screening approach utilised in this chapter and in order to create super secretor strain. It could be investigated by either independently over-expressing into pSC 1 harbouring Δ*flgKL* Δ*clpX* strain or removing from it with mutagenesis and simultaneously checking the protein secretion with respect to that of the control pSC 1-Δ*flgKL* Δ*clpX*.

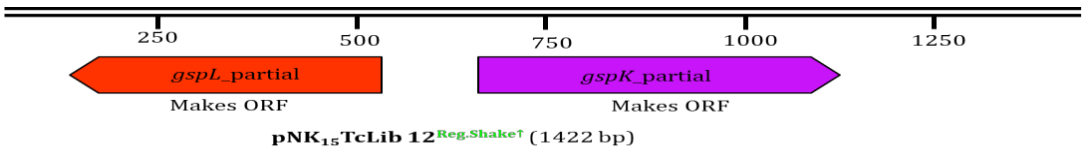


Figure 5.19: pNK$_{15}$TcLib$^{Reg.shake↑}$ clone.

4. pNK$_{15}$TcLib 25$^{Reg.Shake\downarrow}$

pNK$_{15}$TcLib$^{Reg.Shake\downarrow}$ clone 25 (Fig. 5.20) had increased the protein secretion in MUB assays (Fig. 5.14-15) but it was significantly decreased in the western immunoassay ((p<0.05) (Fig. 5.16)). Following Sanger sequencing and BLAST searches it was identified that it contained partial sequences of both *adhE* and *ychE* genes that encodes Alcohol dehydrogenase/aldehyde-dehydrogenase (Kessler *et al.,* 1991) and putative inner membrane proteins (Daley, 2005), respectively. It was also found that there was a large non-coding region in between *dhaE* and *ychE* genes. Overexpression of this complete sequence had decreased the protein secretion as mentioned before. However, it is not clear if the partial truncated protein or the regulatory region had affected the protein secretion of pSC 1 (Cutinase) *via* the FT3SS of Δ*flgKL* Δ*clpX* strain. It could be investigated by quantifying the expression levels of this transcript by using qPCR.

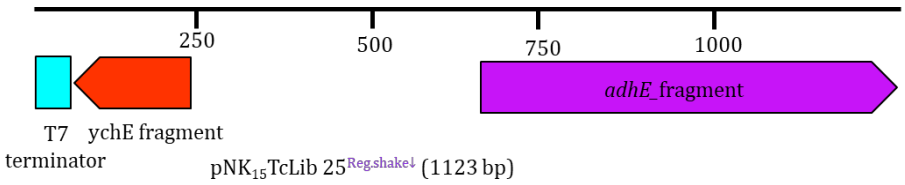


Figure 5.20: pNK$_{15}$TcLib$^{Reg.Shake\downarrow}$ clone.

5. pNK$_{15}$TcLib 90$^{Reg.Shake\downarrow}$

pNK 90$^{stat.\downarrow}$ clone (Fig. 5.21) contains incomplete *gshA* and *yfbV* genes that code for Glutamate-cysteine liase and a UFO028 membrane protein (*yfbV*) on either end of the complete *yqaA* gene, which is a member of a 7 protein DedA family protein, located in the cytoplasm (Daley, 2005; Doerrler *et al.*, 2013). Of interest, it was reported in a genome-wide screening that the *yqaA* mutant has a defect in swarming but not swimming motility (Inoue *et al.*, 2007). However, overexpression of *yqaA* in this work significantly reduced the protein secretion in Δ*CKL* Δ*clpX* strain harbouring pSC 1. It would be interesting to investigate the decrease in protein secretion which might suggest that in response to increased YqaA protein in the medium which signals the onset of swarming the flagella master regulator $FlhD_4C_2$ might induce the flagella gene expression (Fraser and

Hughes, 1999; Inoue *et al.*, 2007; Doerrler *et al.*, 2013). It was also observed that protein secretion was considerably reduced in presence of the YqaA protein. Therefore, it might be worth investigating whether the levels of hook-basal body were increased than control Δ*CKL* Δ*clpX* strain in response to increased YqaA protein. It can be tested by cloning of *yqaA* gene and monitoring the number of flagella formation maybe by electron microscopy or gene mutagenesis could be utilised to investigate the correlation between number of flagella hook-basal body formation and the protein secretion *via* the FT3SS (Turner *et al.*, 2010).

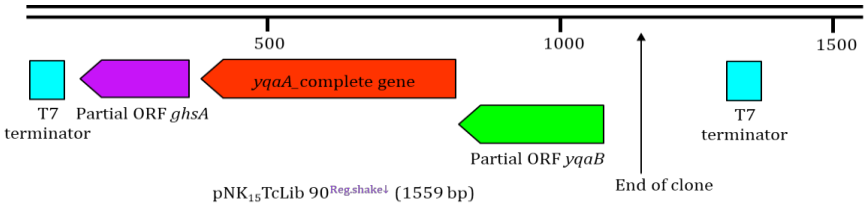


Figure 5.21: pNK$_{15}$TcLib 90$^{stat.\downarrow}$ clone.

As discussed in Section 5.6, several potential protein secretion affecting genes loci were identified. In order to verify the genomic screening approach utilised here for genomic library construction and to further engineer the secretor strain for improving protein secretion, the findings of this chapter need to be investigated. However, due to time constraints, only secretion decreasing pNK 37$^{stat.\downarrow}$ that encoded Crl and secretion improving pNK 12$^{Reg.shake\uparrow}$ that encodes partial GspL were selected for further analysis. These two proteins will be investigated to verify and validate the genomic screening approach utilised for improving protein secretion *via* the FT3SS by means of highlighting the true nature of these two proteins by either expressing into new strain or removing with mutagenesis (See later Chapter 6).

5.7 Discussion

The aim in this chapter was to develop and test the feasibility of a genomic screening method in order to try and identify genes that influence FT3SS secretion using a previously developed secretion assay based on Cutinase secretion. This method could be argued to be key to what has now been termed as an inverse or reverse genetic engineering (Bailey *et al.*, 2002). This approach has been used by others in the past, such as in studies to improve glycosylation

of proteins in *E. coli* (Pandhal *et al.*, 2012). Any genes identified are then targeted for knock-out or knock-ins to improve the desired phenotype.

The approach taken here was to create a random gene library of *E. coli* ligated into a compatible plasmid. As outlined above a bespoke plasmid was designed *in silico* and was successfully tested in this chapter. Specifically, it was shown to co-exist with our secretion plasmid (containing the FT3SS-Cutinase construct) and not affect growth rates- thus suggesting its potential for use for genomic screening. In addition to compatible Ori and antibiotic resistance a bespoke MCS which allowed ligation of genomic DNA cut with high frequency (4-base cutter) *vs* a 6-base cutter with one cut -site in the plasmid was constructed successfully. This approach seemed to be successful given randomly sized clones seem to have been cloned into this vector, and those sequenced to be spread around the genome.

One limitation of the work presented here was that despite some pre-fractionation of DNA fragment size, most of the cloned fragments were in the range 0.8-2.0 kb, whereas it was envisaged to produce a library with 2-5.0 kb fragments. Therefore, in order to improve on the current work, one might improve the size fractionation protocol by repeated purification based on size; however, time precluded this during my study. Overall, it seems though that this plasmid was useful in genomic library screening and also in a dual plasmid system such as the one developed here, but also for other functions such as complementation of mutations, expression of snRNA etc. However, given more time a more thorough characterisation of the plasmid, such as confirming its copy number would be carried out. Although we envisage that the new plasmid pNK$_{15}$TcLib had the same copy number (15 per cell) as its parent plasmid pACYC184, we have not confirmed this experimentally. Similarly, we have not fully established the level of expression from this plasmid's inserts or that the terminators are fully functional-something appropriate qPCR would allow.

Alongside a facile and effective plasmid system, it was essential for our work to establish a mean to screen secretion as effectively, accurately and quickly as possible. To achieve this a high-throughput screen based on Cutinase secretion

and detection by MUB assay was adapted from the work in the previous chapter. While it has its limitations, it proved a useful method to screen large numbers of colonies.

In the preliminary screening presented in this chapter, it was perhaps a surprise that on the whole the screen seemed more adept at identifying clones with reduced secretion levels than increased. There may be several reasons for this, although we have no evidence for any at this time. Firstly, the screen itself may have been limiting *i.e.* the amount of supernatant used might need to be decreased so that the baseline level of secreted enzymes were lower and the chances of using up all the substrate (and hence reaching maximal fluorescence) was reduced. Similarly, the length of time of incubation could be varies- however in both cases time did not allow such tests to be carried out.

Next, the application of this screening approach could be extended for the investigation and lab-based evolution of potential genes for several biological processes, including protein production, glycosylation, motility, stress response, antibiotic resistance (Bailey *et al.*, 2002; Ghosh *et al.,* 2012; Pandhal *et al.*, 2012). It would be very interesting to design an experiment to investigate the correlation between the intracellular expression, protein secretion in western immunoassay and the MUB assay (both in conventional and improvised), which could be then utilised with confidence for the estimation of expression in preliminary analysis itself without the need of performing western immunoassay. Despite these limitations, MUB assays could serve as an appealing method for preliminary screening to narrow down a large number of library clones to a few potential candidates, as it is less cumbersome.

As a further proof of principle in this thesis clones that affected significantly altered FT3SS secretion were sequenced to identify the genome fragments contained therein. This was partly to see if the kinds of genes identified might have a logical explanation with respect to the FT3SS secretion but also would act to test the random nature of any inserts. Sanger sequencing and BLAST searches revealed a number of interesting genes, which were randomly incorporated in the pNK plasmid, irrespective of their nucleotide sequence orientation. These

included a clone that decreased secretion containing the intact Crl gene/protein. In fact, the findings that Crl had an effect that could be predicted partly validates this approach in that it has the ability to identify clones that one might expect to affect flagella secretion but also given several of the other clones did not. One issue however with the data obtained in our small preliminary screen was the size of genomic fragments, which was only (up to 2 kb) and hence contained many gene fragments- as seen with the *gspL* containing clone ($pNK_{15}TcLib^{Reg.shake\uparrow}$). Future studies should involve individual cloning or knockout of identified genes, but also consider and estimate copy numbers of plasmids, internal levels of the Cutinase *etc*. in the cells. This is pertinent as there are a number of points at which secretion from our FT3SS system could be interrupted. This could occur at the level of transcription of flagellar genes (as with Crl) or interactions with the substrate protein in the cytoplasm (although we can rule out FT3SS chaperones as these binding domains are absent in our construct). In addition, proteins could interact with the export apparatus and preclude hook formation for example. This simply means that any post-identification analysis should examine several aspects of the mechanism of nay effects- these might include qPCR of flagella genes, export of other flagella proteins, levels of Cutinase inside the cell, the amount of specific Cutinase and other mRNA- and possibly even RNA half-lives as well as plasmid copy numbers as discussed above. For example, it is likely that extra copies of pNK encoded genes, protein or regulatory regions (and native copy) and their downstream target proteins might create obstacles and obstruct Cutinase from efficiently reaching the export gate for secretion or the proteins might compete for export chaperones (Phillips and Silhavy, 1990; Nishihara *et al.*, 1998; Martínez-Alonso *et al.*, 2010).

It might be worth investigating whether reduction in protein export was due to insufficient expression, but from western blot analysis on representative blots of intracellular proteins it is highly unlikely the case (Fig. 5.13-5.16). Additionally, it was not clear in the library whether the obtained phenotype was a result of a single gene expression or multiple genes, as random genomic DNA fragments were incorporated in the library plasmid (Yuan *et al.*, 2005). Potential

clones harbouring one or more gene obtained following screening could be investigated by independent gene mutagenesis or the entire operon could also be mutated (Datsenko and Wanner, 2000). Furthermore, pNK could also be utilised with independent cloning of potential genes and co-expressing as before with pSC 1 in secretor strain Δ*flgKL* Δ*clpX.* qPCR technique might be worth for these investigations which are less cumbersome and yet accurate than molecular cloning in pNK or clean mutagenesis. qPCR gives relative quantification of mRNA expression levels (Bustin, 2000; Vandesompele *et al.*, 2002). In addition to discussion points made under each heading, general themes were emerged for improving secretion and are reported in Chapter 7.

CHAPTER 6

INVESTIGATION OF THE GENES HIGHLIGHTED FROM THE PROTEOMICS ANALYSIS, GENOMIC LIBRARY AND POTENTIAL REGULATORY NON-CODING RNA

6 Introduction:

As outlined in Chapter 3, the platform secretion strain and signal were established using information based on observations of a previous student (Green, 2016), where it was identified that a *clpX* gene deletion increased protein secretion through the FT3SS. At least in part due to its ability to upregulate flagella gene expression by increasing the *flhDC* levels (Tomoyasu *et al.*, 2002). Additionally, it was observed that exclusion of *fliC*-3'UTR in the pJEXpress 404 significantly improved protein secretion, irrespective of whether an early or late secretion signal was used (Chapter 3). Therefore, based on these observations, to further improve protein secretion from the platform, parallel approaches of both systems biology (proteomic: Chapter 4) and use of an unbiased genetic library (Chapter 5) were employed, which revealed a wealth of information for improving protein secretion *via* the FT3SS, which was now followed up.

Additionally, given that *clpX* mutation probably improves secretion *via* upregulation of FlhDC, other means to achieve this were also investigated. This leads to an investigation of role of small non-coding RNA (snRNA), which may directly or indirectly regulate the $FlhD_4C_2$ complex (Lay and Gottesman, 2012; Fitzgerald, Bonocora and Wade, 2014; Ludwig *et al.*, 2018). A recent literature search revealed ~280 snRNA ranging from 100 to 250 nucleotides long, in the genome of *Salmonella* (Lay and Gottesman, 2012; Miranda-CasoLuengo *et al.*, 2017). However, only less than 2% snRNAs have been studied in detail and the role of the remaining snRNA remains enigmatic (Miranda-CasoLuengo *et al.*, 2017). As the snRNA regulate important cellular processes directly or indirectly including bacterial chemotaxis and motility, it was attempted in this chapter to examine the role of a limited set of snRNAs on protein secretion *via* the modified FT3SS.

The regulation of motility and *flhDC* by snRNA is at least as complex as that seen with other levels of regulation and for the first time it was reported that the FlhDC transcriptional regulators are subject to negative regulation as well as positive regulation (Lay and Gottesman, 2012). The snRNAs such as ArcZ, OmrA, OmrB, OxyS negatively regulate, while McaS, MicA positively regulate motility

and FlhDC expression by base-pairing with 5'UTR of this mRNA (Lay and Gottesman, 2012). Using this information, it was hypothesised that the protein secretion of (Cutinase) pSC 1 *via* the FT3SS could be improved by over-expressing activating snRNAs or by the exclusion of negative regulators of the $FlhD_4C_2$ complex. The positively regulating snRNAs (McaS, MicA) bind to the flagella master regulator *flhDC* mRNA and opens it for flagella gene expression, keeping it in 'ON' mode (Fig. 6.1). These snRNAs could be cloned into previously constructed pNK plasmid for improving protein secretion upon, co-expression with pSC 1 (5'UTR-$FliC_{47aa}$-Cutinase -pJExpress-No-3'UTR (Fig 3.6)). However, the negative regulators (*ArcZ, OmrA*) of *flhDC* prevent transcription and can be removed from the cell by deletion mutagenesis and thus the master regulator can be potentially upregulated (Fig. 6.1). This should increases expression of the FlhDC complex which results in downstream expression of flagella Class II and Class III genes (Aldridge and Hughes, 2002).

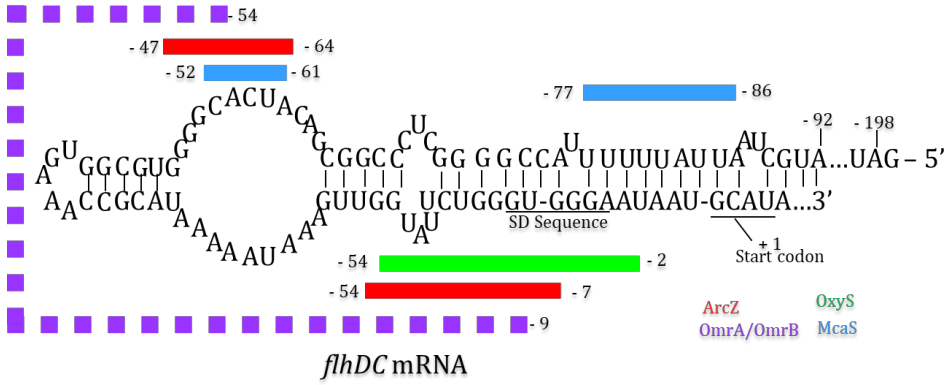


Figure 6.1: Diagram summarising predicted binding sites of snRNA's on the Flagella master regulator $FlhD_4C_2$ mRNA.
Adapted from (Lay and Gottesman, 2012; Mika and Hengge, 2013)

As reported previously, a number of genes, proteins and pathways were highlighted for improving the heterologous protein secretion of (Cutinase) pSC 1 *via* the FT3SS in Chapter 4 (proteomics) and Chapter 5 (Genomic Library). Therefore, *clpB*: an ATPase associated with a variety of cellular activities and *ycgR:* flagellar brake protein were selected from the proteomics analysis while, *crl:* a putative sigma factor binding protein involved in the expression of *csgBA* operon that leads to Curli fibre formation and impacts a regulatory crosstalk with the flagella system and *gspL*: a type II secreton from genomic library were

selected for validation of those findings in this chapter. Also in parallel, two negative regulators of *flhDC*, *arcZ* and *omrA* were selected for the investigation of the potential role snRNA in improving protein secretion *via* the FT3SS.

Aim: The major aim of this chapter was to investigate the non-coding RNA for improving protein secretion *via* the FT3SS and to validate the findings of proteomics (Chapter 4) and preliminary genomic library screening (Chapter 5).

Results:

6.1 Small non-coding RNA (snRNA) regulation of the flagella master regulator $FlhD_4C_2$ Complex- a potential fulcrum for intervention

As mentioned earlier, some snRNAs directly influence the expression of the flagella master regulator $FlhD_4C_2$ complex. In order to investigate the effect of snRNAs on recombinant protein secretion *via* the FT3SS, snRNAs that downregulate the flagella gene expression *via* $FlhD_4C_2$ complex were selected from the literature (Lay and Gottesman, 2012). Thus, the snRNA-*arcZ* and *omrA* were selected which decrease protein expression by binding to *flhDC* mRNA (Lay and Gottesman, 2012). It was hypothesised that the recombinant protein secretion *via* the FT3SS could be significantly improved by the removal of *arcZ* and *omrA* snRNAs, which directly inhibits protein expression by occupying the upstream promoter sequences of the $FlhD_4C_2$ complex or *via* directly binding to *flhDC* mRNA (Fig. 6.1). As explained above, the main rationale here was the dramatic effect of the *clpX* mutation, which probably acts mainly on FlhDC increasing secretion so dramatically. However, another driver was to potentially utilise a means to boost FlhDC levels that did not have such potentially pleiotropic effects on the cell as removal of a major protein quality control system as ClpX. Therefore, snRNAs (*arcZ* and *omrA*), were selected for knockout mutagenesis, along with genes selected from proteomics (*clpB* and *ycgR*: Chapter 4) and the preliminary genomic library screen (*crl* and *gspL*: Chapter 5) with λ-red recombineering in *ΔCKL ΔclpX* strain (Section 2.6) with the stepwise workflow shown in Fig. 6.3 (Datsenko and Wanner, 2000; Thomason *et al.*, 2007). The pKD3 was used as a template for assembling the antibiotic cassette expecting 1.1Kb PCR product (Fig. 6.3).

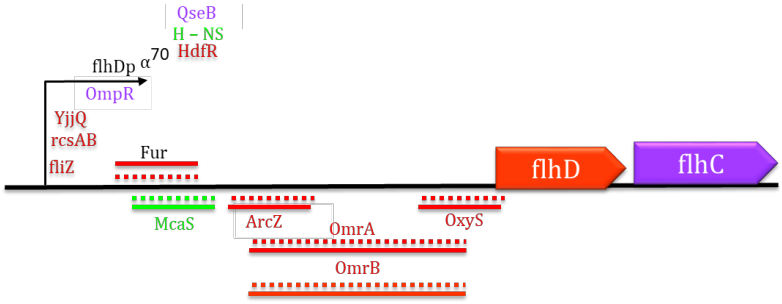


Figure 6.2: Binding of small non-coding RNA to Flagella master regulator
Binding snRNA upstream of FlhDC complex, adapted from (Lay and Gottesman, 2012).

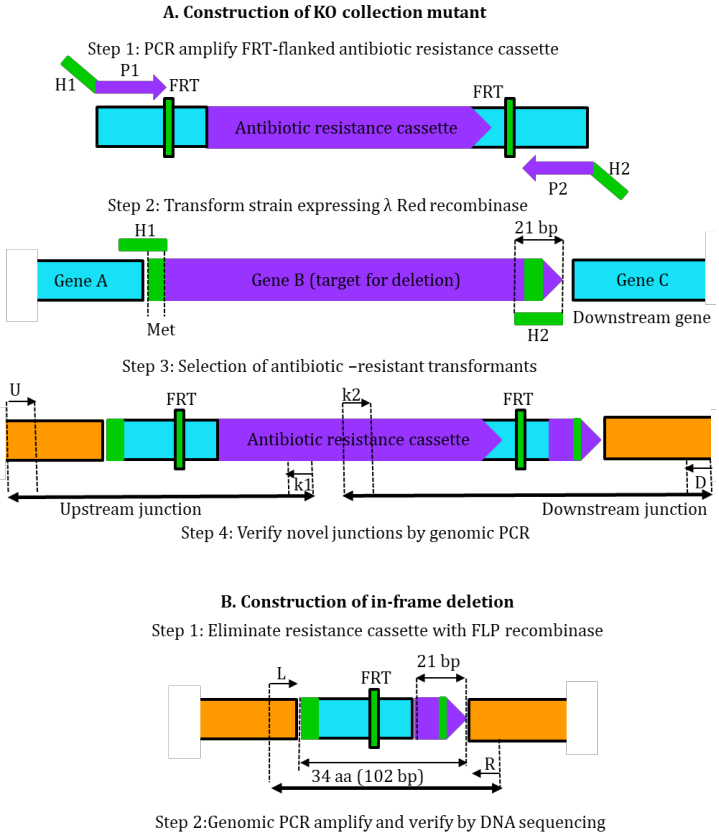


Figure 6.3: PCR gene replacement strategy.
Gene targeting fragments encoding antibiotic resistance with short homology extensions (H1 and H2) were generated by using priming sites P1 and P2 (step 1). The gene targeting fragments were introduced into Δ*CKL* Δ*clpX* strain, expressing the Red recombinase from pKD46 (step 2). Antibiotic-resistant transformants were selected (step 3) and verified with PCR (step 4). The elimination of the resistance cassette with the FLP recombinase plasmid pCP20 was expected to leave behind a 102 bp scar a 34-residue peptide (step 1). The scar region was amplified and sequenced to be sure no mutations occurred. Adapted from (Datsenko and Wanner, 2000; Baba *et al.*, 2006).

PCR products for mutagenesis for all target genes were produced and electroporated into the Δ*CKL* Δ*clpX* strains with successful transformants selected on Cm^R plates. Prior to verification of successful mutagenesis with

colony PCR, plasmids were cured from all the strains. Figure 6.4 shows the presence of the mutagenesis cassette in the chromosome of screened clones from these experiments and verifies successful mutant generation.

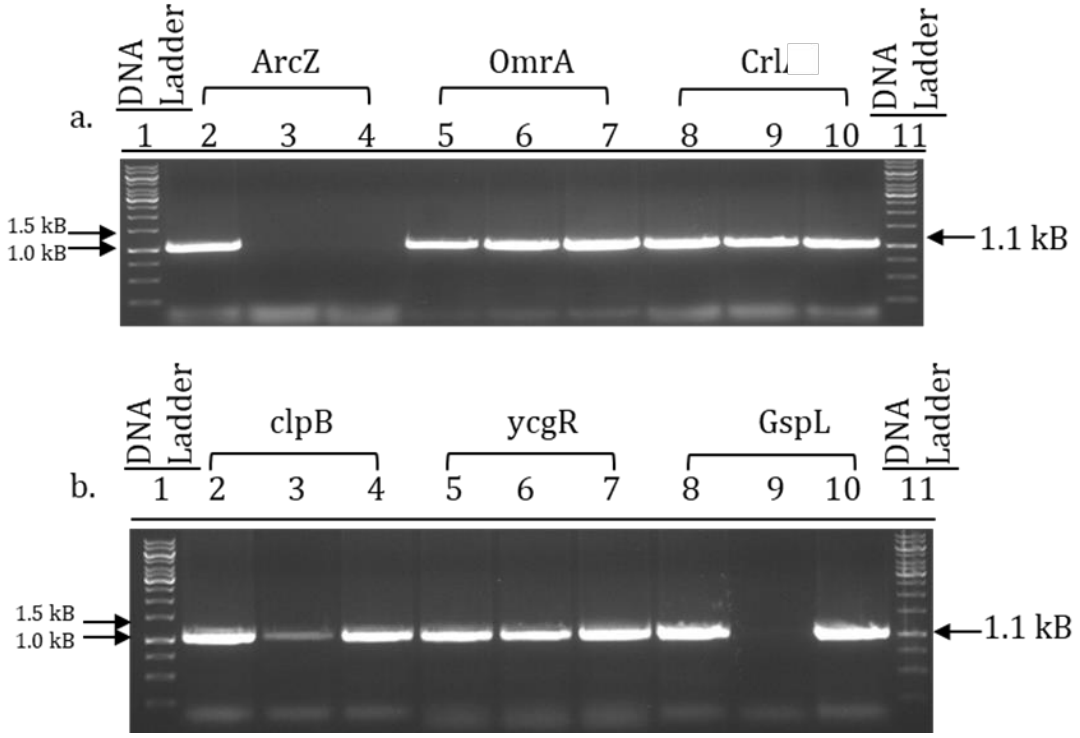


Figure 6.4: Amplification of antibiotic cassette for KO mutagenesis.
The antibiotic resistance cassette was PCR amplified by using pKD3/pKD 4 plasmids as a template, expecting 1.1 kb PCR products as shown. PCR reactions were run with, *arcZ* (Lane 2-4), *omrA* (Lane 5-7), *crl* (Lane 8-10) (a.) and *clpB* (Lane 2-4), *ycgR* (Lane 5-7) and *gspL* (Lane 8–10) (b.). Lane 1 and 11 were loaded with 1.0 KB Ladder (1 Kb EZ Run) (a. and b.). Reactions were run on 1 % agarose gel, at 150 volts for 65 min. Pictures taken by exposing the gel to UV Light, at 302 nm in G-box (SynGene Technologies).

6.2 Growth effects of the selected mutations

The KO mutants created (Section 6.1) were grown in liquid broth culture along with the parent strain *ΔCKL* Δ*clpX* and their overnight growth curves were monitored in order to investigate any growth defects (Fig. 6.5). These data show that all the mutants display a longer lag-phase but then progressed to similar final OD_{600} and with similar growth rates to the parent strain except for the *gspL* strain which had a growth defect (Fig. 6.5f.). These mutants were utilised for the recombinant protein secretion *via* the FT3SS, in Section 6.3.

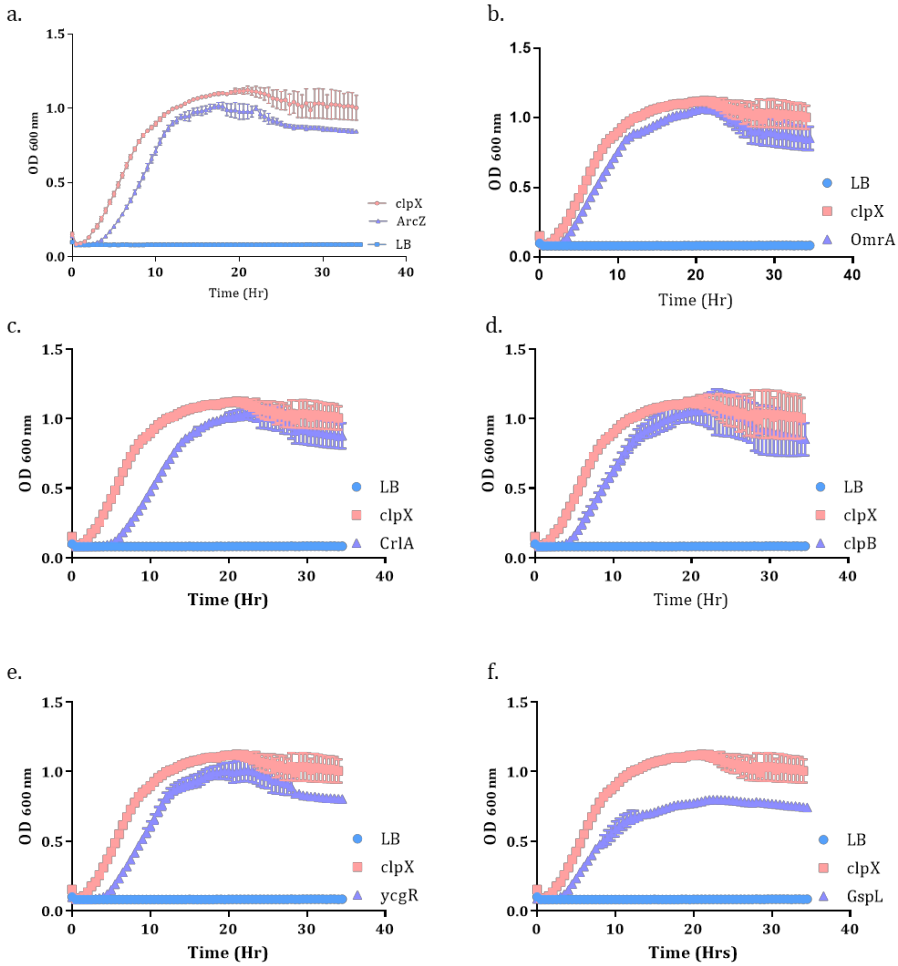


Figure 6.5: Monitoring the growth curves of KO mutant strains.
The Δ*CKL* Δ*clpX* parent strain was used to create the ΔarcZ, Δ*omrA*, Δ*crl*, Δ*clpB*, ΔygcR and Δ*gspL* mutants, as shown in a, b, c, d, e and f, respectively. Their overnight growth curves were observed by reading $O.D._{600}$ with Plate reader (Tecan Magellan, Infinite® 200 PRO), 3 biological and technical repeats were observed, the graph was plotted using mean error bars (±) SEM with Graphpad prism 7.0.

6.3 Protein secretion *via* the FT3SS of KO mutants

The parent strain Δ*CKL* Δ*clpX*, and the KO mutant strains created in Section 6.1 for snRNA ArcZ and OmrA (Δ*CKL* Δ*clpX* Δ*arcZ* and Δ*CKL* Δ*clpX* Δ*omrA*) and ClpB and YcgR (Δ*CKL* Δ*clpX* Δ*clpB* and Δ*CKL* Δ*clpX* Δ*ycgR*) and Crl and GspL (Δ*CKL* Δ*clpX* Δ*crl* and Δ*CKL* Δ*clpX* Δ*gspL*) respectively, were transformed with the secretion construct pSC 1 (5'UTR-$FliC_{47aa}$-Cutinase -pJExpress No 3'UTR (Fig. 3.7)). These Δ*CKL* Δ*clpX* and the KO mutants harbouring pSC 1 were grown in liquid broth culture as before (Section 2.7.2.) and the following day re-inoculated into fresh medium containing 0.05 mM IPTG. The supernatants were then collected and the MUB assay was performed as before (Section 2.10.3 and 3.2.1)

and the results are reported in Fig. 6.6. It can be observed (Fig. 6.6) that snRNA mutant Δ*CKL* Δ*clpX* Δ*arcZ*, expressed with pSC 1, secretes significantly higher Cutinase than the parent strain Δ*CKL* Δ*clpX via* the FT3SS. There was no significant difference in protein secretion of Δ*CKL* Δ*clpX* and the other KO mutants except the Δ*CKL* Δ*clpX* Δ*gspL* strain. The protein secretion in *gspL* KO strain was decreased- however this strain did have a growth defect that might explain any reduction (Fig. 6.6 a. and b).

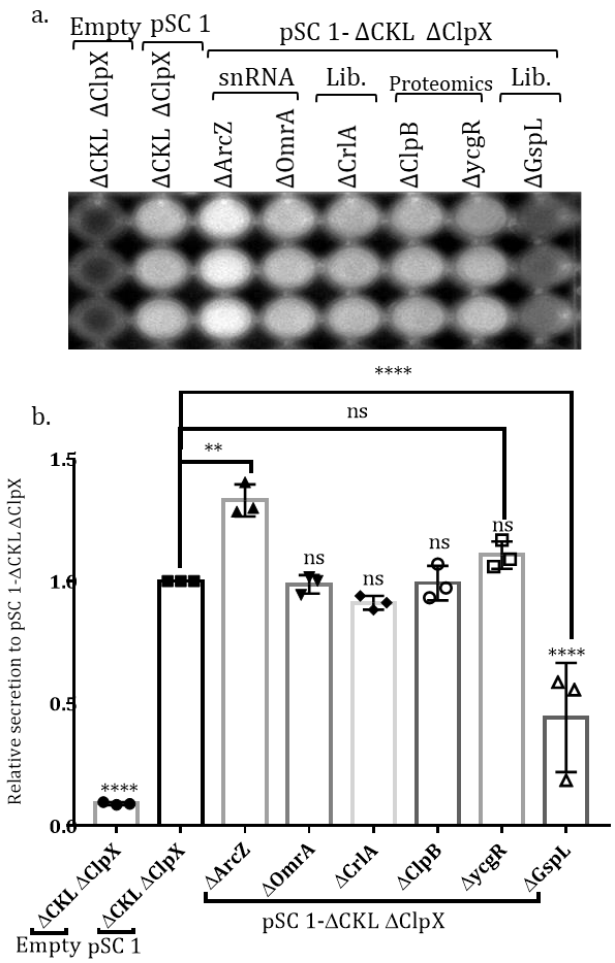


Figure 6.6: Secretion efficiency of KO strains with MUB assay.
(a.) MUB assay with controls PCB, MUB, LB, LB+MUB and Δ*clpX* strain in well 1 serves as a negative control. pSC1 expressed into Δ*flgKL* Δ*clpX* (Positive Control) and (Δ*flgKL* Δ*clpX*), *(*Δ*clpX* Δ*omrA), (*Δ*clpX Δcrl), (*Δ*clpX* Δ*clpB), (*Δ*clpX* Δ*ycgR)*, *(*Δ*clpX* Δ*gspL)* strains in well 2–8, respectively. Fluorescence detected at 302-446 nm and visualised by exposing the plate to U.V. Light at 302 nm. (b) Graph plotted with Graphpad prism 7.0, error bars (±) SEM, statistical significance defined by ****p< 0.0001, determined by One way ANOVA for multiple comparisons, results from 3 biological replicates (n=27).

To follow up the MUB assay data- especially as the MUB assay often underestimates secretion increased, the intracellular protein expression and the

secretion were also verified with western immunoassay as before (Fig. 6.7 and Fig. 6.8). It can be observed (Fig. 6.7) that, recombinant protein-Cutinase was evenly expressed in the Crl and ClpB strains compared with the parent strain. However, the snRNAs, *arcZ* and *omrA* as well as *ycgR* show 2-fold improved protein expression over the parent strain Δ*CKL* Δ*clpX* (p<0.05). It can also be observed that *gspL* KO mutant identified from genomic library shows 3-fold increased intracellular Cutinase expression (p<0.0001) than the Δ*CKL* Δ*clpX* strain.

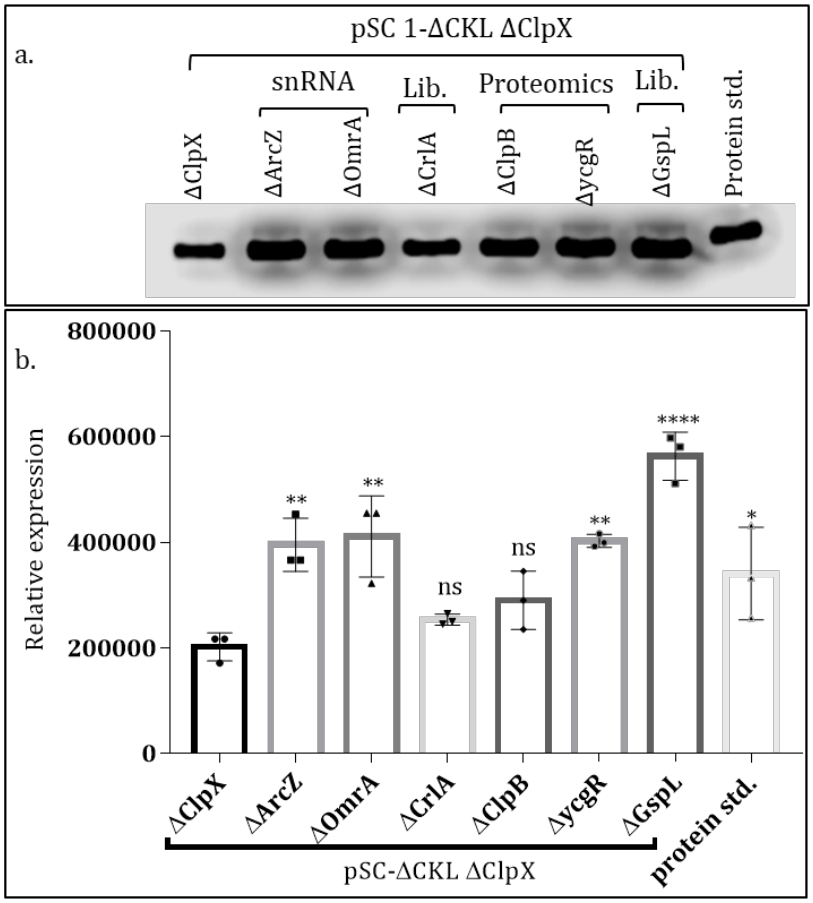


Figure 6.7: Protein expression of KO strains with Western blot analysis.
Intracellular protein expression in KO strains (Lane 2-7) was compared using pSC 1 (5'-C_{47aa}-Cut-pJEx No 3') expressed into Δ*flgKL* Δ*clpX* strain as a Positive control, in Lane 1. Pre-stained protein ladder shown, Lane 10 loaded with protein standard. Mutants *(*Δ*clpX* Δ*arcZ), (*Δ*clpX* Δ*omrA), (*Δ*clpX* Δ*crl), (*Δ*clpX* Δ*clpB), (*Δ*clpX* Δ*ycgR)* and *(*Δ*clpX* Δ*gspL)* were loaded in well 2-7, respectively. Blots were developed on *LI-COR,* C–Digit. Graph plotted with Graphpad prism 7.0, error bars (±) SEM, statistical significance was defined by ****p<0.0001, determined by One way ANOVA for multiple comparisons.

While, in the case of extracellular (secreted) proteins, it was observed that recombinant protein secretion *via* the FT3SS of Δ*flgKL* Δ*clpX* Δ*arcZ* and Δ*flgKL* Δ*clpX* Δ*omrA* strains was significantly increased 4.73-fold, (p<0.0001) and 1.16-fold (p<0.005) than the parent strain, Δ*flgKL* Δ*clpX*. It was also found that the protein secretion in both *clpB* and *ycgR* mutants was also significnatly increased by 1.41-fold and 1.6-fold (p<0.0001), respectively, (Fig. 6.8), however *crl* and *gspL* mutants had significantly decreased protein secretion *via* the modified FT3SS (p<0.05) of the parent Δ*flgKL* Δ*clpX* strain.

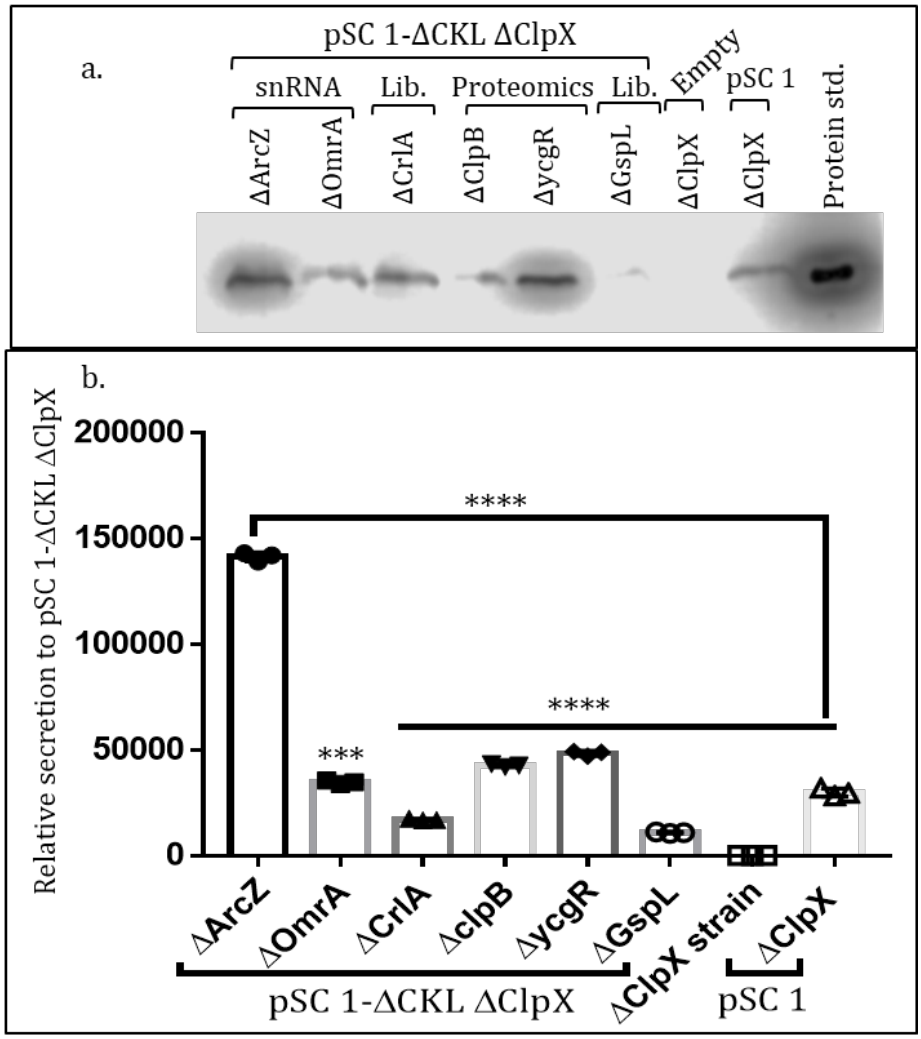


Figure 6.8: Protein secretion of KO strains with Western blot analysis.
Extracellular (secreted) protein expression in the KO strains (Lane 2–7) was compared with pSC 1 expressed in Δ*flgKL* Δ*clpX* in Lane 8 (Positive control). Lane 7 loaded with Δ*flgKL* Δ*clpX* strain harbouring empty plasmid (Negative control), Lane 9 loaded with protein standard. Blots were developed on *LI-COR,* C-Digit. 3 biological repeats observed, graph plotted with Graphpad prism 7.0, error bars (±) SEM, statistical significance defined by ****p<0.0001, determined by One way ANOVA for multiple comparisons.

6.4 Discussion

As previously reported, the snRNAs such as *arcZ* and *omrA*, directly affect $FlhD_4C_2$ complex and significantly regulate downstream flagella genes at both transcriptional and translational level (Lay and Gottesman, 2012). In order to investigate the role of snRNA with respect to recombinant protein secretion *via* the modified FT3SS, again used the combination of our secretor strain, construct (pSC 1) and Cutinase assay.

The significant increase in recombinant protein secretion in the Δ*CKL* Δ*clpX* Δ*arcZ* strain than the parent Δ*CKL* Δ*clpX* strain could be attributed to the removal of *arcZ* snRNA, as it has been reported that the *arcZ* negatively regulates the *flhD* and *flhC* by direct base-pairing to a region upstream of the ribosome binding site (Lay and Gottesman, 2012). The *arcZ* is transcribed as a 121 ntds RNA, which is processed to generate a smaller, stable RNA that consists of the last 56 ntds of the original transcript (Argaman *et al.*, 2001; Papenfort *et al.*, 2009). Additionally, it has been also reported that over-expression of *arcZ* in *S. enterica* lowered the levels of the mRNA encoding *flhD* and *flhC* by threefold, leading to reduction in motility whereas Gottesman *et al.,* (2012) demonstrated the complete elimination of motility with over-expression of *arcZ* (Lay and Gottesman, 2012). Furthermore, ArcZ positively regulates *rpoS* mRNA translation as a result of direct base pairing between a region spanning the first 26 ntds of the smaller, processed snRNA or 66-91 ntds of the full length transcript and a region in the leader of *rpoS* mRNA which in turn, RpoS negatively regulate *flhDC* indirectly by the stimulation of synthesis of SdsR and GadY snRNAs (Fig. 6.9) (Opdyke *et al.,* 2004; Uchiyama *et al.*, 2010; Fröhlich *et al.*, 2012). Similarly, it has been also reported that ArcZ stimulates motility in part *via* ArcZ and in part *via* the regulation of the anti-sigma factor FliA (Kato *et al.*, 2007). Therefore, the exclusion of this negative regulatory snRNA *ArcZ* from Δ*CKL* Δ*clpX* strain significantly improved the protein secretion *via* the FT3SS, in agreement with previous work, as removal of ArcZ keeps *flhDC* mRNA transcript free for binding transcription machinery and the positive regulator snRNA such as *McaS* which shares the same binding sites as *ArcZ* in the *flhDC* mRNA (Lay and Gottesman, 2012; Thomason *et al.*, 2012). Therefore, the exclusion of *ArcZ* might allow the

binding of *McaS* to *flhDC* mRNA more efficiently, leading to the continuous transcription and translation of $FlhD_4C_2$ complex and the downstream Class II and Class III genes of the bacterial flagella biogenesis (Aldridge *et al.,* 2002). However, the mechanistic details of this regulation in this case are not clear, and it would be therefore worth testing the expression levels of *flhDC* and *fliA* RNA by using qPCR and conversely the effect of overexpression *arcZ*. Similarly, one might also consider removing ArcZ from a strain that did not have the ΔClpX allele- to ask the question whether any effect were cumulative and whether fitness or protein secretion might be improved if snRNA mutations were introduced into more wild-type backgrounds. This is worth considering bearing in mind given the pleiotropic effects of *clpX* mutations.

Conversely, only a very small 16% improvement in secretion was seen with the *omrA* mutant. The *omrA* codes for an 88 ntds snRNA that has been shown to negatively regulate the *csgD* and *flhDC* mRNAs (Holmqvist *et al.*, 2010; Lay and Gottesman, 2012). The *OmrA* exhibits its inhibitory action on the master regulator, similar to *ArcZ* by binding upstream of *flhDC* operon and base pairing with the *flhDC* mRNA. However, *OmrA* exhibits its action in association with *OmrB,* but there is redundancy in their action with one able to compensate for the other (Lay and Gottesman, 2012). This is probably the case here but would need to be tested by production of a double *omrAB* mutant. Furthermore, the *arcZ* and *omrAB* mutations could be combined in the creation of new strain that might exhibit the additive phenotype, meaning it would improve secretion *via* the FT3SS. In fact, combination with the positive *McaS* RNA may boost expression still further.

As outlined in this chapter, the *clpB* and *ycgR* genes were highlighted from the proteomics analysis, ClpB protein was significantly upregulated while YcgR was significantly downregulated, as reported (Table 8.4 - 8.8). In order to validate the findings of proteomics and investigate the role of these ClpB and YcgR proteins during protein secretion, their genes were knocked out to create Δ*CKL* Δ*clpX* Δ*clpB* and Δ*CKL* Δ*clpX* Δ*ycgR* mutant strains, respectively (Fig. 6.3-

6.4). These KO strains increased by secretion at relatively low levels of 41 and 61% ($p < 0.0001$) fold respectively, over the parent Δ*CKL* Δ*clpX* strain (Fig. 6.8).

The YcgR protein was significantly downregulated during high secretion comparison in the preliminary proteomics analysis (Table 8.8-8.8.). The YcgR acts as a flagellar brake protein, which regulates the swimming and swarming in a bis-(3-5) cyclic diguanylic acid (c-di-GMP) dependent manner (Ryjenkov *et al.*, 2006; Fang *et al.,* 2010; Paul *et al.*, 2010). However, the mechanism is not fully known, when bound to c-di-GMP the YcgR binds to elements of the Flagellar motor (MotA) (Boehm *et al.*, 2010) and or the FliG and FliM (Paul *et al.*, 2010). The binding of YcgR to FliM occurs in the absence of c-di-GMP, causing the motor to slow down, which suggests the increasing levels of c-di-GMP leads to decreased motility (Fang *et al.,* 2010). Furthermore, entry into stationary phase initiates production of increased levels of c-di-GMP, which binds to YcgR and results into slower flagellar motor speeds and increased counter-clockwise flagellar rotation bias (Fang *et al.,*, 2010; Paul *et al.*, 2010). Additionally, it was reported that a mutation in H-NS that leads to non-flagellation in *E. coli* due to reduced expression of the *flhDC* master operon, YcgR suppresses the motility defect of H-NS-deficient cells (Ko and Park, 2000b). The promoter analysis of *YcgR* revealed that it contains (σ^F) –35 and –10 consensus sequences of (TAAA-N15-GCCGTAA) of class III flagellar genes, consistent with their dependence on FlhDC (Bartlett et al., 1988).

However, during protein secretion *via* the FT3SS, motility was not desired and therefore functional flagella were modified. The significant downregulation of YcgR during protein secretion maybe represents the decreased c-di-GMP levels that correspond to increased recombinant protein secretion as the YcgR-c-di-GMP complex might interfere with proton conducting motor protein MotA rotation by interfering with the proper association of the Mot proteins with FliG (Wolfe and Visick, 2008; Paul *et al.*, 2010). Therefore, to validate these findings, it was hypothesised that the removal of YcgR protein that directly interacts with the flagella proteins MotAB and FliM would increase protein secretion *via* the FT3SS and it was observed that protein secretion was significantly increased

(1.614 fold, $p<0.0001$) over the control Δ*CKL* Δ*clpX* strain, as expected, which supports the above hypothesis (Mogk *et al.*, 2003). Hence, it can be concluded that Δ*CKL* Δ*clpX* Δ*ycgR* strain is more efficient for protein secretion than Δ*CKL* Δ*clpX* strain. The *ycgR* mutation could also be combined with *arcZ* and *omrA* mutants that showed increased protein secretion (Section 6.5.1).

Also, the ClpB protein was significantly upregulated during secretion in preliminary proteomics analysis. ClpB belongs to the ATPase associated with a variety of cellular activities (AAA^+ superfamily), and together with DnaK chaperone system (DnaK/DnaJ/GrpE) it re-solubilises the aggregated proteins required in the flagellum synthesis (Shi *et al.*, 1992). It has been reported that deletion or mutation of the heat shock chaperones results in a 10-20-fold decrease in the rate of synthesis of flagellin and reduced transcription of *flhD* operon encoding FlhD-FlhC (Shi *et al.*, 1992). Therefore, it was hypothesised that if the ClpB protein was necessary during protein secretion to process the aggregated, over-expressed proteins; removal of *clpB* may decrease the protein secretion (Mogk *et al.*, 2003). However, it was not the case; instead, protein secretion was significantly increased (1.41-fold, $p<0.0001$ (Fig. 6.8)) in absence of the *clpB* gene. This may be due to the metabolic stress on the cell during secretion (Eriksson and Clarke, 2000). It is suggeted that increased ClpB production in pSC 1-Δ*CKL* Δ*clpX*, shows the metabolic stress encountered by the cell, as the protein binding stimulates the ATPase activity of ClpB, ATP hydrolysis unfolds the protein aggregates, which probably helps expose the new hydrophobic binding site on the surface of ClpB-bound aggregate, contributing to the solubilisation and refolding of denatured protein aggregates by DnaK (Barnett *et al.,* 2000), although there is no experimental evidence to support this hypothesis. However, it could be tested by over-expressing ClpB protein by cloning *clpB* gene into pNK plasmid (Fig. 5.6) and expressing in *clpB* mutant strain to see whether it restores the secretion or it could be over-expressed in pSC 1-Δ*CKL* Δ*clpX* strain to see if it influences protein secretion *via* the FT3SS to support this hypothesis and validate the findings of the proteomics.

Finally, the effects of deletion mutation of two genes identified from the preliminary genomic screening experiments were examined. It was hypothesised that, if these observations and genomic screening approach holds good, the exclusion of *crl*, which was inhibitory to the protein secretion would increase protein secretion, while the *gspL*, which had improved protein secretion, would decrease protein secretion of pSC 1 (Cutinase) *via* the FT3SS, respectively.

It was observed that protein secretion *via* the FT3SS of both Δ*crl* and Δ*gspL* was significantly decreased ($p<0.0001$) compared to the control Δ*CKL* Δ*clpX* strain. The significant decrease of protein secretion in *gspL* mutant was in agreement with the hypothesis, while the decrease in *crl* mutant was not. The GspL protein is a member of an operon of genes (*gspC-O*) which are generally not expressed, and are homologous to those encoding the secreton or type II secretion machinery in *Klebsiella oxytoca* (Pugsley *et al.*, 1997; Nivaskumar *et al.*, 2014). The GspL is an inner membrane protein that interacts with GspC, GspE and GspM proteins in the heterooligomer formation (Pugsley *et al.*, 1997; Py *et al.,* 2001; Sandkvist *et al.,* 2001). The GspL protein was over-expressed during genomic library screening and its co-expression with pSC 1 had significantly increased the protein secretion *via* the FT3SS, which was validated by the exclusion *gspL* gene with mutagenesis, expecting significant decrease in protein secretion. A significant decrease in protein secretion was obtained following the removal of *gspL*, as expected (Fig. 6.8). It clearly suggests that *gspL* gene helps in protein secretion of pSC 1 (Cutinase), *via* the FT3SS. It has been reported that *E. coli* does not secrete endogenous proteins, although it contains the *gsp* genes that are homologous to those encoding other secretons, as these genes are transcriptionally silent under standard laboratory conditions due to transcription silencing by the H-NS protein (Defez and Felice, 1981). However, secretion was also achieved when the de-repressed *gsp* operon were present on a multiple-copy number plasmid (Francetic *et al.*, 2000). This evidence could be enough to explain the upregulation of protein secretion by the co-expression of *gspL* harbouring pNK clone and pSC 1. However, the mechanism of interaction of *gspL* that encodes for T2SS protein and secretion *via* the FT3SS remains enigmatic. Therefore, it is necessary investigate if the secretion was achieved *via*

the type II secretion system in addition to the FT3SS. However, this has a little possibility as Cutinase has type III secretion signal attached.

It could also be predicted that the GspL over-expression may influence the $FlhD_4C_2$ regulator *via* the H-NS protein cascade (Stojiljkovic *et al.,* 1995). In order to investigate the mechanism of improved secretion, *gspL* could be cloned into pNK plasmid or additional mutagenesis may be required to identify the mechanism whether the GspL is necessary for efficient protein secretion, in the parent strain. It can also be predicted that the overexpression of GspL might lead to increased protein secretion in Δ*CKL* Δ*clpX* strain and it could be achieved by over-expressing the *gspL* clone, as mentioned before.

Next, the Crl gene was targeted which encodes for an anti-sigma factor that affects curli fibre formation, during the stationary growth of the cell, and whose overexpression affected FT3SS secretion in Chapter 5 (Olsén *et al.,* 1989). It is an RNA polymerase holoenzyme assembly factor that changes the dynamics between the σ^s and σ^{70} for binding to RNA polymerase, core enzyme towards the $\sigma^{s,}$ leading to more RNA polymerase sigma S (E σ^s), allowing more transcription from promoters within the σ^s regulon (Pratt and Silhavy, 1998; Typas *et al.*, 2007). It also enhances both promoter binding and opening of E σ^s holoenzyme at certain promoters (Dudin *et al.,* 2013). *Crl* is also required in σ^s -dependent expression of the *csgBA* operon, that encodes for the fibronectin-binding protein involved in curli formation, which in turn is required for the repression of the *bgl* operon by σ^s (Schnetz, 2002).

As seen in Chapter 5, when *crl* was overexpressed with pSC 1 in the Δ*CKL* Δ*clpX* strain, it significantly decreased protein secretion by 2-fold (Fig. 5.11-5.13). However, in the Δ*CKL* Δ*clpX* Δ*crl* mutant, the protein secretion was decreased even in absence of *crl*, which is not completely understood. It was hoped that, if the *crl* has the role in decreasing protein secretion by changing the sigma factor dynamics towards curli fibre over the flagella, its exclusion might have improved the protein secretion. However, it could be inferred from this paradox that the *crl* does have a role in decreasing the protein secretion *via* the FT3SS but with yet unknown mechanism, as it has been reported that Crl

activates discrete sets of promoters even though it does not bind to promoter DNA (Omaya *et al.*, 2014; Hao *et al.*, 2016). This paradox may be appearing due to insufficient concentration of the σ^s factor in the growth phase for Crl to exhibit its action or its effect is nullified by the presence of the other sigma factors (Lelong *et al.*, 2007). It may also suggest that in absence of Crl, proteins related to secretion were functionally inactivated but there is no experimental evidence to support these speculations. It could be tested by either checking motility on motility agar plates, as removal of *crl* might render the cell with motility even in the stationary phase, as it would not form the curli fibre and hence the biofilm (Typas *et al.*, 2007; Banta *et al.*, 2013).

The *crl* gene exclusion from the parent strain marginally decreased (20%) protein secretion, instead of improving it as hypothesised. However, examination of the *crl* genetic locus in our *E. coli* MC1000 strain might explain what occurred. It was found that the *crl* gene was already inactivated by the presence of an insertional element, as shown in Fig. 6.9. This might explain why the removal of *crl* did not affect the protein secretion at a very high level- and might mean that the Crl mutation is already advantageous to MC1000 in terms of FT3SS production. In fact, this may be the case as a Crl is known to negatively regulate flagella genes, albeit indirectly (Collado-Vides *et al.*, 2019). However, in order to fully establish the role of *crl* in improving protein secretion it needs to be removed from the strain that harbours functional allele of the *crl* gene.

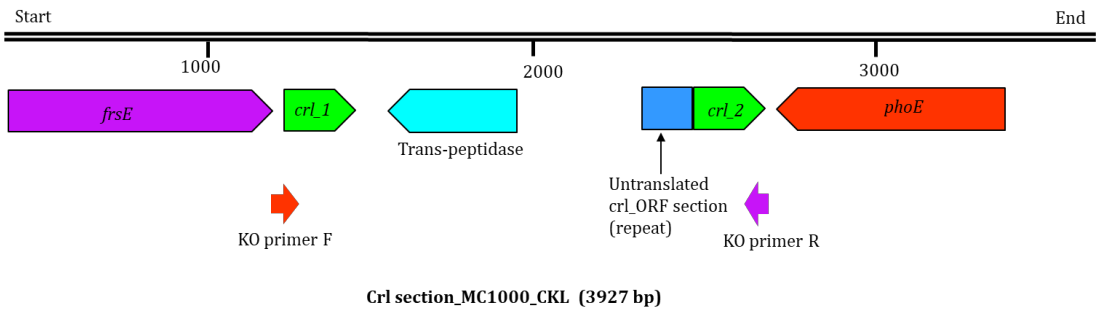


Figure 6.9: Mapping of *crl* gene in the MC1000 Δ*fliC* Δ*flgKL* strain genome. The sequence of pNK clone is shown with partial *frsE* and *phoE* genes in the genome of *Escherichia coli* MC1000 Δ*fliC* Δ*flgKL.* Picture was taken using SnapGene™.

CHAPTER 7

GENERAL DISCUSSION, CONCLUSION AND FUTURE PERSPECTIVES

7 Discussion and Future prospects

7.1 Summary of major findings

The research carried out during this thesis describes the attempts at generating a high efficiency protein secretion-nanomachine by engineering FT3SS of *E. coli* flagella. It was used for the directed secretion of a range of recombinant proteins into the extracellular medium of cell cultures using various approaches. It provided a deeper understanding of the molecular mechanism behind the generation of a high capacity FT3SS for recombinant protein secretion.

7.1.1 Investigation of the use of early, late or RNA-based signals (*fliC* 5'-UTR/ 3'-UTR) for directing secretion *via* the FT3SS

In this chapter, the focus was to identify and confirm, both the secretion signal and *E. coli* strain that would form the basis of a secretion platform. Once this was established, an assay for the detection of protein secretion was optimised. Major findings;

1. The MUB assay for Cutinase secretion was refined for easy quantitative estimation of protein secretion, enabling an effective, accurate and high throughput fluorescent plate reader-based assay.
2. The pJEXpress 404 is an inducible secretion construct containing FT3SS motifs such as 5'UTR, 3'UTR and FliC 47 amino acid secretion signal peptide and characteristic features such as tags to allow detection, isolation and purification of cargo proteins was very effective.
3. Using the MUB and western assay, it was reported that the Hookless (early secretion locked) strain does not allow secretion of either, early (FlgD and FlgE), or late (5'UTR-FliC$_{47aa}$) secretion signals. Only the early (FlgD) secretion signal mediated protein export through the FT3SS of the secretion switched strain.
4. The late secretion signal (5'UTR-FliC$_{47aa}$) and substrate switched HAPless Δ*flgKL* Δ*clpX* strain combination was found to be the most efficient for recombinant protein secretion *via* the FT3SS.

5. Intact late signal, 5'UTR-FliC$_{47aa-}$ (pSC 1) signal was most efficient for directing protein export than the truncated 5'UTR signal, *via* the FT3SS.
6. The 3'UTR was not required for protein export irrespective of the early or late secretion signal, as protein secretion was significantly increased in absence of 3'UTR in secretion construct pJExpress 404-Cutinase.

7.1.2 Mass spectrometry-based proteomics analysis (iTRAQ) of the strains for improving protein secretion *via* the FT3SS

This chapter focused on the investigation of total proteome expression of cells during protein secretion and non-secretion using a mass spectrometry-based iTRAQ analysis, which highlighted the proteins that affected secretion *via* the FT3SS. Major findings;

1. The most efficient secretion signal (5'UTR-FliC$_{47aa}$) expressed in Δ*flgKL* and Δ*flgKL* Δ*clpX* strains, allowed us to investigate the effect of *clpX* gene on the secretion in a strain with *clpX* gene (Δ*CKL*) and without the *clpX* gene (Δ*flgKL* Δ*clpX*).
2. Role of the secretion signal and the corresponding differential protein expression were investigated using pJExpress 404 construct with a secretion signal (5'UTR-FliC$_{47aa}$) and without a secretion signal in the same secretor strain (Δ*flgKL* Δ*clpX*).
3. Following data analyis ClpB and YcgR proteins were investigated for the validation of proteomics analysis with recombineering technique, which significantly improved the protein secretion *via* the FT3SS.

7.1.3 Construction of genomic library for the investigation of novel genes for enhancing protein secretion *via* the FT3SS.

In this chapter, it was attempted to set up and begin a genomic screening method to identify factors affecting FT3SS function in *E. coli* using screens set up in Chapter 3. Major findings;

1. In this chapter, a multi-purpose library plasmid pNK$_{15}$TcLib was constructed, which could be utilised for genome screening in relation to the FT3SS.

2. A genomic library harbouring a random fragment of genomic DNA was constructed.
3. A higher throughput version of the MUB based assay was developed for facile and fast screening.
4. Several potential protein secretion regulating (up and down) clones were identified and analysed (Fig. 5.6). These included a clone containing Crl and a fragment of the GspL operon, which further improved the secretion.

7.1.4 Investigation of the genes highlighted from proteomics analysis, genomic library and the potential small non-coding RNA regulators

Validation of the genes identified from proteomics analysis, genomic library and small non-coding RNA (snRNA) in recombinant protein secretion *via* the FT3SS was carried out with knockout mutagenesis. Major findings;

1. ArcZ and OmrA snRNA's improved the protein secretion following KO mutagenesis as hypothesised.
2. ClpB and YcgR (Chapter 4) marginally improved protein secretion- (Improved protein secretion in ClpB KO mutant was rather a surprise).
3. Crl and GspL (Chapter 5), Crl improved (marginally) while GspL reduced protein export *via* the FT3SS as hypothesised prior to KO mutagenesis.
4. In conclusion, these results and wealth of information obtained from Chapter 3-5, shows that in the near future, the secretion platform could be further engineered to create the potential 'super-secretor strain' for recombinant protein secretion, including those of human origin, in *E. coli*.

7.2 General discussion

7.2.1 Examining the potential use of 5'UTR-*fliC* as an efficient secretion directing signal

In this study, *E. coli* was engineered to secrete recombinant proteins into the extracellular medium in one-step, by modifying FT3SS, which was previously utilised for protein production (Majander *et al.*, 2005; Ve, 2010; Singer *et al.*, 2012). It was aimed to improve through strain engineering and designing of a plasmid-based secretion and purification system. In contrast to previous attempts, the substrate-switched Δ*fliC* Δ*flgKL* strain was engineered as a platform

for secretion with the exclusion of FlgKL, which frees up chaperone FlgN to promote translation coupled secretion of FlgM, further raising FT3SS apparatus level (Karlinsey *et al.*, 2000). The Δ*fliC* Δ*flgKL* strain was further engineered to improve secretion with the exclusion of *clpX* gene which is a component of *clpX*P protease complex that actively degrades $FlhD_4C_2$ complex, thus improves the yield of secreted protein as *clpX* mutant accumulate $FlhD_4C_2$ which in turn expresses Class III genes including *fliC* (Tomoyasu *et al.*, 2002; Guo *et al.*, 2014).

As mentioned, early, late and *fliC*-5'UTR secretion signals were tested for protein secretion in early secretion locked (Δ*fliC* Δ*flgKL* Δ*DE*) and substrate specificity switched isogenic strains (Δ*fliC* Δ*flgKL*). Late signals pSC 1 (5'UTR-$FliC_{47aa}$) were more efficient for protein secretion *via* the FT3SS of the substrate switched Δ*fliC* Δ*flgKL* Δ*clpX* strain. It was also reported that exclusion of *fliC* 3'UTR irrespective of secretion signal improves protein secretion. The *fliC*-3'UTR was incorporated into pJExpress 404, following the complete '5'UTR-$FliC_{47aa}$-Cutinase -Flag-Strep' TAA (stop codon) transcript (Fig. 3.6). The improvement of protein secretion in absence of 3'UTR may be due to mRNA stability and its influence on flagella gene regulation; which, is not clearly understood. Green *et al*., (2018) had reported the yield of recombinant protein secretion for strain Δ*fliC* Δ*flgKL* carrying pSC 1 (5'UTR-$FliC_{47aa}$) was 0.16 mg L^{-1} , which is much lower than yields of 15 mg L^{-1} *via* T1SS (Fernández and Lorenzo, 2001; Specht, 2003; Henderson *et al.*, 2004), 12 mg L^{-1} *via* FT3SS (Δ*fliCD*) (Majander *et al.*, 2005); and 60 mg L^{-1} GFP in the periplasm *via* T2SS (Matos *et al.*, 2012). However, these reports didn't exclude the possibility of cell lysis contributing to the final protein yield. The platform FT3SS system without the requirement for protein extraction was still appealing for direct export of proteins in the extracellular medium, as the possibility of cell lysis was completely negated (Green, 2016), alluding the possibility of achieving continuous protein secretion by stabilising the cell cultures. Secretion capacity of this FT3SS platform could be significantly improved by removing accessory proteins and cellular proteases such as *clpX*P, as illustrated by the significant improvement in protein secretion in Δ*fliC* Δ*flgKL* Δ*clpX* compared to the Δ*fliC* Δ*flgKL* strain. Removal of MotAB from the secretor strain, as motility was not desired during protein production resulted in a

significant increase in secretion (Green, 2016). Additional cellular regulatory targets were identified using a systems biology approach (Chapter 4) and screening a genomic library for secretion improving genes (Chapter 5) and elucidating the role of noncoding RNA on the Flagella master regulator and protein secretion (Chapter 6).

7.2.2 Possibilities for the future improvements in secretion *via* the FT3SS

The main secretion constructs employed for protein secretion in this thesis contains two promoters. The T_5 promoter-based transcript does not seem to be exported *via* the FT3SS while the shorter transcript of native *fliC* promoter may be mainly contributing to the overall secretion, as differential levels of target protein-Cutinase were reported intracellularly and extracellularly. It suggests that maybe the frequency of transcription of longer T_5 transcripts is higher than the shorter, native *fliC* transcripts. It maybe because T_5 promoter engages the transcription machinery rendering limitations on the downstream native FliC promoter for transcription start sites. There is no experimental evidence to support these speculations, however, experiments could be performed as discussed in Section 3.5.3, to test this theory. It might be worth identifying the true nature of these transcripts which could provide necessary information for further improving secretion efficiency. It was also reported that FlgM is transcribed from two different Class II and Class III promoters, however only Class III transcript is secreted while the Class II transcript is retained in the cell (Karlinsey *et al.*, 2000). It was also reported that in absence of FlgN protein, the transcription of (σ^{28})-dependent *fliC* gene was increased by 10-fold in an HBB mutant strain, which can not export FlgM. Therefore, in order to improve protein secretion *via* the FT3SS, the Class II transcript for FlgM, which is an anti-sigma factor (FliA) could be excluded, as it does not allow Class III gene expression. While the Class III FlgM is secreted allowing the substrate specificity switching or the FlgN chaperone (for FlgK and FlgL) could be inactivated as it has been reported that the inactivation of FlgN chaperone significantly reduces the inhibitory activity of the FlgM (Karlinsey *et al.*, 2000) (See Chapter 4, it was significantly downregulated during protein secretion). Additionally, the late secretion signal which contains 5'UTR and N-terminal first 47 amino acids could

be further investigated by truncating the 47 aa or 5'UTR sequence with deletion mutation to further optimise the secretion. It might be worth investigating the 5'UTR with random mutagenesis followed by screening for improved secretion. It is encouraging that attempts have already been made in *Salmonella* protein production by considerably shortening the signal sequence (Végh *et al.*, 2006).

It was also attempted to screen the 5'UTR sequence with random mutagenesis by creating the error-prone clones of 5'UTR. The 5'UTR was amplified with error-prone PCR and all the clones were attached to the $FliC_{47aa}$ signal in pJExpress 404 to improve the secretion efficiency but with a limited success. As 5'UTR does secrete proteins at lower levels in substrate switched strains (Fig. 3.15), the error-prone 5'UTR library may be tested as 5'UTR does not give the extra amino acid stretch following the post-production, which is highly desired from industrial point of view. Therefore, experiment could be performed to identify the optimal 5'UTR signal with this simple, worth pursuing PCR amplification technique. Optimisation of 5'UTR could be performed in order to identify optimal 5'UTR clone from error-prone PCR library for efficient interaction of optimised 5'UTR with the export gate, which can then be attached to the remaining $FliC_{47aa}$ signal. The plasmid-based protein secretion achieved *via* this platform FT3SS was a very attractive and appealing model for protein production, despite the low capacity. It is because the possibility of cell lysis was completely negated, and the protein produced inside the cytoplasm is exported as a linear peptide that gives the functional peptide. The exclusion of *clpXP* and *motAB* has significantly improved the secretion (Green, 2016). The improved rate of protein production and secretion can either be compensated by increasing the number of flagella-basal body units or engineering the FT3SS for more efficient secretion. The transcript produced with late signal in pJExpress 404 contains '5'UTR-$FliC_{47aa}$-TEV-Cutinase -TEV-Fla-Strep-Tag-TAA'. This entire cassette from pJExpress 404 could be excised and integrated into the chromosome of FT3SS engineered strain for protein secretion using mutagenesis to create the desired protein 'super-secretor' strain (Datsenko and Wanner, 2000; Thomason *et al.*, 2007). This is due the fact that plasmid-based production may be not as effective as genome based production and if the cassette is integrated into the high

frequency of transcription promoters such as *fliC* which produce 20-30000 monomers in a matured flagellum (Aldridge and Hughes, 2001), the protein production could be significantly improved over plasmid-based secretion. This rate of increased production could be balanced with export by further engineering strain for more flagella-basal body formation (Schuhmacher *et al.,* 2015). Overall, though this platform FT3SS system gives less yield than anticipated, it is still compelling as it completely eradicates the possibility of contribution of secreted proteins with cell lysis (Green, 2016). This platform system was proposed to extend its application to IB context by investigating its performance for industrially relevant proteins of human origin such as C_H2-IgG antibody fragment and Human growth hormone (hGH), in addition to Cutinase utilised in this thesis. The yields reported for recombinant protein secretion were much lower than hoped; but they were still compelling, given that no leakage in strain or growth defect was observed (Green, 2016).

In order to further improve the efficiency of protein secretion achieved (Chapter 1) and to identify the target regulatory proteins, which influence secretion *via* the FT3SS; complete proteome of the secretor strain (harbouring secretion signal in pJExpress-404: pSC 1) was investigated. It was hoped to identify regulatory proteins, genes and biochemical pathways that changed during secretion and non-secretion to further understand FT3SS. Note here that FT3SS was investigated for the first time with mass spectrometry-based iTRAQ analysis using systems biology approach in order to engineer the host strain for improved protein secretion *via* the FT3SS, since the complete potential of host strain was not fully explored.

The iTRAQ analysis was performed as proteomics makes it possible to generate significant information that can be utilised for the development of metabolic and cellular engineering of the strain (Wu, Zhong and He, 2016). It can be consequently utilised to enhance the yield and productivity of proteins by metabolic engineering of the host (Han and Lee, 2003). Metabolic engineering refers to the directed modification of cellular metabolism by amplification, deletion or substitution of metabolic pathways with molecular biology

techniques such as genetic engineering and with the advancement of new analytical, cloning and biological data analysis tools, proteomics strategies have become more rational (Glick, 1995; Yadav *et al.*, 2012). The iTRAQ analysis reported here allowed us to generate global scale proteome information that can be used to understand the dynamics of metabolic and regulatory network in addition to understanding the physiological changes existed during secretion (Han and Lee, 2003). The operation of a metabolic network in the high secretor strain were the result of coordinated function of enzymes through the reaction pathways and hence proteomics analysis could be utilised to examine the changes in levels of protein (enzyme) expression and consequently to develop a systematic strategy for designing optimal metabolic pathways as it allows profiling of a large number of proteins (Han and Lee, 2003). The wealth of information obtained from these analyses can be utilised to highlight proteins that could be further manipulated in order to achieve high efficiency protein secretion *via* the FT3SS. It was established that the application of proteomics analysis could also be extended to several biological analyses with confidence, as these analyses were performed with independent biological replicates (3x) (Section 4.4, Fig. 4.5). The data obtained from the individual replicates, the relative regulation of proteins in iTRAQ and their mapping on biochemical pathways, supported the credence of intracellular iTRAQ data and also bridged the gap between the understandings of the biochemical regulation in intracellular proteins as hoped in the design phase of the experiment. Information derived from these analyses enhanced confidence of employbility of mass spectrometry-based quantitative techniques for the investigation of FT3SS for improving protein secretion.

The significantly upregulated ClpB and downregulated YcgR proteins were selected for the validation of the applications of mass spectrometry-based approaches for improving secretion *via* the FT3SS, as stated previously. The ClpB protein belongs to the superfamily of AAA^+ (ATPase associated with a variety of cellular activities) proteins, which together with DnaK Chaperone system (DnaK-DnaJ-GrpE) resolubilises the aggregated proteins which was also supported by the upregulation of DnaK and DnaJ (Table 4.2 and 4.3) maybe suggesting the

increased protein transcription and translation rates, that might lead to protein aggregation (Mogk *et al.*, 2003). It suggests that maybe the upregulation in expression of these proteins occurs in order to balance protein expression and secretion by means of resolubilisation of protein aggregates produced during protein overexpression (Shi *et al.*, 1992; Zolkiewski, 1999). While, the flagellar brake protein, YcgR regulates flagellar motility in a c-di-GMP-dependent manner *in vivo* (Ryjenkov *et al.*, 2006). It acts as a molecular brake that interacts with the flagellar switch complex proteins FliG and FliM, however, there is no experimental evidence whether it interacts with MotAB (Boehm *et al.*, 2010; Paul *et al.*, 2010). This upregulated ClpB and downregulated YcgR were therefore further researched to validate the proteomic findings and extend the application of proteomics approaches to future analyses.

In order to validate the findings of these mass spectrometry-based techniques for the improvement of protein secretion *via* the FT3SS and to further extend its application to prokaryotic biology, in addition to ClpB and YcgR, some other proteins need to be studied. The upregulated proteins could be knocked out with recombineering (Chapter 6). These KO mutants when transformed with the secretion construct (pSC 1), should decrease protein secretion, on the contrary, KO mutants of downregulated proteins, when expressed with the secretion construct would increase protein secretion. In another approach, the library plasmid (Chapter 5) could be utilised to overexpress the highlighted protein and co-express along with the secretion construct (pSC 1) to further investigate the role of regulated protein in secretion. Depending upon the data obtained from this analysis, it would be worth investigating highlighted proteins with simultaneous mutagenesis and screening for protein secretion. It might be investigated by creating KO mutagenesis of downregulated proteins from pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* (Table 4.2) and pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* (Table 4.3) and upregulated proteins from pSC 2-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* (Table 4.4) comparison. Additionally, the upregulated proteins from pSC 1-Δ*CKL* Δ*clpX vs* pSC 2-Δ*CKL* Δ*clpX* (Table 4.2) and pSC 1-Δ*CKL* Δ*clpX vs* pSC 1-Δ*CKL* (Table 4.3) comparison could be over-expressed using a library plasmid (Chapter 5). The exclusion of downregulated proteins and overexpression of upregulated

proteins might improve the secretion. It could also be utilised for the characterisation of complex protein mixtures in order to understand the biological system to determine the relationship between proteins, their function and maybe protein-protein interaction (Walker *et al.*, 2010; Zieske *et al*., 2006). For example, it was very appealing to investigate the role of DnaK, DnaJ and GrpE in order to improve the host strain for protein secretion. It might be worth removing the chemotactic proteins on the other hand, as it was not required during protein production. Protein secretion could be further optimised with respect to growth parameters such as medium, as during secretion it was observed that cells were experiencing carbon depletion in the medium.

7.2.3 Genomic screening for further engineering of the FT3SS

In order to employ genomic screening for improving protein secretion *via* the FT3SS, it was a prerequisite to construct a genomic library. Therefore, the pSC 1 (5'UTR-FliC$_{47aa}$-Cutinase -pJExpress-404 (Fig. 3.7)) and the Δ*CKL* Δ*clpX* (Fig. 3.1) secretor platform strains were utilised for the construction and screening of the genomic library, as this combination had shown maximal efficiency of protein secretion *via* the FT3SS (Fig. 4.2 and 4.3) and for which a high throughput Cutinase based assay was developed and refined (Chapter 3).

Host cell engineering has been utilised as a powerful tool for designing microbial platforms for the targeted metabolite production as reported (Floras *et al.*, 1996). The classical method of metabolic engineering with the aim of identifying a rate-determining step in a pathway and alleviating the bottleneck by enzyme over-expression has encouraged much research; which is due to the intervention of limiting steps of counter balancing regulation and often confusing coupled pathways (Bailey *et al.*, 2002). Products like recombinant proteins are intricately coupled to the growth process. Identification of genes that need to be knocked-out or knocked-in remains cryptic to get the desired phenotype (Ghosh *et al.,* 2012). However, a combination of techniques such as knock-ins, knock-outs, promoter and enzyme engineering have been utilised (Ghosh *et al.,* 2012). Just like the strategies utilised to enhance metabolite flux through a pathway such as (i) blocking of the branched chain pathways that lead to by-product

formation, (ii) increase the supply of precursors and flux through the rate limiting steps in the pathway and (iii) removing the feedback controls in the pathway, it is possible to extrapolate the same principle with slight modifications for the designing of Δ*CKL* Δ*clpX* host for the efficient protein secretion *via* the FT3SS (Kucharova *et al.*, 2013). Similar to metabolic synthetic pathways, protein secretion involves multiple steps such as transcription, translation, folding and finally export, but these steps are intricately linked to cellular machinery in which multiple host factors determine the flux through each step of the pathway, unlike the metabolic pathways. Therefore, in order to improve secretion *via* the FT3SS, a slight variant of metabolic engineering was adapted with a novel approach of genomic screening, which involves three strategies (Bailey *et al.*, 2002). Firstly, the identification, construction or calculating a desired phenotype, second, determination of the genetic or environmental condition that confers the desired phenotype and thirdly endowing that phenotype on another strain by directed genetic or environmental manipulations *i.e.* altering phenotype of selected host by genetic manipulations (Ghosh *et al.,* 2012; Pandhal *et al.*, 2013).

7.2.4 Limitation of the genomic library

The genomic library serves as a potential method for the directed large screening of the lab-based gene evolution; however it gives only the preliminary insight into the true nature and functionality of the genes and therefore it needs to be further validated (Bailey *et al.*, 2002; Pandhal *et al.*, 2013). The major limitation of the genomic library constructed was that the plasmid used for genomic library construction contained genomic DNA fragment of 0.8-2.0 Kb, rather than fragments larger than 4.00 Kb. It was also reported that most of the clones contained incomplete genes and without native promoter sequences (Fig. 5.6). Additionally, the shorter fragments gave less probability of the occurrence of the complete gene, promoter and operator sequences in the gene fragment, given the library plasmid did not contain suitable promoter and also insert was delimited with terminators on either end of the insert in MCS (Fig. 5.6). However, during this library construction, the insert DNA fragments of less than 1.0 kb should have been utilised for screening as during library construction, some of the DNA fragments might have been inserted in the reverse orientation

compared to the coding strand in the expression vector (pNK) thereby maybe generating antisense RNA upon induction (Ghosh *et al.,* 2012). These antisense fragments would have hybridised to the sense mRNA or the functional protein of the specific genes leading to may be partial 'gene silencing'. However, the studies on antisense mechanism are limited in prokaryotes than eukaryotes as Dicer and RNAi are absent in prokaryotes (Ghosh *et al.,* 2012).

On the contrary, the larger fragments could have a higher probability of harbouring functional genes possibly with native operator and promoter sequences, however their identification would not have been easy (Yuan *et al.*, 2005). Some genes in an operon execute their function which are usually encoded by the same promoter for the downstream genes (Ermolaeva, 2001). However, the successful implementation of the genomic library for improving protein secretion *via* the FT3SS makes it easy to identify the potential clones depending upon the desired phenotype obtained following genomic screening (Pandhal *et al.*, 2012). It could also be utilised for lab-based evolution of relevant novel genes with desired phenotypes (Ghosh *et al.,* 2012; Pandhal *et al.*, 2012; Ni *et al.*, 2017).

7.2.5 Advantages of the genomic library for the lab-based gene evolution

The genomic library was constructed using pNK plasmid and it was shown that the genomic screening could be utilised for improving protein secretion *via* the FT3SS. Also, its application could be extended to addressing biological questions in prokaryotes by means of coexpressing pNK harbouring random genes with the previously known plasmid-based expression systems (Bailey *et al.*, 2002; Ghosh *et al.,* 2012; Pandhal *et al.*, 2013). The desired phenotype once obtained, the original pNK plasmid could be investigated to establish the true nature of protein encoded by the unknown genomic DNA fragment incorporated in it using bioinformatics and BLAST searches. Protein thus obtained can be further investigated with over-expression or mutagenesis (Datsenko and Wanner, 2000). In conclusion, *E. coli* genes can be collected on a specific plasmid in a genomic library and desired genes can be highlighted by coexpressing with known plasmid-based expression system, as shown by pSC 1 protein secretion *via* the FT3SS. This lab-based gene directed evolution were performed for the

first time with respect to protein secretion *via* the FT3SS. It was demonstrated with change in protein secretion of pSC 1 *via* the FT3SS, suggesting that pNK incorporated genes may encode a novel protein that influences the secretion of pSC 1 (Section 5.6, Fig. 5.12). Thus, it reports the first attempt in the FT3SS for directed evolution of secretion improving genes, by means of this genomic library and screening assay outlined (Section 3.2.2).

7.2.6 Prospects of the genomic library

The genomic library was better suited for the rational directed gene evolution than random mutagenesis approaches since the genes obtained in the library, and further validated with gene knockout or amplification techniques, simultaneously (Yuan *et al.*, 2005; Pandhal *et al.*, 2012). In order to investigate the target in metabolic engineering, metabolic flux analysis and *in silico* reconstructing metabolic model, the genomic library could serve as a valuable cosmopolitan technique (Delgado and Liao, 1997). However, the major disadvantage of this technique is, it gives only preliminary insight into the probability of potentially desired genes, which needs further confirmation with mutagenesis or gene amplification. It is more precise for lab-based gene evolution than commonly utilised random mutagenesis (Cirino *et al.,* 2003; Baneyx and Mujacic, 2004; Ghosh *et al.,* 2012).

In future, the assay needs to be optimised for simple screening maybe by increasing the volume of expression cultures or changing growth parameters (Sivashanmugam *et al.*, 2009; Ukkonen *et al.*, 2011). Also, independent libraries needs to be constructed using random genomic DNA fragments of less than 1.0 kb, in order to achieve the gene silencing, as reported by Ghosh *et al.,* (2012) for improving host for protein production (Ghosh *et al.,* 2012). It also needs to be constructed for larger genomic DNA fragments as the pNK does not have native promoter. It is also desirable to utilise the microscopic techniques e.g. for the investigation of phenotype of pNK $35^{stat.\downarrow}$ for curli fibre formation (Section 5.6).

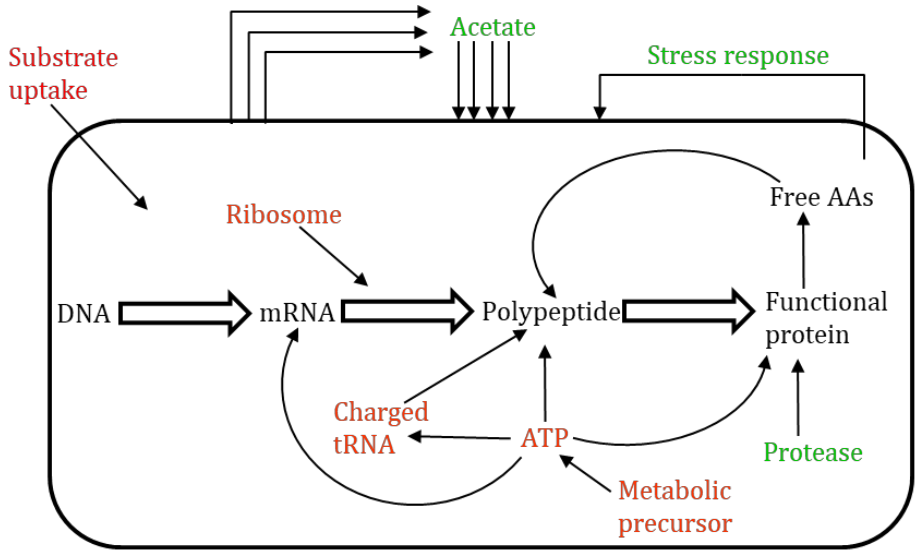


Figure 5.22: Schematic of the cellular stress response of various factors affecting recombinant protein synthesis.
The upregulated pathways are shown in green (Proteases, stress response acetate formation) while downregulated (substrate uptake, ribosome, translation rates, tRNA and ATP) are shown in green. Adapted from (Ghosh *et al.,* 2012).

7.2.7 Rationale of exploiting snRNA for the improvement in protein secretion

As previously stated, ~280 snRNAs have been reported in *Salmonella* with a limited study on their role in bacterial physiology. The significant improvement in protein secretion obtained *via* the modified FT3SS of Δ*CKL* Δ*clpX* Δ*arcZ* mutant (Chapter 6) encourages exploration of their potential role in further improving secretion. The noncoding RNA (snRNA) negatively (*ArcZ* and *OmrA*) and positively (*McaS* and *MicA*) regulate master regulator $FlhD_4C_2$ Complex. It was attempted to overexpress the positive regulatory snRNAs and knockout the negative regulatory snRNAs. The experiment was designed to overexpress snRNA and co-transformed along with pSC 1 (secretion plasmid) as previously used for genomic screening (Chapter 5). It was required to refine the library plasmid (pNK_{15}TcLib) for snRNA expression by incorporating 35 bp Anderson promoter (Bba_J23119: iGEM 2006_Berkeley (2006-08-24)), along with snRNA sequence in the MCS of pNK plasmid (Fig. 5.6) which had terminators on either end of the MCS. However, due to time constraints the construction of new plasmid was not successful. In future, it would be worth investigating the snRNAs such as *McaS* and *MicA* that positively regulate the $FlhD_4C_2$ Complex (Lay and Gottesman, 2012; Jørgensen *et al.*, 2013). It would be interesting to investigate

the role of negative regulatory *OxyS* snRNA as it has been reported to influence the *flhDC* mRNA and the positive regulatory snRNA such as *McaS* and *MicA*, on the secretion *via* FT3SS (Lay and Gottesman, 2012). It would also be interesting to investigate secretion by overexpressing the positive regulatory snRNA in the negative snRNA mutants, such as *ArcZ* and *OmrA* or the additional *OxyS* and *OmrB* mutants and their combined mutant strain. It might be worth perusing the investigation of all the known 280 snRNA by means of creating KO library, as the role of these RNAs remains unknown (Miranda-CasoLuengo *et al.*, 2017). Figure 7.2 below shows additional potential mutagenesis targets that could enable the engineering of the super-secretor strain, as hoped for improving protein export *via* the FT3SS.

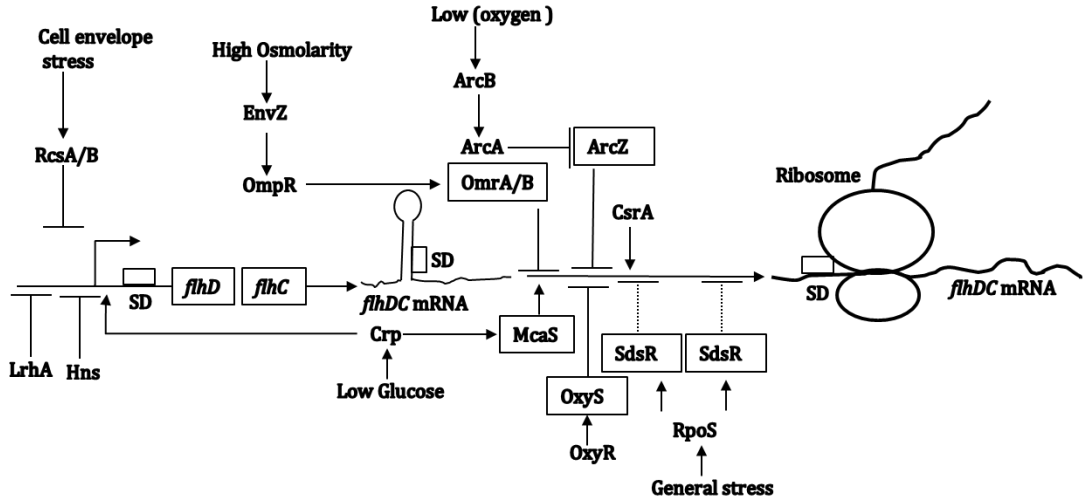


Figure 7.2: Model for the regulation of FlhD and FlhC expression.
The arrows and blocked arrows indicate positive and negative regulation, respectively. Solid lines show direct, while dotted lines indirect regulation, snRNA is shown in the box, SD: Shine-Dalgarno sequence. Adapted from (Lay and Gottesman, 2012).

7.2.8 Possibilities for a hyper secretion strain- combining all data

One upregulated and one downregulated protein were selected from proteomics (Chapter 4) and genomic library (Chapter 5) and two negative regulatory snRNAs of $FlhD_4C_2$ complex were used to create KO mutants as previously reported. A significant difference was observed in protein secretion of the mutants and the parent strain Δ*CKL* Δ*clpX* (Fig. 6.8). The insight obtained from each chapter regarding protein secretion *via* the FT3SS is readily accessible for making the further KO mutagenesis to utilise it for improving secretion. Additionally, the screening of all 280 non-coding RNA (snRNA) screening would

be very appealing and worth perusing to achieve the aim of this work. These mutations could be combined with additional mutagenesis screening with the hope that the mutations would be additive and therefore may improve secretion *via* the FT3SS. The combination of secretion improving snRNA and the genes knockout mutants obtained from either library, proteomics or both might be worth perusing for the engineering of parent Δ*CKL* Δ*clpX* strain. In future, it might be worth utilising quantitative PCR (qPCR) to investigate cryptic gene behavior following mutagenesis, e.g *gspL* gene following mutagenesis, to validate and support the predicted theories. It would also require utilising the motility and biofilm assay in case of *crl* a gene analysis. It is utmost necessary to develop the independent gene KO mutagenesis libraries by utilising the proteomics, library screening and snRNA prior to developing a strategy to combine the potential mutants to engineer 'Super-secretor-strain'.

7.3 Improved secretion capacity of the FT3SS

With an IB goal, FT3SS had been previously modified for the secretion of recombinant proteins into the medium with mixed success, however, the research carried out in this thesis is maybe far more extensive, diverse and cutting-edge, in terms of engineering both the secretion strain and the secretion construct, than any other previously reported work (Majander *et al.*, 2005; Widmaier *et al.*, 2009; Singer *et al.*, 2012; Green, 2016). Most of the above mentioned studies investigated different peptide signals, some of them modified the strain to enable increased protein secretion, which resulted in 2-fold increase (Majander *et al.*, 2005) and over a 10-fold increase in secreted proteins (Green, 2016), however, the value surpassed in this thesis by over 5-fold to that obtained by Green et al. (2016). Additionally, apart from Majander *et al.* (2005) and Green *et al.* (2016) studies, which concerned *E. coli,* however all other investigations involved *Salmonella* and therefore work in this thesis carried out in *E. coli* has greatly furthered the understanding of heterologous protein secretion *via* the FT3SS of an industrially relevant organism (Yoon *et al.,* 2010). The strain-based improvements were considerable in this study to secretion capacity, it would be beneficial to continue to investigate if this could be further improved by utilising wealth of information obtained from proteomics analysis and genomic library

screening. The MUB assay would serve as a great tool, as previously reported for screening of many mutants once derived from these analyses to investigate the secretion capacity of a wide range of mutant strains in *ΔCKL* background. The mutations could be implemented in a direct manner informed by the information derived from proteomics or genomic library or from literature (e.g. snRNA), which could be utilised for the identification of bottlenecks in recombinant protein secretion or FT3SS secretion apparatus. The modification of the secretion construct could allow multiple traits to be selected. Additionally, a library of secretion plasmids could be generated as reported (Chapter 5) to investigate variation in secretion based on modification of genetic elements such as promoters, ribosomal binding site (RBS) and a secretion signals such as 5'UTR which was shown to be important for efficient secretion (Chapter 3). Additionally, as the presence and absence of tags, UTRs and secretion peptides affected both expression and secretion -it would be beneficial to investigate optimum combinations of these to increase the secretion efficiency. It may also involve utilisation of sections of these components in varying combinations or utilising other secretion signals such as FlgM or version of FliC.

In order to investigate the most effective cell culturing methods for improved secretion, a range of media, induction and incubation protocols could be assessed, such as growing cells at higher cell density or growing at lower temperature or growing in bioreactor with monitored aeration and temperature conditions as depletion in energy source was observed in proteomics (Soini *et al.,* 2008; Rosano and Ceccarelli, 2014). It would be beneficial to use a parallel bioreactor system for multiple experiments that could be implemented in tandem to investigate these effects. This would have an added benefit of being more like actual bioreactor of industrial process. The assay could be further investigated in order to optimise expression condition for increased yields of secreted proteins, as the expression and secretion capacity must be equally matched for optimum secretion. In this thesis most of the work was carried out using Cutinase. Use of only Cutinase secretion assay to screen for effective secretion, may not result in an accurate optimisation strategy for every industrially relevant protein. Additionally, Green *et al.,* (2018) reported that not

all proteins were secreted at the same capacity in all strains. This could potentially be overcome by modifying the secretion construct by replacing the Cutinase by another enzymes such as Sialidase, which hydrolyses terminal sialic acid residues on polysaccharide chains and therefore can be linked to a range of substrates (Chokhawala *et al*., 2007). Also, proteomics screening in this thesis was performed using Cutinase, it might be worth testing with other proteins known to be secreted by the FT3SS such as Human growth hormone which has been already cloned into secretion platform (5'UTR-FliC$_{47aa}$) pJExpress 404.

It could also be possible to harness the FT3SS to implement the strain-based improvements to protein secretion *via* selective directed pressure as shown by genomic library in this thesis. The phenomenon, that cells responds to mild toxins by exporting it out of the cell (Saida *et al.*, 2006) could be utilised and fine tuned with protein secretion by inserting a mildly toxic protein into secretion construct, as differential intracellular concentrations of toxin would have a variable effect on cell survival (Pandhal *et al.*, 2012). If the toxin was fine-tuned with secretion signal and expression, this could result in a population of cells which survive based on their ability to secrete the toxin (Doherty *et al.,* 1993; Saida *et al.*, 2006). Also, snRNA has shown great potential for improving protein secretion. It would be appealing to over-express *mcaS* in *arcZ* mutant, which can be predicted to reinforce the secretion by the positive regulatory *mcaS* and negative regulatory *arcZ* (Miranda-CasoLuengo *et al.*, 2017). It might be worth over-expressing positive regulatory while mutating negative regulatory snRNA, as only 280 snRNAs have been reported (Lay and Gottesman, 2012).

It is likely that combinations of these approaches could be incorporated to result in a strain better designed to secrete proteins *via* the FT3SS than the current Δ*CKL* Δ*clpX* Δ*arcZ* strain, provided a wealth of proteomics and genomic library information and all the necessary tools such as optimised MUB and western immunoassay are readily available. Given the time if mutants reported in Chapter 6 such as *crl, clpB, gspL, ycgR, omrA* were combined, this might have further improved the protein secretion than the Δ*CKL* Δ*clpX* strain. Furthermore, formulation of enriched media might help to boost secretion as it was observed

during proteomics analysis that high protein secreting cells were experiencing energy deficiency and in response had upregulated the alternate energy pathways. This could be teamed-up with a high efficiency secretion construct, which in turn would carry a favourable combination of secretion signal peptide residue and the UTR to ensure an efficient secretion *via* the FT3SS.

7.4 The modified FT3SS as a tool

The modified FT3SS was investigated in order to secreting IB relevant proteins (proteins of human origin) out of bacterial cell directly into the medium. During the investigation of this system, the findings of this work which included the role of 5' and 3' UTR on protein secretion, proteomics analysis during protein secretion and non-secretion, the genomic screening and the role of non-coding RNA in protein secretion *via* the FT3SS have provided the novel insight on the characteristics of FT3SS. Following the insight derived from this work, it would be possible to utilise platform to further the understanding of the FT3SS.

The FT3SS was thoroughly investigated in this thesis but the throughput is still low, therefore, a combination of gene knockouts is logical. This could be utilised to investigate a wide range of combinations of FTSS gene knockouts thereby increasing the potential to discover a new features, functions or interactions of proteins in the FT3SS. Furthermore, essentiality of secretion signal and their 5' and 3'UTRs could also be investigated in a high throughput manner. Also, as the investigation of flagella phenotype is related to motility on a motility agar plates, it might provide a more logical representation of FT3SS activity of cells in liquid broth culture (Minamino *et al,.* 2008). Furthermore, the Cutinase secretion assay (MUB) could also be utilised to investigate and quantify proteins secreted *via* the FT3SS without accessory proteins of tags, harboured in secretion construct, as this is desired from industrial point of view and the perspective that it might improve the protein secretion as well. Furthermore, for accurate quantification of protein secretion, it might need to utilise fermenter tank for growing cells that would allow protein purification from supernatant and would mimic industrial process.

7.5 Role of modified FT3SS from industrial biotechnology perspective

The work carried out during this project has demonstrated the suitability of FT3SS platform for use in IB industry. Secretion of several proteins has been achieved *via* this method. Majander *et al.,* achieved 12 mgL^{-1} of protein secretion following high density culture. Green *et al.,*(2019), achieved 23.5 mgL^{-1} of native E2 protein secretion *via* this method, by growing cells at an OD_{600} of 1.5, with further improvement in FT3SS during this work it was managed to improve the efficiency by at least 6-8 fold, as only few knockouts were generated due to time constraint. However, the information obtained has great potential to improve the FT3SS to out compete the current industry yields by many folds. The yields did not match those reported for secreted proteins in the expression systems used in industry (e.g. in *Pichia* up to 15 g/L, rodent collagen fragments) or in *E. coli* following secretion into the periplasm *via* TAT system (60 mgL^{-1} of GFP in a batch culture or 1 gL^{-1} in fed batch, high cell density culture (Matos *et al.*, 2012)). However, due to several benefits of using an extracellular secretion in *E. coli*-based expression system, it makes an appealing model for IB use. In terms of comparison of this system to others in *E. coli* extracellular secretion systems (T1SS) employed in IB use yields 2 mgL^{-1}, however the yields recorded in this work far surpassed the previously reported yields. This demonstrates that FT3SS may be an amenable platform that could be utilised for technological purposes.

Also, the design of secretion construct in secretion platform was such that, it could be adapted to fit with IB industry specification utilised for protein secretion and purification due to tag or protease cleavage sites. It might be a convincing point to the industrial collaborators that continuous protein expression is possible with simultaneous protein secretion, at least for a few generations considering the stability and continuously evolving mutations in *E. coli*. It would be interesting to investigate the efficiency of the FT3SS for the biotechnologically important proteins, such as Human growth hormone (hGH), Insulin, C_H2-IgG antibody, TrxA. Some of these such as hGH and C_H2-IgG have been already cloned in to pJExpress 404 secretion platform and ready to be tested in most efficient Δ*CKL* Δ*clpX* Δ*arcZ* strain while in parallel, continuously further improving secretion platform.

7.6 Conclusion

Here, the amenability of *E. coli* to produce recombinant proteins of industrial importance such as Cutinase was established. A combinatorial approach was utilised where a strain design and a secretion signal in modular secretion plasmid were incorporated for achieving extracellular protein export *via* a truncated *E. coli* FT3SS. This direct export may enable streamlined downstream processing of high value-added products which are free of cellular contaminants. This might lead to the possibility of continuous protein secreting cell cultures which may not require sacrificing the cultures in order to retrieve the protein product. As reported, during the development of this truncated platform, several findings provided significant information about secretion signal peptides and UTRs of FliC monomers. The representative basal level secretion achieved, showed great potential in terms of future improvements in secretion capacity. FT3SS once established, it was thoroughly investigated using three main approaches (1) Systems biology -proteomics analysis performed to understand protein regulation existed, (2) Genomic screening performed to identify secretion improving genes, and (3) Role of small non-coding RNA (snRNA) was investigated; while these findings were validated with recombineering. About 5-8-fold improvement in protein export was obtained in Δ*arcZ* strain than what we had reported previously, eventhough only few genes were tested due to time constraint. As reported, if the knowledge of FT3SS regulation derived from these approaches and a mutation combining strategy is used, it would potentially improve protein export by many-fold and yield would be much greater. Apart from a wealth of information obtained, tools such as a high throughput assay - MUB was refined that enabled quick measurements of protein export output of the truncated FT3SS strain. While the yields were comparable, it could be considerably improved using information (reported in Chapter 4, 5, 6), assays and recombineering techniques reported.

I am very confident that with the integration of the wealth of information and tools reported, the platform FT3SS would become a forerunner for use in biopharmaceuticals production. It might even exceed the yields achieved by any other expression systems currently utilised in biotech industry.

References

Abby, S. S. and Rocha, E. P. C. (2012) The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems, *PLoS Genetics*, 8; e1002983. doi: 10.1371/journal.pgen.1002983.

Abby, S. S. and Rocha, E. P. C. (2017) Identification of protein secretion systems in bacterial genomes using MacSyFinder, *Methods in Molecular Biology*, 1615; 1–21. doi: 10.1007/978-1-4939-7033-9_1.

Abdel, H. A. M., Attwood, M. M. and Guest, J. R. (2001) Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*, *Microbiology* 8; 1-15. doi: 10.1099/00221287-147-6-1483.

Abrusci, P. *et al.* (2013) Architecture of the major component of the type III secretion system export apparatus, *Nature Structural & Molecular Biology*, 20; 99-104. doi: 10.1038/nsmb.2452.

Acharya, A. *et al.* (2005) Production of constitutively acetylated recombinant p53 from yeast and *Escherichia coli* by tethered catalysis, *Protein Expression and Purification*, 41; 417-425. doi: 10.1016/j.pep.2005.01.015.

Adams, D. E., Bliska, J. B. and Cozzarelli, N. R. (1992) Cre-lox recombination in *Escherichia coli* cells mechanistic differences from the in Vitro reaction, *Journal of Molecular Biology*, 226; 661-673. doi: 10.1016/0022-2836(92)90623-R.

Adler, J. (1966) Chemotaxis in Bacteria, *Science*, 153; 708-716. doi: 10.1126/science.153.3737.708

Adler, J. (1969) Chemoreceptors in bacteria, *Science*, 166; 1588-1597. doi: 10.1126/science.166.3913.1588.

Ahmed A.T. K. *et al.* (2016) Expression and characterization of a Cutinase (AnCUT2) from *Aspergillus niger*, *Open Life Sciences*, 11; 29-38. doi: 10.1515/biol-2016-0004.

Aizawa, S. I. (2001) Bacterial flagella and type III secretion systems, *FEMS Microbiology Letters*, 202; 157-164. doi: http://dx.doi.org/10.1016/S0378-1097(01)00301-9.

Ajikumar, P. K. *et al.* (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*, *Science*, 330; 70-4. doi: 10.1126/science.1191652.

Akeda, Y. and Galán, J. E. (2005) Chaperone release and unfolding of substrates in type III secretion, *Nature*, 437; 911-915. doi: 10.1038/nature03992.

Alcock, F. *et al.* (2013) Live cell imaging shows reversible assembly of the TatA component of the twin-arginine protein transport system, *Proceedings of the National Academy of Sciences of the United States of America*, 110; 38 E3650-E3659. doi: 10.1073/pnas.1306738110.

Aldridge, P. *et al.* (2006) Transcriptional and translational control of the *Salmonella fliC* gene, *Journal of Bacteriology*, 188; 4487-4496. doi: 10.1128/JB.00094-06.

Aldridge, P. and Hughes, K. T. (2001) How and when are substrates selected for type III secretion?, *Trends in Microbiology*, 9; 209-214. doi: 10.1016/S0966-842X(01)02014-5.

Aldridge, P. and Hughes, K. T. (2002) Regulation of flagellar assembly, *Current Opinion in*

Microbiology, 5; 160-165. doi: 10.1016/S1369-5274(02)00302-8.

Alimi, J. P. *et al.* (2000) Reverse transcriptase-polymerase chain reaction validation of 25 orphan genes from *Escherichia coli* K-12 MG1655, *Genome Research*, 10; 959-966. doi: 10.1101/gr.10.7.959.

Almiron, M. *et al.* (1992) A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*, *Genes and Development*, 6; 2646-2654. doi: 10.1101/gad.6.12b.2646.

Amsterdam, V. U. *et al.* (2014) A structurally informed autotransporter platform for efficient heterologous protein secretion and display, *Microbial Cell Factories,* 18; 1-11. doi: 10.1186/1475-2859-11-85.

Andersen, C. *et al.* (2002) Transition to the open state of the TolC periplasmic tunnel entrance, *Proceedings of the National Academy of Sciences of the United States of America*, 99; 11103-11108. doi: 10.1073/pnas.162039399.

Anderson, D. M., Ramamurthi, K. S. and Tam, C. (2002) YopD and LcrH Regulate Expression of *Yersinia enterocolitica* YopQ by a Posttranscriptional Mechanism and Bind to *yopQ* RNA', *Journal of Bacteriology,* 184; 1287-1295. doi: 10.1128/JB.184.5.1287.

Argaman, L. *et al.* (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*, *Current Biology*, 11; 941-950. doi: 10.1016/S0960-9822(01)00270-6.

Arnosti, D. N. and Chamberlin, M. J. (1989) Secondary sigma factor controls transcription of flagellar and chemotaxis genes in *Escherichia coli*, *Proceedings of the National Academy of Sciences of the United States of America*, 3; 830-4. doi: 10.1073/pnas.86.3.830.

Arnqvist, A. *et al.* (1992) The Crl protein activates cryptic genes for curli formation and fibronectin binding in *Escherichia coli* HB101, *Molecular Microbiology*, 17; 2443-52. doi: 10.1111/j.1365-2958.1992.tb01420.x.

Ates, L. S., Houben, E. N. G. and Bitter, W. (2016) Type VII Secretion: A Highly Versatile Secretion System, *Virulence Mechanisms of Bacterial Pathogens,* 5; 357-384. doi: 10.1128/microbiolspec.VMBF-0011-2015.

Auvray, F. *et al.* (2001) Flagellin polymerisation control by a cytosolic export chaperone, *Journal of Molecular Biology*, 308; 221-229. doi: 10.1006/jmbi.2001.4597.

Baba, T. *et al.* (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, *Molecular Systems Biology*, 2; 2006.0008. doi: 10.1038/msb4100050.

Baeshen, M. N. *et al.* (2015) Production of biopharmaceuticals in *E. Coli*: Current scenario and future perspectives, *Journal of Microbiology and Biotechnology*, 25; 953-962. doi: 10.4014/jmb.1412.12079.

Baig, F. *et al.* (2014) Dynamic transcriptional response of *Escherichia coli* to inclusion body formation, *Biotechnology and Bioengineering*, 111; 980-99. doi: 10.1002/bit.25169.

Bailey, J. E. *et al.* (2002) Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes, *Biotechnology and Bioengineering*, 79; 568-79. doi: 10.1002/bit.10441.

Baker, J. L., Çelik, E. and DeLisa, M. P. (2013) Expanding the glycoengineering toolbox:

The rise of bacterial N-linked protein glycosylation, *Trends in Biotechnology*, 31; 313-323. doi: 10.1016/j.tibtech.2013.03.003.

Baker, T. A. and Sauer, R. T. (2012) ClpXP, an ATP-powered unfolding and protein-degradation machine, *Biochimica et Biophysica Acta - Molecular Cell Research*, 1823; 15-28. doi: 10.1016/j.bbamcr.2011.06.007.

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

Baneyx, F. and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli*, *Nature Biotechnology*, 22; 1399-408. doi: 10.1038/nbt1029.

Bange, G. *et al.* (2010) FlhA provides the adaptor for coordinated delivery of late flagella building blocks to the type III secretion system, *Proceedings of the National Academy of Sciences of United States of America*, 107; 11295-11300. doi: 10.1073/pns.1001383107.

Banta, A. B. *et al.* (2013) Key features of S required for specific recognition by Crl, a transcription factor promoting assembly of RNA polymerase holoenzyme, *Proceedings of the National Academy of Sciences of United States of America*, 110; 15956-60, doi: 10.1073/pnas.1311642110.

Baquero, M. R. *et al.* (1995) *sbmC*, a stationary-phase induced SOS *Escherichia coli* gene, whose product protects cells from the DNA replication inhibitor microcin B17, *Molecular Microbiology*, 18; 301-311. doi: 10.1111/j.1365-2958.1995.mmi_18020301.x.

Barembruch, C. and Hengge, R. (2007) Cellular levels and activity of the flagellar sigma factor FliA of *Escherichia coli* are controlled by FlgM-modulated proteolysis, *Molecular Microbiology*, 65; 76-89. doi: 10.1111/j.1365-2958.2007.05770.x.

Barnett, M. E., Zolkiewska, A. and Zolkiewski, M. (2000) Structure and activity of ClpB from *Escherichia coli*: Role of the amino- and carboxyl-terminal domains, *Journal of Biological Chemistry*, 275; 37565-71. doi: 10.1074/jbc.M005211200.

Bartlett, D. H., Frantz, B. B. and Matsumura, P. (1988) Flagellar transcriptional activators FlbB and FlaI: gene sequences and 5' consensus sequences of operons under FlbB and FlaI control.', *Journal of Bacteriology*, 170; 1575-1581. doi: 10.1128/jb.170.4.1575-1581.1988.

Basler, M., Ho, B. T. and Mekalanos, J. J. (2013) Tit-for-tat: type VI secretion system counterattack during bacterial cell-cell interactions, *Cell*, 152; 884-94. doi: 10.1016/j.cell.2013.01.042.

Beckwith, J. (2013) The Sec-dependent pathway, *Research in Microbiology*, 164; 497-504. doi: 10.1016/j.2013.03.007.

Belasco, J. G. (2010) All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay, *Nature Reviews-Molecular Cell Biology* 11; 467-478. doi: 10.1038/nrm2917.

Bendtsen, J. D. *et al.* (2004) Improved prediction of signal peptides: SignalP 3.0, *Journal of Molecular Biology*, 340; 783-795. doi: 10.1016/j.jmb.2004.05.028.

Berg, H. C. (2008) Bacterial flagellar motor, *Current Biology*, 18; 103-132. doi: 10.1016/j.cub.2008.07.015.

Berlec, A. and Štrukelj, B. (2013) Current state and recent advances in biopharmaceutical production in *Escherichia coli*, yeasts and mammalian cells, *Journal of Industrial Microbiology and Biotechnology*, 40; 257-274. doi: 10.1007/s10295-013-1235-0.

Berry, R. M. (2001) Bacterial Flagella : Flagellar Motor, *Life Sciences*, 41; 1-8. doi: 10.1895/wormbook.1.146.1.

Bhowmick, S. and Tripathy, S. (2014) A Tale of effectors; their secretory mechanisms and computational discovery in pathogenic, non-pathogenic and commensal microbes, *Molecular Biology*, 03; 1-14. doi: 10.4172/2168-9547.1000118.

Bieker, K. L. and Silhavy, T. J. (1990) The genetics of protein secretion in *E. coli*, *Trends in Genetics*, 6; 329-334. doi: 10.1016/0168-9525(90)90254-4.

Blattner, F. R. *et al.* (1997) The Complete Genome Sequence of *Escherichia coli* K-12, *Science,* 277; 1453-62. doi: 10.1126/science.277.5331.1453.

Blight, M. A., Chervaux, C. and Holland, I. B. (1994) Protein secretion pathways in *Escherichia coli*, *Current Opinion in Biotechnology*, 5; 468-474. doi: 10.1016/0958-1669(94)90059-0.

Boehm, A. *et al.* (2010) Second Messenger-Mediated Adjustment of Bacterial Swimming Velocity, *Cell*, 141; 107-16. doi: 10.1016/j.cell.2010.01.018.

Bollinger, B. L. and Stover, J. (1999) The Economic Impact of AIDS, *World Journal of Aids*, 4; 1-12. doi: 10.4236/wja.2014.41013.

Bongaerts, J. *et al.* (2001) Metabolic engineering for microbial production of aromatic amino acids and derived compounds, *Metabolic Engineering*, 3; 289-300. doi: 10.1006/mben.2001.0196.

Boyer, P. D. (1993) The binding change mechanism for ATP synthase-Some probabilities and possibilities, *Bioenergetics*, 1140; 215-50. doi: 10.1016/0005-2728(93)90063-L.

Branston, S. D. *et al.* (2012) Investigation of the impact of Tat export pathway enhancement on *E. coli* culture, protein production and early stage recovery, *Biotechnology and Bioengineering*, 109; 983-991. doi: 10.1002/bit.24384.

Brockmeier, U. *et al.* (2006) Systematic Screening of All Signal Peptides from *Bacillus subtilis* : A Powerful Strategy in Optimizing Heterologous Protein Secretion in Gram-positive Bacteria, *Journal of Molecular Biology,* 362; 393-402. doi: 10.1016/j.jmb.2006.07.034.

Brooks, S. A. (2004) Appropriate glycosylation of recombinant proteins for human use: Implications of choice of expression system, *Applied Biochemistry and Biotechnology,* 28; 241-256. doi: 10.1385/MB:28:3:241.

Brown, J. D. *et al.* (2008) The rate of protein secretion dictates the temporal dynamics of flagellar gene expression, *Molecular Microbiology,* 70; 924-937. doi: 10.1111/j.1365-2958.2008.06455.x.

Burdette, L. A. *et al.* (2018) Developing Gram-negative bacteria for the secretion of heterologous proteins, *Microbial Cell Factories*, 196; 1-16. doi: 10.1186/s12934-018-1041-5.

Bustin, S. A. (2000) Absolute quantification of mRNA using real-time reverse

transcription polymerase chain reaction assays, *Journal of Molecular Endocrinology*, 25; 169-93. doi: 10.1677/jme.0.0250169.

Camberg, J. L., Hoskins, J. R. and Wickner, S. (2009) ClpXP protease degrades the cytoskeletal protein, FtsZ and modulates FtsZ polymer dynamics, *Proceedings of National Academy of Science, USA,* 106; 10614-10619. doi: 10.1073/pnas.0904886106.

Carvalho, C. M. L., Aires-Barros, M. R. and Cabral, J. M. S. (1998) Cutinase structure, function and biocatalytic applications, *Electronic Journal of Biotechnology*, 1; 91-113. doi: 10.2225/vol1-issue3-fulltext-8.

Casali, N. (2003) *Escherichia coli* host strains, *Methods in Molecular Miology* 235; 27-48. doi: 10.1385/1-59259-409-3:27.

Cascales, E. (2008) The type VI secretion toolkit, *EMBO Reports*, 9; 735-41. doi: 10.1038/embor.2008.131.

Cascales, E. and Christie, P. J. (2004) *Agrobacterium* VirB10, an ATP energy sensor required for type IV secretion, *Proceedings of National Academy of Science, USA,* 101; 17228-33. doi: 10.1073/pnas.0405843101.

Chadsey, M. S., Karlinsey, J. E. and Hughes, K. T. (1998) The flagellar anti-sigma factor FlgM actively dissociates *Salmonella typhimurium* sigma28 RNA polymerase holoenzyme, *Genes & Development*, 12; 3123-3136. doi: 10.1101/gad.12.19.3123.

Chambers, G. *et al.* (2015) Biotech Britain, *Capital Economics*, 2484735; 1-43.

Chan, W. *et al.* (2013) A comparison and optimization of methods and factors affecting the transformation of *Escherichia coli*, *Bioscience Reports*, 33; 931-937. doi: 10.1042/BSR20130098.

Chatzi, K. E. *et al.* (2013) Breaking on through to the other side : protein export through the bacterial, *Biochemical Journal*, 37; 25-37. doi: 10.1042/BJ20121227.

Chatzi, K. E. *et al.* (2014) SecA-mediated targeting and translocation of secretory proteins, *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843; 1466-1474. doi: 10.1016/j.bbamcr.2014.02.014.

Chen, M. *et al.* (2012) Generation of diploid *Pichia pastoris* strains by mating and their application for recombinant protein production, *Microbial Cell Factories,* 1; 1-18. doi: 10.1186/1475-2859-11-91.

Chevance, F. F. and Hughes, K. T. (2008) Coordinating assembly of a bacterial macromolecular machine, *Nature Review in Microbiology*, 6; 455-465. doi: nrmicro1887 [pii]\r10.1038/nrmicro1887.

Chilcott, G. S. and Hughes, K. T. (1998) The type III secretion determinants of the flagellar anti-transcription factor, FlgM, extend from the amino-terminus into the anti-sigma28 domain, *Molecular Microbiology*, 30; 1029-1040. doi: 10.1046/j.1365-2958.1998.01131.x.

Chilcott, G. S. and Hughes, K. T. (2000) Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar typhimurium and *Escherichia coli*, *Microbiology and Molecular Biology Reviews*, 64; 694-708. doi: 10.1128/MMBR.64.4.694-708.2000.

Chiverton, L. M. *et al.* (2016) Quantitative definition and monitoring of the host cell

protein proteome using iTRAQ- a study of an industrial mAb producing CHO-S cell line, *Biotechnology Journal*, 11; 1014-1024. doi: 10.1002/biot.201500550.

Choi, J. H. and Lee, S. Y. (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*, *Applied Microbiology Biotechnology*, 64; 625-635. doi: 10.1007/s00253-004-1559-9.

Chokhawala, H. A., Yu, H. and Chen, X. (2007) High-throughput substrate specificity studies of sialidases by using chemoenzymatically synthesized sialoside libraries, *European Journal of Chemical Biology,*, 8; 194-201. doi: 10.1002/cbic.200600410.

Chuang, S. E. and Blattner, F. R. (1993) Characterization of twenty-six new heat shock genes of *Escherichia coli*, *Journal of Bacteriology*, 175; 5242-52. doi: 10.1128/jb.175.16.5242-5252.1993.

Chung, C. H. and Goldberg, A. L. (1981) The product of the lon (*capR*) gene in *Escherichia coli* is the ATP-dependent protease, protease La, *Proceedings of the National Academy of Sciences of the United States of America*, 78; 4931-5. doi: 10.1073/pnas.78.8.4931.

Chung, C. W. *et al.* (2009) Export of recombinant proteins in *Escherichia coli* using ABC transporter with an attached lipase ABC transporter recognition domain (LARD), *Microbial Cell Factories*, 12; 1-12. doi: 10.1186/1475-2859-8-11.

Cianfanelli, F. R., Monlezun, L. and Coulthurst, S. J. (2016) Aim , Load , Fire : The Type VI Secretion System: a Bacterial Nanoweapon, *Trends in Microbiology*, 24; 51-62. doi: 10.1016/j.tim.2015.10.005.

Cirino, P. C., Mayer, K. M. and Umeno, D. (2003) Generating mutant libraries using error-prone PCR, *Methods in Molecular Biology,* 231; 3-9. doi: 10.1385/1-59259-395-X:3.

Claret, L. and Hughes, C. (2000) Rapid turnover of FlhD and FlhC, the flagellar regulon transcriptional activator proteins, during Proteus swarming, *Journal of Bacteriology*, 182; 833-836. doi: 10.1128/JB.182.3.833-836.2000.

Clark, A. J. (1998) The mammary gland as a bioreactor: expression, processing, and production of recombinant proteins, *Journal of Mammary Gland Biology and Neoplasia*, 3; 337-50. doi: 10.1023/a:1018723712996.

Clemmer, K. M. and Rather, P. N. (2008) The Lon protease regulates swarming motility and virulence gene expression in *Proteus mirabilis*, *Journal of Medical Microbiology*, 57; 931-937. doi: 10.1099/jmm.0.47778-0.

Cong, Y. *et al.* (2012) Site-specific PEGylation at histidine tags, *Bioconjugate Chemistry*, 23, 248-263. doi: 10.1021/bc200530x.

Cornelis, G. R. (2006) The type III secretion injectisome, *Nature Reviews- Microbiology*, 4; 811-825. doi: 10.1038/nrmicro1526.

Costa, T. R. D. *et al.* (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights, *Nature Reviews Microbiology*, 13; 343-59. doi: 10.1038/nrmicro3456.

Coulthurst, S. J. (2013) The Type VI secretion system e-a widespread and versatile cell targeting system, *Research in Microbiology*, 164; 640-654. doi: 10.1016/j.resmic.2013.03.017.

Courchesne, D. *et al.* (2017) Scalable Production of Genetically Engineered Nano fibrous Macroscopic Materials via Filtration, *ACS Biomaterial Science and Engineering,* 5; 733-741. doi: 10.1021/acsbiomaterials.6b00437.

Cozzani, I. *et al.* (1980) Separation and characterization of NAD- and NADP-specific succinate-semialdehyde dehydrogenase from *Escherichia coli* K-12 3300, *BBA-Enzymology*, 613; 309-317. doi: 10.1016/0005-2744(80)90085-6.

Crago, A. M. and Koronakis, V. (1998) *Salmonella* InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization, *Molecular Microbiology* , 30; 47-56. doi: 10.1046/j.1365-1958.1988.01036.x.

Cui, Y. Y. *et al.* (2014) Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering, *Microbial Cell Factories*, 21; 1-11. doi: 10.1186/1475-2859-13-21.

Daley, D. O. (2005) Global Topology Analysis of the *Escherichia coli* Inner Membrane Proteome, *Science*, 308; 1321-1323. doi: 10.1126/science.1109730.

Darbari, V. C. and Waksman, G. (2002) Structural Biology of Bacterial Type IV Secretion Systems, *Annual Review of Biochemistry,* 84; 603-29. doi: 10.1146/annualrev-biochem-062911-102821.

Datsenko, K. A. and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proceedings of the National Academy of Sciences of the United States of America*, 97; 6640-5. doi: 10.1073/pnas.120163297.

Defez, R. and De Felice, M. (1981) Cryptic operon for beta-glucoside metabolism in *Escherichia coli* K12: genetic evidence for a regulatory protein, *Genetics,* 97; 11-25. doi:PMC1214377.

DeFrees, S. (2006) GlycoPEGylation of recombinant therapeutic proteins produced in *Escherichia coli*, *Glycobiology*, 16; 833-843. doi: 10.1093/glycob/cwl004.

Delalez, N. J. *et al.* (2010) Signal-dependent turnover of the bacterial flagellar switch protein FliM, *Proceedings of the National Academy of Sciences of the United States of America*, 107; 11347-11351. doi: 10.1073/pnas.1000284107.

Delepelaire, P. (2004) Type I secretion in gram-negative bacteria, *Biochimica et Biophysica Acta Molecular Cell Research*, 1694; 149-161. doi: 10.1016/j.bbamcr.2004.05.001.

Delgado, J. and Liao, J. C. (1997) Inverse flux analysis for reduction of acetate excretion in *Escherichia coli*, *Biotechnology Progress*, 13; 361-7. doi: 10.1021/bp970047x.

Denks, K. *et al.* (2014) The Sec translocon mediated protein transport in prokaryotes and eukaryotes, *Molecular Membrane Biology*, 7688; 58-84. doi: 10.3109/09687688.2014.907455.

Desvaux, M. *et al.* (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue, *Trends in Microbiology*, 17; 139-145. doi: 10.1016/j.tim.2009.01.004.

Desvaux, M., Parham, N. J. and Henderson, I. R. (2004) Type V protein secretion: simplicity gone awry, *Current Issues in Molecular Biology*, 6; 111-124. doi: 10.1128?MMBR.68.4.692-744.2004.

Diepold, A. and Wagner, S. (2014) Assembly of the bacterial type III secretion machinery, *FEMS Microbiology Reviews*, 38; 802-822. doi: 10.1111/1574-6976.12061.

Dietsche, T. *et al.* (2016) Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus, *PLoS Pathogens*, 12; 1-25. doi: 10.1371/journal.ppat.1006071.

Dietsche, T. *et al.* (2017) Assembly and stoichiometry of the core structure of the bacterial flagellar type III export gate complex, *PLOS Biology,* 5; 1-15. doi: 10.1371/journal.pbio.200228110.1371

Dijl, J. M. *et al.* (2002) Functional genomic analysis of the *Bacillus subtilis* Tat pathway for protein secretion, *Journal of Biotechnology*, 98; 243-254. doi: 10.1016/S0168-1656(02)00135-9.

Doerrler, W. T. *et al.* (2013) New functions for the ancient DedA membrane protein family, *Journal of Bacteriology*, 195; 3-11. doi: 10.1128/JB.01006-12.

Doherty, A. J., Connolly, B. A. and Worrall, A. F. (1993) Overproduction of the toxic protein, bovine pancreatic DNaseI, in *Escherichia coli* using a tightly controlled T7-promoter-based vector, *Gene*, 136; 337-40. doi: 10.1016/0378-1119(93)90491-K.

Donath, M. J., Dominguez, M. A. and Withers, S. T. (2011) Development of an automated platform for high-throughput P1-phage transduction of *Escherichia coli*, *Journal of Laboratory Automation*, 16, 141-147. doi: 10.1016/j.jala.2010.08.005.

Dover, S. and Halpern, Y. S. (1972) Utilization of -aminobutyric acid as the sole carbon and nitrogen source by *Escherichia coli* K-12 mutants, *Journal of Bacteriology*, 109, 835-43. doi: 10.1111/zph.12473.

Dudin, O., Lacour, S. and Geiselmann, J. (2013) Expression dynamics of RpoS/Crl-dependent genes in *Escherichia coli*, *Research in Microbiology*, 164; 838-847. doi: 10.1016/j.resmic.2013.07.002.

Dudin, O. *et al.* (2014) Repression of flagellar genes in exponential phase by CsgD and CpxR, two crucial modulators of *Escherichia coli* biofilm formation, *Journal of Bacteriology*, 196; 707-715. doi: 10.1128/JB.00938-13.

Dukan, S. and Touati, D. (1996) Hypochlorous acid stress in *Escherichia coli:* Resistance, DNA damage, and comparison with hydrogen peroxide stress, *Journal of Bacteriology*, 178; 6145-50. doi: 10.1128/jb.178.21.6145-6150.1996.

Dunn, S. D. *et al.* (2000) The b subunit of *Escherichia coli* ATP synthase, *Journal of Bioenergetics and Biomembranes*, 1857; 129-140. doi: 10.1023/A:1005571818730.

Endy, D. (2005) Foundations for engineering biology, *Nature,* 438; 449-453. doi: 10.1038/nature04342.

Erhardt, M. and Hughes, K. T. (2010) C-ring requirement in flagellar type III secretion is bypassed by FlhDC upregulation, *Molecular Microbiology*, 75; 376-393. doi: 10.1111/j.1365-2958.2009.06973.x.

Erhardt, M., Namba, K. and Hughes, K. T. (2010) Bacterial nanomachines: the flagellum and type III injectisome, *Cold Spring Harbor Perspectives in Biology*, 2; 1-23. doi: 10.1101/cshperspect.a000299.

Eriksson, M. J. and Clarke, A. K. (2000) The *Escherichia coli* heat shock protein ClpB restores acquired thermotolerance to a cyanobacterial ClpB deletion mutant, *Cell Stress and Chaperones*, 5; 255-264. doi: 10.1379/1466-1268(2000)005<0255:TECHSP>2.0.CO;2.

Ermolaeva, M. D. (2001) Prediction of operons in microbial genomes, *Nucleic Acids Research*, 29; 1216-1221. doi: 10.1093/nar/29.5.1216.

Eser, M. *et al.* (2009) Disulfide bond formation by exported glutaredoxin indicates glutathione's presence in the *E. coli* periplasm, *Proceedings of the National Academy of Sciences of the United States of America*, 106; 1572-1577. doi: 10.1073/pnas.0812596106.

Evans, L. D. B. *et al.* (2013) A chain mechanism for flagellum growth, *Nature*, 504; 287-290. doi: 10.1038/nature12682.

Fabiani, F. D. *et al.* (2017) A flagellum-specific chaperone facilitates assembly of the core type III export apparatus of the bacterial flagellum, *PLOS Biology*, 15; 1-22. doi: 10.1371/journal.pbio.2002267.

Fahrner, K. A. and Berg, H. C. (2015) Mutations That Stimulate *flhDC* Expression in *Escherichia coli* K-12, *Journal of Bacteriology,* 197; 3087-96. doi: 10.1128/JB.00455-15.

Fang, X. and Gomelsky, M. (2010) A post-translational, c-di-GMP-dependent mechanism regulating flagellar motility, *Molecular Microbiology*, 76; 1295-305. doi: 10.1111/j.1365-2958.2010.07179.x.

Farewell, A., Kvint, K. and Nyström, T. (1998) Negative regulation by RpoS: A case of sigma factor competition, *Molecular Microbiology*, 29; 1039-51. doi: 10.1046/j.1365-2958.1998.00990.x.

Farmer, W. R. and Liao, J. C. (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control, *Nature Biotechnology*, 18; 533-7. doi: 10.1038/75398.

Feldman, M. F. and Cornelis, G. R. (2003) The multitalented type III chaperones: All you can do with 15 kDa, *FEMS Microbiology Letters*, 219; 151-158. doi: 10.1016/S0378-1097(03)00042-9.

Ferna, L. A. (2004) Prokaryotic expression of antibodies and affibodies, *Current Opinions in Biotechnology*, 15; 364-373. doi: 10.1016/j.copbio.2004.06.004.

Fernández, L. A. and De Lorenzo, V. (2001) Formation of disulphide bonds during secretion of proteins through the periplasmic-independent type I pathway, *Molecular Microbiology*, 40; 332-346. doi: 10.1046/j.1365-2958.2001.02410.x.

Ferrer-Miralles, N. and Villaverde, A. (2013) Bacterial cell factories for recombinant protein production; expanding the catalogue, *Microbial Cell factories*, 12; 1-4. doi: 10.1186/1475-2859-12-113.

Fidan, O. and Zhan, J. (2015) Recent advances in engineering yeast for pharmaceutical protein production, *RSC Advances*, 5; 86665-86674. doi: 10.1039/c5ra13003d.

Filloux, A. (2004) The underlying mechanisms of type II protein secretion, *Biochimica et Biophysica Acta*, 1694; 163-179. doi: 10.1016/j.bbamcr.2004.05.003.

Fisher, A. C. *et al.* (2011) Production of secretory and extracellular N-linked

glycoproteins in *Escherichia coli*, *Applied and Environmental Microbiology*, 77; 871-81. doi: 10.1128/AEM.01901-10.

Fitzgerald, D. M., Bonocora, R. P. and Wade, J. T. (2014) Comprehensive Mapping of the *Escherichia coli* Flagellar Regulatory Network, *PLoS Genetics*, 10; 1-24. doi: 10.1371/journal.pgen.1004649.

Floras, N. *et al.* (1996) Pathway engineering for the production of aromatic compounds in *Escherichia coli*, *Nature Biotechnology*, 14; 620-3. doi: 10.1038/nbt0596-620.

Flores, S. *et al.* (2004) Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway, *Biotechnology and Bioengineering*, 87; 485-94. doi: 10.1002/bit.20137.

Fojan, P. *et al.* (2000) What distinguishes an esterase from a lipase: A novel structural approach, *Biochimie*, 82; 1033-1041. doi: 10.1016/S0300-9084(00)01188-3.

Francetic, O. *et al.* (2000) Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to Chitinase secretion, *The EMBO Journal*, 19; 6697-703. doi: 10.1093/emboj/19.24.6697.

Francetic, O. and Pugsley, A. P. (1996) The cryptic general secretory pathway (gsp) operon of *Escherichia coli* K-12 encodes functional proteins, *Journal of Bacteriology*, 178; 3544-9. doi: 10.1128/jb.178.12.3544-3549.1996.

Fraser, G. M. *et al.* (2003) Substrate specificity of type III flagellar protein export in *Salmonella* is controlled by subdomain interactions in FlhB, *Molecular Microbiology*, 48; 1043-1057. doi: 10.1046/j.1365-2958.2003.03487.x.

Fraser, G. M. and Hughes, C. (1999) Swarming motility, *Current Opinion in Microbiology*, 2; 630-635. doi: 10.1016/S1369-5274(99)00033-8.

Frenzel, A., Hust, M. and Schirrmann, T. (2013) Expression of recombinant antibodies, *Frontiers in Immunology*, 4, 1-20. doi: 10.3389/fimmu.2013.00217.

Freudl, R. (2013) Leaving home ain't easy: Protein export systems in Gram-positive bacteria, *Research in Microbiology*, 164; 664-674. doi: 10.1016/j.resmic.2013.03.014.

Freudl, R. (2018) Signal peptides for recombinant protein secretion in bacterial expression systems, *Microbial Cell Factories*, 17; 1-10. doi: 10.1186/s12934-018-0901-3.

Fröhlich, K. S. *et al.* (2012) A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD, *Nucleic Acids Research*, 40; 3623-40. doi: 10.1093/nar/gkr1156.

Fronzes, R. *et al.* (2009) Structure of a type IV secretion system core complex, *Science*, 323; 266–268. doi: 10.1126/science.1166101.

Fronzes, R., Christie, P. J. and Waksman, G. (2009) The structural biology of type IV secretion systems, *Nature Reviews-Microbiology*, 7; 703-714. doi: 10.1038/nrmicro2218.

Gafken, P. R. and Lampe, P. D. (2006) Methodologies for characterizing phosphoproteins by mass spectrometry, *Cell Communication and Adhesion*, 13; 249-62. doi: 10.1080/15419060601077917.

Galán, J. E. *et al.* (2014) Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells, *Annual Review of Microbiology*, 68; 415-38. doi: 10.1146/annurev-micro-092412-155725.

Galan, J. E. and Wolf-Watz, H. (2006) Protein delivery into eukaryotic cells by type III secretion machines, *Nature*, 444; 567-573. doi: 10.1038/nature05272.

Georgiou, G. and Segatori, L. (2005) Preparative expression of secreted proteins in bacteria: Status report and future prospects, *Current Opinion in Biotechnology*, 16; 538-545. doi: 10.1016/j.copbio.2005.07.008.

Gerber, B. R., Asakura, S. and Oosawa, F. (1973) Effect of temperature on the in vitro assembly of bacterial flagella, *Journal of Molecular Biology*, 16; 538-45. doi: 10.1016/0022-2836(73)90040-5.

Ghosh, C., Gupta, R. and Mukherjee, K. J. (2012) An inverse metabolic engineering approach for the design of an improved host platform for over-expression of recombinant proteins in *Escherichia coli*, *Microbial Cell Factories*, 11; 1-9. doi: 10.1186/1475-2859-11-93.

Glick, B. R., *et al*. (1995) Metabolic load and heterogenous gene expression, *Biotechnology Advances*, 13; 247-261. doi: 10.1016/0734-9750(95)00004-A.

Glick, A. L., *et al.* (2002) Global Support Grows For Afghan Restoration Mammalian Cells Spin, *Science*, 295; 419-421. doi: 10.1126/science.295.5554.419b.

Gophna, U., Ron, E. Z. and Graur, D. (2003) Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events, *Gene*, 312; 151-163. doi: 10.1016/S0378-1119(03)00612-7.

Gottesman, S. *et al.* (1993) ClpX, an Alternative Subunit for the ATP-dependent Clp Protease of *Escherichia coli*, *The Journal of Biological Chemistry*, 268; 22618-26. doi: 268. 22618–22626.

Green, C. A. (2016) Re-engineering the bacterial flagella secretion system : whipping it into shape, uk.bl.ethos.703371.

Green, C. A. *et al.* (2019) Engineering the flagellar type III secretion system: improving capacity for secretion of recombinant protein, *Microbial Cell Factories*, 18; 1-27. doi: 10.1186/s12934-019-1058-4.

Green, E. R. and Mecsas, J. (2015) Bacterial Secretion Systems – An overview, *American Society for Microbiology*, 4; 1-32. doi: 10.1128/MBF-0012-2015.

Green, E. R. and Mecsas, J. (2016) Bacterial Secretion Systems: An Overview, *Microbiology Spectrum*, 4, 1-32. doi: 10.1128/microbiolspec.VMBF-0012-2015.

Grigorova, I. L. *et al.* (2006) Insights into transcriptional regulation and competition from an equilibrium model of RNA polymerase binding to DNA, *Proceedings of the National Academy of Sciences*, *USA,* 103; 5332-7. doi: 10.1073/pnas.0600828103.

Guest, J. R. and Russell, G. C. (1992) Complexes and Complexities of the Citric Acid Cycle in *Escherichia coli*, *Current Topics in Cellular Regulation*, 33; 231-247. doi: 10.1016/B978-0-12-152833-1.50018-6.

Guo, S. *et al.* (2014) Analysis of factors that affect FlgM-dependent type III secretion for

protein purification with *Salmonella enterica* serovar Typhimurium, *Journal of Bacteriology*, 196; 2333-47. doi: 10.1128/JB.01572-14.

Gupta, P. *et al.* (2009) Enhancement of over expression and chaperone assisted yield of folded recombinant aconitase in *Escherichia coli* in bioreactor cultures, *Journal of Bioscience and Bioengineering*, 107; 102-107. doi: 10.1016/j.jbiosc.2008.10.020.

Gurramkonda, C. *et al.* (2010) Application of simple fed-batch technique to high-level secretory production of insulin precursor using *Pichia pastoris* with subsequent purification and conversion to human insulin, *Microbial Cell Factories*, 9; 1-11. doi: 10.1186/1475-2859-9-31.

Habdas, B. J. *et al.* (2010) The LysR-type transcriptional regulator QseD alters type three secretion in enterohemorrhagic *Escherichia coli* and motility in K-12 *Escherichia coli*, *Journal of Bacteriology*, 192; 3699–3712. doi: 10.1128/JB.00382-10.

Han, M. J. and Lee, S. Y. (2003) Proteome profiling and its use in metabolic and cellular engineering, *Proteomics*, 3; 2317-2324. doi: 10.1002/pmic.200300508.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids, *Journal of Molecular Biology*, 166; 557-580. doi: 10.1016/S0022-2836(83)80284-8.

Hannig, G. and Makrides, S. C. (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*, *Trends in Biotechnology*, 16; 54-60. doi: 10.1016/S0167-7799(97)01155-4.

Hao, Y. *et al.* (2016) Protection against deleterious nitrogen compounds: Role of σs-dependent small RNAs encoded adjacent to *sdiA*, *Nucleic Acids Research*, 44; 6935-48. doi: 10.1093/nar/gkw404.

Hastjarjo, T., Silberberg, A. and Hursh, S. R. (1990) Risky choice as a function of amount and variance in food supply, *Journal of the Experimental Analysis of Behavior*, 53; 155-161. doi: 10.1901/jeab.1990.53-155.

Hahn P. H. and Specht, B. (2003) Secretory delivery of recombinant proteins in attenuated *Salmonella* strains : potential and limitations of Type I protein transporters, *FEMS Immunology and Medical Microbiology*, 37; 87-98. doi: 10.1016/S0928-8244(03)00092-0.

Heel, T. *et al.* (2013) FlgM as a Secretion Moiety for the Development of an Inducible Type III Secretion System, *PLoS ONE*, 8; e59034. doi: 10.1371/journal.pone.0059034.

Heinemann, M. and Panke, S. (2006) Synthetic biology-putting engineering into biology, *Bioinformatics* 22; 2790-9. doi: 10.1093/bioinformatics/btl469.

Hellwig, S. *et al.* (2004) Plant cell cultures for the production of recombinant proteins, *Nature Biotechnology*, 22; 1415-1422. doi: 10.1038/nbt1027.

Henderson, G. E., Isett, K. D. and Gerngross, T. U. (2011) Site-specific modification of recombinant proteins: A novel platform for modifying glycoproteins expressed in *E. coli*, *Bioconjugate Chemistry*, 22; 903-912. doi: 10.1021/bc100510g.

Henderson, I. R. *et al.* (2004) Type V Protein Secretion Pathway: the Autotransporter Story, *Microbiology and Molecular Biology Reviews*, 68; 692-744. doi: 10.1128/MMBR.68.4.692-744.2004.

Heng, C. *et al.* (2005) Expression and secretion of an acid-stable α-amylase gene in *Bacillus subtilis* by SacB promoter and signal peptide, *Biotechnology Letters*, 27; 1731-1737. doi: 10.1007/s10529-005-2743-4.

Herlihey, F. A., Moynihan, P. J. and Clarke, A. J. (2014) The essential protein for bacterial flagella formation FlgJ functions as a-N-acetylglucosaminidase, *Journal of Biological Chemistry*, 289; 31029-31042. doi: 10.1074/jbc.M114.603944.

Hevesi, A. and Bleiwas, K. (2005) The Economic Impact of the Biotechnology and Pharmaceutical Industries in New York, *Report of the State Comptroller of New York*,1-8.

Heyland, J., Blank, L. M. and Schmid, A. (2011) Quantification of metabolic limitations during recombinant protein production in *Escherichia coli*, *Journal of Biotechnology*, 155; 178-84. doi: 10.1016/j.jbiotec.2011.06.016.

Higashitani, A. *et al.* (1997) Functional dissection of a cell-division inhibitor, SulA, of *Escherichia coli* and its negative regulation by Lon, *Molecular and General Genetics*, 254; 351-7. doi: 10.1007/s004380050426.

Hirano, T. *et al.* (2003) Substrate specificity classes and the recognition signal for *Salmonella* type III flagellar export, *Journal of Bacteriology*, 185; 2485-2492. doi: 10.1128/JB.185.8.2485-2492.2003.

Hirano, T. *et al.* (2009) Mutations in Flk, FlgG, FlhA, and FlhE that affect the flagellar type III secretion specificity switch in *Salmonella enterica*, *Journal of Bacteriology*, 191; 3938-3947. doi: 10.1128/JB.01811-08.

Ho, B. T., Dong, T. G. and Mekalanos, J. J. (2014) A view to a kill: The bacterial type VI secretion system, *Cell Host and Microbe*, 15; 9-21. doi: 10.1016/j.chom.2013.11.008.

Hoffmann, F. and Rinas, U. (2001) On-line estimation of the metabolic burden resulting from the synthesis of plasmid-encoded and heat-shock proteins by monitoring respiratory energy generation, *Biotechnology and Bioengineering*, 76; 333-40. doi: 10.1002/bit.10098.

Holmqvist, E. *et al.* (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis, *EMBO Journal*, 29; 1840-50. doi: 10.1038/emboj.2010.73.

Homma, M. *et al.* (1990) FlgB, FlgC, FlgF and FlgG. A family of structurally related proteins in the flagellar basal body of *Salmonella typhimurium*, *Journal of Molecular Biology*, 211; 465-477. doi: 10.1016/0022-2836(90)90365-S.

Homma, M., Kutsukake, K. and Iino, T. (1985) Structural genes for flagellar hook-associated proteins in *Salmonella typhimurium*, *Journal of Bacteriology*, 163; 464-471. doi:0021.9193/85/080464-08$02.00/0

Hosking, E. R. *et al.* (2006) The *Escherichia coli* MotAB Proton Channel Unplugged, *The Journal of Molecular Biology,* 1; 921-937. doi: 10.1016/j.jmb.2006.09.035.

Houston, A. I. (1991) Risk-sensitive foraging theory and operant psychology, *Journal of the Experimental Analysis of Behavior*, 56; 585-589. doi: 10.1901/jeab.1991.56-585.

Hu, B., Lara-tejero, M. and Kong, Q. (2002) In Situ Molecular Architecture of the *Salmonella* Type III Secretion, *Cell*, 168; 1065-1074. doi: 10.1016/j.cell.2017.02.022.

Huang, C.J., Lin, H. and Yang, X. (2012) Industrial production of recombinant therapeutics

in *Escherichia coli* and its recent advancements, *Journal of Industrial Microbiology & Biotechnology*, 39; 383-399. doi: 10.1007/s10295-011-1082-9.

Huang, D., Sherman, B. and Lempicki, R. (2009) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Rescources, *Nature protocols*, 4, 44-57. doi: 10.1038/nprot.2008.211.

Hubbard, B. K. *et al.* (1998) Evolution of enzymatic activities in the enolase superfamily: Characterization of the (D)-glucarate/galactarate catabolic pathway in *Escherichia coli*, *Biochemistry*, 41; 14369-14375. doi: 10.1021/bi981124f.

Hughes, K. T. (2014) YdiV: a dual function protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-bound FlhDC complex, *Molecular Microbiology,* 83; 1268-1284. doi: 10.1111/j.1365-2958.2012.08007.x.YdiV.

Hughes, K. T. (2017) Flagellum Length Control: How Long Is Long Enough?, *Current Biology*, 27; 413-415. doi: 10.1016/j.cub.2017.04.008.

Huisman, G. W. and Kolter, R. (1994) Sensing Starvation : A Homoserine Lactone-Dependent Signaling Pathway in *Escherichia coli*, *Science,* 265; 537-540. doi: 10.1126/science.7545940..

Hunt, I. (2005) From gene to protein: A review of new and enabling technologies for multi-parallel protein expression, *Protein Expression and Purification*, 40; 1-22. doi: 10.1016/j.pep.2004.10.018.

Idicula T., S. and Balaji, P. V. (2007) Protein aggregation: A perspective from amyloid and inclusion-body formation, *Current Science*, 92, 758-767.

Ikeda, T., Oosawa, K. and Hotani, H. (1996) Self-assembly of the filament capping protein, FliD, of bacterial flagella into an annular structure, *Journal of Molecular Biology*, 259; 679-686. doi: 10.1006/jmbi.1996.0349.

Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *Journal of Molecular Biology*, 151; 389-409. doi: 10.1016/0022-2836(81)90003-6.

Ikeno, S. and Haruyama, T. (2013) Boost protein expression through co-expression of LEA-like peptide in *Escherichia coli*, *PLoS ONE*, 8, e82824. doi: 10.1371/journal.pone.0082824.

Inaba, S. *et al.* (2013) Exchangeability of the flagellin (FliC) and the cap protein (FliD) among different species in flagellar assembly, *Biopolymers*, 99; 63-72. doi: 10.1002/bip.22141.

Inoue, T. *et al.* (2007) Genome-wide screening of genes required for swarming motility in *Escherichia coli* K-12, *Journal of Bacteriology*, 189; 950-957. doi: 10.1128/JB.01294-06.

Ishihama, A. (2000) Functional Modulation of *Escherichia Coli* RNA Polymerase, *Annual Review of Microbiology*, 54; 499-518. doi: 10.1146/annurev.micro.54.1.499.

Iulia, L. *et al.* (2013) The evidence of contaminant bacterial DNA in several commercial Taq polymerases, *Romanian Biotechnological Letters*, 18; 8007-8012.

Jarvis, K. G. *et al.* (1995) Enteropathogenic *Escherichia coli* contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation, *Proceedings of the National Academy of Sciences the United States of America,* 92, 7996-8000. doi: 10.1073/pnas.92.17.7996.

Jermy, A. (2012) Bacterial secretion: Sec and Tat collaborate in a Rieske business, *Nature Reviews Microbiology*, 10; 801. doi: 10.1038/nrmicro2912.

Jhamb, K. and Sahoo, D. K. (2012) Production of soluble recombinant proteins in *Escherichia coli:* Effects of process conditions and chaperone co-expression on cell growth and production of xylanase, *Bioresource Technology*, 123; 135-143. doi: 10.1016/j.biortech.2012.07.011.

Jishage, M. and Ishihama, A. (1999) Transcriptional organization and in vivo role of the *Escherichia coli rsd* gene, encoding the regulator of RNA polymerase sigma D, *Journal of Bacteriology,* 181; 3768-76. doi: 0021-9193/99/$04.00+0.

Jones, K. L., Kim, S. W. and Keasling, J. D. (2000) Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria, *Metabolic Engineering*, 2; 328-338. doi: 10.1006/mben.2000.0161.

Jørgensen, M. G. *et al.* (2013) Dual function of the McaS small RNA in controlling biofilm formation Dual function of the McaS small RNA in controlling biofilm formation, *Genes and Development*, 27; 1132-1145. doi: 10.1101/gad.214734.113.

Joseph, B. C., Pichaimuthu, S. and Srimeenakshi, S. (2015) An Overview of the Parameters for Recombinant Protein Expression in *Escherichia coli*, *Journal of Cell Science & Therapy*, 06. doi: 10.4172/2157-7013.1000221.

Juhas, M. *et al.* (2014) *Escherichia coli* flagellar genes as target sites for integration and expression of genetic circuits, *PLoS ONE*, 9; 1-7. doi: 10.1371/journal.pone.0111451.

Kanonenberg, K., Schwarz, C. K. W. and Schmitt, L. (2013) Type I secretion systems - a story of appendices, *Research in Microbiology*, 164; 596-604. doi: 10.1016/j.resmic.2013.03.011.

Karlinsey, J. E. *et al.* (2000) Completion of the hook ± basal body complex of the *Salmonella typhimurium* flagellum is coupled to FlgM secretion and fliC transcription, *Molecular Microbiology,* 37; 1220-1231. doi: 10.1046/j.1365-2958.2000.02081.x.

Karlinsey, J. E. *et al.* (2000) Translation/secretion coupling by type III secretion systems, *Cell*, 102; 487-497. doi: S0092-8674(00)00053-2 [pii].

Kato, Y. *et al.* (2007) Effect of the *arcA* Mutation on the Expression of Flagella Genes in *Escherichia coli*, *Bioscience, Biotechnology, and Biochemistry*. 71; 77-83. doi: 10.1271/bbb.60375.

Kawamoto, A. *et al.* (2013) Common and distinct structural features of *Salmonella* injectisome and flagellar basal body, *Scientific Reports*, 3; 1-7. doi: 10.1038/srep03369.

Kedzierska, S. *et al.* (1999) The role of DnaK/DnaJ and GroEL/GroES systems in the removal of endogenous proteins aggregated by heat-shock from *Escherichia coli* cells, *FEBS Letters*, 446; 331-337. doi: 10.1016/S0014-5793(99)00154-4.

Keseler, I. M. *et al.* (2017) The EcoCyc database: Reflecting new knowledge about *Escherichia coli* K-12, *Nucleic Acids Research*, 45; D543-D550. doi:

10.1093/nar/gkw1003.

Kessler, D., Leibrecht, I. and Knappe, J. (1991) Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by *adhE*, *FEBS Letters*, 281; 59-63. doi: 10.1016/0014-5793(91)80358-A.

Khokhlatchev, A. *et al.* (1997) Reconstitution of Mitogen-activated Protein Kinase Phosphorylation Cascades in Bacteria, *Journal of Biological Chemistry*, 27; 1-7. doi: 10.1074/jbc.272.17.11057.

Kim, H. J. *et al.* (2012) An engineered *Escherichia coli* having a high intracellular level of ATP and enhanced recombinant protein production, *Applied Microbiology and Biotechnology*, 94; 1079-86. doi: 10.1007/s00253-011-3779-0.

Kim, J. Y., Kim, Y. G. and Lee, G. M. (2012) CHO cells in biotechnology for production of recombinant proteins: Current state and further potential, *Applied Microbiology and Biotechnology*, 93; 917-930. doi: 10.1007/s00253-011-3758-5.

Kirkpatrick, C. *et al.* (2001) Acetate and formate stress: Opposite responses in the proteome of *Escherichia coli*, *Journal of Bacteriology*, 183; 6466-6477. doi: 10.1128/JB.183.21.6466-6477.2001.

Ko, M. and Park, C. (2000) H-NS-dependent regulation of flagellar synthesis is mediated by a LysR family protein, *Journal of Bacteriology*, 182; 4670-4672. doi: 10.1128/JB.182.16.4670-4672.2000.

Ko, M. and Park, C. (2000) Two novel flagellar components and H-NS are involved in the motor function of *Escherichia coli*, *Journal of Molecular Biology*, 303; 371-82. doi: 10.1006/jmbi.2000.4147.

Kojima S. and Blair D. (2004) The bacterial flagellar motor: structure and function of a complex molecular machine, *International Review of Cytology*, 233; 93-134. doi: 10.1016/s0074-7696(04)33003-2.

Kojima, S. (2015) Dynamism and regulation of the stator, the energy conversion complex of the bacterial flagellar motor, *Current Opinion in Microbiology*, 28; 66-71. doi: 10.1016/j.mib.2015.07.015.

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; 336-51. doi: 10.1038/nrmicro2762.

Kosarewicz, A., Königsmaier, L. and Marlovits, T. C. (2012) The blueprint of the type-3 injectisome, *Philosophical Transactions of the Royal Society Biological Sciences*, 367, 1140-1154. doi: 10.1098/rstb.2011.0205.

Koster, M. *et al.* (1997) The outer membrane component , YscC , of the Yop secretion machinery of *Yersinia enterocolitica* forms a ring-shaped multimeric complex, *Molecular Microbiology,* 26; 789-797. doi: 10.1046/j.1365-2958.1997.6141981.x.

Kotzsch, A. *et al.* (2011) A secretory system for bacterial production of high-profile protein targets, *Protein Science*, 20; 597-609. doi: 10.1002/pro.593.

Kourennaia, O. V., Tsujikawa, L. and DeHaseth, P. L. (2005) Mutational analysis of *Escherichia coli* heat shock transcription factor sigma 32 reveals similarities with sigma 70 in recognition of the -35 promoter element and differences in promoter DNA melting

and -10 recognition, *Journal of Bacteriology*, 187; 6762-6769. doi: 10.1128/JB.187.19.6762-6769.2005.

Kranen, E. *et al.* (2014) Autodisplay for the co-expression of lipase and foldase on the surface of *E. coli*: Washing with designer bugs, *Microbial Cell Factories,* 13, 1-12. doi: 10.1186/1475-2859-13-19.

Kucharova, V. *et al.* (2013) Design and optimization of short DNA sequences that can be used as 5' fusion partners for high-level expression of heterologous genes in *Escherichia coli*, *Applied and Environmental Microbiology*,79; 6655-64. doi: 10.1128/AEM.01676-13.

Kulish, O., Wright, A. D. and Terentjev, E. M. (2016) F1 rotary motor of ATP synthase is driven by the torsionally-asymmetric drive shaft, *Scientific Reports*, 6, 1-14. doi: 10.1038/srep28180.

Kutsukake, K. and Doi, H. (1994) Nucleotide sequence of the flgD gene of *Salmonella typhimurium* which is essential for flagellar hook formation, *BBA-Gene Structure and Expression*, 1218, 443-446. doi: 10.1016/0167-4781(94)90202-X.

Lacour, S. and Landini, P. (2004) σS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: Function of σS-dependent genes and identification of their promoter sequences, *Journal of Bacteriology*, 186; 7186-7195. doi: 10.1128/JB.186.21.7186-7195.2004.

Lasica, A. M. *et al.* (2017) The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function, *Frontiers in Cellular and Infection Microbiology*, 7; 1-19. doi: 10.3389/fcimb.2017.00215.

Lawley, T. D. *et al.* (2003) F factor conjugation is a true type IV secretion system, *FEMS Microbiology Letters*, 224; 1-15. doi: 10.1016/S0378-1097(03)00430-0.

De Lay, N. and Gottesman, S. (2012) A complex network of small non-coding RNAs regulate motility in *Escherichia coli*, *Molecular Microbiology*, 86; 524-538. doi: 10.1111/j.1365-2958.2012.08209.x.

Lee, D. J. *et al.* (2009) Gene doctoring: A method for recombineering in laboratory and pathogenic *Escherichia coli* strains, *BMC Microbiology*, 9; 1-14. doi: 10.1186/1471-2180-9-252.

Lelong, C. *et al.* (2007) Mutual Regulation of Crl and Fur in *Escherichia coli* W3110, *Molecular & Cellular Proteomics*, 6; 660-668. doi: 10.1074/mcp.M600192-MCP200.

Lesley, S. A. *et al.* (2002) Gene expression response to misfolded protein as a screen for soluble recombinant protein, *Protein Engineering, Design and Selection*, 15; 153-160. doi: 10.1093/protein/15.2.153.

Lessard, I. A. D. and Walsh, C. T. (1999) VanX, a bacterial D-alanyl-D-alanine dipeptidase: Resistance, immunity, or survival function?, *Proceedings of the National Academy of Sciences*, 96; 11028-11032. doi: 10.1073/pnas.96.20.11028.

Li, C. *et al.* (1993) Adverse Conditions Which Cause Lack of Flagella in *Escherichia coli*, *Journal of Bacteriology,* 175; 2229-2235. doi: 10.1128/jb.175.8.2229-2235.1993.

Li, W. *et al.* (2009) Role of α-subunit VISIT-DG sequence residues Ser-347 and Gly-351 in the catalytic sites of *Escherichia coli* ATP synthase, *Journal of Biological Chemistry*, 284; 10747-10754. doi: 10.1074/jbc.M809209200.

Linke, D. *et al.* (2006) Trimeric autotransporter adhesins: variable structure, common function, *Trends in Microbiology*, 14; 264-70. doi: 10.1016/j.tim.2006.04.005.

Lino, T. (1977) Genetics of structure and function of bacterial flagella, *Annual Review of Genetics*, 11; 161-182. doi: 10.1146/annurev.ge.11.120177.001113.

Liu, L. *et al.* (2013) Developing *Bacillus spp.* as a cell factory for production of microbial enzymes and industrially important biochemicals in the context of systems and synthetic biology, *Applied Microbiology and Biotechnology*, 97; 6113-6127. doi: 10.1007/s00253-013-4960-4.

Liu, M. *et al.* (2005) Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*, *Journal of Biological Chemistry*, 280; 15921-7. doi: 10.1074/jbc.M414050200.

Liu, X. *et al.* (1995) The C-terminal region of the alpha subunit of *Escherichia coli* RNA polymerase is required for transcriptional activation of the flagellar level II operons by the FlhD/FlhC complex, *Journal of Bacteriology*, 177; 5186-5188. doi: 10.1128/jb.177.17.5186-5188.1995.

Lizak, C. *et al.* (2011) *N* -Linked Glycosylation of Antibody Fragments in *Escherichia coli*, *Bioconjugate Chemistry*, 22; 488-496. doi: 10.1021/bc100511k.

Lloyd, R. G. and Buckman, C. (1991) Overlapping functions of *recD*, *recJ* and *recN* provide evidence of three epistatic groups of genes in *Escherichia coli* recombination and DNA repair, *Biochimie*, 73; 313–320. doi: 10.1016/0300-9084(91)90218-P.

Lobstein, J. *et al.* (2012) Shuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm, *Microbial Cell Factories*, 11; 1–16. doi: 10.1186/1475-2859-11-56.

Lorenz, C. and Büttner, D. (2011) Secretion of early and late substrates of the type III secretion system from *Xanthomonas* is controlled by HpaC and the C-terminal domain of HrcU, *Molecular Microbiology*, 79; 447–67. doi: 10.1111/j.1365-2958.2010.07461.x.

Love, K. R. *et al.* (2012) Systematic Single-Cell Analysis of *Pichia pastoris* Reveals Secretory Capacity Limits Productivity, *PLOS One,* 7; 1-11. doi: 10.1371/journal.pone.0037915.

Lubas, W. and Hanover, J. (2000) Functional Expression of O-linked GlcNAc Transferase, *The Journal of Biological Chemistry*, 275; 10983-10988. doi: 10.1074/jbc.275.15.10983.

Ludwig, P. *et al.* (2018) Non-canonical *Escherichia coli* transcripts lacking a Shine–Dalgarno motif have very different translational efficiencies and do not form a coherent group, *Microbiology*, 164; 646-658. doi: 10.1099/mic.0.000619.

Macnab, R. (1992) Genetics and biogenesis of bacterial flagella, *Annual Review of Genetics*, 261; 131-158. doi: 10.1146/annurev.genet.26.1.131.

Macnab, R. and Koshland, D. E. (1974) Bacterial motility and chemotaxis: Light-induced tumbling response and visualization of individual flagella, *Journal of Molecular Biology*, 84; 399-406. doi: 10.1016/0022-2836(74)90448-3.

Macnab, R. M. (2003) How bacteria assemble flagella, *Annual Review of Microbiology*, 57; 77-100. doi: 10.1146/annurev.micro.57.030502.090832.

Macnab, R. M. (2004) Type III flagellar protein export and flagellar assembly, *Biochimica et Biophysica Acta*, 1694; 207-17. doi: 10.1016/j.bbamcr.2004.04.005.

Macnab, R. and Parkinson, J. (1991) Genetic analysis of the bacterial flagellum, *Trends in Genetics*, 7; 196-200. doi: 10.1016/0168-9525(91)90436-T.

Maffei, B., Francetic, O. and Subtil, A. (2017) Tracking Proteins Secreted by Bacteria : What 's in the Toolbox ?, *Frontiers in Cellular and Infection Microbiology,* 7; 1-17. doi: 10.3389/fcimb.2017.00221.

Mahalik, S., Sharma, A. K. and Mukherjee, K. J. (2014) Genome engineering for improved recombinant protein expression in *Escherichia coli*, *Microbial Cell Factories*, 13; 1-13. doi: 10.1186/s12934-014-0177-1.

Majander, K. *et al.* (2005) Extracellular secretion of polypeptides using a modified *Escherichia coli* flagellar secretion apparatus, *Nature Biotechnology*, 23; 475-481. doi: 10.1038/nbt1077.

Majander, K., Korhonen, T. K. and Westerlund W. B. (2005) Simultaneous display of multiple foreign peptides in the FliD capping and FliC filament proteins of the *Escherichia coli* flagellum, *Applied and Environmental Microbiology*, 71; 4263-4268. doi: 10.1128/AEM.71.8.4263-4268.2005.

Mandin, P. and Gottesman, S. (2010) Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA, *EMBO Journal,* 29; 3094-3107. doi: 10.1038/emboj.2010.179.

De Marco, A. and De Marco, V. (2004) Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning, *Journal of Biotechnology*, 109; 45-52. doi: 10.1016/j.jbiotec.2003.10.025.

Marschall, L., Sagmeister, P. and Herwig, C. (2016) Tunable recombinant protein expression in *E. coli:* enabler for continuous processing?, *Applied Microbiology and Biotechnology*, 100; 5719-28. doi: 10.1007/s00253-016-7550-4.

Martínez-Alonso, M. *et al.* (2008) The functional quality of soluble recombinant polypeptides produced in *Escherichia coli* is defined by a wide conformational spectrum, *Applied and Environmental Microbiology*, 74; 7431-7433. doi: 10.1128/AEM.01446-08.

Martínez-Alonso, M. *et al.* (2010) Side effects of chaperone gene co-expression in recombinant protein production, *Microbial Cell Factories*, 9, 1-6. doi: 10.1186/1475-2859-9-64.

Martinez, R. J. (1966) The formation of bacterial flagella, *Journal of Molecular Biology*, 17, 10–17. doi: 10.1016/S0022-2836(66)80090-6.

Matos, C. F. R. O. *et al.* (2012) High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in *Escherichia coli*, *Biotechnology and Bioengineering*, 109, 2533-2542. doi: 10.1002/bit.24535.

Matsunami, H. *et al.* (2016) Complete structure of the bacterial flagellar hook reveals extensive set of stabilizing interactions, *Nature Communications,* 7, 13425. doi: 10.1038/ncomms13425.

McCarter, L. L. (2006) Regulation of flagella, *Current Opinion in Microbiology*, 9; 180-186. doi: 10.1016/j.mib.2006.02.001.

McDaniel, L. E., Bailey, E. G. and Zimmerli, A. (1965) Effect of oxygen supply rates on growth of *Escherichia coli*: Studies in unbaffled and baffled shake flasks, *Applied Microbiology*, 13, 109-14. doi:14264837.

McLaughlin, L. S., Haft, R. J. F. and Forest, K. T. (2012) Structural insights into the Type II secretion nanomachine, *Current Opinion in Structural Biology*, 22, 208-216. doi: 10.1016/j.sbi.2012.02.005.

McMurry, J. L. *et al.* (2015) Weak interactions between *Salmonella enterica* FlhB and other flagellar export apparatus proteins govern type III secretion dynamics, *PLoS ONE*, 10, e0134884. doi: 10.1371/journal.pone.0134884.

Mergulhão, F. J. M., Summers, D. K. and Monteiro, G. A. (2005) Recombinant protein secretion in *Escherichia coli*, *Biotechnology Advances*, 23; 177-202. doi: 10.1016/j.biotechadv.2004.11.003.

Merino, S. and Tomás, J. M. (2014) Gram-negative flagella glycosylation, *International Journal of Molecular Sciences*, 15; 2840-57. doi: 10.3390/ijms15022840.

Meshcheryakov, V. A. *et al.* (2013) Function of FlhB, a Membrane Protein Implicated in the Bacterial Flagellar Type III Secretion System, *PLoS ONE*, 8; 1-13. doi: 10.1371/journal.pone.0068384.

Mijakovic, I. *et al.* (2006) Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine, *Nucleic Acids Research*, 34; 1588-1596. doi: 10.1093/nar/gkj514.

Mika, F. and Hengge, R. (2013) Small regulatory RNAs in the control of motility and biofilm formation in *E. Coli* and *Salmonella*, *International Journal of Molecular Sciences*, 14, 4560-4579. doi: 10.3390/ijms14034560.

Minamino, T. *et al.* (2003) The ATPase FliI can interact with the type III flagellar protein export apparatus in the absence of its regulator, FliH, *Journal of Bacteriology*, 185; 3983-3988. doi: 10.1128/JB.185.13.3983-3988.2003.

Minamino, T. (2014) Protein export through the bacterial flagellar type III export pathway, *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843; 1642-1648. doi: 10.1016/j.bbamcr.2013.09.005.

Minamino, T. *et al.* (2014) The bacterial flagellar protein export apparatus processively transports flagellar proteins even with extremely infrequent ATP hydrolysis, *Scientific Reports*, 4, 7579-87. doi: 10.1038/srep07579.

Minamino, T. *et al.* (2016) The Bacterial Flagellar Type III Export Gate Complex Is a Dual Fuel Engine That Can Use Both H+ and Na+ for Flagellar Protein Export, *PLoS Pathogens*, 12, e1005495. doi: 10.1371/journal.ppat.1005495.

Minamino, T, Imada, K. and Namba, K. (2008) Mechanisms of type III protein export for bacterial flagellar assembly, *Molecular Biosystems*, 4; 1105-1115. doi: 10.1039/b808065h.

Minamino, T., Imada, K. and Namba, K. (2008) Molecular motors of the bacterial flagella, *Current Opinion in Structural Biology*, 18; 693-701. doi: 10.1016/j.sbi.2008.09.006.

Minamino, T. and Macnab, R. M. (1999) Components of the *Salmonella* flagellar export apparatus and classification of export substrates, *Journal of Bacteriology*, 181, 1388-94.

doi: PMC93525.

Miranda-CasoLuengo, A. A. *et al.* (2017) Small RNA, Big Effect: Control of Flagellin Production, *Trends in Microbiology*, 25; 953-954. doi: 10.1016/j.tim.2017.10.003.

Mironova, R. *et al.* (2005) Evidence for non-enzymatic glycosylation of *Escherichia coli* chromosomal DNA, *Molecular Microbiology*, 55; 1801-1811. doi: 10.1111/j.1365-2958.2005.04504.x.

Mizuno, S. *et al.* (2011) The NMR Structure of FliK , the Trigger for the Switch of Substrate Specificity in the Flagellar Type III Secretion Apparatus, *Journal of Molecular Biology*, 409; 558-573. doi: 10.1016/j.jmb.2011.04.008.

Moens, S. and Vanderleyden, J. (1996) Functions of bacterial flagella, *Critical Reviews in Microbiology*, 22, 67-100. doi: 10.3109/10408419609106456.

Mogk, A. *et al.* (2003) Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation, *Molecular Microbiology*, 50; 585-95. doi: 10.1046/j.1365-2958.2003.03710.x.

Monteiro, C. *et al.* (2012) Hfq and Hfq-dependent small RNAs are major contributors to multicellular development in *Salmonella enterica* serovar typhimurium, *RNA Biology*. 9; 489-502. doi: 10.4161/rna.19682.

Moriya, N. *et al.* (2006) The Type III Flagellar Export Specificity Switch is Dependent on FliK Ruler and a Molecular Clock, *Journal of Molecular Biology*, 359, 466-477. doi: 10.1016/j.jmb.2006.03.025.

Mukherji, S. and van Oudenaarden, A. (2009) Synthetic biology: understanding biological design from synthetic circuits, *Nature Reviews -Genetics*, 10, 859-871. doi: 10.1038/nrg2697.

Munera, D. *et al.* (2010) N-Terminal Type III Secretion Signal of Enteropathogenic *Escherichia coli* Translocator Proteins, *Journal of Bacteriology,* 192, 3534-3539. doi: 10.1128/JB.00046-10.

Murata, T. *et al.* (2008) Phosphorylation of two eukaryotic transcription factors, Jun dimerization protein 2 and activation transcription factor 2, in *Escherichia coli* by Jun N-terminal kinase 1, *Analytical Biochemistry* 376, 115-121. doi: 10.1016/j.ab.2008.01.038.

Murphy, K. C. (1998) Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*, *Journal of Bacteriology*, 180, 2063-2071. doi: 0021-9193/98/$04.00+0.

Muskotál, A. *et al.* (2006) Interaction of FliS flagellar chaperone with flagellin, *FEBS Letters*, 580, 3916-20. doi: 10.1016/j.febslet.2006.06.024.

Nahku, R. *et al.* (2010) Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures, *Journal of Biotechnology*, 145; 60-5. doi: 10.1016/j.jbiotec.2009.10.007.

Nair, S. and Finkel, S. E. (2004) Dps protects cells against multiple stresses during stationary phase, *Journal of Bacteriology*, 186; 4192-8. doi: 10.1128/JB.186.13.4192-4198.2004.

Nakanishi, A. *et al.* (1998) Identification of DNA gyrase inhibitor (GyrI) in *Escherichia*

coli, *Journal of Biological Chemistry*, 273; 1933-1938. doi: 10.1074/jbc.273.4.1933.

Namba, K. and Vonderviszt, F. (1997) Molecular architecture of bacterial flagellum, *Quarterly Reviews of Biophysics*, 30; 1-65. doi: 10.1017/S0033583596003319.

Nandagopal, N. and Elowitz, M. B. (2011) Synthetic biology: integrated gene circuits, *Science,* 333; 1244-1248. doi: 10.1126/science.1207084.

Narayanan, N., Khan, M. and Chou, C. P. (2010) Enhancing functional expression of heterologous lipase B in *Escherichia coli* by extracellular secretion, *Journal of Industrial Microbiology and Biotechnology*, 37, 349-361. doi: 10.1007/s10295-009-0680-2.

Neuwald, A. F. *et al.* (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes, *Genome Research*, 9; 27-43. doi: 10.1101/gr.9.1.27.

Ni, B. *et al.* (2017) Evolutionary Remodeling of Bacterial Motility Checkpoint Control, *Cell Reports*, 18; 866-877. doi: 10.1016/j.celrep.2016.12.088.

Ni, Y and Chen, R. (2009) Extracellular recombinant protein production from *Escherichia coli*, *Biotechnology Letters*, 31; 1661-1670. doi: 10.1007/s10529-009-0077-3.

Nielsen, H. and Engelbrecht, J. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, *Protein Engineering*, 10; 1-6. doi: 10.1142/S0129065797000537.

Nishihara, K. *et al.* (1998) Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*, *Applied and Environmental Microbiology*, 64; 1694-9. doi: 10.1111/j.1749-6632.1956.tb40107.x.

Nishihara, K. *et al.* (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*, *Applied and Environmental Microbiology*, 66; 884-889. doi: 10.1128/AEM.66.3.884-889.2000.

Nivaskumar, M. and Francetic, O. (2014) Type II secretion system: A magic beanstalk or a protein escalator, *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843; 1568-77. doi: 10.1016/j.bbamcr.2013.12.020.

Oberhettinger, P. *et al.* (2012) Intimin and Invasin Export Their C-Terminus to the Bacterial Cell Surface Using an Inverse Mechanism Compared to Classical Autotransport, *PLoS ONE*, 7; e47069. doi: 10.1371/journal.pone.0047069.

Oh, T. J., Jung, I. L. and Kim, I. G. (2001) The *Escherichia coli* SOS gene sbmC is regulated by H-NS and RpoS during the SOS induction and stationary growth phase, *Biochemical and Biophysical Research Communications*, 288; 1052-1058. doi: 10.1006/bbrc.2001.5872.

Ohnishi, K. *et al.* (1994) FlgD is a scaffolding protein needed for flagellar hook assembly in *Salmonella typhimurium*, *Journal of Bacteriology*, 176; 2272-81. doi: 10.1128/jb.176.8.2272-2281.1994.

Olivares, A. O., Baker, T. A. and Sauer, R. T. (2015) Mechanistic insights into bacterial AAA + proteases and protein- remodelling machines, *Nature Reports,* 14; 33-44. doi: 10.1038/nrmicro.2015.4.

Olmos S. J. and Contreras S. R. (2003) Genetic system constructed to overproduce and secrete proinsulin in Bacillus subtilis, *Applied Microbiology and Biotechnology*, 62, 369-373. doi: 10.1007/s00253-003-1289-4.

Olsén, A., Jonsson, A. and Normark, S. (1989) Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*, *Nature*, 338; 652-5. doi: 10.1038/338652a0.

Opdyke, J. A., Kang, J. G. and Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in *Escherichia coli*, *Journal of Bacteriology*, 186; 6698-6705, doi: 10.1128/JB.186.20.6698-6705.2004.

Paddon, C. J. and Keasling, J. D. (2014) Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development, *Nature Reviews Microbiology*, 186; 6698-705. doi: 10.1038/nrmicro3240.

Pallen, M. J., Chaudhuri, R. R. and Henderson, I. R. (2003) Genomic analysis of secretion systems, *Current Opinion in Microbiology*, 6, 519-527. doi: 10.1016/j.mib.2003.09.005.

Pandhal, J. *et al.* (2012) Systematic metabolic engineering for improvement of glycosylation efficiency in *Escherichia coli*, *Biochemical and Biophysical Research Communications*, 419, 472-476. doi: 10.1016/j.bbrc.2012.02.020.

Pandhal, J. *et al.* (2013) Inverse metabolic engineering to improve *Escherichia coli* as an N-glycosylation host, *Biotechnology and Bioengineering*, 110, 2482-2493. doi: 10.1002/bit.24920.

Papenfort, K. *et al.* (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA, *Molecular Microbiology*, 74; 139-158. doi: 10.1111/j.1365-2958.2009.06857.x.

Park, C. (1995) Modulation of Flagellar Expression in *Escherichia coli* by Acetyl Phosphate and the Osmoregulator OmpR, *Journal of Bacteriology*, 177, 4696-4702. doi: 10.1128/jb.177.16.4696-4702.1995.

Patridge, E. V. and Ferry, J. G. (2006) WrbA from *Escherichia coli* and *Archaeoglobus fulgidus* is an NAD(P)H:quinone oxidoreductase', *Journal of Bacteriology*, 188; 3498-506. doi: 10.1128/JB.188.10.3498-3506.2006.

Paul, K. *et al.* (2008) Energy source of flagellar type III secretion, *Nature*, 451, 489-492. doi: 10.1038/nature06497.

Paul, K. *et al.* (2010) The c-di-GMP Binding Protein YcgR Controls Flagellar Motor Direction and Speed to Affect Chemotaxis by a "Backstop Brake" Mechanism, *Molecular Cell*, 38; 128-39. doi: 10.1016/j.molcel.2010.03.001.

Peebo, K. and Neubauer, P. (2018) Application of Continuous Culture Methods to Recombinant Protein Production in Microorganisms, *Microorganisms*, 6; 56-64. doi: 10.3390/microorganisms6030056.

Peng, S. *et al.* (2014) Evaluation of three reference genes of *Escherichia coli* for mRNA expression level normalization in view of salt and organic acid stress exposure in food, *FEMS Microbiology Letters*, 355; 78-82. doi: 10.1111/1574-6968.12447.

Petsch, D. and Anspach, F. B. (2000) Endotoxin removal from protein solutions, *Journal of Biotechnology*, 76; 97-119. doi: 10.1016/S0168-1656(99)00185-6.

Pham, T. K. *et al.* (2010) A quantitative proteomic analysis of biofilm adaptation by the periodontal pathogen *Tannerella forsythia*, *Proteomics*, 10; 3130-41. doi: 10.1002/pmic.200900448.

Phillips, G. J. and Silhavy, T. J. (1990) Heat-shock proteins DnaK and GroEL facilitate export of LacZ hybrid proteins in *E. coli*, *Nature*, 344; 882-4. doi: 10.1038/344882a0.

Pina, A. S., Lowe, C. R. and Roque, A. C. A. (2014) Challenges and opportunities in the purification of recombinant tagged proteins, *Biotechnology Advances*, 32; 366-381. doi: 10.1016/j.biotechadv.2013.12.001.

Pinkas, D. M. *et al.* (2011) Tunable, post-translational hydroxylation of collagen domains in *Escherichia coli*, *ACS Chemical Biology*, 6; 320-324. doi: 10.1021/cb100298r.

Piper, S. E. *et al.* (2009) A global view of *Escherichia coli* Rsd protein and its interactions, *Molecular BioSystems*, 5; 1943-7. doi: 10.1039/b904955j.

Pitera, D. J. *et al.* (2007) Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*, *Metabolic Engineering*, 9; 193-207. doi: 10.1016/j.ymben.2006.11.002.

Portaliou, A. G. *et al.* (2016) Type III Secretion: Building and Operating a Remarkable Nanomachine, *Trends in Biochemical Sciences*, 4; 175-189. doi: 10.1016/j.tibs.2015.09.005.

Pratt, L. A. and Silhavy, T. J. (1998) Crl stimulates RpoS activity during stationary phase, *Molecular Microbiology*, 29; 1225-36. doi: 10.1046/j.1365-2958.1998.01007.x.

Prinz, W. a *et al.* (1997) The Role of the Thioredoxin and Glutaredoxin Pathways in Reducing Protein Disulfide Bonds in the *Escherichia coli* Cytoplasm, *Cell Biology and Metabolism*, 272; 15661-15667. doi: 10.1074/jbc.272.25.15661.

Promega (2012) Cloning', *Protocols & Applications Guide*, 23.

Pru, B. M. *et al.* (2001) FlhD / FlhC-regulated promoters analyzed by gene array and lacZ gene fusions, *FEMS Microbiology Letters,* 197, 91-97. doi: 10.1111/j.1574-6968.2001.tb10588.x.

Pru, B. M. and Matsumura, P. (1996) A Regulator of the Flagellar Regulon of *Escherichia coli*, flhD, also affects cell division, *Journal of Bacteriology,* 178; 668-674. doi: 10.1128/jb.178.3.668-674.1996.

Pugsley, A. P. *et al.* (1997) Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in Gram-negative bacteria - A review, *Gene*, 192; 13-19. doi: 10.1016/S0378-1119(96)00803-7.

Py, B., Loiseau, L. and Barras, F. (2001) An inner membrane platform in the type II secretion machinery of Gram-negative bacteria, *EMBO Reports*, 2; 244-248. doi: 10.1093/embo-reports/kve042.

Records, A. R. (2011) The Type VI Secretion System: A Multipurpose Delivery System with a Phage-Like Machinery, *Molecular Plant-Microbe Interactions*, 24; 751-757. doi: 10.1094/MPMI-11-10-0262.

Renault, T. T. *et al.* (2017) Bacterial flagella grow through an injection-diffusion mechanism, *eLife*, 6, e23136. doi: 10.7554/eLife.23136.

Retallack, D. M., Jin, H. and Chew, L. (2012) Reliable protein production in a Pseudomonas fluorescens expression system, *Protein Expression and Purification*, 81; 157-165. doi: 10.1016/j.pep.2011.09.010.

Riley, M. *et al.* (2006) *Escherichia coli* K-12: A cooperatively developed annotation snapshot - 2005, *Nucleic Acids Research*, 34; 1-9. doi: 10.1093/nar/gkj405.

Roblin, P. *et al.* (2015) A Salmonella type three secretion effector/chaperone complex adopts a hexameric ring-like structure, *Journal of Bacteriology*, 197; 688-98. doi: 10.1128/JB.02294-14.

Rodríguez-Limas, W. A., Tannenbaum, V. and Tyo, K. E. J. (2015) Blocking endocytotic mechanisms to improve heterologous protein titers in *Saccharomyces cerevisiae*, *Biotechnology and Bioengineering*, 112; 376-385. doi: 10.1002/bit.25360.

Rohe, P. *et al.* (2012) An automated workflow for enhancing microbial bioprocess optimization on a novel microbioreactor platform, *Microbial Cell Factories* 11, 1-14. doi: 10.1186/1475-2859-11-144.

Rosano, L. and Ceccarelli, E. A. (2014) Recombinant protein expression in *Escherichia coli*: Advances and challenges, *Frontiers in Microbiology*, 5; 1-17. doi: 10.3389/fmicb.2014.00172.

Ross, P. L. *et al.* (2004) Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents, *Molecular & Cellular Proteomics*, 3; 1154-69. doi: 10.1074/mcp.M400129-MCP200.

Rudolph, R. and Lilie, H. (1996) In vitro folding of inlusion body proteins, *FASEB Journal*, 10, 49-56. doi: 10.1021/bi2012965.

Ryjenkov, D. A. *et al.* (2006) The PilZ domain is a receptor for the second messenger c-di-GMP: The PilZ domain protein YcgR controls motility in enterobacteria, *Journal of Biological Chemistry*, 281; 30310-4. doi: 10.1074/jbc.C600179200.

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; 29096-104. doi: 10.1074/jbc.270.49.29096.

Sahdev, S., Khattar, S. K. and Saini, K. S. (2008) Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies, *Molecular and Cellular Biochemistry*, 307; 249-264. doi: 10.1007/s11010-007-9603-6.

Saida, F. *et al.* (2006) Expression of Highly Toxic Genes in *E. coli*: Special Strategies and Genetic Tools, *Current Protein and Peptide Science*, 7; 47-56. doi: 10.2174/138920306775474095.

Saier, M. H. (2006) Protein secretion and membrane insertion systems in gram-negative bacteria, *Journal of Membrane Biology*, 214; 75-90. doi: 10.1007/s00232-006-0049-7.

Saijo-Hamano, Y. *et al.* (2004) In Vitro characterization of FlgB, FlgC, FlgF, FlgG, and FliE, flagellar basal body proteins of *Salmonella*, *Journal of Molecular Biology*, 339; 423-435. doi: 10.1016/j.jmb.2004.03.070.

Saito, T. *et al.* (1998) Flagellar filament elongation can be impaired by mutations in the hook protein FlgE of *Salmonella typhimurium* : a possible role of the hook as a passage for the anti-sigma factor FlgM, *Molecular Microbiology,* 27, 1129-1139. doi:

10.1046/j.1365-2958.1998.00738.x.

Sajó, R. *et al.* (2014) Soluble components of the flagellar export apparatus, FliI, FliJ, and FliH, do not deliver flagellin, the major filament protein, from the cytosol to the export gate, *Biochimica et Biophysica Acta Molecular Cell Research*, 1843; 2414-23. doi: 10.1016/j.bbamcr.2014.07.004.

Samatey, F., *et al.* (2004) Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism, *Nature*, 431; 1062-1068. doi: 10.1038/nature02997.

Sandkvist, M., Cross, A. R. and Branch, C. (2001) MicroReview Biology of type II secretion, *Molecular Microbiology,* 40; 271-83. doi: 10.1046/j.1365-2958.2001.02403.x.

Saraogi, I. and Shan, S. (2014) Co-translational protein targeting to the bacterial membrane, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843; 1433-1441. doi: 10.1016/j.bbamcr.2013.10.013.

Sato, Y. *et al.* (2014) FliT selectively enhances proteolysis of FlhC Subunit in FlhD4C2 complex by an ATP-dependent protease, ClpXP, *Journal of Biological Chemistry*, 289; 33001–33011. doi: 10.1074/jbc.M114.593749.

Saunders, S. E. and Burke, J. F. (1990) Rapid isolation of miniprep DNA for double strand sequencing, *Nucleic Acid Research,* 18; 4948. doi: 10.1093/nar/18.16.4948.

Schade, S. Z., Adler, J. and Ris, H. (1967) How bacteriophage chi attacks motile bacteria, *Journal of Virology*, 1, 599-609. doi: PMC375288.

Schädlich, L. *et al.* (2009) Refining HPV 16 L1 purification from E. coli: Reducing endotoxin contaminations and their impact on immunogenicity, *Vaccine*, 27, 1511-1522. doi: 10.1016/j.vaccine.2009.01.014.

Schaerlaekens, K. *et al.* (2004) Comparison of the Sec and Tat secretion pathways for heterologous protein production by *Streptomyces lividans*, *Journal of Biotechnology*, 112; 279-288. doi: 10.1016/j.jbiotec.2004.05.004.

Schellekens, H. (2002) Immunogenicity of therapeutic proteins: Clinical implications and future prospects, *Clinical Therapeutics*, 24; 1720-1740. doi: 10.1016/S0149-2918(02)80075-3.

Schlegel, S., Rujas, E., Ytterberg, A. J., Zubarev, R. A., Luirink, J., Gier, J., *et al.* (2013) Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels, *Microbial Cell Factories*, 12, 1-12. doi: 10.1186/1475-2859-12-24.

Schnetz, K. (2002) Silencing of the *Escherichia coli* bgl operon by RpoS requires Crl, *Microbiology*, 148; 2573-8. doi: 10.1099/00221287-148-8-2573.

Schuhmacher, J. S., Thormann, K. M. and Bange, G. (2015) How bacteria maintain location and number of flagella, *FEMS Microbiology Reviews*, 39; 812-22. doi: 10.1093/femsre/fuv034.

Serrano, L. (2007) Synthetic biology: promises and challenges, *Molecular Systems Biology*, 3, 158-63. doi: 10.1038/msb4100202.

Sheu, D.-S., Wang, Y. T. and Lee, C.-Y. (2017) Rapid detection of polyhydroxyalkanoate-

accumulating bacteria isolated from the environment by colony PCR, *Microbiology*, 1612; 42–2019. doi: 10.1099/00221287-146-8-2019.

Shi, W. *et al.* (1992) DnaK, DnaJ, and GrpE are required for flagellum synthesis in *Escherichia coli*, *Journal of Bacteriology*, 174; 6256-6263. doi: 10.1128/jb.174.19.6256-6263.1992.

Shin, J. and Noireaux, V. (2010) Study of messenger RNA inactivation and protein degradation in an *Escherichia coli* cell-free expression system, *Journal of Biological Engineering*, 1; 4-9. doi: 10.1186/1754-1611-4-9.

Singer, H. M. *et al.* (2012) Selective purification of recombinant neuroactive peptides using the flagellar type III secretion system, *mBio*, 3; 115-12. doi: 10.1128/mBio.00115-12.

Singer, H. M., Erhardt, M. and Hughes, K. T. (2014) Comparative analysis of the secretion capability of early and late flagellar type III secretion substrates, *Molecular Microbiology*, 93; 505-520. doi: 10.1111/mmi.12675.

Sivashanmugam, A. *et al.* (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*, *Protein Science,* 18; 936-48. doi: 10.1002/pro.102.

Smith, H. E. (2007) The transcriptional response of *Escherichia coli* to recombinant protein insolubility, *Journal of Structural and Functional Genomics*, 8; 27-35. doi: 10.1007/s10969-007-9030-7.

Smith, R. W. and Koffler, H. (1971) Bacterial Flagella, *Advances in Microbial Physiology*, 6; 219-339. doi: 10.1016/S0065-2911(08)60070-3.

Soini, J., Ukkonen, K. and Neubauer, P. (2008) High cell density media for Escherichia coli are generally designed for aerobic cultivations- Consequences for large-scale bioprocesses and shake flask cultures, *Microbial Cell Factories*, 7, 1-11. doi: 10.1186/1475-2859-7-26.

Sørensen, H. P. and Mortensen, K. K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*, *Journal of Biotechnology*, 115; 113-128. doi: 10.1016/j.jbiotec.2004.08.004.

Soutourina, O. *et al.* (1999) Multiple Control of Flagellum Biosynthesis in *Escherichia coli:* Role of H-NS Protein and the Cyclic AMP-Catabolite Activator Protein Complex in Transcription of the flhDC Master Operon, *Journal of Bacteriology,* 181; 7500-7508. odi: PMC94207.

Soutourina, O. A. and Bertin, P. N. (2003) Regulation cascade of flagellar expression in Gram-negative bacteria, *FEMS Microbiology Reviews*, 27; 505-523. doi: 10.1016/S0168-

Sowa, Y. and Berry, R. M. (2008) Bacterial flagellar motor, *Quarterly Reviews of Biophysics*, 41; 103-132. doi: 10.1017/S0033583508004691.

Springer, M. S., Goy, M. F. and Adler, J. (1977) Sensory transduction in *Escherichia coli:* two complementary pathways of information processing that involve mehtylated proteins, *Proceedings of the National Academy of Sciences, USA*, 74; 3312-6. doi: 10.1073/pnas.74.8.3312.

Stafford, G. P., *et al.* (2007) Sorting of Early and Late Flagellar Subunits After Docking at

the Membrane ATPase of the Type III Export Pathway, *Journal of Molecular Biology*, 374; 877-882. doi: 10.1016/j.jmb.2007.09.080.

Stafford, G. P., Ogi, T. and Hughes, C. (2005) Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator FlhD2C2, *Microbiology*, 151; 1779-1788, doi: 10.1099/mic.0.27879-0.

Van D. S. *et al.* (1998) Concepts and principles of O-linked glycosylation, *Critical Reviews in Biochemistry and Molecular Biology*, 33; 151-208. doi: 10.1080/10409239891204198.

Stojiljkovic, I., Schonherr, R. and Kusters, J. G. (1995) Identification of the hopG gene, a component of *Escherichia coli* K-12 type II export system, and its conservation among different pathogenic *Escherichia coli* and *Shigella* isolates, *Journal of Bacteriology*, 177; 1892-5. doi: 10.1128/jb.177.7.1892-1895.1995.

Strandberg, L. and Enfors, S. O. (1991) Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*, *Applied and Environmental Microbiology*, 57; 1669-1674. doi: 10.3389/fcimb.2012.00161.

Surette, M. G. and Bassler, B. L. (1998) Quorum sensing in *Escherichia coli* and *Salmonella typhimurium*, *Proceedings of the National Academy of Sciences of the United States of America*, 95; 7046-50. doi: 10.1073/pnas.95.12.7046.

Swartz, J. R. (2001) Advances in *Escherichia coli* production of therapeutic proteins, *Current Opinion in Biotechnology*, 12; 195-201. doi: 10.1016/S0958-1669(00)00199-3.

Terashima, H., Kojima, S. and Homma, M. (2008) Chapter 2: Flagellar Motility in Bacteria. Structure and Function of Flagellar Motor, *International Review of Cell and Molecular Biology*, 270; 39-85. doi: 10.1016/S1937-6448(08)01402-0.

Thomas, J. G. and Baneyx, F. (2000) ClpB and HtpG facilitate de novo protein folding in stressed *Escherichia coli* cells, *Molecular Microbiology*, 36; 1360-70. doi: 10.1046/j.1365-2958.2000.01951.x.

Thomas, J., Stafford, G. P. and Hughes, C. (2004) Docking of cytosolic chaperone-substrate complexes at the membrane ATPase during flagellar type III protein export, *Proceedings of the National Academy of Sciences of the United States of America*, 101; 3945-3950. doi: 10.1073/pnas.0307223101.

Thomas, S., Holland, I. B. and Schmitt, L. (2014) The Type 1 secretion pathway - The hemolysin system and beyond, *Biochimica et Biophysica Acta - Molecular Cell Research*., 1843; 1629-1641. doi: 10.1016/j.bbamcr.2013.09.017.

Thomason, L. C. *et al.* (2007) Multicopy plasmid modification with phage Red recombineering, *Plasmid*, 58; 148-158. doi: 10.1016/j.plasmid.2007.03.001.

Thomason, L. C., Costantino, N. and Court, D. L. (2007) *E. coli* Genome Manipulation by P1 Transduction, *Current Protocols in Molecular Biology*, 1; 1.17.1-1.17.8. doi: 10.1002/0471142727.mb0117s79.

Thomason, M. K. *et al.* (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*, *Molecular Microbiology*, 84; 17-35. doi: 10.1111/j.1365-2958.2012.07965.x.

Tolia, N. H. and Joshua-Tor, L. (2006) Strategies for protein coexpression in *Escherichia coli*, *Nature Methods*, 3; 55-64. doi: 10.1038/nmeth0106-55.

Tomoyasu, T. *et al.* (2002) The C1pXP ATP-dependent protease regulates flagellum synthesis in *Salmonella enterica* serovar Typhimurium, *Journal of Bacteriology*, 184; 645-653. doi: 10.1128/jb.184.3.645-653.2002.

Tomoyasu, T. *et al.* (2003) Turnover of FlhD and FlhC, master regulator proteins for Salmonella flagellum biogenesis, by the ATP-dependent ClpXP protease, *Molecular Microbiology*, 48; 443-452. doi: 10.1046/j.1365-2958.2003.03437.x.

Tsao, C. Y. *et al.* (2011) LuxS coexpression enhances yields of recombinant proteins in *Escherichia coli* in part through posttranscriptional control of GroEL, *Applied and Environmental Microbiology*, 77; 2141-52. doi: 10.1128/AEM.02347-10.

Tseng, T. T., Tyler, B. M. and Setubal, J. C. (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology, *BMC Microbiology*, 9; 1-9. doi: 10.1186/1471-2180-9-S1-S2.

Turner, L. *et al.* (2010) Visualization of flagella during bacterial swarming, *Journal of Bacteriology*, 192; 3259-3267. doi: 10.1128/JB.00083-10.

Turner, L., Stern, A. S. and Berg, H. C. (2012) Growth of flagellar filaments of *Escherichia coli* is independent of filament length, *Journal of Bacteriology*, 194; 2437-42. doi: 10.1128/JB.06735-11.

Tyanova, S. *et al.* (2016) The Perseus computational platform for comprehensive analysis of proteomics data, *Nature Methods*, 13; 731-40. doi: 10.1038/nmeth.3901.

Typas, A. *et al.* (2007) Stationary phase reorganisation of the *Escherichia coli* transcription machinery by Crl protein, a fine-tuner of activity and levels, *EMBO Journal,* 26; 1569-1578. doi: 10.1038/sj.emboj.7601629.

Uchiyama, J. *et al.* (2010) Involvement of σS accumulation in repression of the flhDC operon in acidic phospholipid-deficient mutants of *Escherichia coli*, *Microbiology*, 156, 1650-1660. doi: 10.1099/mic.0.036749-0.

Ukkonen, K. *et al.* (2011) High-yield production of biologically active recombinant protein in shake flask culture by combination of enzyme-based glucose delivery and increased oxygen transfer, *Microbial Cell Factories*, 107, 1-9. doi: 10.1186/1475-2859-10-107.

Ulfig, A. and Freudl, R. (2018) The early mature part of bacterial twin-arginine translocation (Tat) precursor proteins contributes to TatBC receptor binding, *Journal of Biological Chemistry*, 293; 7281-7299. doi: 10.1074/jbc.RA118.002576.

Ulsen, P. Van *et al.* (2014) Type V secretion : From biogenesis to biotechnology, *BBA - Molecular Cell Research*, 1843; 1592-1611. doi: 10.1016/j.bbamcr.2013.11.006.

Vandesompele, J. *et al.* (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, *Genome Biology*, 18, 1-12, doi: 10.1186/gb-2002-3-7-research0034.

Ve, B. M. (2010) Application of a Short, Disordered N-Terminal Flagellin Segment , a Fully Functional Flagellar Type III Export Signal to Expression of Secreted Proteins, *Applied and Environmental Microbiology,* 76; 891-899. doi: 10.1128/AEM.00858-09.

Végh, B. M. *et al.* (2006) Localization of the flagellum-specific secretion signal in *Salmonella* flagellin, *Biochemical and Biophysical Research Communications*, 345; 93-98.

doi: 10.1016/j.bbrc.2006.04.055.

Vemuri, G. N. *et al.* (2005) Physiological response of central metabolism in *Escherichia coli* to deletion of pyruvate oxidase and introduction of heterologous pyruvate carboxylase, *Biotechnology and Bioengineering*, 90; 64-76, doi: 10.1002/bit.20418.

Villaverde, A. and Carrió, M. M. (2003) Protein aggregation in recombinant bacteria: Biological role of inclusion bodies, *Biotechnology Letters*, 25; 1385-1395. doi: 10.1023/A:1025024104862.

Wacker, M., Linton, D., Hitchen, Paul G, *et al.* (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*, *Science,*, 298; 1790-1793. doi: 10.1126/science.298.5599.1790.

Wacker, M., Linton, D., Hitchen, Paul G., *et al.* (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*, *Science*, 298; 1790-3. doi: 10.1126/science.298.5599.1790.

Van Der Wal, F. J. *et al.* (1998) Optimization of bacteriocin release protein (BRP)-mediated protein release by *Escherichia coli:* Random mutagenesis of the pCloDF13-derived BRP gene to uncouple lethality and quasilysis from protein release, *Applied and Environmental Microbiology*, 64, 392-398. doi: 0099-2240/98/$04.00+0.

Walia, R., Deb, J. K. and Mukherjee, K. J. (2007) Development of expression vectors for *Escherichia coli* based on the pCR2 replicon, *Microbial Cell Factories*, 14, 1-9. doi: 10.1186/1475-2859-6-14.

Walsh G (2006) Biopharmaceutical benchmarks 2006, *Nature Biotechnology*, 24; 831-833. doi: http://dx.doi.org/10.1038/nbt0706-769.

Walther, J. *et al.* (2015) The business impact of an integrated continuous biomanufacturing platform for recombinant protein production, *Journal of Biotechnology*, 213; 3-1. doi: 10.1016/j.jbiotec.2015.05.010.

Wang, Y. *et al.* (2012) T3DB : an integrated database for bacterial type III secretion system, *BMC Bioinformatics*, 66, 1-6. doi: 10.1186/1471-2105-13-66..

Warikoo, V. *et al.* (2012) Integrated continuous production of recombinant therapeutic proteins, *Biotechnology and Bioengineering*, 109; 3018-29. doi: 10.1002/bit.24584.

Warming, S. *et al.* (2005) Simple and highly efficient BAC recombineering using *galK* selection, *Nucleic Acids Research*, 33; 1-12. doi: 10.1093/nar/gni035.

Weber, J., Hoffmann, F. and Rinas, U. (2002) Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production: Redirection of metabolic fluxes, *Biotechnology and Bioengineering*, 80; 320-30. doi: 10.1002/bit.10380.

Wei, B. L. *et al.* (2001) Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*, *Molecular Microbiology*, 40; 245-256. doi: 10.1046/j.1365-2958.2001.02380.x..

Weikert, C. *et al.* (2000) Co-overexpression of RspAB Improves Recombinant Protein Production in *Escherichia coli*, *Metabolic Engineering*, 2; 293-9. doi: 10.1006/mben.2000.0163.

Weinacker, D. *et al.* (2013) Applications of recombinant *Pichia pastoris* in the healthcare

industry, *Brazilian Journal of Microbiology*, 44; 1043-1048. doi: 10.1590/S1517-83822013000400004.

Werten, M. W. T. *et al.* (1999) High-yield secretion of recombinant gelatins by *Pichia pastoris*, *Yeast*, 15; 1087-1096. doi: 10.1002

Widmaier, D. M. *et al.* (2009) Engineering the *Salmonella* type III secretion system to export spider silk monomers, *Molecular Systems Biology* 5; 1-9. doi: 10.1038/msb.2009.62.

Williams, A. W. *et al.* (1996) Mutations in *fliK* and *flhB* affecting flagellar hook and filament assembly in *Salmonella typhimurium*, *Journal of Bacteriology*, 178; 2960-2970. doi: 10.1128/jb.178.10.2960-2970.1996.

Wolfe, A. J. and Visick, K. L. (2008) Get the message out: Cyclic-Di-GMP regulates multiple levels of flagellum-based motility, *Journal of Bacteriology*, 190; 463-75. doi: 10.1128/JB.01418-07.

Wu, F., Zhong, F. and He, F. (2016) Microbial Proteomics : Approaches , Advances , and Applications, *Journal of Bioinformatics and Proteomics Review*, 2, 1-7. doi: 10.5194/acp-18-16033-2018.

Yadav, V. G. *et al.* (2012) The future of metabolic engineering and synthetic biology: Towards a systematic practice, *Metabolic Engineering*, 14; 233-41. doi: 10.1016/j.ymben.2012.02.001.

Yamamoto, S. and Kutsukake, K. (2006) FliT acts as an anti-FlhD2C2 factor in the transcriptional control of the flagellar regulon in *Salmonella enterica* serovar typhimurium, *Journal of Bacteriology*, 188; 6703–6708. doi: 10.1128/JB.00799-06.

Yang, W. *et al.* (2001) A new method for protein coexpression in *Escherichia coli* using two incompatible plasmids, *Protein Expression and Purification*, 22; 472-8. doi: 10.1006/prep.2001.1453.

Yang, W., Ni, L. and Somerville, R. L. (1993) A stationary-phase protein of *Escherichia coli* that affects the mode of association between the trp repressor protein and operator-bearing DNA, *Proceedings of the National Academy of Sciences of the United States of America*, 90; 5796-5800. doi: 10.1073/pnas.90.12.5796.

Yeom, S. J. *et al.* (2016) Long-term stable and tightly controlled expression of recombinant proteins in antibiotics-free conditions, *PLoS ONE*, 11; e166890. doi: 10.1371/journal.pone.0166890.

Yokoseki, T., Iino, T. and Kutsukake, K. (1996) Negative regulation by FliD, FliS, and FliT of the export of the flagellum-specific anti-sigma factor, FlgM, in *Salmonella typhimurium*, *Journal of Bacteriology*, 178, 899-901. doi: 10.1128/jb.178.3.899-901.1996.

Yonekura, K., Maki-Yonekura, S. and Namba, K. (2001) Structure analysis of the flagellar cap-filament complex by electron cryomicroscopy and single-particle image analysis, *Journal of Structural Biology*, 133, 246-253. doi: 10.1006/jsbi.2000.4345.

Yonekura, K., Maki-Yonekura, S. and Namba, K. (2002) Growth mechanism of the bacterial flagellar filament, *Research in Microbiology*, 153; 191-197. doi: 10.1016/S0923-2508(02)01308-6.

Yonekura, K., Maki-Yonekura, S. and Namba, K. (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy, *Nature*, 424, 643-650. doi: 10.1038/nature01830.

Yoon, S. H., Kim, S. K. and Kim, J. F. (2010) Secretory production of recombinant proteins in *Escherichia coli*, *Recent Patents on Biotechnology*, 4; 23-9. doi; 10.2174/18722081079006955..

Young, E. and Alper, H. (2010) Synthetic biology: Tools to design, build, and optimize cellular processes, *Journal of Biomedicine and Biotechnology*, 2010, 130781. doi: 10.1155/2010/130781.

Yuan, L. *et al.* (2005) Laboratory-Directed Protein Evolution, *Microbiology and Molecular Biology Reviews*, 69; 373-92. doi: 10.1128/MMBR.69.3.373-392.2005.

Yue, B. G. *et al.* (2000) Functional coexpression of serine protein kinase SRPK1 and its substrate ASF/SF2 in *Escherichia coli*, *Nucleic Acids Research*, 28, e14. doi: 10.1093/nar/28.5.e14.

Zafar, M. A. *et al.* (2014) Transcriptional occlusion caused by overlapping promoters, *Proceedings of the National Academy of Sciences*, 111; 1557-61. doi: 10.1073/pnas.1323413111.

Zechner, E. L., Lang, S. and Schildbach, J. F. (2012) Assembly and mechanisms of bacterial type IV secretion machines, *Physiological Transactions of the Royal Society of Biology,* 19; 1073-1087. doi: 10.1098/rstb.2011.0207.

Zhang, F. *et al.* (1995) Secretion and Circular Dichroism Analysis of the C-Terminal Signal Peptides of HlyA and LktA, *Biochemistry*, 34; 4193-4201. doi: 10.1021/bi00013a007.

Zhao, K., Liu, M. and Burgess, R. R. (2007) Adaptation in bacterial flagellar and motility systems: From regulon members to 'foraging'-like behavior in *E. coli*, *Nucleic Acids Research*, 35; 4441-4452. doi: 10.1093/nar/gkm456.

Zhao, X. *et al.* (2013) Cryoelectron tomography reveals the sequential assembly of bacterial flagella in Borrelia burgdorferi, *Proceedings of the National Academy of Sciences of the United States of America*, 110; 14390-5. doi: 10.1073/pnas.1308306110.

Zhiqiang, L. *et al.* (2009) Structural and functional studies of *Aspergillus oryzae* Cutinase : Enhanced thermostability and hydrolytic activity of synthetic ester and polyester degradation, *Journal of the American Chemical Society*, 131; 15711-15716. doi: 10.1021/ja9046697.

Zieske, L. R. (2006) A perspective on the use of iTRAQ™ reagent technology for protein complex and profiling studies', *Journal of Experimental Botany*, 57; 1501-8. doi: 10.1093/jxb/erj168.

Zolkiewski, M. (1999) ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation: A novel multi-chaperone system from *Escherichia coli*, *Journal of Biological Chemistry*, 247; 28083-6, doi: 10.1074/jbc.274.40.28083.

8 Appendix

8.1 Appendix 1: Genes and proteins involved in FT3SS formation.

Classified into operon class, cellular localisation, assembly method, function, size of the proteins. Data collated from Aizawa 1999, Macnab 2003; Berg 2003; Aldridge and Hughes 2002; Stafford and Hughes 2007; Raha *et al.* 1994; Mytelka & Chamberlin, 1996; Namba *et al*. 2000; Minamino and Erhardt 2010.

Protein	Operon Class	Cellular Location	Function	Stoichi-ometry	Size (kDa)
FlhA	2	Centre of MS ring	Protein export component	~ 2	75
FlhB	2	Centre of MS ring	Protein export component	~ 2	42
FlhC	1	Cytoplasm	Class 2 master regulator		22
FlhD	1	Cytoplasm	Class 2 master regulator		14
FlhE	2	Cytoplasm	Role swarming motility		12
FliA	2	Cell envelope protein Cytoplasm	Sigma factor 28, for class 3 operons		27
FliC	3	Cell exterior	Filament (Flagellin)	~30000	53
FliD	3	Cell exterior	Filament cap	10	50
FliE	2	Periplasm	MS– ring junction, export gate	~ 9	11
FliF	2	Cytoplasmic membrane	MS– ring, mount for rotor switch and rod, housing for export apparatus	26	61
FliG	2	Peripheral	Rotor / switch protein, torque generation, binds MS ring	26	37
FliH	2	Peripheral	Negative regulator of FliI, protein export, ATPase complex		26
FliI	2	Cytoplasm	Drives type III protein export ATPase		49
FliJ	2	Cytoplasm	Rod, hook, filament chaperone, ATPase complex		17
FliK	2	Cytoplasm	Hook–length control		39
FliL	2	Cell exterior	Motor / Switch associated protein?		17
FliM	2	Peripheral	C ring rotor / switch protein		38
FliN	2	Peripheral	C ring, switch component		38
FliO	2	Centre of MS ring	Export component	~ 37	11
FliP	2	Centre of MS ring			

FliQ	2	Centre of MS ring	Export component	110	27
FliR	2	Centre of MS ring	Export component	~ 1	10
FliS	3	Cytoplasm	Export component FliC Chaperone	~ 4	29
FliT	3	Cytoplasm	FliD Chaperone	~ 1	15
FliY	2	Cytoplasm	Regulates class III transcription (?)	~ 1	14
FliZ	2	Cytoplasm	σ^s antagonist		
FlgA	2	Periplasm	Chaperone for P ring protein		
FlgB	2	Periplasm	Rod protein	7	
FlgC	2	Periplasm	Rod protein	6	
FlgD	2	Cell exterior	Hook capping protein	~ 5	
FlgE	2	Cell exterior	Hook protein	132	
FlgF	2	Periplasm	Rod protein	6	
FlgG	2	Periplasm	Distal rod protein	26	
FlgH	2	Outer membrane	L– ring	28	
FlgI	2	Periplasm	P – ring	24	
FlgJ	2	Periplasm	Rod capping protein	5?	
FlgK	3	Cell exterior	HAP1: Hook filament junction protein	13	
FlgL	3	Cell exterior	HAP2: Hook filament junction protein	~ 10	
FlgM	3	Cytoplasm	Anti – sigma factor 28	~ 32	
FlgN	3	Cytoplasm	FlgK, FlgL specific chaperone	~ 16	32
MotA	3	Cytoplasmic Membrane	Force generation, stator protein		34
MotB	3	Cytoplasmic Membrane	Force generation, stator protein, converts proton energy into torque		

8.2 Appendix 2: Gene knockouts, Flagella gene and phenotypic effect

Gene	Organism	Function in WT	Reported gene and phenotypic effect in mutant	Reference
A. Knockout mutations result in increased flagella gene expression and motility				

fliT	*Salmonella enterica*	Prevents $FlhD_4C_2$ from binding to Class II promoters, also releases DNA – bound $FlhD_4C_2$ resulting in increased *clpX*P proteolysis	Higher Claa II gene expression and increased motility, less $FlhD_4C_2$	Yamamoto & Kutsukake, 2006
fliZ	*E. coli*	Binds to the *flhDC* promoter region	Increased *flhDC* expression in the *fliZ* knockout mutant	Pesavento & Hengge, 2012
IrhA	*E. coli*	Binds to *flhD*	3.5 reduction in *FlhDC* expression and hypermotile cells	Lehnen *et al.,* 2002
dksA	*E. coli*	Enhances activity of ribosome synthesis moleculaes pppGpp and ppGpp, which reduces flagella gene expression	Increased *flhDC* expression (Lemke et al., 2009), hyperflagelated & hypermotile cells (Aberg *et al.,* 2009) or less motile cells (Magnusson *et al.,* 2007)	Lemke et al, 2009, Aberg *et al.,* 2009; MAgnusson *et al.*, 2007
clpXP	*E. coli*	Degrades $FlhD_4C_2$ protein: protease	Hyperflagellated	Kitagawa *et al.,* 2011; Tomoyasu et al. 2002
B. Knockout mutations result in decreased flagella gene expression or motility				
rscB	*E. coli & salmonella enterica*	Binds to the *flhDC* regulon & negatively regulates *flhDC*, positively regulates *fliPQR*	Decreased flhDC gene expression when over expressed in *E. coli*, increased motility in *Salmonella* knockout	Wang et al. 2007; Francez – Charlot *et al.,* 2004
hdfR	*E. coli*	Binds $FlhD_4C_2$	Seen in H–NS deficient mutants. HdfR is usually negatively controlled by H – NS, hence H – NS mutants have more active HdfR & mutants are nonflagellated	Ko & Park, 2000
sirA, csrB, csrC	*E. coli*	CsrA stabilises $FlhD_4C_2$ & increases its expression, csrB & csrC are antagonists of csrA. sirA activates csrBC translation	Reduced flhDC gene expression & motility & less flhDC mRNA in late exponential phase	Wei *et al.,* 2001
OmpR	*E. coli*		*flhDC* expression decreased with increased	Shin & Park 1995

EnvZ	*E. coli*	Phosphorylated OmpR negatively regulates flhDC EnvZ phosphorylates Omp^R. Phosphorylated Omp^R binds to flhDC, repressing translation. However, growth is low.	osmolality in WT, this was not seen in the Omp^R knockout mutant Slow cell division in *the* Omp^R mutant, flhDC expression reduced	Pruss, 1998.
DnaK	*Salmonella enterica*	This chaperone aids correct assembly of $FlhD_4C_2$ allowing it to interact with promoters.	Lack flagella & Class II and III gene expression.	Takaya *et al.,* 2006
CRP	*E. coli*	Increases flhD operon transcription in the presence of *FliA*	Crp mutant was non – motile	Soutouurina *et al*. 1999; Zhao *et al*., 2007
QseB	*E. coli*	Increases flhD operon transcription in the presence of *FliA*	Less flagellin & less transcription of *flhD, fliA, motA* & *fliC*	Sperandio *et al.*, 2002
H – NS	*E. coli*	Increased flhD operon transcription when all H – NS binding sites are bound by H – NS	Hns mutant was non – motile	Soutourino *et al.,* 1999

8.3 Appendix 3: List of Primers used during this project

Primer name	**Oligonucleotide sequences**
flgD 100 aa F	CATATGAGGAGGTAGTCCAGATGTCCATTGCGGT
flgD 100 aa R	CTCGAGCACGCCGTGACCGATCAGGTTACTGGCC
flgD 50 aa R	AAAACTCGAGGGTTTTTCAGCTGCGCCACC
flgE 100 aa F	AAACATATGAGGAGGTAGTCCAGATGGCCTTTTCTCAAGCGG
flgE 100 aa R	AAACTCGAGAATTGTCCGTTACGGCT
flgE 50 aa R	AAACTCGAGCAGTCCCACTTTCGAACCGG
pJExpress F	TAATATGTGGAATTGTGAGC
pJExpress R 4 SS	AAACTCACCTTGCATTTGTACTCTTTGCCGTTC
pJEXpress C_H2 R	.AAACTCACCTTGCATTTGTACTCTTTGCCGTTC
pJExpress empty R	CCGTCAGTCTCAGTTAATC
5'UTR atg F	AAACATATGGCGGGAATAAGGGGCAGAGAA
5'UTR atg R	ACAGAATTCCATGATTCGTTATCCTATATTGC
hGH MfeI F (Our style)	GACCGACAATTGTTCCCGACGATTCCGCTGTC
hGH NsiI R (Our Style)	ACGCTGATGCATGAAGCCGCAAGAACCCTCAA
hGH F (Fuji Style)	GACCGACATATGAACTTTAAGAAGGAGATATAATGAAAAA GACTGCTATTGC

hGH R (Fuji Style)	ACGCTGATGCATGAAGCCGCAAGAACCCTCAA
clpX F	ATGACAGATAAACGCAAAGA
clpX R	TTCACCAGATGCCTGTTGCGCT
Cm^R R	TCATCGCTCTGGAGTGAATACCACGA
C_1 (Datsenko Wanner)	TTATACGCAAGGCGACAAGG
C_2 (Datsenko Wanner)	GATCTTCCGTCACAGGTAGG
pJExpress F 4 SS	GAGCGCTCACAATTCCACAACGGT
FlgE SDM F	5'TTCGAACCGGCAGCCATATCGGCAGCAGAGGCCGTGCCTGATTTAAAGC-3'
FlgE SDM R	5'GCTTTAAATCAGGCACGGCCTCTGCTGCCGATATGGCTGCCGGTTCGAA-3'
FlgD SDM F	5'CTGCGCCACCAGCGCAGTCAGAGCACTGCTTTGTAAATCTGCGGCGTTG-3'
FlgD SDM R	5'CAACGCCGCAGATTTACAAAGCAGTGCTCTGACTGCGCTGGTGGCGCAG-3'
5'UTR Error prone PCR F	AACCATATGGCGGGAATAAGGGGCAGAGAAAAGA
5'UTR Error prone PCR R	AACCTCGAGACCCGCTGCGTCATCCTTCGCGCT
pNK_{15}TcLib F	CTTACCAGCCTAACTTCGATCA
pNK_{15}TcLib R	ACTGCCCGCTTTCCATATAA
pNK_{15}TcLib F (Colony PCR)	AAGTTTATATGGAAAGCGGGCAGT
pNK_{15}TcLib R(colony PCR)	AAGCGCGCTGATCAACTAGTCCATGG
pJExpress R for 3' UTR confo R	CGCCCGGGCTAATTATGGGGTGTCGCCCTTG
pNK Tet^R F	CTTACCAGCCTAACTTCGATCA
pNK Tet^R R	ACTGCCCGCTTTCCATATAA
pNK p15A ori R	GGGGGGCGGAGCCTATGGAAA
pNK col.PCR F	AAGTTTATATGGAAAGCGGGCAGT
pNK col.PCR R	AAGCGCGCTGATCAACTAGTCCATGG
Promoter (snRNA) F (Anderson pNK)	ACGCGTttgacagctagctcagtcctaggtataatAGATCT
Promoter (snRNA) R (Anderson pNK)	AGATCTAattatacctaggactgagctagctgtcaaACGCGT
McaS F	ACTAGATCTACCGGCGCAGAGGAGACAATG
McAS R	CgtCCATGGTGCAAAGTTAAAACTGCATAaaaAAATAGA
MicA F	ATCAGATCTGAAAGACGCGCATTTGTTATC
MicA R	CTGCCATGGAAAAGGCCACTCGTGAGTGGC
Knockout mutagenesis primers	
Mot-cheZ F	GCCTGACGACTGAACATCCTGTCATGGTCAACAGTGGAAGGATGATGTC
Mot-cheZ R	AGACCGCCTGATATGACGTGGTCACGCCACATCAGGCAATACAAAATTCCGGG GATCCGTCGAC
Mot-cheZ C	ATTTTATTACCCACGCTCACCAGCCTGTTGGC
Tgr F	GTGCCGATGACTTTCTATCAGGAGTAAACCTGGACGAGAGACAACGGTAGTGT AGGCTGGAGCTGCTTC
Trg R	TGCCGTTCTGAAAGGTGAAGGGATCTGTCGATCCCTCCTTGAACATTTATTCC GGGGATCCGTCGAC
Trg confo	ATAGCCTTTCTCCTTGCCGGATGGCGGGTAAGAGGCTAAG
Tsr F	TTCCTTTCCAGGCCGAAAATCTTGCATCGGTCCACAGGAAAGAGAAACCGTGT AGGCTGGAGCTGCTTC
Tsr R	TCCTTATGCCCGATAACATTTTGCTTATCGGGCATTTTCATGGCGAATTCCGG GGATCCGTCGAC
Tsr confo	TTGAATATCTTTTTAGTATCCATAGTAAAACCTGGC
ArcZ F	GTGCGGCCTGAAAAACAGTGCTGTGGTGTAGGCTGGAGCTGCTTC
ArcZ KO R	AAAAAATGACCCCGGCTAGACCGGGGTGCGCATTCCGGGGATCCGTCGACc
ArcZ KO F Confo	AGTTTTGCTATCTTAACTGC
ArcZ KO R Confo	CGTGGGTGGCAAAAGCCACT
OmrA KO F	GATCCCAGAGGTATTGATTGGTGAGATTATTCGGTGTAGGCTGGAGCT GCTTC

OmrA KO R	GTAAATTAGGTGCGAAAAAAAACCT ATTCCGGGGATCCGTCGACCTG
OmrA KO F confo	GTGCAGACCACAATCAAGAT
OmrA KO R confo	GCGACAGTAAATTAGGTGCG
Crl KO F	ATGACGTTACCGAGTGGACACCCGAAGAGCGTGTAGGCTGGAGCTGCTTC
Crl KO R	TCACGCCGTTAACTTCACCGGCTCGTCACGAAAATTCCGGGGATCCGTCGACC
Crl KO F confo	CATCACAACAGGAGATAGCA
Crl KO R confo	ATGCATCCGGCACATTTCAC
GspL KO F	TCGTGATGCCTGAATCGTTGATGGTCATACGTTCTTCCT GTGTAGGCTGGAGCTGCTTC
GspL KO R	TTTAATCATTTCCCTGAGTTCCTTACTGTCCATACATTCCGGGGATCCG TCGACC
clpB KO F	ATGCGTCTGGATCGTCTTACTAATAAATTCCAGCTTGCTCGTGTAGGCT GGAGCTGCTTC
clpB KO R	TTACTGGACGGCGACAATCCGGTCTTCATTAACTTCCAGGCATTCCGGG GATCCGTCGACCTG
ycgR KO F	CGCGTGAGTCATTACCATGAGCAGTTCCTGAAACAAGTGTAGGCTGGA GCTGCTTC
ycgR KO R	GTAAATCAGTCGCGCACTTTGTCCGCTTTTTCCC ATTCCGGGGATCCGTCGACCTG
Kam^R R	TCAGAAGAACTCGTCAAGAAGGCGAT
pKD3 ArcZ R	AAAAAATGACCCCGGCTAGACCGGGGTGCGCCATATGAATATCCTCCTT AGTTC
pKD3 OmrA R	GTAAATTAGGTGCGAAAAAAAACCTCATATGAATATCCTCCTTAGTTC
pKD3 Crl R	TCACGCCGTTAACTTCACCGGCTCGTCACGAAACATATGAATATCCTCC TTAGTTC
pKD3 clpB R	TTACTGGACGGCGACAATCCGGTCTTCATTAACTTCCAGGCCATATGAA TATCCTCCTTAGTTC
pKD3 ycgR R	GTAAATCAGTCGCGCACTTTGTCCGCTTTTTCCCCATATGAATATCCTC CTTAGTTC
pKD3 GspL R	TTTAATCATTTCCCTGAGTTCCTTACTGTCCATACCATATGAATATCCT CCTTAGTTC

8.4 Appendix 4: Secretion construct derived from pJExpress 404

pJEXpress 404 Cutinase : Nucleotide and Amino Acid Sequence

```
CA TAT GGC GGG AAT AAG GGG CAG AGA AAA GAG TAT TTC GGC GAC T
   Y   G   G   N   K   G   Q   R   K   E   Y   F   G   D   *
AA CAA AAA ATG GCT GTT TTT GAA AAA AAT TCT AAA GGT TGT TTT A
    Q   K   M   A   V   F   E   K   N   S   K   G   C   F   T
CG ACA GAC GAT AAC AGG GTT GAC GGC GAT TGA GCC GAC GGG TGG A
T   D   D   N   R   V   D   G   D   *   A   D   G   W   K
AA CCC AAT ACG TAA TCA ACG ACT TGC AAT ATA GGA TAA CGA ATC A
    P   N   T   *   S   T   T   C   N   I   G   *   R   I   M
TG GCA CAA GTC ATT AAT ACC AAC AGC CTC TCG CTG ATC ACT CAA A
    A   Q   V   I   N   T   N   S   L   S   L   I   T   Q   N
AT AAT ATC AAC AAG AAC CAG TCT GCG CTG TCG AGT TCT ATC GAG C
    N   I   N   K   N   Q   S   A   L   S   S   S   I   E   R
GT CTG TCT TCT GGC TTG CGT ATT AAC AGC GCG AAG GAT GAC GCA G
    L   S   S   G   L   R   I   N   S   A   K   D   D   A   A
```

```
CG GGT CTC GAG GAG AAT CTG TAT TTT CAG GGC GAA TTC GGT CGT A
   G   L   E   E   N   L   Y   F   Q   G   E   F   G   R   T
CC ACC CGT GAT GAT CTG ATT AAT GGT AAT AGC GCA AGC TGC CGT G
   T   R   D   D   L   I   N   G   N   S   A   S   C   R   D
AT GTG ATT TTT ATC TAT GCA CGT GGT AGC ACC GAA ACC GGT AAT C
   V   I   F   I   Y   A   R   G   S   T   E   T   G   N   L
TG GGC ACC CTG GGT CCG AGC ATT GCA AGC AAT CTG GAA AGC GCA T
   G   T   L   G   P   S   I   A   S   N   L   E   S   A   F
TT GGT AAA GAT GGT GTT TGG ATT CAG GGT GTT GGT GGT GCA TAT C
   G   K   D   G   V   W   I   Q   G   V   G   G   A   Y   R
GT GCA ACC CTG GGT GAT AAT GCA CTG CCT CGT GGC ACC AGC AGC G
   A   T   L   G   D   N   A   L   P   R   G   T   S   S   A
CA GCA ATT CGT GAA ATG CTG GGT CTG TTT CAG CAG GCA AAT ACC A
   A   I   R   E   M   L   G   L   F   Q   Q   A   N   T   K
AA TGT CCG GAT GCC ACC CTG ATT GCC GGT GGT TAT AGC CAG GGT G
   C   P   D   A   T   L   I   A   G   G   Y   S   Q   G   A
CA GCA CTG GCA GCA GCA AGC ATT GAA GAT CTG GAT AGT GCC ATT C
   A   L   A   A   A   S   I   E   D   L   D   S   A   I   R
GT GAT AAA ATT GCA GGC ACC GTT CTG TTT GGC TAT ACC AAA AAT C
   D   K   I   A   G   T   V   L   F   G   Y   T   K   N   L
TG CAA AAT CGT GGT CGC ATT CCG AAT TAT CCG GCA GAT CGT ACC A
   Q   N   R   G   R   I   P   N   Y   P   A   D   R   T   K
AA GTT TTT TGC AAT ACC GGT GAT CTG GTT TGT ACC GGT AGC CTG A
   V   F   C   N   T   G   D   L   V   C   T   G   S   L   I
TT GTT GCA GCA CCG CAT CTG GCC TAT GGT CCT GAT GCC CGT GGT C
   V   A   A   P   H   L   A   Y   G   P   D   A   R   G   P
CG GCA CCG GAA TTT CTG ATT GAA AAA GTT CGT GCA GTT CGT GGT A
   A   P   E   F   L   I   E   K   V   R   A   V   R   G   S
GC GCA CTG CAG GAA AAC CTG TAC TTC CAA GGT GAC TAC AAG GAC G
   A   L   Q   E   N   L   Y   F   Q   G   D   Y   K   D   D
AT GAC GAT AAG TGG AGC CAT CCG CAG TTT GAG AAA TCT AGA AAG C
   D   D   K   WW  S   H   P   Q   F   E   K   S   R   K   L
TT TAA TCG TTG TAA CCT GAT TAA CTG AGA CTG ACG GCA ACG CCA A
   *   S   L   *   P   D   *   L   R   L   T   A   T   P   N
AT TGC CTG ATG CGC TGC GCT TAT CAG GCC TAC AAG TTG AAT TGC A
   C   L   M   R   C   A   Y   Q   A   Y   K   L   N   C   N
AT TTA TTG AAT TTG CAC ATT TTT GTA GGC CGG ATA AGG CGT TTA C
   L   L   N   L   H   I   F   V   G   R   I   R   R   L   R
GC CGC ATC CGG CAA CAT AAA GCG CAA TTT GTC AGC AAC GTG CTT C
   R   I   R   Q   H   K   A   Q   F   V   S   N   V   L   P
CC CGC CAC CGG CGG GGT TTT TTT CTG CCT GGA ATT TAC CTG TAA C
   R   H   R   R   G   F   F   L   P   G   I   Y   L   *   P
CC CCA AAT AAC CCC TCA TTT CAC CCA CTA ATC GTC CGA TTA AAA A
   P   N   N   P   S   F   H   P   L   I   V   R   L   K   T
CC CTG GAG AAA CGG ATA ATC ATG CCG ATA ACT CAT ATA ACG CAG G
   L   E   K   R   I   I   M   P   I   T   H   I   T   Q   G
GC TGT TTA TCG TGG GAT CC
   C   L   S   W   D   X
```

8.5 Appendix 5: List of Amino Acids

Amino Acid	3-Letter	1-Letter	Amino Acid	3-Letter	1-Letter
Alanine	Ala	A	Leucine	Leu	L

Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic Acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Iso Leucine	Ile	I	Valine	Val	V
Selenocysteine	Sec	U			

Key for side chain properties

Basic (Positively charged)
Acidic (Negatively charged)
Polar (uncharged)
Hydrophobic
Special cases

8.6 Table 8.1: Absorbance of proteins for Bradford Asssay.

Sample /Absorbance (595 nm)	**Sample 1**	**Sample 2**	**Sample 3**
5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *Escherichia coli MC1000 ΔfliC ΔflgKL (ΔCKL)*	0.898	0.922	0.931
No 5′^{UTR}-No FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *Escherichia coli MC1000 ΔfliC ΔflgKL ΔclpX (ΔclpX)*	0.733	0.895	0.870
5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *Escherichia coli MC1000 ΔfliC ΔflgKL ΔclpX (ΔclpX)*	1.047	0.851	0.905

8.7 Table 8.3: Bradford Assay protein conc and OD.

iTRAQ Sample (3 x replicates)	**Abs. 595 nm**	**Conc. µg µl^{-1}**	**Conc. X 5**	**To 100 µg**	**For 20 µl**
5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli ΔfliC ΔflgKL*	0.898	1.2268	6.13	16.30	3.26
5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli ΔfliC ΔflgKL*	0.922	1.26	6.32	15.80	3.16
5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli ΔfliC ΔflgKL*	0.931	1.27	6.39	15.62	3.12
No 5′^{UTR}-No-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E.coli ΔfliC* ΔflgKL *ΔclpX*	0.733	0.9615	4.80	20.79	4.16
No 5′^{UTR}-No-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli ΔfliC ΔflgKL ΔclpX*	0.895	1.22	6.10	16.36	3.27
No 5′^{UTR}-No-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli ΔfliC ΔflgKL ΔclpX*	0.870	1.18	5.90	16.92	3.38
5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli ΔfliC ΔflgKL ΔclpX*	1.047	1.46	7.33	13.63	2.73
5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli ΔfliC* ΔflgKL *ΔclpX*	0.851	1.15	5.75	17.37	3.5

5′^{UTR}-FliC$_{47aa}$-Cutinase pJExpress No 3′^{UTR} in *E. coli* Δ*fliC* Δ*flgKL* Δ*clpX*	0.905	1.23	6.19	16.15	3.27

Table 8.4: Upregulated protein from pSC 1-Δ*CKL* Δ*clpX* and pSC 2-Δ*CKL* Δ*clpX* comparison.

Sr No	Protein ID	Gene name	Protein names	Fold change	*p*-Value
1	P08622	*dnaJ*	Chaperone protein DnaJ	1.30	7.4E-05
2	P09372	*grpE*	Protein GrpE	1.34	0.00684
3	P0A6B7	*iscS*	Cysteine desulfurase IscS	1.16	0.00128
4	P0A6F5	*groL*	60 kDa chaperonin	1.36	1.4E-09
5	P0A6H5	*hslU*	ATP-dependent protease ATPase subunit HslU	1.34	0.00010
6	P0A6Y8	*dnaK*	Chaperone protein DnaK	1.39	7.7E-14
7	POA6Z3	*htpG*	Chaperone protein HtpG	1.32	2.2E-09
8	P0A8J4	*ybeD*	UPF0250 protein YbeD	1.42	0.00062
9	P0A9B2	*gapA*	Glyceraldehyde-3-phosphate dehydrogenase A	1.22	0.00702
10	P0A9M0	*Lon*	Lon protease	1.20	5.6E-05
11	P0AFF6	*nusA*	Transcription termination/ant termination protein NusA	1.10	0.01952
12	P27298	*prlC*	Oligopeptidase A	1.19	2.0E-05
13	P30958	*mfd*	Transcription-repair-coupling factor	1.15	0.00122
14	P63284	*clpB*	Chaperone protein *ClpB*	1.32	1.0E-09
15	P69910	*potA*	Spermidine/putrescine import ATP-binding protein PotA	1.31	9.5E-06
16	P77395	*ybbN*	Uncharacterized protein YbbN	1.23	0.00895
17	P9999	*cutA*	Cutinase	1.57	4.4E-06

Table 8.5: Protein downregulation identified from pSC 1-Δ*CKL* Δ*clpX* and pSC 2-Δ*CKL* Δ*clpX* comparison of iTRAQ.

Sr No	Protein ID	Gene name	Protein names	Fold change	*p*-value
1	P02924	*araF*	L-arabinose-binding periplasmic protein	- 1.22	0.00041
2	P07330	*cheB*	Chemotaxis response regulator protein-glutamate methylesterase	- 1.24	0.00843
3	P07363	*cheA*	Chemotaxis protein CheA	- 1.28	8.9E-06
4	P0A8G3	*uxaC*	Uronate isomerase	- 1.16	0.00103
5	P0A964	*cheW*	Chemotaxis protein CheW	- 1.31	0.00193
6	P0AD59	*ivy*	Inhibitor of vertebrate lysozyme	- 1.14	0.00327
7	P0AE67	*cheY*	Chemotaxis protein CheY	- 1.43	0.00389
8	P0AEM6	*fliA*	RNA polymerase sigma factor FliA	- 1.33	0.00010
9	P0AEX9	*malE*	Maltose-binding periplasmic protein	- 1.17	0.00656
10	P0AGF6	*tdcB*	L-threonine dehydratase catabolic TdcB	- 1.31	0.01802
11	P11407	*fumB*	Fumarate hydratase class I, anaerobic	- 1.18	0.01203

12	P21367	*ycaC*	Uncharacterized protein YcaC	- 1.18	0.00478
13	P37329	*modA*	Molybdate-binding periplasmic protein	- 1.31	1.2E-05
14	P39160	*uxuB*	D-mannonate oxidoreductase	- 1.20	0.0036
15	P42632	*tdcE*	PFL-like enzyme TdcE	- 1.28	0.0025
16	P43533	*flgN*	Flagella synthesis protein FlgN	- 1.32	0.01258
17	P63883	*amiC*	N-acetylmuramoyl-L-alanine amidase AmiC	- 1.21	0.0077
18	P75937	*flgE*	Flagellar hook protein FlgE	- 1.31	4.0E-05
19	P76010	*ycgR*	Flagellar brake protein *YcgR*	- 1.54	0.0259
20	P76440	*preT*	NAD-dependent dihydropyrimidine dehydrogenase subunit PreT	- 1.12	0.0028
21	P77674	*prr*	Gamma-aminobutyraldehyde dehydrogenase	- 1.28	0.00132

Table 8.6: Protein upregulation identified from pSC 1-Δ*CKL* Δ*clpX* and pSC 1-Δ*CKL* comparison of iTRAQ analysis.

Sr No	Protein ID	Gene name	Protein names	Fold change	*p*-Value
1	P07003	*poxB*	pyruvate dehydrogenase, thiamine triphosphate binding, FAD binding	1.20786	0.0012
2	P07363	*cheA*	fused chemotactic sensory histidine kinase in the two-component regulatory system with cheB and cheY, sensory histidine kinase/signal sensing protein (cheA)	1.1165	0.0067
3	P08622	*dnaJ*	chaperone Hsp40, DnaK co-chaperone	1.192	0.0032
4	P0A6F5	*groL*	Cpn60 chaperonin GroEL, the large subunit of GroESL	1.192	2.5E-05
5	P0A6Y8	*dnaK*	chaperone Hsp70, with co-chaperone DnaJ	1.239	7.6E-06
6	P0A8G6	*wrbA*	NAD (P) H: quinone oxidoreductase	1.275	0.0001
7	P0A8J4	*ybeD*	UPF0250 family protein	1.3363	0.0036
8	P0A991	*fbaB*	fructose-bisphosphate aldolase class I	1.174	0.0022
9	P0A9B2	*gapA*	glyceraldehyde-3-phosphate dehydrogenase A	1.273	0.0011
10	P0AB14	*yccJ*	uncharacterized protein	1.205	0.0021
11	P0ABD3	*bfr*	bacterioferritin, iron storage and detoxification protein	1.236	0.0013
12	P0ABT2	*dps*	Fe-binding and storage protein; stress-inducible DNA-binding protein	1.577	9.7E-06
13	P0AC59	*grxB*	glutaredoxin 2 (Grx 2)	1.195	0.0130
14	P0AC62	*grxC*	glutaredoxin 3	1.191	0.0039
15	P0ACC3	*erpA*	iron-sulfur cluster insertion protein	1.231	0.0215
16	P0AD59	*ivy*	inhibitor of c–type lysozyme, periplasmic	1.158	0.0016
17	P0AEH5	*elaB*	putative membrane-anchored DUF883 family ribosome-binding protein	1.786	0.0167
18	P0AEM9	*fliY*	cystine transporter subunit	1.133	0.0098

19	P0AFH8	*osmY*	salt-inducible putative ABC transporter periplasmic binding protein	1.5060	0.0047
20	P13445	*rpoS*	RNA polymerase, sigma S (sigma 38) factor	1.430	0.0004
21	P21179	*katE*	catalase HPII, heme d-containing	1.180	0.0038
22	P22256	*gabT*	4-aminobutyrate aminotransferase, PLP-dependent	1.234	8.6E-05
23	P23847	*dppA*	dipeptide/heme ABC transporter periplasmic binding protein; dipeptide chemotaxis receptor	1.1454	0.0051
24	P25526	*gabD*	succinate-semialdehyde dehydrogenase I, NADP-dependent	1.233	0.0036
25	P42588	*patA*	putrescine:2-oxoglutaric acid aminotransferase, PLP-dependent	1.250	0.0265
26	P42620	*yqjG*	putative S-transferase	1.1884	0.0181
27	P45578	*luxS*	S-ribosyl homocysteine lyase	1.1102	0.0023
28	P63284	*clpB*	protein disaggregation chaperone	1.137	0.0011
29	P69910	*gadB*	glutamate decarboxylase B, PLP-dependent	1.575	4.4E-08
30	P77499	*sufC*	SufBCD Fe-S cluster assembly scaffold protein, ATP-binding protein	1.156	0.010
31	P9999	*cutA*	Cutinase	1.468	1.1E-05
32	Q46857	*dkgA*	2,5-diketo-D-gluconate reductase A	1.165	0.0114

Table 8.7: Protein downregulation from pSC 1-Δ*CKL* Δ*clpX* and pSC 1-Δ*CKL* comparison.

Sr No	Protein ID	Gene name	Protein names	Fold change	*p*-Value
1	P08331	*cpdB*	2':3'-cyclic nucleotide 2' phosphodiesterase	- 1.107	0.0030
2	P0A6F1	*carA*	carbamoyl phosphate synthetase small subunit, glutamine amidotransferase	- 1.117	0.0043
3	P0A6H1	*clpX*	ATPase and specificity subunit of *clpX*-ClpP ATP-dependent serine protease	- 2.388	7.5E-08
4	P0A786	*pyrB*	aspartate carbamoyltransferase, catalytic subunit	- 1.493	0.0049
5	P0A7F3	*pyrI*	aspartate carbamoyltransferase, regulatory subunit	- 1.239	0.0041
6	P0A8G3	*uxaC*	uronate isomerase	- 1.112	0.0036
7	P0A9G2	*nhaA*	transcriptional activator of nhaA	- 1.170	0.0124
8	P0ABQ2	*garR*	tartronate semialdehyde reductase	- 1.333	3.4E-05
9	P0AEG4	*dsbA*	periplasmic protein disulfide isomerase I	- 1.260	0.0008
10	P0AEM6	*fliA*	RNA polymerase, sigma 28 (sigma F) factor	- 1.161	0.010
11	P0AFX9	*rseB*	anti-sigma E factor binds RseA	- 1.183	0.016
12	P38104	*rspA*	bifunctional D-altronate/D-mannonate dehydratase	- 2.033	2.1E-07
13	P38105	*rspB*	putative Zn-dependent NAD(P)-binding oxidoreductase	- 2.498	7.0E-05

14	P39160	*uxuB*	D-mannonate oxidoreductase, NAD-dependent	- 1.189	0.0004
15	P76440	*preT*	dihydropyrimidine dehydrogenase, NADH-dependent, subunit N	- 1.113	0.0076
16	P77739	*yniA*	fructosamine kinase family protein	- 1.151	0.0167
17	Q46856	*yqhD*	aldehyde reductase, NADPH-dependent	- 1.136	0.0024
18	Q46915	*gudX*	glutamate dehydratase-related protein, substrate unknown	- 1.363	0.0035

Table 8.8: Upregulated proteins in pSC 2-*ΔCKL* Δ*clpX vs* pSC 1-*ΔCKL* comparison

Sr No	Protein ID	Gene name	Protein names	Fold chain	P-Value
1	P02924	*araF*	L-arabinose ABC transporter periplasmic binding protein	1.203	0.0023
2	P05704	*trg*	methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	1.20	0.006
3	P06610	*btuE*	glutathione peroxidase	1.25	0.0004
4	P07003	*poxB*	pyruvate dehydrogenase, thiamine triphosphate-binding, FAD-binding	1.35	7.7E-06
5	P07024	*ushA*	bifunctional UDP-sugar hydrolase/5'-nucleotidase	1.11	0.005
6	P07330	*cheB*	fused chemotaxis regulator: protein-glutamate methylesterase in a two-component regulatory system with CheA	1.28	0.016
7	P07363	*cheA*	fused chemotactic sensory histidine kinase in a two-component regulatory system with CheB and CheY: sensory histidine kinase/signal sensing protein	1.43	1.25E-07
8	P09551	*argT*	lysine/arginine/ornithine transporter subunit	1.125	0.0009
9	P0A6K6	*deoB*	phosphopentomutase	1.11	0.001
10	P0A867	*talA*	transaldolase A	1.18	0.011
11	P0A8G6	*wrbA*	NAD(P)H: quinone oxidoreductase	1.35	6.80E-05
12	P0A964	*cheW*	purine-binding chemotaxis protein	1.28	0.014
13	P0A991	*fbaB*	fructose-bisphosphate aldolase class I	1.17	0.006
14	P0A9H9	*Chez*	Chemotaxis regulator, protein phosphatase for CheY	1.24	0.006
15	P0AB14	*yccJ*	uncharacterized protein	1.27	0.002
16	P0ABD3	*bfr*	Bacterioferritin, iron storage and detoxification protein	1.27	1.43E-05
17	P0ABT2	*dps*	Fe-binding and storage protein; stress-inducible DNA-binding protein	1.40	1.26E-06
18	P0AC59	*grxB*	glutaredoxin 2 (Grx2)	1.23	0.0008

19	P0ACY3	*yeaG*	protein kinase, endogenous substrate unidentified; autokinase	1.27	2.91E-05
20	P0AD59	*ivy*	inhibitor of c-type lysozyme, periplasmic	1.32	3.92E-05
21	P0AE28	*aroM*	AroM family protein	1.14	0.038
22	P0AE67	*cheY*	chemotaxis regulator transmitting signal to flagellar motor component	1.46	0.0006
23	P0AEE5	*mglB*	methyl-galactoside transporter subunit	1.11	0.002
24	P0AEH5	*elaB*	putative membrane-anchored DUF883 family ribosome-binding protein	1.61	0.010
25	P0AEM9	*fliY*	cystine transporter subunit	1.25	7.50E-06
26	P0AEU0	*his*	histidine ABC transporter periplasmic binding protein	1.12	0.014
27	P0AEX9	*malE*	maltose transporter subunit	1.33	5.6E-06
28	P0AFH8	*osmY*	salt-inducible putative ABC transporter periplasmic binding protein	1.58	0.0005
29	P0AG80	*ugpB*	sn-glycerol-3-phosphate ABC transporter periplasmic binding protein	1.19	0.016
30	P0AG82	*pstS*	phosphate ABC transporter periplasmic binding protein	1.13	0.0019
31	P0C0L2	*osmC*	lipoyl-dependent Cys-based peroxidase, hydroperoxide resistance; salt-shock inducible membrane protein; peroxiredoxin	1.29	0.006
32	P13445	*rpoS*	RNA polymerase, sigma S (sigma 38) factor	1.48	0.0001
33	P13482	*treA*	periplasmic trehalase	1.220	0.017
34	P18843	*nadE*	NAD synthetase, NH3/glutamine-dependent	1.14	7.6E-05
35	P21179	*katE*	catalase HPII, heme d-containing	1.25	0.0001
36	P21362	*yciF*	putative rubrerythrin/ferritin-like metal-binding protein	1.16	0.023
37	P22256	*gabT*	4-aminobutyrate aminotransferase, PLP-dependent	1.23	1.04E-05
38	P23847	*dppA*	dipeptide/heme ABC transporter periplasmic binding protein; dipeptide chemotaxis receptor	1.22	3.6E-06
39	P23917	*mak*	manno(fructo)kinase	1.24	0.010
40	P25516	*acnA*	aconitate hydratase 1; aconitase A	1.150	0.003
41	P25526	*gabD*	succinate-semialdehyde dehydrogenase I, NADP-dependent	1.22	0.003
42	P25665	*metE*	5-methyltetrahydropteroyltri-glutamate- homocysteine S-methyltransferase	1.30	0.003
43	P26612	*amyA*	cytoplasmic alpha-amylase	1.26	0.003
44	P29013	*ycgB*	SpoVR family stationary phase protein	1.19	0.017
45	P30859	*artI*	arginine transporter subunit	1.23	0.0002
46	P31130	*ydeI*	hydrogen peroxide resistance OB fold protein; putative periplasmic protein	1.42	0.015

47	P31678	*otsB*	trehalose-6-phosphate phosphatase, biosynthetic	1.37	0.012
48	P37329	*modA*	molybdate ABC transporter periplasmic binding protein; chlorate resistance protein	1.30	3.7E-06
49	P37685	*aldB*	aldehyde dehydrogenase B	1.28	0.002
50	P39325	*ytfQ*	galactofuranose ABC transporter periplasmic binding protein	1.30	0.0009
51	P45578	*luxS*	S-ribosyl homocysteine lyase	1.13	0.004
52	P52647	*pfo*	pyruvate-flavodoxin oxidoreductase	1.15	0.006
53	P52697	*pgl*	6-phosphogluconolactonase	1.10	0.013
54	P56262	*ysgA*	putative carboxymethylenebutenolidase	1.22	0.007
55	P68066	*grcA*	autonomous glycyl radical cofactor	1.30	0.0005
56	P69811	*fruB*	fused fructose-specific PTS enzymes: IIA component/HPr component	1.078	0.003
57	P69910	*gadB*	glutamate decarboxylase B, PLP-dependent	1.20	0.0002
58	P75937	*flgE*	flagellar hook protein	1.42	2.8E-07
59	P76010	*ycgR*	flagellar velocity braking protein, c-di-GMP-regulated	1.416	0.0077
60	P76108	*ydcS*	putative ABC transporter periplasmic binding protein	1.29	0.003
61	P76177	*ydgH*	DUF1471 family periplasmic protein	1.123	0.008
62	P76193	*ldtE*	murein L, D-transpeptidase	1.22	0.0098
63	P76227	*ynjH*	DUF1496 family protein	1.78	0.027
64	P76329	*yedP*	putative mannosyl-3-phosphoglycerate phosphatase	1.29	0.018
65	P76621	*csiD*	carbon starvation protein	1.374	0.0002
66	P77499	*sufC*	SufBCD Fe-S cluster assembly scaffold protein, ATP-binding protein	1.17	0.0018
67	P77735	*yajO*	2-carboxybenzaldehyde reductase	1.35	0.16

Table 8.9: Downregulated proteins in pSC 2-Δ*CKL* Δ*clpX* *vs* pSC 1-Δ*CKL* comparison

Sr No	Protein ID	Gene name	Protein names	Fold chain	P-Value
1	P00805	*ansB*	periplasmic L-asparaginase 2	- 1.15	0.0003
2	P00968	*carB*	carbamoyl-phosphate synthase large subunit	- 1.17	0.0005
3	P04982	*rbsD*	D-ribose pyranase	- 1.29	0.009
4	P06983	*hemC*	hydroxymethylbilane synthase	- 1.08	0.016
5	P09152	*narG*	nitrate reductase 1, alpha subunit	- 1.12	0.0075
6	P09372	*grpE*	heat shock protein	- 1.28	0.0005
7	P09546	*putA*	fused DNA-binding transcriptional regulator/proline dehydrogenase/pyrroline-5-carboxylate dehydrogenase	- 1.11	0.0005
8	P0A6F1	*carA*	carbamoyl phosphate synthetase small subunit, glutamine amidotransferase	- 1.14	0.0001

9	P0A6F5	*groL*	Cpn60 chaperonin GroEL, large subunit of GroESL	- 1.13	6.7E-06
10	P0A6H1	*clpX*	ATPase and specificity subunit of *clpX*-ClpP ATP-dependent serine protease	- 2.18	8.6E-09
11	P0A6L4	*nanA*	N-acetylneuraminate lyase	- 1.12	0.0125
12	P0A6Y5	*hslO*	heat shock protein Hsp33	- 1.17	0.004
13	P0A6Y8	*dnaK*	chaperone Hsp70, with co-chaperone DnaJ	- 1.12	2.8E-05
14	P0A6Z3	*htpG*	protein refolding molecular co-chaperone Hsp90, Hsp70-dependent; heat-shock protein; ATPase	- 1.20	3.6E-05
15	P0A786	*pyrB*	aspartate carbamoyltransferase, catalytic subunit	- 1.3	0.006
16	P0A7J3	*rplJ*	50S ribosomal subunit protein L10	- 1.08	0.017
17	P0A836	*sucC*	succinyl-CoA synthetase, beta subunit	- 1.08	0.006
18	P0A9M0	*lon*	DNA-binding ATP-dependent protease	- 1.11	0.0006
19	P0A9P4	*trxB*	thioredoxin reductase, FAD/NAD(P)-binding	- 1.10	0.02
20	P0ABB0	*atpA*	F1 sector of membrane-bound ATP synthase, alpha subunit	- 1.10	0.002
21	P0ABB4	*atpD*	F1 sector of membrane-bound ATP synthase, beta subunit	- 1.13	8.6E-06
22	P0ABH7	*gltA*	citrate synthase	- 1.10	0.019
23	P0AC33	*fumA*	fumarate hydratase (fumarase A), aerobic Class I	- 1.08	0.003
24	P0ACE0	*hybC*	hydrogenase 2, large subunit	- 1.10	0.006
25	P0AES2	*gudD*	D-glucarate dehydratase 1	- 1.19	0.00
26	P0AG67	*rpsA*	30S ribosomal subunit protein S1	- 1.17	0.013
27	P11349	*narH*	nitrate reductase 1, beta (Fe-S) subunit	- 1.17	0.013
28	P25553	*aldA*	aldehyde dehydrogenase A, NAD-linked	- 1.08	0.013
29	P25714	*yidC*	membrane protein insertase	- 1.14	0.018
30	P30958	*mfd*	transcription-repair coupling factor	- 1.10	0.008
31	P33599	*nuoC*	NADH:ubiquinone oxidoreductase, fused CD subunit	- 1.10	0.002
32	P33602	*nuoG*	NADH:ubiquinone oxidoreductase, chain G	- 1.08	0.010
33	P36683	*acnB*	aconitate hydratase 2; aconitase B; 2-methyl-cis-aconitate hydratase	- 1.09	0.004
34	P38104	*rspA*	bifunctional D-altronate/D-mannonate dehydratase	- 1.93	4.5E-08
35	P38105	*rspB*	putative Zn-dependent NAD(P)-binding oxidoreductase	- 2.28	4.4E-05
36	P39829	*garD*	D-galactarate dehydrogenase	- 1.17	0.0035
37	P56580	*srlE*	glucitol/sorbitol-specific enzyme IIB component of PTS	- 1.16	0.007
38	P63284	*clpB*	protein disaggregation chaperone	- 1.16	1.3E-07
39	P76558	*maeB*	malic enzyme: putative oxidoreductase/phosphotransacetylase	- 1.11	0.0001
40	P77212	*rclA*	reactive chlorine stress species (RCS) resistance protein; pyridine nucleotide-dependent disulfide oxidoreductase family	- 1.21	0.003
41	Q46915	*gudX*	glucarate dehydratase-related protein, substrate unknown	-1.309	0.003

Appendix 6: List of softwares used

List of Software	Details
GraphPad Prism	Graphpad prism software, CA, USA.
Cytoscape	Open source bioinformatics software
LO-COR C - Digit	C-Digit® Image studio Light *ver 5.2*
Pymol	Schrodinger, LLC, NY, USA
Magellan™ Data analysis software	Tecan UK Ltd, Reading, UK
FinchTV 1.4.0	Geospiza Inc.,
Gene Designer DNA 2.0	ATUM Corporate HQ, Newark, California
SnapGene® 4.1	GSL Biotech LLC, Chicago, IL
Genome compiler 2.2.88	Genome compiler corporation, Los Altos, CA

Appendix 7: List of Reagents:

List of Reagents	Catalogue #; Manufacturers details
1 kb DNA Ladder	cat#: N3232L; New England Biolabs, Hitchin, UK
EZ-Run protein ladder (EZ-Run™ Prestained *Rec* protein ladder)	cat#: BP3603500; Fisher BioReagents Fisher scientific UK Ltd, Loughbourough
Adenine	cat#: A8626; Sigma Aldrich, Dorset, UK
Dimethyl sulfoxide	cat#: D8418; Sigma Aldrich, Dorset, UK
Isopropanol	cat#: 10674732; Fisher Chemicals, Loughborough, UK
Phosphate buffered saline (PBS)	cat#: D8537; Sigma Aldrich, Dorset, UK
PierceTM Protease Inhibitor Mini Tablets	cat#: 88665; Life Technologies, Paisley UK

Appendix 8: List of Kit used

List of kit	Catalogue #; Manufacturers details
ISOLATE II PCR and Gel kit	cat#; Bio-52060; Bioline USA Inc, Tauton MA 02780
Isolate II Plasmid Mini kit	cat#; Bio-52057; Bioline USA Inc, Tauton MA 02780
Wizard® Genomic DNA purification kit	Promega Corporation, Madison WI 53711 USA
Pierce™ BCA Protein Assay kit	cat#; 23225; ThermoFisher, Paisley, UK

Appendix 9: List of equipments

List of Equipments	Details
Infinite® M 200 Pro series	Tecan UK Ltd, Reading, UK
C-Digit blot scanner	C-digit; Leusden; Netherlands
Jenway 7300 UV / Visible, Spectrophotometer	Cole-Palmer Ltd, Staffordshire
Frech® Press, Thermo scientific sc	Thermo Fisher, Paisley, UK
HPLC	Dionex, U.K.
Mass Spec	Thermo scientific, Germany